Identification of Natural Metabolites in Mixture: A Pattern Recognition Strategy Based on C-13 NMR

ARTICLE *in* ANALYTICAL CHEMISTRY · FEBRUARY 2014

Impact Factor: 5.64 · DOI: 10.1021/ac403223f · Source: PubMed

7 AUTHORS, INCLUDING:

Jean-Marc Nuzillard
French National Centre for Scientific Research
168 PUBLICATIONS **2,048** CITATIONS
SEE PROFILE

Mahmoud Hamzaoui
Celabor, Scrl
8 PUBLICATIONS **51** CITATIONS
SEE PROFILE

Nicolas Borie
Université de Reims Champagne-Ardenne
9 PUBLICATIONS **27** CITATIONS
SEE PROFILE

Romain Reynaud
Soliance SA
12 PUBLICATIONS **45** CITATIONS
SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

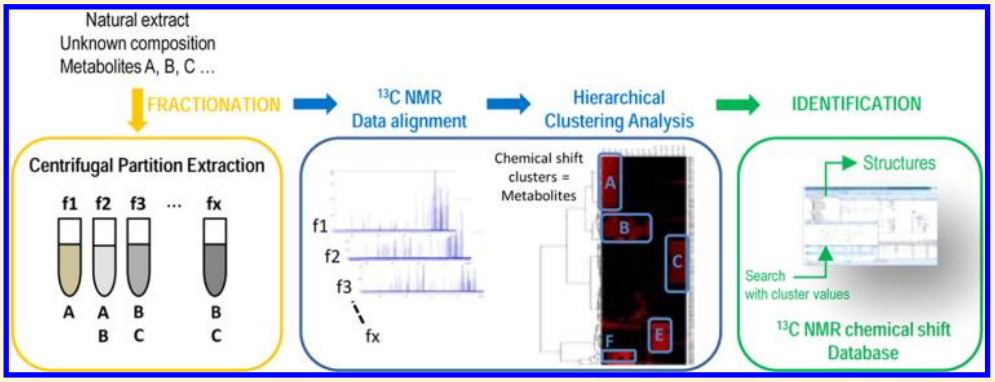
Identification of Natural Metabolites in Mixture: A Pattern Recognition Strategy Based on ^{13}C NMR

Jane Hubert,*[,†] Jean-Marc Nuzillard,[†] Sylvain Purson,[†,‡] Mahmoud Hamzaoui,[§] Nicolas Borie,[†] Romain Reynaud,[‡] and Jean-Hugues Renault[†]

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

[‡]Soliance S.A., Pomacle, France

[§]Division of Pharmacognosy and Natural Product Chemistry, Department of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece



ABSTRACT: Because of their highly complex metabolite profile, the chemical characterization of bioactive natural extracts usually requires time-consuming multistep purification procedures to achieve the structural elucidation of pure individual metabolites. The aim of the present work was to develop a dereplication strategy for the identification of natural metabolites directly within mixtures. Exploiting the polarity range of metabolites, the principle was to rapidly fractionate a multigram quantity of a crude extract by centrifugal partition extraction (CPE). The obtained fractions of simplified chemical composition were subsequently analyzed by ^{13}C NMR. After automatic collection and alignment of ^{13}C signals across spectra, hierarchical clustering analysis (HCA) was performed for pattern recognition. As a result, strong correlations between ^{13}C signals of a single structure within the mixtures of the fraction series were visualized as chemical shift clusters. Each cluster was finally assigned to a molecular structure with the help of a locally built ^{13}C NMR chemical shift database. The proof of principle of this strategy was achieved on a simple model mixture of commercially available plant secondary metabolites and then applied to a bark extract of the African tree *Anogeissus leiocarpus* Guill. & Perr. (Combretaceae). Starting from 5 g of this genuine extract, the fraction series was generated by CPE in only 95 min. ^{13}C NMR analyses of all fractions followed by pattern recognition of ^{13}C chemical shifts resulted in the unambiguous identification of seven major compounds, namely, sericoside, trachelosperogenin E, ellagic acid, an epimer mixture of (+)-gallocatechin and (−)-epigallocatechin, 3,3′-di-*O*-methylellagic acid 4′-*O*-xylopyranoside, and 3,4,3′-tri-*O*-methylflavellagic acid 4′-*O*-glucopyranoside.

Natural extracts from plants and microorganisms still constitute invaluable sources of biologically active metabolites for the development of drugs or cosmetics.[1−3] The major challenge in the search for such metabolites arises from the extreme complexity of plant extracts or culture media which contain a wide diversity of molecules with distinct physical and chemical properties. At present, even if modern analytical and purification techniques are routinely available in most laboratories, a considerable work taking several days even several weeks or years is still necessary to isolate and elucidate individual metabolite structures from crude natural extracts. In some cases, time-consuming multistep purification procedures are unavoidable, for instance when the objective is to elucidate the complex molecular structure of a novel compound. In numerous other cases, the systematic purification of individual constituents results in a considerable waste of time. Bioactivity-guided fractionation procedures have been developed to focus only on the fractions or metabolites with a defined biological activity. However, often such approaches are applied to finally rediscover already known compounds.

In view of these observations, new methods enabling the identification of natural metabolites directly within mixtures would be very useful.

Received: October 7, 2013
Accepted: February 20, 2014
Published: February 20, 2014

