Environ. Sci. Technol. **2003**, *37,* 1783−1793

3-D Structural Modeling of Humic Acids through Experimental Characterization, Computer Assisted Structure Elucidation and Atomistic Simulations. 1. Chelsea Soil Humic Acid

MAMADOU S. DIALLO,*,†,‡ ANDRE SIMPSON,§ PAUL GASSMAN,∥ JEAN LOUP FAULON,⊥ JAMES H. JOHNSON, JR.,‡ WILLIAM A. GODDARD III,† AND PATRICK G. HATCHER#

Materials and Process Simulation Center, Beckman Institute 139-74, California Institute of Technology, Pasadena, California 91125, Department of Civil Engineering, Howard University, Washington, D.C., 20059, Department of Physical Sciences, University of Toronto, Scarborough College, 1265 Military Trail, Toronto M1C 1A4, Canada, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352, Computational Biology & Evolutionary Computing Department, Sandia National Laboratories, P.O. Box 969, MS 9951, Livermore, California 94551-0969, and Department of Chemistry, The Ohio State University, Columbus, Ohio 43210

This paper describes an integrated experimental and computational framework for developing 3-D structural models for humic acids (HAs). This approach combines *experimental characterization*, *computer assisted structure elucidation* (CASE), and *atomistic simulations* to generate all 3-D structural models or a representative sample of these models consistent with the analytical data and bulk thermodynamic/structural properties of HAs. To illustrate this methodology, structural data derived from elemental analysis, diffuse reflectance FT-IR spectroscopy, 1-D/2-D ^{1}H and ^{13}C solution NMR spectroscopy, and electrospray ionization quadrupole time-of-flight mass spectrometry (ESI QqTOF MS) are employed as input to the CASE program SIGNATURE to generate all 3-D structural models for Chelsea soil humic acid (HA). These models are subsequently used as starting 3-D structures to carry out constant temperature-constant pressure molecular dynamics simulations to estimate their bulk densities and Hildebrand solubility parameters. Surprisingly, only a few model isomers are found to exhibit molecular compositions and bulk thermodynamic properties consistent with the experimental data. The simulated ^{13}C NMR spectrum of an equimolar mixture of these model isomers compares favorably with the measured spectrum of Chelsea soil HA.

* Corresponding author phone: (626)395-2730; fax: (626)585-0918; e-mail: diallo@wag.caltech.edu and mdiallo@howard.edu. Present address: Materials and Process Simulation Center, Beckman Institute 139-74, California Institute of Technology, Pasadena, CA 91125.

† California Institute of Technology.

‡ Howard University.

§ University of Toronto.

∥ Pacific Northwest National Laboratory.

⊥ Sandia National Laboratories.

The Ohio State University.

Introduction

Humic substances (HS) are the most abundant reservoir of carbon on earth (*1*). Humic acids (HAs), the initial focus of this research, are operationally defined as the fraction of HS that is insoluble in water at low pH (<2) and soluble at higher pH (>2). They have a significant impact on a variety of biogechemical and environmental processes (*1*−*10*). They act as (i) soil stabilizers, (ii) nutrient and water reservoirs for plants, (iii) sorbents for toxic metal ions, radionuclides, and organic pollutants, and (iv) chemical buffers with catalytic activity. Because HAs are operationally defined through their solubility in aqueous solutions, the development of 2-D and 3-D structural models for these compounds has remained an outstanding problem in environmental chemistry, soil chemistry, and organic geochemistry (*1*−*14*). The conventional approach is commonly used to elucidate the structure of an unknown compound. With the conventional approach, a structural model is inferred from a set of analytical data through a repetitive trial-and-error process that consists of matching the postulated structure with the available analytical data. Over the last two decades, several investigators have used this approach to generate 2-D structural models for HAs. Stevenson (*1*) has proposed a 2-D model that incorporates many of the key structural features for a "typical" soil HA. Schluten and Schnitzer (*11*) have combined elemental analysis, ^{13}C NMR, pyrolysis mass spectrometry and oxidative degradation data to develop a "state-of-the art structural concept" for a soil HA. A number of investigators have also used the conventional approach to develop structural building blocks for HAs. Steelink (*12*) combined elemental analysis data with titration data to develop his HA building block. Jansen and co-workers (*13*, *14*) have derived a structural HA building block that incorporates "more fully the results of experimental data and retro-biosynthetic analyses". The proposed Temple-Northeastern-Birmingham (TNB) HA building block is a modified version of Steelink's HA "monomer". There are, however, two major problems associated with this conventional approach to HA model building. First, the process is carried out manually in most cases; thus, it is time-consuming and not very reliable for multifunctional geomacromolecules such as HAs. Second and most importantly, the conventional approach does not provide any means of selecting the appropriate isomers when numerous structural models can be inferred from the same set of analytical data. Thus, reliable results may be difficult to achieve when structural models of HAs generated with the conventional approach are used in subsequent calculations of their physicochemical properties by computational chemistry.

This paper describes an integrated experimental and computational framework for developing 3-D structural models for HAs (Figure 1). It combines *experimental characterization*, *computer assisted structure elucidation* (CASE), and *atomistic simulations* to generate all the 3-D structural models or a representative sample of these models that are consistent with the analytical data and bulk thermodynamic/structural properties of HAs. This approach is predicated upon four basic premises: (1) HAs from different sources (e.g., soils, plants, sediments, and streams) have different structural characteristics. No single structural model can be used to describe HAs from different sources. (2) Given a set of reliable structural data, the hierarchical approach shown

10.1021/es0259638 CCC: $25.00
Published on Web 04/04/2003

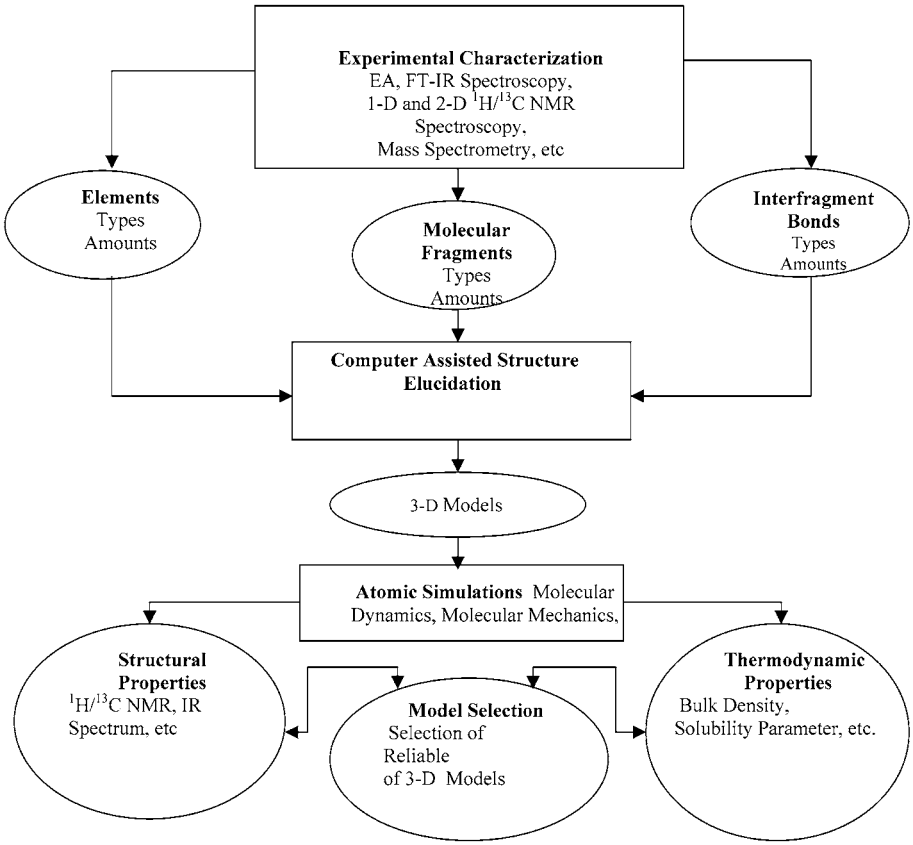

FIGURE 1. The hierarchical approach for modeling the 3-D structures of humic substances.

in Figure 1 can be used to generate all the 3-D models or a representative sample of these models that best match the structural data for the HA of interest. (3) These models can then be used in subsequent calculations of their bulk thermodynamic and structural properties (e.g., density, solubility parameter, ^{13}C NMR spectrum, etc.) by standard and validated methods of computational chemistry. (4) Only models that yield bulk thermodynamic and structural properties in agreement with the experimental data can be considered as reliable 3-D structural models for the HA of interest.

Experimental Methods and Procedures

Chelsea soil HA was selected as a model compound to illustrate this new methodology. The HA samples were extracted from Houghton muck, a Histosol soil widely found in the Great Lakes region of the United States [Michigan, Wisconsin, Minnesota, Illinois, Indiana, and Ohio] (*15*). The selection of Houghton muck as the HA source sample for this study was partially motivated by the availability of data on its origin and insight into the mechanisms of formation of Chelsea soil HA. Houghton muck consists of "very deep, very poorly drained soils formed in herbaceous organic deposits more than 51 in. thick" that usually occupy "closed depressions" in flood plains and morraines (*15*). The native vegetation that led to its formation consisted predominantly of grasses, sedges, reeds, buttonbrush, and cattails (*15*). The poor drainage of Houghton muck, the characteristics of its native vegetation, and the relatively large mean residence time of organic matter in Histosol soils (*1*) suggest that the condensation of plant and microbial degradation products (e.g., lignin degradation products, polyphenols, sugars, amino acids, etc.) was a major formation pathway for Chelsea soil HA.

Chelsea soil HA extraction was based on a standard procedure developed by the International Humic Substances Society (*16*, *17*). The extracted Chelsea soil HA sample was characterized by elemental analysis, diffuse reflectance FT-IR spectroscopy, 1-D ^{1}H and ^{13}C solution NMR spectroscopy, 2-D solution NMR (TOCSY and HMQC) spectroscopy, and ESI QqTOF MS. Detailed descriptions of the extraction and characterization methods are given in the section "Experimental Methods and Procedures" of the Supporting Information.

