analytical chemistry

Article

pubs.acs.org/ac

2D NMR Barcoding and Differential Analysis of Complex Mixtures for Chemical Identification: The *Actaea* Triterpenes

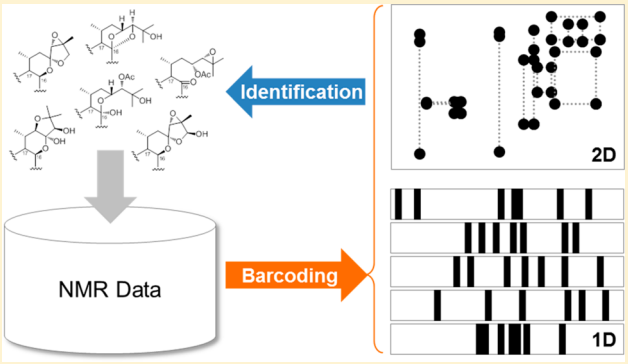
Feng Qiu,[†,#] James B. McAlpine,[†] David C. Lankin,[†] Ian Burton,[‡] Tobias Karakach,[‡] Shao-Nong Chen,[†] and Guido F. Pauli*[,†]

[†]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois 60612, United States

[‡]Institute for Marine Biosciences, National Research Council, Halifax, Nova Scotia B3H 3Z1, Canada

Supporting Information

ABSTRACT: The interpretation of NMR spectroscopic information for structure elucidation involves decoding of complex resonance patterns that contain valuable molecular information (δ and J), which is not readily accessible otherwise. We introduce a new concept of 2D-NMR barcoding that uses clusters of fingerprint signals and their spatial relationships in the δ–δ coordinate space to facilitate the chemical identification of complex mixtures. Similar to widely used general barcoding technology, the structural information of individual compounds is encoded as a specifics pattern of their C,H correlation signals. Software-based recognition of these patterns enables the structural identification of the compounds and their discrimination in mixtures. Using the triterpenes from various *Actaea* (syn. *Cimicifuga*) species as a test case, heteronuclear multiple-bond correlation (HMBC) barcodes were generated on the basis of their structural subtypes from a statistical investigation of their δ_H and δ_C data in the literature. These reference barcodes allowed in silico identification of known triterpenes in enriched fractions obtained from an extract of *A. racemosa* (black cohosh). After dereplication, a differential analysis of heteronuclear single-quantum correlation (HSQC) spectra even allowed for the discovery of a new triterpene. The 2D barcoding concept has potential application in a natural product discovery project, allowing for the rapid dereplication of known compounds and as a tool in the search for structural novelty within compound classes with established barcodes.



NMR spectroscopy has been used not only routinely in structural elucidation of organic compounds, but also increasingly in qualitative and quantitative analysis of complex mixtures, such as combinatorial libraries,[1,2] plant metabolomes,[3−5] and physiological fluids.[6−8] Interpretation of their NMR spectra, particularly when analyzing metabolomic samples, is usually challenging due to spectral complexity. While it might represent hidden information, the unique structure of a single chemical entity is still represented by a distinctive albeit overlapping pattern of NMR signals. In fact, these characteristic resonances can be used as fingerprints to facilitate a rapid and accurate identification of different chemical species.[9]

Moreover, our previous studies have developed ^{1}H iterative Full Spin Analysis (HiFSA) methodology which reproduces the experimental ^{1}H NMR spectra by an iterative QM-based approach.[10−13] HiFSA uses a combination of spectral simulation and iterative permutation of starting values from experimental ^{1}H NMR spectra to generate universal, tabulated data consisting of full sets of ^{1}H NMR spin parameters, i.e., all chemical shifts (δ) and coupling constants (J). This produces highly characteristic HiFSA fingerprints for in silico identification and quantification of the target analytes, including those in complex mixtures. Much like biometric recognition, a small portion of these fingerprints can be sufficient to distinguish different chemical species. The identification of individual components in a mixture can then be focused on the more characteristic but less complicated subregions of the NMR spectra. This concept was recently demonstrated with the development of ActaPredict, a computational tool which utilizes only the easily accessible ^{1}H NMR signals of skeletal methyls as structural identifiers of the nearly 200 known, closely related cycloartane triterpenes from *Actaea*.[14] This approach also exhibited high efficiency and accuracy in the chemical identification for moderately complex mixtures.

As both HiFSA and ActaPredict are based on an analysis of 1D ^{1}H NMR spectra, the 1D approach may become less efficient when analyzing increasingly complex mixtures. In such a situation, 2D-NMR spectra can be more powerful in that they

Received: January 19, 2014
Accepted: March 7, 2014
Published: March 27, 2014

   dx.doi.org/10.1021/ac500188j | *Anal. Chem.* 2014, 86, 3964−3972

provide improved peak resolution via 2D chemical shift and/or signal dispersion.[15,16]

The present study introduces the methodology of 2D-NMR barcoding for the accelerated chemical identification of components in complex mixtures. The barcoding concept has been implemented previously using Raman spectroscopy for the classification and identification of complex biological materials.[17−19] These techniques have employed 1D binary barcodes to highlight the Raman shift fingerprints of each sample. In the present study, matrix (2D) barcodes were used to represent the characteristic patterns produced by the fingerprinting cross-peaks in the 2D-NMR spectra of each analyte. This approach adopts the theory of a popular 2D barcode technique that uses a variety of symbols, such as dots and rectangles in two dimensions, to represent the data relating to a specific object. Similarly, we hypothesize that the combination of cross peaks and their spatial patterns in the 2D-NMR spectra can be regarded as "chemical barcodes" which can be scanned in silico, recognized, and decoded to the structures of chemical substances. As an extension of the previously described methyl-based approach,[14] the present study continues to use the *Actaea* cycloartane triterpenes as the model compounds to validate the concept of 2D-NMR barcoding. Initially, the 2D heteronuclear multiple-bond correlation (HMBC) barcodes for each type of the *Actaea* triterpenes were generated from an analysis of their spectral data found in the literature. On the basis of these barcodes, a VBA macro was written within Microsoft Excel that enables the reading and deciphering of the 2D HMBC barcodes. This in silico tool allowed efficient dereplication of the known triterpenes in complex fractions of *A. racemosa* (black cohosh).

With the aim of searching for new compounds, the present study also employed an NMR-based method that virtually separates the compounds from the complex mixtures by a differential analysis of their heteronuclear single-quantum correlation (HSQC) spectra. Combined with HMBC barcoding, both known and unknown triterpenes were accurately identified from their mixtures without the need for repetitive purification. Combined with cryo-microprobe NMR technology, this provides a powerful toolset for the identification of both minor and difficult-to-separate compounds from complex metabolomic samples.

