Lehrstuhl für Ökologische Chemie und Umweltanalytik der Technischen Universität München in Freising-Weihenstephan

# Molecular level structural analysis of natural organic matter and of humic substances by NMR spectroscopy

## Habilitationsschrift

vorgelegt von

Dr. Norbert Hertkorn

Zur Erlangung der Lehrbefähigung für das Fachgebiet

## Ökologische und analytische Chemie

Am Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München-Weihenstephan

Januar 2006

*your true ′land′ is the place where you are going, not the place where you are now*
*(Rumi)*

# Molecular level structural analysis of natural organic matter and of humic substances by NMR spectroscopy

## 1. ROLE AND SIGNIFICANCE OF NATURAL ORGANIC MATTER (NOM)

Prebiotic organic molecules are generally considered essential precursors of biological life (Anet et al., 2004; Hughes et al., 2004). Complex mixtures of organic molecules have been formed ever since the earliest stages of the earth's history and they were quickly distributed across all environmental compartments (Figure 1); analogous molecular assemblies have also been observed in various extraterrestrial systems. Any knowledge of chemical structures of this complex natural organic matter (NOM) depends on an educated application of modern analytical chemistry. From that basis, concepts could be elucidated and advanced, which depicted suitable conditions for the origin and development of primitive and higher forms of life, in which NOM in its various appearances has always played a crucial role (Figure 2); NOM occurrence, structure and interactions will remain forever indispensable to maintain and support diverse and highly variable ecosystems and habitats on earth.

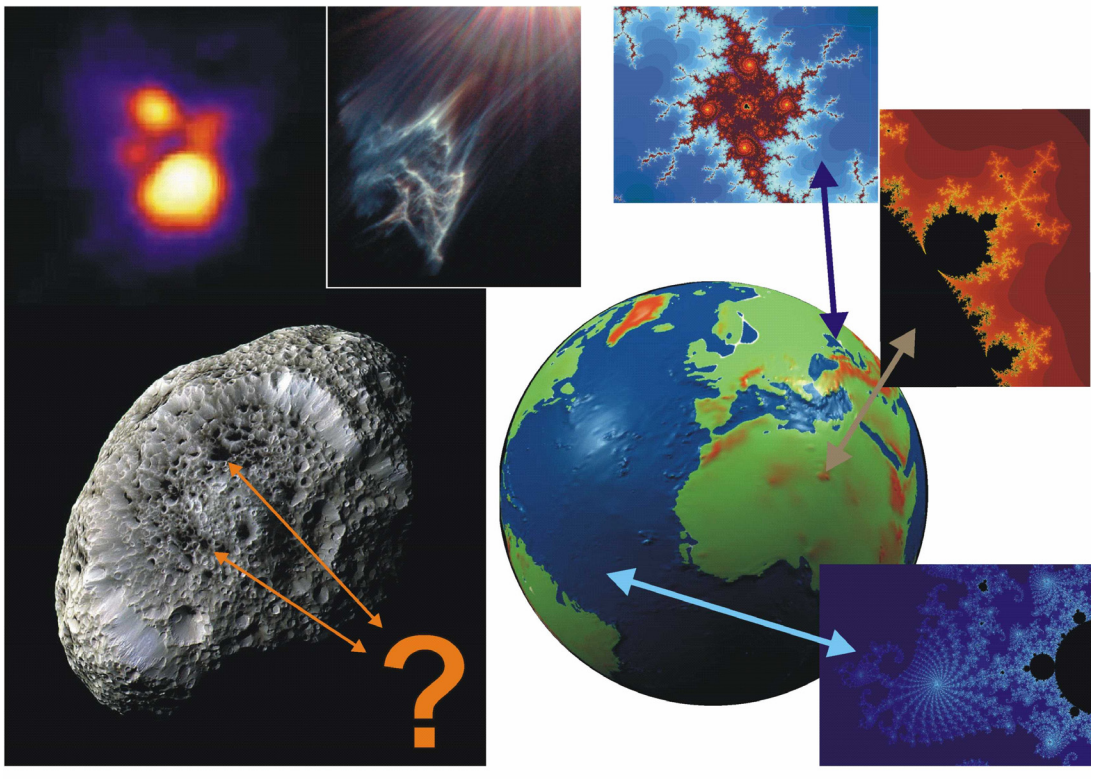



**Figure 1.** Natural organic matter (NOM) on earth is abundant in marine, humic and terrestrial ecosystems; in all these environments, NOM is comprised of an extremely complex array of chemical structures and interactions across a large range of time- and size- scales. Extraterrestrial complex organic molecules have been recently identified, e.g. on the surface of the Saturn moon Hyperion in the solar system (lower left; image from the space-craft Cassini), in interstellar molecular clouds [top middle; reflection nebula IC 349 (Barnard´s Merope nebula), Pleiades star cluster; Hubble space telescope], and huge amounts of carbon- rich molecules are expelled from giant red 'carbon stars' with how surface temperature (carbon star ICR+10216; CW Leo); dust layers around such celestial bodies have been very recently resolved with high-resolution interferometry (Weigelt et al., 1998).

The formation of NOM on earth preceded the evolution of life, initially under conditions of intense high-energy radiation and subsequently beneath a reducing atmosphere. From early stages in the earth's history, coevolution occurred between prebiotic/abiotic molecules, NOM and primitive and higher forms of life.

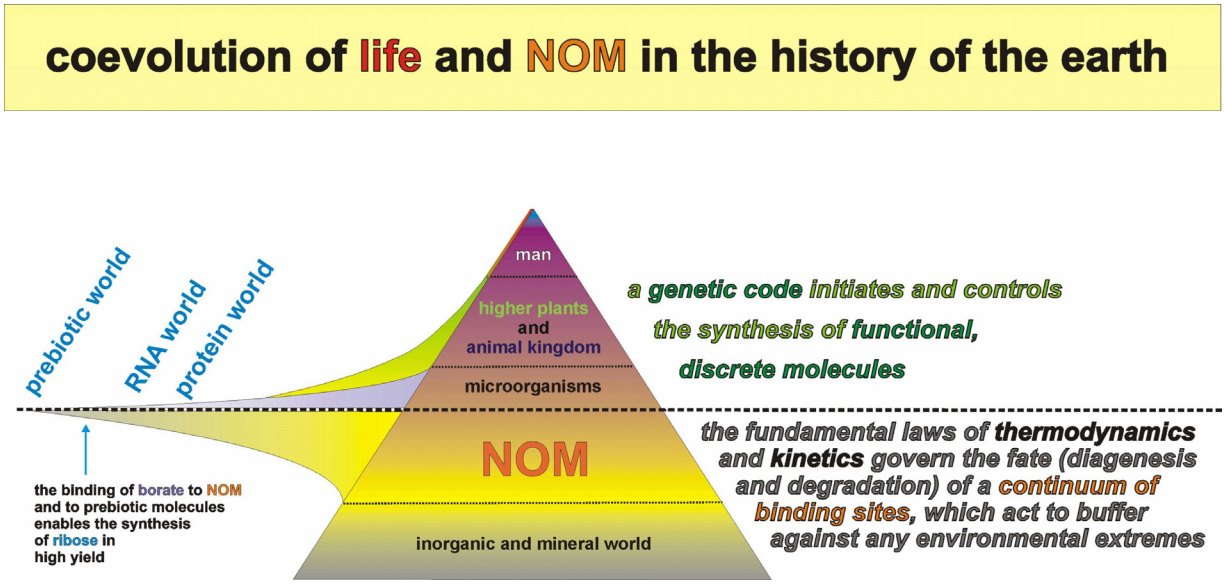



**Figure 2.** The formation of NOM on earth preceded the evolution of life; the binding of NOM and prebiotic molecules to borate contributed to the synthesis of ribose, a crucial precursor of nucleotides, in good yield, (Ricardo et al., 2004). Later in the earth's history, coevolution occurred between prebiotic/abiotic molecules, NOM and primitive and higher forms of life. The ecological function of NOM in the geo- and biosphere is composed of acting as buffer against any environmental extremes and to furnish metal ions and organic molecules, which are essential for living organisms.

NOM is generated through biotic and abiotic pathways and consequently, the composition of NOM reflects that of the living world. The extremely complex structure of NOM stipulates an immense diversity of binding sites in NOM, and the resulting near continuum of reactivities towards inorganic ions and organic molecules acts to buffer against any, potentially damaging, environmental extremes in the geo- and biosphere (MacCarthy and Rice 1988; Perminova et al., 2005). In addition to this primary function of NOM in the ecosphere, another key supportive role of NOM to life on earth resides in the provision of inorganic and organic nutrients to plants and microorganisms (Sanudo-Wilhelmy et al., 2002; Hertkorn, 2005, Qualls, 2005).

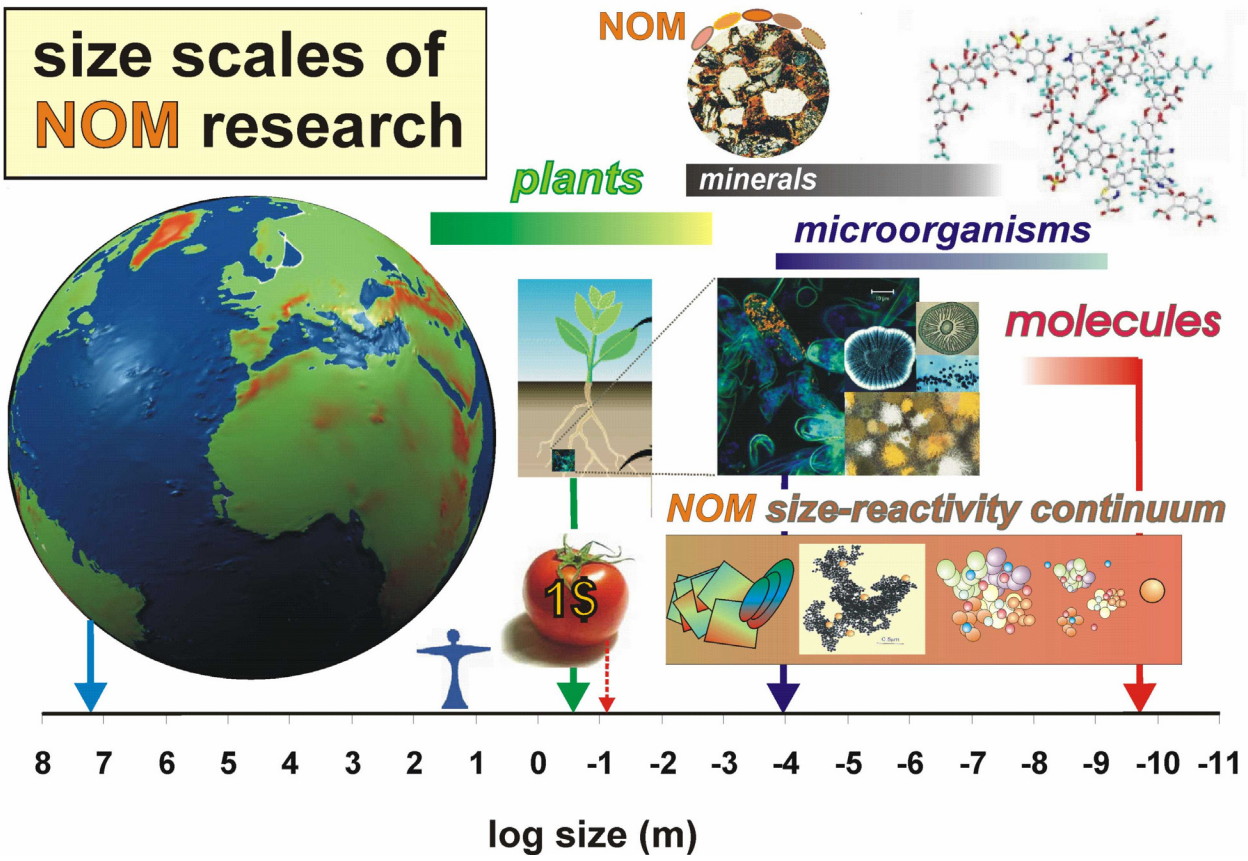



**Figure 3.** NOM research issues encompass time (Figures 2 and 4) and size-scales across several orders of magnitude. Even if global change is concerned with the entire earth, it is nevertheless the consequence of a multitude of small-scale reactions. Similarly, agricultural productivity and plant growth ultimately relate to chemical reactions on the molecular-level. Metabolism of microbial communities is responsible for the bulk of biotic NOM synthesis and turnover; in surface environments, photoreactions of NOM open very prominent abiotic degradation pathways of NOM (Figure 4).

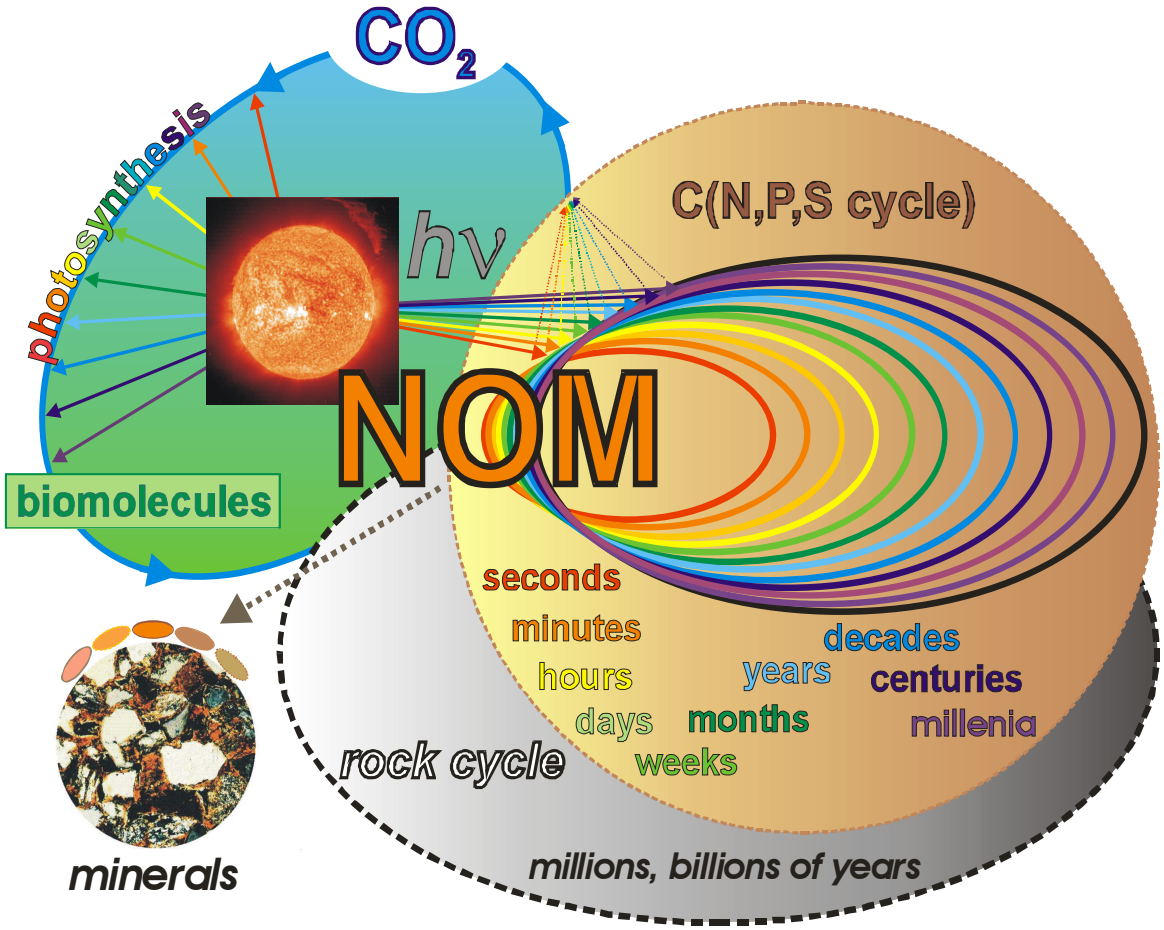



**Figure 4.** Natural organic matter (NOM) continuously interacts with a broad range of terrestrial, limnic and marine ecosystems. Common to all these environments are the fundamental molecular aspects of life, and an availability of extended mineral surfaces for interaction and binding of NOM. The dynamic equilibrium of NOM generation and decomposition spans timescales across many orders of magnitude (microseconds to hundred-thousands of years) and it results from a combination of biotic and abiotic reactions. NOM may be intrinsically recalcitrant because of the chemical structure of its organic molecules; alternatively, strong NOM-mineral interaction may alter the reactivity of these organic molecules towards increased resistance against or acceleration of degradation. Physical protection of organic matter at inner mineral surfaces provides another option to induce recalcitrance of NOM. Photochemical degradation, one of the most significant abiotic reactions of NOM, often results in small molecules like $CO_2$ which are mobile and are easily distributed within various ecosystems. Biomolecules from photosynthesis or otherwise originating from a genetic code are decomposed according to the general laws and constraints of thermodynamics and kinetics. At very long timescales, NOM interaction with minerals at elevated temperatures results in the formation of geopolymers, like kerogen, coal, and oil. These ancient materials participate in bio- and geochemical cycling by natural and anthropogenic combustion and by wheathering (Dickens et al., 2004).

At present, the most convincing direct experimental evidence of the immense complexity of NOM is derived from ultrahigh-resolution mass spectroscopy, namely FTICR (Fourier-Transform Ion Cyclotron Resonance) mass spectrometry, (Kujawinski et al., 2004, Kujawinski et al., 2002; Kujawinski 2002; Stenson et al., 2003). This is an intrinsic consequence of

NOM properties and the quality of FTICR mass spectra, which display by far the largest peak capacity of any method of organic structural spectroscopy (cf. case study IV, Figures 5, 84) and therefore the most highly resolved (and consequently, most information-rich) spectra.

**peak capacity of *organic structural spectroscopy* in the analysis of NOM**

| *method* | bandwidth | linewidth | peak capacity |
|---|---|---|---|
| *UV/VIS spectroscopy* | 600 nm | 50 nm | 12 |
| *FTIR spectroscopy* | 4000 $cm^{-1}$ | 10 $cm^{-1}$ | 400 |
| *$^{1}$H NMR spectroscopy* | 6000 Hz | 2 Hz | 3000 |
| *$^{13}$C NMR spectroscopy* | 27500 Hz | 1 Hz | 27500 |
| *$^{1}$H,$^{13}$C HSQC NMR spectroscopy* | 9 $10^{7}$ $s^{-2}$ | 25 $s^{-2}$ | 3.6 $10^{6}$ |
| *FTICR mass spectrometry* | 4000 Da | 0.5 mDa | 8 $10^{6}$ |
| *CE ESI/FTICR MS* | 1800 s/4000 Da | 0.2 s/ 3 mDa | 1.2 $10^{10}$ |

**Figure 5.** Peak capacity (total bandwidth / individual linewidth) for various methods of organic structural spectroscopy used in the analysis of NOM

FTICR mass spectrometry allows to establish molecular formulas $C_lH_nO_mN_o$ (P, S,…) for thousands of individual molecules present in NOM, which can be, most instructively, displayed by van-Krevelen diagrams, (Kim et al., 2003; Visser, 1983; Wu et al., 2004). The appearance of these plots varies with the source of NOM; however, in all high quality mass spectra of NOM obtained so far, many hundreds to several ten-thousands of different molecular formulas appear as resolved dots in (H/C versus O/C atomic ratio) van-Krevelen diagrams of NOM (Kramer et al., 2004; Kim et al., 2004).

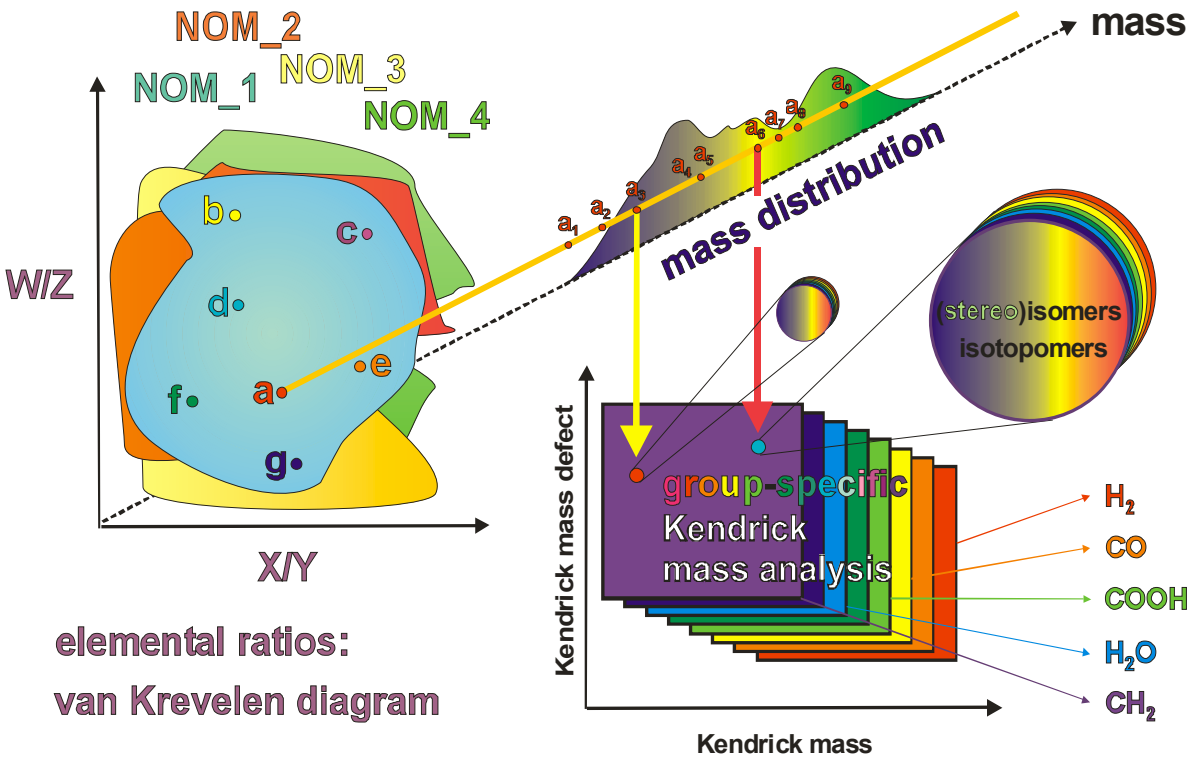



**Figure 6.** Ultrahigh resolution (FTICR) mass spectra provide the most direct experimental evidence of the enormous complexity of NOM molecular composition. Any of the several thousands of dots observed in van-Krevelen diagrams of high-quality FTICR mass spectra of NOM (examples given here: a – g) indicates two elemental ratios (usually H/C versus O/C) in a given molecular formula and exhibits a mass distribution (example given here: $a_1$ – $a_9$), which can serve as a third axis in a three-dimensional van-Krevelen plot. Any of these molecular compositions occupies a specific spot in an n-dimensional Kendrick mass-defect analysis, which allows to define classes of compounds. Here, any of these individual dots stands for an immense diversity of isomeric molecules and even isotopomers (molecules of ultimately biogenic origin, which are synthesized by different enzymes or organisms and therefore show a characteristic stable- and radio-isotope distribution, which allows to determine their origin and diagenesis). The number of feasible isomers increases dramatically with molecular size!
The spacing between individual molecular compositions in a van-Krevelen plot is at first defined by the number of chemical bonds between atoms: in organic molecules, carbon offers four valences, nitrogen three, oxygen two and hydrogen one, and these are arranged in single, double, and triple bonds. The other key factor is the smallest increment of difference in moleular composition, which is commensurate with the basic rules of chemical bonds, described above (cf. Figure 87, 88). Hence, with increasing mass of the molecule (i.e. with an increase of the number of constituent atoms and chemical bonds) the spacing between available molecular compositions in a van-Krevelen diagram declines ever more.

Most importantly, a very large area of elemental ratios (the typical display shows the H/C against the O/C atomic ratios) is near uniformly covered by the molecular compositions found in NOM. This extensive and near continuous coverage of a large area implies an extreme variety of covalent chemical bonds between the different fundamental precursors and constituents of NOM (like lipids, proteins, carbohydrates and lignins), which occupy each clearly distinct and rather limited regions in a van-Krevelen diagram (Visser, 1983; Kramer et

al., 2004); i.e. the signature of these precursors has been nearly lost during the formation of NOM. This is a very important result with far-reaching implications, both for our understanding of NOM diagenesis and for the design and interpretation of analytical data obtained by any method.

Any of the thousands of individual dots in a van-Krevelen diagram represents two elemental ratios, and for every single of those, a range of different molecular mass values is found in NOM, which can be displayed as a third dimension in a three-dimensional van-Krevelen diagram (cf. Figure 6). The Kendrick mass defect analysis (Hughey et al., 2001; Hsu et al., 1992) is another ingenious and complementary approach to visualize and analyze complex mass spectra in n-dimensions. Here, classes of compounds are defined, which nevertheless represent only a superposition of all feasible isomers; the number of these isomers easily accounts for many millions even in moderately sized molecules of a few hundred Daltons when some double bond equivalents and heteroatoms are present (which is in line with the molecular formulae observed in NOM).

Considering that manifest complexity of high-quality mass spectra of NOM, where any single of the several thousands of peaks represent a range of, possibly wildly different, molecular structures, it becomes immediately obvious, that owing to extremely severe overlap of signals, any spectrum or any chromatogram/electropherogram of NOM will appear near featureless and with inferior resolution.

Extensive superposition of signals will be most strongly marked in NMR spectra, in which any single (NMR-active) atom will produce an individual resonance with a specific chemical shift. In contrast, in a mass spectrum under conditions of non-fragmentation, any combination of isomers will produce only one single peak. For that reason of sheer numbers, the most extreme overlap of signals is expected to occur in NMR spectra, making those at first glance a seemingly less desirable choice for high-resolution and -precision molecular level structural analysis. On the contrary, the specificity of the NMR resonances of the “small” nuclei $^{1}H$, $^{13}C$, $^{15}N$ and $^{31}P$ provides unsurpassed information referring to the short range molecular order around these nuclei, which are those of prime importance in defining the backbone of NOM. The chemical shift of a NMR resonance is indicative of the chemical environment of the observed nucleus up to four bonds away in any direction (Cavanagh et al., 1996; Croasmun and Carlson, 1996; Becker, 2000; Zerbe, 2003. Therefore, the quantitative one-dimensional NMR spectrum provides the most instructive and detailed molecular level structural

information concerning the quantity and structural detail of the fundamental building blocks of NOM.

This information is supplemented and confirmed by a large array of specialized NMR spectra and pulse sequences, like spectral editing (cf. chapters 2.4.1. and 8.1.3.1.3.) and higher-dimensional NMR spectroscopy, which all serve to enhance the reliability of NMR resonance assignments, (Cavanagh et al., 1996; Croasmun et al., 1996; Becker,2000; Zerbe, 2003; Reid, 1997).

Owing to the extreme complexity of NOM, a reliable molecular-level precision structural analysis necessitates a combination of complementary analytical methods (Diallo, 2003; Ussiri and Johnson, 2003; Francioso et al., 2003; Lead et al., 1999; Christl et al., 2000; Abbt-Braun et al., 2004; Zbytniewski and Buszewski 2005). Bulk data of NOM, like physical parameters, total acidity, elemental analysis, at first seem more precisely described, but certainly lack molecular level structural resolution. However, any valid structural model of NOM has to fit into the constraints defined by these “hard” bulk data. An illustrative example of serious obstacles encountered during an ostensibly facile analytical problem concerns the determination of the oxygen content of NOM, a very crucial parameter, for instance for the determination of the functional group composition. Oxygen in NOM is in the form of various organic functional groups (ether, esters, amides, keto, carboxylic acids, alcohols and phenols, sulphate, amides, among others) and minerals (chiefly metal oxides and carbonates). Furthermore, dry isolates of NOM are very hygroscopic and take up to 12 weight percent of water. The mineral salts in NOM are complex coordination compounds with water, terminal and bridging hydroxyl- and oxo-ligands and other oxygen-containing anions, like borate, nitrate, carbonate and sulphate, some of which feature considerable thermal lability. In this respect, oxygen-containing compounds in NOM form a quasi reactivity continuum against thermal degradation, which makes the unambiguous discrimination of organic and mineral oxygen in NOM by standard elemental analysis techniques a very difficult analytical task (Gèlinas et al., 2001; Giovanela et al., 2004).

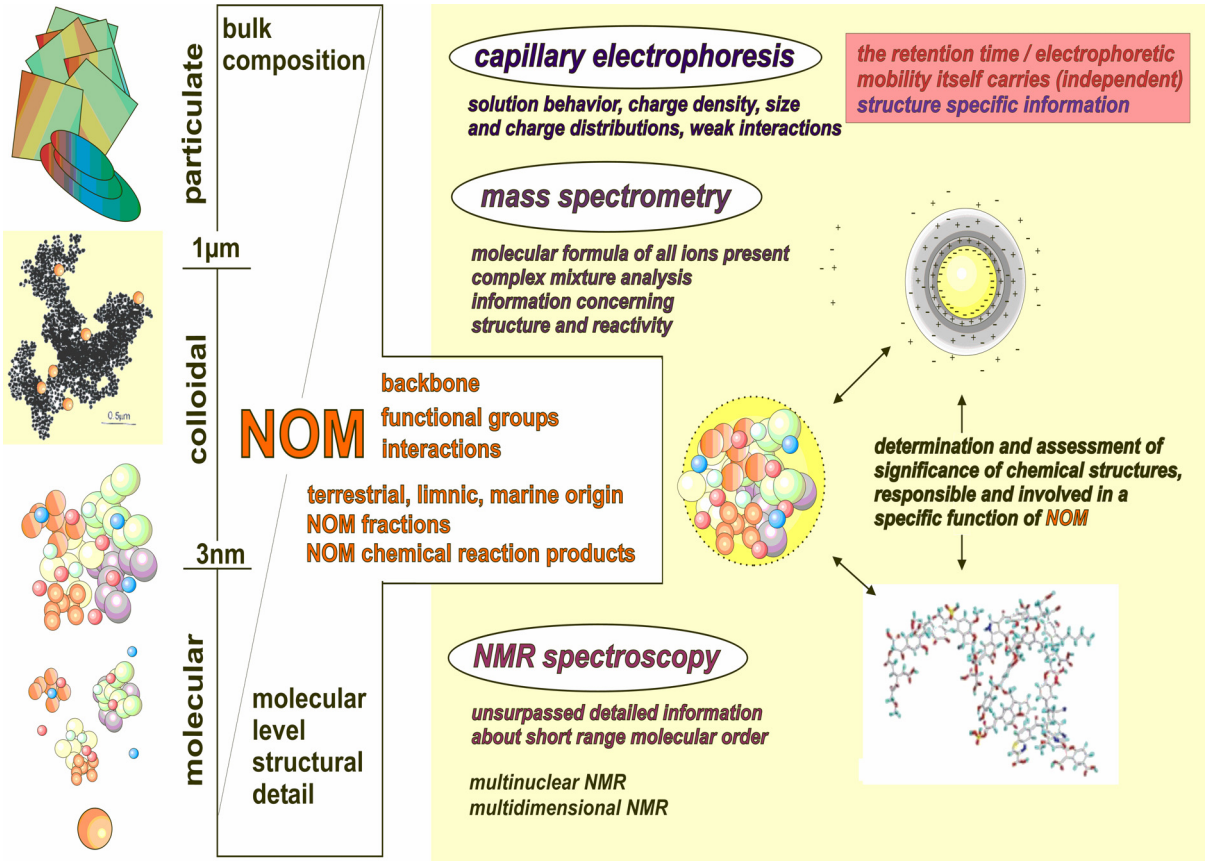



**Figure 7.** NOM displays a size and reactivity continuum and is in most of their natural environments strongly bound to minerals. For molecular-level precision determination of chemical structures, which are involved in specific functions of NOM, a combination of high-performance capillary separation techniques (like HPLC, SEC, capillary electrophoresis CE), ultrahigh resolution mass spectrometry and NMR spectroscopy is most promising.

The quantitative determination of metal contents of NOM faces analogous difficulties (Heumann et a., 2002), and the validity of analytical data ever more depends on careful experimental protocols and data treatment schemes when reaching out to molecular level structural detail with intrinsically complex techniques.

In our view, three highly complementary methods in combination provide the most detailed and precise molecular level structural detail of NOM currently available: capillary electrophoresis CE/high performance liquid chromatography HPLC (high-performance capillary separation techniques), mass-spectrometry and NMR spectroscopy (Figure 7). High performance separations can be highly automated and experimental conditions can be specifically adjusted to probe pH, concentration and temperature dependence; passive (e.g. non-reactive molecules with specified size and shape) and reactive compounds (e.g. metal ions, organic molecules)

can be added to probe the reactivity of NOM (Schmitt-Kopplin and Kettrup, 2003; Schmitt-Kopplin and Junkers, 2003; Schmitt-Kopplin and Frommberger, 2003; Ping et al., 2003; Junkers et al., 2002). Here, NOM molecules are viewed as entities from the outside and solution behaviour, charge density, size and charge distributions and weak interactions can be determined under a wide range of conditions. The retention time (LC) and the electrophoretic mobility (CE) itself carry structure-specific information independent of the detection mode (which is typically UV/fluorescence or mass spectrometry); this is an immensely important supplement to MS and NMR derived structural information. NMR spectroscopy on the other hand views bond electrons from the perspective of atomic nuclei and even a few chemical bonds represent a long distance in the NMR world. But within that range, NMR spectroscopy provides isotope-specific, reasonably quantitative, unsurpassed detailed and precise information about short range molecular order (arrangement of chemical bonds including stereochemistry) and dynamics (enabling e.g. reactivity studies). Mass spectrometry is especially suitable for the analysis of complex mixtures and allows the determination of molecular formulae of all ions present (at least up to several hundred Daltons for molecules with non-repetitive structures and up to several ten-thousands for (bio)molecules with repetitive structures, like proteins and nucleic acids (Xiong et al., 2004). Fragmentation provides deep insights into structural detail of NOM and allows to some extent the discrimination of isomeric molecules. FTICR mass spectrometry also allows (gas-phase) reactivity studies of NOM with a wide variety of molecules (Yu et al., 2004; Li et al., 2005). Owing to an unsurpassed peak-capacity (ratio of total mass range to individual linewidth) in comparison with all other methods of organic structural spectroscopy (cf. figure 5), the relevance of ultrahigh resolution mass spectrometry in the molecular-level structural analysis of NOM is expected to increase markedly in the future.

Natural organic matter (NOM) is an operationally defined, enormously complex mixture of organic, and some inorganic, constituents occurring in terrestrial, limnic and marine ecosystems (Derenne and Largeau, 2001; Gaffney et al., 1996; De Leeuw and Largeau, 1993; Hedges and Oades, 1997; Stevenson, 1994; Hansell and Carlson, 2002; Perdue and Ritchie, 2003; Hedges, 2002; Steinberg, 2003). On a global basis, NOM on average greatly outweighs biochemicals in the living organisms from which they derive. The origins, reactions and fates of these ubiquitous and inconspicuous materials are relatively obscure, in large part because the rich vein of geochemical information that typically derives from detailed structural and

stereochemical analysis is yet to be tapped (Hedges, 2002). In recent years, the perception of NOM has evolved from an emphasis of a largely separate pool of remarkably old and static substances to the current view of a dynamic assemblage of organic molecules that interact with each other, trace metals and living organisms over a broad continuum of space and time scales (Hedges, 2002).

NOM plays immensely important roles in the natural world, and it is a key refractory constituent of the global carbon (cf. Figures 4, 8) and other element cycles (Hedges and Oades, 1997; Stevenson 1994; Perakis and Hedin, 2002). To a significant extent, NOM determines the binding and bioavailability of toxic and nutrient metal ions and organic xenobiotics and it is involved in such key processes as modulating temperatures at the globe's surface, weathering rocks to soils, stabilization of atmospheric levels of oxygen over geological time scales (Hedges, 2002), and composing precursors for eventual formation of coal and petroleum. Organic compounds embedded in marine sediments and paleosols also provide exquisitely detailed records of natural history, even where macroscopic physical fossils are rare or absent (as in petroleum) (Hedges, 2002).

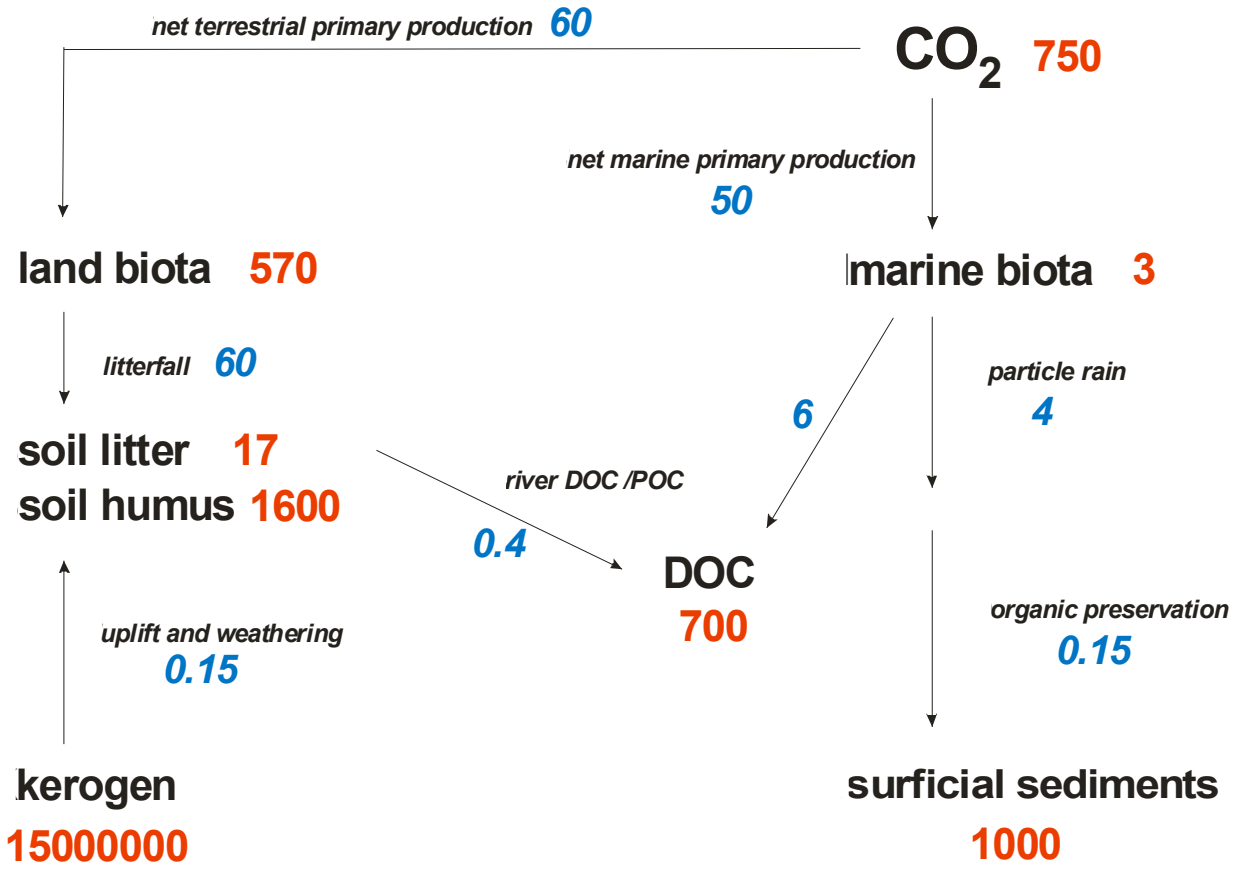



**Figure 8.** Selected global reservoirs (red) and fluxes (*blue italics*) of various NOM reservoirs in the geosphere [Gt or Gt/a] (Hedges and Oades, 1997).

## 1.1. CHARACTERISTICS AND CONSTRAINTS OF THE SPECTROSCOPIC CHARACTERIZATION OF NOM AND HS

From the perspective of analytical chemistry, both NOM and HS represent the structurally most diverse class of natural organic polymers, and they are characterized by an extensive polydispersity and a pronounced irregularity of their molecular level structure (MacCarthy and Rice, 1988; Hedges et al., 2000; Perminova et al., 2003). The stochastic synthesis of NOM/HS out of biotic and abiotic precursors is governed by fundamental restraints of thermodynamics and kinetics and follows no genetic code.

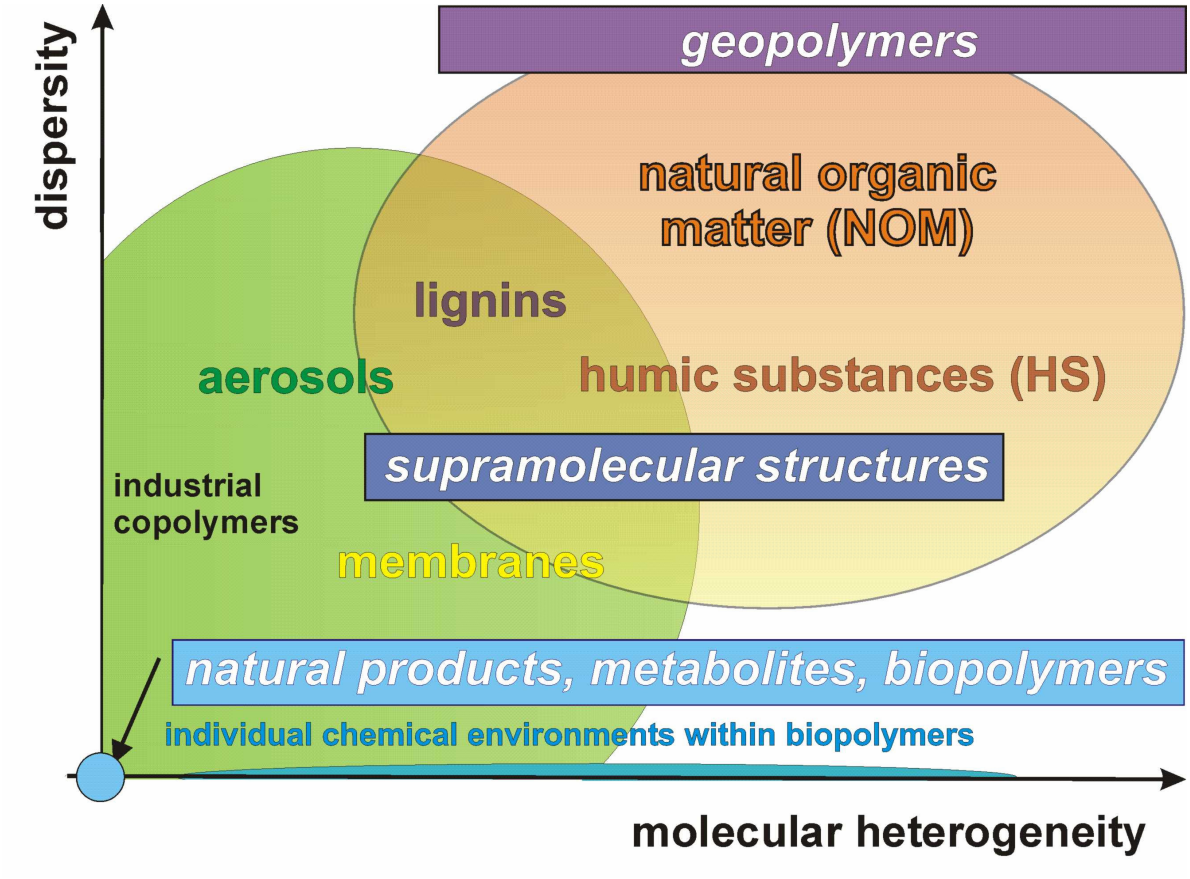



**Figure 9.** Hierarchical order of intricacy in the structural analysis of materials in terms of polydispersity and molecular heterogeneity (cf. text).

The level of intricacy in the analysis of large molecules and of mixtures can be classified according to their polydispersity and heterogeneity (cf. Figure 9). The structures of complicated, but monodisperse, natural products and biopolymers are readily accessible (provided that sufficient amounts of materials are available) by a combination of analytical methods, which primarily rely on NMR spectroscopy and mass spectrometry (Kujawinski, 2002; Barrientos et al., 2003; Glore and Gronenborn, 1998; Glore and Gronenborn, 1998). Supramolecular structures (Takahashi et al., 2000), composed of (modified) biopolymers aligned in aggregates, which are supported and defined by weak interactions, require a more elaborate characterization, which requires an adequate definition of covalently bonded molecules and of their non covalent interactions (Barrientos et al., 2003). Consequently, the characterization and structural analysis (Hertkorn et al., 2004; Burdon, 2001) of geopolymers,

which feature a substantial extent of both polydispersity and molecular heterogeneity (MacCarthy and Rice, 1988), is most demanding with respect to methodology and concepts.

A highly resolved three–dimensional structure of a monodisperse biopolymer is based on a precise description of the chemical environment of any single atom (Gratias et al., 1998). As a consequence, e.g. in a NMR data set of a protein, any atom will show unique values of (multinuclear) chemical shifts δ and (homo- and heteronuclear) coupling constants (*D*, *J*), which are defined by the respective short and long range chemical environment. Currently, the molecular level structural analysis of NOM/HS is primarily focused on the definition of the covalent bonds. In an ongoing evolution, future high quality structural analyses of NOM/HS will have to provide a characterization of individual molecules and a description of the extent and mechanisms of their interactions.

Any adequate understanding of geophysical and biological processes requires a sufficient characterization of the participating and constituent materials. Since diffraction methods do not provide usable structural data of mixtures, structural analysis of NOM/HS has to rely on data from complementary spectroscopic methods (Christl et al., 2000; Kujawinski, 2002; Diallo, 2003; Kögel-Knabner, 2002; Thomsen et al., 2002), which – in combination – can lead to the deduction of precisely defined chemical (sub)structures.

Many methods of structural spectroscopy rely upon transitions between energy levels in atoms and molecules, and the respective spectra are defined by a substance specific set of energy and intensity values (Figure 10). The interaction of atoms and molecules with light across the electromagnetic spectrum induces an array of physical processes, which can be utilized to reveal structural information. High energy spectroscopic techniques are capable to sensitively determine the content of elements in NOM/HS (AAS, AES) and, with some uncertainty, also their oxidation state [XANES]. When proceeding to ever diminutive splittings of energy levels (XANES, Lowe, 1992; Morra et al., 1997; Xia et al., 1999; Hertkorn et al., 2006; UV, Stevenson, 1994; Bloom and Leenheer, 1989; IR, Bloom and Leenheer, 1989; Hayes et al.,1989; NMR, Preston, 1996; Mahieu et al., 1999; Cook, 2004; Simpson, 2002), a continually enhanced resolution of organic structures will become accessible at the expense of sensitivity.

NMR spectroscopy is a low energy method and non destructive; the NOM/HS samples can be used for several NMR analyses and subsequently investigated with other analytical methods.

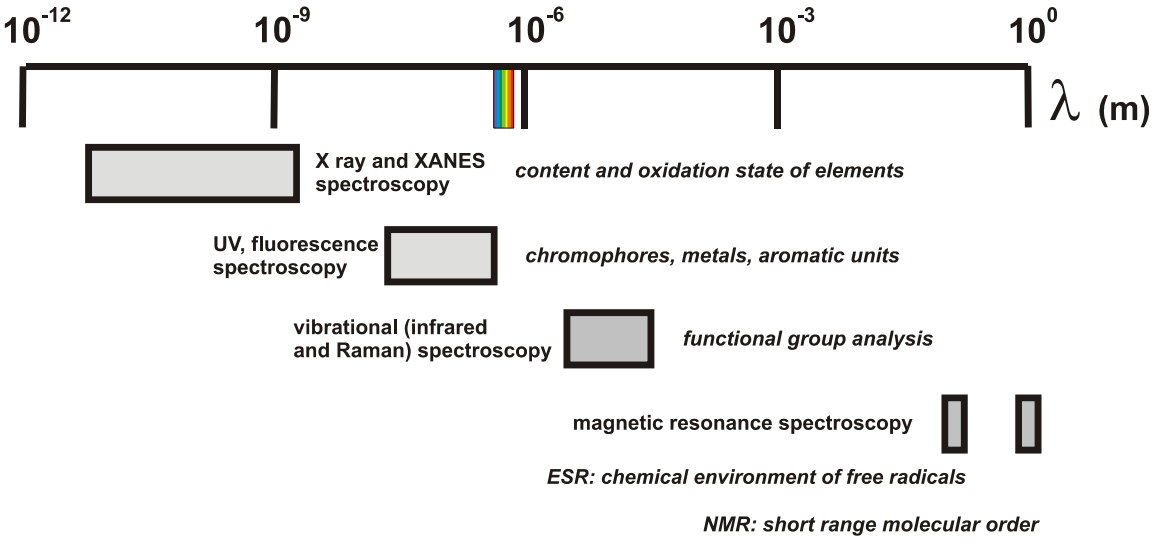



**Figure 10.** Methods and significance (italics) of structural spectroscopy of NOM/HS across the electromagnetic spectrum.

Analytical methods have recently become capable to separate nearly identical large molecules (Hong et al., 1998); however the analytical chemistry of inseparable mixtures has been neglected and remains significantly underdeveloped in terms of practical methods and concepts even today. Attempts to develop a precise molecular level structural analysis of NOM/HS will therefore address several independent and important topics:

1) An accurate description of the molecular level structure of NOM/HS constitutes the fundamental requirement to understand the mechanisms and significance of its interaction with minerals, organic and inorganic compounds and living organisms in the eco- and geosphere.
2) An exhaustive analysis of data of NOM/HS characterization derived from complementary analytical techniques, aided by mathematical analysis, will contribute significantly to an evaluation of the scope and accuracy of each individual method.
3) A hierarchical order of carefully tested analytical methods and data treatment schemes will contribute to develop and evaluate concepts of the analytical chemistry of inseparable mixtures.

NOM/HS represents an enormously complex mixture of organic and inorganic materials and any spectra of these substances will feature a very significant overlap from individual species present and, consequently, a highly degraded resolution. A significant degree of the uncer-

tainty in the spectral assignment of NOM/HS originates from the fact that both positions of spectral lines (energy levels) and intensities (transition integral values) of a highly complex array of structures and interactions cannot be calculated with an accuracy anywhere near realistic values directly from quantum mechanical *ab-initio* or semi empirical methods (Helgaker et al., 1999), (even more so, as precise structural models of NOM/HS are not yet available).

Quantification remains one of the key unsatisfactorily resolved research issues in the analysis of NOM/HS. In any method, complicated non linear and unknown transfer functions are caused by a near continuum of differential responses arising from small and large molecules and strong and/or weak interactions within the NOM/HS system (cf. Figure 11). Even otherwise well established methods, such as elemental analysis, or the determination of the content of heavy metals by atomic absorption spectroscopy, often produce a very wide range of data values (Giovanela et. al., 2004), when applied in a routine fashion to the analysis of NOM/HS.

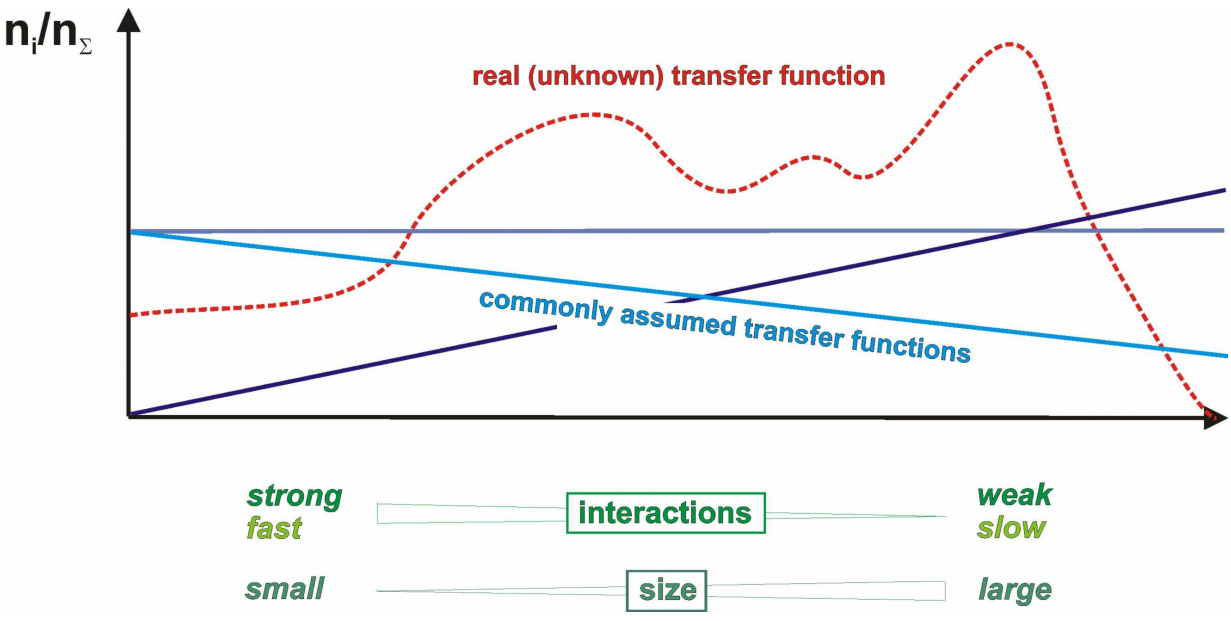



**Figure 11.** Commonly assumed (linear) and unknown real (non linear) transfer functions in the analytical characterization of NOM/HS.

Therefore, every detail of any spectral analysis of NOM/HS has to be very carefully evaluated and calibrated, both intrinsically and by comparison with realistic reference materials, which are not routinely available on a commercial basis. The International Humic Substances Society IHSS (www.ihss.gatech.edu) provides an ever increasing selection of humic substances reference and standard materials, which are valuable for referencing and testing in method development.

## 2. THE SPECIAL ROLE OF NMR SPECTROSCOPY IN THE MOLECULAR LEVEL CHARACTERIZATION OF NOM/HS

The two most significant and most highly resolving methods of organic structural spectroscopy are magnetic resonance spectroscopy and mass spectrometry. Mass spectrometry involves measurement of mass numbers rather than energy absorption and is not (directly) associated with a region of the electromagnetic spectrum (Kujawinski, 2002). Magnetic resonance spectroscopy is concerned with the splitting of magnetic spins of electrons (electron spin resonance, ESR) and that of atomic nuclei (nuclear magnetic resonance, NMR) in an external magnetic field $B_0$. The splitting of the NMR transition is not solely an intrinsic molecular property, but also depends on the magnitude of an external magnetic field $B_0$. The special role of NMR spectroscopy in the molecular level characterization of NOM/HS resides in its ability to provide unsurpassed in-depth, isotope specific information about short range molecular order.

NMR spectroscopy is arguably the single most important method of molecular level precision structural analysis of NOM, and combines far reaching, isotope-specific definition of chemical environments with the most reliable quantification of any organic structural spectroscopy method (Pauli et al., 2005). The last-mentioned subject seems remarkable, considering the broad range of complex NMR relaxation mechanisms operating in NOM, which establish highly non-linear dependencies between numbers of chemical environments and NMR resonance integrals (cf. below) and the diversity of “indirect” effects, which affect NMR chemical shifts and lineshapes and obscure facile relationships between molecule structures and chemical environments (table 1). Nevertheless, in NMR under ideal conditions, the integrated intensity (signal area) of a NMR resonance is proportional to the number of nuclei giving rise to

that NMR signal (Heikkinen et al., 2003; Keeler and Maciel, 2003; Tenailleau et al., 2004; Henderson, 2004); in any other method of organic structural spectroscopy (like UV/VIS, infrared and Raman spectroscopy and mass spectrometry), the transfer functions between the molecular level transition integral values and the response of the analytical instrument (Figure 11) are much less understood than in NMR. A convincing illustration of that issue is the enormous dependency of the ionization efficiency of molecules in mixtures on the experimental conditions in LC/MS hyphenation. So far, the most authentic results have been obtained with highly miniaturized nano-spray systems under very well controlled flow conditions, but clearly, much research is still required to address the topic of quantitative mass spectrometry of mixtures (Schmidt et al., 2003; Zamfir et al., 2004).

Owing to the nature of NOM/HS as a complex mixture of organic and inorganic compounds, any NMR parameter will be characterized by a weighed average and by a distribution of values. *NMR shieldings* and *resonance integrals* of individual chemical environments in NOM/HS will be superimposed to produce a broad envelope of overlapping resonances in the NMR spectra, resulting in a rather low signal to noise ratio with respect to weight unit (when compared, e.g. to natural products). The overall resolution will become insufficient to clearly resolve *J-couplings* under routine conditions in one dimensional NMR spectra. *NMR relaxation parameters* and, in result, the *linewidths* also will be described by distribution functions; the current literature is deficient in addressing these issues (Preston 1996; Cook, 2004; Wang et. al., 2003; Flynn et al., 2000; Fründ and Lüdeman, 1989; Preston and Blackwell, 1985).

The minute spacing of energy levels in NMR transitions (cf. Figures 10, 12) causes profound effects on the relaxation in NMR spectra as spontaneous emission is virtually absent. The spin-lattice relaxation, which is responsible to establish thermal equilibrium between the two spin states after the application of the static magnetic field $B_0$ and of high frequency pulses, is associated with transfer of Zeeman energy (enthalpic effect); the longitudinal relaxation in NOM/HS is described by a function of the longitudinal relaxation times $T_1$ of the individual nuclei. Efficient longitudinal relaxation requires coupling to molecular processes (mostly random reorientation of atoms) occurring at frequencies close to the NMR frequencies $\omega_0$ or $2\omega_0$.

One major obstacle in obtaining quantitative NMR spectra of NOM/HS is related to differential and slow longitudinal relaxation (Tang et al., 2004), which is a characteristic property of nuclei of restricted (region of spin diffusion) and of very high flexibility (extreme narrowing limit). Diamonds are rather extreme examples of a rigid lattice and their purity (i. e. absence

of lattice defects) can be determined from $T_1$ values in their $^{13}C$ NMR spectra (Terblanche et al., 2001; Reynhardt, 2003). In a crude approximation, the NMR longitudinal relaxation of heteronuclei within NOM/HS is inherently related to their NMR frequencies and to their distance to adjacent protons and free electrons. It is a peculiar feature of NOM/HS that both heteronuclei behaving according to the extreme narrowing limit and to spin diffusion (or in between) may be present in the same material.

Transverse (or spin-spin) relaxation involves mutual exchange of spin energy without altering the total Zeeman energy of the nuclear spins. The net effect of this energy transfer is to cause a loss of phase coherence (entropic process), equivalent to a decay of bulk magnetization in the plane of the detection (xy-plane). The linewidth $\Delta\nu$ in NMR spectra is governed by the rate of transverse relaxation (under the assumption of single exponential decay): $\Delta\nu = 1/2\pi T_2$, with $T_2$ representing a function of the transverse relaxation times of individual nuclei.

Very fast transverse relaxation, as induced by paramagnetic centres (metal ions, Gèlinas et al., 2001; Gratias et al., 1998; Fan, et al., 2000; Smernik and Oades, 1999): especially $Fe^{2+}$ and $Fe^{3+}$, and organic radicals) and by the limited flexibility of large size molecules and aggregates, may broaden NMR resonances of NOM/HS beyond recognition.

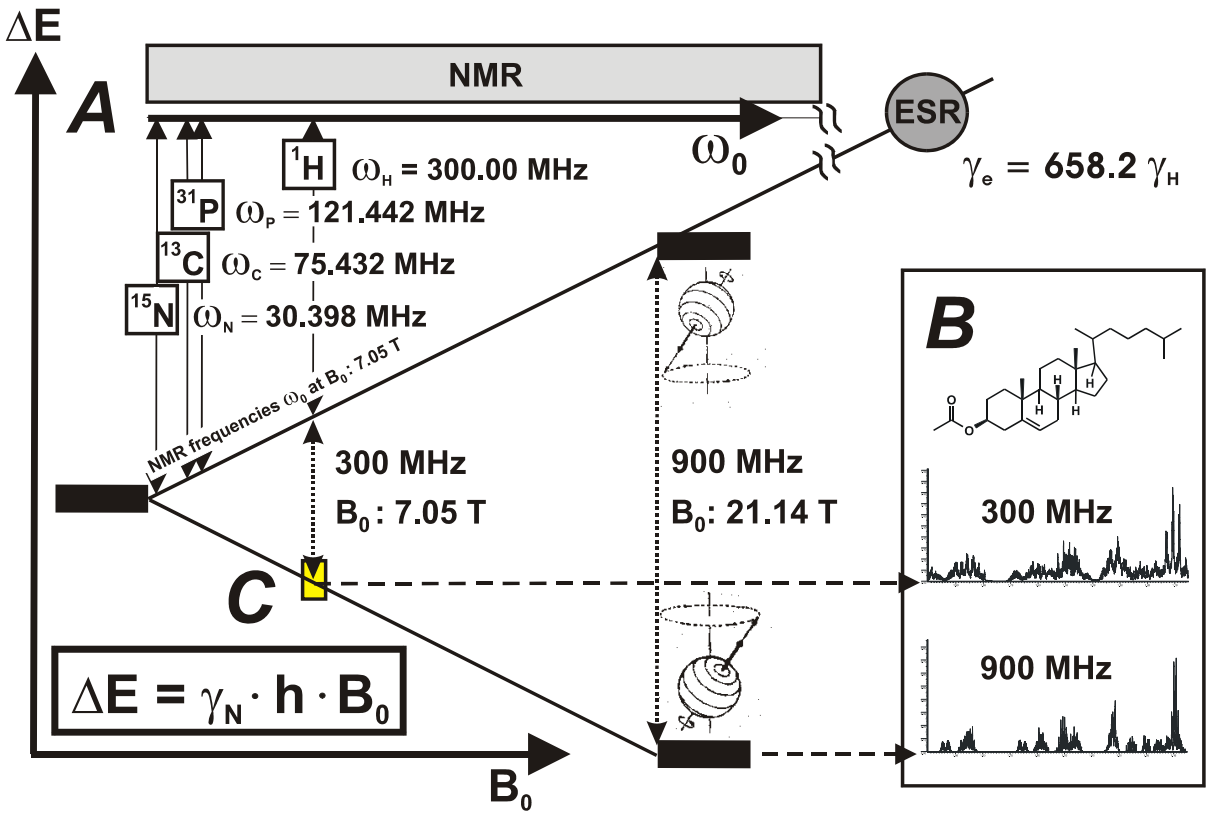



**Figure 12.** Dependence of some magnetic resonance spectral characteristics on the magnetic field $B_0$; the electron-proton magnetic moment ratio is 658.2 (**A**) At constant magnetic field $B_0$, the resonance frequency depends on the relative magnetogyric ratios (or magnetic moments) of the nuclei $\gamma_N$ (NMR) and the electron $\gamma_e$ (ESR); (**B**) sections of proton NMR spectra of cholesterol acetate at $B_0$ = 7.05 T (300 MHz $^1$H frequency) and at $B_0$ = 21.14 T (900 MHz $^1$H frequency); note the *qualitative* difference of high resolution proton NMR spectra acquired at various $B_0$ values. This variation remains the most significant obstacle for automated NMR resonance assignment in natural products. As coupling constants have less significance in one dimensional NMR spectra of NOM/HS, the variance in the NMR spectra of NOM/HS is much less pronounced, and common assignment procedures, based on integration (cf. case study 2) are valid. Pattern recognition in 2D NMR spectra can be automated; NMR spectra of NOM/HS acquired at increased $B_0$ show enhanced resolution and improved assignment options. (**C**) The relative energy, represented by the chemical shift range, covers a miniscule ["ppm"] range of the (already tiny) NMR energy transition energy; the ratio of total chemical shift range to total NMR transition energy ranges from 20 ppm ($^1$H, diamagnetic molecules) up to 20000 ppm ($^{59}$Co, $^{195}$Pt NMR). Owing to the near equality of the Boltzmann factors for the NMR energy levels, out of 2000000 proton nuclei, only 81 participate in the $^1$H NMR experiment at $B_0$ = 11.7 T and 283 in the NMR experiment at $B_0$ = 21.14 T (at room temperature: 300 K). All other proton nuclei remain silent throughout the NMR experiment. This ratio is even worse for other nuclei, explaining the relative insensitivity of NMR spectroscopy when compared to higher energy spectroscopic methods.

NOM/HS as a complex assemblage of small and large molecules, related by a broad range of various interactions, enacts strong effects on the local mobility of its individual atoms in their respective chemical environments; therefore, the above mentioned fundamental relaxation processes are composed of a broad range of mechanisms (Case, 2002; Becker, 2000; Murali

and Krishnan, 2003; Vugmeyster et al., 2004). In the available literature, the NMR relaxation times determined within high resolution NMR spectra of NOM/HS represent weighed average values without providing their distribution function. However, several studies, concerned with a careful analysis of optimum conditions for quantitative solid state $^{13}C$ NMR spectra of NOM/HS (Peuravuori et al., 2003; Keeler and Maciel, 2003; Mao and Schmidt-Rohr, 2004; Smernik and Oades, 2003; Mao et al., 2002; Preston, 2001; Smernik and Oades, 2001; Mao et al., 2000; Cook and Langford, 1998; Smernik, 2005), have indicated rather complex time dependencies of various relaxation mechanisms. It is expected that the relaxation mechanisms operative in any NMR experiment will exhibit non-linear and non-exponential behaviour and will impose selective effects in different sections of chemical shift within the NMR spectra of NOM/HS.

Any meaningful NMR spectrum of NOM/HS has to show a decent S/N ratio to allow to some degree valid conclusions about the chemical environments; in practice, a compromise between the contradicting requirements of large S/N ratio (requires many scans and a long total acquisition time) and non-selectivity (requires rather long inter pulse intervals of at least 5 times $T_1$ of the most slowly relaxing nucleus) has to be reached.

The latter condition is exceedingly difficult to satisfy because $T_1$ values of the most slowly relaxing nuclei in NOM/HS are commonly not known. Intensity variations within the aliphatic region of $^{13}C$ NMR spectra of a soil humic acid, when acquired with different relaxation delays indicate, that even protonated carbon atoms in NOM/HS exhibit rather long longitudinal relaxation times (Peuravuori et al., 2003; Kovalevskii et al., 2000). From an analysis of the S/N ratio, normalized with respect to the number of scans (which is further increasing, even when the relative proportions of the fundamental building blocks have reached a plateau (Hertkorn et al., 2002), it can be deduced that ($^{13}C$) $T_1$ values of individual nuclei in the order of dozens of seconds are rather common in NOM/HS. The parameters used in our laboratory for the acquisition of routine quantitative NMR spectra of NOM/HS with a modern 500 MHz ($B_0$ = 11.7 T) spectrometer, representing the practical compromise between a meaningful S/N ratio and total acquisition time, are given in Table 28 (cf. chapter 7).

## 2.1. NMR SENSITIVITY OF NUCLEI OF INTERST WITHIN NOM/HS RESEARCH

The NMR receptivity D is proportional to the signal to noise ratio (S/N) and it is defined as the product of the NMR sensitivity and the number of nuclei (Mason, 1987), and depends on the following variables:

$$D \sim S/N \sim N\, \gamma_{exc}\, \gamma_{det}^{3/2}\, B_0^{3/2}\, (NS)^{1/2}\, T_2^{*}/T\, [I\,(I+1)] \qquad (1)$$

Here, N is the number of nuclei (corresponding either to natural or to a (de)enriched isotopic abundance), $\gamma_{exc}$ is the magnetogyric ratio of the excited spin [rad $s^{-1}T^{-1}$]; $\gamma_{det}$ is the magnetogyric ratio of the detected spin [rad $s^{-1}T^{-1}$]; $B_0$ is the static external magnetic field [T]; NS is the number of experiments; $1/T_2^{*}$ is the homogeneous line width [Hz]; T is the absolute temperature [K]; I is the isotope spin quantum number. Other equations, relating D to spectrometer-linked parameters (Webb, 1997), are useful to further optimise acquisition and sample conditions.

The NMR nuclei of primary interest in NOM/HS research are depicted in Figure 13 according to their significance: *constituent nuclei* form the backbone or the primary functional groups of NOM/HS; *tag nuclei* serve to characterize functional groups in NOM/HS after chemical derivatization (cf. Figure 43); *metal nuclei* are those with favourable NMR characteristics due to high NMR receptivity with a modest ($^{27}$Al, $^{51}$V) or absent (I = ½ : $^{113}$Cd, $^{195}$Pt, $^{199}$Hg, $^{205}$Tl, $^{207}$Pb) quadrupole moment, and the *"pore" nuclei* are the noble gases $^{3}$He and $^{129}$Xe, which can be spin polarized (Brunner, 1998), to sensitively detect pore size distributions by means of solid – state NMR spectroscopy of NOM/HS (Magusin et al., 1997), but which also show potential to efficiently transfer magnetization to dissolved NOM/HS (Navon et al., 1996).

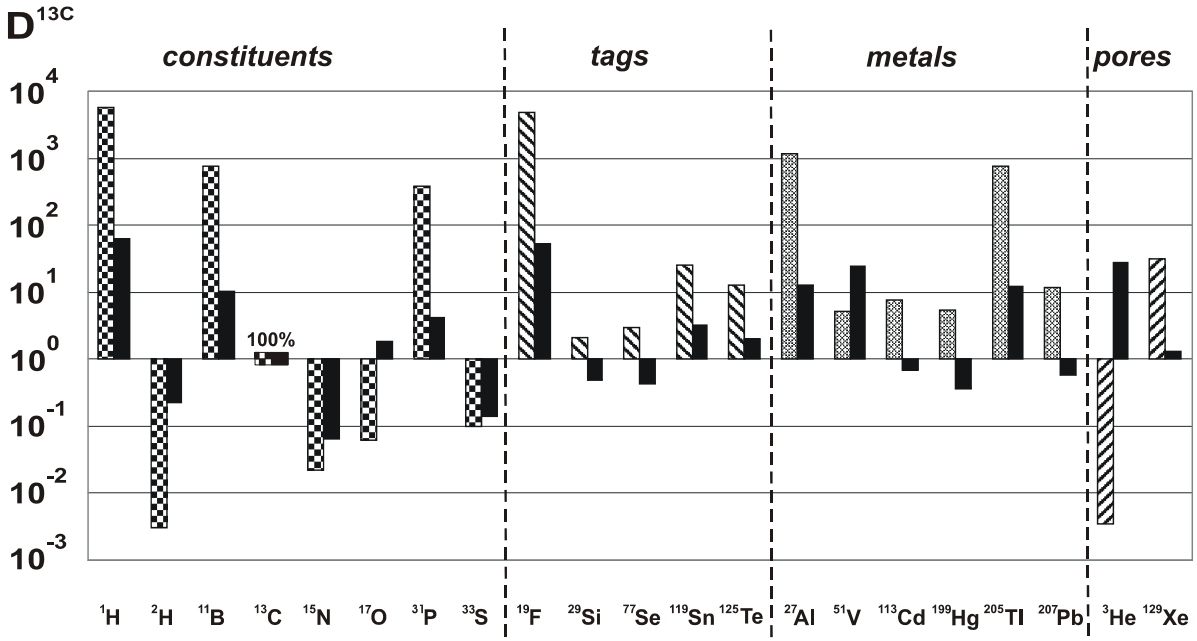



**Figure 13.** NMR receptivity of nuclei of primary interest in NOM/HS research at natural abundance, normalized to $^{13}C_{nat}$ = 1 (left columns) and at 100% isotopic abundance, normalized to $^{13}C_{100\%}$ = 1 [right (black) columns] (cf. text); $^{31}P$ is a constituent (functional group) (Paytan et al., 2003; Makarov et al., 1997; Cade-Menun et al., 2002), and a tag, (Crestini and Argyropoulos, 1997; Tohmura and Argyropoulos, 2001), nucleus.

## 2.2. RELATIONSHIP OF NMR CHEMICAL SHIFT VALUES AND RESONANCE INTEGRALS TO STRUCTURAL PARAMETERS IN NOM/HS

In the NMR based structural analysis of NOM/HS, the most widely used approach is to relate chemical shifts and NMR resonance integrals as observed in a NMR spectrum to structural features in NOM/HS. This direct approach, however, when applied without appropriate care, can lead to serious misinterpretations, because the observed resonance positions in NMR spectra of NOM/HS are a result of various "direct" and "indirect" effects and interactions. The so-called "direct" effects directly, understandably and quantitatively relate chemical shifts and resonance integrals in experimental NMR spectra with chemical environments in NOM/HS. The so-called "indirect" effects are caused by additional interactions occurring in NOM/HS, which affect chemical shifts and resonance intensities in experimental NMR spectra and therefore obscure the facile conclusions drawn by the sole analysis of "direct" effects. A list of "indirect" effects is provided in Table 1 together with the NMR nuclei most likely to become affected; in chapters 2.3. to 2.5, a more in-depth outline will be provided. The peculiar nature of NOM/HS as a complex mixture of small and large molecules in association with minerals and related by a diversity of interactions requires that a dependable NMR analysis has to properly address both "direct" and "indirect" effects on the chemical shift.

**Table 1.** "indirect" effects on the chemical shift in NOM/HS (cf. text).

| **name of effect** | **NMR nuclei most likely to become affected** |
|---|---|
| association / aggregation | $^{1}$H, $^{19}$F |
| differential longitudinal relaxation | low frequency X-nuclei and all quadrupole nuclei |
| differential transverse relaxation | $^{1}$H |
| charge effects | $^{1}$H, $^{15}$N, $^{17}$O, $^{33}$S, all nuclei with free electron pairs |
| hydrogen bonding | $^{1}$H, $^{15}$N, $^{17}$O, $^{33}$S, all nuclei with free electron pairs |
| chemical shift anisotropy | $^{199}$Hg (linear coordination), X-nuclei with a very large range of chemical shift |
| coordination chemical shifts | $^{1}$H, $^{15}$N, $^{17}$O, $^{33}$S, all nuclei with free electron pairs |
| paramagnetics (radicals and metal ions) | $^{1}$H, $^{15}$N, $^{17}$O, $^{33}$S, any nucleus |
| chemical exchange | $^{1}$H, $^{17}$O, metals |
| severe background resonance from probe | $^{11}$B, $^{17}$O, $^{23}$Na, $^{27}$Al, $^{29}$Si |

## 2.3. PROTON NMR SPECTROSCOPY OF NOM/HS

Hydrogen in NOM/HS is bonded to carbon atoms, forming stable covalent bonds (NOM/HS backbone), and to heteroatoms (O > N > S), forming labile bonds (NOM/HS functional groups), which are susceptible to chemical exchange on the NMR timescale (cf. chapter 2.6 and 3.1.2.). Proton NMR spectra of NOM/HS in NaOD, acquired under routine conditions, combine all of the labile proton species into a single NMR resonance, which represents the weighed average chemical shift of all chemically exchanging protons in NOM/HS. A highly efficient suppression of this very strong NMR resonance is required and various experimental protocols have been applied (Lee et al., 1998), to obtain the relative quantities of non-exchangeable hydrogen in various substructures of NOM/HS.

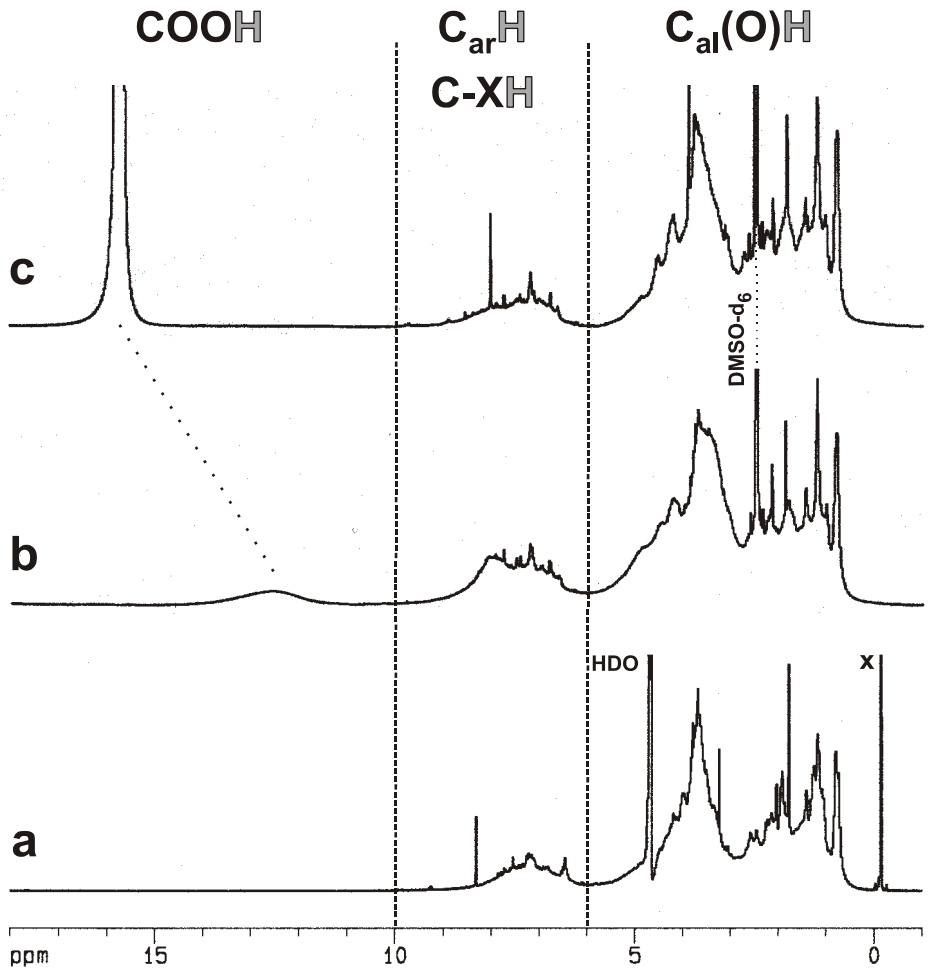



**Figure 14.** $^1$H NMR spectra of a soil humic substance Bouzule FA 3; (a) in 0.1N NaOD (x: reference $(H_3C)_3Si$-$CD_2$-$CD_2$-COONa: –0.14 ppm) with solvent suppression; (b) in DMSO-$d_6$ under total exclusion of moisture; single pulse NMR spectrum; (c) in DMSO-$d_6$ with some $CF_3COOD$ added; single pulse NMR spectrum.

In DMSO-$d_6$ under total exclusion of moisture, the chemical exchange of labile protons is slowed down sufficiently to allow the observation of different forms of these proton species within NOM/HS, which appear in distinctly separate regions of chemical shift (cf. figure 31), (Beck and Reuter, 1974). Consecutive addition of trifluoroacetic acid (preferably $CF_3COOD$) combines all labile proton NMR resonances into a single far downfield signal ($\delta(^1H)$ = 13-17 ppm). Taking into account subtle variations of chemical shifts caused by effects of pH variation on the charge and the aggregation of NOM/HS, a comparison of the three proton NMR spectra of NOM/HS acquired in NaOD, DMSO-$d_6$ and DMSO-$d_6$ / $CF_3COOD$ produces a very detailed assessment of the relative proportions of stable and labile proton species in NOM/HS; the proposed NMR resonance assignments can be corroborated by various forms of homonuclear 2D NMR spectroscopy (cf. chapter 3.1).

## 2.4. NMR INVESTIGATIONS OF "SMALL" SPIN-1/2 NUCLEI, CONSTITUTING THE HUMIC BACKBONE STRUCTURE

In the case of the "small sized" spin-1/2 nuclei $^1H$, $^{13}C$, $^{15}N$ and $^{31}P$, which constitute the backbone of HS/NOM, the relationship between chemical shift values and the corresponding structural features of NOM/HS is rather well understood (Preston, 1996; Mahieu et al., 1999; Makarov et al., 1997; Cade-Menun et al., 2002). In routine NMR analysis of NOM/HS, these NMR spectra are divided into several regions of chemical shift, which are considered to represent coarse substructure regimes as shown in Figures 15, 16.

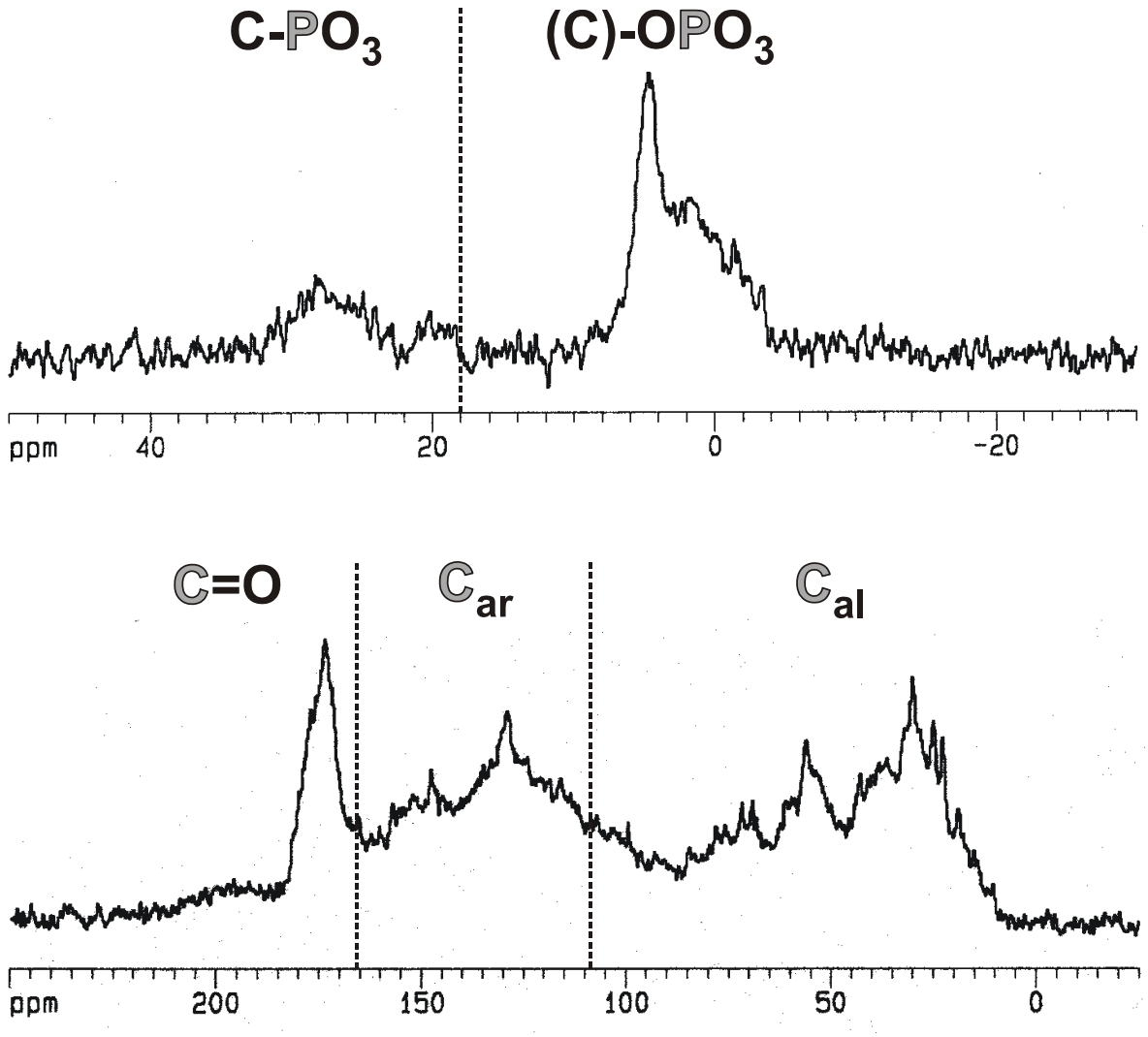



**Figure 15.** $^{13}C$ NMR (bottom) of a municipal solid waste humic acid, and $^{31}P$ NMR (top) spectrum of a marine humic substance, with the chemical shift ranges of the fundamental building blocks indicated.

These quantitative 1D NMR spectra define the relative proportions of the fundamental building blocks of NOM/HS and therefore provide a key and indispensable information and reference against which any structural model of NOM/HS has to be judged (Preston, 1996; Mahieu et al., 1999). Within these coarse substructure regimes, a detailed analysis of chemical shift distribution can reveal considerable detail, especially when used in combination with spectral editing ($^{13}C$ and $^{15}N$ NMR) and higher dimensional NMR spectroscopy (cf. chapters 2.4.1 and 3.).

### 2.4.1. SPECTRAL EDITING OF CARBON NMR SPECTRA

Spectral editing and, in particular, sorting carbon atoms according to their number of directly bound protons (Shin et al., 1996; Shin and Moon, 1996; Buddrus et al., 1989), is a very powerful assignment tool, primarily within the aliphatic and aromatic sections of the $^{13}C$ NMR spectra of NOM/HS. The subspectra of methyl, methylene and methine carbon atoms are calculated from linear combinations of three individual DEPT spectra (Sorensen and Jakobsen, 1988).

The two main applications of spectral editing for the structural analysis of NOM/HS are the evaluation of branching in aliphatic units and the determination of the average degree of carbon substitution in aromatic rings. In aliphatic compounds $C_nH_{2n+2}$, proton chemical shifts are only marginally affected by branching; extensive steric hindrance causes a minor degree of deshielding, which is not clearly related to simple structural details. However, in the carbon frequency, both $\alpha$- and $\beta$-aliphatic substitution cause a deshielding of ~10 ppm, while $\gamma$-aliphatic substitution induces an upfield shift of ~4 ppm (Breitmaier and Voelter, 1990). Multiplicity information in conjunction with chemical shift data allows to draw detailed conclusions about the remote aliphatic substitution in NOM/HS, which can be further refined by the analysis of HSQC NMR spectra (cf. chapters 3.1.4 and 8.3.1.3.).

In absence of very fast transverse relaxation (Buddrus et al., 1989), the average number of protons per aromatic ring in NOM/HS is obtained by comparing the ratios of the (normalized) areas of a $^{13}C$ NMR DEPT spectrum (showing protonated carbon only) and the single pulse $^{13}C$ NMR spectrum (showing all carbon atoms). For the analysis of NOM/HS, it is recommended to use short DEPT pulse sequences, since they are the least susceptible to differential relaxation. However, the subspectral editing according to quaternary carbon atoms via routine QUAT-sequences visibly suffers from the variance of the $^{1}J(CH)$ values in NOM/HS; within the aliphatic section of the $^{13}C$ NMR spectrum, crosstalk from protonated carbon atoms obscures the unambiguous definition of genuine quaternary carbon atoms in NOM/HS.

## 2.5. NMR OF METAL NUCLEI WITH SPIN ½, COORDINATED TO NOM/HS

With a very few exceptions, all metals feature at least one NMR active isotope and thereby provide the opportunity to directly observe metal binding to NOM/HS by metal NMR spectroscopy (Li et al., 1998; Lambert et al., 1995; Lu et al., 1998; Howe et al., 1997). However, the understanding of the NMR chemical shift - structure relationship of coordination and organometallic compounds, where alterations in coordination number, symmetry and oxidation state can readily occur (Öz et al., 1998; Benn and Rufinska, 1986), remains rather restricted for all NMR spectra of metal nuclei until today. Any of these above mentioned effects and, in addition, pronounced concentration (Howe et al., 1997; Harrison et al., 1983); pH (Howe et al., 1997; Nakashima and Rabenstein, 1983) and temperature dependencies (Harrison et al., 1983; Medek et al., 1997; Claudio et al., 2000), very strongly influence the metal NMR chemical shift – and all of these effects are highly specific for any metal NMR nucleus. In the case of paramagnetic complexes, metal NMR resonances show large shift displacements from those of diamagnetic counterparts (Walker, 2003; Walker, 1999) and are commonly broadened beyond detection due to very efficient electron-nuclear dipole-dipole interactions. The current lack of understanding how all these processes will translate into chemical shifts of metal nuclei impedes current research: so far many of the metal nuclei with high NMR receptivity have never been used in the investigation of metal binding to NOM/HS.

Chemical processes, like internal rotations in molecules, valence isomerization, chemical exchange and chemical reactions can lead to exchange of nuclei between non-equivalent electronic surroundings. Kinetic processes on time scales ranging from μs to s are effective in causing chemical exchange effects in NMR spectra (Palmer 2002; Palmer et al., 2001). Naturally, the near continuous distribution of binding constants and rates of chemical exchange of humic and NOM ligands to metals will impose a strong predominance of effects of chemical exchange on the metal NMR spectra. For that very reason, the chemical shifts and resonance integrals in NMR spectra of metal nuclei as obtained by direct analysis of experimental NMR spectra of metal ions coordinated to NOM/HS do not directly indicate the content and the nature of the chemical environment of the respective nuclei. Instead, these NMR spectra most likely will represent a weighed average of superimposed slow, intermediate and fast chemical exchange characteristics, produced by groups of coordination environments, which are related

by chemical shift differences and exchange rates. In metal NMR spectra, which will invariably show broadened resonances and a limited signal-to-noise ratio when coordination of metals to NOM/HS comes into effect, it is not feasible to define an unambiguously defined status of slow chemical exchange, which is an initial requirement to allow any exhaustive mathematical analysis of chemically exchanging systems (Hertkorn et al., 2004). Since metal ions are partners in the chemical exchange, the regime of slow chemical exchange is reached only at extremely low metal ion concentrations, when the S/N ratio of the metal NMR spectra becomes prohibitively low. Even conditions of pure fast chemical exchange in metal-NOM/HS systems are very difficult to prove experimentally, as conditions of a dynamic equilibrium, defined by the superposition of the temperature dependent binding constants and exchange rates of each individual metal coordination environment, are reached (Hertkorn et al., 2004). It is interesting to note that at a very low metal ion concentration, only the strongest binding sites within NOM/HS will be occupied. This may induce a selectivity of binding vastly different from that observed at elevated metal to NOM/HS ratio (Drexel et al., 2002; Haitzer et al., 2002).

But with these particulars under consideration, metal NMR spectroscopy offers the most direct and most detailed information concerning structure and dynamics of coordination compounds in NOM/HS. Any reliable analysis of metal binding of NOM/HS by metal NMR spectroscopy requires a systematic mapping of the exchanging system by variation of the metal to NOM/HS ratio, pH, temperature and – possibly – concentration in conjunction with an independent determination of the total metal content and the fraction of the (free) metal ions (Hertkorn et al., 2004). The acquisition of the metal NMR spectra at different field strengths $B_0$ is advantageous to further characterize the NMR exchange characteristics (Hertkorn et al., 2004; Millet et al., 2000).

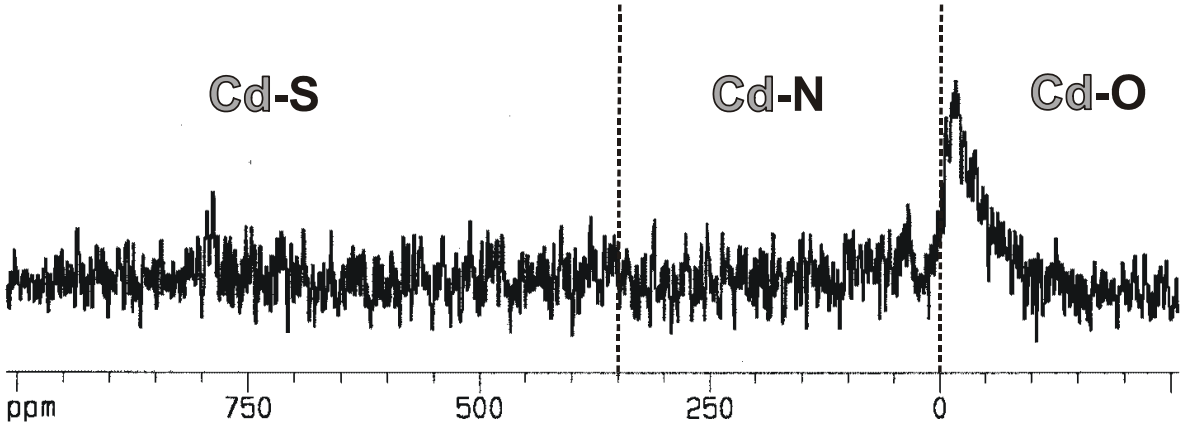



**Figure 16.** $^{113}Cd$ NMR spectrum of Suwannee River NOM with $Cd(ClO_4)_2$ at a Cd/C ratio of $4.16.10^{-3}$, pH = 10.2, $B_0$ = 11.7 T, (Hertkorn et al., 2004).

Several of the environmentally significant and toxic metal isotopes are spin ½ nuclei ($^{113}Cd$, $^{195}Pt$, $^{199}Hg$, $^{203/205}Tl$, $^{207}Pb$) with a reasonable NMR receptivity (cf. Figure 13) and therefore are suitable targets for direct observation of metal coordination to NOM/HS.

## 2.5.1. CASE STUDY 1: A Potentiometric and $^{113}$Cd NMR Study of Cadmium Complexation by Natural Organic Matter at Two Different Magnetic Field Strengths

### *2.5.1.1. Abstract*

***The binding of cadmium to Suwannee river natural organic matter (NOM) has been investigated across a broad range of Cd/C ratios (0.00056 – 0.0056) and pH values (3.5 - 11) by $^{113}$Cd NMR spectroscopy at two magnetic field strengths ($B_0$ = 9.4 and 11.7 T). Caused by the very peculiar and highly complex nature of the Cd-NOM exchanging system, these $^{113}$Cd NMR spectra are characterized by a pH and concentration dependent superposition of slow, intermediate and fast chemical exchange. The complex interplay of solution chemistry and chemical exchange requires a thorough mapping of this Cd-NOM chemically exchanging system through NMR acquisition at two magnetic field strengths and a systematic variation of Cd/C ratios and pH values. The interpretation of $^{113}$Cd NMR spectra is greatly facilitated and constrained by simultaneous measurements of pH and pCd, which allows a model-independent calculation of organically bound $Cd^{2+}$ under all experimental conditions. Within the range of chemical conditions applied in this study, $^{113}$Cd NMR spectrometric evidence is consistent with coordination of cadmium by oxygen, nitrogen and sulfur ligands in NOM. Under all experimental conditions, cadmium is primarily coordinated to oxygen; however, several lines of evidence point to the participation of nitrogen ligands, even in acidic solutions where nitrogen ligands are primarily bound to protons. Under alkaline conditions, up to one third of cadmium may be coordinated to nitrogen, and a small, but unquantifiable, percentage of cadmium is coordinated to sulfur ligands, as evidenced by far low-field $^{113}$Cd NMR resonances.***

### *2.5.2. Introduction*

Cadmium occurs ubiquitously in soils, sediments, and natural waters, and it has long been recognized that the chemistry of cadmium is modified through its interactions with both inorganic and organic constituents in those environments. In the last three decades, numerous papers have been published on cadmium binding by humic substances (Oste et al., 2002; Otto et al., 2001; Otto et al., 2001), soil organic matter (Balabane and van Oort, 2002), natural organic matter – NOM (Lowe, 1992), mineral surfaces (Randall et al., 1999; Dong et. al.,

2000; Arias et al., 2000), sediments (Huang, 2001; Eimers et al., 2002; De La Torre and Tessier, 2002; Desy et al., 2002), and a variety of biological materials (Adhiya et al., 2002; Carrilho et al., 2002; Plette et al., 1996; Xia and Rayson, 2002; Grassi and Mingazzini, 2001; Bhanoori and Venkateswerlu, 2000). The latter studies include efforts to understand the toxic effects of cadmium to biota (Griffin et al., 2000), and efforts to utilize biomass as an adsorbent for remediation of cadmium-contaminated waters, (Xia and Rayson, 2000; Keun et al., 2002).

Studies of cadmium binding by NOM, which in this paper is understood to include aquatic humic acids, aquatic fulvic acids, and other organic constituents of natural waters, have relied mainly on potentiometric measurements (Gardiner, 1974; Holmes et al., 2001) to distinguish between free and complexed cadmium in aqueous samples. In the study of multiphasic systems (Lee et al., 1996), bound cadmium can be physically separated from free cadmium, and the two can be quantified separately. Such studies can quantify the relative proportions of free and complexed cadmium in the system under investigation, but they cannot provide substantial insight into the nature of the donor atom(s) responsible for binding cadmium. Such insight is needed to understand the competition, or lack thereof, between cadmium and other metal ions that have some affinity for NOM. The present study is restricted to the competitive binding of $H^+$ and $Cd^{2+}$ to binding sites in NOM.

In the last few years, several studies have described the use of $^{113}$Cd nuclear magnetic resonance spectroscopy to more clearly elucidate the nature of metal binding sites in NOM. (Otto et al., 2001; Otto et al., 2001; Larive et al., 1996; Chung et al., 1996; Kulikova and Perminova, 2002). Although some of the first studies were unfortunately flawed because the effect of added $Cl^-$ on the solution chemistry of $Cd^{2+}$ was overlooked (Larive et al., 1996), a relatively consistent description of the complexation of $Cd^{2+}$ by NOM is beginning to emerge from $^{113}$Cd NMR studies (Lowe, 1992; Otto et al., 2001; Otto et al., 2001; Larive et al., 1996; Chung et al., 1996; Kulikova and Perminova, 2002) . Most recent studies strongly support the hypothesis that $Cd^{2+}$ reacts primarily with O-ligands (Al Anet, 2004; Ricardo et al., 2004; Sanudo-Wilhelmy et al., 2002; Oste et al., 2002; Dong et al., 2000). Only at alkaline pH do N-ligands appear to play a role in the complexation of $Cd^{2+}$ (Lowe, 1992).Most recent studies have been conducted at mildly acidic pH (usually 6.0-6.4), where only a single, rather broad peak is observed in the $^{113}$Cd NMR spectrum. These observations have led to the conclusion that $Cd^{2+}$ is undergoing relatively rapid chemical exchange under the conditions of these

experiments (Otto et al., 2001; Larive et al., 1996). In contrast, broad, complex and poorly resolved spectra were observed at alkaline pH (Lowe, 1992). Those authors concluded that fast exchange of $Cd^{2+}$ does not occur under alkaline conditions.

Upon closer examination, some of the reported observations, even if substantially correct, may not provide significant insight into the actual coordination chemistry of $Cd^{2+}$ in natural environments. The observation that nearly all complexation of $Cd^{2+}$ involves O-ligands, for example, is an inescapable consequence of conducting $^{113}$Cd NMR experiments at a Cd/C molar ratio that exceeds the N/C and S/C ratios of the NOM. Of the published results that are known to the authors (Otto et al., 2001; Otto et al., 2001; Larive et al., 1996; Chung et al., 1996; Kulikova and Perminova, 2002; Chung and Moon, 1994; Perminova et al., 2001; Chung and Moon, 1996; Balcke et al., 2002) at least half of the measurements were made under such conditions. Specifically, Cd/C ratios were so high that Cd/N ratios were greater than one in 50% of the experiments, and Cd/S ratios were greater than one in 92% of the measurements. The situation is made even worse by the fact that not all N and S occur as amines and thiols – perhaps the optimum forms for metal complexation. For instance, XANES spectroscopy suggests that most of the sulfur in Suwannee River fulvic acid occurs as organic sulfonates and organic sulfates (Morra et al., 1997). Likewise, $^{15}$N NMR studies have shown that most nitrogen in humic substances occurs as amides, (Bortiatynski et al., 1996; Knicker et al., 1997; Kögel-Knabner, 2000).

On the issue of the participation of $Cd^{2+}$ in fast exchange reactions, a variety of mechanistic explanations have been offered. Chung et al., 1996 believe that their $^{113}$Cd NMR spectra "reflect the rapidly exchanging conditions in aqueous solution and involve many species that [are] formed by a diversity of functional groups of fulvic acid. Larive et al., 1996, state that their spectra "are consistent with the fast exchange model". They describe the interaction between $Cd^{2+}$ and Suwannee River fulvic acid with a one-site model of fulvic acid, which contrasts sharply with the "diversity of functional groups of fulvic acid" that are suggested by Chung et al., 1996*;* Otto et al.,2001, on the basis of measurements of $T_1$ and $T_2$ relaxation times for solutions containing $Cd^{2+}$ and Suwannee River fulvic acid, concluded that fulvic acids contain two major types of binding sites – a strong site in which bound $Cd^{2+}$ relaxes rapidly and a weaker site in which bound $Cd^{2+}$ relaxes more slowly. They also state that, under their experimental conditions, chemical exchange appears to be a minor contributor to $^{113}$Cd relaxation, at least for strongly bound $Cd^{2+}$. In the only study that has reported $^{113}$Cd

NMR spectra of Cd-NOM systems outside the pH range of 6.0-6.4 (Lowe, 1992), the remarkable changes in the $^{113}$Cd NMR spectra with increasing pH have been attributed to a decreased rate of chemical exchange. It was suggested that the rates of complexation reactions between $Cd^{2+}$ and binding sites in NOM should decrease at higher pH because of the lower activity of $Cd^{2+}$ at higher pH.

This case study presents a concerted effort to combine a detailed analysis of the chemical speciation of $Cd^{2+}$ over a range of pH and Cd/C ratios with an equally detailed $^{113}$Cd NMR study of those solutions. To provide further insight into chemical exchange in this system, $^{113}$Cd NMR spectra were acquired at two field strengths (9.4 T and 11.7 T), and the temperature dependence of chemical shift and linewidth was analyzed briefly.

### *2.5.3. Technical background*

2.5.3.1. Complexation of Cadmium by NOM: The reaction of $Cd^{2+}$ with NOM, and especially its reactions with fulvic acids and humic acids, have been studied and modeled for many years (Oste et al., 2002; Gardiner, 1974; Holmes et al., 2001; Brady and Pagenkopf, 1978; Saar and Weber, 1982; Xue and Sigg, 1998; Antunes et al.,2002; Liu and Gonzalez, 2000; John et al., 1988). Early modeling efforts, which were conducted before the advent of computers, were mainly restricted to fitting simple one-site or two-site models to specific data sets, (Brady and Pagenkopf, 1978; Saar and Weber, 1982; John et al., 1988). Such modeling efforts often used Scatchard plots, double-reciprocal plots, log-log plots, and other algebraic manipulations of experimental data to facilitate the graphical estimation of binding site concentrations and equilibrium constants. Such graphical approaches can be highly misleading when applied to systems that contain more than one type of binding site (Klotz, 1982). The first efforts to model the binding of $Cd^{2+}$ as deduced from $^{113}$Cd NMR spectra have relied upon such a modeling approach (Otto et al., 2001; Larive et al., 1996).

Over the years, analytical methodologies have advanced and, with the advent of computers, more robust numerical models have emerged. Today, the binding of $Cd^{2+}$ by NOM can be modeled as a function of pH, competition by other metal ions, etc. A particularly excellent and appropriate example is the work of Kinniburgh et al., 1999, who used the thermodynamically consistent NICCA-Donnan model to describe a variety of metal binding equilibria. Among the examples in that paper is a beautiful set of data and model curves for binding of

$Cd^{2+}$ by a purified peat humic acid over a wide range of pH. The binding of $Cd^{2+}$ was very well described by their model in the absence and presence of competition by $Ca^{2+}$.

The objectives of this case study do not include mathematical modeling of chemical speciation in the Cd-NOM system. A brief discussion of the conceptual framework of the most popular models of chemical speciation is nonetheless warranted. Three modern families of models are the Gaussian models (Perdue and Lytle, 1983; Perdue et al., 1984; Perdue and Parrish, 1987; Allison and Perdue, 1994; Dobbs et al., 1989; Perdue and Carreira, 1997) Model V and Model VI (Tipping, 1994; Tipping and Hurley, 1992; Tipping, 1998), and the aforementioned NICA group of models (Tipping and Hurley, 1992; Benedetti et al., 1995; Benedetti et al., 1996; Kinniburgh et al., 1996; Koopal et al.,1994; Milne et al., 2001). All of these models treat NOM, fulvic acids, humic acids, etc. as highly complex mixtures that contain a continuum, or effectively so, of non-identical binding sites whose equilibrium constants for binding of protons or metal cations (such as $Cd^{2+}$) span a range of several orders-of-magnitude. Although each family of models has certain unique attributes, the similarities between the models are much greater than the differences (Perdue, 1998; Perdue, 2001). If these modern models of chemical speciation are to be believed, then the overly simplified models that have been used previously (Otto et al., 2001; Larive et al., 1996), to aid in the interpretation of $^{113}$Cd NMR spectra in the presence of NOM, fulvic acid, etc. are unlikely to yield meaningful results. All subsequent presentation and interpretation of chemical and spectrometric data in this paper are consistent with modern view of metal complexation by NOM, i.e., that metal cations are complexed by a highly complex mixture of non-identical binding sites.

Even without sophisticated models, an understanding of the complexity of the distribution of binding sites in NOM is a useful guide in planning and interpreting metal binding experiments. For example, the experimental protocol for this study calls for the use of relatively low Cd/C molar ratios, even lower than the S/C ratio of the NOM in two of the three experiments. Under these conditions, binding sites are present in great excess relative to added $Cd^{2+}$, and the degree of complexation of $Cd^{2+}$ will be determined mainly by its competition with $H^+$ for binding sites in the NOM. Naturally, the degree to which $H^+$ competes with $Cd^{2+}$ for a binding site depends on the acidic strength of that site. It follows that both the compositional diversity of potential binding sites and the "available" fraction of each type of binding site are functions of pH. At low pH, strongly acidic carboxylic acids, i.e., those that are in close proximity to

polar electron-withdrawing groups and/or functional groups bearing a formal positive charge, are more fully dissociated than are weaker acids. As pH increases into the alkaline pH range, however, weakly acidic carboxyl groups, organic ammonium ions, thiols, and phenols are all more fully dissociated and potentially more involved in the complexation of $Cd^{2+}$. As $Cd^{2+}$ reacts with these weaker acids at higher pH, the concentration of bound $Cd^{2+}$ should increase and the concentration of free $Cd^{2+}$ should decrease. It is thus expected that pCd will increase as pH is increased. At any given pH, solutions containing a higher total concentration of cadmium (higher Cd/C ratio) should have a higher concentration of free $Cd^{2+}$ (lower pCd).

2.5.3.2. Effects of Chemical Exchange on NMR Spectra: The chemical shift of the $^{113}$Cd nucleus is remarkably sensitive to the nature, number and geometric arrangement of ligands (Bloom and Leenheer, 1989; Megumu et al.,1986; Summers, 1988), and the observed signal dispersion of 950 ppm for coordination compounds of cadmium with oxygen, nitrogen and sulfur donor atoms can potentially provide well-resolved $^{113}$Cd NMR spectra, even for closely related types of coordination. The $^{113}$Cd NMR chemical shifts of representative Cd-humic species are expected to resemble those of related low, (Chung and Moon, 1996; Summers, 1988; Wang and Gilpin, 1983) and higher molecular, (Summers, 1988; Bloom and Leenheer, 1989) coordination compounds. This otherwise excellent spectral resolution can be diminished, however, if chemical exchange of $Cd^{2+}$ between a number of non-identical binding sites leads to substantial broadening of $^{113}$Cd resonances. Such exchange-broadened, weighted-average peaks have been observed in the $^{113}$Cd NMR spectra of some simple ligands and biological macromolecules, and in the spectra of NOM and humic substances, (Lowe, 1992; Larive et al., 1996; Chung et al., 1996; Kulikova and Perminova, 2002). When $Cd^{2+}$ undergoes chemical exchange during an NMR measurement, the exchange rate is conventionally referred to as slow, intermediate, or fast exchange, depending on whether the rate of the exchange reaction (in Hz) is much less than, approximately equal to, or much greater than the difference in resonance frequencies of the exchanging species (in Hz).

Even for well-defined chemical systems, NMR parameters such as chemical shift ($\delta_{CdL}$), relaxation times ($T_{1(CdL)}$ and $T_{2(CdL)}$), lineshape, and linewidth are affected by chemical exchange. When dealing with simple, well-defined chemical systems, rigorous numerical methods are available for interpreting results that are obtained under conditions of intermediate or fast exchange, if spectra can be obtained at the limit of slow exchange (Bain, 2003).

