ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/5665079

Microbially Derived Inputs to Soil Organic Matter: Are Current Estimates Too Low?

ARTICLE *in* ENVIRONMENTAL SCIENCE AND TECHNOLOGY · DECEMBER 2007

Impact Factor: 5.33 · DOI: 10.1021/es071217x · Source: PubMed

CITATIONS
160

READS
164

4 AUTHORS, INCLUDING:

Myrna J Simpson
University of Toronto
139 PUBLICATIONS **3,607** CITATIONS
SEE PROFILE

Emma Smith
University of the West Indies at Cave Hill, Bar…
16 PUBLICATIONS **485** CITATIONS
SEE PROFILE

Brian Kelleher
Dublin City University
44 PUBLICATIONS **809** CITATIONS
SEE PROFILE

Available from: Emma Smith
Retrieved on: 01 February 2016

Unraveling the Structural Components of Soil Humin by Use of Solution-State Nuclear Magnetic Resonance Spectroscopy

ANDRE J. SIMPSON,*,† GUIXUE SONG,‡ EMMA SMITH,† BUUAN LAM,† ETELVINO H. NOVOTNY,‡,§ AND MICHAEL H. B. HAYES‡

Department of Chemistry, University of Toronto, Scarborough Campus, Toronto, Ontario, M1C 1A4 Canada, Department of Chemical and Environmental Sciences, University of Limerick, Limerick, Ireland, and Embrapa Solos, R. Jardim Botânico, 1024, CEP 22460-000, Rio de Janeiro-RJ-Brazil

Humin is the most recalcitrant and least understood fraction of soil organic matter. By definition, humin is that fraction not extracted by traditional aqueous alkaline soil extractants. Here we show that ≥70% of the traditional humin fraction is solubilized when 0.1 M NaOH + 6 M urea and dimethyl sulfoxide (DMSO) + 6% H_2SO_4 are used in series after conventional extraction. Multidimensional solution-state NMR is applied in this study to gain an understanding of the major constituents present in these "solubilized humin fractions". The spectra indicated strong contributions from five main categories of components, namely. peptides, aliphatic species, carbohydrates, peptidoglycan, and lignin. Diffusion edited spectroscopy indicated that all species are present as macromolecules (or stable aggregate species). Although the distribution of the components is generally similar, peptidoglycan is present at significant levels supporting a higher microbial contribution to humin than to humic and fulvic fractions. The abundance of plant- and microbial-derived materials found does not exclude "humic" materials (e.g., oxidized lignin) or the presence of novel compounds at lower concentrations but suggests that a large proportion of humin is formed from classes of known compounds and parent biopolymers.

Introduction

The global soil carbon pool is 3.3 times greater than the atmospheric pool and 4.5 times greater than that of the biotic pool (*1*). At least 50% of soil organic carbon can be categorized as humic substances (HS) (*2*). Humin, by definition the fraction of HS that is insoluble in traditional aqueous alkaline extractants (NaOH in the vast majority of humic extractions), represents more than 50% of the organic carbon in mineral soils (*3*) and more than 70% of that in lithified sediments (*4*). Humin is regarded as the most recalcitrant soil organic fraction and thus remains longest in the soil environment. It is difficult both to isolate and to analyze by conventional wet chemical analyses because of its associations with the inorganic (mainly clays) and organic components. Thus humin is the least characterized fraction of soil organic matter. Because of the lack of structural information, it is practically impossible to predict how this huge reserve of sequestered carbon will respond in soil chemical interactions and in future climatic shifts.

Thus far, approaches based on solid-state NMR have arguably been the most successful in understanding the general composition of humin (*5*, *6*). In a comprehensive review of humin, Rice (*7*) has pointed out that "the work to date, has not been able to establish the nature of the binding of the OM to mineral matter or to show how the organic components interact with each other." In this study we take a step toward understanding the major structural components of humin, itself a preliminary but key step toward ultimately understanding its reactivity in the environment. This, to our knowledge, is the first time that multidimensional solution-state NMR has been applied for studies of structural components in humin. By use of a sequence of solvents, urea + base increased the amount extracted in base by 25%, and then DMSO/H_2SO_4 isolated an additional 60% of the residual materials in association with the clays, making these "soluble humin" fractions amenable to solution-state NMR studies.

Materials and Methods

Humin Isolation. Two soils were used: a grassland site at the Teagasc Research Farm, Clonroche, County Wexford, Ireland, and an adjacent plot that had been subjected for over 20 years to annual plowing and cultivation for wheat crops. Soils were first extracted exhaustively with a series of aqueous solvents of increasing pH culminating with 0.1 M sodium hydroxide, pH 12.6 [details are provided elsewhere (*8*)]. The residue was then extracted exhaustively with NaOH/urea (0.1 M/6 M), and subsequently the residue from that extraction (after dialysis to remove urea) was exhaustively extracted with DMSO + 6% (v/v) H_2SO_4. In the case of the urea extract, humic acid (HA) and fulvic acid (FA)-type fractions were separated by precipitation of the HAs at pH 1.5 with 6 M HCl, and the FAs were then isolated by use of XAD-8 resin. The HAs were dissolved in 0.1 M NaOH and diluted to a concentration <40 mg L^{-1}, and the pH was lowered to 2.5. The solution was applied to XAD-8 and treated as for the FA fraction. In the case of the DMSO/H_2SO_4 isolates, the extracts were separated by precipitation after dilution with distilled water (to pH 1.5) and the FAs were passed on to XAD-8 resin. The acid and DMSO were eluted in distilled water, and the resin retentates were back eluted in 0.1 M NaOH. The precipitated HAs were dialyzed and freeze-dried.

In this paper the isolates from the humin samples will be categorized into four types: urea FA, urea HA, DMSO FA, and DMSO HA. The labels C and UC are used to indicate origins in cultivated and uncultivated (or grassland) sites. As an example, the DMSO HA UC sample is that isolated with DMSO/H_2SO_4 from the uncultivated site and is that fraction which precipitated at low pH in aqueous solution.

Solution-State NMR Spectroscopy. Samples (20−60 mg, dependent on the amount of material available) were dissolved in DMSO-d_6 (0.6 mL) and transferred to a 5 mm NMR tube for analysis. In the case of DMSO HA, 4% D_2SO_4 (deuterated sulfuric acid) was added to the sample to ensure complete solubility; all other samples were soluble in DMSO without addition of the acid. Details of the NMR experiments can be found in Supporting Information.