EXPERIMENTAL SECTION

Samples and Sample Preparation. Two triterpene-enriched samples, A and B, were acquired by VLC fractionation of the EtOAc extracts of black cohosh (*Actaea racemosa* roots and rhizomes) as described in S1 of the Supporting Information. Prior to NMR analysis, the samples were desiccated over anhydrous $CaSO_4$ in vacuo to eliminate chemical shift variations resulting from differences in residual water content.

Mining, Acquisition, and Barcoding of the NMR Data. Initially, an extensive literature survey was conducted to locate and mine the NMR data of known *Actaea* triterpenes, including their δ_H for H-26/27 and δ_C for C-24/25. These NMR data from the literature, associated with other information pertaining to the compound including common names, systematic names, structural types, and biogenetic sources, were then entered into a Microsoft Excel spreadsheet (see S4, Supporting Information). Details of the in-house data acquisition are described in S6, Supporting Information. The spectral barcoding was based on a virtual HMBC spectrum generated by averaging the δ_H and δ_C values of the corresponding nuclei in all known triterpenes of a particular structural subtype. Thus, the characteristic correlations in the virtual HMBC spectra, which include $^3J_{H\text{-}26,C\text{-}24}$, $^2J_{H\text{-}26,C\text{-}25}$, $^3J_{H\text{-}27,C\text{-}24}$, and $^2J_{H\text{-}27,C\text{-}25}$, were used as reference barcodes for both manual and in silico identification.

Matching of the Correlation Patterns. The matching of the correlation patterns to the reference barcodes was carried out by visual comparison between the investigative and reference spectra. This procedure was automated with a Visual Basic for Applications 7.0 in Microsoft Excel 2010 (S7 and S8, Supporting Information).

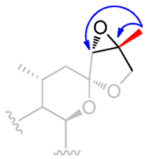
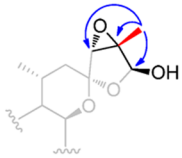
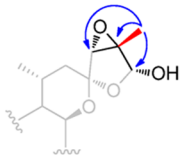
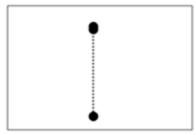
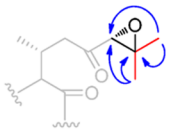
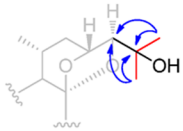
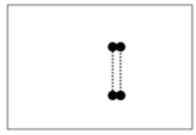
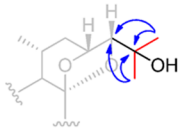
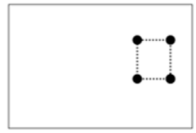
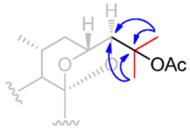
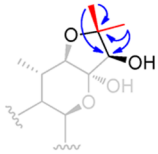
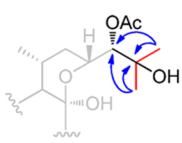
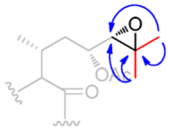
RESULTS AND DISCUSSION

Selection of NMR Experiments. In principle, due to their dimensionality, any type of 2D NMR experiment can be used for the barcoding of target analytes for chemical identification. In practice, however, the appropriate NMR spectra should be selected on the basis of the following four factors: (i) the structural characteristics of the target analytes; (ii) the chemical complexity of the target samples; (iii) the ease of spectral acquisition and barcoding; and (iv) the uniformity of the NMR solvents. Regarding the last point, as chemical shift variations are counterproductive, the data should be acquired not only uniformly but also as much as possible in commonly used NMR solvents. For small molecules with relatively simple proton spin systems, 1D ^{1}H NMR spectra can be the general first choice because of the relative ease of spectral acquisition and interpretation. However, the additional signal dispersion of 2D NMR spectra can be required when analyzing complicated spin systems (including small molecules containing magnetically nonequivalent protons and/or elements of coupling asymmetry) or larger molecules with isochronic spins such as peptides, carbohydrates, glycosides, or even macromolecules. In addition, analysis of mixtures or residually complex samples[20] (see also http://go.uic.edu/residualcomplexity) such as reference materials can benefit from the 2D approach. Final consideration of most suitable experiments for NMR barcoding relate to sample quantity and available NMR instrumentation. Recent developments in high sensitivity probe technology, e.g., cryo- and/or microprobes, have dramatically improved NMR detection sensitivity, thus reducing data collection times and making the acquisition of 2D NMR data of mass-limited samples much more efficient.[21,22]

As recently established,[14] the *Actaea* triterpenes can be divided into several subtypes based on their structural skeletons (S2 and S3, Supporting Information). All consist of the same type of pentacyclic core but differ in the side chain at C-17, which is often biosynthetically cyclized via oxygens located at C-16 to C-23, C-24, and/or C-25. As a result of these structural characteristics, the high field region of the ^{1}H NMR spectra (0.0−2.0 ppm) of individual *Actaea* triterpenes is already complex, exhibiting close similarities for the methylene and methine protons of the pentacyclic core. While subtle differences exist, the signals of these protons are less intense (highly coupled) and severely overlapped due to the underlying convoluted spin systems. By comparison, most of the methyl signals in this region of the spectrum are singlets with much greater resolution and intensities and are, thus, more suitable for use in 1D fingerprinting. By using these signals as descriptors, our previous studies have developed a decision tree model that enabled structural classification of the *Actaea* triterpenes.[14] The results showed that, among the skeletal

dx.doi.org/10.1021/ac500188j | *Anal. Chem.* 2014, 86, 3964−3972

Table 1. The Reference HMBC Barcodes for the Ten Principal Types of *Actaea* Triterpenes