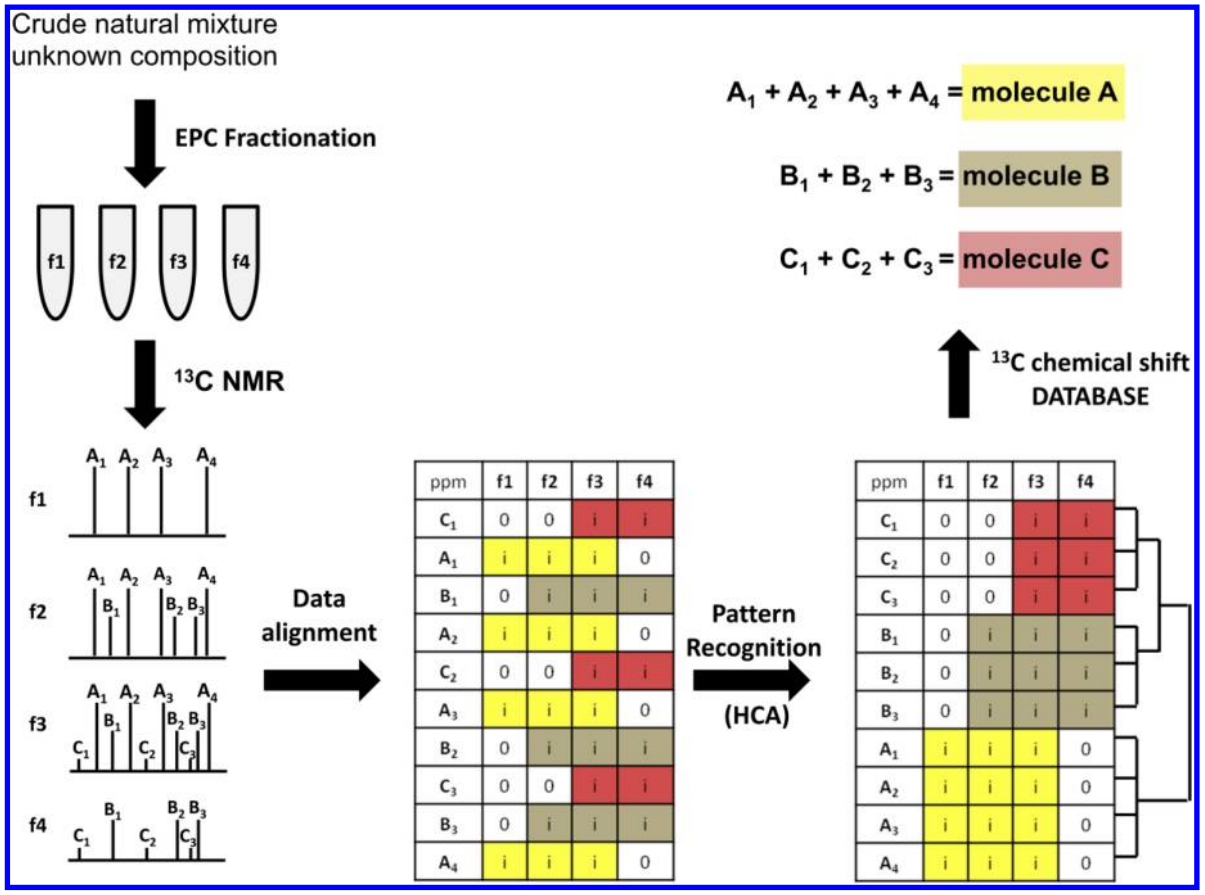


Figure 1. Global dereplication strategy workflow.

Over the past few years, powerful "-omics" analytical platforms combining state-of-the-art spectroscopic instruments, bioinformatics tools, and metabolite databases have been developed to qualitatively or quantitatively analyze the whole set of metabolites present in biological systems.[4,5] Metabolomics has already demonstrated strong potentialities for the metabolic profiling or fingerprinting of plant extracts, mainly in an attempt to improve the quality control of herbal medicines, to evaluate the chemical differences between plant species, or to assess correlations between bioactivity and composition.[4−7] By combining multivariate statistical treatments and metabolic network modeling with high-resolution analytical systems, metabolomic studies include a set of powerful tools that can also be used to determine the overall composition of complex natural mixtures.[8,9]

In the present work, a dereplication strategy was developed to determine the global composition of natural extracts without purifying each component individually. This approach was designed specifically for an early application in cosmetics where the chemical characterization of natural active ingredients is subjected to increasing regulatory constraints. In this context, an efficient method enabling the identification of major metabolites in crude extracts without huge resource implications would make a real difference.

The procedure is based on a particular metabolomics-derived workflow, as described in Figure 1. The first goal consists in producing a series of successive fractions from a crude natural sample by centrifugal partition extraction, a separation technique that can selectively produce simplified mixtures of natural compounds from crude extracts on a multigram scale and in a short time.[10,11] In a second step, the whole set of metabolites contained in the CPE-generated fractions are analyzed by ^{13}C NMR. Because a given compound will be necessarily detected as an entire and well-defined set of ^{13}C resonances, hierarchical clustering analysis (HCA) is performed on aligned ^{13}C chemical shifts across spectra of the fraction series in order to highlight the statistical correlations within the whole data set produced. The idea is to directly visualize individual metabolites in the two-dimensional map resulting from HCA as "chemical shift clusters" within the fraction series. These clusters are assigned to a molecular structure using a locally built ^{13}C NMR chemical shift database of natural metabolites.

In this paper, the proof of principle of this new strategy is first demonstrated on a model mixture of pure molecules representative of the main chemical classes of plant secondary metabolites (experiment 1). Then, the whole procedure is applied on an authentic crude extract (experiment 2) obtained from the bark of the African tree *Anogeissus leiocarpus* Guill. & Perr. (Combretaceae).

EXPERIMENTAL SECTION

Chemicals, Reagents, And Plant Material. Methyl-*tert*-butyl ether (MTBE), ethyl acetate (EtOAc), *n*-heptane, acetonitrile (CH_3CN), and sodium hydroxide were purchased from Carlo Erba Reactifs SDS (Val de Reuil, France). Quercetin, (+)-catechin, glycyrrhetic acid, *p*-hydroxycinnamic acid, and rutoside were purchased from Acros Organics. Deionized water was used to prepare all aqueous solutions. The crude bark extract (ethanol 99%) of *A. leiocarpus* Guill. & Perr. (Combretaceae) was provided by Soliance (Pomacle, France).