* Corresponding author tel: 1-416-287-7547; fax: 1-416-287-7279; e-mail: andre.simpson@utoronto.ca.
† University of Toronto.
‡ University of Limerick.
§ Embrapa Solos.

10.1021/es061576c CCC: $37.00 © xxxx American Chemical Society
Published on Web 12/15/2006

Results and Discussion

Samples and Assignments. The organic fraction that did not dissolve in the exhaustive extraction sequence culminating in 0.1 M NaOH is considered to be humin (*2*). In order to solubilize these residual materials, further exhaustive extractions were carried out, first with a mixture of 0.1 M NaOH + 6 M urea and finally with a mixture of DMSO + 6% (v/v) H_2SO_4. The extract in urea + base amounted to an additional 25% of the organic matter that was extracted in the basic media. The fine clay trapped on the filter pad had 16% organic C (negligible amounts of organic matter were found to be associated with the residual sand and silt fractions). Extraction with DMSO/ H_2SO_4 removed ~60% of the organic matter associated with the clay. Combined, both extractants solubilized ≥70% of the traditional humin fraction (that remaining after exhaustive extraction with sodium hydroxide). This paper focuses on this soluble fraction only, and no information can be drawn as to the similarities and differences that may be found in the remaining <30% material.

All assignments offered in this paper have been made by ^{1}H NMR spectroscopy, total correlation spectroscopy (TOCSY), heteronuclear multiple quantum coherence spectroscopy (HMQC), and distortionless enhancement by polarization transfer−heteronuclear single quantum coherence spectroscopy (DEPT-HSQC), as well as extensive predictions and simulations (*9*). In specific cases, where additional confirmation was required, pulsed-field gradient spin echo−nuclear Overhauser effect spectroscopy (PFGSE-NOESY), total correlation spectroscopy (PFGSE-TOCSY), and truncated driven nuclear Overhauser effect (TNOE)−with diffusion editing (TOE-DE) were performed (see Supporting Information for experimental details). Note it has been shown that the same techniques employed here can distinguish between the presence of hypothetical "cross-linked" humic structures and structures derived from biopolymers (*10*). However, although the presence of quaternary carbons in proximity to protonated groups, for example in lipids, proteins, etc., can be inferred, none of the techniques used herein can detect quaternary carbons directly (due to lack of sample and fast relaxation making heteronuclear multiple bond correlation, HMBC, and direct carbon analysis challenging). Thus if large unbroken networks of quaternary carbons are present in the sample, as would be the case for domains in black carbon and is possible in regions of novel "humic" structures that contain no protons, these may be underestimated or not detected by the techniques employed.

Quantification of the NMR data in this paper has not been provided. The humin samples are very complex mixtures and the major components present display significant overlap. Attempts were made to deconvolute the spectral profile into its corresponding spectral components. While simple models can be constructed that describe the NMR profile very well, none of these models was able to account for any smaller peaks present (for example, the range of side-chain protons that are present in peptides (Figure 1C, region 0−3 ppm) that are masked by the more intense signals from other components. Furthermore, as the exact component mixture of the classes of compounds present is unknown, for example, it is impossible to predict the exact peptide mixtures present in the humin, it would be unfeasible to describe their contribution to the overall 1D NMR signal. The same applies for all the major categories of components present (e.g., carbohydrates, lipids). As humin is a mixture that contains components whose precise spectral envelopes cannot be determined accurately and with spectral profiles that overlap considerably (and to an unknown degree), any attempt to quantify the data would lead to inaccurate and potentially misleading results.

Where possible, in order to make this work more accessible, assignments are given on the 1D spectra, in terms of the major structural categories present (for example, peptide, lipids, carbohydrates, peptidoglycan). However, it should be emphasized that humin is a complex mixture and there are considerable overlaps in the 1D NMR spectra. In particular, the carbohydrate and aspects of the lignin signals superimpose on each other. These structures are, however, distinguishable in multidimensional NMR and thus can be assigned with confidence. The region labeled "lignin" in the 1D spectrum will also contain some contributions from carbohydrates, and vice versa. The same is true for many regions of the 1D humin spectra. Thus it is important to bear in mind that although the assignments provided on the 1D spectra will represent the predominant species present, lesser contributions from other species (determined from multidimensional NMR) will also contribute in many instances. Specific notification is given where contributions from species other than those represented by the main assignment are significant and/or of high scientific importance.

Urea Extractable Fractions. Figure 1A shows the ^{1}H NMR spectrum for the urea FA C sample. The proton spectrum of the FA is dominated by contributions from peptides (see Figure 1C; the double "hump" at ~4−4.4 ppm is indicative of peptide, especially when large methyl and H−N resonances are also present) and from cuticles, waxes, and/or lipids and lignin, all of which have previously been identified in classic humic and fulvic fractions (*10*, *11*). Signals from larger macromolecular/rigid species can be further emphasized by the use of diffusion editing. Diffusion editing "spatially encodes" molecules at the start of the experiment and then "refocuses" these at the end of the experiment. Species that diffuse or exhibit a high degree of motion during the experiment are not refocused and are essentially gated from the final spectrum (*12*). In essence, the spectrum produced will contain only signals from species that undergo little or no self-diffusion; hence structures identified will be in macromolecular and/or rigid domains. The majority of the humin signals remain after diffusion editing (see Figure 1B). The "diffusion gate" employed in this study was carried out with 53.5 gauss/cm encoding and decoding gradients (see Supporting Information for details). In other studies employing similar experimental conditions, nearly all signals are destroyed and only those from macromolecules survive (*13*). Similarly, in classic humic and fulvic materials, the vast majority of signals would be greatly attenuated under these conditions (*12*). This disparity tells us that the components in the humin fraction are of much greater size that those previously seen in FA and HA extracts and likely are in the form of macromolecules and/or very stable aggregates (*11*). There are no major changes in the relative intensities of various humin components after diffusion editing (some signal from DMSO and urea remains after diffusion editing, presumably from species in close association or trapped within the humin structures), yet signals 8, 10, and 11 are slightly attenuated (use, for example, the top of the lignin methoxyl resonance 6 as a reference to observe the attenuation of 8, 10, and 11). These signals are consistent with aliphatic structures. Previously waxes and cutins derived from plants have been identified in abundance in humic extracts (*14*), and these are likely to be preserved because of their cross-linked recalcitrant structure, hydrophobicity, and potential associations with clays (*15*). However, when cutin components are dominant, the main chain (CH_2) is generally very large (i.e., it dominates the spectrum), whereas peaks from linkages such as esters and ethers are generally small (*16*). This was not the case here, and advanced experiments indicate that some of the lipids present are associated with peptide, thus supporting the presence of lipoprotein in the sample (see next paragraph for discussion). It is important

