MixONat, a Software for the Dereplication of Mixtures Based on 13 C NMR Spectroscopy

Antoine Bruguière, Séverine Derbré, Joël Dietsch, Jules Leguy, Valentine Rahier, Quentin Pottier, Dimitri Bréard, Sorphon Suor-Cherer, Guillaume Viault, Anne-Marie Le Ray, et al.

MixONat, a software for the dereplication of mixture based on ^{13}C NMR spectroscopy

Antoine Bruguière[†], Séverine Derbré[†*], Joël Dietsch[†,‡], Jules Leguy[§], Valentine Rahier[§], Quentin Pottier[†], Dimitri Bréard[†], Sorphon Suor-Cherer[†], Guillaume Viault[†], Anne-Marie Le Ray[†], Frédéric Saubion[§], Pascal Richomme[†*]

† SONAS, EA921, UNIV Angers, SFR QUASAV, Faculty of Health Sciences, Dpt Pharmacy, 16 Bd Daviers, 49045 Angers cedex 01, France

‡ JEOL Europe SAS, 1 Allée de Giverny, 78290 Croissy-sur-Seine, France

§ LERIA, EA2645, UNIV Angers, SFR MathSTIC, Faculty of Sciences, 2 boulevard Lavoisier, 49045 Angers cedex 01, France

* Corresponding authors: severine.derbre@univ-angers.fr; pascal.richomme@univ-angers.fr

ABSTRACT: Whether chemists or biologists, researchers dealing with metabolomics require tools to decipher complex mixtures. As a part of metabolomics and initially dedicated to identifying bioactive natural products, dereplication aims at reducing the usually time-consuming process of known compounds isolation. Mass spectrometry and nuclear magnetic resonance are the most commonly reported analytical tools during dereplication analysis. Though low sensitive, ^{13}C-NMR has many advantages for such a study. Noteworthy, it is nonspecific allowing simultaneous high-resolution analysis of any organic compounds including stereoisomers. Since NMR spectrometers nowadays provide useful dataset in a reasonable time frame, we have embarked upon writing a software dedicated to ^{13}C-NMR dereplication. The present study describes the development of a freely distributed algorithm, namely MixONat and its ability to help researchers decipher complex mixtures. Based on Python 3.5, MixONat analyses a $\{^{1}H\}$-^{13}C NMR spectrum optionally combined with DEPT-135 and 90 data - to distinguish carbon types (*i.e.* CH_3, CH_2, CH and C) - as well as a MW filtering. The software requires predicted or experimental carbon chemical shifts (δc) databases and displays results that can be refined based on user interactions.

As a proof of concept, this ^{13}C-NMR dereplication strategy was evaluated on mixtures of increasing complexity and exhibiting pharmaceutical (poppy alkaloids), nutritional (rosemary extracts) or cosmetics (mangosteen peel extract) applications. Associated results were compared with other methods commonly used for dereplication. MixONat gave coherent results that rapidly oriented the user towards the correct structural types of secondary metabolites, allowing the user to distinguish between structurally close natural products, including stereoisomers.

Whether for natural products (NPs) identification from extracts or fractions,[1,2] for predicting the composition of biological samples[3] or for crude reaction analysis in organic synthesis,[4] chemists require tools to decipher complex mixtures. Since 1990, regarding secondary metabolites in the context of drug discovery, the idea of making structural assumptions about the composition of a complex mixture without any NPs isolation have emerged as the so-called dereplication concept.[5] Initially, dereplication aimed at avoiding bioguided fractionation steps as well as isolation and structural determination of well-known NPs. More generally, it now means to avoid spending time to determine the molecular structure of previously reported NPs. As a part of metabolomics,[6] dereplication usually requires mass spectrometry (MS) or nuclear magnetic resonance (NMR) detection together with analytical and preparative chromatographic separation, most often in a hyphenated manner.[7] Then, comparison between spectral data and NPs databases (DBs) allows formulating hypotheses about the mixture composition.[8] More recently, in order to avoid long chromatographic steps, dereplication was successfully performed directly on crude extracts and fractions.[9-11]

Nowadays, LC-MS^2 and the molecular networking initiative seem to gain importance in NPs dereplication[12] though actually both LC-MS and NMR exhibit advantages and drawbacks. The LC step before MS analyses requires an optimization of suitable chromatographic conditions that can waste valuable time. Moreover, while MS provides a higher sensitivity, a standard ionization protocol may not be suitable for a wide range of structurally different NPs. However one molecular formula deduced from high-resolution mass spectrometry is generally associated with a large number of isomers. MS/MS experiments then allow their differentiation according to specific fragmentation patterns. Introduced in 2013 by the Dorrestein group, molecular networking based on the LC-MS^2 analysis of structurally related compounds is now widely used as an efficient dereplication method.[13] During this process, similar MS^2 behavior of structural analogs leads to their gathering in clusters of nodes based on reference standards fragmentations.[14] The related Global Natural Product Social Molecular Networking (GNPS) is a web-based facility allowing sharing of raw, processed or identified MS^2 data.[15] Conversely, if a few milligrams of one mixture are available, NMR is easy to implement as it does not need any specific sample preparation, except for solubilization in a suitable deuterated sol-

vent. Last but not least, NMR detects all organic compounds and allows differentiation of stereoisomers, which may be very challenging or even impossible through MS analysis[2, 16-18].

That is why both 1D and 2D NMR experiments are often used in metabolomics[16] and dereplication applications, in addition or as an alternative to MS analysis.[19, 20] Metabolomics routinely uses ^{1}H NMR together with 2D-NMR experiments and/or spectral deconvolution to make hypotheses about the attendance of primary and, from time to time, secondary metabolites in a given organism, organ or tissue.[21, 22] The good sensitivity of ^{1}H spectra allows short acquisition times, but complex mixtures generate ^{1}H chemical shifts (δ_H) overlapping because of the low spectral range of ^{1}H-NMR which impairs correct spectra interpretation. Even if $\{^{1}H\}$-^{13}C NMR exhibits many advantages including a large spectral dispersion almost preventing from overlapping signals, it is hardly ever used for metabolomics purposes as longer acquisition times are required due to ^{13}C low natural abundance.[23, 24] Eventually on-line experimental NMR DBs such as HMDB[25] or BMRB[26] may be searched either manually or automatically in order to identify metabolites through 1D δ associations and/or 2D correlations matching.

In the field of NPs research, NMR-based dereplication processes have actually emerged together with MS-based methods. Indeed, hyphenated LC-NMR was first applied to NPs mixtures in the early 90s.[27] In this particular case, NPs were not characterized in crude mixtures strictly speaking but -after an appropriate HPLC chromatography- on the basis of 1D and 2D NMR data of an isolated compound. To our knowledge, dereplication using ^{13}C NMR was initiated even earlier, *i.e.* in 1982, on essential oils[28], the process being automated later in 1995.[29] The associated algorithm worked with an in-house DB containing the δc of volatile mono- and sesquiterpenes recorded in $CDCl_3$. Then, the SISCONST program was developed and applied for the dereplication of a mixture of terpenes in 2001, using an experimental DB of 3800 mono- and sesquiterpenes. SISCONST required δ_c and carbon multiplicities deduced from DEPT experiments. Evaluated on volatile oils, it correctly identified major compounds (> 2.2%).[30, 31] Indeed, as far as ^{13}C NMR dereplication is concerned, carbon multiplicity appears as quite a discriminant filter. Based on a *Garcinia* genus (Clusiaceae) DB (772 NPs)[32] this funnel effect is illustrated in Table 1.

Table 1. Funnel effect of multiplicities obtained by DEPT-135 and -90 additional experiments on NPs from Garcinia DB (772 NPs). **A.** Number of NPs sharing the same combination of carbon type. **B.** Number of times that one combination leads to a group of the size displayed on the right column.