NOM is vastly more complex than the simple ligands and biomolecules that are often the focus of $^{113}$Cd NMR studies, (Summers, 1988). The postulated continuum of binding sites with respect to the equilibrium constants for complexation reactions (Allison and Perdue, 1994; Perdue, 1998) that is the foundation of modern models of chemical speciation leads directly to the prediction that the $^{113}$Cd NMR spectra of the Cd-NOM system will be a function of such compositional variables as pH and Cd/C ratio, simply because of the availability of a wide variety of non-identical binding sites (with respect to both $H^+$ and $Cd^{2+}$ binding). Assuming for the moment that Cd(II) can exist only as the free ion, a group of hydroxyl complexes, and as organically bound Cd, the relationship between the chemical speciation of $Cd^{2+}$ and the observed chemical shift ($\delta_{obs}$) of $^{113}$Cd can be more fully developed. Under the limiting condition of fast chemical exchange, the observed chemical shift of $Cd^{2+}$ is expected to be a weighted-average of the chemical shifts of all forms of Cd(II) in the solution, i.e.:

$$\delta_{obs} = \delta_{Cd^{2+}}\left(\frac{[Cd^{2+}]}{Cd_{Total}}\right) + \sum_i \delta_{Cd(OH)_i^{2-i}}\left(\frac{[Cd(OH)_i^{2-i}]}{Cd_{Total}}\right) + \sum_i \delta_{CdL_i}\left(\frac{[CdL_i]}{Cd_{Total}}\right) \quad (2)$$

Alternatively, the contribution of organically complexed $Cd^{2+}$ to $\delta_{obs}$ can be expressed in terms of an average chemical shift ($\delta_{org}$) of the organically complexed forms of $Cd^{2+}$.

$$\delta_{obs} = \delta_{Cd^{2+}}\left(\frac{[Cd^{2+}]}{Cd_{Total}}\right) + \sum_i \delta_{Cd(OH)_i^{2-i}}\left(\frac{[Cd(OH)_i^{2-i}]}{Cd_{Total}}\right) + \delta_{org}\sum_i\left(\frac{[CdL_i]}{Cd_{Total}}\right) \quad (3)$$

It is readily shown that $\delta_{org}$ is related to the chemical shifts of the individual organic complexes of $Cd^{2+}$ by the equation:

$$\delta_{org} = \frac{\sum_i \delta_{CdL_i}[CdL_i]}{\sum_i [CdL_i]} \quad (4)$$

In other words, $\delta_{org}$ is a weighted-average chemical shift for organically complexed $Cd^{2+}$, and the weighting factors are the concentrations of the individual complexes. The dependence of $\delta_{org}$ on pCd and pH, which is not immediately evident in Eq. 4, lies in the variation of $[CdL_i]$ on these parameters. If both $Cd^{2+}$ and $H^+$ react in a 1:1 mole ratio with each binding site in NOM, as is assumed in modern metal speciation models such as MINTEQA2, (Perdue and Carreira, 1997), then $[CdL_i]$ is given by:

$$[CdL_i] = L_{i,Total}\left(\frac{K_{i,Cd}[Cd^{2+}]}{1+K_{i,H}[H^+]+K_{i,Cd}[Cd^{2+}]}\right) \quad (5)$$

where $L_{i,Total}$ is the total concentration of the $i^{th}$ binding site in NOM, and $K_{i,H}$ and $K_{i,Cd}$ are equilibrium constants for formation of 1:1 complexes between the $i^{th}$ binding site and $H^+$ and $Cd^{2+}$ ions, respectively. In this simplified representation, the distinction between activity and concentration is ignored. It follows from Eq. 4-5 that $\delta_{org}$ is given by:

$$\delta_{org} = \frac{\sum_i \delta_{CdL_i} L_{i,Total}\left(\frac{K_{i,Cd}}{1+K_{i,H}[H^+]+K_{i,Cd}[Cd^{2+}]}\right)}{\sum_i L_{i,Total}\left(\frac{K_{i,Cd}}{1+K_{i,H}[H^+]+K_{i,Cd}[Cd^{2+}]}\right)} \quad (6)$$

This equation will not actually be employed in this paper, because neither the number or nature of the non-identical ligands in NOM is known, including their respective equilibrium constants for binding $H^+$ and $Cd^{2+}$. It is derived here just to emphasize the direct dependence of $\delta_{org}$ on the chemical composition of a Cd-NOM system.

Another less obvious consequence of the wide variety of non-identical binding sites in NOM is the expected variation of $^{113}$Cd NMR spectra with NMR field strength, even when the composition of a test solution is held constant. If the complexation reactions between $Cd^{2+}$ and the wide variety of ligands in NOM are elementary process, then there must exist a

comparably wide variety of rate constants for association and/or dissociation of Cd-NOM complexes. That being the case, it is quite unlikely that there can exist any set of chemical and spectroscopic conditions where all possible exchange reactions could be classified as fast exchange reactions (or, for that matter, intermediate or slow exchange reactions). If intermediate and slow chemical exchange are significant processes in a Cd-NOM system, then $\delta_{obs}$ will vary with field strength even when solution chemistry is invariant. To further complicate the situation, the aforementioned pH-dependent competition between $Cd^{2+}$ and $H^{+}$ for binding sites in NOM should cause the relative proportion of binding sites undergoing fast, intermediate, or slow exchange of $Cd^{2+}$ to be pH-dependent. Given all of these consequences that arise uniquely from the high complexity of the mixture of binding sites in NOM, it is not anticipated that valid results can be obtained if $^{113}$Cd NMR spectra in the Cd-NOM system are analyzed using methods normally applied to pure compounds, (Jensen et al., 1981).

### *2.5.4. Experimental section*

2.5.4.1. Chemicals: Cadmium oxide (CdO, 95.05 % $^{113}$Cd-enriched) was obtained from Chemotrade, Leipzig, Germany and used as received. Analytical grade perchloric acid (70% $HClO_4$) was obtained from Fluka, sodium deuteroxide (40% NaOD) and $D_2O$ (99.95% $^{2}H$) were obtained from Merck, Darmstadt, Germany.

Natural organic matter (NOM) was concentrated by reverse osmosis from the Suwannee River in southeastern Georgia, U.S.A. The concentrated sample was desalted and freeze-dried. The dry, ash-free elemental composition of Suwannee River NOM (as molar ratios) is: H/C = 0.950; O/C = 0.622; N/C = 0.017; S/C = 0.004, (Gao, 1994). A more complete description of the methodology and the chemical composition of this sample have been published previously, (Serkiz and Perdue, 1990). In all following references to NOM, its concentration will be given as the molar concentration of dissolved organic carbon (DOC). More recently, the International Humic Substances Society (IHSS) used the same field and laboratory experimental procedures to collect a reference NOM sample (1R101N) from the same site on the Suwannee River. That reference sample has a very similar elemental composition of: H/C = 0.959; O/C = 0.610; N/C = 0.018; S/C = 0.005).

2.5.4.2. Preparation of stock solutions: Stock solutions of Suwannee River NOM were prepared as needed by dissolving 625 mg of NOM in about 15 mL of $D_2O$ that contained 144 mg of

40% (w/w) NaOD/$D_2O$. When the sample was apparently dissolved, it was transferred quantitatively to a volumetric flask and diluted to 25 mL with $D_2O$. The added NaOD was sufficient to neutralize about 50% of the COOH groups of the NOM, which enabled such a concentrated stock solution to be prepared. The final solution contained 1.04 M dissolved organic carbon (DOC) and 0.055 M $Na^+$.

The stock solution of $^{113}Cd(ClO_4)_2$ was prepared by reacting 75.2 mg of $^{113}$CdO(s) with 100 µL of 70% (w/w) $HClO_4$ in 5 mL of $D_2O$ in a closed vessel at 40ºC for several hours. The solution was transferred quantitatively to a 10-mL volumetric flask and diluted with $D_2O$. This stock solution contained 0.0583 M $^{113}Cd(ClO_4)_2$.

2.5.4.3. NMR spectroscopy: $^{113}$Cd NMR spectra were acquired with a Bruker AC 400 NMR spectrometer ($B_0$: 9.4 T; frequency: $^{113}$Cd: 88.76 MHz) and a Bruker DMX 500 spectrometer ($B_0$: 11.7 T, frequency: $^{113}$Cd: 110.95 MHz) at 303 K, unless otherwise specified. At 9.4 T, a 15-mm broadband observe probehead was employed with a $^{113}$Cd 90-deg pulse length of 35 µs. At 11.7 T, a 10-mm probehead was used with a $^{113}$Cd 90-deg pulse length of 13.5 µs. Single pulse experiments were performed without proton decoupling with typical acquisition conditions: 60-deg pulses with relaxation delays (d1) ranging from 10 ms to 5 s and acquisition times (aq) ranging from 20.3 ms to 1.35 s at a sweep width of 100 kHz (9.4 T) and 125 kHz (11.7 T). The initial delays (de) ranged from 6 µs to 55 µs. Smaller "de" values were used at low pH values; however, longer "de" values produced a more even baseline at pH > 7. The chemical shift reference was external 0.1 M $Cd(ClO_4)_2$ at $\delta(^{113}Cd) = 0$ ppm. The $^{113}$Cd NMR spectra required between 30 minutes and several days to obtain.

The authenticity of the NMR spectra was checked by repetitive measurements of sample spectra at various d1 and aq values. Under conditions of single Lorentzian resonances, no deviation of lineshape and line position could be detected across the whole range of d1 and aq that were employed. Under conditions of prevailing intermediate and slow chemical exchange (at higher pH), the signal-to-noise ratio per scan (at lower signal-to-noise ratio per time unit) increased as d1 + aq was increased from 120 ms up to several seconds. Using the relative signal-to-noise ratios at various delay times, longitudinal $^{113}$Cd NMR relaxation times were estimated to be in the range of 0.75 to 2.5 s, in line with the values obtained by Otto et al., 2001. There was no measurable effect on the relative distribution of resonance positions. The observed distribution of resonances was not affected by the pulsewidth-dependent excitation

profile, (van de Ven, 1995), as demonstrated by sample spectra taken at different center-frequency offset values. These preliminary tests indicate that the NMR operating parameters used in this study yield quantitative $^{113}$Cd NMR spectra. Owing to their reduced effective bandwidth of excitation, anti-ring sequences could not be used in these series of experiments.

Variable temperature studies were performed with 5 mm broadband observe probeheads showing 90-deg pulses of 15 μs (9.4 T) and 8.2 μs (11.7 T). The temperature during these series of measurements was calculated from a 5 mm $CH_3OH$ sample tube according to, (van Geet, 1970). It was noted that linewidths measured in 5 mm tubes were a few percent smaller than linewidths acquired with 10-15 mm tubes, most probably indicating a more uniform temperature distribution within the smaller tube.

A sine modulation (less than one period) was observed in some of the $^{113}$Cd NMR spectra that were obtained at higher pH, and such spectra were manually baseline-corrected using a sine function. Curve-fitting of the NMR spectra was performed with Bruker standard XWin-NMR 3.0 software on a Unix workstation (AC 400 spectra were converted from the ASPECT 3000 format to the XWin-NMR format). Lorentzian lineshapes were used, with a peak picking constant (pc) of 0.8 and an integration processing parameter (azfw) of 200 ppm. The Lorentzian linewidths $\Delta\nu_L$, which were obtained from Lorentzian fits of experimental $^{113}$Cd NMR spectra that have been calculated with 50 Hz exponential broadening, were not corrected for effects of apodization (van de Ven, 1995).

2.5.4.4. Potentiometric measurements: All potentiometric measurements of the activities of $H^+$ and $Cd^{2+}$ were obtained using a Metrohm 713 pH meter. For pH measurements (pH = -$\log_{10}${$H^+$ activity}), a Metrohm combination pH electrode (6.0219.110) was used. The inner and outer chambers of the pH electrode were filled with 3 M KCl and 3 M $NaNO_3$, respectively. The pH electrode was calibrated at room temperature using standard Fluka pH buffers (pH 4.008, 6.865, and 9.180 at 25°C). For pCd measurements (pCd = -$\log_{10}${$Cd^{2+}$ activity}), a Metrohm cadmium ion selective electrode (6.0502.110) and a Metrohm double-junction Ag|AgCl reference electrode (6.0726.110) were used. The inner and outer chambers of the reference electrode were filled with 3 M KCl and 3 M $NaNO_3$, respectively. The cadmium electrode was calibrated at room temperature using $10^{-3}$ M, $10^{-4}$ M, and $10^{-5}$ M solutions of $Cd(ClO_4)_2$ that were adjusted to pH 4 with $HClO_4$. Using the Davies equation to calculate the activity coefficient of $Cd^{2+}$ in these solutions, their computed pCd values are 3.103, 4.034, and 5.011, respectively.

All of the pH buffers and standard solutions of $Cd(ClO_4)_2$ were $H_2O$ solutions; however, all of the experimental samples used in this research were $D_2O$ solutions. Accordingly, the primary data from the Metrohm pH meter were corrected as needed. The actual pH ($pH = -\log_{10}\{D^+$ activity$\}$) is reportedly 0.4 units higher than the measured pH (Milne et al., 2001); so all tabulated pH values in this paper were obtained by adding 0.4 to the value displayed by the pH meter.

It was anticipated the response of the $Cd^{2+}$ ion selective electrode might be different in $H_2O$ and $D_2O$. Duplicate series of $Cd^{2+}$ standard solutions were thus prepared in these solvents. Plots of voltage versus log $[Cd_{total}]$ were linear in both solvents, and the slopes and intercepts were nearly identical. Assuming that the activity coefficient of $Cd^{2+}$ ion was not significantly different in $H_2O$ and $D_2O$, the pCd values of samples in $D_2O$ were thus estimated directly from calibration curves based on standard solutions in $H_2O$.

2.5.4.5. Preparation of samples: The samples that were actually used in the NMR titration experiments were prepared by diluting 5 mL of the NOM stock solution and 50-500 μL of the $^{113}Cd(ClO_4)_2$ stock solution with $D_2O$ to a final volume of 7 mL. The compositions of these samples varied slightly as new stock solutions of NOM were prepared, but typical compositions are given in Table 2.

**Table 2.** Typical initial compositions of solutions at the three Cd/C molar ratios which were used in this study.

| Cd/C | DOC, M | $Na^+$, M | $Cd^{2+}$, M | $ClO_4^-$, M | pH | pCd |
|---|---|---|---|---|---|---|
| 0.00056 | 0.744 | 0.0392 | $4.16\times10^{-4}$ | $8.32\times10^{-4}$ | 4.34 | 4.27 |
| 0.00112 | 0.744 | 0.0392 | $8.32\times10^{-4}$ | $1.66\times10^{-3}$ | 4.23 | 3.84 |
| 0.00559 | 0.744 | 0.0392 | $4.16\times10^{-3}$ | $8.32\times10^{-3}$ | 4.15 | 2.86 |

2.5.4.6. Overall strategy of titration/NMR experiments: NOM contains a wide variety of binding sites in which O-, N-, and S-donor atoms can bond directly to cations such as $Cd^{2+}$. Those binding sites also have an affinity for $H^+$, so both the strength of metal-NOM interactions and the nature of the most active binding sites in NOM are pH dependent. As the pH

and Cd/C ratio are systematically increased, $Cd^{2+}$ is expected to react with progressively more weakly acidic binding sites. Titration/NMR experiments were conducted at three different Cd/C ratios, two of which are well below the S/C ratio of the NOM, and at two different magnetic field strengths $B_0$ = 9.4 T and 11.7 T.

Each of the six titration/NMR experiments consisted of numerous repetitions of the following five steps:

1. Weigh the sample in a 30-mL polyethylene bottle.
2. Add an aliquot of 1.55 M $NaOD/D_2O$, and then re-weigh the sample in its bottle.
3. Measure pH and pCd, and then re-weigh the sample in its bottle.
4. Transfer an aliquot of the sample to an appropriate NMR tube (15-mm and 10-mm tubes, respectively, for the 400 MHz and 500 MHz NMR spectrometers). Measure the $^{113}$Cd NMR spectrum
5. Transfer the sample back to its bottle, and re-measure pH and pCd.

The same 30-mL polyethylene bottle and NMR tube were used throughout each titration/NMR experiment to minimize loss of the sample. There were small but unavoidable losses of sample each time the sample was transferred to/from the NMR tube and each time pH and pCd were measured. These losses were quantified from the mass of the sample before and after each transfer or measurement. The addition of the $NaOD/D_2O$ titrant also diluted the sample. All subsequent calculations of the chemical composition of each sample made appropriate corrections for dilution and for losses of sample.

During the course of this study, there were numerous minor variations from the procedure given here. Occasionally, stronger or weaker NaOD titrants were used to adjust pH. From time to time, pH and pCd were not measured. In some experiments, only the overall mass gain/loss for the experiment was measured, and that mass gain/loss was apportioned to each measurement step in the experiment to approximately correct the raw data.

### *2.5.5. Results and discussion*

2.5.5.1. Chemical Speciation of Cadmium: As described in the Experimental section of this paper, potentiometric measurements of pH and pCd were usually, but not always, made before and after each NMR experiment. All measured values of pCd and pH for experiments that were conducted at Cd/C ratios of 0.00056, 0.00112, and 0.00559 are given in Figure 17.

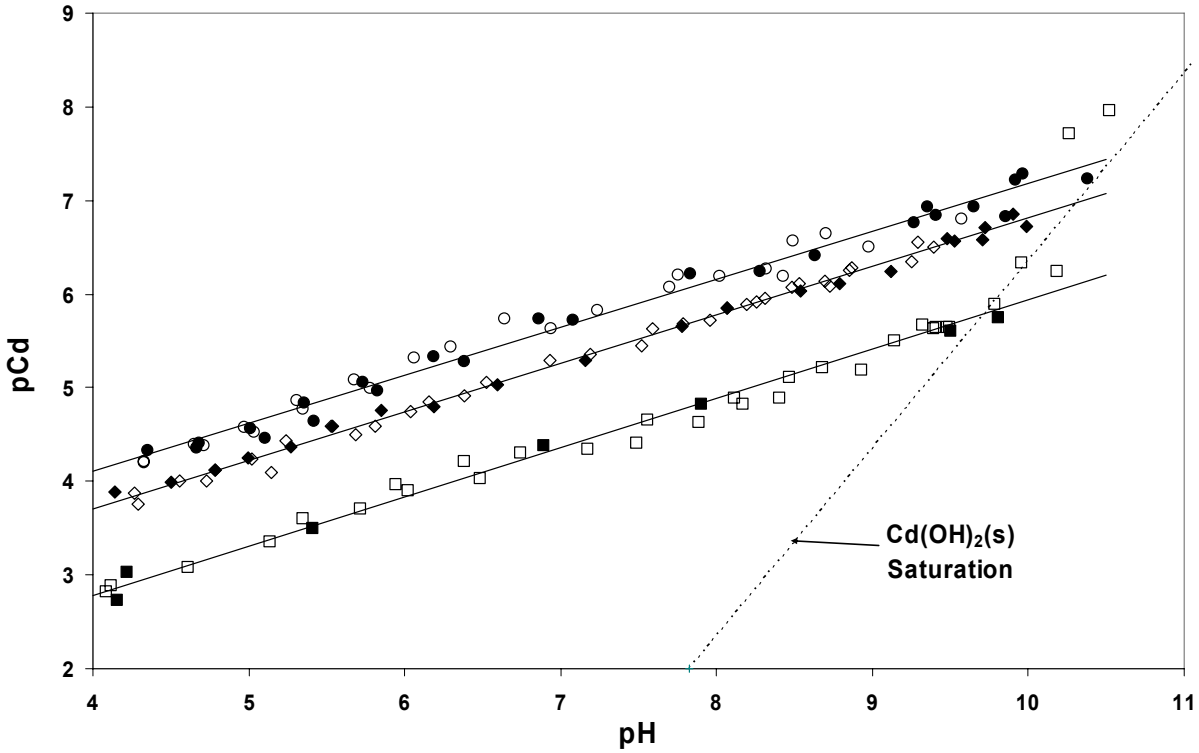



**Figure 17.** pCd versus pH at Cd/C molar ratios of 0.00056 (●, ○), 0.00112 (♦, ◊), and 0.00559 (■, □); solid/open symbols: measured before/after NMR experiments**.**

The trends are remarkably linear. In the experiments at a Cd/C ratio of 0.00559, the two observations at the highest pH values deviate markedly from the general trend, and those points were omitted from the linear regression analysis of the three groups of measurements, the results of which are summarized in Table 3.

**Table 3.** Linear regression analysis of pCd versus pH at Cd/C molar ratios of 0.00056, 0.00112, and 0.00559.

| Cd/C | Observations | Slope | Intercept | $R^2$ |
|---|---|---|---|---|
| 0.00056 | 48 | 0.513 + 0.009 | 2.05 + 0.07 | 0.985 |
| 0.00112 | 55 | 0.517 + 0.005 | 1.64 + 0.04 | 0.995 |
| 0.00559 | 37 | 0.527 + 0.011 | 0.67 + 0.09 | 0.985 |

It is unclear why the plots of pCd versus pH in Figure 17 all have a slope of about 0.5. Other studies of pH-dependent binding of $Cd^{2+}$ to complex natural materials have yielded comparable results. Several representative results are summarized in Table 4.

**Table 4.** Linear regression analysis of pCd versus pH for binding of $Cd^{2+}$ by several natural materials.

| Substrate | pH range | Slope | Intercept | $R^2$ | Reference |
|---|---|---|---|---|---|
| Aldrich humic acid | 4-9 | 0.652 | 2.74 | 0.988 | T4_1 |
| Peat humic acid | 4-10 | 0.705 | 1.44 | 0.998 | T4_2 |
| Forest soil humic acid | 4-8 | 0.590 | 2.03 | 0.999 | T4_3 |
| Bacterial cell walls | 5-8 | 0.367 | 1.62 | 0.976 | T4_4 |
| Soil organic matter | 3-10 | 0.477 | 1.08 | 0.927 | T4_5 |

T4_1: Xia and Raysom et al., 2002; T4_2: Kinniburgh et al., 1996; T4_3: Oste et al., 2002; T4_4: Plette et al., 1996; T4_5: Lee et al., 1996.

(Kinniburgh et al., 1999), state that the $H^+/Cd^{2+}$ exchange coefficient for binding of $Cd^{2+}$ to a substrate such as a humic substance can be obtained graphically from iso-pH plots of log (Bound Cd) versus pCd. The locus of [pCd, pH] points at some arbitrary degree of loading (Bound Cd), when plotted as pCd versus pH, are linearly correlated. This method was applied to the first four substrates in Table 4, at a constant loading of 0.06 mmol/g of bound Cd, to obtain the linear regression results in that Table. In the fifth study in Table 4, the authors normalized soil-water distribution coefficients ($K_d$) to the percent organic matter in soils to obtain $K_{om}$ values. Log $K_{om}$ values were, in turn, found to be linear functions of pH. The results in Table 4 for soil organic matter pertain to linear plots of log $K_{om}$ versus pH. It is readily shown that a plot of pCd versus pH should have the same slope but a different intercept. The results in Table 4 adequately demonstrate that the slopes of 0.513-0.527 that were obtained in the present study for plots of pCd versus pH are consistent with the pH-dependent binding of $Cd^{2+}$ to other substrates.

As is seen in Figure 17, the data for the three Cd/C ratios are displaced vertically from one another. At a given pH, the concentration of $Cd^{2+}$ increases (and pCd decreases) with increasing Cd/C ratio. From the intercepts in Table 3, it is readily calculated that the concentration of free $Cd^{2+}$ ion increases by about 1.4 orders of magnitude when the Cd/C ratio is increased by one order of magnitude. Not surprisingly, as the Cd/C ratio increases, $Cd^{2+}$ reacts with progressively weaker binding sites, leaving a larger fraction of the added $Cd^{2+}$ in the form of the uncomplexed $Cd^{2+}$ ion.

As a general rule, both pCd and pH decreased during the time required to acquire the NMR spectrum of a sample. The amount of drift (and the time required to acquire the NMR spectrum) increased with increasing pH. Even so, for a given Cd/C ratio, nearly all pCd and pH values plot on the same line, whether they were measured before or after the NMR experiment. Such non-equilibrium behavior is often encountered in studies of proton or metal binding by humic substances, and it has been attributed both to slow chemical reactions such as hydrolysis of esters, which generate new acidic sites in NOM, (Antweiler, 1991; Bowles et al., 1989), and to physical changes in the solution chemistry of NOM that do not actually generate new acidic sites but instead gradually expose existing sites to the bulk solution through pH-induced changes in the average conformation or degree of association of NOM molecules, (Varney et al., 1983).

The dashed line in Figure 17, which has a slope of +2, is the saturation boundary for $Cd(OH)_2(s)$, based on its solubility in $H_2O$. Data points lying on or to the right of this line are either saturated or supersaturated with respect to this solid phase. It appears that all data points at pH 10 or higher lie close to this saturation boundary. It is therefore possible that a portion of the $Cd^{2+}$ in these solutions is in the form of $Cd(OH)_2(s)$, and this might explain why the two data points at the highest pH values deviate so strongly from the generally linear trends of the three data sets.

Alternatively, during the five-step experimental protocol that was described earlier in this paper, samples were repeatedly exposed to the atmosphere. If sufficient $CO_2(g)$ were to dissolve in the samples, $CdCO_3(s)$ might be expected to form, thus compromising the NMR experiments. It is also possible to introduce $CO_2$ during addition of the base titrant, if the titrant contains a significant concentration of $CO_2$. This is actually the most likely explanation for the deviation of the last two data points of the data series with Cd/C = 0.00559 from the general trend. The pH of those solutions was adjusted by addition of a $NaOD/D_2O$ solution

that was 100-fold stronger than the normal NaOD titrant. That more concentrated NaOD titrant might have been contaminated with $CO_2$. Whatever the case, if $CdCO_3(s)$ is formed (e.g., by the reaction: $Cd^{2+} + H_2O + CO_2 \rightarrow CdCO_3(s) + 2\ H^+$), then the relationship between pCd and pH becomes:

$$pCd = (\text{Log } K + \text{Log } (P_{CO2})) + 2\ pH \quad (7)$$

Although the Log K for this reaction in $D_2O$ is not known and the $P_{CO2}$ at any point in these experiments is unknown, it is clear that the slope of the relationship between pCd and pH must be +2. With the exception of the data points at pH of 10 or higher, all data points lie along trend lines with slopes that vary from 0.513 to 0.527, so it is reasonable to conclude that, in general, $CdCO_3(s)$ and $Cd(OH)_2(s)$ are not formed in these experiments.
Having ruled out the presence of $CdCO_3(s)$ and $Cd(OH)_2(s)$ in nearly all of the solutions, the mass balance equation at any point in a titration experiment is:

$$Cd_{total} = [Cd^{2+}] + \Sigma_i[Cd(OH)_i] + [\text{Cd-NOM}] \quad (8)$$

where $\Sigma_i[Cd(OH)_i]$ is the sum of the concentrations of all hydroxy complexes of $Cd^{2+}$, and [Cd-NOM] is the concentration of organically complexed $Cd^{2+}$. Both $[Cd^{2+}]$ and $\Sigma_i[Cd(OH)_i]$ are readily calculated from pCd and pH at a given ionic strength, using the Davies equation to obtain activity coefficients. Equilibrium constants for the hydrolysis of $Cd^{2+}$ in $H_2O$ were used, because the corresponding equilibrium constants in $D_2O$ were unavailable. The error thus introduced is unknown, but probably relatively small. Organically complexed $Cd^{2+}$ is obtained by difference:

$$[\text{Cd-NOM}] = Cd_{total} - [Cd^{2+}] - \Sigma_i[Cd(OH)_i] \quad (9)$$

Figure 18 shows the distribution of $Cd^{2+}$ among these three forms as a function of pH for all of the NMR experiments. A few clearly anomalous points occur at the low- and high-pH ends of this plot. It is also evident, as was the case in Figure 17, that the relative concentration of free $Cd^{2+}$ at a given pH decreases with decreasing Cd/C ratio, as do the concentrations of the

corresponding hydroxy complexes of $Cd^{2+}$. Most importantly, nearly all $Cd^{2+}$ is organically complexed above pH 5.0, and hydroxy complexes are never major species. So, for most of the experimental conditions used in this study, the observed $^{113}$Cd NMR spectra are almost entirely attributable to organically complexed forms of $Cd^{2+}$. At pH > 5, any variation in spectra with experimental conditions must be attributed to internal changes in the distribution and/or rate of chemical exchange of $Cd^{2+}$ among organically complexed forms.

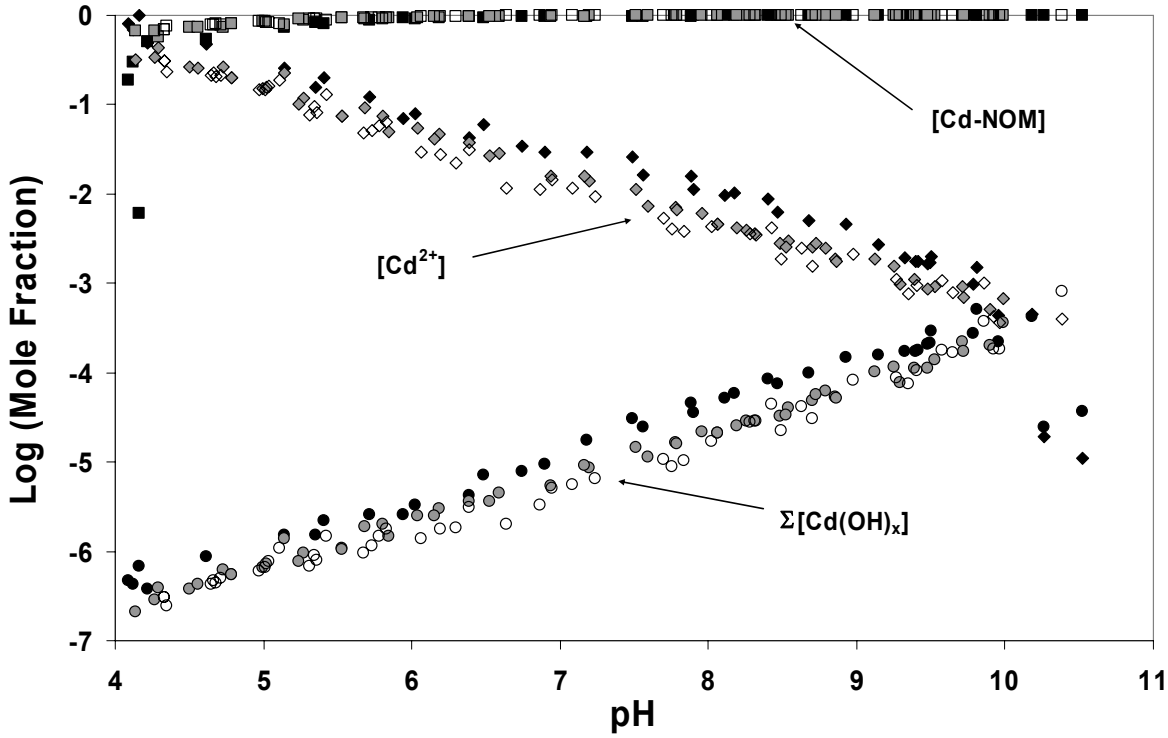



**Figure 18.** Chemical speciation of $Cd^{2+}$ versus pH at Cd/C molar ratios of 0.00056 (white), 0.00112 (grey), and 0.00559 (black); all results of measurements taken before/after NMR experiments are included.

2.5.5.2. Overview of $^{113}$Cd NMR Spectra: A total of six solutions of $Cd(ClO_4)_2$ and Suwannee River NOM, having Cd/C molar ratios of 0.00559, 0.00111 and 0.00056, were titrated from approximately pH 4 to pH 10 by stepwise addition of aliquots of $NaOD/D_2O$. At each Cd/C ratio, $^{113}$Cd NMR spectra were acquired at a magnetic field strength of 9.4 T for one sample and at a magnetic field strength of 11.7 T for the other sample. All spectra were acquired at a constant temperature (303 K). Altogether, more than 100 spectra were acquired

and analyzed. In addition, a large number of ancillary spectra were acquired using specialized NMR parameters to test a variety of secondary questions regarding experimental protocol, temporal variations of solution chemistry, etc.

Figure 19 contains a representative subset of the $^{113}$Cd NMR spectra from this study. These spectra were chosen to illustrate the effects of pH, Cd/C, and $B_0$ on chemical shift, line width, and line shape in the $^{113}$Cd NMR spectra of the solutions used in this study.

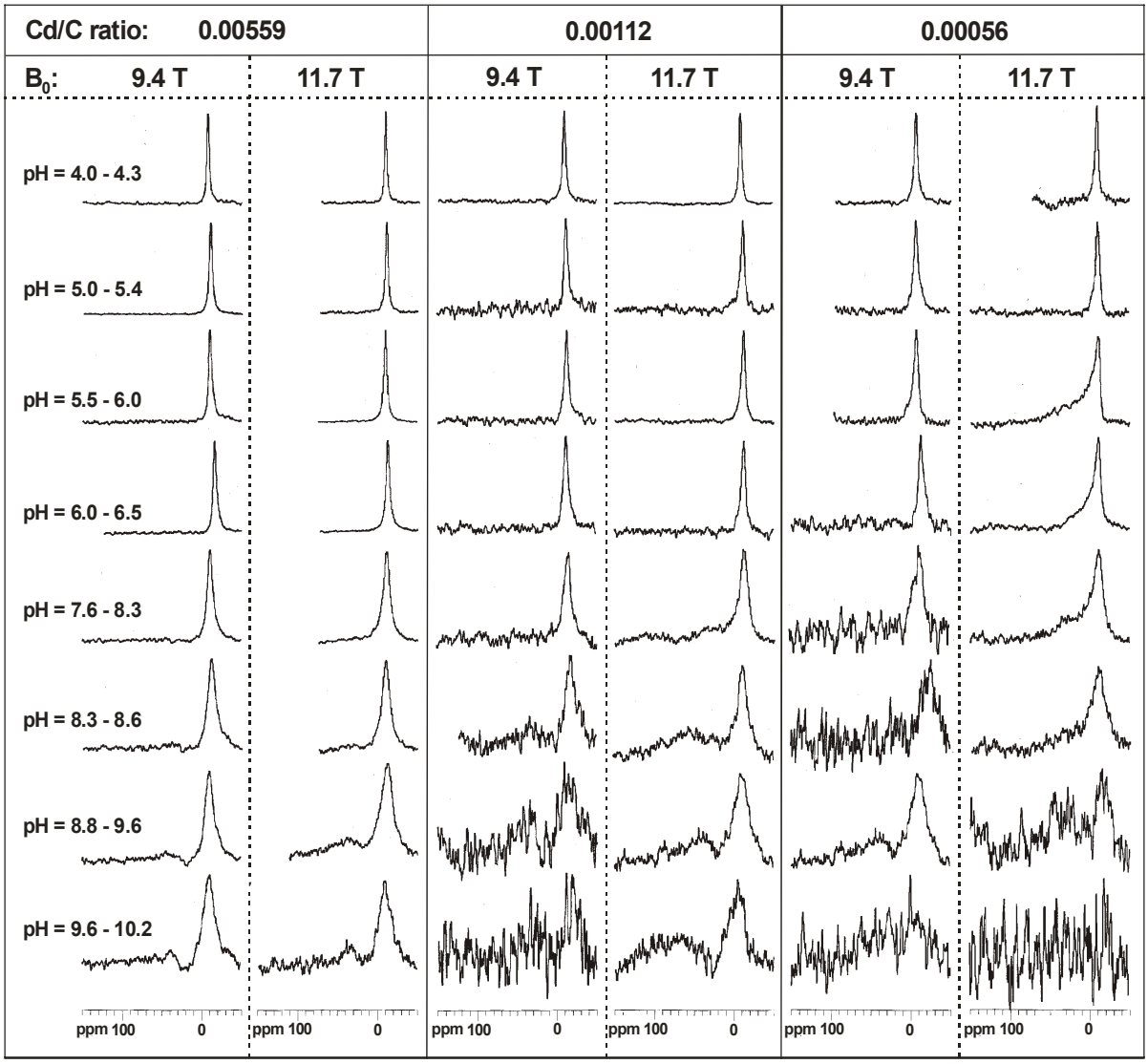



**Figure 19.** Selected $^{113}$Cd NMR spectra at 303K of $^{113}Cd(ClO_4)_2$ bound to Suwannee River organic matter at various Cd/C ratios, $^{113}$Cd NMR frequencies, and pH. All spectra are calculated with exponential line broadening (lb = 50 Hz), but are not normalized with respect to acquisition time.

A cursory examination of these spectra reveals several general trends:

- the line widths of the Lorentzian peaks that are observed at low pH, which are already broadened by the effects of coordination and chemical exchange, increase steadily as pH is increased.
- the observed chemical shifts of the Lorentzian peaks move initially upfield with increasing pH, reach a maximum upfield shift at circumneutral pH, and then migrate downfield at even higher pH.
- line shapes gradually deviate from Lorentzian shape as pH is increased, leading ultimately to the breakup of the Lorentzian peaks into more complex spectra.
- superimposed on these general trends are second-order effects that arise due to Cd/C ratio and $B_0$, all of which provide additional insight into the interaction of $Cd^{2+}$ with binding sites of NOM.

2.5.5.3 pH Dependence of $^{113}$Cd NMR Spectra: A more complete series of $^{113}$Cd NMR spectra, at constant Cd/C ratio (0.0056) and $B_0$ = 11.7 T, is shown in Figure 20.

With the exception of the uppermost two spectra, the numeric labels on the spectra are average pH values, calculated from measured pH values before and after acquisition of a spectrum. These spectra, which were calculated using a line broadening of 50 Hz, are fully consistent with the above general trends. For comparison, both the experimental spectra and the Lorenztian fits, from which average chemical shifts and linewidths were obtained, are shown in Figure 20. Below pH 8, where fast chemical exchange is more important than intermediate or slow chemical exchange, the $^{113}$Cd NMR spectra nicely indicate the steadily growing linewidth of the single Lorentzian resonances and the gradual alteration of their average $^{113}$Cd NMR chemical shift. The tiny computed resonances in the Lorentzian fits are most likely not significant in this pH range. Above about pH 8, the signal-to-noise ratios of the $^{113}$Cd NMR spectra deteriorate rapidly, as a larger fraction of the chemical exchange becomes dominated by intermediate and slow exchange. Above pH 9, multiple resonances are always observed, being especially evident in the fitted spectra. The breakup of the single Lorentzian peak signifies a further shift into the regime that is dominated by intermediate and slow chemical exchange. The initial breakup of the Lorentzian peak reveals a variety of oxygen-coordinated Cd species. At higher pH, however, an ever increasing fraction of $Cd^{2+}$ is coordinated to nitrogen.

The spectra at pH 10.3 and pH 9.0 are actually sub-spectra collected during the first and second halves of the total acquisition time on a single sample whose average pH is 9.65, and the two pH values are the pH values before and after acquisition of the NMR spectrum. These sub-spectra are included in Figure 20 to illustrate the effects of drifting pH on the general appearance of $^{113}$Cd NMR spectra – a phenomenon that will be examined more completely in the next section. It is especially noteworthy that the single broad upfield Lorentzian peak collapses into several sharper peaks, as is commonly observed when a system undergoing rapid/intermediate chemical exchange enters the regime of intermediate/slow rates of chemical exchange.

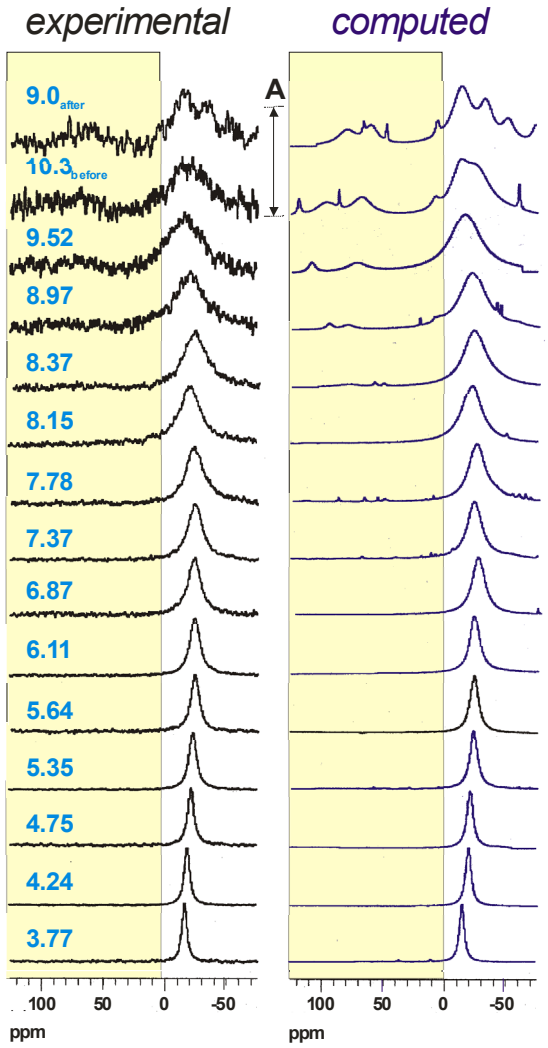



**Figure 20**. Experimental (left: exponential line broadening = 50 Hz) and computed (right: azfw = 200 ppm; pc = 0.8) $^{113}$Cd NMR spectra of $^{113}Cd(ClO_4)_2$ bound to Suwannee River organic matter at a Cd/C ratio = 0.0056 and $^{113}$Cd NMR frequency: 110.95 MHz ($B_0$ = 11.7 T). The numbers given correspond to the average pH value {[pH(before) + pH(after)]/2} during the NMR acquisition. The two top spectra (series A) are two subspectra (1$^{st}$ and 2$^{nd}$ half of total acquisition time), corresponding to pH(before) = 10.3 and pH(after) = 9.0 to illustrate the effect of pH drift on the $^{113}$Cd NMR spectra (cf. Text).

2.5.5.4. The Effect of Drifting pCd and pH on $^{113}$Cd NMR Spectra: As mentioned in the earlier discussion of the chemical speciation of $Cd^{2+}$, both pCd and pH tended to drift to lower values during the time required to acquire an NMR spectrum, the drift being most noticeable at high pH. At high pH, the loss of Lorentzian line shapes was accompanied by decreased signal-to-noise ratio in the spectra (see Figure 19), thus requiring many more scans to acquire an acceptable NMR spectrum. It is clear from the discussion of Figure 17 and Figure 18 that changes in pCd and pH modify the chemical speciation of $Cd^{2+}$, and those changes were undoubtedly occurring as the NMR spectra were being acquired. In the regions of slow and intermediate chemical exchange, subtle variations of pH impose strong effects on the $^{113}$Cd NMR spectra, as already seen in Figure 20. Further evidence of the effects of pH drift and line broadening on the appearance of $^{113}$Cd NMR spectra are given in Figure 21, which includes both experimental spectra and Lorentzian fits.

The lowest spectrum in Figure 21 was obtained under the same conditions used for the spectra in Figure 20 (line broadening of 50 Hz). The immediately overlying spectrum differs only in the use of a line broadening of 250 Hz and a much wider spectral window. The effect is not surprising – the signal-to-noise ratio is improved by increased line broadening at the expense of resolution. Indeed, the complex, broad upfield resonances collapse into two Lorentzian peaks.

Figure 21 also includes three series of sub-spectra that were collected at slightly different initial pH values and over different total acquisition times. Series A consists of three sub-spectra that were collected over equal fractions of the total acquisition time, during which pH drifted from 9.6 to 9.0. Series B consists of five sub-spectra that were collected over equal fractions of the total acquisition time, during which pH drifted from 9.7 to 9.4. Series C consists of two sub-spectra that were collected over equal fractions of the total acquisition time, during which pH drifted from 10.3 to 9.8. These three series of sub-spectra lead to the important observation that $^{113}$Cd NMR spectra which were acquired over extended times will therefore represent an average, but changing composition of Cd-containing species throughout the total acquisition time, especially under conditions of intermediate or slow exchange. The combined effects of the rather low sensitivity of $^{113}$Cd NMR spectroscopy, extensive resonance broadening, increased $T_1$ relaxation times at higher pH and the unavoidable alteration of solution chemistry during NMR acquisition time impose restrictions on the

"chemical" resolution of cadmium coordination to humic substances that can be obtained by NMR spectroscopy.

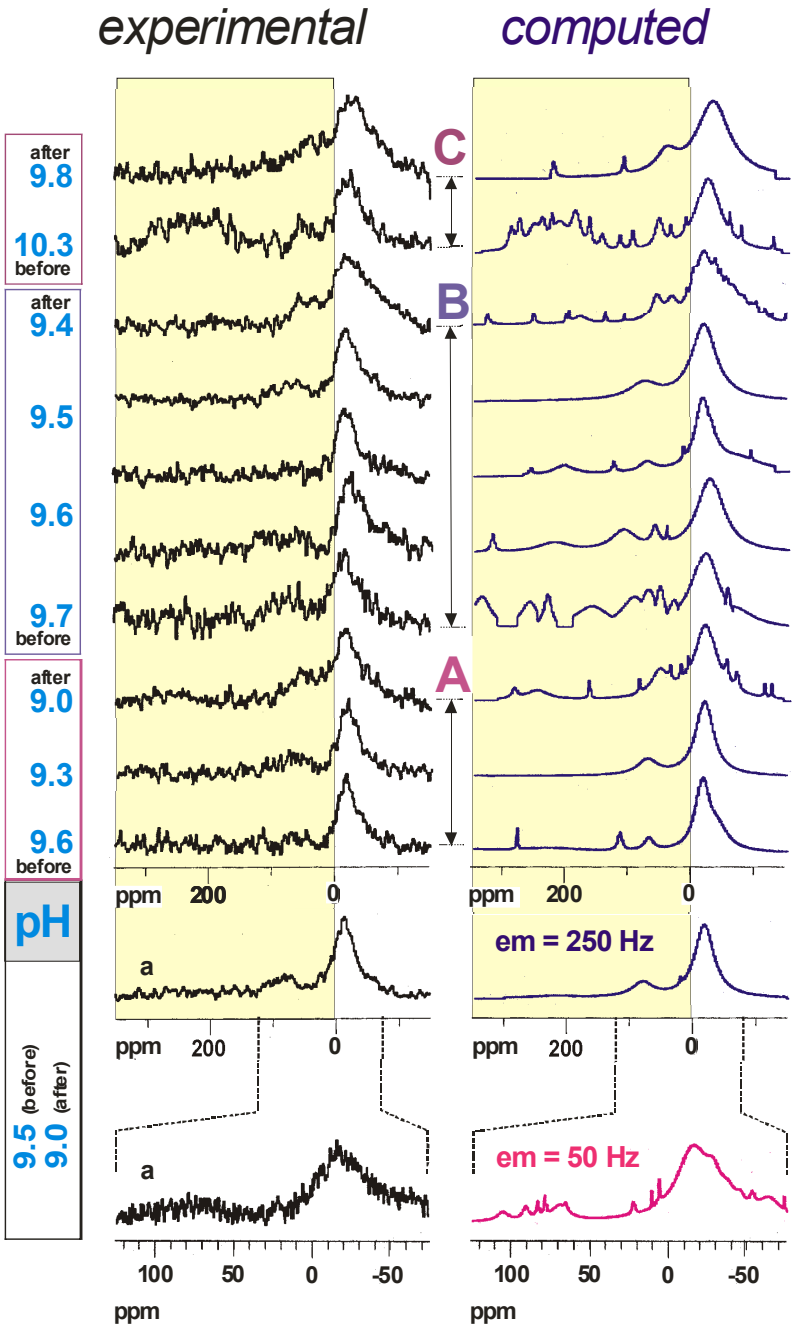



**Figure 21.** Experimental (left: exponential line broadening: 250 Hz) and computed (right: azfw = 200 ppm; pc = 0.8) $^{113}$Cd NMR spectra of $^{113}Cd(ClO_4)_2$ bound to Suwannee River organic matter at a Cd/C ratio = 0.0056 and $^{113}$Cd NMR frequency: 110.95 MHz ($B_0$ = 11.7 T) shown at various stages of NMR acquisition, in the course of which alterations of pH occurred (cf. Text); left column indicates initial, final and interpolated pH during NMR experiments; *bottom:* $^{113}$Cd NMR spectrum with scale, exponential broadening (50Hz) and computation analogous to Figure 20, *2$^{nd}$ from bottom:* the same spectrum calculated with a 250 Hz exponential line broadening; letters A,B,C denote series of NMR spectra acquired without sample change between measurements of individual NMR spectra (acquisition order of all NMR spectra shown: from bottom to top).

2.5.5.5. Comparison of Observed $^{113}$Cd Chemical Shifts with Literature Values: With the exception of the work of Li et al., 1998; Lowe, 1992, which laid the groundwork for the present study, all published results of $^{113}$Cd NMR studies of $Cd^{2+}$ binding by NOM and related materials (Lowe, 1992; Otto et al., 2001; Otto et al., 2001; Larive et al., 1996; Chung et al., 1996; Kulikova and Perminova, 2002) were obtained using solutions in which pH was held constant at pH 6.0-6.4 and Cd/C ratios were varied over a wide range (collectively, from 0.0021 to 0.0960). The results of the six titration series in this study are compared in Table 5 with published results that were obtained under comparable conditions.

**Table 5.** Comparison of observed $^{113}$Cd chemical shifts in the presence of NOM, soil fulvic acids, and aquatic fulvic acids at pH 6.0-6.4 and at Cd/C molar ratios of 0.0011 to 0.011.

| Cd/C Range | pH Range | $\delta_{obs}$ Range | Reference |
|---|---|---|---|
| 0.0021 – 0.0070 | 6.40 | -11.9 to –7.4 | T5_1 |
| 0.0041 – 0.0061 | 6.00 | -15.6 to –20.4 | T5_2 |
| 0.0068 | 6.40 | -22.6 | T5_3 |
| 0.0109 | 6.40 | -19.8 | T5_4 |
| 0.0048 – 0.0096 | 6.40 | -21.6 to –20.3 | T5_5 |
| 0.0011 – 0.0056 | 6.40 – 6.50 | -21.8 to –24.1 | This study |

T5_1: Larive et al., 1996; T5_2: Chung et al., 1996; Lowe et al., 1992; T5_4: Otto et al., 2001a; T5_5: Otto et al., 2001b;

As discussed by Li et al. (1998), the $\delta_{obs}$ values reported by Larive et al., 1996 were affected by the presence of chloro complexes of $Cd^{2+}$ in their samples. With the exception of that data set, the results of this study are consistent with published $\delta_{obs}$ values that were obtained under comparable conditions.

Figure 22 shows the variation of observed chemical shift ($\delta_{obs}$) of $^{113}$Cd with pH at the three Cd/C ratios and at the two field strengths.

Significantly, different chemical shifts were obtained at different fields strengths for solutions with the same pH and Cd/C ratio. This dependence of $\delta_{obs}$ on field strength results from some

proportion of the chemical exchange occurring at rates that are intermediate or slow on the NMR time scale, even at the lowest pH range of these experiments, where the most narrow Lorentzian lines were observed. Although there is clearly some scatter in the data, the trend is observed at all three Cd/C ratios. Average chemical shifts are used in Figure 22 only within the pH range of Lorentzian peaks, and this pH range increases dramatically with increasing Cd/C ratio (see later discussion).

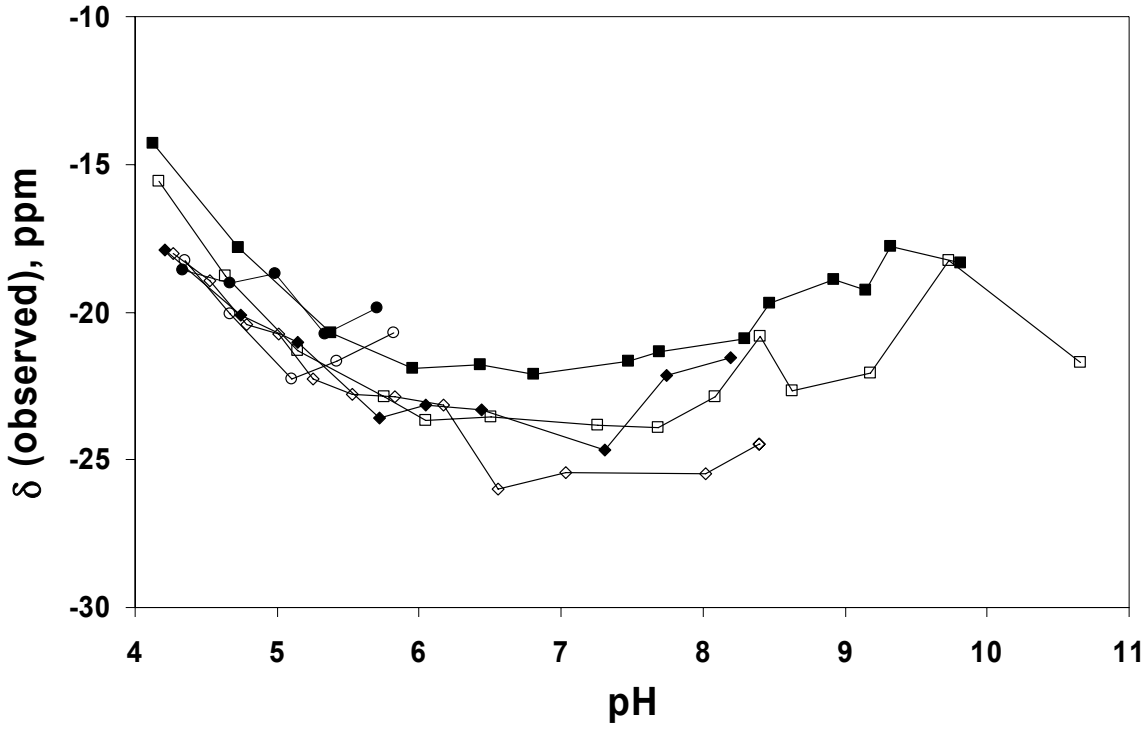



**Figure 22.** Observed $^{113}$Cd chemical shifts versus average pH (before/after) at Cd/C molar ratios of 0.00056 (●, ○), 0.00112 (♦, ◊), and 0.00559 (■, □); solid/open symbols: measured at magnetic field strengths of 9.4 T and 11.7 T, respectively**.**

2.5.4.6. The Chemical Shift of Organically Bound $Cd^{2+}$: As stated earlier in the discussion of the chemical speciation of $Cd^{2+}$, free $Cd^{2+}$ and organically bound $Cd^{2+}$ are the dominant forms of $Cd^{2+}$ in nearly all the solutions that were studied. Neglecting the contribution of hydroxy complexes of Cd(II), the observed chemical shift ($\delta_{obs}$) under conditions of fast exchange (see Eq. 3) is approximately given by:

$$\delta_{obs} = \delta_{Cd^{2+}}\left(\frac{[Cd^{2+}]}{Cd_{Total}}\right) + \delta_{org}\sum_i\left(\frac{[CdL_i]}{Cd_{Total}}\right) \tag{10}$$

Given that the chemical shift of the free $Cd^{2+}$ ion is zero ppm, by definition, it follows that the average chemical shift of organically complexed Cd(II) is given by:

$$\delta_{org} = \frac{\delta_{obs}}{\sum_i\left(\frac{[CdL_i]}{Cd_{Total}}\right)} \tag{11}$$

Equation 11 has been used to calculate $\delta_{org}$ from $\delta_{obs}$ and the known concentration of organically bound $Cd^{2+}$ in each solution. It is assumed in these calculations that $\delta_{Cd} = 0$ ppm. The results obtained at 11.7 T and at a Cd/C ratio of 0.00559 are given in Figure 23, in which both $\delta_{obs}$ and $\delta_{org}$ are shown over the pH range of about pH 4 to pH 10. At acidic pH values, where $Cd^{2+}$ is a significant fraction of total Cd, $\delta_{org}$ values are –38 to –39 ppm, which is considerably farther upfield than the corresponding $\delta_{obs}$ values shown in Figure 22. As pH is increased into the pH 6-8 range, $\delta_{org}$ values are in the range of –24 to –26 ppm, only 1-2 ppm upfield from $\delta_{obs}$ values. As pH is increased beyond pH 8, virtually all $Cd^{2+}$ is organically complexed, and $\delta_{obs}$ and $\delta_{org}$ are essentially equal. Under such conditions, this analysis, which is based on the assumption of fast exchange, may no longer be appropriate.

It is tempting to interpret the results in Figure 23 in terms of the results of Chung et al., 1996, who reported that 1:1 and 1:2 complexes of $Cd^{2+}$ with a few simple monocarboxylic acids have chemical shifts of –22 to –24 ppm and –39 to –40 ppm, respectively. Such a comparison is not valid, however, because their analysis, restated here in terms of the parameters used in this paper, is based on the variation of $\delta_{obs}$ with Cd/C ratio at constant pH in solutions of pure ligands. The results in Figure 23 depict the variation in $\delta_{obs}$ and $\delta_{org}$ with pH at constant Cd/C ratio in solutions containing a highly complex mixture of binding sites.

Unlike the simple carboxylic acids studied by Chung et al., Wu et al., 2004, NOM is a highly complex mixture that contains not only carboxyl groups but also amino and sulfhydryl groups that can potentially bind $Cd^{2+}$. Recalling that $\delta_{org}$ itself is a weighted-average chemical shift for organically bound $Cd^{2+}$, the downfield shift in $\delta_{org}$ as pH is increased from pH 4 to pH 8 can readily be explained by increased participation of a small proportion of amino ligands in the fast chemical exchange of $Cd^{2+}$ (Wang and Gilpin, 1983; Jacobsen and Ellis, 1981).

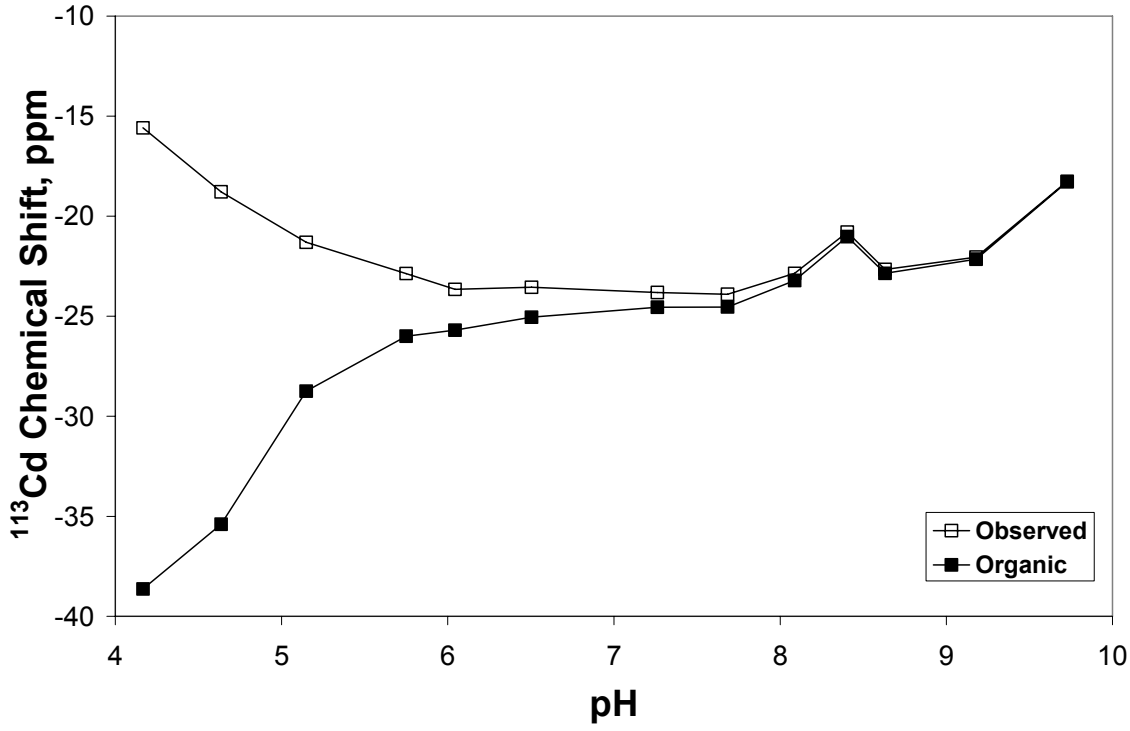



**Figure 23.** Comparison of the observed $^{113}$Cd chemical shift and the calculated chemical shift of organically complexed $Cd^{2+}$ versus average pH (before/after) at a magnetic field strength of 11.7 T and a Cd/C molar ratio of 0.00559.

Comparable, but less well-defined, trends are found in the other five titration series (not shown in Figure 23). For solutions with Cd/C ratios of 0.00056 and 0.00112, extrapolated $\delta_{org}$ values at pH 4 are approximately –26 and -28 ppm, respectively. So, at lower Cd/C ratios, $\delta_{org}$ lies farther downfield in the direction of the resonance frequency range of nitrogen-coordinated $Cd^{2+}$. This observation is consistent with the hypothesis that amino ligands are parti-

cipating in the fast chemical exchange of $Cd^{2+}$, and it indicates further that those ligands have a rather high affinity for $Cd^{2+}$ - high enough to compete effectively at low pH with the much more abundant carboxyl groups responsible for complexation of $Cd^{2+}$.

2.5.5.7. The Variation of Line Width with Chemical Speciation of $Cd^{2+}$: Line widths ($\Delta\nu_L$) of the Lorentzian peaks in the six titration series are plotted as a function of pH in Figure 24.

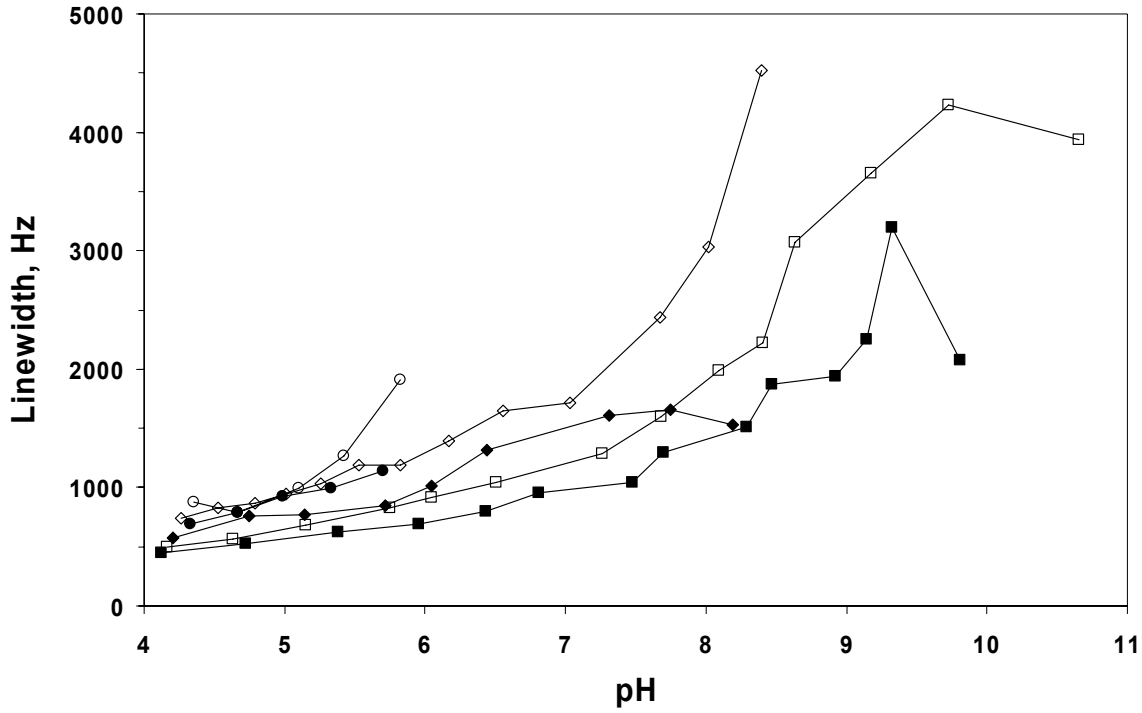



**Figure 24.** Line widths (in Hz) of Lorentzian peaks in the $^{113}$Cd NMR spectra versus average pH (before/after) at Cd/C molar ratios of 0.00056 (●, ○), 0.00112 (♦, ◊), and 0.00559 (■, □); solid/open symbols: measured at magnetic field strengths of 9.4 T and 11.7 T, respectively**.**

All titration series exhibit the same general behavior – a gradually accelerating increase in $\Delta\nu_L$ with increasing pH until the spectra become obviously non-Lorentzian. That transition occurs at higher pH in solutions with higher Cd/C ratios. At any given pH, the line width is less at $B_0$ = 9.4 T than at $B_0$ = 11.7 T. Additionally, at any given pH, $\Delta\nu_L$ values increase with decreasing Cd/C ratio. All of these trends find their origin in the complex interplay between

the chemistry of the solutions and the magnetic field strengths at which this chemistry is observed.

Consider the effect of increased pH on $\Delta\nu_L$ for one of the titration series in Figure 24, e.g., the solution with a Cd/C ratio of 0.00559 whose $^{113}$Cd NMR spectra were acquired at $B_0$ = 9.4 T. As pH is increased, the concentration of $Cd^{2+}$ ion decreases (see Figure 17 and Figure 18). The lowered concentration of $Cd^{2+}$ should decrease the rates of bimolecular chemical exchange reactions, and slower chemical exchange will cause line broadening. The other effect is more subtle, but nonetheless significant. As pH is increased, a larger fraction of the heterogeneous distribution of binding sites forms complexes with $Cd^{2+}$. It is reasonable to expect a broader distribution of complexes to have a wider range of NMR resonance frequencies for bound Cd. For example, as discussed in the previous section, the formation of an increasing proportion of amino complexes with increasing pH is consistent with trends in $\delta_{org}$ with increasing pH. The participation of complexes with substantially different resonance frequencies in chemical exchange also results in line broadening. It is believed that the gradually accelerating increase in $\Delta\nu_L$ with increased pH is due to a gradual transition from a low-pH regime where line broadening is caused mainly by decreased chemical exchange rates to a higher-pH regime where line broadening is caused mainly by the more heterogeneous population of complexes contributing to chemical exchange.

These same factors contribute to the other five curves in Figure 24. If the curves for the two titration series with Cd/C ratios of 0.00559 are compared, it is evident that $\Delta\nu_L$ is less at all pH for spectra that were acquired at $B_0$ = 9.4 T. In the pH range of 4.0-8.5, a well-defined Lorentzian peak was observed in each data series, and six pairs of observations lie within 0.1 pH units of one another. For those pairs of observations, the average ratio of $\Delta\nu_L$ for spectra at 11.7 T and 9.4 T is 1.20 ± 0.10. Similar observations of paired data for the data sets with Cd/C ratios of 0.0011 and 0.00056 yielded $\Delta\nu_L$ ratios of 1.31 ± 0.11 and 1.26 ± 0.26, respectively. At the three Cd/C ratios, the ratios of $\Delta\nu_L$ at 11.7 T and 9.4 T are not statistically different from the ratio of field strengths.

It is also evident in Figure 24 that $\Delta\nu_L$ values increase with decreasing Cd/C ratio. As stated earlier in the analysis of the effects of increased pH on $\Delta\nu_L$, two general effects – slower chemical exchange at lower [$Cd^{2+}$] and increased heterogeneity of exchanging species at higher pH (but constant Cd/C ratio) – are believed to be principally responsible for the

observed line broadening. At any given pH, as the Cd/C ratio is decreased, the concentration of free $Cd^{2+}$ decreases. The result is slower chemical exchange and the resultant line broadening.

The data in Figure 24 show clearly that the rate of increase in $\Delta\nu_L$ with increasing pH is much greater for solutions having low Cd/C ratios. The transition to non-Lorentzian peak shapes also occurs at much lower pH in these solutions. For example, in the $^{113}$Cd NMR spectra which were obtained at $B_0$ = 11.7 T for the titration series with Cd/C = 0.00056, $\Delta\nu_L$ increases from about 1000 Hz to about 2000 Hz between pH 5 and pH 6. Comparable increases in $\Delta\nu_L$ are reached at pH 7.3 and pH 8.1 in the titration series with Cd/C ratios of 0.00112 and 0.00559, respectively. These observations suggest that the population of complexes contributing to chemical exchange is more heterogeneous in solutions with lower Cd/C ratios, which at first seems inconsistent with the fact that solutions with lower Cd/C ratios actually occupy a much smaller fraction of binding sites of NOM at a given pH. If a small fraction of the total population of binding sites consists of strongly coordinating amino ligands, their proportional contribution to the exchange process might be much greater at the lowest overall loading of $Cd^{2+}$, i.e., in solutions having a lower Cd/C ratio. This is believed to be the case, and this interpretation is supported by the gradual downfield displacement of $\delta_{org}$ values with decreasing Cd/C ratio at any given acidic pH (see earlier discussion of the pH dependence of $\delta_{org}$ values).

An instructive way to visualize the mutual dependencies of pH, pCd and $\Delta\nu_L$ in the regime of Lorentzian line shape is to indicate $\Delta\nu_L$ using symbols whose size is scaled to $\Delta\nu_L$ in a plot of pH against pCd (Figure 25).

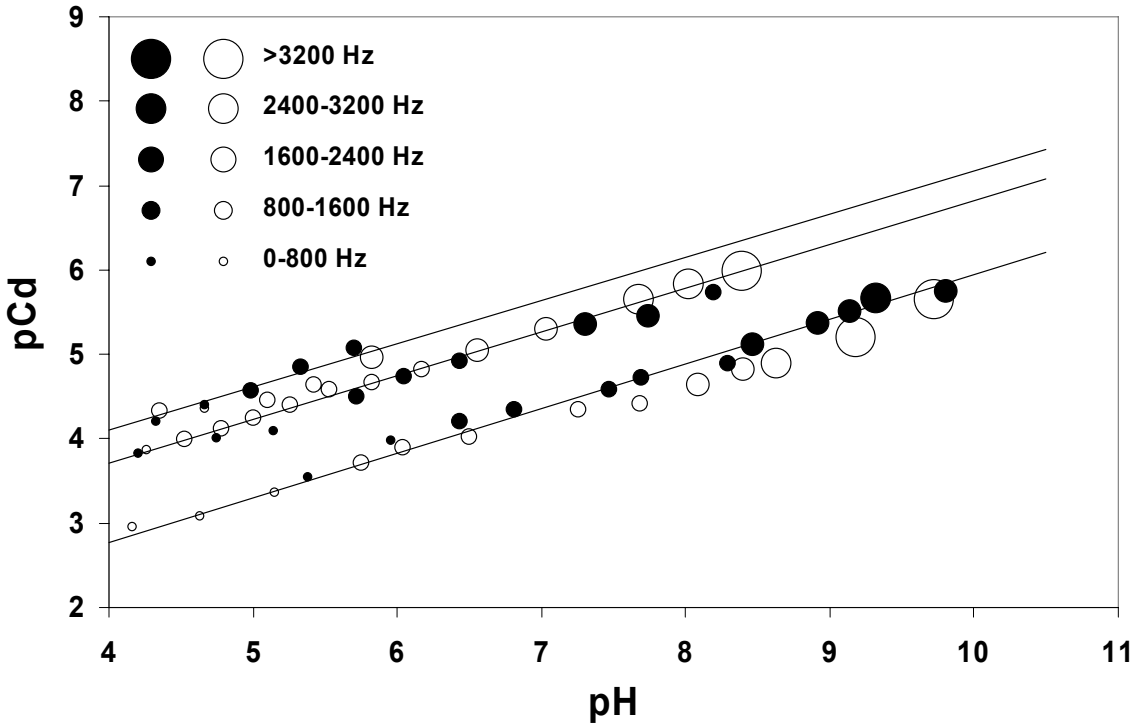



**Figure 25.** A subset of data from Figure 14, for which the $^{113}$Cd NMR spectra are adequately described by a single Lorentzian line, used here to illustrate the variations in line width (in Hz) with average pH and pCd (before/after) at magnetic field strengths of 9.4 T (solid circles) and 11.7 T (open circles), including measurements taken at Cd/C molar ratios of 0.00056 (top line), 0.00112 (middle line), and 0.00559 (bottom line).

By comparison with Figure 17, for which plots of pCd versus pH extended from pH 4 to pH 10 at all Cd/C ratios, the data series in Figure 25 all start at pH 4, but they end wherever Lorentzian line shapes were no longer observed. This transition occurred at pH values of about 6, 8.5, and 10 for the titration series with Cd/C ratios of 0.00056, 0.00112, and 0.00559, respectively. Along the regression line of each titration series, it is also evident that, at a given pCd and pH, line widths are larger at $B_0$ = 11.7 T than at $B_0$ = 9.4 T. When changing $B_0$ from 9.4 to 11.7 T, the range of comparable $\Delta\nu_L$ values is displaced to lower values of pH and pCd, and the highest measured $\Delta\nu_L$ values are only observed at $B_0$ = 11.7 T (Figure 25). The sharp increase of $\Delta\nu_L$ at $B_0$ = 11.7 T observed at higher pH indicates cadmium binding of a large variety of ligands differing widely in their chemical shift.

2.5.5.8. Non-Lorentzian Peak Shapes: At pH values in the range of 6-9, an asymmetric distortion of the Lorentzian line shape occurred, the details of which depended strongly on the

Cd/C ratio and $B_0$. The curve-fitting Bruker software resolved the experimental $^{113}$Cd NMR spectra into multiple Lorentzian peaks (cf. Figure 20 and Figure 21). The line width of the main resonance apparently decreased, in conjunction with the appearance of several other resolved, but overlapping resonances, each of which presumably includes contributions from a group of complexes undergoing slow and intermediate chemical exchange, because $\delta_{obs}$ values of the series differing in Cd/C ratio and $B_0$ never precisely correspond to each other (see Figure 22). As pH was increased further, a transient upfield shoulder split off the main peak to form distinct resonances in the range of -40 to -80 ppm, leaving the main resonance with a reduced line width centered about -25 ppm. The well-resolved upfield resonances are an indication of an increased proportion of slow exchange, even between oxygen species separated by less than 20-40 ppm. With increasing pH, the superimposed resonances underwent several episodes of further broadening and breakup, producing combinations of resonances with more symmetrical lineshapes. Further increases in pH led to the emergence of low-field resonances in the range of 0-200 ppm, providing clear, unequivocal evidence of an ever-growing proportion of nitrogen-coordinated $Cd^{2+}$.

2.5.5.9. Coordination of cadmium to sulfur: In an earlier study, (Li et al., 1998, Lowe, 1992), reported that they observed no sulfur-coordinated $Cd^{2+}$, even at high pH and low Cd/C ratio. The very wide resonance frequency range of sulfur-coordinated $Cd^{2+}$ in a system dominated by chemical exchange and heterogeneity of binding sites, together with the relatively low abundance of reduced sulfur in this NOM sample, virtually guarantee that the signal-to-noise ratio will be unfavorable in the frequency range where sulfur-bound $Cd^{2+}$ should be observed. It is expected that occurrence and disappearance of these resonances will be strongly affected by subtle changes in the relative ratio of these coordination compounds and their exchange characteristics in relation to their NMR spectral parameters.

In the present study, in which two of the Cd/C ratios were lower than any used by Li et al., 1998, reproducible broadened $^{113}$Cd NMR resonances in the range of 300 - 800 ppm, which require at least one sulfur within the coordination sphere of the $Cd^{2+}$ ion, have been observed at specific combinations of Cd/C ratio, pH and $B_0$. An example is shown in Figure 26, in which a $^{113}$Cd NMR resonance is found at $\delta$ = 780 ppm – a chemical shift range typical of $CdS_4$-type coordination (Summers, 1988).

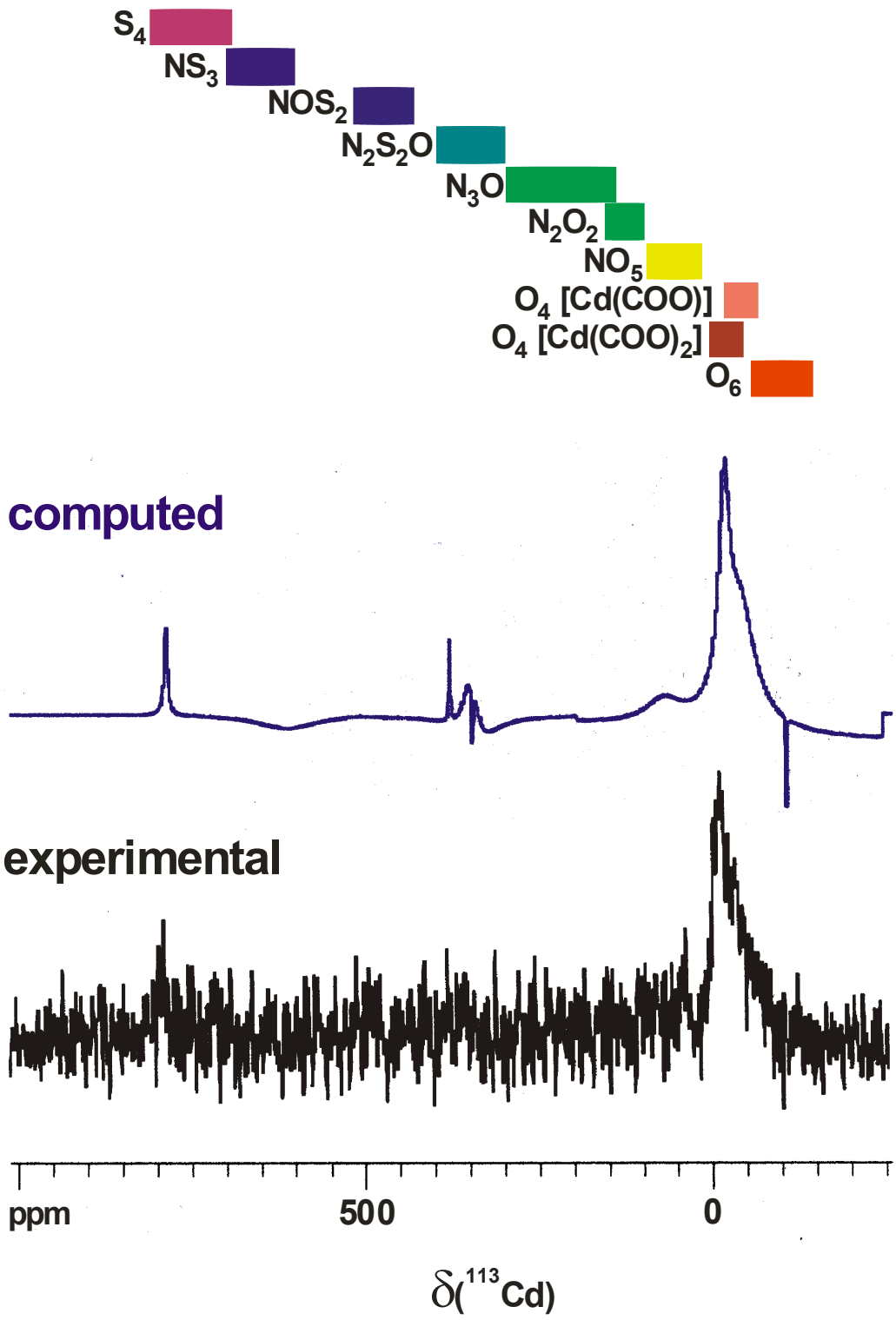



**Figure 26.** Experimental (bottom) and computed (top: azfw = 200 ppm; pc = 0.8) $^{113}$Cd NMR spectra of $^{113}Cd(ClO_4)_2$ bound to Suwannee River organic matter, indicating coordination to sulfur (Cd/C = 0,0056; pH = 9.0; $^{113}$Cd NMR frequency: 110.95 MHz, aq = 233 ms, d1 = 3.0 s, ns = 53340); together with typical chemical shift ranges of cadmium coordination compounds, (Koopal et al., 1994; Milne et al., 2001).

Other low field resonances in the region of $\delta$ = 325 – 375 ppm may indicate single sulfur coordination, but they can also be attributed to multiple chelating amine ligands, (Summers, 1988). The occurrence of very low field $^{113}$Cd NMR resonances within narrow windows of Cd/C ratio, pH and field strength $B_0$ has been reproducible and indicates that at least a fraction (<5%) of Cd is coordinated to sulfur under these experimental conditions.

2.5.5.10. Effect of Temperature on $^{113}$Cd NMR Spectra: The effect of temperature on $\delta_{obs}$ and $\Delta\nu_L$ was examined briefly using a sample with a Cd/C ratio of 0.00559 and pH of approximately pH 4. Spectra were acquired at temperatures ranging from 280 K to 363 K. As temperature increased from 280 K to 303 K, $\Delta\nu_L$ values dropped significantly from about 536 Hz to 456 Hz. Further increases in temperature up to 363 K caused only a marginal decrease of about 5% in $\Delta\nu_L$ values. The lack of sensitivity of $\Delta\nu_L$ to increased temperature above 303 K indicates that a dynamic equilibrium, characterized by a nearly constant line width in the $^{113}$Cd NMR spectrum, has been reached. It was also noted that $\delta_{obs}$ values migrated upfield, but the change was less than 1 ppm. Because $Cd^{2+}$ is exchanging within a highly complex mixture of binding sites, changes in $\delta_{obs}$ and $\Delta\nu_L$ are the result of the combined effects of faster chemical exchange and changes in the temperature-dependent equilibrium constants for binding of $Cd^{2+}$ to the mixture of binding sites. Further speculation is probably not warranted. Otto et al., 2000 have also briefly examined the effects of temperature on $^{113}$Cd NMR spectra in solutions of $Cd^{2+}$ and fulvic acid. They also found that the line width decreases by only a few percent as temperature is raised from 291 K to 330 K; however, they concluded that exchange broadening does not make a major contribution to relaxation. They observed a very minor downfield migration of $\delta_{obs}$ with increasing temperature, and they concluded that the change in line width was more likely due to a decrease in “the binding constant” than to a change in the rate of chemical exchange. Their experiment was conducted at an unspecified pH (perhaps pH 6.4) and at a Cd/C ratio that is more than three times higher than the Cd/C ratio used in the present study. Given the fact that the heterogeneity of the distribution of actively exchanging complexes in a mixture of binding sites is strongly influenced by both the Cd/C ratio and the pH, it is probably not possible to further compare these two studies of temperature effects.

2.5.5.11. General Comments on Chemical Exchange in the Cd-NOM System and the Appearance of $^{113}$Cd NMR Spectra: The work presented here has emphasized the complex interplay between NMR field strength and solution chemistry on the appearance of $^{113}$Cd NMR spectra in the Cd-NOM system. As stated earlier in this paper, when $Cd^{2+}$ undergoes chemical exchange during an NMR measurement, the exchange rate is conventionally referred to as slow, intermediate, or fast, depending on whether the rate of the exchange reaction (in Hz) is much less than, approximately equal to, or much greater than the difference in resonance frequencies of the exchanging species (in Hz). The difference in resonance frequencies of exchanging species (in Hz) is proportional to the magnetic field strength ($B_0$) and is independent of solution chemistry. The rates of chemical exchange reactions (in Hz) depend on the nature and concentration of exchanging ligands and on the concentration of uncomplexed $Cd^{2+}$ ion, all of which are themselves functions of solution chemistry but not of magnetic field strength. In principle, these basic relationships between spectroscopy and solution chemistry apply to the Cd-NOM system; however, their application is confounded by the sheer complexity of NOM, which is best described as a complex, heterogeneous mixture having a wide range of molecular weights, equilibrium constants for binding of protons and metal cations, and rate constants for formation and dissociation of metal-ligand complexes. The observations presented in this paper are consistent with the intuitive expectation that there is no accessible, unique combination of magnetic field strength and solution chemistry for which the exchange rates of all forms of cadmium are simultaneously fast, intermediate, or slow. In fact, irrespective of $B_0$, Cd/C ratio, pH, and temperature, $^{113}$Cd NMR spectra of cadmium ions interacting with NOM will always reflect a superposition of slow, intermediate and fast exchange characteristics. The relative proportion of slow chemical exchange, however, will generally increase with increasing pH, higher $B_0$, decreasing Cd/C ratio and lower temperature.

It follows from these considerations that the presence of a particular form of complexed cadmium in the Cd-NOM system is not sufficient to assure its observation in a $^{113}$Cd NMR spectrum. Depending on the interplay between solution chemistry and magnetic field strength, that particular species may or may not be visible in the spectrum. When a relatively narrow peak is observed, it can be concluded that the species, representing a narrowly defined range of geometries of donor atoms in the coordination sphere of the cadmium ion, is not only present but that its rate of chemical exchange is much smaller than the difference in resonance

frequency (in Hz) of the observed species and its potentially unobservable exchange partner(s). An identification of the unknown interacting species taking part in the chemical exchange would require the acquisition of two dimensional $^{113}$Cd,$^{113}$Cd exchange NMR spectra (Summers, 1988; Vasak, 1998), which is beyond the scope of this research. In our opinion, it is absolutely crucial to emphasize these very peculiar and highly complex characteristics of the metal-NOM exchanging system, as these tend to be overlooked in most of the existing literature covering metal NMR spectroscopy of metal interactions with humic substances and NOM.

### *2.5.6. Conclusions*

$^{113}$Cd NMR spectroscopy offers the most direct information about the cadmium binding to natural organic matter and humic substances, which is characterized by an enormous variety of ligands and by a *quasi*-continuous distribution of binding constants. Owing to the prevalence of chemical exchange in the Cd/NOM system a NMR analysis based solely of chemical shift may result in erroneous results. In general, metal NMR studies of metal binding to NOM require a thorough mapping of the chemically exchanging system following the experimental, procedures developed in this publication; this is of special importance, when the chemical shift ranges of O-, N- and S-coordination are less clearly separated than in the case of $^{113}$Cd NMR spectra.

A meaningful assessment of the effects of chemical exchange on the $^{113}$Cd NMR spectra required NMR studies at two different magnetic field strengths and a series of NMR spectra acquired with a systematic variation of Cd/C ratios and pH values. In contrast to binding of cadmium to peptides and proteins (Vasak, 1998), clean slow exchange conditions will remain elusive in any room temperature NMR study of cadmium binding to NOM; this precludes the determination of kinetic and thermodynamic data of individual coordination sites.

The interpretation of $^{113}$Cd NMR spectra is greatly facilitated and constrained by simultaneous measurements of pH and pCd, which allows a model-independent calculation of organically bound $Cd^{2+}$ under all experimental conditions. In addition, the average chemical shift of organically bound $Cd^{2+}$ ($\delta_{org}$) can be readily calculated from the observed chemical shift, simply by removing the contribution of free $Cd^{2+}$ ion. Trends in $\delta_{org}$ with pH and pCd

can be interpreted directly in terms of actively exchanging complexes within a diverse pool of potential binding sites.

The interpretation of the effects of pH, Cd/C and $B_0$ on line widths and chemical shifts was also greatly facilitated by independent measurements of the concentration of free $Cd^{2+}$. This analysis suggests strongly that, at any combination of pH, Cd/C, and $B_0$ used in this study, line widths and chemical shifts reflect a complex superimposition of slow, intermediate and fast chemical exchange among a heterogeneous population of binding sites. Somewhat surprisingly, some of these trends were most readily explained by the participation of a minor, but strongly binding, fraction of binding sites in which $Cd^{2+}$ is coordinated by nitrogen ligands such as amino groups. Oxygen derived functional groups (aliphatic, aromatic and carboxylic hydroxyl groups) are the most numerous ligands of cadmium bound to NOM, but at higher pH and lower Cd/C ratio a remarkably larger proportion of the nitrogen present in NOM is coordinated to Cd, compared to oxygen.

The NMR conditions used in this study (e.g., $B_0$ = 11.7 T, 10 mm broadband probe, 95 % enriched $^{113}$Cd) would, in principle, actually allow $^{113}$Cd NMR spectra with acceptable signal-to-noise ratio to be acquired in the regime of fast chemical exchange at Cd/C ratios that are up to two orders of magnitude lower than the lowest Cd/C ratio used in this study. In practice, the pH (or pCd) window in which Lorentzian peaks are observed is greatly reduced, and shifted to low pH as the Cd/C ratio decreases. Extrapolating from the results of this research, fast exchange at a Cd/C ratio of $10^{-5}$ or less might only be observed below pH 3 - outside the range expected to be of environmental relevance.

At high pH, where intermediate and slow chemical exchange are more prevalent, it is probably not likely that lower Cd/C ratios can be used, because of temporal changes in pCd and pH during the extended times necessary for the acquisition of NMR spectra with meaningful S/N ratio. These chemical changes cause the chemical speciation of $Cd^{2+}$ to change continuously as a spectrum is being acquired, which results in substantial broadening of the spectrum and a loss of signal intensity.

## 2.6. NMR SPECTRA OF QUADRUPOLE NUCLEI WITH SPIN QUANTUM NUMBERS I > ½.

All nuclei with a spin quantum number I>1/2 exhibit an ellipsoidal charge distribution and an electric quadrupole moment eQ and a quadrupole coupling constant ($e^2q_{zz}Q/h$) (Mason, 1987). Q is positive, if the nucleus is prolate (lengthened) in the direction of its spin angular momentum, negative if it is oblate (flattened). In a molecule, an electric field gradient (efg) at the nucleus is caused because of asymmetry in the local charge distribution, which is imposed by electrons and other nuclei. The energy of a nuclear quadrupole is quantized according to its orientation in the efg even in the absence of an external magnetic field $B_0$ (Mason, 1987). Like the chemical shift δ, the scalar coupling *J* and the dipolar coupling *D*, efg is a tensor and like *D* it is traceless: the isotropic average of energy terms involving the efg is zero.
Therefore, in the liquid phase, the resonance positions in the NMR spectrum are not affected by the nuclear quadrupole coupling and only the averaged value of the shielding tensor δ and the (indirect) spin-spin coupling constants *J* contribute to the positions of the NMR spectral lines (Harris et al., 2002). In the liquid phase, the changes in the local efg with molecular motions induce transitions between different states of the magnetic quantum number, and this is the quadrupole relaxation, which is a very efficient intramolecular NMR relaxation mechanism, depending on the asymmetry parameter η (Pregosin, 1991), and the quadrupole coupling constant $e^2Qq_{zz}/h$ according to Eq. (12)

$$1/T_1 = 1/T_2 = 3/40\ \{(2I+3)/[I^2\,(2I-1)]\}\ [1+\eta^2/3\ [(e^2Qq_{zz}/\,h]^2\ \tau_c \qquad (12)$$

I: nuclear spin quantum number; Q: quadrupole moment [$fm^2$]; η: asymmetry parameter ($0 < \eta < 1$; $\eta = (q_{xx}-q_{yy})/q_{zz}$); $q_{zz}$: maximum component of the efg tensor with principal components $|q_{zz}| > |q_{yy}| > |q_{xx}|$.

Owing to pronounced differential line broadening and quadrupole relaxation (Medek et al., 1997), the observed resonance integrals at sections of chemical shift in NMR spectra of quadrupole nuclei, which are attributed to defined substructures, may deviate from the actual content of nuclei in the respective chemical environments (Lambert et al., 1995; Lu et al., 1998). While $^{11}B$ NMR spectra of NOM/HS are readily accessible (Schmitt-Kopplin et al., 1998), (cf. Figure 27), reports on oxygen-17 NMR spectroscopy of NOM/HS remain scarce.

This is at first glance a surprising fact, considering the primary importance of oxygen containing functional groups, which define a significant proportion of NOM/HS elemental composition, reactivity and properties, in combination with a reasonable receptivity of $^{17}O$ NMR spectroscopy (cf. Figure 13) and a considerable range of $^{17}O$ NMR chemical shift of organic molecules, which allows to discriminate rather similar chemical environments (Kähligand Robien, 1994; Grandy et al., 1984).

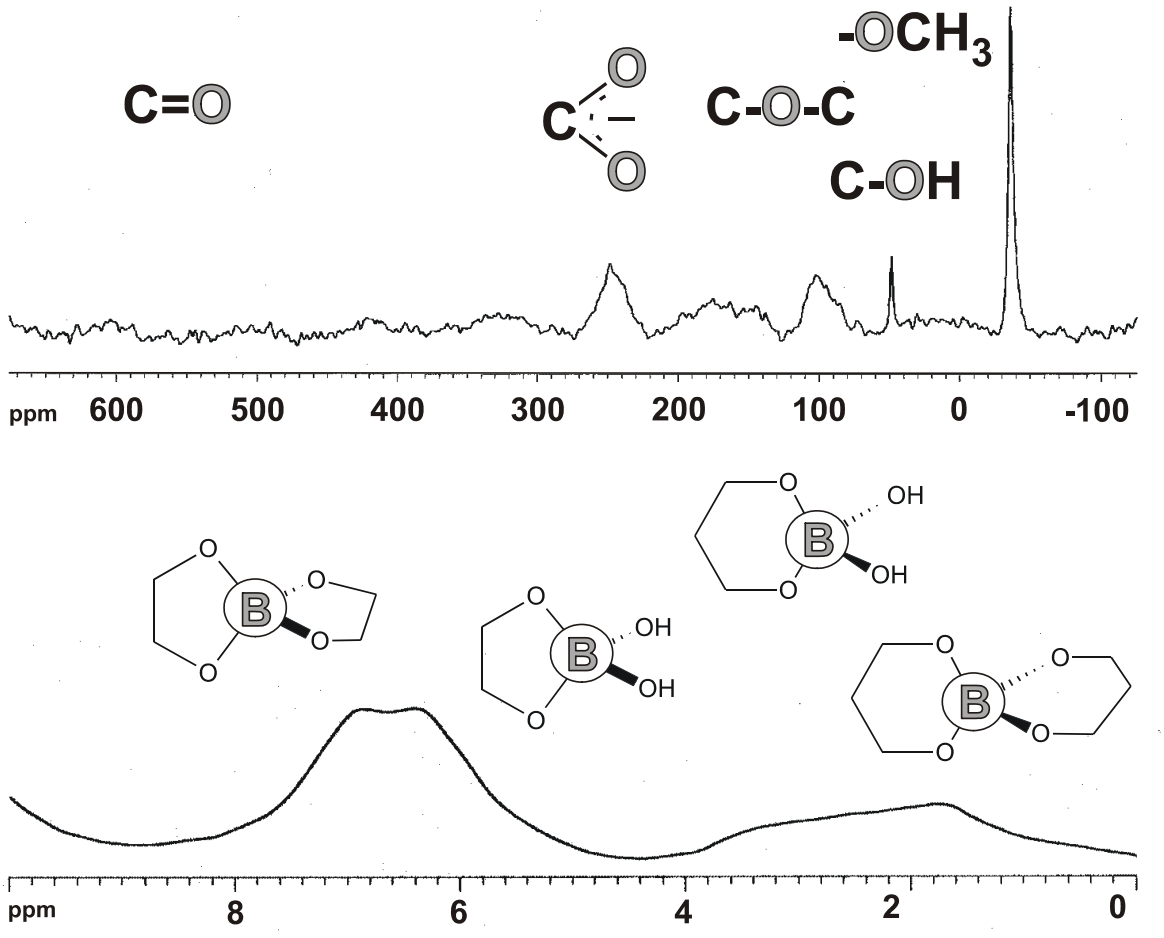



**Figure 27.** $^{11}B$ NMR spectrum of a fresh water humic acid (bottom), and a $^{17}O$ NMR spectrum (top) of a bog lake humic substance (sealed in dry pyridine-$d_5$), together with main substructure regimes.

The content of oxygen derived functional groups in NOM/HS (e. g. aliphatic, aromatic and carboxylic hydroxy, keto, aldehyde, methoxy) can be readily estimated from $^{1}H$ and $^{13}C$ NMR spectra in combination with infrared and titration studies. Based upon this information, reasonable estimates of the appearance of the $^{17}O$ NMR spectra of NOM/HS can be derived (Hertkorn, 2006). In practice, the combined effects of variable quadrupole coupling constants of individual coordination sites (Kählig and Robien, 1994; Schumacher and Lauterwein 1989), nuclear Overhauser effects (Sergeyev et al., 1999), differential local symmetry (Paytan et al., 2003), and, most prominently, the rate of chemical exchange in relationship to the spread in $^{17}O$ NMR frequency of the respective chemical environments in NOM/HS will lead to severe differential line broadening and intensity distortions (Hertkorn, 2006). Lack of systematic studies aimed at a better understanding of these effects so far have impeded the use of $^{17}O$ NMR spectroscopy for the meaningful structural analysis of NOM/HS.

## 2.7. NMR INVESTIGATIONS OF INTERACTIONS OF ORGANIC AND INORGANIC COMPOUNDS WITH NOM/HS

Accurate molecular level information about mechanisms and dynamics of the interactions of NOM/HS with organic molecules can be obtained largely from the NMR spectra of the main constituent nuclei by methods of chemical shift mapping, nuclear-spin relaxation, nuclear Overhauser effect, and study of exchange and diffusion phenomena (Price, 2002; Morris et al., 1999; Pellecchia et al., 2002; Zuiderweg, 2002). When labels of $^{2}H$, $^{13}C$, $^{15}N$ in small molecules are used, the low natural abundance of $^{2}H$ (0.00015%), $^{13}C$ (1.1%) and $^{15}N$ (0.37%) in NOM/HS is of considerable advantage, since the behaviour of specific chemical environments within these small molecules can be probed by an array of one, two and three dimensional NMR spectra at rather low levels of concentrations. The relaxation in $^{2}H$ (Nanny and Maza, 2001), $^{13}C$ (Nanny et al., 1997), and $^{19}F$ (Dixon et al., 1999), NMR spectra of small aromatic molecules has been used to analyse the dynamics of their interaction with NOM/HS.

Another promising technique to probe the interaction of NOM/HS with fluorinated organic compounds is $^{19}F$ NMR spectroscopy (Campos-Olivas et al., 2002). Owing to a very strong back bonding contribution in the carbon-fluorine bond, the effects of fluorination upon representative bulk properties of the organic molecule and its reactivity to NOM/HS are rather small (Campos-Olivas et al., 2002). The $^{19}F$ NMR label has exceptional sensitivity (Mortimer and Dawson, 1991; cf. Figure 13); this is even more remarkable as organically bound fluorine is absent in natural NOM/HS: environmentally significant trace levels of fluorinated organic molecules can be readily investigated by $^{19}F$ NMR spectroscopy (Mortimer and Dawson, 1991; Ellis et al., 2000).

Interaction of NOM/HS with non metal (e. g. $^{11}B$, $^{77}Se$) and metal ions (most appropriately: $^{27}Al$, $^{51}V$, $^{113}Cd$, $^{199}Hg$, $^{203/205}Tl$, $^{207}Pb$) can be followed by direct observation of these NMR active nuclei (Li, 1998; Lambert et al., 1995; Lu et al., 1998; Hertkorn et al., 2004). In NMR spectra of quadrupole nuclei, both chemical exchange and quadrupole relaxation obscure direct relationships between the appearance of the metal NMR spectra and the coordination geometry and the nature of the NOM/HS ligands (cf. chapter 2.6.).

## 3. General Characteristics of Two Dimensional NMR Spectra of NOM/HS

The invaluable contribution of quantitative one dimensional NMR spectra to NOM/HS structural analysis resides in the ability to define the *relative amounts of fundamental building blocks*. The main significance of the uniquely versatile 2D NMR spectroscopy of NOM/HS is to *enhance the reliability of NMR resonance assignments*, as cross peaks in 2D NMR spectra indicate molecule fragments rather than individual atoms. A combined analysis of several 2D NMR spectra ultimately leads to the assemblage of extended substructures of NOM/HS (cf. Figure 28, Table 6).

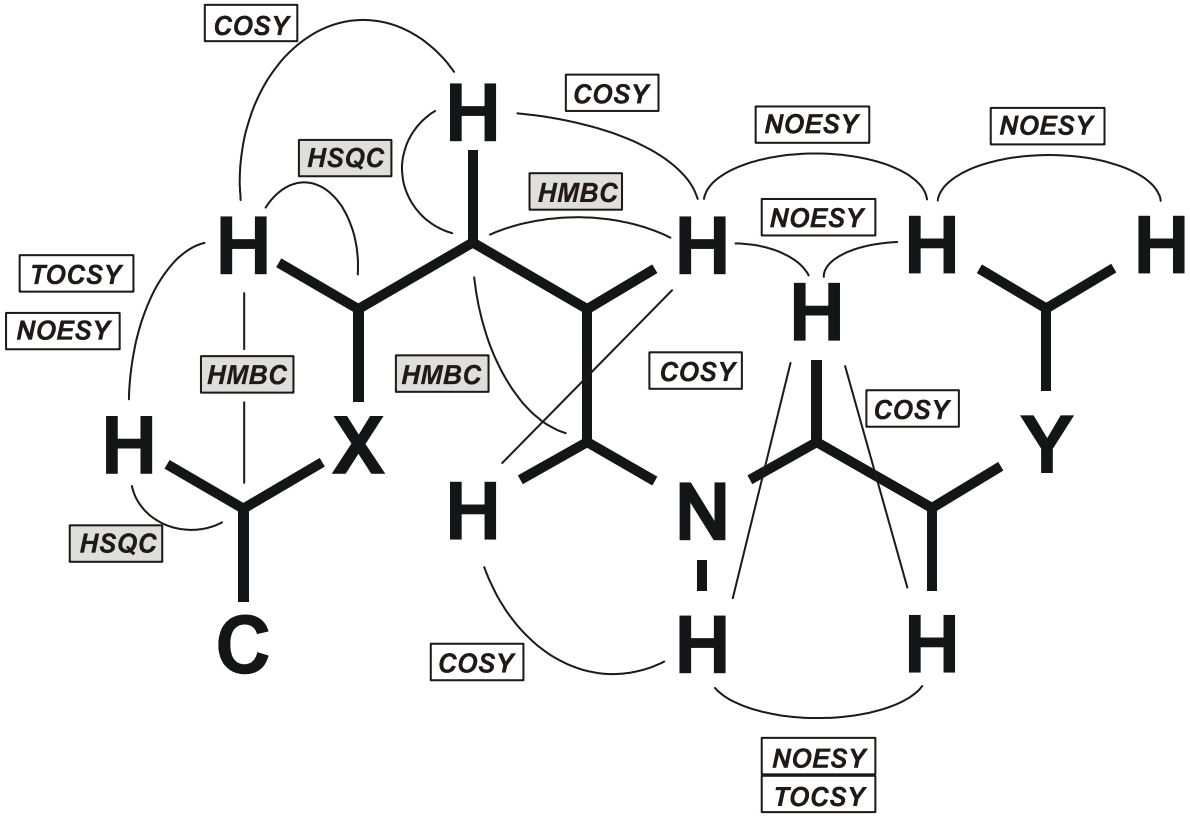



**Figure 28.** The contributions of different 2D NMR experiments in the assembly of extended substructures of NOM/HS.

**Table 6.** A selection of 2D NMR experiments, useful for the structural analysis of NOM/HS.

| NMR acronym | selected interaction | remarks |
|---|---|---|
| JRES | $^{n}J(H,H)$ | sensitive, but long duration of F1 increments may interfere with fast transverse relaxation at higher resolution (F1) |
| COSY | $^{2,3}J(H,H)$ | antiphase magnetization, susceptible to self-cancellation of cross peaks; optimum transfer: 1/J |
| TOCSY | $^{n}J(H,H)$ | extended spin systems; optimum transfer: 1/2J (n = 2-6) |
| HSQC, HMQC, HETCOR | $^{1}J(X,H)$ | highly informative for X: $^{13}C$ |
| HMBC | $^{n}J(X,H)$ | most useful for X: $^{13}C$ and $^{15}N$ |
| NOESY, HOESY | spatially proximate H…H and H…X interactions | for NOM/HS often accompanied or dominated by chemical exchange |

An increased signal dispersion of NMR cross peaks into two frequency dimensions greatly reduces resonance overlap, especially in the case of heteronuclear correlated NMR spectra, where the resonance frequencies of heteronuclei cover a substantial range (Cavanagh et al., 1996; Croasmun and Carlson, 1996). In addition, 2D NMR experiments act as filters, which emphasize specific forms of binding (cf. Table 6; case study 2) and strongly discriminate against other resonances, thereby allowing to selectively probe *bonding interactions*, *spatial relationships* and *chemical exchange*, respectively (Cavanagh et al., 1996; Croasmun and Carlson, 1996; Hertkorn et al., 2002; Hertkorn et al., 2001). This simplification of spectra enables a detailed analysis of heavily crowded regions of chemical shift. The sensitivity of proton detected heteronuclear 2D NMR spectra is commonly much higher than that of heteronuclear 1D NMR spectra (cf. eq. 13).

In comparison with classic inverse geometry probes, cryogenic NMR probes provide a distinct increase in signal to noise ratio of proton detected NMR spectra (Russel et al., 2000; Crouch et al., 2001; Martin, 2002). The relative S/N enhancement depends on sample composition, reaching a factor in excess of four in case of organic solvents, which is declining to near unity for solutions with very high ionic strength (Flynn et al., 2000; Kelly et al., 2002).

Therefore NOM/HS, dissolved in DMSO-$d_6$ under complete exclusion of air and moisture are promising samples for NMR spectra of very high resolution (cf. Figure 29). The exclusion of oxygen obviates longitudinal relaxation from paramagnetic oxygen. The very large intrinsic sensitivity of the cryogenic probes allows the acquisition of meaningful NMR spectra from samples of low concentration, when the tendency of NOM/HS to align and aggregate in

solution is minimized. In addition, hydrogen bonding is severely reduced in dipolar aprotic solvents. Both effects enhance the molecular mobility, increase the longitudinal (Wang et al., 2003), and spin-spin relaxation times and consequently the attainable resolution in 2D NMR spectra (Bovey and Mirau, 1996).

In extension of the various aspects of high-resolution 2D and 3D (Simpson et al., 2003), NMR spectroscopy of NOM/HS, which have been highlighted in several recent publications (Fan et al., 2000; Buddrus et al., 1989; Hertkorn et al., 2002; Hertkorn et al., 2001; Chien and Bleam, 1998; Haiber et al., 1999; Kingery et al., 2000; Schmitt-Kopplin et al., 1998; Simpson et al., 1997; Wang et al., 1998; Simpson, 2001; Simpson et al., 2001; Kingery et al., 2000; Simpson et al., 2002), the emphasis of the following section is to provide representative examples how different two dimensional NMR spectra contribute to the structural analysis of NOM/HS. In practice, 2D NMR spectroscopy of NOM/HS will frequently become an exercise in optimisation of acquisition and calculation parameters for individual samples to enhance NMR sensitivity and resolution, because interaction of NOM/HS constituent molecules – and in result: all relaxation dependent NMR parameters strongly depend on sample conditions, like concentration, solvent, pH and temperature.

## 3.1. HOMONUCLEAR 2D NMR SPECTRA OF NOM/HS

### *3.1.1. Introduction; differences between COSY and TOCSY NMR spectra of NOM/HS*

Proton homonuclear shift correlated 2D NMR spectra offer high sensitivity and long range connectivity information. The latter is instrumental to assemble NOM/HS substructures (cf. Figure 28). One method is to relate heteronuclear single bond [$^{1}$J(C,H)] cross peak (which represent isolated pairs of directly connected carbon and proton atoms; cf. chapter 3.1.4.) via a combined analysis with homonuclear (COSY and TOCSY) cross peaks. The main potential of COSY spectra resides in the structural analysis of vicinal and geminal proton-proton connectivities, which represent substructures of organic molecules with a minimum size of five and six atoms (cf. Table 7).