structural type [common name]	HMBC correlations	HMBC cross peaks (δ_H, δ_C)[a]	HMBC reference barcode
(24*R*,25*R*)-24,25-epoxyacta-(16*S*,23*S*)-16,23;23,26-binoxol [26-deoxyacteol]		H-27 → C-24: (1.46, 62.4) H-27 → C-25: (1.46, 68.1)	
(24*R*,25*S*)-24,25-epoxy-(26*S*)-26-hydroxyacta-(16*S*,23*R*)-16,23;23,26-binoxoside [26*β*-hydroxyacteol]		H-27 → C-24: (1.79, 63.5) H-27 → C-25: (1.79, 65.7) H-27 → C-26: (1.79, 98.5)	
(24*R*,25*S*)-24,25-epoxy-(26*R*)-26-hydroxyacta-(16*S*,23*R*)-16,23;23,26-binoxoside [26*α*-hydroxyacteol]		H-27 → C-24: (1.56, 62.9) H-27 → C-25: (1.56, 64.1) H-27 → C-26: (1.56, 97.8)	
(11*S*,15*R*)-11,15-dihydroxy-(24*R*)-24,25-epoxy-16,23-dioxoactanoside [cimicidanol]		H-26 → C-24: (1.32, 65.7) H-26 → C-25: (1.32, 60.7) H-27 → C-24: (1.36, 65.7) H-27 → C-25: (1.36, 60.7)	
(15*R*)-15,25-dihydroxyacta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside [cimigenol]		H-26 → C-24: (1.46, 89.8) H-26 → C-25: (1.46, 71.1) H-27 → C-24: (1.49, 89.8) H-27 → C-25: (1.49, 71.1)	
(15*R*)-15,25-dihydroxyacta-(16*S*,23*R*,24*R*)-16,23;16,24-binoxoside [24-*epi*-cimigenol]		H-26 → C-24: (1.28, 83.8) H-26 → C-25: (1.28, 68.7) H-27 → C-24: (1.40, 83.8) H-27 → C-25: (1.40, 68.7)	
25-acetoxy-(15*R*)-15-hydroxy-acta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside [25-*O*-acetylcimigenol]		H-26 → C-24: (1.67, 86.7) H-26 → C-25: (1.67, 83.2) H-27 → C-24: (1.69, 86.7) H-27 → C-25: (1.69, 83.2)	
(23*R*,24*R*)-23,24-dihydroxy-acta-(16*S*,22*R*)-16,23;22,25-binoxoside [cimiracemoside]		H-26 → C-24: (1.68, 83.4) H-26 → C-25: (1.68, 83.8) H-27 → C-24: (1.77, 83.4) H-27 → C-25: (1.77, 83.8)	
(24*S*)-24-acetoxy-(15*R*,16*R*)-15,16,25-trihydroxyacta-(23*R*)-16,23-monoxoside [hydroshengmanol]		H-26 → C-24: (1.41, 80.5) H-26 → C-25: (1.41, 73.3) H-27 → C-24: (1.45, 80.5) H-27 → C-25: (1.45, 73.3)	
(23*R*)-23-acetoxy-(24*S*)-24,25-epoxy-(15*R*)-15-hydroxy-16-oxoactanoside [23-*O*-acetylshengmanol]		H-26 → C-24: (1.27, 65.1) H-26 → C-25: (1.27, 58.5) H-27 → C-24: (1.41, 65.1) H-27 → C-25: (1.41, 58.5)	

[a]The HMBC cross-peaks were denoted by the corresponding $^1H-^{13}C$ correlations and discriminated by their spatial locations in the $\delta_H-\delta_C$ coordinate of the HMBC spectrum, using the average δ_H and δ_C values listed in S4, Supporting Information.

dx.doi.org/10.1021/ac500188j | *Anal. Chem.* 2014, 86, 3964−3972

methyl groups, CH_3-26 and CH_3-27 are highly significant and accessible structural indicators.

However, in addition to the methyl signals, substantial differences can also be observed for the signals of the methylene and methine protons of the side chains. Owing to the deshielding effects of the neighboring oxygens, the chemical shifts of these protons appear at lower fields (5.0−3.0 ppm), where signals are more dispersed. Similarly, the signals of the characteristic carbons of the side chains, i.e., C-23, C-24, and C-25, appear at lower field in the ^{13}C NMR spectra (50−90 ppm) and can help differentiate the structures of the side chains. Despite these dispersive characteristics, the use of only 1D ^1H or ^{13}C NMR spectra is often insufficient to resolve the chemical complexity of triterpene mixtures and identify individual components with certainty. In comparison, 2D HMBC spectra are more effective in this task, e.g., when building on correlations between the methyl protons of CH_3-26/CH_3-27 and their proximal carbons C-25/C-24. Compared with the 1D NMR signals of these characteristic protons and carbons, their correlations in the 2D HMBC spectra provide a greater degree of certainty for structural identification. Even more importantly for the aim of barcoding, an added advantage of the use of such HMBC correlations is that the methyl signals have greater intensity, while the signals of the carbons have greater dispersion. This makes the 2D HMBC experiment particularly suitable for the barcoding analysis of minor components in complex mixtures.

Generation and Use of HMBC Barcodes. In widely adopted earlier barcoding technology, barcodes are used to represent data by varying the widths and spacings of parallel lines which are referred to as linear or one-dimensional. More recently, barcodes have evolved into two dimensions using rectangles, dots, hexagons, and other geometric patterns, allowing the encoding of additional information. On the basis of this technique, e.g., molecular biology has developed the taxonomic method of DNA barcoding, which utilizes short genetic markers in an organism's DNA to identify *biological* species and subspecies. In the present study, we are applying a similar approach to spectral analysis to perform structural identification of *chemical* species. The basic concept is that the spectral signals and their topological patterns are reconstructed into reference barcodes that represent the 2D or 3D structures of their respective chemical species. These barcodes can then be used to aid the in silico translation of the investigative spectra into the human-readable format as either the names or images of the chemical identities.

After the HMBC had been selected as the appropriate NMR experiment, the next step was to determine which signals and what patterns of these signals are to be used for barcoding of the corresponding compounds. In the HMBC spectra of *Actaea* triterpenes, both methyl signals of CH_3-26 and CH_3-27 show correlations to their proximal carbons C-25 and C-24. This results in four HMBC correlations, i.e., $^3J_{H\text{-}26,C\text{-}24}$, $^2J_{H\text{-}26,C\text{-}25}$, $^3J_{H\text{-}27,C\text{-}24}$, and $^2J_{H\text{-}27,C\text{-}25}$, which altogether produce a rectangular pattern with the four cross-peaks as the four vertices. These rectangular HMBC patterns are differentiated by their widths ($\Delta\delta_C$ and $\Delta\delta_H$) and spacings (δ_H and δ_C). Thus, each of these specific patterns can be assigned a unique "HMBC barcode" which represents a specific type of *Actaea* triterpene. Practically speaking, recognition of these HMBC barcodes identifies the presence of related chemical species.