Computational Methods and Procedures

Computer Assisted Structure Elucidation. The CASE software SIGNATURE (*18*) was employed to generate all the 3-D structural models of Chelsea soil HA consistent with the quantitative/qualitative analytical data. SIGNATURE performs three basic tasks. First, it calculates an exhaustive and nonoverlapping list of molecular fragments and associated interfragment bonds that best match the structural input data for the HA of interest. In the second task, the software evaluates the total number of structural models that are consistent with the list of molecular fragments and interfragment bonds found in step 1. Finally, SIGNATURE generates all 3-D models of the HA of interest or a statistically representative sample of these models by randomly connecting the molecular fragment and interfragment bonds found in step 1. This CASE program is based on the *signature* descriptor (*18*). The signature of an atom can be viewed as a string of characters over an alphabet of atom types as defined in molecular modeling software such as Cerius^{2} (*19*). Atomic signatures can be defined at the 0, 1, 2, and h levels. The h-signature of an atom *x* in any given molecular group G is a tree rooted in x that describes its bonding environment up to a distance h. The h-signature of a molecular group or a chemical bond is readily expressed as a linear combination of its h-atomic signatures. For complex organic geomacromolecules such as humic acids, the signature descriptor provides a simple and robust means of coding (i) elemental

analysis data as 0 level atomic signatures, (ii) quantitative $^1H/^{13}C$ NMR data as 1 or 2 level atomic signatures, and (iii) qualitative data (e.g., molecular fragments and interfragment bonds from FT-IR spectroscopy, qualitative 1-D/2-D NMR spectroscopy, ESI mass spectrometry, etc.) as 1, 2, or higher level molecular signatures. Once these qualitative and quantitative data have been coded into the pertinent signatures for the HA of interest, the following conservation law provides the conceptual framework for the use of the *signature* molecular descriptor in structure elucidation:

sum of *h-signatures* of molecular fragments + sum of *h-signatures* of interfragment bonds = sum of *h-signatures* of the HA of interest

Let ${}^h\sigma(S)$ and ${}^h\sigma_\epsilon(S)$ be the set of experimentally derived input h-signatures and associated standard errors for the HA of interest. The quantity x_i of each molecular fragment (MF) f_i ($1 < i < I$), and the quantity y_j of each interfragment bond (IB) b_j ($1 < i < J$) can be calculated by solving the following system of equations:

$$\begin{vmatrix} {}^0\sigma(S) - {}^0\sigma_\epsilon(S) \le \sum_{i=1}^{I} x_i\,{}^0\sigma(f_i) + \sum_{j=1}^{J} y_j\,{}^0\sigma(b_j) \le {}^0\sigma(S) + {}^0\sigma_\epsilon(S) \\ {}^1\sigma(S) - {}^1\sigma_\epsilon(S) \le \sum_{i=1}^{I} x_i\,{}^1\sigma(f_i) + \sum_{j=1}^{J} y_j\,{}^1\sigma(b_j) \le {}^1\sigma(S) + {}^1\sigma_\epsilon(S) \\ \cdots\cdots\cdots\cdots\cdots \\ {}^h\sigma(S) - {}^h\sigma_0{}^\epsilon(S) \le \sum_{i=1}^{I} x_i\,{}^h\sigma(f_i) + \sum_{j=1}^{J} y_j\,{}^h\sigma(b_j) \le {}^h\sigma(S) + {}^h\sigma_\epsilon(S) \end{vmatrix} \quad (1)$$

where I and J are the total numbers of molecular fragments and interfragment bonds. Since the purpose of the SIGNATURE program is to construct molecular models, x_i and y_j are always positive integer numbers. Because of limited experimental data, the linear system given in eq 1 is generally undetermined and has more than one solution. However, for the purpose of HA structure elucidation, we seek the best solution (eq 2), i.e., that which minimizes the difference between the sum of the *signatures* of the molecular fragments and interfragment bonds, and the *signature* of the HA of interest:

$$\min\{|\Sigma X - \sigma(S)|\},\ \Sigma X \le \sigma(S) + \sigma_\epsilon(S),\ \Sigma X \le \sigma(S) - \sigma_\epsilon(S),\ X \text{ integral} \quad (2)$$

$$\sum = \begin{vmatrix} {}^0\sigma(f_1) & {}^1\sigma(f_1) & \cdots & {}^h\sigma(f_1) \\ \vdots & \vdots & & \vdots \\ {}^0\sigma(f_I) & {}^1\sigma(f_I) & \cdots & {}^h\sigma(f_I) \\ {}^0\sigma(b_1) & {}^1\sigma(b_1) & \cdots & {}^h\sigma(b_1) \\ \vdots & \vdots & & \vdots \\ {}^0\sigma(b_J) & {}^1\sigma(b_J) & \cdots & {}^h\sigma(b_J) \end{vmatrix} \quad (3)$$

where $\sigma(S) = \{{}^0\sigma(S), \ldots, {}^h\sigma(S)$, and $\sigma_\epsilon(S) = ({}^0\sigma_\epsilon(S), \ldots, {}^h\sigma_\epsilon(S)\}$ are the vectors of input atomic/molecular signatures and associated standard errors, Σ is the matrix of signatures for the selected input molecular fragments and interfragment bonds, and $X = (x_1,\ldots,x_I, y_1,\ldots,y_J)$ is the solution vector. Equation 2 formally describes an *integer linear programming* problem. SIGNATURE uses two basic techniques to solve this problem: systematic enumeration and simulated annealing (*18*). Once the optimal molecular building blocks (i.e., types and amounts of MFs and IBs) have been determined, SIGNATURE generates all the 3-D models that are consistent with the input data by randomly connecting the pertinent molecular fragments and interfragment bonds for the HA of interest. The users of SIGNATURE can also impose structural constraints such as generating 3-D models with number average molecular weights within a specified range. Thus, SIGNATURE has the inherent capability to generate representative 3-D models for complex organic geomacromolecules such as lignin (*20*) and asphaltenes (*21*) if the pertinent analytical data is available. A summary of the mathematical theory behind the SIGNATURE software is given in the section "Computer Assisted Structure Elucidation Procedures" of Supporting Information.

Atomistic Simulations. Molecular mechanics (i.e., energy minimization) and NPT MD simulations were used to calculate the bulk thermodynamic properties (molar volume, bulk density, and solubility parameter) of the SIGNATURE generated 3-D structural models of Chelsea soil HA. Each SIGNATURE generated structural model was first energy minimized and subsequently subjected to three series of 15 ps of constant volume-constant temperature (NVT) MD simulations at 3000 K followed by energy minimization (with final rms force of 0.1 Kcal mol^{-1} A^{-1}). Each annealed model was subsequently placed in a 3-D cell with periodic boundary conditions and packed to a bulk density of 1.0 g/cm^3 using the Amorphous Builder of Cerius^2 (*19*). The models were then subjected to energy minimization to remove the packing-induced bad contacts. Each minimized 3-D periodic model was subsequently subjected to 25 ps of NPT MD simulations at $T = 300$ K followed by energy minimization until its bulk density remained constant. Only two cycles of MD simulations followed by energy minimization were needed in most cases to achieve this goal. The Dreiding force field (*22*) [with an EXP 6 Lennard-Jones potential for the van der Waals interaction] was employed in all the MD simulations and energy minimization. The charge equilibration (Qeq) procedure (*23*) was used to evaluate all partial atomic charges. Ewald summation (*24*) was employed to calculate the long range electrostatic and van der Waals interactions for all the periodic systems. Conversely, these interactions were treated directly with a cutoff radius of 30 Å for the nonperiodic systems. The Berendsen thermal coupling method [time constant of 0.1 ps] (*25*) and the Andersen pressure control method [cell mass prefactor of 0.04] (*26*) were employed in all NPT MD simulations. After completion of the MD simulations and energy minimization runs, the cell volume (V_p), condensed phase strain energy (E_p), and gas-phase strain energy (E_{np}) were calculated for each model. The molar volume (V_m), bulk density (ρ), and cohesive energy (E_c) for each 3-D period Chelsea HA model were expressed as

$$V_m = N_a V_p \quad (4)$$

$$\rho = \frac{V_m}{M_n} \quad (5)$$

$$E_c = -(E_p - E_{np}) \quad (6)$$

where N_a is Avogadro's number and M_n is the molar mass of the SIGNATURE generated Chelsea HA model. Following Barton (*27*), the Hildebrand solubility parameter (δ) is expressed as

$$\delta = \sqrt{\frac{E_c}{V_m}} \quad (7)$$

Equations 4−7 can be used to calculate the bulk density, cohesive energy, and Hildebrand solubility parameter of HAs by MD simulations once a sample of representative 3-D structural models have been generated.

