Anal. Chem. **2007,** *79,* 1010−1021

Substitution Patterns in Aromatic Rings by Increment Analysis. Model Development and Application to Natural Organic Matter

E. M. Perdue,*,† N. Hertkorn,‡,§ and A. Kettrup‡

School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, 30332, and GSF-Forschungszentrum für Umwelt und Gesundheit, Institut für Ökologische Chemie, 85758 Neuherberg, Germany

The aromatic region of two-dimensional heteronuclear ^{1}H, ^{13}C NMR spectra of natural organic matter and related materials (e.g., ^{1}H and ^{13}C chemical shifts ranging from approximately 5 to 10 and 80 to 140 ppm, respectively) is highly complex and difficult to interpret using conventional approaches. In principle, this region of the NMR spectrum should be amenable to detailed analysis, because the effects of many common substituents on the chemical shifts of aromatic carbon and hydrogen are well documented. This paper describes the development of a model for prediction of *s*ubstitution *p*atterns in *a*romatic *r*ings by *i*ncrement *a*nalysis (SPARIA). In the forward mode, SPARIA is used to predict the chemical shifts of ^{1}H and ^{13}C on aromatic moieties containing every possible combination of eight common substituents that are likely to be representative of substituents on aromatic moieties in natural organic matter. The accuracy of SPARIA in the forward mode is evaluated for 29 aromatic compounds (100 peaks) by comparison of predicted chemical shifts for ^{1}H and ^{13}C with experimental values and with predictions of commercially available software for prediction of NMR spectra. The most important development in this paper is the inverse mode that is built into SPARIA. Given chemical shifts for ^{1}H and ^{13}C (such as may be obtained from a two-dimensional, heteronuclear NMR spectrum), the inverse mode of SPARIA calculates all possible combinations of the eight selected substituents that yield chemical shifts within a specified window of chemical shift for both ^{1}H and ^{13}C. Both the distribution of possible substitution patterns and simple descriptive statistics of the distribution are thus obtained. The inverse mode of SPARIA has been tested on the 29 aromatic compounds (100 peaks) that were used to evaluate its forward mode, and the dependence of the inverse process on the size of the chemical shift window has been evaluated. Finally, the inverse mode of SPARIA has been applied to selected peaks from the two-dimensional heteronuclear HSQC spectrum of a sample of natural organic matter that was isolated by reverse osmosis from the Suwannee River in southeastern Georgia.

Among the properties of natural organic matter (NOM) that have received considerable attention in the past 30 years is its aromaticity. The aromaticity of NOM is generally attributed to two major sources—aromatic compounds derived more-or-less directly from biodegradation of biomass[1,2] and aromatic compounds formed during combustion of biomass and fossil fuels[3] or by weathering of graphitic materials.[4] The aromaticity of the former materials is most likely present in simple benzene rings; however, condensed aromatic rings are likely to contribute substantially to the aromaticity of the latter materials. The greater mobility of simple benzene rings in NOM results in relatively sharper cross-peaks than for condensed ring systems in both homonuclear and heteronuclear two-dimensional NMR spectroscopy.[5] The analysis of substitution patterns on these simple benzene rings is the subject of this paper. Although the motivation for and application of this analysis lies in the study of NOM, the approach developed here is broadly applicable to the interpretation of the two-dimensional NMR spectra of organic compounds containing benzene rings.

If the pattern of substituents on an aromatic ring is known, it is a relatively straightforward task to predict the chemical shifts of ^{1}H and ^{13}C on that aromatic ring. This is a deterministic process, and several commercial software products are commonly used for this purpose. When working with NOM, where substitution patterns are not known, this forward process is of little practical use. It is the goal of this paper to provide a probabilistic analysis of the substitution patterns on an aromatic ring that can account for observed chemical shifts of ^{1}H and ^{13}C, as observed in a two-dimensional heteronuclear NMR experiment. This process is the inverse of the far more common forward process. The conceptual model that is described here enables the prediction of *s*ubstitution *p*atterns in *a*romatic *r*ings by *i*ncrement *a*nalysis (SPARIA). At

* To whom correspondence should be addressed. E-mail: mperdue@eas.gatech.edu. Phone: 1-404-894-3942. FAX: 1-404-894-5638.

† Georgia Institute of Technology.

‡ Institut für Ökologische Chemie.

§ E-mail: hertkorn@gsf.de. Phone: +4989-3187-2834. FAX: +4989-3187-2705.

(1) Hedges, J. I.; Oades, J. M. *Org. Geochem.* **1997**, *27*, 319−361.
(2) Kögel-Knabner, I. *Soil Biol. Biochem.* **2002**, *34*, 139−162.
(3) Hockaday, W. C.; Grannas, A. M.; Kim, S.; Hatcher, P. G. *Org. Geochem.* **2006**, *37*, 501−510.
(4) Dickens, A. F.; Gelinas, Y.; Masiello, C. A.; Wakeham, S.; Hedges, J. I. *Nature* **2004**, *427*, 336−339.
(5) Fan, T. W.-M.; Higashi, R. M.; Lane, A. N. *Environ. Sci. Technol.* **2000**, *34*, 1636−1646.

10.1021/ac061611y CCC: $37.00 © 2007 American Chemical Society
Published on Web 12/23/2006

Table 1. Incremental Chemical Shifts for Selected Substituents on Aromatic Rings[a]

		^{1}H incremental chemical shift			^{13}C incremental chemical shift		
		ortho	meta	para	ortho	meta	para
electron-withdrawing	COOH	0.85	0.18	0.25	1.6	−0.1	4.8
groups (COR)	$COOCH_3$	0.71	0.11	0.21	1.0	0.0	4.5
	COC_2H_5	0.63	0.13	0.20	0.2	0.2	4.2
neutral groups (R)	H	0.00	0.00	0.00	0.0	0.0	0.0
	C_2H_5	−0.15	−0.06	−0.18	−0.6	−0.1	−2.8
	$CH{=}CH_2$	0.06	−0.03	−0.10	−1.8	−1.8	−3.5
electron-donating	OH	−0.56	−0.12	−0.45	−12.6	1.6	−7.6
groups (OR)	OCH_3	−0.48	−0.09	−0.44	−15.0	0.9	−8.1

[a] ^{1}H and ^{13}C incremental chemical shifts are relative to 7.26 and 128.5 ppm, respectively.

present, the model is implemented partially in computer code and partially in Microsoft Excel.

Technical Background. Whenever a chemical substituent is attached to an aromatic ring, the distribution of electron density on the aromatic ring is perturbed in a systematic, predictable manner. Rates of reactions, chemical equilibria, and even NMR chemical shifts respond in a predictable manner to the perturbation caused by a substituent. The oldest quantitative expression of the effects of chemical substituents on the properties of aromatic rings is the Hammett equation[6]—the predecessor of all other linear free energy relationships and quantitative structure−activity relationships.

In the case of NMR spectrometry, the effect of a chemical substituent on the chemical shift of ^{1}H and ^{13}C is dependent on both the nature of the substituent and its position (ortho, meta, or para) relative to the ^{1}H or ^{13}C atom. The incremental chemical shifts of many common substituents have been compiled, after analyzing the effects of those substituents on the chemical shifts of ^{1}H and ^{13}C in a large number of aromatic compounds. In this paper, several "representative" substituents have been chosen and their incremental chemical shifts have been used throughout the paper.[7]

NOM contains primarily C, H, and O, so substituents containing only those elements were considered. Naturally, one of the substituents must be H. Saturated and unsaturated alkyl side chains are represented by C_2H_5 and $CH{=}CH_2$, respectively. Aromatic carbon can bond directly to the oxygen atoms of OH and OCH_3 groups. Finally, the variety of carbonyl-containing groups that may exist in NOM is represented by COOH, CO_2CH_3, and COC_2H_5. This group of substituents collectively reflects the range of possibilities for polar, mesomeric, and steric effects in substituted aromatic compounds. The concept and list of substituents could easily be extended, but this minimal set of substituents is adequate for purposes of developing, testing, and applying SPARIA.