Next, the reference HMBC barcodes that represent each type of *Actaea* triterpene (Table 1) were generated by a comparative analysis of the NMR data of purified reference compounds. The data sets of ^1H and ^{13}C NMR spectra of approximately 150 known *Actaea* triterpenes of 10 structural subtypes, predominantly measured in pyridine-d_5, were initially collected in an in-house database (S4, Supporting Information). For the triterpenes in each subtype, the standard deviations (σ) of $\delta_{H\text{-}26}$ and $\delta_{H\text{-}27}$ were within the range of 0.01−0.10 ppm (S5, Supporting Information), indicating a small variance in the chemical shifts of these two methyl protons. A small variation was also observed for $\delta_{C\text{-}24}$ and $\delta_{C\text{-}25}$, showing a σ of 0.1−2.0 ppm. Accordingly, using the average values of δ_H and δ_C, a virtual HMBC spectrum containing only the characteristic correlations could be constructed to generate reference barcodes for each structural subtype of the *Actaea* triterpenes (Figure 1). Each of these barcodes consists of a cluster of cross peaks which form distinct lines and/or rectangular patterns that can be readily recognized.

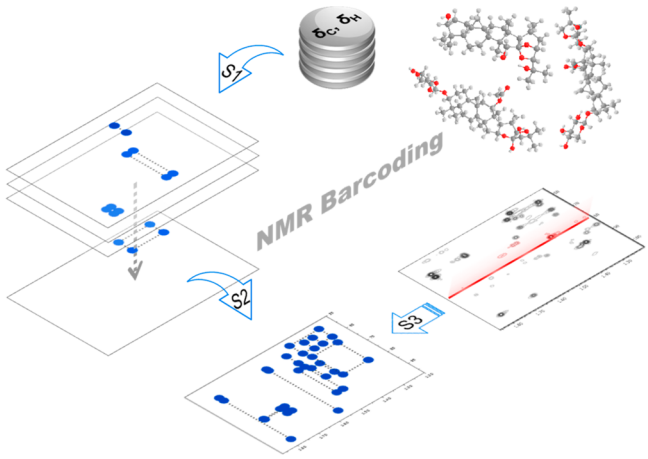


Figure 1. The basic workflow of 2D NMR barcoding. S1: The 2D NMR data of reference compounds are statistically analyzed to generate a virtual spectrum, representing a particular type of structure. S2: The distribution and correlation patterns of signals in the virtual 2D NMR spectrum are used as reference barcodes for the structural identification. S3: The chemical components of a (residually) complex sample are identified by in silico matching of their experimental NMR signals with the reference NMR barcodes.

According to these reference barcodes, the particular types of *Actaea* triterpenes may already be identified by a visual inspection of their 2D HMBC spectra. In addition, a VBA program named NMR Barcode Reader was developed within Microsoft Excel (S7 and S8, Supporting Information) to facilitate the automated batch identification of the more complex samples. After processing and peak picking in any third-party NMR processing software, the 2D HMBC spectrum can be saved or exported as a delimited text file in the format of CSV or TSV, consisting of two columns which are the δ_H and δ_C values of the cross-peaks. Upon loading this file in the Excel spreadsheet, the VBA program searches and locates the line and rectangular patterns of the correlations (investigative barcodes) in the peak listing of the 2D HMBC spectrum. Finally, these candidate patterns are matched with the reference barcodes by comparing their shape similarity and spatial locations in the δ_H−δ_C coordinate.

Chemical Identification by HMBC Barcoding. The reference HMBC barcodes were used to analyze the chemical constituents of two complex mixtures of *Actaea* triterpenes,

dx.doi.org/10.1021/ac500188j | Anal. Chem. 2014, 86, 3964−3972

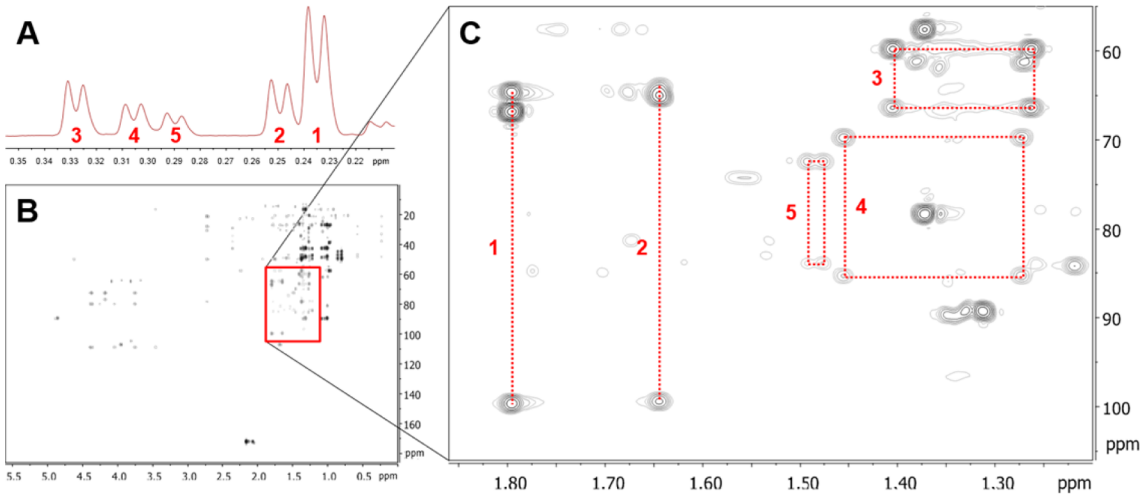


Figure 2. The triterpene components in sample A were identified by matching their HMBC correlations with the reference HMBC barcodes. (A) The ^{1}H NMR signals of cyclopropane methylene protons indicated the significant (residual) chemical complexity of sample A, containing at least five *Actaea* triterpenes. (B) The HMBC spectrum of sample A was acquired and used for chemical identification. (C) By matching with the reference HMBC barcodes, these five *Actaea* triterpenes were identified as: (24*R*,25*S*)-24,25-epoxy-(26*S*)-26-hydroxyacta-(16*S*,23*R*)-16,23;23,26-binoxoside (26*β*-hydroxyacteol, **1**), (24*R*,25*S*)-24,25-epoxy-(26*R*)-26-hydroxyacta-(16*S*,23*R*)-16,23;23,26-binoxoside (26*α*-hydroxyacteol, **2**), (23*R*)-23-acetoxy-(24*S*)-24,25-epoxy-(15*R*)-15-hydroxy-16-oxoactanoside (23-*O*-acetylshengmanol, **3**), (15*R*)-15,25-dihydroxyacta-(16*S*,23*R*,24*R*)-16,23;16,24-binoxoside (24-*epi*-cimigenol, **4**), and (24*S*)-24-acetoxy-(15*R*,16*R*)-15,16,25-trihydroxyacta-(23*R*)-16,23-monoxoside (hydroshengmanol, **5**).