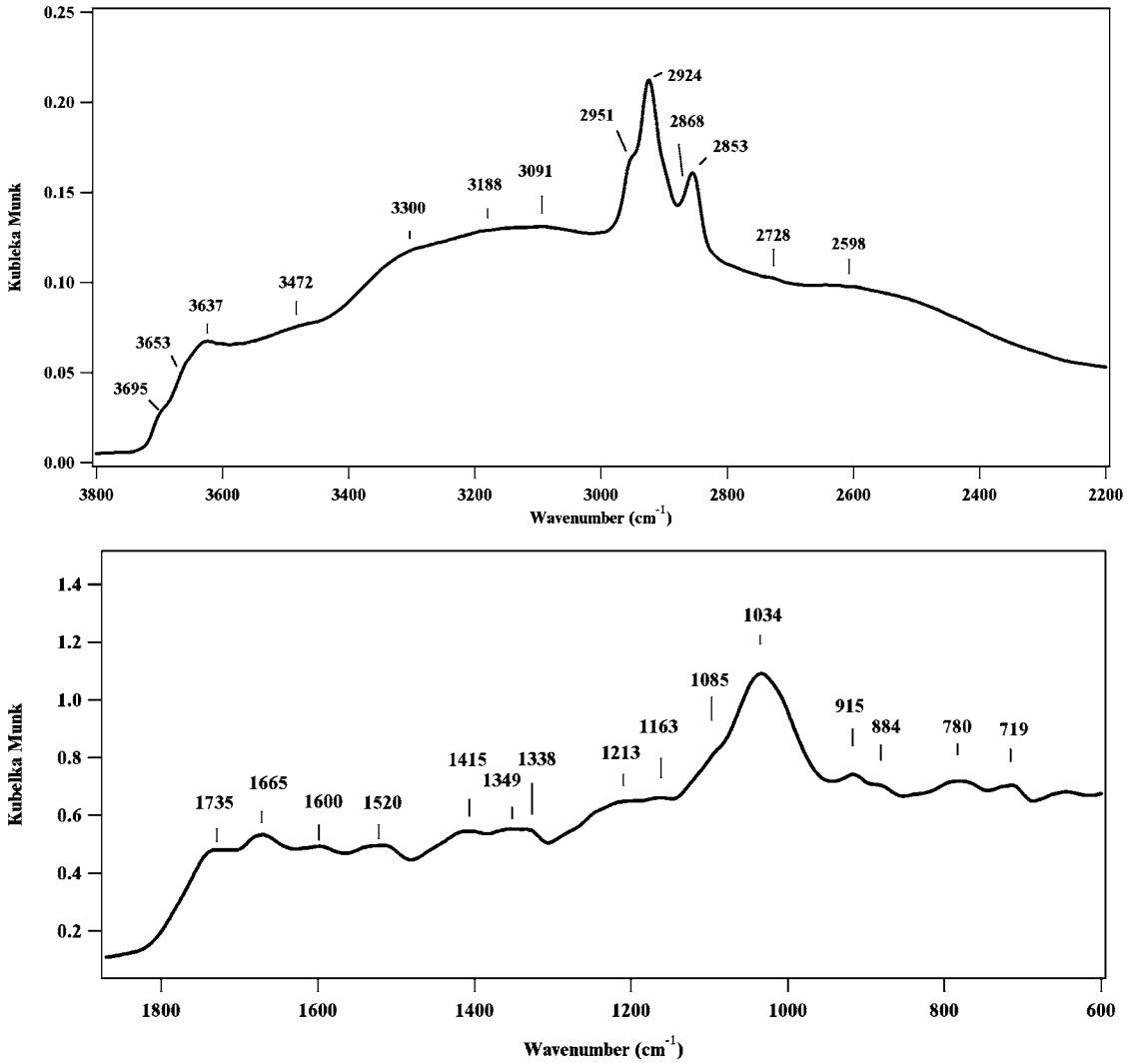

FIGURE 2. Diffuse reflectance FT-IR spectrum for Chelsea soil humic acid.

1-D ^{13}C NMR Simulation. The chemical shift values for the selected 3-D models for Chelsea soil HA were calculated using the NMR module of Chem Draw Pro (*28*) and used as input to the commercial NMR simulation package "NMRSIM" distributed by Bruker Analytik (29). The ^{13}C NMR spectrum of an equimolar mixture of the selected Chelsea soil HA 3-D models was calculated using relaxation parameters T1/T2 of 40 ms as measured experimentally for the Chelsea soil HA sample. The resulting spectrum was processed in "XWIN NMR" using an exponential multiplication with a 100 Hz line broadening prior to Fourier transformation as applied to the measured FID ^{13}C NMR spectrum.

Results

Diffuse Reflectance Fourier Transform Infrared Spectroscopy. The organic normalized weight fractions for C (51.31%), H (4.00%), O (39.67%), N (4.12%), and S (0.93%) along with the O/C atomic ratio (0.58) of Chelsea soil HA are typical of soil humic acids (*1*). Its DRIFT spectrum (Figure 2) exhibits typical broad bands and shoulders found in the IR spectra of many soil HAs (*1, 30*−*32*). The region between 3800 and 2200 cm^{-1} exhibits very broad bands with four distinct frequency ranges. The high-frequency modes above 3500 cm^{-1} are assigned to nonbonded OH stretches. Most of these occur as broad shoulders of poorly resolved aromatic C−H stretches in the 3400−3000 cm^{-1} region. The 3000−2800 cm^{-1} region exhibits high-intensity bands characteristic of symmetric and asymmetric aliphatic CH_2 and CH_3 stretches. Conversely, the weak bands in the 2800−2400 cm^{-1} region are assigned to OH stretches from COOH groups. In the 1800−1300 cm^{-1} region, we observe several peaks including the following: (i) C═O stretches from COOH groups, (ii) aromatic C═C stretches, (iii) CH deformation of CH_3 groups, and (iv) CH bending of CH_2 groups. In the 1270−760 cm^{-1} region, we observe several peaks including the following: (i) aromatic C−H and C−OH stretches and (ii) out-of-plane C−H bends. These are assigned to mono, di, and tri hydroxyl substituted aromatics. Characteristic aliphatic C−O and C−OH stretches of carbohydrates are also observed in the 1270−760 cm^{-1} region.

1-D and 2-D Solution NMR Spectroscopy. The ^{13}C NMR spectrum of Chelsea soil HA (Figure 3A) also exhibits the typical broad features found in the ^{13}C NMR spectra of many soil HAs (*33*−*42*). The aliphatic region (>50 ppm and centered around peak 1 can arise from (i) carbon atoms in the side chains of peptides or amino acids, (ii) carbon atoms in straight or branched hydrocarbon chains, or (iii) from any carbon atom once or twice removed from electron withdrawing functional groups such as ester, carboxylic acid, ether, or hydroxyl. Peak 2 is assigned to methoxy aromatic carbon and is a good indicator for the presence of lignin derived aromatic compunds. This assignment is consistent with the cross-peak 1 of the HMQC spectrum (Figure 5) which indicates that the protons attached to this carbon interact

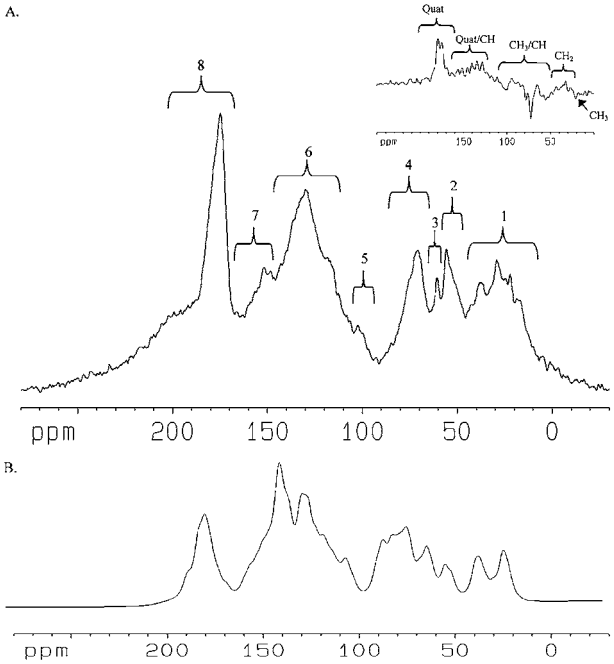

FIGURE 3. 1-D Solution ^{13}C NMR spectra for Chelsea soil humic acid. A. Measured carbon spectrum acquired under quantitative experimental conditions in D2O/NaOD. The assignments are consistent with the following: 1. aliphatic, 2. methoxy, 3. α carbon in amino acids and peptides, 4. sugars carbon, 5. anomeric sugar carbon, 6. aromatic carbon, 7. oxygen substituted aromatic carbon, and 8. carboxylic acid and carbonyl carbon. Assignments are confirmed by the PENDENT edited carbon spectrum (insert) which is phased so that quaternary carbon point up, CH carbons point down, CH_2 carbon point up and CH_3 carbon point down. B. Simulated ^{13}C NMR spectrum of an equimolar mixture of the SIGNATURE generated Chelsea soil humic acid models from Figure 10.

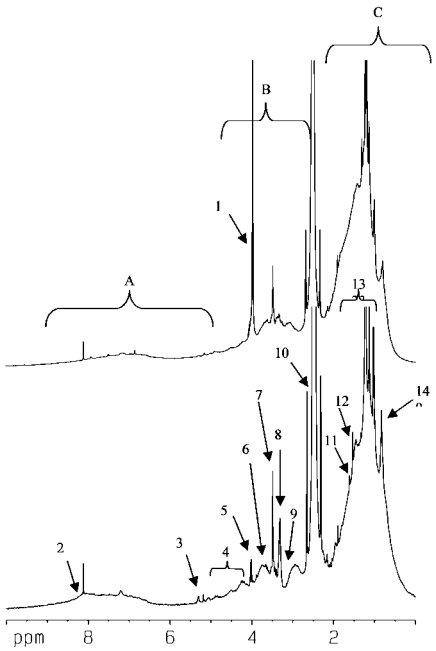

FIGURE 4. 1-D Solution ^{1}H NMR spectra for Chelsea soil humic acid at a concentration of 1 mg/mL, in DMSO-d_6 (top) and DMSO-d_6 with the addition of 100 μL of D_2O (bottom). The major structural categories are labeled (top) as protons in aromatic and amides (A), sugars, amino acids, methoxy, aliphatic hydroxy (B) and aliphatic, amine protons (C). More detailed assignments are consistent with 1. H_2O at high concentration in D_2O (results from the addition of the D_2O), 3. double bond protons, 4. protons in sugars or on the α carbon of peptides/amino acids and lignin bridging units, 5. quartet multiplet, 6. broad "hump" from methoxy, sugar protons,or α carbon of peptides/ amino acids, 7. strong resonance likely to be from hydroxylated aliphatic carbon, 8. water in DMSO at low concentration, 9. distortion of baseline from presaturation pulse, 10. DMSO-d_6, 11 and 12 exchangeable amine protons such as those in amino acids or terminal residues of peptides, 13 various methylene resonances and 14 methyl units.

through space with the aromatic protons as observed in lignin and HAs of similar origin (*43*). Peak 3 is assigned to α carbon from amino acids or peptides. The TOCSY spectrum (Figure 6) is consistent with the presence of such components or building blocks in the Chelsea soil HA sample. Peak 4 is often attributed to sugars, although ether, ester, hydroxylated carbon may also resonate in this region. In hexose sugars, the ratio of anomeric sugar carbon (peak 5) to the "ring" carbon atoms is 1:5. For Chelsea soil HA, the ratio of peak 4 to peak 5 is approximately 3%: 11; thereby suggesting that most of the broad band centered around peak 4 results from sugars. Peak 6 is assigned to nonoxygenated aromatic carbon, whereas peak 7 is attributed to oxygen substituted aromatic carbon from polyphenolic structures. Peak 9 results from carbonyl or carboxylic bearing compounds (e.g., aliphatic acids, amino acids, sugars, and lignin derived aromatics).