The standard increments of chemical shift for the substituents used in SPARIA are tabulated in Table 1 and plotted as vectors in chemical shift space in Figure 1. In Figure 1, vectors radiate out from the chemical shifts of ^{1}H and ^{13}C in benzene to the calculated chemical shifts of the corresponding monosubstituted benzenes. The practical advantage of grouping the eight substituents into

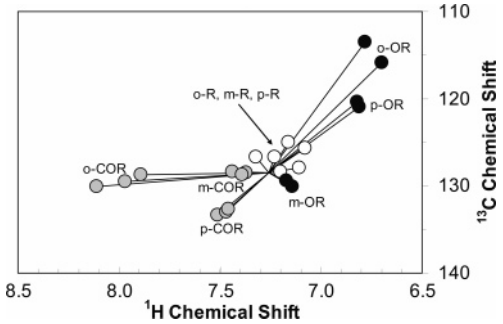


Figure 1. Chemical shifts of monosubstituted benzenes containing the eight substituents in three classes of substituents used in SPARIA.

three classes (COR, R, OR) is immediately evident in Figure 1. As classes of substituents, *o*-COR, *p*-COR, *o*-OR, and *p*-OR have large characteristic effects on the chemical shifts of ^{1}H and ^{13}C. Within any class of substituents at a given ring position, the effects of the individual substituents on chemical shift are very similar. It is also evident that *o*-R, *m*-R, and *p*-R have small and rather similar effects on chemical shift. In addition, the effects of *m*-COR, *m*-R, and *m*-OR are all small and rather similar. It can be anticipated that the inverse mode of SPARIA will more readily detect *o*-COR, *p*-COR, *o*-OR, and *p*-OR than any other combination of classes of substituents and ring positions.

The use of incremental chemical shifts in the forward mode of SPARIA is illustrated in Figure 2 for three substituted benzenes. The chemical shifts of ^{1}H and ^{13}C at the top position on each ring are calculated by starting with the chemical shifts of ^{1}H and ^{13}C in benzene (7.26 and 128.5 ppm, respectively) and adding vectors that represent the incremental chemical shifts of ^{1}H and ^{13}C for the five other substituents. Mathematically, the vectors can be added in any order, but they are consistently added here in a counterclockwise progression around the benzene ring. (A numerical example for this process is given in Table S-1, Supporting Information, for 3,5-dimethoxy-4-hydroxybenzoic acid.) The three structures in Figure 2 are quite different from one another; however, their predicted chemical shifts of ^{1}H and ^{13}C are nearly identical. It follows logically that the existence of a peak in a two-dimensional NMR spectrum at the coordinates of a known substance is a necessary but insufficient condition for making an accurate peak assignment. Figure 2 thus anticipates the need for the inverse mode of SPARIA, which, if searching the rectangular space that is centered on an observed peak, should find these

(6) Hammett, L. P. *Chem. Rev.* **1935**, *17*, 125−136.
(7) Hesse, M.; Meier, H.; Zeeh, B. *Spektroskopische Methoden in der organischen Chemie*; Thieme: Stuttgart, 1991; pp 156−157.

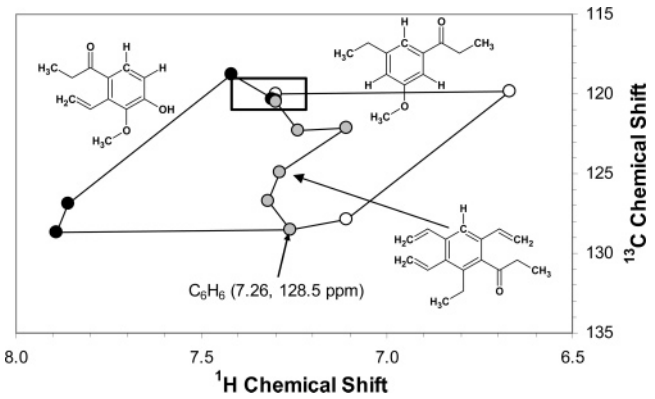


Figure 2. Illustrating the forward and inverse modes of SPARIA.

three and possibly many other structures. Peak assignments can thus only be made in a probabilistic sense.

In the analysis and discussion that follows, it will be necessary to refer specifically to each of the five ring positions on a benzene ring. Because individual aromatic C−H groups are observed in the HSQC NMR experiment, it is logical to use a referencing scheme that is oriented relative to the C−H group that is under observation. Moving counterclockwise around the benzene ring, the ring positions are designated as ortho, meta, para, meta', and ortho' (o-, m-, p-, m'-, and o'). The scheme used here makes unconventional use of the o' and m' symbols (ortho' and meta'), which are commonly used to refer to ring positions on the second benzene ring in molecules that contain two or more benzene rings. Nonetheless, this notation provides a succinct and easily understood means of distinguishing between the two ortho or two meta positions on a single benzene ring, and it will be used in the subsequent sections of this paper.

It will also be necessary to distinguish between the total number of possible substitution patterns and the number of unique substitution patterns on an aromatic ring. For example, if the five ring positions are substituted with all possible combinations of two substituents, then there are 2^5 (32) possible substitution patterns. Because a plane of symmetry and a rotational axis of symmetry pass through the C−H group and the para substituent, the number of unique substitution patterns will be less than the total possible number of possible substitution patterns. Among the 32 substitution patterns are 8 (2^3) patterns that have identical ortho groups and identical meta groups (internal plane of symmetry). The remaining 24 substitution patterns consist of 12 rotationally equivalent pairs. When half of the rotationally equivalent substitution patterns are removed, the number of unique substitution patterns is reduced to 20 ($= 8 + {}^1/_2(32 - 8)$).

Forward Mode of SPARIA—Predicting Chemical Shifts. At present, the algorithm used in the forward mode of SPARIA uses incremental chemical shifts of isolated substituents and does not consider substituent−substituent interactions. Specifically, torsional strain between bulky substituents that are ortho to one another can limit π-bond conjugation between substituents and the aromatic ring, thus reducing the mesomeric effects of those groups. Mesomerically active groups that are ortho or para to one another may interact with each other through the aromatic ring, thus increasing or decreasing the mesomeric effects of those substituents.[8−10] A future version of SPARIA will incorporate these second-order effects and improve the accuracy of the forward mode.

Using standard incremental values for substituent-induced modification of the chemical shifts of aromatic ^{1}H and ^{13}C in Table 1, together with the reference chemical shifts of ^{1}H and ^{13}C in benzene, the chemical shifts of ^{1}H and ^{13}C at an unsubstituted ring position have been predicted for all possible combinations of the eight substituents on the other five ring positions. The total number of combinations is $8^5 = 32\,768$. The forward mode calculations were accomplished with a user-written software called SPARIA.EXE (a very simple computer program that generates the entire database of substitution patterns and chemical shifts of ^{1}H and ^{13}C is given in Table S-2). The output file contains 32 768 data records, each of which lists a substitution pattern (ortho, meta, para, meta', ortho') and the predicted chemical shifts of ^{1}H and ^{13}C for that substitution pattern. The symmetry of the benzene ring leads to 512 (8^3) patterns having an internal plane of symmetry and 16 128 rotationally equivalent pairs of substitution patterns (${}^1/_2(32\,768 - 512)$). The number of unique substitution patterns is predicted to be 16 640 (512 + 16 128). To locate and remove the redundant substitution patterns, the output file was imported into a Microsoft Excel spreadsheet, and each of the 32 768 substitution patterns was processed as follows:

1. Whenever the ortho' group had a higher alphabetical rank than the ortho group, the substitution pattern was not modified.

2. Whenever the ortho' group had a lower alphabetical rank than the ortho group, the entire substitution pattern was rotated 180°.

3. Whenever the ortho' and ortho groups were identical, (a) if the meta' and meta groups were identical or if the meta' group had a higher alphabetical rank than the meta group, the substitution pattern was not modified and (b) if the meta' group had a lower alphabetical rank than the meta group, the entire substitution pattern was rotated 180°.

Once each substitution pattern had been processed as described, the filter tool in Excel was used to extract unique records from the list. The removal of mirror images by this process yielded 16 640 unique substitution patterns. These unique substitution patterns will yield only 10 368 unique peaks, however, because the incremental chemical shifts in Table 1 cannot differentiate between relative orientations of ortho and meta substituents. For example, if the C−H group that is observed by NMR is at ring position 1, increment analysis (as implemented in SPARIA) yields identical predictions for a 2,3-disubstituted compound and a 2,5-disubstituted compound.

The resulting distribution of chemical shifts for ^{1}H and ^{13}C is shown in Figure 3. The eight labeled peaks in Figure 3 are calculated for aromatic compounds containing five identical SPARIA substituents on an aromatic ring. The peak positions of these end members are grouped spatially in the same manner as they are classified in Table 1 and Figure 1, with electron-withdrawing (COR) groups in the lower left region of the plot, neutral substituents (R) near the middle of the plot, and electron-

(8) Thomas, S.; Ströhl, D.; Kleinpeter, E. *J. Chem. Comput. Sci.* **1994**, *34*, 725−729.
(9) Holik, M. *J. Mol. Struct.* **1999**, *482−483*, 347−351.
(10) Meiler, J.; Maier, W.; Will, M.; Meusinger, R. *J. Magn. Reson.* **2002**, *157*, 242−252.