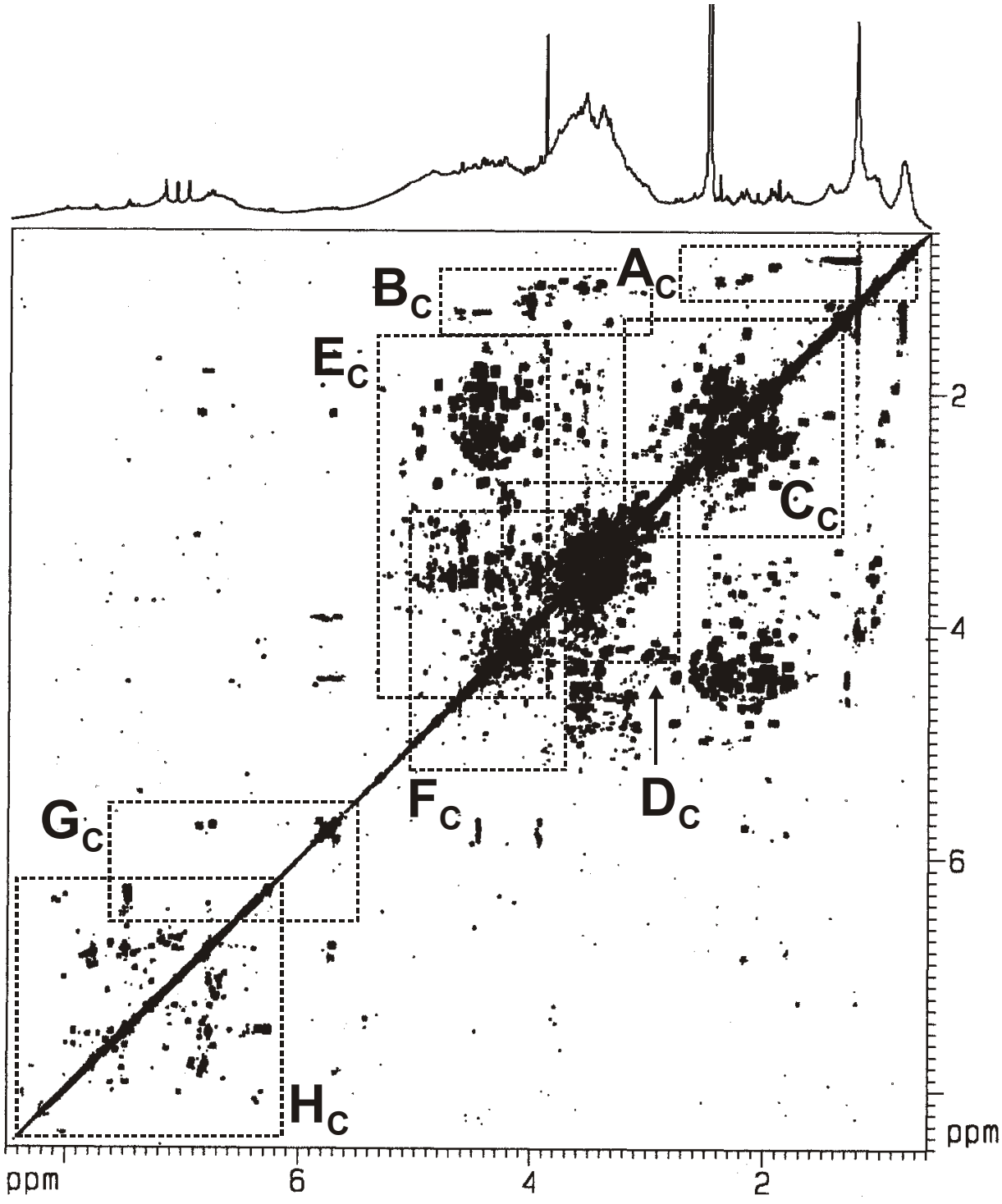



**Figure 29.** 2Q-COSY spectrum of a peat fulvic acid in dry DMSO-$d_6$ (for assignment of cross peak substructure regimes: cf. Table 7).

The rather long transfer times in COSY and TOCSY NMR spectra (Cavanagh et al., 1996; Croasmun and Carlson, 1996; case study 2), imply loss of magnetization caused by fast transverse relaxation. Owing to different mechanisms of magnetization transfer (Ernst et al., 1987), both common and different resonances will appear in COSY and TOCSY NMR spectra of NOM/HS (Hertkorn et al., 2002). TOCSY cross peaks represent net transfer of in-phase magnetization with near absorption line shapes for diagonal and cross peaks and they indicate extended spin systems rather than vicinal and geminal connectivities alone as provided in standard COSY NMR spectra. COSY cross peaks represent coherence transfer by scalar couplings (*J*): antiphase magnetization is created without transfer of net magnetization. Therefore, COSY cross peaks are susceptible to self-cancellation, when the linewidth $\Delta\nu$ exceeds the coupling constant *J* (Figure 38). However, COSY spectra of dilute solutions of soil NOM/HS in dry DMSO-$d_6$ under total exclusion of moisture exhibit exceptionally well resolved (cf. Figure 29) cross peaks.

**Table 7.** Regions of chemical shift in the $^1$H,$^1$H COSY NMR spectrum (Figure 29).

| region | assignment | humic constituent |
|---|---|---|
| $A_C$ | -C-CH-$CH_3$ | methyl groups, bound to purely aliphatic carbon |
| $B_C$ | -C-CH-CH-O- | desoxy sugars, ethers, esters |
| $C_C$ | -$C_f$-CH-$C_f$H-C- | intra functionalised aliphatic, ß to heteroatoms |
| $D_C$ | -CH(O)-CH(O)- | intra carbohydrate, without anomerics |
| $E_C$ | -CH(X)-CH-$C_f$ | functionalised aliphatic unit with one heteroatom (X = O, N) |
| $F_C$ | -O-CH(O)H-CH | intra carbohydrate, with anomerics |
| $G_C$ | -CH=CH- | double bonds, five-membered heterocycles |
| $H_C$ | -$C_{ar}$H-$C_{ar}$H- | ortho protons in aromatic rings |

Fast and differential transverse relaxation within the polydisperse and molecularly inhomogeneous NOM/HS attenuates COSY cross peaks (optimum transfer time at 1/J) more severely than TOCSY (optimum transfer time at 1/2J) cross peaks (Hertkorn et al., 2002). Therefore cross peaks visible solely in TOCSY spectra represent long range couplings in extended spin systems, but also may indicate short range couplings in spin systems of restricted flexibility, where self-cancellation of COSY cross peaks has occurred (at linewidths > 15 Hz). Cross

peaks showing up solely in COSY spectra indicate either small spin systems of very high flexibility (with long transverse relaxation times, sharp lines and large S/N ratio) or parts of extended spin systems; in this case the proton resonance integrals can be distributed across many TOCSY cross peaks and then, the S/N ratio of individual TOCSY cross peaks may become attenuated below recognition.

Cross peaks in highly resolved, phase-sensitive COSY NMR spectra of NOM/HS show a fine structure, which allows to determine (under consideration of the effects of the line width on the anti-phase lineshape) the values of $^{n}J(H,H)$ coupling constants. In phase sensitive COSY NMR spectra, the determination of active (spins, whose frequencies are connected by the cross peaks, produce anti-phase doublets) and passive spins (spins, which are not directly involved in the magnetization transfer, split the cross peaks into in-phase doublets) allows to specify the coupling network of any given cross peak, a major leap to assemble and confirm extended substructures in NOM/HS (cf. Figure 30).

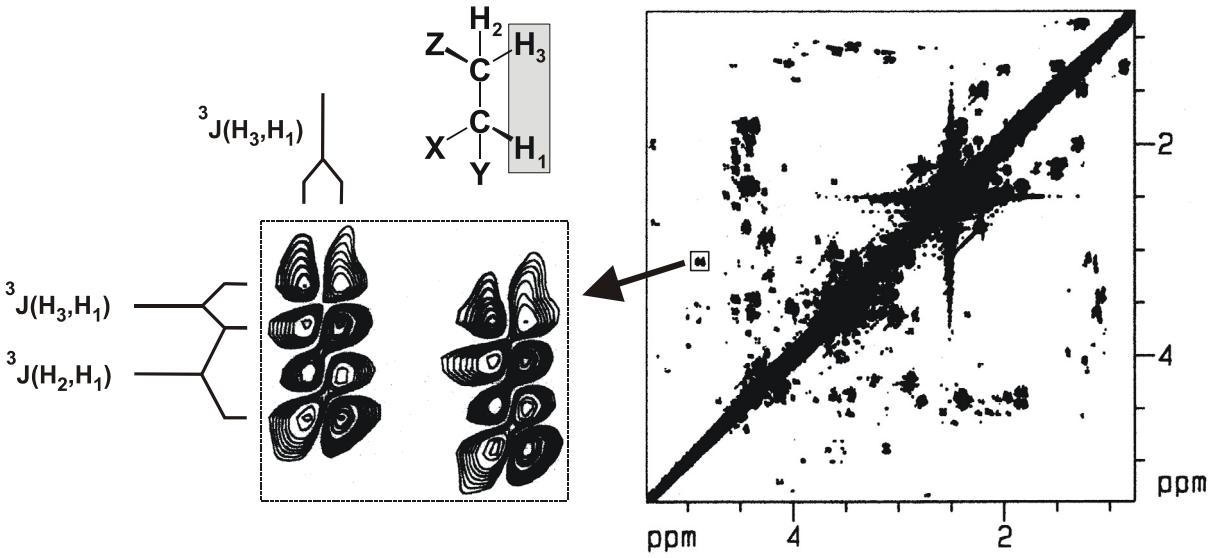



**Figure 30.** Evaluation of spin systems by analysis of the fine structure of COSY cross peaks in a soil humic substance; the selected cross peaks show splitting by active coupling [$^{3}J(H_3\text{-}H_1)$] and by passive coupling [$^{3}J(H_2\text{-}H_1)$]; the nature of this AMX-type spin system indicates non carbohydrate origin (anomeric protons commonly show AB-spin systems).

### *3.1.2. NOESY/EXSY NMR spectra: identification of exchangeable protons in NOM/HS*

NOESY/EXSY NMR spectra rely on the transfer of longitudinal magnetization to establish connectivities between spins either by the nuclear Overhauser effect (nOe) or by chemical exchange. NOESY NMR spectra of NOM/HS exhibit relatively intense cross peaks resulting from chemical exchange and relatively weak cross peaks resulting from spatial proximity (nOe effects; Hertkorn et al., 2001; Chien and Bleam, 1998), their relative phase with respect to diagonal peaks depends on the molecular size and correlation time $\tau_c$. In NOM/HS the magnetization transfer caused by cross relaxation and by chemical exchange often results in NOESY cross peaks of alike signs as diagonal peaks, a behaviour typical of large molecules (spin diffusion limit).

**Table 8.** Regions of chemical shift in the $^1H,^1H$ NOESY/EXSY NMR spectrum shown in Figure 31.

| region | assignment | origin of cross peak |
|---|---|---|
| $A_N$ | -$C_{al}$H…HO-C-, -X-CH and -$C_f$-CH…H$C_f$-C, -X-CH…HC-X | chemical exchange, nOe |
| $B_N$ | -$C_{al}$-OH…$H_2O$/HO-$C_{al}$ , $C_{al}$-H…H-C-X- | chemical exchange, nOe |
| $C_N$ | -(C=O)NH…H-$C_{al}$- | nOe |
| $D_N$ | -$C_{ar}$-OH…HO-$C_{ar}$ , -$C_{ar}$H-$C_{ar}$H- | chemical exchange, nOe |
| $E_N$ | -COOH…HO-$C_{al}$- | chemical exchange |
| $F_N$ | -COOH…HO-$C_{ar}$- | chemical exchange |

The significance of the NOESY/EXSY spectra for the characterization of NOM/HS resides in the ability to identify exchangeable protons and to determine their exchange rates, which can be calculated from the build-up of the cross peak integral at different mixing times $\tau_m$ (Kumar et al., 1981; Xiang and Markham, 1996). The location and intensity distribution of the cross peaks (cf. Figure 31; Table 8) indicates that selective chemical exchange takes place between different labile proton species in NOM/HS (Hertkorn et al., 2001).

At very short mixing times ($\tau_m$ < 10 ms) only cross peaks originating from chemical exchange are observed in NOESY/EXSY spectra of NOM/HS. Under conditions of slow chemical exchange (such as in dry DMSO-$d_6$ solution under careful exclusion of moisture), spin systems containing HX-C-H units (X: N, S, O) produce cross peaks in COSY and TOCSY NMR spectra, which vanish under conditions of fast chemical exchange. Therefore, a

combined analysis of NOESY/EXSY, and COSY/TOCSY cross peaks is required to provide an in depth analysis of non-exchangeable and labile protons in NOM/HS (Hertkorn et al., 2006).

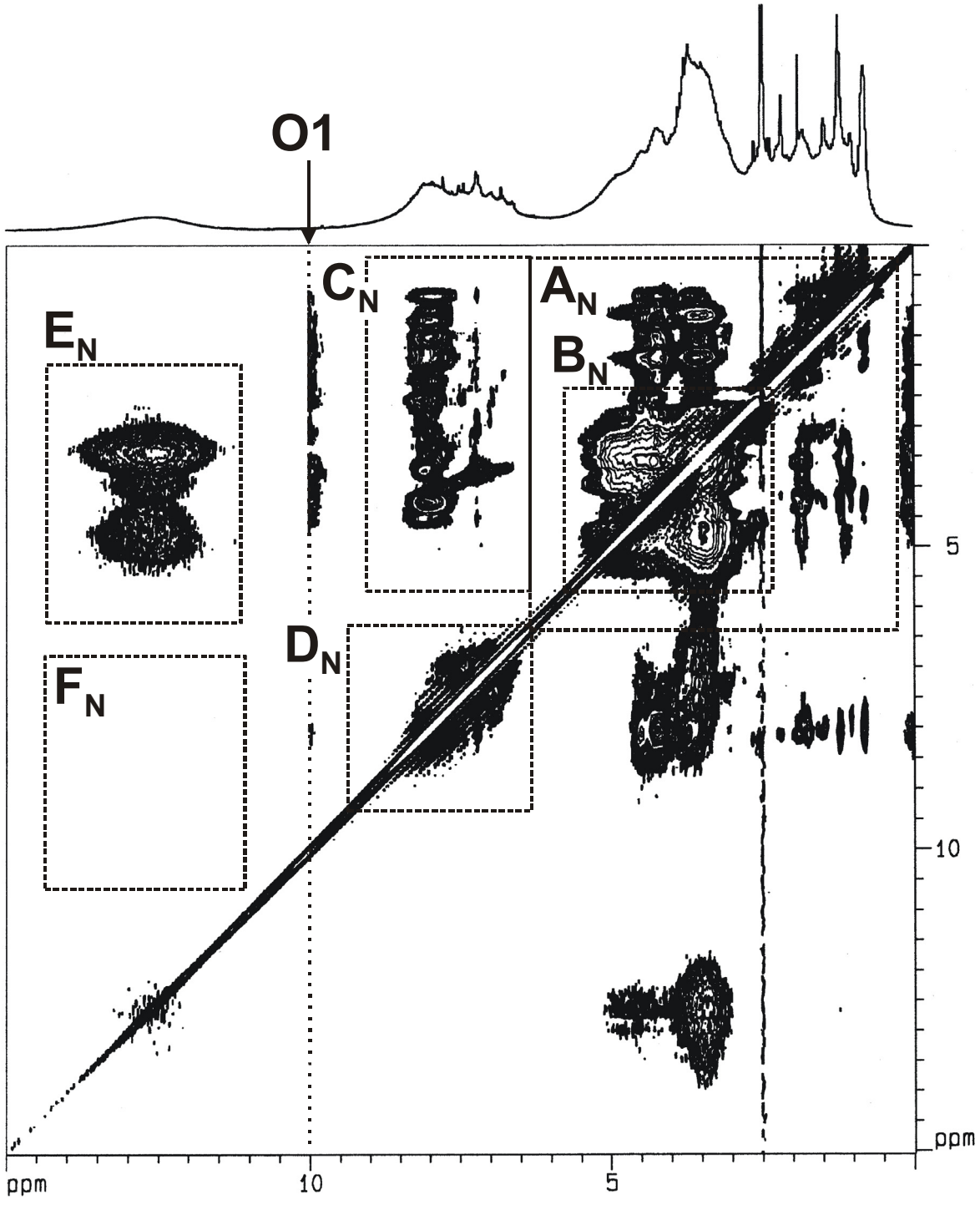



**Figure 31.** EXSY/NOESY spectrum of a soil humic acid in dry DMSO-$d_6$ ($\tau_m$ = 250 ms); O1: excitation offset frequency.

### *3.1.3. Homonuclear J-spectroscopy of NOM/HS*

Among the simplest 2D NMR experiments to understand and implement is J-spectroscopy, which uses spin echo modulation to provide a separation of chemical shifts and spin-spin coupling effects. Homonuclear J-spectroscopy enables the determination of proton-proton coupling patterns and –constants across the entire $^{1}$H NMR spectrum (Tomlins et al., 1998; Viant, 2003).

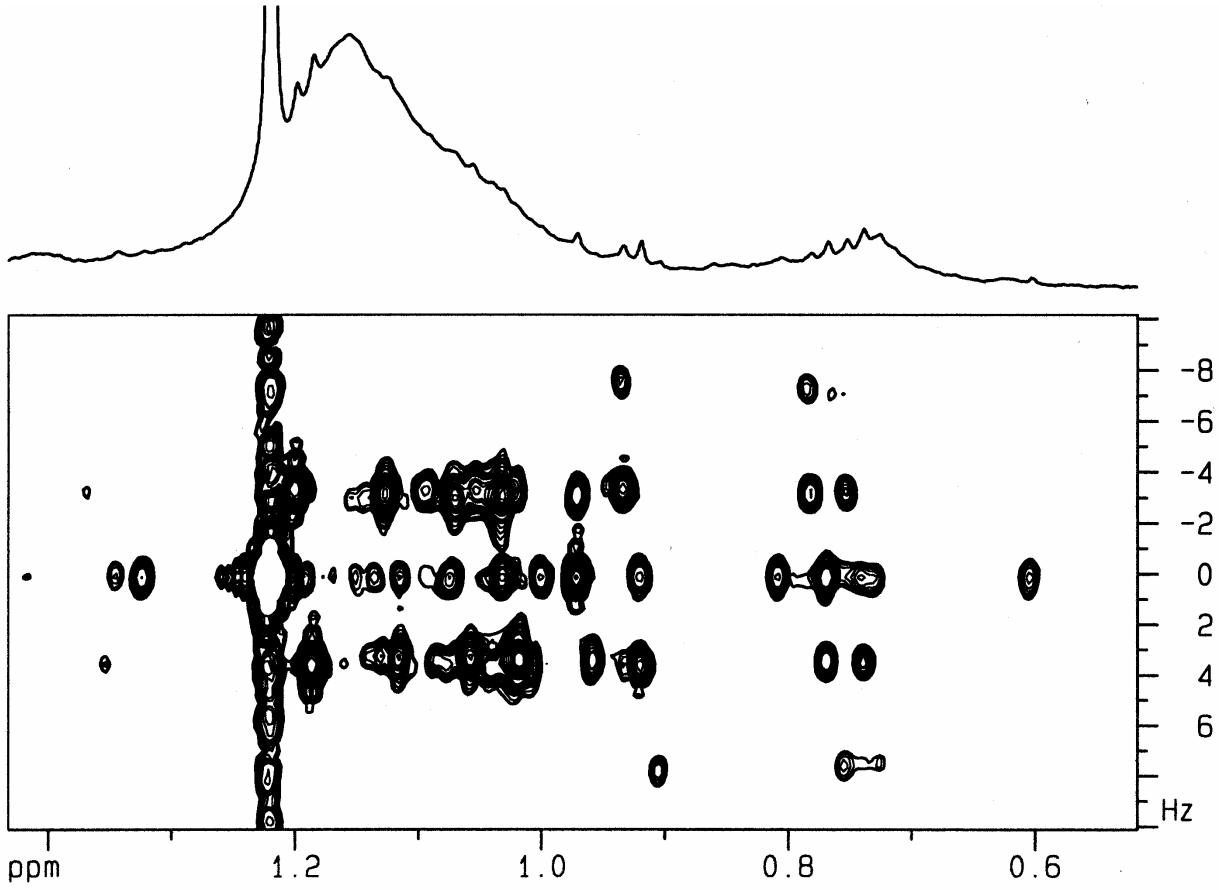



**Figure 32.** $^{1}$H,$^{1}$H-J-resolved spectrum of a soil leachate fraction, obtained by free flow electrophoresis.

The most significant and useful assignment options in the analysis of NOM/HS are (a) the determination of aromatic substitution patterns, (b) the discrimination of (typically more complex) carbohydrate and (commonly less complex) other chemical environments in the middle region of the spectrum, (c) the assessment of branching in the upfield region of functionalised aliphatics and, perhaps most informative, (d) the determination of the number

of protons vicinal to methyl protons (cf. Figure 32). Heteronuclear variants of J-spectroscopy are considered less informative, because carbon multiplicity information is readily available from $^{13}C$ DEPT NMR spectra (cf. chapter 2.4.1.). The sensitivity of homonuclear J-spectroscopy is excellent; however, the attainable resolution of coupling constants strongly depends on the relaxation characteristics of the respective NOM/HS.

### 3.1.4. HETERONUCLEAR SHIFT CORRELATED 2D NMR SPECTRA

Multidimensional heteronuclear NMR experiments correlate heteronuclear with proton resonances by transfer of coherence (or polarization) between the heteronuclear (S) and the proton (I) spins, with initial excitation of either I or S spin polarization and detection of either I or S spin magnetization. The overall sensitivity of the heteronuclear correlation NMR experiments is proportional (Croasmun and Carlson, 1996), to

$$S/N \sim \gamma_{ex} \gamma_{det}^{3/2} [1 - \exp(-R_{1,ex}T_C)] \qquad (13)$$

$\gamma_{ex}$ and $\gamma_{det}$ are the magnetogyric ratios of each nucleus excited at the begin and detected at the end of the pulse sequence; $T_C$ is the recycle time of the experiment; $R_{1,ex}$ is the spin-lattice relaxation rate constant of the excited nucleus [cf. equation (1)].

NMR pulse sequences, which correlate directly bonded proton and carbon atoms with respect to their chemical shift, show optimum transfer amplitudes in the 5-20 ms range (HSQC at $^{1}J(CH)$ = 150 Hz: 10.9 ms; Heikkinen et al., 2003). Their relatively short duration with respect to average $T_1$ and $T_2$ relaxation times in humic substances (Fründ and Lüdeman, 1989; Preston and Blackwell, 1985; Hertkorn et al., 2002), is more favourable to prevent signal loss in case of fast transverse relaxation, when compared to homonuclear 2D NMR sequences. The large spread of frequency in F1 ($^{13}C$) dimension and the exclusive discrimination in favour of proton-carbon single bonds greatly reduces resonance overlap. These effects in combination make HSQC NMR spectra very auspicious candidates to reveal structural information of NOM/HS in great detail.

**Table 9.** Regions of chemical shift in the $^{1}H,^{13}C$ HSQC NMR spectrum Figure 33.

| region | assignment | partial structure |
|---|---|---|
| $A_H$ | -C-$CH_3$ | methyl groups, terminating aliphatic chains |
| $B_H$ | -$C_f$-CH | aliphatic CH, β to heteroatoms |
| $C_H$ | -X-CH- | single heteroatom substituted aliphatics |
| $D_H$ | C-N(H)-CH- | amines, bound to aliphatic carbon |
| $E_H$ | -(C=O)NH-CH- | amides, e. g. in proteins |
| $F_H$ | -$OCH_3$ | methoxyl (aliphatic and aromatic esters and ethers) |
| $G_H$ | -O-CH- | single heteroatom (oxygen) substituted aliphatics |
| $H_H$ | O-CH-X | double heteroatom substituted aliphatics (e.g., acetals, anomeric CH) |
| $I_H$ | -$C_{ar}$H | aromatic and heterocyclic systems |

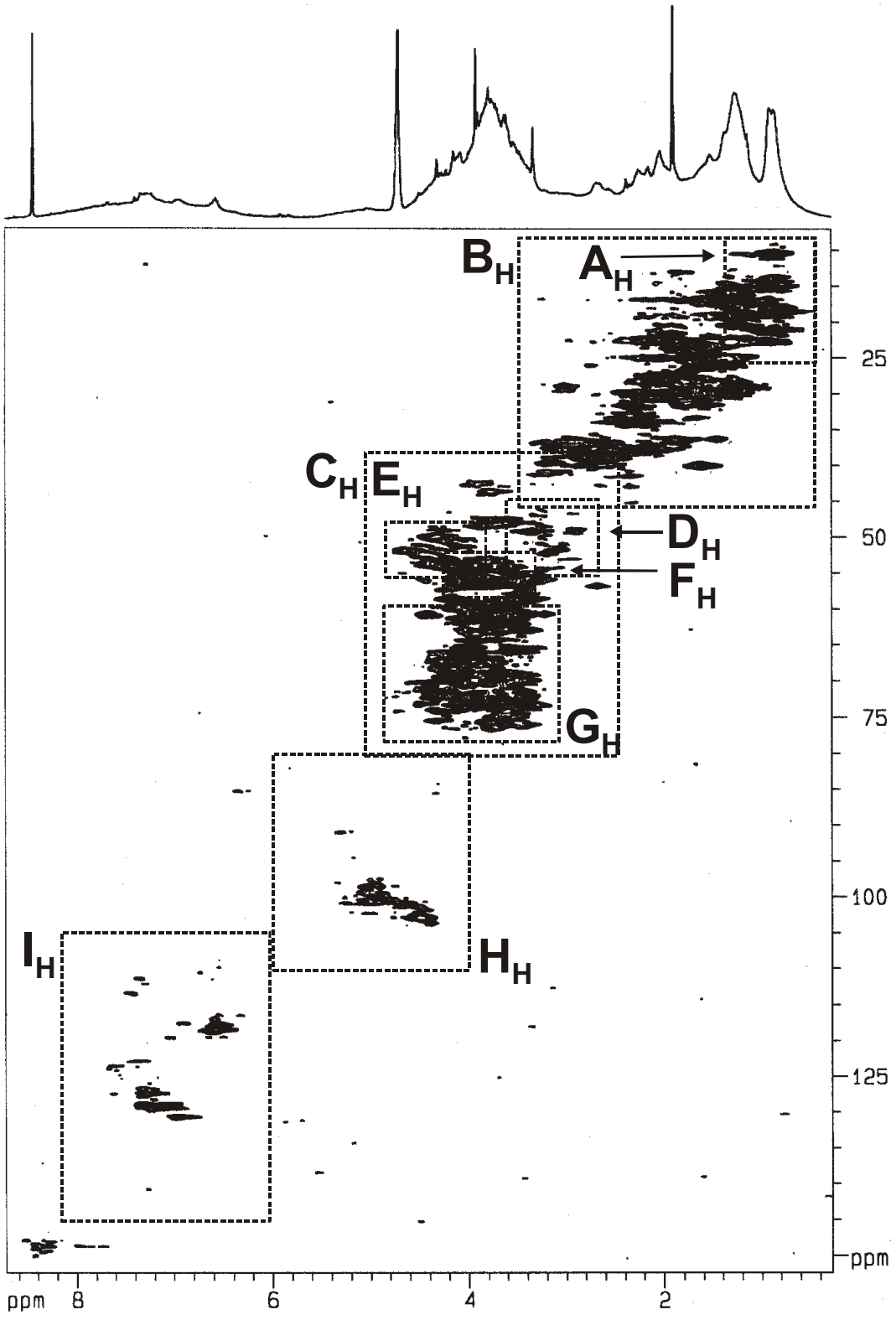



**Figure 33.** $^{1}H$, $^{13}C$ -HSQC NMR spectrum of a soil humic acid in 0.1 N NaOD with key structures as given in Table 9.

### *3.1.4.1. Cross peak fine structure of one bond $^{13}C,^{1}H$ correlation spectra of NOM/HS*

In heteronuclear NMR correlation experiments, high resolution in F2 as defined by the digitisation can be easily achieved in the detection period $t_2$. During the evolution time $t_1$, the resolution in F1 is limited by the number of experiments in the total acquisition time.

The resolution in $^{13}C$ NMR spectra of samples with natural isotopic abundance is rather high under conditions of proton decoupling due to rather long transverse relaxation times and the absence of homonuclear $^{1}J(C,C)$ couplings (Cavanagh et al., 1996; Croasmun and Carlson, 1996). The requirements for the digitisation of the proton NMR spectra are less stringent as the resonances are broadened by fast transverse relaxation and homonuclear $^{n}J(H,H)$ couplings, and because the overall bandwidth of proton NMR spectra [in (Hz)] is smaller. However, carbon detection schemes have considerable sensitivity disadvantages [cf. equations (1) and (13)]. Owing to different physical mechanisms operating within the pulse sequences, the four fundamental single bond carbon proton [$^{1}J(C,H)$] correlation NMR spectra of NOM/HS show a variation in relative cross peak amplitudes and resolution (Reynolds et al., 1997; Bax et al., 1990; Norwood et al., 1990); an in-depth analysis of these (relaxation) effects provides substantial information about structure and dynamics of NOM/HS, (Hertkorn, 2006).

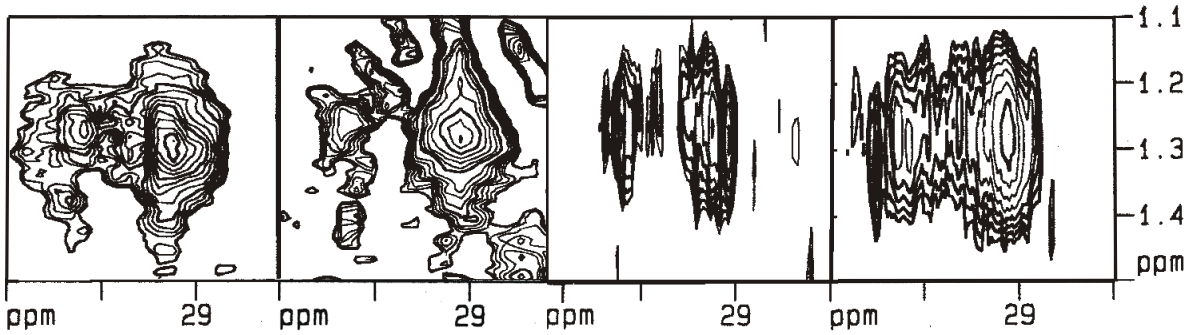



**Figure 34.** Cross peaks of polymethylene section in IHSS reference material 1R106H, acquired with four different $^{1}J(CH)$ correlation NMR pulse sequences (all four phase sensitive NMR spectra represent identical total acquisition times, cf. Table 29): $^{1}H,^{13}C$-HMQC (left), $^{1}H,^{13}C$-HSQC, (second from left), $^{13}C$, $^{1}H$-DEPT-HETCOR (second from right), $^{13}C,^{1}H$-INEPT-HETCOR (right).

## 3.2. CASE STUDY 2: Comparative analysis of partial structures of a peat humic and fulvic acid using one and two dimensional NMR spectroscopy

### *3.2.1. Abstract*

***NMR resonance integrals obtained from one dimensional NMR spectra provide semiquantitative contents of humic constituents with limited resolution in structural detail. When supplemented by connectivity information available from homo- and heteronuclear two dimensional NMR spectra a more reliable assignment of humic substructures becomes available. This is demonstrated with a comparative one and two dimensional NMR analysis of a fulvic and a humic acid obtained from Eriophorum peat. An example of a detailed analysis of the proton chemical shift region normally attributed to carbohydrates shows substantial contributions from amino acids, amino and desoxy sugars and highly oxidized aliphatic chains of intermediate length. The very good resolution of structural detail by a combined analysis of all NMR spectra shows that the effect of the fractionation procedure on the composition and chemical structure of humic materials is very significant. The comparison of the partial structures comprising HA and FA of the peat humic materials studied indicates that FA is diagenetically downstream of HA favouring the biopolymer degradation (BD) model of humification.***

### *3.2.2. Introduction*

Humic substances (HS) constitute the most important pool of transient refractory organic carbon in the geosphere. They occur in soil, aquatic and marine environments and compose an important cross-link of natural and anthropogenic fluxes of carbon, nitrogen and phosphorus. Humification, with respect to the contribution to the global carbon cycle (20 Gt/a) is the second important process in turnover after photosynthesis (60 Gt/a) (Hedges and Oades, 1997). In contrast to biochemical pathways of synthesis, the formation of HS out of the plant and animal debris follows no genetic code and occurs stochastically, governed by kinetic and thermodynamic restraints so that only refractory structures are likely to survive being not susceptible to further microbial and chemical degradation. The stochastic nature of humification is expected to produce an immense structural heterogeneity and polydispersity of HS (MacCarthy and Rice, 1988). As a result, up to date HS are defined purely operational as

alkali extractable fraction of soil and solid fuel organic matter and their most common classification is based on the solubility at different pH values (Stevenson, 1982). In general, HS are classified into humic acids (HA), insoluble at pH < 1-2, and fulvic acids (FA), soluble across the whole range of pH values. HA represent a higher molecular weight, more condensed fraction of HS, whereas FA are of lower molecular weight and more oxidized than HA. The two most common hypotheses of humification are the biopolymer degradation (BD) and the abiotic condensation (AC) models. The AC hypothesis implies the complete disintegration of the non-living organic matter as the first stage of humification (Hatcher and Spiker, 1988) and considers FA as precursors of HA. In contrast, the BD hypothesis considers the humification as a slow oxidation, condensation and decomposition process of plant and animal biopolymers, in which the integrity of the parent materials is not completely destroyed by the microbial degradation (Hedges and Keil, 1999; Hedges, 1988). To assess the validity of both hypotheses, a detailed information on the partial structures of HA and FA can be of particular importance.

NMR spectroscopy over last decades has provided key insight into structural details of humic substances (Ourisson et al., 1979). Both the polydispersity and structural heterogeneity of HS induce specific effects on the key NMR parameters chemical shift $\delta$, coupling constant J, amplitude A and resonance width at half height $\Delta\nu_{1/2}$. The heterogeneity of HS causes overlap of resonances in their NMR spectra, resulting in rather broad signal envelopes. Variation in aggregate size, shape (molecular conformation), charge and mutual interaction of humic molecules affect the NMR relaxation times $T_1$ and $T_2$. Large molecular and aggregate size and strong interaggregate interactions induce a low segmental mobility within the humic molecules and lead to slow longitudenal ($T_1$) and fast transverse ($T_2$) relaxation (Bovey and Mirau, 1996). The effects on NMR resonances are a decrease in intensity A and increase in half width $\Delta\nu_{1/2}$, possibly beyond recognition. Differential relaxation within the polydisperse humic matrix interferes with a quantitative NMR analysis, which is based on the determination of resonance position and intensity in both one- and two dimensional NMR spectra. The extended duration of the two dimensional NMR experiments makes them more susceptible to effects of variance in relaxation times $T_1$ and $T_2$. Owing to extensive signal overlap and to fast relaxation, coupling constants are not commonly utilized in the analysis of one dimensional (1D) NMR spectra of humic substances. Instead, calculation of partial integrals

according to coarse regions of substructures is usually performed (Preston, 1996; Bortiatynski et al., 1996). This modus of interpretation resembles the state of the art of protein $^{1}$H NMR spectra analysis in the late fifties (Saunders et al., 1957).

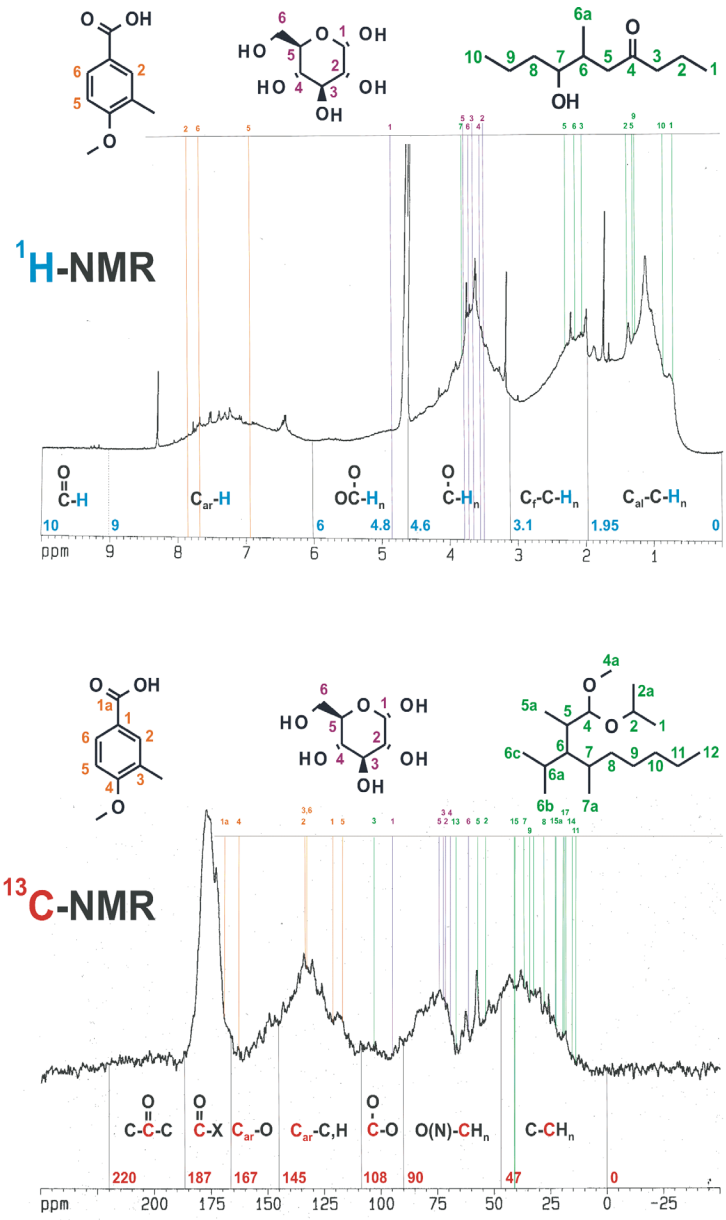



**Figure 35.** $^{1}$H NMR spectrum (top) and $^{13}$C NMR spectrum of a terrestrial NOM material with section integral limits as shown. The superior resolution of different chemical environments as indicated by the chemical shift values of typical small organic molecules is lost in NOM due to severe resonance overlap. The capability to directly observe quaternary carbon atoms in combination with an increased peak capacity of $^{13}$C NMR spectroscopy (ratio of total bandwidth / individual linewidth) relative to that of $^{1}$H NMR (in the analysis of NOM almost one order of magnitude larger) allows to observe better resolved structural detail in $^{13}$C NMR spectra of NOM, albeit at a lesser signal to noise ratio (at natural abundance, the NMR receptivity of $^{13}$C nuclei is 1/5682 of proton nuclei).

Two dimensional (2D) NMR spectroscopy offers several significant advantages in the analysis of the complex humic structure. An increased signal dispersion into two frequency dimensions greatly reduces resonance overlap, especially in the heteronuclear correlated NMR spectra with their large spread of frequency in F1 (e.g. $^{13}C$) dimension (Cavanagh et al., 1996; Croasmun and Carlson, 1996). In addition, 2-D NMR experiments act as filters which emphasize specific forms of binding. So, heteronuclear single bond correlated spectra like HSQC and HMQC show exclusively directly bonded pairs of carbon and hydrogen atoms, whereas standard homonuclear COSY spectra exhibit geminal (two-bond) and vicinal (three-bond) proton-proton couplings only. The sensitivity of proton detected homo- and hetero-nuclear 2-D NMR spectra in general is much higher than that of X-nuclei 1-D NMR spectra. Above effects combined enable a detailed analysis of even the most crowded regions of chemical shift.

Cross peaks in two dimensional NMR spectra indicate a range of connectivities defined by the kind of NMR experiment performed, allowing to probe bonding interactions, spatial relation-ships and chemical exchange (Cavanagh et al., 1996; Croasmun and Carlson, 1996; Hertkorn, et al., 2001). Various aspects of high-resolution 2D NMR spectroscopy of HS have been highlighted in several recent publications. With one publication focusing on the use of NOESY spectra to provide information about spatial arrangement of aromatic and aliphatic units in HS (Chien and Bleam, 1998) another one (Nanny and Maza, 2001) describing homo-nuclear J-resolved spectra of HS, most of the other publications demonstrate that a considera-ble range of shift correlated 2D NMR spectra is applicable to the analysis of HS. So, (Schmitt-Kopplin et al., 1998) demonstrated that information from 2D NMR spectra (HMQC and COSY) can reveal structural information about the course of photodegradation of a soil HS. (Wang et al., 1998) showed a range of well resolved 2D NMR spectra (MAXY, TOCSY) with some preliminary assignments. Haiber et al., 1999 showed HMQC and COSY NMR spectra of ultrafiltrate fractions of an aquatic humic substance, which were differing in cross peak location, but were characterized by limited resolution of structural detail. Kingery et al., 2000 focused on HMQC and TOCSY NMR spectra of an IHSS soil HA and tentatively assigned selected HMQC and TOCSY cross peaks to fatty acid derivatives. They demon-strated that additional TOCSY cross peaks could be observed when using DMSO-$d_6$ instead of HDO as a solvent and attributed those to C$\alpha$-NH correlations within amino acid side chains. Fan et al., 2000 performed NOESY, TOCSY and HSQC NMR spectra for a mollisol

HA treated with the chelating agent Tiron and provided substantial assignment details of peptidic and some carbohydrate cross peaks. They compared these results with TOCSY cross peaks appearing in the acid digest of the HA used. Hertkorn et al., 2001 demonstrated the applicability of a range of homo- (COSY, TOCSY, EXSY) and heteronuclear NMR experiments (HSQC, DEPT-HSQC) for the study of HS structure and provided initial cross peak assignments. It was concluded that the NOESY spectrum in dry DMSO-$d_6$ could be regarded rather as an EXSY-NMR-spectrum, since contributions from chemical exchange significantly outperformed those from spatial relationships.

The objectives of the presented study are to (1) assess the information content on the partial structures of HS available from the interpretation of single 1D or 2D NMR spectra and from a combined analysis, (2) identify and compare the partial structures of FA and HA isolated from the same peat source, (3) assess the suitability of the "biopolymer degradation" (BD) and the "abiotic condensation" (AC) models to explain the process of humification

### *3.2.3. Results and Discussion*

#### *3.2.3.1. 1D NMR spectra*

The proton NMR spectra of FA and HA (Figure 36) at first glance closely resemble each other. Both FA and HA exhibit prominent resonances in the carbohydrate region and the overall lineshape within this region is rather similar with an apparently reduced average linewidth of the individual resonances in FA. A more distinct variation of the resonance pattern is obvious in the region of functionalized aliphatic resonances. The pronounced resonance maximum observed in the aromatic region of FA (6.77 ppm) is also present in the HA spectrum, but is strongly attenuated. HA features a prominent upfield resonance (0.84 ppm) displayed with much less intensity in FA, probably representative of methyl groups terminating aliphatic units. Strong upfield signals around 1.2 ppm occur both in FA and HA, but with a slightly displaced shift to higher field in HA which also appears to be composed of a larger variety of several overlapping resonances.

Integration of the $^1$H NMR spectra of FA and HA according to specific regions of chemical shift provides the amounts of non-exchangeable protons shown in Table 10. About 25 % of the non-exchangeable protons of HA are present in non-substituted aliphatic groups, whereas in FA this value reaches only 14 %.

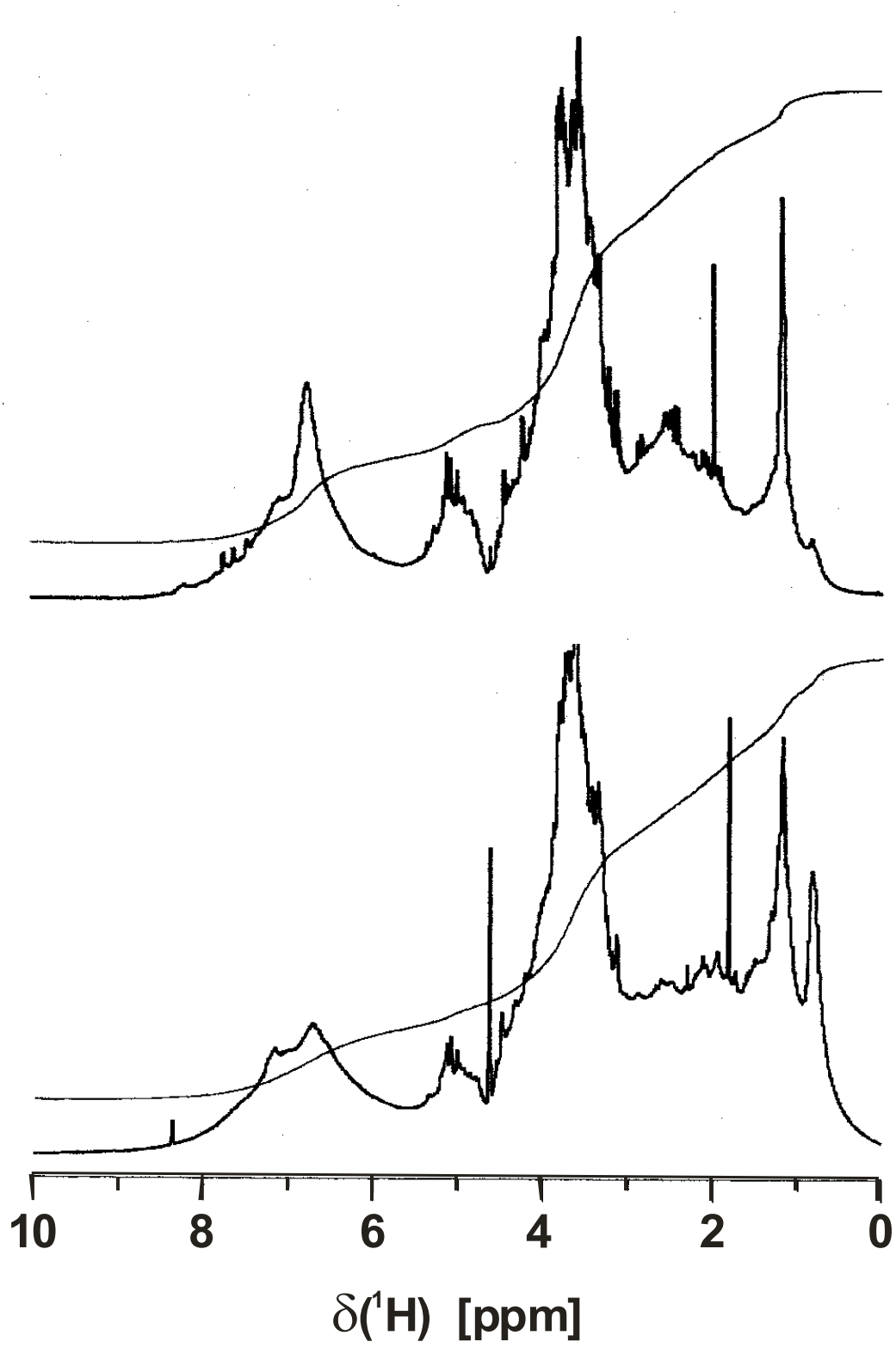



**Figure 36.** $^{1}$H NMR spectrum of FA (top) and HA (bottom).

**Table 10.** Distribution of non-exchangeable hydrogen among different structural groups as determined from $^{1}$H NMR spectra of peat HA and FA

| range of δ($^{1}$H) [ppm] | assignment | proportion of hydrogen in structural groups [%] | |
|---|---|---|---|
| | | **FA** | **HA** |
| 10.0 - 6.0 | $C_{ar}H$ | 18.2 | 15.6 |
| 6.0 - 4.7 | Acetal | 4.5 | 6.8 |
| 4.6 - 3.1 | $CH_nO$ | 45.0 | 35.6 |
| 3.1 - 1.95 | $C_f$-$CH_n$ | 18.1 | 17.1 |
| 1.95 - -0.5 | $C_{al}$-$CH_n$ | 14.2 | 24.9 |

In general, lesser-substituted aromatics and carbohydrates appear to be more relevant constituents of FA than of HA, while unfunctionalized aliphatics are more prominent features of HA. This is confirmed by the appearance of the 1D $^{13}$C NMR spectra (Figure 37).

The $^{13}$C NMR spectra of FA and HA are more distinctively different than the corresponding $^{1}$H NMR spectra. The spectra differ significantly in all regions except the (so called) carbohydrate section ranging from 60-90 ppm. HA shows a higher fraction of purely aliphatic resonances (δ < 50 ppm) and also more detail and resolution. Strong methoxyl resonances at δ = 53 (methyl esters) and 57 ppm (aliphatic methyl ethers) occur in HA, while in FA only a faint resonance at δ = 57 ppm is visible. The aromatic section within $^{13}$C NMR spectrum of FA exhibits two relatively sharp resonances at δ = 117 and 132 ppm and a relatively narrow phenolic resonance at δ = 156 ppm, while in HA these two high field resonances are superimposed to a relatively broad hump ranging from 95-160 ppm. In HA a distinct range of resonances appears downfield of a intensity minimum at 143 ppm in the phenolic region from 150-160 ppm. The intensity distribution within the carboxylic region of HA and FA is different. Due to low intensity and dependence on small base line distortions the integration of the carbonyl region (220-187 ppm) cannot be performed with very good accuracy.

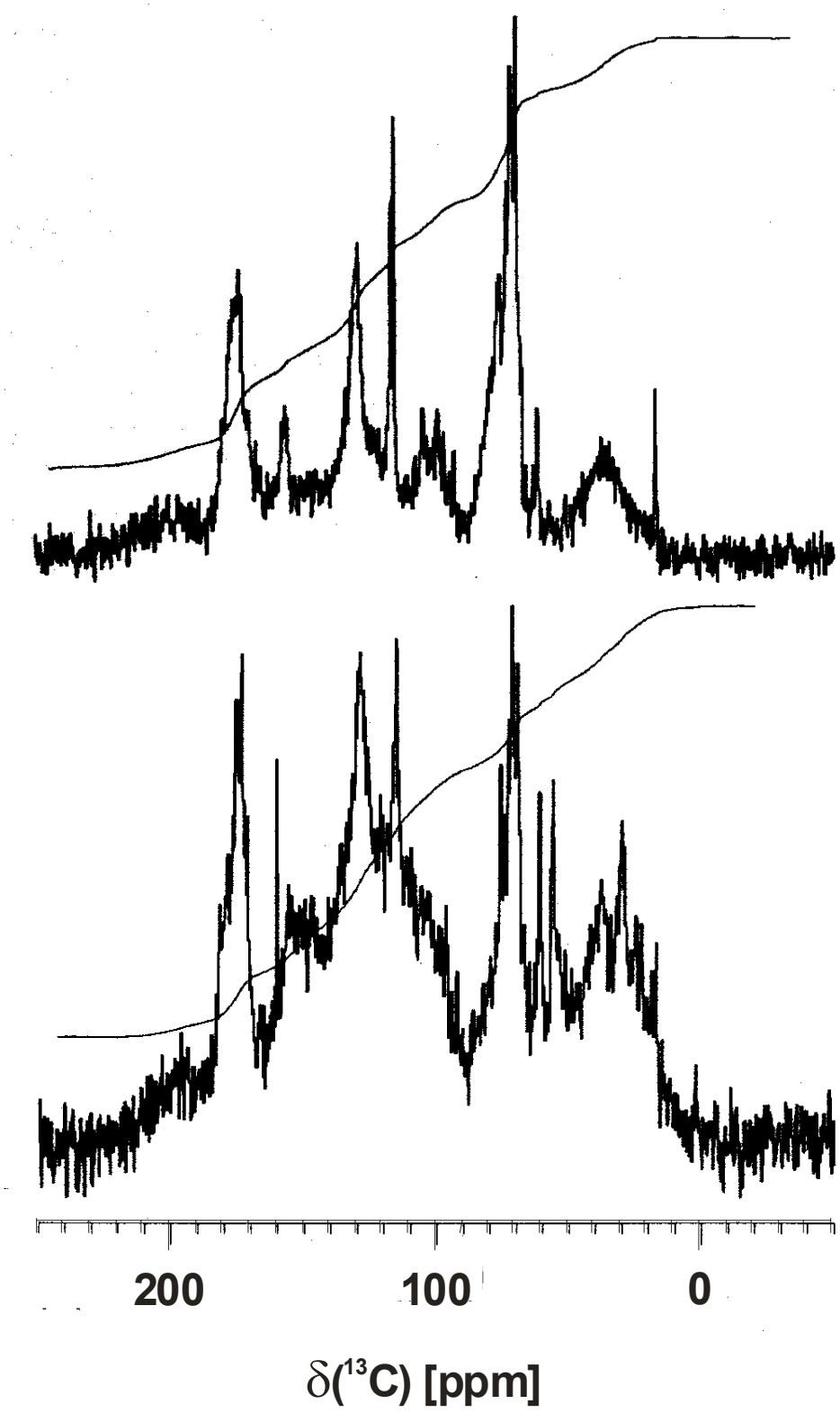



**Figure 37.** $^{13}C$ NMR spectrum of FA (top) and HA (bottom).

The resonance integrals of seven common spectral regions (Table 11) indicate for HA a reduced content of carboxylic carbons and an enhanced content of phenolic groups. In HA the integral value of the anomeric region is affected by contributions from overlapping high-field aromatic resonances, and therefore there is no good correlation with the CHO carbohydrate integral. As expected, the amount of non-polar aliphatic and C–H and C–C aromatic carbons is much higher in HA than in FA. The carbohydrate content determined on the basis of the 47-90 ppm region is found to be substantially higher in FA. These $^{13}C$ NMR integral differences agree well with data of elemental analysis, which show a higher content of carbon for HA and a nearly equal content of hydrogen. This uniformity in hydrogen content is due to compensating effect when in (e.g.) HA both contents of aromatics (with a low H/C ratio) and pure aliphatics with a high H/C ratio are increased relative to FA. The found content of carbohydrates is considered as relatively high for peat (Preston et al., 1989), this is indicative of highland peats.

**Table 11.** Distribution of carbon among the main structural groups of peat FA and HA as determined from their 1D $^{13}C$ NMR spectra

| range of $\delta(^{13}C)$, [ppm] | assignment | Proportions of carbon in structural groups [%] | |
|---|---|---|---|
| | | **FA** | **HA** |
| 220-187 | C=O | 5.9 | 3.8 |
| 187-167 | (C=O)-X- | 14.2 | 10.8 |
| 167-145 | $C_{ar}$-O,N | 8.6 | 10.1 |
| 145-108 | $C_{ar}$-C,H | 25.3 | 28.7 |
| 108- 90 | O-C-O | 8.4 | 8.4 |
| 90- 47 | $O(N_f)$-$CH_n$ | 27.4 | 22.2 |
| 47- 0 | $CH_n$ | 10.2 | 16.0 |

As seen from Table 11, the relative content of aromatic carbon in HA is 1.15 referred to that of FA, but the ratio of aromatic protons in the $^{1}H$ NMR spectra of HA and FA is 0.85 (Table 10). That implies that the average degree of non hydrogen substitution of aromatics in FA is by one quarter less than in HA. From the intensity distribution within the proton NMR spectrum it can be deduced that phenols (and possibly five membered heterocyclic rings) in FA are in the average even less substituted.

### *3.2.3.2. COSY NMR spectra*

The COSY NMR spectrum of FA exhibits more numerous cross peaks than that of HA (Figure 38). With a few exceptions all of the COSY cross peaks visible in HA also occur in FA. Most of the total cross peak intensity shows up in regions of chemical shift indicated in Table 12.

**Table 12.** Regions of chemical shift in the COSY NMR spectra of FA and HA

| region | F2 [ppm] | F1 [ppm] | Assignment | humic constituent |
|---|---|---|---|---|
| $A_C$ | 4.4 - 3.0 | 1.4 - 1.0 | -C-CH-CH-O- | desoxy sugars, ethers, esters |
| $B_C$ | 3.2 - 1.8 | 3.2 - 1.8 | $-C_f$-CH-$C_f$H-C- | intra functionalized aliphatic, ß to heteroatoms |
| $C_C$ | 4.5 - 3.0 | 4.5 - 3.0 | -CH(O)-CH(O)- | intra carbohydrate, without anomerics |
| $D_C$ | 5.0 - 4.4 | 4.4 - 1.8 | -CH(O)-CH-$C_f$ | functionalized aliphatic unit with one heteroatom |
| $E_C$ | 5.6 - 4.8 | 4.4 - 3.0 | -O-CH(O)H-CH | intra carbohydrate, with anomerics |
| $F_C$ | 8.5 - 6.3 | 8.5 - 6.3 | -$C_{ar}$H-$C_{ar}$H- | ortho protons in aromatic rings |

FA shows significant cross peak intensity in the region $D_C$ with corresponding cross peaks barely visible in HA. These cross peaks show up in the same region as those of Hα-Hβ in peptides (cf. Figure 39), but follow a distinctively different pattern. Therefore, they are expected to result not from peptides, but from highly oxygenated side chains of ether and ester type. Correlations in the F1 upfield section originate from α-carbonyl and other functionalized aliphatics while those at higher values of F1 chemical shift are in α-positions to heteroatoms. Strong cross peak intensity resulting from intra carbohydrate coupling is found in the regions $C_C$ both in FA and HA.

Although typical HS show a high degree of aromatic substitution (4-5 per ring), COSY cross peaks in region $E_C$ indicate the occurrence of less substituted aromatics with vicinally coupled protons (1, 1,2; 1,3; 1,4; 1,2,3; 1,2,4; 1,2,5 and 1,2,3,4) which are characteristic of plant phenolics and have been found also in materials at early stages of humification (Simpson et al., 1997).

The COSY cross peak intensity per weight unit sample of FA, normalized to identical number of scans, is considerably larger than that of HA. This less pronounced self cancellation of antiphase cross peak intensity is expected when individual resonances exhibit larger $T_2$ values and narrower line widths. FA apparently exhibits an extended transverse relaxation time when

compared with HA, which is probably caused by the presence of smaller, more mobile aggregates. In region $A_C$ a range of cross peaks is observed only in HA and not found in FA. This is indicative of a class of alkylated carbohydrates occurring only in HA.

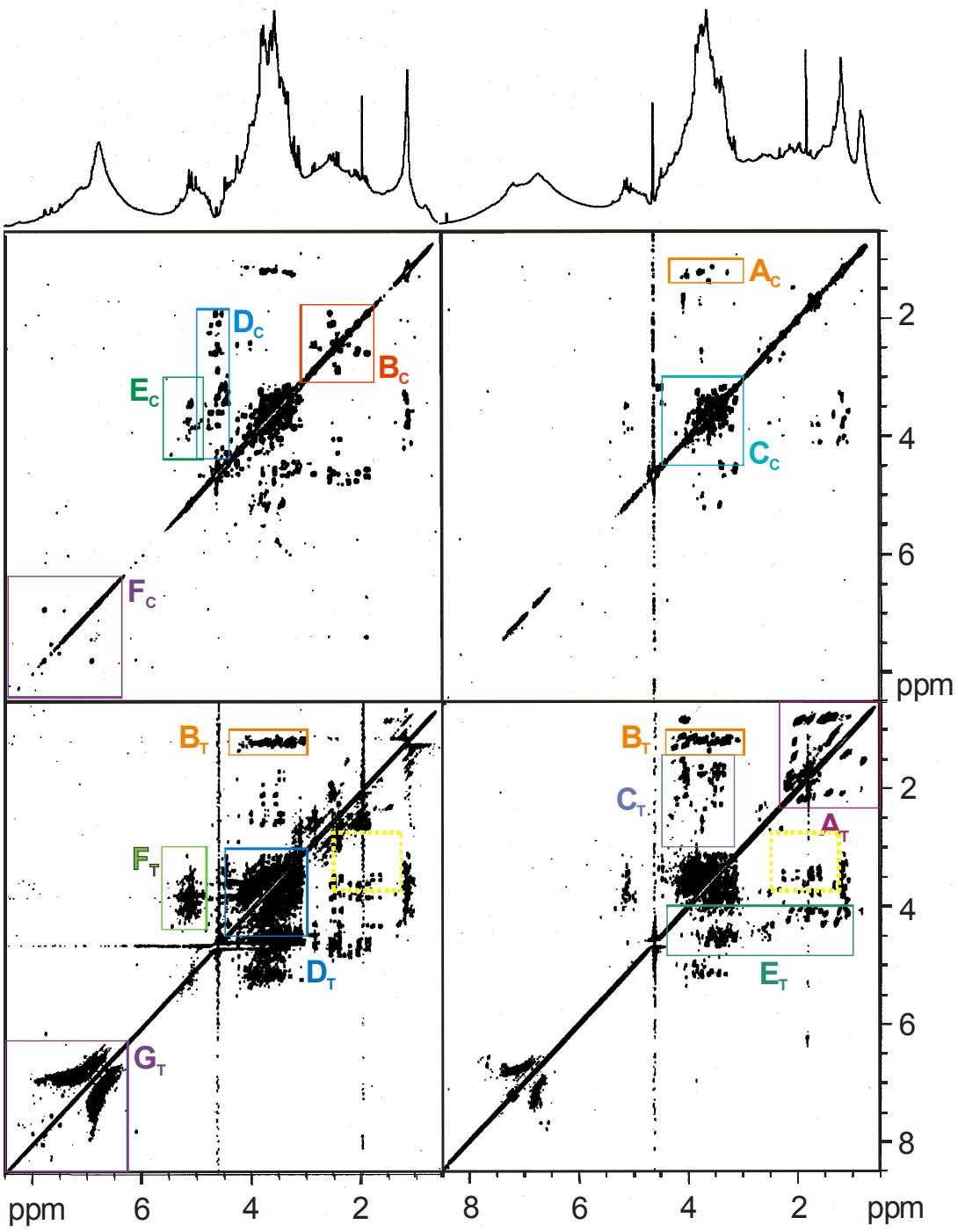



**Figure 38.** COSY (top) and TOCSY (bottom: $\tau_{mix}$ = 70 msec) NMR spectra of FA (left) and HA (right) with sections of chemical shift indicated in tables 12 (COSY) and 13 (TOCSY); for reasons of clarity only one set of boxes is shown (exception $B_T$); the dotted box indicated in the TOCSY spectra has identical chemical shift range to the dotted box within Figure 39 (cf. text).

### *3.2.3.3. TOCSY NMR spectra*

Both TOCSY NMR spectra of FA and HA (Figure 38) exhibit a higher number of cross peaks than their respective COSY spectra. Within the region $A_T$ of HA cross peaks indicate about a dozen correlations between the terminal methyl resonance ($\delta$ = 0.81 ppm) and aliphatic positions. Analogous cross peaks become visible in the COSY spectra only when calculated with soft truncating window functions in F2 and F1 such as $\pi/2$-shifted sine bells; then also aromatic cross peaks in the COSY spectra of FA and HA emerge. The cross peak location closely resembles those of intra aliphatic correlations within amino acids in peptidic bonds (cf. below). Other cross peaks not visible in COSY spectra show up in region $C_T$.

**Table 13.** Regions of chemical shift in the TOCSY NMR spectra of FA and HA

| region | F2 (ppm) | F1(ppm) | assignment | partial structure |
|---|---|---|---|---|
| $A_T$ | 2.3 - 0.5 | 2.3 - 0.5 | $-CH_n-CH_n-CH_n-CH_n$ | intra aliphatic chains (n = 1-3) |
| $B_T$ | 4.4 - 3.0 | 1.4 - 1.0 | -O-CH-CH-C- | desoxy sugars, ethers, esters |
| $C_T$ | 4.5 – 3.2 | 3.0 – 1.4 | $-N,O-CH-CH-CH_f$ | functionalized aliphatic chain with one heteroatom |
| $D_T$ | 4.5 - 3.0 | 4.5 - 3.0 | -CH(O)-CH(O)- | intra carbohydrate, without anomerics |
| $E_T$ | 4.4 – 1.0 | 4.8 - 4.0 | $-(C=O)NH-CH-CH-C_f$ | Hα-Hβ in peptides |
| $F_T$ | 5.6 - 4.8 | 4.4 - 3.0 | -O-CH(O)-CH(O)- | intra carbohydrate, with anomerics |
| $G_T$ | 8.5 - 6.3 | 8.5 - 6.3 | $-C_{ar}H-C_{ar}H-$ | ortho protons in aromatic rings |

These cross peaks correspond to functionalized aliphatic chains of intermediate length ($C_{3-5}$). In contrast to COSY cross peaks, several of these new cross peaks indicate more extended spin systems and can be recognized as ridges of resonances with equal frequencies F1 (or F2). The average number of 3-4 cross peaks per ridge points towards highly functionalized aliphatic chains with heteroatom substitution terminated by simple and functionalized aliphatics. A range of cross peaks within $C_T$ is found only in HA, but not in FA. Taking into account the improved relaxation characteristics of FA compared with those of HA, these new cross peaks represent structural features characteristic of HA which are virtually absent in FA.

In the region $E_T$, a series of cross peaks of peptide origin occurs. This region is blanked in F2 dimension by the non excitation of the NaOD resonance (with limited bandwidth), but cross peaks are clearly visible along the F1 axis. The location of these peptide cross peaks is quite different from those in the $D_C$ region in the COSY spectra of FA.

Compared with the appearance of the COSY cross peaks in HA, many more (about >30 instead of <10) correlations of anomeric positions (H1 with H2 protons) in carbohydrates are

found in the section $F_T$ within the TOCSY spectrum; FA behaves in an analogous manner. The relative intensity of these cross peaks increased when changing the TOCSY mixing time from 35 ms over 75 ms to 100 ms, showing that many of these correlations are characterized by rather small vicinal coupling constants.

In the aromatic region $G_T$ a broad range of cross peaks is found across the whole range of aromatic chemical shifts. Cross peaks with upfield chemical shifts ($\delta(^1H) < 7$ ppm) belong to a class of mono- and doubly oxygenated (ortho and para hydroxyl or methoxyl) rings. Another prominent type of substitution patterns, which is ranging from aliphatic ($\delta(^1H)$: 7…7.4 ppm) up to deshielding ortho carbonyl or carboxyl substitution ($\delta(^1H) > 7.4$ ppm) is observed both in FA and HA. The individual cross peaks present in FA and HA differ, but clearly are indicative of multiple oxygen substitution in conjunction with shift neutral aliphatic and ortho deshielding substituents such as (substituted) carbonyl groups. In general, even while scaled with the relatively low signal intensity from the 1D proton NMR spectra only marginal cross peak intensity is found in the aromatic systems. This is indicative of highly substituted aromatic systems in HS with only a minor fraction of ortho proton substituted aromatic rings present.

### *3.2.3.4. Comparative analysis of COSY and TOCSY spectra*

The COSY and TOCSY sequences represent the most important class of homonuclear 2D NMR experiments in which cross peaks arise through the interaction of the J-coupled spins. However, the different mechanisms of magnetization transfer acting in COSY and TOCSY NMR sequences (Cavanagh et al., 1996; Croasmun and Carlson, 1996; Hertkorn et al.,2001) affect cross peak intensity and lineshape in molecularly irregular and polydisperse HS. These proton 2D-NMR pulse sequences are intrinsically sensitive but characterized by rather long transfer times (TOCSY: 1/2J; COSY: 1/J) due to small geminal (two bond) and vicinal (three bond) H,H-coupling constants ($J < 15$ Hz). This extended duration makes them susceptible to signal loss in the case of fast transverse relaxation. In COSY spectra overlap of antiphase coherence causes an attenuation of cross peak intensity when J is less than or equal to the line width. Thus the disappearance of corresponding FA/HA cross peaks in the COSY relative to TOCSY spectra is thought to be caused by self cancellation of antiphase magnetization at increased linewidth ($\Delta\nu_{1/2} > 10$-15 Hz).

In TOCSY spectra phase sensitive detection of net magnetization transfer and choice of variable mixing times allow differentiation of small and extended spin systems. TOCSY spectra of HS typically show an enhanced off diagonal cross peak intensity when compared with the respective COSY spectra. However, standard COSY spectra represent vicinal and geminal coupled spin pairs only and frequently produce better resolved cross peaks in the heavily overlapping ”carbohydrate region” of HS than TOCSY spectra. Under standard conditions COSY cross peaks indicate almost exclusively vicinal couplings in aromatic rings. However, it is expected that many of the cross peaks visible only in TOCSY spectra also represent geminal and vicinal couplings between protons. Clear exceptions are cross peaks within section $C_T$ (Figure 38), and the intense TOCSY cross peak at F2/F1: 4.1/0.8 ppm, which appears at extended mixing times only. It most likely represents a $^4J$ long range correlation between a heteroatom substituted position and a methyl group.

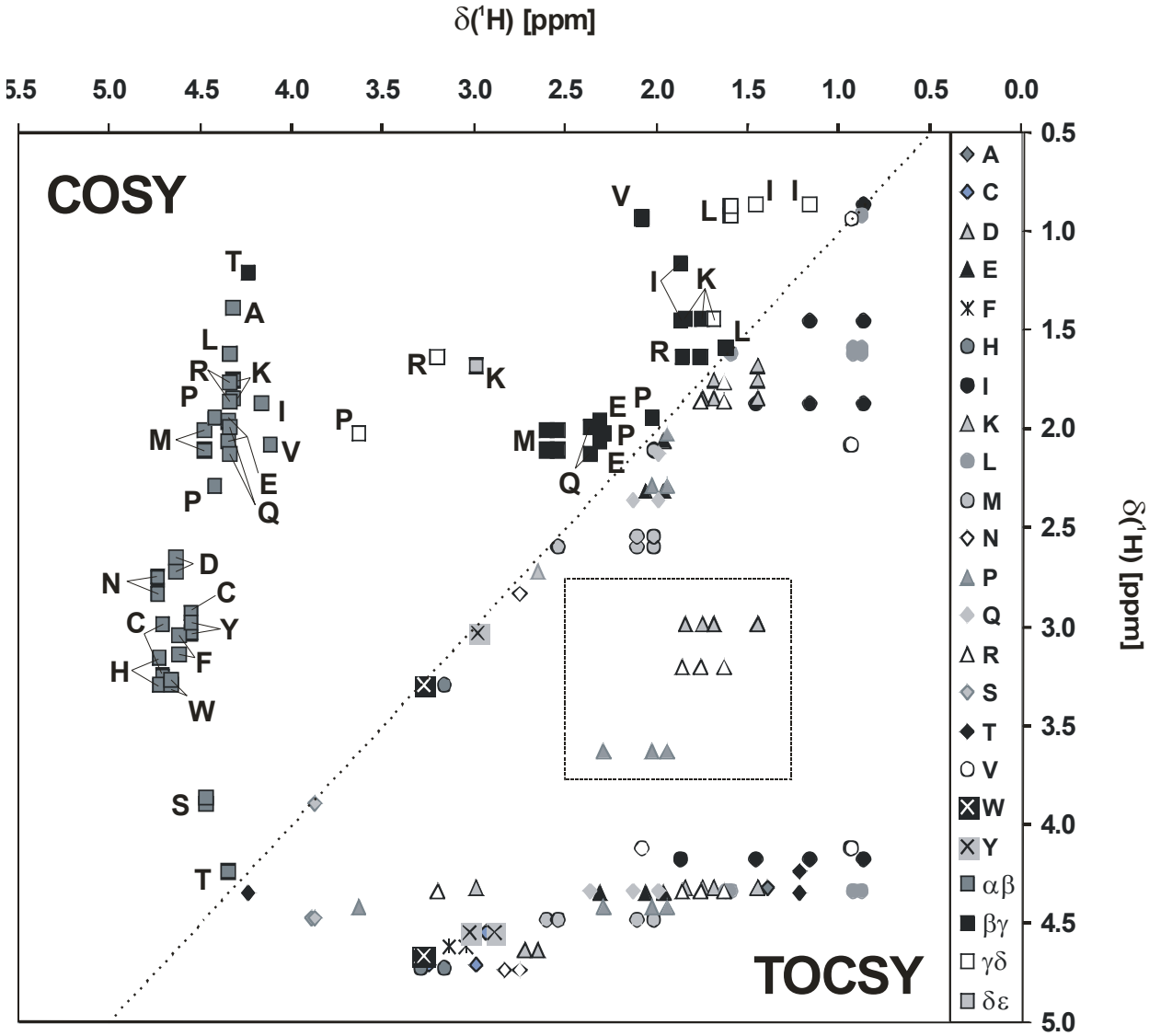



**Figure 39.** COSY (upper left half) and TOCSY cross peaks (lower left half) of proteinaceous amino acids in proteins following Alanine (A) according to chain position (COSY) and amino acid (TOCSY); cf. text (Wishart et al., 1995).

To enable an assignment of the complex spectral patterns within the sections $D_C$ and $E_T$ of the homonuclear 2D NMR spectra, which both represent heteroatom substituted CH-groups, figure 39 has been assembled from literature data (Wishart et al., 1995). It indicates both COSY and TOCSY cross peaks of the non exchangeable protons of the proteinaceous amino acids at a chemical shift, when the respective amino acid in proteins is followed by alanine (A). Along the diagonal this figure is split into two sections of equal size. The upper left half shows the peptidic cross peaks, caused only by geminal and vicinal couplings as observed in the standard COSY spectra. They are classified according to the positions within the amino

acid side chain. On the lower right half the respective TOCSY cross peaks are indicated for the complete spin systems excluding aliphatic/aromatic and γε-methionine interactions which are absent due to low transfer efficiency. The TOCSY cross peaks are classified according to the amino acids. The TOCSY spectrum of HA shows most of these peptidic amino acid cross peaks at positions significantly different from the chemical shift values of free amino acids, which are not observed. In the COSY NMR spectrum of HA these cross peaks are strongly attenuated and mostly missing, probably due to self cancellation (cf. above). Similar patterns in the $D_C$ section of the COSY spectrum of FA, at first glance looking like αβ- and βγ-amino acid correlations (cf. Figures 38 and 39), occur at significantly displayced chemical shift values. These cross peaks then most probably represent strongly oxidized and functionalized aliphatic chains with no aliphatic CH bonded solely to purely aliphatic units.

The dotted box within figure 39 is characterized by TOCSY cross peaks of the long chain amino acids lysine (K), proline (P) and arginine (R). The TOCSY and COSY cross peaks of FA and HA appearing in this section of the spectrum do not follow these patterns and are therefore not expected to originate from peptides but rather from other functionalized aliphatic groups. Of importance is that within this section a few cross peaks appear both in FA and HA, a feature not observed otherwise for non carbohydrate (and possibly a few aromatic) cross peaks. Some of these cross peaks apparently exhibit identical F1 frequencies and indicate coupled systems mostly of intermediate chain length ($C_{3-5}$). The range of chemical shift displayed by this series of cross peaks requires a single directly bonded heteroatom and other remotely functionalized groups in the chain (like, e.g. in oxopropyl groups).

In the carbohydrate region $F_T$ a redistribution of cross peak intensity occurs towards more intense H1-H2 correlations when compared with the COSY section $E_C$. The increased ash content of HA may account for some loss of cross peak intensity, especially in the COSY spectra due to increased paramagnetic relaxation and induced aggregation.

### *3.2.3.5. HSQC NMR spectra*

NMR pulse sequences which correlate directly bonded proton and carbon atoms with respect to their chemical shift show optimum transfer amplitudes in the 5-20 ms range (HSQC at $^{1}J(CH)$ = 150 Hz: 10.9 ms; Heikkinen et al., 2003) and their relatively short duration with respect to average $T_1$ and $T_2$ relaxation times in humic substances (Fründ and Lüdeman, 1989, Preston and Blackwell, 1985) is more favorable to prevent signal loss in case of fast transverse relaxation. The large spread of frequency in F1 ($^{13}C$) dimension greatly reduces resonance overlap. These effects combined make HSQC NMR spectra very auspicious candidates to reveal structural information of HS in great detail (Figure 40).

**Table 14.** Main regions of chemical shift in the HSQC NMR spectra of peat FA and HA

| region | F2($^{1}$H), ppm | F1($^{13}$C), ppm | assignment | partial structure |
|---|---|---|---|---|
| $A_H$ | 3.4 – 0.4 | 45 - 5 | C-CH, $C_f$-CH | aliphatic groups |
| $B_H$ | 5.3 – 2.6 | 87 - 37 | X-CH | single heteroatom substituted aliphatics |
| $C_H$ | 5.6 – 4.2 | 107 - 87 | O-CH-X | double heteroatom substituted aliphatics (e.g., acetals) |
| $D_H$ | 9.0 – 6.0 | 150 - 105 | $C_{ar}$-H | aromatic and heterocyclic systems |

The cross peak locations of FA and HA closely correspond to each other in the region $C_H$. More substantial variation of cross peak location and relative intensity distribution is observed in regions $D_H$ and $B_H$; while the $^{13}C$ low field section of $B_H$ ($\delta > 60$ ppm) bears significant resemblance in FA and HA most of the HA cross peaks at $\delta < 60$ ppm are missing in FA. Within section $A_H$, the cross peaks in FA and HA follow entirely different patterns. To understand the nature of the differences, the model HSQC cross peaks originating from proteinaceous amino acids in peptide bonds (Wishart et al., 1995) are displayed in Figure 41 classified according to signal multiplicity and position within the side chain.

The HSQC NMR spectra of FA and HA exhibit a substantial number of cross peaks which can be divided into 4 main regions of chemical shift according to Table 14.

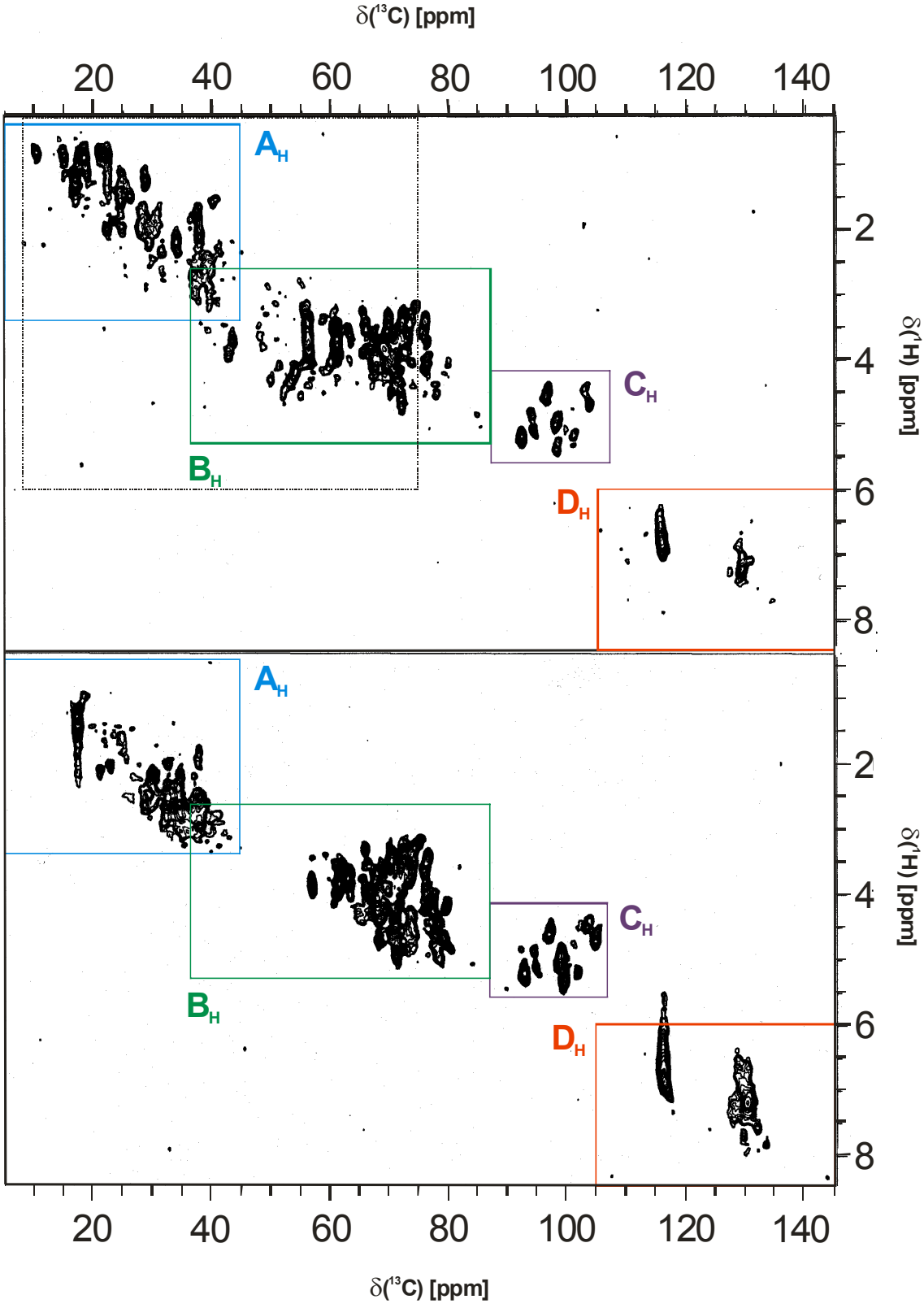



**Figure 40.** HSQC NMR spectra of HA (top=) and FA (down) with sections of chemical shift indicated in table 6; the dotted box corresponds to the range of chemical shifts shown in figure 41.

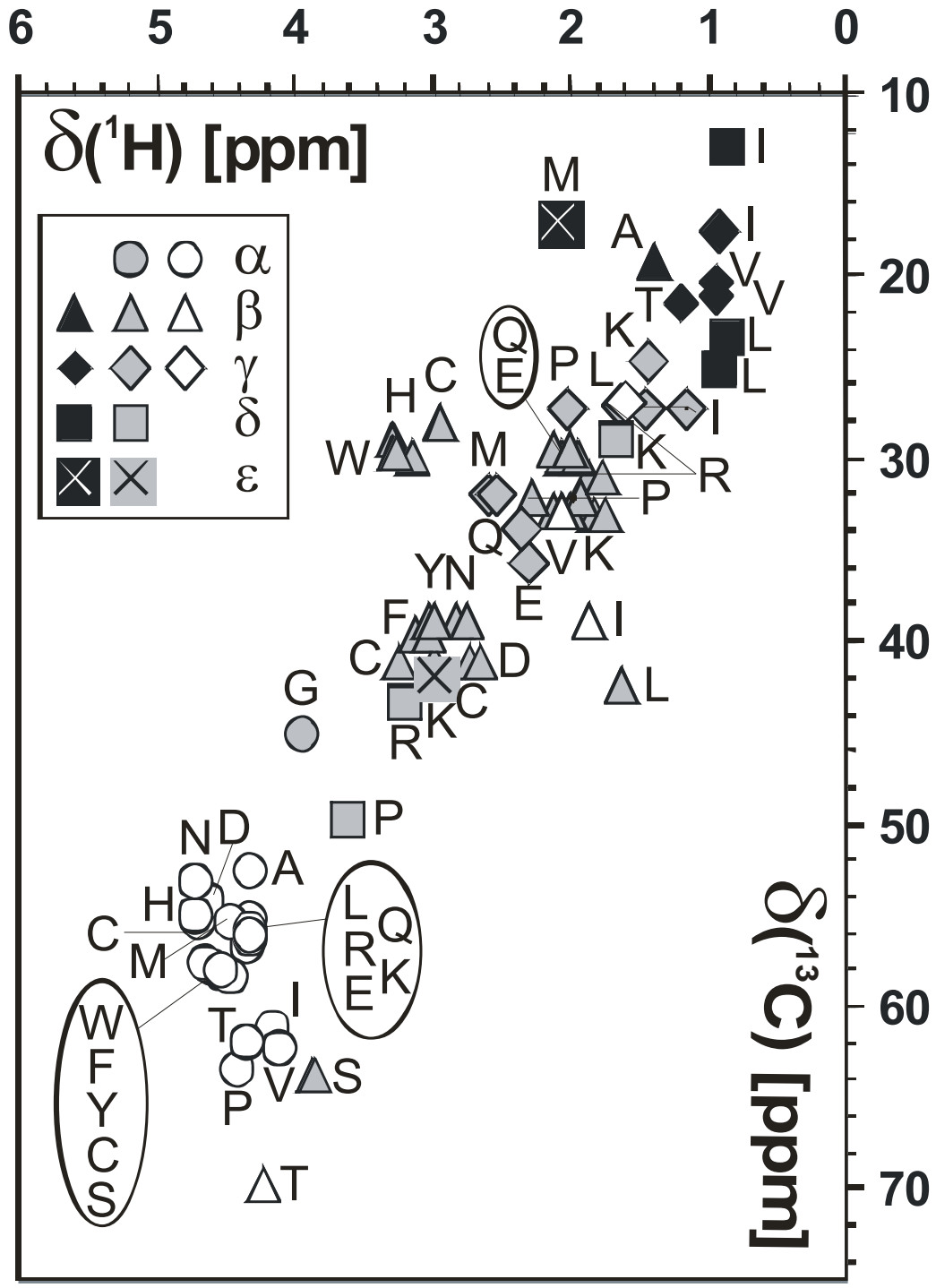



**Figure 41.** HSQC cross peaks of proteinaceous amino acids in proteins following Alanine (A), edited according to signal multiplicity (black: methyl; grey: methylene; white: methine) and position within the side chain (Wishart et al., 1995).

The dotted box in Figure 40 corresponds to the range of chemical shift shown in Figure 41. Within the section $A_H$ strong resonances of all amino acid methyl positions except methionine (M) –show up in HA, but not in FA. With the exception of the β cross peaks of the amino acids tryptophan (W), histidine (H) and cystine (C) almost all of the other β-ε cross peaks occur in HA as well as all of the α cross peaks. However, some additional HA cross peaks observed within the section $A_H$ deviate substantially from the amino acid cross peaks given in Figure 41. Those in the center of section $A_H$ are likely to originate from terpenoid hydrocarbons, but also lipids, fatty acid derivatives and aliphatic residues of desoxy sugars appear in this region of chemical shift (cf. Figure 42).

Despite the considerable nitrogen content of FA peptidic amino acid cross peaks are apparently missing in FA. Polymethylene (1.2/29 ppm) is prominent in HA but almost absent in FA. Aliphatic methoxy (3.8/57 ppm) is more pronounced in HA than in FA. Within the section $B_H$ a group of upfield $^{13}C$ cross peaks with chemical shift values around 50 pm and with proton chemical shifts lower than 3 ppm not found in FA is present in HA. Among them are strong cross peaks originating from Cα in peptides (cf. Figure 41), but also more feeble ones appear. Apart from the strong glycine CH-α cross peak (3.8/42 ppm (Hertkorn et al., 2001) these cross peaks are assigned to amino sugar moieties containing a single amino group within a hexose unit according to their characteristic position in the proton and carbon frequencies.

In the section $B_H$ the HSQC spectrum of FA and HA contains a group of cross peaks, assignable to CHOH fragments, usually considered as carbohydrates ($\delta(^{13}C)$: 67 - 85 ppm), including a significant cross peak at 3.9/62 ppm, representing C6 in hexoses. This group presents the most significant proportion of the visible resonances in FA and it is accompanied with intense and sharp correlations in the section $C_H$ corresponding to anomeric $(CH)_1$ pairs of the carbohydrates. In the 1D $^1H$ NMR spectrum these signals show partial overlap with CHOH and peptide resonances and the residual water resonance. The presence of the anomeric carbons indicates that the carbohydrates in FA and HA remain mainly intact. The location and intensity of cross peaks within the anomeric region $C_H$ as well as those within the CHOH section of $B_H$ are very similar in the peat FA and HA studied. The tentative assignments provided in table 15 are based on the location of the HSQC cross peaks within the anomeric region, which show well resolved resonances.

**Table 15.** proposed assignment of anomeric region $C_H$ of the HSQC resonances

| T5 FA | | T5 HA | | |
|---|---|---|---|---|
| $\delta(^{13}C)$ [ppm] | $\delta(^{1}H)$ [ppm] | $\delta(^{13}C)$ [ppm] | $\delta(^{1}H)$ [ppm] | **tentative assignment** |
| 91.7 | 4.83 | | | ribopyranose |
| 92.4 | 5.19 | 92.2 | 5.17 | glucuronic acid; xylose; β-glucose in oligomers; terminal (reducing end) α-glucose in oligomers |
| 93.8 | 4.85 | 93.7 | 4.81 | oligomers of α-glucose |
| 93.9 | 5.09 | 94.1 | 5.10 | fucose; glucose; ribopyranose; galactose |
| 96.6 | 4.56 | 96.1 | 4.58 | glucuronic acid; fucose; xylose; mannose; terminal β-glucose in polymers |
| | | 96.5 | 4.51 | arabinose; rhamnose; mannose; terminal (reducing end) β-glucose in oligomers |
| 98.0 | 5.06 | 97.8 | 4.99 | α-galactose and α-fucose in oligomers; oligomers of α-glucose |
| 98.7 | 5.36 | 97.9 | 5.38 | ribopyranose |
| 98.7 | 5.22 | 98.0 | 4.89 | lyxose |
| | | 98.3 | 5.23 | α-galactose in oligomers |
| 100.5 | 4.75 | 99.5 | 5.07 | mannose and fucose in oligomers |
| 101.0 | 5.19 | 100.9 | 5.13 | galactose and fucose in oligomers |
| | | 100.7 | 5.31 | α-glucose at branch point in polymers |
| 102.0 | 4.48 | 102.7 | 4.44 | cellulose fragments*; β-galactose and β-glucose in oligomers |
| 104.3 | 4.71 | 103.5 | 4.67 | cellulose fragments* |

*(Nehls et al., 1994)

The positions of the cross-peaks slightly deviate from the published values for the free carbohydrates, since the carbohydrate moieties are bound to the humic core and apparently comprise a part of oligomeric chains. The intense and sharp resonances of rhamnose and fucose C1 carbon atoms are in good agreement with the presence of correlations assigned to $CH_3CHO–$ fragments, found in the $A_C/B_T$ section within the COSY/TOCSY spectra of FA and HA. Additional small variations in cross peak location can be attributed to variable contributions from overlapping resonances.

Within the aromatic section two major cross peaks occur: an upfield resonance (6.8/116 ppm) representative of an ortho or para oxygen substituted and a set of a di-ortho aliphatic or hydrogen substituted aromatic ring (7.15/130 ppm). The latter cross peak is split into various resonances, at different positions in FA and HA. Downfield proton resonances ($\delta(^{1}H)$ >7.5 ppm), which are prominent features in the HSQC NMR spectrum of FA indicate at least one deshielding ortho carbonyl derivative substituent in the ring. In HA a class of upfield $^{13}C$ resonances at $\delta(^{13}C)$ < 114 ppm not visible in FA occurs. Oxygenated aromatic rings, like phenols and oxygenated heterocycles show a substantial fraction of aromatic carbon upfield resonances.

*3.2.3.6. Comparison of the experimental and model HSQC spectra of HS*

To support the given detailed assignments, the model HSQC spectrum of HS was calculated using a priori information on the molecular building blocks of HS deriving from the original natural products: aliphatic and terpenoid hydrocarbons, alkyl aromatics, carbohydrates, fatty acids and their esters, lignin and tannin derivatives (polyphenols), lipids, aromatic nitrogen compounds, peptides and proteins, phenols, suberins, respectively (De Leeuw and Largeau, 1993; Kögel-Knabner et al., 1992; Jahnel et al., 1998; Schulten and Gleixner, 1999).

All of these building blocks are characterized by specific ranges of chemical shift in their 1D $^{1}H$ and $^{13}C$ NMR spectra as shown in figure 42. In many regions at a given chemical shift significant overlap is observed from at least several constituents both in $^{1}H$ and $^{13}C$ NMR spectra. This superposition is more pronounced in the center and the shielded (upfield) sections. This ambiguity precludes an unequivocal resonance assignment according to the humic constituent specified above based solely on the chemical shift in 1D NMR spectra. So, the so called "carbohydrate" region of proton chemical shift (cf. number seven in Figure 42) features resonances originating from carbohydrates, lignin side chains, peptides, ethers and esters. The significant spread in $^{13}C$ frequency allows a separate identification of several classes of these constituents according to specific ranges within the $^{1}H$, $^{13}C$ HSQC NMR spectrum. As an example, CH-α positions of peptides exhibit significant deshielding in proton chemical shift but relative shielding in $\delta(^{13}C)$ and show therefore a clustering in a region which is not occupied by other humic constituents. Methoxy resonances from lignin occupy a rather restricted region of average proton chemical shift and are displaced to higher field in the carbon frequency, when compared to peptidic CH-α positions. Compared to normal carbohydrates, CH in α-position to amino groups in amino sugars are significantly shielded both in $^{1}H$ and $^{13}C$ frequencies. Phenylpropanoid lignin side chains occupy a chemical shift range between peptidic CH-α and carbohydrate methine positions and another extended region between carbohydrate ring and anomeric resonances. Heterocyclic nitrogen compounds are characterized by maximal downfield shift of $^{1}H$ and $^{13}C$ frequencies. Depending on the degree of α, β, and γ carbon substitution, terpenoid hydrocarbons occupy a considerable range of upfield carbon shifts, but retain a most upfield position in the proton frequency. Strong clustering of specific chemical environments is expected for tannin (C ring

linkage: 4.8/37 and C5 in A ring: 5.9/96 ppm), methyl ether (3.8/55 ppm), anomeric positions in carbohydrates (105-96/5-4 ppm) and suberin (2.55/45 ppm), to name a few.

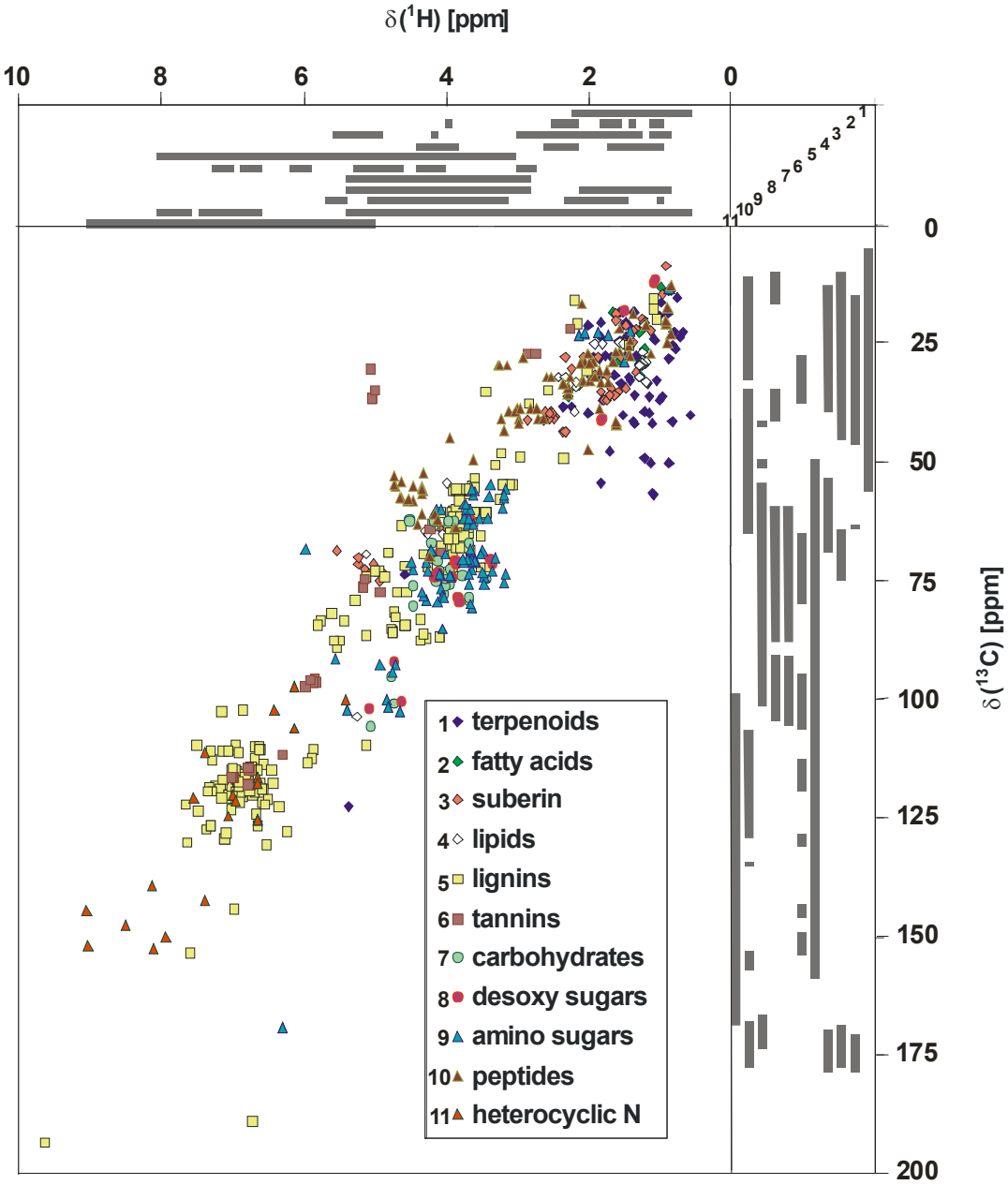



**Figure 42.** chemical shift ranges of eleven humic constituents with respct to their one dimensional $^{1}H$ and $^{13}C$ NMR and their $^{1}H,^{13}C$ HSQC NMR spectra.

Compared to the appearance of the model HSQC NMR spectrum (Figure 42), the HSQC NMR spectrum of HA (Figure 40) is dominated by resonances from a few humic constituents, namely by those of peptides (sections $A_H$ and $B_H$), carbohydrates (sections $B_H$ and $C_H$), polymethylene (section $A_H$) and aliphatic methoxyl (section $B_H$). In the aromatic section $D_H$ of both FA and HA two major groups of cross peaks dominate of non protein origin, which most probably represent an assembly of oxygenated aromatic systems. While aliphatic methoxy and carbohydrate resonances almost exclusively make up the section $B_H$ of FA, the aliphatic section $A_H$ exhibits cross peaks resembling lipid structures. At lower countour levels many more cross peaks appear, which can be traced to other humic constituents.

### *3.2.4. Conclusions*

The proportion of the various sources of organic matter in humic substances is related to maturity, depositional environment and degree of degradation (Lu et al., 2000). The combined analysis of several 2D NMR spectra provides superior resolution of structural detail when compared with an exhaustive analysis of one dimensional NMR spectra alone. Here, differences in structure and composition are more clearly revealed in $^{13}C$ NMR spectra than in $^{1}H$ NMR spectra.

The very good resolution of structural detail by a combined analysis of all NMR spectra shows that the effect of the fractionation procedure on the composition and chemical structure of humic materials is very significant.

Both homo and heteronuclear 2D NMR spectra demonstrate that the carbohydrate structures in FA and HA are remarkably similar, but carbohydrates occupy an increased share within FA. A more detailed resonance assignment is possible starting with an analysis of anomeric cross peaks in the HSQC NMR spectrum.

As visible in the $^{13}C$ NMR spectrum, the aromatic carbon fractions of FA and HA are clearly different, but the aromatic part of HA contains features found in FA, such as the rather strong resonances at 117 and 132 ppm and the strong HSQC cross peaks derived therefrom which resemble resonances typical of lignin units. The phenolic section in FA appears to be specific of FA and those resonances are not prominent in HA. The aromatic region $D_H$ in HA and FA is not dominated by cross peaks from aromatic amino acids in peptidic bonds.

However, despite differences in cross-peak location and intensity similarity of the basic aromatic substitution patterns of HA and FA is obvious from the 2D NMR spectra: highly substituted aromatic rings dominate, such as multiply oxygenated rings, which induce strong upfield shifts from ortho and para positions in $\delta(^1H)$ and $\delta(^{13}C)$. Alkyl substituted rings generate the strong HSQC crosspeak at $\delta$ = 7.2/129 ppm. Cross peaks with strongly deshielded proton resonances ($\delta(^1H)$ > 7.5 ppm) are caused by ortho substitution with carbonyl derivatives which are characterized by a chemical shift anisotropy with strong proton deshielding effect (Günther, 1992). Cross peaks originating from the broad hump, observed in the $^{13}C$ NMR spectrum of HA will be difficult to observe under standard conditions, since the average degree of proton substitution of aromatic rings within this section will be disproportionally low and so will be the scaling factor of 2D NMR cross peaks. Furthermore, the transverse relaxation time within these extended aromatic systems is expected to be fast (Fan et al., 2000).

The composition of the aliphatic part in FA and HA is fundamentally different. Differences in lineshape are already visible from the one dimensional NMR spectra even when showing almost coinciding amounts of functionalized aliphatic compounds, but the HSQC cross peak location within the whole range of the aliphatic section has almost no correspondence. FA is deprived in the typical aliphatic constituents of HA such as polymethylene, peptide side chains and terpenoid hydrocarbons and does not exhibit readily identifiable components. FA also lacks other methyl groups in extended aliphatic systems and the aliphatic region is mostly comprised of functionalized aliphatic units. Long chain alkyl and highly branched unsubstituted alkyls are nearly absent in FA; i.e. when branching occurs in aliphatic units within FA, a heteroatom or an aromatic ring will be typically not further than 2-3 bonds away from the terminal, but also from any other aliphatic carbon. The very strong proton NMR signal at $\delta$ = 1.17 in FA is not caused by polymethylene (which is present only in HA at $\delta$ = 1.22/29.6 ppm) since it is connected to a carbon atom with $\delta$ = 18 ppm. This aliphatic resonance shows a series of strong cross peaks in the COSY ($A_C$) and TOCSY ($B_T$) spectra, indicating that this resonance could be caused by a series of methylated carbohydrates. A part of the low field $^1H$ and $^{13}C$ cross peaks within section $A_H$ (centered around 2.7/37 ppm) is possibly caused by lipids, which then also would produce a share of the cross peaks within section $D_C$ of the COSY spectrum.

HA shows prominent peptidic cross peaks both in homo- and heteronuclear 2D NMR spectra and so, a significant fraction of the upfield resonances in the NMR spectra of HA can be attributed to peptide side chains. Terpenoid hydrocarbons, which are characterized by a considerable fraction of isolated methyl groups not coupling to other protons (Buddrus and Lambert, 1995; Lambert and Buddrus, 1996) and carbon resonances shifted to higher field ($\delta$ ($^{13}$C) < 25 ppm (Hertkorn et al., 2001); are found mainly in HA and not in FA. Aliphatic methoxyl resonances, which also do not couple to other protons in standard homonuclear shift correlated 2D NMR spectra exhibit a much stronger cross peak in the HSQC spectrum of HA than in that of FA.

With respect to the chemical environment of nitrogen an analysis of the TOCSY and HSQC spectra reveals pronounced differences in FA and HA even when those are characterized by similar values of nitrogen content and C/N ratio (FA: 28.0, HA: 27.1). Peptidic amino acid cross peaks present in the TOCSY and HSQC NMR spectra of HA are not found in FA. Some HSQC cross peaks of minor intensity in the 3.3/50 range may be representative of amino sugars and are present in HA only. With clear signs of amino sugars and amino acid residues being absent in the NMR spectra the chemical environment of nitrogen in FA is not readily recognizable. In absence of amino sugars and peptides, amine and heterocyclic nitrogen atoms (indol, pyrrol and pyridine derivatives) are likely candidates for nitrogen structures in FA.

From the obtained NMR results on the peat humic materials studied FA appears to be diagenetically downstream of HA in line with the biopolymer degradation (BD) model of humification. HA consists of an aromatic core with some distinct features resembling lignine structures. Both carbohydrates and peptides are part of this structure, possibly enriched at the surface and therefore more mobile and more easily to detect by NMR (Fan et al., 2000). Carbohydrate-lignin complexes (Shevchenko and Bailey, 1996) intimately associated by strong hydrogen bonds may survive partly during degradation; this could explain the close resemblance of carbohydrates observed in FA and HA. Polymethylene and terpenoid hydrocarbons are other aliphatic constituents of HA. In FA both nitrogen and aliphatic structures are more difficult to evaluate on the basis of molecularly characterized structures. FA seems to comprise a very far processed stage of humification.

## 3.2.5. MATERIALS AND METHODS

### *3.2.5.1. Isolation of the humic materials used*

The peat HA and FA were isolated from *Eriophorum* highland peat (*Eriophorum vaginatum*: 80%, pine skin: 10%, *Sphagnum* moss: 10%, decomposition rate: 40%) according to the procedure described by Lowe (Lowe, 1992). The homogenized peat sample was debituminized by treating 3 - 4 times with benzene:ethanol mixture (1:1, v/v) and air–dried. The humic fraction was extracted with 0.1 M NaOH and desalinated using cation-exchangers. To isolate HA, the obtained solution was acidified until pH 2. The precipitated HA were centrifuged, dialysed and lyophylized. To isolate FA, the supernatant was discharged through the Amberlite XAD-2 resin. The sorbed FA were eluted with 0.1 M NaOH, desalinated and lyophylized.

### *3.2.5.2. Elemental analysis*

Elemental analyses were performed on the elemental analyzer Carlo Erba Strumentazione 1106 and calculated on ash- and water-free basis. The contents of elements were [%, mass.]: FA: C - 49.0, H - 4.2, N - 2.0, O - 44.8; HA: C - 55.2, H - 4.3; N - 2.4, O - 38.1.

### *3.2.5.3. NMR spectroscopy*

NMR spectra were acquired with a Bruker DMX 500 NMR spectrometer (Rheinstetten, Germany) operating at 500.13 MHz proton frequency. The spectra were recorded on peat HA and FA samples dissolved in 700 μL 0.1 M NaOD at concentrations of 25 mg/mL. All proton detected NMR spectra were acquired by use of a 5 mm inverse geometry broadband probehead equipped with an actively shielded z-gradient coil ($90^o(^1H)$ = 9.3 μs; $90^o(^{13}C)$ = 9.8 μs). 1-D $^1H$ NMR spectra were recorded using the first increment of the presat-NOESY sequence (aq: 4.68 s, relaxation delay: 1.5 s, mixing time: 0 ms, exponential line broadening: 0.3 Hz). $^{13}C$ NMR spectra were acquired with a 5 mm broadband probe, using inverse gated WALTZ-16 decoupling (relaxation delay: 8 s at $90^o(^{13}C)$ = 9.4 μs; 28672 (FA), 51968 (HA) scans) with an acquisition time of 250 ms (SW = 65360 Hz) and an exponential line broadening of 25 Hz. The duration of relaxation time was set to 8 s on the basis of our previous investigations (Kovalevskii et al., 2000) which showed that time delays of 8 s or longer provided

complete relaxation of quaternary carbon nuclei in HS. Hence, $^{13}C$ NMR spectra acquired under such conditions can be considered as quantitative.

HSQC (F2 ($^{1}H$): acquisition time AQ = 205 ms at spectral width SW = 5000 Hz, one bond coupling constant $^{1}J(CH)$ = 150 Hz, relaxation delay D1 = 1.5 s; number of scans NS = 256; F1 ($^{13}C$): SW = 22637 Hz (180 ppm), 256 increments) spectra were decoupled [$^{13}C$: GARP (70 μs)] and calculated to a 2k x 512 matrix with typical window functions in F2 (exponential line broadening: 30 Hz) and shifted sine bell (π/6) in F1. Gradient (length: 1 ms; recovery: 450 μs), but not sensitivity enhanced sequences were used for all indirectly detected spectra. The TOCSY spectra (AQ = 205 ms at SW = 5000 Hz) used MLEV-17 mixing (70 ms) and solvent non excitation (3-9-19 binominal pulse sequence; d19 = 222 μs) and 384 TPPI increments. The COSY spectra (2k x 1k matrix) were multiplied by a squared unshifted sine bell in F2 and F1; the TOCSY spectra (2k x 1k matrix) by a squared sine bell, shifted by π/6 in F2 and by a sine bell, shifted by π/6 in F1. Reference for $^{1}H$-NMR was the residual HDO resonance at 4.63 ppm and $^{13}C$ was referenced to external $CH_3OH$ in $D_2O$ (49.00 ppm) at 303 K (for 1D and 2D NMR spectra).