The ^{1}H NMR spectrum of Chelsea HA (Figure 4) is very broad. Due to the near continuous overlap, the integration of the individual proton signals is virtually impossible. However, it is possible to integrate the broad regions highlighted in Figure 4. Region A consists predominantly of aliphatic protons. The 2-D HMQC NMR spectrum of Chelsea HA (Figure 5) exhibits several cross-peaks consistent with CH_2 groups from straight or branched hydrocarbon chains. With the exception of small exchangeable signals from free amino acids (peaks 11 and 12 on the lower part of Figure 4), region B consists predominantly of nonexchangeable aliphatic protons. Conversely, region C exhibits a variety of signals from both nonexchangeable i.e., $CH_2-CO-O-R$, CH_2-O-R, amino acids, sugars, etc., and exchangeable functional groups mainly OH from sugars. More detailed assignments for the 1-D and 2-D NMR spectra are provided in the corresponding figure captions. Overall, the results of the DRIFT and NMR spectroscopic experiments are consistent with the hypothesis that the condensation of plant and microbial degradation products (e.g., lignin degradation products, polyphenols, sugars, amino acids, etc.) was a major formation pathway for Chelsea soil HA.

Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry. The availability of reliable molar mass data is critical to the development of 3-D structural models for HAs. During the earlier 1970s, Swift and co-workers carried out ultracentrifugation and gel permeation chromatography (GPC) studies of aqueous solutions of humic substances (*43*, *44*). They reported mass averaged molar masses of humic acids ranging from 10 000 to 100 000 Dalton. During the last two decades, virtually all standard methods of macromolecular physical chemistry [e.g., vapor pressure osmometry (VPO), high-pressure size exclusion chromatography (HP−SEC), flow field fractionation (FFF), etc.] have been used to determine the number average molar mass (M_n) and weight average molar mass (M_w) of humic substances (*45*−*48*). While there are some discrepancies among the reported measured values, a consensus is emerging that fulvic acids (FAs) have much lower molar masses than humic acids (HAs). VPO studies in tetrahydrofuran (*45*), HP−SEC (*46*−*47*), and FFF (*48*) measurements have established that FAs exhibit M_n values ranging from 600 to 1500 Dalton. Similarly, FFF studies (*48*) have established that HAs isolated from aquatic, soil, peat, and coal samples have M_n values ranging from 800 to 2500 Dalton. Recently, more advanced analytical tools such as ESI Fourier Transform ion cyclotron (FT-ICR) mass

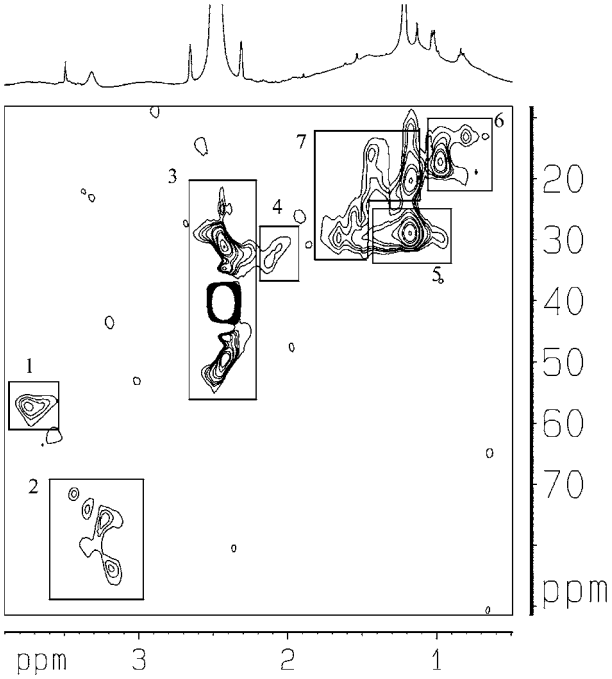

FIGURE 5. Region of the HMQC spectrum (in DMSO-d_6) of Chelsea soil humic acid with cross-peaks. Cross-peaks are consistent with the following assignments: 1. aromatic methoxy, 2. hydroxylated carbon, 3. DMSO-d_6 solvent peak, 4. CH_2 adjacent to C═C or COOH, 5. straight chain CH_2 groups that are least two bonds removed from electron withdrawing functionalities such as oxygen, 6. CH_3 groups, and 7. various aliphatic groups that are consistent with CH_2 adjacent to hydroxylated carbon or once removed from double bonds, carboxylic acids or amino side chain carbons.

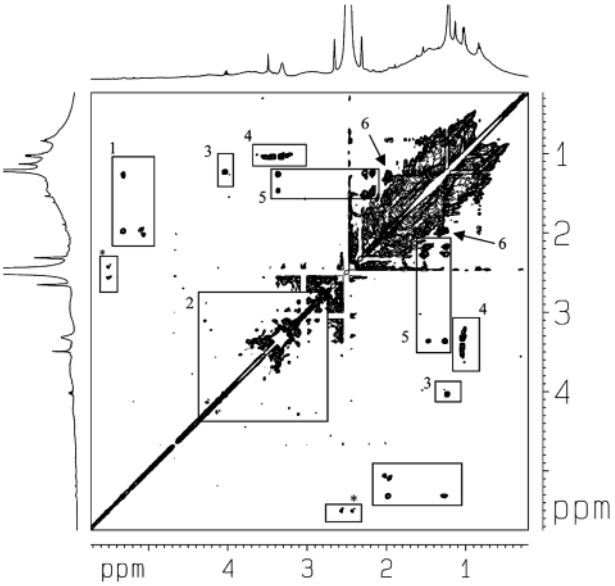

FIGURE 6. Expanded region of the TOCSY spectrum of Chelsea soil humic acid in DMSO-d_6. * Represents an artifact resulting from the symmetrisation of the spectrum. The labels are consistent with the following assignments: 1. couplings from double bond protons to CH_2 protons in aliphatic chains, 2. couplings in sugar moieties, 3. $\alpha-\beta$ couplings in alanine or possible in an acetyl structure, 4. couplings from hydroxylated aliphatic carbon to carbon in aliphatic chains ($-(CH_2)_nCOH(CH_2)_n-$), 5. $\alpha-\beta-\gamma$ couplings in amino acid residues, and 6. a strong coupling that is likely to result from aliphatic CH_2 groups once removed from a double bond (i.e., $CH{=}CH-CH_2-CH_2$) or aliphatic acids components ($HOOC-CH_2-CH_2-CH_2$). The intense region from 1 to 2.5 ppm contains a multitude of partially resolved couplings that are difficult to observe in the printed spectrum. These are consistent with a distribution of couplings within a series of slightly different aliphatic compounds that may be variously substituted and branched.

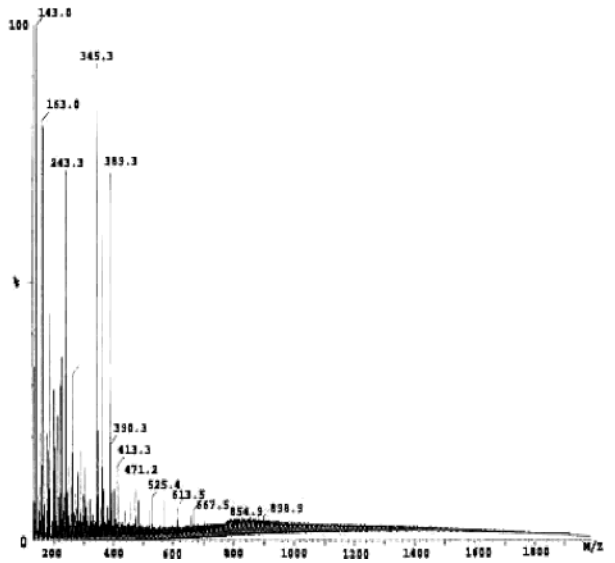

FIGURE 7. Electrospray ionization quadrupole time-of-flight (ESI QqToF) mass spectrum for Chelsea soil humic acid.

spectrometry (MS), ESI Qq TOF MS, and ESI multistage tandem MS have been used to characterize humic substances (*49*−*55*). All the MS spectra reported in these studies showed a broad distribution of peaks extending to 1000−2000 Dalton. However, the majority of the most intense peaks occurred below a m/z ratio of 1000 Dalton thereby suggesting that the HAs tested do not have very large molecular weights.