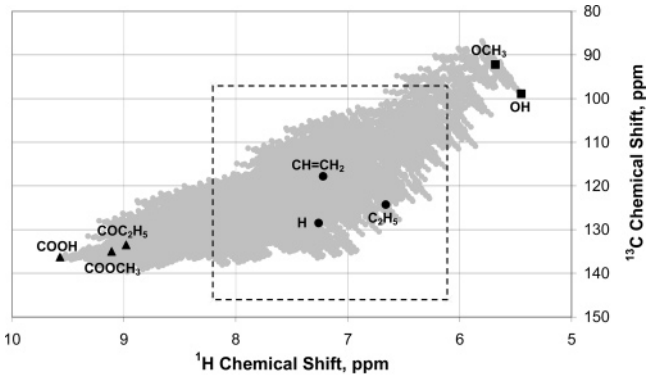


Figure 3. Chemical shifts of HSQC cross-peaks for aromatic C−H in 16 640 unique combinations of eight substituents on the other five ring positions. Labeled peaks correspond to structures in which all five substituents are identical. The dominant cross-peaks for aromatic C−H in natural organic matter generally fall within the rectangular area in this figure.

donating substituents (OR) in the upper right region of the plot. The most generally observed peaks in NOM and related materials lie within the rectangular area in Figure 3.[11−13] Aromatic rings substituted exclusively with COR or OR groups are unlikely to be major constituents of NOM. Aromatic rings in NOM are more likely to contain a mixture of R, COR, and OR groups.

Testing the Accuracy of Predicted Chemical Shifts. The database of substitution patterns and chemical shifts that were generated using the forward mode of SPARIA will be used by the inverse mode to convert measured chemical shifts into probabilistic substitution patterns. The inverse mode thus depends on the accuracy of those predicted chemical shifts. The accuracy of chemical shifts that are obtained using SPARIA in the forward mode was evaluated by comparison with experimental chemical shifts of ^{1}H and ^{13}C in 29 aromatic compounds containing collectively 100 aromatic C−H groups. For comparative purposes, chemical shifts were also predicted using ACD/CNMR Predictor 5.0 and ACD/HNMR Predictor 5.0 from Advanced Chemistry Development, Inc. The structures of the 29 compounds used for this test are given in Figure 4, and their common names are given in Table S-3 (Supporting Information). (The experimental chemical shifts of the 100 aromatic C−H groups are given in Figure S-1.)

Collectively, the 29 compounds used for this test contain most of the substituents that are used in SPARIA, and they are often invoked as likely structural subunits in natural organic matter. This concept and list of compounds could easily be extended for a more rigorous test of SPARIA. Surrogate structures containing only the eight SPARIA substituents were used to approximate the actual structures of more than half of the compounds in Figure 4. (Examples that illustrate the differences between actual structures and SPARIA surrogate structures are given in Figure S-2.) In contrast, the ACD software uses the correct structure of a compound when calculating chemical shifts of ^{1}H and ^{13}C. Clearly, both the use of incremental chemical shifts and the use of surrogate substituents are evaluated by the test of the accuracy of the forward mode of SPARIA. For this reason, SPARIA is not likely to match the performance of ACD software.

Plots of predictions of the SPARIA and ACD models versus experimental data for ^{13}C and ^{1}H chemical shifts are shown in Figure 5. There is considerably greater relative scatter in the prediction of the chemical shifts of ^{1}H than of ^{13}C. The accuracy of the predictions of chemical shift was quantified as the root-mean-square error (RMSE) between experimental and calculated chemical shifts and by linear regression of predicted chemical shift against experimental chemical shift. Results for these assessments of the accuracy of the models are given in Table 2. As anticipated, the predictions of the commercial software from ACD are somewhat better than those of the forward mode of SPARIA, but not greatly so.

In the course of using SPARIA in the forward mode, it was found that predictions of the chemical shift of ^{1}H were often poor in molecules containing COOH groups that are ortho to each other (see **9**, **16**, and **29** in Figure 4). The problem arises in the use of standard incremental chemical shifts for both COOH groups when steric crowding may force one or both of them out of coplanarity with the benzene ring. This torsional strain diminishes the mesomeric interaction between the COOH group and the benzene ring. Similarly, it was observed that predictions of the chemical shift of ^{13}C were poor in structures containing the CH=CH−COOH group (see **3**, **12**, **13**, **14**, **19**, and **23** in Figure 4). In this instance, the problem may lie in the use of the vinyl group as a surrogate for this more complex, conjugated side chain. Future refinements of the forward mode of SPARIA will address these issues, but for now, only the standard incremental chemical shifts in Table 1 are used in this paper. (Optimized incremental chemical shifts for these substituents are given in Table S-4, and linear regression fits and RMSE values using these optimized parameters are given in Table S-5.)

The forward mode of SPARIA is not particularly innovative, but it is an essential tool for rapidly generating the database of 16 640 unique substitution patterns that are required to implement the much more innovative inverse mode of SPARIA. For the compounds and peaks used in this analysis, the forward mode of SPARIA performs surprisingly well, even with the use of surrogate substituents and without optimization of standard incremental chemical shifts, in comparison with state-of-the-science software such as ACD/CNMR Predictor 5.0 (2001) and ACD/HNMR Predictor 5.0 (2001) from Advanced Chemistry Development, Inc. The RMSEs for predicted chemical shifts of ^{1}H and ^{13}C using SPARIA are both within a factor of 1.5 of the RMSEs obtained when the ACD software is used.

Inverse Mode of SPARIA—Predicting Substitution Patterns. The inverse mode of SPARIA is used to predict substitution patterns from the chemical shifts of aromatic ^{1}H and ^{13}C. This is accomplished by searching the database of substitution patterns that are generated by the forward mode of SPARIA for all substitution patterns whose predicted chemical shifts of ^{1}H and ^{13}C lie within a stipulated range of the observed chemical shifts and then using those data to predict the probability of occurrence of each class of substituent (COR, R, OR) at each of the five ring positions. The inverse mode is illustrated in Table 3 for one of the aromatic C−H groups in 3,4,5-trimethoxyacetophenone (**28**

(11) Hertkorn, N.; Permin, A.; Perminova, I.; Kovalevskii, D.; Yudov, M.; Petrosyan, V.; Kettrup, A. *J. Environ. Qual.* **2002**, *31*, 375−387.
(12) Simpson, A. J.; Kingery, W. L.; Hatcher, P. G. *Environ. Sci. Technol.* **2003**, *37*, 337−342.
(13) Cook, R. L.; McIntyre, D. D.; Langford, C. H.; Vogel, H. J. *Environ. Sci. Technol.* **2003**, *37*, 3935−3944.