Table 2. The Individual Triterpenes Identified in the Residually Complex Samples A and B by Means of in Silico Matching of Their 2D HMBC Correlations with the Reference HMBC Barcodes

sample	cpd	HMBC matches	structural types, new systematic name [common name]
A	**1**	(1.79, 63.6), (1.79, 65.8), (1.79, 98.7)	(24*R*,25*S*)-24,25-epoxy-(26*S*)-26-hydroxyacta-(16*S*,23*S*)-16,23;23,26-binoxoside [26*β*-hydroxyacteol]
	2	(1.64, 63.1), (1.64, 64.0), (1.64, 98.4)	(24*R*,25*S*)-24,25-epoxy-(26*R*)-26-hydroxyacta-(16*S*,23*R*)-16,23;23,26-binoxoside [26*α*-hydroxyacteol]
	3	(1.25, 65.4), (1.25, 58.8), (1.41, 65.4), (1.41, 58.8)	(23*R*)-23-acetoxy-(24*S*)-24,25-epoxy-(15*R*)-15-hydroxy-16-oxoactanoside [23-*O*-acetylshengmanol]
	4	(1.26, 84.3), (1.26, 68.7), (1.45, 84.3), (1.45, 68.7)	(15*R*)-15,25-dihydroxyacta-(16*S*,23*R*,24*R*)-16,23;16,24-binoxoside [24-*epi*-cimigenol]
	5	(1.47, 82.9), (1.47, 71.4), (1.48, 82.9), (1.48, 71.4)	(24*S*)-24-acetoxy-(15*R*,16*R*)-15,16,25-trihydroxyacta-(23*R*)-16,23-monoxoside [hydroshengmanol]
B	**6**	(1.71, 83.6), (1.79, 83.6), (1.71, 83.6), (1.79, 83.6)	(23*R*,24*R*)-23,24-dihydroxyacta-(16*S*,22*R*)-16,23;22,25-binoxoside [cimiracemoside]
	7	(1.71, 83.6), (1.79, 83.6), (1.71, 83.6), (1.79, 83.6)	(23*R*,24*R*)-23,24-dihydroxyacta-(16*S*,22*R*)-16,23;22,25-binoxoside [cimiracemoside]
	8	(1.53, 63.0), (1.53, 68.5)	(24*R*,25*R*)-24,25-epoxyacta-(16*S*,23*R*)-16,23;23,26-binoxol [26-deoxyacteol]
	9	(1.51, 63.0), (1.51, 68.5)	(24*R*,25*R*)-24,25-epoxyacta-(16*S*,23*R*)-16,23;23,26-binoxol [26-deoxyacteol]
	10	no matches	previously unknown

which were generated by chromatographic fractionation of black cohosh extracts (S1, Supporting Information). As these two samples exhibited significant chemical complexity, containing at least five triterpene components of different aglycone types, it was rather challenging to identify their structures by a full interpretation of the 1D/2D NMR spectra. In such a situation, the recognition of barcodes as specific 2D NMR correlation patterns simplified the spectral analysis and led to a much more rapid chemical identification. As shown in Figure 2A, the doublet ^{1}H signals of the cyclopropane methylene protons (H-19), resonating in the range of δ_H 1.00−0.20, indicated the presence of five *Actaea* triterpenes in sample A. Upon the acquisition of the HMBC spectrum, the chemical identification was carried out by the in silico search of the specific correlations based on the reference HMBC barcodes. As a result, five investigative barcodes were located in the HMBC spectrum (Figure 2C). By comparison with the reference barcodes (Table 2), these patterns were identified as belonging to (24*R*,25*S*)-24,25-epoxy-(26*S*)-26-hydroxyacta-(16*S*,23*R*)-16,23;23,26-binoxoside (commonly known as 26*β*-hydroxyacteol, **1**), (24*R*,25*S*)-24,25-epoxy-(26*R*)-26-hydroxyacta-(16*S*,23*R*)-16,23;23,26-binoxoside (26*α*-hydroxyacteol, **2**), (23*R*)-23-acetoxy-(24*S*)-24,25-epoxy-(15*R*)-15-hydroxy-16-oxoactanoside (23-*O*-acetylshengmanol, **3**), (15*R*)-15,25-dihydroxyacta-(16*S*,23*R*,24*R*)-16,23;16,24-binoxoside (24-*epi*-cimigenol, **4**), and (24*S*)-24-acetoxy-(15*R*,16*R*)-15,16,25-trihydroxyacta-(23*R*)-16,23-monoxoside (hydroshengmanol, **5**).

Similarly, sample B also contained at least five *Actaea* triterpenes, as indicated by the number of doublet ^{1}H signals of H-19 observed in the high field region (Figure 3A). Further in silico searching and matching of HMBC barcodes indicated only two structural subtypes, i.e., acteol and cimiracemoside derivatives, were present in the sample (Figure 3C, Table 2). For an individual triterpene, the ^{1}H NMR signals of one cyclopropane methylene proton (1H) and one methyl group (3H) should have an integration ratio of 1:3. Therefore, by calculating this value, four of the five triterpenes were determined to be two (23*R*,24*R*)-23,24-dihydroxyacta-(16*S*,22*R*)-16,23;22,25-binoxosides (cimiracemosides, **6** and