### *3.2.5.4. Assignments used for calculation of the model 2-D NMR spectrum of HS*

$^{1}H$ and $^{13}C$ chemical shift values of the humic constituents indicated in Figure 42 were obtained from literature values or calculated by the ACD/Labs software (Advanced Chemistry Development, Pegnitz, Germany) NMR Proton and Carbon Predictor, Version 4.0, and put into display with the EXCEL format. The prediction of chemical shift proceeds with sufficient accuracy; this has been tested for a range of model compounds (polyphenols, terpenoid hydrocarbons and peptides). Peptide chemical shifts of the 20 proteinaceous amino acids were compiled from a data base on protein chemical shifts (Wishart et al., 1991, 1995), while the shifts of the other amino acids occurring in humic substances (Sparks, 1995; Szajdak and Österberg, 1996) were calculated. Reference compounds used for the calculation of the chemical shifts included a softwood lignin (Sakakibara, 1990) and tannin structure (Hemingway and Karchesy, 1989); terpenoid hydrocarbons were composed as a sum of cholesterinacetate and hopane (Ourisson et al., 1979). Lipids used were phosphoenolpyruvate, lipoic acid and dipalmitoyl-lecithine (Breitmaier and Voelter, 1990); amino sugars (Zhang et al., 1997; Jahnel et al., 1998) were composed of the repeating units of various chitin

derivatives (Stankiewicz and Bergen, 1998). Heterocyclic nitrogen included porphine as a pyrrol derivative, the four main nucleobases (Kuzyakov, 1997) and indol. Fatty acids and their esters were composed of the saturated long chain, unbranched type, and suberin (Doi, 1990) consisted of a polybutanoic and polypropanoic ester with α-methylene group with respect to carbonyl and an unbranched side chain in α-position to oxygen incremented from $C_1$ to $C_5$.

The assignments in Table 15 were given by the Bruker (Rheinstetten, Germany) reference 2D HSQC spectra database SBASE and corroborated by calculation using the ACD predictor (cf. above), if not specified otherwise.

## 4. NMR Analysis of Functional Groups in NOM/HS

With a weight percentage ranging from 35 to 50%, the functional groups determine the physicochemical properties and, to a large extent, the ecological behaviour of NOM/HS (Perdue, 1985; Ritchie and Perdue, 2003). The combination of chemical derivatization with one and two dimensional NMR spectroscopy is a powerful new method to determine the composition of functional groups in humic substances (Hertkorn et al., 2002; Hertkorn et al., 1996; Thorn et al., 1989). Acidic protons of functional groups are reacted with reagents composed of a suitable NMR-active label and a leaving group (cf. Figure 43); ideally, the tag would substitute any labile proton of a given (class of) functional groups without side reaction – a condition not necessarily met, when NOM/HS is derivatized.

A very wide range of reagents can be reacted with NOM/HS for a NMR-based analysis of functional groups (cf. Figure 13). This method complements existing techniques of functional group analysis in NOM/HS, but provides unprecedented detail with respect to the structural details of their environment. Caused by the very high sensitivity of proton detected NMR experiments, these NMR analyses of functional groups in NOM/HS can be realized with sub-mg amounts of humic materials.

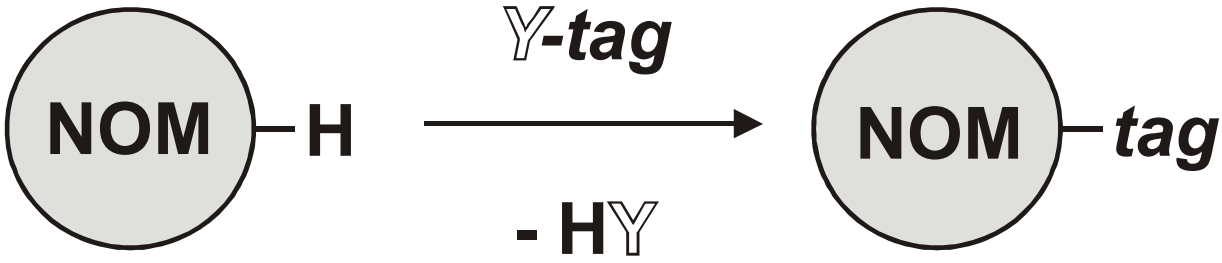



**Figure 43.** Substitution of an acidic proton in NOM/HS with a NMR active functional group ("tag").

## 4.1. THE SILYLATION OF NOM/HS

(Trimethyl)silylation substitutes the exchangeable protons in NOM/HS with rather unpolar trimethylsilyl groups (Schraml, 1990); carboxylic acids are transformed into silyl esters, alcohols into silyl ethers and amides into silyl amides.

$$-C\text{-}X\text{-}H + (H_3C)_3Si\text{-}Y = -C\text{-}X\text{-}Si(CH_3)_3 + HY,\ (X = O, N, S) \qquad (14)$$

In $^{1}H$, $^{29}Si$ HSQC NMR spectra of silylated NOM/HS, the aromatic silyl esters exhibit downfield proton chemical shift compared with aliphatic silyl esters. Furthermore, $\delta(^{29}Si)$ of the silyl esters correlates with the $pK_a$ of the parent carboxylic acid (Schraml, 1990; Schraml et al., 1986) The aliphatic and aromatic silyl esters, the primary, secondary and tertiary silyl ethers as well as phenol-derived silyl ethers and silyl amides (Schraml, 1990; Herzog et al.,1996) cover distinct areas in the $^{2}J(^{1}H,^{29}Si)$- and $^{1}J(^{1}H,^{13}C)$-HSQC NMR spectra and show within their range of chemical shift in the 2D matrix very much detail and resolution. Structural information of this kind is currently not accessible by any other analytical method.

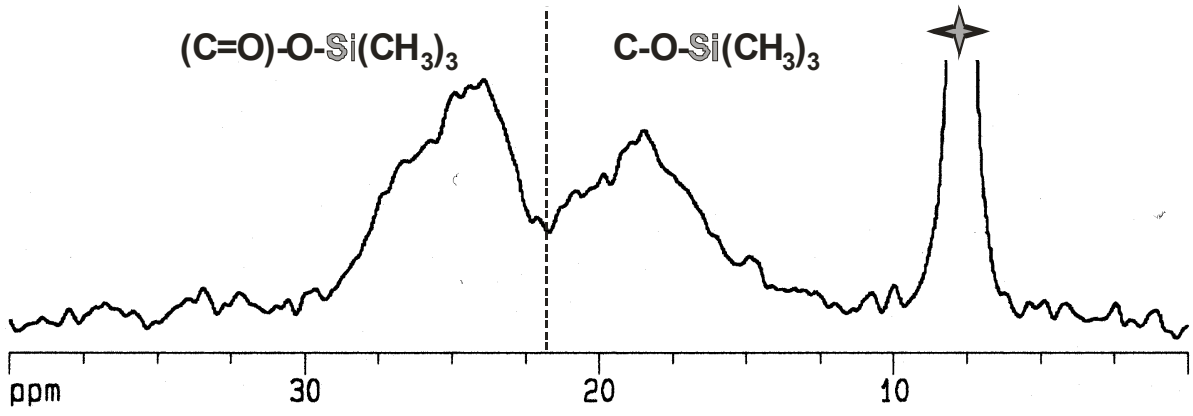



**Figure 44.** $^{2}J$ ($^{29}Si$, $^{1}H$) INEPT NMR spectrum of a silylated Suwannee river organic matter (upfield resonance: excess silylating agent BSTFA).

With a mathematical procedure of 2D NMR subtraction, the *minimum* and *difference* NMR spectra can be calculated from experimental NMR spectra (cf. Figure 45). Minimum spectra (HA = FA) show the common structural elements within two and more humic materials; difference NMR spectra show their specific characteristics, respectively (Hertkorn et al., 2002).

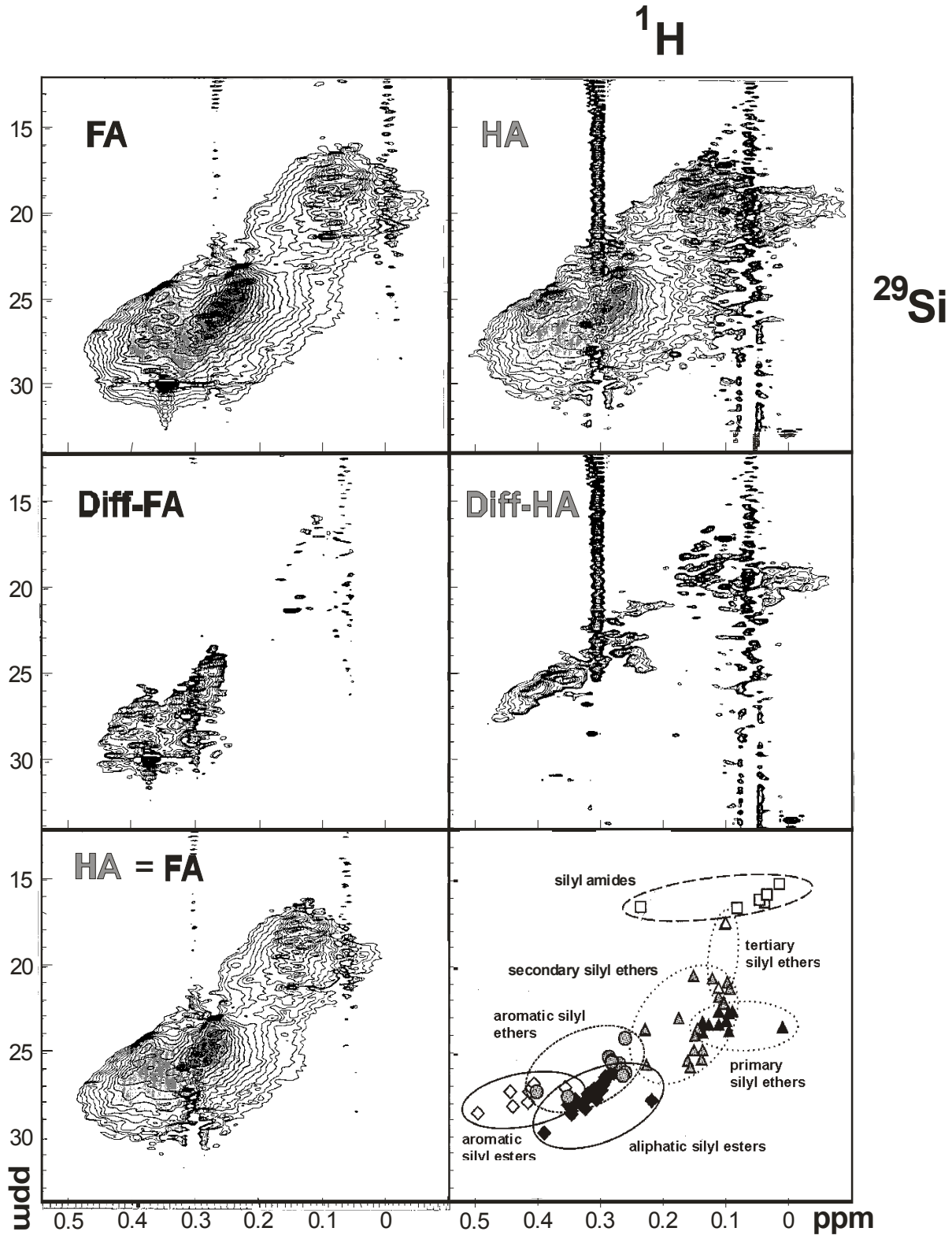



**Figure 45.** $^{2}J$ ($^{29}Si$,$^{1}H$) HSQC NMR spectra of a silylated fulvic (FA, left) and a humic acid (HA, right) from a bog lake (Holohsee HO10FA, HO10HA, Black forest, Germany) ; *upper trace:* full spectra, *middle trace:* difference NMR spectra of HA and FA showing preferential substructures in FA (left) and HA (right); *lower left:* conforming structures in FA and HA; *lower right:* $^{1}H$,$^{29}Si$-HSQC-NMR chemical shift data of model compounds.

## 4.2. THE METHYLATION OF NOM/HS

Methylation of NOM/HS in conjunction with one- and two dimensional NMR spectroscopy is another independent method for the characterization of its functional groups according to structural details. The methylation of NOM/HS effects a functionalisation of acidic protons, but also side reactions, like selective cleavages, C-methylation and the 1,3-dipolar cycloaddition, occur (Thorn et al., 1987; Mikita et al., 1981) that allow a better comprehension of the substitution patterns of the aliphatic backbone structure of NOM/HS (Lambert and Buddrus, 1996). The methyl group is sterically less demanding than any silyl group and may therefore be more appropriate to derivatize less accessible acidic protons in NOM/HS. $^{13}C$ labelling of methylating agents offers a very high sensitivity in proton detected (and) carbon nD NMR spectra (n = 1, 2, 3), and additional NMR assignment options via $^{13}C,^{13}C$ coupling constants.

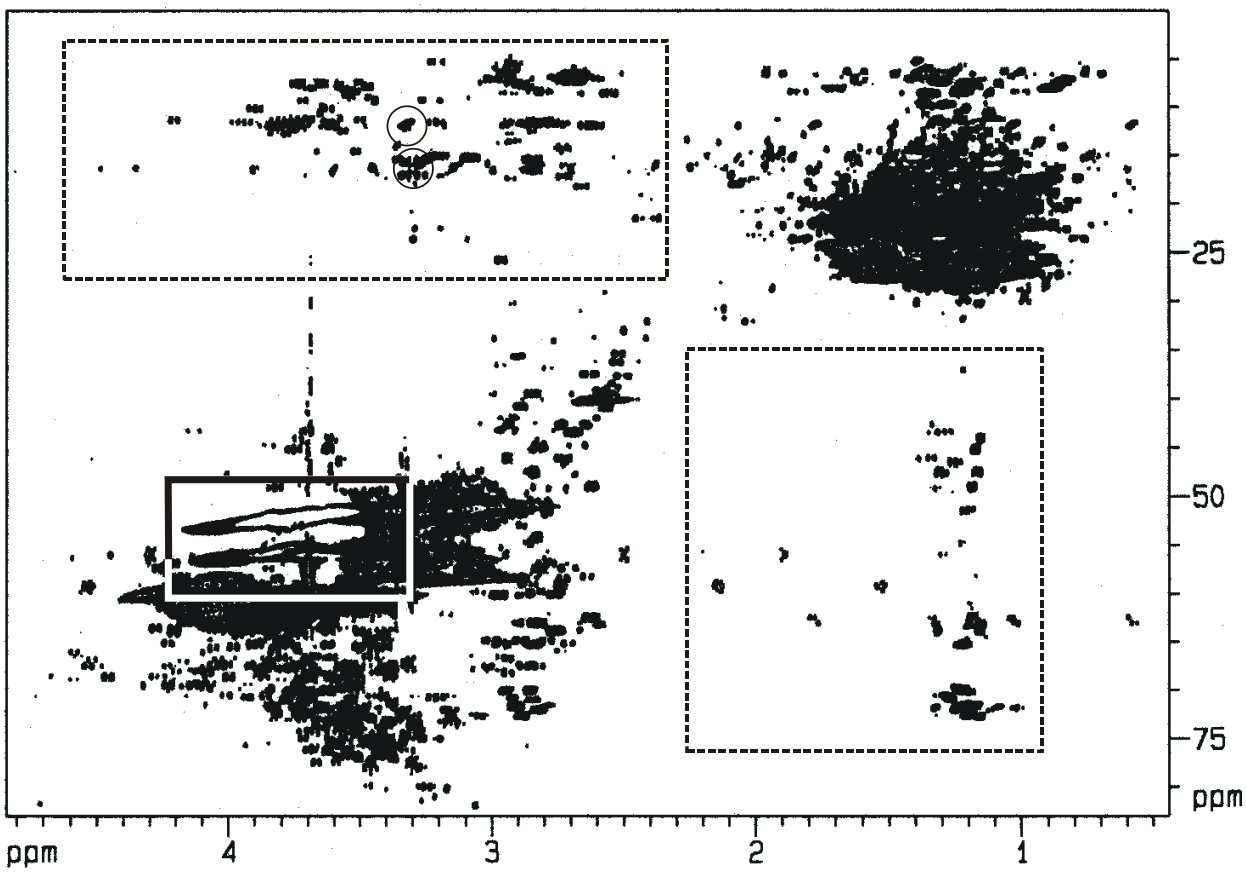



**Figure 46.** Section of the aliphatic cross peaks in a $^{1}H,^{13}C$ HSQC-TOCSY NMR spectrum of a $^{13}C$-methylated aquatic fulvic acid HO13FA from a bog lake Holohsee, black forest, Germany), representing $^{1}J(CH)$ and $^{n}J(CH)$ correlations; $^{n}J(CH)$ correlations are indicated with a dotted box; circles represent HSQC cross peaks, and the solid box indicates chemical shift region of methoxyl cross peaks.

Compared to the NMR spectra of the unmethylated native NOM/HS materials, the NMR spectra of the methylated NOM/HS show an improved resolution as well as a significant increase of the number of resonances and cross peaks with a simultaneous change of the average signal position. The $^{1}H$, $^{13}C$ HSQC-TOCSY NMR spectra of a methylated NOM/HS shows one bond carbon-proton couplings and long range carbon-proton correlations that allow the assemblage of extended partial structures (cf. Figure 46). In the methoxyl region of the HSQC NMR spectrum, aromatic and aliphatic methyl esters and methyl ethers are represented in four distinctly separated regions of chemical shift.

# 5. NMR Data Processing and Evaluation Schemes

## 5.1. PROCESSING PARAMETERS FOR THE CALCULATION OF NMR SPECTRA FROM FIDS

The NMR signal typically is acquired in the form of a discrete time domain sampling sequence in digital form (interferogram; free induction decay: FID), and several schemes of numerical digital signal processing are available to extract the information content of the signal (Rutledge, 1996; Hoch and Stern, 1996). The most commonly used approach to convert the time-domain signal into a frequency-domain NMR spectrum is by applying a Fourier transformation. Alternative processing techniques, like (forward and backward) linear prediction, maximum entropy reconstruction and Bayesian analysis, are not treated in this concise introduction to signal processing (Rutledge, 1996; Hoch and Stern, 1996).

Experimental NMR spectra commonly exhibit a range of shortcomings, like truncation artefacts, low S/N ratios, limited resolution, or undesirable line shapes. Multiplication of the interferogram with an adjusted time-domain filter function prior to Fourier transformation is equivalent to a convolution of the spectrum with a lineshape function. This process is called apodization, windowing, or filtering in the time domain and can be used for the improvement of the S/N ratio, resolution enhancement, lineshape conversion, or a combination of those. Typically, resolution enhancement causes a degrading of the S/N ratio, and vice versa. One of the most widely known apodization function, considered to constitute the optimum balance between noise reduction and the effects of line broadening, is the matched filter: here the window function doubles the linewidth in the frequency domain.

NMR spectra of NOM/HS are composed of superimposed resonances, showing a considerable range of linewidths; therefore, no single window function will be capable to display optimum results across the entire spectral range. However, some general rules apply (cf. Figure 47):

(1) In one dimensional NMR spectra of heteronuclei, an optimum S/N ratio is of primary interest and the corresponding noisy FIDs will be multiplied by rather strongly decaying exponential functions (lb = 4-150 Hz).

(2) The analysis of sections within proton NMR spectra, acquired in dilute solution at optimum $B_0$ homogeneity, may in special cases benefit from a modest resolution enhancement by a Lorentz-to-Gaussian transformation. More common are modest exponential broadening functions, e.g. in the range of lb = 0.2-1.0 Hz, to reduce baseline

noise near the high and low field limits. This improves the reliability and robustness, when the similarity of NMR spectra will be assessed by mathematical methods, like e.g. bucket analysis (Hertkorn et al., 2004; Brindle et al., 2002).

(3) Heteroatom NMR spectra of NOM/HS showing groups of resonances, differing widely in their linewidth – a common feature e.g. in $^{13}C$ NMR spectra of the lower mass fractions of DOC – may require several different apodization functions to separately analyse slow and fast relaxing components in this mixture.

(4) The information content of two dimensional spectra of NOM/HS even more depends on judiciously applied window functions (Hertkorn et al., 2002), which have to be applied separately in F2 and F1 dimension. Two dimensional NMR spectra of NOM/HS with a substantial S/N ratio and a very high digital resolution in F2 and F1 can be calculated with some emphasis on resolution enhancement; this is not feasible for standard resolution 2D NMR spectra of NOM/HS.

(a) High quality absolute value COSY NMR spectra of NOM/HS can withstand unshifted sine bell functions to produce highly resolved cross peaks (cf. Figure 29).

(b) Any phase sensitive heteronuclear 2D NMR spectrum of NOM/HS is most advantageously calculated with adjusted exponential broadening in F2, and with modest exponential broadening or shifted sine bell functions in F1.

(c) Phase sensitive homonuclear 2D NMR spectra of NMO/HS are best calculated with adjusted exponential broadening or shifted sine bells in F2, and – again – with modest exponential broadening or shifted sine bell functions in F1.

(d) The analysis of cross peaks in phase sensitive COSY NMR spectra of NOM/HS may benefit from Lorentz-to-Gaussian resolution enhancement in F2 and F1, but a very considerable digital resolution is required in both F2 and F1.

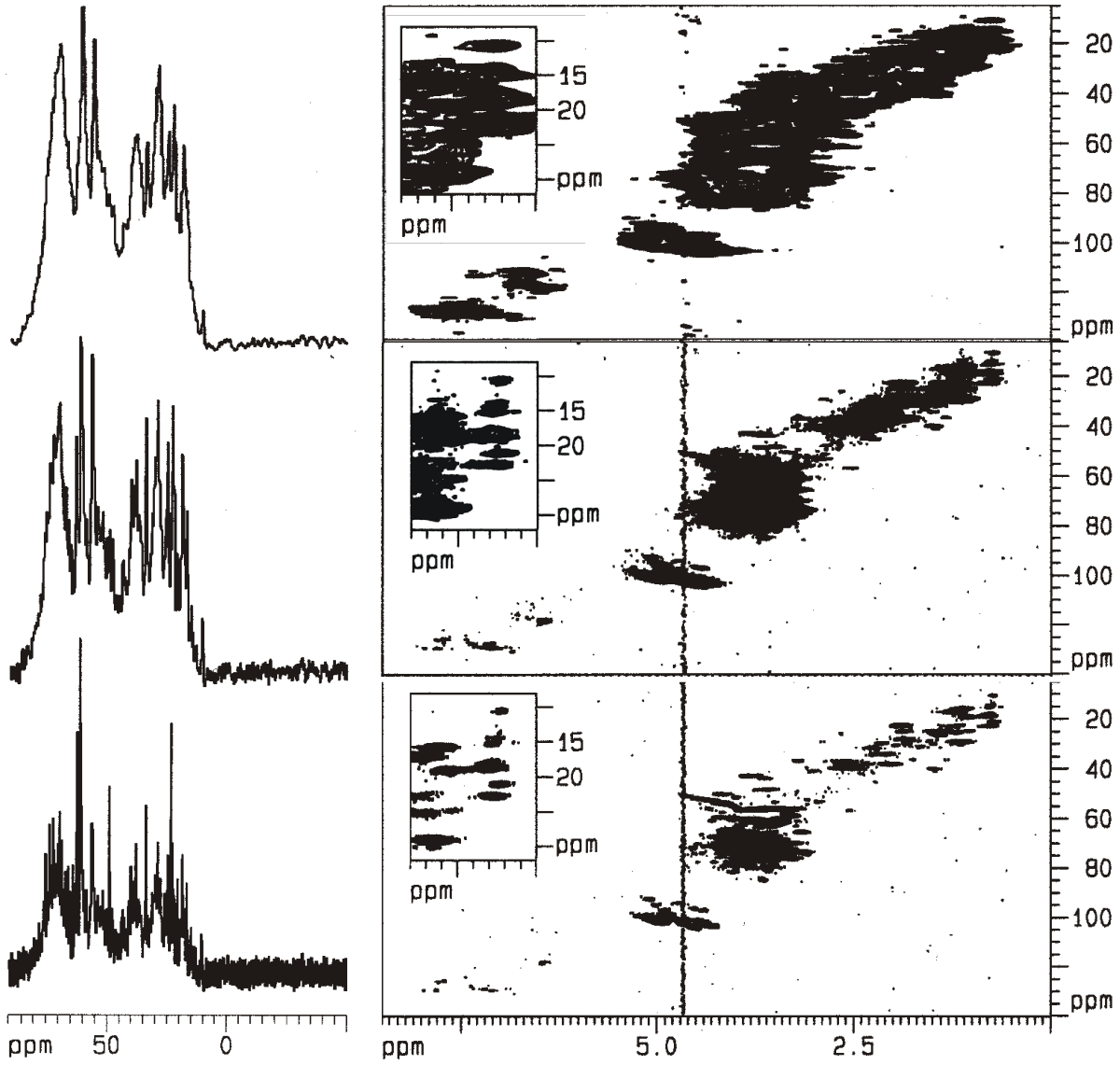



**Figure 47.** Left: DEPT-90 $^{13}$C NMR spectrum (upfield section) of the IHSS reference soil humic acid 1R106H, calculated with different exponential line broadening factors: bottom/middle/top: 3/25/100 Hz; right: gs-$^{1}$H,$^{13}$C-HSQC NMR spectrum of the aromatic section of an Italian soil humic acid (Sobretta), calculated with different window functions: bottom/middle/top (F2; F1) EM 3; SSB 6 / EM 10; GM +0.5, -0.5 / EM 50; SSB 2.

## 5.2. THE USE OF SIMULATION OF NMR PARAMETERS IN THE ASSIGNMENT OF NMR RESONANCES AND CROSS PEAKS

The combined analysis of several one and two dimensional NMR spectra of NOM/HS offers a broad range of options for resonance assignment and mutual evaluation. Nowadays, proton and carbon NMR spectra of extended spin systems can be computed, based on empirical correlations, with reasonable accuracy on desktop computers. An iterative adjustment of model and experimental NMR spectra, corroborated by the current understanding of humic bio- and geosynthetic pathways (Kögel-Knabner, 2002; Hatcher and Spiker, 1988; Hedges and Keil, 1999; Lu et al., 2000; Hedges et al., 2002; Hertkorn et al., 2002; Hatcher and Clifford, 1997), allows to propose model structures (Simpson et al., 2002) conforming to spectral data. In this respect, measurement and data base search of NMR parameters of model compounds (Glore and Gronenborn, 1998; Wishart et al., 1995; Wishart et al., 1991) and calculation of parameters of NMR spectra of extended organic structures (Hertkorn et al., 2002) are powerful tools to discriminate among several proposed humic substructures. Even when used only as initial step, which requires verification by independent analytical data, this approach provides very significant resolution in structural detail at atomic resolution, which is not available by other analytical techniques.

## 5.3. ANALYSIS OF AROMATIC SUBSTITUTION PATTERNS

For specific regions of chemical shift, adapted assignment schemes can be developed (Simpson et al., 2004); an example is the *increment analysis* of 2D NMR spectra to elucidate the aromatic substitution patterns in NOM/HS (Hertkorn et al., 2004). The substitution of aromatic rings in HS/NOM, which carry from one to five hydrogen substituents, can be deduced from the increment analysis of homo- and heteronuclear 2D NMR spectra. The $\delta(^{1}H)$ and $\delta(^{13}C)$ values of any $C_{ar}$-H bond are most strongly affected by the nature of ortho and para substituents. Oxygen substitution in ortho and para positions causes proton upfield shift, while carbonyl derivatives induce downfield shifts of ortho protons caused by the chemical shift anisotropy of the carbonyl group. However, any cross peak within the aromatic section of a COSY NMR spectrum indicates (at least) two ortho protons; the chemical shift of these two protons is affected by all other substituents of the aromatic ring. Probable aromatic substitution patterns can be deduced from increment analysis of the chemical shift values on a

probabilistic and individual structure level (cf. Figure 48; case study 3). The distribution of cross peaks in homonuclear 2D NMR spectra indicates a broad range of oxygen, carbon and carbonyl derivative substitution in aromatic substructures of NOM/HS; there is no preference for (only) a few specific substitution patterns.

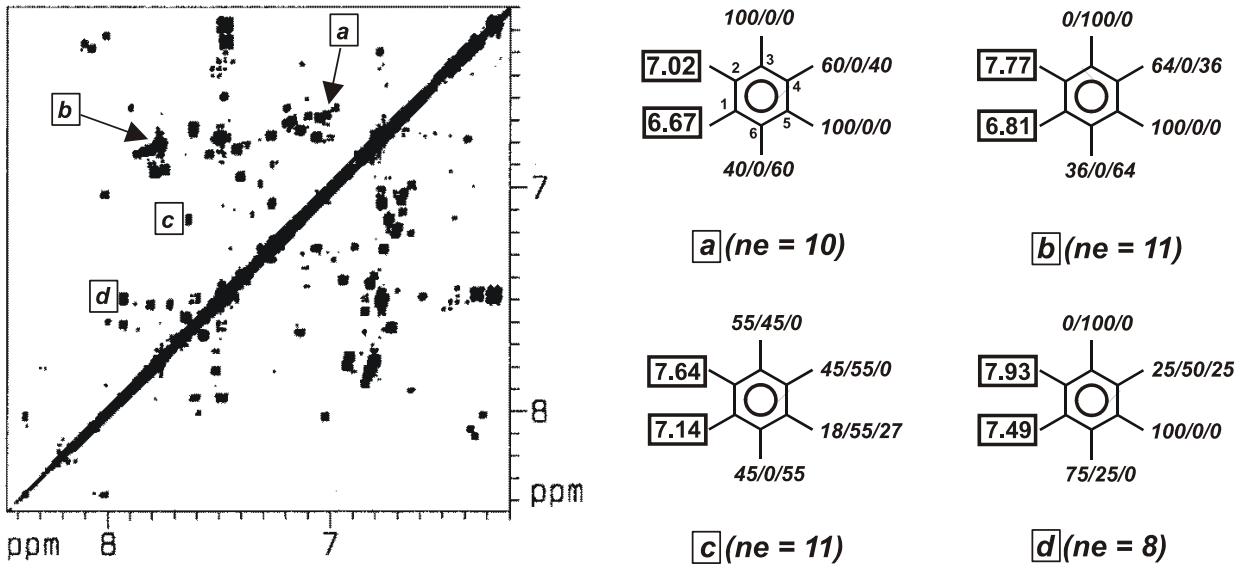



**Figure 48.** Section of aromatic cross peaks in a 2Q COSY NMR spectrum of a soil humic acid, with a probabilistic level ($\Sigma_R + \Sigma_{COR} + \Sigma_{OR} = 100$ %) of substituent distribution for classes of substituents *R / COR / OR*, indicated for each position [ne = number of hits (out of $2^{16}$ = 32768 combinations of substituent patterns, resulting in 16640 unique substitution patterns) within a chemical shift window of 0.1 ppm in both F1 and F2 around the center of the cross peak indicated].

# 5.4. CASE STUDY 3: SUBSTITUTION PATTERNS IN AROMATIC RINGS BY INCREMENT ANALYSIS (SPARIA) – MODEL DEVELOPMENT AND APPLICATION TO NATURAL ORGANIC MATTER

## *5.4.1. Abstract*

***The aromatic region of two-dimensional heteronuclear $^{1}H$, $^{13}C$ NMR spectra of natural organic matter and related materials (e.g., $^{1}H$ and $^{13}C$ chemical shifts ranging from approximately 5 – 10 ppm and 80 – 140 ppm, respectively) is highly complex and generally not interpreted in any detail. In principle, this region of the NMR spectrum should be amenable to detailed analysis, because the effects of many common substituents on the chemical shifts of aromatic carbon and hydrogen are well documented. This case study describes the development of a model for prediction of substitution patterns in aromatic rings by increment analysis (SPARIA). In the forward mode, SPARIA is used to predict the chemical shifts of $^{1}H$ and $^{13}C$ on aromatic moieties containing every possible combination of eight common substituents that are likely to be representative of substituents on aromatic moieties in natural organic matter. The accuracy of SPARIA in the forward mode is evaluated for 29 aromatic compounds (100 peaks) by comparison of predicted chemical shifts for $^{1}H$ and $^{13}C$ with experimental values and with predictions of commercially available software for prediction of NMR spectra. The most important development in this case study is the inverse mode that is built into SPARIA. Given chemical shifts for $^{1}H$ and $^{13}C$ (such as may be obtained from a two-dimensional, heteronuclear NMR spectrum), the inverse mode of SPARIA calculates all possible combinations of the eight selected substituents that yield chemical shifts within a specified window of chemical shift for both $^{1}H$ and $^{13}C$. Both the distribution of possible substitution patterns and simple descriptive statistics of the distribution are thus obtained. The inverse mode of SPARIA has been tested on the 29 aromatic compounds (100 peaks) that were used to evaluate its forward mode, and the dependence of the inverse process on the size of the chemical shift window has been evaluated. Finally, the inverse mode of SPARIA has been applied to selected peaks from the two-dimensional heteronuclear HSQC spectrum of sample of natural organic matter that was isolated from the Suwannee River in southeastern Georgia, U.S.A. by reverse osmosis, and to an individual molecele-level analysis of an alkaline soil extract.***

### *5.4.2 Introduction*

Among the properties of natural organic matter (NOM) that have received considerable attention in the past 30 years is its aromaticity. First of all, aromaticity (or the lack thereof) in NOM provides some insight into its biochemical precursors and about its most likely pathways of degradation. Additionally, the attenuation of light in natural waters, the generation of disinfection byproducts during chlorination of natural waters, and the solubilization and transport of hydrophobic organic chemicals in natural waters have all been related to the aromaticity of NOM. The most direct insight into the aromaticity of NOM is provided by $^{1}H$ and $^{13}C$ NMR spectrometry. Even using those methods, very little detailed structural information has been elucidated from NMR measurements on the aromatic moieties in NOM.

Two major groups of molecules of very different diagenetic origin supposedly contribute most to the aromatic substructures of NOM:

a) The main precursors of aromatic NOM constituents are phenolics derived from vascular plants, like lignins and tannins, aromatic amino acids and other aromatic secondary metabolites of plant and microbial origin, and oxidized aromatic products from photochemical and metal-catalyzed reactions. These aromatic molecules do not typically represent basic substitution patterns of prominent biochemicals.
b) The other major synthetic route to aromatic NOM constituents is initiated by the incomplete combustion of organic compounds, resulting in a near continuum of materials commonly referred to as black carbon (BC). Black carbon is probably comprised of an enormous range of extended polycyclic aromatic systems with variable degrees of oxygenation and oxidation, and the occurrence of oxygen-containing heterocycles is very likely. Recently, it has been found, that a fraction of marine BC is also derived from weathering of graphite minerals (Dickens et al., 2004).

The elucidation of molecular level structural detail of BC will remain a difficult task without resorting to a combination of chemical reactions and organic structural spectroscopy (namely ultrahigh resolution FTICR mass spectrometry). The aromatic fraction derived from BC is expected to contribute to the broad hump in the $^{13}C$ NMR spectra of NOM, which is characterized by short transverse NMR relaxation times (Fan, et al., 2000).

In contrast, the class of aromatic systems in NOM, which is derived from natural products, frequently comprises non-condensed ring systems. Owing to their enhanced flexibility in the

molecular framework of NOM, these compounds are likely to generate more intense cross peaks in the homo- and heteronuclear 2D NMR spectra of NOM than those derived from BC. At present, the most instructive method to determine the aromatic constituents of NOM is the joint analysis of one dimensional proton and carbon NMR spectra. Since these NMR spectra are comprised of heavily overlapping sets of resonances, an assignment of these spectra on the basis of chemical shift alone will remain severely ambiguous. Connectivity information, obtained from homo- and heteronuclear 2D NMR spectra, significantly improves the reliability of the resonance assignments in NMR spectra of NOM (Hertkorn, 2002, 2005).

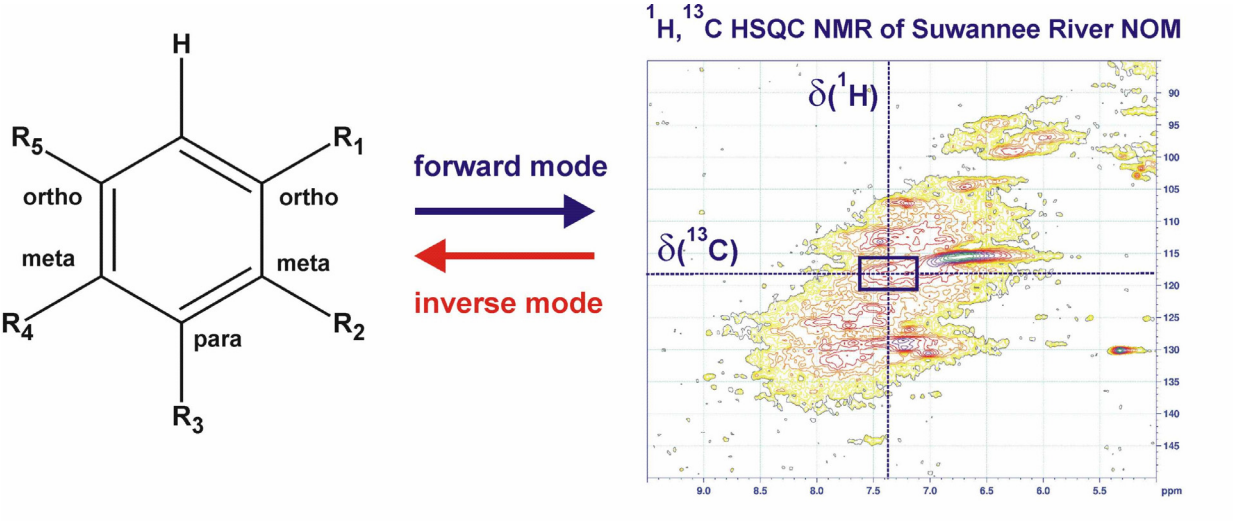



**Figure 49:** The inverse mode of SPARIA computes all possible combinations of the eight selected substituents that yield chemical shifts within a specified window of chemical shift for both $^{1}H$ and $^{13}C$. Both the distribution of possible substitution patterns and simple descriptive statistics of the distribution are this obtained.

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

aromatic rings by increment analysis (SPARIA). At present, the model is implemented partially in computer code and partially in Microsoft Excel.

### *5.4.3. Technical background*

Whenever a chemical substituent is attached to an aromatic ring, the distribution of electron density on the aromatic ring is perturbed in a systematic, predictable manner. Rates of reactions, chemical equilibria, and even NMR chemical shifts respond in a predictable manner to the perturbation caused by a substituent. The oldest quantitative expression of the effects of chemical substituents on the properties of aromatic rings is the Hammett equation – the predecessor of all other linear free energy relationships (LFER's).

In the case of NMR spectrometry, the effect of a chemical substituent on the chemical shift of $^{1}H$ and $^{13}C$ is dependent both on the nature of the substituent and its position (ortho, meta, or para) relative to the $^{1}H$ or $^{13}C$ atom. The incremental chemical shifts of many common substituents have been compiled (Thomas et al., 1994; Hesse et al., 1991) after analyzing the effects of those substituents on the chemical shifts of $^{1}H$ and $^{13}C$ in a large number of aromatic compounds. In this case study, several "representative" substituents have been chosen and their incremental chemical shifts have been used throughout this case study.

1. NOM contains primarily C, H, and O, so substituents containing only those elements were considered. Naturally, one of the substituents must be –H. It is further known that the biological precursors of NOM contain COOH, R—OH, Ar—OH, ester, ether, saturated and unsaturated alkyl groups, etc. Saturated and unsaturated alkyl side chains are represented by $–C_2H_5$ and $–CH{=}{=}CH_2$, respectively. Aromatic carbon can bond directly to the oxygen atoms of –OH and $–OCH_3$ groups. Finally, the variety of carbonyl-containing groups that may exist in NOM is represented by –COOH, $–CO_2CH_3$, and $–COC_2H_5$.

**Table 16.** Substituents Used in the SPARIA Model

| Neutral Groups (R) | Electron-Withdrawing Groups (COR) | Electron-Donating Groups (OR) |
|---|---|---|
| —H | —COOH | —OH |
| —$C_2H_5$ | —$COOCH_3$ | —$OCH_3$ |
| —$CH=CH_2$ | —$COC_2H_5$ | |

Substituent effects causing displacement of NMR chemical shifts can be classified according to steric (active primarily between ortho-substituents), polar (acting through space and chemical

bonds) and mesomeric (primarily affecting electron densities at ortho and para positions) interactions. Mesomeric substituent effects are considered rather decisive with respect to both reactivity and NMR chemical shift effects and accordingly, the eight substituents used in the standard version of SPARIA were classified according to neutral (R), mesomeric (and inductive) electron - withdrawing (COR) and mesomeric electron-donating (OR) – the latter two oxygen substituents also exhibit inductive electron-withdrawing properties.

2. This group of substituents collectively reflects the range of possibilities for polar, mesomeric, and steric effects in substituted aromatic compounds. The concept and list of substituents could easily be extended, but this minimal set of substituents is adequate for purposes of developing, testing, and applying SPARIA. The standard increments of chemical shift for the substituents used in SPARIA are presented in Table 1 (Hesse et al., 1991).

**Table 17.** Incremental chemical shifts for selected substituents on aromatic rings – $^{1}H$ and $^{13}C$ incremental chemical shifts are relative to 7.26 ppm and 128.5 ppm, respectively (from Hesse et al, 1991).

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

In the course of using SPARIA in the forward mode, it became clear that predictions of the chemical shift of $^{1}H$ were often poor in molecules containing COOH groups that were ortho to each other. The incremental chemical shifts of COOH were optimized for such groups using the solver in Microsoft Excel. Similarly, it was observed that predictions of the chemical shift of $^{13}C$ were poor in structures containing the –CH=CH—COOH group, which was being modeled using the $–CH=CH_2$ substituent. A new set of incremental chemical shifts was obtained using the solver in Microsoft Excel for such structures.

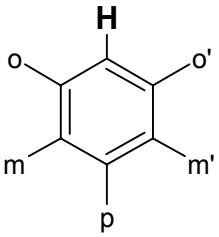
As a simple illustration of the use of the data in Table 1 for prediction of chemical shifts of $^{1}H$ and $^{13}C$ in aromatic compounds, consider 3,5-dimethoxy-4-hydroxy-benzoic acid, which contains two equivalent aromatic C—H bonds. The calculated chemical shifts of $^{1}H$ and $^{13}C$, using data from Table 17, are shown in Table 18:

**Table 18.** Forward prediction of the chemical shifts of $^{1}H$ and $^{13}C$ in 3,5-dimethoxy-4-hydroxy-benzoic acid using increment analysis.

| | **Reference** | **o–COOH** | **m-H** | **p–$OCH_3$** | **m–OH** | **o–$OCH_3$** | **Prediction** |
|---|---|---|---|---|---|---|---|
| $^{1}H$ | 7.26 | +0.85 | 0.00 | -0.44 | -0.12 | -0.48 | 7.07 |
| $^{13}C$ | 128.5 | +1.6 | 0.0 | -8.1 | +1.6 | -15.0 | 108.6 |

The observed chemical shifts of $^{1}H$ and $^{13}C$ are 7.19 ppm and 106.9 ppm, respectively. Table 18 is intended to illustrate simply the process by which increment analysis yields predicted chemical shifts. A comprehensive analysis of the predictive abilities of SPARIA in both the forward and inverse modes follows later in this case study.

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




It will also be necessary to distinguish between the total number of possible substitution patterns and the number of unique substitution patterns on an aromatic ring. For example, if the five ring positions are substituted with all possible combinations of two substituents, then there are $2^5$ (32) possible substitution patterns. Because a plane of symmetry and a rotational axis of symmetry pass through the C—H group and the substituent that is para to that group, the number of unique substitution patterns will be less than the total possible number of possible substitution patterns. Among the 32 substitution patterns are 8 patterns that have identical ortho groups and identical meta groups (internal plane of symmetry). The remaining 24 substitution patterns include 12 rotationally equivalent pairs. When half of the rotationally equivalent substitution patterns are removed, the number of unique substitution patterns is reduced to 20 (= 8 + ½(32 – 8)).

### *5.4.4. SPARIA Forward Mode – Predicting Chemical Shifts*

Using standard incremental values for substituent-induced modification of the chemical shifts of aromatic $^{1}H$ and $^{13}C$ in Table 1, together with the reference chemical shifts of $^{1}H$ and $^{13}C$ in benzene, the chemical shifts of $^{1}H$ and $^{13}C$ at an unsubstituted ring position have been predicted for all possible combinations of the eight substituents on the other five ring

positions. The total number of combinations is $8^5$ = 32,768. This process was accomplished with a user-written software called SPARIA.EXE. The output file contains 32,768 data records, each of which lists a substitution pattern (ortho, meta, para, meta', ortho') and the predicted chemical shifts of $^1$H and $^{13}$C for that substitution pattern. The symmetry of the benzene ring leads to 512 ($8^3$) patterns having an internal plane of symmetry and 16,128 rotationally equivalent pairs of substitution patterns (½(32,768 – 512)). The number of unique substitution patterns is predicted to be 16,640 (512 + 16,128). To locate and remove the redundant substitution patterns, the output file was imported into a Microsoft Excel spreadsheet, and each of the 32,768 substitution patterns was processed as follows:

1. Whenever the right-ortho group had a lower alphabetical rank than the left-ortho group, the entire substitution pattern was rotated 180 degrees.
2. Whenever the left- and right-ortho groups were identical,

a. if the right-meta group had a lower alphabetical rank than the left-meta group, the entire substitution pattern was rotated 180 degrees.

b. if the left- and right-meta groups were identical or if the right-meta group had a higher alphabetical rank than the left-meta group, the substitution pattern was not modified.

3. Whenever the right-ortho group had a higher alphabetical rank than the left-ortho group, the substitution pattern was not modified.

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

The resulting distribution of chemical shifts for $^1$H and $^{13}$C and the peaks for compounds containing only one of the eight substituents in all five positions of substitution are shown in Figure 50.

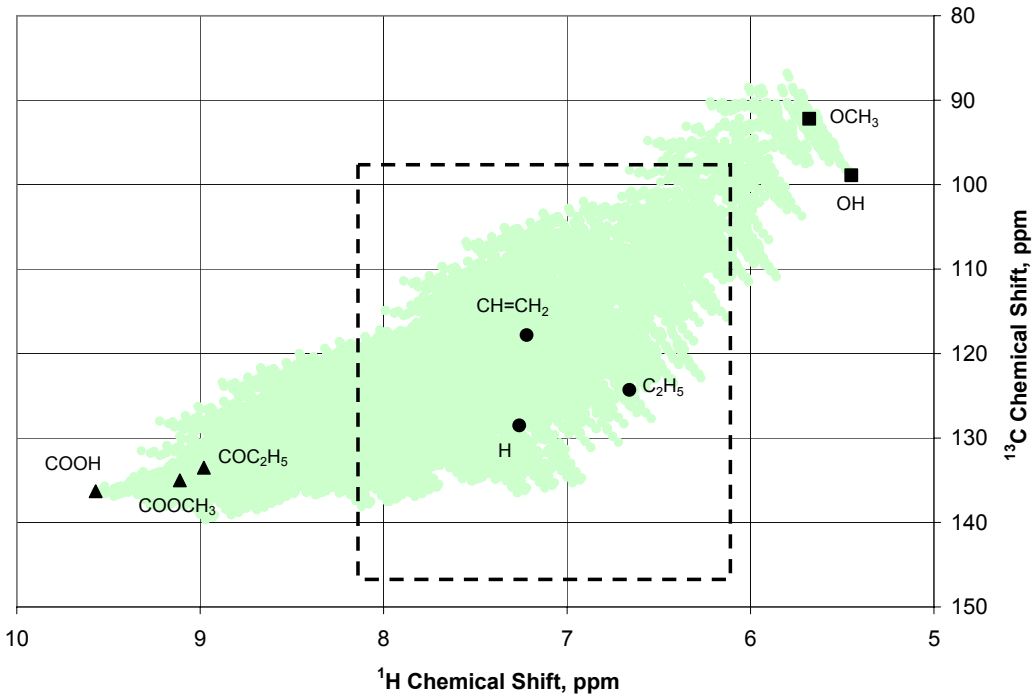



**Figure 50.** Chemical shifts of HSQC cross-peaks for aromatic C—H in 16,640 unique combinations of eight substituents on the other five ring positions. Labeled peaks correspond to structures $C_6HS_5$in which all five substituents are identical. The dominant cross-peaks for aromatic C—H in natural organic matter generally fall within the rectangular area in this Figure.

The proton and carbon NMR chemical shift space occupied by the HSQC cross peaks of the 10368 unique molecules is defined by all feasible combinations of chemical shift increments and shown in Figure 50. Here, the endmembers $C_6HS_5$ (S stands for any of the 8 substituents which are sorted according to classes R, COR, OR as shown in Table 1) which identical SPARIA substituents (see Table 1) on an aromatic ring are indicated. The C-H cross peak position depends on the combined inductive and mesomeric effects of the five other substituents in the molecule $C_6HR_l(COR)_m(OR)_n$ [l + m + n =5]. The strongly (ortho and para) mesomeric electron donating OR-substituents and the strongly electron-withdrawing COR-substituents occupy the opposite up- and downfield regions of the chemical shift space, while the neutral R-substituents occupy the center section. Accordingly, any combination of substituents from only one of these three groups of substituents occupies a rather confined space, which is defined more or less by a triangle (groups R and COR) or by a straight line (group OR; cf. Figure 55).

Interestingly, the most generally observed peaks in NOM and related materials lie within the rectangular area in Figures 50, 60. Aromatic rings substituted exclusively with COR or OR groups are unlikely to be major constituents of NOM. Aromatic rings in NOM are more likely to contain a mixture of R, COR, and OR groups.

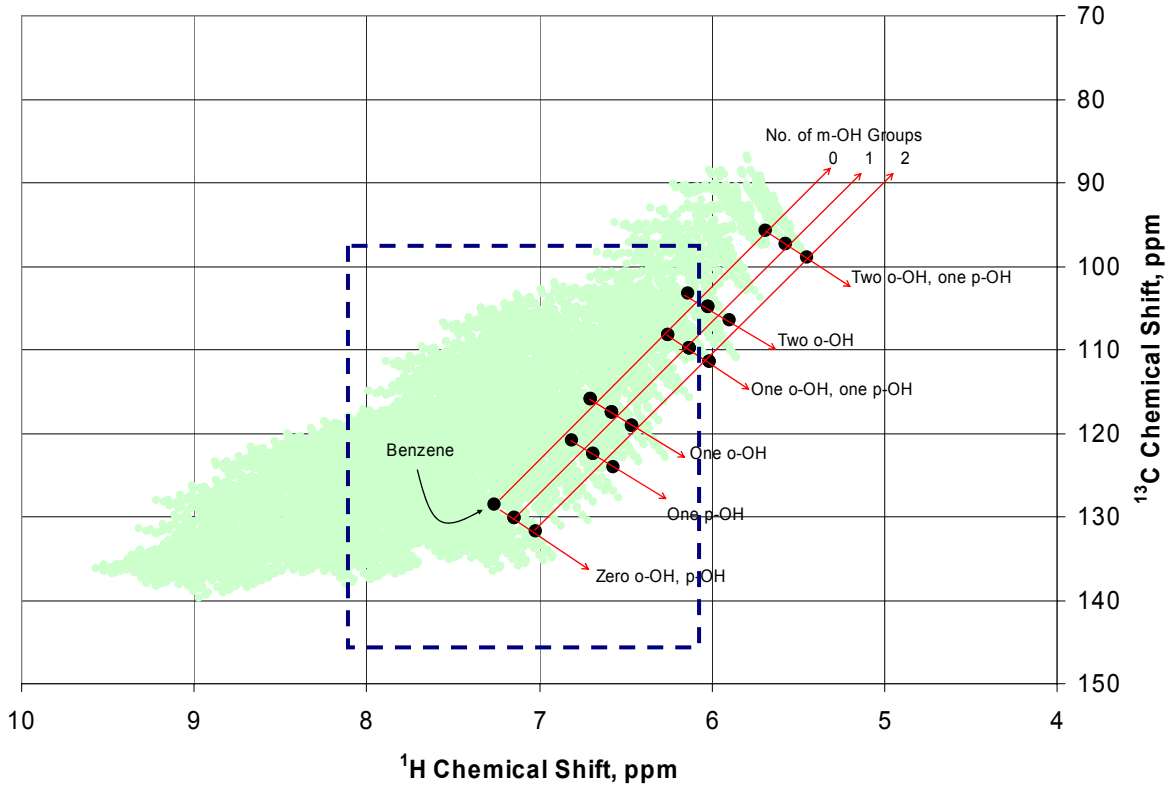



**Figure 51:** example of additive propagation of chemical shift increments for molecules $C_6H_n(OH)_{6-n}$(n = 0-6)

In Figure 51, the substituent effects of OH and H on the chemical shift of the HSQC cross peaks in the molecules $C_6H_n(OH)_{6-n}$ (n = 0 – 6) are displayed. First, it is noteworthy, that the increment analysis does not discriminate between the two possible ortho and meta positions – therefore SPARIA recognizes only 18 different combinations of substituents (out of 22 different aromatic substitution patterns) resulting in 18 calculated different positions for HSQC crosspeaks as shown in Figure 51. The additivity of individual substituent increment values results in the regular patterns observed in Figure 51; here the strongly shielding ortho and para increments for OH substituents (cf. Table 17) induce comparatively larger displacements of HSQC cross peaks than meta increments.

The first example shown here to justify the necessity of a SPARIA analysis in its inverse mode is concerned with the propagation of HSQC cross peak location for a more general combination of aromatic substituents in molecules $C_6R_l(COR)_m(OR)_n$ [l + m + n = 6] with increasing counts of COR substituents (m = 0…..5; Figure 52). Molecules $C_6H(COR)_5$ (m = 5) with any combination of five COR substituents occupy a rather confined triangular space (cf. above) in the far low field section of proton and carbon chemical shift. Molecules $C_6HR_l(COR)_4(OR)_n$ [l + n = 1; m = 4] do not show HSQC cross peak in that very extreme downfield section of chemical shift since any neutral R or oxygen substituent OR will reduce mesomeric and inductive electron-withdrawal in the aromatic ring and the C-H bond referred to in the SPARIA analysis. The chemical shift space of the molecules $C_6HR_l(COR)_4(OR)_n$ [l + n = 1; m = 4] is spreading into two lobes, and the total chemical shift space accessible to these molecules is more than ten times larger than that of $C_6H(COR)_5$ molecules. The upper lobe with its shielding in both proton and carbon frequencies is caused by OR substituents (i.e. OH and $OCH_3$), while the deshielded carbon lobe is caused by molecules of the type $C_6HR(COR)_4$. Further expansion of the accessible HSQC cross peak shift range by another factor of almost three is observed for molecules $C_6HR_l(COR)_3(OR)_n$ [l + n = 2; m = 3]. At this degree of COR substitution (m = 3), HSQC cross peak patterns are still recognizable, and very often represent groups of structurally related molecules. More than half of the entire HSQC cross peaks of molecules $C_6HR_l(COR)_2(OR)_n$ [l + n = 3; m= 2] occupy a vast chemical shift range; however, any of the very low and high field proton and carbon chemical sections remain unoccupied as high counts of either OR or COR substituents are required to place any HSQC cross peak into those areas. In the case of $C_6HR_l(COR)(OR)_n$ [l + n = 4; m= 1] molecules, HSQC cross peaks are only observed at proton upfield shifts $\delta_H < 8$ ppm, and the overall accessible chemical shift space, although still large, begins to decline. Not any single HSQC cross peak with $\delta_H >$ 7.5 ppm is found for molecules $C_6HR_l(OR)_n$ [l + n = 5; m = 0]. The distribution of cross peaks from molecules $C_6H(COR)_m(OR)_n$ [m = 0 – 5] provides interesting clues to what extent carboxyl derivative substitution affects proton chemical shift in aromatic rings. Any NMR signal with $\delta_H > 7.4$ ppm almost surely requires at least one (ortho) carbonyl substituent, and signals at $\delta_H > 8$ ppm most likely indicate two (ortho) COR groups. Conversely, any aromatic system with at least four COR groups in any order will exhibit NMR resonances with $\delta_H > 8$ ppm.

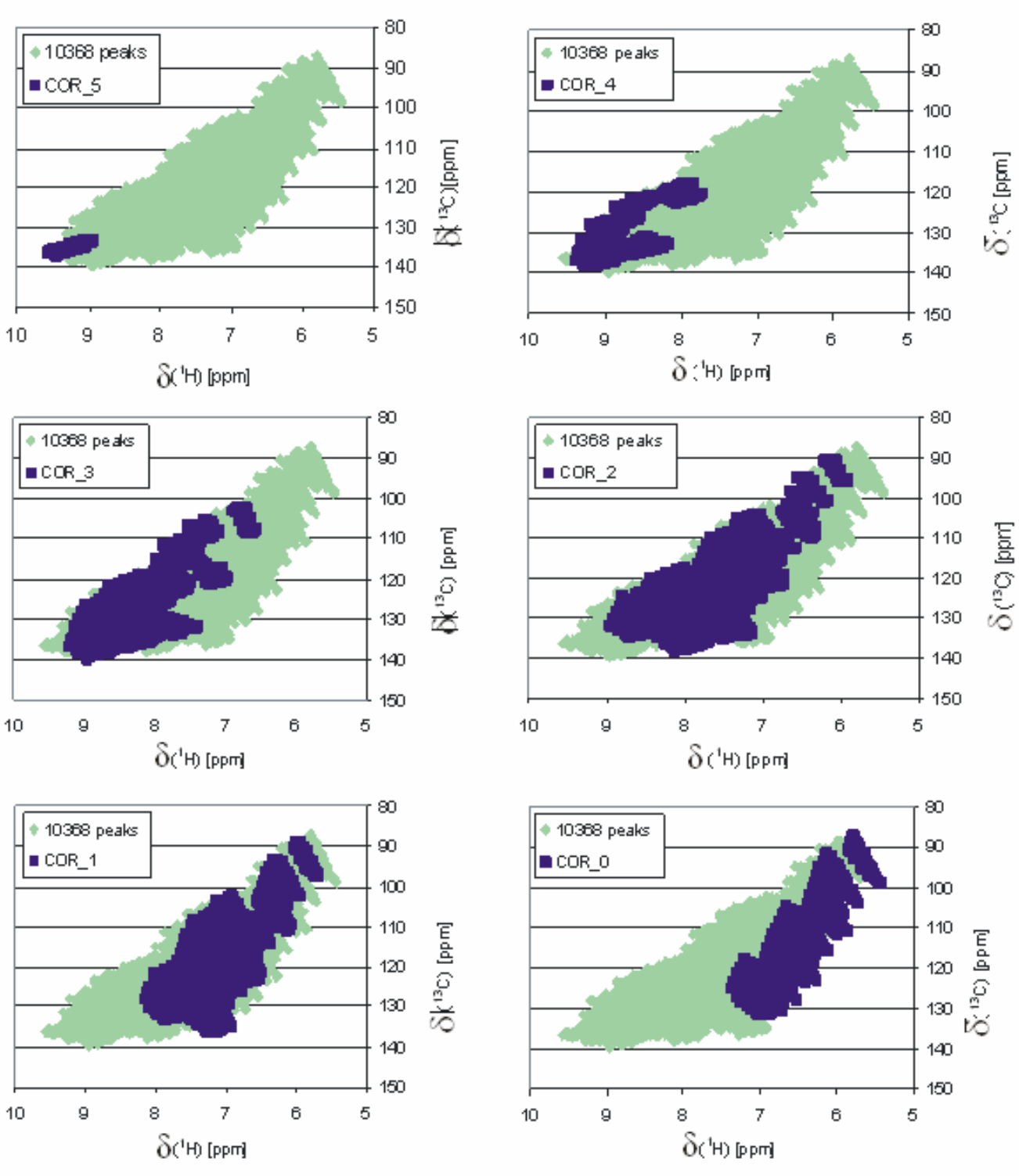



**Figure 52.** Propagation of the chemical shift space occupied by compounds $C_6HR_l(COR)_{0\text{-}5}(OR)_n$

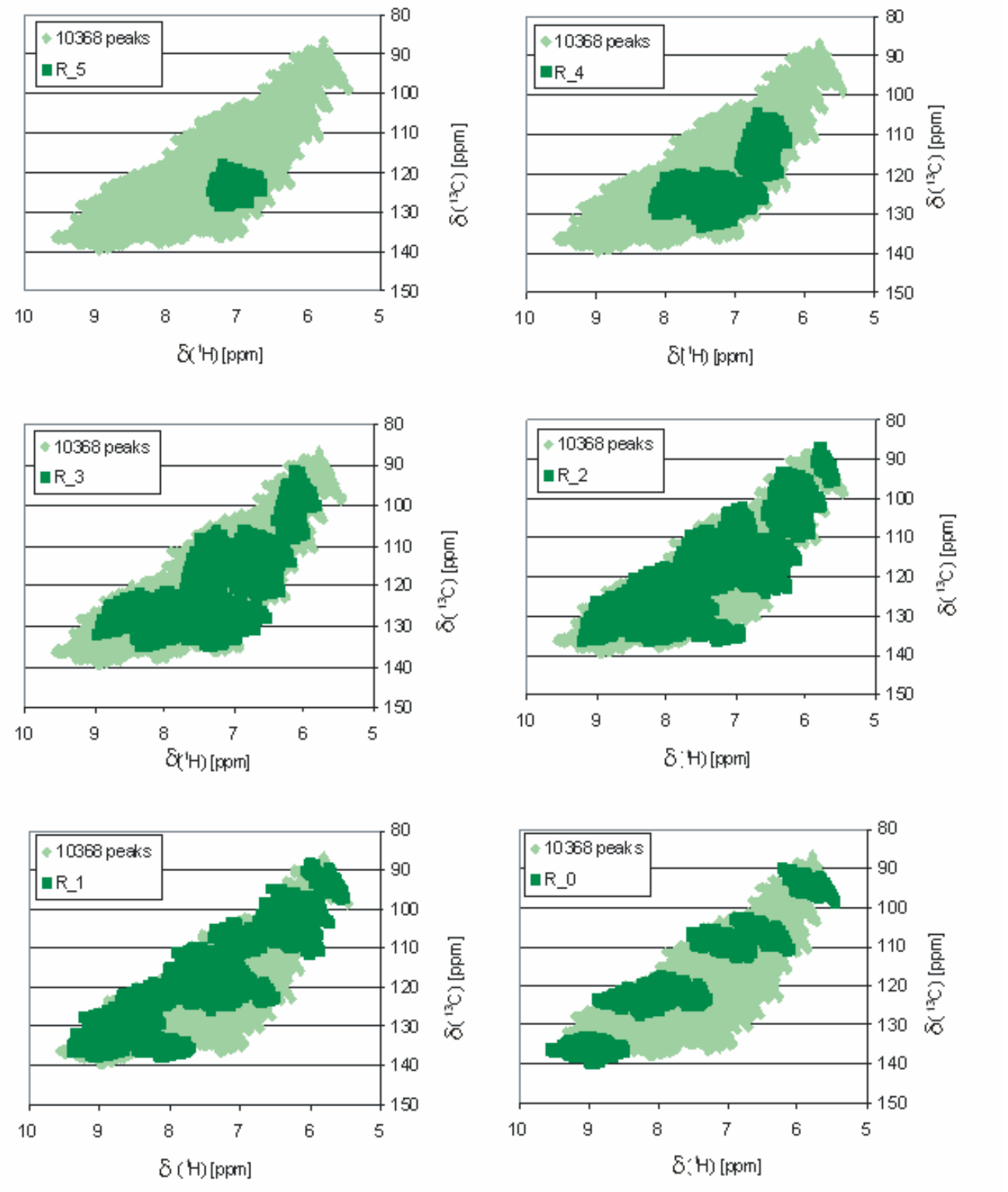



**Table 53.** Propagation of the chemical shift space occupied by compounds $C_6HR_{0\text{-}5}(COR)_m(OR)n$

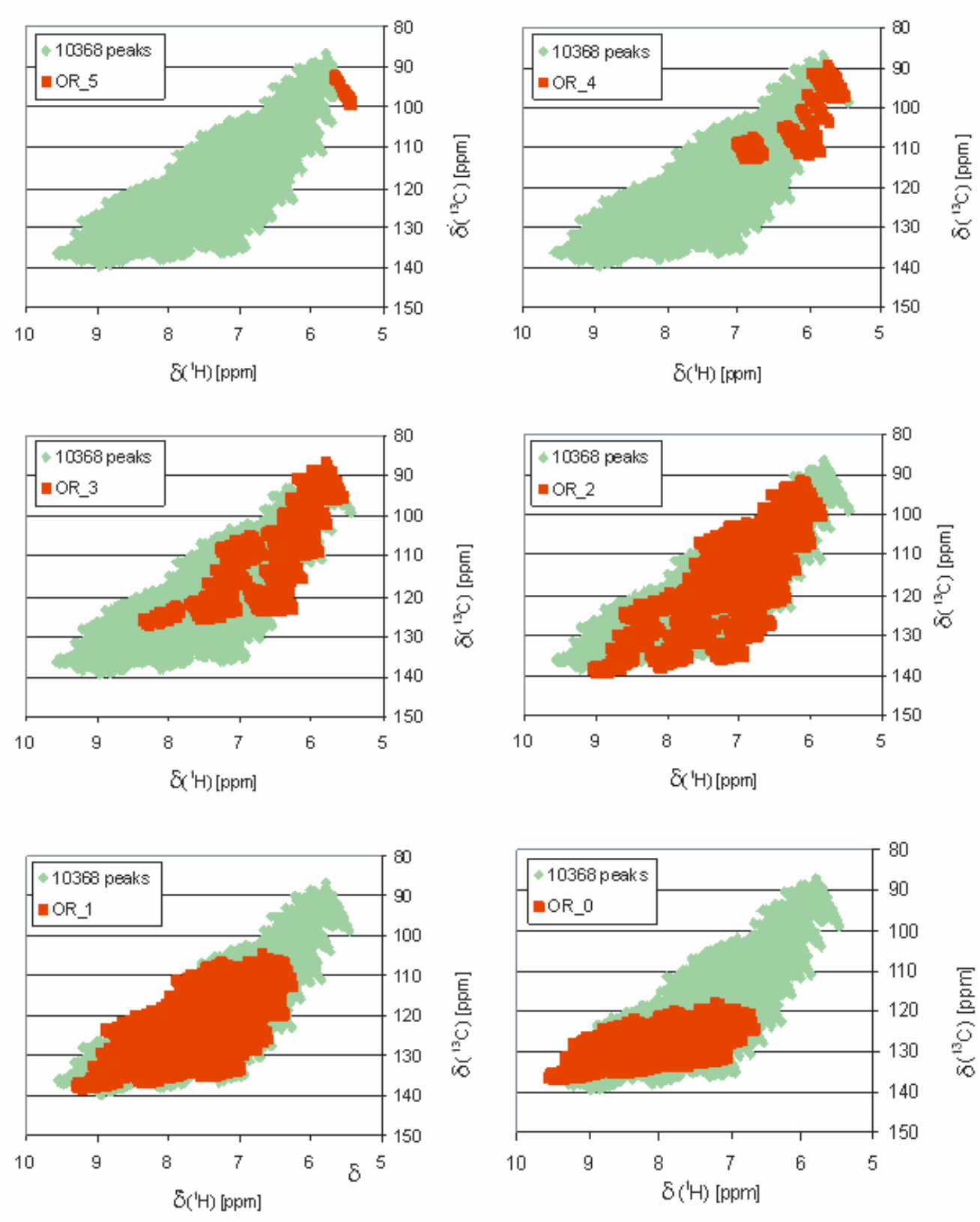



**Table 54.** Propagation of the chemical shift space occupied by compounds $C_6HR_l(COR)_m(OR)_{0\text{-}5}$

Analogous trends of HSQC cross peak distribution are observed for molecules $C_6HR_l(COR)_m(OR)_n$ [l + m + n = 5 and l = 5….0; Figure 53) and $C_6HRe(COR)_m(OR)_n$ [l + m + n = 5 and n = 5….0; Figure 54).

The most highly oxygenated compounds $C_6HR_l(COR)_m(OR)_{4,5}$ occupy the most shielded section in both proton and carbon chemical shift space. Aromatics without any OR substituent still occupy a fairly wide chemical shift space which is limited by the endmember groups of $C_6H(COR)_5$ and $C_6HR_5$ molecules. A proton chemical shift $\delta_H < 7$ ppm and similarly, carbon resonances below $\delta_C = 120$ ppm, almost definitively indicate the presence of oxygen (likely in ortho- or para position) (Figure 54). With lesser degree of R-type substitution, the contiguous space occupied by molecules $C_6HR_5$ breaks up into two lobes, representative of compounds $C_6HR_4(COR)$ [downfield] and $C_6HR_4(OR)$ [upfield], respectively (Figure 53).

Molecules devoid of R-substituents show an interesting pattern of isolated islands, representative of the six various groups $C_6H(COR)_m(OR)_n$ [m + n = 5: m = 0 – 5].

An alternative way to display the necessity of an inverse SPARIA analysis to determine the aromatic substitution pattern for a given HSQC cross peak position is provided in figure 55, in which the accessible chemical shift spaces for HSQC cross peaks are shown in color code (S = R, green; S = COR, blue; S = OR, red) for molecules with five substituents of a given class ($C_6HS_5$; top left), four ($C_6HS_4X$; top right), three ($C_6HS_3X_eY_m$ with l + m = 2; lower left) and two ($C_6HS_2X_eY_m$ with l + m = 3; lower right).

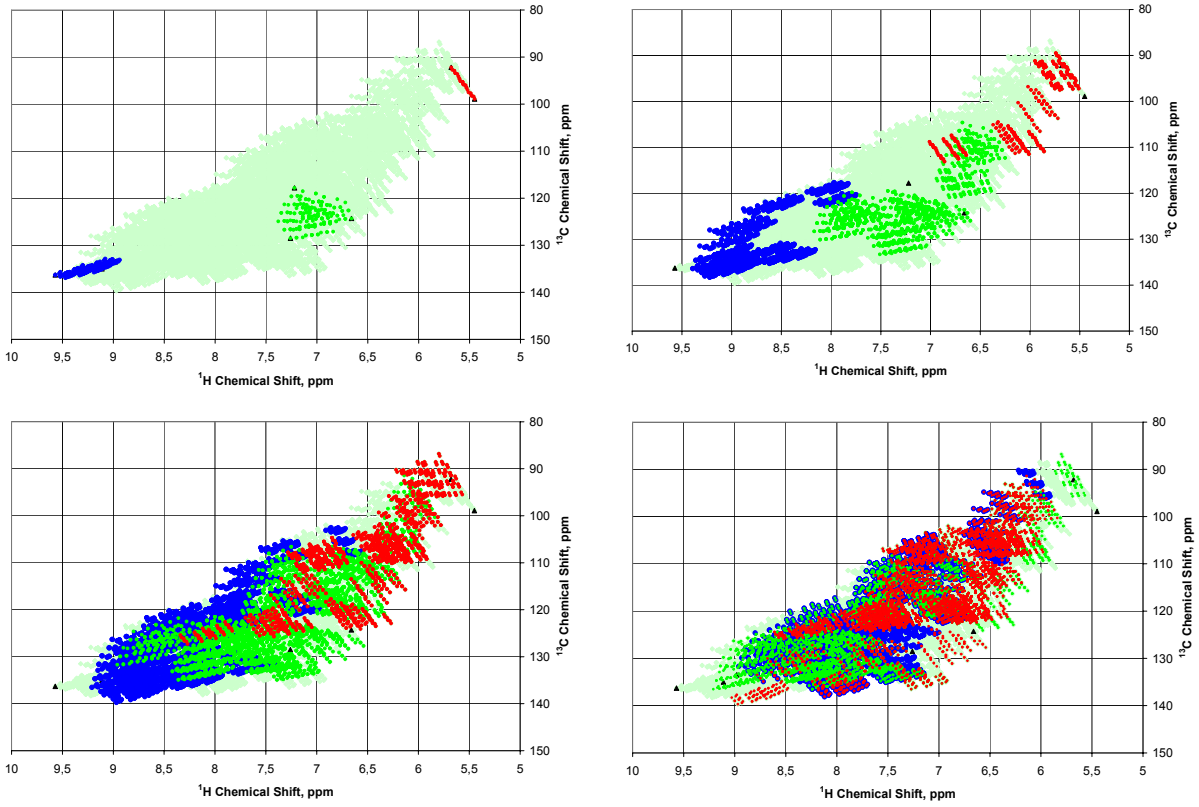



**Figure 55:** Propagation of the chemical shift space occupied by compounds $C_6HR_l(COR)_m(OR)_n$ (l+m+n=5; to left: all substituents in molecules $C_6HS_5$ are from a single class; top right: a single substituent X is from anothe class in molecules $C_6HXS_4$; lower left/right: compounds $C_6HX_2Q_qS_s$ / $C_6HX_3Q_qS_s$ contain two/three X-substituent according to color code: green: R, blue: COR, red: OR red;

The HSQC cross peaks of molecules $C_6HS_5$ fall into three cleary distinct areas of chemical shift (Figure 55, top right); these HSQC cross peaks for the molecules $C_6HR_4(COR)_m(OR)_n$ [m + n = 1; green], $C_6HR_l(COR)_4(OR)_n$ [l + n = 1; blue] and $C_6HR_l(COR)_m(OR)_4$ [l + m = 1; red] show overlap (and therefore ambiguity of substituent pattern assignment) in only a few restricted sectors ($\delta_{H/C}$ ~ 7.8/120 ppm and 6.7/110 ppm, respectively). Much more extensive overlap occurs for HSQC cross peaks of molecules $C_6HR_3(COR)_m(OR)_n$ [m + n = 2; green], $C_6HR_l(COR)_3(OR)_n$ [l + n = 2; blue] and $C_6HR_l(COR)_m(OR)_3$ [l + m = 2; red] and HSQC cross peaks for molecules $C_6HR_3(COR)_m(OR)_n$ [m + n = 3; green], $C_6HR_l(COR)_2(OR)_n$ [l + n = 3; blue], $C_6HR_l(COR)_m(OR)_2$ [l + m = 3; red], occupy near contiguous areas across a sizable portion of the entire proton and carbon chemical shift space. Overlap of HSQC cross peaks

implies ambiguity in the assignment of the respective aromatic substitution patterns since many very differently substituted aromatics can result in identical HSQC cross peak positions. It is therefore obvious, that a reliable prediction of aromatic substitution patterns is only feasible with a stringent SPARIA analysis in its inverse mode (i.e. calculation of all feasible substitution patterns from a given HSQC cross peak position).

### *5.4.4.1. Testing the accuracy of predicted chemical shifts*

The accuracy of chemical shifts that are obtained using SPARIA in the forward mode was evaluated by comparison with experimental chemical shifts of $^{1}H$ and $^{13}C$ in 29 aromatic compounds containing collectively 100 aromatic C—H groups. For comparative purposes, chemical shifts were also predicted using ACD/CNMR Predictor 5.0 and ACD/HNMR Predictor 5.0 from Advanced Chemistry Development, Inc. The 29 compounds used for this test are listed in Table 19 and their structures are given in Figure 56.

**Table 19.** Compounds Used to Test Forward and Inverse Predictions of the SPARIA Model

| | | | | | |
|---|---|---|---|---|---|
| **1** | 4-hydroxy-3-methoxyacetophenone | **11** | terephthalic acid | **21** | 2,4-dihydroxybenzaldehyde |
| **2** | 3,5-dimethoxy-4-hydroxyacetophenone | **12** | 1,2,4-trimellitic acid | **22** | m-hydroxybenzoic acid |
| **3** | 2-methoxyphenol | **13** | pyromellitic acid | **23** | protocatechoic acid |
| **4** | 3,4,5-trimethoxyacetophenone | **14** | p-hydroxybenzoic acid | **24** | gallic acid |
| **5** | 3,4-dimethoxyacetophenone | **15** | o-hydroxybenzoic acid | **25** | alpha-methylcinnamic acid |
| **6** | 4-methoxyacetophenone | **16** | ferulic acid | **26** | coumarin-3-carbonic acid |
| **7** | p-coumaric acid | **17** | vanillin | **27** | R-mandelic acid |
| **8** | m-coumaric acid | **18** | syringic acid | **28** | caffeic acid |
| **9** | o-coumaric acid | **19** | benzoic acid | **29** | homovanillic acid |
| **10** | phthalic acid | **20** | resorcinol | | |

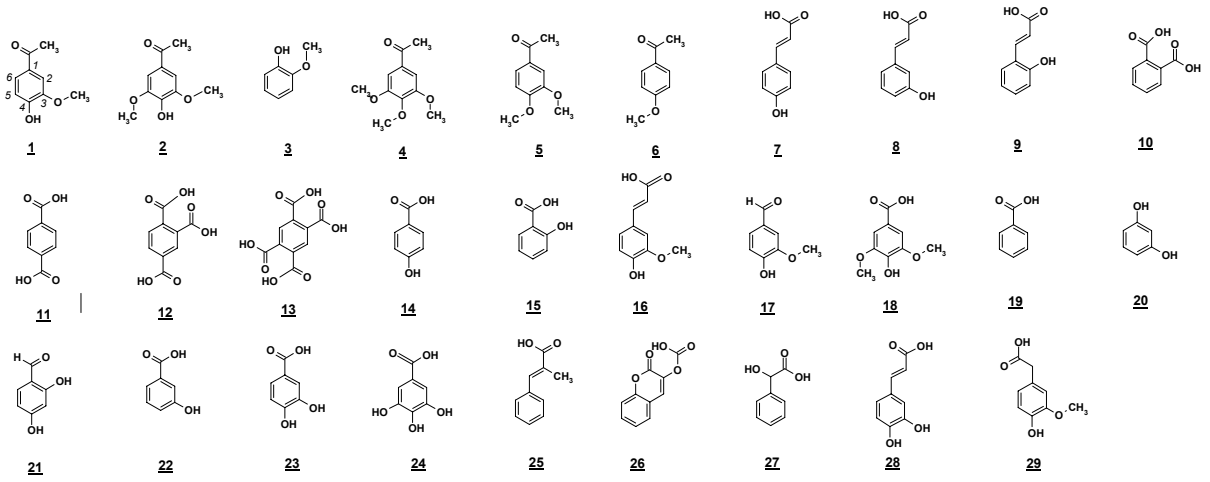


**Figure 56.** Compounds Used to Test Forward and Inverse Predictions of the SPARIA Model

In Figure 57, the 100 experimental NMR peaks for these 29 compounds are superimposed on the predicted peaks for the 16,640 unique substitution patterns. The subset of 10 peaks in Figure 57 is used later in this case study to optimize the inverse mode of SPARIA – specifically the size of the target window of chemical shift for $^{1}H$ and $^{13}C$ (Figures 60, 61, 62). Collectively, the 29 compounds used for this test contain most of the substituents that are used in SPARIA, and they are often invoked as likely structural subunits in natural organic matter. This concept and list of compounds could easily be extended for a more rigorous test of SPARIA. The accuracy of the predictions of chemical shift in the forward mode of SPARIA was quantified as the root mean square error (RMSE) between experimental and calculated chemical shifts. A similar analysis was conducted to evaluate the accuracy of the predictions of the well-known ACD models. Plots of predictions of the SPARIA and ACD models versus experimental data for $^{13}C$ and $^{1}H$ chemical shifts are shown in Figure 58, and summary results for the accuracy of the models are given in Table 21.

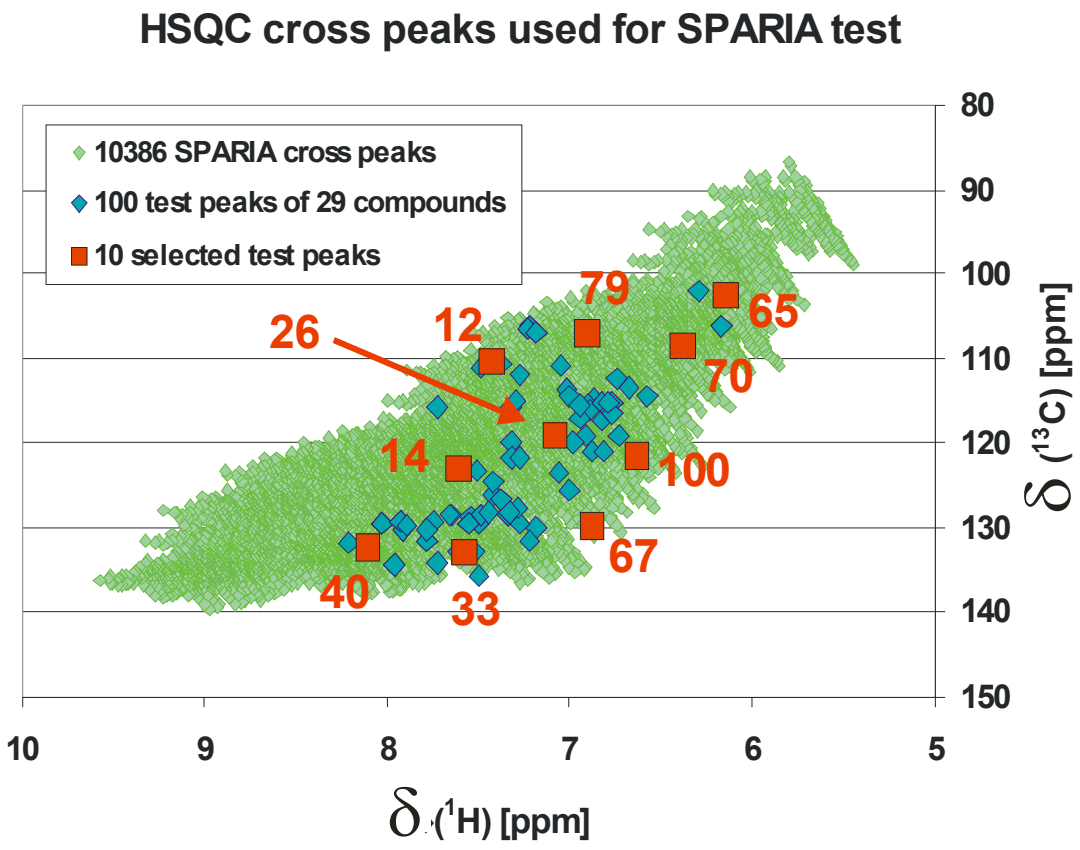



**Figure 57.** The 100 NMR peaks of 29 compounds listed in Table 20, including the subset of 10 peaks that were used to optimize the size of the peak window, both superimposed on the peaks of the 16,640 possible substitution patterns (see Figure 56).

**Table 20:** chemical shift data of 100 $^{1}H,^{13}C$-HSQC cross peaks from 29 compounds (Figure 56 and Table 19); Cmpd. ID: compound number; Peak ID: continual peak ID numbers; Position: position in aromatic ring according to Figure 56; $^{1}H$ and $^{13}C$ chemical shifts, Exp: experimental chemical shifts (in DMSO-$D_6$: 2.49/39.50 ppm), ACD: chemical shift as predicted with ACD HNMR and CNMR Software, version 5.0, INCR: chemical shift as predictet, using standard substituent increment values, listed in Table 17, RSS_A: residual sum of squares ACD $(ACD - Exp)^2$, RSS_I: residual sum of squares INCR $(INCR - Exp)^2$; ten selected HSQC cross peaks used for testing of SPARIA are highlighted in grey color.

| | | | $^{1}H$ Chemical Shift | | | | | $^{13}C$ Chemical Shift | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cmpd.ID | Peak ID | Position | Exp | ACD | INCR | RSS_A | RSS_I | Exp | ACD | INCR | RSS_A | RSS_I |
| 1 | 1 | 2 | 7,49 | 7,53 | 7,29 | 0,00 | 0,04 | 111,16 | 110,78 | 115,30 | 0,14 | 17,14 |
| 1 | 2 | 5 | 6,87 | 6,95 | 6,74 | 0,01 | 0,02 | 114,82 | 114,69 | 117,00 | 0,02 | 4,75 |
| 1 | 3 | 6 | 7,51 | 7,53 | 7,33 | 0,00 | 0,03 | 123,30 | 123,87 | 122,20 | 0,32 | 1,21 |
| 2 | 4 | 2 | 7,22 | 7,24 | 6,85 | 0,00 | 0,14 | 106,23 | 106,30 | 107,20 | 0,00 | 0,94 |
| 2 | 5 | 6 | 7,22 | 7,24 | 6,85 | 0,00 | 0,14 | 106,23 | 106,30 | 107,20 | 0,00 | 0,94 |
| 3 | 6 | 3 | 6,74 | 6,79 | 6,66 | 0,00 | 0,01 | 112,42 | 111,92 | 115,10 | 0,25 | 7,18 |
| 3 | 7 | 4 | 6,73 | 6,78 | 6,72 | 0,00 | 0,00 | 119,24 | 120,25 | 121,80 | 1,02 | 6,55 |
| 3 | 8 | 5 | 6,88 | 6,84 | 6,70 | 0,00 | 0,03 | 120,96 | 121,43 | 122,00 | 0,22 | 1,08 |
| 3 | 9 | 6 | 6,90 | 6,86 | 6,61 | 0,00 | 0,08 | 115,64 | 115,23 | 116,80 | 0,17 | 1,35 |
| 4 | 10 | 2 | 7,23 | 7,29 | 6,88 | 0,00 | 0,12 | 106,40 | 106,22 | 106,50 | 0,03 | 0,01 |
| 4 | 11 | 6 | 7,23 | 7,29 | 6,88 | 0,00 | 0,12 | 106,40 | 106,22 | 106,50 | 0,03 | 0,01 |
| 5 | 12 | 2 | 7,43 | 7,52 | 7,32 | 0,01 | 0,01 | 110,22 | 110,44 | 114,60 | 0,05 | 19,18 |
| 5 | 13 | 5 | 7,05 | 6,90 | 6,82 | 0,02 | 0,05 | 110,78 | 110,71 | 114,60 | 0,00 | 14,59 |
| 5 | 14 | 6 | 7,61 | 7,57 | 7,36 | 0,00 | 0,06 | 122,99 | 123,35 | 121,50 | 0,13 | 2,22 |
| 6 | 15 | 2 | 7,92 | 7,89 | 7,80 | 0,00 | 0,01 | 130,37 | 132,68 | 129,60 | 5,34 | 0,59 |
| 6 | 16 | 3 | 7,02 | 6,78 | 6,91 | 0,06 | 0,01 | 113,75 | 114,73 | 113,70 | 0,96 | 0,00 |
| 6 | 17 | 5 | 7,02 | 6,78 | 6,91 | 0,06 | 0,01 | 113,75 | 114,73 | 113,70 | 0,96 | 0,00 |
| 6 | 18 | 6 | 7,92 | 7,89 | 7,80 | 0,00 | 0,01 | 130,37 | 132,68 | 129,60 | 5,34 | 0,59 |
| 7 | 19 | 2 | 7,50 | 7,58 | 7,20 | 0,01 | 0,09 | 129,64 | 130,44 | 128,30 | 0,64 | 1,80 |
| 7 | 20 | 3 | 6,77 | 6,82 | 6,67 | 0,00 | 0,01 | 115,27 | 116,19 | 117,70 | 0,85 | 5,90 |
| 7 | 21 | 5 | 6,77 | 6,82 | 6,67 | 0,00 | 0,01 | 115,27 | 116,19 | 117,70 | 0,85 | 5,90 |
| 7 | 22 | 6 | 7,50 | 7,58 | 7,20 | 0,01 | 0,09 | 129,64 | 130,44 | 128,30 | 0,64 | 1,80 |
| 8 | 23 | 2 | 7,00 | 8,26 | 6,76 | 1,59 | 0,06 | 114,50 | 114,39 | 114,10 | 0,01 | 0,16 |
| 8 | 24 | 4 | 6,82 | 6,95 | 6,60 | 0,02 | 0,05 | 117,43 | 117,73 | 112,40 | 0,09 | 25,30 |
| 8 | 25 | 5 | 7,19 | 7,25 | 7,11 | 0,00 | 0,01 | 129,96 | 130,69 | 131,90 | 0,53 | 3,76 |
| 8 | 26 | 6 | 7,08 | 7,29 | 6,87 | 0,04 | 0,04 | 119,19 | 119,74 | 119,10 | 0,30 | 0,01 |
| 9 | 27 | 3 | 6,89 | 6,49 | 6,67 | 0,16 | 0,05 | 116,14 | 116,80 | 117,70 | 0,44 | 2,43 |
| 9 | 28 | 4 | 7,22 | 7,52 | 7,04 | 0,09 | 0,03 | 131,47 | 131,77 | 126,60 | 0,09 | 23,72 |
| 9 | 29 | 5 | 6,81 | 6,92 | 6,78 | 0,01 | 0,00 | 120,93 | 121,30 | 122,70 | 0,14 | 3,13 |
| 9 | 30 | 6 | 7,55 | 7,41 | 7,20 | 0,02 | 0,12 | 128,71 | 128,87 | 128,30 | 0,03 | 0,17 |
| 10 | 31 | 3 | 7,65 | 7,80 | 8,29 | 0,02 | 0,41 | 128,37 | 128,60 | 130,00 | 0,05 | 2,66 |
| 10 | 32 | 4 | 7,57 | 7,59 | 7,69 | 0,00 | 0,01 | 132,78 | 132,90 | 133,20 | 0,01 | 0,18 |
| 10 | 33 | 5 | 7,58 | 7,59 | 7,69 | 0,00 | 0,01 | 132,78 | 132,90 | 133,20 | 0,01 | 0,18 |
| 10 | 34 | 6 | 7,66 | 7,80 | 8,29 | 0,02 | 0,40 | 128,37 | 128,60 | 130,00 | 0,05 | 2,66 |
| 11 | 35 | 2 | 8,03 | 8,06 | 8,29 | 0,00 | 0,07 | 129,52 | 131,31 | 130,00 | 3,20 | 0,23 |
| 11 | 36 | 3 | 8,03 | 8,06 | 8,29 | 0,00 | 0,07 | 129,52 | 131,31 | 130,00 | 3,20 | 0,23 |
| 11 | 37 | 5 | 8,03 | 8,06 | 8,29 | 0,00 | 0,07 | 129,52 | 131,31 | 130,00 | 3,20 | 0,23 |
| 11 | 38 | 6 | 8,03 | 8,06 | 8,29 | 0,00 | 0,07 | 129,52 | 131,31 | 130,00 | 3,20 | 0,23 |
| 12 | 39 | 3 | 8,21 | 8,55 | 9,14 | 0,12 | 0,86 | 131,76 | 132,12 | 131,60 | 0,13 | 0,03 |
| 12 | 40 | 5 | 8,11 | 8,25 | 8,54 | 0,02 | 0,18 | 132,42 | 134,37 | 134,80 | 3,80 | 5,66 |

| | | | $^{1}$H Chemical Shift | | | | | $^{13}$C Chemical Shift | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cmpd ID | Peak ID | Position | Exp | ACD | INCR | RSS_A | RSS_1 | Exp | ACD | INCR | RSS_A | RSS_I |
| 12 | 41 | 6 | 7,75 | 7,99 | 8,47 | 0,06 | 0,52 | 129,29 | 130,07 | 129,90 | 0,61 | 0,37 |
| 13 | 42 | 3 | 7,96 | 8,24 | 9,32 | 0,08 | 1,85 | 134,54 | 138,30 | 131,50 | 14,14 | 9,24 |
| 13 | 43 | 6 | 7,96 | 8,24 | 9,32 | 0,08 | 1,85 | 134,54 | 138,30 | 131,50 | 14,14 | 9,24 |
| 14 | 44 | 2 | 7,78 | 7,92 | 7,99 | 0,02 | 0,04 | 131,62 | 132,49 | 131,70 | 0,76 | 0,01 |
| 14 | 45 | 3 | 6,82 | 6,91 | 6,88 | 0,01 | 0,00 | 115,18 | 115,86 | 115,80 | 0,46 | 0,38 |
| 14 | 46 | 5 | 6,82 | 6,91 | 6,88 | 0,01 | 0,00 | 115,18 | 115,86 | 115,80 | 0,46 | 0,38 |
| 14 | 47 | 6 | 7,78 | 7,92 | 7,99 | 0,02 | 0,04 | 131,62 | 132,49 | 131,70 | 0,76 | 0,01 |
| 15 | 48 | 3 | 6,95 | 7,01 | 6,88 | 0,00 | 0,00 | 117,10 | 117,80 | 115,80 | 0,49 | 1,69 |
| 15 | 49 | 4 | 7,50 | 7,53 | 7,39 | 0,00 | 0,01 | 135,69 | 136,30 | 134,90 | 0,37 | 0,62 |
| 15 | 50 | 5 | 6,92 | 6,96 | 6,99 | 0,00 | 0,00 | 119,21 | 119,80 | 120,80 | 0,35 | 2,53 |
| 15 | 51 | 6 | 7,78 | 7,93 | 7,99 | 0,02 | 0,04 | 130,30 | 131,10 | 131,70 | 0,64 | 1,96 |
| 16 | 52 | 2 | 7,27 | 7,32 | 6,72 | 0,00 | 0,30 | 111,89 | 111,22 | 113,30 | 0,45 | 1,99 |
| 16 | 53 | 5 | 6,78 | 6,88 | 6,58 | 0,01 | 0,04 | 116,22 | 115,80 | 118,60 | 0,18 | 5,66 |
| 16 | 54 | 6 | 7,06 | 7,15 | 6,76 | 0,01 | 0,09 | 123,50 | 123,07 | 120,20 | 0,18 | 10,89 |
| 17 | 55 | 2 | 7,38 | 7,45 | 7,29 | 0,00 | 0,01 | 110,68 | 110,25 | 115,30 | 0,18 | 21,34 |
| 17 | 56 | 5 | 6,95 | 7,06 | 6,74 | 0,01 | 0,04 | 115,39 | 115,37 | 117,00 | 0,00 | 2,59 |
| 17 | 57 | 6 | 7,42 | 7,48 | 7,33 | 0,00 | 0,01 | 126,11 | 127,01 | 122,20 | 0,81 | 15,29 |
| 18 | 58 | 2 | 7,19 | 7,38 | 7,07 | 0,04 | 0,01 | 106,90 | 107,72 | 108,60 | 0,67 | 2,89 |
| 18 | 59 | 6 | 7,19 | 7,38 | 7,07 | 0,04 | 0,01 | 106,90 | 107,72 | 108,60 | 0,67 | 2,89 |
| 19 | 60 | 2 | 7,93 | 8,08 | 8,11 | 0,02 | 0,03 | 129,30 | 130,27 | 130,10 | 0,94 | 0,64 |
| 19 | 61 | 3 | 7,49 | 7,46 | 7,44 | 0,00 | 0,00 | 128,53 | 129,15 | 128,40 | 0,38 | 0,02 |
| 19 | 62 | 4 | 7,61 | 7,57 | 7,51 | 0,00 | 0,01 | 132,81 | 133,98 | 133,30 | 1,37 | 0,24 |
| 19 | 63 | 5 | 7,49 | 7,46 | 7,44 | 0,00 | 0,00 | 128,53 | 129,15 | 128,40 | 0,38 | 0,02 |
| 19 | 64 | 6 | 7,93 | 8,08 | 8,11 | 0,02 | 0,03 | 129,30 | 130,27 | 130,10 | 0,94 | 0,64 |
| 20 | 65 | 2 | 6,15 | 6,46 | 6,14 | 0,10 | 0,00 | 102,60 | 104,30 | 103,30 | 2,89 | 0,49 |
| 20 | 66 | 4 | 6,17 | 6,55 | 6,25 | 0,14 | 0,01 | 106,30 | 109,10 | 108,30 | 7,84 | 4,00 |
| 20 | 67 | 5 | 6,88 | 7,01 | 7,02 | 0,02 | 0,02 | 129,80 | 132,00 | 131,70 | 4,84 | 3,61 |
| 20 | 68 | 6 | 6,17 | 6,55 | 6,25 | 0,14 | 0,01 | 106,30 | 109,10 | 108,30 | 7,84 | 4,00 |
| 21 | 69 | 3 | 6,30 | 6,17 | 6,27 | 0,02 | 0,00 | 102,10 | 103,15 | 103,50 | 1,10 | 1,96 |
| 21 | 71 | 5 | 6,38 | 6,44 | 6,38 | 0,00 | 0,00 | 108,57 | 109,63 | 108,50 | 1,12 | 0,00 |
| 21 | 70 | 6 | 7,52 | 7,61 | 7,65 | 0,01 | 0,02 | 132,86 | 131,19 | 131,90 | 2,79 | 0,92 |
| 22 | 72 | 2 | 7,32 | 7,51 | 7,55 | 0,04 | 0,05 | 115,83 | 118,30 | 117,50 | 6,10 | 2,79 |
| 22 | 73 | 4 | 6,98 | 7,09 | 6,95 | 0,01 | 0,00 | 119,83 | 121,15 | 120,70 | 1,74 | 0,76 |
| 22 | 74 | 5 | 7,27 | 7,33 | 7,32 | 0,00 | 0,00 | 129,55 | 130,40 | 130,00 | 0,72 | 0,20 |
| 22 | 75 | 6 | 7,32 | 7,51 | 7,66 | 0,04 | 0,12 | 120,00 | 122,70 | 122,50 | 7,29 | 6,25 |
| 23 | 76 | 2 | 7,32 | 7,58 | 7,43 | 0,07 | 0,01 | 121,63 | 118,30 | 119,10 | 11,09 | 6,40 |
| 23 | 77 | 5 | 6,77 | 6,94 | 6,76 | 0,03 | 0,00 | 116,52 | 117,30 | 117,40 | 0,61 | 0,77 |
| 23 | 78 | 6 | 7,28 | 7,52 | 7,54 | 0,06 | 0,07 | 121,85 | 122,90 | 124,10 | 1,10 | 5,06 |
| 24 | 79 | 2 | 6,90 | 7,09 | 6,98 | 0,04 | 0,01 | 107,05 | 111,23 | 111,50 | 17,47 | 19,80 |
| 24 | 80 | 6 | 6,90 | 7,09 | 6,98 | 0,04 | 0,01 | 107,05 | 111,23 | 111,50 | 17,47 | 19,80 |
| 25 | 81 | 2 | 7,56 | 7,87 | 7,32 | 0,10 | 0,06 | 129,43 | 129,69 | 126,70 | 0,07 | 7,45 |
| 25 | 82 | 3 | 7,34 | 7,04 | 7,23 | 0,09 | 0,01 | 128,41 | 128,96 | 130,30 | 0,30 | 3,57 |
| 25 | 83 | 4 | 7,44 | 7,18 | 7,16 | 0,07 | 0,08 | 128,30 | 128,88 | 125,00 | 0,34 | 10,89 |
| 25 | 84 | 5 | 7,34 | 7,04 | 7,23 | 0,09 | 0,01 | 128,41 | 128,96 | 130,30 | 0,30 | 3,57 |
| 25 | 85 | 6 | 7,56 | 7,87 | 7,32 | 0,10 | 0,06 | 129,43 | 129,69 | 126,70 | 0,07 | 7,45 |
| 26 | 86 | 3 | 7,89 | 7,33 | 7,23 | 0,31 | 0,44 | 129,90 | 126,71 | 127,60 | 10,18 | 5,29 |
| 26 | 87 | 4 | 7,42 | 7,14 | 6,79 | 0,08 | 0,40 | 124,60 | 125,16 | 122,20 | 0,31 | 5,76 |

| | | | $^{1}$H Chemical Shift | | | | | $^{13}$C Chemical Shift | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CmpdID | Peak ID | Position | Exp | ACD | INCR | RSS_A | RSS_1 | Exp | ACD | INCR | RSS_A | RSS_I |
| 26 | 88 | 5 | 7,73 | 7,23 | 7,07 | 0,25 | 0,44 | 134,10 | 127,50 | 125,90 | 43,56 | 67,24 |
| 26 | 89 | 6 | 7,73 | 7,20 | 6,75 | 0,28 | 0,96 | 115,80 | 116,26 | 115,30 | 0,21 | 0,25 |
| 27 | 90 | 2 | 7,38 | 7,66 | 7,11 | 0,08 | 0,07 | 126,57 | 125,60 | 127,90 | 0,94 | 1,77 |
| 27 | 91 | 3 | 7,33 | 7,51 | 7,20 | 0,03 | 0,02 | 128,18 | 128,01 | 128,40 | 0,03 | 0,05 |
| 27 | 92 | 4 | 7,29 | 7,21 | 7,08 | 0,01 | 0,04 | 127,80 | 128,69 | 125,70 | 0,79 | 4,41 |
| 27 | 93 | 5 | 7,33 | 7,51 | 7,20 | 0,03 | 0,02 | 128,18 | 128,01 | 128,40 | 0,03 | 0,05 |
| 27 | 94 | 6 | 7,38 | 7,66 | 7,11 | 0,08 | 0,07 | 126,57 | 125,60 | 127,90 | 0,94 | 1,77 |
| 28 | 95 | 2 | 7,30 | 7,32 | 6,64 | 0,00 | 0,44 | 115,04 | 115,63 | 115,70 | 0,35 | 0,44 |
| 28 | 96 | 5 | 6,58 | 6,82 | 6,55 | 0,06 | 0,00 | 114,56 | 113,78 | 119,30 | 0,61 | 22,47 |
| 28 | 97 | 6 | 7,00 | 6,99 | 6,75 | 0,00 | 0,06 | 125,66 | 124,16 | 120,70 | 2,25 | 24,60 |
| 29 | 98 | 2 | 6,68 | 6,97 | 6,51 | 0,08 | 0,03 | 113,57 | 114,00 | 114,50 | 0,18 | 0,86 |
| 29 | 99 | 5 | 6,79 | 7,17 | 6,55 | 0,14 | 0,06 | 115,23 | 116,00 | 116,70 | 0,59 | 2,16 |
| 29 | 100 | 6 | 6,63 | 6,89 | 6,55 | 0,07 | 0,01 | 121,61 | 122,80 | 121,40 | 1,42 | 0,04 |

**Table 20 (continued):** chemical shift data of 100 $^{1}$H,$^{13}$C-HSQC cross peaks from 29 compounds (Figure 56 and Table 19); Cmpd. ID: compound number; Peak ID: continual peak ID numbers; Position: position in aromatic ring according to Figure 56; $^{1}$H and $^{13}$C chemical shifts, Exp: experimental chemical shifts (in DMSO-$D_6$: 2.49/39.50 ppm), ACD: chemical shift as predicted with ACD HNMR and CNMR Software, version 5.0, INCR: chemical shift as predictet, using standard substituent increment values, listed in Table 2, RSS_A: residual sum of squares ACD $(ACD - Exp)^2$, RSS_I: residual sum of squares INCR $(INCR - Exp)^2$; ten selected HSQC cross peaks used for testing of SPARIA are highlighted in grey color.

There is considerably greater relative scatter in the prediction of the chemical shifts of $^{1}$H than of $^{13}$C. Linear regression fits (not shown) of predicted versus experimental chemical shifts for the $^{1}$H data have $r^2$ values of 0.73 (SPARIA) and 0.79 (ACD). Similar fits for $^{13}$C have $r^2$ values of 0.94 (SPARIA) and 0.98 (ACD).

**Table 21.** Linear Regression Analysis and Root Mean Square Error (in ppm) for Chemical Shifts of $^{1}$H and $^{13}$C in the 29 Compounds in Table 19.

| | Regression Parameters for $^{1}$H | | | | Regression Parameters for $^{13}$C | | | |
|---|---|---|---|---|---|---|---|---|
| Model | RMSE | Intercept | Slope | $R^2$ | RMSE | Intercept | Slope | $R^2$ |
| ACD | 0.23 | 0.69 | 0.92 | 0.79 | 1.52 | 2.32 | 0.99 | 0.98 |
| SPARIA - Standard | 0.35 | -1.55 | 1.21 | 0.73 | 2.27 | 13.27 | 0.89 | 0.94 |
| SPARIA - COOX | 0.25 | -0.31 | 1.03 | 0.81 | 2.22 | 13.91 | 0.89 | 0.94 |
| SPARIA - CH=CHX | 0.33 | -1.41 | 1.19 | 0.75 | 1.94 | 12.77 | 0.90 | 0.96 |
| SPARIA – Both X | 0.22 | -0.17 | 1.02 | 0.83 | 1.91 | 12.45 | 0.90 | 0.96 |

- COOX represents o-dicarboxylic acids and CH=CHX represents –CH=CH—COOH in limited tests to improve the forward predictions of SPARIA. These modifications were conducted separately and together (Both X). See discussion in the text.

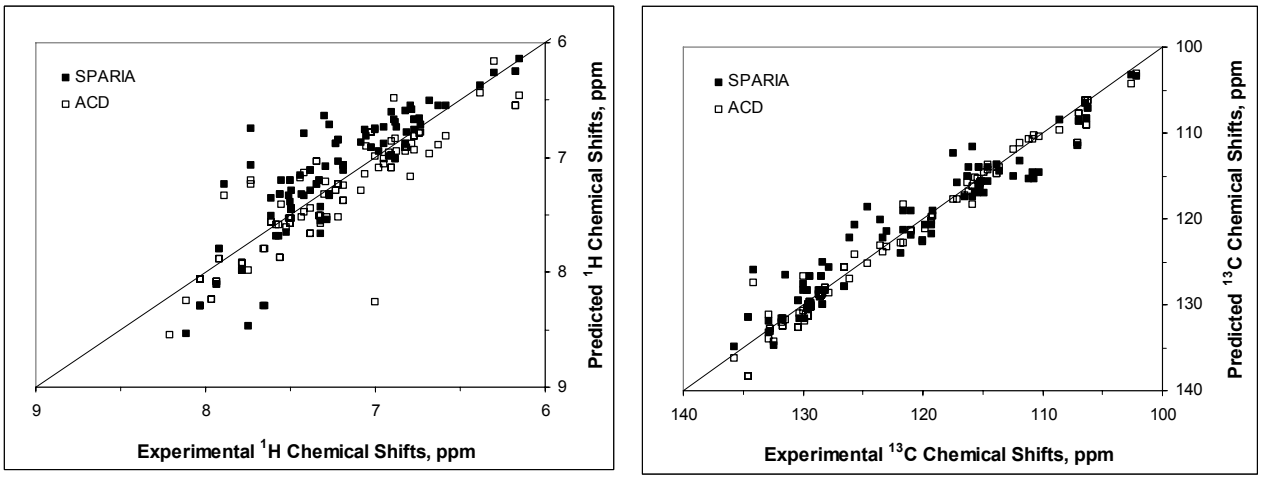



**Figure 58.** Predicted versus Experimental Chemical Shifts for $^{1}H$ (left) and $^{13}C$ (right)

In the course of using SPARIA in the forward mode, it became clear that predictions of the chemical shift of $^{1}H$ were often poor in molecules containing COOH groups that were ortho to each other (see **10**, **12** and **13** in Table 20). The incremental chemical shifts of COOH were optimized for such groups using the solver tool in Microsoft Excel.

In these molecules, torsional strain from steric interaction between ortho-carboxylic groups weakens mesomeric effects while polar electron withdrawal is affected to a lesser extent. Furthermore, a variable alignment of the carbonyl (C=O) bond vector alters the chemical shift anisotropy experienced by neighbouring atoms (Abraham et al., 2003) , and both these interactions more strongly affect aromatic protons than carbon atoms. Because dipolar interactions act through space rather than through δ-bonds, the polar electron-withdrawal, which affects carbon more strongly than proton chemical shifts, declines in the order –COOH > -CH=CH-COOH > -CH=$CH_2$, therefore these effects diminish in the order ortho > meta > para. In contrast, mesomeric effects are most pronounced in the extensively conjugated π-electrons of the cinnamic acid derivatives. The interplay of these individual mechanisms results in corrected (designed to minimize RMSE) increments, named –CH=CHX in Table 21, whose values fall in between those of vinyl (-CH=$CH_2$) and carboxyl (-COOH).

With that alteration alone, the overall RMSE for proton chemical shifts drops from 0.35 ppm (SPARIA-Standard) to 0.25 ppm, while RMSE for carbon chemical shifts remains nearly unaltered (cf. Table 21). This behaviour indicates that steric interactions referring to the distance and orientation of the carboxylic groups in relation to the aromatic protons are most

significant; however a reduced mesomeric electron withdrawal, caused by out-of-plane bending of the carbonyl group is indicated by the para 0.12 ppm upfield proton NMR chemical shift.

Mesomeric substituent effects, causing a disturbance of ring currents, more strongly affect proton than carbon chemical shifts in aromatic rings.

