Computer-Aided 13 C NMR Chemical Profiling of Crude Natural Extracts without Fractionation

Ali Bakiri, Jane Hubert, Romain Reynaud, Sylvie Lanthony, Dominique Harakat, Jean-Hugues Renault, Jean-Marc Nuzillard

Computer-aided ^{13}C NMR chemical profiling of crude natural extracts without fractionation

Ali Bakiri[†,§], Jane Hubert[†,*], Romain Reynaud[§], Sylvie Lanthony[†], Dominique Harakat[†], Jean-Hugues Renault[†], Jean-Marc Nuzillard[†,§]

[†] Institut de Chimie Moléculaire de Reims, UMR CNRS 7312, SFR CAP'SANTE, Université de Reims Champagne-Ardenne, Reims, France

[§] Soliance-Givaudan, Pomacle, France

ABSTRACT

A computer-aided, ^{13}C NMR-based dereplication method is presented for the chemical profiling of natural extracts without any fractionation. An algorithm was developed in order to compare the ^{13}C NMR chemical shifts obtained from a single routine spectrum with a set of predicted NMR data stored in a natural metabolite database. The algorithm evaluates the quality of the matching between experimental and predicted data by calculating a score function and returns the list of metabolites which are the most likely to be present in the studied extract. The proof of principle of the method is demonstrated on a crude alkaloid extract obtained from the leaves of *Peumus boldus*, resulting in the identification of eight alkaloids including isocorydine, rogersine, boldine, reticuline, coclaurine, laurotetanine, *N*-methylcoclaurine and norisocorydine, as well as three monoterpenes including *p*-cimene, eucalyptol, and α-terpinene. The results were compared to those obtained with other methods, either involving a fractionation step before the chemical profiling process or using mass spectrometry detection in the infusion mode or coupled to gas chromatography.

KEYWORDS

Natural products, mixture analysis, dereplication, ^{13}C NMR, mass spectrometry, metabolite database, *Peumus boldus*

INTRODUCTION

Natural product (NP) studies usually deal with samples of very complex chemical composition. 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. One other major bottleneck of NP studies when searching for novel biologically active substances and/or original chemical structures is the frequent rediscovery of very common or already known metabolites, which causes a great waste of time. A good knowledge of the presence of known molecules within natural samples is therefore of great interest to speed up the chemical profiling of natural mixtures. This process of rapidly identifying known chemotypes, known as dereplication, has emerged over the last years as an essential approach to prevent duplication of isolation efforts.[1–3]

The most common detection techniques for the dereplication of NPs are mass spectroscopy (MS) and nuclear magnetic resonance (NMR), each one being very complementary to the other.[1,2] Liquid or gas chromatography coupled to high-resolution MS are currently the most powerful "high-throughput screening" strategies for the on-line identification of metabolites in natural resources.[4] Nevertheless, the important variability observed between MS datasets obtained from one mass analyzer to another and the tedious interpretation of MS data only based on elemental composition and fragmentation pattern are critical issues that limit the efficiency of MS-based déréplication. [5] Efforts are currently underway to develop automatic processing methods, mainly through the implementation of computational treatments. Molecular networking in particular was recently validated as a promising computer-based approach to visualize and organize tandem mass spectrometry datasets and automate database search for metabolite identification within complex mixtures.[6,7] Conversely, NMR is much less sensitive than MS, but remains, by far, the most efficient detection technique to unambiguously elucidate the molecular structures of secondary plant metabolite.[8] A range of 1D and 2D NMR data acquisition sequences can be used to chemically profile natural samples via the interpretation of ^{1}H-^{1}H and ^{1}H-^{13}C coupling patterns.[9–13] Direct ^{13}C NMR detection provides a large chemical shift dispersion (240 ppm) that reduces signal overlaps and results in spectra containing well-resolved individualized peaks due to broadband decoupling. The main drawback of this technique is its low sensitivity due to a low natural abundance of the ^{13}C nuclei (≈ 1.1 %) and its low magnetogyric ratio (≈ 25 % of the one of the ^{1}H nucleus). The detection of quaternary carbon resonances which gives access to the whole carbon skeleton of metabolites is another strong advantage of ^{13}C NMR, even though they cannot benefit from sensitivity enhancement by coherent or incoherent magnetization transfer from ^{1}H nuclei.[14] The relevance of ^{13}C NMR for mixture analysis has already been suggested in previous studies. For instance, a very interesting algorithm was proposed in 1986 for the identification and quantification of organic mixture components using a quantitative ^{13}C NMR spectrum as input data.[15] Unfortunately the method was tested on a model mixture of standard compounds and has never been applied for the analysis of genuine mixtures of natural products. Another procedure based on ^{13}C NMR and computer tools was developed in 1994 to identify the most abundant phenolic derivatives in liquids produced by pyrolysis of biomass.[16] The same method was applied to analyze the major chemical constituents of essential oils and was suggested as a useful complementary tool to GC/MS for the correct identification of isomers.[17] In another work, high-quality ^{13}C NMR data obtained with a ^{13}C-optimized NMR probe were used to establish ^{13}C-^{13}C statistical correlations in order to enhance the accuracy of metabolite identification at natural abundance.[14] Two years ago our laboratory developed a dereplication method based on the combination of ^{13}C NMR with support-free liquid-liquid chromatography and Hierarchical Clustering Analysis for the fast identification of the major metabolites in natural extracts.[18] This approach has since been successfully applied to several extracts containing different chemical classes of secondary metabolites.[19–22] With the same goal of maximizing the efficiency of NP dereplication, this paper presents a computer-aided ^{13}C NMR dereplication workflow that does not require any pre-fractionation of the investigated sample in order to rapidly identify the major constituents of a natural mixture while reducing as much as possible isolation steps. The proof of principle of this algorithm is demonstrated on a crude alkaloid extract obtained from the leaves of *Peumus boldus* (Monimiaceae), a small tree native to the central region of Chile that is traditionally used for its digestive and hepato-biliary protective effects.[23]

RESULTS AND DISCUSSION

As the ^{13}C NMR fingerprint of a metabolite corresponds to an invariant set of chemical shifts that corresponds to its carbon skeleton, the ^{13}C NMR signals that originate from a single compound can be assigned to the corresponding metabolite even if this metabolite is mixed with other constituents, provided of course that there is a way to aggregate these signals. In this context an algorithm was developed in order to identify the major constituents of a crude natural mixture from a single ^{13}C NMR analysis on the basis of chemical shifts comparison between experimental and predicted datasets. This dereplication process involving the calculation of score values to evaluate each match between database records and the crude extract spectrum was tested on a crude alkaloid extract obtained from *Peumus boldus* leaves.

For data acquisition, ^{13}C NMR signals were recorded with broadband proton decoupling mode, in order to obtain a simplified ^{13}C NMR spectrum where each carbon atom of the metabolite mixture corresponds to a single sharp peak with the greatest possible intensity even for the quaternary resonances. Multiplicity edition by the J-modulated spin echo method (J-MOD) or the Distortionless Enhancement by Polarisation Transfer method (DEPT) would have indicated the parity of the number of protons attached to each carbon atom, and helped confirming or excluding proposals from database at the end of the dereplication process. J-MOD and DEPT pulse sequence rely on a spin echo module tuned for an average coupling constant ^{1}J (^{1}H-^{13}C)

value and are therefore less sensitive than the simple ^{1}H-decoupled ^{13}C sequence that was retained for this study

The ^{13}C NMR spectrum of the crude *Peumus boldus* extract is presented in Figure 1. The automatically produced peak list was directly used as an input file submitted to the algorithm. The principle was to perform a search over a spectral database containing the structures and predicted chemical shifts of all metabolites already reported in the genus *Peumus* (n=58) and to compare the ^{13}C NMR chemical shifts of the crude extract spectrum to the predicted chemical shifts of the database records. For practical reasons, we used for this purpose the NMR Workbook Suite commercial software from ACD/Labs comprising in a single toolbox a chemical structure drawing interface, a ^{1}H/^{13}C NMR prediction software and a spectral database. The ACD/Labs prediction software is based on both the Hierarchical Organisation of Spherical Environments (HOSE) code[24] and neural network algorithms and takes advantage of a considerable amount of literature data (more than 200,000 chemical structures for ^{13}C NMR prediction with more than 2,000,000 ^{13}C NMR chemical shifts). To our three year experience, the predicted values obtained with this software are quite reliable and the user interfaces are easy to manage routinely. However, this software remains primarily a commercial product with very limited access to the calculation details to obtain the predicted values. A few other efficient and open source prediction tools would probably have been of great value and possibly implemented in the proposed dereplication workflow, provided that some parts of the database search algorithm be adapted accordingly. One can mention the NMRShiftDB web database[25] which also involves the use of increment methods for NMR chemical shift prediction, or several quantum chemical methods based for instance on the *ab initio* Hartree-Fock theory, second-order Moller-Plesset theory or Density Functional theories (DFTs) which have been recognized to yield good matching between experimental and computed values.[25-27]