dx.doi.org/10.1021/ac500188j | *Anal. Chem.* 2014, 86, 3964−3972

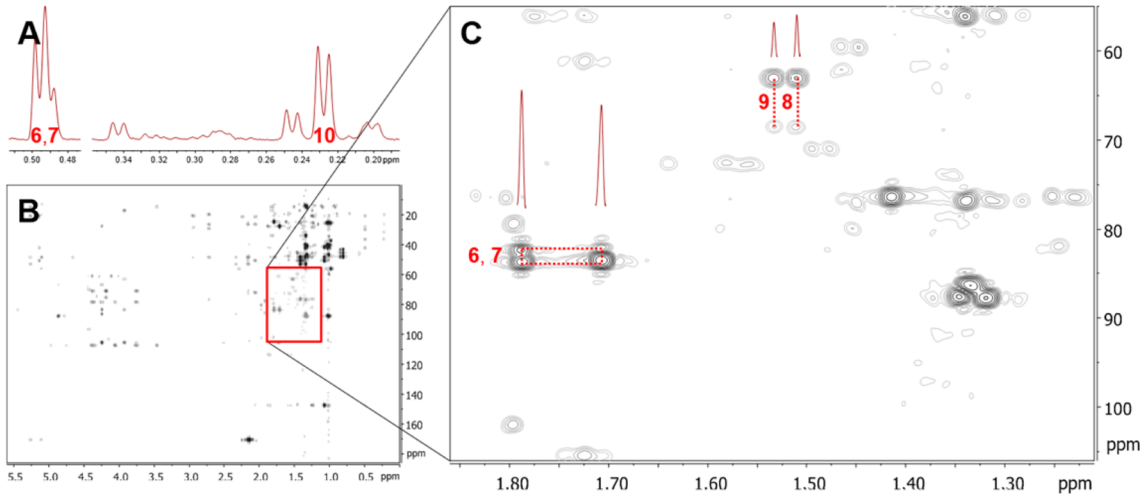


Figure 3. The triterpene components in sample B were identified by matching their HMBC correlations with the reference HMBC barcodes. (A) The ^{1}H NMR signals of cyclopropane methylene protons suggested the presence of at least five *Actaea* triterpenes in sample B. (B) The full HMBC spectrum of sample B. (C) By matching with the reference HMBC barcodes, only two different skeleton types of triterpenes were found to be present, in contrast to sample A which contained five (Figure 2). The integral ratio between methyl and cyclopropane methylene protons further confirmed four of the five triterpenes as being two (23*R*,24*R*)-23,24-dihydroxyacta-(16*S*,22*R*)-16,23;22,25-binoxosides (cimiracemosides, **6** and **7**) and two (24*R*,25*R*)-24,25-epoxyacta-(16*S*,23*S*)-16,23;23,26-binoxosides (26-deoxyacteols, **8** and **9**). However, one triterpene (**10**) could not be matched to a known triterpene skeleton and was subsequently identified by differential analysis of HSQC spectra (Figure 4).

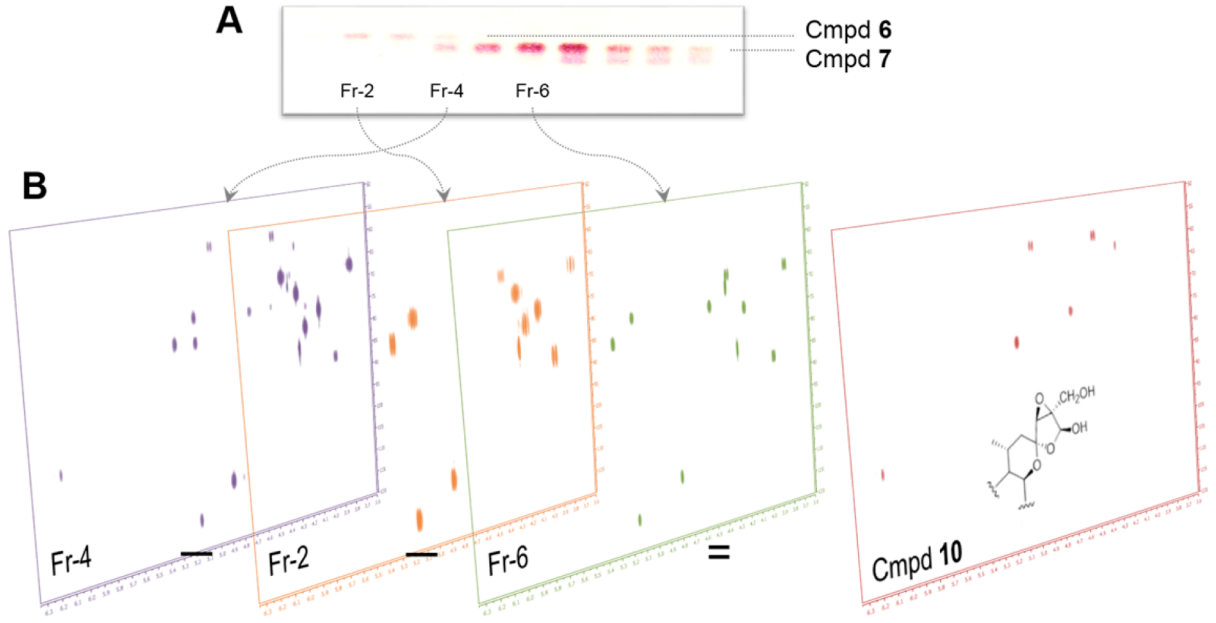


Figure 4. The structural elucidation of compound **10** by differential analysis of HSQC spectra. (A) The NP-TLC analysis of the corresponding VLC fractions indicated that Fr-4 contained the same components as both nearby fractions Fr-2 and Fr-6, albeit in different ratios. (B) After subtracting the HSQC signals of Fr-2 and Fr-4 from Fr-6, a residual HSQC spectrum was obtained, showing only the signals belonging to **10**.

7) as the major components and two additional minor (24*R*,25*R*)-24,25-epoxyacta-(16*S*,23*S*)-16,23;23,26-binoxosides (26-deoxyacteols, **8** and **9**). However, one triterpene (**10**), with the chemical shifts of H-19a/b appearing at δ_H 0.576 and 0.228 ppm, remained unknown due to the absence of the corresponding 2D HMBC correlations that matched any of the reference barcodes. This suggested that **10** might have an unusual (new) structure in the side chain, resulting from the biosynthetic modification of the methyl groups, CH_3-26 and CH_3-27, leading to a major change in the HMBC correlation patterns. This assumption suggested further investigation to characterize the structure by the following differential analysis of the HSQC spectra.

Differential Analysis of the HSQC Spectra. Owing to the chemical complexity and limited mass of sample B, further purification and isolation of **10** was not feasible. As sample B resulted from repeated fractionation of black cohosh extracts, the signals of **10** in the spectrum of sample B might be distinguished from signals common to the spectra of the nearby fractions. By such a differential analysis of several HSQC spectra, **10** might be structurally identified from the mixture without a physical separation. Recently, the differential analysis of 2D-NMR spectra has been shown to be powerful in the

dx.doi.org/10.1021/ac500188j | *Anal. Chem.* 2014, 86, 3964−3972

Table 3. The δ_H and δ_C (in ppm) Values of the Side Chains of the Cycloartane Triterpenes, Actein (11)[a] and the Newly Identified 27-Hydroxyactein (10)[b]

skeleton	position	actein (26*S*) [**11a**] δ_H, mult	δ_C	27-hydroxyactein (26*S*) [**10a**] δ_H, mult	δ_C	actein (26*R*) [**11b**] δ_H, mult	δ_C	27-hydroxyactein (26*R*) [**10b**] δ_H, mult	δ_C
	16	4.630, m	73.10	4.660, m	72.75	4.630, m	73.00	4.660, m	72.23
	24	3.948, s	62.93	4.253, s	60.85	3.803, s	63.46	4.153, s	60.87
	26	5.771, s	98.20	6.144, s	96.57	5.748, s	98.45	6.189, s	95.81
	27a	1.799, s	13.15	4.437, d	57.78	1.645, s	13.06	4.444, d	57.38
	27b			4.949, d				4.606, d	