A.

C	CH	CH_2	CH_3	**Nb of NPs**	C	CH	CH_2	CH_3	**Nb of NPs**
13	4	2	5	**19**	15	7	8	8	**7**
15	8	6	9	**15**	16	9	8	10	**7**
11	4	1	3	**15**	16	6	4	7	**7**
15	4	3	6	**14**	14	5	4	6	**7**
13	4	2	4	**12**	11	4	1	2	**7**
12	5	2	4	**10**	11	5	4	10	**7**
13	5	1	5	**10**	11	4	1	4	**6**
12	6	1	4	**9**	10	5	1	2	**6**
15	5	2	6	**9**	12	5	3	4	**6**
16	7	6	9	**9**	15	6	1	6	**6**
8	5	0	0	**9**	9	4	0	1	**6**
13	5	1	4	**9**	15	3	4	6	**5**
14	5	4	5	**8**	11	6	3	3	**5**
13	4	2	6	**8**	…	…	…	…	**…**

B.

Occurrences	**NPs in the group**
218	**1**
66	**2**
27	**3**
13	**4**
12	**5**
6	**7**
5	**6**
5	**9**
2	**8**
2	**10**
2	**15**
1	**12**
1	**14**
1	**19**

Most recently, in 2014, an efficient strategy based on the fractionation of a crude extract leading to a dataset of ^{13}C-NMR spectra analyzed by a hierarchical clustering analysis was proposed. The clusters of δ_c are those of the major NPs from the mixture and may be identified using either an experimental or a predicted DB.[33] The same authors later published a freely available algorithm aiming at comparing δ_c of NPs from a crude alkaloid extract of boldo with those of a predicted DB, taking into account signals intensities as well.[9]

Considering this, we describe here the development of a freely distributed algorithm, namely MixONat (Mixture of Natural Products, available at https://sourceforge.net/projects/mixonat), and its ability to dereplicate mixtures of NPs using a $\{^{1}H\}$-^{13}C NMR spectrum optionally combined with DEPT-135 and 90 data for either experimental or predicted δ_C comparisons. As a proof of concept, the procedure was applied to various vegetal fractions or crude extracts of pharmaceutical (poppy alkaloids), nutritional (rosemary extracts) or cosmetics (mangosteen peel extract) interest.

EXPERIMENTAL SECTION

Chemical and reagents. Papaverine hydrochloride was purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). It was solubilized in water and the solution was alkalinized to pH 12 using a 1M NaOH solution. It was then extracted 3 times with dichloromethane (DCM). DCM phases were gathered, dried and further analyzed. Rosmarinic acid standard was bought from Molekula (St Jean de Soudain, France).

Plant material, extraction and fractionation and analyses. *Papaver somniferum*: Poppy pods (Sample PS-201810, confidential breeding) were extracted by an aqueous solution (acetic acid 1%). The extract was filtrated on paper then alkalinized to pH 12 using a 1 M NaOH solution. It was then extracted 3 times with DCM. DCM phases were gathered, dried and analyzed by HPLC-UV (Agilent HP 1100 Series, Agilent Technologies, Les Ulis, France) using a Gemini C18 column (150 x 4.6 mm, 3µm, 100Å, Phenomenex, Le Pecq, France) with mobile phase A = 0.005 M sodium 1-heptanesulfonate buffer adjusted to pH 2.6; phase B = methanol. With a flow of 1.2 mL/min, the gradient was programmed as follows: t = 0 min, 75% A, 25% B; t = 15 min, 45% A, 55% B; t = 18 min, 45% A, 55% B; t = 20 min, 75% A, 25% B. The extract was also analyzed using LDI-MS in positive reflectron mode on a Biflex III time of flight (TOF) mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with a 337 nm pulsed nitrogen laser (model VSL-337i, Laser Sciences Inc., Boston, MA). A mass range of 40-2000 Da was chosen for spectra acquisition. Acceleration voltage was set to 19 kV, pulse ion extraction was 200 ns and laser frequency was 5 Hz. Applied laser energy ranged from 65 to 75% (86.5-93.1 µJ). Additionally, ^{13}C-NMR (17 408 scans), DEPT-135 (8 704 scans) and DEPT-90 (8 704 scans) spectra were obtained in $CDCl_3$.

Rosmarinus officinalis: Rosemary sample was bought from IPHYM laboratories (Jonage, France) in October 2018. 10 g of plant material was extracted using pressurized liquid extraction (Speed Extractor E-914, Büchi, Essen, Germany) first with DCM (3 cycles, 10 min each) under 100°C and 100 bars, to obtain 1.8 g of extract (18%), and then with methanol, during (3 cycles, 10 min each) under 100°C and 100 bars, to obtain 1.7 g of extract (17%). The E392 was obtained following one of the protocols described by the European Food Safety Authority (EFSA).[34] 2 g of dried rosemary leaves were extracted by 50 mL of cyclohexane (ultrasonic bath 10 min) and then by 50 mL of ethanol (ultrasonic bath 10 min). 45 mg (2%) of cyclohexanic extract and 180 mg (9%) of ethanolic extract were obtained. The methanol extract was analyzed by HPLC (Prominence-i LC-2030C, Shimadzu, Noisiel, France) coupled with ELSD (SEDEX 90 LT-ELSD, SEDERE) using a Luna C18 column (250 x 4.6 mm, 5µm, 100Å, Phenomenex) with mobile phase A = water + 0.1% formic acid; phase B = methanol. With a flow of 1 mL/min, the gradient was programmed as follows: t = 0 min, 70% A, 30% B; t = 5 min, 65% A, 35% B; t = 10 min, 55% A, 45% B; t = 30 min, 30% A, 70% B; t = 31 min, 0% A, 100%; t = 36 min, 0% A, 100% B. The sample was also analyzed using HPLC-MS (Esquire 3000 Plus, Ion trap, Bruker) in the same LC conditions. Furthermore, ^{13}C-NMR (2 048 scans), DEPT-135 (1 024 scans) and DEPT-90 (1 024 scans) spectra of the methanol extract (50 mg) were obtained in MeOD. The dichloromethanic extract and E392 extracts were analyzed with the same HPLC-ELSD instrument, on the same Luna C18 column but with mobile phase A = water + 0.1% formic acid; phase B = acetonitrile. With a flow of 1 mL/min, the gradient was programmed as follows: t = 0 min, 40% A, 60% B; t = 25 min, 25% A, 75% B; t = 40 min, 25% A, 75% B; t = 41 min, 0% A, 100% B; t = 46 min, 0% A, 100% B. The dichloromethanic extract was also analyzed using the same HPLC-MS instrument, in the LC conditions previously mentioned. Furthermore, ^{13}C-NMR (10 000 scans), DEPT-135 (5 000 scans) and DEPT-90 (5 000 scans) spectra of the dichloromethanic extract (50 mg) were obtained in DMSO-d6. The E392 extract (50 mg) was also analyzed by NMR, ^{13}C-NMR (1 024 scans), DEPT-135 (512 scans) and DEPT-90 (512 scans) in DMSO-d6.