dx.doi.org/10.1021/ac403223f | *Anal. Chem.* 2014, 86, 2955−2962

Centrifugal Partition Extraction. The fractionation process was developed on a lab-scale CPE column of 303.5 mL capacity (FCPE300, Kromaton Technology, Angers, France) containing 7 circular partition disks and engraved with a total of 231 oval partition twin cells (≈ 1 mL per twin cell). The liquid phases were pumped by a KNAUER Preparative 1800 V7115 pump (Berlin, Germany). The system was coupled to a UVD 170S detector set at 210, 254, 280, and 366 nm (Dionex, Sunnivale, CA). Fractions were collected by a Pharmacia Superfrac collector (Uppsala, Sweden). Two experiments were performed as described in a previous study,[10] starting from either 1 g of a synthetic mixture of (+)-catechin, quercetin, glycyrrhetic acid, *p*-hydroxycinnamic acid, and rutoside (200 mg each) or 5 g of the crude bark extract of *A. leiocarpus*. Briefly, a three-phase solvent system composed of *n*-heptane (700 mL), M*t*BE (700 mL), CH_3CN (700 mL), and water (700 mL) was thoroughly equilibrated in a separatory funnel. After separation of the *n*-heptane rich upper phase, one equivalent volume of M*t*BE (700 mL) was added to the mixture of middle and lower phases in order to slightly reduce the polarity of the middle phase. After decantation, the final middle and lower phases were separated. For each CPE run, samples were dissolved in a mixture of lower/middle/upper phases (45:10:5 v/v) to ensure that the three liquid phases were also in equilibrium in the sample solution. The lower phase was used as a stationary phase and maintained inside the column by application of a constant centrifugal force field. The column was filled at 200 rpm, and the rotation speed was then increased to 1000 rpm. The sample solution was loaded into the column by progressively pumping the less polar upper phase through the stationary phase from 0 to 20 mL/min in 3 min in the ascending mode. The flow rate of the mobile phase was then maintained at 20 mL/min. The *n*-heptane rich upper phase was pumped for 50 min to ensure the elution of all hydrophobic compounds. The moderately polar middle phase was then pumped for 33 min to elute compounds of medium hydrophobicity. Finally, the most hydrophilic compounds retained inside the column were recovered by pumping the aqueous phase in the descending mode at 20 mL/min. In total, 15 fractions of 120 mL and 19 fractions of 100 mL were collected over the whole experiments 1 and 2, respectively.

^{13}C NMR Analyses and Data Processing. All samples were analyzed using the same acquisition and processing parameters. Fractions were dried under vacuum, and about 20 mg of each (for both experiments 1 and 2) were dissolved in 500 μL DMSO-*d6*. NMR analyses were performed at 298 K on a Bruker Avance AVIII-600 spectrometer (Karlsruhe, Germany) equipped with a cryoprobe optimized for ^{1}H detection and with cooled ^{1}H, ^{13}C, and two-dimensional (2D) coils and preamplifiers. The ^{13}C S/N calculated from the standard ASTM test sample (60% C_6D_6, 40% dioxane) is 1271:1. ^{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 coadded 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. The spectra were manually phased and baseline-corrected using TOPSPIN 2.1 (Bruker) and calibrated on the central resonance (δ 39.80 ppm) of DMSO-*d6*. The S/N of all spectra was determined with the standard Bruker calculation method. Noise regions were selected from 205 to 225 ppm in all spectra of the fraction series, and signal regions were selected differently for each spectrum as the 10 ppm spectral width centered on the most intense signal (DMSO-*d6* signal apart). Since the ^{13}C signals of quaternary carbon atoms generally have an intensity 1/3 lower than that of the other carbon types for a single molecule, we arbitrarily considered that minor compounds within the mixtures correspond to all ^{13}C signals lower than 1/3 of the major signal intensity. In this way, the S/N ratio of minor compounds for each fraction of experiment 2 (genuine extract) was also calculated by selecting a 10 ppm spectral region where the most intense signal was lower than 1/3 of the major signal of the spectra. A minimum intensity threshold of 0.05 was then used to automatically collect all positive ^{13}C NMR signals, while avoiding potential noise artifacts. Each peak list was converted into a text file. Absolute intensities of the collected peaks in the fraction series were then aligned by using an in-house algorithm written in the python language. The principle was to divide the ^{13}C spectral width (from 0 to 200 ppm) into regular bins of 0.2 ppm and to place the absolute intensity of each ^{13}C peak into the corresponding bin. The bins for which no signal was detected in any fraction were removed from the bin list. The resulting table was imported into PermutMatrix, version 1.9.3 (LIRMM, Montpellier, France) for clustering analysis. Hierarchical clustering analysis was directly applied on raw peak intensity values. The classification was performed on the rows only (i.e., on the chemical shift bins). The Euclidian distance was used to measure the proximity between samples, and the Ward's method was performed to agglomerate the data. The resulting ^{13}C chemical shift clusters were visualized as dendrograms on a 2D map. The higher the intensity of ^{13}C NMR peaks, the brighter the color was in the map (see Figures 2, 3, and 5).

Database of *A. leiocarpus* Metabolite Structures. A literature survey was performed to obtain names and/or structures of the metabolites already described in *A. leiocarpus*. In total, 28 metabolites were found. Each metabolite data record was then stored in a ^{13}C chemical shift database by means of ACD/NMR Workbook Suite 2012 (ACD/Laboratories, Ontario, Canada). Each structure was drawn with ChemSketch, and the chemical shifts were assigned to the corresponding carbon positions. When ^{13}C chemical shifts were not available in the literature, a predicted spectrum was calculated with ACD/Laboratories CNMR Predictor, and the predicted ^{13}C chemical shifts were supplied to the database. For metabolite identification, each ^{13}C chemical shift cluster obtained from HCA was submitted to the structure search engine of the database management software. A ^{13}C NMR chemical shift error interval of ±1 ppm was used, and the minimum number of ^{13}C query shift to match was set at ≈80% of the number of chemical shifts in the given cluster.

RESULTS AND DISCUSSION

General Strategy. The aim of the present work was to develop an efficient dereplication strategy for the characterization of natural metabolite mixtures without purifying each individual compound. In the first step, a series of simplified fractions was produced from the initial mixture by CPE. CPE is a recent solid support-free liquid−liquid separation technique directly inspired from centrifugal partition chromatography. It involves the distribution and the transfer of solutes between at least two immiscible liquid phases according to their partition coefficient. The lower number and higher volume of partition cells in CPE as compared to CPC make it possible to pump the mobile phase at higher flow rates ranging from 10 to 50 mL/min and to handle samples on a multigram scale.[12,13] When applied in a sequential elution mode with a three-phase solvent system,