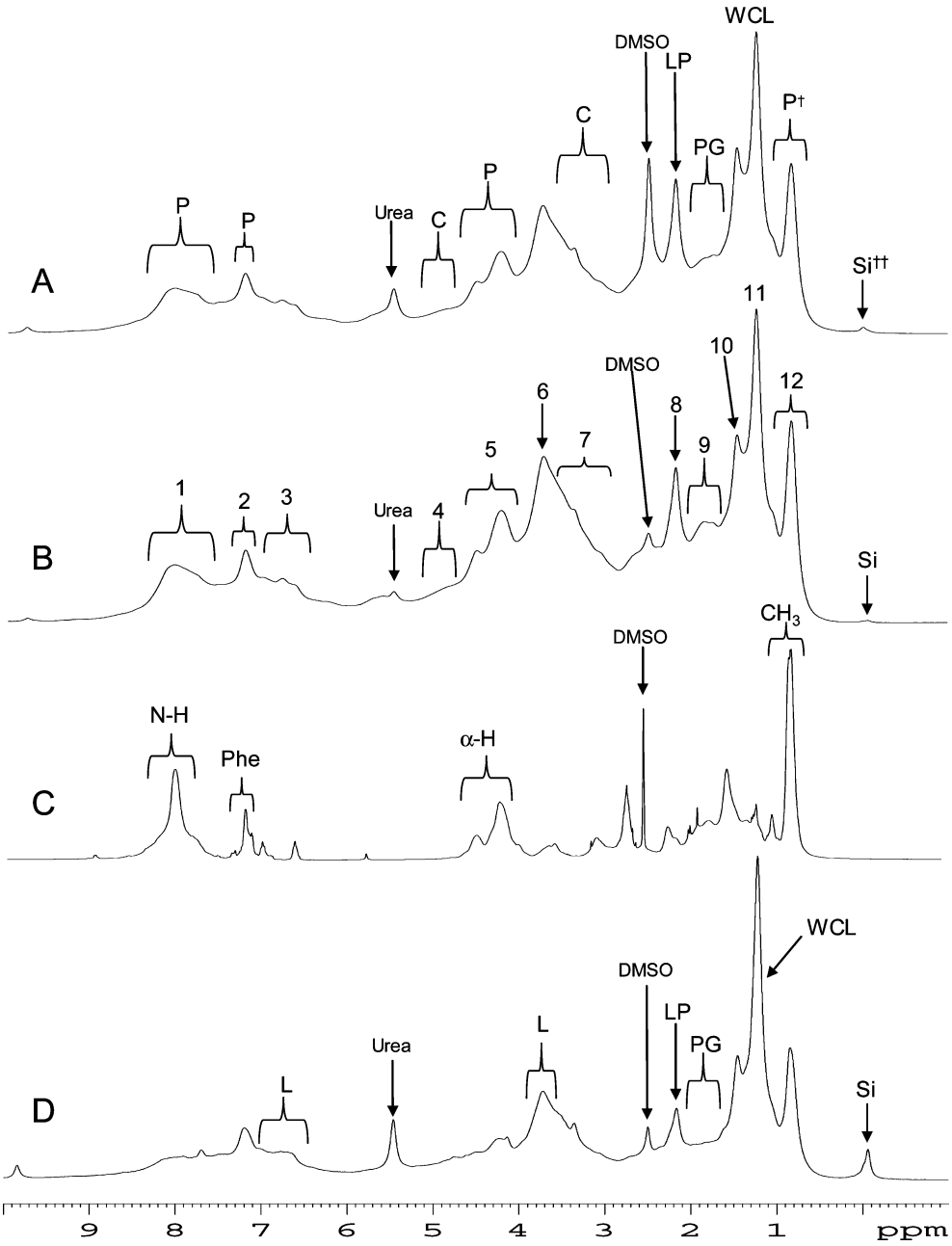

FIGURE 1. ^{1}H NMR spectra for (A) urea FA C, (B) diffusion edited ^{1}H spectrum of urea FA C, (C) bovine serum albumin, and (D) diffusion edited ^{1}H spectrum of urea HA C. Simple assignments, shown for spectrum A indicate strong contributions from P, proteins/peptides; L, lignin; C, carbohydrate; WCL, waxes, cutins, and lipids; PG, peptidoglycan; and LP, lipoprotein/lipids (note: while there is strong evidence to support the presence of lipoprotein in the sample, it is not clear how much of the signal is due to lipoprotein and how much from lipids). More specific assignments shown for spectrum B refer to 1, amide; 2, phenylalanine; 3, aromatics in lignin; 4, anomeric protons (carbohydrates); 5, α-protons (peptides); 6, methoxyl (lignin); 7, other carbohydrate protons; 8, $R_2-OCO-\boldsymbol{CH_2}-R_1$ methylene adjacent to a carbonyl (some of which appears to be in the form of lipoprotein, in which case R_2 would be a protein); 9, *N*-acetyl group in peptidoglycan; 10, aliphatic methylene units β to an acid or ester; 11, aliphatic methylene $(CH_2)_n$; 12, CH_3 (note: when this peak is large relative to 11, it often indicates strong contributions from peptides as in these examples; see spectrum C). In spectrum C only the dominant peptide signals are labeled. In spectrum D, assignments are as above; only key resonances are labeled. Si indicates a natural silicate species and not TMS (a commonly used NMR reference standard).

to note that while the presence of lipoprotein is supported, it is not clear exactly how much of the R−O−OC−CH_2−R peak in the various humins is present as lipoprotein. Furthermore, the presence of some plant-derived waxes/cutins in lower abundance cannot be ruled out, and at potentially higher abundances in other samples (e.g., DMSO HA fraction, Figure 3B).