Garcinia mangostana: 1 kg of fresh mangosteen fruits, imported from Vietnam, were purchased in an Asian grocery store (Tang Frères, Paris, France) in September 2017. A sample was identified by PR as *Garcinia mangostana* fruits and a voucher specimen (GM-201701) was kept at the laboratory's herbarium. The pericarps were separated from the rest of the fruit, dried and grinded. 45 g of grinded pericarps were extracted using pressurized liquid extraction (Speed Extractor E-914, Büchi) with cyclohexane (3 cycles, 8 min each) under 100°C and 100 bars, to obtain 5.6 g of extract (12.4%). The cyclohexane extract was analyzed by HPLC-UV (Waters 2695 with a diode array detector Waters 2996, Guyancourt, France) using a Gemini C18 column (150 x 4.6 mm, 3µm, 100Å, Phenomenex) with mobile phase A = water + 0.1% formic acid; phase B = acetonitrile. With a flow of 0.75 mL/min, the gradient was programmed as follows: t = 0 min, 40% A, 60% B; t = 30 min, 10% A, 90% B; t = 32 min, 0% A, 100% B; t = 37 min, 0% A, 100% B. Additionally, ^{13}C-NMR (10 000 scans), DEPT-135 (8 000 scans) and DEPT-90 (8 000 scans) spectra were obtained in $CDCl_3$.

4 g of the cyclohexane extract were fractionated with a CombiFlash Rf-200 system (Teledyne Isco, Lincoln, NE, USA) equipped with binary pumps, multiwavelength UV detectors, and fraction collectors using a Si-HC 50µm Si-OH 120g column (Interchim, Montluçon, France) with mobile phase A = petrol ether; phase B = 60% petrol ether, 30% chloroform, 10% acetone. With a flow of 60 mL/min, the gradient was programmed as follows: t = 0 min, 100% A, 0% B; t = 2 min, 100% A, 0% B; t = 7 min, 80% A, 20% B; t = 12 min, 80% A, 20% B; t = 17 min, 60 % A, 40% B; t = 22 min, 60% A, 20% B; t = 27 min, 40 % A, 60% B; t = 42 min, 40% A, 60% B; t = 50 min, 0% A, 100% B; t = 60 min, 0% A, 100% B. 7 fractions were obtained: F1 (52.4 mg, 9-hydroxycalabaxanthone[35]), F2 (90.9 mg, garcinone E[36]), F3 (1.8 g, α-mangostin[35]), F4 (483.0 mg, γ-mangostin[35]), F5 (105.0 mg), F6 (381.2 mg) and F7 (312.3 mg). F6 was analyzed by HPLC-UV, in the same conditions as the crude extract. Additionally, ^{13}C-NMR (12 000 scans), DEPT-135 (6 000 scans) and DEPT-90 (6 000 scans) spectra were obtained in $CDCl_3$.

310.7 mg of F6 were purified by flash chromatography (CombiFlash Rf), using a Si-OH 40µm 40g column (Macherey-Nagel, Hoerdt, France) with mobile phase A = petrol ether; phase B = 50% chloroform, 50% ethyl acetate. With a flow of 30 mL/min, the gradient was programmed as follows: t = 0 min, 100% A, 0% B; t = 5 min, 100% A, 0% B; t = 10 min, 90% A, 10% B; t = 20 min, 90% A, 10% B; t = 22 min, 86% A, 14% B; t = 33 min, 86% A, 14% B; t = 35 min, 80% A, 20% B; t = 45 min, 80% A, 20% B; t = 50 min, 70% A, 30% B; t = 60 min, 70% A, 30% B; t = 65 min, 0% A, 100% B. From t = 70 min to t = 75 min, the solvent system was switched to 100% ethyl acetate.

8-deoxygartanin[37] (23 mg), gartanin[37] 49 mg), β-mangostin[38] (25 mg) and gudraxanthone[35] (12 mg) were purified.

NMR analyses and data processing. The extracts or fractions (30 mg) were dissolved in 600 µL of the chosen deuterated solvent. NMR analyses were performed at 298 K on a JEOL 400MHz YH spectrometer (JEOL Europe, Croissy-sur-Seine, France) equipped with an inverse 5 mm probe (ROYAL RO5). For ^{13}C NMR (100 MHz) spectra, a WALTZ-16 decoupling sequence was used with an acquisition time of 1.04 s (32768 data points) and a relaxation delay of 2 s. 12 000 scans were collected for 30 mg of extract to obtain a satisfactory S/N ratio. For *G. mangostana* crude extract, a Bruker Avance HD 700 MHz spectrometer (Bruker, Palaiseau, France) equipped with a 5 mm cryoprobe (ATMA gradZ) was also used. ^{13}C NMR spectra were acquired at 175 MHz. A standard zgdc2 sequence was used with similar acquisition and relaxation times. The spectral width was 250 ppm. A 1 Hz exponential line broadening filter was applied to each FID prior to Fourier transformation. Spectra were manually phased and baseline corrected using the MestReNova software (Mestrelab Research, Santiago de Compostela, Spain) and referenced on the central resonance of the deuterated solvent.[39] For DEPT experiments, 6 000 scans were required for 30 mg of extract and alignment on ^{13}C spectrum were made using a selected δ_C. A minimum intensity threshold was then used to automatically collect positive ^{13}C NMR and DEPT-90 signals and positive and negative DEPT-135 signals while avoiding potential noise artifacts. The peak list and intensity data obtained from each spectrum were exported as a .csv file using Excel Microsoft Office (Microsoft, Redmond, WA, USA) software and used as an input file in MixONat software.

DBs: general information. To create a DB of NPs and their δ_C that can be used by the dereplication program, the first step is to gather the structures of the compounds of interest (*e.g.* NPs previously identified in a genus or a botanical family), either by drawing them with a dedicated software (*e;g.* ChemDraw[40], ChemSketch[41]) or downloading them from various DBs accessible through subscription (*e.g.* SciFinder[42], Dictionary of Natural Products[32]). Once the individual files (.mol, .cdx, .sk2) are gathered in a structure data file (.sdf), their δ_C are predicted using a NMR prediction software (*i.e.* ACD/Labs Spectrus processor and C/H-NMR Predictor). [43, 44] The C-typeGen program included in the software (Figure S1) creates a suitable DB: it reads the SDF and sorts each chemical shift by carbon type. A new SDF is then created. The latter contains, for each compound of the DB, the predicted δ_C values organized as methyl, methylene, methine or quaternary carbons. The creation of such a DB is required for the MixONat algorithm to work properly.[45]

Specific DBs. Garcinia DB was created by gathering the structures of molecules described in the *Garcinia* genus from the Dictionary of Natural Products[32], leading to a SDF containing 718 NPs. Lamiaceae DB was built by searching for compounds described in the Lamiaceae family on SciFinder, resulting to a database of 982 NPs. Papaveraceae DB was also created using SciFinder, gathering molecules isolated from the Papaveraceae and resulting in a 174 NPs DB. The CH-NMR-NP DB containing the experimental δ_C of 32 854 NPs was created from the data available (*i.e.* structure, name, molecular formula, molecular weight, source, ^{13}C chemical shifts, deuterated solvent, reference) on the dedicated JEOL website[46].

MixONat software. The algorithm was implemented in the Python 3.5 programming language. [47] The open source cheminformatics package RDKit was used to draw the molecular structures and read SDFiles[48]. The freely distributed software can be found at https://sourceforge.net/projects/mixonat.

MixONat software: Inputs and parameters. A graphical user interface (GUI) was designed with Kivy, an open Python library source compatible with Linux, Windows and OS X.

The home tab of the MixONat software (Figure S2) allows the selection of the input files by the user, *i.e.* at least a DB (.sdf), sorted by the C-typeGen program and ^{13}C-NMR data, imported as a table (.csv) of δ_C values and intensities. The users are also encouraged to provide DEPT-135 and 90 data (.csv).