[a]The ^{1}H and ^{13}C chemical shifts of actein were taken from ref 25. [b]The δ_C values were determined from the residual HSQC spectra (see main text); $J_{H27a,H27b}$ was 12.6 Hz.

identification of new compounds from (residually) complex natural product mixtures.[23,24] In the present study, the HSQC spectra were considered to be highly appropriate for the differential analysis, because they provide both ^{1}H and ^{13}C chemical shift information but yield much less complicated cross peak patterns than HMBC due to the restriction to only one-bond ^{1}H−^{13}C correlations. As the major structural differences of *Actaea* triterpenes arise from modifications of their side chains at C-17, the differential analysis of their HSQC spectra can be focused on one specific region, i.e., δ_H 5.50/3.30 and δ_C 100.0/50.0, where the fingerprint signals of the ether, oxirane, and acetoxy groups in the side chains are commonly observed.

As indicated by the NP-TLC analysis in Figure 4A, sample B contained the same components as both nearby fractions Fr-2 and Fr-4. The HSQC spectra of these three fractions were initially acquired under identical experimental conditions (S6, Supporting Information). Figure 4B shows the aligned spectra in stacked mode, with an expansion of the fingerprint region for the side chains and sugar moieties. After subtracting the cross peaks in Fr-2 and Fr-4 from Fr-6, a residual HSQC spectrum was obtained, showing only the cross peaks of **10**. The characteristic signal at δ_H 4.660 and δ_C 72.75 (H-16) indicated that the structure of **10** is similar to (12*R*)-12-acetoxy-(24*S*,25-*R*)-24,25-epoxy-(26*R*&*S*)-26-hydroxy-3-*O*-*β*-D-xylopyranosylacta-(16*S*,23*R*)-16,23;23,26-binoxoside (actein, **11**), a major acteol-type triterpene in black cohosh. However, comparative analysis of the ^{1}H NMR spectra of **10** and actein (Table 3) showed that one methyl signal (CH_3-27) was missing in **10**.

Further investigation of the lower-field ^{1}H NMR spectra found that the H-24 protons in **10** resonate at δ_H 4.153/4.253 (26*R*/26*S*), compared to those in actein at δ_H 3.803/3.948. Similarly, protons H-26 in actein are observed at δ_H 6.189/6.144, compared to those in **10** at δ_H 5.771/5.748. Interestingly, in **10**, two pairs of geminal protons were observed at δ_H 4.949/4.437 and δ_H 4.606/4.444, respectively, both giving rise to a pair of doublet signals (J = 12.6 Hz; Figure 5). These data suggested that compound **10** was a 7:3 mixture of 26*β*−OH and 26*α*−OH isomers, possibly formed by the oxidation of CH_3-27 in actein to −CH_2OH. The observed HMBC correlation between H-27a and C-26 (δ_C 96.57) further

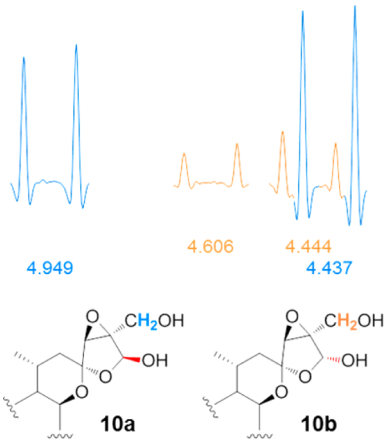


Figure 5. The H-27 signals and side-chain structures of the 26-hydroxyactein (**10**) newly identified in Sample B. This triterpene was present as an epimeric mixture containing a 7:3 ratio of 26*β*−OH (**10a**) and 26*α*−OH (**10a**) as indicated by two pairs of doublet signals (J 12.6 Hz) of their geminal protons H-27 at δ_H 4.949/4.437 and δ_H 4.606/4.444, respectively.

confirmed that the structure of **10** is a new type of *Actaea* triterpene, namely, (12*R*)-12-acetoxy-(24*S*,25*R*)-24,25-epoxy-(26*R*&*S*)-26,27-dihydroxy-3-*O*-*β*-D-xylopyranosylacta-(16*S*,23*R*)-16,23;23,26-binoxoside. As **6** and **7** are known glycosides, it was readily determined that **10a**/**b** are xylosides based on the residual sugar signals in the HSQC spectrum. Finally, it is worth noting that the content of both **10** isomers together was only ∼20 mol %, i.e., ∼80 μg in the ∼400 μg sample. This demonstrates the power (i.e., the sensitivity and resolution) of the NMR cryogenic microprobe used in this study for the identification of compounds in chemically complex and mass-limited samples.

CONCLUSIONS

The present study, using *Actaea* triterpenes as the model, demonstrates 2D NMR barcoding as an efficient tool for dereplication of (residually) complex mixtures of small molecule analogues. This approach requires the identification of structurally discriminating signal patterns for the target

dx.doi.org/10.1021/ac500188j | Anal. Chem. 2014, 86, 3964−3972

analytes. The clear advantage of 2D barcodes over 1D signal patterns is evident as recognizing not only distinctive clusters but also spatial relationships. Nonetheless, HiFSA(-based) fingerprints and qHNMR methods[10−13] provide confident compound identification because: (i) HiFSA fingerprints are highly congruent (RMS ≪ 1%) matches of complex ^{1}H resonance patterns; (ii) the iterative nature of the HiFSA process extends well to mixtures; (iii) multiple (or even all) signals of an analyte are fingerprinted. Recent results from *Silybum*[11] and *Ginkgo*[12,13] indicated that the intrinsic limitations from the 1D nature of HiFSA remain to be explored.