Figure 2 depicts spectra from some of the advanced experiments used to support the presence of lipoprotein. Resonance 1 arises from the methylene CH_2 adjacent to an ester/acid (*16*). This can be confirmed through a 1D pulsed-field gradient spin echo−total correlation experiment. (PFGSE-TOCSY). This 1D analogue of the 2D NMR TOCSY experiment results in much higher sensitivity and resolution per unit time and provides information on the moieties that are directly bonded to the resonance of interest. Figure 2B shows results of this experiment and clearly indicates that unit 1 (Figure 1B, peak 8) is part of an aliphatic chain, as would be the case for a lipoprotein and lipid. It is important to note that the 1D PFGSE-TOCSY can only identify couplings

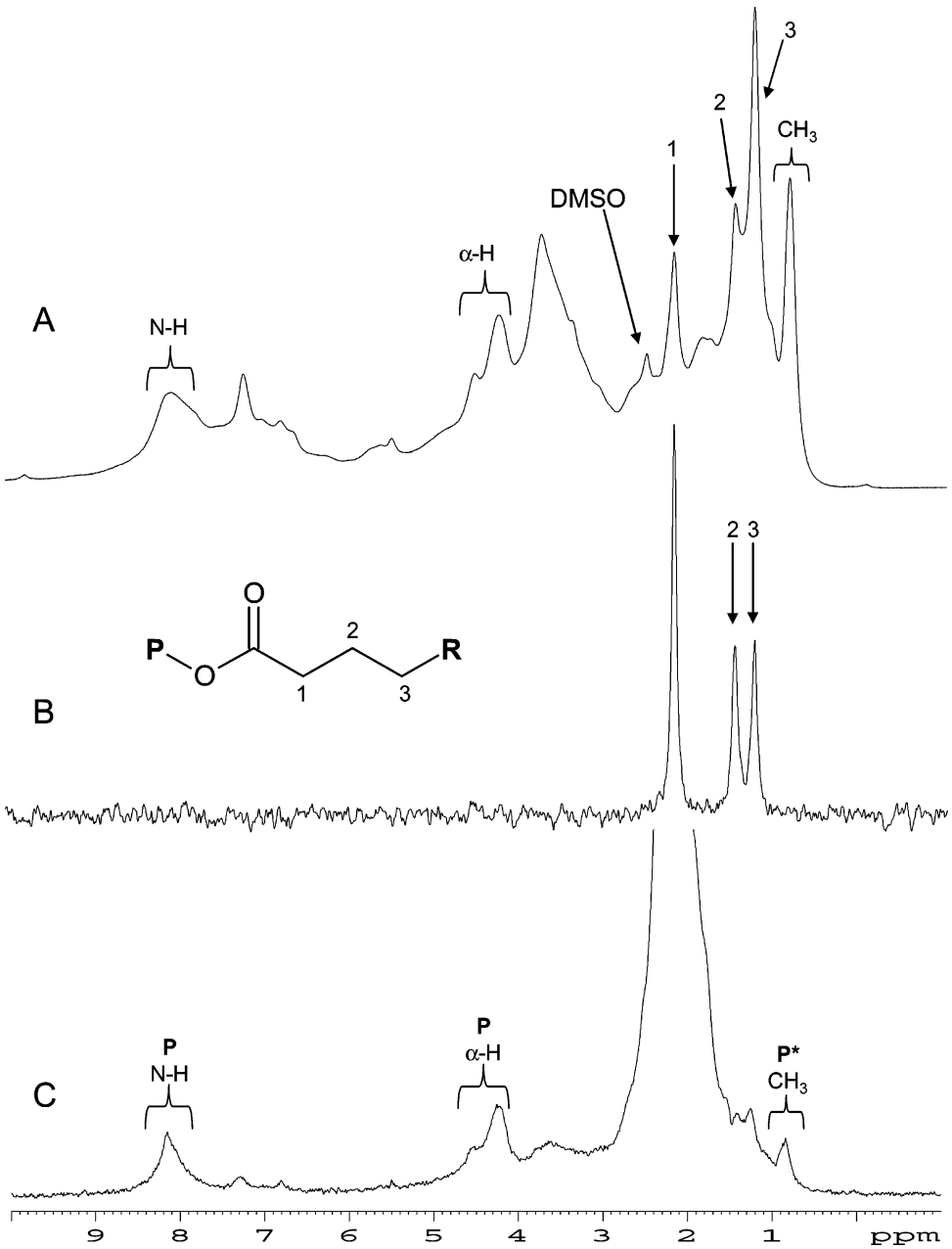

FIGURE 2. (A) ^{1}H diffusion edited NMR of urea FA C; (B) 1D PFGSE-TOCSY with selective excitation of peak 1; (C) 1D TOE-DE spectrum with selective saturation of peak 1. Assignments in spectra A and B are as follows: 1, R−OCO−*CH*$_2$−R methylene unit adjacent to the carbonyl in lipids (including lipoproteins and cutins); 2, aliphatic methylene units β to an acid or ester; 3, aliphatic methylene $(CH_2)_n$. Assignments in spectrum C highlight the major units from peptide (see Figure 1C). P in the inset represents a peptide, as would be the case for a lipoprotein (in the case of a lipid this P would be a proton).

between bonded protons in the same spin system. As methylene group 1 (Figure 2B, inset structure) is adjacent to an ester (which interrupts the spin system) it is not possible to determine what this ester is connected to by use of 1D PFGSE-TOCSY. For this the truncated driven nuclear Overhauser effect−diffusion edited (TOE-DE) experiment is well suited. The TOE-DE experiment efficiently detects spatial interactions between units in macromolecules. When selectively saturated (50 ms saturation period), the methylene shows strong correlations to the α-protons, CH_3 (in select amino acid side chains), and amide protons in peptides consistent with the presence of lipoprotein. Lipoprotein is structurally diverse, is a key component of bacterial cells, and is released during bacterial growth (*17*). The exact structure (or contributions) of the lipoprotein cannot be determined, but it is likely that a portion of this material may be from soil microbes. Microbial contributions are supported by the presence of signal 9 in Figure 1B. This can be assigned to peptidoglycan (see Figure 4B and Supporting Information, Figure S1) that comprises up to 90% by weight of gram-positive bacteria and is the key structural component in all microbial cell walls. It is important to note that peptidoglycan does contain a small "peptide linker" in its structure. It is not possible to accurately quantify contributions of peptide in the form of peptidoglycan in the various samples due to spectral overlap. Rough estimations, based on deconvolution of the spectral profile (which by its very nature ignores smaller hidden resonances), indicate only a small fraction of the peptide is in the form of peptidoglycan and the majority of the peptide in humin (likely >80% in most samples) will be from sources other than peptidoglycan. It is logical that humin will contain contributions from peptidoglycan, as this material is by nature's design (as microbe cell walls) resistant to many soil chemical and biological processes. Furthermore,

it is well-known that microbes have a strong association with soil minerals and are often in close association with the clay components (*18*). Thus it is not surprising that the signature from their cell walls will be apparent in the clay-rich humin fractions.