The ESI QqTOF mass spectrum of Chelsea HA is shown in Figure 7. It is qualitatively similar to the ESI QqTOF mass spectrum of Mount Rainier soil HA recently reported by Kujawinski et al. (*55*). This sample is representative of degraded wood HAs from forest soils of the Pacific Northwest regions of Canada and the United States. Both the ESI Q-TOF mass spectra of Chelsea soil HA (see the section "Experimental Methods and Procedures" of Supporting Information) and Mount Rainier soil HA (*55*) were acquired under identical experimental conditions using the Micromass Q-Tof II mass spectrometer of the Campus Chemical Instrument Facility at The Ohio State University. Through subsequent characterization by ultrahigh resolution ESI FT-ICR MS, Kujawinski et al. (*55*) confirmed the presence of lignin degradation products in the Mount Rainier soil HA sample. Both the ESI QqTOF and FT-ICR mass spectra also confirmed the predominance of relatively low molecular weight compounds in the Mount Rainer soil HA sample (*55*). The ESI QqTOF of Chelsea soil HA (Figure 7) tails at approximately 1200 Da thereby suggesting that higher molecular weight compounds are also not significant components or building blocks of Chelsea soil HA.

Computer Assisted Structure Elucidation and Model Generation. The generation of the 3-D molecular models for Chelsea HA by SIGNATURE was carried out in two steps: (i) determination of the optimal set of molecular building blocks and interfragment bonds and (ii) generation of all the 3-D structural models consistent with this set of building blocks and interfragment bonds. The first step was carried out by running the "elucidation" mode of the SIGNATURE program. In this mode, the program solves eq 2 to determine the optimal amounts of the molecular fragment and interfragment bonds that best match the input data. The input atomic/molecular signatures for Chelsea soil HA are given in Table 1. They were derived from elemental analysis and quantitative

TABLE 1: Experimentally Derived Input Atomic Signatures ($^h\sigma(S)_{exp}$) for Chelsea Soil HA[a]

atom type	signature	$^h\sigma(S)_{exp}$
C		
H	h_	93.40
Osp^3	o_	43.50
Osp	o′	14.50
Nsp^3	n	6.90
Ssp^3	s	0.66
aliphatic C	c_	17.00
aromatic C	cp	34.00
methyl C	c_(h_h_h_*_)	3.00
α C amino acid	c_(n_c_h_*_)	3.00
anomeric sugar C	c_(o_c_o_h_)	2.00
carbonyl + carboxyl C	c=(o′*_*_*_)	24.00
O substituted aromatic C	cp(cpcpo_*_)	9.00
methoxy aromatic C	o_(cp(cpcp*_)c_(h_h_h_)*_*_)	6.00
CA hexose sugar C	c_(o_(h_*_*_)c_(c_o_h_)o_(c_*_*_)h_(*_*_*_))	2.00
CB hexose sugar C	c_(o_(h_*_*_)c_(c_o_h_)c_(o_o_h_)h_(*_*_*_))	2.00
CC hexose sugar C	c_(o_(h_*_*_)c_(c_o_h_)c_(c_o_h_)h_(*_*_*_))	3.19
CE hexose sugar C	c_(c_(o_h_h_)c_(c_o_h_)o_(c_*_*_)h_(*_*_*_))	2.00
CF hexose sugar C	c_(o_(h_*_*_)c_(c_o_h_)h_(*_*_*_)h_(*_*_*_))	2.00

[a] The o-level atomic signatures (h_, o_, o′, n, s, c_, and cp) were derived from elemental analysis and/or quantitative ^{13}C NMR spectroscopy. All the remaining 1- and 2-level atomic signatures were derived from quantitative ^{13}C NMR spectroscopy. NMR integrals for the sugar C atoms were calculated assuming that carbohydrates in humic acids typically consist of 85% hexoses and 15% pentoses (*1*). To ensure the generation of 3-D structural models with a number average molecular weights that are consistent with the ESI Q-TOF mass spectrum (Figure 7) of Chelsea soil humic acid, the number of C atoms of the 3-D structural models was constrained to vary between 40 and 75. All input atomic signatures are expressed on 100 C atom basis. A standard error of 20% was assumed in all input atomic signatures.

TABLE 2. Qualitative Input Parameters for Chelsea Soil Humic Acid: Selected List of "Precursor Molecules" and Interfragment Bonds for Chelsea Soil Humic Acid (HA)[a]

lignin derived fragments	amino acids	polyphenols	sugars	aliphatic acids	bonds
1-(4-hydroxy-3,5-dimethoxy-phenyl) ethanol	aspartic acid	1,2,3 trihydrox benzoic acid	galacturonic acid	undecanoic acid	C_{aro}_C_{aro}
1-(3,4-dimethoxyphenyl) ethanol	glutamic acid	2,3,4 trihydrox benzoic acid	gluconic acid	dodecanoic acid	C_{aro}_H
3,4,5-trimethoxy cinnamic acid	alanine	2,3,6 tricarboxy phenol	mannuronic acid	tridecanoic acid	C_{aro}_O
1-(4-hydroxyphenyl) ethanol	asparagine	2,4 dicarboxy phenol	allose	tetradecanoic acid	C_{aro}_N
3,4-dimethoxy cinnamyl alcohol	arginine	2,4 dihydroxy benzoic acid	arabinose	pentadecanoic acid	C_{ali}_ C_{aro}
3,4-dimethoxy cinnamic acid	cysteine	3.4,5 trihydroxy benzoic acid	fucose	hexadecanoic acid	C_{ali}_H
3,4-dimethoxybenzoic acid	glutamine	3.4 dihydroxy benzoic acid	galactose	heptadecanoic acid	C_{ali}_O
4-methoxy cinnamic acid	glycine	3.5 dihydroxy benzoic acid	glucose	octadecanoic acid	C_{ali}_N
4-hydroxy benzoic acid	histidine	3 hydroxy benzoic acid	gulose	nonadecanoic acid	
apocynol	isoleucine	4 hydroxy benzoic acid	idose	eicosanoic acid	
cinnamyl alcohol	leucine	phenol	mannose	ethanoic acid	
coniferyl alcohol	lysine	o-creosol	rhamnose	propanioc acid	
dihydroferulic acid	methionine	m-creosol	ribose	butanoic acid	
dihydrocoumaric acid	phenylalanine	p-creosol	xylose	pentanoic acid	
eugenol	proline	phloroglucinol		hexanoic acid	
ferulic acid	serine	resorcinol		heptanoic acid	
gallic acid	threonine			octanoic acid	
guaiacol	trytophan			nonanoic acid	
guaiacyl propionic acid	tyrosine			decanoic acid	
isoeugenol	valine				
protocatechuic acid					
sinapyl alcohol					
sinapinic acid					
syringyl alcohol					
syringic acid					
syringol					
syringyl propionic acid					
vanylic acid					
veratric acid					
vinyl guaiacol					
cis-ferulic acid					
p-anisic acid					
p-coumaric acid					
p-coumaryl alcohol					

[a] The "precursor molecules" were selected from a library of molecular fragments derived from the qualitative FT-IR and 1-D/2-D $^{13}C/^1H$ NMR spectroscopic data. The bonding sites in the "precursor molecules" and the corresponding interfragment bonds were selected to ensure that they could be covalently linked through the condensation mechanisms known to mediate HA formation from plant biodegradation products in Hitosol soils [e.g., oxidative coupling, sugar-amine condensation, alky-aromatic substitution (*1, 57, 58*)].

^{13}C NMR data (Figure 3A) assuming a standard error of 20%. The set of molecular fragments (i.e., "precursor" molecules) and interfragment bonds used as input to SIGNATURE are given in Table 2. Consistent with the DRIFT and NMR spectroscopic data along with the hypothesized formation pathway for Chelsea soil HA, we selected 34 lignin degradation products (*56*), 20 amino acids, 14 carbohydrates, 16 polyphenols, and 19 aliphatic acids as "precursor molecules" (Table 2). Similarly, the bonding sites in the "precursor molecules" and the corresponding interfragment bonds (Table 2) were selected to ensure that they could be covalently linked through various condensation mechanisms known to medi-

TABLE 3. SIGNATURE Output Parameters for Chelsea Soil Humic Acid:[a] List of Molecular Fragments and Interfragment Bonds that Best Match the Structural Input Data from Tables 1 and 2

molecular fragments	no. of fragments	interfragment bonds	no. of bonds
p-coumaric acid	1	C_{aro}_O	3
p-anisic acid	1	C_{aro} _N	1
2,3,6 tricarboxyl phenol	1	C_{ali}_O	1
resorcinol	1		
methionine	1		
galacturonic acid	1		

[a] These molecular fragments and interfragment bonds are the solution of eq 2. They were derived by asking SIGNATURE to find the optimum list and amounts of "precursor molecules" and corresponding interfragments bonds from the selected "pool" given in Table 2 subject to the structural constraints given in Table 1.

TABLE 4. Atomic Signatures for Chelsea Soil Humic Acid:[a] Model Predictions versus Experimentally Derived Input Data

signature	$^{a}\lvert {}^{h}\sigma(S)_{exp} - {}^{h}\sigma(S)_{pred}\rvert$
h_	2.20
o_	5.70
o'	1.10
n	4.70
s	1.60
c_	9.70
cp	19.30
c_(h_h_h_*_)	1.40
c_(n_c_h_*_)	0.80
c_(o_c_o_h_)	0.20
c=(o'*_*_*_)	8.40
cp(cpcpo_*_)	2.10
o_(cp(cpcp*_)c_(h_h_h_)*_*_)	0.70
c_(o_(h_*_*_)c_(c_o_h_)o_(c_*_*_)h_(*_*_*_))	0.20
c_(o_(h_*_*_)c_(c_o_h_)c_(o_o_h_)h_(*_*_*_))	0.20
c_(o_(h_*_*_)c_(c_o_h_)c_(c_o_h_)h_(*_*_*_))	0.90
c_(c_(o_h_h_)c_(c_o_h_)o_(c_*_*_)h_(*_*_*_))	0.20
c_(o_(h_*_*_)c_(c_o_h_)h_(*_*_*_)h_(*_*_*_))	0.20
average atomic signature error	3.00

[a] $^{a}\lvert {}^{h}\sigma(S)_{exp} - {}^{h}\sigma(S)_{pred}\rvert$: signature error per 100 C atoms. It is equal to the absolute value of the difference between the experimentally derived input atomic/molecular signature and the corresponding predicted model signature. In all cases, the signature errors, including that in the aromatic carbon signature (19.30%), are smaller or comparable to the assumed standard error (20%) in the signature input data (Table 1). The average signature error is 3%. That is, approximately three atoms are missing or are in excess for every 100 carbon atoms of the SIGNATURE generated 3-D Chelsea soil HA models.

ate HA formation in Histosol soils including biotic and abiotic oxidative coupling, sugar-amine condensation, alkyl-aromatic substitution, etc. (*1, 57, 58*). To ensure the generation of 3-D structural models with molar mass that are consistent with the ESI QqTOF mass spectrum of Chelsea soil HA (Figure 7), we constrained the number of C atoms of the candidate models to vary between 40 and 75.