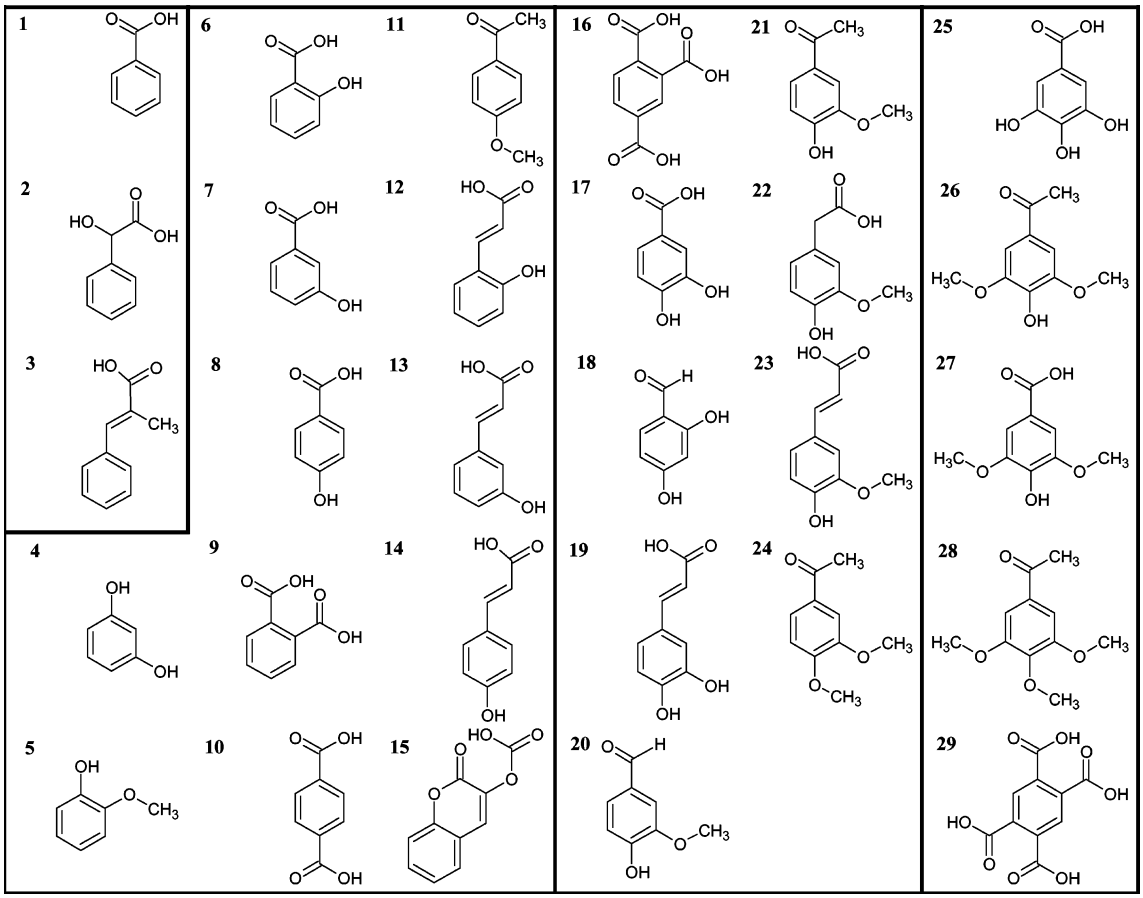


Figure 4. Compounds used to test forward and inverse predictions of the SPARIA model (mono-, di-, tri-, and tetrasubstituted benzenes).

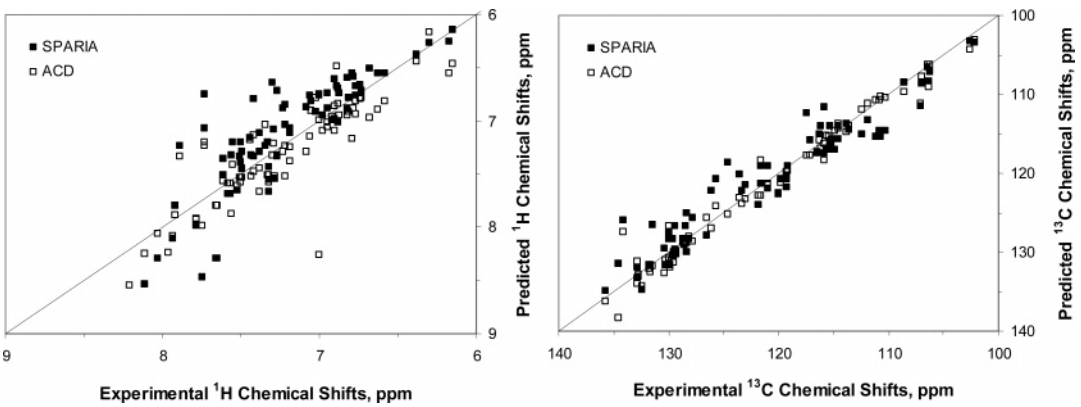


Figure 5. Predicted versus experimental chemical shifts for ^{1}H (left) and ^{13}C (right).

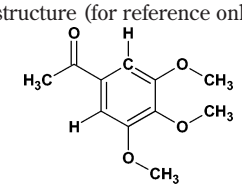
Table 2. Linear Regression Analysis and Root-Mean-Square Error (in ppm) for Chemical Shifts of ^{1}H and ^{13}C in the 29 Compounds in Figure 4

	regression parameters for ^{1}H				regression parameters for ^{13}C			
model	RMSE	intercept	slope	R^2	RMSE	intercept	slope	R^2
ACD	0.23	0.69	0.92	0.79	1.52	2.32	0.99	0.98
SPARIA	0.35	−1.55	1.21	0.73	2.27	13.27	0.89	0.94

in Figure 4), using an arbitrary window size of ±0.1 ppm for ^{1}H and ±1.0 ppm for ^{13}C. Using the experimentally determined chemical shifts of ^{1}H and ^{13}C for this compound (7.23 and 106.4 ppm, respectively), together with the aforementioned window size, the target window is 7.13−7.33 by 105.4−107.4 ppm. The inverse mode of SPARIA finds only 80 substitution patterns (from the total of 16 640) for which predicted chemical shifts fall within the target window. Ironically, the SPARIA substitution pattern that most closely matches the chemical structure of 3,4,5-trimethoxyacetophenone (COC_2H_5, H, OCH_3, OCH_3, OCH_3) is not among

Table 3. Inverse Mode of Sparia (Illustrated Using 11 of 80 Matching Substitution Patterns for an Aromatic C−H Group in 3,4,5-trimethoxyacetophenone (*28*)

structure (for reference only)



	observed peak, ppm		peak window, ppm				
	δ ^{1}H	δ ^{13}C	$\Delta(\delta$ ^{1}H)	$\Delta(\delta$ ^{13}C)			
input to SPARIA	7.23	106.40	0.10	1.00			
	calculated peak, ppm		substitution pattern used for the calculation				
output from SPARIA	δ ^{1}H	δ ^{13}C	ortho	meta	para	meta'	ortho'
	7.13	106.9	COOH	C_2H_5	OCH_3	H	OCH_3
	7.13	106.9	COOH	H	OCH_3	C_2H_5	OCH_3
	7.17	106.2	$COOCH_3$	C_2H_5	OCH_3	COOH	OCH_3
	7.17	106.2	$COOCH_3$	COOH	OCH_3	C_2H_5	OCH_3
	7.21	105.8	COC_2H_5	COC_2H_5	OCH_3	$COOCH_3$	OCH_3
	7.21	105.8	COC_2H_5	$COOCH_3$	OCH_3	COC_2H_5	OCH_3
	7.25	106.6	COC_2H_5	$CH{=}CH_2$	$CH{=}CH_2$	$CH{=}CH_2$	OCH_3
	7.28	105.9	COOH	$CH{=}CH_2$	OH	COC_2H_5	OCH_3
	7.28	105.9	COOH	COC_2H_5	OH	$CH{=}CH_2$	OCH_3
	7.31	106.8	COOH	C_2H_5	OCH_3	COOH	OCH_3
	7.31	106.8	COOH	COOH	OCH_3	C_2H_5	OCH_3
			probability of occurrence by class of substituents and position				
calculated probability[a]		class	ortho	neta	para	meta'	ortho'
		COR	1.00	0.45	0.00	0.45	0.00
		R	0.00	0.55	0.09	0.55	0.00
		OR	0.00	0.00	0.91	0.00	1.00
			error of the prediction by class of substituents and position				
calculated error[b]		class	ortho	meta	para	meta'	ortho'
		COR	0.00	0.45	0.00	0.45	0.00
		R	0.00	−0.45	0.09	0.55	0.00
RMSE = 0.36[c]		OR	0.00	0.00	−0.09	−1.00	0.00

[a] Probability is calculated as the number of occurrences of each class of substituents (COR, R, OR) divided by the total number of hits (e.g., the probability of having OR group in the para position is 10/11 = 0.91). [b] Error is calculated as the difference between calculated and actual probability (e.g., the error for OR groups in the para position is (0.91−1.00 = −0.09). [c] The RMSE is calculated from the 15 separate errors for probability of occurrence of each class of substituents at each ring position.

the 80 solutions obtained from the inverse mode of SPARIA. A forward mode calculation for that substitution pattern yields predicted chemical shifts of 6.88 and 106.5 ppm for ^{1}H and ^{13}C, respectively, so the predicted ^{1}H chemical shift lies slightly outside the target window. This outcome illustrates the consequences of cross-ring mesomeric interactions between the OCH_3 and $COCH_3$ groups that are para to each other—an interaction that is not considered in the current implementation of the forward mode of SPARIA.

A subset of 11 representative substitution patterns is given in Table 3, which includes the probability of occurrence of each class of substituents (COR, R, OR) on each of the five ring positions and an error analysis of the predictions. The *calculated* probability of occurrence is assumed to equal the frequency of occurrence of that class of substituents at that ring position. The *actual* probability of occurrence, which can be determined from the known structure, is either 0 or 1 for each of the three classes of substituents at each of the five ring positions. The error of each prediction is calculated as the difference between the calculated probability and the actual probability. The overall error in the predicted substitution pattern is expressed as the root-mean-square of those 15 separate errors (RMSE). (The statistical summary for all 80 "hits" is given in Table S-6).