The components in the urea HA are similar to those in the urea FA, although the ratios are different (see Figure 1D). The contribution from peptide is less prominent (note the lack of the clear "double hump" at ~4−4.4 ppm), and the lower contributions from amide protons. This is consistent with the solubility of peptide in aqueous urea solutions. Urea is known to be able to lyse bacterial cells and solubilize cellular proteins (*19*). Once released from the cells, many peptides/proteins are soluble at all pH values and will thus remain in the FA fraction. The same applies for many lipoproteins, and while lipoprotein is present in the HA fraction it is likely less than observed for the FA (indicated by the resonance labeled LP; compare with the same resonance in Figure 1D). The levels of lignin in the humic and fulvic acids appear to be relatively similar, and their distributions will likely be determined by sizes, degrees of oxidation, and states of aggregation. Conversely, the contribution from main chain (CH_2) is much higher, but this could simply reflect that aliphatic materials accumulate in the HA fraction, consistent with their solubility. It is not possible to determine whether these are simply longer chain lipids derived from microbial sources or additional contributions from aliphatic biomaterials from plants. Throughout this paper it is generally possible to identify the main types of structures present, but it is much more difficult to identify the sources of these materials in terms of microbial or plant sources. This will require more advanced NMR (e.g., H−C−P experiments to identify microbial phospholipid signatures), combined with liquid or gas chromatography−mass spectrometry (LC-MS and GC-MS) based biomarker analysis. These will be the focus of future studies.

A detailed comparison between Figures 1D and 3A will not be undertaken since the aim of this paper is to identify the structural components in the humin isolates, not to determine the effects of cultivation on the humin. Briefly, the relative increase in peptidoglycan in the uncultivated site supports higher contributions from microbial biomass in this site. This is to be expected as inputs from plant biomass were lower, and the soil organic matter was less than 50% of that in the uncultivated (grassland). Thus, the inputs from the plant biomass were gradually depleted by the soil microbial activity, and hence the contributions from microbial tissues increased. In addition there is also a large contribution from polymethylene in the cultivated site. Accumulations of cutins and waxes with a higher resistance to microbial degradation may account for these increases.

DMSO/H_2SO_4-Extractable Fraction. Figure 3B shows the DMSO HA fraction isolated from the clay fraction of the uncultivated soil. The large contributions from long-chain CH_2 are clear, and the ratio observed between the main chain $(CH_2)_n$ and functionalities (R−CH_2−COOR) is closer to that in aliphatic biopolymers (such as cutins) than is observed for the other humin fractions. Additional 1D PFGSE-TOCSY experiments were carried out to study the functionalities in close association with the long-chain material (data not shown). Clear couplings were seen to esters, ethers, and double bonds, and ratios were similar to those present in cutin material. However, the data are not definitive, as similar functionalities could exist in a microbial-derived mixture (containing lipids, phospholipids, etc.). Thus it is only possible to state that *if* cutin contributions are significant in humin they are most likely to be concentrated in the DMSO HA fraction. It has recently been shown that aliphatic species in humic and fulvic mixtures preferentially bind to clay surfaces (*15*), which may partially explain the highly aliphatic nature of clay−organo complexes seen in soils (*20*). Thus, irrespective of source of the aliphatic materials, it would appear that their interactions with clay surfaces are likely to be at least partially responsible for their preservation in the soil environment. Peptides/proteins also contribute to this fraction, clearly indicated by the double hump at ~4−4.4 ppm, and the large CH_3 resonance (too large to be from aliphatic material alone). Because deuterated acid was added to dissolve the material, the N−H signals are not evident as these were exchanged to N−D. In another recent publication it has been shown that proteins/peptides can also exhibit a strong association with clay surfaces (*21*). Again the source of the peptide is unclear, but it is likely that the clay provides a protective surface for the preservation of this labile material. Further work is needed to confirm this as the peptide materials may arise from the release of cellular proteins/peptides as a result of lysis by DMSO.

Figure 3C shows the spectrum for DMSO FA for the grassland soil. This fraction is dominated by lignin residues, which are superimposed onto a strong contribution from carbohydrates. Figure 4A shows the HMQC spectrum of the DMSO FA sample. For more information on general HMQC assignments see Simpson 2001 (*22*); for specific aromatic lignin assignments see Simpson et al. (*9*) and Lu and Ralph (*23*); and for specific aliphatic assignments see Deshmukh et al. (*16*). Region 9 results from the very strong Ar−O−CH_3 residue in lignin. Back projections and summations across the corresponding 2D planes (not shown) indicate that the apex of the large "triangular" resonance at ~3.7 ppm in the 1D spectrum is mainly from the methoxyl of lignin (although it is superimposed upon strong carbohydrate contributions). This is supported by 1D data (see Figure 3C,D). Through close comparison of the humin and lignin spectra, it is clear that the apex of the central region (L, Figure 3A,C) in humin is from methoxyl in lignin and that the right-hand shoulder is mainly from carbohydrates. Clearly the presence of lignin is a strong indicator of plant inputs. We can only assume the large contribution in the DMSO FA fraction is due to partially oxidized and functionalized lignin. Once released from the soil, this material is relatively soluble in aqueous basic and acidic solutions. It can be assumed that the lignin components are less oxidized/degraded than those isolated in the aqueous base solvents, and these would be expected to be of higher molecular size than those in the aqueous extracts. Alternatively, the lignin components could be associated with, or trapped by, the components associated with the clays. It is not clear whether this lignin was oxidized naturally (purely by humification processes) or by the cleavage of labile bonds by the DMSO/H_2SO_4 solvent system. Further work will be needed to resolve this.