A simulated annealing search (*18*) of 10 cycles ($T_{initial}$ = 100 K, T_{final} = 1000 K, and $T_{increment}$ = 100 K) was used to solve eq 2. The optimal molecular fragments for Chelsea soil HA (Table 3) consist of 2 lignin-derived aromatics (p-coumaric acid and p-anisic acid), 2 polyphenols (2,3,6-tricarboxylphenol and resorcinol), 1 amino acid (methionine), and 1 carbohydrate (galacturonic acid). Its optimal interfragment bonds consist of 3 aromatic C_{aro}_O bonds, 1 aromatic C_{aro} _N bond, and 1 aliphatic C_{ali}_O bond. The combination of these molecular fragments and interfragment bonds yields a total number of 18 3-D structural models for Chelsea soil HA. Each model has a molecular formula of $C_{45}H_{43}O_{24}N_1S_1$ with a molar average molecular weight of 1016 Dalton. This value is within the mass range of the ESI Qq TOF mass spectrum

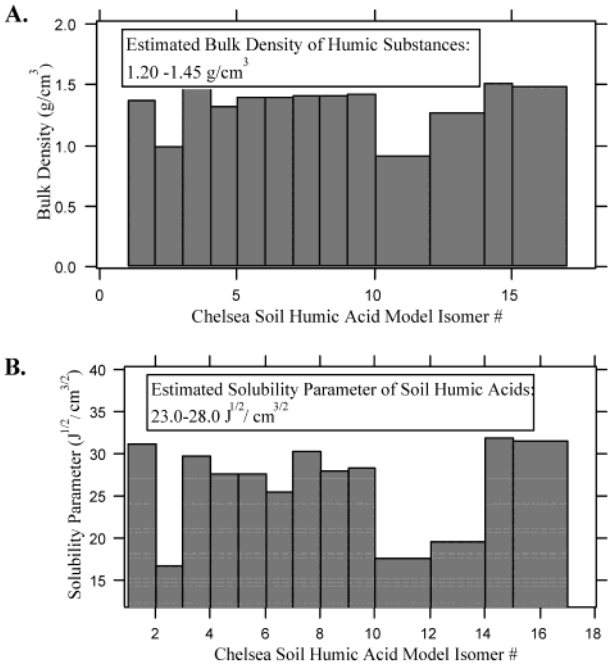

FIGURE 8. Calculated bulk densities (A) and solubility parameters (B) of the SIGNATURE generated 3-D models for Chelsea soil humic acid.

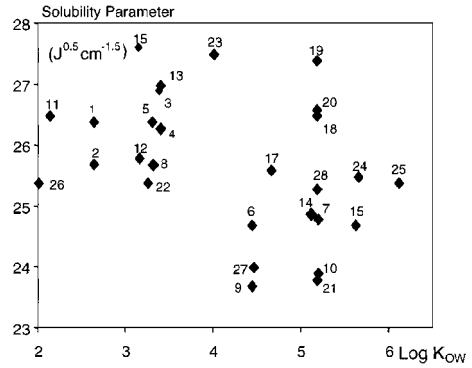

FIGURE 9. Solubility parameters for "humic organic matter" (HOM) estimated by Poerschman and Kopinke (*62*) from literature sorption data. The solubility parameter (δ) is defined as the square root of the cohesive density (see eq 7) and thus is readily calculated from heat of vaporization and molar volume data. Because direct measurements of these thermodynamic properties for humic acids are not feasible in most cases, fitting measured organic normalized soil binding constants (K_{OC}) of hydrophobic organic compounds (HOCs) to a Flory−Huggins model has become the "standard" method for determining the solubility parameters of humic substances in the environmental science literature (*62*). The solubility parameters data shown in this figure were calculated by Poerschman and Kopinke (*62*) by fitting literature data of HOC sorption onto HOM to a Flory−Huggins model. They found their range of values to be "quite narrow" (i.e., 23.0−28.0 $J^{1/2}/cm^{3/2}$). Reprinted with permission from Poerschman, J.; Kopinke, F. D. *Environ. Sci. Technol.* 2001, *35*, 1142. Copyright 2001 American Chemical Society.

of Chelsea soil HA (Figure 7). As shown in Table 4, the signature errors (i.e., differences between the experimentally derived and predicted atomic signatures) of the Chelsea soil HA models are relatively low. In all cases, the signature errors, including that in the aromatic carbon signature (19.30%), are smaller or comparable to the assumed standard error (20%) in the signature input data (Table 1). The average signature error is 3%. That is, approximately three atoms are missing or are in excess for every 100 carbon atoms of the SIGNATURE generated Chelsea soil HA models.

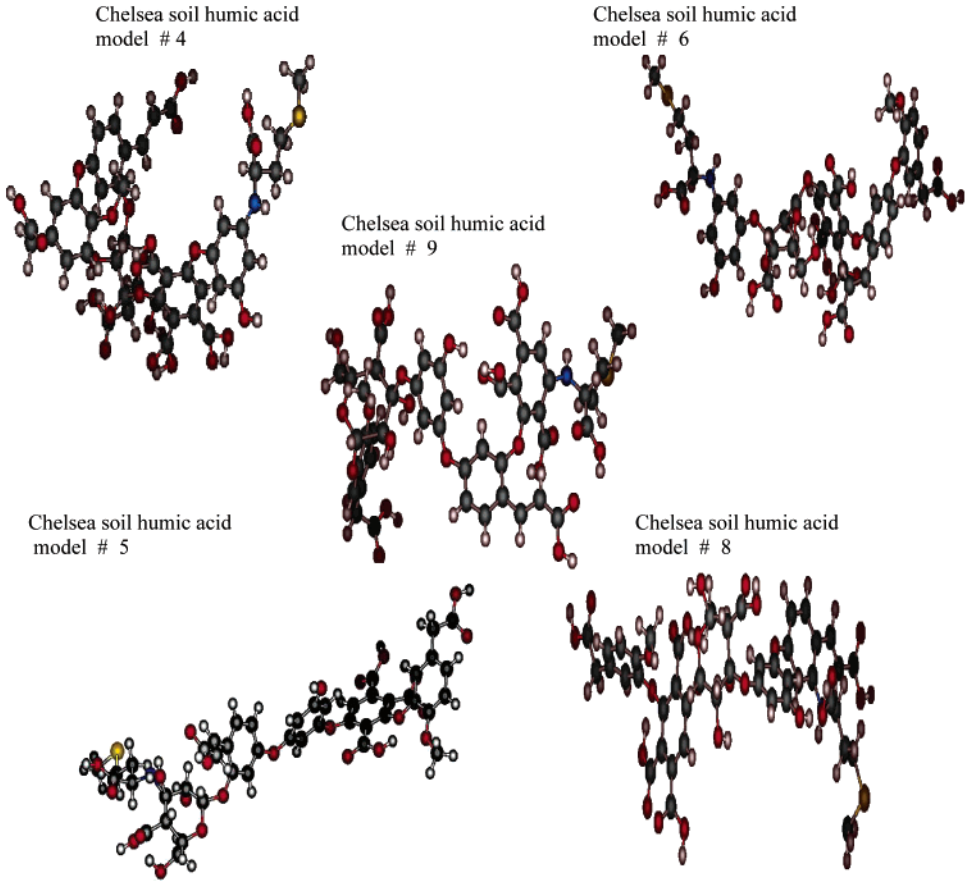

FIGURE 10. 3-D structures and bulk densities (ρ) and solubility parameters (δ) of the selected SIGNATURE generated model isomers for Chelsea humic acid (HA). C atoms are in black, O atoms are in red, N atoms are in blue, S atoms are in yellow, and the remaining atoms are H atoms.

Atomistic Simulations and Model Selection. In the third phase of this study, SIGNATURE was used to generate all the 18 3-D structural models model isomers for Chelsea soil HA by randomly connecting the optimal "precursor molecules" and corresponding interfragment bonds given in Table 3. The strain energy, molar volume, density, and solubility parameter of each of these models was subsequently calculated using the procedures described in the section "Computational Methods and Procedures". Three of the Chelsea soil HA model isomers with unacceptably high strain energies ($>10^{12}$ kcal/mol) were discarded. A fourth model with unacceptably low bulk density (0.74 g/cm^3) was also discarded. The bulk densities (ρ) and solubility parameters (δ) of the remaining Chelsea soil HA model isomers are shown in Figure 8. The bulk densities of humic substances have been estimated to range from 1.20 to 1.45 g/cm^3 (*59*−*61*). Poerschman and Kopinke (*62*) have recently compiled solubility parameter data for "humic organic matter" (HOM) from a variety sources (Figure 9). They found their range of values to be "quite narrow" (i.e., 23.0−28.0 J$^{1/2}$/cm$^{3/2}$). Based on these literature values, surprisingly, only five of the Chelsea soil HA model isomers exhibit bulk density and solubility parameters comparable in magnitude to experimental estimates from the literature. The 3-D structure, bulk density, and solubility parameter of each of the Chelsea soil HA model isomers are shown in Figure 10. These models have the following characteristics: (1) They are structurally different model isomers that have molecular composition consistent with the experimental characterization data. (2) They are "molecularly heterogeneous" compounds that incorporate key building blocks of HAs (e.g., lignin derived aromatic compounds, sugar, amino acid and polyphenols) typically found in HAs extracted from Histosol soils (*1*). (3) They consist of relatively small compounds (i.e., number average molecular weight of 1016 Dalton). (4) They possess multiple functional groups (e.g. COOH and OH groups) that can interact with organic/inorganic solutes, metal ions and mineral surfaces (*1*). (5) They have the structural features needed to form high molecular weight noncovalently bonded supramolecular aggregates in aqueous solutions through hydrogen bonding, van der Waals forces, etc. (*1*). (6) They also exhibit the structural features needed to form high molecular weight supramolecular assemblies in soils through matrix assisted covalent interactions (e.g., chemical cross linking). (7) Their calculated bulk densities and solubility parameters agree very well with experimentally derived estimates from the literature (*59*−*62*).