With the database search algorithm proposed in the present work and applied to the ^{13}C NMR peak list of the crude extract spectrum of *Peumus boldus* leaves, a total of 33 metabolites were proposed with score values higher than 0.9 (Table 1). As the score value is defined as the ratio between the numbers of chemical shifts of the crude extract that match a molecular structure of the database and the total number of carbon positions of this structure, a score value higher than 0.9 means that more than 90 % of the carbon positions of the metabolites proposed by the database matched the experimental NMR signals of the crude extract spectrum. Among these metabolites, a range of alkaloids including isocorydine (**1**), rogersine (**2**), boldine (**3**), reticuline (**4**), coclaurine (**5**), laurotetanine (**6**), norisocorydine (**7**), and N-methylcoclaurine (**8**) were found with a score value of 1, indicating a complete matching of their predicted ^{13}C NMR chemical shifts with the ^{13}C NMR signals detected in the crude extract spectrum. An alkaloid, pronuciferine, showed a score value of 0.94 with only one mismatching chemical shift value (Table 1). A mismatch may arise from important chemical shift differences between predicted and experimental values or to non-homogenous peak intensity values resulting for instance from quaternary or symmetrical carbon positions, therefore all chemical structures displaying score values ranging from 0.9 and 1 were examined to confirm or not the presence of the proposed metabolites. This step was performed by manually checking the ^{13}C NMR chemical shifts of the matching chemical structures in the crude extract spectrum. The major alkaloids **1-8** were easily confirmed. Pronuciferine was rapidly eliminated from the list of proposals because this alkaloid contains a carbonyl group for which the characteristic resonance around 185 ppm was not found in the crude extract spectrum. For such metabolites displaying score values comprised between 0.9 and 1, further refinement of the approach including for instance Density Functional Theory (DFT) calculations would be useful to minimize bias and improve the reliability of the prediction of NMR chemical shifts.[25] Sans opinion… Rev2 sera content…

In order to compare the results obtained by the ^{13}C NMR-based database search algorithm with another method, the crude extract was solubilized in methanol and directly infused in a QToF mass spectrometer. The resulting MS spectrum is presented in Figure 2. Several intense even molecular ions corresponding to nitrogen-containing compounds were detected. The ion at *m/z* 330.1 was attributed to the parent ion $[M+H]^+$ of reticuline (**4**) and the ion at *m/z* 286.1 was attributed to the parent ion $[M+H]^+$ of coclaurine (**5**), thus confirming their presence in the crude extract as revealed by the database search algorithm. Another intense ion detected at *m/z* 328.1 could be attributed to the parent ion $[M+H]^+$ of norisocorydine, boldine or laurotetanine, all sharing the same molecular formula $C_{19}H_{21}NO_4$. Unfortunately, it was impossible to discriminate between these three isomers by this method. Similarly, it was impossible to distinguish between isocorydine and rogersine for the attribution of the molecular ion detected at *m/z* 342.1 because these two alkaloids share the $C_{20}H_{23}NO_4$ molecular formula. Yet mass spectrometry was here a good method to rapidly and sensitively detect a pool of alkaloids in the crude extract of *Peumus boldus* leaves. But, in total, only two alkaloids of the eight proposed by the ^{13}C NMR-based database search algorithm were confirmed. However, without more detailed structural data, it remained very difficult to discriminate between the different other molecular structures, and even hyphenated to liquid chromatography, standards molecules would have been necessary to unambiguously identify the individual alkaloids present in the extract. The database search algorithm described in this work thus appears as a highly useful complementary tool to MS for the chemical profiling of natural extracts.

In order to go further, another ^{13}C NMR-based dereplication method including a fractionation step was applied on the same extract. For this purpose, a pH-zone refining Centrifugal Partition Chromatography (CPC) (REF J.-H. Renault, G. Le Crouerour, P. Thépenier, J.-M. Nuzillard, M. Zèches-Hanrot, L. Le Men-Olivier, Isolation of indole alkaloids from *Catharanthus roseus* by centrifugal partition chromatography in the pH-zone refining mode, J. Chromatogr. A 849

(1999) 421-431.) method was developed to fractionate *Peumus boldus* alkaloids on the basis of their proton affinity and distribution coefficient. It may be mentioned here that it is the first time that *Peumus boldus* alkaloids were separated by CPC. The experiment was performed using a biphasic solvent system primarily composed of M*t*BE and water while CH_3CN was added as bridge solvent to reduce the polarity difference between the stationary and mobile phases and ensure the solubility of all substances. The aqueous stationary phase was supplemented with methane sulfonic acid (MSA) in order to ensure the protonation of all alkaloids and therefore to promote their trapping into the aqueous stationary phase. Triethylamine (TEA) was added to the organic mobile phase in order to progressively neutralize the retainer and to selectively displace the alkaloids from the stationary to the mobile phase according to their partition coefficients and acidity constants. As a result, the column effluent remained at pH 2 for 45 minutes, then increased to pH 6 during the displacement of alkaloids from 45 to 90 minutes, and finally increased up to pH 10 when the CPC fractionation was completed. Eleven fraction pools P_I-P_{XI} were obtained and analyzed by ^{13}C NMR. Automatic peak picking and alignment of ^{13}C NMR signals across spectra of the fraction series resulted in a table with 11 columns (one per fraction) and 193 rows (one per chemical shift bin containing at least one ^{13}C NMR signal in at least one fraction). The table was submitted to Hierarchical Clustering Analysis (HCA) on the rows. As a result, statistical correlations between ^{13}C NMR resonances belonging to a single structure within the fraction series were visualized as “chemical shift clusters” in front of the corresponding dendrograms (Figure 3). Cluster A corresponded to an intense cluster of 18 ^{13}C NMR chemical shifts. After entering these chemical shifts into the database, the structure of isocorydine **(1)** was proposed as the first hit of over 21 proposals. This structure was confirmed by checking all chemical shifts of isocorydine in raw NMR data of P_I where the intensity of cluster 1 was predominant. Using the same method, clusters B+B’ were identified as a mixture of *N*-methylcoclaurine (**8**) and reticuline (**4**). These two metabolites have very similar molecular structures and were detected collected? in the same fraction pools, thus their ^{13}C NMR chemical shifts were aggregated under the same clusters. The other clusters highlighted on the HCA correlation map were identified as coclaurine (**5**), rogersine (**2**), norisocorydine (**7**), laurotetanine (**6**), and boldine (**3**), as indicated in Figure 3. The molecular structures of all these compounds were confirmed by checking NMR data of the corresponding fractions. Yet, the eight alkaloids identified by the database search algorithm directly from the ^{13}C NMR analysis of the crude *Peumus boldus* extract without separation were also identified by the ^{13}C NMR-based dereplication method involving a fractionation step. On the one hand, this means that time consuming separation procedures are not always necessary to identify the major constituents of a crude mixture. Applying the database search algorithm just after a single ^{13}C NMR analysis of an extract can help to initiate a good chemical profiling process while saving time, materials and solvents and giving useful information for further investigations such as biological or toxicological evaluation, fractionation and purification process development. On the other hand, it must be noted that the signal-to-noise (S/N) ratios of all ^{13}C NMR spectra recorded after CPC fractionation were much higher than that of the crude extract spectrum (Figure 5This brings us to the question of the minor constituents that would be difficult to unambiguously identify by applying the database search algorithm just after a single ^{13}C NMR analysis. Nevertheless, there is nothing to prevent the use of such database search algorithm on fractions of simplified composition enriched with the minor constituents of the extract. Mais Caramel ne permet pas de trouver d'autres constituants minoritaires dans ce cas…