Assignment of Peptidoglycan. There is a larger contribution in the DMSO FA sample from *N*-acetyl in peptidoglycan (compared to the urea FA). Peptidoglycan is a strong indicator of microbial inputs. Figure 4B shows that in two dimensions the *N*-acetyl group can be clearly separated in an HMQC and correlates very well with 2D predictions of both the ^1H and ^{13}C chemical shifts (2D simulations not shown). DEPT-HSQC (which simply tell us whether peaks in a 2D experiment are from CH_3, CH_2, or CH) further confirms this to be an *N*-acetyl group, and 1D PFGSE-TOCSY shows no correlations, as expected (note the spin system of the CH_3 in the *N*-acetyl is isolated by the adjacent quaternary carbon and so no correlation is expected in a TOCSY) (data sets not shown). Both 1D TOE-DE and 1D PFGSE-NOESY with selective irradiation at the *N*-acetyl group show a strong correlation with the carbohydrate region, indicating that the N−CO−CH_3 is part of a rigid system and the carbohydrate structures are in close proximity. Supporting Information, Figure S1-C, shows the PFGSE-NOESY as an example. Additionally, Figure S1-B shows the 1D NMR simulation for the carbohydrate

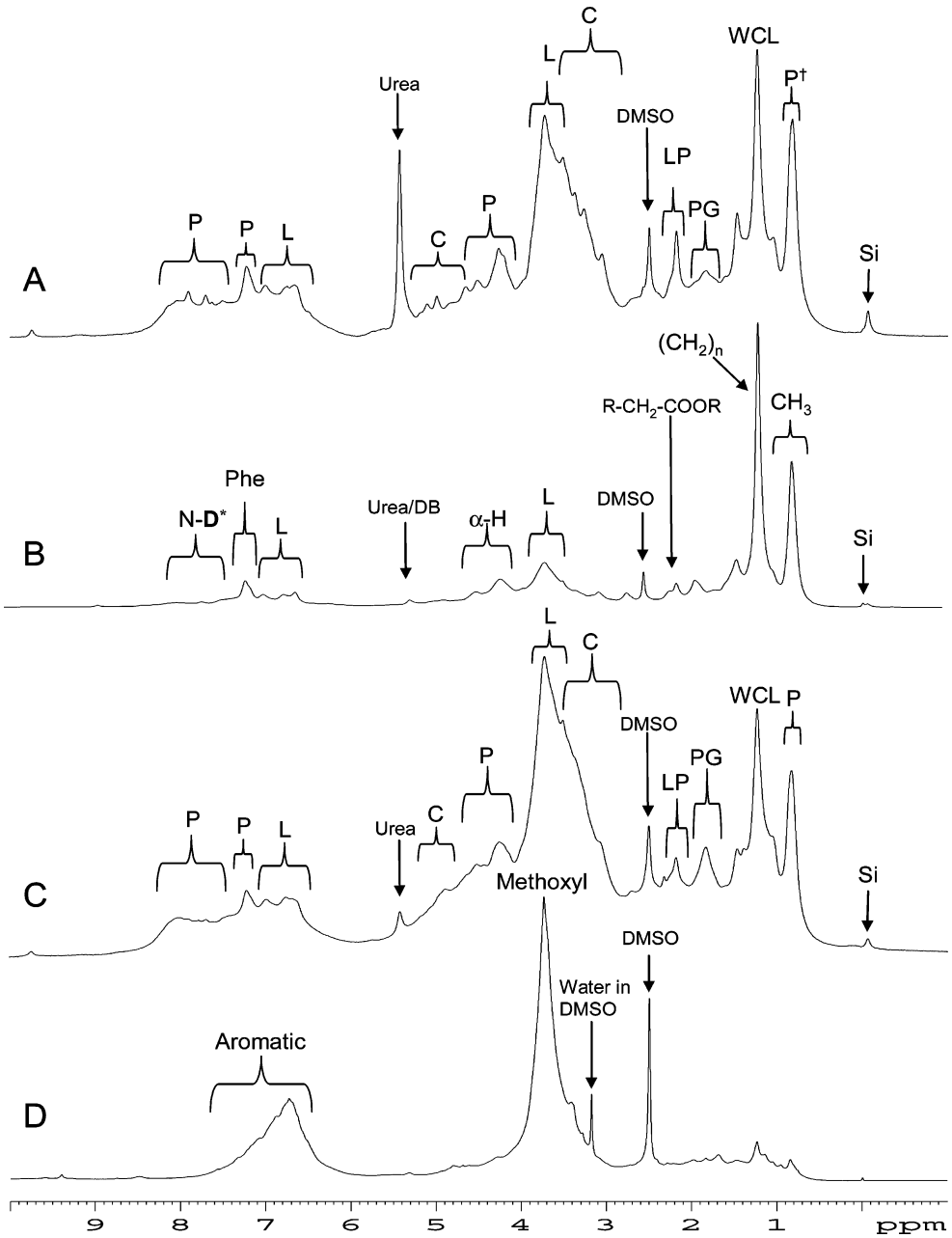

FIGURE 3. ^{1}H diffusion edited NMR spectra of (A) urea HA UC, (B) DMSO HA UC, (C) DMSO FA UC, and (D) KRAFT Lignin (Aldrich). Unless indicated otherwise, assignments are the same as shown in Figure 1. Note that the peak marked with a dagger is too large to be attributed to aliphatic species alone and indicates a substantial contribution from peptide; see Figure 1C. It is very important to note that humin HA is only soluble in DMSO in the presence of a strong acid. D_2SO_4 was added, which in addition to solubilizing the humin also deuterium-exchanged the N−H (to N−D) so the amide resonance in humin HA is strongly attenuated.

portion of the carbohydrate backbone of peptidoglycan (see Figure S1-B inset for structure). Note this simulation is only for six residues and approximates the carbohydrate portion only. As different bacteria have different peptide residues attached to the *N*-acetylmuramic acid residue (in the PEP position; see Figure S1-B inset), it is impossible to accurately predict all the peptide side arms that could potentially be present in soil humin, and thus the simulations concentrate only on the carbohydrate portion that is consistent in the vast majority of bacterial peptidoglycan. It is clear from Figure S1-C that there is a "through-space" correlation between the *N*-acetyl and the carbohydrate residues, strongly supporting the assignment of peptidoglycan. The N−H correlation in the PFGSE NOESY (Figure S1-C) will partially result from spatial interactions between the N−H group in the sugar residue (see structure in Figure S1-B) and the N−H groups in the peptide chain (see PEP in structure, Figure S1-B). Note in DMSO all N−H groups are under constant exchange (with other exchangeable groups in the sample and water in the solvent) and tend to resonate at ~8.1 ppm. Correlations between the *N*-acetyl and the α-protons in the peptide linker are seen in the TOD-DE experiment (see Figure S1-D).