The inverse mode of SPARIA predicts a 100% probability of a COR group at the ortho position and an OR group at the ortho' position, and it predicts a 91% probability of an OR group at the para position. These predictions match the substitution pattern of 3,4,5-trimethoxyacetophenone. Both R and COR groups have relatively high probabilities of occurrence at both of the meta positions, but R groups are slightly more probable. This prediction is correct for one of the meta positions (H), but it is incorrect for the other (OCH_3). Although it is not the goal of the inverse mode to predict the actual substituent at each ring position, the most likely substituent at the ortho' and para positions is correctly predicted to be OCH_3.

The relatively poor quality of predicted substitution patterns in the meta position is inherent in the relative insensitivity of the NMR peak positions of ^{1}H and ^{13}C to meta substituents. In Table 1, for example, the entire range of ^{1}H incremental chemical shifts for meta substituents is only 0.3 ppm, which is slightly larger than the width of the target window used in this example. The corresponding range for ^{13}C is only 1.9 ppm, which is actually smaller than the height of the target window. This means that many combinations of meta substituents can lead to peaks that fall within the target window, making it rather difficult to obtain accurate predictions. In contrast, the entire range of ^{1}H incre-

mental chemical shifts for ortho substituents is 1.4 ppm, and the corresponding range for ^{13}C is 16.6 ppm, both of which are 7−8 times larger than the window size used here. For para substituents, the corresponding ranges of chemical shift are 0.7 and 12.9 ppm, respectively, which are about 3−6 times greater than the window size used here. For both ortho and para substituents, only limited combinations of substituents can lead to peaks that fall within the target window, and the predictions for those positions are thus more reliable.

Although the substitution patterns in Table 3 exhibit a strong preference for substituents in one ortho position versus the other, no such preference is found for the meta positions. In both cases, the effect of a given substituent on chemical shift is the same from either ortho position or from either meta position. The strong positional preference of ortho substituents in Table 3 is a direct consequence of the process by which redundant substitution patterns were removed from the original database of 32 768 substitution patterns to obtain the final list of 16 640 unique substitution patterns (see earlier discussion).

Optimizing Window Size in the Inverse Mode of SPARIA. In the preceding example, an arbitrary window size of ±0.1 ppm for ^{1}H and ±1.0 ppm for ^{13}C was used. If the size of the window approaches zero in both dimensions, it is unlikely that any predicted peak will fall within that window. Conversely, if the window approaches the full range of possible chemical shifts for aromatic ^{1}H and ^{13}C, then it is likely that all possible peaks will fall within that window. In either extreme case, no useful information can be obtained. The size of the window must therefore be optimized to obtain a statistically reasonable number of possible substitution patterns and to yield results of the greatest possible accuracy. To optimize the standard window size for the inverse mode of SPARIA, 10 of the 100 peaks for the 29 compounds in Figure 4 were selected to provide a variety of substitution patterns and a broad range of chemical shift for both ^{1}H and ^{13}C (see Figure S-1). The half-width of the chemical shift window for ^{13}C ($\Delta(\delta\ ^{13}C)$) was varied from 0.25 to 5.00 ppm, and $\Delta(\delta\ ^{1}H)$ was fixed at $\Delta(\delta\ ^{13}C)/10$. For each window size, the inverse mode of SPARIA was used to generate a probabilistic substitution pattern and table of calculated errors for each peak. The overall error at a given window size is calculated as the RMSE for 150 separate errors (3 classes of substituents, 5 ring positions, and 10 peaks).

A plot of the overall RMSE for the 10 selected peaks versus $\Delta(\delta\ ^{13}C)$ is given in Figure 6. (A similar plot for all 10 peaks used to optimize the size of the target window is given in Figure S-3.) At very small half-widths of the $\delta\ ^{13}C$ window, the number of likely substitution patterns is too small to yield reliable statistics. This problem can be overcome by increasing $\Delta(\delta\ ^{13}C)$; however, the discriminating power of SPARIA gradually declines as $\Delta(\delta\ ^{13}C)$ increases. The minimum RMSE occurs at a $\Delta(\delta\ ^{13}C)$ value of 1 ppm. The corresponding standard value of $\Delta(\delta\ ^{1}H)$ is 0.1 ppm. These values will be used henceforth as the standard window size for the inverse mode of SPARIA. These window sizes are slightly smaller than the RMSE values for predicted chemical shifts of ^{1}H (0.35 ppm) and ^{13}C (2.27 ppm), which were obtained using the forward mode of SPARIA with standard incremental chemical shifts (see Table 2).

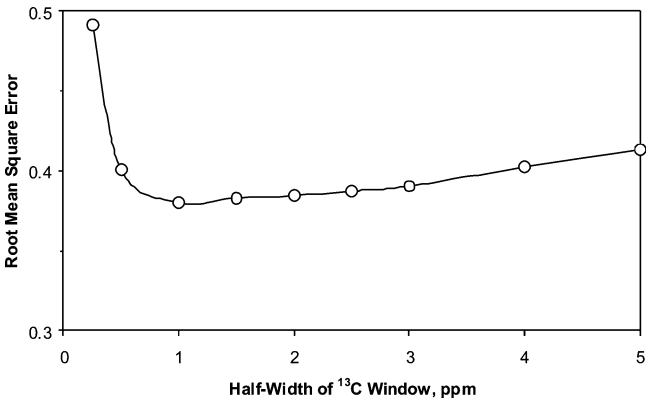


Figure 6. Error analysis of predicted substitution patterns versus half-width of the chemical shift window of ^{13}C ($\Delta(\delta\ ^{13}C$).

Testing the Inverse Mode of SPARIA. Having selected standard half-widths of 1.0 and 0.1 ppm for ^{13}C and ^{1}H, respectively, the chemical shifts of the 100 peaks for the 29 compounds in Table 3 were used to test the predictive capabilities of the inverse mode of SPARIA. Starting with a given pair of chemical shifts for aromatic ^{1}H and ^{13}C, the entire database of 16 640 unique substitution patterns was searched for those substitution patterns whose chemical shifts for aromatic ^{1}H and ^{13}C were within the specified target window of chemical shift centered around the given peak position. The number of matching substitution patterns ranged from 4 to 285 and averaged 109 for the 100 peaks that were analyzed. The most probable fractions of R, COR, and OR groups were calculated for the ortho, meta, para, meta', and ortho' positions on the ring containing the aromatic C−H group whose peak was being analyzed. The predicted distribution was then compared with the known distribution of these classes of substituents in the compound whose peak was being analyzed.

For each predicted substitution pattern, the error of the prediction was calculated at each of the five ring positions for each of the three classes of substituents. The cumulative distributions of errors for the three classes of groups at the ortho, meta, and para positions are given in Figure 7. (A more comprehensive analysis of the distribution of errors from the inverse mode of SPARIA is given in Figure S-4.) Data for ortho and ortho' positions have been combined, as have data for meta and meta' positions. All classes of substituents, but most especially OR, are very well predicted at ortho positions. Furthermore, errors are more-or-less symmetrically distributed around zero. For the para position, the occurrence of OR groups is generally well predicted and errors are symmetrical around zero. There is a strong tendency, however, for R groups to be underestimated and for COR groups to be correspondingly overestimated.

For OR, COR, and R groups in meta positions, the predictions of SPARIA are generally unreliable. The very narrow range of incremental chemical shifts for both ^{1}H and ^{13}C, relative to the size of the standard target window used in the inverse mode of SPARIA, and its impact on the predictions of the inverse mode of SPARIA have already been discussed. If the predictions of substitution patterns in the meta positions were entirely random, the probabilities of occurrence of OR, COR, and R groups would be 2/8, 3/8, and 3/8, respectively (see Table 1). Given that the actual abundance of a particular class of substituent at any ring position is either zero or one, random SPARIA errors for OR

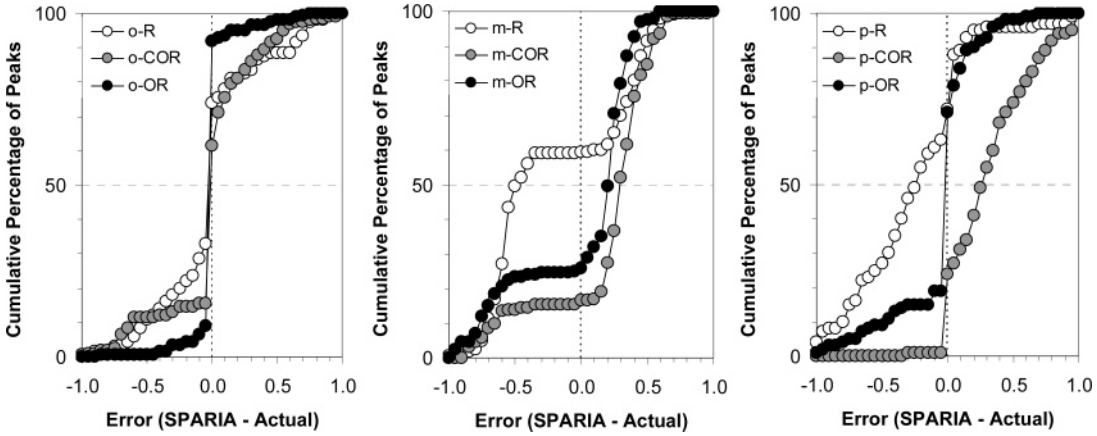


Figure 7. Overall summary of the performance of the inverse mode of SPARIA for three classes of substituents at ortho, meta, and para ring positions.