In addition to the eight alkaloids reported above, 24 monoterpenes were also proposed by the database search algorithm, all with a score value of 1, excepted for α-copaene and β-oploplenone, δ-cadinene and bornyl acetate which were effectively revealed absent from the *Peumus boldus* extract after checking their chemical shifts in the experimental ^{13}C NMR spectrum. As carried out for alkaloids, molecular structure validation of the 20 monoterpenes showing a score value of 1 was tentatively performed with the NMR dataset obtained after CPC fractionation of the *Peumus boldus* extract. Unfortunately, only NMR signals of very low intensity that could potentially correspond to monoterpenes were detected in the spectra of the CPC fractions, thus making their interpretation impossible. The very low intensity of monoterpene signals in the NMR spectra of the CPC fractions was probably due to the volatility of these compounds, which resulted in their loss during fractionation and under-vacuum mobile phase elimination. As an alternative, the crude *Peumus boldus* extract was analyzed by GC/MS. As illustrated in Figure 4, several well-resolved monoterpene signals were obtained on the resulting chromatogram. The retention times, molecular formula and electron ionization spectra of these GC/MS-detected monoterpenes are given in Table 2. By comparing the MS spectra of the different peaks of the chromatogram to the NIST MS Search library, six monoterpenes were putatively identified among which *p*-cimene (**9**), eucalyptol (**10**), and α-terpinene (**11**) were common to the list of hits proposed by our database search algorithm. The identity of the other monoterpenes proposed by the score function approach were not confirmed, either because the electron ionization MS spectra of the GC/MS-detected monoterpenes did not exactly match the spectra of the NIST MS Search library, or because the monoterpenes of the list of hits were absent from this library.

In summary, a database search algorithm was developed in order to chemically profile the major metabolites of a natural extract from a ^{13}C NMR analysis and without physical separation of the constituents. This algorithm was tested on a crude extract obtained from *Peumus boldus* leaves, leading to the successful identification of eight alkaloids including isocorydine (**1**), rogersine (**2**), boldine (**3**), reticuline (**4**), coclaurine (**5**), laurotetanine (**6**), norisocorydine (**7**), and N-methylcoclaurine (**8**) as well as a range of monoterpenes, among which *para*-cymene (**9**), eucalyptol (**10**), and α-terpinene (**11**) were unambiguously confirmed by GC/MS analysis of the same extract. The strong point of this approach relies on the possibility to work with the NMR tool directly on complex samples, which can be very useful to recover information about natural extract composition at a very early stage of the chemical profiling process. This can help to design most reliable and appropriate experimental conditions for further metabolite separation and to focus only on the relevant compounds depending on the target before engaging in time-consuming multi-step purification procedures. The method can also be considered as a rapid and highly complementary approach to mass spectrometry-based analytical tools such as MS infusion, LC/MS or GC/MS from which only molecular formula and fragmentation patterns are recovered. Additional work is currently taking place in our laboratory to further improve this dereplication process using not only 1D ^{13}C NMR chemical shifts data, but also 2D HSQC and HMBC NMR data from which ^{1}H-^{13}C connectivity information should greatly enhance the performance of the algorithm to identify more constituents in natural extracts.

EXPERIMENTAL SECTION

Chemicals, reagents, plant material. Methyl *tert*-butyl ether (M*t*BE), methanol, ethyl acetate, petroleum ether, chloroform, and acetonitrile (CH_3CN) were purchased from Carlo Erba Reactifs SDS (Val de Reuil, France). Sodium hydroxide (NaOH), methane sulfonic acid (MSA) and triethylamine (TEA) were purchased from PROLABO (Paris, France). Deuterated methanol (methanol-*d4*) was obtained from Sigma-Aldrich (Saint-Quentin, France). Deionized water was used to prepare all aqueous solutions.

Sample preparation. *Peumus boldus* dry leaves (500 g) purchased from Cailleau Herboristerie (Chemillé-Melay, France) were ground into a fine powder and macerated for 24 hours in 10 L of petroleum ether for delipidation. Then, the marcs were alkalinized with 180 mL of NH_3

(15 % in water) and macerated in ethyl acetate (12 L) for 24 hours. After leaching, three liquid-liquid extractions were successively performed with 2 L of H_2SO_4 (0.5 N) each time. The recovered aqueous phases were pooled and the pH was increased up to 9 with NH_3 (15 % in water). Three successive extractions were again performed with 2 L of chloroform each time. The organic phases were pooled and washed with H_2O (2 L). Chloroform was removed under vacuum to recover 3.5 g of a crude alkaloid extract.

Database search algorithm. The algorithm was implemented in the Python 2.7 programming language. The open source cheminformatics package RDKit[28] was used to draw the molecular structures and to read SDFiles. A literature survey was performed in order to obtain the names and the chemical structures of all the metabolites already described in the literature for the genus *Peumus*. A total of 58 molecular structures and their predicted ^{13}C NMR chemical shifts were added to the spectral database (ACD/Labs, Ontario, Canada). Starting from the ^{13}C NMR peak list of the crude extract spectrum as input file, the principle of the algorithm is to perform a search over a part of the spectral database containing the metabolites of the genus under examination and to compare the ^{13}C NMR chemical shifts of the crude extract spectrum to the predicted chemical shifts of the database records. For each record, the algorithm constructs a list of NMR signals from the crude extract spectrum that matches signals of the record under examination (matching list). NMR signals are added to the matching list only if they satisfy two main conditions:

- ^{13}C NMR signals detected in the crude extract spectrum must have chemical shift values equal to the database record chemical shifts within a user defined tolerance of typically 2 ppm (+/- 1 ppm). This tolerance value is used in order to avoid false negatives due to variations between experimental and predicted chemical shifts.

- ^{13}C NMR signals must have intensity values close to the mean intensity of all ^{13}C NMR signals to be considered as belonging to the same molecular structure. All signals with an intensity greater or lower than the mean intensity of the matching list (^{13}C NMR peak list) by more than a user defined tolerance, typically 2 times the standard deviation around the mean intensity, are discarded.

Each match between database records and the crude extract spectrum is evaluated using a score value calculated as followed: Score = $\frac{\sum \text{matching signals from the database molecule}}{\text{total number of the database molecule signals}}$

The score value is defined as the ratio between the numbers of chemical shifts of the crude extract that match a molecular structure of the database and the total number of carbon positions of this same record. Thus, the score values are comprised between 0 (no matching peaks between the crude extract spectrum and the database records) and 1 (a set of signals in the crude extract ^{13}C NMR spectrum matches all predicted chemical shifts of a database record).

The algorithm returns a list of metabolites sorted in the order of decreasing score. We assume that the records with the highest scores have the highest probability to be present in the sample, however this assumption is accepted only if all NMR signals of the proposed record are found in the mixture spectrum. Therefore a second search is performed for the top ranked records using a larger error tolerance on the chemical shifts (+/- 1.5 ppm) in order to check the presence of potentially missing peaks. The workflow of the algorithm is illustrated in Figure 1.

Fractionation of the crude alkaloid extract by pH-zone refining Centrifugal Partition Chromatography (CPC). The performance of the database search algorithm was evaluated by comparing the results to those obtained with a previously developed ^{13}C NMR based dereplication method involving a fractionation step.[18] The crude alkaloid extract of *Peumus boldus* leaves was fractionated by pH-zone refining[29] CPC on a laboratory-scale FCPE300® (Kromaton Technology, Angers, France) equipped with a rotor made of 7 circular partition disks containing a total of 231 partition twin cells. The CPC column of 303.5 mL capacity was connected to a KNAUER Preparative Pump 1800® V7115 (Berlin, Germany). Fractions of 20 mL were collected by a Pharmacia Superfrac collector (Uppsala, Sweden). A biphasic solvent system (3 L) was prepared by mixing into a separatory funnel M*t*BE, CH_3CN, and water in the proportions 4/1/5 (v/v). After solvent system equilibration, the two liquid phases were separated. Methane sulfonic acid (MSA) was used as retainer in the aqueous stationary phase (4 mM, pH ≈ 2) and triethylamine (TEA) was used as displacer in the mobile organic phase (5 mM, pH ≈ 10). The column was filled at 200 rpm with the acid aqueous stationary phase and the rotation speed was then adjusted at 1200 rpm. The crude alkaloid extract of the leaves of *Peumus boldus* (2 g) was dissolved in a 30 mL mixture of aqueous phase/TEA-free organic phase in the proportions 1/1 (v/v). The sample was adjusted to pH 2 with MSA before injection to ensure alkaloid protonation. The sample solution was then loaded into the column through a 30 mL sample loop. The fractionation procedure was initiated by pumping progressively the alkaline organic mobile phase in the ascending mode from 0 to 20 mL/min in 5 min. The flow rate was then maintained at 20 mL/min for 90 minutes. The collected fractions (20 mL each) were checked by thin-layer chromatography (TLC) on Merck 60 F254 pre-coated silica gel plates and developed with chloroforme/methanol (93:7, v/v). Detection was performed under UV light (254 and 366 nm) and by spraying with the Dragendorff reagent. Fractions were then pooled on the basis of their TLC profile similarities, resulting in 11 pools noted from P_I to P_{XI}.