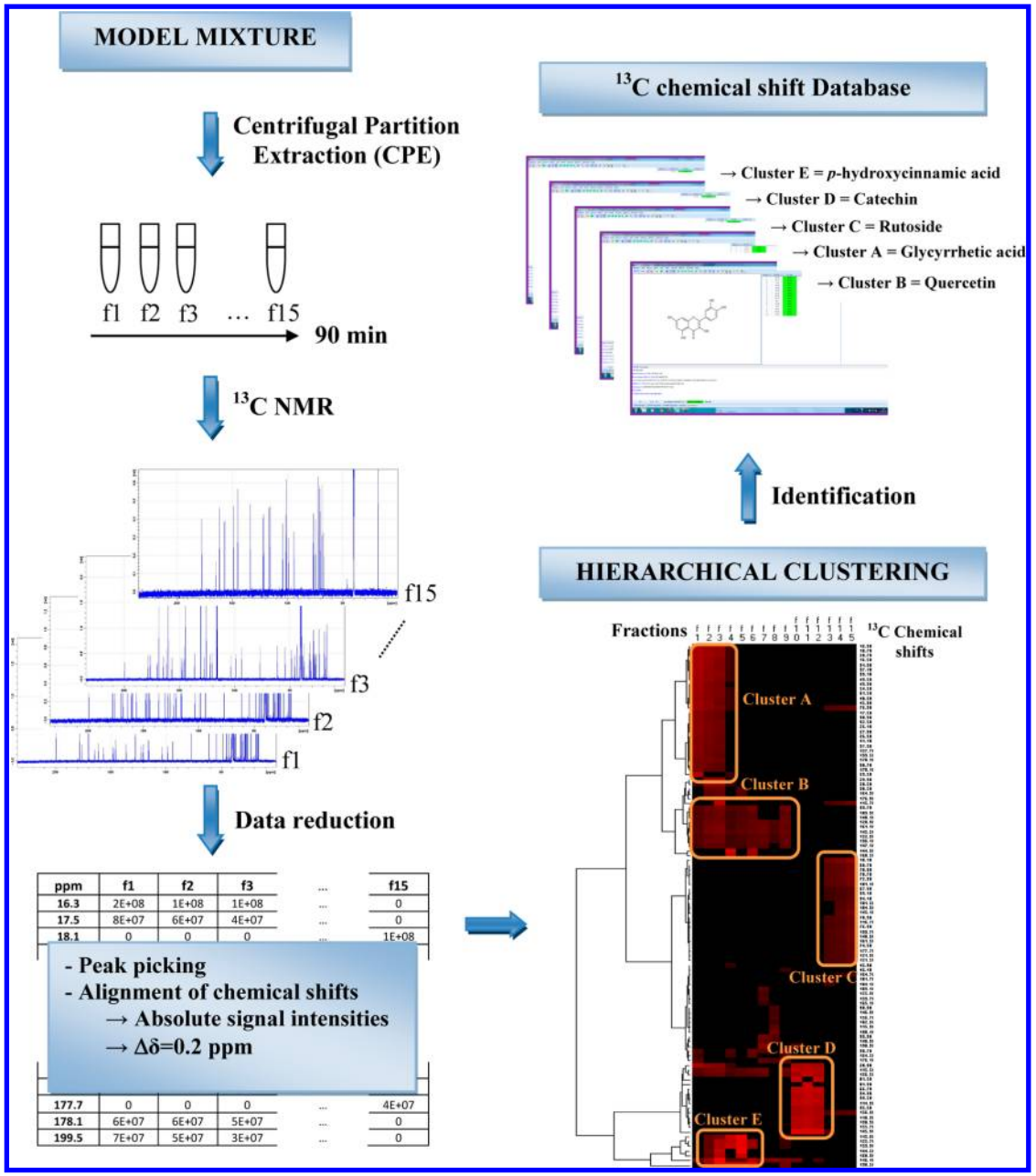


Figure 2. Dereplication strategy applied on a synthetic standard mixture.

CPE has recently proven to be an efficient and cost-effective alternative for the rapid and selective fractionation of natural extracts.[10] We thus have prioritized this technique for the fractionation step.

The second step is the detection of the whole set of metabolites present in the successive CPE-generated fractions. ^{1}H NMR currently remains the main profiling technique for plant extracts due to its high sensitivity, and many applications have been extensively discussed in recent reviews.[14−18] However, even when hyphenated to a chromatographic separation device, ^{1}H NMR signals of compounds in complex mixtures are poorly dispersed, resulting in strong overlaps in most spectral regions. Although natural organic compounds also contain carbon atoms, the use of ^{13}C NMR for the profiling of plant extracts remains largely unexplored, mainly because the low abundance of the ^{13}C isotope (1%) and its low gyromagnetic ratio (25% of the one of ^{1}H) considerably reduce its detection sensitivity. Today, high magnetic field NMR spectrometers and cryogenic probes enable the acquisition of ^{13}C spectra with high-resolution and good sensitivity in a short time. ^{13}C NMR provides strong advantages for the analysis of complex mixtures. Carbon atoms represent a significant part of all organic molecules, and each ^{13}C position in a structure corresponds to a single resonance on a broadband ^{1}H decoupled ^{13}C NMR spectra (degeneracy by symmetry excluded). In addition, the ^{13}C NMR spectral width is significantly higher than that of ^{1}H (220 ppm for ^{13}C and 12 ppm for ^{1}H), which significantly reduces signal overlaps. For these reasons, our approach was based on the use of ^{13}C NMR. As a given compound will be detected as an entire and well-defined set of ^{13}C resonances, the statistical correlations within this set of resonances will be high if the structure of this compound is detected several times in the fraction series.

The third step consists in the automatic peak picking and alignment of ^{13}C NMR signals. From the ^{13}C peak lists of the

dx.doi.org/10.1021/ac403223f | Anal. Chem. 2014, 86, 2955−2962

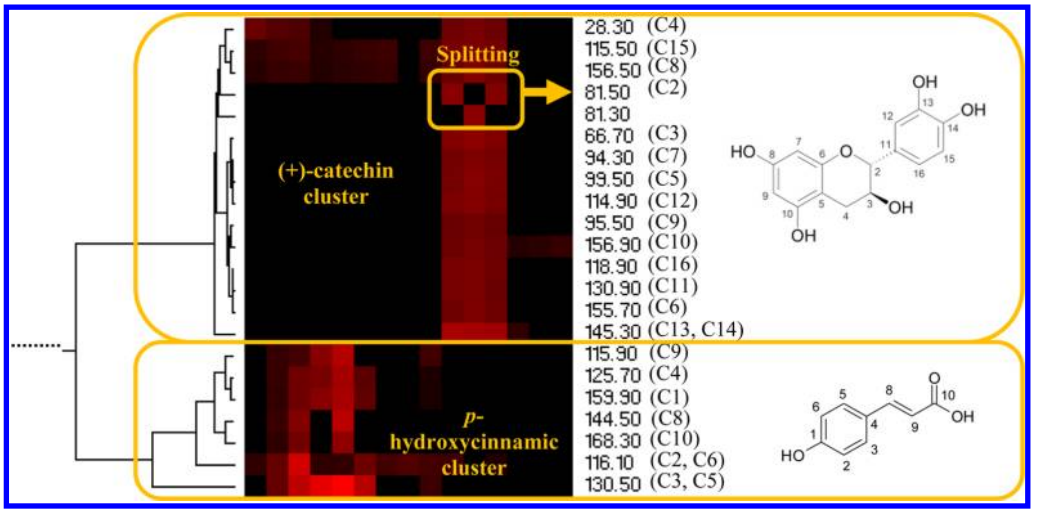


Figure 3. Extract of the HCA obtained after applying the whole dereplication strategy on a standard mixture. ^{13}C NMR chemical shifts belonging to the same structure are strongly correlated. An exemple is provided here for (+)-catechin and *p*-hydroxycinnamic carbon skeletons.