**Figure 59:** main resonance structures of vinylbenzene (left) and cinnamic acid (right); vinyl substituents exhibit considerable electron-donating properites which do not appropriately reflect the rather electron-withdrawing capacity of the vinyl-carboxylic acid .

Similarly, it was observed that predictions of the chemical shift of $^{13}C$ were poor in structures containing the –CH=CH—COOH group (see **7**, **8**, **9**, **16**, **25** and **28** in Table 20). This resonance electron-withdrawing substituent is not adequately represented by the resonance electron-donating $–CH=CH_2$ substituent, which is used in SPARIA. A new set of incremental chemical shifts was obtained using the solver in Microsoft Excel for such structures. The optimized incremental chemical shifts for torsionally strained –COOH and for –CH=CH—COOH are given in Table 22.

**Table 22.** Incremental chemical shifts for selected substituents on aromatic rings – $^{1}H$ and $^{13}C$ incremental chemical shifts are relative to 7.26 ppm and 128.5 ppm, respectively.

| | $^{1}H$ Incremental Chemical Shift | | | $^{13}C$ Incremental Chemical Shift | | |
|---|---|---|---|---|---|---|
| | **Ortho** | **meta** | **para** | **ortho** | **meta** | **para** |
| **—COOX** * | 0.46 | -0.13 | 0.34 | 2.4 | -0.2 | 3.8 |
| **—CH=CHX** * | 0.39 | 0.08 | 0.13 | -0.4 | -0.5 | 0.9 |

* COOX represents o-dicarboxylic acids and CH=CHX represents –CH=CH—COOH in limited tests to improve the forward predictions of SPARIA.

When SPARIA uses the modified incremental chemical shifts from Table 22, the agreement between observed and calculated chemical shifts improves considerably (see Table 21). When compared with the RMSE values for the standard SPARIA model, it is seen in Table 21 that the RMSE of predicted $^{1}H$ chemical shifts is improved dramatically by use of the optimized parameters for –COOX; however, this adjustment had little effect on the RMSE of predicted

$^{13}C$ chemical shifts. Torsional strain from steric interaction between ortho-COOH groups weakens mesomeric effects but has less impact on polar electron withdrawal. Furthermore, a variable alignment of the carbonyl (C=O) bond vector alters the chemical shift anisotropy experienced by neighboring atoms, and both these interactions more strongly affect aromatic protons than carbon atoms. Conversely, the RMSE of predicted $^{13}C$ chemical shifts is improved dramatically by use of the optimized parameters for –CH=CH–COOX; however, this adjustment had little effect on the RMSE of predicted $^{1}H$ chemical shifts. Given that the ultimate goal is to apply SPARIA to natural organic matter, for which it is impossible to know if/when standard incremental chemical shifts of substituents should be modified, only the standard incremental chemical shifts of substituents in Table 17 will be used subsequently in this case study.

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

### *5.4.5. Inverse Mode – Predicting Substitution Patterns*

The inverse mode of SPARIA is used to predict substitution patterns from the chemical shifts of aromatic $^{1}H$ and $^{13}C$. A rectangular window of arbitrary size, which is centered around the peak of interest, is first defined. If the size of the window approaches zero in both dimensions, it is unlikely that any predicted peak will fall within that window. Conversely, if the window approaches the full range of possible chemical shifts for aromatic $^{1}H$ and $^{13}C$, then it is likely that all possible peaks will fall within that window. In either extreme case, no useful information can be obtained. The size of the window must therefore be optimized to obtain a statistically reasonable number of possible substitution patterns and to yield results of the greatest possible accuracy.

Before optimizing the window size in the inverse mode of SPARIA, it is instructive to demonstrate the inverse mode for an arbitrary window size of ±0.1 ppm for $^{1}H$ and ±1.0 ppm for $^{13}C$. The overall process is illustrated in Table 23 for one of the aromatic C—H groups in 3,4,5-trimethoxyacetophenone (**4** in Figure 56). Using the experimentally determined chemical shifts of $^{1}H$ and $^{13}C$ for this compound (7.23 ppm and 106.4 ppm, respectively), together with the aforementioned window size, the target window is 7.13 – 7.33 ppm by 105.4 – 107.4 ppm. The inverse mode of SPARIA finds only 80 substitution patterns (from the total of 16,640) for which predicted chemical shifts fall within the target window. Ironically, the SPARIA substitution pattern that most closely matches the chemical structure of 3,4,5-trimethoxyacetophenone ($COC_2H_5$, H, $OCH_3$, $OCH_3$, $OCH_3$) is not among the 80 solutions obtained from the inverse mode of SPARIA. A forward mode calculation for that substitution pattern yields predicted chemical shifts of 6.88 and 106.5 ppm for $^{1}H$ and $^{13}C$, respectively, so the predicted $^{1}H$ chemical shift for that substitution pattern lies slightly outside the target window. In the interest of brevity, only a subset of eleven representative substitution patterns is given in Table 23. Basic statistics are provided for that subset of substitution patterns.

Two statistical analyses of the substituent patterns are provided in Table 23 – an analysis of the probability distribution of substituents on each of the five ring positions and a prediction of the number of each *class* of substituents (COR, R, and OR) on each of the five ring positions. The latter prediction is compared with the actual substitution pattern of 3,4,5-trimethoxy-acetophenone. It is clear, in the case of this example, that one ortho position will almost certainly contain an electron-withdrawing COR group ($COCH_3$), and the other ortho position will almost certainly contain an electron-donating OR group ($OCH_3$). The para position is highly likely to contain an electron-donating OR group ($OCH_3$). Both of the meta positions on the ring are predicted to have a neutral R group or an electron-withdrawing COR group (R is slightly more probable). This prediction is correct for one of the meta positions (H), but it is incorrect for the other ($OCH_3$). In summary, the most probable substitution pattern is correct for four of the five ring positions (80% accuracy).

Errors in the predictions of the inverse mode of SPARIA are calculated by position and class of substituent as (SPARIA – Actual), and the overall error in the predicted substitution pattern is expressed as the root mean square of those 15 separate errors (RMSE).

Although the substitution patterns in Table 23 exhibit a strong preference for substituents in one ortho position versus the other, no such preference is found for the meta positions. In both

cases, the effect of a given substituent on chemical shift is the same from either ortho position or from either meta position. The strong positional preference of ortho substituents in Table 23 is a direct consequence of the process by which redundant substitution patterns were removed from the original database of 32,768 substitution patterns to obtain the final list of 16,640 unique substitution patterns (see earlier discussion).

**Table 23.** Inverse mode of SPARIA (illustrated using 11 of 80 matching substitution patterns for an aromatic C—H group in 3,4,5-trimethoxyacetophenone (**4** in Figure 56).

| | Observed Peak, ppm | | Peak Window, ppm | | | | |
|---|---|---|---|---|---|---|---|
| **Input to SPARIA** | **δ $^{1}$H** | **δ $^{13}$C** | **Δ(δ $^{1}$H)** | **Δ(δ $^{13}$C)** | | | |
| | 7.23 | 106.40 | 0.10 | 1.00 | | | |
| **Output from SPARIA** | **δ $^{1}$H** | **δ $^{13}$C** | **Ortho** | **Meta** | **Para** | **Meta'** | **Ortho'** |
| | 7.13 | 106.9 | COOH | $C_2H_5$ | $OCH_3$ | H | $OCH_3$ |
| | 7.13 | 106.9 | COOH | H | $OCH_3$ | $C_2H_5$ | $OCH_3$ |
| | 7.17 | 106.2 | $COOCH_3$ | $C_2H_5$ | $OCH_3$ | COOH | $OCH_3$ |
| | 7.17 | 106.2 | $COOCH_3$ | COOH | $OCH_3$ | $C_2H_5$ | $OCH_3$ |
| | 7.21 | 105.8 | $COC_2H_5$ | $COC_2H_5$ | $OCH_3$ | $COOCH_3$ | $OCH_3$ |
| | 7.21 | 105.8 | $COC_2H_5$ | $COOCH_3$ | $OCH_3$ | $COC_2H_5$ | $OCH_3$ |
| | 7.25 | 106.6 | $COC_2H_5$ | $CH=CH_2$ | $CH=CH_2$ | $CH=CH_2$ | $OCH_3$ |
| | 7.28 | 105.9 | COOH | $CH=CH_2$ | OH | $COC_2H_5$ | $OCH_3$ |
| | 7.28 | 105.9 | COOH | $COC_2H_5$ | OH | $CH=CH_2$ | $OCH_3$ |
| | 7.31 | 106.8 | COOH | $C_2H_5$ | $OCH_3$ | COOH | $OCH_3$ |
| | 7.31 | 106.8 | COOH | COOH | $OCH_3$ | $C_2H_5$ | $OCH_3$ |
| **Convert Substituents to Counts** | | | **Ortho** | **Meta** | **Para** | **Meta'** | **Ortho'** |
| | | **COOH** | 6 | 2 | 0 | 2 | 0 |
| | | **$COOCH_3$** | 2 | 1 | 0 | 1 | 0 |
| | | **$COC_2H_5$** | 3 | 2 | 0 | 2 | 0 |
| | | **H** | 0 | 1 | 0 | 1 | 0 |
| | | **$C_2H_5$** | 0 | 3 | 0 | 3 | 0 |
| | | **$CH=CH_2$** | 0 | 2 | 1 | 2 | 0 |
| | | **OH** | 0 | 0 | 2 | 0 | 0 |
| | | **$OCH_3$** | 0 | 0 | 8 | 0 | 11 |
| **Reduce to Classes of Substituents** [a] | | | **Ortho** | **Meta** | **Para** | **Meta'** | **Ortho'** |
| | | **COR** | 1.00 / 1 | 0.45 / 0 | 0.00 / 0 | 0.45 / 0 | 0.00 / 0 |
| | | **R** | 0.00 / 0 | 0.55 / 1 | 0.09 / 0 | 0.55 / 0 | 0.00 / 0 |
| | | **OR** | 0.00 / 0 | 0.00 / 0 | 0.91 / 1 | 0.00 / 1 | 1.00 / 1 |
| **Error (SPARIA – Actual)** | | | **Ortho** | **Meta** | **Para** | **Meta'** | **Ortho'** |
| | | **COR** | 0.00 | 0.45 | 0.00 | 0.45 | 0.00 |
| **RMSE =** 0.36 | | **R** | 0.00 | -0.45 | 0.09 | 0.55 | 0.00 |
| | | **OR** | 0.00 | 0.00 | -0.09 | -1.00 | 0.00 |

[a] Overall distribution of classes of substituents (see Tables 16, 17) for the subset of 11 matching substitution patterns and for the actual structure of 3,4,5-trimethoxyacetophenone (SPARIA / Actual).

The inverse mode of SPARIA leads to the correct prediction of the substitution patterns for the ortho and para positions of 3,4,5-trimethoxyacetophenone, but it fails to predict the presence of an OR group in one of the meta positions. The relatively poor quality of predicted substitution patterns in the meta position is inherent in the relative insensitivity of the NMR peak positions of $^{1}H$ and $^{13}C$ to meta substituents. In Table 17, for example, the entire range of $^{1}H$ incremental chemical shifts for meta substituents is only 0.3 ppm, which is slightly larger than the width of the target window used in this example. The corresponding range for $^{13}C$ is only 1.9 ppm, which is actually smaller than the height of the target window. This means that many combinations of meta-substituents can lead to peaks that fall within the target window, making it rather difficult to obtain accurate predictions. In contrast, the entire range of $^{1}H$ incremental chemical shifts for ortho substituents is 1.4 ppm, and the corresponding range for $^{13}C$ is 16.6 ppm, both of which are 7 –8 times larger than the window size used here. For para-substituents, the corresponding ranges of chemical shift are 0.7 ppm and 12.9 ppm, respectively, which are about 3 – 6 times greater than the window size used here. For both ortho- and para-substituents, only limited combinations of substituents can lead to peaks that fall within the target window, and the predictions for those positions are thus more reliable.

A closer inspection of the actual groups that are predicted in each ring position offers additional insight into the predictive abilities and limitations of the inverse mode of SPARIA. In the ortho position, the probability of occurrence of a COR group is 100%; however, there is a reasonable probability of occurrence of any of the three groups of that type (COOH, $COOCH_3$, $COC_2H_5$), with COOH being somewhat more likely to occur. (In the full list of 80 matching substitution patterns (not shown), the three substituents occur with almost equal probability.) In the other ortho position, the probability of occurrence of an OR group is 100%, and only $OCH_3$ is found among the matching substitution patterns in Table 23. Likewise, the probability of occurrence of an OR group in the para position is 91%, and $OCH_3$ appears in 8 of the 11 matching substitution patterns. At least in the case of this example, SPARIA more easily distinguishes among OR groups than among COR groups.

#### *5.4.5.1. Optimizing Window Size in the Inverse Mode of SPARIA*

To optimize the standard window size for the inverse mode of SPARIA, ten of the 100 peaks for the 29 compounds in Table 20 were selected to provide a variety of substitution patterns

and a broad range of chemical shift for both $^{1}H$ and $^{13}C$ (see Figure 57). For each selected peak, the half-width of the chemical shift window for $^{13}C$ ($\Delta(\delta\ ^{13}C)$) was varied from 0.25 to 5.00 ppm, and $\Delta(\delta\ ^{1}H)$ was fixed at $\Delta(\delta\ ^{13}C)/10$.

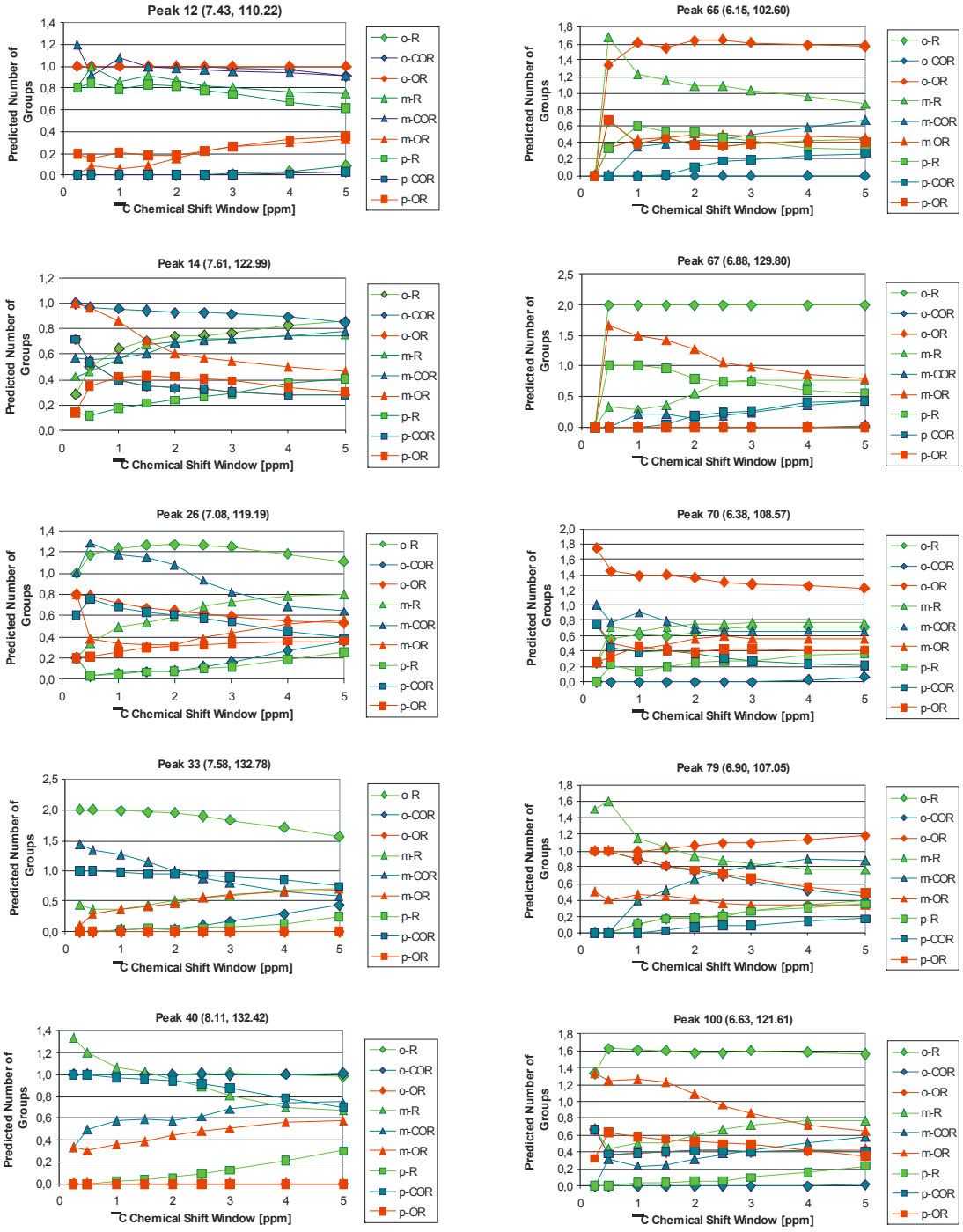



**Figure 60:** Location of 10 HSQC cross peaks given in Table 20 and within here selected for testing the inverse mode of SPARIA together with 100 experimental HSQC cross peaks of 29 compounds (Figure 57).

More detailed results are shown in Figure 60. In general, the evenness of the predicted substitution patterns varies according to substituent position within the aromatic ring rather than with the location of the HSQC cross peak relative to the entire chemical shift space (Figure 57). For example, the central peak 14 shows a strong decrease of meta-OR-substitution with increasing window size, while ortho-R and meta-COR counts are rising. Peak 33, located near the carbon low field limit, exhibits moderate decline of ortho-OR and meta-COR and increases meta-OR, ortho-COR and para-R counts of substituents predicted at increasing window size. In contrast, meta-COR strongly rises in the upfield-carbon peak 79, while ortho-COR and para-OR substituent counts decline with increased window size.

At very small half-widths of the δ $^{13}$C window, the number of likely substitution patterns is too small to yield reliable statistics. This problem can be overcome by increasing Δ(δ $^{13}$C); however, if Δ(δ $^{13}$C) becomes too large, the discriminating power of SPARIA will gradually decline and the predicted substitution pattern will mathematically approach the average substitution pattern of the entire database.

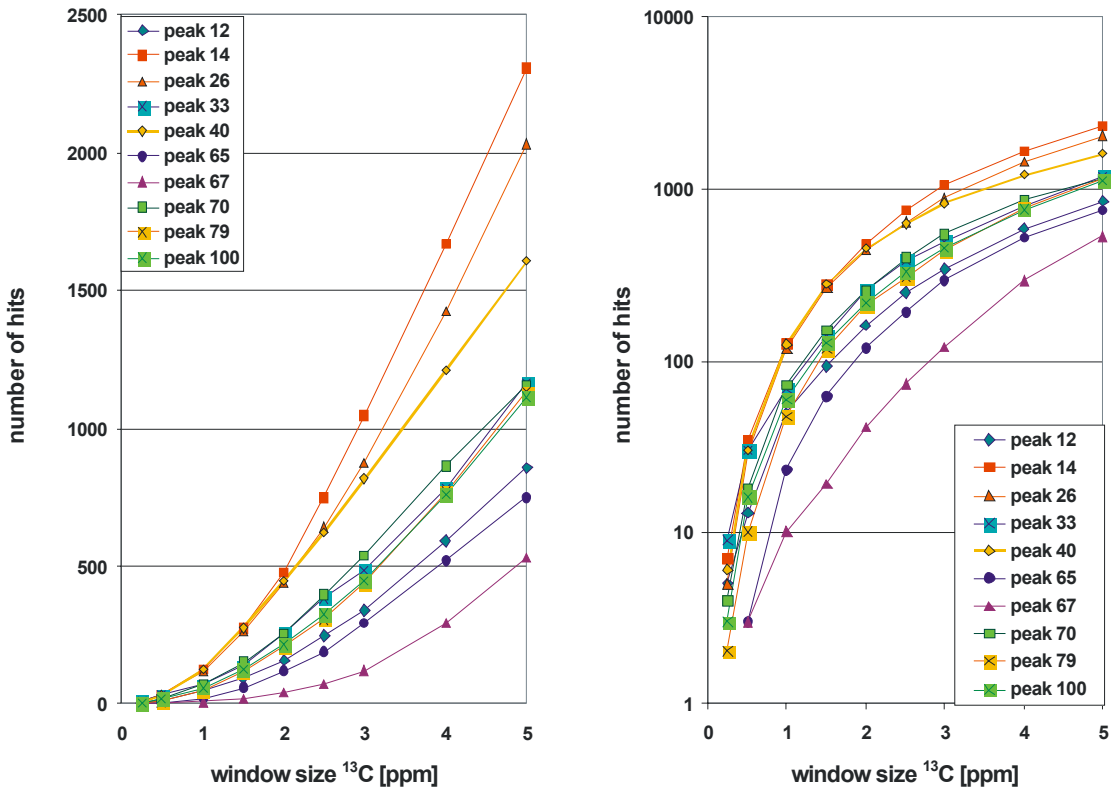



**Figure 61:** number of hits in SPARIA inverse mode depending on window size (given: half-width of the chemical shift window of $^{13}$C [Δ(δ$^{13}$C) = 10 Δ(δ$^{1}$H)]; left/right: standard/logarithmic scale

The number of hits not only depends on the window size, but also on the position of the HSQC cross peak within the entire chemical shift space (Figure 61). Centrally located peaks, like peak 14, peak 26 and also peak 40, consistently feature larger numbers of hits than HSQC cross peaks located near the (carbon) upfield (peak 12) or lowfield (peak 67) edge of the accessible shift space. It is interesting to note, that the rather centered peak 65 also exhibits low counts across the whole range of shift windows. As seen from Figure 52, only aromatics with a single or zero carbonyl derivative substituents will occupy this section of chemical shifts – analogously, highly alkylated aromatics will also not fall into that range, leaving a lesser number of feasible molecular compositions to fit into the described chemical shift range. Another similar reason for the lower count numbers in the upfield sector is because only two oxygen-containing substituents (OH and $OCH_3$) contribute to extreme upfield increments (instead of three COR,R substituents each in the central and downfield section, respectively; cf. Table 17).

For each window size, the inverse mode of SPARIA was used to generate an average substitution pattern for each peak, and the overall accuracy of each average substitution pattern was expressed as the root mean square error (RMSE) of predicted versus actual numbers of –R, –COR, and –OR groups at the ortho, meta, and para ring positions. Results are given in Figure 62.

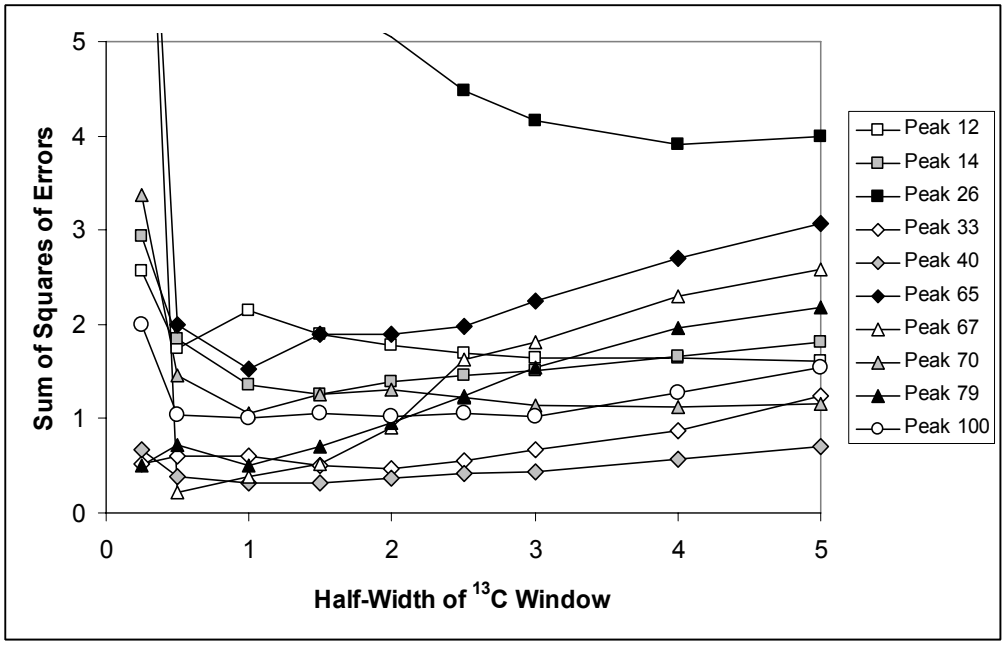



**Figure 62.** Error Analysis of Predicted Substitution Patterns Versus Half-Width of the Chemical Shift Window of $^{13}C$ ($\Delta(\delta\ ^{13}C)$.

At very small half-widths of the $\delta\ ^{13}C$ window, the number of likely substitution patterns is too small to yield reliable statistics. This problem can be overcome by increasing $\Delta(\delta\ ^{13}C)$; however, if $\Delta(\delta\ ^{13}C)$ becomes too large, the discriminating power of SPARIA will gradually decline and the predicted substitution pattern will mathematically approach the average substitution pattern of the entire database. With the exception of Peaks 26 and 40, whose substitution patterns are not well predicted at any window size, the general pattern shows the expected behavior. Although different results would likely be obtained on any other arbitrarily selected set of peaks, the average trend in Figure 62 (- - - -) suggests that a $\Delta(\delta\ ^{13}C)$ value of 1 ppm should be a suitable choice as the standard window size for the inverse mode of SPARIA. The corresponding standard value of $\Delta(\delta\ ^{1}H)$ is 0.1 ppm. These window sizes are slightly smaller than the RMSE values (see Table 21) for predicted chemical shifts of $^{1}H$ (0.35 ppm) and $^{13}C$ (2.27 ppm), which were obtained using the forward mode of SPARIA with standard incremental chemical shifts (see Table 17).

*5.4.5.2. Testing the Inverse Mode of SPARIA.*

Having selected standard half-widths of 1.0 and 0.1 ppm for $^{13}C$ and $^{1}H$, respectively, the chemical shifts of the 100 peak for the 29 compounds in Table 19 were used to test the predictive capabilities of the inverse mode of SPARIA. Starting with a given pair of chemical shifts for aromatic $^{1}H$ and $^{13}C$, the entire database of 16,640 unique substitution patterns for eight substituents on five ring positions was searched for those substitution patterns whose chemical shifts for aromatic $^{1}H$ and $^{13}C$ were within the specified window of chemical shift centered around the given peak position. The average fractions of neutral, electron-donating, and electron-withdrawing substituents were calculated for the ortho, meta, para, meta', and ortho' positions on the ring containing the aromatic C—H group whose peak was being analyzed. The predicted distribution was then compared with the known distribution of these classes of substituents in the compound whose peak was being analyzed.

For each predicted substitution pattern, the error of the prediction is calculated at each of the five ring positions for each of the three classes of substituents, i.e., a total of fifteen individual errors, and the overall error in the prediction is expressed as the root mean square of the 15 individual errors. The frequency and cumulative distributions of errors from the inverse mode

of SPARIA for the 100 peaks used previously in this case study to test the forward mode of SPARIA are given in Figure 23.

All classes of substituents, but most especially –OR, are very well predicted at ortho positions. Some additional details are visible in the cumulative error plot for ortho substituents. Because of the highly accurate predictions of –OR substituents in the ortho and ortho' positions, compensatory errors are found in the predictions of –R and –COR groups in the ortho and ortho' positions. The moderate underestimation of –COR groups in the ortho position, for example, is almost exactly balanced by a corresponding overestimation of –R groups in that position. An opposite, but also compensatory, pattern of errors is found in the predictions of –R and –COR groups in the ortho' position.

For –OR, –COR, and –R groups in meta positions, the predictions of SPARIA are generally unreliable. The very narrow range of incremental chemical shifts for both $^{1}H$ and $^{13}C$, relative to the size of the standard target window used in the inverse mode of SPARIA, and its impact on the predictions of the inverse mode of SPARIA have already been discussed. The errors for the three classes of substituents (–OR, –COR, and –R) in the meta and meta' positions are clustered around -0.6 and +0.3. If the predictions were entirely random, the most probable errors would be -2/3 and +1/3, corresponding to a prediction of 0.33 by the inverse mode of SPARIA when the correct value is either one or zero.

For the para position, the occurrence of –OR groups is generally well predicted, as stated previously. There is a strong tendency, as is evident in the cumulative plot, for –R groups to be underestimated and for –COR groups to be correspondingly overestimated.

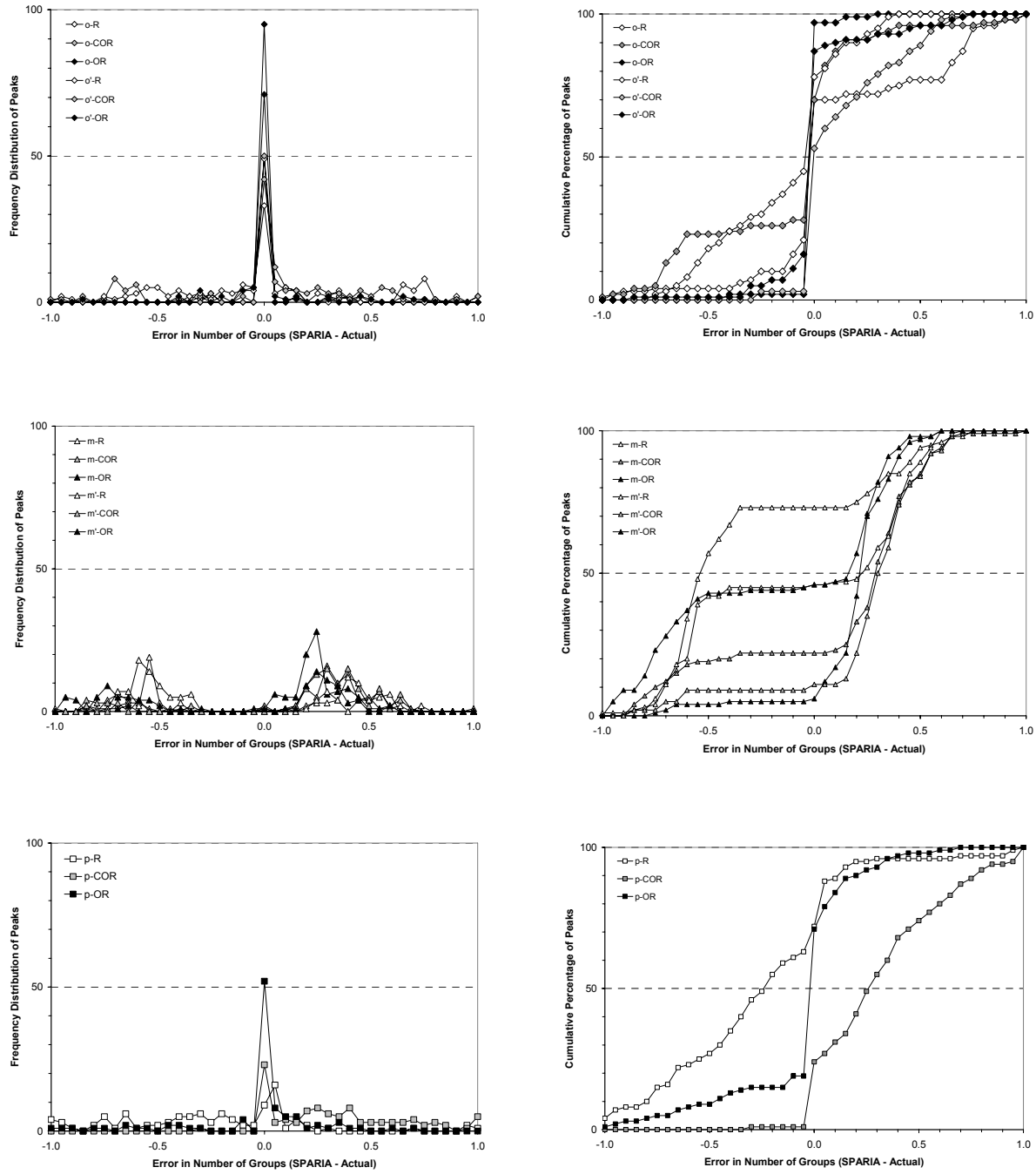



**Figure 63.** Errors in predictions of the inverse mode of SPARIA, including frequency distributions and cumulative distributions of error.

Another view of the predictive capabilities of the inverse mode of SPARIA is obtained by rounding each error according to the following rules:

- If -1 <= Error <= -0.5, then Error = -1
- If -0.5 < Error <= 0.5, then Error = 0
- If 0.5 < Error <= 1, then Error = 1

The predictions of the inverse mode of SPARIA for all three classes of substituents at the ortho, meta, and para ring positions were plotted versus this "rounded" error, and the results are given in Figure 64.

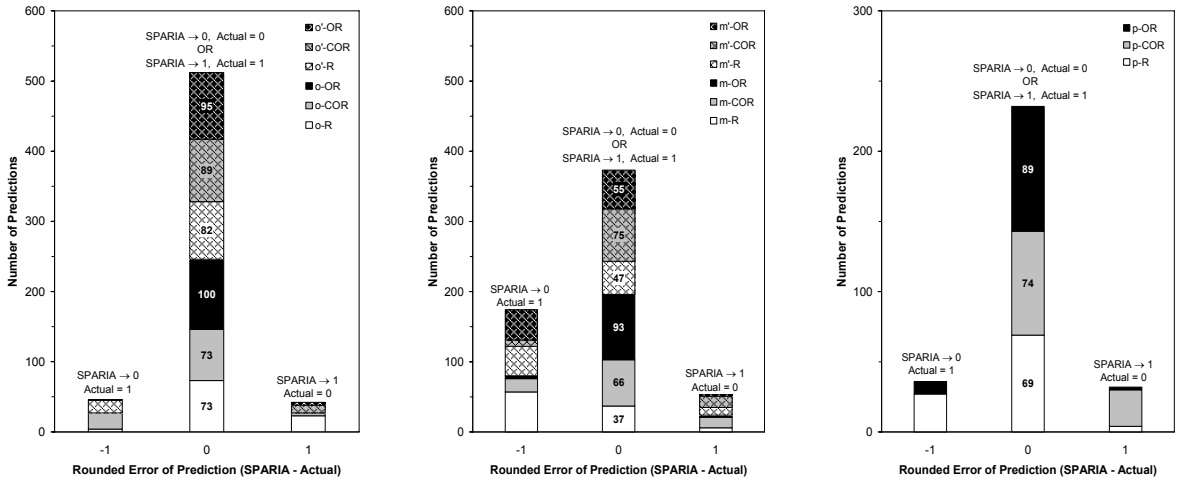



**Figure 64.** Overall summary of the performance of the inverse mode of SPARIA for three classes of substituents at ortho, meta, and para ring positions, when errors are rounded (see text for discussion). Numbers inside the data bars represent the number of correct predictions for a particular class of substituent and ring position (maximum = 100).

The overall percentages of correct predictions (using rounded errors) from the inverse mode of SPARIA are 85%, 62%, and 77%, respectively, at the ortho, meta, and para ring positions. When all such results are combined, the global percentage of correct predictions is 74%. In the case of predictions of substitution patterns at ortho and para positions, the false positive and false negative predictions are very nearly equal. Because of the nearly random nature of predicted substituents at meta positions, as has been discussed earlier in this case study, false

negative predictions are more abundant than are false positive predictions, many of which were converted into "correct" values by the error-rounding process. When the inverse mode of SPARIA is applied to natural organic matter (next section), it is anticipated that predictions of substitution patterns will be more accurate for ortho and para substituents than for meta substituents. Having said that, it remains the case that 74% of the predictions of the presence or absence of –OR groups in both meta positions were correct, equaling the overall performance of the inverse mode of SPARIA for all classes of substituents and ring positions.

The correct prediction by the inverse mode of SPARIA of 74% of the substitution patterns for the three classes of substituents at five ring positions is a significant development that makes possible a deeper, probabilistic interpretation of the heteronuclear two-dimensional $^{1}H$, $^{13}C$ NMR spectra of complex natural samples such as NOM. It is fortuitous that, among all possible substitution patterns, aromatic C—H groups having *only* meta substituents occur only in monosubstituted and 1,3-disubstituted benzenes. So, in most instances, aromatic C—H groups will have ortho and/or para substituents that are more accurately predicted by the inverse mode of SPARIA. The situation is further aided by the fact that aromatic C—H groups having para substituents *but no ortho substituents* occur only in monosubstituted, 1,2-disubstituted, and 1,2,3-trisubstituted benzenes. For all other substitution patterns, an aromatic C—H group will have one or more ortho substituents. The very strong effect of ortho substituents on chemical shift (see Table 17) and their near-ubiquitous occurrence in substituted benzenes accounts largely for the predictive capability of the inverse mode of SPARIA.

In summary, many different aromatic chemical environments can contribute to a given HSQC cross peak, and the inverse SPARIA analysis of a very helpful tool for the elucidation of the respective substitution patterns. The SPARIA analysis complements the pattern-matching approach of NMR spectral prediction in aromatic systems (Simpson et al., 2004a), and indicates a more intricate situation, when working with complex systems such as NOM.

## 5.5. APPLICATION TO NOM

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

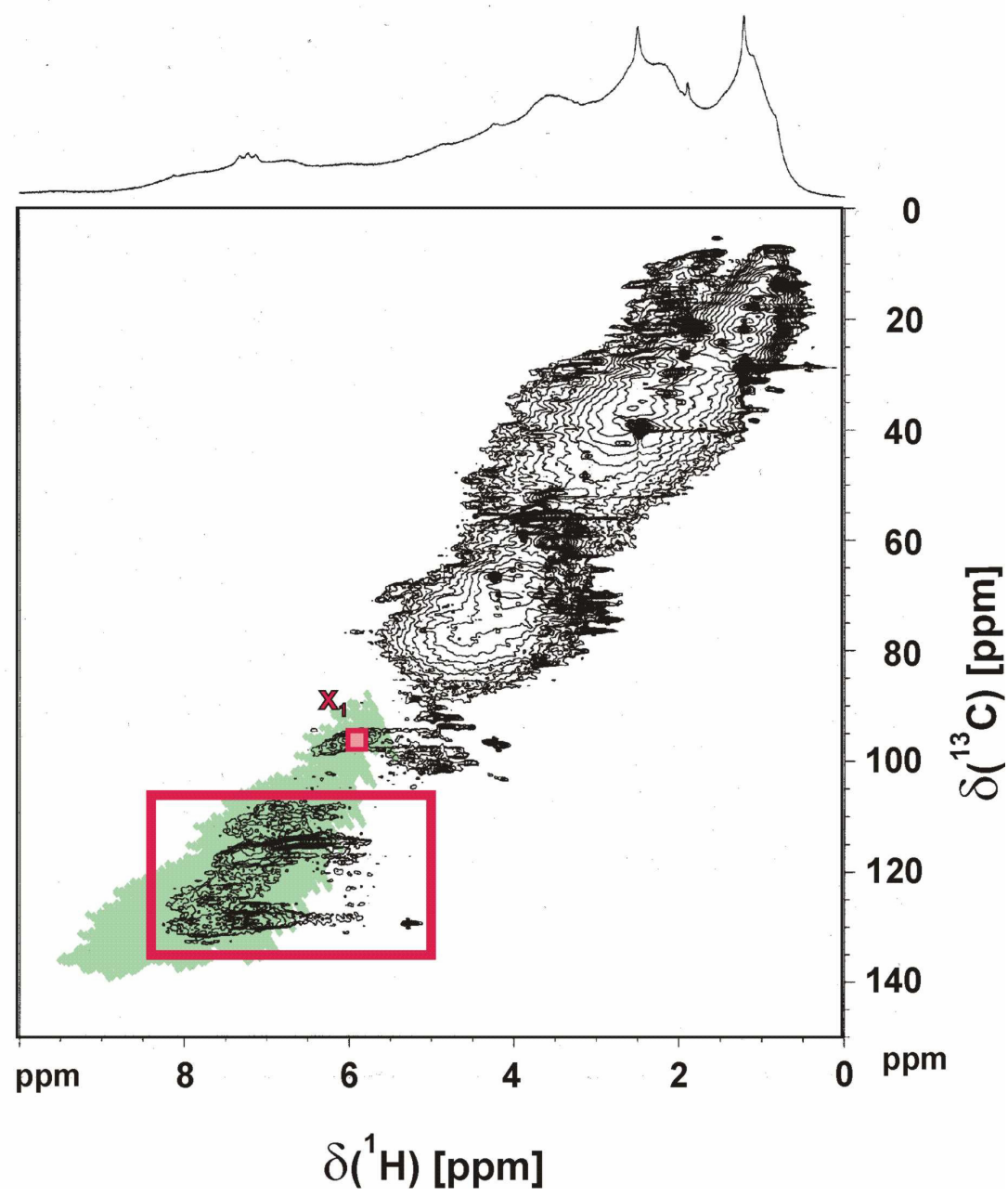



**Figure 65.** $^{1}H$, $^{13}C$-HSQC NMR spectrum of Suwannee River NOM, together with the aromatic section of chemical shift (red box, represent chemical shift range of Figures 50, 70) and cross peak $x_1$ (cf. Table 24; IHSS ID 1R101N; 99.5 mg in 750 µL of DMSO-$d_6$)

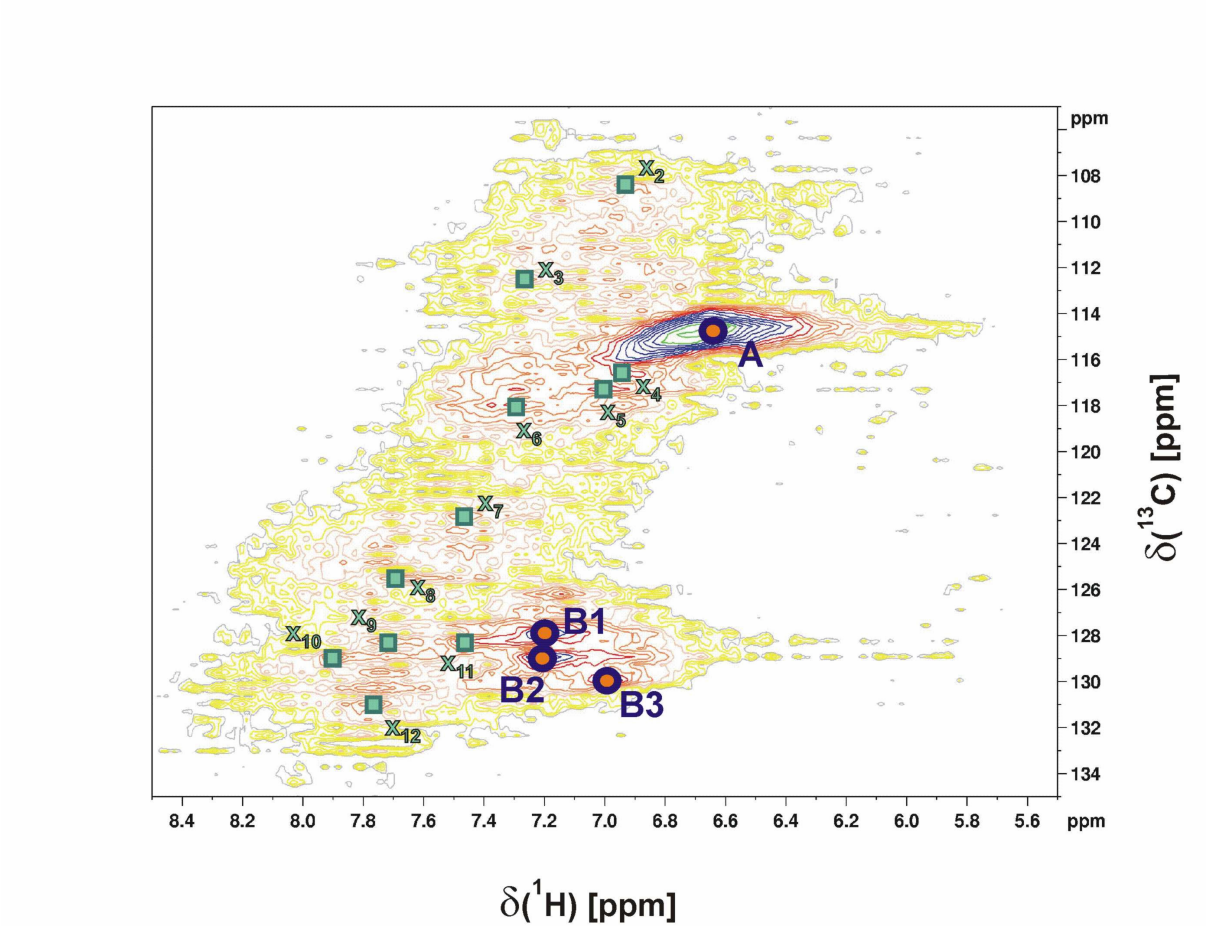



**Figure 66.** The aromatic region of the $^{1}H$, $^{13}C$-HSQC NMR spectrum of Suwannee River NOM (IHSS ID 1R101N; 99.5 mg in 750 µL of DMSO-$d_6$)

The $^{1}H$, $^{13}C$-HSQC spectrum of Suwannee River NOM was obtained using a Bruker DMX-500 spectrometer on a solution of 99.5 mg of NOM in 750 µL of DMSO-$d_6$ (gradient enhaced $^{1}H$, $^{13}C$ HSQC NMR spectrum at 500.13 MHz proton NMR frequency: 512 scans, 512 increments in F1 for a 175 ppm $^{13}C$ NMR bandwidth, acquisition time: 250 ms, relaxation delay: 1.25 s). The aromatic region of this spectrum is given in Figure 66.

The total expansion of the aromatic $^{1}H$, $^{13}C$ HSQC cross peaks of Suwannee River NOM is very considerable and represents the superposition of many hundreds of individual resonances; within the specified window (Figure 66) , these cross peaks cover a sizable area of the chemical shift space of the 16640 unique combinations of eight substituents in five ring positions provided in Figures 50 and 65. The very upfield cross peak $x_1$ (Figure 65) is most likely not of aromatic origin and appears to result from anomeric C-H bonds (Mao and Schmidt-Rohr, 2004).

**Table 24.** Most probable substitution patterns for the 16 most intense peaks in the HSQC spectrum (cf. Figure 66) of Suwannee River NOM, derived from an inverse mode SPARIA analysis with a window size ± 0.1 ppm in $^{1}H$ NMR and ± 1.0 ppm in $^{13}C$ NMR chemical shift

| HSQC cross peak | δ (1H) [ppm] | δ (13C) [ppm] | class | *Hits* | Ortho | Meta | Para | Meta' | Ortho' |
|---|---|---|---|---|---|---|---|---|---|
| **PEAK A** | 6,63 | 114,47 | | ***77*** | | | | | |
| **main peak** | | | ***COR*** | | 0% | 29% | 0% | 29% | 0% |
| | | | ***R*** | | 100% | 34% | 97% | 34% | 3% |
| | | | ***OR*** | | 0% | 38% | 3% | 38% | 97% |
| **PEAK B1** | 7,198 | 127,73 | | ***60*** | | | | | |
| **main peak** | | | ***COR*** | | 0% | 37% | 13% | 30% | 0% |
| | | | ***R*** | | 100% | 40% | 87% | 38% | 100% |
| | | | ***OR*** | | 0% | 23% | 0% | 32% | 0% |
| **PEAK B2** | 7,211 | 128,76 | | ***52*** | | | | | |
| **main peak** | | | ***COR*** | | 0% | 27% | 29% | 29% | 0% |
| | | | ***R*** | | 100% | 48% | 71% | 40% | 100% |
| | | | ***OR*** | | 0% | 25% | 0% | 31% | 0% |
| **PEAK B3** | 6,993 | 129,77 | | ***19*** | | | | | |
| **main peak** | | | ***COR*** | | 0% | 16% | 16% | 0% | 0% |
| | | | ***R*** | | 100% | 37% | 84% | 26% | 100% |
| | | | ***OR*** | | 0% | 47% | 0% | 74% | 0% |
| **PEAK x1** | 5,973 | 96,08 | | ***30*** | | | | | |
| **auxiliary peak** | | | ***COR*** | | 0% | 40% | 0% | 37% | 0% |
| probably anomeric HSQC cross peak | | | ***R*** | | 0% | 53% | 60% | 40% | 0% |
| | | | ***OR*** | | 100% | 7% | 40% | 23% | 100% |
| **PEAK x2** | 6,922 | 108,28 | | ***120*** | | | | | |
| **auxiliary peak** | | | ***COR*** | | 78% | 33% | 0% | 33% | 0% |
| | | | ***R*** | | 23% | 44% | 23% | 44% | 0% |
| | | | ***OR*** | | 0% | 23% | 78% | 23% | 100% |
| **PEAK x3** | 7,389 | 112,4 | | ***139*** | | | | | |
| **auxiliary peak** | | | ***COR*** | | 100% | 35% | 0% | 35% | 0% |
| | | | ***R*** | | 0% | 42% | 100% | 42% | 0% |
| | | | ***OR*** | | 0% | 22% | 0% | 22% | 100% |
| **PEAK x4** | 6,94 | 116,42 | | ***153*** | | | | | |
| **auxiliary peak** | | | ***COR*** | | 5% | 23% | 69% | 26% | 0% |
| | | | ***R*** | | 95% | 63% | 11% | 59% | 20% |
| | | | ***OR*** | | 0% | 14% | 20% | 14% | 80% |
| **PEAK x5** | 6,99 | 117,14 | | ***186*** | | | | | |
| **auxiliary peak** | | | ***COR*** | | 3% | 38% | 77% | 37% | 0% |
| | | | ***R*** | | 97% | 47% | 3% | 46% | 19% |
| | | | ***OR*** | | 0% | 15% | 19% | 17% | 81% |
| **PEAK x6** | 7,289 | 117,86 | | ***112*** | | | | | |
| **auxiliary peak** | | | ***COR*** | | 25% | 61% | 49% | 61% | 17% |
| | | | ***R*** | | 75% | 24% | 23% | 24% | 12% |
| | | | ***OR*** | | 0% | 15% | 28% | 15% | 71% |
| **PEAK x7** | 7,466 | 122,63 | | ***220*** | | | | | |
| **auxiliary peak** | | | ***COR*** | | 55% | 27% | 25% | 27% | 34% |
| | | | ***R*** | | 45% | 27% | 15% | 26% | 40% |
| | | | ***OR*** | | 0% | 46% | 60% | 46% | 25% |

**Table 24 (continued).** Most probable substitution patterns for the 16 most intense peaks in the HSQC spectrum (cf. Figure 66) of Suwannee River NOM, derived from an inverse mode SPARIA analysis with a window size ± 0.1 ppm in $^{1}$H NMR and ± 1.0 ppm in $^{13}$C NMR chemical shift

| HSQC cross peak | δ (1H) [ppm] | δ (13C) [ppm] | class | *Hits* | Ortho | Meta | Para | Meta' | Ortho' |
|---|---|---|---|---|---|---|---|---|---|
| PEAK x8 | 7,696 | 125,38 | | *147* | | | | | |
| auxiliary peak | | | *COR* | | 19% | 26% | 0% | 26% | 77% |
| | | | *R* | | 81% | 50% | 100% | 50% | 23% |
| | | | *OR* | | 0% | 24% | 0% | 24% | 0% |
| PEAK x9 | 7,77 | 128,44 | | *151* | | | | | |
| auxiliary peak | | | *COR* | | 34% | 25% | 15% | 27% | 52% |
| | | | *R* | | 66% | 37% | 85% | 34% | 48% |
| | | | *OR* | | 0% | 38% | 0% | 38% | 0% |
| PEAK x10 | 7,904 | 128,8 | | *144* | | | | | |
| auxiliary peak | | | *COR* | | 36% | 39% | 15% | 38% | 53% |
| | | | *R* | | 64% | 33% | 85% | 33% | 47% |
| | | | *OR* | | 0% | 28% | 0% | 28% | 0% |
| PEAK x11 | 7,466 | 128,11 | | *73* | | | | | |
| auxiliary peak | | | *COR* | | 10% | 30% | 48% | 27% | 18% |
| | | | *R* | | 90% | 44% | 52% | 41% | 82% |
| | | | *OR* | | 0% | 26% | 0% | 32% | 0% |
| PEAK x12 | 7,762 | 130,84 | | *99* | | | | | |
| auxiliary peak | | | *COR* | | 15% | 53% | 64% | 46% | 29% |
| | | | *R* | | 85% | 22% | 36% | 23% | 71% |
| | | | *OR* | | 0% | 25% | 0% | 30% | 0% |

In the case of four main HSQC cross peaks (A, B1, B2, B3) and twelve minor ones ($x_1$…$x_{12}$), the inverse SPARIA analysis (with a window size ± 0.1 ppm in $^{1}$H NMR and ± 1.0 ppm in $^{13}$C NMR chemical shift) as been utilized to compute the distribution of the three classes of substituents (R, COR, OR) across the five remaining ring positions as described in chapter 5.4.5. (Table 24).

The sizable number of hits for any cross peak position provided a realistic representation of the average aromatic substitution patterns. In Figure 68, the classes of substituents in the selected 16 HSQC cross peaks have been displayed according to probability ranges in color code (100….90%: orange; 9….75%: light green; 75….50%: light blue; question mark: < 50% probability for any substituent class in these positions). In addition, these cross peaks have been related to lignin-derived aromatic rings, and the three main groups of derivatives (p-hydroxy-benzoic acid, vanillic acid, syringic acid) were chosen for this analysis (Figure 67). Six HSQC cross peaks do not directly relate to lignins [$x_1$ (most likely anomeric CH-derived), $x_4$, $x_5$, $x_6$, $x_8$ and $x_{12}$]; all others, including the most intense ones (A, B1, B2, B3) could be explained by main classes of lignin aromatic molecules.

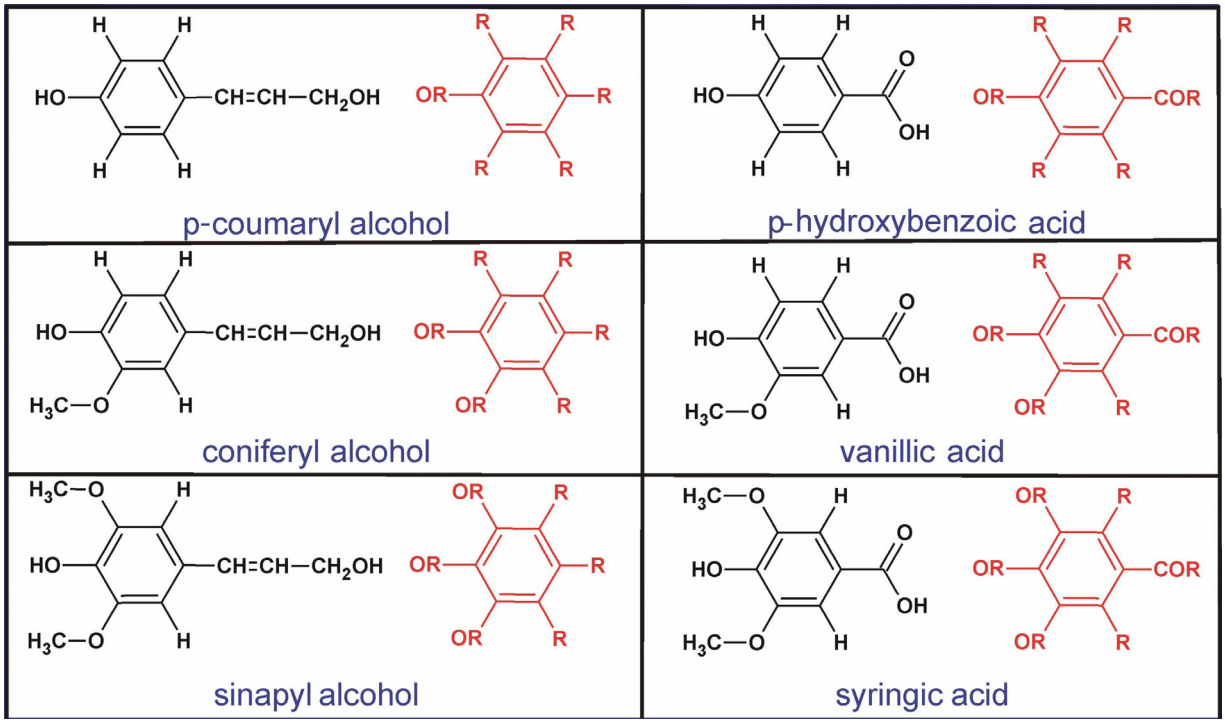



**Figure 67:** three main lignin-derived phenolic compounds and their representation according to classes of substituents (R, COR, OR) as used in the SPARIA analysis

The main HSQC cross peaks A, B1, B2, B3 and cross peaks $x_9$, $x_{10}$, $x_{11}$ relate to p-hydroxy benzoic acids; cross peaks A, $x_3$ and $x_7$ could represent vanillyl derivatives. Cross peaks $x_2$ and $x_7$ as well fall within the chemical shift window of syringyl derivatives. Future work intended to confirm these preliminary assignments is attempted by fractionation studies and high-resolution homonuclear 2D-NMR spectroscopy (cf. Figure 30) and homonuclear cross peak patterns analysis and homonuclear SPARIA analysis, respectively.

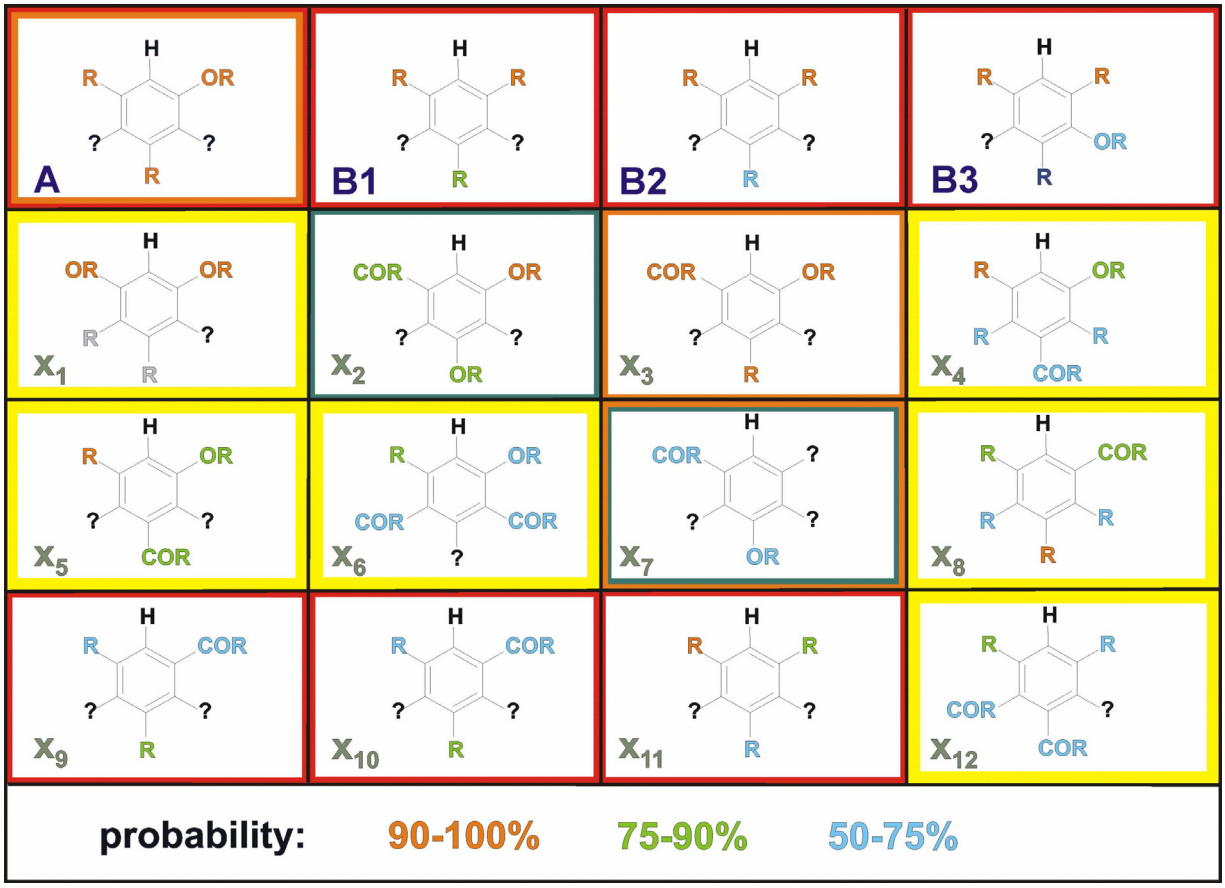



**Figure 68.** Most probable substitution patterns for the 16 most intense peaks in the HSQC spectrum (cf. Table 24) of Suwannee River NOM, given in relationship to lignin-derived aromatic structures (color of boxes: red, para-hydroxy benzoic acid derivative; green: syringyl-derivative; orange: vannilyl derivative; yellow: no plausible main lignin-derivative)

## 5.6. INVERSE SPARIA ANALYSIS OF HSQC CROSS PEAKS OF A SOIL EXTRACT

Alkaline soil extracts are capable to produce high quality HSQC NMR spectra, if paramagnetic metal ions are not abundant. The numerous cross peaks cover a fairly broad range of the HSQC chemical shift space and are rather evenly distributed. These features are useful and aid in the analysis of NOM spectra. In this example analysis, 120 aromatic cross peak positions in a high quality $^{1}H$, $^{13}C$ HSQC NMR spectrum (cf. Figure 69) of a freeze-dried alkaline soil extract, which encompass a large area of chemical shift space, have been analyzed with SPARIA in its inverse mode and the results are given in Table 25.

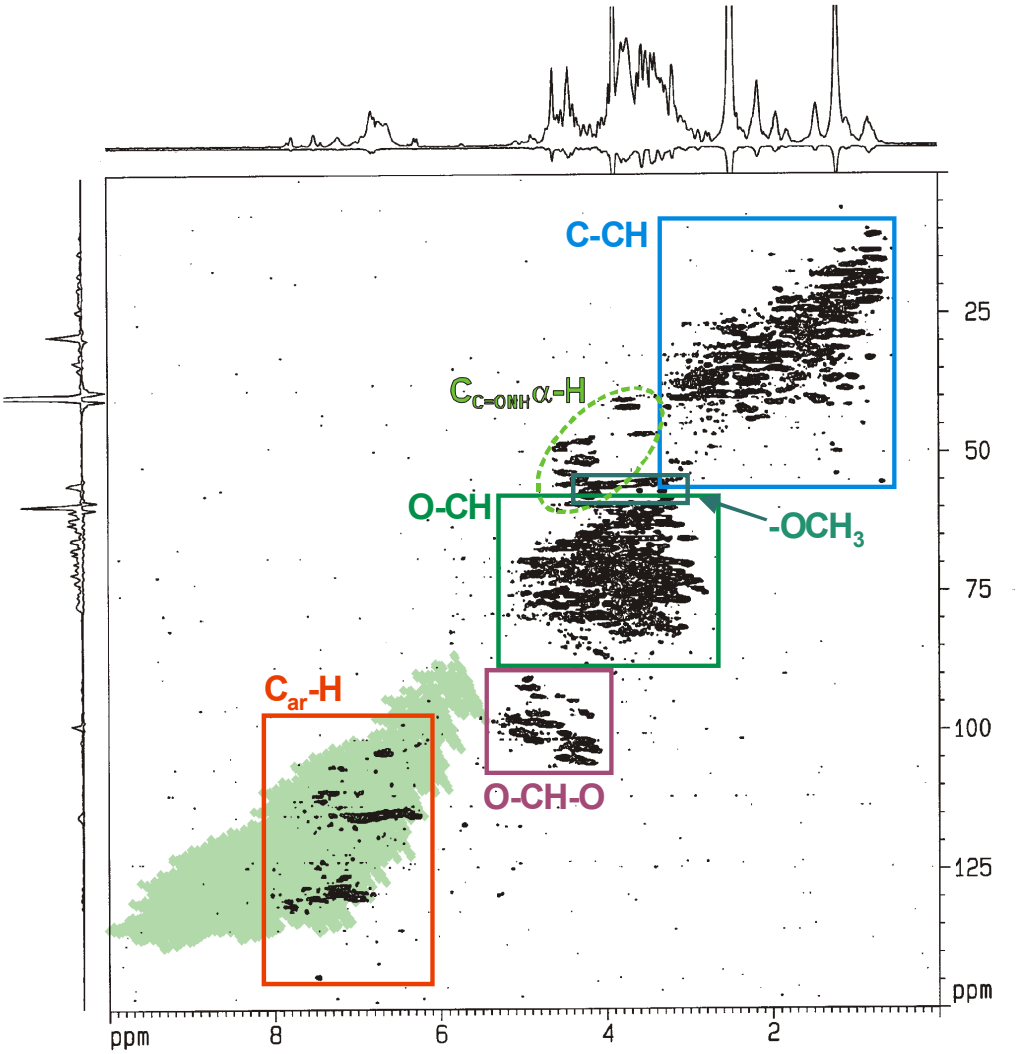



**Figure 69.** $^{1}H,^{13}C$ HSQC NMR spectrum of a freeze-dried alkaline soil extract (15.8 mg in 750 µl DMSO-$d_6$, sealed under exclusion of moisture) with aromatic region used for SPARIA analysis in its inverse mode (red, cf. Figure 70) together with the chemical shift space of the 10368 unique structures (faint green) and other main substructures

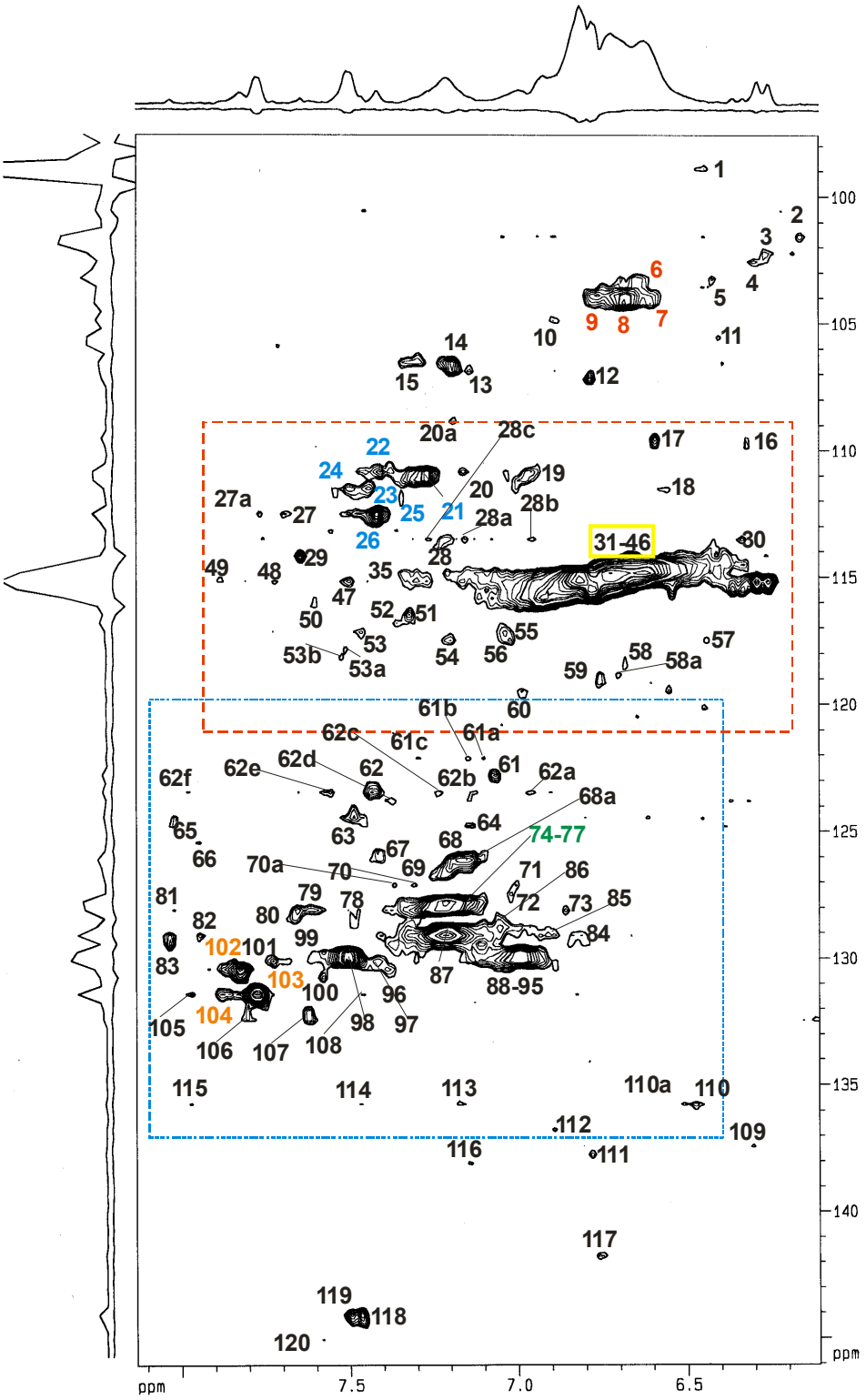



**Figure 70.** aromatic section of the $^{1}H,^{13}C$ HSQC NMR spectrum of an alkaline soil extract (cf. Figure 69) with 120 cross peak positions indicated for SPARIA analysis in its inverse mode; five groups, selected for individual molecule level analysis (cf. Tables 25, 26, 27 and Figure 74 are highlighted); $^{1}H$ and $^{13}C$ NMR spectra are *internal* projections.

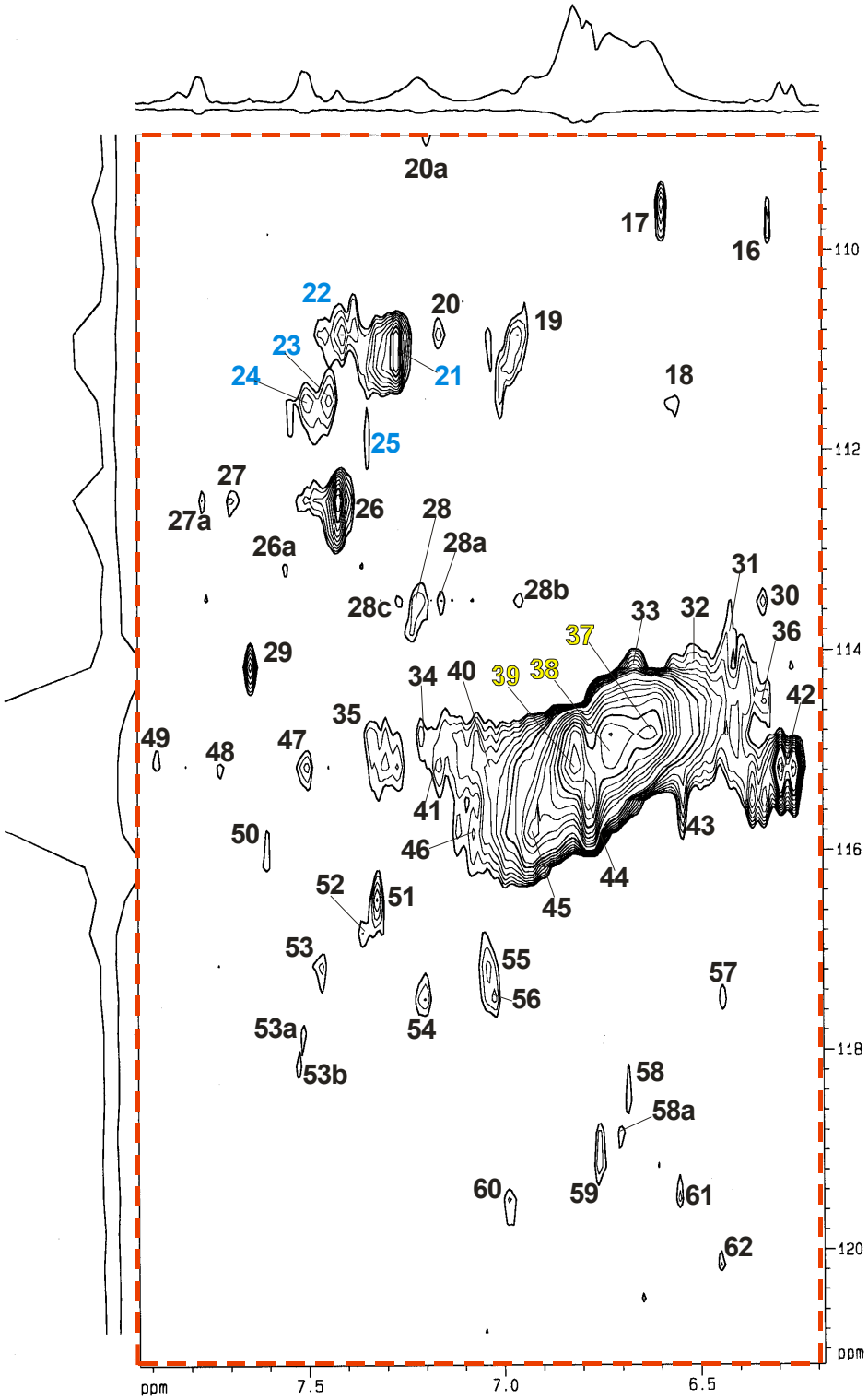



**Figure 71.** aromatic upfield section of the $^{1}H,^{13}C$ HSQC NMR spectrum of an alkaline soil extract (cf. Figure 69) with cross peak positions 16 through 62 indicated for SPARIA analysis in its inverse mode; cross peaks from groups II and III, selected for individual molecule level analysis (cf. Tables 25, 26, 27 and Figure 74 are colored); $^{1}H$ and $^{13}C$ NMR spectra are *internal* projections.

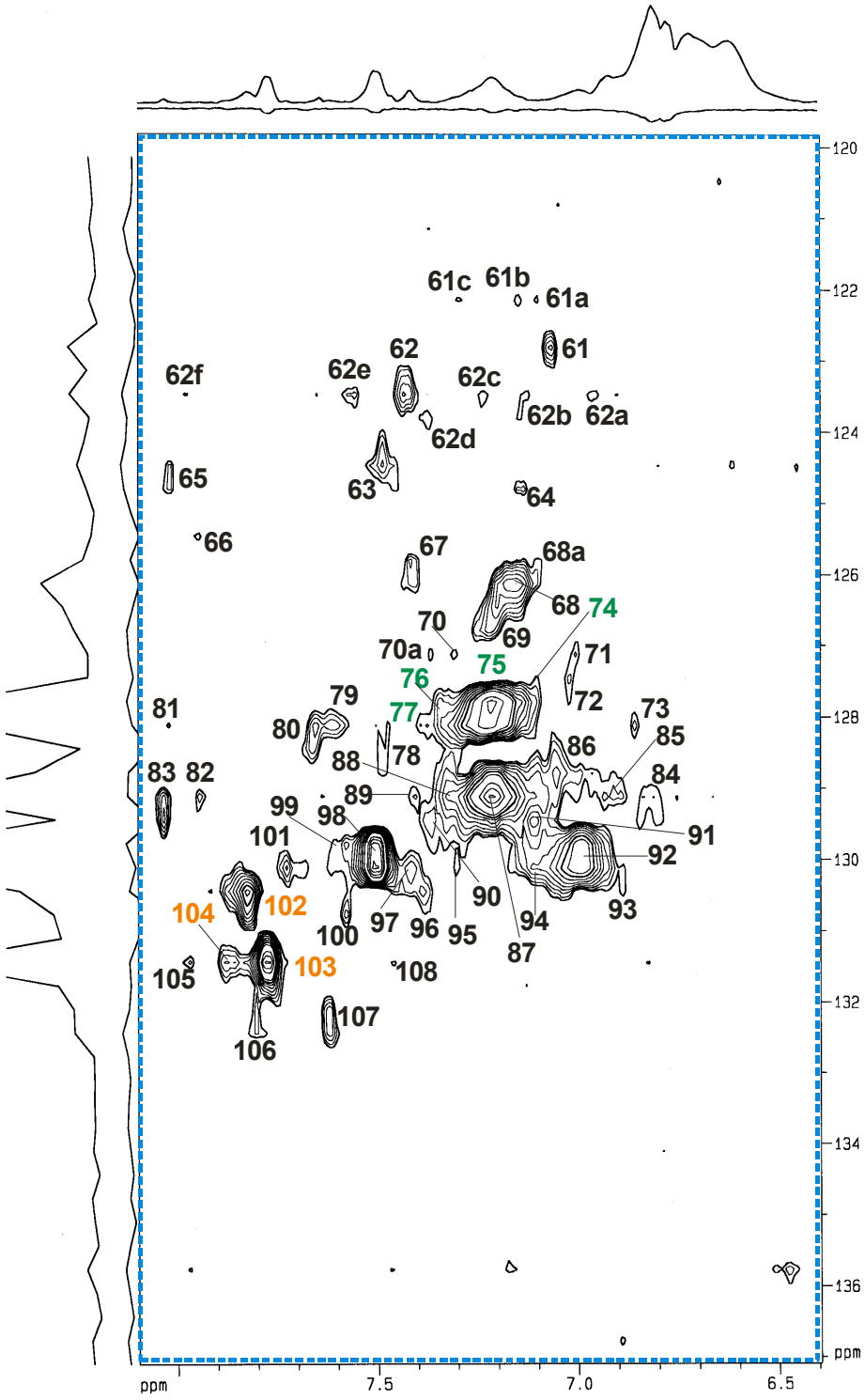



**Figure 72.** aromatic upfield section of the $^{1}H,^{13}C$ HSQC NMR spectrum of an alkaline soil extract (cf. Figure 69) with cross peak positions 61 through 108 indicated for SPARIA analysis in its inverse mode; cross peaks from groups IV and V, selected for individual molecule level analysis (cf. Tables 25, 26, 27 and Figure 74 are colored); $^{1}H$ and $^{13}C$ NMR spectra are *internal* projections.

| Peak No. | δ $^{1}$H | δ $^{13}$C | Hits | ortho %_R | ortho %_COR | ortho %_OR | meta %_R | meta %_COR | meta %_OR | para %_R | para %_COR | para %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 6,448 | 98,736 | 17 | 0,000 | 0,000 | 1,000 | 0,382 | 0,588 | 0,029 | 0,824 | 0,176 | 0,000 |
| 2 | 6,16 | 101,42 | 25 | 0,060 | 0,000 | 0,940 | 0,420 | 0,320 | 0,260 | 0,880 | 0,000 | 0,120 |
| 3 | 6,264 | 102,15 | 23 | 0,217 | 0,000 | 0,783 | 0,543 | 0,261 | 0,196 | 0,435 | 0,130 | 0,435 |
| 4 | 6,296 | 102,43 | 39 | 0,205 | 0,000 | 0,795 | 0,603 | 0,231 | 0,167 | 0,333 | 0,256 | 0,410 |
| 5 | 6,422 | 103,14 | 43 | 0,116 | 0,000 | 0,884 | 0,314 | 0,616 | 0,070 | 0,000 | 0,767 | 0,233 |
| 6 | 6,644 | 103,21 | 43 | 0,116 | 0,000 | 0,884 | 0,314 | 0,616 | 0,070 | 0,000 | 0,767 | 0,233 |
| 7 | 6,604 | 104,06 | 51 | 0,176 | 0,000 | 0,824 | 0,275 | 0,578 | 0,147 | 0,020 | 0,647 | 0,333 |
| 8 | 6,683 | 104,05 | 33 | 0,197 | 0,000 | 0,803 | 0,167 | 0,758 | 0,076 | 0,030 | 0,606 | 0,364 |
| 9 | 6,761 | 103,77 | 26 | 0,135 | 0,038 | 0,827 | 0,173 | 0,808 | 0,019 | 0,038 | 0,654 | 0,308 |
| 10 | 6,89 | 104,76 | 18 | 0,000 | 0,444 | 0,556 | 0,778 | 0,111 | 0,111 | 0,000 | 0,111 | 0,889 |
| 11 | 6,403 | 105,44 | 99 | 0,298 | 0,000 | 0,702 | 0,424 | 0,399 | 0,177 | 0,010 | 0,404 | 0,586 |
| 12 | 6,782 | 107,06 | 34 | 0,103 | 0,338 | 0,559 | 0,544 | 0,206 | 0,250 | 0,206 | 0,118 | 0,676 |
| 13 | 7,14 | 106,72 | 53 | 0,000 | 0,500 | 0,500 | 0,509 | 0,406 | 0,085 | 0,019 | 0,000 | 0,981 |
| 14 | 7,191 | 106,44 | 42 | 0,000 | 0,500 | 0,500 | 0,393 | 0,595 | 0,012 | 0,048 | 0,000 | 0,952 |
| 15 | 7,305 | 106,38 | 34 | 0,000 | 0,500 | 0,500 | 0,221 | 0,779 | 0,000 | 0,029 | 0,000 | 0,971 |
| 16 | 6,32 | 109,64 | 40 | 0,275 | 0,000 | 0,725 | 0,300 | 0,275 | 0,425 | 0,275 | 0,450 | 0,275 |
| 17 | 6,591 | 109,44 | 52 | 0,490 | 0,000 | 0,510 | 0,567 | 0,279 | 0,154 | 0,962 | 0,019 | 0,019 |
| 18 | 6,562 | 111,41 | 71 | 0,486 | 0,014 | 0,500 | 0,451 | 0,239 | 0,310 | 0,972 | 0,000 | 0,028 |
| 19 | 6,96 | 110,74 | 60 | 0,250 | 0,250 | 0,500 | 0,192 | 0,533 | 0,275 | 0,517 | 0,000 | 0,483 |
| 20 | 7,159 | 110,73 | 48 | 0,031 | 0,469 | 0,500 | 0,510 | 0,250 | 0,240 | 0,646 | 0,000 | 0,354 |
| 21 | 7,268 | 110,9 | 51 | 0,010 | 0,490 | 0,500 | 0,569 | 0,275 | 0,157 | 0,863 | 0,000 | 0,137 |
| 22 | 7,407 | 110,73 | 55 | 0,000 | 0,500 | 0,500 | 0,491 | 0,436 | 0,073 | 0,891 | 0,000 | 0,109 |
| 23 | 7,439 | 111,39 | 52 | 0,000 | 0,500 | 0,500 | 0,519 | 0,404 | 0,077 | 1,000 | 0,000 | 0,000 |
| 24 | 7,497 | 111,42 | 49 | 0,000 | 0,500 | 0,500 | 0,429 | 0,520 | 0,051 | 1,000 | 0,000 | 0,000 |

**Table 25/1.** SPARIA inverse mode analysis of aromatic substitution patterns for 120 aromatic cross peak positions [window size: ± 1 ppm ($^{13}$C) and ± 0.1 ppm ($^{1}$H)] according to R, COR, OR substituent distributions in ortho-, meta- and para positions; five groups of closely spaced cross peaks are singles out for individual molecule level analysis (cf. Figure 70 and Table 26)

| Peak No. | δ $^{1}$H | δ $^{13}$C | Hits | ortho %_R | ortho %_COR | ortho %_OR | meta %_R | meta %_COR | meta %_OR | para %_R | para %_COR | para %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 7,343 | 111,75 | 64 | 0,000 | 0,500 | 0,500 | 0,516 | 0,313 | 0,172 | 1,000 | 0,000 | 0,000 |
| 26 | 7,417 | 112,4 | 70 | 0,000 | 0,500 | 0,500 | 0,421 | 0,393 | 0,186 | 1,000 | 0,000 | 0,000 |
| 27 | 7,69 | 112,38 | 29 | 0,000 | 0,500 | 0,500 | 0,224 | 0,741 | 0,034 | 1,000 | 0,000 | 0,000 |
| 28 | 7,216 | 113,44 | 62 | 0,056 | 0,444 | 0,500 | 0,460 | 0,234 | 0,306 | 0,919 | 0,081 | 0,000 |
| 29 | 7,642 | 114,05 | 43 | 0,000 | 0,500 | 0,500 | 0,291 | 0,698 | 0,012 | 0,930 | 0,070 | 0,000 |
| 30 | 6,333 | 113,38 | 31 | 0,500 | 0,000 | 0,500 | 0,290 | 0,097 | 0,613 | 1,000 | 0,000 | 0,000 |
| 31 | 6,42 | 113,71 | 42 | 0,500 | 0,000 | 0,500 | 0,310 | 0,131 | 0,560 | 1,000 | 0,000 | 0,000 |
| 32 | 6,512 | 113,93 | 39 | 0,500 | 0,000 | 0,500 | 0,308 | 0,205 | 0,487 | 1,000 | 0,000 | 0,000 |
| 33 | 6,664 | 114,02 | 47 | 0,521 | 0,000 | 0,479 | 0,383 | 0,277 | 0,340 | 0,957 | 0,000 | 0,043 |
| 34 | 7,205 | 114,74 | 59 | 0,093 | 0,407 | 0,500 | 0,364 | 0,237 | 0,398 | 0,831 | 0,169 | 0,000 |
| 35 | 7,298 | 114,99 | 58 | 0,043 | 0,457 | 0,500 | 0,362 | 0,241 | 0,397 | 0,914 | 0,086 | 0,000 |
| 36 | 6,332 | 114,39 | 25 | 0,500 | 0,000 | 0,500 | 0,240 | 0,100 | 0,660 | 1,000 | 0,000 | 0,000 |
| 37 | 6,626 | 114,7 | 40 | 0,525 | 0,000 | 0,475 | 0,375 | 0,250 | 0,375 | 0,950 | 0,000 | 0,050 |
| 38 | 6,724 | 114,71 | 44 | 0,545 | 0,000 | 0,455 | 0,466 | 0,307 | 0,227 | 0,773 | 0,136 | 0,091 |
| 39 | 6,814 | 115,01 | 50 | 0,570 | 0,000 | 0,430 | 0,490 | 0,360 | 0,150 | 0,600 | 0,260 | 0,140 |
| 40 | 7,029 | 115,08 | 56 | 0,384 | 0,170 | 0,446 | 0,455 | 0,268 | 0,277 | 0,464 | 0,429 | 0,107 |
| 41 | 7,161 | 115,01 | 56 | 0,188 | 0,330 | 0,482 | 0,348 | 0,286 | 0,366 | 0,679 | 0,286 | 0,036 |
| 42 | 6,272 | 115,05 | 10 | 0,500 | 0,000 | 0,500 | 0,100 | 0,050 | 0,850 | 1,000 | 0,000 | 0,000 |
| 43 | 6,514 | 115,28 | 27 | 0,556 | 0,000 | 0,444 | 0,444 | 0,093 | 0,463 | 0,889 | 0,000 | 0,111 |
| 44 | 6,779 | 115,34 | 42 | 0,583 | 0,000 | 0,417 | 0,464 | 0,357 | 0,179 | 0,619 | 0,214 | 0,167 |
| 45 | 6,923 | 115,68 | 61 | 0,516 | 0,057 | 0,426 | 0,557 | 0,270 | 0,172 | 0,311 | 0,541 | 0,148 |
| 46 | 7,073 | 115,71 | 64 | 0,453 | 0,125 | 0,422 | 0,445 | 0,359 | 0,195 | 0,297 | 0,547 | 0,156 |
| 47 | 7,5 | 115,05 | 64 | 0,039 | 0,484 | 0,477 | 0,391 | 0,414 | 0,195 | 0,859 | 0,094 | 0,047 |
| 48 | 7,721 | 115,06 | 33 | 0,000 | 0,500 | 0,500 | 0,348 | 0,621 | 0,030 | 0,818 | 0,182 | 0,000 |
| 49 | 7,883 | 114,99 | 8 | 0,000 | 0,500 | 0,500 | 0,063 | 0,938 | 0,000 | 1,000 | 0,000 | 0,000 |

**Table 25/2.** SPARIA inverse mode analysis of aromatic substitution patterns for 120 aromatic cross peak positions [window size: ± 1 ppm ($^{13}$C) and ± 0.1 ppm ($^{1}$H)] according to R, COR, OR substituent distributions in ortho-, meta- and para positions; five groups of closely spaced cross peaks are singles out for individual molecule level analysis (cf. Figure 70 and Table 26).

| | | | | ortho | ortho | ortho | meta | meta | meta | para | para | para |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Peak No. | δ $^{1}$H | δ $^{13}$C | Hits | %_R | %_COR | %_OR | %_R | %_COR | %_OR | %_R | %_COR | %_OR |
| 50 | 7,603 | 115,91 | 52 | 0,038 | 0,500 | 0,462 | 0,481 | 0,404 | 0,115 | 0,673 | 0,250 | 0,077 |
| 51 | 7,32 | 116,36 | 66 | 0,273 | 0,303 | 0,424 | 0,288 | 0,462 | 0,250 | 0,530 | 0,318 | 0,152 |
| 52 | 7,356 | 116,69 | 59 | 0,263 | 0,305 | 0,432 | 0,305 | 0,483 | 0,212 | 0,508 | 0,356 | 0,136 |
| 53 | 7,462 | 117,04 | 59 | 0,127 | 0,475 | 0,398 | 0,568 | 0,220 | 0,212 | 0,508 | 0,288 | 0,203 |
| 54 | 7,2 | 117,37 | 94 | 0,495 | 0,106 | 0,399 | 0,234 | 0,590 | 0,176 | 0,181 | 0,628 | 0,191 |
| 55 | 7,036 | 117,04 | 108 | 0,597 | 0,019 | 0,384 | 0,398 | 0,468 | 0,134 | 0,056 | 0,713 | 0,231 |
| 56 | 7,025 | 117,33 | 118 | 0,623 | 0,013 | 0,364 | 0,403 | 0,462 | 0,136 | 0,025 | 0,703 | 0,271 |
| 57 | 6,44 | 117,33 | 14 | 0,679 | **0,000** | 0,321 | 0,429 | 0,036 | 0,536 | 0,643 | **0,000** | 0,357 |
| 58 | 6,681 | 118,34 | 73 | 0,795 | **0,000** | 0,205 | 0,568 | 0,151 | 0,281 | 0,041 | 0,370 | 0,589 |
| 59 | 6,756 | 118,91 | 124 | 0,766 | **0,000** | 0,234 | 0,456 | 0,226 | 0,319 | 0,000 | 0,468 | 0,532 |
| 60 | 6,984 | 119,37 | 139 | 0,683 | **0,000** | 0,317 | 0,266 | 0,540 | 0,194 | 0,022 | 0,633 | 0,345 |
| 61 | 7,066 | 122,67 | 53 | 0,915 | 0,057 | 0,028 | 0,528 | 0,198 | 0,274 | 0,830 | 0,057 | 0,113 |
| 62 | 7,43 | 123,31 | 98 | 0,531 | 0,362 | 0,107 | 0,194 | 0,316 | 0,490 | 0,306 | 0,214 | 0,480 |
| 63 | 7,485 | 124,3 | 68 | 0,559 | 0,346 | 0,096 | 0,294 | 0,279 | 0,426 | 0,500 | 0,191 | 0,309 |
| 64 | 7,138 | 124,65 | 82 | 0,963 | 0,037 | **0,000** | 0,323 | 0,409 | 0,268 | 0,927 | 0,000 | 0,073 |
| 65 | 8,016 | 124,51 | 94 | 0,399 | 0,601 | **0,000** | 0,362 | 0,441 | 0,197 | 0,798 | **0,000** | 0,202 |
| 66 | 7,943 | 125,33 | 109 | 0,468 | 0,532 | **0,000** | 0,358 | 0,486 | 0,156 | 0,936 | **0,000** | 0,064 |
| 67 | 7,417 | 125,69 | 50 | 0,840 | 0,140 | 0,020 | 0,280 | 0,490 | 0,230 | 0,900 | 0,040 | 0,060 |
| 68 | 7,152 | 125,99 | 49 | **1,000** | **0,000** | **0,000** | 0,357 | 0,408 | 0,235 | 1,000 | **0,000** | **0,000** |
| 69 | 7,218 | 126,57 | 47 | **1,000** | **0,000** | **0,000** | 0,351 | 0,394 | 0,255 | 0,979 | 0,021 | **0,000** |
| 70 | 7,306 | 126,99 | 42 | 0,952 | 0,048 | **0,000** | 0,345 | 0,452 | 0,202 | 0,929 | 0,071 | **0,000** |
| 71 | 7,004 | 126,97 | 37 | **1,000** | **0,000** | **0,000** | 0,378 | 0,203 | 0,419 | 0,973 | 0,027 | **0,000** |
| 72 | 7,02 | 127,33 | 38 | **1,000** | **0,000** | **0,000** | 0,382 | 0,184 | 0,434 | 0,947 | 0,053 | **0,000** |
| 73 | 6,858 | 127,99 | 16 | **1,000** | **0,000** | **0,000** | 0,313 | 0,031 | 0,656 | **1,000** | **0,000** | **0,000** |
| **74** | 7,103 | 127,46 | 43 | **1,000** | **0,000** | **0,000** | 0,395 | 0,279 | 0,326 | 0,953 | 0,047 | **0,000** |

**Table 25/3.** SPARIA inverse mode analysis of aromatic substitution patterns for 120 aromatic cross peak positions [window size: ± 1 ppm ($^{13}$C) and ± 0.1 ppm ($^{1}$H)] according to R, COR, OR substituent distributions in ortho-, meta- and para positions; five groups of closely spaced cross peaks are singles out for individual molecule level analysis (cf. Figure 70 and Table 26)

| | | | | ortho | ortho | ortho | meta | meta | meta | para | para | para |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Peak No. | δ $^{1}$H | δ $^{13}$C | Hits | %_R | %_COR | %_OR | %_R | %_COR | %_OR | %_R | %_COR | %_OR |
| 75 | 7,212 | 127,7 | 39 | 1,000 | 0,000 | 0,000 | 0,346 | 0,372 | 0,282 | 0,897 | 0,103 | 0,000 |
| 76 | 7,339 | 127,7 | 38 | 0,908 | 0,092 | 0,000 | 0,316 | 0,421 | 0,263 | 0,842 | 0,158 | 0,000 |
| 77 | 7,382 | 127,98 | 44 | 0,898 | 0,102 | 0,000 | 0,341 | 0,364 | 0,295 | 0,682 | 0,318 | 0,000 |
| 78 | 7,487 | 128,38 | 56 | 0,857 | 0,143 | 0,000 | 0,438 | 0,250 | 0,313 | 0,482 | 0,518 | 0,000 |
| 79 | 7,609 | 127,99 | 64 | 0,641 | 0,359 | 0,000 | 0,344 | 0,188 | 0,469 | 0,797 | 0,203 | 0,000 |
| 80 | 7,652 | 128,02 | 61 | 0,639 | 0,361 | 0,000 | 0,369 | 0,213 | 0,418 | 0,770 | 0,230 | 0,000 |
| 81 | 8,018 | 127,97 | 79 | 0,506 | 0,494 | 0,000 | 0,342 | 0,481 | 0,177 | 0,962 | 0,038 | 0,000 |
| 82 | 7,941 | 128,98 | 81 | 0,574 | 0,426 | 0,000 | 0,364 | 0,438 | 0,198 | 0,765 | 0,235 | 0,000 |
| 83 | 8,024 | 129,28 | 76 | 0,520 | 0,480 | 0,000 | 0,434 | 0,382 | 0,184 | 0,763 | 0,237 | 0,000 |
| 84 | 6,823 | 128,96 | 9 | 1,000 | 0,000 | 0,000 | 0,222 | 0,000 | 0,778 | 1,000 | 0,000 | 0,000 |
| 85 | 6,909 | 128,95 | 16 | 1,000 | 0,000 | 0,000 | 0,188 | 0,125 | 0,688 | 1,000 | 0,000 | 0,000 |
| 86 | 7,053 | 128,67 | 29 | 1,000 | 0,000 | 0,000 | 0,414 | 0,138 | 0,448 | 0,862 | 0,138 | 0,000 |
| 87 | 7,212 | 128,97 | 31 | 1,000 | 0,000 | 0,000 | 0,500 | 0,226 | 0,274 | 0,613 | 0,387 | 0,000 |
| 88 | 7,313 | 128,97 | 34 | 0,971 | 0,029 | 0,000 | 0,529 | 0,294 | 0,176 | 0,500 | 0,500 | 0,000 |
| 89 | 7,405 | 128,99 | 48 | 0,938 | 0,063 | 0,000 | 0,490 | 0,281 | 0,229 | 0,438 | 0,563 | 0,000 |
| 90 | 7,365 | 129,34 | 47 | 0,968 | 0,032 | 0,000 | 0,574 | 0,223 | 0,202 | 0,298 | 0,702 | 0,000 |
| 91 | 7,107 | 129,31 | 24 | 1,000 | 0,000 | 0,000 | 0,438 | 0,146 | 0,417 | 0,708 | 0,292 | 0,000 |
| 92 | 6,991 | 129,8 | 13 | 1,000 | 0,000 | 0,000 | 0,308 | 0,077 | 0,615 | 0,769 | 0,231 | 0,000 |
| 93 | 6,887 | 130,22 | 5 | 1,000 | 0,000 | 0,000 | 0,100 | 0,000 | 0,900 | 1,000 | 0,000 | 0,000 |
| 94 | 7,106 | 129,96 | 19 | 1,000 | 0,000 | 0,000 | 0,474 | 0,105 | 0,421 | 0,579 | 0,421 | 0,000 |
| 95 | 7,299 | 129,93 | 41 | 1,000 | 0,000 | 0,000 | 0,646 | 0,195 | 0,159 | 0,146 | 0,854 | 0,000 |
| 96 | 7,385 | 130,3 | 63 | 0,992 | 0,008 | 0,000 | 0,603 | 0,190 | 0,206 | 0,063 | 0,937 | 0,000 |
| 97 | 7,416 | 130,01 | 65 | 0,985 | 0,015 | 0,000 | 0,623 | 0,192 | 0,185 | 0,062 | 0,938 | 0,000 |
| 98 | 7,504 | 129,96 | 73 | 0,945 | 0,055 | 0,000 | 0,548 | 0,247 | 0,205 | 0,110 | 0,890 | 0,000 |
| 99 | 7,576 | 129,67 | 69 | 0,891 | 0,109 | 0,000 | 0,420 | 0,319 | 0,261 | 0,246 | 0,754 | 0,000 |

**Table 25/4.** SPARIA inverse mode analysis of aromatic substitution patterns for 120 aromatic cross peak positions [window size: ± 1 ppm ($^{13}$C) and ± 0.1 ppm ($^{1}$H)] according to R, COR, OR substituent distributions in ortho-, meta- and para positions; five groups of closely spaced cross peaks are singles out for individual molecule level analysis (cf. Figure 70 and Table 26)

| Peak No. | δ $^{1}$H | δ $^{13}$C | Hits | ortho %_R | ortho %_COR | ortho %_OR | meta %_R | meta %_COR | meta %_OR | para %_R | para %_COR | para %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 7,576 | 130,62 | 85 | 0,941 | 0,059 | 0,000 | 0,341 | 0,441 | 0,218 | 0,118 | 0,882 | 0,000 |
| 101 | 7,724 | 129,97 | 81 | 0,809 | 0,191 | 0,000 | 0,253 | 0,432 | 0,315 | 0,383 | 0,617 | 0,000 |
| 102 | 7,823 | 130,32 | 65 | 0,708 | 0,292 | 0,000 | 0,231 | 0,477 | 0,292 | 0,508 | 0,492 | 0,000 |
| 103 | 7,774 | 131,3 | 56 | 0,768 | 0,232 | 0,000 | 0,223 | 0,482 | 0,295 | 0,339 | 0,661 | 0,000 |
| 104 | 7,876 | 131,29 | 55 | 0,636 | 0,364 | 0,000 | 0,327 | 0,345 | 0,327 | 0,436 | 0,564 | 0,000 |
| 105 | 7,966 | 131,29 | 61 | 0,533 | 0,467 | 0,000 | 0,459 | 0,238 | 0,303 | 0,393 | 0,607 | 0,000 |
| 106 | 7,802 | 132,23 | 45 | 0,733 | 0,267 | 0,000 | 0,267 | 0,456 | 0,278 | 0,156 | 0,844 | 0,000 |
| 107 | 7,619 | 132,13 | 76 | 0,974 | 0,026 | 0,000 | 0,191 | 0,625 | 0,184 | 0,053 | 0,947 | 0,000 |
| 108 | 7,457 | 131,3 | 107 | 0,991 | 0,009 | 0,000 | 0,383 | 0,388 | 0,229 | 0,019 | 0,981 | 0,000 |
| 109 | 6,305 | 137,28 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 110 | 6,472 | 135,66 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 111 | 6,783 | 137,64 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 112 | 6,893 | 136,62 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 113 | 7,172 | 135,59 | 10 | 1,000 | 0,000 | 0,000 | 0,000 | 0,000 | 1,000 | 0,000 | 1,000 | 0,000 |
| 114 | 7,463 | 135,63 | 4 | 1,000 | 0,000 | 0,000 | 0,125 | 0,375 | 0,500 | 0,000 | 1,000 | 0,000 |
| 115 | 7,968 | 135,63 | 49 | 0,500 | 0,500 | 0,000 | 0,214 | 0,061 | 0,724 | 0,000 | 1,000 | 0,000 |
| 116 | 7,141 | 137,96 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 117 | 6,76 | 141,6 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 118 | 7,46 | 144,06 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 119 | 7,501 | 143,97 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 120 | 7,576 | 144,93 | 0 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |

**Table 25/5:** SPARIA inverse mode analysis of aromatic substitution patterns for 120 aromatic cross peak positions [window size: ± 1 ppm ($^{13}$C) and ± 0.1 ppm ($^{1}$H)] according to R, COR, OR substituent distributions in ortho-, meta- and para positions; five groups of closely spaced cross peaks are singles out for individual molecule level analysis (cf. Figure 70 and Table 26)

Although interesting and revealing trends can be derived from the entire data set, *five groups* of closely spaced cross peaks have been selected for an in-depth analysis of aromatic substitution patterns by SPARIA in its inverse mode.

A notable consequence of the standard SPARIA analysis and its dependency on substituent increments (cf. Table 17, which only distinguish ortho-, meta- and para-positions, is the inability to differentiate between certain isomers as indicated in Figure 73.

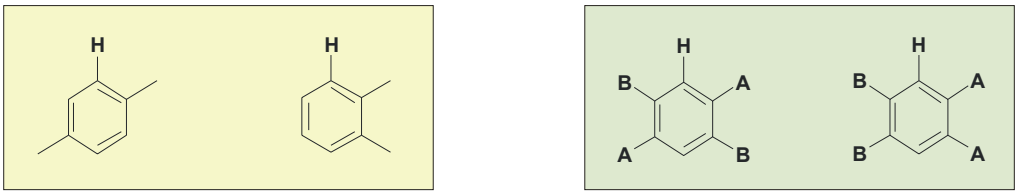



**Figure 73.** different aromatic molecules, which cannot be distinguished in the heteronuclear SPARIA standard model

At this stage of analysis (cf. Table 26), only the distribution of –R, -COR and –OR substituents among the positions –ortho, -meta and -para is evaluated; i.e. at this level, the two different ortho- and meta positions are not distinguished.

In short, *group I* cross peaks owe their upfield proton and carbon chemical shift to their large percentage of –(di)ortho –OR substituents (> 80%); interestingly, a substantial percentage (> 65%) of –COR substituents are predicted in para positions.

*Group II* cross peaks exhibit negligible counts for –R (-OR) substituents in ortho-positions and zero probability for –COR substitution in –para position. *Group III* cross peaks exhibit zero counts for ortho-COR substituents and high probability of para-R substitution. *Group IV and V* cross peaks exhibit zero –OR substituents in ara position; *group IV* cross peaks almost exclusively have –R substituents in ortho- and para positions.

| Peak No. | δ $^{1}$H | δ $^{13}$C | Hits | ortho %_R | ortho %_COR | ortho %_OR | meta %_R | meta %_COR | meta %_OR | para %_R | Para %_COR | para %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GROUP I | | | | | | | | | | | | |
| 6 | 6,644 | 103,213 | 43 | 0,116 | 0,000 | 0,884 | 0,314 | 0,616 | 0,070 | 0,000 | 0,767 | 0,233 |
| 7 | 6,604 | 104,064 | 51 | 0,176 | 0,000 | 0,824 | 0,275 | 0,578 | 0,147 | 0,020 | 0,647 | 0,333 |
| 8 | 6,683 | 104,047 | 33 | 0,197 | 0,000 | 0,803 | 0,167 | 0,758 | 0,076 | 0,030 | 0,606 | 0,364 |
| 9 | 6,761 | 103,769 | 26 | 0,135 | 0,038 | 0,827 | 0,173 | 0,808 | 0,019 | 0,038 | 0,654 | 0,308 |
| average | 6,673 | 103,773 | 38,3 | 0,156 | 0,010 | 0,835 | 0,232 | 0,690 | 0,078 | 0,022 | 0,669 | 0,310 |
| GROUP II | | | | | | | | | | | | |
| 21 | 7,268 | 110,903 | 51 | 0,010 | 0,490 | 0,500 | 0,569 | 0,275 | 0,157 | 0,863 | 0,000 | 0,137 |
| 22 | 7,407 | 110,725 | 55 | 0,000 | 0,500 | 0,500 | 0,491 | 0,436 | 0,073 | 0,891 | 0,000 | 0,109 |
| 23 | 7,439 | 111,392 | 52 | 0,000 | 0,500 | 0,500 | 0,519 | 0,404 | 0,077 | 1,000 | 0,000 | 0,000 |
| 24 | 7,497 | 111,422 | 49 | 0,000 | 0,500 | 0,500 | 0,429 | 0,520 | 0,051 | 1,000 | 0,000 | 0,000 |
| 25 | 7,343 | 111,748 | 64 | 0,000 | 0,500 | 0,500 | 0,516 | 0,313 | 0,172 | 1,000 | 0,000 | 0,000 |
| 26 | 7,417 | 112,4 | 70 | 0,000 | 0,500 | 0,500 | 0,421 | 0,393 | 0,186 | 1,000 | 0,000 | 0,000 |
| average | 7,395 | 111,432 | 56,8 | 0,002 | 0,498 | 0,500 | 0,491 | 0,390 | 0,119 | 0,959 | 0,000 | 0,041 |
| GROUP III | | | | | | | | | | | | |
| 37 | 6,626 | 114,697 | 40 | 0,525 | 0,000 | 0,475 | 0,375 | 0,250 | 0,375 | 0,950 | 0,000 | 0,050 |
| 38 | 6,724 | 114,712 | 44 | 0,545 | 0,000 | 0,455 | 0,466 | 0,307 | 0,227 | 0,773 | 0,136 | 0,091 |
| 39 | 6,814 | 115,009 | 50 | 0,570 | 0,000 | 0,430 | 0,490 | 0,360 | 0,150 | 0,600 | 0,260 | 0,140 |
| average | 6,721 | 114,806 | 44,7 | 0,547 | 0,000 | 0,453 | 0,444 | 0,306 | 0,251 | 0,774 | 0,132 | 0,094 |
| GROUP IV | | | | | | | | | | | | |
| 74 | 7,103 | 127,46 | 43 | 1,000 | 0,000 | 0,000 | 0,395 | 0,279 | 0,326 | 0,953 | 0,047 | 0,000 |
| 75 | 7,212 | 127,702 | 39 | 1,000 | 0,000 | 0,000 | 0,346 | 0,372 | 0,282 | 0,897 | 0,103 | 0,000 |
| 76 | 7,339 | 127,702 | 38 | 0,908 | 0,092 | 0,000 | 0,316 | 0,421 | 0,263 | 0,842 | 0,158 | 0,000 |
| 77 | 7,382 | 127,981 | 44 | 0,898 | 0,102 | 0,000 | 0,341 | 0,364 | 0,295 | 0,682 | 0,318 | 0,000 |
| average | 7,259 | 127,711 | 41,0 | 0,952 | 0,049 | 0,000 | 0,350 | 0,359 | 0,292 | 0,844 | 0,157 | 0,000 |
| GROUP V | | | | | | | | | | | | |
| 102 | 7,823 | 130,316 | 65 | 0,708 | 0,292 | 0,000 | 0,231 | 0,477 | 0,292 | 0,508 | 0,492 | 0,000 |
| 103 | 7,774 | 131,297 | 56 | 0,768 | 0,232 | 0,000 | 0,223 | 0,482 | 0,295 | 0,339 | 0,661 | 0,000 |
| 104 | 7,876 | 131,291 | 55 | 0,636 | 0,364 | 0,000 | 0,327 | 0,345 | 0,327 | 0,436 | 0,564 | 0,000 |
| average | 7,824 | 130,968 | 58,7 | 0,704 | 0,296 | 0,000 | 0,260 | 0,435 | 0,305 | 0,428 | 0,572 | 0,000 |

**Table 26:** SPARIA inverse mode analysis of aromatic substitution patterns in five groups of $^{1}$H,$^{13}$C HSQC cross peaks of an alkaline soil extract (cf. Figures 70. 71. 72)

In general, meta-positions show considerable contributions from all substituents, while several ortho- and para-positions within *groups I – V* exhibit strong preference or exclusiveness for one specific group of substituents.