groups would cluster around −0.750 and +0.250. For COR and R groups, random SPARIA errors would cluster around −0.625 and +0.375. In the cumulative frequency distribution of errors for predicted substitution patterns at the meta positions in Figure 7, the most frequent errors for the OR substituents occur near −0.750 and +0.250. The most frequent errors for COR and R substituents are clustered around −0.7 to −0.6 and +0.3 to +0.4. It thus appears that SPARIA predictions of substitution patterns at meta positions are very nearly random.

It is also possible to recast the error analysis using the simple rule that the most probable class of substituents at a ring position is assigned a probability of 1 and the other two classes of substituents are assigned a probability of 0. Then the predictions of the inverse mode of SPARIA can be classified as follows: if error = −1, false negative; if error = 0, correct; if error = 1, false positive.

Using this type of analysis of errors, the overall percentages of correct predictions from the inverse mode of SPARIA are 85, 60, and 77%, respectively, at the ortho, meta, and para positions (see Figure S-5). When all such results are combined, the global percentage of correct predictions is 74%. When the inverse mode of SPARIA is applied to natural organic matter (next section), it is anticipated that predictions of substitution patterns will be more accurate for ortho and para substituents than for meta substituents. Having said that, it remains the case that 74% of the predictions of the presence or absence of OR groups in both meta positions were correct, equaling the overall performance of the inverse mode of SPARIA for all classes of substituents and ring positions.

The correct prediction by the inverse mode of SPARIA of 74% of the substitution patterns for the three classes of substituents at five ring positions is a significant development that makes possible a deeper, probabilistic interpretation of the heteronuclear two-dimensional ^{1}H, ^{13}C NMR spectra of complex natural samples such as NOM. It is fortuitous that, among all possible substitution patterns, aromatic C−H groups having only meta substituents occur only in monosubstituted and 1,3-disubstituted benzenes (**1**, **2**, **3**, **4**, **7**, and **13** in Figure 4). So, in most instances, aromatic C−H groups will have ortho and/or para substituents that can be more accurately predicted by the inverse mode of SPARIA. The situation is further aided by the fact that aromatic C−H groups having para substituents but no ortho substituents occur only in monosubstituted, 1,2-disubstituted, and 1,2,3-trisubstituted benzenes (**1**, **2**, **3**, **5**, **6**, **9**, **12**, and **15** in Figure 4). For all other substitution patterns, an aromatic C−H group will have one or more ortho substituents. The very strong effect of ortho substituents on chemical shift (see Table 1) and their near-ubiquitous occurrence in substituted benzenes accounts largely for the predictive capability of the inverse mode of SPARIA.

Application of SPARIA to NOM. The inverse mode of SPARIA has been applied to a reference sample of NOM from the International Humic Substances Society (sample ID 1R101N). A team led by one of the authors (E.M.P.) collected this sample by reverse osmosis from the Suwannee River in southeastern Georgia, in 1999. The dry, ash-free elemental composition of this NOM is as follows: 52.47% C, 4.19% H, 42.69% O, 1.10% N, and 0.65% S.

The HSQC NMR spectrum that is analyzed in this paper has been acquired at 303 K with a Bruker DMX 500 spectrometer from 99.6 mg of Suwannee River NOM (IHSS material 1R101N), dissolved in 750 μL of DMSO-d_6 (reference: 2.49/39.50 ppm for ^{1}H/^{13}C, respectively) under total exclusion of moisture, using vacuum line techniques and a sealed 5-mm tube. A 5-mm dual ^{13}C, ^{1}H cryogenic probe using 90° excitation pulses (90°(^{1}H) = 10.0 μs; 90°(^{13}C) = 11.5 μs) was employed for the acquisition of the carbon decoupled ^{1}H,^{13}C-HSQC NMR spectrum under the following conditions: ^{13}C-90-deg decoupling pulse, GARP (70 μs); F2 (^{1}H), acquisition time, 250 ms at a spectral width of 7002 Hz, ^{1}J(CH) = 150 Hz, 1.25-s relaxation delay; F1 (^{13}C), SW = 22 014 Hz (175 ppm); number of scans(F2)/F1-increments (^{13}C frequency), 512/512. The HSQC spectrum was computed to a 4096 × 512 matrix with exponential line broadening of 15 Hz in F2 and a shifted sine bell (π/4) in F1. A gradient, but not sensitivity-enhanced sequence (echo−antiecho; *z*-gradient: 1-ms length, 250-μs recovery) was used.

In Figure 8, the aromatic region of the NMR spectrum is superimposed on the pattern of 16 640 peaks that were predicted previously using the forward mode of SPARIA for eight substituents on five ring positions. The degree of overlap is impressive, and this image affirms quite clearly and convincingly that NOM is a very complex mixture. The locus of peak intensities covers most of the aromatic region of the spectrum. Only the regions corresponding to ring substitution patterns consisting almost entirely of COR groups or of OR groups are not well represented

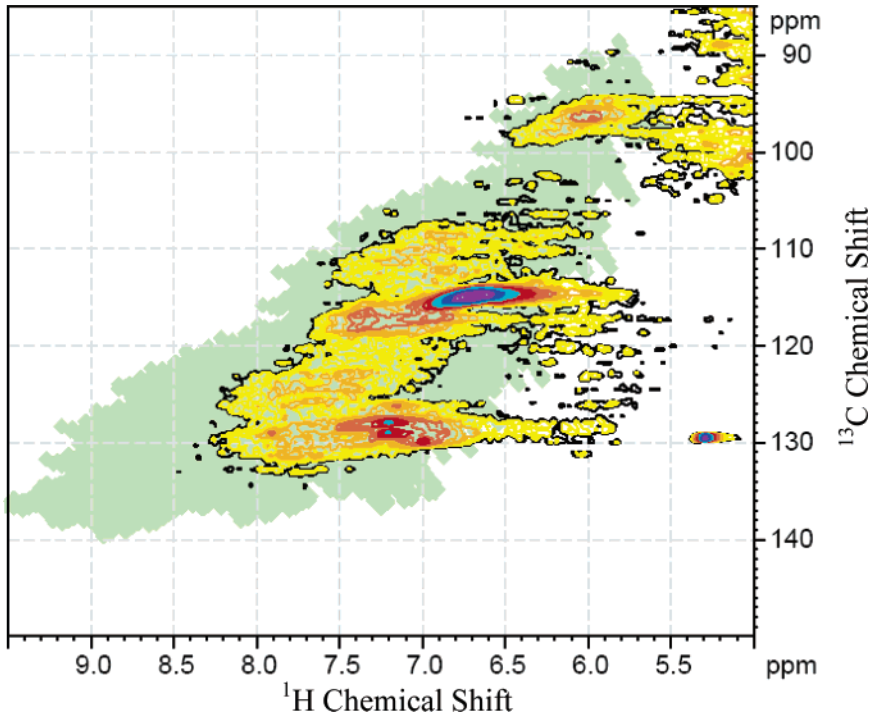


Figure 8. Aromatic region of the ^{1}H, ^{13}C-HSQC NMR spectrum of Suwannee River NOM (IHSS ID 1R101N)—99.5 mg in 750 μL of DMSO-d_6, superimposed on the 16 640 peak positions that are predicted by SPARIA for eight substituents in five ring positions.