different fractions, a 2D matrix was built. Each column corresponded to a fraction of the series, and each row corresponded to a chemical shift windows containing at least one ^{13}C signal in at least one fraction (Figure 1). In a fourth step, the table content was submitted to hierarchical clustering analysis. HCA is an unsupervised pattern recognition method enabling the detection of chemical shift similarities within the fraction series. As a result, statistically correlated "^{13}C signal groups" belonging to the same molecular structure were visualized on a dendrogram in the form of clusters.

Finally, these chemical shift clusters were assigned to molecular structures with the help of a locally built ^{13}C NMR chemical shift database containing the structures and predicted ^{13}C NMR chemical shifts of all the compounds already described in the literature for the natural extract under examination.

Proof of Principle on a Model Mixture. The global procedure was first tested on a model mixture made of glycyrrhetic acid, *p*-hydroxycinnamic acid, (+)-catechin, quercetin, and rutoside (Figure 2). These substances were selected for four main reasons: (i) all are secondary metabolites frequently identified in crude extracts of plant origin; (ii) they belong to various chemical classes (triterpenes, phenolic acids, flavonoid aglycones, and heterosides; (iii) they exhibit a large polarity range from the less polar triterpenic structure of glycyrrhetic acid to the highly polar structure of rutoside, and (iv) some exhibit real similarities in their molecular backbone. For instance rutoside is the diglycosylated derivative of quercetin; phenolic acids and flavonoids contain similar hydroxylated aromatic cores.

This model mixture was fractionated by CPE in a single run of 90 min by using a three-phase solvent system in a sequential elution mode.[10] As a result, 15 adjacent fractions (f1, f2..., f15) containing simplified mixtures of metabolites were obtained in a decreasing order of polarity.

The fraction series produced by CPE was then analyzed by ^{13}C NMR. The average S/N ratio of the spectra was 120 ± 70, indicating a good resolution. Automatic peak picking and alignment of the ^{13}C signals across spectra resulted in a table with 15 columns (one per fraction) and 110 rows (one per chemical shift window that contains at least one ^{13}C signal in at least one fraction). This matrix was submitted to hierarchical clustering analysis (HCA) on the rows. The statistical correlations between ^{13}C NMR chemical shifts within the fraction set were easily visualized in dendrograms. As presented in Figure 2, five well-defined clusters were intensely colored in red on the 2D correlation matrix. These clusters corresponded to the respective carbon skeletons of glycyrrhetic acid (cluster A), quercetin (cluster B), rutoside (cluster C), (+)-catechin (cluster D), and *p*-hydroxycinnamic acid (cluster E). As a result, 28 of 30 ^{13}C signals for glycyrrhetic acid, 7 of 7 signals for *p*-hydroxycinnamic acid, 14 of 14 signals for (+)-catechin, 13 of 15 signals for quercetin, and 23 of 27 signals for rutoside were correctly clustered. This example illustrates the unambiguous identification of quercetin resonances in the spectrum, even in the presence of rutoside, which is its glycoside derivative. The small number of classification errors mainly resulted from two particular situations. First, data alignment might have failed even by using a chemical shift window of 0.2 ppm, resulting in the splitting of a single signal into two adjacent sections. For instance, the signal C2 of (+)-catechin was placed in two consecutive bins at 81.3 and 81.5 ppm (Figure 3). Second, very close ^{13}C chemical shift values can fall into the same bin, resulting in a loss of information. Different bin widths were tested, but as illustrated in Figure 4, we found that 0.2 ppm was the best compromise to avoid splitting of ^{13}C chemical shifts of a single structure into two adjacent sections from one spectrum to another or conversely to avoid the merging into the same bin of close ^{13}C chemical shifts belonging to different molecules. These data confirm one more time that data alignment remains highly challenging in NMR-based metabolite profiling methods. In addition to solvent effects, many NMR chemical shifts are also dependent on sample composition and concentration. For instance, aromatic compounds like *p*-hydroxycinnamic acid exhibited chemical shift deviations up to ≈0.1 ppm, depending on the presence of other constituents in the fractions under examination (Figure 5). Of course this "matrix effect" might be accentuated in our case, since each fraction exhibits a different chemical profile. It will be essential to investigate this aspect on the real extract.

To conclude on this part, the results obtained after applying the global dereplication strategy on a model mixture are satisfactory as 85 of the 93 automatically collected signals were assigned to the right carbon skeleton.

Proof-of-Principle on an Authentic Natural Extract. *A. leiocarpus* Guill. & Perr. (Combretaceae) is a sub-Saharian African tree whose bark and leaves exhibit antimalarial and trypanocidal activities.[19,20] The ethanol extract of the bark, highly rich in tannins (≈ 80% w/w), contains also a diversity of