On close inspection of the results obtained by SPARIA in its inverse mode, a more detailed and specific picture emerges. SPARIA at first calculates the distribution of 8 substituents among the five remaining positions in molecules $C_6HR_l(COR)_m(OR)_n$ [l + m + n = 5] and provides individual substituent assignments for both ortho- and meta- positions and one para position. In all previous analyses, these individual counts (cf. Tab. 25, 26) have been merged into average group counts for ortho-, meta- and para positions. When the entire hit list of individual SPARIA substituent assignments for any given cross peak is considered, the single sets of proposed molecular structures can be decomposed into two groups, which normally show a much more decisive and discriminating pattern of predicted substituent distribution. This behavior is illustrated for the *group I* cross peaks 6 – 9 in figure 74 and Table 27. In the standard SPARIA inverse analysis, peak 6 shows 43 hits and here, the only exclusive count is the predicted absence of –COR substituents in ortho positions. These 43 molecules can be sorted into two groups of 33 and 10 molecules; here, the sorting was performed according to classes of substituents in ortho > para > meta positions. While the probabilistic analysis of peak 6 merely indicates preference for –OR substitution in ortho- and –COR substitution in meta- and para positions, the individual molecule level analysis yields two groups of structures, in which all ortho- and para positions are assigned to one specific class of substituents. While the major group of molecules (33 hits / 77%) features two –OR ortho- and one para-COR-substituent, the minor group (10 hits / 23 %) exhibits one –OR substituent each in ortho- and para-positions, one ortho-R substituent and one COR in meta-position. Ambiguity remains solely for the prediction of the meta substituents for reasons described in chapter 5.4.5). Likewise, breakdown into two or, alternative three groups (peak 9), provides specific assignments for a minimum of three ring positions (table 27). It is obvious, that the superposition of these subgroups, which in cases represent fairly different structures, will result in the considerable scatter observed in the standard SPARIA model.

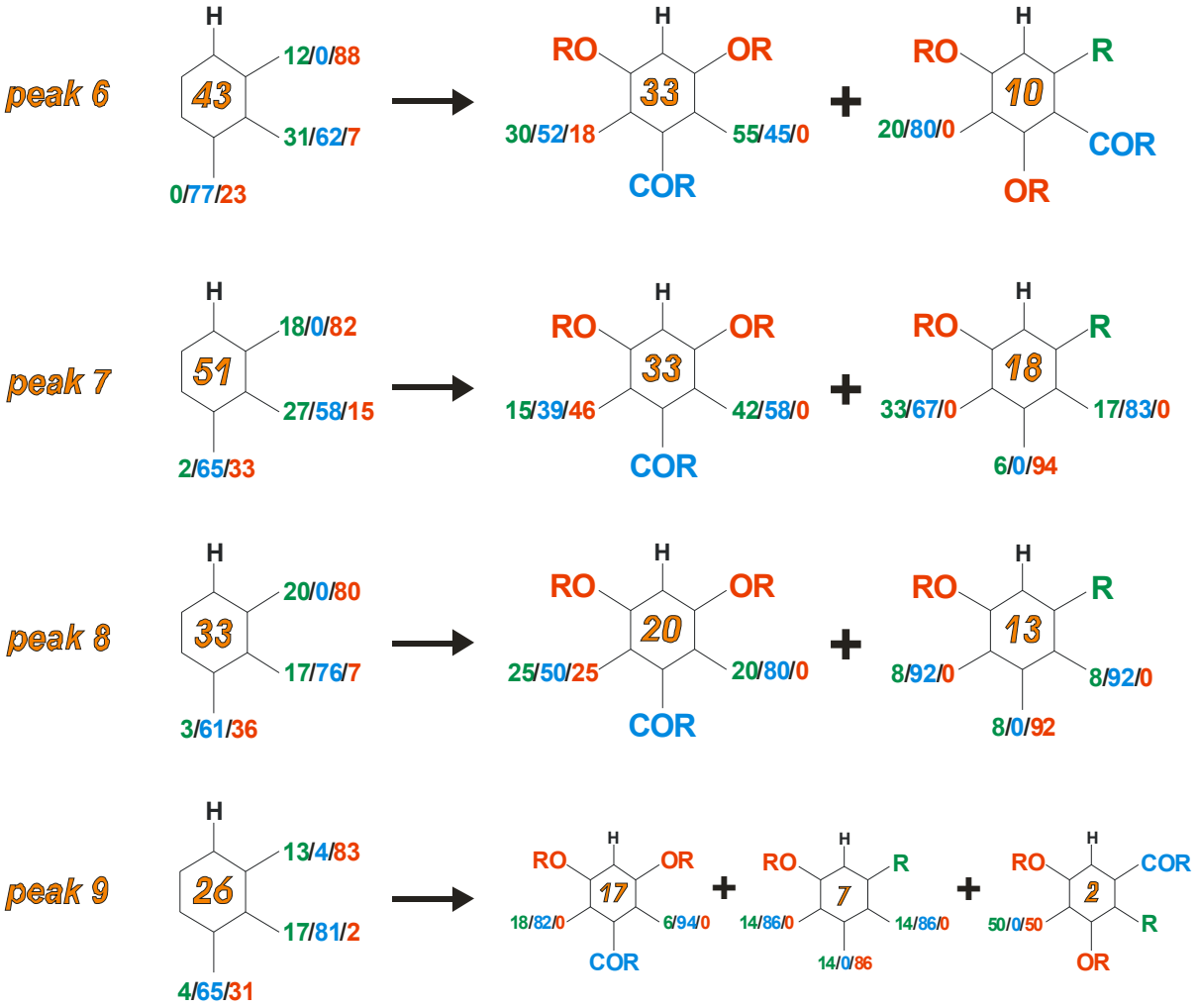



**Figure 74.** individual molecule analysis of aromatic substitution patterns for group I HSQC cross peaks 6-9, starting from standard SPARIA inverse mode analysis (cf. Table 25/1 and 27/1); letters denote 100% probability, numbers denote percent probability of classes of substituents R (green), COR (blue), OR (red).

| Peak No. | δ $^{1}$H | δ $^{13}$C | hits | hits_% | 1 ortho %_R | 1 ortho %_COR | 1 ortho %_OR | 2 meta %_R | 2 meta %_COR | 2 meta %_OR | 3 para %_R | 3 para %_COR | 3 para %_OR | 4 meta %_R | 4 meta %_COR | 4 meta %_OR | 5 ortho %_R | 5 ortho %_COR | 5 ortho %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GROUP I | | | | | | | | | | | | | | | | | | | |
| 6 | 6,64 | 103,2 | 43 | | 0,116 | 0,000 | 0,884 | 0,314 | 0,616 | 0,070 | 0,000 | 0,767 | 0,233 | | | | | | |
| 6 | | | 10 | 23 | R | 0 | 0 | 0 | COR | 0 | 0 | 0 | OR | 20 | 80 | 0 | 0 | 0 | OR |
| 6 | | | 33 | 77 | 0 | 0 | OR | 30 | 60 | 0 | 0 | COR | 0 | 30 | 52 | 18 | 0 | 0 | OR |
| 7 | 6,60 | 104,1 | 51 | | 0,176 | 0,000 | 0,824 | 0,275 | 0,578 | 0,147 | 0,020 | 0,647 | 0,333 | | | | | | |
| 7 | | | 18 | 35 | R | 0 | 0 | 17 | 83 | 0 | 6 | 0 | 94 | 33 | 66 | 0 | 0 | 0 | OR |
| 7 | | | 33 | 65 | 0 | 0 | OR | 42 | 58 | 0 | 0 | COR | 0 | 15 | 39 | 46 | 0 | 0 | OR |
| 8 | 6,68 | 104,0 | 33 | | 0,197 | 0,000 | 0,803 | 0,167 | 0,758 | 0,076 | 0,030 | 0,606 | 0,364 | | | | | | |
| 8 | | | 13 | 39 | R | 0 | 0 | 8 | 92 | 0 | 8 | 0 | 92 | 8 | 92 | 0 | 0 | 0 | OR |
| 8 | | | 20 | 61 | 0 | 0 | OR | 20 | 80 | 0 | 0 | COR | 0 | 25 | 50 | 25 | 0 | 0 | OR |
| 9 | 6,76 | 103,8 | 26 | | 0,135 | 0,038 | 0,827 | 0,173 | 0,808 | 0,019 | 0,038 | 0,654 | 0,308 | | | | | | |
| 9 | | | 7 | 27 | R | 0 | 0 | 14 | 86 | 0 | 14 | 0 | 86 | 14 | 86 | 0 | 0 | 0 | OR |
| 9 | | | 2 | 8 | 0 | COR | 0 | R | 0 | 0 | 0 | 0 | OR | 50 | 0 | 50 | 0 | 0 | OR |
| 9 | | | 17 | 65 | 0 | 0 | OR | 6 | 94 | 0 | 0 | COR | 0 | 18 | 82 | 0 | 0 | 0 | OR |

**Table 27/1:** individual molecule analysis of aromatic substitution patterns for group I HSQC cross peaks, starting from standard SPARIA inverse mode analysis (cf. Table 25/1); letters in color boxes denote 100% probability, numbers denote percent probability (zero probability in box and in color), hit_% denotes statistical weight of the respective subclass;

| Peak No. | δ $^{1}$H | δ $^{13}$C | hits | hits_% | 1 ortho %_R | 1 ortho %_COR | 1 ortho %_OR | 2 meta %_R | 2 meta %_COR | 2 meta %_OR | 3 para %_R | 3 para %_COR | 3 para %_OR | 4 meta %_R | 4 meta %_COR | 4 meta %_OR | 5 ortho %_R | 5 ortho %_COR | 5 ortho %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GROUP II | | | | | | | | | | | | | | | | | | | |
| 21 | 7,27 | 110,9 | 51 | | 0,010 | 0,490 | 0,500 | 0,569 | 0,275 | 0,157 | 0,863 | 0,000 | 0,137 | | | | | | |
| 21 | | | 7 | 14 | 0 | COR | 0 | 14 | 86 | 0 | 0 | 0 | OR | 14 | 43 | 43 | 0 | 0 | OR |
| 21 | | | 44 | 86 | 2 | 98 | 0 | 82 | 18 | 0 | R | 0 | 0 | 45 | 25 | 30 | 0 | 0 | OR |
| 22 | 7,41 | 110,7 | 55 | | 0,000 | 0,500 | 0,500 | 0,491 | 0,436 | 0,073 | 0,891 | 0,000 | 0,109 | | | | | | |
| 22 | | | 6 | 11 | 0 | COR | 0 | 0 | COR | 0 | 0 | 0 | OR | 0 | COR | 0 | 0 | 0 | OR |
| 22 | | | 49 | 89 | 0 | COR | 0 | 82 | 18 | 0 | R | 0 | 0 | 45 | 39 | 16 | 0 | 0 | OR |
| 23 | 7,44 | 111,4 | 52 | | 0,000 | 0,500 | 0,500 | 0,519 | 0,404 | 0,077 | 1,000 | 0,000 | 0,000 | | | | | | |
| 23 | | | 20 | 38 | 0 | COR | 0 | 0 | COR | 0 | R | 0 | 0 | 45 | 25 | 30 | 0 | 0 | OR |
| 23 | | | 32 | 62 | 0 | COR | 0 | R | 0 | 0 | R | 0 | 0 | 41 | 53 | 6 | 0 | 0 | OR |
| 24 | 7,50 | 111,4 | 49 | | 0,000 | 0,500 | 0,500 | 0,429 | 0,520 | 0,051 | 1,000 | 0,000 | 0,000 | | | | | | |
| 24 | | | 23 | 47 | 0 | COR | 0 | R | 0 | 0 | R | 0 | 0 | 35 | 48 | 17 | 0 | 0 | OR |
| 24 | | | 26 | 53 | 0 | COR | 0 | 0 | COR | 0 | R | 0 | 0 | 31 | 65 | 4 | 0 | 0 | OR |
| 25 | 7,34 | 111,7 | 64 | | 0,000 | 0,500 | 0,500 | 0,516 | 0,313 | 0,172 | 1,000 | 0,000 | 0,000 | | | | | | |
| 25 | | | 37 | 58 | 0 | COR | 0 | R | 0 | 0 | R | 0 | 0 | 40 | 49 | 11 | 0 | 0 | OR |
| 25 | | | 27 | 42 | 0 | COR | 0 | 19 | 81 | 0 | R | 0 | 0 | 33 | COR | 66 | 0 | 0 | OR |
| 26 | 7,42 | 112,4 | 70 | | 0,000 | 0,500 | 0,500 | 0,421 | 0,393 | 0,186 | 1,000 | 0,000 | 0,000 | | | | | | |
| 26 | | | 32 | 46 | 0 | COR | 0 | 0 | COR | 0 | R | 0 | 0 | 28 | 22 | 50 | 0 | 0 | OR |
| 26 | | | 38 | 54 | 0 | COR | 0 | 97 | 0 | 3 | R | 0 | 0 | 34 | 42 | 24 | 0 | 0 | OR |

**Table 27/2:** individual molecule analysis of aromatic substitution patterns for group II HSQC cross peaks, starting from standard SPARIA inverse mode analysis (cf. Table 25/2); letters in color boxes denote 100% probability, numbers denote percent probability (zero probability in box and in color), hit_% denotes statistical weight of the respective subclass;

| Peak No. | δ $^{1}$H | δ $^{13}$C | hits | hits_% | 1 ortho %_R | 1 ortho %_COR | 1 ortho %_OR | 2 meta %_R | 2 meta %_COR | 2 meta %_OR | 3 para %_R | 3 para %_COR | 3 para %_OR | 4 meta %_R | 4 meta %_COR | 4 meta %_OR | 5 ortho %_R | 5 ortho %_COR | 5 ortho %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GROUP III | | | | | | | | | | | | | | | | | | | |
| 37 | 6,63 | 114,7 | 40 | | 0,525 | 0,000 | 0,475 | 0,375 | 0,250 | 0,375 | 0,950 | 0,000 | 0,050 | | | | | | |
| 37 | | | 38 | 95 | R | 0 | 0 | 50 | 42 | 8 | R | 0 | 0 | 19 | 10 | 71 | 0 | 0 | OR |
| 37 | | | 2 | 5 | R | 0 | 0 | R | 0 | 0 | 0 | 0 | OR | R | 0 | 0 | R | 0 | 0 |
| 38 | 6,72 | 114,7 | 44 | | 0,545 | 0,000 | 0,455 | 0,466 | 0,307 | 0,227 | 0,773 | 0,136 | 0,091 | | | | | | |
| 38 | | | 40 | 91 | R | 0 | 0 | 50 | 48 | 2 | 0 | 0 | OR | 32 | 20 | 48 | 0 | 0 | OR |
| 38 | | | 4 | 9 | R | 0 | 0 | R | 0 | 0 | 0 | 0 | OR | R | 0 | 0 | R | 0 | 0 |
| 39 | 6,81 | 115,0 | 50 | | 0,570 | 0,000 | 0,430 | 0,490 | 0,360 | 0,150 | 0,600 | 0,260 | 0,140 | | | | | | |
| 39 | | | 43 | 86 | R | 0 | 0 | 47 | 53 | 0 | 70 | 30 | 0 | 37 | 30 | 33 | 0 | 0 | OR |
| 39 | | | 7 | 14 | R | 0 | 0 | R | 0 | 0 | 0 | 0 | OR | 86 | 0 | 14 | 0 | 0 | OR |

**Table 27/3:** individual molecule analysis of aromatic substitution patterns for group III HSQC cross peaks, starting from standard SPARIA inverse mode analysis (cf. Table 25/3); letters in color boxes denote 100% probability, numbers denote percent probability (zero probability in box and in color), hit_% denotes statistical weight of the respective subclass;

| Peak No. | δ $^{1}$H | δ $^{13}$C | hits | hits_% | 1 ortho %_R | 1 ortho %_COR | 1 ortho %_OR | 2 meta %_R | 2 meta %_COR | 2 meta %_OR | 3 para %_R | 3 para %_COR | 3 para %_OR | 4 meta %_R | 4 meta %_COR | 4 meta %_OR | 5 ortho %_R | 5 ortho %_COR | 5 ortho %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GROUP IV | | | | | | | | | | | | | | | | | | | |
| 74 | 7,10 | 127,5 | 43 | | 1,000 | 0,000 | 0,000 | 0,395 | 0,279 | 0,326 | 0,953 | 0,047 | 0,000 | | | | | | |
| 74 | | | 41 | 95 | R | 0 | 0 | 51 | 42 | 7 | R | 0 | 0 | 22 | 17 | 61 | R | 0 | 0 |
| 74 | | | 2 | 5 | R | 0 | 0 | R | 0 | 0 | 0 | COR | 0 | R | 0 | 0 | R | 0 | 0 |
| 75 | 7,21 | 127,7 | 39 | | 1,000 | 0,000 | 0,000 | 0,346 | 0,372 | 0,282 | 0,897 | 0,103 | 0,000 | | | | | | |
| 75 | | | 35 | 90 | R | 0 | 0 | 34 | 57 | 9 | R | 0 | 0 | 20 | 26 | 54 | R | 0 | 0 |
| 75 | | | 4 | 10 | R | 0 | 0 | R | 0 | 0 | 0 | COR | 0 | R | 0 | 0 | R | 0 | 0 |
| 76 | 7,34 | 127,7 | 38 | | 0,908 | 0,092 | 0,000 | 0,316 | 0,421 | 0,263 | 0,842 | 0,158 | 0,000 | | | | | | |
| 76 | | | 31 | 82 | R | 0 | 0 | 39 | 61 | 0 | 81 | 19 | 0 | 35 | 42 | 23 | R | 0 | 0 |
| 76 | | | 7 | 18 | R | 0 | 0 | 0 | 0 | OR | R | 0 | 0 | 0 | 0 | OR | 0 | COR | 0 |
| 77 | 7,38 | 128,0 | 44 | | 0,898 | 0,102 | 0,000 | 0,341 | 0,364 | 0,295 | 0,682 | 0,318 | 0,000 | | | | | | |
| 77 | | | 35 | 80 | R | 0 | 0 | 49 | 51 | 0 | 60 | 40 | 0 | 34 | 40 | 26 | R | 0 | 0 |
| 77 | | | 9 | 20 | R | 0 | 0 | 0 | 0 | OR | R | 0 | 0 | 0 | 0 | OR | 0 | COR | 0 |

**Table 27/4:** individual molecule analysis of aromatic substitution patterns for group IV HSQC cross peaks, starting from standard SPARIA inverse mode analysis (cf. Table 25/4); letters in color boxes denote 100% probability, numbers denote percent probability (zero probability in box and in color), hit_% denotes statistical weight of the respective subclass;

| Peak No. | δ $^{1}$H | δ $^{13}$C | hits | hits_% | 1 ortho %_R | 1 ortho %_COR | 1 ortho %_OR | 2 meta %_R | 2 meta %_COR | 2 meta %_OR | 3 para %_R | 3 para %_COR | 3 para %_OR | 4 meta %_R | 4 meta %_COR | 4 meta %_OR | 5 ortho %_R | 5 ortho %_COR | 5 ortho %_OR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GROUP V | | | | | | | | | | | | | | | | | | | |
| 102 | 7,82 | 130,3 | 65 | | 0,708 | 0,292 | 0,000 | 0,231 | 0,477 | 0,292 | 0,508 | 0,492 | 0,000 | | | | | | |
| 102 | | | 52 | 80 | R | 0 | 0 | 23 | 67 | 10 | 38 | 62 | 0 | 17 | 46 | 37 | 52 | 48 | 0 |
| 102 | | | 13 | 20 | 0 | COR | 0 | 62 | 23 | 15 | R | 0 | 0 | 8 | 0 | 92 | R | 0 | 0 |
| 103 | 7,77 | 131,3 | 56 | | 0,768 | 0,232 | 0,000 | 0,223 | 0,482 | 0,295 | 0,339 | 0,661 | 0,000 | | | | | | |
| 103 | | | 47 | 84 | R | 0 | 0 | 28 | 66 | 6 | 21 | 79 | 0 | 17 | 49 | 34 | 64 | 36 | 0 |
| 103 | | | 9 | 16 | 0 | COR | 0 | 44 | 0 | 56 | R | 0 | 0 | 0 | 0 | OR | R | 0 | 0 |
| 104 | 7,88 | 131,3 | 55 | | 0,636 | 0,364 | 0,000 | 0,327 | 0,345 | 0,327 | 0,436 | 0,564 | 0,000 | | | | | | |
| 104 | | | 43 | 78 | R | 0 | 0 | 44 | 47 | 9 | 30 | 70 | 0 | 23 | 35 | 42 | 35 | 65 | 0 |
| 104 | | | 12 | 22 | 0 | COR | 0 | 58 | 25 | 17 | 92 | 8 | 0 | 0 | COR | 0 | R | 0 | 0 |

**Table 27/5:** individual molecule analysis of aromatic substitution patterns for group V HSQC cross peaks, starting from standard SPARIA inverse mode analysis (cf. Table 25/5); letters in color boxes denote 100% probability, numbers denote percent probability (zero probability in box and in color), hit_% denotes statistical weight of the respective subclass;

The standard SPARIA inverse mode analysis has been developed to analyze cross peaks in heteronuclear two-dimensional NMR spectra. At present, the homonuclear version of SPARIA is developed, which is capable to analyze homonuclear 2D-NMR spectra (cf. Figure 48). Cross peaks in these spectra require at least two hydrogen atoms in the aromatic ring, whose chemical shift is strongly affected by all remaining substituents. A combined homo- and heteronuclear SPARIA inverse analysis of aromatic substitution patterns in NOM and other complex materials will resolve many of the remaining ambiguities of the current approach; nevertheless, even in its current version, SPARIA provides unprecedented molecular level detail in the analysis of aromatic systems in NOM and other complex materials.

## 6. The Role of NMR-based Structural Analysis of NOM/HS in Remediation

NMR spectroscopy and other molecular level precision analytical methods, like Fourier transform ion cyclotron mass spectrometry FTICR-MS, are indispensable for the structural analysis of NOM/HS and for the elucidation of reaction mechanisms in which these materials participate. However, remediation technologies are rather concerned with an expedient, affordable and swift solution of a practical problem, often within the framework of non – scientific legal and political regulations. In this respect the scope of fundamental research regarding NOM/HS structural analysis is to advance and to distribute knowledge in science and in the public and, on the practical level, to provide guidance for optimising experiments and results. For that reason, a firm relationship needs to be established between the information rich and often highly correlated analytical data and the reactivity of the respective NOM/HS materials towards pollutants. The NMR-based protocols developed for this purpose typically require an initial data reduction step, like bucket analysis or segmental integration, followed by pattern recognition or chemometric methods for classification and maximum information recovery. Various methods of multivariate analysis, like principal component analysis PCA (Brindle et al., 2002; Dong et al., 2000; Keun, et al., 2002), non-linear mapping analysis and probabilistic neural networks PNN (Holmes et al., 2001) are used to interpret NMR data. Very frequently a set of structure-based descriptors is obtained which allows to predict the behaviour and the reactivity of NOM/HS in a specific environment. Along these lines, NMR data of NOM/HS have been employed to classify NOM/HS according to origin (Visser, 1983; Thomsen et al., 2002), performance in analytical methods (Ussiri and Johnson, 2003; Fansioso et al., 2003) and reactivity towards organic (Kulikova and Perminova, 2002; Perminova et al., 2001) and inorganic (Balcke et al., 2002) materials.

These studies furnish compelling evidence that carefully developed protocols of NMR characterization of NOM/HS in conjunction with mathematic analysis provide a robust and reliable guidance to optimise the experimental design in various stages of remediation.

## 7. Experimental

All NMR spectra shown have been acquired by the author with a Bruker DMX 500 NMR spectrometer (Rheinstetten, Germany), operating at 500.13 MHz proton frequency from various NOM/HS materials, according to NMR acquisition and processing parameters as given in Table 29; referencing is performed according to Harris et al., 2002. Minimum S/N ratio corresponds to good visibility of main resonances; fair S/N ratio allows for the detection of minor compononents, however with insufficient S/N ratio for quantification or lineshape assessment; NMR spectra of good S/N ratio can be analyzed with respect to content and structural detail of minor NOM constituents.

**Table 28.** substance amounts and key acquisition parameters for routine quantitative NMR spectra of NOM/HS (90-degree excitation pulses $B_0$ = 11.7 T; 5 mm NMR sample tubes; cf. chapters 2.3., 2.4., 2.5., 2.6., 3.1.-3.1.4.1., 4., 5.)

| nucleus | minimum S/N | fair S/N | good S/N | aq [sec] | d1 [sec] |
|---|---|---|---|---|---|
| 1D $^{1}$H NMR | 25 µg | 500 µg | 5 mg* | 4.5 | 16 |
| 1D $^{13}$C NMR | 1 mg | 7 mg | 40 mg | 0.25 | 8 |
| 1D $^{15}$N NMR | 5 mg | 50 mg | 135 mg | 0.25 | 45 |
| 1D $^{31}$P NMR | 1 mg | 7 mg | 40 mg | 0.3 | 4 |
| $^{1}$H,$^{1}$H-COSY | 150 µg | 3 mg | 10 mg | 0.4 | 2 |
| $^{1}$H,$^{1}$H-TOCSY | 75 µg | 3 mg | 10 mg | 0.2 | 2 |
| $^{1}$H,$^{13}$C-HSQC | 1 mg | 7 mg | 35 mg | 0.2 | 2 |

* higher amounts of NOM / HS (750 µl solvent) would lead to intensity distortions and line broadening, caused by extensive interaction.

**Table 29.** Key acquisition and processing parameters given for NMR spectra displayed in Figures 14 - 48.

| Figure | PK [a] | NS | AQ | D1 | NE | WDW1 | WDW2 | PR1 | PR2 |
|---|---|---|---|---|---|---|---|---|---|
| **14a-c** | CR | 256 | 4360 | 15400 | | EM | | 1 | |
| **15 ($^{13}$C)** | B5 | 17684 | 263 | 8000 | | EM | | 50 | |
| **15 ($^{31}$P)** | B5 | 57241 | 400 | 3600 | | EM | | 35 | |
| **16** | B10 | 54155 | 233 | 3000 | | EM | | 100 | |
| **27 ($^{11}$B)** | B10 | 26594 | 341 | 1000 | | EM | | 50 | |
| **27 ($^{17}$O)** | B5 | 2734808 | 38 | 40 | | EM | | 150 | |
| **29** | CR | 160 | 341 | 1329 | 1779 | QS | SI | 0 | 0 |
| **30** | CR | 80 | 1507 | 100 | 4096 | EM | EM | 4 | 4 |
| **31** | CR | 48 | 137 | 1800 | 400 | QS | SI | 3 | 3 |
| **32** | CR | 640 | 585 | 1215 | 164 | QS | SI | 0 | 0 |
| **33** | I5 | 416 | 250 | 2750 | 706 | EM | SI | 35 | 2.5 |
| **34 ($^{1}H_{det}$)** | I5 | 416 | 250 | 2750 | 706 | EM | SI | 35 | 2.5 |
| **34 ($^{13}C_{det}$)** | B5 | 4096 | 250 | 2840 | 67 | EM | GM | 5 | –1/0.3 |
| **44 ($^{29}$Si)** | B5 | 22849 | 1101 | 1500 | | EM | | 35 | |
| **45 ($^{29}$Si)** | I5 | 256 | 250 | 1750 | 512 | QS | SI | 3 | 3 |
| **45 ($^{13}$C)** | I5 | 256 | 250 | 1750 | 512 | QS | SI | 3 | 3 |
| **46** | I5 | 32 | 187 | 1500 | 848 | QS | SI | 6 | 6 |
| **47a** | B5 | 79552 | 250 | 2750 | | EM | | * | |
| **47b** | I5 | 512 | 171 | 1500 | 512 | EM | SI | given | given |
| **48** | CR | 160 | 341 | 1329 | 1779 | QS | SI | 0 | 0 |

* cf. Figure 47; [a] probeheads used for acquisition of NMR spectra, B5: 5 mm broadband observe; CR: cryogenic 5 mm $^{1}$H, $^{13}$C, $^{15}$N inverse geometry TXI/z-gradient; B10: 10 mm broad band observe; I5: 5 mm inverse geometry broad band/z-gradient; NS: number of scans (for 2D NMR: F2); AQ: acquisition time [ms]; D1: relaxation delay [ms]; NE: number of F1 increments in 2D NMR spectra; WDW1, WDW2 windowing function in F1/F2: EM: exponential multiplication [Hz], GM: Lorentz-to-Gaussian transformation (EM/GM: line broadening factor [Hz] Gaussian multiplication factor GM [Hz]); QS: shifted square sine bell; SI: sine bell; PR1, PR2 coefficients used for windowing functions WDW1, WDW2, EM/GM are given in [Hz], SI/QS derived functions indicate shift by $^{\pi}$/n (n is given).

# 8. CASE STUDY 4: Characterization of a major refractory component of marine dissolved organic matter

## 8.1. ABSTRACT

***Refractory carboxyl-rich alicyclic molecules (CRAM) are characterized in marine dissolved organic matter (DOM) using nuclear magnetic resonance spectroscopy and ultrahigh resolution mass spectrometry. CRAM are distributed throughout the water column and are the most abundant components of deep ocean DOM ever characterized. CRAM are comprised of a complex mixture of carboxylated and fused alicyclic structures with a carboxyl-C:aliphatic-C ratio of 1:2 to 1:7. CRAM are expected to constitute a strong ligand for metal binding, and multiple coordination across cations could promote aggregation and marine gel formation thereby affecting CRAM reactivity and the bioavailability of nutrients and trace metals. It appears CRAM are ultimately derived from biomolecules with structural similarities to sterols and hopanoids. The occurrence of CRAM in freshwater and terrestrial environments seems likely, considering the global distribution of biomolecules and the similarities of biogeochemical processes among environments.***

### *8.1.1. Introduction*

The oceans are a major global reservoir of reduced carbon (700 Pg), most of which occurs as dissolved organic matter (DOM) (Hedges, 1992). Much of the DOM in the ocean is refractory and has an average radiocarbon age of several millennia (Druffel et al., 1992). Little is known about the chemical composition of DOM and the reason for its refractory nature (Hedges et al., 2000; Benner, 2002). At present, in-depth knowledge about the origin, structure and function of DOM in the global carbon cycle remains elusive. Molecular analyses combined with degradative techniques have identified specific carbohydrates, amino acids and lipids in marine DOM, but these compounds only account for a small fraction of dissolved organic carbon (DOC) (Benner, 2002). The concentration and isolation of DOM from seawater by ultrafiltration facilitates its characterization by spectroscopic techniques. This approach, in combination with molecular analyses, has identified complex heteropolysaccharides (HPS) as major constituents of marine ultrafiltered DOM (UDOM) (Benner et al., 1992; Vernonclark 1994; Aluwihare et al., 1997). A diverse suite of neutral, amino and acidic carbohydrates is found in HPS (McCarthy et al., 1996; Boon et al., 1998; Benner and Kaiser, 2003). The

abundance of HPS is maximal in surface waters and declines sharply with depth, indicating HPS is a rapidly cycling component of DOM in the upper ocean. However, the most prominent contributors to the total $^{13}C$ and $^{1}H$ nuclear magnetic resonance (NMR) integral in UDOM from deep water remain undefined and have been recognized only as a humic-like background in $^{1}H$-NMR spectra (Aluwihare et al., 2002). Tentative assignments derived from cross polarization / magic angle spinning $^{13}C$-NMR spectra of marine dissolved humic substances indicated extensively branched, interlinked and possibly cyclic aliphatic carbon (Hedges et al., 1992).

Two previously-characterized UDOM samples collected from surface (2 m) and deep (4000 m) waters of the mid Pacific Ocean were selected for this study (Benner et al., 1997). Additional samples from other locations were analyzed (data not shown), and all of the major structural features presented herein are generally representative of marine DOM and UDOM. The novel combination of $^{1}H$-NMR, $^{13}C$-NMR and Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) reveals the structural details of a major refractory constituent of marine UDOM. This newly characterized component is referred to as carboxyl-rich alicyclic molecules (CRAM). CRAM are slowly cycling components of marine UDOM, and they are the most abundant, identified components of DOM in the deep ocean.

## 8.2. MATERIALS AND METHODS

### *8.2.1. Sample collection*

Surface (2 m) and deep (4000 m) water samples were collected from the Pacific Ocean (6$^{o}$N, 140$^{o}$W and 2$^{o}$N, 140$^{o}$W) using Niskin bottles. Samples were passed through a Nitex sieve (60 μm mesh size) and a 0.1 μm pore-size filter before tangential-flow ultrafiltration with a 1 kDalton molecular-weight cutoff membrane to isolate ultrafiltered dissolved organic matter (UDOM; Benner et al. 1997). Ultrafiltered concentrates were diafiltered with deionized water to remove sea salts and freeze dried. The concentrations of dissolved organic carbon (DOC) in surface and deep waters were 72 μM and 44 μM, respectively. Ultrafiltration retained 26% of the DOC from surface water and 20% from deep water (Benner et al. 1997). The molar C:N ratios of the UDOM samples were 16.1 and 18.4, and the stable carbon isotopic compositions were –21.4‰ and –21.8‰ indicating a predominantly marine origin (Benner et al. 1997). Samples were also analyzed for hydrolysable amino acids (McCarthy et al., 1996), neutral

sugars (McCarthy et al., 1996; Skoog and Benner 1997) and amino sugars (Benner and Kaiser 2003) by high performance liquid chromatography.

### *8.2.2. NMR Spectroscopy*

All experiments in this study were performed with a Bruker DMX 500 spectrometer at 303 K. The samples were dissolved in 0.1 N NaOD for $^{1}$H-NMR, $^{13}$C-NMR and 2D NMR spectroscopy. No fractionation was observed in the samples. The reference for $^{1}$H-NMR was $(H_3C)_3Si\text{-}CD_2\text{-}CD_2\text{-}COONa$ (-0.14 ppm), and for $^{13}$C an external reference ($CH_3OH$ in $D_2O$: 49.00 ppm) was used. All proton detected NMR spectra were acquired with a 5 mm z-gradient $^{1}$H/$^{13}$C/$^{15}$N TXI cryogenic probe using 90$^{o}$ excitation pulses (90$^{o}$($^{1}$H) = 8.3 µs; 90$^{o}$($^{13}$C) = 19 µs). 1D $^{1}$H-NMR spectra were recorded using the first increment of the presat-NOESY sequence (solvent suppression with presaturation and spin-lock, 4.7 s acquisition time, 15.3 s relaxation delay, 320 scans, 1 ms mixing time, 1 Hz exponential line broadening). $^{13}$C-NMR spectra were acquired with a 5 mm broad-band probe (90$^{o}$($^{13}$C) = 9.2 µs), using inverse gated WALTZ-16 decoupling (8 s relaxation delay; 49152 scans for surface UDOM, 68542 scans for deep UDOM) with an acquisition time of 263 ms and an exponential line broadening of 35 Hz.

The one bond coupling constant $^{1}$J(CH) used in 1-D $^{13}$C DEPT and 2D $^{1}$H,$^{13}$C DEPT-HSQC spectra was set to 145 Hz [DEPT-HSQC NMR spectra were acquired with 320 scans and 312 increments (Willker, 1992)]. Carbon decoupled $^{1}$H,$^{13}$C-HSQC NMR spectra were acquired under the following conditions: $^{13}$C-90-deg decoupling pulse, GARP (70µs); F2 ($^{1}$H): acquisition time: 183 ms at spectral width of 5482 Hz, $^{1}$J(CH) = 150 Hz, 1.82 s relaxation delay; F1 ($^{13}$C): SW = 22637 Hz (180 ppm); number of scans(F2)/F1-increments ($^{13}$C frequency) for surface UDOM (1920/800) and deep UDOM (238/191), respectively. HSQC and DEPT-HSQC spectra were calculated to a 2048 x 512 matrix with exponential line broadening of 30 Hz in F2 and a shifted sine bell (π/6) in F1 (Figure 79, 80). Gradient, but not sensitivity-enhanced, sequences (1 ms length, 450 µs recovery) were used for all proton detected spectra.

#### *8.2.2.1. Analysis of NMR spectra.*

NMR integrals were measured manually from printed spectra. Difference NMR spectra were obtained using XWinNMR 3.0 software from spectra rather than from FIDs. In $^{13}$C NMR,

subtraction (d_UDOM – $x_n$ • s_UDOM) was performed to attempt removal of the $F1_C$-component (i.e. the N-acetyl group, which is supposedly a substituent of HPS) to result in CRAM (CRAM: $x_1 = 0.76$). Removal of the $F2_C$-component (supposedly the carboxylic acid group of CRAM) resulted in HPS (with negative amplitude at $x_2 = 2.22$). Conversely, the manual subtraction (s_UDOM – $x_n$ • d_UDOM) removed the $F2_C$ component to result in HPS ($x_3 = 0.44$) or, alternatively, the $C_C$ component (carbohydrate methine carbon atoms and $HOCH_2$-groups) to result in CRAM ($x_4 = 1.17$; initial negative amplitude).

Bucket analysis (Brindle, et al., 2002) was performed on the experimental $^{13}C$ NMR spectra of surface and deep UDOM, and of the computed difference NMR spectra, representing HPS and CRAM (Figure 75); these were decomposed into 47 equidistant integral segments with 5 ppm bandwidth, ranging from 0-235 ppm (Figure 77). Using the composition of a typical marine surface UDOM amino acid hydrolysate (with an average number of 2.83 side chain carbon atoms per CONH unit, which resulted, in conjunction with the atomic C/N ratio of 16.1, in 23.6% peptide carbon content in surface UDOM; Kaiser, 2005) and a typical marine deep UDOM amino acid hydrolysate (average number of 2.98 side chain carbon atoms per CONH unit, C/N atomic ratio: 18.4; resulting in 21.7 percent peptide carbon in deep UDOM; Kaiser, 2005), data reduced $^{13}C$-NMR spectra were computed via ACD CNMR predictor software, version 5.0 (Figure 77).

From a combination of these data-reduced $^{13}C$-NMR spectra (Figure 77), a rather conservative estimate for the minimum content of CRAM in UDOM was calculated, by use of a basic 3-component reverse mixing model (Nelson and Baldock, 2005) "UDOM = $n_1$ • HPS + $n_2$ • CRAM + $n_3$ • peptides" , based on the assumption that all nitrogen in UDOM is proteinaceous; linear regression analysis with equal weighting factors for all buckets provided the respective coefficients $n_i$ for the computed 3-component surface and deep UDOM shown in Figure 77. That procedure does not recognize inorganic nitrogen, nucleic acids and other heterocyclic nitrogen, amino- and N-acetyl sugars, which certainly are constituents of marine UDOM, and therefore provides an upper limit for the content of peptides.

#### *8.2.2.2. Computer generation and computing of NMR chemical shift data of an aliphatic model polycarboxylic acid $C_{644}H_{1142}O_{146}$*

The computer generation of an aliphatic model polycarboxylic acid $C_{644}H_{1142}O_{146}$ was initiated with a long straight chain aliphatic carbon backbone (polymethylene), carboxylic

groups were then introduced, and the spacing in between those was varied from one (like in polyacrylic acid) to four carbon atoms (like in polyamylic acid) in increasing order. A systematic pattern of increasing methylation throughout the entire chain was then applied to account for a variable degree of branching, resulting in 83 aliphatic methylene carbon (with no methyl attached), 83 methine carbon (with a single methyl attached), and 83 quaternary carbon (doubly methylated) atoms; in total, 249 methyl groups have been placed between the terminal carboxylic groups. The chemical environment of the 73 carboxylic groups necessary to accomplish all feasible substitution patterns (cf. Figure 81) was defined as H-C-COOH (resulting in another 73 methine carbon atoms), which act as spacers in between the methylated aliphatic chain units. The entire assembly results in a molecular formula $C_{644}H_{1142}O_{146}$ (excluding two terminal methyl groups), corresponding to a H/C atomic ratio of 1.773 and a O/C atomic ratio of 0.227 (cf. Figures 81, 85)

The individual proton and carbon chemical shifts of $C_{644}H_{1142}O_{146}$ were then computed [by means of Advanced Chemistry Development HNMR and CNMR predictor software (version 5.0)] from three sections ($C_{154}H_{310}O_{31}$; $C_{164}H_{330}O_{25}$; $C_{216}H_{434}O_{24}$, cf. Figure 81) with sufficient overlap to avoid effects of chain termination on the proton and/or carbon NMR chemical shifts for any given atom position.

### *8.2.3. Capillary electrophoresis*

Capillary zone electrophoresis (CZE) was performed using a PA/CE 5000 Beckman-Coulter instrument with a 57 cm fused silica capillary (50 cm length to the detector) and UV-detection (214 nm) as well as electrospray ionization negative mode MS detection (LCQ-Duo Finnigan ion trap mass spectrometer under conditions analogous to those in Schmitt-Kopplin and Kettrup, 2003). CZE separations were carried out at 30°C in 25 mM ammonium carbonate buffers at pH 9.3 and pH 11.4 under 25 kV. For capillary gel electrophoresis (CGE), 0.3% methylcellulose was added to the pH 9.3 carbonate buffer.

### *8.2.4. FTICR mass spectrometry*

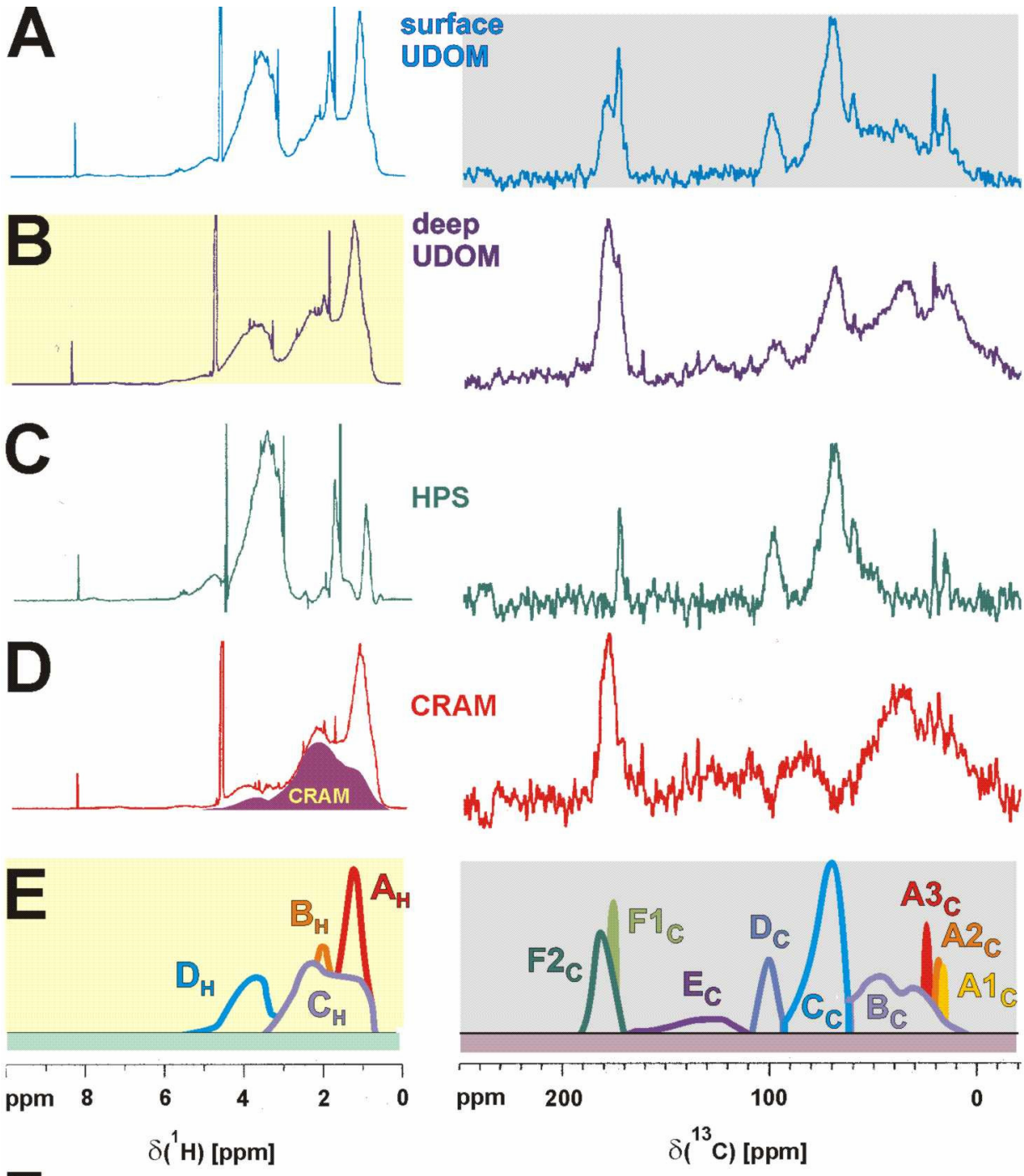
High-resolution mass spectra for molecular formula assignment were performed with a Bruker APEX-Q III Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) equipped with a 7 Tesla superconducting magnet using the electrospray positive ion mode of the ESI/MALDI combi source. The sample solution (1 mg UDOM in 1 ml methanol)

was diluted 1:1 with aqueous 0.2% formic acid and introduced into the electrospray source by infusion with a flow of 120 μL $h^{-1}$. The spectra were externally calibrated with the collision induced dissociation spectrum of the peptide luteinizing hormone-releasing hormone (Sigma, [35263-73-1]). The spectra were acquired with a time domain of 1 megaword with a mass range of 200-2000 m/z. The spectra were zero filled to a processing size of 2 megawords. Before Fourier transformation of the time-domain transient, a sine apodization was performed. The ions were mass selected between ion source and FTICR-MS analyzer with a quadrupole using a setting for ion selection of about 100 mass units. The ion accumulation time in the ion source was set to 3 s for each scan. 60 scans were added for each mass spectrum.
Standard $CH_2$-based Kendrick mass defect analysis has been performed on 613 CRAM molecular compositions $C_nH_mO_q$ derived from positive ions $C_nH_{m+1}O_q^+$ as obtained in 7 Tesla FTICR mass spectra. The $CO_2$-based Kendrick mass has been computed by multiplying the IUPAC mass of CRAM by the factor (44/43.989828). In order to keep the nominal Kendrick masses for given molecular compositions identical for both $CH_2$- and $CO_2$-based Kendrick mass analyses (Figure 84), we have subtracted one mass unit from the computed nominal $CO_2$-based Kendrick mass.

## 8.3. RESULTS AND DISCUSSION

### *8.3.1. analysis of one-dimensional $^1H$ and $^{13}C$ NMR spectra*

Four major functional forms of non-exchangeable hydrogen were recognizable in $^1$H-NMR spectra of surface and deep UDOM acquired in NaOD (Figures 75A, 75B). Strong resonances were observed in the 0.6-1.6 ppm region, which is characteristic of hydrogens bound to saturated carbon with heteroatoms three or more bonds away. Resonances in the 1.8-2.8 ppm region, which are characteristic of hydrogens bound to saturated carbon with heteroatoms two bonds away, and those in the 3.0-5.5 ppm region, which are characteristic of H-C bound to oxygen, were also prevalent. In contrast, only weak resonances (<1.5 % of total $^1$H NMR integral) were found in the olefinic and aromatic regions (6-9 ppm; Figure 76).




F

| NMR resonance | $D_H$ | $C_H$ | $B_H$ | $A_H$ | $F1_C$ | $F_C$ | $E_C$ | $D_C$ | $C_C$ | $B_C$ | $A3_C$ | $A2_C$ | $A1_C$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\delta(^1H/^{13}C)$ [ppm] | 3.0-6.0 | 0.6-3.6 | 1.95 | 1.19 | 175 | 168-187 | 108-168 | 90-108 | 60-90 | 0-60 | 22.5 | 17.3 | 15.5 |
| surface UDOM | 56 | 30 | 6 | 8 | 4.2 | 17.7 | 2.0 | 8.5 | 36.1 | 35.7 | 2 | 0 | 0 |
| deep UDOM | 30 | 62 | 3 | 5 | 5.8 | 21.4 | 6.4 | 5.7 | 21.3 | 45.2 | 1.6 | 0 | 0 |
| HPS | 83 | 0 | 9 | 6 | 5.8 | 5.8 | 0 | 20.5 | 73.7 | 0 | 2.4 | 1.1 | 0.9 |
| CRAM | 0 | 100 | 0 | 0 | 0 | 19.9 | 11.8 | 5.2 | 9.6 | 53.5 | 0 | 0 | 0 |

**Figure 75.** $^1$H- and $^{13}$C-NMR data of marine UDOM and its major constituents: acquired $^1$H (left) and $^{13}$C (right)

**Figure 75 (continued).** NMR spectra of surface (**A**) and deep (**B**) marine UDOM and, calculated from difference NMR spectra, of its two major constituents HPS (**C**) and CRAM (**D**). The $^{1}$H NMR line spectrum (**D**) represents CRAM plus contributing genuine peptides; the purple area indicates the CRAM component as obtained from difference NMR spectra after subtraction of peptides (artificial mixture of polypeptides, which is representative of marine UDOM hydrolysates); (**E**) Modeled spectra of deep UDOM ($^{1}$H NMR) and surface UDOM ($^{13}$C NMR) to indicate the NMR integral regions given in (**F**): $^{1}$H NMR integrals (left) of four major resonances $A_H$ – $D_H$ ($A_H$: methyl bound to purely aliphatic carbon, $B_H$: acetyl, $C_H$: functionalized CRAM and peptide aliphatics, $D_H$: heteroatom substituted protons (mostly carbohydrates and CH-α of peptides). $A_H$+$B_H$+$C_H$+$D_H$ are normalized to 100%, the total aromatic proton NMR integral for all materials is less than 1.5% (Figure 76). $^{13}$C NMR integrals of nine major resonances $A_C$ – $F_C$ ($A1_C$: (poly)alanine-$CH_3$; $A2_C$: methylated carbohydrates; $A3_C$: N-acetyl carbohydrate; $B_C$: all aliphatics, aliphatic side chains of CRAM and peptides, including Cα; $C_C$: single oxygen substituted carbon; $D_C$: anomeric and other doubly oxygenated carbon; $E_C$: aromatics; $F_C$: carbonyl derivative (carboxylic, amide, ester) carbon, with $F1_C$: carboxylic acids and N-acetyl from HPS, $F2_C$: carboxylic acids and peptides from CRAM (with $F_C$ = $F1_C$ + $F2_C$). $B_C$+$C_C$+$D_C$+$E_C$+$F_C$ are normalized to 100%; grey underlined resonances $A1_C$ -$A3_C$ and $F1_C$ (italics) are parts of aliphatic ($B_C$) and carbonyl derivative carbons ($F_C$), respectively.

Four major functional forms of carbon were recognizable in the $^{13}$C-NMR spectra (Figure 75A, 75B). In the 0-60 ppm region ($A_C$, $B_C$ in Figure 75E) carbon is mainly bound to carbon and hydrogen atoms only; however, aliphatic carbon of amines, amides (C-α of peptides; cf. Figure 79) and methoxyl groups also contribute here. The 60-108 ppm region ($C_C$ and $D_C$ in Figure 75E) is characteristic of carbon bound to one or more heteroatoms, and the 108-168 ppm region ($E_C$ in Figure 75E) is characteristic of olefinic and aromatic carbon. Carbonyl derivatives (esters, amides and carboxylic acids) resonate from 168-225 ppm ($F_C$ in Figure 75E). Strong resonances were observed in all these regions except that of olefinic and aromatic C.

The NMR spectra of surface and deep UDOM closely resembled each other in the positions of their relative intensity minima and maxima, but differ vastly in their relative amplitudes (Figure 75A, 75B). This indicated surface and deep UDOM were composed of variable amounts of preserved substructures, which can be visualized by difference NMR spectra (Figure 75C, 75D). Two major UDOM constituents were observed in difference spectra. A complex mixture of heteropolysaccharides (HPS) was represented by surface minus deep UDOM (Figure 75C), with major resonances in $D_H$ and $C_C$+$D_C$ (Figure 75E). A mixture, we refer to as carboxyl-rich alicyclic molecules (CRAM) was represented by deep minus surface UDOM (Figure 75D), with major resonances in $C_H$, $B_C$ and $F2_C$ (Figure 75E).

The larger peak capacity of $^{13}$C NMR spectroscopy (ratio of total bandwidth / individual linewidth; Figure 5) relative to that of $^{1}$H NMR, in conjunction with the capability to directly observe quaternary carbon atoms, allows to detect finer resolved structural detail in $^{13}$C NMR spectra of UDOM (Figure 75E), albeit at a lesser signal to noise ratio (at natural abundance,

the NMR receptivity of $^{13}C$ nuclei is 1/5682 compared with that of proton nuclei). Alternatively, the very high sensitivity of high-field proton NMR spectra allows to determine rather tiny variations in UDOM composition in minute detail (Figure 76), providing crucial (and quantitative) structural information not available by any other means at present.

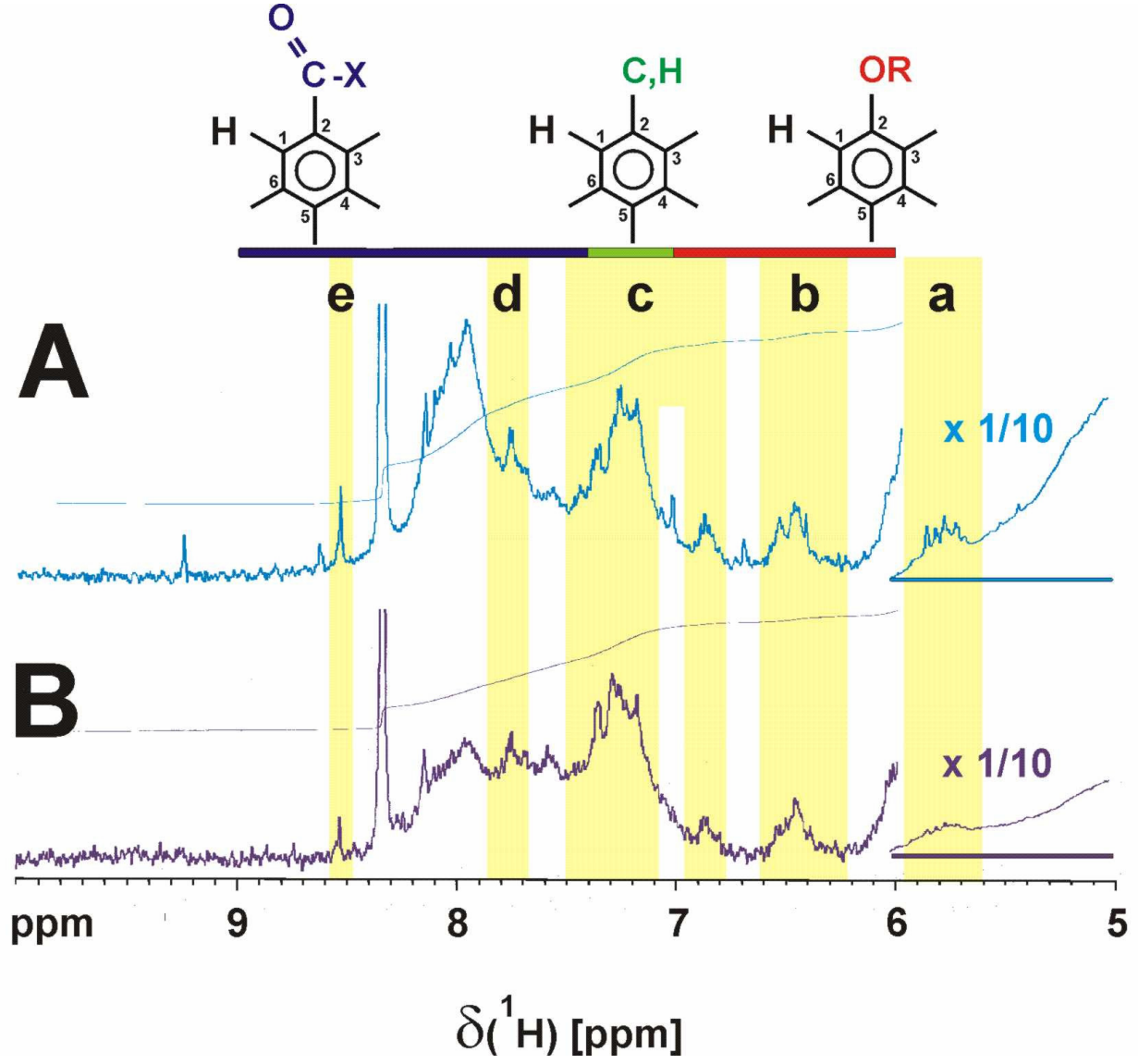



**Figure 76.** low field section of the $^{1}H$-NMR spectra of (A) surface (blue) and (B) deep (purple) UDOM (expansion of Figure 75A,B), indicating resonances of non-exchangeable protons (solvent: 0.1 N NaOD) occurring in both UDOM spectra (yellow boxes a–e), suggesting conformity of five-membered heterocyclic ring (a), phenolic (b), C-substituted (c), and carbonyl-substituted (d,e) aromatic protons in both environments. In contrast, low field ($\delta$ ($^{1}H$) > 7.8 ppm) aromatic carbonyl derivatives are much more abundant in surface than in deep UDOM. Small amounts of formiate ($\delta$ ($^{1}H$) = 8.31 ppm), methanol ($\delta$ ($^{1}H$) = 3.24 ppm; Figure 75) and acetate ($\delta$ ($^{1}H$) = 1.82 ppm; Figure 75) are formed by hydrolysis of UDOM in NaOD.

Difference NMR spectra emphasize components that have varying concentrations. In the spectra (Figures 75C and 75D), they describe HPS and CRAM, because peptides comprise a similar fraction of carbon in surface and deep UDOM (McCarthy et al., 1996; Hedges et al., 2001).

*8.3.1.1. Computation of the CRAM content in UDOM from difference NMR spectra*

The main UDOM constituents HPS and CRAM were visualized by difference NMR spectra (cf. Figure 75), which were created manually from spectra rather than from FIDs. Poor resolution and S/N ratio ($^{13}$C NMR) of UDOM NMR spectra, even after extended acquisition time, made it advisable to avoid negative signal amplitudes in the difference NMR spectra (realized by manual adjustment of the relative multiplication factors; cf. experimental section).

Difference NMR spectra, derived from poorly resolved spectra of complex unknowns should be regarded as rather qualitative guidance tools. Hence, they emphasize UDOM constituents with variable concentration (like HPS and CRAM, respectively) and tend to attenuate or even ignore constituents with a more uniform distribution (like, e.g. proteins/peptides; McCarthy et al., 1996; Hedges et al., 2001). In addition, the S/N-ratio of difference spectra is intrinsically rather poor, and noisy sections are insufficiently defined. More importantly, global weighting factors employed across the entire spectral bandwidth do not recognize the specific distribution of NMR resonances of the individual molecular NOM constituents. Regarding the level of amino acid composition and protein content, surface and deep UDOM appear remarkably similar, if the amino acid composition of hydrolyzates (and its amino sugar content/composition), their C/N ratio, and NMR properties are considered.

Data reduction schemes decrease the dimensionality of spectral data to a more readily interpretable form, and they are also useful to illustrate the compositional relationships of a marine surface and deep UDOM with respect to its two major non-protein constituents, namely CRAM and HPS. The $^{13}$C NMR spectra of UDOM and its main constituents were at first decomposed into 47 equidistant integral segments with 5 ppm bandwidth, ranging from 0-235 ppm; the resulting histograms are in essence very low resolution NMR spectra (Figure 77).

From these data-reduced $^{13}$C-NMR spectra (Figure 77), a very conservative estimate for the minimum content of CRAM in UDOM was computed, by use of a basic 3-component reverse mixing model (Nelson and Baldock, 2005) "UDOM = $n_1$ • HPS + $n_2$ • CRAM + $n_3$ •

peptides", based on the assumption that all nitrogen in UDOM is proteinaceous. The resulting histogram $^{13}C$ NMR spectra of a 3-component surface UDOM ($n_1 = 0.601$, $n_2 = 0.229$, $n_3 = 0.236$) and a 3-component deep UDOM ($n_1 = 0.271$, $n_2 = 0.509$, $n_3 = 0.217$) demonstrated, that CRAM is a major constituent of both surface and deep UDOM. Improved accordance of experimental and computed, data reduced, $^{13}C$ NMR spectra of surface and deep UDOM was obtained at smaller values of $n_3$, suggesting (plausibly) that a sizable fraction of nitrogen in UDOM does not occur in the form of proteins / peptides.

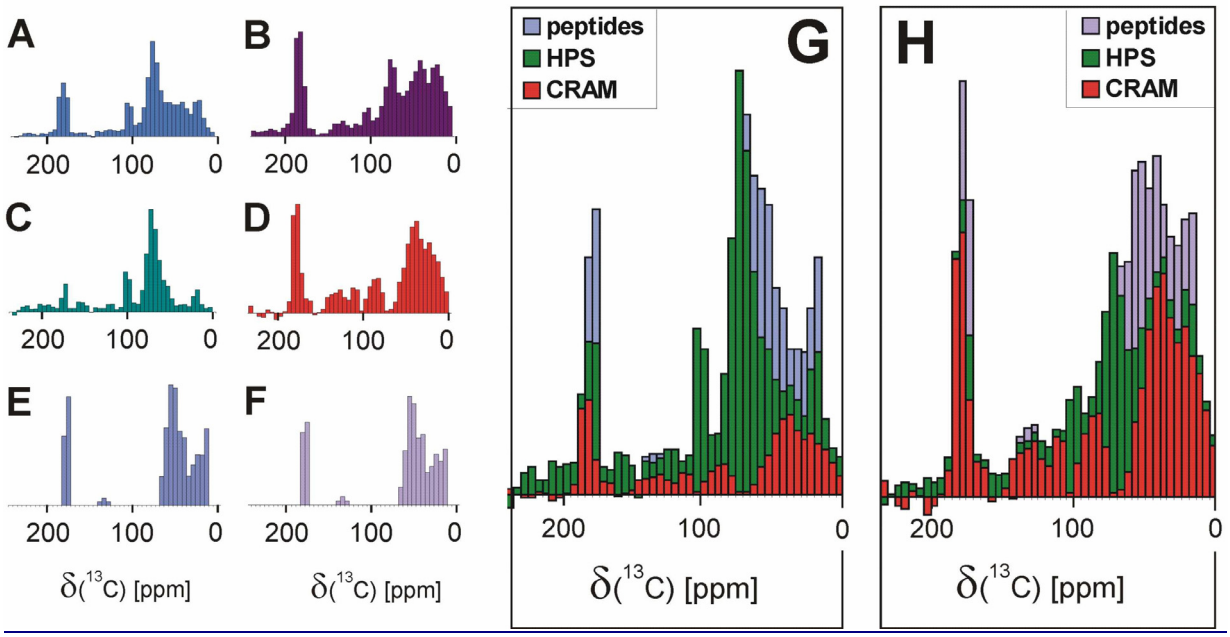



**Figure 77.** data reduced $^{13}C$ NMR spectra (47 buckets of integrated $^{13}C$ NMR intensity with 5 ppm width each, ranging from 0-230 ppm) of (A) surface and (B) deep UDOM, as computed from the experimental NMR spectra, and of the main UDOM constituents (C) HPS and (D) CRAM, as computed from the difference NMR spectra (cf. Figure 75). The histogram spectra of the proteins in surface UDOM (G) and deep UDOM (H) were computed from chemical shifts of a protein with the composition of a surface (E) and deep (F) marine NOM hydrolysate via ACD software (cf. experimental).
From these data-reduced $^{13}C$-NMR spectra, a very conservative estimate for the minimum content of CRAM in UDOM was computed, by use of a basic 3-component reverse mixing model "UDOM = $n_1$•HPS + $n_2$•CRAM + $n_3$•peptides", based on the assumption that all nitrogen in UDOM is proteinaceous, resulting in a (G) data-reduced $^{13}C$ NMR spectrum of a 3-component surface UDOM ($n_1 = 0.601$, $n_2 = 0.229$, $n_3 = 0.236$) and a (H) data-reduced $^{13}C$ NMR spectrum of a 3-component deep UDOM ( $n_1 = 0.271$, $n_2 = 0.509$, $n_3 = 0.217$). Improved accordance of experimental and computed, data reduced, $^{13}C$ NMR spectra of surface and deep UDOM is obtained at smaller values of $n_3$, suggesting (plausibly) that a sizable fraction of nitrogen in UDOM does not occur in the form of proteins / peptides.

*8.3.1.2. key NMR data to establish the existence of CRAM in UDOM*

Two key observations were critical for establishing the existence of CRAM in UDOM. The first observation was based on the nitrogen content of UDOM relative to the quantity of the carbonyl derivative (i.e. carboxylic acid, ester, amide) resonance $F_C$. The C/N atomic ratios of surface and deep UDOM were 16.1 and 18.4, respectively. Assuming all nitrogen is amide, the N/C atomic ratio defines the fraction of the total $^{13}C$ NMR integral in amide form. Calculations indicated 6.2% and 5.4% of the total $^{13}C$ NMR resonances of surface and deep UDOM were represented by amides; i.e. a C/N atomic ratio of (16.1 s_UDOM/18.4 d_UDOM) corresponds to a percent N/C ratio of 6.4 for s_UDOM and 5.7 for d_UDOM, respectively (cf. Figure 75). This is considerable less carbon NMR resonance integral than found in $F2_C$ of s_UDOM (17.7 %) and d-UDOM (21.4 %), respectively. Amide only accounted for about one third of the $F2_C$ resonance in surface UDOM and one fourth of it in deep UDOM. Consequently, a substantial excess of carboxyl (ester or acid) resonance was present in the $^{13}C$ NMR spectra of both surface and deep UDOM.

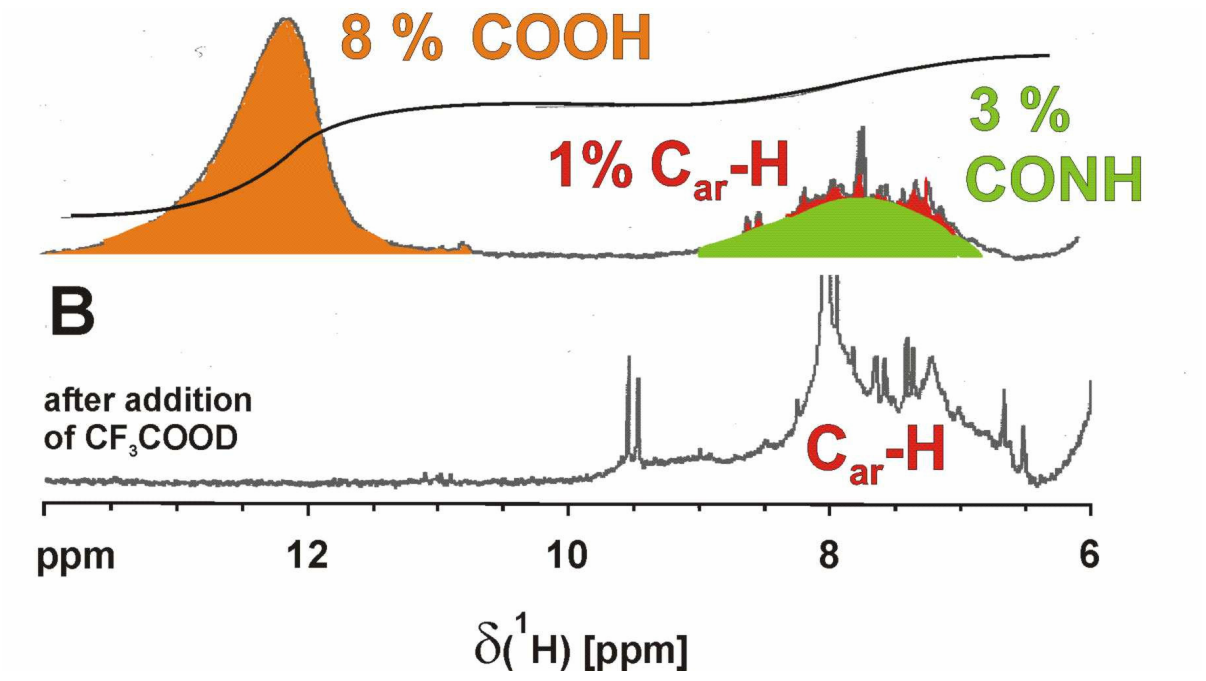



**Figure 78.** The chemical exchange of labile protons in deep UDOM is slowed down sufficiently in DMSO-$d_6$ under total exclusion of moisture to allow the observation (**A**) of both exchangeable [carboxyl: 8 % of total $^1H$ NMR integral (orange); peptide: 3 % of total $^1H$ NMR integral (green)] and non exchangeable [$C_{ar}$–H: 1 % of total $^1H$ NMR integral (red)] protons. Consecutive addition of $CF_3COOD$ combines all labile proton NMR resonances into a single far downfield signal (**B**; here: $\delta$ ($^1H$) = 14.7 ppm). The chemical shift of non-exchangeable aromatic protons in deep UDOM is affected by pH variation, which causes alterations in charge, conformation and relative orientation of UDOM constituent molecules.

The second key observation came from $^{1}H$ NMR spectra of deep UDOM acquired under conditions of slow chemical exchange in dry DMSO-$d_6$ under total exclusion of moisture, using vacuum-line techniques. Here both non-exchangeable (typically protons attached to carbon atoms) and exchangeable protons (typically protons attached to heteroatoms, like O, N, S) produce recognizable NMR resonances. Under these conditions, NMR peaks of exchangeable protons show appreciable line-broadening (10-200 Hz). Accordingly, the direct observation of a sizeable carboxylic acid resonance (Figure 78) demonstrated the presence of carboxylic acids rather than esters in deep UDOM.

#### *8.3.1.3 Comparative analysis of one- and two-dimensional NMR spectra of UDOM*

Cross peaks in two-dimensional heteronuclear single quantum coherence (HSQC) NMR spectra relate directly bound carbon and hydrogen atoms. In analogy to procedures established in 1D NMR spectra of NOM and UDOM (Figure 75), chemical shift ranges (areas) corresponding to fundamental UDOM substructures can be also defined in HSQC NMR spectra (Hertkorn et al., 2002); here, seven areas of chemical shift were used to define major UDOM substructures (Figure 79). In order of increasing chemical shift, area 1 defined methyl bound to carbon and sulfur (dotted circle), and in its lower left corner, branched purely aliphatic CH pairs and polymethylene ($\delta_{H/C}$ = 1.2/29 ppm). Area 2 represented a complex set of methylene and methine cross peaks without direct bonds to heteroatoms (except purely aliphatic amines, which are considered as rather elusive constituents of UDOM). Low intensity methoxyl cross peaks occur in area 3, and cross peaks representative mainly of CH-$\alpha$ in proteins and vicinal dicarboxylic acids occurred in area 4. Very highly superimposed carbohydrate methylene cross peaks in area 5, carbohydrate methine cross peaks in area 6 and better resolved, but still strongly overlapping anomeric CH cross peaks in area 7 all indicated the complex structure of HPS.

This analysis is corroborated by one and two dimensional multiplicity edited $^{13}C$ NMR spectra of deep UDOM, in which the superposition of the $^{13}C$-DEPT NMR spectra provided a ratio of methyl:methylene:methine:quaternary C of 7:20:46:27 (Figure 80). The large chemical shift ranges of methyl, methylene and methine in deep UDOM provided further indications of the complexity of CRAM.

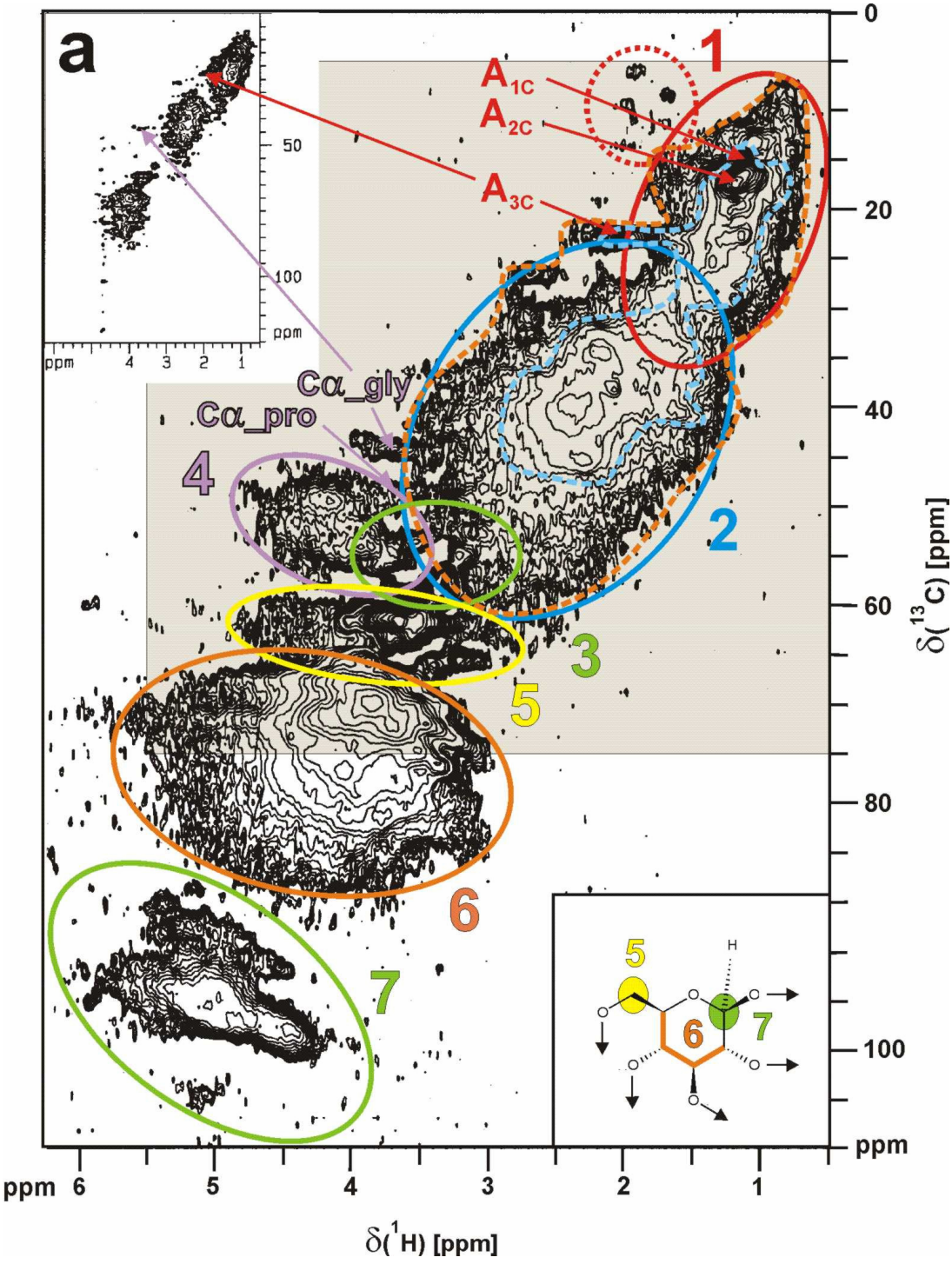



**Figure 79.** Two dimensional NMR spectra of UDOM: $^{1}H$, $^{13}C$ HSQC NMR spectra of surface UDOM with seven groups of major constituents (cf. text) and of deep UDOM (insert **a**) top left): the relative cross peak amplitudes within surface- and deep UDOM HSQC NMR spectra reflect both the variable relative fractions and NMR relaxation characteristics of UDOM constituents. A faster transverse relaxation of CRAM as compared to HPS is observed as well as a faster transverse relaxation of the HPS and peptide components in deep UDOM compared to that in surface UDOM (e. g.: disappearance of section 4 and 7 cross peaks in deep UDOM). That indicates a restricted flexibility and possibly an enhanced contribution of cross linkage at depth as observed with recent mass spectrometric results (Boon et al., 1998). The grey shaded section corresponds to the spectral range shown in figure 82.

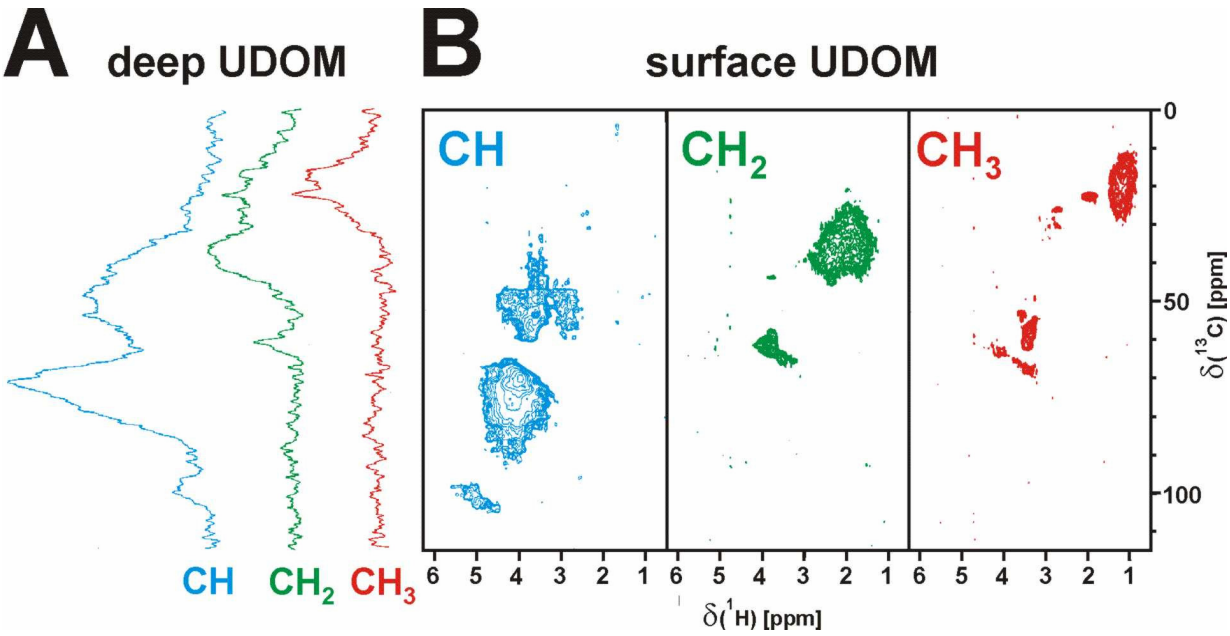



**Figure 80.** Scaled multiplicity edited (**A**) one dimensional (DEPT) $^{13}$C NMR subspectra of deep UDOM and (**B**) two dimensional (DEPT-HSQC) NMR spectra of surface UDOM (exhibit slower transverse relaxation): methyl (red), methylene (green) and methine (blue).

The $^{13}$C NMR chemical shifts are sensitive to aliphatic branching in a predictable way. Methylation in α- and β-positions induces a downfield $^{13}$C NMR chemical shift of ~8 ppm each, while γ-methylation induces an upfield chemical shift of ~4 ppm (Kalinowski et al., 1984). δ- and more remote alkylation has marginal effects and can be neglected (Kalinowski et al., 1984). In contrast, the effects of aliphatic branching on proton chemical shifts are considerably less pronounced and less regular; in general, steric crowding in aliphatic compounds induces downfield proton chemical shift. A substantial degree of branching can be deduced from the considerable range (>25 ppm) of methyl chemical shifts in the edited one dimensional $^{13}$C-NMR spectrum of deep marine UDOM (Figure 80). A similar reasoning applies to methylene and methine chemical environments in both one and two dimensional NMR spectra. The near Gaussian distribution of the $^{13}$C NMR chemical shifts within the HSQC cross peak area 2 (Figure 79) indicated broad substitution patterns in CRAM.

Another indication of the high variability of chemical environments in CRAM was the considerable range of chemical shift (~20 ppm) and the near Gaussian distribution of intensity of the carboxyl resonance $F2_C$ in one-dimensional $^{13}$C NMR spectra of UDOM (Figure 75). Extensively α- and β-methylated chemical environments in aliphatic carboxylic acids were

not common in CRAM, as indicated by the absence of extreme $^{13}C$ NMR downfield HSQC cross peaks which would fall outside the boundaries observed for HSQC area 2 in Figure 79. It appeared that moderate branching and a fairly even distribution of carboxyl groups most accurately described the chemical environment in CRAM.

The positions and amplitudes of HSQC cross peaks allowed discrimination between peptide and aliphatic polycarboxylic acid chemical environments, both of which contributed to the considerable expansion of HSQC area 2 and methylene cross peaks (Figures 79, 80, 82A). The HSQC cross peaks of peptides occupied only the $^{13}C$-NMR upfield part of area 2 HSQC cross peaks (Figures 79, 82B; Hertkorn et al., 2002). Therefore another material was required to complement the remaining downfield $^{13}C$-NMR area 2 HSQC cross peaks in UDOM. This material was consistent with an aliphatic, polycarboxylic-acid structure, such as CRAM.

*8.3.1.4. Assessment of the proton and carbon NMR chemical shift space of a computer-generated aliphatic model polycarboxylic acid $C_{644}H_{1142}O_{146}$ , a necessary prerequisite to evaluate the chemical environments of CRAM within UDOM.*

A trustworthy resonance assignment in NMR spectra of complex unknowns typically requires a comparative assessment of experimental and calculated NMR data, based on judiciously selected model structures (nowadays, proton and carbon NMR spectra of extended spin systems can be computed, based on empirical correlations, with reasonable accuracy on desktop computers). An iterative analysis of model and experimental NMR spectra, corroborated by the current understanding of NOM bio- and geosynthetic pathways, allows to propose model structures conforming to spectral data and to discriminate among several proposed NOM substructures (cf. Figures 42, 86). Although this initial procedure requires verification by independent and complementary analytical data, it nevertheless provides very significant structural detail at atomic resolution, which is not available by any other analytical technique at present.

The very low resolution NMR signature of CRAM, which covers a substantial chemical shift range in both $^{1}H$ and $^{13}C$ NMR spectra (Figures 75, 79, 80), already indicates a fairly complex aliphatic polycarboxylic acid structure with considerable intrinsic heterogeneity. For the evaluation of $^{1}H$ and $^{13}C$ NMR properties of CRAM, a model open chain aliphatic polycarboxylic acid $C_{644}H_{1142}O_{146}$ with no unsaturation other than derived from carboxylic

groups was computer-generated in a way to allow for systematic increase in aliphatic branching as well as increase in the spatial separation of carboxylic groups (Figure 81).

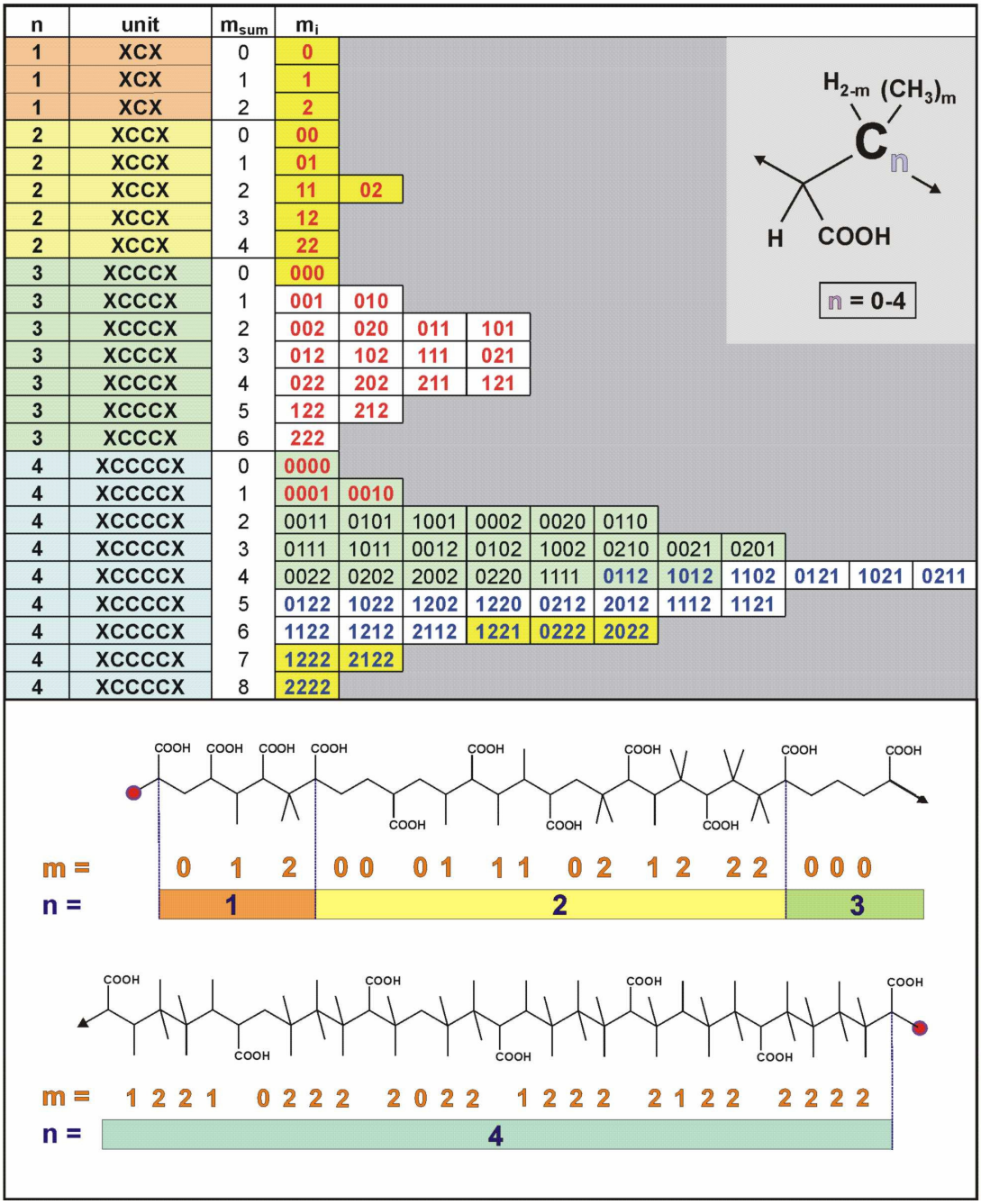



**Figure 81**. model open chain aliphatic polycarboxylic acid $C_{644}H_{1142}O_{146}$, with a variable spacing between H-C-COOH groups, $XC_nX$, X = H-C-COOH, n = 1-4 (left column) and variable patterns of methylation ($m_i$ = 0,1,2; $m_{sum}$ equals sum of methyl groups between adjacent X).
$^{1}H$ and $^{13}C$ chemical shifts were computed from three overlapping sections (red: $C_{214}H_{366}O_{62}$, green: $C_{212}H_{374}O_{50}$, blue: $C_{262}H_{476}O_{48}$), since the entire molecule was too large for direct calculation. The two terminal sections of the molecule, which have been highlighted in yellow, are shown in the bottom part for better visualization and understanding.

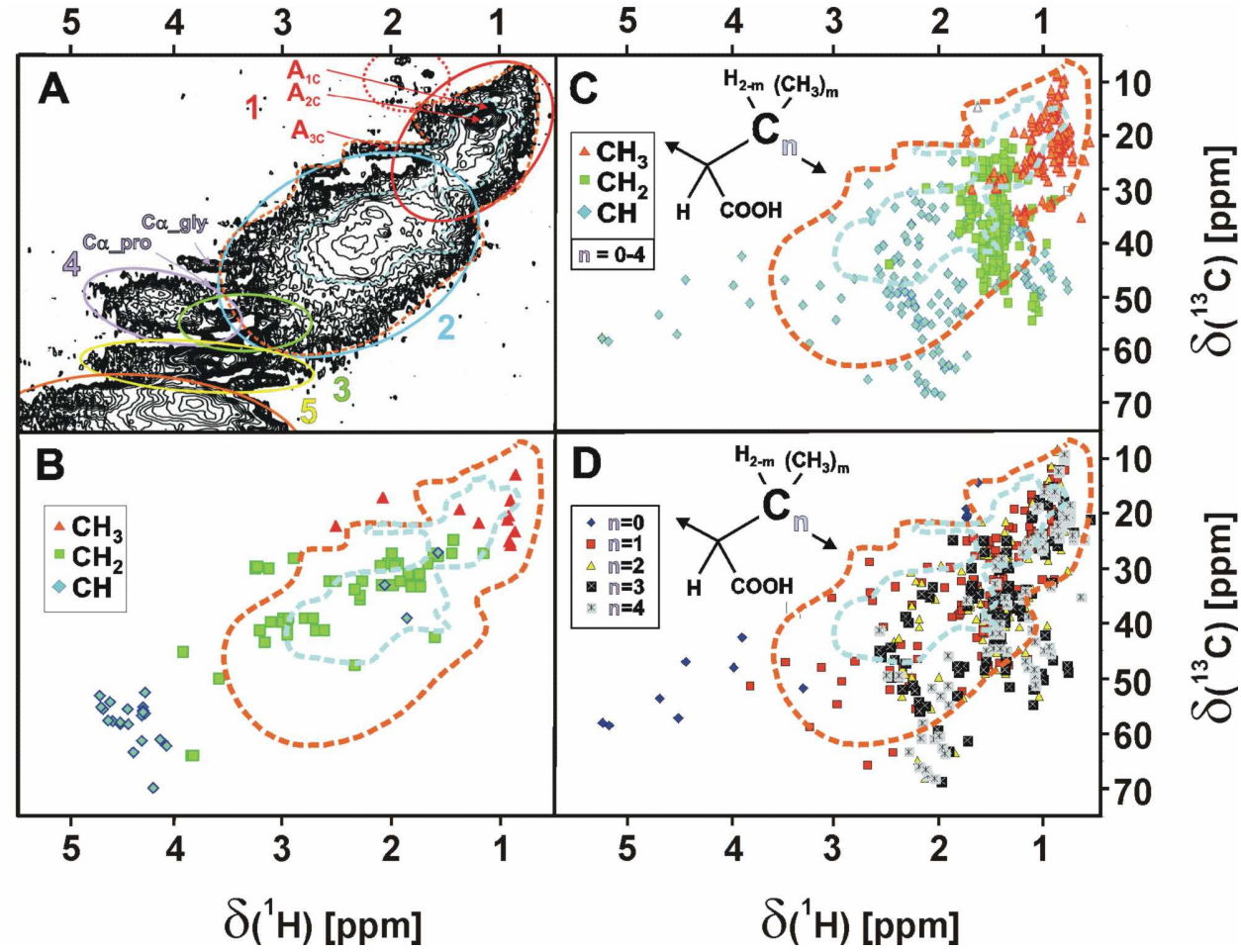



**Figure 82.** (A) Upfield section of the $^{1}H,^{13}C$ HSQC NMR spectrum of surface UDOM (partial plot of Figure 79) (B)*: multiplicity edited chemical shifts of $^{1}H,^{13}C$ HSQC cross peaks of all 20 proteinaceous amino acids in peptides, following alanine (Hertkorn et al., 2002); (C)*: HSQC cross peaks of a randomly branched model aliphatic polycarboxylic acid $C_{644}H_{1142}O_{146}$ ({-[$(CH_3)_l$-C-COOH]-$C_n(CH_3)_m$-} with l = 0, 1; m = 0-2, and n = 0-4, cf. text below), displayed according to (C) carbon multiplicity and (D) according to the separation of carboxylic substituents. (* contour lines indicate aliphatic HSQC cross peaks of surface UDOM at unit and at five-fold intensity; Figure 79).

That artificial model molecule $C_{644}H_{1142}O_{146}$ represents a structure and chemical shift space for open chain branched polycarboxylic acids and serves to evaluate NMR features of UDOM, which are attributed to CRAM (cf. Figure 81). Full methylation (m = 2) leads to the disappearance of HSQC cross peaks for the respective chain positions, but induces strong deshielding on $\delta(^{13}C)$ and minor effects on $\delta(^{1}H)$ to neighboring positions up to three bonds away. Within a NMR analysis at this level of connectivity, the absence of model acid HSQC cross peaks in the HSQC NMR spectra of s_UDOM strongly indicates non-occurrence or non-significance of respective model acid partial structures in the NOM material. Accordingly, multiply methylated α- and β-carbon atoms (representing extensively branched

non functionalized aliphatics) with characteristic downfield carbon chemical shift and highly shielded protons, can be discarded as prominent constituents of CRAM (cf. Figure 82). Extensive branching is also not consistent with the low abundance of purely aliphatic (i.e. non carboxylated) carbon atoms in CRAM and with the limited occurrence of methyl groups (Figure 80).

Instead, carboxylic groups (and possibly less abundant auxiliary functional groups causing downfield proton and carbon NMR chemical shift, respectively) appear typically not more than 2-3 positions distant from aliphatic carbon (i.e. in α-,β-,γ- positions, respectively). A near random occurrence of aliphatic branching in CRAM accounts most appropriately for the observed ~ 25 ppm carbon chemical shift range within the methylene trace of the multiplicity-edited 1D and 2D NMR spectra (Figure 80) and a similarly large spread of δ ($^{13}$C) across the entire area 2 HSQC cross peaks in s_- and d_UDOM; respectively. The relatively minor content of methyl carbon in deep UDOM (7%, Figure 80) relative to the substantial fractions of methylene and methine carbon (20% and 40%, respectively) convincingly suggests the presence of cyclic structures already at this level of NMR analysis (alicyclic ring topology causes removal of methyl groups compared with open chain analogues). However, the overall complexity of CRAM chemical structures, and the presence of other key structures in UDOM (peptides and carbohydrates in particular) precludes overly quantitative statements at this stage of NMR analysis. However, peptides are in average characterized by a rather large proportion of methyl groups; according to the composition of a marine UDOM hydrolysate, the ratio of Cq/CH/$CH_2$/$CH_3$ in the peptide fraction is 30/32/22/16 in surface UDOM, and 30/33/23/14 in deep UDOM, respectively.

### *8.3.2. Capillary electrophoretic separation of UDOM*

NMR provides quantitative information about the composition of CRAM, indicating the near absence of double bond and aromatic hydrogen atoms. However, it cannot unequivocally discriminate between alicyclic and open chain aliphatic environments (cf. Figure 82). The NMR-derived structural information about short range molecular order in CRAM is complemented by electrospray ionization (ESI) mass spectrometry data, which provide ultra-high resolution and mass accuracy, leading to the assignment of molecular formulae for hundreds to thousands of UDOM constituent molecules. However, the variable ionization efficiency of UDOM molecules, which is a crucial and structure-dependent parameter in the generation of

UDOM ions in a mass spectrometer, has the potential to severely bias mass spectra-derived data about UDOM composition and structure towards efficiently ionisable UDOM constituents.

Here, highly complementary data from capillary electrophoretic separation of UDOM offer genuine options to validate mass spectral data of UDOM ion composition and structure against NMR derived molecular-level structural information. The detection of a molecule after chromatographic or electrophoretic separation offers intrinsic structural information, since the electrophoretic mobility as well as the chromatographic retention time depend on molecule properties (which are exploited to effect separation).

The surface and deep UDOM samples showed structure-specific bimodal capillary zone electrophoretic (CZE) mobility profiles in both CZE/UV hyphenation with short-wave length (214 nm) detection, which is considered as rather unselective to structural details, and in CZE/ESI-MS with mass spectrometric detection (Figure 83). A larger fraction of highly charged aliphatic molecules was observed in deep UDOM (Schmitt-Kopplin and Junkers, 2003; Schmitt-Kopplin and Kettrup, 2003). The effective mobilities for deep UDOM fell in the range observed for aliphatic polycarboxylic acids and were consistent with its higher content of CRAM, as indicated by NMR data. The close similarities of the electropherograms obtained at pH 9.4 and pH 11.4 confirmed the near absence of phenols in UDOM (cf. Figure 76) and other weakly acidic compounds in both surface and deep UDOM (data not shown). The congruence of electrophoretic mobility profiles in non-selective CZE/UV and mass selective CZE/MS hyphenation indicated only slight variance of electrospray ionization efficiencies within UDOM constituents (Figure 83). Low mobility constituents, like carbohydrates, showed somewhat lower ionization efficiencies than carboxylic acids, as indicated by the lesser relative intensity of these constituents in surface UDOM compared with deep UDOM. Nevertheless, owing to the overall correspondence of CZE/UV and CZE/MS electropherograms, the mass spectra obtained from marine UDOM were considered representative of UDOM composition.

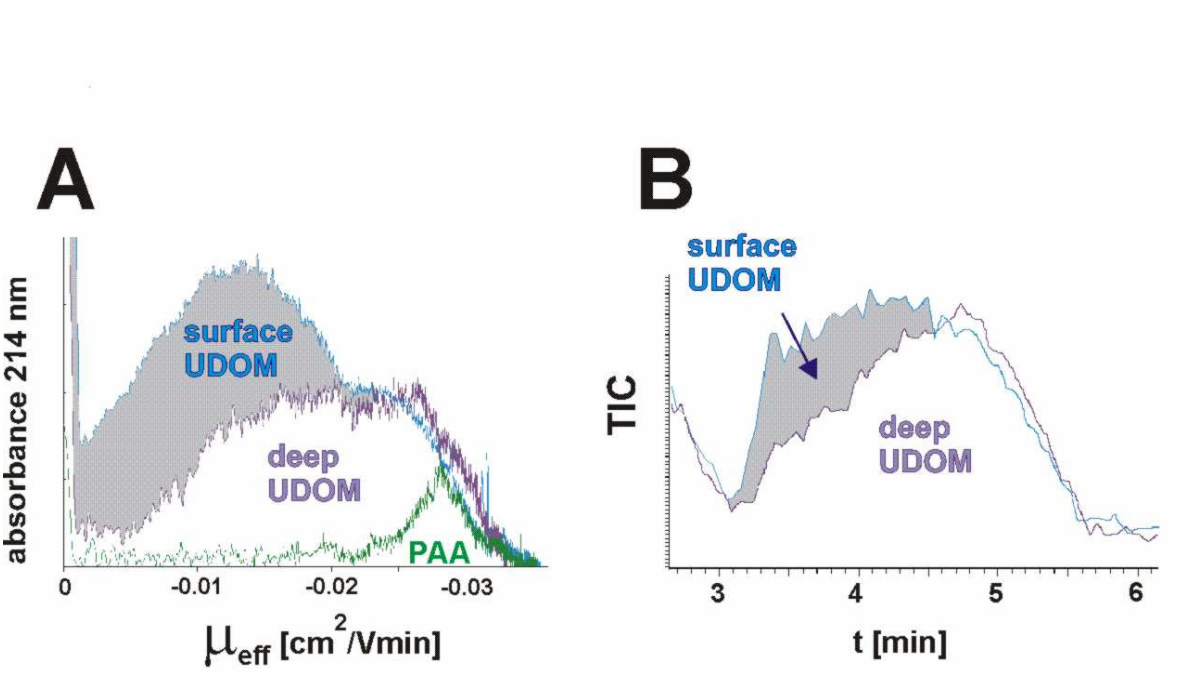



**Figure 83.** Electropherograms of surface and deep marine UDOM: (**A)** Capillary zone electropherogram (CZE; UV detection at 214 nm; effective mobility scale in $cm^2\ V^{-1}\ min^{-1}$) of surface and deep UDOM and a polyacrylic acid standard (2 kDa); **(B)** superimposed CE-ESI/MS electropherograms of surface and deep UDOM (migration time in min). The congruence of (**A**) and (**B**) indicates near uniform ionization efficiency of NOM constituents in marine UDOM.

### *8.3.3. FTICR mass spectrometry of UDOM*

Electrospray ionization (ESI) mass spectra of UDOM indicated, depending on resolution, hundreds to thousands of individual signals and conspicuous series of mass spacing patterns separated by 2 and 14 Da (Figure 84). These patterns are most pronounced in deep UDOM at negative ion ESI with both ion trap (data not shown) and Fourier transformation ion cyclotron resonance (FTICR) mass spectrometry systems. Based on accurate mass measurements obtained from ultrahigh-resolution FTICR mass spectra, the most prominent recurring mass differences between these ions were defined as 14.0156 Da [variation in methylene count $(CH_2)_n$], 2.0157 Da (variation in double bond equivalents/$H_2$), 1.0034 Da (mass difference between $^{13}C$ and $^{12}C$), and 0.0364 Da (formal exchange of $CH_4$ versus oxygen) (Stenson et al., 2003). A representative subset of the most intense peaks, which all were observable as prominent members of these mass spacing patterns in both positive and negative ionization modi, was attributed to CRAM, based on the finding (cf. above), that prominent NMR-detectable features of CRAM would produce strong mass spectroscopic signatures as well. For an in-depth analysis; 613 positive ions $XH^+$ were selected and the molecular formulae X were

determined. These were found to be most accurately represented by the composition $C_nH_mO_q$, indicating the absence of nitrogen in CRAM and a substantial degree of unsaturation in excess of carboxyl groups. The $CH_2$-specific Kendrick mass defect analysis (Hughey et al., 2001; Hsu et al., 1992) of CRAM showed classes of compounds varying only in the number of methylene groups and, almost exclusively, relative Kendrick mass defects (Δm) corresponding to variance in unsaturation (double bond equivalents (DBE), Δm = -0.013 Da) and exchange of oxygen vs. $CH_4$ (Δm = -0.0364 Da; Figure 84).

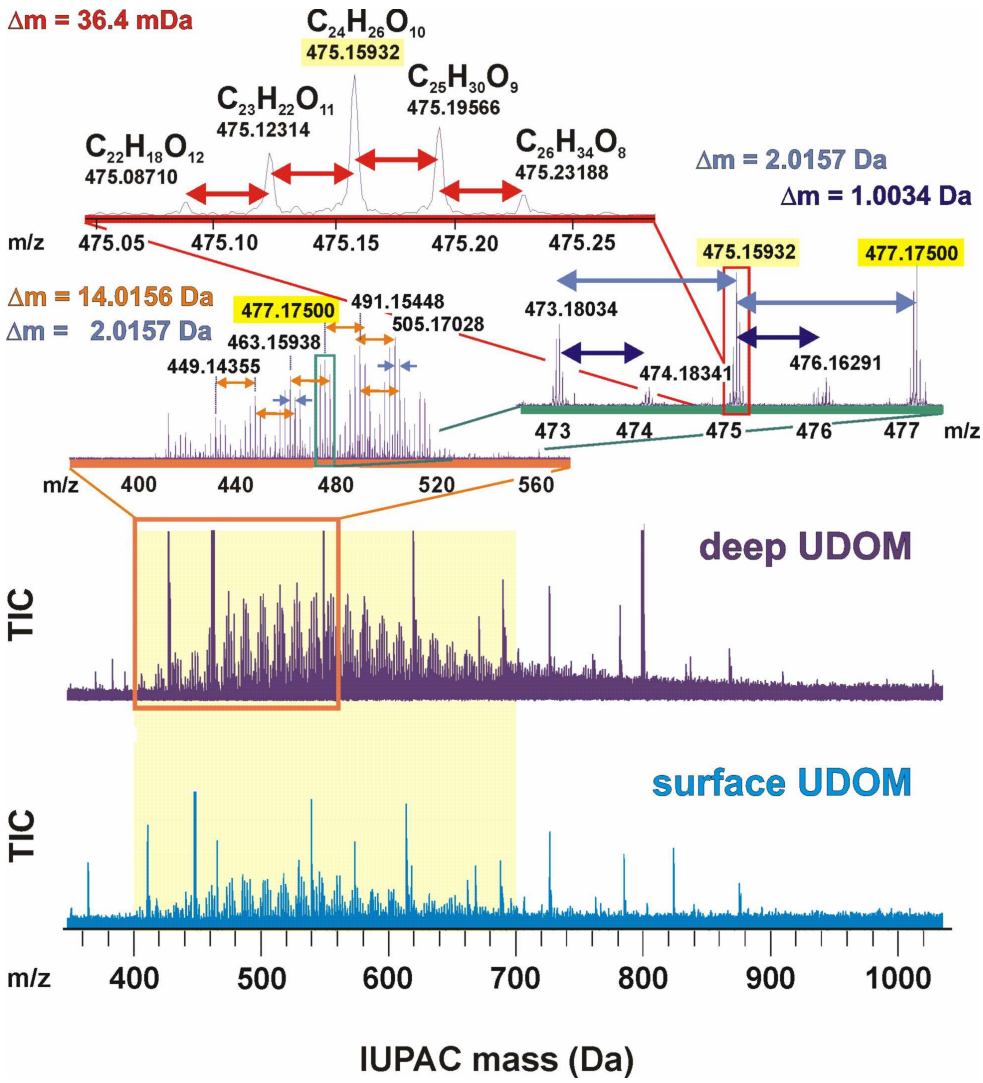



**Figure 84.** ESI positive ion 7 Tesla FTICR mass spectra of (*bottom, left*) surface and (*middle left*) deep UDOM with selected expanded sections (*top left*), given to demonstrate the very high resolution in these spectra; the most significant recurring mass differences (Δm = 14.0156 Da, 2.0157 Da, 1.0034 Da, 36.4 mDa) have been explained in the text. In the uppermost expansion, molecular formulae $C_nH_mO_q$ are given for prominent peaks. The yellow shaded mass section equals the mass range used for the Kendrick mass analysis.

In a van Krevelen diagram (Kim et al., 2003; Visser, 1983), the area of CRAM-derived elemental compositions was centered along the line connecting the virtual endmembers $CH_2$ and CO, respectively, and was clearly distanced from the region of peptides (Figure 85). CRAM exhibited an O/C ratio in excess of 0.25 and was devoid of extended aromatic systems and carboxylated, saturated hydrocarbons. Lower-molecular-weight CRAM exhibited a higher degree of oxidation than higher-molecular-weight molecules (Figure 85), suggesting a size-reactivity continuum similar to that observed previously in marine DOM (Benner, 2002; Amon and Benner, 1996). All molecular formulas derived from these ions exhibited a higher degree of unsaturation than could be introduced by carboxyl groups alone, even if all oxygen was assumed in carboxyl groups (cf. purple line in Figure 85). NMR data confirm the very low abundance of hydrogen in aromatics and double bonds in both surface and deep UDOM (< 1.5% of total $^1$H NMR integral; Figures 75, 76).