In summary, the humin samples studied contain varying contributions from five main categories of structures, namely, peptide, aliphatic species (including contributions from lipoproteins, lipids, and potentially cuticular species), carbohydrates, peptidoglycan (the main structural components in bacterial cell walls), and lignin. Lignin is clearly identifiable as a plant input in humin materials, while peptidoglycan is indicative of microbial inputs. It is not possible in the present

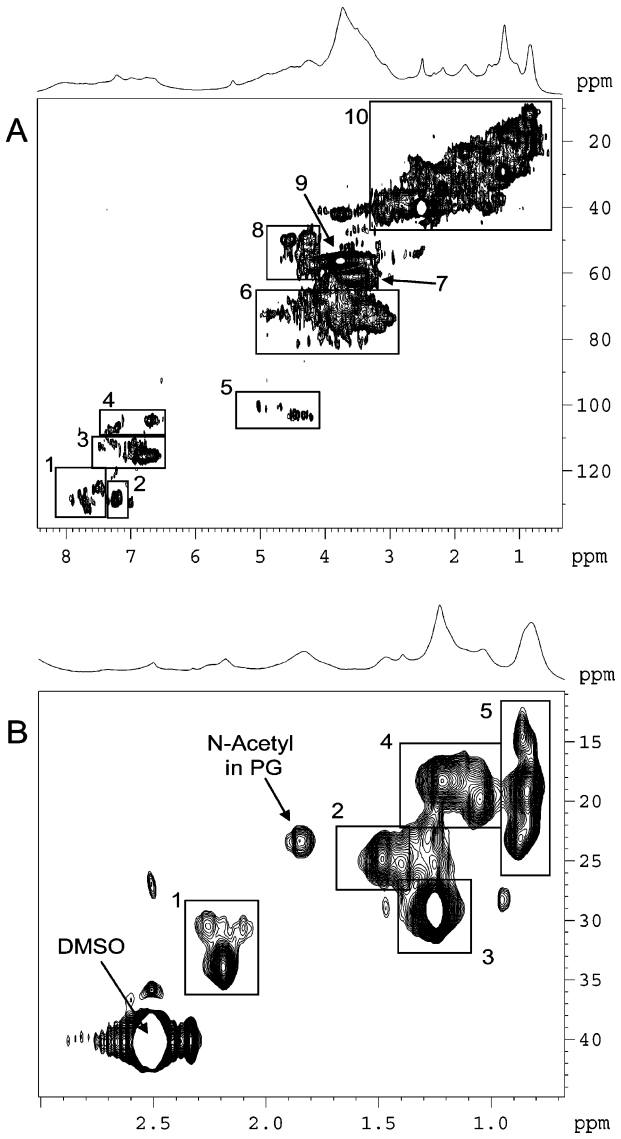

FIGURE 4. HMQC of the DMSO FAFA UC: (A) complete spectrum; (B) expansion of the aliphatic region. Assignments in spectrum A are as follows: 1, aromatic protons in *p*-hydroxybenzoates (lignin); 2, phenylalanine (in peptide); 3, mainly aromatic protons adjacent to an Ar−OR functionality in lignin; 4, units in syringyl units (lignin); 5, anomeric protons (carbohydrates); 6, other CH in carbohydrates; 7, CH_2 in carbohydrates; 8, α-protons in peptides and proteins; 9, methoxyl in lignin; 10, aliphatic linkages including signals from various lipids, and side-chain protons in peptides. Assignments in spectrum B are as follows 1, R−OCO−CH_2−R methylene unit adjacent to the carbonyl in lipids (including lipoproteins and cutins); 2, methylene units in an aliphatic chains β to an acid or ester; 3, methylene $(CH_2)_n$ in aliphatic chains, 4, aliphatic methylene γ to an acid or ester; 5, CH_3 (a small contribution in this region will be from terminal CH_3 from lipids; however, the majority of signals are from peptides [indicated by the distribution of ^{13}C shifts common in proteins (*9*)].

studies to positively identify the source of the peptide, aliphatic or carbohydrate species. From the size of the *N*-acetyl peak in peptidoglycan, it is clear this peptidoglycan alone cannot account for all the carbohydrates present. The solubilized humin materials have components similar to those in the more traditional humic and fulvic acid fractions (*10*), with the exception that peptidoglycan is present at significant levels and these humin fractions appear to be more macromolecular in nature than their traditional HA and FA counterparts. The presence of bacterial peptidoglycan is not surprising because the humin fraction is known to be strongly associated with clays and because bacteria are known to associate strongly with clay minerals. With this consideration, this it is conceivable that the clay component provides protection for microbial species that ultimately contribute, in terms of biomass, to the operationally defined humin fraction. In addition, it is feasible that some of the more labile components, especially proteins/peptides, may be preserved through association with the clays. However, further work will be needed to determine if that is the case or if the peptide is present due to the release of cellular proteins during lysis by DMSO and urea.

These initial studies clearly demonstrate that solution-state techniques have a key role to play in understanding the structural domains in humin. However, extracting humin with harsh solvents (for example, DMSO/H_2SO_4) can potentially functionalize and modify the humin structure. While it is clear the main categories of biopolymers identified in this study are major constituents in humin, the exact state in which they exist is not clear (for example, are they highly oxidized, and if so, is this from humification processes or modification during extraction?). Furthermore, extraction disrupts clay−organic associations, producing material with reactivities different than those found in nature. However, assignments of the major components in solution-state NMR spectra of the extractable components are critical. Once assigned, this information can be used in combination with state-of-the-art high-resolution (HR-MAS) NMR (which utilizes solution-state experiments to study "swellable" materials) to study humin in situ. Such applications have already been central to understanding clay−organic associations (*24*) and interfacial chemistry and reactivity in whole soils (*21*, *25*). The studies outlined here assign the major resonances in the solution-state NMR of humin and form the first stepping stone in a long path toward understanding the environmental reactivity of humin in the environment.

Acknowledgments

We thank the Canadian Foundation of Climate and Atmospheric Sciences (Fund GR-520), the Science Foundation Ireland, and Environmental Protection Agency—Ireland for funding this research. We thank ACD Labs for the donation of the prediction software that was employed extensively in these studies.