dx.doi.org/10.1021/ac403223f | *Anal. Chem.* 2014, 86, 2955−2962

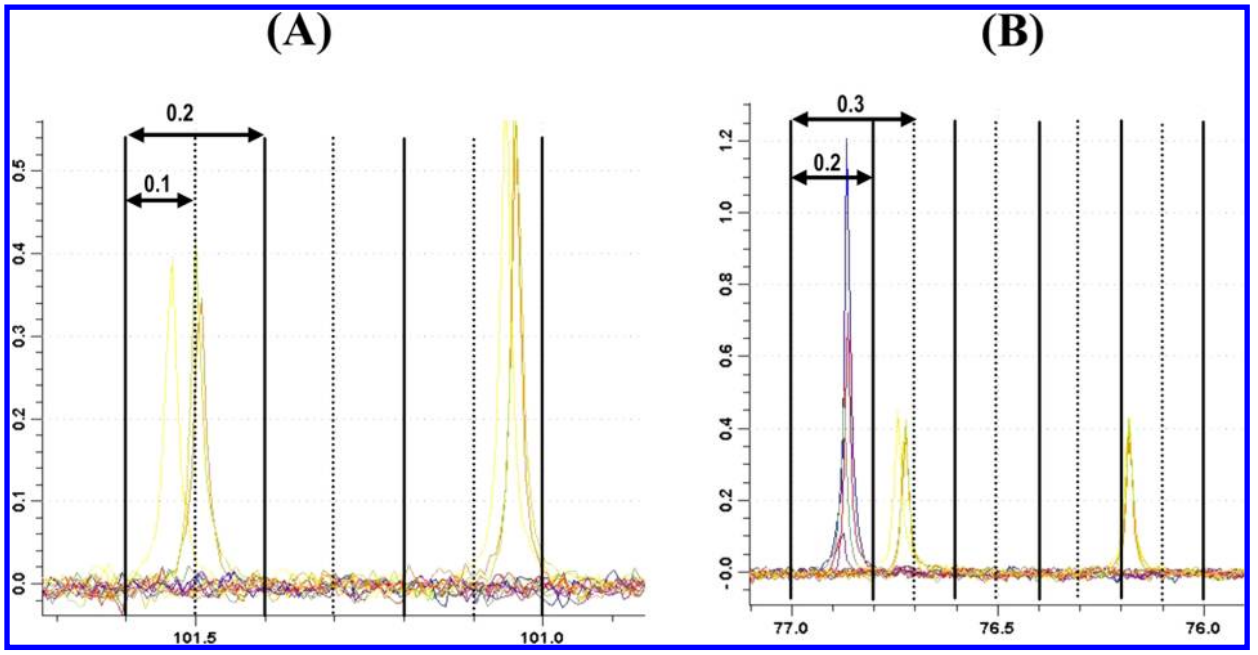


Figure 4. ^{13}C NMR spectra overlay of 15 fractions recovered from the CPE fractionation of a standard mixture. To define the optimum chemical shift window for data alignment, different cases must be considered: (A) If $\Delta\delta < 0.2$ ppm, a unique ^{13}C signal can split into different windows from one fraction to another. (B) If $\Delta\delta > 0.2$ ppm, two different ^{13}C signals can merge in a same window. The chemical shift window for ^{13}C NMR data alignment was therefore set at 0.2 ppm.

moderately polar secondary metabolites derived from ellagic acid, flavonoids, triterpenes, and saponins. As the background of this work was to characterize these other potentially interesting compounds, we have applied a CPE method allowing the total removal of tannins while fractionating the less polar constituents.[10] Each recovered fraction exhibited a chemical diversity, including different molecular classes and covering a large polarity range. In a single run of 95 min, 5 g of the crude bark extract were fractionated by CPE, resulting in a series of 20 simplified fractions of decreasing polarity. The most hydrophobic compounds were eluted by the *n*-heptane-rich mobile phase (from t_0 to 58 min), and the moderately polar compounds were eluted with the M*t*BE-rich mobile phase (from 59 to 92 min). The water-soluble tannins, fully retained inside the aqueous stationary phase, were all recovered in the last fraction, which was not taken into account for the rest of the study. The mass of each recovered fraction was largely sufficient to achieve ^{13}C NMR analyses (with 20 mg each), while leaving material for further biological evaluation. The average S/N ratio of the whole spectra within the fraction series was 46 ± 21 and that of minor compounds was 4.8 ± 1.2, depending on sample composition. Running HCA on the aligned ^{13}C NMR data, exactly as described in the previous section, resulted in the formation of several well-defined chemical shift clusters corresponding to the main structures (besides the water-soluble tannins) of the crude bark extract of *A. leiocarpus*. The HCA representation of the main clusters is given in Figure 5. The 19 columns correspond to the successive CPE fractions analyzed by NMR and the 241 rows correspond to the chemical shift intervals for which at least one ^{13}C NMR signal was collected in at least one fraction. Cluster A corresponds to a major cluster of 26 ^{13}C chemical shifts. After entering these chemical shifts into the database, a single structure was proposed, corresponding to sericoside. In order to evaluate the confidence we could have in this finding, further 2D HSQC, HMBC, and COSY NMR analyses were performed on fraction f12, where the intensities of cluster A chemical shifts were predominant. The identity of sericoside was confirmed. By applying exactly the same approach, five other intense ^{13}C chemical shift clusters B, C, D, E, and F were identified as trachelosperogenin E, ellagic acid, an epimer mixture of (+)-gallocatechin and (−)-epigallocatechin, 3,3′-di-*O*-methylellagic acid 4′-*O*-xylopyranoside, and 3,4,3′-tri-*O*-methylflavellagic acid 4′-*O*-glucopyranoside, respectively. Each molecular structure proposed by the database was the right structure as confirmed by additional 2D NMR analyses.

Several key aspects regarding the global procedure have to be discussed. First, we observed that the performance of data alignment was significantly dependent on the molecular class. As illustrated in Figure 6, the ^{13}C chemical shifts of triterpene derivatives were for instance much better aligned than the chemical shifts of aromatic structures, including flavonoids or ellagic acid derivatives. As an example, the chemical shifts of the carbon C3/C3′ (140.1 ppm), C5/C5′ (110.7 ppm), and C7/C7′ (159.5 ppm) of ellagic acid were split into adjacent sections at 140.3, 110.5, and 159.7 ppm, respectively (region C″ on the dendogram instead of region C). These sample-to-sample chemical shift variations were already observed with the model mixture and can be explained by $\pi-\pi$ stacking interactions between aromatic rings.