The second tab (Figure S3) displays all the different parameters that can be adjusted:

- The tolerance (ε) reflects the accuracy of the used database, as the program compares each chemical shift in the experimental spectrum (δ_{13C}) with each chemical shift in the SDF (δ_{SDF}) for each molecule. It considers that δ_{13C} matches with δ_{SDF} if $\delta_{SDF} - \varepsilon < \delta_{13C} < \delta_{SDF} + \varepsilon$. The default value for ε is 1.3.[44,49]
- The tolerance incrementation can be turned ON or OFF. If ON, the program starts the matching process with ε = 0.0 and then increments this value by steps of 0.1 ppm, until it reaches the ε value. This matches first the chemical shifts that are closest together. If this parameter is OFF, the algorithm matches a δ_{SDF} with the first δ_{13C} falling into the ± ε interval. It is recommended to leave this parameter ON, especially when using experimental DBs.
- The DEPT alignment parameter corresponds to the tolerance on alignment of δ_C values from ^{13}C NMR and DEPT spectra. As a perfect alignment really never occurs, this parameter allows a correct association of a δ_{13C} to its corresponding carbon in the DEPT-135 and 90 spectra. The default value is set as an usual digital resolution of 0.02 ppm.
- The equivalent carbon factor can be turned ON or OFF. Turning it ON allows same δ_{13C} to be matched multiple times if several identical δ_{SDF} are found (equivalent carbons in the database).
- The molecular weight filter only shows results if they correspond to the ones requested by the user.

The third and last tab is the C-typeGen program that allows the user to create DBs compatible with the MixONat dereplication process. It sorts δ_{SDF} by carbon types.

MixONat algorithm: Matching process (Figure 1). First, the program starts sorting each carbon of the ^{13}C NMR spectrum depending on its type. This sorting is different, depending on the DEPT files provided by the user and according to the chosen DEPT alignment. If there is no DEPT data, then the carbon types are not differentiated. If a DEPT 135 has been given, carbons exhibiting negative intensities are considered as methylenes whereas other ones are considered as methines or methyls. Carbons of the ^{13}C NMR spectrum not detected in DEPT experiments are considered as quaternaries. Adding a DEPT-90 spectrum allows to distinguish methines from methyls.

In a second step, the matching process consists, for each compound of the DB, in the comparison of δ_{13C} with δ_{SDF}. It is done by

list of carbon types. All δ are first sorted by descending numerical order before the matching process. During the latter, the algorithm allows or not multiple uses of δ_{13C}, depending on the equivalent carbon parameters. When all the δ_{SDF} have been considered, the score and error of the molecule are calculated and stored. The score is defined as the number of δ_{13C} matched with δ_{SDF} out of the number of carbons of the compound. The error is the cumulated absolute difference between matched signals (*i.e.* $\Sigma\,|\delta_{SDF}-\delta_{13C}|$).

MixONat software: Interactive results. At the end of the matching process, a fixed number of molecules is displayed provided they reach a minimal score set by the user. Compounds are ranked by decreasing score. Those with similar score are secondly ranked by increasing error. Their structure, name, molecular weight, score and error are indicated (Figure S4). On the structure, matched carbons are highlighted in red. For each molecule, it is possible to open a window showing the numbered structure with matched δ_{SDF}. Matched δ_{SDF} and intensities are also shown along with different colors in a graphical representation of a reconstructed ^{13}C-NMR spectrum. Signals are numbered according to the structure whilst a chemical shift list is displayed. This easily shows if intensities of matched carbons are homogenous, hence hypothetically being signals belonging to the same molecule (Figure S5). It is thus possible for the user to link the information gathered from the structure, the spectrum and the chemical shifts and eventually decide to remove or add chemical shifts, then modifying the score of the selected molecule. This function can be, for example, used to remove a carbon matched with an abnormal intensity (meaning it probably belongs to another molecule), or to add a quaternary carbon that was not matched because it was not picked on the spectrum due to its low intensity, or predicted a bit too far away, etc.[50] If a carbon is added or removed, the spectrum and the highlighted carbons on the structure will be updated accordingly. It is also possible to delete a molecule if necessary (*e. g.* for chemotaxonomy considerations). Once the results have been checked by the user, they can be saved as a text and image file (Figure S6).

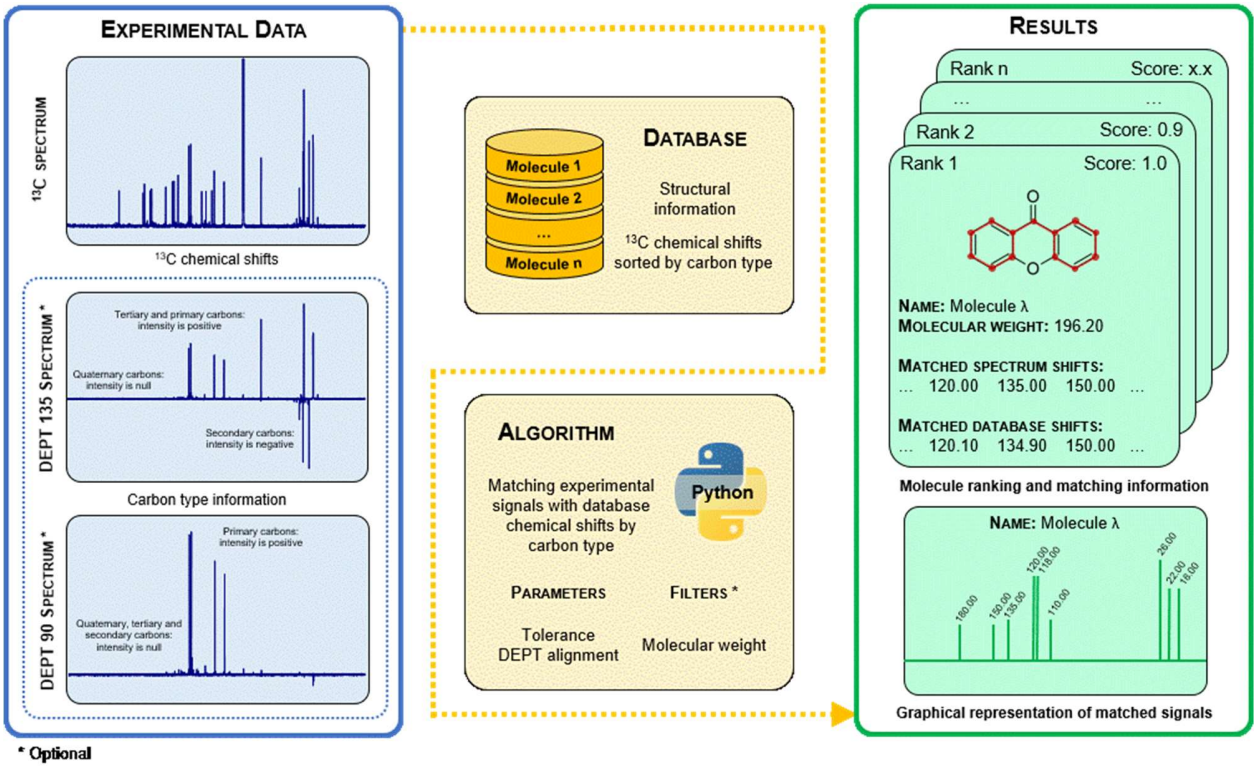


Figure 1. General process of the MixONat program. The program matches chemical shifts from the user's experimental (^{13}C and optionally DEPT-135 and DEPT-90) with those of a database collecting selected molecules δ_C.

RESULTS AND DISCUSSION

In order to evaluate the relevance of the present ^{13}C-NMR dereplication strategy using the MixONat software and the pertinence of the results, we have applied the whole process to various herbal extracts. A comparison was also done with other methods commonly used for dereplication.