Table 4. Predicted Distribution of Neutral, Carbonyl, and Oxygen Groups in Sixteen Prominent Peaks of the ^{1}H, ^{13}C HSQC NMR Spectrum of Suwannee River NOM in DMSO-d_6[a]

				ortho			meta			para			meta'			ortho'		
peak no.	δ ^{1}H	δ ^{13}C	hits	R	COR	OR	R	COR	OR	R	COR	OR	R	COR	OR	R	COR	OR
1	6.63	114.5	77	1.00	0.00	0.00	0.34	0.29	0.38	0.97	0.00	0.03	0.34	0.29	0.38	0.03	0.00	0.97
2	7.20	127.7	60	1.00	0.00	0.00	0.40	0.37	0.23	0.87	0.13	0.00	0.38	0.30	0.32	1.00	0.00	0.00
3	7.21	128.8	52	1.00	0.00	0.00	0.48	0.27	0.25	0.71	0.29	0.00	0.40	0.29	0.31	1.00	0.00	0.00
4	6.99	129.8	19	1.00	0.00	0.00	0.37	0.16	0.47	0.84	0.16	0.00	0.26	0.00	0.74	1.00	0.00	0.00
5	5.97	96.1	30	0.00	0.00	1.00	0.53	0.40	0.07	0.60	0.00	0.40	0.40	0.37	0.23	0.00	0.00	1.00
6	6.92	108.3	120	0.23	0.78	0.00	0.44	0.33	0.23	0.23	0.00	0.78	0.44	0.33	0.23	0.00	0.00	1.00
7	7.39	112.4	139	0.00	1.00	0.00	0.42	0.35	0.22	1.00	0.00	0.00	0.42	0.35	0.22	0.00	0.00	1.00
8	6.94	116.4	153	0.95	0.05	0.00	0.63	0.23	0.14	0.11	0.69	0.20	0.59	0.26	0.14	0.20	0.00	0.80
9	6.99	117.1	186	0.97	0.03	0.00	0.47	0.38	0.15	0.03	0.77	0.19	0.46	0.37	0.17	0.19	0.00	0.81
10	7.29	117.9	112	0.75	0.25	0.00	0.24	0.61	0.15	0.23	0.49	0.28	0.24	0.61	0.15	0.12	0.17	0.71
11	7.47	122.6	220	0.45	0.55	0.00	0.27	0.27	0.46	0.15	0.25	0.60	0.26	0.27	0.46	0.40	0.34	0.25
12	7.70	125.4	147	0.81	0.19	0.00	0.50	0.26	0.24	1.00	0.00	0.00	0.50	0.26	0.24	0.23	0.77	0.00
13	7.77	128.4	151	0.66	0.34	0.00	0.37	0.25	0.38	0.85	0.15	0.00	0.34	0.27	0.38	0.48	0.52	0.00
14	7.90	128.8	144	0.64	0.36	0.00	0.33	0.39	0.28	0.85	0.15	0.00	0.33	0.38	0.28	0.47	0.53	0.00
15	7.47	128.1	73	0.90	0.10	0.00	0.44	0.30	0.26	0.52	0.48	0.00	0.41	0.27	0.32	0.82	0.18	0.00
16	7.76	130.8	99	0.85	0.15	0.00	0.22	0.53	0.25	0.36	0.64	0.00	0.23	0.46	0.30	0.71	0.29	0.00

[a] See text for details.

in NOM. Some relatively intense peaks do not overlie the peaks associated with substitution patterns on benzene rings, and those peaks, which may be attributed to heterocyclic aromatic rings, the anomeric C−H in sugars, etc., cannot be analyzed by the inverse mode of SPARIA.

Even before proceeding with application of the inverse mode of SPARIA to this spectrum, it is evident that a very large number of substitution patterns would be required to account for all the peak intensity in this figure. The most probable substitution patterns for the aromatic moieties giving rise to 16 prominent peaks in the spectrum of Suwannee River NOM have been generated using the inverse mode of SPARIA. Collectively, the selected peaks span a range of 5.97−7.90 ppm for ^{1}H chemical shift and a range of 96.1−130.8 ppm in ^{13}C chemical shift.

Starting with a given pair of chemical shifts for aromatic ^{1}H and ^{13}C, the entire database of 16 640 unique substitution patterns for eight substituents on five ring positions was searched for those substitution patterns whose chemical shifts for aromatic ^{1}H and ^{13}C were within the standard window of chemical shift for the inverse mode of SPARIA ($\Delta(\delta\ ^{13}C) = 1$ ppm, $\Delta(\delta\ ^{1}H) = 0.1$ ppm) centered around the given peak position. The matching substitution patterns (hits) were processed as described earlier in Table 3 to calculate the average fractions of COR, R, and OR substituents for the ortho, meta, para, meta', and ortho' positions on the ring

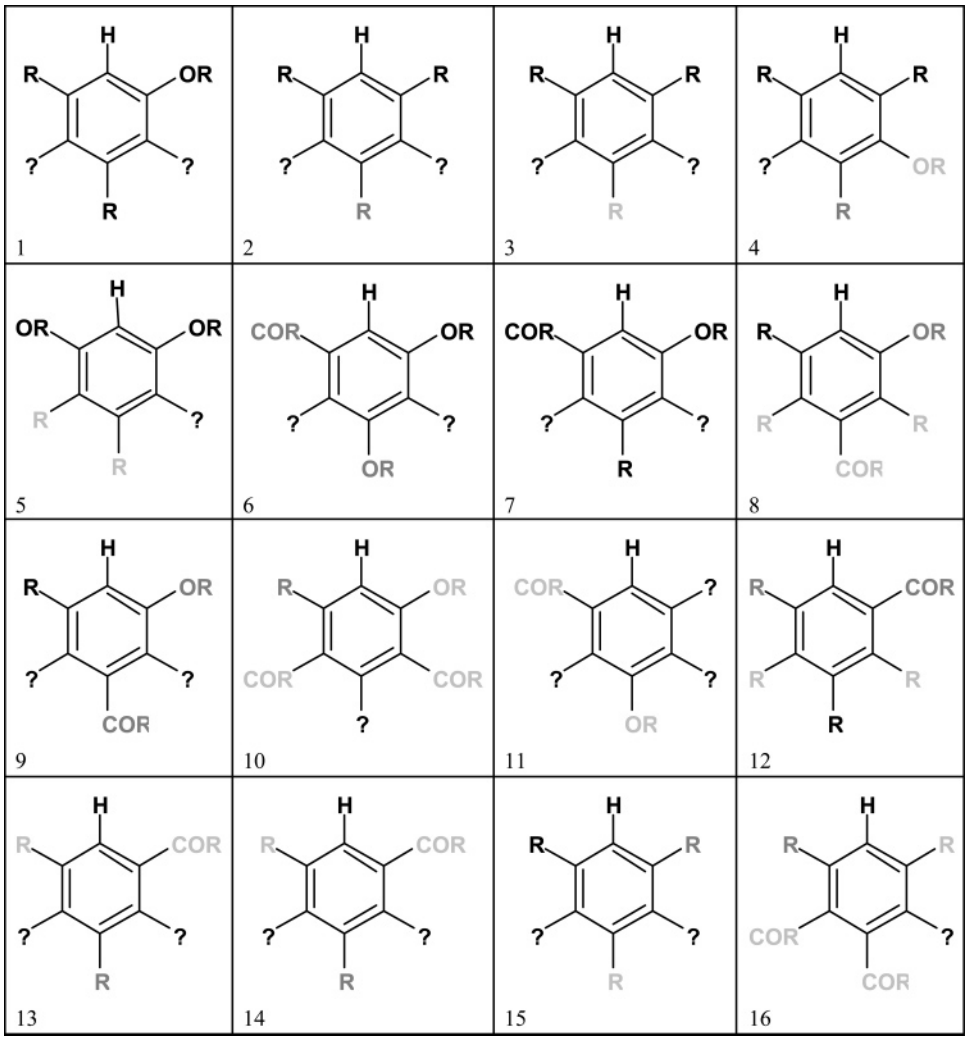


Figure 9. Most probable substitution patterns for the 16 most intense peaks in the HSQC spectrum of Suwannee River NOM (R denotes electron-neutral carbon substituents and hydrogen).

containing the aromatic C−H group whose peak was being analyzed. The resulting probabilistic substitution pattern is given in Table 4 for each of the selected peaks.

For the selected peaks, the number of hits ranged from 19 to 220 and averaged 111 hits per peak. These results are comparable to those obtained during the previously discussed test of the inverse mode of SPARIA, for which the number of hits ranged from 4 to 285 and averaged 109. It is thus anticipated that the inverse mode of SPARIA will perform comparably on the test peaks and the actual peaks in the spectrum of Suwannee River NOM.

In Table 4, the probabilities of occurrence of the three classes of substituents are calculated at five ring positions in 16 structures. An error analysis such as that which was conducted during the test of the inverse mode of SPARIA is impossible here, because the actual substitution patterns are unknown. The corresponding predicted most probable substitution patterns are shown in Figure 9. The peak being modeled arises from the C and H at the top of each structure. Ring positions are labeled in black, medium gray, light gray, or with the "?" character to indicate the most probable class of substituents and its probability of occurrence at each ring position. The code is as follows: 90% or greater probability (black); 75%−90% probability (medium gray); 50%−75% probability (light gray); less than 50% probability ("?").