The simulated ^{13}C NMR spectrum of an equimolar mixture of these model isomers compares favorably with the measured spectrum (Figure 3B).

Discussion

The impact of HAs on key environmental and biogeochemical processes cannot be overstated. Yet, despite two centuries of investigations, the fundamental question of the 3-D structures of HAs remains unresolved (*1*−*14*). This is the primary reason behind the ongoing debate about the molecular and supramolecular structures of HAs in aqueous solutions, in soils and at mineral−water interfaces. Swift and co-workers (*43, 44*) believe that HAs consist of high molecular weight macromolecules that assume random coil conformation in aqueous solutions. Conversely, Piccolo et al. (*63, 64*) view HAs as nonconvalently bonded aggregates of small molecules in aqueous solutions, whereas Wershaw and coworkers (*65, 66*) hypothesize that HAs form "micelle-like" aggregates in aqueous solutions and "membrane-like aggregates" on mineral surfaces. MacCarthy and Rice (*67, 68*), on the other hand, believe that HAs are "complex" and

"heterogeneous" mixtures of organic compounds. Although the 3-D structural models of Chelsea soil HA shown in Figure 10 do not support the viewpoints that HAs are (i) high molecular weight macromolecules (*43*, *44*) or (ii) complex mixtures of numerous organic compounds (*67*, *68*), at the present time we are not able to provide definite answers to the fundamental questions of the 3-D structures of HAs in aqueous solutions, soils, and at mineral−water interfaces.

Two key issues that need to be evaluated are the sensitivity of the SGNATURE generated models to (i) the input list of "precursor molecules" and associated interfragment bonds and (ii) the constraint on their number of carbon atoms. As previously stated, the input "precursor molecules" and interfragment bonds for Chelsea soil HA (Table 2) were selected from a library of molecular fragments derived from the qualitative DRIFT and 1-D/2-D $^{13}C/^{1}H$ NMR spectroscopic data to ensure that they could be covalently linked through condensation mechanisms known to mediate HA formation in Histosol soils. Although microbial degradation products such as polyphenols and aliphatic acids were included in the list of precursor "molecules" for Chelsea soil HA, these might not be significant building blocks for HAs isolated from the "deeper" organic layers of other Histosol soils where microbial degradation products such as peptidoglycans and cutin acids could be more prevalent. Thus, the SIGNATURE generated 3-D structural models shown in Figure 10 might be only representative of HAs isolated from the top layers of Histosol soils such as Houghton muck (*15*). Due to the need for generating 3-D structural models of Chelsea soil HA with molar mass within the observed m/z range of the ESI Qq TOF mass spectrum of Chelsea soil HA (Figure 7), the number of carbon atoms of the candidate models was constrained to vary between 40 and 75. This constraint can be readily relaxed to generate an ensemble of 3-D structural models of Chelsea soil HA of different molar mass that satisfy the structural input data given in Table 1 within specified signature errors (see Table 4). In fact, this is one the key strength of our methodology that will be used in subsequent studies to explore the "polydispersity" issue in 3-D structural modeling of HAs.

The impact of measurement uncertainties on the use of the density of bulk HAs as a criterion for model selection needs also to be assessed. From a practical, as well as fundamental, point of view, HAs can be described as amorphous materials. The bulk density (ρ) and the solubility parameter (δ) are among the most important thermodynamic properties of amorphous material systems. The bulk properties of amorphous compounds are primarily controlled by their ability to form energetically favorable 3-D close-packed molecular arrangements. Thus, 3-D structural models of HAs that do not yield accurate density and solubility parameter are questionable. Whereas the compilation of literature estimates of "humic organic matter" solubility parameters by Poerschman and Kopinke (*62*) was statistically significant (Figure 9), we found no data set of measured or estimated densities of bulk HAs large enough to enable meaningful statistical estimates of uncertainty. Consequently, the density selection criterion for the SIGNATURE generated Chelsea model isomers was relaxed by considering a relatively large interval of "acceptable" densities for bulk HAs (i.e., 1.20−1.45 g/cm^3).

We also need to assess the reliability of the computational procedures used to simulate the ^{13}C NMR spectrum of the equimolar mixture of the SIGNATURE generated Chelsea soil HA models (Figure 3B). First, the NMR module of Chem Draw Pro was employed to estimate the pertinent chemical shifts. We are not quite sure how reliable this approach is when applied to larger "molecules" such as the SIGNATURE generated structural models. Second, the ^{13}C NMR spectrum of the mixture of the Chelsea soil models was calculated using relaxation parameters T1/T2 of 40 ms as measured experimentally. Although this computational procedure seems reasonable, it has not been validated on well defined mixtures of smaller compounds: Despite these simplifying assumptions, the simulated ^{13}C NMR of the equimolar mixture of the SIGNATURE generated Chelsea soil HA models (Figure 3B) compares favorably with the measured spectrum of Chelsea soil HA (Figure 3A) except for the feature centered at 80 pm in Figure 3B. The development of improved and validated procedures for simulating the 1-D and /2-D NMR of SIGNATURE generated 3-D models of HAs will be addressed in subsequent studies.

Because our methodology is predicated upon the knowledge of the origin of the HA of interest and insight into its mechanisms of formation, HAs predominantly formed through the condensation of plant and/or microbial degradation products appear to be the ideal candidates for assessing the applicability and limitations of this methodology. Thus, its systematic application to bulk HA samples and well resolved fractions from Histosol, Mollisol, degraded wood, and peat spoils should result in the development of representative 3-D structural models. Such models could then be used in subsequent integrated experimental and computational studies to address some key questions regarding the molecular and supramolecular structures of HAs: (1) To what extent could bulk HAs be described as a polydisperse mixture of a limited number of "molecularly heterogeneous" compounds? (2) Do organic geomacromolecules such as HAs with no well-defined head and tail self-assemble in ordered micelles (*65*, *66*) or form noncovalently bonded fractal like aggregates in aqueous solutions (*5*, *10*, *69*)? (3) To what extent could HAs form high molecular weight covalently bonded supramolecular assemblies with thermodynamic and structural properties consistent with those of bulk soil HAs?

The integrated experimental and computational approach (Figure 1) described in this paper also provides means of generating the 3-D structural models needed to estimate the thermodynamic and physicochemical properties of HAs by MD simulations. Diallo et al. (*70*) have recently combined this methodology with the Flory−Huggins solution theory to obtain good estimates of the binding constants of 18 hydrophobic organic compounds (HOCs) to dissolved Chelsea soil HA *without using any empirically derived adjustable parameters* (70). We are currently assessing the extent to which the 3-D structural models of Chelsea soil HA (Figure 10) can be used in an integrated multiscale modeling framework (i.e., atomistic simulations combined with mean field statistical thermodynamics) to predict HOC sorption onto Chelsea soil HA *without using any empirically derived adjustable parameter*.

Acknowledgments

M.S.D. dedicates this paper to his wife Laura for encouraging him to tackle difficult problems in environmental science. We thank Prof. Weilin Huang of Drexel University (Philadelphia, PA) for providing the Chelsea soil humic acid samples. The theoretical component of this work (computer assisted structure elucidation and atomistic simulations) was conducted in the Department of Civil Engineering and the Institute of Multimedia Applications at Howard University and at the Materials and Process Simulation of the Beckman Institute at the California Institute of Technology. Funding for this work was provided to Howard University by the Great Lakes and Mid-Atlantic Hazardous Substance Research Center under Grant R-825540 from the USEPA Office of Research and Development. Partial funding for this research was also provided to Howard University and Caltech by the US Department of Commerce under Cooperative Agreement 7NANB8HO102. The 1-D and 2-D $^{1}H/^{13}C$ solution NMR and ESI Q-ToF MS experiments were carried out in the Department of Chemistry at The Ohio State University (OSU). Funding was provided by the National Science Foundation to the OSU Environmental Molecular Science Institute (CHE-

0089147) and Collaborative Research Agreement In Environmental Molecular Science (CHE-0089173). The DRIFT spectroscopic investigations were carried out at the W. R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at the Pacific Northwest National Laboratory (PNNL). PNNL is operated for the Department of Energy by Battelle.