Poppy alkaloids. Poppy alkaloids are widely used all over the world for pharmaceutical applications, mainly as painkillers. They are extracted from various varieties of *Papaver somniferum*. If opium was initially the raw material, poppy straw is nowadays preferred by industrials. Different chemical varieties are exploited, containing mainly either morphine, thebaine, codeine, noscapine or other isoquinoline alkaloids.[51] The present ^{13}C-NMR dereplication process was thus evaluated for its ability to discriminate between different chemotypes obtained during a confidential breeding experiment. The chosen extract contained papaverine and noscapine as major NPs (Figures S7). At first, only δ_C from the ^{13}C spectrum were considered (Figure S8) and "equivalent carbons" were authorized. Using Papaveraceae DB (174 NPs), papaverine was predicted at rank 1 (Score 1.00) when (-)-noscapine [syn. (-)-narcotine] was suggested at the 4th position (score 0.77) (Figure S9). Indeed, for this alkaloid,

5 signals did not match as their chemical shifts were predicted more than 1.3 ppm away from their expected value in $CDCl_3$ (Table S1). The way the algorithm works sometimes hinders matching. For example, as far as δ_{13C} at 62.5 and 61.0 ppm in the extract were concerned, the algorithm started to match the higher one at δ_C = 62.5 ppm with the closer predicted δ_{SDF} at 62.0 ppm. Then, the second one at 61.0 ppm was too far away from the predicted δ_{SDF} at 63.1 ppm. However, as MixONat software offers an interactive interface, the user may correct the matching after careful analysis of results. In this example, dehydrocavidine (score 0.81) and berberrubine were proposed at positions 2 and 3 respectively (Figure S9). Additional DEPT-135 and 90 experiments were thus registered and used during the dereplication process to discriminate carbons type. Papaverine remained at the 1st position and (-)-noscapine moved to the 3rd one. Dehydrocavidine switched to rank 4 just after the latter with a score decreasing to 0.71 because of 2 more unmatched carbons (Figure S10). Finally, (-)-eschscholtzine appeared at rank 2. However, the authorization of "equivalent carbon" by the software artificially doubled the score of this symmetrical NP. Together with chemotaxonomic consideration, the user would eliminate such a hypothesis. Finally, as expected, MixONat software successfully helped the user to identify papaverine and (-)-noscapine in this poppy extract.

One can wonder about the use of predicted chemical shifts. In this example, the comparison of papaverine δ_{13C} with the ones described in literature ($CDCl_3$) was not convincing.[52] Thus, the ^{13}C-NMR spectrum of a commercial reference was recorded in $CDCl_3$, leading to a perfect matching (Table S2).

Rosemary phenols. Rosemary is a medicinal plant which leaves are traditionally used to relieve symptoms of dyspepsia and treat mild spasmodic disorders of the gastrointestinal tract.[53] According to the European pharmacopeia, it contains more than 3% of hydroxycinnamic derivatives such as rosmarinic acid. Triterpenes and tricyclic phenolic diterpenes (*e.g.* carnosic acid and carnosol) are also described as major NPs. The latter are found in large amount in extracts used in Europe,as food antioxidants (*i.e.* E392 additive).[34] After successive extraction of *Rosmarinus officinalis* leaf by DCM and MeOH, their major NPs were investigated using the Lamiaceae DB, ^{13}C-NMR spectrum, DEPT-135 and 90 experiments and the MixONat software. "Equivalent carbons" were authorized. Results were compared with LC-ELSD-DAD-MS^n data.

The MeOH extract contained rosmarinic acid and a disaccharide, suggested as sucrose (Figure S11 and Table S3). Out of 982 compounds in the DB, the ^{13}C-NMR dereplication process predicted rosmarinic acid (score 0.83) and sucrose (score 0.83) at ranks 4 and 5 respectively (Figures S12-13). It should be noted that rosmarinic acid is an ester including a caffeic acid moiety, predicted at rank 1. The NP proposed in position 3 is an ester of caffeic acid and sucrose. 3 δ_{13C} of rosmarinic acid were not matched by the algorithm. First, the methine at 117.1 ppm was set aside due to the way the algorithm works, as explained earlier. However, the user might correct the matching after ^{13}C-NMR spectrum examination and associate the signals at 115.1, 115.2, 116.2, 116.4 and 117.5 together with the predicted ones at 115.1, 116.2, 116.5, 117.1, and 117.5. Then, the signal at 129.3 ppm was a little bit too far from its prediction at 130.6 ppm. The use of the interactive interface would also correct this to reach as score of 0.94 and the 2nd rank. As far as the predicted δ_{SDF} at 173.8 was concerned, it was absent from the spectra (Figure S13), probably due to a matrix effect, as already observed.[54] Finally, with the help of the MixONat software, the user would easily identify the two major compounds of this extract.

Both the rosemary DCM extract and the so-called E392 antioxidant additive share similar major metabolites, *i.e.* the triterpenes ursolic, oleanolic, betulinic and micromeric acids and the phenolic diterpenes carnosic acid and its derivatives (Figure S14-15 and S19, Table S4). Indeed, while triterpenes are stable NPs, carnosic acid will spontaneously oxidized in carnosol and degradation products,[55] impacting the anti-oxidant activity of the whole extract.[56] Thus, the present ^{13}C-NMR dereplication process was investigated as a mean to predict the presence of carnosic acid or its degradation products.

The E392 extract obtained by ethanolic extraction of *R. officinalis* leaves after a delipidating step contains mainly carnosic acid together with aforementioned triterpenes (Figure S14). As Lamiaceae plants (*e.g.* mint, lavender, sage, thyme) are well-known to biosynthesize mono- and sesquiterpenes from essential oils, such low molecular weight NPs constitute a large part of the Lamiaceae DB. To focus on the non-volatile NPs from the rosemary dry extracts, a molecular weight filter was used and only compounds beyond 250 Da were selected. Amongst the 982, MixONat predicted the presence of betulinic, ursolic, micromeric, oleanolic and carnosic acids in ranks 4, 6, 11, 17 and 22 respectively, with scores ranging from 0.90 to 0.75 (Figure S16). Among the first 24 suggested NPs, 10 are triterpenes from either ursane, lupane or oleanane types. At ranks 1-3, were respectively suggested (+)-lupeol, viminalol and (+)-betulin whose structure are very close to betulinic and ursolic acids. 10 are diterpenes, 6 of them sharing the same abietane skeleton as carnosic acid and its derivatives. At this stage, the user can opt for a careful comparison between E392 δ_C and literature data for these first ranked NPs. However, LC-MS^2 analysis of E392 gave us the molecular weight of the major NPs, *i.e.* 330, 332, 454 and 456 Da (Table S4). Finally, using these specific values as a filter, MixONat software managed to successfully identify betulinic acid (score 0.9), ursolic acid (score 0.87), micromeric acid (score 0.87), oleanolic acid (score 0.83) and carnosic acid (score 0.75) in the first 5 ranks (Figure S17). However none of these compounds reached a score of 1. Proposed in the first position, betulinic acid (C_{30}) had 2 carbons that were incorrectly predicted ($\Delta\delta_C$ > 1.3 ppm): C-3 (78.4) and C-19 (48.6 ppm). The quaternary C-4 was predicted at 38.1 ppm but its signal was hidden by the ones from C-1 and C-13. C-3 and C-4 were not matched neither in ursolic, micromeric and oleanolic acids for the same raisons. For the latter, other unmatched δc were either due to missing signals or inaccurate predictions (Table S5). Back to the antioxidant diterpenes, carnosic acid reached position 5 because, as a minor compound, signal for quaternaries C-9, C-11 and C-12 were missing. Furthermore, the δ_{SDF} of C-14 and C-20 were overpredicted (> 1.3 ppm). Nevertheless, it came in ahead the DB's diterpenes, *i.e.* 6,7-dehydrocarnosic acid (rank 6, 0.7), carnosol (rank 8, 0.6) or carnosic acid quinone (rank 10, 0.6) demonstrating that this process allows to decipher carnosic acid, carnosol or their degradation products in such a complex mixture.