Consistent with earlier testing of the inverse mode of SPARIA, predictions are much more robust at ortho, para, and ortho' positions. Considering only the 48 predictions of substitution patterns at those positions, 40% are made at a probability level of 0.90−1.00, 29% are made at a probability level of 0.75−0.89, 33% are made at a probability level of 0.50−0.74, and 4% are made at a probability level of less than 0.50. For the remaining 32 predictions at the meta and meta' positions, no predictions at all could be made at a probability level of 0.75 or higher, 28% are made at a probability level of 0.50−0.74, and 72% are made at a probability level of less than 0.50.

Having obtained most probable substitution patterns from the inverse mode of SPARIA, those patterns can be compared with known compounds and structural moieties to provide evidence for or against their presence in the NOM sample. As a simple illustration, the substitution patterns in Figure 9 can be compared with the known substitution patterns of lignin-derived phenols.[14] In this comparison, "?" may represent either R, COR, or OR, and R represents not only C_2H_5 and $CH{=}CH_2$ but also H (an

(14) Hedges, J. I.; Ertel, J. R. *Anal. Chem.* **1982**, *54*, 174−178.

unsubstituted position). The three main phenols used in the biosynthesis of lignins are as follows:[15,16] 4-(3-hydroxy-1-propenyl)-phenol (coumaryl alcohol); 4-(3-hydroxy-1-propenyl)-2-methoxy-phenol (coniferyl alcohol); 4-(3-hydroxy-1-propenyl)-2,6-dimethoxy-phenol (sinapyl alcohol).

During early diagenesis, the 3-hydroxy-1-propenyl side chain is typically oxidized and cleaved to a shorter side chain; however, the OCH_3 groups on the benzene ring are more resistant to modification. Assuming this behavior, then structural moieties that are derived from coumaryl alcohol must have only one OR group, and they must have only one other substituent—either an R or COR group that is para to the OR group. Possible matches are found in **1**−**4** and **13**−**15**. Structural moieties that are derived from coniferyl alcohol must have two adjacent OR groups, one of which is para to either an R or COR group, and this pattern is only possible in **1**, **7**, and **11**. Structural moieties that are derived from sinapyl alcohol must have three adjacent OR groups, the middle one of which is para to either an R or COR group, and this pattern is only possible in **6** and **11**.

Even though there is a considerable flexibility within these general structural requirements, six of the substitution patterns in Figure 9 cannot be lignin-derived phenols. Structure **5** contains an OR group on each side of the observed C−H group, structures **8** and **9** contain a COR group para to the observed C−H group, structures **10** and **16** contain two COR groups on the benzene ring, and structure **12** does not contain an OR group at all.

The 16 peaks that were selected for analysis represent only a small percentage of the total peak intensity in the spectrum shown in Figure 8. Even so, it can be concluded that structural moieties that are unrelated to lignin contribute significantly to the aromaticity of NOM. Given the inherently stochastic nature of the inverse mode of SPARIA, a more robust and detailed analysis of substitution patterns of aromatic rings in NOM awaits further refinement of the forward mode of SPARIA, because the ability of the inverse mode to extract more accurate probabilistic substitution patterns from chemical shift data depends ultimately on the accuracy of the chemical shift tables that are generated by the forward mode of SPARIA.

An alternative approach to the inverse problem of generating likely substitution patterns from two-dimensional NMR spectra was published recently by Simpson et al.[17] They matched multiple peaks in COSY spectra of a pine forest fulvic acid with a custom database of chemical shifts (containing COSY and HSQC spectra) for lignin-related fragments that was generated using ACD software. The reliance on a two-step procedure, starting with COSY spectra (restricted to vicinal couplings of adjacent pairs of protons), followed by HSQC data, and the nature of the sample itself (from pine—a woody gymnosperm), implies fragments that were derived from sinapyl alcohol would not have been detected. Of the 20 peaks that were analyzed, three peaks were potentially derived from coniferyl alcohol, 9 were potentially derived from coumaryl alcohol, and 9 were unrelated to known precursors of lignin. This distribution of results is very similar to that obtained when the inverse mode of SPARIA is used to analyze Suwannee River NOM.

CONCLUSIONS

SPARIA was used to generate a database of 16 640 unique substitution patterns from eight substituents on five ring positions. The forward mode of SPARIA was tested on 29 structures containing 100 aromatic H by comparing predicted chemical shifts of ^1H and ^{13}C with actual chemical shifts and with the predictions of optimized commercial software. Even though SPARIA calculates chemical shifts for surrogate structures that contain only its eight substituents, those results are only slightly less accurate than the predictions of commercial software.

In the inverse mode of SPARIA, a target window of ±1 ppm for $\delta^{13}C$ and ±0.1 ppm for δ^1H allows both a focused search for potentially valid substitution patterns and a sufficient number of "hits" for robust statistical analyses. The inverse mode of SPARIA was tested on 100 NMR peaks (^1H and ^{13}C) by comparing predicted substitution patterns with actual structures of 29 compounds for which the peaks were measured. SPARIA predicted the presence or absence of each of three classes of substituents at each of five ring positions. The overall accuracy of the predicted substitution patterns at ortho, meta, and para ring positions was 85, 62, and 77%, respectively, resulting in an overall accuracy of 74%.

The HSQC ^1H, ^{13}C NMR spectrum of Suwannee River NOM contains detectable resonances over much of the chemical shift space that is occupied by the 16 640 unique substitution patterns which were generated by the forward mode of SPARIA, which certainly demonstrates the immense complexity of this material. Sixteen of the more prominent peaks were analyzed using the inverse mode of SPARIA. Assuming that a probability of 50% or greater is likely to be correct, then 1−2 peaks are potentially syringyl phenols, 1−3 peaks are potentially vanillyl phenols, 6−7 additional peaks are potentially lignin-derived, and 6 of the 16 peaks are most probably generated from structural moieties that are not derived from lignin.

SUPPORTING INFORMATION AVAILABLE

Six tables, five figures, and associated discussion are provided to elaborate on the methods used to develop, test, and apply SPARIA. These include the following: Table S-1, forward prediction of the chemical shifts of ^1H and ^{13}C in 3,5-dimethoxy-4-hydroxybenzoic acid using increment analysis. Table S-2, a Pascal program for generating the database of substitution patterns and chemical shifts that are used in SPARIA. Table S-3, compounds used to test forward and inverse predictions of the SPARIA model. Table S-4, incremental chemical shifts for selected substituents on aromatic rings—^1H and ^{13}C incremental chemical shifts are relative to 7.26 and 128.5 ppm, respectively. Table S-5, linear regression analysis and root-mean-square error (in ppm) for chemical shifts of ^1H and ^{13}C in the 29 compounds in Figure 4 and in Table S-3. Table S-6, inverse mode of SPARIA using all 80 matching substitution patterns for an aromatic C−H group in 3,4,5-trimethoxyacetophenone (**28** in Figure 4). Figure S-1, the NMR peaks of 29 compounds used to evaluate the forward and inverse modes of SPARIA (see Figure 4 and Table S-3). Figure S-2, actual and surrogate structures used to test the accuracy of the forward

(15) Freudenberg, K.; Neish, A. C. *Constitution and Biosynthesis of Lignin*; Springer-Verlag: Berlin, 1968.
(16) Sakakibara, A. Chemistry of lignin, In *Wood and Cellulosic Chemistry*; Hon, D. N.-S., Shiraishi, N., Eds.; Marcel Dekker: New York, NY, 1991; Chapter 4, pp 113−175.
(17) Simpson, A. J.; Lefebvre, B.; Moser, A.; Williams, A.; Larin, N.; Kvasha, M.; Kingery, W. L.; Kelleher, B. *Magn. Reson. Chem.* **2004**, *42*, 14−22.

mode in SPARIA. Figure S-3, error analysis of predicted substitution patterns versus half-width of the chemical shift window of ^{13}C ($\Delta(\delta\ ^{13}C)$). Figure S-4, errors in predictions of the inverse mode of SPARIA, including frequency distributions and cumulative distributions of error. Figure S-5, overall summary of the performance of the inverse mode of SPARIA for three classes of substituents at ortho, meta, and para ring positions, when errors are rounded (see text for discussion). Numbers inside the data bars represent the number of correct predictions for a particular class of substituent and ring position (maximum = 100). This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review August 29, 2006. Accepted October 31, 2006.

AC061611Y