Second, if the fractionation method applied on the initial crude extract is efficient (i.e., if structurally related compounds are not recovered in the same fractions), even very close molecular structures with several common identical ^{13}C chemical shifts can be distinguished. This is illustrated for instance in Figure 6 between the clusters of sericoside (cluster A) and trachelosperogenin E (cluster B), having 12 nearly identical ^{13}C chemical shifts. Both clusters were clearly separated from each other because, even if mixed with compounds from other chemical classes, their two related structures were detected in different fractions. This was also observed for 3,3′-di-*O*-methylellagic acid 4′-*O*-xylopyranoside (cluster E) and 3,4,3′-tri-*O*-methylflavellagic acid 4′-*O*-glucopyranoside (cluster F) sharing 4 identical

dx.doi.org/10.1021/ac403223f | *Anal. Chem.* 2014, 86, 2955−2962

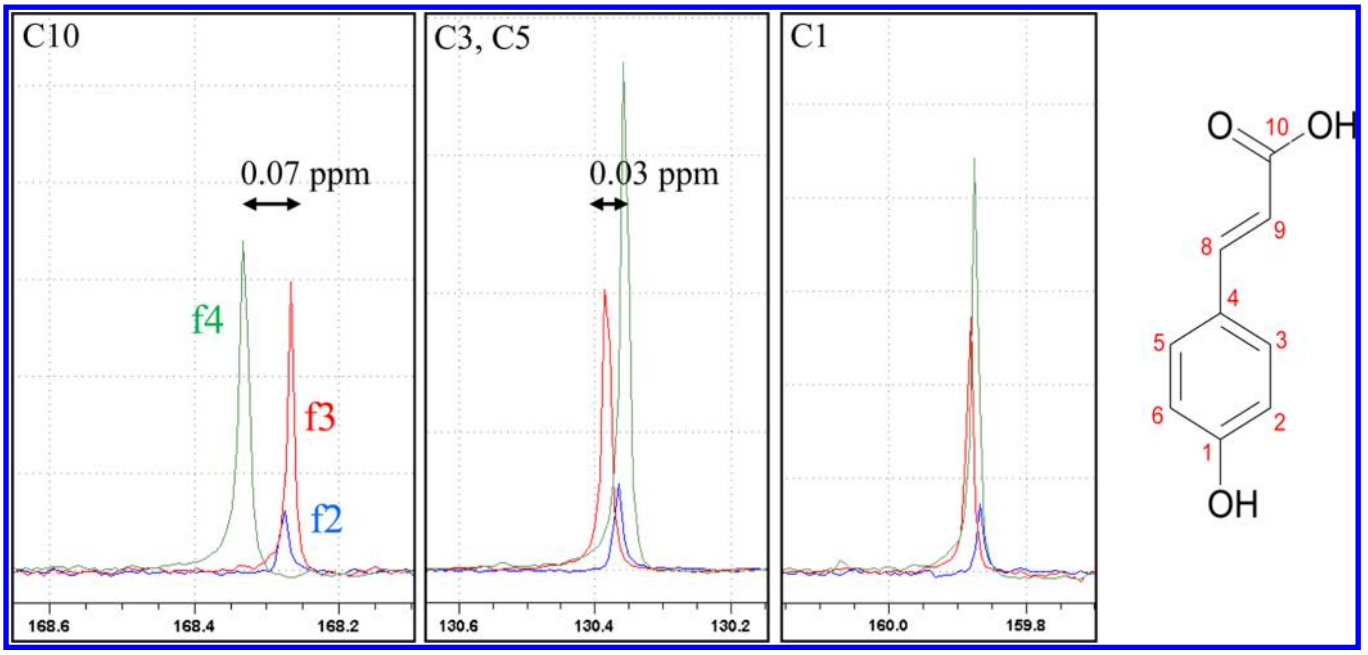


Figure 5. ^{13}C NMR spectra overlay of fractions F2, F3, and F4 (experiment 1). Significant chemical shift deviations can result from variations in the chemical composition (matrix effect) of fractions. As an example, the chemical shifts of *p*-hydroxycinnamic acid can deviate up to 0.07 ppm between the spectra of three successive fractions, depending on the carbon position and on the presence of other constituents: glycyrrhetic acid for instance was also present in F2 and F3 but absent from F4 and quercetin was present in the three fractions but not at the same concentration.

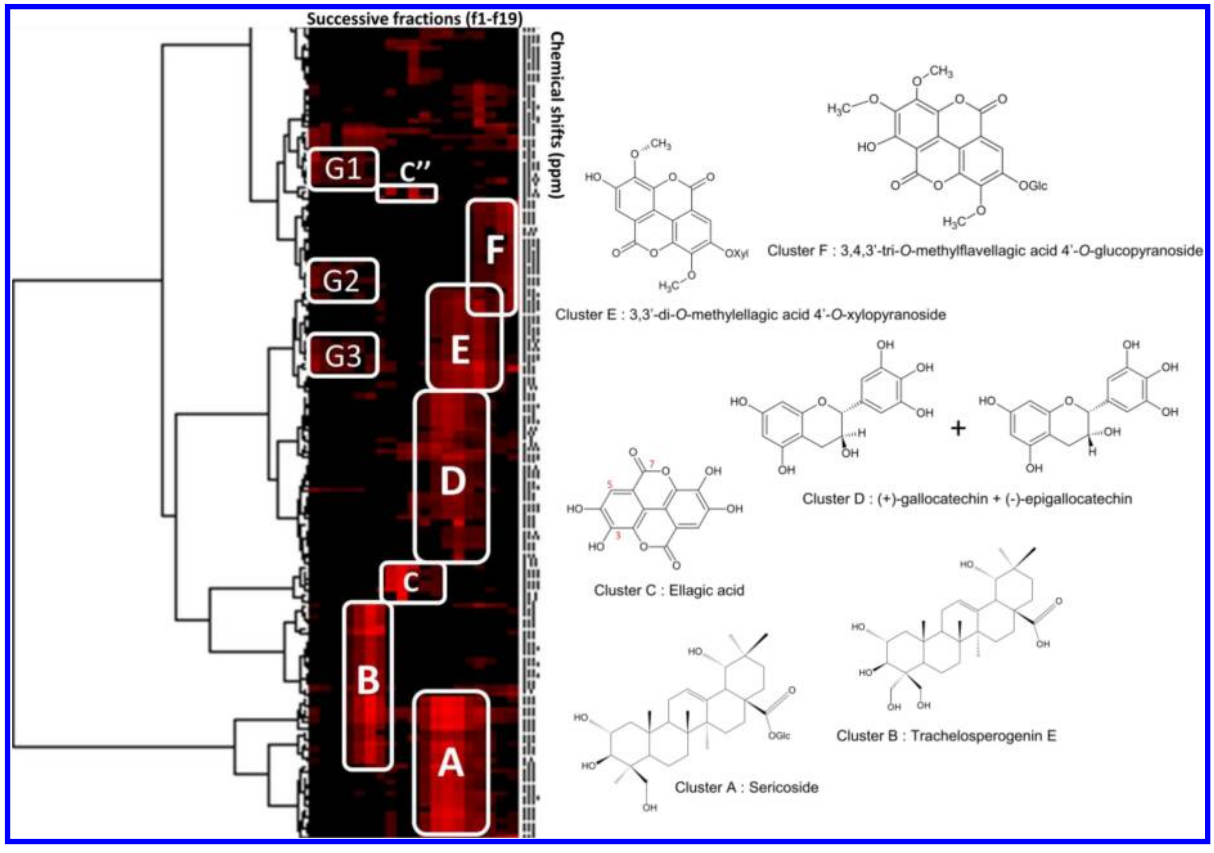


Figure 6. Chemical shift clusters obtained by applying the dereplication strategy on the crude bark extract of *A. leiocarpus*.