The same method applied on the rosemary DCM extract containing the same triterpenes but carnosol instead of carnosic acid (Figure S19, Table S4) led to the same conclusions as carnosol reached rank 1 (Score 0.95, Figure S20) followed by the triterpenes. Only C-13,

predicted at 135.1 ppm, was not matched with δ_{13C} at 134.3 ppm in the spectrum as the latter was previously suggested as C-8 (δ_{SDF} 133.6 ppm). But both C-8 and C-13 chemical shifts (δ_{13C} 131.6 and 134.3 ppm) appeared in the ^{13}C-NMR spectrum (Figure S21, Table S6).

Mangosteen peel xanthones. Regarding the former example, one may wonder if a fractionation step is required before ^{13}C-NMR dereplication of complex mixtures, including compounds exhibiting similar backbones. Thus, in the framework of a project that aimed at discovering new Unfolded Protein Response (UPR) modulators,[57-60] the same methodology was finally applied to a *G. mangostana* fruit peel apolar extract together with one chromatography fraction. Mangosteen is well-known to contain bioactive prenylated xanthones. Indeed, one LC-UV analysis revealed α-mangostin, γ-mangostin, gartanin, garcinone E and 8-deoxygartanin as major NPs in the cyclohexanic crude extract whereas gudraxanthone, 9-hydroxycalabaxanthone and β-mangostin appeared as minor products (Figure S22). After acquisitions of the required NMR spectra using a routine 400 MHz NMR spectrometer, the ^{13}C-NMR dereplication of this extract was then undertaken using the aforementioned *Garcinia* DB. Equivalent carbons were allowed. As a result, out of the 718 NPs, 23/25 suggested compounds with scores > 0.9 appeared as xanthones bearing at least one prenylated side chain (Figure S23). It should be noted that all these proposals share either the 1,3-dihydroxy,2-prenylxanthone or either the 1,3,6,7-tetrahydroxy,8-prenylxanthone scaffold of the major products α-mangostin (68% of the extract, Figures S22-S23). The latter reached the 1st position as all the predicted δ_{SDF} were matched with a δ_C by the MixONat software (Table S7). Then garcinone E, γ-mangostin, gartanin and 8-deoxygartanin were suggested in position 4, 10, 11 (score 0.96) and 20 (score 0.93) respectively. For all of them, the predicted δ_{SDF} that were not matched were either missing quaternary carbons (C-7 of garcinone E) or a CH not picked in the DEPT-90 experiment (C-2" of γ-mangostin, C-7 of gartanin, C-6 and C-8 of 8-deoxygartanin) (Table S7 and Figure S25). Indeed, during the methines peak picking process, the threshold was set as to avoid signals due to the methyl groups (*i.e.* OMe at 62.2 ppm) of the major α-mangostin. This exemplifies the limitation of the present ^{13}C-NMR method to dereplicate the minor NPs in crude extracts.

To benchmark the ability of MixONat to identify minor compounds, the same mangosteen extract was analyzed on a 700 MHz NMR spectrometer. The same experimental conditions were used allowing the detection of chemical shifts due to minor xanthones (Figure S26, amount ≤ 1.0%). Except for gudraxanthone (see below) the eight xanthones of the mixture (Figure S22) were ranked between 1 and 22. The major α-mangostin reached the 2nd position with all predicted δ_{SDF} matched, followed by garcinone E, β-mangostin, 9-hydroxycalabaxanthone, 8-deoxygartanin, gartanin and γ-mangostin at position 12, 14, 17-19 and 22 respectively (score 0.96) (Table S8, Figure S27). Subsequently, their actual presence in the extract was confirmed by a careful comparison of the experimental chemical shifts with those found in literature.

Another way to identify minor NPs consists in a coarse fractionation of complex crude extracts. The mangosteen peel extract was thus fractionated and the dereplication process applied again on a fraction containing minor xanthones, *i.e.* gartanin, 8-deoxygartanin, β-mangostin and gudraxanthone (Figure S28). After such a concentration step, ^{13}C-NMR dereplication showed that 8-deoxygartanin and β-mangostin reached the two first position with a perfect match, whereas gartanin and gudraxanthone were suggested at positions 10 and 24 respectively (score 0.96 and 0.89, Figure S29-30). Concerning gartanin, its quaternary C-8a predicted at 107.8 ppm was not matched (δ_C 109.3 ppm). With the same causes producing the same effects, C-5 and C-8a of gudraxanthone predicted at 146.4 and 122.7 ppm from experimental data in DMSO-d6 were not matched either. Finally, a DB of experimental chemical shifts, namely CH-NMR-NP DB was also used.[46] Amongst 32854 entries, β-mangostin, 8-deoxygartanin and gartanin reached the three first position with a perfect match whilst gudraxanthone was not suggested (Figure S31). The latter was suggested in the position 122 with a score of 0.79. Indeed, the NMR was described in DMSO-d6 only[35,61], inducing discrepancies in δ_C values. This demonstrates, however, that when DBs containing NPs of interest and their experimental chemical shifts are available, their use obviously increases chances for better matches. Even if the MixONat program gives satisfactory results with predicted DBs, dereplication based on NMR would then greatly benefit from a repository of publicly accessible raw NMR data for all published NPs.[62-64]

CONCLUSION

To decipher complex mixtures using ^{13}C-NMR data, we propose here the freely available software MixONat, a complementary tool to those already existing, notably based on LC-MS profiling. Although initiated at the same time as MS methodologies, ^{13}C-NMR-based dereplications remained less popular, probably due to their lower sensitivity. However, NMR spectrometers nowadays provide valuable dataset within a reasonable amount of time on quantities of the order of 10-30 mg. Trying to integrate the best of former works,[9, 29, 31, 33] the present methodology allows to take into account the type of carbon through DEPT experiments and classifies compounds from a specific DB according to decreasing scores. Signal intensities are eventually monitored through an interactive interface. The MixONat software also requires DBs that may contain predicted δ_{SDF}. Therefore, in the event of pre-profiling analyses, neither reference compounds nor published data are initially required.

For all analyzed mixtures, MixONat suggested the correct NPs in first ranks with coherent results that rapidly direct the user towards a particular structural type. Moreover, the software is able to distinguish structurally close NPs, including stereoisomers. Then, manual comparison of the best-ranked hypotheses with literature data may confirm identifications. Most of the time, no more than the twenty first compounds need to be checked, out of several hundreds of molecules present in the DBs. Interactive results greatly facilitate the work of finding the right NPs. Finally, one can imagine the interest to associate ^{13}C-NMR data and MixONat software to LC-MS^2 data to assert the major compound of a mixture, including stereoisomers.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI:

Additional information as noted in the text (PDF): Screen shots of MixONat software, HPLC chromatograms and ^{13}C NMR spectra of studied

extracts and fractions, structures of cited NPs, parts of displayed results, tables to follow the matching processes as well as details on MixONat software (input and parameters, matching process and interactive results)

AUTHOR INFORMATION

Corresponding Authors

* E-mail: severine.derbre@univ-angers.fr and pascal.richomme@univ-angers.fr

Author Contributions

The manuscript was written through contributions of all authors.

ACKNOWLEDGMENT

AB gratefully acknowledges the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI) for his PhD grant.
The authors warmly thank Serge Akoka, Patrick Giraudeau and the CEISAM laboratory (University of Nantes) for performing the spectral analyzes on the *G. mangostana* crude extract using a 700 MHz NMR spectrometer.