Supporting Information Available

NMR experimental details and a figure supporting the assignment of peptidoglycan. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Lal, R. Soil carbon sequestration impacts on global climate change and food security. *Science* **2004**, *304* (5677), 1623−1627.

(2) Hayes, M. H. B.; McCarthy, P.; Malcolm, R. L.; Swift, R. S. *Humic Substances II in search of structure*; Wiley: New York, 1989.

(3) Kononova, M. M. *Soil organic matter. Its nature, its role in soil formation and in soil fertility*, 2nd ed.; Pergamon Press: Oxford, U.K., 1966.

(4) Hedges, J. I.; Keil, R. G. Sedimentary Organic-Matter Preservation—an Assessment and Speculative Synthesis. *Mar. Chem.* **1995**, *49* (2−3), 81−115.

(5) Almendros, G.; Polo, A.; Dorado, E. Isolation and Characterization of the Non-Extractable Humin Fraction from Different Soils. *Agrochimica* **1981**, *25* (5−6), 479−491.

(6) Preston, C. M.; Newman, R. H. Demonstration of Spatial Heterogeneity in the Organic-Matter of De-Ashed Humin Samples by Solid-State C-13 CPMAS NMR. *Can. J. Soil Sci.* **1992**, *72* (1), 13−19.

(7) Rice, J. A. Humin. *Soil Sci.* **2001**, *166* (11), 848−857.
(8) Simpson, A. J.; Watt, B. E.; Graham, C. L.; Hayes, M. H. B. Humic substances from podzols under oak forest and a cleared forest site. I. Isolation and characterization. In *Humic Substances in Soils, Peats, and Waters*; *Spec. Publ.—R. Soc. Chem.* **1997**, *172*, 73−82.
(9) Simpson, A. J.; Lefebvre, B.; Moser, A.; Williams, A.; Larin, N.; Kvasha, M.; Kingery, W. L.; Kelleher, B. Identifying residues in natural organic matter through spectral prediction and pattern matching of 2D NMR datasets. *Magn. Reson. Chem.* **2004**, *42* (1), 14−22.
(10) Kelleher, B. P.; Simpson, A. J. Humic substances in soils: Are they really chemically distinct? *Environ. Sci. Technol.* **2006**, *40* (15), 4605−4611.
(11) Simpson, A. J.; Kingery, W. L.; Hayes, M. H. B.; Spraul, M.; Humpfer, E.; Dvortsak, P.; Kerssebaum, R.; Godejohann, M.; Hofmann, M. Molecular structures and associations of humic substances in the terrestrial environment. *Naturwissenschaften* **2002**, *89* (2), 84−88.
(12) Simpson, A. J. Determining the molecular weight, aggregation, structures and interactions of natural organic matter using diffusion ordered spectroscopy. *Magn. Reson. Chem.* **2002**, *40*, S72−S82.
(13) Kelleher, B. P.; Simpson, M. J.; Simpson, A. J. Assessing the fate and transformation of plant residues in the terrestrial environment using HR-MAS NMR spectroscopy. *Geochim. Cosmochim. Acta* **2006**, *70* (16), 4080−4094.
(14) Simpson, A. J.; Kingery, W. L.; Hatcher, P. G. The identification of plant derived structures in humic materials using three-dimensional NMR spectroscopy. *Environ. Sci. Technol.* **2003**, *37* (2), 337−342.
(15) Simpson, A. J.; Simpson, M. J.; Kingery, W. L.; Lefebvre, B. A.; Moser, A.; Williams, A. J.; Kvasha, M.; Kelleher, B. P. The Application of ^{1}H High-Resolution Magic-Angle Spinning NMR for the Study of Clay−Organic Associations in Natural and Synthetic Complexes. *Langmuir* **2006**, *22* (10), 4498−4503.
(16) Deshmukh, A. P.; Simpson, A. J.; Hatcher, P. G. Evidence for cross-linking in tomato cutin using HR-MAS NMR spectroscopy. *Phytochemistry* **2003**, *64* (6), 1163−1170.
(17) Zhang, H. W.; Niesel, D. W.; Peterson, J. W.; Klimpel, G. R. Lipoprotein release by bacteria: Potential factor in bacterial pathogenesis. *Infect. Immun.* **1998**, *66* (11), 5196−5201.
(18) Marshall, K. C. Clay Mineralogy in Relation to Survival of Soil Bacteria. *Annu. Rev. Phytopathol.* **1975**, *13*, 357−373.
(19) Hum, W.; Xu, Z.; Malakian, K.; Wolfrom, S.; Bapat, A.; Stahl, M. Extraction of *E. coli* expressed ER-alpha-LBD inclusion bodies with low concentration of urea produces homogenous protein that is easily crystalized. *Protein Sci.* **2004**, *13*, 227−227.
(20) Schnitzer, M.; Ripmeester, J. A.; Kodama, H. Characterization of the organic matter associated with a soil clay. *Soil Sci.* **1988**, *145* (6), 448−54.
(21) Feng, X. J.; Simpson, A. J.; Simpson, M. J. Chemical and mineralogical controls on humic acid sorption to clay mineral surfaces. *Org. Geochem.* **2005**, *36* (11), 1553−1566.
(22) Simpson, A. Multidimensional solution state NMR of humic substances: a practical guide and review. *Soil Sci.* **2001**, *166* (11), 795−809.
(23) Lu, F. C.; Ralph, J. Non-degradative dissolution and acetylation of ball-milled plant cell walls: high-resolution solution-state NMR. *Plant J.* **2003**, *35* (4), 535−544.
(24) Simpson, A. J.; Simpson, M. J.; Kingery, W. L.; Lefebvre, B. A.; Moser, A.; Williams, A. J.; Kvasha, M.; Kelleher, B. P. The application of H-1 high-resolution magic-angle spinning NMR for the study of clay−organic associations in natural and synthetic complexes. *Langmuir* **2006**, *22* (10), 4498−4503.
(25) Simpson, A. J.; Kingery, W. L.; Shaw, D. R.; Spraul, M.; Humpfer, E.; Dvortsak, P. The application of H-1 HR-MAS NMR spectroscopy for the study of structures and associations of organic components at the solid−aqueous interface of a whole soil. *Environ. Sci. Technol.* **2001**, *35* (16), 3321−3325.

Received for review July 3, 2006. Revised manuscript received November 2, 2006. Accepted November 6, 2006.

ES061576C