The triterpenes serve as an excellent example for the 2D barcoding approach. However, the method is suitable for the analysis of other molecular classes, especially secondary metabolites, where a large number of close structural analogues often result in residual complexity presenting a dereplication challenge. 2D NMR enhances spectral barcoding with greater signal resolution and more definitive signal patterns, altogether leading to greater differentiation. Moreover, as shown by the differential analysis using 2D HSCQ, the improved sensitivity of cryogenic and/or micro probes make NMR barcoding a powerful tool for the analysis of minor components that are embedded in complex mixtures and/or difficult to separate by chromatography. As shown, 2D HMBC and HSQC barcodes can be integrated into discovery armamentarium to expedite both the dereplication of known and detection/identification of unknown chemicals. This barcoding is not limited to empirical approaches but can be integrated with computational approaches to establish structure/spectral correlations. Keys to developing this approach for new target compounds are: (i) sufficient reliable NMR data for the target analytes; (ii) comprehensive literature mining; (iii) establishment and maintenance of spectral databases; and (iv) validation of the spectral predictions. Implementation of measurements of shape similarity and related algorithms will enhance the accuracy and robustness of pattern matching and enable automation of NMR barcoding. Altogether, these improvements can expand applicability and augment the performance of NMR barcoding for chemical identification even within metabolomic mixtures.

■ ASSOCIATED CONTENT

Supporting Information

The fractionation procedure for triterpene-enriched samples (S1); the structures and newly established nomenclature of the *Actaea* triterpenes used as model compounds in this study (S2); the distribution of structural types (S3); the NMR data of the *Actaea* triterpenes in the in-house database (S4); the average chemical shifts of key signals (S5); the NMR experimental details (S6); the algorithm used for HMBC barcode search and matching (S7); the user interface of the NMR Barcode Reader (S8). The latest version of the VBA program for the in silico identification of *Actaea* triterpenes (NMR-BRAT v1.0) is made available on the corresponding author's website (http://tigger.uic.edu/~gfp/). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gfp@uic.edu. Tel: (312) 355-1949. Fax: (312) 355-2693.

Present Address

#F.Q.: The Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, U.S.A.

Notes

The authors declare no competing financial interest.

This paper represents part 22 of the series on *Residual Complexity and Bioactivity* (see http://go.uic.edu/residualcomplexity).

■ ACKNOWLEDGMENTS

This study was supported in part by Grant P50 AT000155 from the Office of Dietary Supplements (ODS) and the National Center for Complementary and Alternative Medicine (NCCAM), both of the National Institutes of Health (NIH), as well as Grant RC2 AT005899 from NCCAM/NIH. We thank Dr. Tanja Gödecke at UIC for preparing black cohosh extracts and Dr. Ayano Imai for her assistance in mining the literature. The authors also gratefully acknowledge helpful discussions with Dr. John Walter, IMB/NRC, Halifax (Canada). The latest version of the VBA program for the in silico identification of *Actaea* triterpenes (NMR-BRAT v1.0) is made available on the corresponding author's Web site (http://tigger.uic.edu/~gfp/).

■ REFERENCES

(1) Wu, B.; Zhang, Z.; Noberini, R.; Barile, E.; Giulianotti, M.; Pinilla, C.; Houghten, R. A.; Pasquale, E. B.; Pellecchia, M. *Chem. Biol.* **2013**, *20*, 19−33.

(2) Rizzo, V.; Pinciroli, V. *J. Pharm. Biomed. Anal.* **2005**, *38*, 851−857.

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

(4) Schripsema, J. *Phytochem. Anal.* **2010**, *21*, 14−21.

(5) Kim, H. K.; Choi, Y. H.; Verpoorte, R. *Nat. Protoc.* **2010**, *5*, 536−549.

(6) Zhang, S.; Nagana Gowda, G. A.; Ye, T.; Raftery, D. *Analyst* **2010**, *135*, 1490−1498.

(7) Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *NMR Biomed.* **2005**, *18*, 143−162.

(8) Solanky, K. S.; Bailey, N. J.; Beckwith-Hall, B. M.; Bingham, S.; Davis, A.; Holmes, E.; Nicholson, J. K.; Cassidy, A. *J. Nutr. Biochem.* **2005**, *16*, 236−244.

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

(10) Napolitano, J. G.; Lankin, D. C.; McAlpine, J. B.; Niemitz, M.; Korhonen, S.-P.; Chen, S.-N.; Pauli, G. F. *J. Org. Chem.* **2013**, *78*, 9963−9968.

(11) Napolitano, J. G.; Lankin, D. C.; Graf, T. N.; Friesen, J. B.; Chen, S.-N.; McAlpine, J. B.; Oberlies, N. H.; Pauli, G. F. *J. Org. Chem.* **2013**, *78*, 2827−2839.

(12) Napolitano, J. G.; Lankin, D. C.; Chen, S.-N.; Pauli, G. F. *Magn. Reson. Chem.* **2012**, *50*, 569−575.

(13) Napolitano, J. G.; Gödecke, T.; Rodríguez-Brasco, M. F.; Jaki, B. U.; Chen, S.-N.; Lankin, D. C.; Pauli, G. F. *J. Nat. Prod.* **2012**, *75*, 238−248.

(14) Qiu, F.; Imai, A.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Farnsworth, N. R.; Chen, S.-N.; Pauli, G. F. *J. Nat. Prod.* **2012**, *75*, 432−443.

(15) Xia, J.; Bjorndahl, T.; Tang, P.; Wishart, D. *BMC Bioinf.* **2008**, *9*, 1−16.

(16) Xi, Y.; Ropp, J.; Viant, M.; Woodruff, D.; Yu, P. *Metabolomics* **2006**, *2*, 221−233.

(17) Patel, I. S.; Premasiri, W. R.; Moir, D. T.; Ziegler, L. D. *J. Raman Spectrosc.* **2008**, *39*, 1660−1672.

(18) Watson, D. A.; Brown, L. O.; Gaskill, D. F.; Naivar, M.; Graves, S. W.; Doorn, S. K.; Nolan, J. P. *Cytometry, Part A* **2008**, *73A*, 119−128.
(19) Wang, Q.; Lonergan, S. M.; Yu, C. *Meat Sci.* **2012**, *91*, 232−239.
(20) Pauli, G. F.; Chen, S.-N.; Friesen, J. B.; McAlpine, J. B.; Jaki, B. U. *J. Nat. Prod.* **2012**, *75*, 1243−1255.
(21) Dalisay, D. S.; Molinski, T. F. *J. Nat. Prod.* **2009**, *72*, 739−744.
(22) Martin, G. E. *eMagRes* **2012**, *1*, 883−894.
(23) Hu, P. J.; Sherman, D. H. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7685−7686.
(24) Schröder, F. C.; Gibson, D. M.; Churchill, A. C. L.; Sojikul, P.; Wursthorn, E. J.; Krasnoff, S. B.; Clardy, J. *Angew. Chem., Int. Ed.* **2007**, *46*, 901−904.
(25) Kusano, A.; Takahiro, M.; Shibano, M.; In, Y.; Ishida, T.; Miyase, T.; Kusano, G. *Chem. Pharm. Bull.* **1998**, *46*, 467−472.