REFERENCES

1. Wolfender, J.-L.; Litaudon, M.; Touboul, D.; Queiroz, E. F., Innovative omics-based approaches for prioritisation and targeted isolation of natural products – new strategies for drug discovery. *Nat. Prod. Rep.* **2019,** *36* (6), 855-868.
2. Hubert, J.; Nuzillard, J.-M.; Renault, J.-H., Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochem. Rev.* **2015,** *16* (1), 55-95.
3. Singh, U.; Baishya, B., DQF J-RES NMR: Suppressing the singlet signals for improving the J-RES spectra from complex mixtures. *J. Magn. Reson.* **2019,** *301*, 19-29.
4. Kumar, N.; Devineni, S. R.; Singh, G.; Kadirappa, A.; Dubey, S. K.; Kumar, P., Identification, isolation and characterization of potential process-related impurity and its degradation product in vildagliptin. *J. Pharm. Biomed. Anal.* **2016,** *119*, 114-121.
5. Beutler, J. A.; Alvarado, A. B.; Schaufelberger, D. E.; Andrews, P.; McCloud, T. G., Dereplication of phorbol bioactives: *Lyngbya majuscula* and *Croton cuneatus*. *J. Nat. Prod.* **1990,** *53* (4), 867-74.
6. Robinette, S. L.; Brüschweiler, R.; Schroeder, F. C.; Edison, A. S., NMR in metabolomics and natural products research: Two sides of the same coin. *Acc. Chem. Res.* **2012,** *45* (2), 288-297.
7. Gaudencio, S. P.; Pereira, F., Dereplication: racing to speed up the natural products discovery process. *Nat. Prod. Rep.* **2015,** *32* (6), 779-810.
8. Allard, P. M.; Peresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V. C.; Roussi, F.; Litaudon, M.; Wolfender, J. L., Integration of molecular networking and *in-silico* MS/MS fragmentation for natural products dereplication. *Anal. Chem.* **2016,** *88* (6), 3317-23.
9. Bakiri, A.; Hubert, J.; Reynaud, R.; Lanthony, S.; Harakat, D.; Renault, J. H.; Nuzillard, J. M., Computer-aided ^{13}C NMR chemical profiling of crude natural extracts without fractionation. *J. Nat. Prod.* **2017,** *80* (5), 1387-1396.
10. Le Pogam, P.; Schinkovitz, A.; Legouin, B.; Le Lamer, A.-C.; Boustie, J.; Richomme, P., Matrix-free UV-laser desorption ionization mass spectrometry as a versatile approach for accelerating dereplication studies on lichens. *Anal. Chem.* **2015,** *87* (20), 10421-10428.
11. Schinkovitz, A.; Boisard, S.; Freuze, I.; Osuga, J.; Mehlmer, N.; Brück, T.; Richomme, P., Matrix-free laser desorption ionization mass spectrometry as a functional tool for the analysis and differentiation of complex phenolic mixtures in propolis: a new approach to quality control. *Anal. Bioanal. Chem.* **2018,** *410* (24), 6187-6195.
12. Wang, M., et al., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnol.* **2016,** *34*, 828.
13. 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., Molecular networking as a dereplication strategy. *J. Nat. Prod.* **2013,** *76* (9), 1686-1699.
14. Fox Ramos, A. E., et al., Collected mass spectrometry data on monoterpene indole alkaloids from natural product chemistry research. *Sci. Data* **2019,** *6* (1), 15.
15. Global Natural Products Social Molecular Networking, https://gnps.ucsd.edu/ (Accessed July 2019).
16. Marti, G.; Eparvier, V.; Moretti, C.; Susplugas, S.; Prado, S.; Grellier, P.; Retailleau, P.; Gueritte, F.; Litaudon, M., Antiplasmodial benzophenones from the trunk latex of *Moronobea coccinea* (Clusiaceae). *Phytochemistry* **2009,** *70* (1), 75-85.
17. Přichystal, J.; Schug, K. A.; Lemr, K.; Novák, J.; Havlíček, V., Structural analysis of natural products. *Anal. Chem.* **2016,** *88* (21), 10338-10346.
18. Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C., High-throughput metabolomics by 1D NMR. *Angew. Chem. Int. Ed.* **2019,** *58* (4), 968-994.
19. Simmler, C.; Graham, J. G.; Chen, S.-N.; Pauli, G. F., Integrated analytical assets aid botanical authenticity and adulteration management. *Fitoterapia* **2018,** *129*, 401-414.
20. Nardella, F.; Margueritte, L.; Lamure, B.; Viéville, J. M. P.; Bourjot, M., Targeted discovery of bioactive natural products using a pharmacophoric deconvolution strategy: Proof of principle with eleganolone from *Bifurcaria bifurcata* R. Ross. *Phytochem. Lett.* **2018,** *26*, 138-142.
21. Simmler, C.; Kulakowski, D.; Lankin, D. C.; McAlpine, J. B.; Chen, S.-N.; Pauli, G. F., Holistic analysis enhances the description of metabolic complexity in dietary natural products. *Adv. Nutr.* **2016,** *7* (1), 179-189.
22. Rolin, D.; Deborde, C.; Maucourt, M.; Cabasson, C.; Fauvelle, F.; Jacob, D.; Canlet, C.; Moing, A., Chapter 1. High-resolution ^{1}H-NMR spectroscopy and beyond to explore plant metabolome. In *Advances in Botanical Research*, Rolin, D., Ed. Academic Press: 2013; Vol. 67, pp 1-66.
23. Clendinen, C. S.; Lee-McMullen, B.; Williams, C. M.; Stupp, G. S.; Vandenborne, K.; Hahn, D. A.; Walter, G. A.; Edison, A. S., ^{13}C NMR metabolomics: applications at natural abundance. *Anal. Chem.* **2014,** *86* (18), 9242-50.
24. Clendinen, C. S.; Stupp, G. S.; Ajredini, R.; Lee-McMullen, B.; Beecher, C.; Edison, A. S., An overview of methods using ^{13}C for improved compound identification in metabolomics and natural products. *Front. Plant Sci.* **2015,** *6*, 611.
25. The Human Metabolome Database, http://www.hmdb.ca/ (Accessed October 2019).
26. Biological Magnetic Resonance Data Bank, http://www.bmrb.wisc.edu/ (Accessed October 2019).
27. Hostettmann, K.; Wolfender, J.-L.; Rodriguez, S., Rapid detection and subsequent isolation of bioactive constituents of crude plant extracts. *Planta Med.* **1997,** *63* (1), 2-10.
28. Formacek, V.; Kubeczka, K. H., ^{13}C-NMR analysis of essential oils. In *Aromatic Plants: Basic and Applied Aspects*, Margaris, N.; Koedam, A.; Vokou, D., Eds. Springer Netherlands: Dordrecht, 1982; pp 177-181.
29. Tomi, F.; Bradesi, P.; Bighelli, A.; Casanova, J., Computer aided identification of individual components of essential oils using carbon ^{13} NMR spectroscopy. *Magn. Reson. Anal.* **1995,** *1*, 25–34.
30. Fromanteau, D. L. G.; Gastmans, J.-P.; Vestri, S. A.; De P. Emerenciano, V.; Borges, J. H. G., A constraints generator in structural determination by microcomputer. *Comput. Chem.* **1993,** *17* (4), 369-378.
31. Ferreira, M. J. P.; Costantin, M. B.; Sartorelli, P.; Rodrigues, G. V.; Limberger, R.; Henriques, A. T.; Kato, M. J.; Emerenciano, V. P.,

Computer-aided method for identification of components in essential oils by ^{13}C NMR spectroscopy. *Anal. Chim. Acta* **2001,** *447* (1-2), 125-134.

32. Dictionary of Natural Products, http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml (accessed March 2019).

33. Hubert, J.; Nuzillard, J. M.; Purson, S.; Hamzaoui, M.; Borie, N.; Reynaud, R.; Renault, J. H., Identification of natural metabolites in mixture: a pattern recognition strategy based on ^{13}C NMR. *Anal. Chem.* **2014,** *86* (6), 2955-62.