NMR analyses and data processing. All NMR analyses were performed using identical acquisition and processing parameters. The crude alkaloid extract from *Peumus boldus* (15 mg) and the fractions P_I to P_{XI} obtained by CPC were dissolved in 600 µL of DMSO-*d6*. NMR analyses were performed at 298 K on a Bruker Avance AVIII-600 spectrometer (Karlsruhe, Germany) equipped with a cryoprobe optimized for ^1H detection and with cooled ^1H, ^{13}C et ^2D

coils and preamplifiers. ^{13}C NMR spectra were acquired at 150.91 MHz. A standard zgpg pulse sequence was used with an acquisition time of 0.909 s and a relaxation delay of 3 s. For each sample, 1024 scans were co-added to obtain a satisfactory signal-to-noise (S/N) ratio. The spectral width was 238.9070 ppm and the receiver gain was set to the highest possible value. A 1 Hz line broadening filter was applied to each FID prior to Fourier transformation. Spectra were manually phased and baseline corrected using the TopSpin 3.2 software (Bruker) and calibrated on the central resonance (δ 47.60 ppm) of methanol-*d4*. The S/N ratio of all spectra was determined with the standard Bruker calculation method. Noise regions were selected from 205 to 225 ppm in all spectra and signal regions were selected differently for each spectrum as the 10 ppm spectral width centered on the most intense signal (methanol-*d4* apart). A minimum intensity threshold of 0.05 was then used to automatically collect all positive ^{13}C NMR signals while avoiding potential noise artifacts. The peak list obtained from the crude extract analysis was exported as a text file and used as input file of the database search algorithm. The 11 peak lists obtained from the CPC fraction series were also exported as text files and processed exactly as previously described [18].

GC/MS analysis. The crude alkaloid extract from *Peumus boldus* leaves was analyzed by GC/MS. Chromatographic separation was carried out using a Trace 1300 gas chromatograph (Thermo Scientific, Villebon sur Yvette, France) equipped with an AI 1310 injector. A TR-5MS (Thermo Scientific) capillary column (30 m length, 0.25 mm internal diameter, and 0.25 µm film thickness) was installed in the GC oven. The injection volume was 1 µL. Initially the temperature was set at 50 °C for 5 min followed by a 10 °C/min ramp to 300 °C. After 5 min at 300 °C the temperature was increased up to 310 °C with a ramp of 30 °C/min and stayed for 2 min. The detector was an ISQ Single Quadrupole mass spectrometer (Thermo Scientific). Helium was used as carrier gas at a flow rate of 1 mL/min. Data acquisition was performed using the electron ionization (EI). The temperature of the transfer line and that of the ion source were held at 250 °C and 230 °C, respectively. The mass range *m/z* 50-900 amu was scanned in the full scan acquisition mode with a dwell time of 0.2 s. Data were processed using the Xcalibur software. The oil components were identified by comparing the MS spectra with Libraries of NIST MS Search 2.0 program.

MS infusion analysis. An aliquot of the crude alkaloid extract from *Peumus boldus* leaves was solubilized in methanol and directly infused in a quadrupole time-of-flight hybrid mass spectrometer (QTOF micro$^{®}$, Waters, Manchester, UK) equipped with an electrospray source. The mass range of the instrument was set at m/z 100-1200 and scan duration was set at 1 s in the positive ion mode. The capillary voltage was 3000 V, the cone voltage was 35 V, and the temperature was 80 °C.

ASSOCIATED CONTENT

Supporting Information. Pseudocode of the database search algorithm

AUTHOR INFORMATION

Corresponding Author: *Jane Hubert

Jane.hubert@univ-reims.fr

NOTES

The authors declare no conflict of interest.

ACKNOWLEDGMENT

The authors thank the Givaudan Company, the CNRS and the Ministry of Higher Education and Research. The EU-programme FEDER for the PlAneT CPER project and the EU-Marie-Curie Action for the IAPP/FP7-PEOPLE “Natprotec” project are also gratefully acknowledged.

REFERENCES

(1) Gaudêncio, S. P.; Pereira, F. *Nat. Prod. Rep.* **2015**, *32*, 779–810.

(2) Hubert, J.; Nuzillard, J.-M.; Renault, J.-H. *Phytochem. Rev.* **2015**, doi:10.1007/s11101-015-9448-7.

(3) Wishart, D. S. *Bioanalysis* **2009**, *1*, 1579–1596.

(4) Carter, G. T. *Nat. Prod. Rep.* **2014**, *31*, 711–717.

(5) Nadia B, C.; Yu, K. *Chromatogr. Online*. 2013, pp 3–13.

(6) Garg, N.; Kapono, C. A.; Lim, Y. W.; Koyama, N.; Vermeij, M. J. A.; Conrad, D.; Rohwer, F.; Dorrestein, P. C. *Int. J. Mass Spectrom.* **2015**, *377*, 719–727.

(7) Allard, P. M.; Péresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V. C.; Roussi, F.; Wolfender, J. L. *Anal. Chem.* **2016**, *88*, 3317–3323.

(8) Armitage, E. G.; Barbas, C. *J. Pharm. Biomed. Anal.* **2014**, *87*, 1–11.

(9) Smolinska, A.; Blanchet, L.; Buydens, L. M. C.; Wijmenga, S. S. *Anal. Chim. Acta* **2012**, *750*, 82–97.

(10) Schripsema, J. *Phytochem. Anal.* **2010**, *21*, 14–21.

(11) Kim, H. K.; Choi, Y. H.; Verpoorte, R. *Trends Biotechnol.* **2011**, *29* , 267–275.

(12) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. *J. Exp. Bot.* **2005**, *56*, 255–265.

(13) Robinette, S. L.; Brüschweiler, R.; Schroeder, F. C.; Edison, A. S. *Acc. Chem. Res.* **2012**, *45*, 288–297.

(14) Clendinen, C. S.; Lee-McMullen, B.; Williams, C.; Stupp, G. S.; Vandenborne, K.; Hahn, D. a; Walter, G. a; Edison, A. S. *Anal. Chem.* **2014**, *86*, 9242−9250.

(15) Laude, D. A.; Wilkins, C. L. *Anal. Chem.* **1986**, *58*, 2820–2824.

(16) Bighelli, A.; Tomi, F.; Casanova, J. *Biomass and Bioenergy* **1994**, *6* , 461–464.

(17) Ferreira, M. J. .; Costantin, M. B.; Sartorelli, P.; Rodrigues, G. V; Limberger, R.; Henriques, A. T.; Kato, M. J.; Emerenciano, V. P. *Anal. Chim. Acta* **2001**, *447* , 125–134.

(18) Hubert, J.; Nuzillard, J.-M.; Purson, S.; Hamzaoui, M.; Borie, N.; Reynaud, R.; Renault, J.-H. *Anal. Chem.* **2014**, *86* , 2955–2962.

(19) Oettl, S. K.; Hubert, J.; Nuzillard, J.-M.; Stuppner, H.; Renault, J.-H.; Rollinger, J. M. *Anal. Chim. Acta* **2014**, *846*, 60–67.

(20) Abedini, A.; Chollet, S.; Angelis, A.; Borie, N.; Nuzillard, J. M.; Skaltsounis, A. L.; Reynaud, R.; Gangloff, S. C.; Renault, J. H.; Hubert, J. *J. Chromatogr. B* **2016**, *1029–1030*, 121–127.

(21) Sientzoff, P.; Hubert, J.; Janin, C.; Voutquenne-Nazabadioko, L.; Renault, J.-H.; Nuzillard, J.-M.; Harakat, D.; Magid, A. *Molecules* **2015**, *20*, 14970–14984.