chemical shifts. Both molecular structures only differ from each other by the number of methyl groups and nature of the sugar on the ellagic acid skeleton; nevertheless, their discrimination by HCA was possible because they were not recovered in the same fractions after CPE fractionation. It should also be mentioned that different molecular structures detected in a same fraction will be especially discriminated as they belong to different molecular classes. This results directly from the low probability of collecting

dx.doi.org/10.1021/ac403223f | *Anal. Chem.* 2014, 86, 2955−2962

similar chemical shift values for molecules exhibiting different carbon skeletons. As an example, the saponin skeleton of sericoside (cluster A) was strongly discriminated from flavonoids (cluster D) and from the ellagic acid-derived skeletons (clusters E and F) even when detected together in the same fractions ranging from f12 to f17. Small molecules ($n_{carbons} < 10$) can also be clearly recognized if their concentration is higher than that of the other constituents in the same fraction. For instance, the ^{13}C chemical shifts of ellagic acid ($n_{carbons} = 7$), largely predominant in fractions f8, f9, and f10, were strongly correlated in the HCA dendogram (cluster C). Third, since the correlation trees obtained from HCA are in part dependent on absolute signal intensities, the clusters corresponding to minor metabolites were hierarchically organized after the most abundant metabolites. As a result, the clusters of minor metabolites may be divided into several "sub-clusters", as observed in Figure 6 for G1, G2, and G3, which correspond to 3,4,3′-tri-*O*-methylflavellagic acid. We thus should bear in mind that minor metabolites can potentially be identified by using the values of these sub-clusters when consulting the ^{13}C NMR chemical shift database. Of course the present strategy will be refined through future research efforts, especially regarding the optimization of ^{13}C chemical shift alignment and identification of minor metabolites.

CONCLUSION

The combination of an efficient multigram-scale fractionation method with ^{13}C NMR analyses and HCA for pattern recognition of ^{13}C signals across spectra of the fraction series enabled the direct identification of the main metabolites present in a crude natural extract without purification. The proof-of-principle of this approach was successfully achieved on a simple synthetic model mixture. Then, starting from 5 g of a crude bark extract of *A. leiocarpus*, application of the whole procedure resulted in the simultaneous identification of seven secondary metabolites from different chemical classes. In the future, research efforts must be sustained to refine the limiting steps, especially ^{13}C chemical shift alignment, minor compound identification, and ^{13}C chemical shift database implementation.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Jane.hubert@univ-reims.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the Soliance society, the CNRS, the Ministry of Higher Education and Research, and the "Champagne-Ardenne DRRT" for financial support. 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. We also thank Pr. Catherine Lavaud from ICMR for stimulating scientific discussions.

REFERENCES

(1) Imhoff, J. F.; Labes, A.; Wiese. *Biotechnol. Adv.* **2011**, *29*, 468−482.
(2) Harvey, A. L. *Drug Discovery Today* **2008**, *13*, 894−901.
(3) Mishra, B. B.; Tiwari, V. K. *Eur. J. Med. Chem.* **2011**, *46*, 4769−4807.
(4) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155−171.
(5) Zhang, A. H.; Sun, H.; Wang, P.; Han, Y.; Wang, X. J. *Analyst* **2012**, *137*, 293−300.
(6) Sheridan, H.; Krenn, L.; Jiang, R. W.; Sutherland, I.; Ignatova, S.; Marmann, A.; Liang, X. M.; Sendker, J. *J. Ethnopharmacol.* **2012**, *140*, 482−491.
(7) Yuliana, N. D.; Khatib, A.; Verpoorte, R.; Choi, Y. H. *Anal. Chem.* **2011**, *83*, 6902−6906.
(8) Rochfort, S. *J. Nat. Prod.* **2005**, *68*, 1813−1820.
(9) Ngo, L. T.; Okogun, J. I.; Folk, W. R. *Nat. Prod. Rep.* **2013**, *30*, 584−592.
(10) Hamzaoui, M.; Renault, J. H.; Nuzillard, J. M.; Reynaud, R.; Hubert, J. *Phytochem. Anal.* **2013**, *24*, 367−373.
(11) Hamzaoui, M.; Renault, J. H.; Reynaud, R.; Hubert, J. *J. Chromatogr., B* **2013**, *937*, 7−12.
(12) Hamzaoui, M.; Hubert, J.; Hadj-Salem, J.; Richard, B.; Harakat, D.; Marchal, L.; Foucault, A.; Lavaud, C.; Renault, J. H. *J. Chromatogr., A* **2011**, *1218*, 5254−5262.
(13) Hamzaoui, M.; Hubert, J.; Reynaud, R.; Marchal, L.; Foucault, A.; Renault, J. H. *J. Chromatogr., A* **2012**, *1247*, 18−25.
(14) Schripsema, J. *Phytochem. Anal.* **2010**, *21*, 14−21.
(15) Kim, H. K.; Choi, Y. H.; Verpoorte, R. *Trends Biotechnol.* **2011**, *29*, 267−275.
(16) Heyman, H. M.; Meyer, J. J. M. *S. Afr. J. Bot.* **2012**, *82*, 21−32.
(17) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. *J. Exp. Bot.* **2005**, *56*, 255−265.
(18) Robinette, S. L.; Bruschweiler, R.; Schroeder, F. C. *Acc. Chem. Res.* **2012**, *45*, 288−297.
(19) Vonthron-Senecheau, C.; Weniger, B.; Ouattara, M.; Bi, F. T.; Kamenan, A.; Lobstein, A.; Brun, R.; Anton, R. *J. Ethnopharmacol.* **2003**, *87*, 221−225.
(20) Shuaibu, M. N.; Wuyep, P. T. A.; Yanagi, T.; Hirayama, K.; Ichinose, A.; Tanaka, T.; Kouno, I. *Parasitol. Res.* **2008**, *102*, 697−703.