34. EFSA, Use of rosemary extracts as a food additive. Scientific opinion of the panel on food additives, flavourings, processing aids and materials in contact with food. *EFSA journal* **2008,** *721*, 1-29.

35. Ryu, H. W.; Curtis-Long, M. J.; Jung, S.; Jin, Y. M.; Cho, J. K.; Ryu, Y. B.; Lee, W. S.; Park, K. H., Xanthones with neuraminidase inhibitory activity from the seedcases of *Garcinia mangostana*. *Bioorg. Med. Chem.* **2010,** *18* (17), 6258-64.

36. Suksamrarn, S.; Suwannapoch, N.; Ratananukul, P.; Aroonlerk, N.; Suksamrarn, A., Xanthones from the green fruit hulls of *Garcinia mangostana*. .*J Nat. Prod.* **2002,** *65* (5), 761-3.

37. Groweiss, A.; Cardellina, J. H.; Boyd, M. R., HIV-Inhibitory prenylated xanthones and flavones from *Maclura tinctoria*. *J. Nat. Prod.* **2000,** *63* (11), 1537-9.

38. Likhitwitayawuid, K.; Phadungcharoen, T.; Krungkrai, J., Antimalarial xanthones from *Garcinia cowa*. *Planta Med.* **1998,** *64* (1), 70-2.

39. Gottlieb, H. E.; Kotlyar, V.; Nudelman, A., NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem.* **1997,** *62* (21), 7512-7515.

40. PerkinElmer ChemDraw, http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemDraw/ChemDrawProfessional/Default.aspx (accessed March 2019).

41. Softonic ChemSketch, https://chemsketch.fr.softonic.com/ (accessed March 2019).

42. CAS SciFinder, https://www.cas.org/products/scifinder (accessed March 2019).

43. ACD/Labs NMR Spectroscopy Software, https://www.acdlabs.com/products/adh/nmr/ (accessed March 2019).

44. Bruguière, A.; Derbré, S.; Coste, C.; Le Bot, M.; Siegler, B.; Leong, S. T.; Sulaiman, S. N.; Awang, K.; Richomme, P., ^{13}C-NMR dereplication of *Garcinia* extracts: Predicted chemical shifts as reliable databases. *Fitoterapia* **2018,** *131*, 59-64.

45. Please note that the program has been optimized for DBs created with ACD/Labs and hence may not work properly with a different type of DB.

46. JEOL. Natural product NMR-DB "CH-NMR-NP", https://www.j-resonance.com/en/nmrdb/ (Accessed October 2019).

47. Rossum, G. *Python reference manual*; CWI (Centre for Mathematics and Computer Science): 1995.

48. Landrum, G. An overview of the RDKit, https://www.rdkit.org/docs/Overview.html (accessed May 2019).

49. The looseness factor can be lowered when working with experimental DBs or increased if the DB was created with a less accurate prediction software.

50. On ^{13}C-NMR spectra, the intensity of the ^{13}C chemical shifts is roughly proportional to the concentration of the compound that contains these carbons. In a mixture, if the concentration of species are different enough, it seems easy to differenciate major and minor compounds. However, their intensity depends also on the type of carbons (i.e. methyle, methylene, methine, quaternary): it makes more difficult to distinguish quaternary carbons from major compounds from methyle from minor ones. All this without taking into account intra- or intermolecular equivalences. On a simple mixture of 3 compounds at various concentrations, we failed to find an automated universal statistical method to discriminate δc from individual compound. A paramagnetic relaxation agent [i.e. Cr(Acac)] was added to the mixture: it polluted the sample without any improvement.

51. International narcotics control board. Narcotic drugs, https://www.incb.org/documents/Narcotic-Drugs/Technical-Publications/2018/INCB-Narcotics_Drugs_Technical_Publication_2018.pdf (accessed March 2019).

52. Gilmore, C. D.; Allan, K. M.; Stoltz, B. M., Orthogonal synthesis of indolines and isoquinolines via aryne annulation. *J. Am. Chem. Soc.* **2008,** *130* (5), 1558-1559.

53. European Medicines Agency. *Rosmarini folium*, https://www.ema.europa.eu/en/medicines/herbal/rosmarini-folium (accessed April 2019).

54. Akoury, E., Isolation and structural elucidation of rosmarinic acid by nuclear magnetic resonance spectroscopy. *Am. Res. J. Chem.* **2017,** *1* (1), 17-23.

55. Zhang, Y.; Smuts, J. P.; Dodbiba, E.; Rangarajan, R.; Lang, J. C.; Armstrong, D. W., Degradation study of carnosic acid, carnosol, rosmarinic acid, and rosemary extract (*Rosmarinus officinalis* L.) assessed using HPLC. *J. Agric. Food Chem.* **2012,** *60* (36), 9305-9314.

56. Frankel, E. N.; Huang, S.-W.; Aeschbach, R.; Prior, E., Antioxidant activity of a rosemary extract and its constituents, carnosic acid, carnosol, and rosmarinic acid, in bulk oil and oil-in-water emulsion. *J. Agric. Food Chem.* **1996,** *44* (1), 131-135.

57. Bruguière, A.; Ray, A.-M. L.; Bréard, D.; Blon, N.; Bataillé, N.; Guillemette, T.; Simoneau, P.; Richomme, P., Identifying Natural Products (NPs) as potential UPR inhibitors for crop protection. *Planta Med.* **2016,** *82*, S1-S381.

58. Li, G.; Petiwala, S. M.; Pierce, D. R.; Nonn, L.; Johnson, J. J., Selective modulation of endoplasmic reticulum stress markers in prostate cancer cells by a standardized mangosteen fruit extract. *PLoS One* **2013,** *8* (12), e81572.

59. Xu, X. H.; Liu, Q. Y.; Li, T.; Liu, J. L.; Chen, X.; Huang, L.; Qiang, W. A.; Chen, X.; Wang, Y.; Lin, L. G.; Lu, J. J., Garcinone E induces apoptosis and inhibits migration and invasion in ovarian cancer cells. *Sci. Rep.* **2017,** *7* (1), 10718.

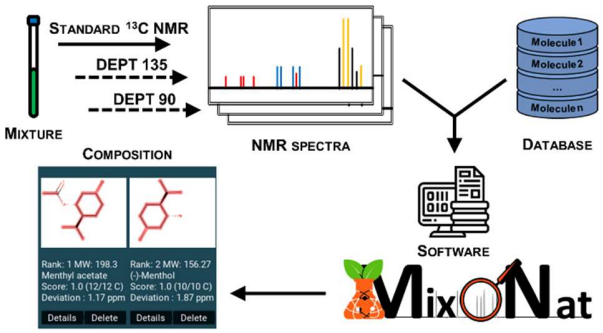
60. Li, G.; Petiwala, S. M.; Nonn, L.; Johnson, J. J., Inhibition of CHOP accentuates the apoptotic effect of alpha-mangostin from the mangosteen fruit (*Garcinia mangostana*) in 22Rv1 prostate cancer cells. *Biochem. Biophys. Res. Commun.* **2014,** *453* (1), 75-80.

61. Gudraxanthone is not soluble in CDCl3, methanol-d4 nor acetone-d6.

62. McAlpine, J. B., et al., The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. *Nat. Prod. Rep.* **2019,** *36* (1), 35-107.

63. NMReDATA initiative. http://nmredata.org/ (Accessed October 2019).

64. Pupier, M., et al., NMReDATA, a standard to report the NMR assignment and parameters of organic compounds. *Magn. Reson. Chem.* **2018,** *56* (8), 703-715.



Graphical abstract