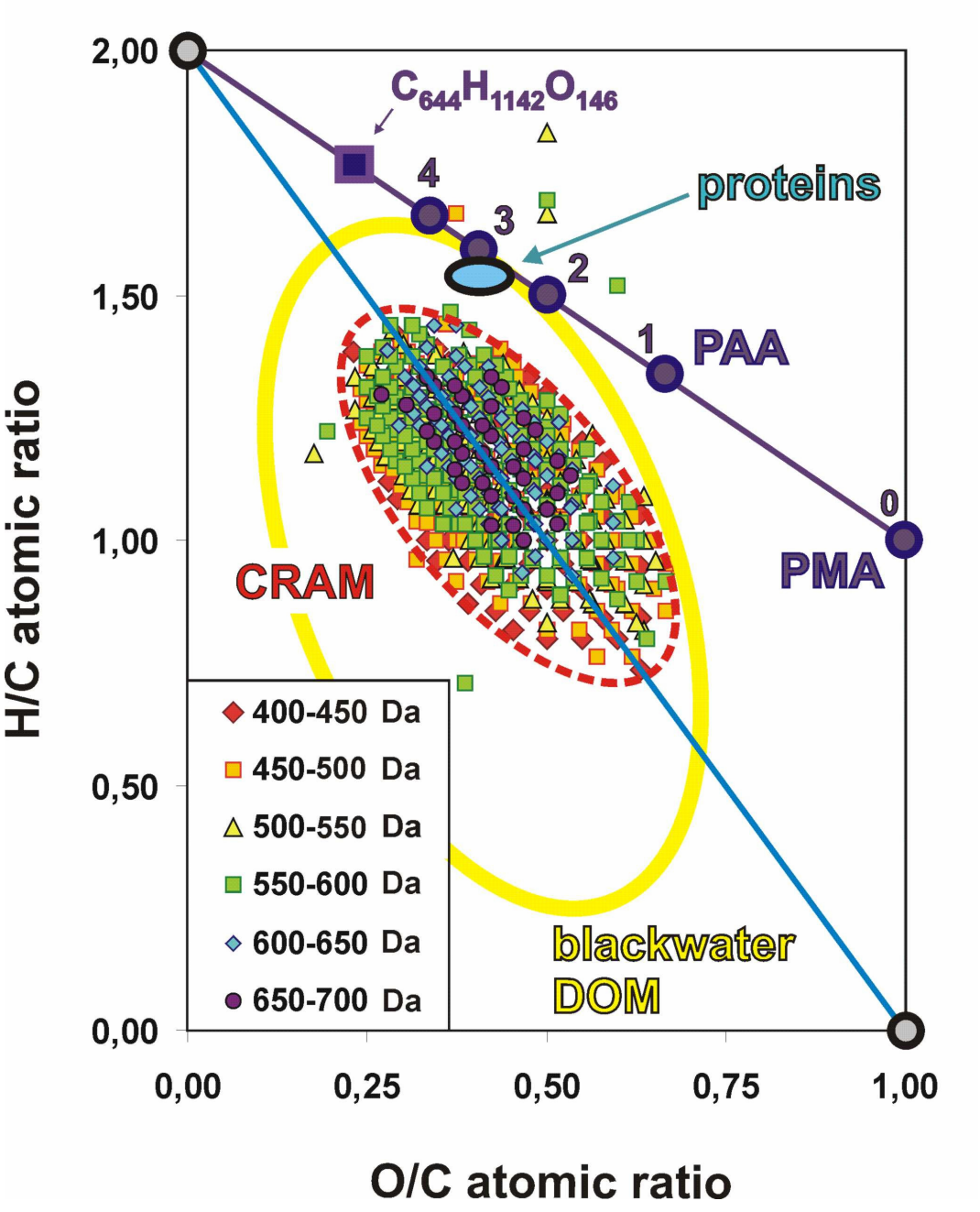



**Figure 85.** Mass sorted van Krevelen diagram of 613 ions attributed to CRAM (red ellipsoid, broken line) with region of proteins indicated in blue. The purple line refers to *any* fully saturated open chain aliphatic carboxylic acid, with aliphathic polycarboxylic acids indicated according to spacing between carboxyl groups (cf. Figure 81; e.g. polymaleic acid (PMA; n = 0) polyacrylic acid (PAA; n = 1), higher polycarboxylic acids (n = 2-4) and model aliphatic polycarboxylic acid $C_{644}H_{1142}O_{146}$ (Figure 81). The proportion of unsaturation and oxidation in CRAM increases slightly with decreasing mass of the molecular ions. The total area occupied by the 50 Dalton mass increments shown decreases with increasing mass, indicating an increased diversity of molecular compositions at lower masses. The blue line connects the virtual endmembers $CH_2$ and CO and bisects the compositional space of CRAM nearly in half (Figure 89). That positioning of CRAM within the van Krevelen diagram is the consequence of a complex relationship between the number of feasible isomers, which can be assembled from molecular formulae $C_nH_mO_q$, the occurrence of these isomers in UDOM, and the ionization efficiency of any single isomer under the given experimental conditions (Hertkorn, 2006; Figure 89). CRAM occupies a minor section of elemental compositions as compared to a blackwater DOM (Hughey, 2001).

### *8.3.3.1. $CH_2$ and $CO_2$ based Kendrick mass analysis of CRAM*

Kendrick mass defect analyses of complex mixtures are not restricted to $CH_2$ units, but are also applicable to any other fragment. Necessarily, as imposed by the large amplitude of the $F2_C$ resonance in $^{13}C$ NMR spectra of UDOM and CRAM (which has been established to represent mainly carboxylic acids, cf. NMR discussion, Figures 75, 76), $CO_2$ has to be another key building block of CRAM.

It is therefore reasonable to expand the Kendrick mass defect sorting scheme (Figurere 86) to both key fragments of CRAM, namely $CH_2$ and $CO_2$. Since the IUPAC mass scale is based on $^{12}C$ (IUPAC mass: 12.00000), the presence of hydrogen ($^1H$) induces positive mass defects and the presence of $^{16}O$ causes negative mass defects in both IUPAC and Kendrick systems, respectively.

CRAM, which belong to identical $CH_2$-based Kendrick mass defect classes, produce linear patterns of $CO_2$-based Kendrick mass defects. Conversely, CRAM with identical $CO_2$-based Kendrick mass defects – which represent then molecules belonging to the same $CO_2$-based class - produce analogous linear patterns in the $CH_2$-based Kendrick mass analysis (Figure 86). However, the significantly larger mass spacing (44 Dalton for $CO_2$ instead of 14 Dalton for $CH_2$) precludes the occurrence of very extended $CO_2$-based series within the observed mass range of CRAM. Nevertheless, the fundamental importance of $CO_2$ as a fundamental building block of CRAM is well established by this $CO_2$-based Kendrick mass analysis, which specifically indicates four series of common $CO_2$-based Kendrick mass defects (KMD_$CO_2$: 0.73303 Da, 0.69663 Da, 0.67774 Da, 0.67219 Da, respectively; cf. Figure 86, and experimental section) with five members (Figure 86) each.

However, while the occurrence of certain molecular backbones with variable degrees of carboxylation is certainly a realistic option for CRAM, it has to be emphasized, that many isomers may contribute to a given molecular composition; these are not resolved by any Kendrick mass sorting scheme.

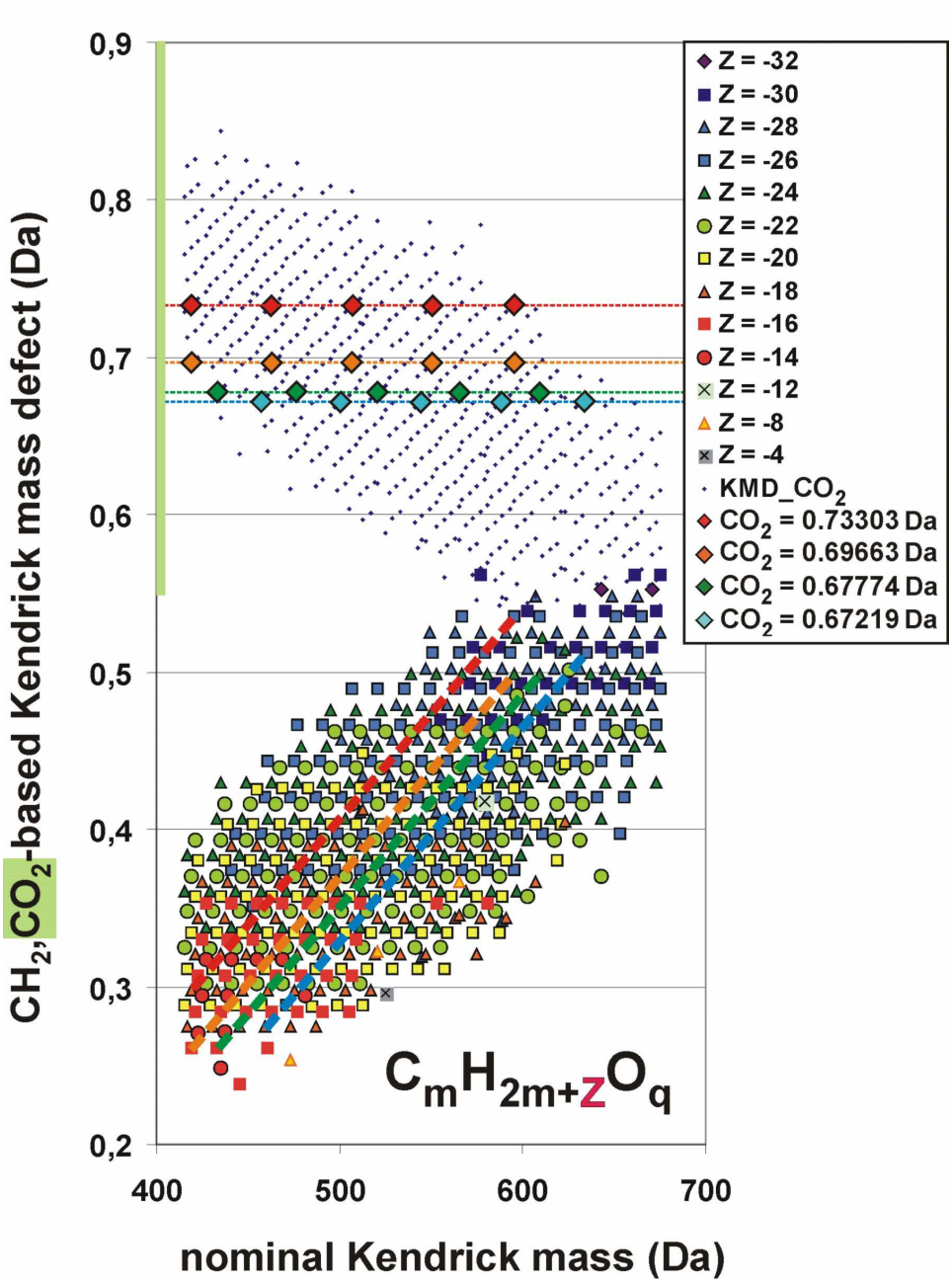



**Figure 86.** Combined $CH_2$-based (bottom) and $CO_2$-based (top) Kendrick mass defect analysis of 613 CRAM ions in deep UDOM in the mass range of 415-675 Da. The $CH_2$-based analysis is displayed according to Z-values of the molecular formulae $C_mH_{2m+Z}O_q$, resulting in 13 series of Z-values (Z = -32 to 4: Z (type): $C_cH_{2c+Z}O_o$) in 156 homologous series $C_cH_hO_o$ (c = 19-37, h = 14-50, o = 6-16, DBE = 3-21).
In the $CO_2$-based Kendrick mass effect analysis, several series of related CRAM with identity in $CO_2$-based Kendrick mass defects are recognizable, each showing a mass difference of the key unit $CO_2$ (44 Dalton). Caused by the larger mass spacing compared with $CH_2$ (14 Dalton), the $CO_2$-based series in CRAM are more truncated than the $CH_2$-based series. CRAM are composed of four five-membered series of molecular compositions $C_nH_mO_q$ with identical $CO_2$-based Kendrick mass defects, indicated by coloured diamonds (cf. text), 27 series with 4 members each, 70 series with 3 members each, and shorter series, respectively, which all are represented by dots.
Molecules with identical $CO_2$-based Kendrick mass defects differ in $CH_2$-based mass defects, and vice versa; this has been indicated by the four pairs of dotted lines, that represent the four $CO_2$-based KMD-series indicated.

The Kendrick mass analysis solely relies on molecular formulae (i.e. it is insensitive to isomeric composition) and accordingly, the Kendrick mass defect analysis is governed by the nature and abundance of different atoms present in a molecule. A considerable regularity of CRAM can be deduced as more than 95 percent of the $CH_2$-based Kendrick mass defect differences are defined by 13.6 mDa (variation in double bond equivalents BDE) and 36.4 mDa (formal exchange of oxygen and methane in the molecular formula) mass differences.

### *8.3.4. NMR properties of CRAM serve to discriminate between (classes of) isomers and allow to propose prominent substructures of CRAM.*

For any given $C_uH_mO_q$ molecular formula, isomers will produce different NMR signatures, which provide extremely useful and indispensable guidance for a molecular-level structural analysis of CRAM. For illustration of the opportunities provided by NMR to assess the occurrence of certain substructures within CRAM, the molecular formula $C_{28}H_{32}O_{13}$ has been selected, which exhibits a H/C and O/C elemental ratio (H/C = 1.15; O/C = 0.42) close to the locus of all 613 CRAM molecules identified (cf. Figure 85). Furthermore, $C_{28}H_{32}O_{13}$ represents the center molecule of the singular seven membered $CH_4$/O isobaric series identified within the 7T FTICR mass spectrum of CRAM.

Analogous considerations as described below would apply for all 613 CRAM molecular compositions identified in CRAM, which have been found to occupy certain compositional limits (DBE/C = 0.30 – 0.68; DBE/H = 0.20 – 0.95; DBE/O = 0.77 – 1.75), that exclude extremely condensed or otherwise non-conceivable structures, which could reasonably generated from biosynthesis, abiotic and biotic degradation and diagenesis from reasonable NOM precursor molecules.

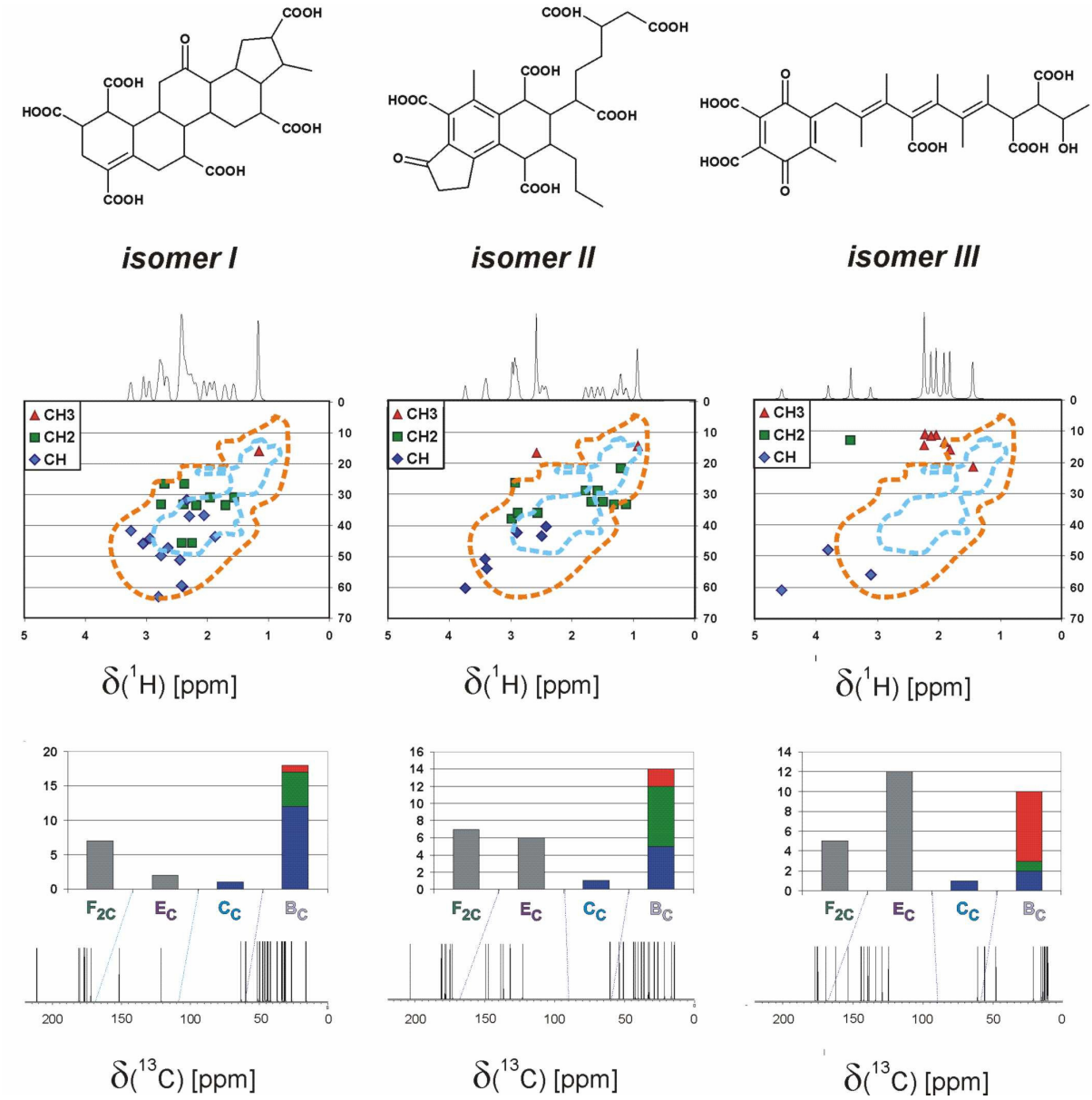



**Figure 87.** Three $C_{28}H_{32}O_{13}$ isomers (top row; IUPAC mass 576.546) together with computed $^{1}H$ NMR (linewidth 10 Hz) and $^{1}H$, $^{13}C$ HSQC-spectra (middle row with HSQC contour lines as given in Figure 79: cross peaks are labelled according to carbon multiplicity) and computed carbon NMR spectra (bottom) with numbers of methyl (red), methylene (green), methane (blue) and quaternary (grey) carbon atoms provided, according to chemical shift ranges given in Figure 75 (cf. text).
*Isomer I* and *isomer II* much better conform to the NMR characteristics of UDOM than *isomer III* [(in one-dimensional $^{1}H$ (Figure 75), non-edited (Figure 75) and multiplicity-edited $^{13}C$ NMR spectra (Figure 80) and in the $^{1}H$, $^{13}C$ HSQC NMR spectra (Figure 79)].

Three molecules $C_{28}H_{32}O_{13}$ with a IUPAC mass of 576.546, which all do not contain aromatic or olefinic protons, a crucial constraint as derived from the proton NMR spectra of surface and deep marine UDOM (Figures 75, 76), serve to illustrate the importance of exhaustive

NMR evaluation in the molecular level analysis of very complex systems such as marine UDOM (or NOM in general).

*Isomer I* represents a terpenoid-derived assortment of condensed alicyclic six- and five-membered rings, dotted with rather evenly distributed carboxylic acid units (to fulfil the restraints defined by Figures 80, 81 and 82). Such a topology corresponds to a minimum occurrence of methyl groups (which terminate any branched alkane unit).

*Isomer II* contains a fully substituted aromatic ring with no hydrogen atom attached; this feature fulfils the key NMR requirement of (virtually) absent aromatic protons in marine UDOM (Figure 76), while allowing for some aromatic carbon NMR integral. The absence of aromatic ring protons would also assist in recalcitrance against biodegradation.

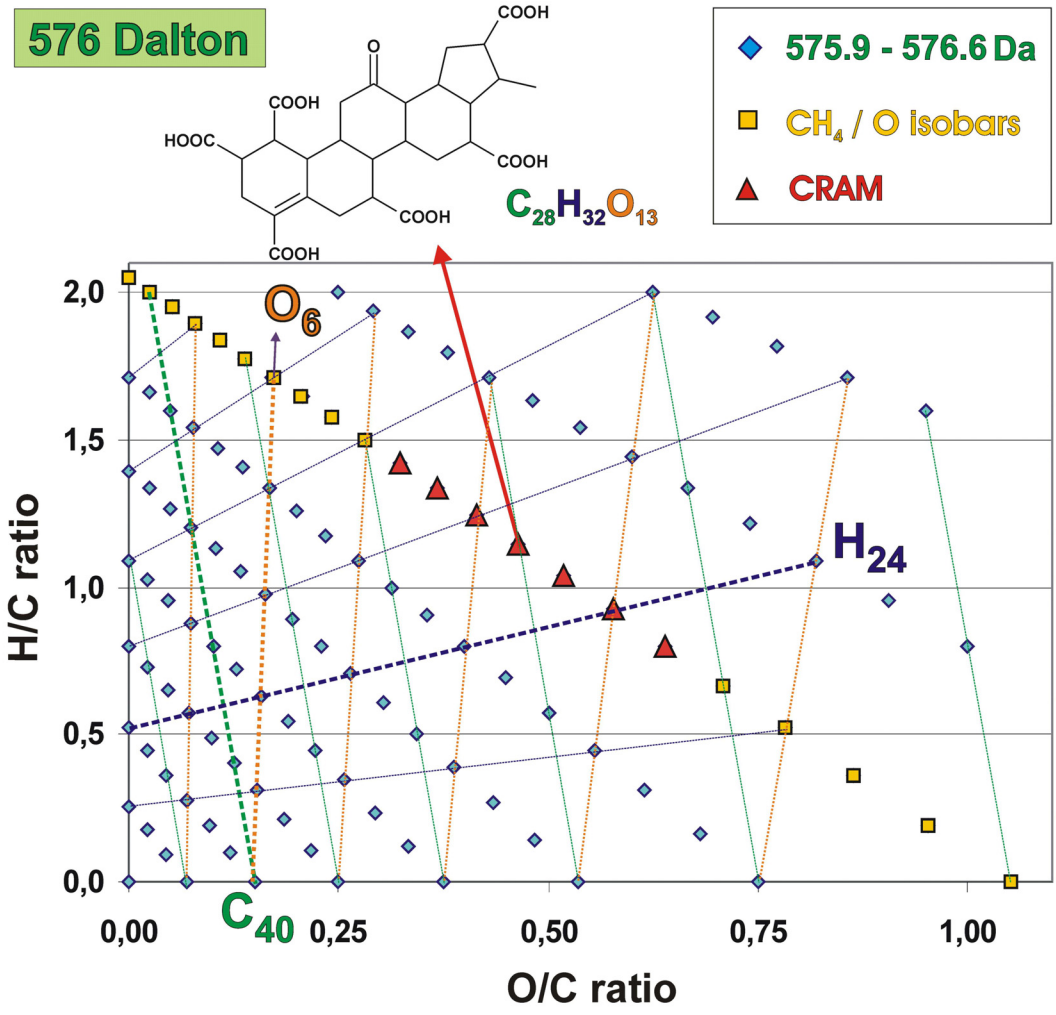
A characteristic NMR feature of *isomer II* is the considerable amount of aromatic carbon NMR resonances (section $E_C$, Figure 75) as opposed to *isomer I*; the fraction of methyl groups can be adjusted by a variable degree of branching more easily than in molecules of type *isomer I* (i.e. isomer II is less defined in its relationship to known classes of biomolecules than both isomers I and III). However, in molecules resembling isomer II, the maximum feasible fraction of methyl carbon still remains rather limited.

*Isomer III*, which is assembled according to partial structures in resemblance to carotenoids, is almost exclusively composed of methyl and quaternary carbon atoms and exhibits NMR properties, which all are very divergent from those observed in UDOM.

An interesting consequence of the considerable degree of unsaturation and oxygenation in $C_{28}H_{32}O_{13}$ molecules is a relative downfield proton NMR shift of the methyl resonances, especially in structures related to *isomer III*, compared to methyl resonances found in peptides and proteins or other alkane environments. This observation, which refers also to other conceivable structures derived from the 613 CRAM ions, implies that a sizable proportion of the methyl resonances in UDOM is derived from peptides rather than from CRAM; i.e. CRAM structures related to *isomers I* and *II* are likely to be more common than those derived from *isomers III*. Both *isomers I and II* also much better conform to the $^{1}H$, $^{13}C$ HSQC NMR cross peak pattern observed in marine UDOM than *isomer III* does (areas 1 and 2; Figure 79).

### *8.3.5. Positioning of CRAM within van-Krevelen diagrams*

Within the given range of H/C and O/C elemental ratios (H/C: 0-2; O/C: 0-1; cf. Figure 88), and the IUPAC mass range from 575.9 – 576.6 Dalton, a continual series of 108 feasible combinations of $C_uH_mO_q$ molecules occurs (n = 20-48; m = 0-80; q = 0-20). These molecules (apart from $C_{48}$) are arranged into 9 isobaric series, which are related by a formal exchange of $CH_4$ against an oxygen atom.




cdr

**Figure 88**. Van Krevelen diagram of all feasible $C_nH_mO_q$ isomers within the provided limits of H/C and O/C elemental ratios defined in this figure, which conform to a single unbroken series of nominal IUPAC mass 576 Dalton. 108 isomers are (apart from $C_{48}$) arranged in 9 series of molecules, which are related by a formal exchange of $CH_4$ versus oxygen. This is the least drastic alteration of molecular structure possible within that diagram (cf. text). Lines are shown, which connect molecules with identical numbers of carbon (green), hydrogen (purple), and oxygen (orange); see also text. In the 613 molecular compositions of CRAM, which have been identified in the FTICR mass spectrum, only the seven isomers indicated as red triangles have been observed; the center one with the molecular formula $C_{28}H_{32}O_{13}$ has been selected to evaluate relationships between molecular composition and NMR properties of CRAM (Figure 87). Molecules with only a very few or no oxygen atoms may become less efficiently ionized than those with a larger number of oxygen atoms in ESI FTICR mass spectra (see also Figure 90).

This alteration results in a mass loss of 36.4 mDalton and an increase of the number of double bond equivalents (DBE) by one unit. Molecules $C_nH_mO_q$ within one of these isobaric series seem to bear more resemblance in composition, but also with respect to chemical structure, than those belonging to different $CH_4$/O-isobaric series, because the minimum other alterations, which would maintain the molecular mass of a $C_nH_mO_q$ molecule within the given unit mass range, require either an exchange of one carbon with 12 hydrogen atoms ($\Delta m$ = 93.90 mDa; $\Delta$DBE = - 7) or the exchange of one oxygen with 16 hydrogen atoms ($\Delta m$ = 130.29 mDa; $\Delta$DBE = - 8). This inevitably would impose a comparatively large alteration of composition and chemical structure. Similarly, the (formal) exchange of 4 carbon with 3 oxygen atoms ($\Delta m$ = 15.26 mDa; $\Delta$ DBE = - 4) in any given chemical environment will result in molecules, which differ rather extensively in chemical structures and therefore, also in its relationship within the diagenesis and reactivity space of NOM. In the (H/C versus O/C) van Krevelen diagram of $C_nH_mO_q$ molecules, all the $CH_4$/O-isobaric series converge at the numerical value of O/C = -1 and H/C = +4; similarly, molecules with identical numbers of hydrogen converge at O/C = -0.75 and H/C = 0, while molecules with identical counts of oxygen atoms converge at H/C = -12 and O/C = 0. These relationships apply not only for the IUPAC nominal mass 576, but are mass independent. Molecules with identical number of carbon atoms occupy positions on parallel lines with a slope of H/C = - …..O/C.

The $C_nH_mO_q$ molecules with the unit mass of 576 Dalton represent, with seven members, the most extended identified isobaric $CH_4$/O-series within the 613 CRAM ion analyzed in the 7T FTICR mass spectrum of deep UDOM (another 12 series of $CH_4$/O-isobars with six members each have also been identified). This series intersects the O/C-axis at a H/C value of 2.06, and the H/C – axis at a O/C value of 1.06 (slope: H/C = - …..O/C + 2.06); i.e. this series (slightly) deviates from the conspicuous line $CH_2$ – CO, which appears to divide the area of identified prominent 613 CRAM ions in the van Krevelen diagram (Figures 85, 89) nearly in half.

This preferred positioning of CRAM molecules in the vicinity of that $CH_2$-CO diagonal line is most likely a consequence of the occurrence of maximum numbers of feasible isomers for a given composition of $C_nH_mO_q$ molecules: both (very) highly unsaturated and near fully saturated $C_nH_m$-skeletons (corresponding to a given mass) exhibit a lesser number of possible isomers than those of intermediate degree of unsaturation. Fully saturated alkanes show a considerable variety of branching options, but lack the additional degrees of freedom, which come along by placing double bonds in various positions or by forming a considerable variety of ring

numbers, -sizes and -junctions. On the other end, highly unsaturated $C_nH_mO_q$ molecules show severe restraints in the number of feasible isomers – more so, if (reasonably) submits like peroxides, cumulated triple bonds or condensed three- and four-membered rings, which are found in natural products, but, are not likely to represent prominent NOM constituents will be excluded from the isomer count. Accordingly, constraints provided by the molecular formulae of highly unsaturated $C_nH_mO_q$ molecules recently identified in deep sea NOM, have led to meaningful proposals of structures from molecular formulae alone.

The presence and occurrence of oxygen in a $C_nH_mO_q$ molecule, which can be formally inserted in any C-C or C-H bond, and is capable to from carbonyl bonds (in keto, ester and carboxylic acid environments) adds enormously to the numbers of feasible $C_nH_mO_q$ isomers (especially when compared to $C_nH_m$ "parent" molecules) – at least up to a O/C ratio of 0.5. In the O/C region of 0.5 – 1.0, the numbers of mathematically feasible isomers remain high, but a lesser degree of meaningful chemical structures (i.e. conceivable from biosynthesis, abiotic and biotic degradation and diagenesis from reasonable NOM precursor molecules) is found.

Therefore, we observe the maximum number of the 613 identified CRAM ions in a region of the van-Krevelen diagram, where also the maximum number of feasible isomers for a given $C_nH_mO_q$ composition is expected. Further studies, which enable the discrimination of isomeric molecules, and which have to rely on $MS^n$-spectra, hyphenated separation (CE, LC)/FTICR mass spectra and NMR spectroscopy are necessary to gain further understanding of marine UDOM composition (and that of NOM in general).

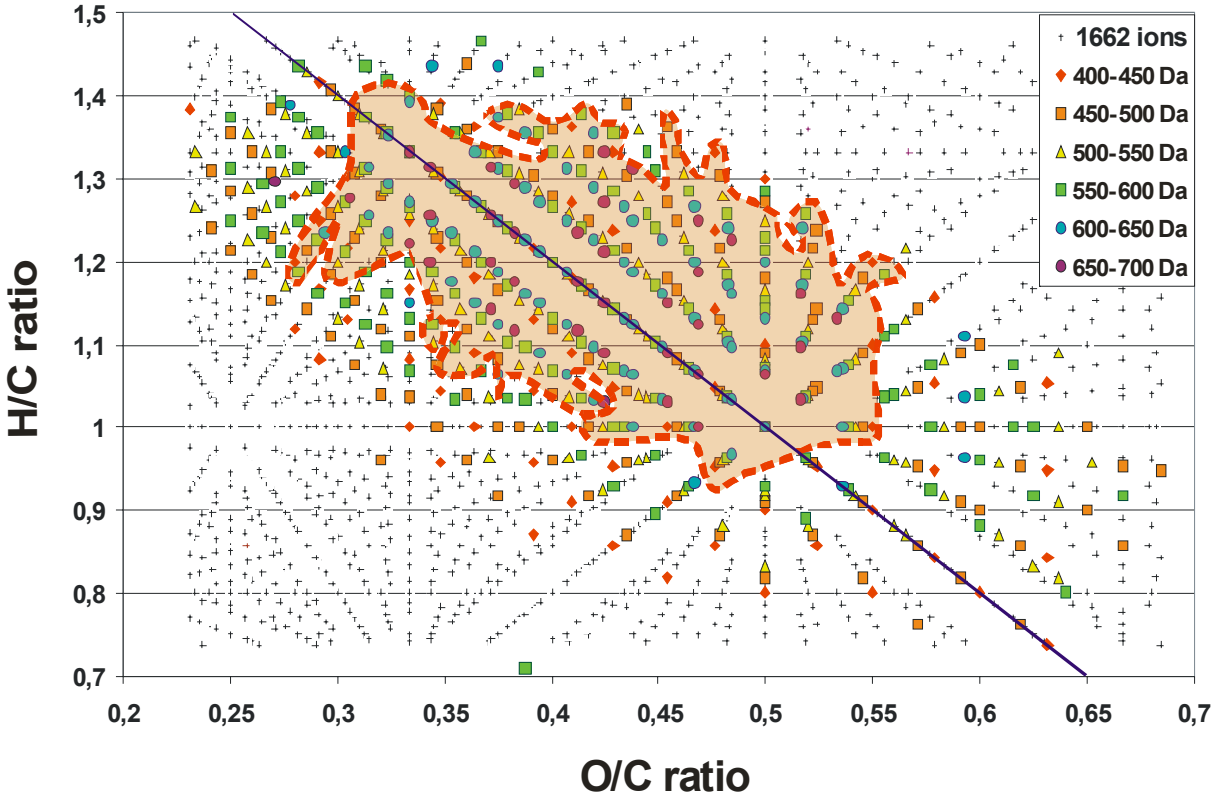



**Figure 89**: van Krevelen diagram of identified 613 CRAM molecules and feasible $C_nH_mO_q$ compositions. A total of 1662 theoretically feasible molecular compositions $C_nH_mO_q$ is found within the nominal IUPAC mass range 415-675 Dalton and the H/C range (0.710 – 1.467) and O/C range (0.231-0.684, which is here defined by the range of CRAM compositions. Within the orange area, CRAM molecules cover *the entire possible* compositional space of $C_nH_mO_q$ molecules. While molecules with additional heteroatoms (e.g. N, S, P….) in UDOM could also contribute to these patterns, all identified 613 CRAM molecules are solely composed of carbon, hydrogen, and oxygen as deduced from the FTICR mass spectra. The blue line connects the virtual endmembers methylene $CH_2$ and carbon monoxide CO; it also bisects the orange area (100 % coverage of compositional space) nearly in half; this is most likely a consequence of a maximum number of feasible isomers for a given molecular formula $C_nH_mO_q$ (cf. Figures 88, 90 and text).

### *8.3.6 The relationship between ion count in FTICR mass spectra and the number of feasible isomeric CRAM*

For an extremely complex material such as marine UDOM it is reasonable to postulate, that many isomers will contribute to any given molecular formula. Analogously, the intensity of mass spectral peaks of UDOM will be a complex function of the number of feasible isomers, which can be derived from a given molecular formula, the occurrence of these isomers in UDOM, and the ionization efficiency of any single isomer under the given experimental conditions.

Two major and independent trends are expected to define the number of feasible $C_nH_mO_q$ isomers for a given mass. First, the maximum number of feasible $C_nH_mO_q$ isomers is expected at intermediate numbers of double-bond equivalents in a molecule, because double-bond displacement and ring-formation, which both greatly enhance the number of feasible isomers, depend on the occurrence of DBE (which solely depend on the H/C-ratio). However, at very low H/C ratios, only highly condensed structures can be formed, this restriction very greatly diminishes the number of feasible $C_nH_mO_q$ isomers (if mathematically possible, but chemically unlikely isomers are excluded; cf. below).

Second, the presence of oxygen allows for greatly enlarged numbers of feasible isomers at low O/C ratios; in the case of available DBE, carbonyl derivatives (C=O) can be also constructed. At higher O/C ratios however, the number of feasible isomers declines for two reasons: oxygen provides only two options to form (single) bonds to other partners, while carbon has four options; in addition, the higher mass of oxygen compared with carbon reduces the total number of "heavy" atoms on oxygen-rich molecules $C_nH_mO_q$ of a given mass.

These tentative considerations lead to the expectations, that the number of feasible $C_nH_mO_q$ isomers for a given mass will reach maximum values at intermediate H/C and O/C ratios and will (sharply) decline at extreme (high and low) H/C and O/C ratios, respectively.

In order to illustrate these dependencies, it is instructive to display the numbers of feasible isomers for a given molecular composition $C_nH_mO_q$ in a van-Krevelen diagram. For practical reasons, we have selected $C_nH_mO_q$ compositions with a nominal IUPAC mass of 178, for which the number of isomers can be calculated within reasonable times on desktop computers: 11 feasible molecular compositions are found, grouped into 3 series of $CH_4$/O isobaric molecules (Figure 90).

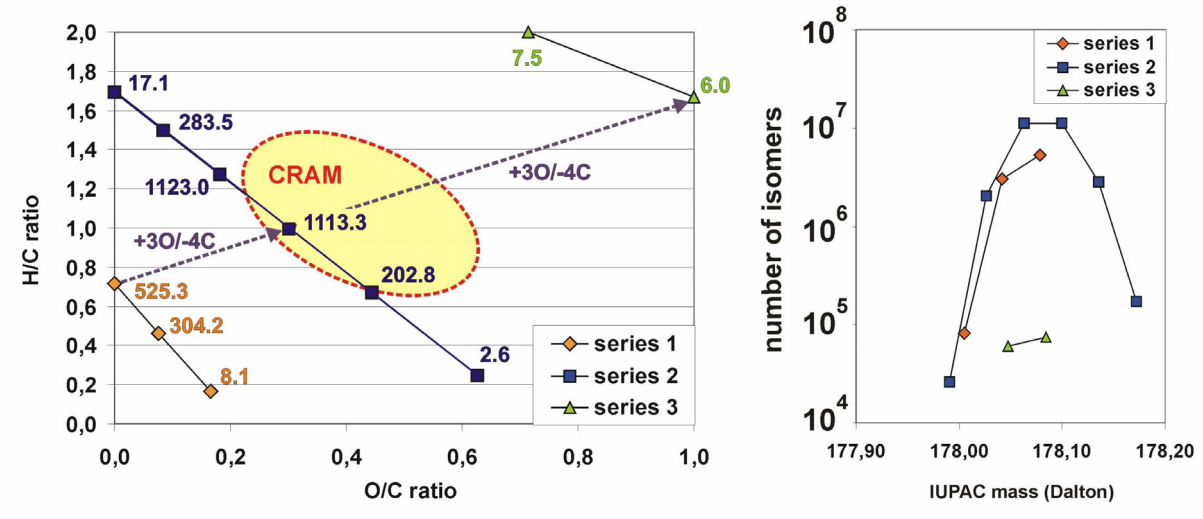



**Figure 90:** Van Krevelen diagram (left) of 11 molecular compositions $C_nH_mO_q$ (3 series of $CH_4$/O isobaric molecules) with a IUPAC nominal mass of 178 Dalton, with numbers of feasible isomers given (unit count = 10000 isomers). Certain restrictions have been applied to exclude mathematically possible, but chemically unlikely structures of UDOM-like materials: (1) no peroxides; i.e. no –O–O– connectivities; (2) no triple bonds (–C≡C–); (3) no three- or four-membered rings, and (4) no carbon with two double bonds (═C═) are allowed. The dotted line relates isobaric molecules, in which four carbon atoms have been exchanged by three oxygen atoms (cf. text). Right: numbers of calculated $C_nH_mO_q$ isomers according to series 1-3 and IUPAC mass (cf. text). The numbers of isomers have been calculated, using the software MOLGEN (Department of Mathematics, University of Bayreuth, D-95440 Bayreuth, Germany) by M. Meringer (Kiadis, Groningen, NL), R. Gugisch (Bayreuth) and C. Ruecker (Bayreuth).

In the highly unsaturated series 1, the number of isomers sharply declines with decreasing H/C ratio, because the considerable to extreme hydrogen-deficiency allows mainly (or only) highly condensed structures. Series 2 shows maximum numbers of isomers at intermediate H/C (and O/C) ratios and declining numbers of isomers in both high and low H/C (and O/C) ratios. Again, in the case of very low H/C ratio, which corresponds to extreme hydrogen deficiency and unsaturation, only very highly condensed molecules can be realized; this severely limits the total number of feasible isomers. In contrast, at very high H/C ratios, only a limited number of double-bond equivalents (DBE) is available. However, the availability of DBE in a molecule $C_nH_mO_q$ greatly enlarges the options to construct isomers by shifting double bonds or/and by the formation of a huge variety of rings. Here, for molecules with a IUPAC mass of 178 Dalton, the maximum H/C ratio in series 2 is 1.692 and the corresponding molecule $C_{13}H_{22}$ features 3 DBE. This already allows for a huge array of unsaturation-related isomers, compared with a fully saturated molecule. [The fully saturated analogue $C_{13}H_{28}$ (184 Dalton)

of the series 2 "endmember" $C_{13}H_{22}$ (178 Dalton, $1.7 \cdot 10^5$ isomers) shows only 802 isomers; analogous behaviour is found for the series 1 "endmember" $C_{14}H_{10}$ (178 Dalton, $5.3 \cdot 10^6$ isomers) in relationship to the fully saturated molecule $C_{14}H_{30}$ (198 Dalton, 1858 isomers), respectively].

The isobaric "neighbour" molecule $C_{12}H_{18}O$ exhibits 4 DBE, but the number of isomers related to $C_{13}H_{22}$ is by a factor of 16.6 higher. A very substantial fraction of that increase is caused by the availability of the single oxygen atom, which can be (formally) inserted in any C-C or C-H bond to produce new isomers with singly bonded oxygen; in addition, DBE can be expressed in the form of C═O bonds (carbonyl derivatives). The maximum number of isomers (approx. $1.1 \cdot 10^7$ each) is reached for the molecules $C_{11}H_{14}O_2$ (5 DBE) and $C_{10}H_{10}O_3$ (6 DBE), respectively. Further exchange of $CH_4$ against oxygen again sharply declines the number of feasible isomers (by a factor of 5.5 when proceeding from $C_{10}H_{10}O_3$ to $C_9H_6O_4$, and by another factor of 78 for changing from $C_9H_6O_4$ to $C_8H_2O_5$).

Series 3 molecules $C_7H_{14}O_5$ and $C_6H_{10}O_6$ feature very limited numbers of isomers because of their large O/C ratio (cf. above). The comparison of $C_8H_2O_5$ and $C_7H_{14}O_5$ demonstrates, that extreme hydrogen-deficiency more severely restricts the feasible number of isomers than a large O/C ratio. Molecules $C_{14}H_{10}$ (10 DBE, $5.3 \cdot 10^6$ isomers; series 1), $C_{10}H_{10}O_3$ (6 DBE, $1.1 \cdot 10^7$ isomers; series 2) and $C_6H_{10}O_6$ (2 DBE, $6 \cdot 10^4$ isomers; series 3) are all related by a formal exchange of four carbon against three oxygen atoms. The introduction of oxygen at the expense of available DBE at first clearly outweighs the lesser number of carbon atoms and DBE because of the (1) reduced severity in unsaturation and (2) the availability of oxygen in the construction of isomers (cf. above). On the transition from $C_{10}H_{10}O_3$ to $C_6H_{10}O_6$ however, both the reduced options of oxygen to take part in chemical binding (two bonds for any oxygen instead of four for any carbon) and the decline in available DBE lead to the drastic decline in the numbers of accessible isomers.

Under the assumption of a very complex chemical structure of UDOM, in which any given molecular formula is represented by a (potentially huge) number of isomers, mass spectral intensities should coarsely correlate with the feasible numbers of isomers. However, oxygen-depleted molecules $C_nH_m$ are less likely to become ionized in ESI-FTICR mass spectra under standard conditions than oxygenated molecules $C_nH_mO_q$. Carbohydrates, which are oxygen-rich, also are less efficiently ionized under standard ESI-FTICR mass spectral conditions than

carboxylic-rich molecules like CRAM, which are the most likely candidates to produce strong signals in ESI-FTICR mass spectra.

Recently, it has been demonstrated, that chemical and photo-ionization of NOM produce ions in FTICR mass spectra, which less strongly depend on the presence of ESI ionizable groups, such as carboxylic acids (see also Figure 88). However, the very large carboxylic acid $^{13}C$ NMR resonance in both surfaces and deep UDOM, and the considerable molecular weight of CRAM, established by FTICR mass spectra, make the occurrence of carboxylic acid functional groups in many molecular compositions $C_nH_mO_q$ of CRAM almost certain. The polyfunctionality of CRAM could also explain the relative insensitivity of the appearance of its ESI FTICR mass spectra, when acquired under conditions of positive and negative ionization (in sharp contrast to mixtures of more clearly classified (bio)molecules).

## 8.4. CONCLUSIONS

Based on all these data, it appears the overwhelming majority of CRAM-ions is represented by carboxylated and fused alicyclic rings with very few hydrogen atoms in double bonds. The biochemical origins and mechanisms of formation of CRAM are unclear, but CRAM shares some structural characteristics found in terpenoids, a diverse class of biochemicals (e.g. isoprenoids, sterols, hopanoids, crenarchaeols) occurring as membrane constituents and secondary metabolites in a wide range of prokaryotic and eukaryotic organisms (Ourisson et al., 1987). Several studies have noted the occurrence of prokaryotic cell wall and membrane components in marine DOM (McCarthy et al., 1998; Benner and Kaiser, 2003; Wakeham et al., 2003; Tanoue et al., 1995), so they are likely sources of CRAM as well. The molecular formulae determined from the FTICR mass spectra data indicated the occurrence of polycarboxylated fused-ring systems rather than open chain isomers. These polycarboxylated fused-ring systems form the general structures of sterols and hopanoids with oxidized side chains.

Similar structures could also be derived from other biochemical precursors that have undergone more extensive alterations. Acrylic acid is a decomposition product of dimethylsulphopropionate, which is produced by several abundant marine phytoplankton (Dacey and Wakeham, 1986). Acrylic acid is susceptible to addition reactions and photo-polymerization in seawater and could also be a precursor of CRAM. Furthermore, radical mediated photo-oxidation of any functionalized organic molecule in oxygenated surface ocean waters is

expected to produce oxidation products with non-repetitive structures (Schmitt-Kopplin et al., 1998).

CRAM are the most abundant identified component of DOM in the deep ocean. Previous studies characterized only ~3% of DOC as specific biochemicals in the deep ocean (Benner, 2002) whereas CRAM accounts for ~8% of the DOC. The structural diversity found within CRAM and its substantial content of alicyclic rings and branching contribute to its resistance to biodegradation and refractory nature. It appears CRAM are largely comprised of the decomposition products of biomolecules, as indicated by its prevalence of carboxyl groups and pattern of increasing oxidation with decreasing molecular size. CRAM are expected to constitute a strong ligand for metal binding, and multiple coordination across calcium cations could promote aggregation and gel formation (Chin et al., 1998), thereby affecting the reactivity of organic matter and the bioavailability of associated nutrients and trace metals.

The occurrence of CRAM in freshwater and terrestrial environments seems likely, considering the global distribution of biomolecules and the similarities of biogeochemical processes among environments. In addition to its abundance in the ultrafiltered fraction of DOM in the ocean, CRAM also appears to be an important component of the humic fraction of marine (Hedges et al., 1992) and freshwater DOM (Leenheer et al., 2003).

The very low resolution signature of CRAM in both separation and spectroscopic techniques has precluded so far its recognition as a relevant constituent of natural organic matter, although the occurrence of sizable amounts of branched aliphatic structures in marine (Sardessai and Wahidullah, 1998, Hedges et al., 1992, Hatcher et al., 1979, Hatcher et al., 1981) and freshwater (Lambert and Buddrus, 1996, Buddrus and Lambert, 1995, Wilson et al, 1981) NOM materials has been proposed in numerous earlier NMR studies. Terpenoid structures have been characterized in fossil organic matter, such as kerogen and petroleum, indicating their ubiquitous and refractory nature (Ourisson et al., 1987; Summons, 1993). Unraveling the mysteries about its formation and cycling will provide critical insights into the global carbon cycle.

## 8.5. APPENDIX

**Appendix 1.** experimentally determined and theoretical mass numbers from positive ion 7 Tesla Fourier transform ion cyclotron mass spectra of deep marine UDOM together with molecular formula $C_cH_hO_o$ , representing the minimal deviation from the theoretical mass values with the options $C_{0\text{-}50}H_{0\text{-}100}O_{0\text{-}50}N_{0\text{-}50}S_{0\text{-}10}$ provided prior to calculation.

**A** experimentally determined mass numbers (based on IUPAC nominal mass; $^{12}C$ = 12.000000) from the positive ion in atomic mass units; these ions are protonated adducts

**B** theoretical mass number (based on IUPAC nominal mass) of the respective best hit (range allowed: C: 0-50, H: 0-100, O: 0-50, N: 0-50; other elements allowed) given in atomic mass units

**C** difference between A and B

**D** number of carbon atoms in the most likely molecular formula (calculated hit with the least deviation from theoretical values)

**E** number of hydrogen atoms in the most likely molecular formula (calculated hit with the least deviation from theoretical values; the original ion is a protonated adduct)

**F** number of oxygen atoms in the most likely molecular formula (calculated hit with the least deviation from theoretical values)

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 1 | 415,10141 | 415,10236 | 2,29 | 21 | 18 | 9 |
| 2 | 415,13797 | 415,13874 | 1,85 | 22 | 22 | 8 |
| 3 | 415,17405 | 415,17513 | 2,6 | 23 | 26 | 7 |
| 4 | 417,08070 | 417,08162 | 2,21 | 20 | 16 | 10 |
| 5 | 417,11720 | 417,11801 | 1,94 | 21 | 20 | 9 |
| 6 | 417,15349 | 417,15439 | 2,16 | 22 | 24 | 8 |
| 7 | 417,19014 | 417,19078 | 1,53 | 23 | 28 | 7 |
| 8 | 419,09650 | 419,09727 | 1,84 | 20 | 18 | 10 |
| 9 | 419,13277 | 419,13366 | 2,12 | 21 | 22 | 9 |
| 10 | 419,16910 | 419,17004 | 2,24 | 22 | 26 | 8 |
| 11 | 419,20525 | 419,20643 | 2,81 | 23 | 30 | 7 |
| 12 | 421,07604 | 421,07654 | 1,19 | 19 | 16 | 11 |
| 13 | 421,11206 | 421,11292 | 2,04 | 20 | 20 | 10 |
| 14 | 421,14848 | 421,14931 | 1,97 | 21 | 24 | 9 |
| 15 | 421,18488 | 421,18569 | 1,92 | 22 | 28 | 8 |
| 16 | 423,09139 | 423,09219 | 1,89 | 19 | 18 | 11 |
| 17 | 423,12766 | 423,12857 | 2,15 | 20 | 22 | 10 |
| 18 | 423,16415 | 423,16496 | 1,91 | 21 | 26 | 9 |
| 19 | 423,20011 | 423,20194 | 4,32 | 22 | 30 | 8 |
| 20 | 425,10680 | 425,10784 | 2,45 | 19 | 20 | 11 |
| 21 | 425,14331 | 425,14422 | 2,14 | 20 | 24 | 10 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 22 | 425,17971 | 425,18061 | 2,12 | 21 | 28 | 9 |
| 23 | 427,12243 | 427,12349 | 2,48 | 19 | 22 | 11 |
| 24 | 427,13816 | 427,13874 | 1,36 | 23 | 22 | 8 |
| 25 | 427,15946 | 427,15987 | 0,96 | 20 | 26 | 10 |
| 26 | 427,17417 | 427,17513 | 2,25 | 24 | 26 | 7 |
| 27 | 429,11749 | 429,11801 | 1,21 | 22 | 20 | 9 |
| 28 | 429,15372 | 429,15493 | 2,82 | 23 | 24 | 8 |
| 29 | 429,18993 | 429,19078 | 1,98 | 24 | 28 | 7 |
| 30 | 431,09658 | 431,09727 | 1,6 | 21 | 18 | 10 |
| 31 | 431,13293 | 431,13366 | 1,69 | 22 | 22 | 9 |
| 32 | 431,16917 | 431,17004 | 2,02 | 23 | 26 | 8 |
| 33 | 431,20541 | 431,20643 | 2,37 | 24 | 30 | 7 |
| 34 | 433,07524 | 433,07654 | 3 | 20 | 16 | 11 |
| 35 | 433,11211 | 433,11292 | 1,87 | 21 | 20 | 10 |
| 36 | 433,14847 | 433,14931 | 1,94 | 22 | 24 | 9 |
| 37 | 433,18493 | 433,18569 | 1,75 | 23 | 28 | 8 |
| 38 | 433,22145 | 433,22208 | 1,45 | 24 | 32 | 7 |
| 39 | 435,05420 | 435,05580 | 3,68 | 19 | 14 | 12 |
| 40 | 435,09141 | 435,09219 | 1,79 | 20 | 18 | 11 |
| 41 | 435,12783 | 435,12857 | 1,7 | 21 | 22 | 10 |
| 42 | 435,16413 | 435,16496 | 1,91 | 22 | 26 | 9 |
| 43 | 435,20050 | 435,20134 | 1,93 | 23 | 30 | 8 |
| 44 | 435,23671 | 435,23773 | 2,34 | 24 | 34 | 7 |
| 45 | 437,07072 | 437,07145 | 1,67 | 19 | 16 | 12 |
| 46 | 437,10700 | 437,10784 | 1,92 | 20 | 20 | 11 |
| 47 | 437,14342 | 437,14422 | 1,83 | 21 | 24 | 10 |
| 48 | 437,17964 | 437,18061 | 2,22 | 22 | 28 | 9 |
| 49 | 437,21682 | 437,21699 | 0,39 | 23 | 32 | 8 |
| 50 | 439,08647 | 439,08710 | 1,43 | 19 | 18 | 12 |
| 51 | 439,12268 | 439,12349 | 1,84 | 20 | 22 | 11 |
| 52 | 439,15890 | 439,15987 | 2,21 | 21 | 26 | 10 |
| 53 | 439,19500 | 439,19626 | 2,87 | 22 | 30 | 9 |
| 54 | 441,10196 | 441,10275 | 1,79 | 19 | 20 | 12 |
| 55 | 441,11771 | 441,11801 | 0,68 | 23 | 20 | 9 |
| 56 | 441,13824 | 441,13914 | 2,04 | 20 | 24 | 11 |
| 57 | 441,15386 | 441,15439 | 1,2 | 24 | 24 | 8 |
| 58 | 441,17399 | 441,17552 | 3,47 | 21 | 28 | 10 |
| 59 | 441,18987 | 441,19078 | 2,06 | 25 | 28 | 7 |
| 60 | 443,09775 | 443,09727 | -1,08 | 22 | 18 | 10 |
| 61 | 443,13321 | 443,13366 | 1,02 | 23 | 22 | 9 |
| 62 | 443,16940 | 443,17004 | 1,44 | 24 | 26 | 8 |
| 63 | 443,20580 | 443,20643 | 1,42 | 25 | 30 | 7 |
| 64 | 445,11250 | 445,11292 | 0,94 | 22 | 20 | 10 |
| 65 | 445,14867 | 445,14931 | 1,44 | 23 | 24 | 9 |
| 66 | 445,18507 | 445,18569 | 1,39 | 24 | 28 | 8 |
| 67 | 445,22147 | 445,22208 | 1,37 | 25 | 32 | 7 |
| 68 | 445,25768 | 445,25846 | 1,75 | 26 | 36 | 6 |
| 69 | 447,09165 | 447,09219 | 1,21 | 21 | 18 | 11 |
| 70 | 447,12791 | 447,12857 | 1,48 | 22 | 22 | 10 |
| 71 | 447,16426 | 447,16496 | 1,57 | 23 | 26 | 9 |
| 72 | 447,20056 | 447,20134 | 1,74 | 24 | 30 | 8 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 73 | 449,07084 | 449,07145 | 1,36 | 20 | 16 | 12 |
| 74 | 449,10727 | 449,10784 | 1,27 | 21 | 20 | 11 |
| 75 | 449,14355 | 449,14422 | 1,49 | 22 | 24 | 10 |
| 76 | 449,17993 | 449,18061 | 1,51 | 23 | 28 | 9 |
| 77 | 449,21598 | 449,21699 | 2,25 | 24 | 32 | 8 |
| 78 | 451,08662 | 451,08710 | 1,06 | 20 | 18 | 12 |
| 79 | 451,12281 | 451,12349 | 1,51 | 21 | 22 | 11 |
| 80 | 451,15921 | 451,15987 | 1,46 | 22 | 26 | 10 |
| 81 | 451,19554 | 451,19626 | 1,6 | 23 | 30 | 9 |
| 82 | 453,10227 | 453,10275 | 1,06 | 20 | 20 | 12 |
| 83 | 453,13847 | 453,13914 | 1,48 | 21 | 24 | 11 |
| 84 | 453,15372 | 453,15439 | 1,48 | 25 | 24 | 8 |
| 85 | 453,17438 | 453,17552 | 2,52 | 22 | 28 | 10 |
| 86 | 455,08099 | 455,08217 | 2,59 | 19 | 18 | 13 |
| 87 | 455,11765 | 455,11840 | 1,65 | 20 | 22 | 12 |
| 88 | 455,13374 | 455,13366 | -0,18 | 24 | 22 | 9 |
| 89 | 455,15374 | 455,15479 | 2,31 | 21 | 26 | 11 |
| 90 | 455,16949 | 455,17004 | 1,21 | 25 | 26 | 8 |
| 91 | 455,19091 | 455,19117 | 0,57 | 22 | 30 | 10 |
| 92 | 455,20547 | 455,20643 | 2,11 | 26 | 30 | 7 |
| 93 | 457,11283 | 457,11292 | 0,2 | 23 | 20 | 10 |
| 94 | 457,14887 | 457,14931 | 0,96 | 24 | 24 | 9 |
| 95 | 457,18514 | 457,18569 | 1,2 | 25 | 28 | 8 |
| 96 | 457,22123 | 457,22208 | 1,86 | 26 | 32 | 7 |
| 97 | 459,09207 | 459,09219 | 0,26 | 22 | 18 | 11 |
| 98 | 459,12820 | 459,12857 | 0,81 | 23 | 22 | 10 |
| 99 | 459,16443 | 459,16496 | 1,15 | 24 | 26 | 9 |
| 100 | 459,20082 | 459,20134 | 1,13 | 25 | 30 | 8 |
| 101 | 459,23688 | 459,23773 | 1,85 | 26 | 34 | 7 |
| 102 | 461,07133 | 461,07145 | 0,26 | 21 | 16 | 12 |
| 103 | 461,10744 | 461,10784 | 0,87 | 22 | 20 | 11 |
| 104 | 461,14364 | 461,14422 | 1,26 | 23 | 24 | 10 |
| 105 | 461,18001 | 461,18061 | 1,3 | 24 | 28 | 9 |
| 106 | 461,21627 | 461,21699 | 1,56 | 25 | 32 | 8 |
| 107 | 461,25225 | 461,25338 | 2,45 | 26 | 36 | 7 |
| 108 | 463,08673 | 463,08710 | 0,8 | 21 | 18 | 12 |
| 109 | 463,12301 | 463,12349 | 1,04 | 22 | 22 | 11 |
| 110 | 463,15938 | 463,15987 | 1,06 | 23 | 26 | 10 |
| 111 | 463,19570 | 463,19626 | 1,21 | 24 | 30 | 9 |
| 112 | 463,23195 | 463,23264 | 1,49 | 25 | 34 | 8 |
| 113 | 465,10212 | 465,10275 | 1,35 | 21 | 20 | 12 |
| 114 | 465,13850 | 465,13914 | 1,38 | 22 | 24 | 11 |
| 115 | 465,17489 | 465,17552 | 1,35 | 23 | 28 | 10 |
| 116 | 465,21126 | 465,21191 | 1,4 | 24 | 32 | 9 |
| 117 | 467,08175 | 467,08202 | 0,58 | 20 | 18 | 13 |
| 118 | 467,11797 | 467,11840 | 0,92 | 21 | 22 | 12 |
| 119 | 467,15428 | 467,15479 | 1,09 | 22 | 26 | 11 |
| 120 | 467,19052 | 467,19117 | 1,39 | 23 | 30 | 10 |
| 121 | 469,09644 | 469,09767 | 2,62 | 20 | 20 | 13 |
| 122 | 469,13354 | 469,13405 | 1,09 | 21 | 24 | 12 |
| 123 | 469,14928 | 469,14931 | 0,06 | 25 | 24 | 9 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 124 | 469,16973 | 469,17044 | 1,51 | 22 | 28 | 11 |
| 125 | 469,18540 | 469,18569 | 0,62 | 26 | 28 | 8 |
| 126 | 469,20560 | 469,20682 | 2,6 | 23 | 32 | 10 |
| 127 | 469,22126 | 469,22208 | 1,75 | 27 | 32 | 7 |
| 128 | 471,12867 | 471,12857 | -0,21 | 24 | 22 | 10 |
| 129 | 471,16454 | 471,16496 | 0,89 | 25 | 26 | 9 |
| 130 | 471,20077 | 471,20134 | 1,21 | 26 | 30 | 8 |
| 131 | 471,23708 | 471,23773 | 1,38 | 27 | 34 | 7 |
| 132 | 473,10800 | 473,10784 | -0,34 | 23 | 20 | 11 |
| 133 | 473,14410 | 473,14422 | 0,25 | 24 | 24 | 10 |
| 134 | 473,18034 | 473,18061 | 0,57 | 25 | 28 | 9 |
| 135 | 473,21658 | 473,21699 | 0,87 | 26 | 32 | 8 |
| 136 | 473,25286 | 473,25338 | 1,1 | 27 | 36 | 7 |
| 137 | 473,27433 | 473,27451 | 0,38 | 24 | 40 | 9 |
| 138 | 475,08710 | 475,08712 | 0,04 | 22 | 18 | 12 |
| 139 | 475,12314 | 475,12349 | 0,74 | 23 | 22 | 11 |
| 140 | 475,15932 | 475,15984 | 1,09 | 24 | 26 | 10 |
| 141 | 475,19566 | 475,19626 | 1,26 | 25 | 30 | 9 |
| 142 | 475,23188 | 475,23265 | 1,62 | 26 | 34 | 8 |
| 143 | 477,06688 | 477,06637 | -1,07 | 21 | 16 | 13 |
| 144 | 477,10250 | 477,10275 | 0,52 | 22 | 20 | 12 |
| 145 | 477,13876 | 477,13914 | 0,8 | 23 | 24 | 11 |
| 146 | 477,17500 | 477,17552 | 1,09 | 24 | 28 | 10 |
| 147 | 477,21150 | 477,21191 | 0,86 | 25 | 32 | 9 |
| 148 | 477,24772 | 477,24829 | 1,19 | 26 | 36 | 8 |
| 149 | 479,08156 | 479,08202 | 0,96 | 21 | 18 | 13 |
| 150 | 479,11793 | 479,11840 | 0,98 | 22 | 22 | 12 |
| 151 | 479,15439 | 479,15479 | 0,83 | 23 | 26 | 11 |
| 152 | 479,19070 | 479,19117 | 0,98 | 24 | 30 | 10 |
| 153 | 479,22700 | 479,22756 | 1,17 | 25 | 34 | 9 |
| 154 | 481,09746 | 481,09767 | 0,44 | 21 | 20 | 13 |
| 155 | 481,13374 | 481,13405 | 0,64 | 22 | 24 | 12 |
| 156 | 481,17007 | 481,17044 | 0,77 | 23 | 28 | 11 |
| 157 | 481,20644 | 481,20682 | 0,79 | 24 | 32 | 10 |
| 158 | 481,24250 | 481,24321 | 1,48 | 25 | 36 | 9 |
| 159 | 483,11270 | 483,11332 | 1,28 | 21 | 22 | 13 |
| 160 | 483,14943 | 483,14970 | 0,56 | 22 | 26 | 12 |
| 161 | 483,16505 | 483,16496 | -0,19 | 26 | 26 | 9 |
| 162 | 483,18572 | 483,18609 | 0,77 | 23 | 30 | 11 |
| 163 | 483,20090 | 483,20134 | 0,91 | 27 | 30 | 8 |
| 164 | 483,23722 | 483,23773 | 1,06 | 28 | 34 | 7 |
| 165 | 485,14441 | 485,14422 | -0,39 | 25 | 24 | 10 |
| 166 | 485,18052 | 485,18061 | 0,19 | 26 | 28 | 9 |
| 167 | 485,21664 | 485,21699 | 0,72 | 27 | 32 | 8 |
| 168 | 485,25280 | 485,25388 | 2,23 | 28 | 36 | 7 |
| 169 | 487,12357 | 487,12349 | -0,16 | 24 | 22 | 11 |
| 170 | 487,15954 | 487,15987 | 0,68 | 25 | 26 | 10 |
| 171 | 487,19590 | 487,19626 | 0,74 | 26 | 30 | 9 |
| 172 | 487,23223 | 487,23264 | 0,84 | 27 | 34 | 8 |
| 173 | 487,26770 | 487,26903 | 2,73 | 28 | 38 | 7 |
| 174 | 489,10270 | 489,10275 | 0,1 | 23 | 20 | 12 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 175 | 489,13898 | 489,13914 | 0,33 | 24 | 24 | 11 |
| 176 | 489,17533 | 489,17552 | 0,39 | 25 | 28 | 10 |
| 177 | 489,21167 | 489,21191 | 0,49 | 26 | 32 | 9 |
| 178 | 489,24807 | 489,24830 | 0,47 | 27 | 36 | 8 |
| 179 | 491,08234 | 491,08202 | -0,65 | 22 | 18 | 13 |
| 180 | 491,11814 | 491,11840 | 0,53 | 23 | 22 | 12 |
| 181 | 491,15448 | 491,15479 | 0,63 | 24 | 26 | 11 |
| 182 | 491,19085 | 491,19117 | 0,65 | 25 | 30 | 10 |
| 183 | 491,22711 | 491,22756 | 0,92 | 26 | 34 | 9 |
| 184 | 491,26377 | 491,26394 | 0,35 | 27 | 38 | 8 |
| 185 | 493,09760 | 493,09767 | 0,14 | 22 | 20 | 13 |
| 186 | 493,13379 | 493,13405 | 0,53 | 23 | 24 | 12 |
| 187 | 493,17006 | 493,17044 | 0,77 | 24 | 28 | 11 |
| 188 | 493,20635 | 493,20682 | 0,95 | 25 | 32 | 10 |
| 189 | 493,24234 | 493,24321 | 1,76 | 26 | 36 | 9 |
| 190 | 495,07700 | 495,07693 | -0,14 | 21 | 18 | 14 |
| 191 | 495,11324 | 495,11332 | 0,16 | 22 | 22 | 13 |
| 192 | 495,14951 | 495,14970 | 0,38 | 23 | 26 | 12 |
| 193 | 495,18578 | 495,18609 | 0,63 | 24 | 30 | 11 |
| 194 | 495,22194 | 495,22247 | 1,07 | 25 | 34 | 10 |
| 195 | 497,09269 | 497,09258 | -0,22 | 21 | 20 | 14 |
| 196 | 497,12865 | 497,12897 | 0,64 | 22 | 24 | 13 |
| 197 | 497,16495 | 497,16535 | 0,8 | 23 | 28 | 12 |
| 198 | 497,18045 | 497,18061 | 0,32 | 27 | 28 | 9 |
| 199 | 497,20120 | 497,20174 | 1,09 | 24 | 32 | 11 |
| 200 | 497,21671 | 497,21699 | 0,56 | 28 | 32 | 8 |
| 201 | 497,25254 | 497,25338 | 1,69 | 29 | 36 | 7 |
| 202 | 499,16013 | 499,15984 | -0,58 | 26 | 26 | 10 |
| 203 | 499,19612 | 499,19626 | 0,28 | 27 | 30 | 9 |
| 204 | 499,23217 | 499,23265 | 0,96 | 28 | 34 | 8 |
| 205 | 499,26879 | 499,26903 | 0,48 | 29 | 38 | 7 |
| 206 | 501,10332 | 501,10275 | -1,14 | 24 | 20 | 12 |
| 207 | 501,13913 | 501,13918 | 0,1 | 25 | 24 | 11 |
| 208 | 501,17535 | 501,17552 | 0,34 | 26 | 28 | 10 |
| 209 | 501,21164 | 501,21191 | 0,54 | 27 | 32 | 9 |
| 210 | 501,24796 | 501,24830 | 0,68 | 28 | 36 | 8 |
| 211 | 503,11856 | 503,11840 | -0,32 | 24 | 22 | 12 |
| 212 | 503,15462 | 503,15479 | 0,34 | 25 | 26 | 11 |
| 213 | 503,19086 | 503,19117 | 0,62 | 26 | 30 | 10 |
| 214 | 503,22729 | 503,22756 | 0,54 | 27 | 34 | 9 |
| 215 | 503,26311 | 503,26395 | 1,67 | 28 | 38 | 8 |
| 216 | 505,09796 | 505,09757 | -0,77 | 23 | 20 | 13 |
| 217 | 505,13399 | 505,13405 | 0,12 | 24 | 24 | 12 |
| 218 | 505,17028 | 505,17044 | 0,32 | 25 | 28 | 11 |
| 219 | 505,20659 | 505,20682 | 0,46 | 26 | 32 | 10 |
| 220 | 505,24287 | 505,24321 | 0,67 | 27 | 36 | 9 |
| 221 | 505,27948 | 505,27960 | 0,24 | 28 | 40 | 8 |
| 222 | 507,07665 | 507,07693 | 0,55 | 22 | 18 | 14 |
| 223 | 507,11334 | 507,11332 | -0,04 | 23 | 22 | 13 |
| 224 | 507,14965 | 507,14970 | 0,1 | 24 | 26 | 12 |
| 225 | 507,18589 | 507,18609 | 0,39 | 25 | 30 | 11 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 226 | 507,22222 | 507,22247 | 0,49 | 26 | 34 | 10 |
| 227 | 507,25851 | 507,25886 | 0,69 | 27 | 38 | 9 |
| 228 | 509,09263 | 509,09258 | -0,1 | 22 | 20 | 14 |
| 229 | 509,12912 | 509,12897 | -0,29 | 23 | 24 | 13 |
| 230 | 509,16541 | 509,16535 | -0,12 | 24 | 28 | 12 |
| 231 | 509,20169 | 509,20174 | 0,1 | 25 | 32 | 11 |
| 232 | 509,23797 | 509,23812 | 0,29 | 26 | 36 | 10 |
| 233 | 511,10839 | 511,10823 | -0,31 | 22 | 22 | 14 |
| 234 | 511,14460 | 511,14462 | 0,04 | 23 | 26 | 13 |
| 235 | 511,16061 | 511,15987 | -1,45 | 27 | 26 | 10 |
| 236 | 511,18096 | 511,18100 | 0,08 | 24 | 30 | 12 |
| 237 | 511,19629 | 511,19626 | -0,06 | 28 | 30 | 9 |
| 238 | 511,21697 | 511,21739 | 0,82 | 25 | 34 | 11 |
| 239 | 511,23274 | 511,23265 | -0,18 | 29 | 34 | 8 |
| 240 | 511,26817 | 511,26903 | 1,68 | 30 | 38 | 7 |
| 241 | 513,12388 | 513,12388 | 0 | 22 | 24 | 14 |
| 242 | 513,13960 | 513,13914 | -0,9 | 26 | 24 | 11 |
| 243 | 513,16024 | 513,16027 | 0,06 | 23 | 28 | 13 |
| 244 | 513,17546 | 513,17552 | 0,12 | 27 | 28 | 10 |
| 245 | 513,21171 | 513,21191 | 0,39 | 28 | 32 | 9 |
| 246 | 513,24792 | 513,24830 | 0,74 | 29 | 36 | 8 |
| 247 | 513,28415 | 513,28468 | 1,03 | 30 | 40 | 7 |
| 248 | 515,11859 | 515,11840 | -0,37 | 25 | 22 | 12 |
| 249 | 515,15491 | 515,15479 | -0,23 | 26 | 26 | 11 |
| 250 | 515,19116 | 515,19117 | 0,02 | 27 | 30 | 10 |
| 251 | 515,22746 | 515,22756 | 0,19 | 28 | 34 | 9 |
| 252 | 515,26413 | 515,26395 | -0,35 | 29 | 38 | 8 |
| 253 | 517,13427 | 517,13405 | -0,43 | 25 | 24 | 12 |
| 254 | 517,17047 | 517,17044 | -0,06 | 26 | 28 | 11 |
| 255 | 517,20683 | 517,20682 | -0,02 | 27 | 32 | 10 |
| 256 | 517,24318 | 517,24321 | 0,06 | 28 | 36 | 9 |
| 257 | 517,27953 | 517,27960 | 0,14 | 29 | 40 | 8 |
| 258 | 519,11368 | 519,11332 | -0,69 | 24 | 22 | 13 |
| 259 | 519,14982 | 519,14970 | -0,23 | 25 | 26 | 12 |
| 260 | 519,18606 | 519,18609 | 0,06 | 26 | 30 | 11 |
| 261 | 519,22248 | 519,22247 | -0,02 | 27 | 34 | 10 |
| 262 | 519,25884 | 519,25886 | 0,04 | 28 | 38 | 9 |
| 263 | 521,09248 | 521,09258 | 0,19 | 23 | 20 | 14 |
| 264 | 521,12896 | 521,12897 | 0,02 | 24 | 24 | 13 |
| 265 | 521,16586 | 521,16535 | -0,98 | 25 | 28 | 12 |
| 266 | 521,20209 | 521,20174 | -0,67 | 26 | 32 | 11 |
| 267 | 521,23841 | 521,23812 | -0,56 | 27 | 36 | 10 |
| 268 | 521,25915 | 521,25925 | 0,19 | 24 | 40 | 12 |
| 269 | 523,10951 | 523,10823 | -2,45 | 23 | 22 | 14 |
| 270 | 523,14510 | 523,14462 | -0,92 | 24 | 26 | 13 |
| 271 | 523,18113 | 523,18100 | -0,25 | 25 | 30 | 12 |
| 272 | 523,21757 | 523,21739 | -0,34 | 26 | 34 | 11 |
| 273 | 525,12430 | 525,12388 | -0,8 | 23 | 24 | 14 |
| 274 | 525,16078 | 525,16027 | -0,97 | 24 | 28 | 13 |
| 275 | 525,19827 | 525,19665 | -3,08 | 25 | 32 | 12 |
| 276 | 525,21333 | 525,21191 | -2,7 | 29 | 32 | 9 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 277 | 525,24989 | 525,24830 | -3,03 | 30 | 36 | 8 |
| 278 | 525,29167 | 525,29055 | -2,13 | 24 | 44 | 12 |
| 279 | 527,15534 | 527,15479 | -1,04 | 27 | 26 | 11 |
| 280 | 527,19211 | 527,19117 | -1,78 | 28 | 30 | 10 |
| 281 | 527,22854 | 527,22756 | -1,86 | 29 | 34 | 9 |
| 282 | 527,26300 | 527,26395 | 1,8 | 30 | 38 | 8 |
| 283 | 529,13376 | 529,13405 | 0,55 | 26 | 24 | 12 |
| 284 | 529,17101 | 529,17044 | -1,08 | 27 | 28 | 11 |
| 285 | 529,20730 | 529,20682 | -0,91 | 28 | 32 | 10 |
| 286 | 529,24323 | 529,24321 | -0,04 | 29 | 36 | 9 |
| 287 | 529,27869 | 529,27960 | 1,72 | 30 | 40 | 8 |
| 288 | 531,11442 | 531,11332 | -2,07 | 25 | 22 | 13 |
| 289 | 531,15010 | 531,14970 | -0,75 | 26 | 26 | 12 |
| 290 | 531,18664 | 531,18609 | -1,04 | 27 | 30 | 11 |
| 291 | 531,22296 | 531,22247 | -0,92 | 28 | 34 | 10 |
| 292 | 531,26012 | 531,25886 | -2,37 | 29 | 38 | 9 |
| 293 | 533,13024 | 533,12897 | -2,38 | 25 | 24 | 13 |
| 294 | 533,16628 | 533,16535 | -1,74 | 26 | 28 | 12 |
| 295 | 533,20258 | 533,20174 | -1,58 | 27 | 32 | 11 |
| 296 | 533,23918 | 533,23812 | -1,99 | 28 | 36 | 10 |
| 297 | 533,27459 | 533,27451 | -0,15 | 29 | 40 | 9 |
| 298 | 535,10793 | 535,10823 | 0,56 | 24 | 22 | 14 |
| 299 | 535,14548 | 535,14462 | -1,61 | 25 | 26 | 13 |
| 300 | 535,18169 | 535,18100 | -1,29 | 26 | 30 | 12 |
| 301 | 535,21832 | 535,21739 | -1,74 | 27 | 34 | 11 |
| 302 | 535,25532 | 535,25377 | -2,9 | 28 | 38 | 10 |
| 303 | 537,12430 | 537,12388 | -0,78 | 24 | 24 | 14 |
| 304 | 537,16085 | 537,16027 | -1,08 | 25 | 28 | 13 |
| 305 | 537,19693 | 537,19665 | -0,52 | 26 | 32 | 12 |
| 306 | 537,23294 | 537,23304 | 0,19 | 27 | 36 | 11 |
| 307 | 539,10272 | 539,10315 | 0,8 | 23 | 22 | 15 |
| 308 | 539,13962 | 539,13953 | -0,17 | 24 | 26 | 14 |
| 309 | 539,17695 | 539,17592 | -1,91 | 25 | 30 | 13 |
| 310 | 539,21414 | 539,21230 | -3,41 | 26 | 34 | 12 |
| 311 | 539,22775 | 539,22756 | -0,35 | 30 | 34 | 9 |
| 312 | 539,26374 | 539,26395 | 0,39 | 31 | 38 | 8 |
| 313 | 541,17032 | 541,17044 | 0,22 | 28 | 28 | 11 |
| 314 | 541,20747 | 541,20682 | -1,2 | 29 | 32 | 10 |
| 315 | 541,24339 | 541,24321 | -0,33 | 30 | 36 | 9 |
| 316 | 541,28011 | 541,27960 | -0,94 | 31 | 40 | 8 |
| 317 | 543,15041 | 543,14970 | -1,31 | 27 | 26 | 12 |
| 318 | 543,18662 | 543,18609 | -0,98 | 28 | 30 | 11 |
| 319 | 543,22312 | 543,22247 | -1,2 | 29 | 34 | 10 |
| 320 | 543,25982 | 543,25886 | -1,77 | 30 | 38 | 9 |
| 321 | 543,29657 | 543,29524 | -2,45 | 31 | 42 | 8 |
| 322 | 545,13033 | 545,12897 | -2,49 | 26 | 24 | 13 |
| 323 | 545,16683 | 545,16535 | -2,71 | 27 | 28 | 12 |
| 324 | 545,20296 | 545,20174 | -2,24 | 28 | 32 | 11 |
| 325 | 545,23951 | 545,23812 | -2,55 | 29 | 36 | 10 |
| 326 | 545,27471 | 545,27451 | -0,37 | 30 | 40 | 9 |
| 327 | 545,29011 | 545,28976 | -0,64 | 34 | 40 | 6 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 328 | 547,11005 | 547,10823 | -3,33 | 25 | 22 | 14 |
| 329 | 547,14564 | 547,14462 | -1,86 | 26 | 26 | 13 |
| 330 | 547,18189 | 547,18100 | -1,63 | 27 | 30 | 12 |
| 331 | 547,21781 | 547,21739 | -0,77 | 28 | 34 | 11 |
| 332 | 547,25477 | 547,25377 | -1,83 | 29 | 38 | 10 |
| 333 | 547,29101 | 547,29016 | -1,55 | 30 | 42 | 9 |
| 334 | 549,08862 | 549,08750 | -2,04 | 24 | 20 | 15 |
| 335 | 549,12464 | 549,12388 | -1,38 | 25 | 24 | 14 |
| 336 | 549,16067 | 549,16027 | -0,73 | 26 | 28 | 13 |
| 337 | 549,19750 | 549,19665 | -1,55 | 27 | 32 | 12 |
| 338 | 549,23387 | 549,23304 | -1,51 | 28 | 36 | 11 |
| 339 | 549,26958 | 549,26942 | -0,29 | 29 | 40 | 10 |
| 340 | 551,10194 | 551,10315 | 2,2 | 24 | 22 | 15 |
| 341 | 551,13953 | 551,13953 | 0 | 25 | 26 | 14 |
| 342 | 551,17638 | 551,17592 | -0,83 | 26 | 30 | 13 |
| 343 | 551,21312 | 551,21230 | -1,49 | 27 | 34 | 12 |
| 344 | 551,25021 | 551,24869 | -2,76 | 28 | 38 | 11 |
| 345 | 553,11897 | 553,11880 | -0,31 | 24 | 24 | 15 |
| 346 | 553,15550 | 553,15518 | -0,58 | 25 | 28 | 14 |
| 347 | 553,19217 | 553,19157 | -1,08 | 26 | 32 | 13 |
| 348 | 553,20730 | 553,20682 | -0,87 | 30 | 32 | 10 |
| 349 | 553,22778 | 553,22795 | 0,31 | 27 | 36 | 12 |
| 350 | 553,24354 | 553,24321 | -0,6 | 31 | 36 | 9 |
| 351 | 553,26390 | 553,26434 | 0,8 | 28 | 40 | 11 |
| 352 | 553,28027 | 553,27960 | -1,21 | 32 | 40 | 8 |
| 353 | 555,14861 | 555,14970 | 1,96 | 28 | 26 | 12 |
| 354 | 555,18670 | 555,18609 | -1,1 | 29 | 30 | 11 |
| 355 | 555,22287 | 555,22247 | -0,72 | 30 | 34 | 10 |
| 356 | 555,25934 | 555,25886 | -0,86 | 31 | 38 | 9 |
| 357 | 555,29546 | 555,29524 | -0,4 | 32 | 42 | 8 |
| 358 | 557,16619 | 557,16535 | -1,51 | 28 | 28 | 12 |
| 359 | 557,20252 | 557,20174 | -1,4 | 29 | 32 | 11 |
| 360 | 557,23876 | 557,23812 | -1,15 | 30 | 36 | 10 |
| 361 | 557,27509 | 557,27451 | -1,04 | 31 | 40 | 9 |
| 362 | 557,31059 | 557,31090 | 0,56 | 32 | 44 | 8 |
| 363 | 559,14624 | 559,14462 | -2,9 | 27 | 26 | 13 |
| 364 | 559,18204 | 559,18100 | -1,86 | 28 | 30 | 12 |
| 365 | 559,21890 | 559,21739 | -2,7 | 29 | 34 | 11 |
| 366 | 559,25526 | 559,25377 | -2,66 | 30 | 38 | 10 |
| 367 | 559,29190 | 559,29016 | -3,11 | 31 | 42 | 9 |
| 368 | 561,12579 | 561,12388 | -3,4 | 26 | 24 | 14 |
| 369 | 561,16193 | 561,16027 | -2,96 | 27 | 28 | 13 |
| 370 | 561,19840 | 561,19665 | -3,12 | 28 | 32 | 12 |
| 371 | 561,23516 | 561,23304 | -3,78 | 29 | 36 | 11 |
| 372 | 561,27149 | 561,26942 | -3,69 | 30 | 40 | 10 |
| 373 | 563,10396 | 563,10315 | -1,44 | 25 | 22 | 15 |
| 374 | 563,14031 | 563,13953 | -1,39 | 26 | 26 | 14 |
| 375 | 563,17699 | 563,17592 | -1,9 | 27 | 30 | 13 |
| 376 | 563,21347 | 563,21230 | -2,08 | 28 | 34 | 12 |
| 377 | 563,25054 | 563,24869 | -3,28 | 29 | 38 | 11 |
| 378 | 563,28464 | 563,28507 | 0,76 | 30 | 42 | 10 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 379 | 565,11782 | 565,11880 | 1,73 | 25 | 24 | 15 |
| 380 | 565,15630 | 565,15516 | -2,02 | 26 | 28 | 14 |
| 381 | 565,19280 | 565,19157 | -2,18 | 27 | 32 | 13 |
| 382 | 565,22903 | 565,22795 | -1,91 | 28 | 36 | 12 |
| 383 | 565,26533 | 565,26434 | -1,75 | 26 | 44 | 13 |
| 384 | 565,28679 | 565,28547 | -2,34 | 29 | 40 | 11 |
| 385 | 567,09784 | 567,09806 | 0,39 | 24 | 22 | 16 |
| 386 | 567,13473 | 567,13445 | -0,49 | 25 | 26 | 15 |
| 387 | 567,17112 | 567,17083 | -0,51 | 26 | 30 | 14 |
| 388 | 567,20802 | 567,20722 | -1,41 | 27 | 34 | 13 |
| 389 | 567,22236 | 567,22247 | 0,19 | 31 | 34 | 10 |
| 390 | 567,24435 | 567,24360 | -1,32 | 28 | 38 | 12 |
| 391 | 567,25812 | 567,25886 | 1,3 | 32 | 38 | 9 |
| 392 | 569,16570 | 569,16535 | -0,61 | 29 | 28 | 12 |
| 393 | 569,20197 | 569,20174 | -0,4 | 30 | 32 | 11 |
| 394 | 569,23903 | 569,23812 | -1,6 | 31 | 36 | 10 |
| 395 | 569,27504 | 569,27451 | -0,93 | 32 | 40 | 9 |
| 396 | 571,14362 | 571,14462 | 1,75 | 28 | 26 | 13 |
| 397 | 571,18162 | 571,18100 | -1,09 | 29 | 30 | 12 |
| 398 | 571,21806 | 571,21739 | -1,17 | 30 | 34 | 11 |
| 399 | 571,25458 | 571,25377 | -1,42 | 31 | 38 | 10 |
| 400 | 571,29089 | 571,29016 | -1,28 | 32 | 42 | 9 |
| 401 | 573,12262 | 573,12388 | 2,2 | 27 | 24 | 14 |
| 402 | 573,16101 | 573,16027 | -1,29 | 28 | 28 | 13 |
| 403 | 573,19766 | 573,19665 | -1,76 | 29 | 32 | 12 |
| 404 | 573,23382 | 573,23304 | -1,36 | 30 | 36 | 11 |
| 405 | 573,26983 | 573,26942 | -0,72 | 31 | 40 | 10 |
| 406 | 573,30635 | 573,30581 | -0,94 | 32 | 44 | 9 |
| 407 | 575,13984 | 575,13953 | -0,54 | 27 | 26 | 14 |
| 408 | 575,17730 | 575,17592 | -2,4 | 28 | 30 | 13 |
| 409 | 575,21358 | 575,21230 | -2,23 | 29 | 34 | 12 |
| 410 | 575,25001 | 575,24869 | -2,29 | 30 | 38 | 11 |
| 411 | 575,28617 | 575,28507 | -1,91 | 31 | 42 | 10 |
| 412 | 575,32184 | 575,32146 | -0,66 | 32 | 46 | 9 |
| 413 | 577,08381 | 577,08241 | -2,43 | 25 | 20 | 16 |
| 414 | 577,11871 | 577,11880 | 0,16 | 26 | 24 | 15 |
| 415 | 577,15671 | 577,15518 | -2,65 | 27 | 28 | 14 |
| 416 | 577,19275 | 577,19157 | -2,04 | 28 | 32 | 13 |
| 417 | 577,22946 | 577,22795 | -2,62 | 29 | 36 | 12 |
| 418 | 577,26537 | 577,26434 | -1,78 | 30 | 40 | 11 |
| 419 | 577,30022 | 577,30072 | 0,87 | 31 | 44 | 10 |
| 420 | 579,13479 | 579,13445 | -0,59 | 26 | 26 | 15 |
| 421 | 579,17217 | 579,17083 | -2,31 | 27 | 30 | 14 |
| 422 | 579,20901 | 579,20722 | -3,09 | 28 | 34 | 13 |
| 423 | 579,22957 | 579,22835 | -2,11 | 25 | 38 | 15 |
| 424 | 579,24544 | 579,24360 | -3,18 | 29 | 38 | 12 |
| 425 | 581,14974 | 581,15010 | 0,62 | 26 | 28 | 15 |
| 426 | 581,18616 | 581,18648 | 0,55 | 27 | 32 | 14 |
| 427 | 581,20139 | 581,20174 | 0,6 | 31 | 32 | 11 |
| 428 | 581,22408 | 581,22287 | -2,08 | 28 | 36 | 13 |
| 429 | 581,23839 | 581,23812 | -0,46 | 32 | 36 | 10 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 430 | 581,25875 | 581,25925 | 0,86 | 29 | 40 | 12 |
| 431 | 581,27367 | 581,27451 | 1,45 | 33 | 40 | 9 |
| 432 | 581,29416 | 581,29564 | 2,55 | 30 | 44 | 11 |
| 433 | 583,14535 | 583,14462 | -1,25 | 29 | 26 | 13 |
| 434 | 583,18254 | 583,18100 | -2,64 | 30 | 30 | 12 |
| 435 | 583,20355 | 583,20313 | -0,72 | 27 | 34 | 14 |
| 436 | 583,21814 | 583,21739 | -1,29 | 31 | 34 | 11 |
| 437 | 583,25485 | 583,25377 | -1,85 | 32 | 38 | 10 |
| 438 | 583,28889 | 583,29016 | 2,18 | 33 | 42 | 9 |
| 439 | 585,16116 | 585,16027 | -1,52 | 29 | 28 | 13 |
| 440 | 585,19810 | 585,19665 | -2,48 | 30 | 32 | 12 |
| 441 | 585,23446 | 585,23304 | -2,43 | 31 | 36 | 11 |
| 442 | 585,27116 | 585,26942 | -2,97 | 32 | 40 | 10 |
| 443 | 585,30574 | 585,30581 | 0,12 | 33 | 44 | 9 |
| 444 | 587,11861 | 587,11840 | -0,36 | 31 | 22 | 12 |
| 445 | 587,14058 | 587,13952 | -1,81 | 28 | 26 | 14 |
| 446 | 587,17671 | 587,17592 | -1,35 | 29 | 30 | 13 |
| 447 | 587,21343 | 587,21230 | -1,92 | 30 | 34 | 12 |
| 448 | 587,24992 | 587,24869 | -2,09 | 31 | 38 | 11 |
| 449 | 587,28570 | 587,28507 | -1,07 | 32 | 42 | 10 |
| 450 | 587,32168 | 587,32146 | -0,37 | 33 | 46 | 9 |
| 451 | 589,15645 | 589,15518 | -2,16 | 28 | 28 | 14 |
| 452 | 589,19300 | 589,19157 | -2,43 | 29 | 32 | 13 |
| 453 | 589,22949 | 589,22795 | -2,61 | 30 | 36 | 12 |
| 454 | 589,26553 | 589,26434 | -2,02 | 31 | 40 | 11 |
| 455 | 589,30020 | 589,30072 | 0,88 | 32 | 44 | 10 |
| 456 | 589,31508 | 589,31598 | 1,53 | 36 | 44 | 7 |
| 457 | 591,13625 | 591,13445 | -3,04 | 27 | 26 | 15 |
| 458 | 591,17221 | 591,17083 | -2,33 | 28 | 30 | 14 |
| 459 | 591,20852 | 591,20722 | -2,2 | 29 | 34 | 13 |
| 460 | 591,24503 | 591,24360 | -2,42 | 30 | 38 | 12 |
| 461 | 591,28141 | 591,27999 | -2,4 | 31 | 42 | 11 |
| 462 | 591,31638 | 591,31637 | -0,02 | 32 | 46 | 10 |
| 463 | 593,15127 | 593,15010 | -1,97 | 27 | 28 | 15 |
| 464 | 593,18780 | 593,18648 | -2,23 | 28 | 32 | 14 |
| 465 | 593,22372 | 593,22287 | -1,43 | 29 | 36 | 13 |
| 466 | 593,26065 | 593,25925 | -2,36 | 30 | 40 | 12 |
| 467 | 595,12838 | 595,12936 | 1,65 | 26 | 26 | 16 |
| 468 | 595,16610 | 595,16575 | -0,59 | 27 | 30 | 15 |
| 469 | 595,20296 | 595,20213 | -1,39 | 28 | 34 | 14 |
| 470 | 595,24011 | 595,23852 | -2,67 | 29 | 38 | 13 |
| 471 | 595,25475 | 595,25377 | -1,65 | 33 | 38 | 10 |
| 472 | 595,29106 | 595,29016 | -1,51 | 34 | 42 | 9 |
| 473 | 597,14508 | 597,14501 | -0,12 | 26 | 28 | 16 |
| 474 | 597,18310 | 597,18140 | -2,85 | 27 | 32 | 15 |
| 475 | 597,19761 | 597,19665 | -1,61 | 31 | 32 | 12 |
| 476 | 597,21819 | 597,21778 | -0,69 | 28 | 36 | 14 |
| 477 | 597,23416 | 597,23304 | -1,88 | 32 | 36 | 11 |
| 478 | 597,26880 | 597,26942 | 1,04 | 33 | 40 | 10 |
| 479 | 597,30616 | 597,30581 | -0,59 | 34 | 44 | 9 |
| 480 | 599,17644 | 599,17592 | -0,87 | 30 | 30 | 13 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 481 | 599,21351 | 599,21230 | -2,02 | 31 | 34 | 12 |
| 482 | 599,24976 | 599,24869 | -1,79 | 32 | 38 | 11 |
| 483 | 599,28603 | 599,28507 | -1,6 | 33 | 42 | 10 |
| 484 | 601,19331 | 601,19157 | -2,89 | 30 | 32 | 13 |
| 485 | 601,22963 | 601,22795 | -2,79 | 31 | 36 | 12 |
| 486 | 601,26624 | 601,26434 | -3,16 | 32 | 40 | 11 |
| 487 | 603,13578 | 603,13445 | -2,21 | 28 | 26 | 15 |
| 488 | 603,17236 | 603,17083 | -2,54 | 29 | 30 | 14 |
| 489 | 603,20866 | 603,20722 | -2,39 | 30 | 34 | 13 |
| 490 | 603,24512 | 603,24360 | -2,52 | 31 | 38 | 12 |
| 491 | 603,28114 | 603,27999 | -1,91 | 32 | 40 | 11 |
| 492 | 603,31841 | 603,31637 | -3,38 | 33 | 44 | 10 |
| 493 | 605,15160 | 605,15010 | -2,48 | 28 | 28 | 15 |
| 494 | 605,18814 | 605,18648 | -2,74 | 29 | 32 | 14 |
| 495 | 605,22446 | 605,22287 | -2,63 | 30 | 36 | 13 |
| 496 | 605,26105 | 605,25925 | -2,97 | 31 | 40 | 12 |
| 497 | 607,12946 | 607,12936 | -0,16 | 27 | 26 | 16 |
| 498 | 607,16680 | 607,16575 | -1,73 | 28 | 30 | 15 |
| 499 | 607,20291 | 607,20213 | -1,28 | 29 | 34 | 14 |
| 500 | 607,23938 | 607,23852 | -1,42 | 30 | 38 | 13 |
| 501 | 607,27474 | 607,27490 | 0,26 | 31 | 42 | 12 |
| 502 | 607,31231 | 607,31129 | -1,68 | 32 | 46 | 11 |
| 503 | 609,14328 | 609,14501 | 2,84 | 27 | 28 | 16 |
| 504 | 609,18237 | 609,18140 | -1,59 | 28 | 32 | 15 |
| 505 | 609,21871 | 609,21778 | -1,53 | 29 | 36 | 14 |
| 506 | 611,16107 | 611,16066 | -0,67 | 27 | 30 | 16 |
| 507 | 611,21373 | 611,21230 | -2,34 | 32 | 34 | 12 |
| 508 | 611,25062 | 611,24869 | -3,16 | 33 | 38 | 11 |
| 509 | 611,28708 | 611,28507 | -3,29 | 34 | 42 | 10 |
| 510 | 613,15710 | 613,15518 | -3,13 | 30 | 28 | 14 |
| 511 | 613,19369 | 613,19157 | -3,46 | 31 | 32 | 13 |
| 512 | 613,22972 | 613,22795 | -2,89 | 32 | 36 | 12 |
| 513 | 615,17093 | 615,17083 | -0,16 | 30 | 30 | 14 |
| 514 | 615,20900 | 615,20722 | -2,89 | 31 | 34 | 13 |
| 515 | 615,24479 | 615,24360 | -1,93 | 32 | 38 | 12 |
| 516 | 615,28079 | 615,27999 | -1,3 | 33 | 42 | 11 |
| 517 | 617,18711 | 617,18648 | -1,02 | 30 | 32 | 14 |
| 518 | 617,22303 | 617,22287 | -0,26 | 31 | 36 | 13 |
| 519 | 617,25840 | 617,25925 | 1,38 | 32 | 40 | 12 |
| 520 | 617,29684 | 617,29564 | -1,94 | 33 | 44 | 11 |
| 521 | 619,16392 | 619,16575 | 2,96 | 29 | 30 | 15 |
| 522 | 619,20248 | 619,20213 | -0,57 | 30 | 34 | 14 |
| 523 | 619,23954 | 619,23852 | -1,65 | 31 | 38 | 13 |
| 524 | 619,27599 | 619,27490 | -1,76 | 32 | 42 | 12 |
| 525 | 619,31316 | 619,31129 | -3,02 | 33 | 46 | 11 |
| 526 | 621,18238 | 621,18140 | -1,58 | 29 | 32 | 15 |
| 527 | 621,21901 | 621,21778 | -1,98 | 30 | 36 | 14 |
| 528 | 621,25546 | 621,25417 | -2,08 | 31 | 40 | 13 |
| 529 | 621,29127 | 621,29055 | -1,16 | 32 | 44 | 12 |
| 530 | 623,18234 | 623,18140 | -1,51 | 29 | 34 | 15 |
| 531 | 623,21901 | 623,21778 | -1,97 | 30 | 38 | 14 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 532 | 623,25546 | 623,25417 | -2,07 | 31 | 42 | 13 |
| 533 | 623,29127 | 623,29055 | -1,16 | 32 | 46 | 12 |
| 534 | 625,19679 | 625,19705 | 0,42 | 29 | 36 | 15 |
| 535 | 627,20697 | 627,20722 | 0,4 | 32 | 34 | 13 |
| 536 | 627,24355 | 627,24360 | 0,08 | 33 | 38 | 12 |
| 537 | 627,27981 | 627,27999 | 0,29 | 34 | 42 | 11 |
| 538 | 629,18569 | 629,18648 | 1,26 | 31 | 32 | 14 |
| 539 | 629,22295 | 629,22287 | -0,13 | 32 | 36 | 13 |
| 540 | 629,25954 | 629,25925 | -0,46 | 33 | 40 | 12 |
| 541 | 629,29493 | 629,29564 | 1,13 | 34 | 44 | 11 |
| 542 | 631,16507 | 631,16575 | 1,08 | 30 | 30 | 15 |
| 543 | 631,20210 | 631,20213 | 0,05 | 31 | 34 | 14 |
| 544 | 631,23845 | 631,23852 | 0,11 | 32 | 38 | 13 |
| 545 | 631,27504 | 631,27490 | -0,22 | 33 | 42 | 12 |
| 546 | 631,31125 | 631,31129 | 0,06 | 34 | 46 | 11 |
| 547 | 633,18151 | 633,18140 | -0,17 | 30 | 32 | 15 |
| 548 | 633,21794 | 633,21778 | -0,25 | 31 | 36 | 14 |
| 549 | 633,25463 | 633,25417 | -0,73 | 32 | 40 | 13 |
| 550 | 633,28964 | 633,29055 | 1,44 | 33 | 44 | 12 |
| 551 | 635,19707 | 635,19705 | -0,03 | 30 | 34 | 15 |
| 552 | 635,23357 | 635,23343 | -0,22 | 31 | 38 | 14 |
| 553 | 635,26882 | 635,26981 | 1,56 | 32 | 42 | 13 |
| 554 | 637,21097 | 637,21270 | 2,71 | 30 | 36 | 15 |
| 555 | 641,18505 | 641,18648 | 2,23 | 32 | 32 | 14 |
| 556 | 641,22181 | 641,22287 | 1,65 | 33 | 36 | 13 |
| 557 | 641,25860 | 641,25925 | 1,01 | 34 | 40 | 12 |
| 558 | 641,29380 | 641,29564 | 2,87 | 35 | 44 | 11 |
| 559 | 643,16556 | 643,16575 | 0,3 | 31 | 30 | 15 |
| 560 | 643,20264 | 643,20213 | -0,79 | 32 | 34 | 14 |
| 561 | 643,23783 | 643,23852 | 1,07 | 33 | 38 | 13 |
| 562 | 643,27428 | 643,27490 | 0,96 | 34 | 42 | 12 |
| 563 | 643,31246 | 643,31129 | -1,82 | 35 | 46 | 11 |
| 564 | 643,34644 | 643,34767 | 1,91 | 36 | 50 | 10 |
| 565 | 645,18303 | 645,18140 | -2,53 | 31 | 32 | 15 |
| 566 | 645,21829 | 645,21778 | -0,79 | 32 | 36 | 14 |
| 567 | 645,25538 | 645,25417 | -1,88 | 33 | 40 | 13 |
| 568 | 645,29055 | 645,29055 | 0 | 34 | 44 | 12 |
| 569 | 647,19806 | 647,19704 | -1,58 | 31 | 34 | 15 |
| 570 | 647,23523 | 647,23343 | -2,78 | 32 | 38 | 14 |
| 571 | 647,27220 | 647,26982 | -3,68 | 33 | 42 | 13 |
| 572 | 649,17570 | 649,17631 | 0,94 | 30 | 32 | 16 |
| 573 | 649,21404 | 649,21270 | -2,06 | 31 | 36 | 15 |
| 574 | 649,25128 | 649,24908 | -3,39 | 32 | 40 | 14 |
| 575 | 651,19075 | 651,19196 | 1,86 | 30 | 34 | 16 |
| 576 | 651,22930 | 651,22835 | -1,46 | 31 | 38 | 15 |
| 577 | 651,26566 | 651,26473 | -1,43 | 32 | 42 | 14 |
| 578 | 653,33046 | 653,33202 | 2,39 | 37 | 48 | 10 |
| 579 | 655,20084 | 655,20213 | 1,97 | 33 | 34 | 14 |
| 580 | 655,23730 | 655,23851 | 1,85 | 34 | 38 | 13 |
| 581 | 655,27478 | 655,27490 | 0,18 | 35 | 42 | 12 |
| 582 | 655,31179 | 655,31129 | -0,76 | 36 | 46 | 11 |

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| count | exp | theory | error [ppm] | C | H (M) | O |
| 583 | 657,18263 | 657,18140 | -1,87 | 32 | 32 | 15 |
| 584 | 657,22030 | 657,21778 | -3,83 | 33 | 36 | 14 |
| 585 | 657,25644 | 657,25417 | -3,45 | 34 | 40 | 13 |
| 586 | 657,29314 | 657,29055 | -3,94 | 35 | 44 | 12 |
| 587 | 659,19792 | 659,19705 | -1,32 | 32 | 34 | 15 |
| 588 | 659,23479 | 659,23343 | -2,06 | 33 | 38 | 14 |
| 589 | 659,27151 | 659,26982 | -2,56 | 34 | 42 | 13 |
| 590 | 659,30782 | 659,30620 | -2,46 | 35 | 46 | 12 |
| 591 | 661,17492 | 661,17631 | 2,1 | 31 | 32 | 16 |
| 592 | 661,21209 | 661,21270 | 0,92 | 32 | 36 | 15 |
| 593 | 661,24844 | 661,24908 | 0,97 | 33 | 40 | 14 |
| 594 | 661,28545 | 661,28547 | 0,03 | 34 | 44 | 13 |
| 595 | 663,19194 | 663,19196 | 0,03 | 31 | 34 | 16 |
| 596 | 663,22682 | 663,22835 | 2,31 | 32 | 38 | 15 |
| 597 | 663,26385 | 663,26473 | 1,33 | 33 | 42 | 14 |
| 598 | 665,20809 | 665,20761 | -0,72 | 31 | 36 | 16 |
| 599 | 665,24396 | 665,24400 | 0,06 | 32 | 40 | 15 |
| 600 | 665,28262 | 665,28098 | -2,47 | 33 | 44 | 14 |
| 601 | 669,25322 | 669,25417 | 1,42 | 35 | 40 | 13 |
| 602 | 669,29176 | 669,29055 | -1,81 | 36 | 44 | 12 |
| 603 | 671,19723 | 671,19705 | -0,27 | 33 | 34 | 15 |
| 604 | 671,23379 | 671,23343 | -0,54 | 34 | 38 | 14 |
| 605 | 671,27049 | 671,26982 | -1 | 35 | 42 | 13 |
| 606 | 673,21289 | 673,21270 | -0,28 | 33 | 36 | 15 |
| 607 | 673,24986 | 673,24908 | -1,16 | 34 | 40 | 14 |
| 608 | 673,28672 | 673,28547 | -1,86 | 35 | 44 | 13 |
| 609 | 673,32329 | 673,32185 | -2,14 | 36 | 48 | 12 |
| 610 | 675,19279 | 675,19196 | -1,23 | 32 | 34 | 16 |
| 611 | 675,22942 | 675,22835 | -1,58 | 33 | 38 | 15 |
| 612 | 675,26504 | 675,26473 | -0,46 | 34 | 42 | 14 |
| 613 | 675,30185 | 675,30112 | -1,08 | 35 | 46 | 13 |

# 9. State-of-the art of NOM structural analysis and future research needs

The outstanding level of intricacy in the characterization and structural analysis of NOM, and the lack of even meaningful concepts to describe the chemical structure of an operationally defined and highly complex mixture, severely limit our knowledge about origin, structure and function of NOM (Hedges et al., 2000; Hedges and Oades, 1997). Key research issues are the development of NOM isolation procedures, the quantification of analytical data and the avancement of concepts defining the chemical structure of complex, inseparable chemical mixtures.

## 9.1. ISOLATION AND FRACTIONATION OF NOM

A fundamental dilemma of NOM isolation is caused by the strong interaction of organic NOM constituents with the mineral phase in almost any realistic natural environment. The large surface area of the interacting organo-mineral complexes (Figure 4) results in a tight binding between the organic and the inorganic phase even in the absence of covalent bonds, and causes substantial obstacles in the isolation of mineral-free organic NOM materials. Soft, selective extraction procedures provide rather authentic materials, which however may not be overly representative because of the combined effects of limited yield and (chemo) selectivity. The isolation of higher-yield materials typically requires rather harsh extraction conditions and the obtained NOM materials may deviate from the original composition. This challenge will remain an incentive for the improvement of existing NOM isolation protocols also in the future. The pronounced polydispersity and molecular heterogeneity of NOM (Figure 9) will pose independent obstacles and challenges, because those properties imply, that NOM physicochemical characteristics are represented by a near continuous distribution of values. Those distinct features induce a variance in physical and chemical behavior of NOM and these will affect NOM separation and isolation techniques. Standardized, or at least well documented, experimental protocols of NOM isolation and purification, as well as the availability of reference NOM materials from a wide range of terrestrial, limnic and marine ecosystems, will remain indispensable for any further development and assessment of NOM isolation methods. Solid phase extraction with specialized and adapted phases (Kaiser et al., 2003), and reverse osmosis, show promise to assist in NOM isolation from the aqueous phase.

Physical fractionation results in NOM materials, which are less altered from their original configuration than those obtained from chemical fractionation (Dickens et al., 2006; Christl et al., 2000). For instance, drastic pH changes during the IHSS isolation procedure of humic and fulvic acids will cause chemical reactions of certain chemical bonds within NOM (e.g. hydrolysis of esters, decarboxylation of acids, substitution); even during storage of NOM in alcoholic solvents (ethanol, methanol, propanol), self-esterification of NOM was observed by FTICR mass spectrometry (McIntyre and McRae, 2005). Physical fractionation of NOM is feasible according to density, size (GPC, CGE, sieving), charge to mass ratio (electrophoresis: CZE, FFE) in a variety of solvents, with or without the addition of auxiliary substances. The nature of alterations within a range of closely related NOM fractions can be determined via in-depth structural analysis and linked to molecular composition and structure (Simpson et al., 2004b). Any variance within these NOM fractions is a response on a incremental change of "environmental" properties (Namjesnik and Cabaniss, 2004) and a thorough molecular-resolution analysis will assist in advancing a wealth of NOM research topics. On the molecular level of NOM analysis, we rate the identification of covalent intra- and inter-linkages of fundamental NOM constituents as a high priority issue.

Van Krevelen diagrams, derived from FTICR mass spectra of a range of NOM samples, show a loss of the signature of the constituent fundamental biomolecules (Figure 6) and, consequently indicate strong covalent binding between these classes of molecules. But also within a class of biomolecules (e.g. carbohydrates), certain linkages of pronounced stability against enzymatic degradation could play very important roles in defining key characteristics of NOM structure and reactivity, like the recalcitrance of NOM against microbial degradation.