(22) Hubert, J.; Chollet, S.; Purson, S.; Reynaud, R.; Harakat, D.; Martinez, A.; Nuzillard, J.-M.; Renault, J.-H. *J. Nat. Prod.* **2015**, *78*, 1609–1617.

(23) Petigny, L.; Périno, S.; Minuti, M.; Visinoni, F.; Wajsman, J.; Chemat, F. *Int. J. Mol. Sci.* **2014**, *15*, 7183–7198.

(24) Bremser, W. *Anal. Chim. Acta* **1978**, *2*, 355-365.

(25) Buevich, A.V.; Elyashberg, E. *J. Nat. Prod.* **2016**, *79*, 3105-3116.

(26) Cimino, P.; Gomez-Paloma, L.; Duca, D.; Riccio, R.; Bifulco, G. *Magn. Reson. Chem.* **2004**, *42*, S26-S33.

(27) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* **2007**, *107,*

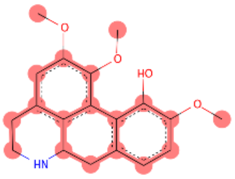
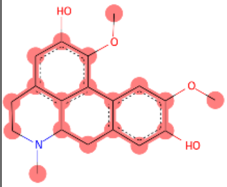
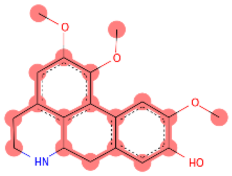
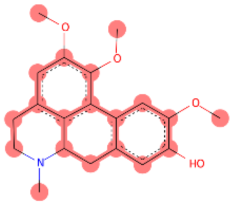
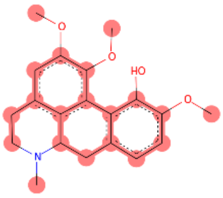
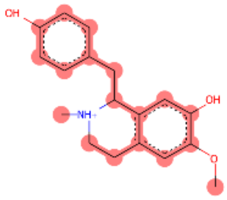
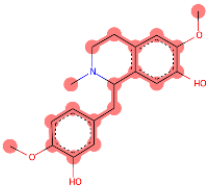
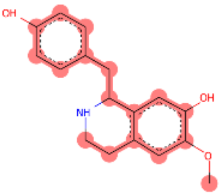
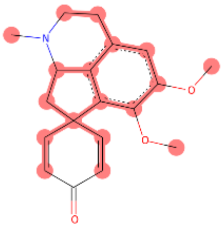
3744-3779.

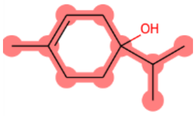
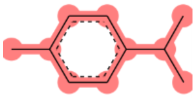
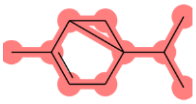
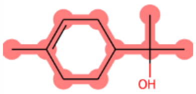
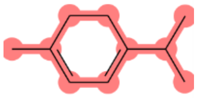
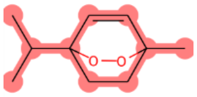
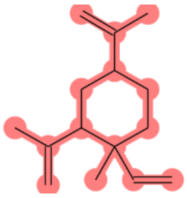
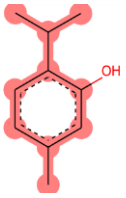
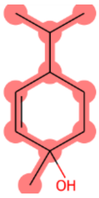
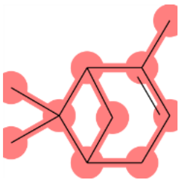
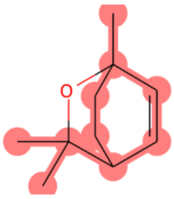
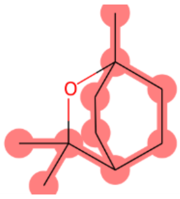
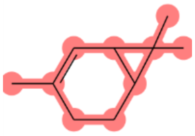
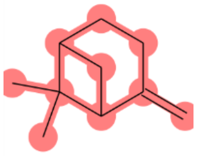
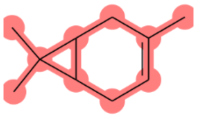
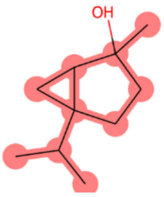
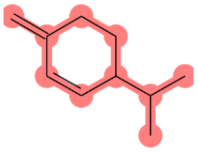
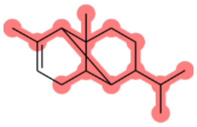
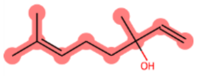
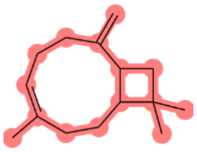
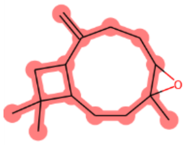
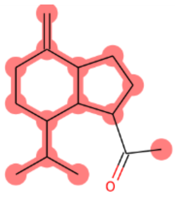
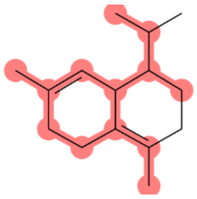
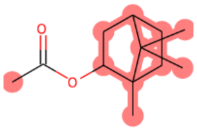
(28) RDKit http://www.rdkit.org/

(29) Okunji, C. O.; Iwu, M. M.; Ito, Y.; Smith, P. L. *J. Liq. Chromatogr. Relat. Technol.* **2005**, *28*, 775–783.

TABLES

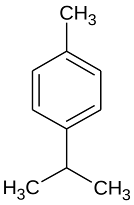
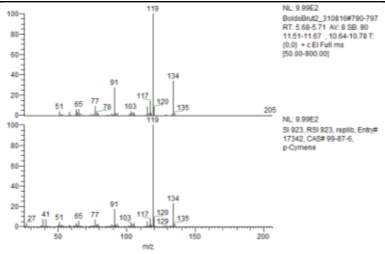
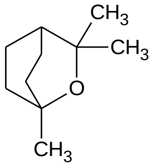
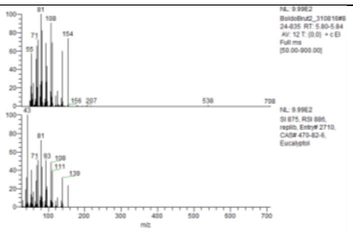
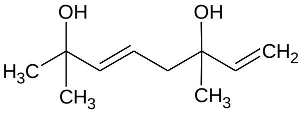
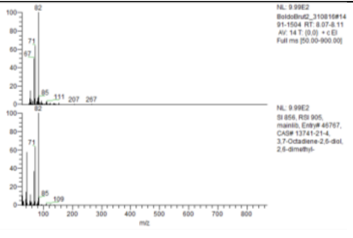
Table 1. Proposals (structures and scores) of the database search algorithm applied to a crude extract of *Peumus boldus* leaves. 33 molecules with a score higher than “0.9”, 28 molecules with a score of “1”. Highlighted carbons correspond to ^{13}C chemical shifts that match peaks from the crude extract spectrum.

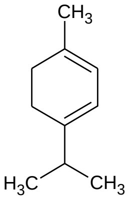
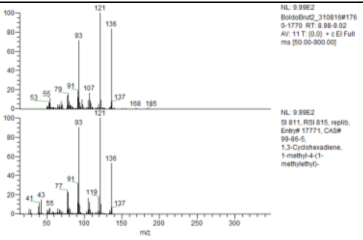
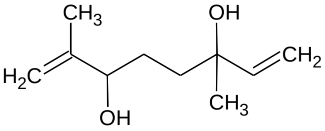
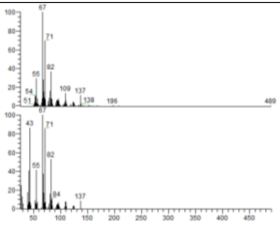
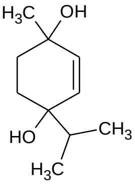
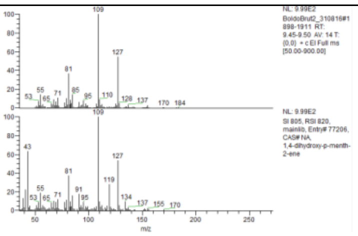
Name	Structure	Name	Structure	Name	Structure
Norisocorydine *Score = 1*		Boldine *Score = 1*		Laurotetanine *Score = 1*	
Rogersine *Score = 1*		Isocorydine *Score = 1*		N-methylcoclaurine *Score = 1*	
Reticuline *Score = 1*		Coclaurine *Score = 1*		Pronuciferine *Score = 0.947*	