Supporting Information Available

Experimental methods and procedures used to extract and characterize Chelsea soil HA and a summary of the mathematical theory behind the SIGNATURE software in the section "Computer Assisted Structure Elucidation". This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Stevenson, F. J. *Humus Chemistry: Genesis, Composition, Reactions*; Wiley: New York 1994.
(2) Jones, M. N.; Bryan, N. D. *Adv. Colloid Interface Sci.* **1998**, *78*, 1.
(3) *Humic and Fulvic Acids*: *Isolation, Structure, and Environmental Role*; Gaffney, J. S., Marley, N. A., Clark, S. B., Eds.; ACS Symposium Series 651; Washington, DC, 996, 2.
(4) *Humic substances in soil, sediment and water*; Aiken, G. R., McKnight, D. M., Wershaw, R. L., MacCarthy, P., Eds.; Wiley-Interscience: New York, 1985.
(5) Tate, R. L., III. *Soil Sci.* **1999**, *164*, 775.
(6) *Humic Substances II: In Search of Structures;* Hayes, M. H. B., MacCarthy, P., Malcolm, P. L., Swift, R. S., Eds.; Wiley: 1989; p 3.
(7) Davies, G.; Fataftah, A.; Cherkasskiy, A.; Ghabbour, E. A.; Radwan, A.; Jansen, S. A.; Kollar, S.; Paciolla, M. D.; Sein, L. T., Jr.; Buermann, W.; Balasubramanian, M.; Budnick, J.; Xing, B. *J. Chem. Soc., Dalton Trans.* **1997**, *21*, 4047.
(8) Davis, J. A. *Geochim. Cosmochim Acta* **1984**, *48*, 679.
(9) *Humic Substances: Structures, Properties and Uses*; Davies, G., Ghabbour, E. A., Eds.; Royal Society of Chemistry: Cambridge, 1998.
(10) Senesi, N. *Soil Sci.* **1999**, *164*, 841.
(11) Schluten, H. R.; Schnitzer, M. *Naturwissenschaften* **1993**, *80*, 29.
(12) Steelink, C. *Humic Substances in Soil, Sediment, and Water. Geochemistry, Isolation and Characterization*; Aiken, G. R., McKnight, D. M., Wershaw, R. L., MacCarthy, P., Eds.; Wiley: New York, 1985; p 457.
(13) Jansen, S.; Malaty, S.; Nwabara, M.; Johnson, S. E.; Ghabbour, E. A.; Davies, G.; Varnum, J. M. *Mater. Sci. Eng.* **1996**, *C4*, 175.
(14) Sein, L. T.; Varnum, J. M.; Jansen, S. A. *Environ. Sci. Technol.* **1999**, *33*, 546.
(15) USDA-NRCS Soil Survey Division. Houghton series: http //www.statlab.iastate.edu/cgi-bin/osd/osdname.cg
(16) International Humic Substances Society. http://www.ihss.gatech.edu/.
(17) Huang, W.; Weber, W. J. *Environ. Sci. Technol.* **1997**, *31*, 2562.
(18) Faulon, J. L. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1204.
(19) *Cerius*^{2} *User Guide*; Molecular Simulations: San Diego, 1997.
(20) Faulon, J. L.; Hatcher, P. G. *Energy Fuels* **1994**, *8*, 402.
(21) Diallo, M. S.; Cagin, T.; Faulon, J. L.; Goddard, W. A., III. *Asphaltenes and Asphalts. II. Developments in Petroleum Science Series, 40 B*; Yen, T. F., Chilingarian, G. V., Eds.; Elsevier Science: Amsterdam, The Netherlands, 2000; p 103.
(22) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. *J. Phys. Chem.* **1989**, *90*, 8897.
(23) Rappe, A. K.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358.
(24) Karasawa, N.; Goddard, W. A., III. *J. Phys. Chem.* **1989**, *93*, 7320.
(25) Bereendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Phys. Chem.* **1984**, *81*, 3684.
(26) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.
(27) Barton, A. F. M. *CRC Handbook of Solubility Parameters and Other Cohesion Parameters*; CRC Press: Boca Raton, FL, 1991.
(28) Cambridge Soft.com. ChemDraw Pro.
(29) Bruker Analytic GmbH. NMR-SIM Software Manual, Version 3.0, 2001.
(30) Baes, A. U.; Bloom, P. R. *Soil Sci. Soc. Am. J.* **1989**, *53*, 695.
(31) Hadzija, O.; Spoljar, B. *Fresenius J. Anal. Chem.* **1995**, *351*, 692.
(32) Inbar, Y., Chen, Y.; Hadar, Y. *Soil Sci. Soc. Am. J.* **1989**, *53*, 1695.
(33) Malcolm, R. L. *Humic Substances II: In Search of Structures*; Hayes, M. H. B., MacCarthy, P., Malcom, R. L., Swift, R. S., Eds.; Wiley: 1989; p 339.
(34) Wilson, M. A.; Plugmire, R. J.; Grant, D. M. *Org. Geochem.* **1984**, *5*, 121.
(35) Wilson, M. A.; Vassalo, A. M.; Perdue, E. M.; Reuter, J. M. *Anal. Chem.* **1987**, *59*, 551.
(36) Hatcher, P. G.; Dennis, L.; Maciel, G. E. *Org. Geochem.* **1981**, *3*, 43.
(37) Hatcher, P. G.; Schnitzer, M.; Dennis, L.; Maciel, G. E. *Soil Sci. Soc. Am. J.* **1981**, *45*, 1089.
(38) Preston, C. M. *Soil Sci.* **1996**, *161*, 144.
(39) Schnitzer, M.; Preston, C. M. *Soil Sci. Soc. Am. J.* **1986**, *50*, 326.
(40) Simpson. A. J,; Burdon, J.; Graham, C. L.; Spencer, N.; Hayes, M. H. B.; Kingery, W. L. *Eur. J. Soil Sci.* **2001**, *52*, 495.
(41) Fan, T. W. M.; Higashi, R. H.; Lane, A. *Environ. Sci. Technol.* **2000**, *34*, 1636.
(42) Kingery, W. L.; Simpson, A. J.; Hayes, M. H. B.; Locke, M. A.; Hicks, R. P. *Soil Sci.* **2000**, *165*, 483.
(43) Swift, R. S.; Posner, A. M. *J. Soil. Sci.* **1971**, *22*, 237.
(44) Swift, R. S. *Soil Sci.* **1999**, *164*, 790.
(45) Aiken, G. R.; Malcom, R. L. *Geochim. Cosmochim. Acta* **1987**, *51*, 2177.
(46) Chin, Y.; Aiken, G.; O'Loughlin, E. *Environ. Sci. Technol.* **1994**, *28*, 1636.
(47) Cabaniss, S. E.; Qunhui, Z.; Maurice, P.; Chin, Y.; Aiken, G. *Environ. Sci. Technol.* **2000**, *34*, 1103.
(48) Beckett, R.; Jue, Z.; Giddings, C. *Environ. Sci. Technol.* **1997**, *21*, 289.
(49) Brown, T. L.; Rice, J. A. *Anal. Chem.* **2000**, *72*, 384.
(50) Kujawinski, E. B.; Hatcher, P. G.; Freitas, M. A. *Anal. Chem.* **2002**, *74*, 413.
(51) Plancque, G.; Amekraz, B.; Moulin, V.; Toulhoat, P.; Moulin, C. *Rapid. Comm. Mass Spectrom.* **2001**, *15*, 827.
(52) Kramer, R. W.; Kujawinski, E. B.; Zhang, X.; Green-Church, K.; Jones, B R. B.; Freitas, M. A.; Hatcher, P. G. *Humic Substances: Structures, Models and Functions*; Davies, G., Ghabbour, E. A., Eds.; Royal Society of Chemistry: 2001; p 95.
(53) Leenheer, J. A.; Rostad, C. E.; Gates, P. M.; Furlong, E. T.' Ferrer, I. *Anal. Chem.* **2001**, *73*, 1461.
(54) Fievre, A.; Solouki, T.; Marshall, A. G.; Cooper, W. T. *Energy Fuels* **1997**, *11*, 554.
(55) Kujawinski, E. B.; Freitas, M. A.; Zhang, X.; Hatcher, P. G.; Green-Church, K. B.; Jones, R. B. *Org. Geochem.* **2002**, *33*, 171.
(56) Ralph, A.; Landucci, L. L. *NMR Database of Lignin and Cell Wall Model Compounds*; 1996. http: //www.dfrc.wisc.edu/software-.html.
(57) Dragun, J.; Helling, C. S. *Soil Sci.* **1988**, *139*, 100.
(58) Tate, R. L., III. *Soil Organic Matter. Biological and Ecological Effects*; Wiley: New York, 1987.
(59) Chiou, C. T.; Porter, P. E.; Schmedding, D. *Environ. Sci. Technol.* **1983**, *17*, 227.
(60) Orlov, D. S. *Humic Substances of Soils and General Theory of Humification*; Russian Translation Series 111, A. A. Balkema Publishers: Brookfield, VT, 1995.
(61) Brown, P. A.; Leenheer, J. A. *Humic Substances in The Suwannee River Georgia: Interactions, Properties, and Proposed Structures*; USGS, Open-File Report 87-557; 311.
(62) Poerschman, J.; Kopinke, F. D. *Environ. Sci. Technol.* **200**, *35*, 1142.
(63) Piccolo, A.; Conte, P.; Cozzolino, A. *Eur. J Soil Sci.* **1999**, *50*, 687.
(64) Piccolo, A.; Conte, P. *Adv. Environ. Res.* **1999**, *3*, 511.
(65) Wershaw, R. L. *Environ. Sci. Technol.* **1993**, *27*, 814.
(66) Wershaw, R. L. *Soil Sci.* **1999**, *164*, 803.
(67) MacCarthy, P.; Rice, J. A. *Humic Substances in Soil, Sediment, and Water. Geochemistry, Isolation and Characterization*; Aiken, G. R., McKnight, D. M., Wershaw, R. L., MacCarthy, P., Eds.; Wiley: New York, 1985; p 275.
(68) MacCarthy, P.; Rice, J. A. *Scientists on Gaia*; Schneider, S., Boston, P. J., Eds.; MIT Press: Cambridge, MA, 1991; p 329.
(69) Myneni, S. C. B.; Brown, J. T.; Martinez, G. A.; Meyer-Ilse, *Science* **1999**, *286*, 1335.
(70) Diallo, M. S.; Faulon, J. L.; Goddard, W. A., III; Johnson, J. H., Jr. *Humic Substances: Structures, Models and Functions*; Davies. G., Ghabbour, E. A., Eds.; Royal Society of Chemistry: Cambridge, 2001; p 221.

Received for review July 11, 2002. Revised manuscript received December 13, 2002. Accepted February 9, 2003.

ES0259638