## 9.2 QUANTIFICATION OF DATA FROM ANALYTICAL CHARACTERIZATION

Quantification remains one of the key unsatisfactorily resolved research issues in the analysis of NOM at all hierarchical levels (elements, fragments, molecules, interactions). A near continuum of differential responses arising from small and large molecules and strong and/or weak interactions within the NOM system (Figures 3, 11) determines the transfer functions from atom/molecule interactions to (instrument) response, which are complicated, non linear, and mostly unknown for any method. For that reason, an in-depth understanding of the

physics and chemistry of the respective technique is indispensable to enable a meaningful and quantitative interpretation of analytical data.

For *any* analytical technique to be used in the characterization of NOM (ranging from bulk characterization to molecular-level structural analysis), reliable means of quantification have to be established by rigorous standardization and testing procedures. Quantification is the single most important unresolved issue in the analytical chemistry of NOM and will remain as such in many years to come. Analogous considerations apply to the generation of clearly reproducible data, which are one step short of reliable quantification, however still of immense value for the precise determination of NOM structure and reactivity. The considerable variability of NOM, depending on origin and isolation, will necessitate the establishment of adapted and verified experimental protocols to characterize NOM structure and reactivity. Each of these may have to refer to a specific and limited range of application and validity.

Bias from over-interpreted data sets can be recognized and addressed by mutual calibration of complementary methods (Diallo et al., 2003; Kögel-Knabner, 2000; Poirer et al., 2005). Future research concerning structural analysis of NOM will emphasize molecular level characterization of extended NOM substructures by means of NMR and MS spectrometry. Crucial auxiliary data from high-performance NOM fractionation and separation will aid in this process. Any structural model of NOM will have to conform to the multiple restrictions induced by complementary bulk analysis and organic structural spectroscopy (Diallo et al., 2003). An exhaustive mathematical analysis of the sizable sets of complementary data will be indispensable to extract and to quantify the relative contributions and the significance of the features, which are used to describe NOM structure and interactions. NOM research will benefit from a broad range of interdisciplinary cooperation combining experimental methodology and ways of addressing scientific issues from the fields of chemistry, physics, biology, geology and mathematics.

From an analytical chemistry point of view, we attribute a substantial potential to several groups of methods, which will alone, or in combination, considerably enhance our knowledge about NOM structure and interactions with molecular-level resolution.

## 9.3. ORGANIC STRUCTURAL SPECTROSCOPY

The two most significant and most highly resolving methods of organic structural spectroscopy, namely magnetic resonance spectroscopy (NMR) (Fan et al., 2000; Hertkorn et al., 2002b; Preston, 1996; Simpson et al., 2002) and mass spectrometry (MS) (Kujawinski, 2002; Kujawinski et al., 2002) already have contributed immensely to the understanding of molecular level structural features of NOM. Future research concerning structural analysis of NOM will also emphasize molecular level characterization of extended NOM substructures by means of NMR and MS techniques. The application bandwidth of various other methods of organic structural organic spectroscopy, like XANES, UV/VIS and infrared spectroscopy is more restricted and will focus on the characterization of specific chemical environments in NOM, like the e.g. functional group analysis (carbonyl-derivatives, aromatics, heterocycles). These complementary data will be valuable for validation purposes and assessment of the transfer functions between molecular processes and the analytical (instrument) response; this knowledge is indispensable to determine in general the quantification potential of any analytical method.

### *9.3.1. NMR spectroscopy:*

***advantages***

*unsurpassed molecular level structural information about short range molecular order;*
*structural analysis and reactivity studies;*
*isotope-specific;*
*non-destructive;*
*very extensive information content even in spectra of medium resolution for "small" nuclei because of plausible correspondence between chemical shift and structure;*
*quantitative relationship between number of spins and NMR signal area;*
*many 1,2,3D NMR experiments are feasible from a single sample to reveal spatial proximity and covalent bonds;*

***weakness***

*relative insensitivity compared with other analytical techniques;*
*complex physics of intermolecular NOM interactions may interfere with direct relationship between chemical shift and molecular structure;*

NMR spectroscopy provides unsurpassed molecular level structural information about short range molecular order in NOM. Quantitative one-dimensional NMR spectra of the main nuclei ($^{1}H$, $^{13}C$, $^{15}N$, $^{31}P$) which provide content and considerable structural detail of fundamental constituents, serve as indispensable frame to constrain any structural model of NOM. Higher dimensional NMR spectra act as filters, which emphasize specific forms of binding (Hertkorn et al., 2002b) (another factor that contributes to increase signal dispersion). Cross peaks in multiple dimensions indicate connectivities (i.e. chemical bonds or spatial proximity) rather than individual atoms (Fan et al., 2000; Hertkorn et al., 2002b; Simpson et al., 2003; Simpson et al., 2002); thereby, the reliability of NMR resonance assignment is enhanced considerably.

The low resolution in NMR spectra of NOM is primarily caused by a very massive overlap of closely spaced NMR resonances (of variable line-width). Owing to a restricted range of chemical shift and enhanced susceptibility to line broadening via strong dipole – dipole interaction, resonance overlap is especially severe in proton NMR spectra of NOM. Line broadening in $^{13}C$ NMR spectra of NOM is at present typically introduced by extensive line broadening functions which are employed, when spectra with "pleasing" signal-to-noise ratio are computed from FIDs; the intrinsic linewidth of many $^{13}C$ NMR resonances in NOM is well in the low Hertz range. Near identity of chemical shift not necessarily implies similarity of chemical structures, as widely different NOM constituents frequently show overlap in their chemical shift values (Hertkorn et al., 2002b).

The dramatic increase in NMR sensitivity and resolution, which is the consequence of various recent developments of instrumentation, like high-field magnets (Chung and Kroon, 2003; Nagai et al., 2001), cryogenic (Hinse et al., 2003) and micro (Schlotterbeck et al., 2002; Walton et al., 2003) probes, all work together to enhance the quality and significance of NMR analysis of NOM. Minute (low μg-range) amounts of NOM (e.g. like available in high-resolution fractionation studies) are amenable to proton NMR characterization in a variety of solvents, thereby addressing issues of association and weak (non covalent) interactions. Cryogenic probes allow the acquisition of higher dimensional NMR spectra of NOM at rather low concentration, when relaxation characteristics of NOM favor enhanced resolution of cross peaks. Proton detected higher dimensional NMR spectra are available from 1-2 mg of NOM, and cryogenic probes with $^{13}C$ detection capability (Keun et al., 2002) will lead to $^{13}C$ NMR spectra of NOM with drastically increased resolution. The unique capability to generate and

analyze data from multiple NMR experiments obtained from a single NOM sample enables an assessment of the significance and authenticity of individual spectra. This also refers to the potential of NMR spectrometry when it is used as reference for other, complementary analytical data.

### *9.3.2. Capillary electrophoresis*

***advantages***

*electrophoretic mobility carries structure-specific information, which can be adapted to a wide range of experimental conditions to probe NOM size, shape, charge characteristics and reactivity;*
*sensitive and versatile suite of methods;*
*can be highly automated;*
*non-destructive;*

***weakness***

*limited structure specific information about short range molecular order;*

The simplicity, high efficiency, selectivity, extreme separation capacity – especially for ionisable molecules – and a relatively low cost have all contributed to the popularity of capillary electrophoresis in environmental science and in metabonomics, proteomics and genomics. Minute amounts of complex samples can be analysed with a variety of structure specific detection systems, such-as (laser-induced) fluorescence, UV/VIS, radioisotope or mass selective detection. CE separation is orthogonal to classical LC-MS techniques and can be adjusted to a wide range of experimental conditions (pH, ionic strength, addition of reactands; experimental variants: CGE, CEC) to provide charge, size and shape distributions obtained from NOM, which take into account intermolecular interactions. In this respect, CE complements NMR information about primary chemical structures (covalent bonds) with conclusions on the corresponding secondary and tertiary structure.

The electrophoretic NOM fractionation, when up-scaled to a preparative level by means of free flow electrophoresis (FFE), is capable to obtain sufficient material for even higher dimensional NMR (and other spectroscopic characterization) and for binding studies. Classical NOM fractionation schemes, like XAD-8 based elution schemes and size-exclusion chromatography, can be used in comparison as complementary separation tools in NOM analysis.

### *9.3.3. High resolution mass spectrometry:*

***advantages***

*unsurpassed combination of resolution and sensitivity;*
*structural analysis and reactivity studies;*
*molecular level structural information of ionisable compounds;*
*molecular formulae from thousands of compounds in a mixture;*

***weakness***

*molecular level structural information is available from ionisable compounds ONLY;*
*differentiation of isomers requires additional experiments (mass spectrometry: fragmentation; separation: retention time, mobility);*
*the ionization efficiency of individual compounds in a mixture may vary widely and very strongly depends on the experimental conditions;*

The ultrahigh mass resolving power, mass resolution and mass accuracy of Fourier transform cyclotron resonance mass spectrometry (FTICR MS) already has been applied to NOM characterization, yielding unprecedented detail in depicting molecular formulae of NOM constituents and opening up entirely novel interpretation capabilities (Kujawinski, 2002; Kujawinski et al., 2002). The ability to resolve and determine molecular formulae of several thousands of components in a complex mixture with limited or without prior wet chemical separation (Kujawinski, 2002) has the potential to change entirely the approach to dealing with chemical and biological complexity in general and with NOM structural analysis in particular. The visualization of FTICR mass spectral analysis with van Krevelen diagrams (Visser 1983, Kim et al., 2003) allows to (1) identify NOM reaction pathways (alterations of NOM structures caused by chemical and biological means) and (2) qualitatively analyze NOM constituent classes that comprise ultrahigh-resolution spectra. Intelligent visualization schemes of extremely well resolved NOM mass spectra will contribute to display an immense wealth of information in an agreeable and informative way. FTICR mass spectral data will produce unprecedented detail information about amino acid and carbohydrate sequences, allowing much more precise assessments of NOM diagenesis than available today. Selective ionization modi, like electron capture dissociation (Cooper et al., 2003), show potential to assign protein-carbohydrate or intra- carbohydrate linkages in NOM; this knowledge will clarify the long-standing issue of recalcitrance of these very important NOM constituents.

Polyphenols, aliphatics, and "exotics", such as oxygenated polycyclic hydrocarbons originating from black carbon, and yet unknowns will provide FTICR MS signatures, which will lead to much higher precision in the description of NOM structure and diagenesis.

#### *9.3.3.1. MS hyphenation with separation techniques*

Owing to its high detection sensitivity, mass spectrometry of NOM can be favorably combined with a high performance separation technique like capillary electrophoresis (CE with mass-selective detection). The highly complementary data set of a combined CE/MS or even NMR + CE/MS characterization of NOM produces an exceeding wealth of information and very massive constraints of feasible NOM molecular structures if mathematical data analysis is performed on these highly complementary data sets (cf. Figure 7 and case study 4). All NMR, MS and CE methods are highly versatile and capable to address a very broad range of topics dealing with the analysis of NOM backbone, functional groups and interactions.
Further miniaturization of separation and detection devices, and especially the coupling with ultrahigh-resolution FTICR mass spectrometry, will provide extreme highly resolved data sets with unsurpassed information about ionisable NOM constituents. A wide range of ionization techniques (electrospray, ESI; chemical ionization, CI; photoionization, PI; and others, all in either positive or/and negative mode) is available to ionize a large fraction of NOM constituents under specifically adapted conditions.

## 9.4. REDEFINING MOLECULAR LEVEL ANALYSIS BY COMBINATION OF CHEMICAL CHARACTERIZATIONS OF NOM AND HIGH PERFORMANCE ANALYTICAL CHEMISTRY

Degradative approaches, which lead to identifiable small molecules like carbohydrates (Aluwihare et al., 2002; Benner and Kaiser, 2003; Jahnel et al., 1998; Kaiser and Benner, 2000; Repeta et al., 2002), lipids and amino acids (Jahnel et al., 1998; Vernonclark, 1994; Yamashita and Tanoue, 2003), only account for less than a quarter of total organic carbon and nitrogen in NOM (Benner, 2002; Hedges and Oades, 1997), while more precisely defined and representative bulk parameters (Benner et al., 1997; Benner et al., 1992) reveal only limited structural information (Hedges, 2002). Small molecules are unambiguously identifiable even at very low concentrations; however, they have lost crucial linkage information.

Selective chemical reactions, like mild hydrolysis, reduction, oxidation and derivatization (Frimmel, 1998; Hertkorn et al., 2002a) of NOM materials and fractions (chapter 4.1.), will result in products with favorable properties for analytical characterization. When soft and selective organic and biochemical reactions are applied to NOM, larger fragments will be obtained, which are amenable to high resolution structural analysis. These molecules will retain linkages with positional and stereo chemical information, which is certainly superior to the knowledge derived from a quantitative analysis of small degradation products of NOM, and of immense value for the assessment of NOM synthesis and degradation.

Chemical transformation of functional groups with NMR- and MS-recognizable labels will enable an advanced functional group analysis of NOM according to structural rather than to behavioral characteristics. Biodegradation studies of NOM and suitable, preferentially stable isotope labeled, precursors will allow to elucidate details of diagenesis and fate of NOM under specific and controlled conditions. The issues of stereochemistry and stable isotope composition of NOM constituents will continue to gain importance in assessing NOM origin and diagenesis. Any progress in the determination of position specific stable isotope composition (e.g. by NMR and MS methods) will be of high significance to further advance this field. Fractionation and physical and chemical separation of NOM will greatly assist in these studies; further miniaturization of will enhance separation capacity and high-resolution fractionation of NOM.

## 9.5. DEVELOPMENT OF CONCEPTS TO DESCRIBE CHEMICAL STRUCTURES OF MIXTURES AND TO ASSESS THEIR SIGNIFICANCE

Theoretically well-founded approaches for a numerical description of the complex, polydisperse and irregular structure of NOM are missing at present. This fact significantly impedes understanding of NOM structure and any application of quantitative structure-activity relationships (QSAR) in modeling their properties. Novel approaches for the quantitative analysis and the numerical description of the NOM structure suitable for a *quantitative* determination of elemental, fragmental and molecular weight composition and predictive modeling have to be developed, based on related constitutive descriptors of various hierarchical levels of molecular organization (e.g. elements, fragments, molecules). This novel combination of analytical chemistry and mathematical data analysis will not only

improve our knowledge about NOM structure and reactivity; any combination of topics 9.1. – 9.5. will lead to strong synergetic effects and further advance of the understanding of any complex natural and living system, whose properties and functioning depend both on strong and weak, non-covalent interactions.

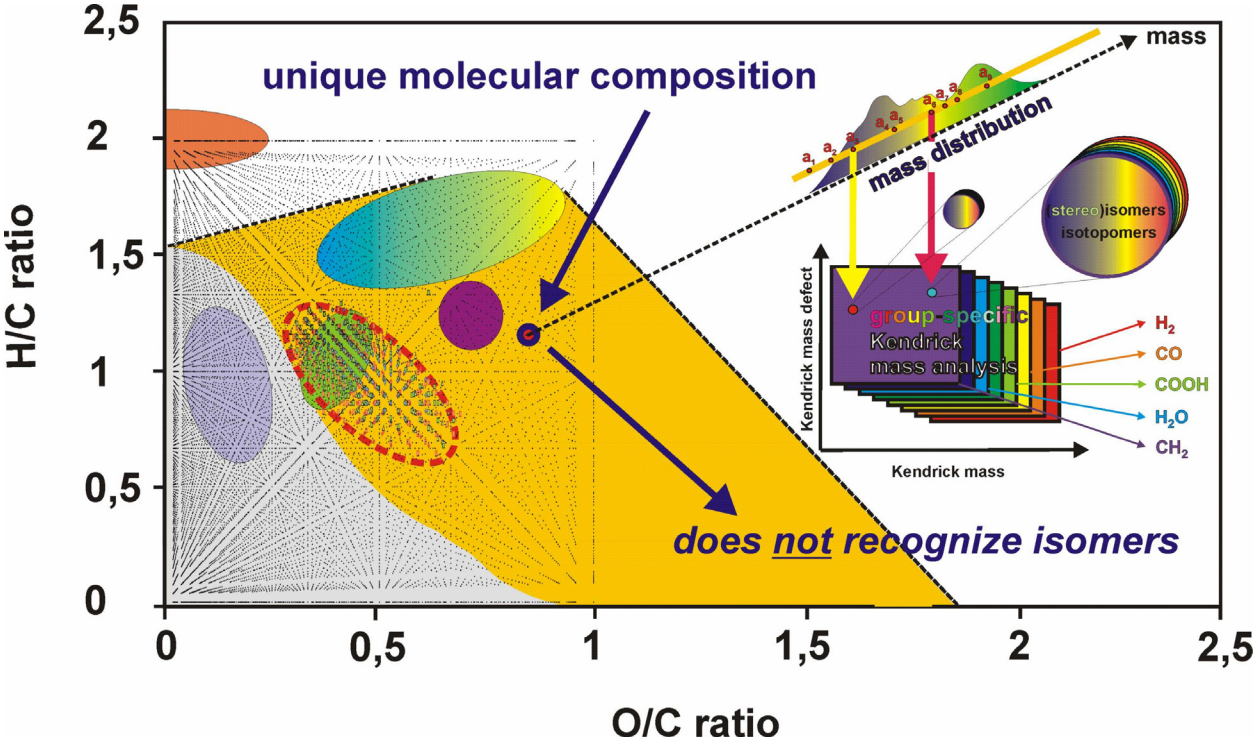



**Figure 91**. Integrated view of the NOM compositional space. The compositional space of NOM is quantized: only specific combinations of $C_nH_mO_q$ elemental compositions of NOM are feasible. This refers to genuine molecules $C_nH_mO_q$ and to molecules $C_nH_mO_qZ_\Sigma$ containing any combination of additional heteroatoms; these molecules would project into the H/C vs. O/C plane in a van-Krevelen diagram. The spacing in between the feasible combinations of atoms declines with increasing mass and can be probed by FTICR mass spectrometry. A combination of alternative ionization methods and high-field FTICR mass spectrometry indicates, that within a fairly large area, all feasible molecular compositions $C_nH_mO_q$ are actually present in e.g. marine UDOM (Figure 89). In van-Krevelen plots of NOM, which are derived from high-quality FTICR mass spectra, the signature of the fairly well known major NOM constituents of biochemical origin, like lipids (orange), peptides and carbohydrates (blue to yellow), nucleic acids (purple), lignins (green), polycyclic aromatics (light violet) and (non-biogenic) black carbon is lost. This indicates, that covalent bonds between these different compound classes in NOM are the rule rather than the exception. Any covalent bond between different classes of NOM precursors will at first produce new compounds with a compositional space in the vicinity of the connecting graphs. CRAM occupy a central position within the van Krevelen plot, and any covalent incorporation of CRAM into other NOM constituent classes will close existing gaps in the accessible compositional space most effectively. Similar considerations apply to reaction products of black carbon (BC), which covers a very large $C_nH_mO_q$ compositional space of unsaturation and oxidation. In this respect, both CRAM and BC most likely are the most important contributors to compositional variation of molecules $C_nH_mO_q$ within the van-Krevelen compositional space of NOM. Van Krevelen diagrams as well as Kendrick mass defect analyses do not distinguish isomers; this level of intricacy in the molecular-level analysis of NOM has to rely on data from separation techniques, MS fragmentation studies, and NMR spectroscopy.

The very substantial progress in the molecular-level characterization of NOM within the last years and the foreseeable improvements of nascent technology and concepts certainly offer exciting prospects for vast advancement in the molecular-level resolution understanding of very complex natural systems in general and of NOM in particular, already in the near future.

## 10. Glossary

| | |
|---|---|
| AAS | atom absorption spectroscopy; |
| AC | abiotic condensation model; |
| AES, | atom emission spectroscopy; |
| $B_0$ | external magnetic field (NMR); |
| BC | black carbon; |
| BD | biodegradation model; |
| BSTFA | *N*,*O*-Bis(trimethylsilyl)trifluoroacetamide; |
| CE | capillary electrophoresis; |
| CEC | capillary electrochromatography; |
| $C_f$ | functionalised aliphatic carbon: functional group, but no heteroatom is directly attached to carbon; |
| CGE | capillary gel electrophoresis; |
| CI | chemical ionization; |
| COSY | correlated spectroscopy; |
| CRAM | carboxyl-rich alicyclic molecules; |
| UDOM | ultrafiltered dissolved organic matter; |
| CZE | capillary zone electrophoresis; |
| D | dipolar coupling constant (NMR); |
| DBE | double bond equivalent; |
| DEPT | distortionless enhancement by polarization transfer; |
| $\Delta\nu$ | linewidth at half of the maximum intensity (NMR); |
| DOC | dissolved organic carbon; |
| ESI | electrospray ionization; |
| ESR | electron spin resonance; |
| F1 | vertical frequency axis in 2D NMR spectra (X nucleus calculated data); |
| F2 | horizontal frequency axis in 2D NMR spectra (proton detected data); |
| FA | fulvid acid; |
| FFE | free-flow electrophoresis; |
| FID | free induction decay; |
| FTICR MS | Fourier-transform ion cyclotron resonance mass spectrometry; |

| | |
|---|---|
| FTICR | Fourier-transform ion cyclotron resonance; |
| GARP | globally optimized alternating-phase rectangular pulses; |
| $\gamma_N$ | gyromagnetic ratio (NMR); |
| γs | gradient enhanced; |
| HA | humic acid; |
| HETCOR | heteronuclear 2D NMR spectroscopy; |
| HMBC | heteronuclear multiple bond correlation; |
| HMQC | heteronuclear multiple quantum correlation; |
| HOESY | heteronuclear Overhauser spectroscopy; |
| HPS | heteropolysaccharide; |
| HS | humic substance; |
| HSQC | heteronuclear single quantum coherence; |
| I | isotope spin quantum number (NMR); |
| IHSS | International Humic Substances Society; |
| INEPT | insensitive nuclei enhancement by polarization transfer; |
| IUPAC | International Union of Pure and Applied Chemistry; |
| J | scalar coupling constant (NMR); |
| JRES | J-resolved spectroscopy; |
| LFER | linear free energy relationships; |
| MALDI | matrix assisted laser desorption ionization; |
| MAXY, | maximum quantum spectroscopy; |
| MINTEQ | chemical equilibrium model for the calculation of metal speciation; |
| MLEV, | composite pulse decoupling sequence, named after Levitt, M.H.; |
| MS | mass spectrometry; |
| NICCA | non-ideal consistent competitive model; |
| NMR | nuclear magnetic resonance; |
| nOe | nuclear Overhauser effect (NMR); |
| NOM | natural organic matter; |
| PI | photoionization; |
| Q | quadrupole moment; |
| QSAR | quantitative structure-activity relationships; |

| | |
|---|---|
| QUAT | quaternary (carbon) only pulse sequence; |
| RMSE | mean square error; |
| S/N | signal-to-noise ratio; |
| SEC | size exclusion chromatography; |
| HPLC, | high-performance liquid chromatography; |
| SPARIA | substituent patterns in aromatic rings by increment analysis; |
| T | Tesla (magnetic field strength unit); |
| $T_1$ | spin-lattice relaxation time (NMR); |
| $T_2$ | spin,spin or transverse relaxation time (NMR); |
| TOCSY | total correlation spectroscopy; |
| UV | ultraviolet; |
| UV/VIS | ultraviolet-visible spectroscopy; |
| WALTZ | wideband alternating-phase low-power technique for zero residual splitting; |
| WHAM | Windermere humic aqueous model; |
| X – nucleus | any NMR active nucleus except protons; |
| XANES | X ray absorption near edge spectroscopy; |
| $\eta$ | asymmetry parameter; |

## 11. Acknowledgements

I sincerely thank Professor Antonius Kettrup with his warm and extremely generous personality for his influential suggestion to work on the characterization of natural organic matter – quite a transition from organometallic chemistry with colourful crystals and sharply defined molecular and physical properties. This, initially timid step transformed itself into one of the biggest eye- and door openers in my entire life and I owe immense gratitude for the unlimited freedom and support, which I consider necessary for the development of such work from scratch; I am aware that this kind of generosity is very rarely found.

Special thanks go also to my good friend Mike Perdue, who has provided me with so many crucial insights, based on his unique knowledge of the fundamentals of physical chemistry and his ability to apply these principles at elevated levels to complex systems. Mike always has condensed my rambling phantasies into meaningful experiments and evaluation schemes. Furthermore, I still admire his supreme mastery of PC computing, which among many other things, put SPARIA on the air – without him, "no way in hell"; even not any trace of this wonderful name in chemistry. Similarly, the seamless integration of metal ion solution chemistry and NMR is credited to Mike´s patience and open-mindedness.

John Hedges, who left us early, was inspirational beyond comprehension, as a scientist and a human being. He has opened to many and myself a whole universe of science and contacts with his personal example. The long anticipated bridge from the terrestrial to the freshwater and marine world brought to us a wealth of indispensable new ideas in a very casual way, and so many groups, we have the privilege to collaborate with by now, originated from that cradle – "beers for John in your world" – I think you can see us here, creeping arkwardly to small insights and honors. I thank you so much for giving me so many new friends in this world, many of them I consider as role models in science and other aspects of my life. Dear Ron and Karl, I still hope for a joint sampling expedition – I cherish your positive and determined attitude in research and the fun capacity in life.

Our great mama Irina Perminova, who opened the Russian Empire to us, provided me with wings to fly out of our caves into a colourful humic world of cultural and scientific world class, not available to us by any other means. I sincerely hope for a good future of our German-Russian friendship for the years to come.

Philippe Schmitt-Kopplin, our ingenious darth vader of the committees and supreme boss-master, handily supplies us with the huge realm of the so versatile and interesting yocto-macro-separation techniques “and mass spectrometry”. Most importantly, without ANY trace of “petty” jealousy, he always ensures a hospitable and relaxed atmosphere of friendship for the whole group to thrive. Eva Holzmann and Silvia Thaller have to cope with me on a daily basis which has to be fun indeed. I thank both of you so immensely much for your sunny temper, your utmost reliability and for the timeliness and accuracy of all your versatile work (and for reminding me of the importance of the Northern Munich hinterland, which is often forgotten on the way to the South). Without the delightful competence of Heinz Frisch, we could have “nevuh” taken off “our way”. I sincerely thank the whole group for such a balanced and fine atmosphere of excitement, fun and inspiration. Gratitude also goes to Sigurd Schulte-Hostede for unlimited support, consultation and encouragement.

Our group is very grateful for generous funding from various sources, at first to the GSF and the Institute of Ecological Chemistry, which provide world-class research environment, the German Ministry of Research BMBF, the NSF/DAAD (PPP), the European Union, the International Humic Substances Society IHSS, the German Research Council DFG, the Alexander von Humboldt-Foundation AvH, the German-Israeli-Foundation GIF, and various other contributors. I personally thank Eva Holzmann, Heidi Neumeir and Silvia Thaller for skilful technical assistance in experiments and the preparation of this thesis. Samples of NOM/HS from different ecosystems have been generously provided by the groups of I. V. Perminova (Lomonosov Moscow State University, Russian Federation), J. I. Hedges (University of Washington, Seattle, WA), Ph. Schmitt-Kopplin, J. Junkers (GSF), with sample inputs from A. Günzl (silylation) and F. Jäkle (methylation), M. Wolf (GSF), R. Benner (University of South Carolina, Columbia, SC), R. Artinger (KfK, Karlsruhe, Germany), E. M. Perdue (GeorgiaTech, Atlanta, GA), M. de Nobili (University of Udine, Italy), Y. Chen (The Hebrew University of Jerusalem, Israel), G. R. Aiken (USGS, Boulder, USA), T. Dittmar (NMHFL Tallahasse and University of Florida). We sincerely thank M. Witt (Bruker Daltonics) for skilful guidance in the acquisition of FTICR mass spectra. The experience, gained during work with these samples, has indispensably contributed to the development of the ideas introduced in this manuscript. The authors gratefully acknowledge valuable discussions with K. Whitehead, Th. Dittmar, and Y. Gelinas, and to many additional crucial partners.

## 12. References

Abbt-Braun, G., Lankes, U., and Frimmel, F.H. (2004) Structural characterization of aquatic humic substances - The need for a multiple method approach. *Aquatic Sciences,* **66**, 151-170.

Abraham, R.J., Mobli, M., and Smith, R.J. (2003) H-1 chemical shifts in NMR: Part 19. Carbonyl anisotropies and steric effects in aromatic aldehydes and ketones. *Magn. Reson. Chem.* **41**, 26-36.

Adhiya, J., Cai, X., Sayre, R.T., and Traina, S.J. (2002) Binding of aqueous cadmium by the lyophilized biomass of Chlamydomonas reinhardtii, *Colloids and Surfaces, A*. **210,** 1-11.

Allison, J.D., and Perdue, E.M. (1994) Modeling metal-humic interactions with MINTEQA2, in N. Senesi, T.M. Miano (eds.), *Humic Substances in the Global Environment and Implications on Human Health*, Elsevier, Amsterdam; pp. 927-942.

Aluwihare, L.I., Repeta, D.J., and Chen, R.F. (2002) Chemical composition and cycling of dissolved organic matter in the Mid-Atlantic Bight, *Deep-Sea Res. II.,* **49**, 4421-4437.

Aluwihare, L.I., Repeta, D.J., and Chen, R.F.A. (1997) A major biopolymeric component to dissolved organic carbon in surface sea water, *Nature,* **387**, 166-169.

Amon, R.M.W., and Benner, R. (1994) Rapid cycling of high-molecular-weight dissolved organic matter in the ocean, *Nature,* **396**, 549-552.

Amon, R.M.W. and Benner, R. (1996) Bacterial utilization of different size classes of dissolved organic matter. *Limnol. Oceanogr.* **41**, 41-51.

Anderson, S.J. (1997) Proton and $^{19}F$ NMR Spectroscopy of Pesticide Intermolecular Interactions, in M. Nanny, R. Minear, J. Leenheer (eds), *Nuclear Magnetic Resonance Spectroscopy in Environmental Chemistry,* pp. 51-71.

Anet, F. (2004) The place of metabolism in the origin of life. *Curr. Opin. Chem. Biol.,* **8**, 654-659.

Antunes, M.C., Simão, J.E., Duarte, A.C., Esteban, M., and Tauler, R. (2002) Application of multivariate curve resolution to the voltammetric study of the complexation of fulvic acids with cadmium(II) ion, *Anal. Chim. Acta.*, **459**, 291-304.

Antweiler, R.C. (1991) The hydrolysis of Suwannee River fulvic acid, in Baker, R.A., (ed.), *Organic Substances and Sediments in Water. Volume 1: Humics and Soils*; Lewis Publishers: Chelsea, MI., pp. 163-177.

Arias, M., Barral, M.T., and Mejuto, J.C. (2000) Enhancement of copper and cadmium adsorption on kaolin by the presence of humic acids, *Chemosphere*, **48**, 1081-1088.

Bain, A. (2003) Chemical exchange in NMR, *Prog. Nucl. Magn. Reson. Spec*., **43**, 63-103.

Balabane, M., and van Oort, F. (2002) Metal enrichment of particulate organic matter in arable soils with low metal contamination, *Soil Biol. Biochem.*, **34**, 1513-1516.

Balcke, G.U., Kulikova, N.A., Hesse, S., Kopinke, F.D., Perminova, I.V., and Frimmel, F.H. (2002) Adsorption of humic substances onto kaolin clay related to their structural features, *Soil Sci. Soc. Am. J.,* **66**, 1805-1812.

Baldock, J.A., Masiello, C.A., Gelinas, Y., and Hedges, J.I. (2004) Cycling and composition of organic matter in terrestrial and marine ecosystems. *Mar. Chem.* **92**, 39-64.

Barrientos, L.G., Louis, J.M., Ratner, D.M., Seeberger, P.H., and Gronenborn, A.M. (2003) Solution structure of a circular-permuted variant of the potent HIV-inactivating protein cyanovirin-N: Structural basis for protein stability and oligosaccharide interaction, *J. Mol. Biol.,* **32**, 211-223.

Bax, A., Ikura, M., Kay, L.E., Torchia, D.A., and Tschudin, R. (1990) Comparison of Different Modes of Two-Dimensional Reverse-Correlation NMR for the Study of Proteins, *J. Magn. Reson*., **86**, 304-318.

Beck, K.C., and Reuter, J.H. (1974) Organic and inorganic geochemistry of some coastal plain rivers of the southeastern United States, *Geochim. Cosmochim. Acta,* **38**, 341-364.

Becker, E.D. (2000) *High Resolution NMR. Theory and Chemical Applications*. Academic Press, San Diego, pp. 205-225.

Benedetti, M.F., Milne, C.J., Kinniburgh, D.G., Van Riemsdijk, W.H., and Koopal, L.K. (1995) Metal Ion Binding to Humic Substances: Application of the Non-Ideal Competitive Adsorption Model, *Environ. Sci. Technol*., **29**, 446-457.

Benedetti, M.F., Van Riemsdijk, W.H., and Koopal, L.K. (1996) Humic Substances Considered as a Heterogeneous Donnan Gel Phase, *Environ. Sci. Technol*., **30**, 1805-1813.

Benn, R., and Rufinska, A. (1986) Hochauflösende Metallkern-NMR-Spektroskopie von Organometallverbindungen, *Angew. Chem.,* **98,** 851-871.

Benner, R. (2002) in *Biogeochemistry of Marine Dissolved Organic Matter,* eds. Hansell, D. A. & Carlson, C.A. (Academic Press, Amsterdam, 2002), pp. 59-151.

Benner, R., and Kaiser, K. (2003) Abundance of amino sugars and peptidoglycan in marine particulate and dissolved organic matter, *Limnol. Oceanog.,* **48**, 118-128.

Benner, R., Biddanda, B., Black, B., and McCarthy, M. (1997) Abundance, size distribution, and stable carbon and nitrogen isotopic compositions of marine organic matter isolated by tangential-flow ultrafiltration, *Mar. Chem.,* **57**, 243-263.

Benner, R., Pakulski, J.D., McCarthy, M., Hedges, J.I. and Hatcher, P.G. (1992) Bulk chemical characterization of dissolved organic matter in the ocean. *Science* **255**, 1561-1564.

Bhanoori, M., and Venkateswerlu, G. (2000) In vivo chitin–cadmium complexation in cell wall of Neurospora crassa, *Biochim. Biophys. Acta*., **1523**, 21-28.

Bloom, P.R., and Leenheer, J.A. (1989) Vibrational, Electronic, High-energy Spectroscopic Methods for Characterizing Humic Substances, in: M.H.B., Hayes, P., Mac Carthy, R.L.,Malcolm, R.S., Swift, (eds.), *Humic Substances II, In Search of Structure*, Wiley, Chichester, pp. 411-446.

Boon, J.J., Klap, V.A., and Eglinton, T.I. (1998) Molecular characterization of microgram amounts of oceanic colloidal organic matter by direct temperature-resolved ammonia chemical ionization mass spectrometry, *Org. Geochem.,* **29**, 1051-1061.

Bortiatynski, J.M., Hatcher, P.G., and Knicker, H. (1996) NMR techniques (C, N, and H) in studies of humic substances. J.S. Gaffney, N.A. Marley, S.B. Clark (eds.) *Humic and Fulvic Acids, American Chemical Society Symposium Series 651*, American Chemical Society: Washington DC, pp. 55-77.

Bovey, F.A., and Mirau, P.A. (1996) *NMR of Polymers*, Academic Press, San Diego, pp. 353-378.

Bowles E.C., Antweiler R.C., and MacCarthy, P. (1989) Acid-base titrations and hydrolysis of fulvic acid from the Suwannee River, in Averett, R.C. *et al.* (eds.), *Humic Substances in the Suwannee River, GA: Interaction, Properties, and Proposed Structures;* USGS Open File Report 87-557, pp. 209-229.

Brady, B., and Pagenkopf, G.K. (1978) Cadmium complexation by soil fulvic acid, *Can. J. Chem.*, **56**, 2331-2336.

Breitmaier, E., and Voelter, W. (1990) *Carbon-13 NMR Spectroscopy*, *High-Resolution Methods and Applications in Organic Chemistry and Biochemistry*, VCH-Verlagsgesellschaft, Weinheim.

Brindle, J.T., Antti, H., Holmes, E., Tranter, G., Nicholson, J.K., Bethell, H.W.L., Clarke, S., Schofield, P.M., McKilligin, E., Mosedale, D.E. and Grainger, D.J. (2002) Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using H-1-NMR-based metabonomics. *Nature Medicine* **8**, 1439-1445.

Brunner, E. (1998) Enhancement of Surface and Biological Magnetic Resonance Using Laser-Polarized Noble Gases, *Concept. Magnetic Res.,* **11,** 313-335.

Buddrus, J., and Lambert, J. (1995) Isolated paraffinic methyl groups in humic substances, *Org. Geochem*., **23,** 269-271.

Buddrus, J., Burba, P., Herzog, P.H., and Lambert, J. (1989) Quantification of Partial Structures of Aquatic Humic Substances by One- and Two-Dimensional Solution $^{13}C$ Nuclear Magnetic Resonance Spectroscopy, *Anal. Chem*., **89,** 628-631.

Burdon, J. (2001) Are the Traditional Concepts of the Structures of Humic Substances Realistic? *Soil Sci*., **166**, 752-769.

Cade-Menun, B.J., Liu, C.W., Nunlist, R., and McColl, J.G. (2002) Soil and litter phosphorus-31 nuclear magnetic resonance spectroscopy: Extractants, metals, and phosphorus relaxation times, *J. Environ. Qual.,* **31**, 457-465.

Campos-Olivas, R., Aziz, R., Helms, G.L., Evans, J.N.S., and Gronenborn, A.M. (2002) Placement of $^{19}F$ into the centre of GB1: effects on structure and stability, *FEBS Letters* **517**, 55-60.

Carrilho, E.N., Ferreira, A.G., and Gilbert, T.R. (2002) Characterization of Sorption Sites on Pilayella littoralis and Metal Binding Assessment Using $^{113}Cd$ and $^{27}Al$ Nuclear Magnetic Resonance, *Environ. Sci. Technol*., **36**, 2003-2007.

Case, D.A. (2002) Molecular Dynamics and NMR Spin Relaxation in Proteins, *Acc. Chem. Res.,* **35**, 325-331.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G., and Skelton, N.J. (1996) *Protein NMR Spectroscopy, Principles and Practice*, Academic Press, London.

Chien, Y.Y., and Bleam, W.F. (1998) Two-Dimensional NOESY Nuclear Magnetic Resonance Study of pH-Dependent Changes in Humic Acid Conformation in Aqueous Solution, *Environ. Sci. Technol*., **32**, 3653-3658.

Chin, W.C., Orellana, M.V., and Verdugo, P. (1998) Spontaneous assembly of marine dissolved organic matter into polymer gels, *Nature,* **391**, 568-572.

Christl I., Knicker H., Kögel-Knabner I., and Kretzschmar R. (2000) Chemical heterogeneity of humic substances: characterization of size fractions obtained by hollow-fibre ultrafiltration. *Eur. J. Soil Sci.* **51**, 617-625.

Chung, K.H., and Moon, C.H. (1994) Selective binding of $Cd^{2+}$ by natural and synthetic organic macromolecules investigated by $^{113}Cd$ NMR spectroscopy, *Environ. Technol*., **15**, 795-800.

Chung, K.H., and Moon, C.H. (1996) Cadmium-113 nuclear magnetic resonance studies of cadmium(II)carboxylate complexes in aqueous solution, *J. Chem. Soc. Dalton Trans.,* 75-78.

Chung, K.H., Rhee, S.W., Shin, H.S., and Moon, C.H. (1996) Probe of cadmium(II)binding on soil fulvic acid investigated by $^{113}Cd$ NMR spectroscopy, *Can. J. Chem*., **74**, 1360-1365.

Claudio, E.S., Horst, M.A., Forde, C.E., Stern, Ch.L., Zart, M.K., and Godwin, H.A. (2000) $^{207}Pb$-$^{1}H$ Two-Dimensional NMR Spectroscopy: A Useful New Tool for Probing Lead (II) Coordination Chemistry, *Inorg. Chem*., **39**, 1391-1397.

Clore, G.M., and Gronenborn, A. M. (1998) Determining the structures of large proteins and protein complexes by NMR, *Tibtech* **16**, 22-34.

Cook, R.L. (2004) Coupling NMR to NOM, *Anal. Bioanal. Chem.,* **378**, 1484-1503.

Cook, R.L., and Langford, C.H. (1998) Structural characterization of a fulvic acid and a humic acid using solid state ramp-CP-MAS C-13 nuclear magnetic resonance, *Environ. Sci. Technol.,* **32**, 719-725.

Crestini, C., and Argyropoulos, D.S. (1997) Structural analysis of wheat straw lignin by quantitative P-31 and 2D NMR spectroscopy. The occurrence of ester bonds and alpha-O-4 substructures, *J. Agric. Food Chem.,* **45**, 1212-1219.

Croasmun, W.R. and Carlson, R.M.K. (1996) *Two-Dimensional NMR Spectroscopy*, Wiley-VCH, Weinheim.

Crouch, R.E., Llanos, W., Mehr, K.G., Hadden, C.E., Russell, D.J., and Martin, G.E. (2001) Applications of cryogenic NMR probe technology to long-range $^{1}H$-$^{15}N$ 2D NMR studies at natural abundance, *Magn. Reson. Chem*., **39**, 555-558.

Dacey, J. W. H., and Wakeham, S. G. (1986) Oceanic dimethylsulfide: production during zooplankton grazing on phytoplankton, *Science,* **233**, 1314-1316.

De La Torre, M.C.A., and Tessier, A. (2002) Cadmium deposition and mobility in the sediments of an acidic oligotrophic lake, *Geochim. Cosmochim. Acta.,* **66**, 3549-3562.

De Leeuw, J.W., and Largeau, C. (1993) A review of macromolecular organic compounds that comprise living organisms and their role in kerogen, coal, and petroleum formation, in M.H. Engle, S. Macko (eds.), *Org. Geochem.*, Plenum, New York, pp. 23-72.

Derenne, S., and Largeau, C. (2001) A Review of Some Important Families of Refractory Macromolecules: Composition, Origin, and Fate in Soils and Sediments, *Soil Sci*. **166**, 833-847.

Desy, J.C., Amyot, M., Pinel-Alloul, B., and Campbell, P.G.C. (2002) Relating cadmium concentrations in three macrophyte-associated freshwater invertebrates to those in macrophytes, water and sediments, *Environ. Poll*., **120**, 759-769.

Diallo, M.S. (2003) 3-D structural modeling of humic acids through experimental characterization, computer assisted structure elucidation and atomistic simulations. 1. Chelsea soil humic acid, *Environ. Sci. Technol.,* **37**, 1783-1793.

Dickens A.F., Baldock J.A., Smernik R.J., Wakeham S.G., Arnarson T.S., Gelinas Y., and Hedges J.I. (2006) Solid-state $^{13}C$ NMR analysis of size and density fractions of marine sediments. *Geochim. Cosmochim. Acta*, in press.

Dickens, A.F., Gelinas, Y., Masiello, C.A., Wakeham, S., and Hedges, J.I. (2004) Reburial of fossil organic carbon in marine sediments, *Nature,* **427**, 336-339.

Dixon, A.M., Mai, M.A., and Larive, C.K. (1999) NMR Investigation of the Interactions between 4-Fluoro-1-acetonaphthone and the Suwannee River Fulvic Acid, *Environ. Sci. Technol., **33***, 958-964.

Dobbs, J.C., Susetyo, W., Carreira, L.A., and Azarraga, L.V. (1989) Competitive binding of protons and metal ions in humic substances by lanthanide ion probe spectroscopy, *Anal. Chem*., **61***,* 1519-1524.

Doi, Y. (1990) *Microbial Polyesters*, VCH Verlagsgesellschaft, Weinheim.

Dong, D., Nelson, Y.M., Lion, L.W., Shuler, M.L., and Ghiorse, W.C. (2000) Adsorption of Pb and Cd onto metal oxides and organic material in natural surface coatings as determined by selective extractions: new evidence for the importance of Mn and Fe oxides, *Wat. Res.*, **34**, 427-436.

Drexel, R.T., Haitzer, M., Ryan, J.N., Aiken, G.R., and Nagy, K.L. (2002) Mercury(II) sorption to two Florida Everglades peats: Evidence for strong and weak binding and competition by dissolved organic matter released from the peat, *Environ. Sci. Technol.,* **36**, 4058-4064.

Druffel, E.R.M., Williams, P.M., Bauer, J.E., and Ertel, J.R. (1992) Cycling of Dissolved and Particulate Organic Matter in the Open Ocean, *J. Geophys. Res.,* **97**, 15639-15659.

Eimers, M.C., Evans, R.D., and Welbourn, P.M. (2002) Partitioning and bioaccumulation of cadmium in artificial sediment systems: application of a stable isotope tracer technique, *Chemosphere,* **46**, 543-551.

Ellis, D.A., Martin, J.W., Muir, D.C.G. and Mabury, S.A. (2000) Development of an $^{19}F$ NMR Method for the Analysis of Fluorinated Acids in Environmental Water Samples, *Anal. Chem*., **72**, 726-731.

Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, Oxford.

Fan, T.W.-M., Higashi, R.M., and Lane, A.N. (2000) Chemical Characterization of a Chelator-Treated Soil Humate by Solution-State Multinuclear Two-Dimensional NMR with FTIR and Pyrolysis-GCMS, *Environ. Sci. Technol*., **34**, 1636-1646.

Flynn, P.F., Mattielo, D.L., Hill, H.D.W. and Wand, A.J. (2000) Optimal Use of Cryogenic Probe Technology in NMR Studies of Proteins, *J. Am. Chem. Soc*., **122**, 4823-4824.

Francioso, O., Ciavatta, C., Montecchio, D., Tugnoli, V., Sanchez-Cortes, S., and Gessa, C. (2003) Quantitative estimation of peat, brown coal and lignite humic acids using chemical parameters, H-1-NMR and DTA analyses, *Biores. Technol.,* **88**, 189-195.

Fründ, R., and Lüdeman, H.D. (1989) The quantitative analysis of solution- and CPMAS-C-13 NMR spectra of humic material, *Sci. Total Environ*., **81/82**, 157-168.

Gaffney, J.S., Marley, N.A., and Clark, S.B., (1996) *Humic and Fulvic Acids*, American Chemical Society Symposium Series 651, Washington.

Gao, H. (1994) *A New Approach to Fractionation of Natural Organic Matter - Using Ion Retardation Resin,* Ph.D. Dissertation, Georgia Institute of Technology.

Gardiner, J. (1974) The chemistry of cadmium in natural water; a study of cadmium complex formation using the cadmium specific-ion electrode, *Wat. Res.*, **8**, 23-30.

Gèlinas, Y., Baldock, J.A., and Hedges, J.I. (2001) Demineralization of marine and freshwater sediments for CP/MAS $^{13}C$ NMR analysis, *Org. Geochem*. **32**, 677-693.

Giovanela, M., Parlanti, E., Soriano-Sierra, E.J., Soldi, M.S., and Sierra, M.M.D. (2004) Elemental compositions, FT-IR spectra and thermal behavior of sedimentary fulvic and humic acids from aquatic and terrestrial environments. *Geochem. J.,* **38**, 255-264.

Glore, G.M., and Gronenborn, A.M. (1998) New methods of structure refinement for macromolecular structure determination by NMR, *Proc. Natl. Acad. Sci. USA* **95,** 5891-5898.

Grandy, D.W., Petrakis, L., Young, D.C., and Gates, B.C. (1984) Determination of oxygen functionalities in synthetic fuels by NMR of naturally abundant $^{17}O$, *Nature,* **308**, 175-177.

Grassi, M., and Mingazzini, M. (2001) $^{113}Cd$-NMR and Fluorescence Studies of the Interactions between Cd(II) and Extracellular Organic Matter Released by Selenastrum capricornutum, *Environ. Sci. Technol*., **35**, 4271-4276.

Gratias, R., Konat, R., Kessler, H., Crisma, M., Valle, G., Polese, A., Formaggio, F., Toniolo, C., Broxterman, Q.B., and Kamphuis, J. (1998) First step toward the quantitative identification of peptide 3(10)-helix conformation with NMR spectroscopy: NMR and X-ray diffraction structural analysis of a fully-developed 3(10)-helical peptide standard, *J. Am. Chem. Soc.,* **120**, 4763-4770.

Griffin, J.L., Walker, L.A., Troke, J., Osborn, D., Shore, R.F., and Nicholson, J.K. (2000) The initial pathogenesis of cadmium induced renal toxicity, *FEBS Letters*., **478**, 147-150.

Günther, H. (1992) *NMR-Spektroskopie*: Grundlagen, Konzepte und Anwendungen der Protonen- und Kohlenstoff-13 Kernresonanz-Spektroskopie in der Chemie, Thieme, Stuttgart. pp. 76-91.

Haiber, S., Burba, P., Herzog, H., and Lambert, J. (1999) Elucidation of aquatic humic partial structures by multistage ultrafiltration and two-dimensional nuclear magnetic resonance spectrometry, *Fresen. J. Anal. Chem.,* **364**, 215-218.

Haitzer, M., Aiken, G.R., and Ryan, J.N. (2002) Binding of mercury(II) to dissolved organic matter: The role of the mercury-to-DOM concentration ratio, *Environ. Sci. Technol.,* **36**, 3564-3570.

Hansell, D.A., and Carlson, C.A. (2002) *Biogeochemistry of Marine Dissolved Organic Matter*, Academic Press, Elsevier Science.

Harris, R.K., Becker, E.D., Cabral de Menzes, S.M., Goodfellow, R., and Granger, P. (2002) Nuclear Spin Properties and Conventions for Chemical Shifts, in D.M. Grant and R.K. Harris (eds.), *Enzyclopedia of Nuclear Magnetic Resonance* **9,** Wiley-VCH, pp. 5-18.

Harrison, P.G., Healy, M.A., and Steel, A.T. (1983) Lead-207 Chemical Shift Data for Bivalent Lead Compounds: Thermodynamics of the Equilibrium $Pb(O_2CCH_3)_2$; $[Pb(O_2CCH_3)]^+ + O_2CCH_3$ in Aqueous Solution in the Temperature Range 303-323 K, *J. Chem. Soc. Dalton Trans*, 1845-1848

Hatcher, P.G., and Clifford, D.J. (1997) The organic geochemistry of coal: from plant materials to coal, *Org. Geochem.,* **27**, 251-274.

Hatcher, P.G., and Spiker, E.C. (1988) Selective Degradation of Plant Biomolecules, in F.H. Frimmel, R.F. Christman (eds.), *Humic Substances and Their Role in the Environment*, Wiley, pp. 58-74.

Hatcher, P.G., Maciel, G.E., and Dennis, L.W. ( 1981) Aliphatic structure of humic acids; a clue to their origin. *Org. Geochem.* **3**, 43-48.

Hatcher, P.G., Rowan, R., and Mattingly M.A. (1979) $^{1}H$ and $^{13}C$ NMR of marine humic acids. *Org. Geochem.* **2,** 77-85.

Hayes, M.H.B., MacCarthy, P., Malcolm, R.L., and Swift, R.S. (1989) *Humic Substances II, In Search of Structure*, John Wiley, Chichester.

Hedges J.I., Eglinton, G., Hatcher, P.G., Kirchman, D.L., Arnosti, C., Derenne, S., Evershed, R.P., Kögel-Knabner, I., de Leeuw J.W. and Littke, R. (2000) The molecularly-uncharacterized component of nonliving organic matter in natural environments. *Org. Geochem.* **31**, 945-958.

Hedges, I.J., and Keil, G.G. (1999) Organic geochemical perspectives on estuarine processes: sorption reactions and consequences, *Mar. Chem*., **65**, 55-65.

Hedges, I.J., and Oades, J.M. (1997) Comparative organic geochemistries of soils and marine sediments, *Org. Geochem*., **27**, 319-361.

Hedges, I.J., Hatcher, P.G., Ertel, J.R., and Meyers-Schulte, K.J. (1992). A comparison of dissolved humic substances from seawater with Amazon River counterparts by $^{13}C$-NMR spectrometry, *Geochim. Cosmochim. Acta.,* **56***,* 1753-1757.

Hedges, J.I. (1988) Polymerization of Humic Substances in Natural Environments, in F.H. Frimmel, R.F. Christman (eds.), *Humic Substances and Their Role in the Environment*, Wiley, pp. 45-58.

Hedges, J.I. (1992) Global biogeochemical cycles: progress and problems, *Mar. Chem.,* **39**, 67-93.

Hedges, J.I. (2002) Sedimentary Organic Matter Preservation and Atmospheric O2 Regulation, *unpublished manuscript.*

Hedges, J.I. (2002) Why Dissolved Organics Matter, in D.A. Hansell and C.A. Carlson (eds.), *Biogeochemistry of Marine Dissolved Organic Matter*, Academic Press, San Diego, pp. 1-33.

Hedges, J.I., Baldock, J.A., Gélinas, Y., Lee, C., Peterson, M., and Wakeham, S.G. (2001) Evidence for non-selective preservation of organic matter in sinking marine particles, *Nature,* **409**, 801-804.

Hedges, J.I., Baldock, J.A., Gelinas, Y., Lee, C., Peterson, M.L., and Wakeham, S.G. (2002) The biochemical and elemental compositions of marine plankton: A NMR perspective, *Mar. Chem.,* **78**, 47-63.

Hedges, J.I., Eglinton, G., Hatcher, P.G., Kirchman, D.L., Arnosti, C., Derenne, S., Evershed, R.P., Kögel-Knabner, I., deLeeuw, J.W., Littke, R., Michaelis, W., and Rullkötter, J. (2000) The molecularly-uncharacterized component of nonliving organic matter in natural environments, *Org. Geochem*., **31**, 945-958.

Heikkinen, S., Toikka, M.M., Karhunen, P.T., and Kilpelainen, I.A. (2003) Quantitative 2D HSQC (Q-HSQC) via suppression of J-dependence of polarization transfer in NMR spectroscopy: Application to wood lignin. *J. Am. Chem. Soc.* **125**, 4362-4367.

Helgaker, T., Jaszunski, M., and Ruud, K. (1999) Ab initio methods for the calculation of NMR shielding and indirect spin-spin coupling constants, *Chem. Rev.,* **99**, 293-352.

Hemingway, R.W., and Karchesy, J.J. (1989) *Chemistry and Significance of Condensed Tannins*, Plenum Press. New York.

Henderson, T.J. (2004) Sensitivity-Enhanced Quantitative $^{13}C$ NMR Spectroscopy via Cancellation of $^{1}JCH$ Dependence in DEPT Polarization Transfers. *J. Am. Chem. Soc*. **126**, 3682-3683.

Hertkorn, N. (2005) Molecular level structural analysis of natural organic matter and of humic substances by multinuclear and higher dimensional NMR spectroscopy; in: I.V. Perminova, N. Hertkorn, and K. Hatfield (2005) *Use of humic substances to remediate polluted environments: From theory to practice*. Kluver Academic Publishers, pp. 391-435.

Hertkorn, N. (2005) unpublished results.

Hertkorn, N. (2006) $^{17}O$ NMR spectroscopy of natural organic matter, *Anal. Chem.* (in preparation).

Hertkorn, N. (2006) Differential responses of substructures in the $^{1}J$ (C,H) correlation NMR spectra of a terrestrial humic substance, *Anal. Chem.* (in preparation).

Hertkorn, N., Claus H., Schmitt-Kopplin Ph., Perdue E.M., and Filip Z. (2002) Utilization and transformation of aquatic humic substances by autochthonous microorganisms, *Environ. Sci. Technol.,* **36**, 4334-4345.

Hertkorn, N., Günzl, A., Freitag, D., and Kettrup, A. (2002) Nuclear Magnetic Resonance Spectroscopy Investigations of Silylated Refractory Organic Substances, in F.H. Frimmel, G. Abbt-Braun, K.G. Heumann, B. Hock, H.-D. Lüdemann and M. Spiteller (eds.), *Refractory Organic Substances in the Environment*, Wiley-VCH, Weinheim, pp. 129-145.

Hertkorn, N., Günzl, A., Wang, C., Freitag, D., and Kettrup, A. (1996) The Role of Humic Substances in the Ecosystems and in Environmental Protection, NMR Investigations of Silylated Humic Substances, in J. Drozd, S.S. Gonet, N.Senesi, J. Weber (eds.), *Proceedings of the 8th Meeting of the International Humic Substances Society*, Wroclaw, Poland, September 9-14, pp. 139-146.

Hertkorn, N., Kovalevskii, D., Perminova, I.V., Schmitt-Kopplin, Ph., Permin, A., Petrosyan, V., and Kettrup, A. (2004) One and two-dimensional proton NMR methods for the identification of labile and non exchangeable protons in humic substances, *Anal. Chem.*, (in preparation).

Hertkorn, N., Perdue, E. M., Schmitt-Kopplin, Ph., and Kettrup, A. (2004) A detailed analysis of aromatic substitution in humic substances by 2D MNR spectroscopy and increment analysis, *Anal. Chem.* (in preparation).

Hertkorn, N., Perdue, E.M., and Kettrup, A. (2004) The binding of Suwannee River organic matter to cadmium. A potentiometric and $^{113}Cd$ NMR study at two different magnetic field strengths, *Anal. Chem.*, **76**, 6327-6341.

Hertkorn, N., Permin, A., Perminova, I., Kovalevskii, D., Yudov, M., Petrosyan, V., and Kettrup, A. (2002) Comparative Analysis of Partial Structures of a Peat Humic and Fulvic Acid Using One- and Two-Dimensional Nuclear Magnetic Resonance Spectroscopy, *J. Environ. Qual.,* **31**, 375-387.

Hertkorn, N., Schmitt-Kopplin, Ph., Artinger, R., Buckau, G., Schäfer, T., Geyer, S., Bleam, W.F., and Jacobson, Ch. (2006) Combined $^{1}H/^{13}C$ NMR, carbon/sulfur K-edge XANES and capillary electrophoresis study on the characterization and fate of aquatic fulvic acids, *Geochim. Cosmochim. Acta,* in preparation.

Hertkorn, N., Schmitt-Kopplin, Ph., Junkers, J., and Kettrup, A. (2004) Die strukturchemische Characterisierung von natürlicher organischer Materie, *GIT Laborfachzeitschrift,* **5**, 487-491.

Hertkorn, N., Schmitt-Kopplin, Ph., Perminova, I.V., Kovalevskii, D., and Kettrup, A. (2001) Two dimensional NMR spectroscopy of humic substances, in R.S. Swift, K.M. Spark (eds.), *Humic Substances Downunder, Understanding and Managing Organic Matter in Soils, Sediments and Waters*, Sept. 21-25, 1998, International Humic Substances Society, University of Adelaide, Australia, pp. 149-158.

Herzog, H., Burba, P., and Buddrus., J. (1996) Quantification of hydroxylic groups in a river humic substance by $^{29}$Si-NMR, *Fresen. J. Anal. Chem*., **354**, 375-377.

Hesse, M., Meier, H., Zeeh, B. (1991) Spektroskopische Methoden in der organischen Chemie, Thieme, Stuttgart, pp.156-157.

Heumann, K.G., Abbt-Braun, G., Behrens, K., Burba, P., Frimmel, F.H., Jakubowski, B., Knöchel, A., Mielcke, J., Rädlinger, G., Marx, G., and Vogl, J. (2002) Element Determination and its Quality Control in Fractions of Refractory Organic Substances and the Corresponding Original Water Samples, in F.H. Frimmel, G. Abbt-Braun, K.G. Heumann, B. Hock, H.-D. Lüdemann, and M. Spiteller (eds.), *Refractory Organic Substances in the Environment*, Wiley-VCH, Weinheim, pp. 39-53.

Hoch, J.C., and Stern, A.S. (1996) *NMR Data Processing*, Wiley-Liss, New York.

Holmes, E., Nicholson, J.K., and Tranter, G. (2001) Metabonomic characterization of genetic variations in toxicological and metabolic responses using probabilistic neural networks, *Chem. Res. Toxicol.,* **14**, 182-191.

Hong, M.F., Sudor, J., Stefansson, M., and Novotny, M.V. (1998) High resolution studies of hyaluronic acid mixtures through capillary gel electrophoresis, *Anal. Chem.,* **7,** 568-573.

Howe, R.F., Lu, X.Q., Hook, J., and Johnson, W.D. (1997) Reaction of aquatic humic substances with aluminium: a Al-27 NMR study, *Mar. Freshwater Res.,* **48**, 377-383.

Hsu, C.S., Qian, K., and Yungning, C. Chen. (1992) An innovative approach to data analysis in hydrocarbon characterization by on-line liquid chromatography-mass spectrometry, *Anal. Chim. Acta.,* **264**, 79-89.

Huang, S.L. (2001) Cadmium adsorption by sediment in a turbulence tank, *Wat. Res., **35***, 2635-2644.

Hughes, R.A., Robertson, M.P., Ellington, A.D., and Levy, M. (2004) The importance of prebiotic chemistry in the RNA World. *Curr. Opin. Chem. Biol.,* **8**, 629-633.

Hughey, C.A., Hendrickson, C.L., Rodgers, P.R., and Marshall, A.G. (2001) Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra, *Anal. Chem.,* **73**, 4676-4681.

Jacobsen, H.J., and Ellis, P.D. (1981) Solution structures of cadmium-glycine complexes probed by cadmium-113 NMR of supercooled aqueous solutions, *J. Phys. Chem.,* **85**, 3367-3369.

Jahnel, J.B., Ilieva, P., Abbt-Braun, G., and Frimmel, F.H. (1998) Aminosäuren und Kohlenhydrate als Strukturbestandteile von refraktären organischen Säuren. *Vom Wasser*, **90**, 205-216.

Jensen, C.F., Deshmukh, S., Jakobsen, H.J., Inners, R.R., and Ellis, P.D. (1981) Cadmium-113 nuclear magnetic resonance studies of cadmium-ethylenediaminetetraacetic acid complexes, *J. Am. Chem. Soc.*, **103**, 3659-3666.

John, J., Salbu, B., Gjessing, E.T., and Bjørnstad, H.E. (1988) Effect of pH, humus concentration and molecular weight on conditional stability constants of cadmium, *Wat. Res.*, **22**, 1381-1388.

Junkers, J., Schmitt-Kopplin, P., Munch, J. C., and Kettrup, A. (2002) Up-scaling capillary zone electrophoresis separations of polydisperse anionic polyelectrolytes with preparative free-flow electrophoresis exemplified with a soil fulvic acid. *Electrophoresis,* **23**, 2872-2879.

Kähling, H., and Robien, W. (1994) $^{17}$O NMR Spectroscopic Investigation of Steroids at Natural Abundance, *Magn. Reson. Chem*., **32**, 608-613.

Kalinowski H.-O., Berger S., and Braun S. (1984) *$^{13}$C-NMR-Spektroskopie*. Georg Thieme Verlag, Stuttgart.

Keeler, C., and Maciel, G.E. (2003) Quantitation in the solid-state C-13 NMR analysis of soil and organic soil fractions, *Anal. Chem.,* **75**, 2421-2432.

Kelly, A.E., Ou, H.D., Withers, R., and Dotsch, V. (2002) Low-conductivity buffers for high-sensitivity NMR measurements, *J. Am. Chem. Soc.,* **124**, 12013-12019.

Keun, H.C., Ebbels, T.M.D., Antti, H., Bollard, M.E., Beckonert, O., Schlotterbeck, G., Senn, H., Niederhauser, U., Holmes, E., Lindon, J.C., and Nicholson, J.K. (2002) Analytical reproducibility in H-1 NMR-based metabonomic urinalysis, *Chem. Res. Toxicol.,* **15**, 1380-1386.

Kim, S., Kaplan, L.A., Benner, R., and Hatcher, P.G. (2004) Hydrogen-deficient molecules in natural riverine water samples - evidence for the existence of black carbon in DOM. *Mar. Chem.,* **92**, 225-234.

Kim, S., Kramer, R.W., and Hatcher, P.G. (2003) Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram, *Anal. Chem.,* **75**, 5336-5344.

Kingery, W.L., Simpson, A.J., Hayes, M.H.B., and Hicks, R.P. (2000) The Application of Multidimensional NMR to the Study of Soil Humic Substances, *Soil Sci.,* **165**, 483-494.

Kingery, W.L., Simpson, A.J., Hayes, M.H.B., Locke, M.A., and Hicks. R.P. (2000) The application of multidimensional NMR to the study of soil humic substances, *Soil Sci*., **165**, 483-494.

Kinniburgh, D.G., Milne, C.J., Benedetti, M.F., Pinheiro, J.P., Filius, J. Koopal, L.K., and Van Riemsdijk, W.H. (1996) Metal Ion Binding by Humic Acid: Application of the NICA-Donnan Model, *Environ. Sci. Technol*., **30**, 1687-1698.

Kinniburgh, D.G., van Riemsdijk, W.H., Koopal, L.K., Borkovec, M., Benedetti, M.F., and Avena, M.J. (1999) Ion binding to natural organic matter: competition, heterogeneity, stoichiometry and thermodynamic consistency, *Colloids Surf., A.*, **151**, 147-166.

Klotz, I.M. (1982) Numbers of Receptor Sites from Scatchard Graphs: Facts and Fantasies, *Science*, **217**, 1247-1249.

Knicker, H., Fründ, R., and Lüdemann, H.-D. (1997) in M. Nanny, R.A. Minear, J.A. Leenheer (eds.) *NMR Spectroscopy in Environmental Chemistry*, Oxford University Press, London, pp. 272-294.

Kögel-Knabner, I. (2000) Analytical approaches for characterizing soil organic matter, *Org. Geochem.,* **31**, 609-625.

Kögel-Knabner, I. (2002) The macromolecular organic composition of plant and microbial residues as inputs to soil organic Matter, *Soil Biol. Biochem*., **34**, 139-162.

Kögel-Knabner, I., de Leeuw, J.W., and Hatcher, P.G. (1992) Nature and distribution of alkyl carbon in forest soil profiles: implications for the origin and humification of aliphatic biomacromolecules, *Sci. Total Environ*. **117/118**, 175-185.

Koopal, L.K., Van Riemsdijk, W.H., de Wit, J.C.M., and Benedetti, M.F. (1994) Analytical Isotherm Equations for Multicomponent Adsorption to Heterogeneous Surfaces, *J. Colloid Interf. Sci*., **166**, 51-60.

Kovalevskii, D.V., Permin, A.B., Perminova, I.V., and Petrosyan, V.S. (2000) Choice of the time of pulse delay for quantitative $^{13}C$ NMR spectroscopy of humic substances, *Bulletin of Moscow University* (*Vestnik MGU*), *Ser. 2 (chem)* **41,** 39-42.

Kramer, R.W., Kujawinski, E.B., and Hatcher, P.G. (2004) Identification of black carbon derived structures in a volcanic ash soil humic acid by Fourier transform ion cyclotron resonance mass spectrometry. *Environ. Sci. Technol.,* **38**, 3387-3395.

Kujawinski, E.B., Del Vecchio, R., Blough, N.V., Klein, G.C., and Marshall, A.G. (2004) Probing molecular-level transformations of dissolved organic matter: insights on photochemical degradation and protozoan modification of DOM from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Mar. Chem.,* **92**, 23-37.

Kujawinski, E.B., Freitas, M.A., Zang, X., Hatcher, P.G., Green-Church, K.B., and Jones, R.B. (2002) The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter. *Org. Geochem.* **33**, 171-180.

Kujawinski, E.B. (2002) Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (ESI FT-ICR MS): Characterization of Complex Environmental Mixtures, *Environ. Forensics,* **3**, 207-216.

Kulikova, N.A., and Perminova, I.V. (2002) Binding of atrazine to humic substances from soil, peat, and coal related to their structure, *Environ. Sci. Technol.,* **36**, 3720-3724.

Kumar, A., Wagner, G., Ernst, R.R., and Wüthrich, K. (1981) Buildup Rates of the Nuclear Overhauser Effect Measured by Two-Dimensional Proton Magnetic Resonance Spectroscopy, *J. Am. Chem. Soc.,* **103**, 3654-3658.

Kuzyakov, Y.V. (1997) The role of amino acids and nucleic bases in turnover of nitrogen and carbon in soil humic fractions. *Europ. J. Soil Sci*., **48**, 121-130.

Lambert, J., and Buddrus, J. (1996) Quantification of Isolated Methyl Groups in Aquatic Humic Substances by Means of $^{1}H$ and $^{13}C$ NMR Spectroscopy, *Magn. Reson. Chem*. **34,** 276-282.

Lambert, J., Buddrus, J., and Burba P. (1995) Evaluation of conditional stability constants of dissolved aluminium/humic substance complexes by means of $^{27}Al$ nuclear magnetic resonance, *Fresenius J. Anal. Chem.,* **351**, 83-87.

Larive, C.K., Rogers, A., Morton, M., and Carper, W.R. (1996) $^{113}Cd$ NMR Binding Studies of Cd-Fulvic Acid Complexes: Evidence of Fast Exchange, *Environ. Sci. Technol.,* **30**, 2828-2831.

Lead, J.R., Balnois, E., Hosse, M., Menghetti, R., and Wilkinson, K. J. (1999) Characterization of Norwegian natural organic matter: Size, diffusion coefficients, and electrophoretic mobilities. *Environment International,* **25**, 245-258.

Lee, G.S.H., Wilson, M.A., and Young, B.R. (1998) The application of the WATERGATE suppression technique for analyzing humic substances by nuclear magnetic resonance, *Org. Geochem*., **28**, 549-559.

Lee, S.-Z., Allen, H.E., Huang, C.P., Sparks, D.L., Sanders, P.F., and Peijnenburg, W.J.G.M. (1996) Predicting Soil-Water Partition Coefficients for Cadmium, *Environ. Sci. Technol*., **30**, 3418-3424.

Leenheer, J.A., Nanny, M.A., and McIntyre, C. (2003) Terpenoids as Major Precursors of Dissolved Organic Matter in Landfill Leachates, Surface Water, and Groundwater, *Environ. Sci. Technol.,* **37**, 2323-2331.

Li Y., Heitz F., Le Grimellec C., and Cole R.B. (2005) Fusion Peptide - Phospholipid Noncovalent Interactions as Observed by Nanoelectrospray FTICR-MS. *Anal. Chem.,* **77**, 1556-1565.

Li, J., Perdue, E.M., and Gelbaum, L.T. (1998) Using Cadmium-113 NMR Spectrometry to Study Metal Complexation by Natural Organic Matter, *Environ. Sci. Technol*., **32**, 483-487.

Liu, A., and Gonzalez, R.D (2000) Modeling adsorption of copper(II), cadmium(II) and lead(II) on purified humic acid, *Langmuir,* **16**, 3902-3909.

Lowe, L.E. (1992) Studies on the nature of sulfur in peat humic acids from the Fraser river delta, British Columbia, *Sci. Total Environ*., **113**, 133-145.

Lu, X., Johnson, W.D., and Hook, J. (1998) Reaction of Vanadate with Aquatic Humic Substances: An ESR and $^{51}$V NMR Study, *Environ. Sci. Technol*., **32**, 2257-2263.

Lu, X.Q., Hanna, J.V., and Johnson, W.D. (2000) Source indicators of humic substances: an elemental composition, solid state $^{13}$C CP/MAS NMR and Py-GC/MS study, *Appl. Geochem*., **15**, 1019-1033.

MacCarthy, P., and Rice, J.A. (1988) An Ecological Rationale for the Heterogeneity of Humic Substances, in S.H. Schneider, P.J. Boston (eds.), *Proceedings of Chapman Conference on the Gaia Hypothesis*, San Diego, CA, March 7-11. MIT Press, Cambridge, pp. 339-345.

Magusin, P.C.M.M., Bolz, A., Sperling, K., and Veeman, W.S. (1997) The use of $^{129}$Xe NMR spectroscopy for studying soils. A pilot study, *Geoderma*, **80**, 449-462.

Mahieu, N., Powlson, D.S., and Randall, E.W. (1999) Statistical analysis of published carbon-$^{13}$C CPMAS NMR spectra of soil organic matter, *Soil Sci. Soc. Am. J.,* **63**, 307-319.

Makarov, M.I., Malysheva, T.I., Haumaier, L., Alt, H.G., and Zech, W. (1997) The forms of phosphorus in humic and fulvic acids of a toposequence of alpine soils in the northern Caucasus, *Geoderma,* **80**, 61-73.

Mao, J.D., and Schmidt-Rohr, K. (2004) Accurate quantification of aromaticity and nonprotonated aromatic carbon fraction in natural organic matter by C-13 solid-state nuclear magnetic resonance, *Environ. Sci. Technol.,* **38**, 2680-2684.

Mao, J.D., and Schmidt-Rohr, K. (2004) Separation of aromatic-carbon $^{13}$C NMR signals from di-oxygenated alkyl bands by a chemical-shift-anisotropy filter, *Solid-State Nucl. Magn. Reson.,* **26**, 36-45.

Mao, J.D., Hu, W.G., Schmidt-Rohr, K., Davies, G., Ghabbour, E.A., and Xing, B., (2000) Quantitative characterization of humic substances by solid-state carbon-13 nuclear magnetic resonance, *Soil Sci. Soc. Am. J.,* **64**, 873-884.

Mao, J.D., Hub, W.G., Ding, G.W., Schmidt-Rohr, K., Davies, G., Ghabbour, E.A., and Xing, B., (2002) Suitability of different C-13 solid-state NMR techniques in the characterization of humic acids, *Int. J. Environ. An. Ch.,* **82**, 183-196.

Martin, G.E. (2002) Cryogenic NMR Probes: Applications, in D.M. Grant, and R.K. Harris (eds.), *Enzyclopedia of Nuclear Magnetic Resonance* **9,** Wiley-VCH, pp. 33-35.

Mason, J. (1987), *Multinuclear NMR*. Plenum Press, New York.

McCarthy, M., Hedges, J.I., and Benner, R. (1996) Major biochemical composition of dissolved high molecular weight organic matter in seawater, *Mar. Chem.,* **55**, 281-297.

McCarthy, M., Hedges, J.I., and Benner, R. (1998) Major Bacterial Contribution to Marine Dissolved Organic Nitrogen, *Science,* **281**, 231-234.

McIntyre, C. and McRae, C. (2005) Proposed guidelines for sample preparation and ESI-MS analysis of humic substances to avoid self-esterification. *Org. Geochem.* **36**, 543-553.

Medek, A., Frydman, V., and Frydman, L. (1997) Solid and liquid phase $^{59}$Co NMR studies of cobalamins and their derivatives, *Proc. Natl. Acad. Sci. USA* **94**, 14237-14242.

Megumu, M., Susume, K., and Fujio, Y. (1986) Cadmium-113 NMR of cadmium(II) complexes with ligands containing N-donor atoms. Dependence of the chemical shift upon the ligand basicity, chelate ring size, counteranion, and cadmium concentration, *Inorg. Chem*., **25**, 964-970.

Mikita, M.A., Steelink, C., and Wershaw, R.L. (1981) Carbon-13 Enriched Nuclear Magnetic Resonance Method for the Determination of Hydroxyl Functionality In Humic Substances, *Anal. Chem.,* **53**, 1715-1717.

Millet, O., Loria, J.P., Kroenke, C.D., Pons, M., and Palmer, A.G. (2000) The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale, J. *Am. Chem. Soc.,* **122**, 2867-2877.

Milne, C.J., Kinniburgh, D.G., and Tipping, E. (2001) Generic NICA-Donnan Model Parameters for Proton Binding by Humic Substances, *Environ. Sci. Technol*., **35,** 2049-2059.

Morra, M.J., Fendorf, S.E., and Brown, P.D. (1997) Speciation of sulfur in humic and fulvic acids using X-ray absorption near-edge structure (XANES) spectroscopy, *Geochim. Cosmochim. Acta*., **61**, 683-688.

Morris, K.F., Cutak, B.J., Dixon, A.M., and Larive, C.K. (1999) Analysis of diffusion coefficient distributions in humic and fulvic acids by means of diffusion ordered NMR spectroscopy, *Anal. Chem.,* **71,** 5315-5321.

Mortimer, R.D., and Dawson, A. (1991) Using $^{19}$F NMR for Trace Analysis of Fluorinated Pesticides in Food Products, *J. Agric. Food Chem*., **39**, 1781-1785.

Murali, N., and Krishnan, V. (2003) A primer for nuclear magnetic relaxation in liquids, *Concepts Magn. Reson. Part A*, **17A**, 86-116.

Nakashima, T.T., and Rabenstein, D.L. (1983) A Lead-207 Nuclear Magnetic Resonance Study of the Complexation of Lead by Carboxylic Acids and Aminocarboxylic Acids, *J. Magn. Reson*., **51**, 223-232.

Namjesnik-Dejanovic, K. and Cabaniss, S.E. (2004) Reverse-phase HPLC method for measuring polarity distributions of natural organic matter. *Environ. Sci. Technol.* **38**, 1108-1114.

Nanny, M.A., and Maza, J.P. (2001) Noncovalent interactions between monoaromatic compounds and dissolved humic acids: A deuterium NMR T-1 relaxation study, *Environ. Sci. Technol.,* **35**, 379-384.

Nanny, M.A., Bortiatynski, J.M., and Hatcher, P.G. (1997) Noncovalent interactions between acenaphthenone and dissolved fulvic acid as determined by C-13 NMR T-1 relaxation measurements, *Environ. Sci. Technol.,* **31**, 530-534.

Navon, G., Song, G.Y., Ròòm, T., Appelt, S., Taylor, A.E., and Pines A. (1996) Enhancement of Solution NMR and MRI with Laser-Polarized Xenon, *Science*, **271**, 1848-1851.

Nehls, I., Wagenknecht, W., Philipp, B., and Stscherbina, D. (1994) Characterization of Cellulose and Cellulose Derivatives in Solution by High Resolution $^{13}$C-NMR Spectroscopy, *Prog. Polym. Sci*., **19,** 29-78.

Nelson, P.N. and Baldock, J.A. (2005) Estimating the molecular composition of a diverse range of natural organic materials from solid-state $^{13}$C NMR and elemental analyses. *Biogeochem.,* **72,** 1-34.

Norwood, T.J., Boyd, J., Heritage, J.E., Soffe, N., and Campbell, I.D. (1990) Comparison of Techniques for $^{1}$H-Detected Heteronuclear $^{1}$H-$^{15}$N Spectroscopy, *J. Magn. Reson*., **87**, 488-501.

Oste, L.A., Temminghoff, E.J.M., Lexmond, T.M., and van Riemsdijk, W.H. (2002) Measuring and Modeling Zinc and Cadmium Binding by Humic Acid, *Anal. Chem*., **74**, 856-862.

Otto, W.H., Burton, S.D., Carper, W.R., and Larive, C.K. (2001) Examination of Cadmium(II) Complexation by the Suwannee River Fulvic Acid Using $^{113}$Cd NMR Relaxation Measurements, *Environ. Sci. Technol*., **35**, 4900-4904.

Otto, W.H., Carper, W.R., and Larive, C.K. (2001) Measurement of Cadmium(II) and Calcium(II) Complexation by Fulvic Acids Using $^{113}$Cd NMR, *Environ. Sci. Technol*., **35**, 1463-1468.

Ourisson, G., Albrecht, P., and Rohmer, M. (1979) The Hopanoids, *Pure Appl. Chem*., **51**, 709-729.

Ourisson, G., Rohmer, M., and Poralla, K. (1987) Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Ann. Rev. Microbiol.* **41**, 301-333.

Öz, G., Pountney, D.L., and Armitage, I.M. (1998) NMR Spectroscopic studies of I = 1/2 metal ions in biological systems, *Biochem. Cell Biol*., **76**, 223-234.

Palmer, A.G. (2002) Chemical exchange effects in biological macromolecules, in D.M. Grant and R.K. Harris (eds.), *Enzyclopedia of Nuclear Magnetic Resonance* **9,** Wiley-VCH, pp. 344-353.

Palmer, A.G., Kroenke, C.D., and Loria, J.P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, *Method. Enzymol*., **339**, 204-238.

Pauli, G.F., Jaki, B.U., and Lankin, D.C. (2005) Quantitative H-1 NMR: Development and potential of a method for natural products analysis. *J. Nat. Prod.,* **68**, 133-149.

Paytan, A., Cade-Menun, B.J., McLaughlin, K., and Faul, K.L. (2003) Selective phosphorus regeneration of sinking marine particles: evidence from P-31-NMR, *Mar. Chem.,* **82**, 55-70.

Pellecchia, M., Sem, D.S., and Wüthrich, K. (2002) NMR in Drug Discovery, *Nature Reviews*. **1**, 211-219.

Perakis, S.S., and Hedin, L.O. (2002) Nitrogen loss from unpolluted South American forest mainly via dissolved organic compounds, *Nature*, **24**, 415-419.

Perdue, E.M. (1985) Acidic Functional Groups of Humic Substances, in G.R. Aiken, D.M. McKnight, R.L. Wershaw, P. MacCarthy (eds.), *Humic Substances in Soil, Sediment, and Water; Geochemistry, Isolation and Characterization*, Wiley, Chichester, pp. 493-526.

Perdue, E.M. (1998) Chemical Composition, Structure, and Metal Binding Properties, in D.O. Hessen, Tranvik (eds.), *Aquatic Humic Substances: Ecology and Biogeochemistry*, Springer: Heidelberg, pp. 41-61.

Perdue, E.M. (2001) Modeling concepts in metal-humic complexation, in C.E. Clapp, M.H.B. Hayes, N. Senesi, P.R. Bloom, P.M. Jardine (eds.), *Humic Substances and Chemical Contaminants*, Soil Science Society of America, Inc., Madison, WI, pp. 305-316.

Perdue, E.M., and Carreira, L.A. (1997) Modeling competitive binding of protons and metal ions by humic substances, in D.C. Adriano, Z.S. Chen, S.S. Yang, Iskandar, I. K. (eds.), *Biogeochemistry of Trace Metals*, in the Advances for Environmental Science series, Applied Science Publishers: Norwood, UK, pp. 381-401.

Perdue, E.M., and Lytle, C.R. (1983) A distribution model for binding of protons and metal ions by humic substances, *Environ. Sci. Technol*., **17**, 654-660.

Perdue, E.M., and Parrish, R.S. (1987) Fitting multisite binding equilibria to statistical distribution models: Turbo Pascal program for Gaussian models, *Comput. Geosci.,* **13**, 587-601.

Perdue, E.M., and Ritchie, J.D. (2003) Dissolved Organic Matter in Fresh Waters, in H.D. Holland and K.K. Turekian (eds.), *Treatise on Geochemistry,* Vol. 5, Surface and Ground Water, Weathering, Erosion and Soils, Chapter 11. pp. 273-318.

Perdue, E.M., Reuter, J.H., and Parrish, R.S. (1984) A statistical model of proton binding by humus, *Geochim. Cosmochim. Acta.*, **48**, 1257-1263.

Perminova, I. V., Hertkorn, N., and Hatfield, K. (2005) *Use of humic substances to remediate polluted environments: From theory to practice*. Kluver Academic Publishers.

Perminova, I.V., Frimmel, F.H., Kudryavtsev, A.V., Kulikova, N.A., Abbt-Braun, G., Hesse, S., and Petrosyan, V.S. (2003) Molecular weight characteristics of humic substances from different environments as determined by size exclusion chromatography and their statistical evaluation, *Environ. Sci. Technol.,* **37**, 2477-2485.

Perminova, I.V., Grechishcheva, N.Y., Kovalevskii, D.V., Kudryavtsev, A.V., Petrosyan, V.S., and Matorin, D.N. (2001) Quantification and prediction of the detoxifying properties of humic substances related to their chemical binding to polycyclic aromatic hydrocarbons, *Environ. Sci. Technol.,* **35**, 3841-3848.

Peuravuori, J., Ingman, P., and Pihlaja, K. (2003) Critical comments on accuracy of quantitative determination of natural humic matter by solid state $^{13}C$ NMR spectroscopy, *Talanta*, **59**, 177-189.

Ping, G., Schmitt-Kopplin, P., Hertkorn, N., Zhang, W. B., Zhang, Y. K., and Kettrup, A. (2003) Separation of selected humic degradation compounds by capillary electro-chromatography with monolithic and packed columns. *Electrophoresis,* **24**, 958-969.

Plette, A.C.C., Benedetti, M.F., and van Riemsdijk, W.H. (1996) Competitive Binding of Protons, Calcium, Cadmium, and Zinc to Isolated Cell Walls of a Gram-Positive Soil Bacterium, *Environ. Sci. Technol*., **30**, 1902-1910.

Poirier, N., Sohi, S.P., Gaunt, J.L., Mahieu, N., Randall, E.W., Powlson D.S., and Evershed, R.P. (2005) The chemical composition of measurable soil organic matter pools. *Org. Geochem.* **36**, 1174-1189.

Pregosin, P.S. (1991) *Transition Metal Nuclear Magnetic Resonance*, Elsevier, Amsterdam.

Preston, C.M. (1996) Applications of NMR to Soil Organic Matter Analysis: History and Prospects, *Soil Sci*., **161**, 144-166.

Preston, C.M. (2001) Carbon-13 solid-state NMR of soil organic matter - using the technique effectively, *Can. J. Soil Sci.,* **81**, 255-270.

Preston, C.M., and Blackwell, B.A. (1985) Carbon-13 nuclear magnetic resonance for a humic and a fulvic acid: signal-to-noise optimisation, quantitation, and spin-echo techniques, *Soil Sci*., **139**, 88-96.

Preston, C.M., Axelson, D.E., Levesque, M., Mathur, S., Dinel, H., and Dudley, R.L. (1989) Carbon-13 NMR and chemical characterization of particle-size separates of peats differing in degree of decomposition, *Org. Geochem*., **14**, 393-403.

Price, W.S. (2002) Diffusion-based Studies of Aggregation, Binding and Conformation of Biomolecules: Theory and Practice, in D.M. Grant and R.K. Harris (eds.), *Enzyclopedia of Nuclear Magnetic Resonance* **9**, Wiley-VCH, pp. 364-374.

Qualls, R.G. (2005) Biodegradability of Fractions of Dissolved Organic Carbon Leached from Decomposing Leaf Litter. *Environ. Sci. Technol.,* **39**, 1616-1622.

Randall, S.R., Sherman, D.M., Ragnarsdottir, K.V., and Collins, C.R. (1999) The mechanism of cadmium surface complexation on iron oxyhydroxide minerals, *Geochim. Cosmochim. Acta*, **63**, 2971-2987.

Reid, D. G. (1997) *Protein NMR Techniques*. Humana Press, Totowa.

Reynhardt, E.C. (2003) Spin-lattice relaxation of spin-1/2 nuclei in solids containing diluted paramagnetic impurity centers. I. Zeeman polarization of nuclear spin system, *Concepts Magn. Res. A* **19A**, 20-35.

Reynolds, W.F., Lean, S.M., Tay, L., Yu, M., Enriquez, R.G., Estwick, D.M., and Pascoe, K.O. (1997) Comparison of $^{13}C$ Resolution and Sensitivity of HSQC and HMQC Sequences and Application of HSQC-Based Sequences to the Total $^{1}H$ and $^{13}C$ Spectral Assignment of Clionasterol, *Magn. Reson. Chem*., **35**, 455-462.

Ricardo, A., Carrigan, M.A., Olcott, A.N., and Benner, S. A. (2004) Borate minerals stabilize ribose. *Science,* **303**, 196.

Ritchie, J.D., and Perdue, E.M. (2003) Proton-binding study of standard and reference fulvic acids, humic acids, and natural organic matter, *Geochim. Cosmochim. Acta,* **67**, 85-96.

Russel, D.J., Hadden, C.E., Martin, G.E., Gibson, A.A., Zens, A.P., and Carolan, J.L. (2000) A Comparison of Inverse-Detected Heteronuclear NMR Performance: Conventional vs Cryogenic Microprobe Performance, *J. Nat. Prod.,* **63**, 1047-1049.

Rutledge, D.N. (1996) *Signal treatment and signal analysis in NMR*, Elsevier, Amsterdam.

Saar, R.A., and Weber, J.H. (1982) Fulvic acid: modifier of metal-ion chemistry, *Environ. Sci. Technol.*, **16**, 510A-517A.

Sakakibara, A. (1990) Chemistry of Lignin*,* in Hon, N.-S., and Shisraishi, N. (Eds.) *Wood and Cellulosic Chemistry,* Marcel Dekker, New York, p. 160.

Sanudo-Wilhelmy S. A., Rossi F. K., Bokuniewicz H., and Paulsen R. J. (2002) Trace metal levels in uncontaminated groundwater of a coastal watershed: Importance of colloidal forms. *Environ. Sci. Technol.,* **36**, 1435-1441.

Sardessai, S. and Wahidullah, S. (1998) Structural characteristics of marine sedimentary humic acids by CP/MAS $^{13}C$ NMR spectroscopy. *Oceanologica Acta*. **21,** 543-550.

Saunders, M., Wishnia, W., and Kirkwood, J.G. (1957). The Nuclear Magnetic Resonance Spectrum of Ribonuclease, *J. Am. Chem*. Soc., **79,** 3289-3290.

Schmidt A., Karas, M., and Dulcks, T. (2003) Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI? *Journal of the American Society for Mass Spectrometry*, **14**, 492-500.

Schmitt-Kopplin, P., and Frommberger, M. (2003) Capillary electrophoresis - electrospray ionisation mass spectrometry: 15 years of development and applications. *Electrophoresis,* **24**, 3837-3867.

Schmitt-Kopplin, P., and Junkers, J. (2003) Capillary zone electrophoresis of natural organic matter, *J. Chromatogr. A.,* **998**, 1-20.

Schmitt-Kopplin, P., and Kettrup, A. (2003) Capillary electrophoresis - electrospray ionization - mass spectrometry for the characterization of natural organic matter: An evaluation with free flow electrophoresis-off-line flow injection electrospray ionization-mass spectrometry, *Electrophoresis,* **24**, 3057-3066.

Schmitt-Kopplin, Ph., Hertkorn, N., Garrison, A.W., Freitag, D., and Kettrup, A. (1998) Influence of Borate Buffers in the Electrophoretical Behaviour of Humic Substances in Capillary Zone Electrophoresis, *Anal. Chem.,* **70,** 3798-3808.

Schmitt-Kopplin, Ph., Hertkorn, N., Schulten, H.R., and Kettrup, A. (1998) Structural Changes in a Dissolved Soil Humic Acid during Photochemical Degradation Processes under $O_2$ and $N_2$ Atmosphere, *Environ. Sci. Technol*., **32,** 2531-2541.

Schraml, J. (1990) $^{29}$Si NMR Spectroscopy of Trimethylsilyl Tags, *Prog. NMR Spectrosc*., **22**, 289-348.

Schraml, J., Blechta, V., Kvicalová, M., Nondek, L., and Chvalovsky, V. (1986) Polar Functional Group Analysis of Mixtures by Silicon-29 Nuclear Magnetic Resonance, *Anal. Chem*., **58**, 1892-1894.

Schulten, H.R., and Gleixner, G. (1999) Analytical Pyrolysis of Humic Substances and Dissolved Organic Matter in Aquatic Systems: Structure and Origin, *Wat. Res*., **33**, 2489-2498.

Schumacher, M., and Lauterwein, J. (1989) The INEPT Experiment for Nonselective Polarization Transfer in $^{17}$O NMR, *J. Magn. Reson*., **83**, 97-110.

Sergeyev, N.M., Sergeyeva, N.D., and Raynes, W.T. (1999) Isotope effects on the O-17, H-1 coupling constant and the O-17-{H-1} nuclear Overhauser effect in water, *J. Magn. Reson.,* **137,** 311-315.

Serkiz, S.M., and Perdue, E.M. (1990) Isolation of dissolved organic matter from the Suwannee river using reverse osmosis, *Wat. Res*., **24**, 911-916.

Shevchenko, S.M., and Bailey, G.W. (1996) The mystery of the lignin-carbohydrate complex: a computational approach, *Theochem*., **364**, 197-208.

Shin, H.S., and Moon, H. (1996) An "average" structure proposed for soil fulvic acid aided by DEPT/QUAT C-13 NMR pulse techniques, *Soil Sci.,* **161**, 250-256.

Shin, H.S., Rhee, S.W., Lee, B., and Moon, H. (1996) Metal binding sites and partial structures of soil fulvic and humic acids compared: aided by Eu(III) luminescence spectroscopy and DEPT/QUAT $^{13}$C NMR pulse techniques, *Org. Geochem*., **24**, 523-529.

Simpson, A.J., Tseng, L.H., Simpson, M.J., Spraul, M., Braumann, U., Kingery, W.L., Kelleher, B.P., and Hayes, M.H.B. (2004b) The application of LC-NMR and LC-SPE-NMR to compositional studies of natural organic matter. *Analyst* **129**, 1216-1222.

Simpson, A. (2001) Multidimensional Solution State NMR of Humic Substances: a Practical Guide and Review, *Soil Sci.,* **166**, 795-809.

Simpson, A.J., Boersma, R.E., Kingery, W.L., Hicks, R.P., and Hayes, M.H.B. (1997) Humic Substances, Peats and Sludges, in M.H.B. Hayes, W.S. Wilson (eds.), *Applications of NMR Spectroscopy for Studies of the Molecular Compositions of Humic Substances*, Royal Society of Chemistry, Cambridge, pp. 46-62.

Simpson, A.J., Burdon, J., Graham, C.L., Hayes, M.H.B., Spencer, N., and Kingery, W.L. (2001) Inerpretation of heteronuclear and multidimensional NMR spectroscopy of humic substances, *Eur. J. Soil Sci.,* **52**, 495-509.

Simpson, A.J., Kingery, W.L., Hayes, M.B., Spraul, M., Humpfer, E., Dvortsak, P., Kerssebaum, R., and Godejohann, M. (2002) Molecular structures and associations of humic substances in the terrestrial environment, *Naturwiss.,* **89**, 84-88.

Simpson, A.J., Lefebvre, B., Moser, A., Williams, A., Larin, N., Kvasha, M., Kingery, W.L., and Kelleher, B. (2004a) Identifying residues in natural organic matter through spectral prediction and pattern matching of 2D NMR datasets, *Magn. Reson. Chem.,* **42**, 14-22.

Simpson, A.J., Salloum, M.J., Kingery, W.L., and Hatcher, P.G. (2003) The Identification of Plant Derived Structures in Humic Materials Using Three-Dimensional NMR Spectroscopy, *Environ. Sci. Technol*., **37**, 337-342.

Simpson, A.J., Salloum, M.J., Kingery, W.L., and Hatcher, P.G. (2002) Improvements in the Two-Dimensional Nuclear Magnetic Resonance Spectroscopy of Humic Substances, *J. Environ. Qual*., **31**, 388-392.

Skoog, A. and Benner, R. (1997) Aldoses in various size fractions of marine organic matter: implications for carbon cycling. *Limnol. Oceanogr.* **42**, 1803-1813.

Smernik, R.J. (2005) Solid-state $^{13}$C NMR spectroscopic studies of soil organic matter at two magnetic field strengths. *Geoderma,* **125**, 249 - 271.

Smernik, R.J., and Oades, J.M. (1999) Effects of added paramagnetic ions on the C-13 CP MAS NMR spectrum of a de-ashed soil, *Geoderma*, **89**, 219-248.

Smernik, R.J., and Oades, J.M. (2001) Solid-state C-13-NMR dipolar dephasing experiments for quantifying protonated and non-protonated carbon in soil organic matter and model systems, *Eur. J. Soil Sci.,* **52**, 103-120.

Smernik, R.J., and Oades, J.M. (2003) Spin accounting and RESTORE – two new methods to improve quantitation in solid-state C-13 NMR analysis of soil organic matter, *Eur. J. Soil Sci.,* **54**, 103-116.

Sorensen, O.W., and Jakobsen, H.J. (1988) Polarization Transfer and Editing Techniques, in. W.S. Brey (ed.), *Pulse Methods in 1D and 2D Liquid-Phase NMR*, Academic Press, San Diego, pp. 149-258.

Sparks, L. (1995) *Environmental Soil Chemistry*. Academic Press, San Diego, pp. 62-64.

Stankiewicz, B.A., and van Bergen, P.F. (1998) *Nitrogen-Containing Macromolecules in the Bio- and Geosphere,* ACS Symposium Series 707, American Chemical Society, Washington, DC, pp. 154.

Steinberg, C.E.W. (2003) *Ecology of Humic Substances in Freshwaters*. Springer Press, Berlin.

Stenson, A.C., Marshall, A.G., and Cooper, W.T. (2003) Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra, *Anal. Chem.,* **75**, 1275-1284.

Stevenson, F.J. (1982) Humus Chemistry. Wiley, New York, pp. 36.

Stevenson, F.J. (1994) *Humus Chemistry: Genesis, Composition, Reactions*, John Wiley, New York.

Summers, M.F. (1988) $^{113}$Cd NMR Spectroscopy of Coordination Compounds and Proteins, *Coord. Chem. Rev.*, **86**, 43-134.

Summons, R.E. (1993) in *Organic Geochemistry: Principles and Applications (Biogeochemical cycles: A review of fundamental aspects of organic matter formation, preservation, and composition)*, eds. Engel, M.H. & Macko, S.A. (Plenum Press, New York), pp. 3-21.

Szajdak, L., and Österberg, R. (1996) Amino acids present in humic acids from soil under different cultivations, *Environ. Int.*, **22**, 3331-3334.

Takahashi, H., Nakanishi, T., Kami, K., Arata, Y., and Shimada, I. (2000) A novel NMR method for determining the interfaces of large protein-protein complexes, *Nat. Struct. Biol.,* **7**, 220-223.

Tang, H.R., Wang, Y.L., Nicholson, J.K., and Lindon, J.C. (2004) Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma, *Anal. Biochem.*, **325**, 260-272.

Tanoue, E., Nishiyama, S., Kamo, M., and Tsugita, A. (1995) Bacterial membranes: possible source of a major dissolved protein in seawater, *Geochim. Cosmochim. Acta.,* **59**, 2643-2650.

Tenailleau, E.J., Lancelin, P., Robins, R.J., and Akoka, S. (2004) Authentication of the Origin of Vanillin Using Quantitative Natural Abundance $^{13}$C NMR. *J. Agric. Food Chem*. **52**, 7782-7787.

Terblanche, C.J., Reynhardt, E.C., and van Wyk, J.A. (2001) C-13 spin-lattice relaxation in natural diamond: Zeeman relaxation at 4.7 T and 300 K due to fixed paramagnetic nitrogen defects, *Solid State Nucl. Mag.,* **20**, 1-22.

Thomas, S., Ströhl, D., and Kleinpeter, E. (1994) Computer Application of an Incremental System for Calculating the $^{13}$C NMR Spectra of Aromatic Compounds., *J. Chem. Comput. Sci.,* **34**, 725-729.

Thomsen, M., Lassen, P., Dobel, S., Hansen, P.E., Carlsen, L., and Bugel-Mogensen, B. (2002) Characterisation of humic materials of different origin: A multivariate approach for quantifying the latent properties of dissolved organic matter, *Chemosphere,* **49**, 1327-1337.

Thorn, K.A., Folan, D.W., Arterburn, J.B., Mikita, M.A., and MacCarthy, P. (1989) Application of INEPT Nitrogen-15 and Silicon-29 Nuclear Magnetic Resonance Spectrometry to Derivatized Fulvic Acids, *Sci. Total Environ*., **81/82**, 209-218.

Thorn, K.A., Steelink, C., and Wershaw, R.L. (1987) Methylation patterns of aquatic humic substances determined by $^{13}$C NMR spectroscopy, *Org. Geochem*., **11,** 123-137.

Tipping, E. (1994) WHAM – a chemical equilibrium and computer code for waters, sediments, and soils incorporating a discrete site/electrostatic model of ion-binding by humic substances, *Comput. Geosci.*, **20**, 973-1023.

Tipping, E. (1998) Humic Ion-Binding Model VI: An Improved Description of the Interactions of Protons and Metal Ions with Humic Substances, *Aquatic Geochem*., **4,** 3-48.

Tipping, E., and Hurley, M.A. (1992) A unifying model of cation binding by humic substances, *Geochim. Cosmochim. Acta*., **56**, 3627-3641.

Tohmura, S., and Argyropoulos, D.S. (2001) Determination of arylglycerol-beta-aryl ethers and other linkages in lignins using DFRC/P-31 NMR, *J. Agric. Food Chem.,* **49**, 536-542.

Tomlins, A.M., Foxall, P.J.D., Lynch, M.J., Parkinson, J., Everett, J.R., and Nicholson, J.K. (1998) High resolution H-1 NMR spectroscopic studies on dynamic biochemical processes in incubated human seminal fluid samples, *Biochim. Biophys. Acta-General Subjects,* **1379**, 367-380.

Ussiri, D.A.N., and Johnson, C.E. (2003) Characterization of organic matter in a northern hardwood forest soil by C-13 NMR spectroscopy and chemical methods, *Geoderma,* **111**, 123-149.

van de Ven, F.J.M. (1995) *Multidimensional NMR in Liquids*, VCH Publishers, Inc. New York, p. 14.

van Geet, A.L. (1970) Calibration of methanol nuclear magnetic resonance thermometer at low temperature, *Anal. Chem.,* **42**, 679-680.

Varney, M.S., Mantoura, R.F.C., Whitfield, M.; Turner, D.R., and Riley, J.P. (1983) Potentiometric and conformational studies of the acid-base properties of fulvic acids from natural waters, in Wong C.S. *et al.* (eds.), *Trace Metals in Seawater*, Plenum: New York, pp. 751-772.

Vasak, M. (1998) Application of Cd-113 NMR to metallothioneins, *Biodegradation*, **9**, 501-512.

Vernonclark, R.N., Goldberg, E.D., and Bertine, K.K. (1994) Organic and inorganic characterization of marine colloids, *Chem. Ecol.,* **11**, 69-83.

Viant, M.R. (2003) Improved methods for the acquisition and interpretation of NMR metabolomic data, *Biochim. Biophys. Res. Commun.,* **310**, 943-948.

Visser, S. A.**,** (1983) Application of Van Krevelen's graphical-statistical method for the study of aquatic humic material, *Environ. Sci. Technol.,* **17,** 412-417.

Vugmeyster, L., Pelupessy, P., Vugmeister, B.E., Abergel, D., and Bodenhausen, G. (2004) Cross-correlated relaxation in NMR of macromolecules in the presence of fast and slow internal dynamics, *C.R. Physique* **5**, 377-386.

Wakeham, S.G., Pease, T.K., and Benner, R. (2003) Hydroxy fatty acids in marine dissolved organic matter as indicators of bacterial membrane material, *Org. Geochem.,* **34**, 857-868.

Walker, F.A. (1999) Magnetic spectroscopic (EPR, ESEEM, Mossbauer, MCD and NMR) studies of low-spin ferriheme centers and their corresponding heme proteins, *Coordin. Chem. Rev.,* **186**, 471-534.

Walker, F.A. (2003) Pulsed EPR and NMR Spectroscopy of Paramagnetic Iron Porphyrinates and Related Iron Macrocycles: How to Understand Patterns of Spin Delocalization and Recognize Macrocycle Radicals, *Inorg. Chem.,* **42**, 4526-4544.

Wang, K., Dickinson, L.C., Ghabbour, E.A., Davies G., and Xing, B. (2003) Proton Spin-Lattice Relaxation Times of Humic Acids as Determined by Solution NMR, *Soil Sci.,* **168**, 128-136.

Wang, L., Mao, X., and Yang, Y. (1998) Application of newly-developed $^{1}$H NMR techniques to the study of humic acids, *Bopuxue Zazhi (Chinese Journal of Magnetic Resonance)* **15**, 411-420.

Wang, S.M., and Gilpin, R.K. (1983) Cadmium-113 and carbon-13 nuclear magnetic resonance spectrometry of cadmium-amino acid complexes, *Anal. Chem*., **55**, 493-497.

Webb, A. G. (1997) Radiofrequency microcoils in magnetic resonance, *Prog. Nucl. Magn. Reson. Spec.,* **31**, 1-42.

Weigelt G., Balega Y., Blocker T., Fleischer A. J., Osterbart R., and Winters J.M. (1998) 76 mas speckle-masking interferometry of IRC +10 216 with the SAO 6 m telescope: Evidence for a clumpy shell structure. *Astron. Astrophys.* **333**, L51-L54.

Willker, W., Leibfritz, D., Kerssebaum, R., and Bermel, W. (1992) Gradient selection in inverse heteronuclear correlation spectroscopy, *Magn. Reson. Chem*., **31**, 287-292.

Wilson, M.A., Barron, P.F., and Gillam, A.H. (1981) The structure of freshwater humic substances as revealed by $^{13}$C-NMR spectrosco. *Geochiom. Cosmochim. Acta.* **45,** 1743-1750.

Wishart, D.S., Bigman, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. (1995) $^{1}H$, $^{13}C$, and $^{15}N$ random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J. Biomol. NMR*., **5**, 67-81.

Wishart, D.S., Sykes, B.D., and Richards, F.M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure, *J. Mol. Biol.,* **222**, 311-333.

Wu, Z.G., Rodgers, R.P., and Marshall, A.G. (2004) Two- and three-dimensional van Krevelen diagrams: A graphical analysis complementary to the Kendrick mass plot for sorting elemental compositions of complex organic mixtures based on ultrahigh-resolution broadband Fourier transform ion cyclotron resonance mass measurements, *Anal. Chem.,* **76**, 2511-2516.

Xia, H., and Rayson, G.D. (2000) Solid-state $^{113}Cd$ NMR studies of metal-binding to a *Datura innoxia* biomaterial, *Adv. Environ. Res*., **4**, 69-77.

Xia, H., and Rayson, G.D. (2002) $^{113}Cd$-NMR spectrometry of $Cd^{2+}$ binding sites on algae and higher plant tissues, *Adv. Environ. Res*., **7**, 157-167.

Xia, K., Skyllberg, U.L., Bleam, W.F., Bloom, P.R., Nater, E.A., and Helmke, P.A. (1999) X-ray absorption spectroscopic evidence for the complexation of Hg(II) by reduced sulfur in soil humic substances, *Environ. Sci. Technol*., **33**, 257-261.

Xia, Z., Akim, L.G., and Dimitris S. Argyropoulos (2001) Quantitative $^{13}C$ NMR Analysis of Lignins with Internal Standards. *J. Agric. Food Chem.*, **49***,* 3573-3578.

Xiang, B.S., and Markham, G.D. (1996) The conformation of inosine 5'-monophosphate (IMP) bound to IMP dehydrogenase determined by transferred nuclear Overhauser effect spectroscopy, *J. Biol. Chem.,* **271**, 27531-27535.

Xiong, Y., Schroeder, K., Greenbaum, N.L., Hendrickson, C.L., and Marshall, A.G. (2004) Improved mass analysis of oligoribonucleotides by C-13, N-15 double depletion and electrospray ionization FT-ICR mass spectrometry. *Anal. Chem.,* **76**, 1804-1809.

Xue, H., and Sigg, L. (1998) Cadmium speciation and complexation by natural organic ligands in fresh water, *Anal. Chim. Acta*., **363**, 249-259.

Yu, Y.H., Kirkup, C.E., Pi, N., and Leary, J.A. (2004) Characterization of noncovalent protein-ligand complexes and associated enzyme intermediates of GlcNAc-6-O-sulfotransferase by electrospray ionization FT-ICR mass spectrometry. *J. Am. Soc. Mass Spectrom.* **15**, 1400-1407.

Zamfir, A., Vukelic, Z., Bindila, L., Peter-Katalinic, J., Almeida, R., Sterling, A., and Allen, M. (2004) Fully-automated chip-based nanoelectrospray tandem mass spectrometry of gangliosides from human cerebellum. *J. Am. Soc. Mass Spectrom.,* **15**, 1649-1657.

Zbytniewski, R. and Buszewski, B. (2005) Characterization of natural organic matter (NOM) derived from sewage sludge compost. Part 2: multivariate techniques in the study of compost maturation. *Bioresour. Technol.,* **96**, 479-484.

Zerbe, O. (2003) *BioNMR in Drug Research*. Wiley-VCH.

Zhang, X., Amelung, W., Yuan, Y., and Zech, W. (1997) Amino sugars in soil of the North American cultivated prairie, *Z. Pflanzenernähr. Bodenk.*, **160**, 533-538.

Zuiderweg, E.R.P. (2002) Mapping protein-protein interactions in solution by NMR Spectroscopy, *Biochemistry* **41**, 1-7.