Terpinen-4-ol *Score = 1*		p-cimene *Score = 1*		α-thujene *Score = 1*	
α-terpinol *Score = 1*		α-terpinene *Score = 1*		Ascaridole *Score = 1*	
β-elemene *Score = 1*		Thymol *Score = 1*		trans-*p*-menth-2-en-ol *Score = 1*	
α-pinene *Score = 1*		Dehydro-1,8-cineole *Score = 1*		Eucalyptol *Score = 1*	
2-carene *Score = 1*		β-Pinene *Score = 1*		3-carene *Score = 1*	
Sabinene hydrate *Score = 1*		Phellandrene *Score = 1*		α-copaene *Score = 0.933*	
Linalol *Score = 1*		Caryophyllene *Score = 1*		β-caryophyllene oxyde *Score = 1*	
β-oploplenone		δ-cadinene *Score = 0.928*		Bornyl acetate *Score = 0.916*	

Score = *0.933*					

1 **Table 2.**

Retention time (min)	Identification (Molecular formula)	MS spectra
3.2	Residual DMSO	
5.7	*p*-cymene ($C_{10}H_{14}$) 	
5.82	Eucalyptol ($C_{10}H_{18}O$) 	
8.1	3,7-octadiene-2,6-diol,2,6-dimethyl ($C_{10}H_{18}O_2$) 	

9.0	α-terpinene ($C_{10}H_{16}$) 	
9.3	1,7-octadiene-3,6-diol-2,6-dimethyl ($C_{10}H_{18}O_2$) 	
9.4	1,4-dihydroxy-*p*-menth-2-ene $C_{10}H_{18}O_2$ 	
9.7	$C_{10}H_{18}O_2$	*not identified*
9.8	$C_{10}H_{18}O_2$	*not identified*
10.1	$C_{10}H_{18}O_2$	*not identified*
10.6	$C_{10}H_{18}O_2$	*not identified*
11.0	$C_{10}H_{16}O$	*not identified*
11.2	$C_{10}H_{18}O_2$	*not identified*
11.3	$C_{10}H_{16}O_2$	*not identified*
13.1	$C_{10}H_{18}O_2$	*not identified*

FIGURES

Figure 1. ^{13}C NMR spectrum of the crude *Peumus boldus* extract with the two major alkaloids norisocorydine and rogersine annotated.

Figure 2. Workflow of the database search algorithm. (1) Chemical shift prediction for known metabolites for the studied genus and creation of the database. (2) ^{13}C NMR analysis of the crude extract and automatic peak picking. (3) Search algorithm: comparison of chemical shifts values of database records to those of the crude extract spectrum and prioritization of a list of putative molecules present within the crude extract. (**) Results confirmed by experimental analysis.

Figure 3. MS spectrum of the crude extract of *Peumus boldus* leaves (MS infusion, positive ionization mode). A) *m/z* 286.1 coclaurine $C_{17}H_{19}NO_3$; B) *m/z* 300.1 N-methylcoclaurine $C_{18}H_{21}NO_3$; C) *m/z* 311.1 not identified; D) *m/z* 328.1 norisocorydine OR boldine OR laurotetanine $C_{19}H_{21}NO_4$; E) *m/z* 330.1 reticuline $C_{19}H_{23}NO_4$; F) *m/z* 342.1 isocorydine OR rogersine $C_{20}H_{23}NO_4$.

Figure 4. ^{13}C NMR chemical shift clusters obtained by applying HCA on CPC fractions of *Peumus boldus*.

Figure 5. Comparison between the ^{13}C NMR profiles of the crude extract of *Peumus boldus* leaves and that of the CPC fractions (from P_I to P_{XI}). S/N: signal-to-noise ratio

Figure 6. GC/MS chromatogram obtained from the analysis of the crude extract of *Peumus boldus* leaves.

Figure 1.

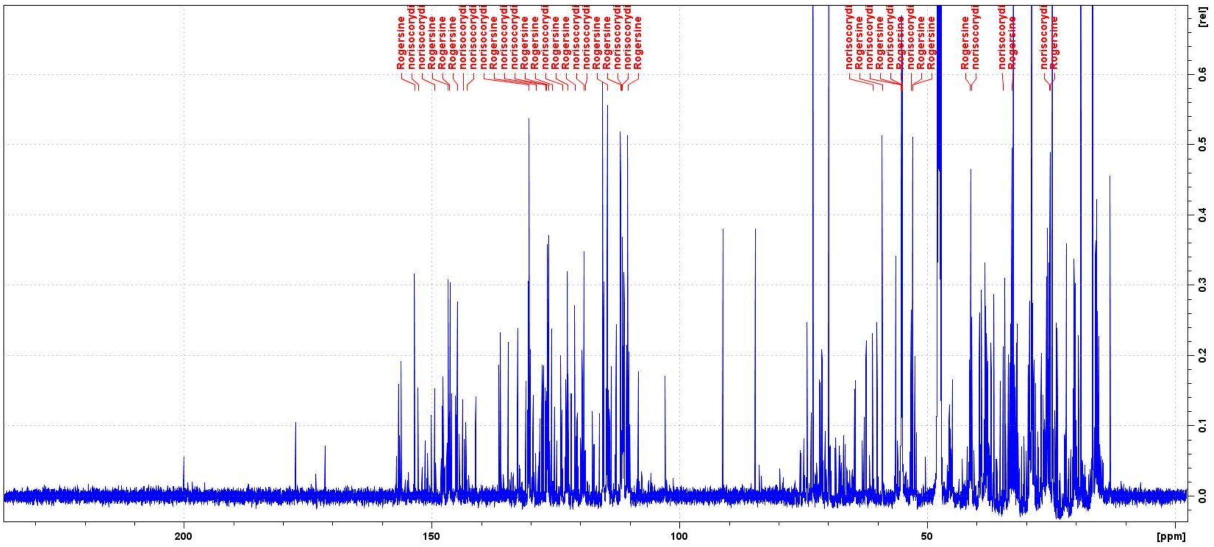
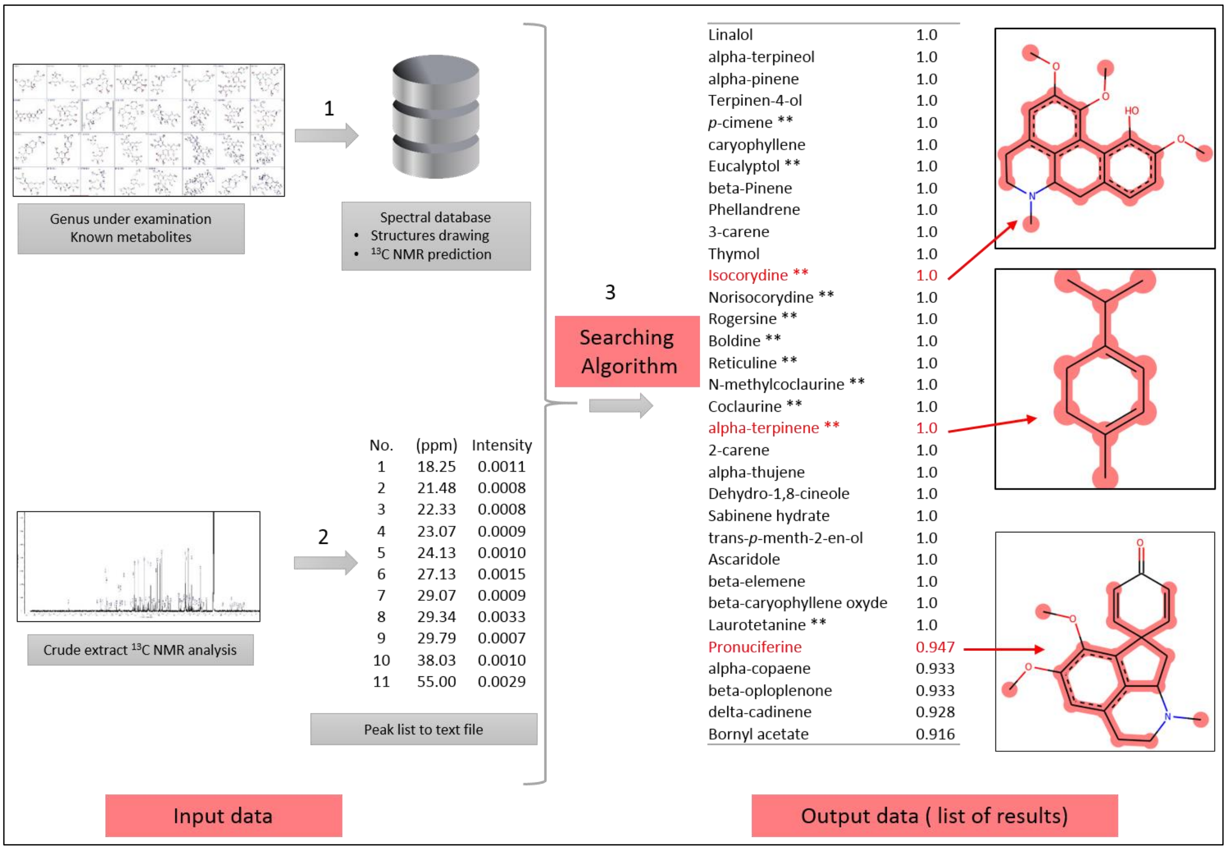

Rogersine
norisocorydi
norisocorydi
Rogersine
Rogersine
Rogersine
norisocorydi
norisocorydi
norisocorydi
Rogersine
norisocorydi
norisocorydi
Rogersine
Rogersine
norisocorydi
Rogersine
Rogersine
norisocorydi
norisocorydi
Rogersine
Rogersine
norisocorydi
norisocorydi
Rogersine
norisocorydi
Rogersine
norisocorydi
Rogersine
norisocorydi
Rogersine
norisocorydi
Rogersine
Rogersine
Rogersine
norisocorydi
norisocorydi
Rogersine
norisocorydi
Rogersine
[rel]
0.6
0.5
0.4
0.3
0.2
0.1
0.0
200
150
100
50
[ppm]

Figure 2.



No.	(ppm)	Intensity
1	18.25	0.0011
2	21.48	0.0008
3	22.33	0.0008
4	23.07	0.0009
5	24.13	0.0010
6	27.13	0.0015
7	29.07	0.0009
8	29.34	0.0033
9	29.79	0.0007
10	38.03	0.0010
11	55.00	0.0029

Linalol	1.0
alpha-terpineol	1.0
alpha-pinene	1.0
Terpinen-4-ol	1.0
p-cimene **	1.0
caryophyllene	1.0
Eucalyptol **	1.0
beta-Pinene	1.0
Phellandrene	1.0
3-carene	1.0
Thymol	1.0
Isocorydine **	1.0
Norisocorydine **	1.0
Rogersine **	1.0
Boldine **	1.0
Reticuline **	1.0
N-methylcoclaurine **	1.0
Coclaurine **	1.0
alpha-terpinene **	1.0
2-carene	1.0
alpha-thujene	1.0
Dehydro-1,8-cineole	1.0
Sabinene hydrate	1.0
trans-*p*-menth-2-en-ol	1.0
Ascaridole	1.0
beta-elemene	1.0
beta-caryophyllene oxyde	1.0
Laurotetanine **	1.0
Pronuciferine	0.947
alpha-copaene	0.933
beta-oploplenone	0.933
delta-cadinene	0.928
Bornyl acetate	0.916

Figure 3.

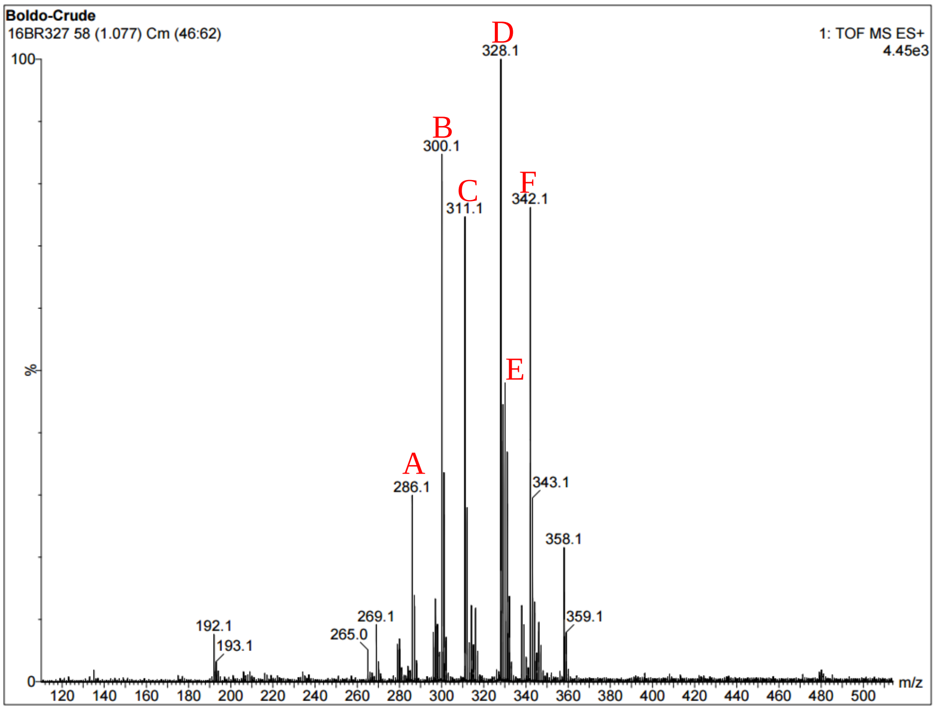

Boldo-Crude
16BR327 58 (1.077) Cm (46:62)
1: TOF MS ES+
4.45e3
100
%
0
A
286.1
B
300.1
C
311.1
D
328.1
E
F
342.1
343.1
358.1
359.1
192.1
193.1
265.0
269.1
120
140
160
180
200
220
240
260
280
300
320
340
360
380
400
420
440
460
480
500
m/z

Figure 4.

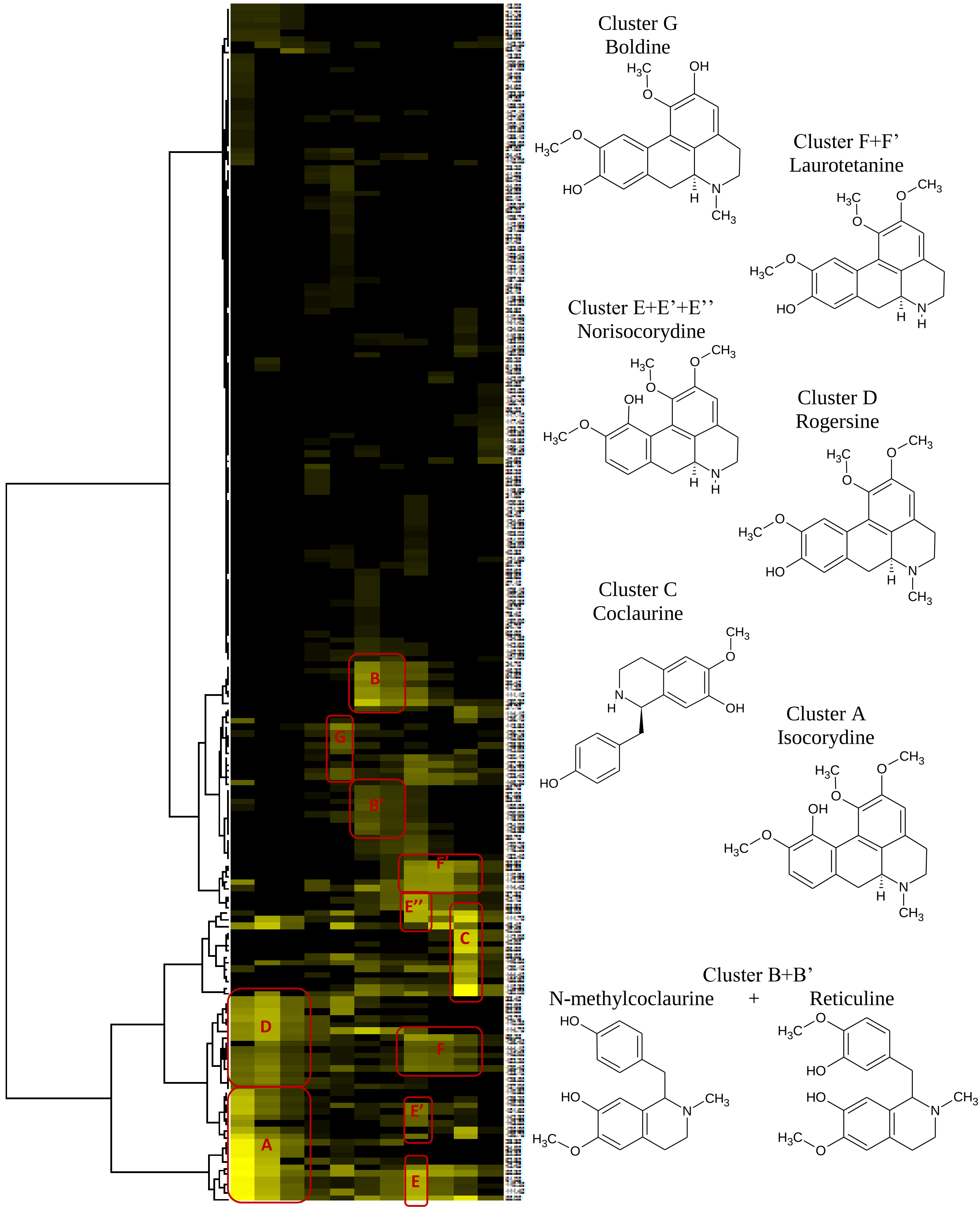
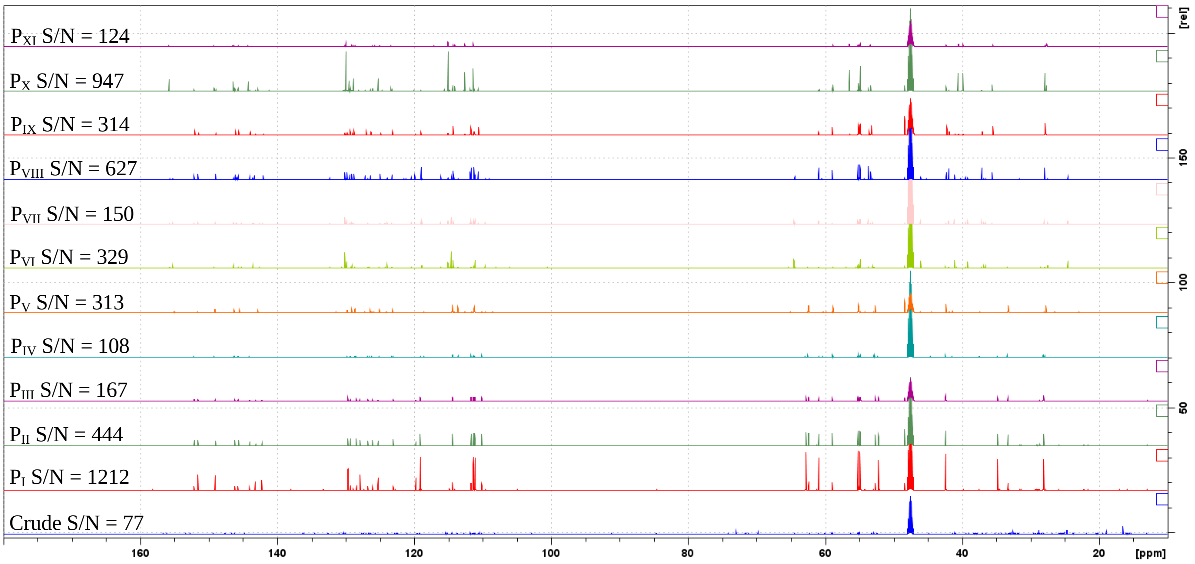
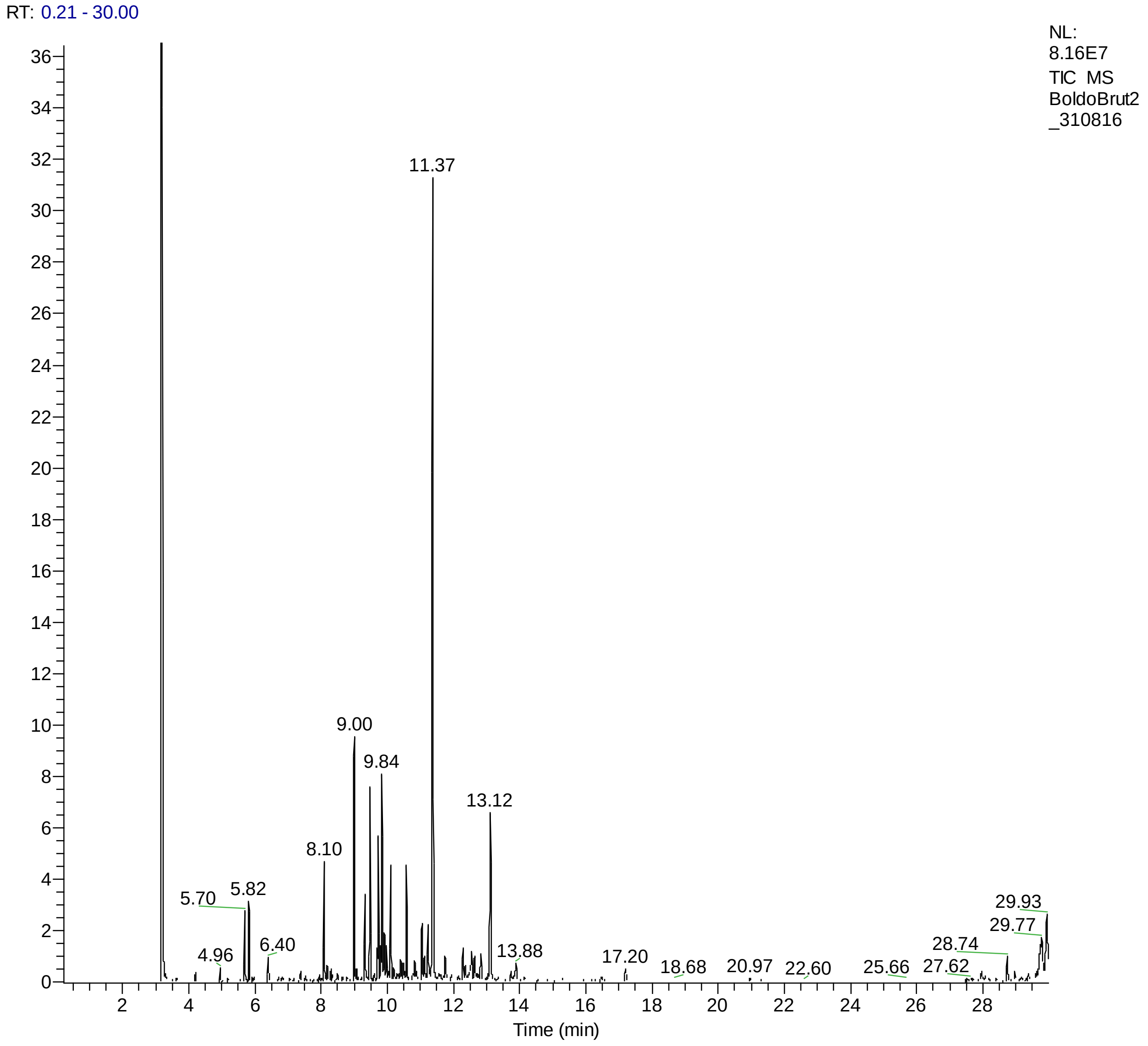

B
G
B'
F'
E''
C
D
F
E'
A
E
Cluster G
Boldine
Cluster F+F'
Laurotetanine
Cluster E+E'+E''
Norisocorydine
Cluster D
Rogersine
Cluster C
Coclaurine
Cluster A
Isocorydine
Cluster B+B'
N-methylcoclaurine
+
Reticuline

Figure 5.



P_{XI} S/N = 124
P_{X} S/N = 947
P_{IX} S/N = 314
P_{VIII} S/N = 627
P_{VII} S/N = 150
P_{VI} S/N = 329
P_{V} S/N = 313
P_{IV} S/N = 108
P_{III} S/N = 167
P_{II} S/N = 444
P_{I} S/N = 1212
Crude S/N = 77
160
140
120
100
80
60
40
20
[ppm]
[rel]
150
100
50

1 **Figure 6.**


RT: 0.21 - 30.00
NL:
8.16E7
TIC MS
BoldoBrut2
_310816
11.37
9.00
9.84
13.12
8.10
5.70
5.82
4.96
6.40
13.88
17.20
18.68
20.97
22.60
25.66
27.62
28.74
29.77
29.93
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
Time (min)