THE UNIVERSITY
of EDINBURGH

Molecular-scale characterisation of humic substances using isotope-filtered nD NMR spectroscopy

Nicholle G. A. Bell

A thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2015

Declaration

I declare that this thesis is my own composition and that the work present herein is my own, with the exception of the contributions clearly stated at the beginning of each chapter. All of the work described was carried out within the School of Chemistry at the University of Edinburgh between September 2011 and August 2015. I affirm that this work has not been previously submitted, either in part or whole, in support of an application for any other degree or qualification at this or any other university or institute of learning.

Signed:

Date:

Acknowledgements

I would like firstly to thank Dušan Uhrin and Margaret Graham for their continued friendship, support and guidance throughout this project. Without their belief and trust in me to get this project off the ground, it would not be possible.

The Graham group past and present for always being very friendly, up for general soil chat and for some very memorable Christmas parties.

Margaret and Natalia for keeping us cultured with many wonderful concerts.

Lorna, Dusan, Juraj, John, Haris, Will, Natalia, Hannah, Adam for all the laughter, craziness, cake and memorable outings. Once part of the NMR group, always part of the NMR group ☺.

Prof Colin Pulham, Dr Jenny Bos and the RSC for giving me the opportunity to run Spectroscopy in a Suitcase along with this project. Everyone should love spectroscopy!

Everyone past and present in stores, finance and admin for making day-to-day life easier.

Davy for the cheerful mail deliveries.

The coffee machine, for turning me from zombie to 'normal' person each day. DropBox for keeping me organised during this writing process.

My family for trying to understand what I am doing, supporting me when things were not going well and for the presents, cards, and celebrations during the more successful moments.

Edinburgh and Scotland, for being the most wonderful place to live and work.

Scottish hills and mountains for inspiring ideas, solutions and providing me with a peace of mind.

Abstract

Humic substances (HS), the complex mixture of organic molecules produced by microbial and abiotic degradation of plant and animals residues, are the major components of organic matter in soil and natural waters. As such, they are vital to the integrity of soil, play important roles in nutrient biogeochemical cycling and determine the mobility and fate of both nutrients and contaminants. How HS enact their varied roles, however, is largely unknown due to the lack of comprehension of their molecular make-up. What is known is that they contain aromatic (e.g. phenols, condensed aromatics) and aliphatic (e.g. carbohydrates, lipids) molecules with prevailing carboxyl and hydroxyl functionalities.

As for other complex chromatographically inseparable mixtures, high-resolution analytical techniques such as Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS) as well as Nuclear Magnetic Resonance (NMR) spectroscopy have been applied to study HS composition at the molecular level. While FT-ICR MS can provide molecular formulae, it is only NMR that can unveil structures. Until now, however, very few unambiguous structures of individual molecules have emerged from standard NMR experiments. This is because purposely designed multidimensional NMR experiments are required to achieve 'spectroscopic separation' where chromatography fails.

In this work, 'spectroscopic separation' was accomplished with the aid of chemical modification in the form of ^{13}C methylation of COOH and OH groups. ^{13}C containing tags allow the observation of signals only from molecules carrying these tags. In combination with purposely-designed isotope-filtered NMR experiments, these tags act as spies reporting on their surrounding structure. This is achieved by utilising scalar and dipolar couplings to transfer the polarisation between protons and carbons of the $^{13}CH_3O$ groups and nuclei of the parent molecules. The necessary spectral resolution is attained using 3D/4D NMR experiments. This approach provides access to an array of correlated chemical shifts of HS molecules and represents a paradigm shift in the use of tags in investigations of complex mixtures: instead of focusing only on the tags, they are used to obtain structural information from the molecules they are attached to.

The compounds at the centre of this investigation are the phenols of HS. These molecules are thought to be important in many of the key processes in organic matter rich soils, particularly

peat. For example, their accumulation is thought to impair the activity of extracellular enzymes essential for the degradation of peat HS. Understanding the nature and source of phenols would provide a more fundamental framework for rationalising their role in peat, as well as other carbon rich soils.

The developed methodology was initially tested on model mixtures containing 3 or 9 phenolic compounds before applying it to a operational fraction of HS, fulvic acid, extracted from a Scottish raised peat bog. For this fulvic acid, over 30 major phenolic molecules/moieties were identified, many of which can be directly traced to the flora prevalent in the vicinity of the sampling site. For the first time in 150 years of HS research, a novel analytical methodology has yielded unambiguous structural information, providing the first steps towards understanding the various roles of HS in natural systems.

Thesis content

Chapter 1 contains an overview of the current understanding of HS in terms of structure, composition and function. It will outline the role and fate of phenolic compounds in peat bogs and phytochemistry of the plants of present at the sampling site of the studied HS sample. It will briefly touch on the application of the FT–ICR MS in structural studies of HS and provide a comprehensive overview of the application of NMR in this area. Chemical modification of HS with the aim of studying their structure is presented next, followed by the theory of NMR spectroscopy and a brief description of the basic building blocks of NMR pulse sequences. The chapter will conclude by presenting the aims and objectives of this study.

Chapter 2 details sample collection, separation of the obtained sample into operational fractions and their chemical modification. Model mixtures used for the development of the designed NMR experiments are introduced together with the applied methylation procedures. Finally, the parameters used for the acquisition of the NMR experiments are listed.

Chapter 3 describes the outcome of the extraction and chemical modification of HS samples and the model mixtures.

Chapter 4 provides a full description of n-dimensional (nD) NMR experiments developed in this study.

Chapters 5, **6** and **7** demonstrate the application of the developed NMR experiments on ^{13}C-methylated model mixtures **I, II** and partially methylated mixture **II**, respectively. Using model mixture **II**, Chapter 6 compares the performance of individual experiments in terms of obtainable Signal-to-Noise (S/N) ratios, outlines their merits, and sets out the way in which the spectra are efficiently analysed.

Chapter 8 provides initial characterisation of the methylated fulvic acid from Red Moss raised peat bog and continues with the analysis of acquired nD NMR spectra. Their analysis, illustrated on one of the compounds, leads to the identification of structures/structural motives of phenolic compounds of the fulvic acid sample. Chapter 8 classifies and compares the identified compounds with those found by others in plants present at the sampling site. A brief discussion of other techniques for the identification of phenolic compounds of HS is also presented.

Chapter 9 concludes the thesis by outlining the significance of the achieved results, future directions and steps required for further molecular- level characterisation of HS. The publications resulting from this work are provided after the References.

Lay Summary

For many of us soil is 'dirt', that brown natural compost that layers the Earth; providing a surface to walk on, an embankment to rivers and lochs, a livelihood for farmers, a lifeline for plants and trees and a habitat for organisms small and large. Would we ever consider soil as a vulnerable complex system struggling to cope with human activities? Our needs are ever increasing and our expectations are becoming more and more refined. We, for example, live in a country where 40% of the food on the supermarket shelves is imported, while the same amount is lost as waste. Do we ever consider that our incessant need to keep supermarkets bountiful is putting unsustainable pressure on crop production and diminishing soil quality?

Behind the scenes, we are facing an imminent crisis; we need to double food production on existing soils without over-using chemical fertilisers. This is impossible without some radical change of farming methods. One such proposed solution is to improve the soil by adding to it more soil. More specifically, the brown organic bit called soil organic matter or SOM. The reasons for this is that SOM contains something that plants need but we do not know exactly what, simply because SOM is a very complex mixture, in fact the most complex mixture on this Earth. SOM is like a jigsaw puzzle where every piece plays a role, but we cannot piece it together as we still do not know what the individual pieces are.

SOM of course plays other roles, keeping the soil together, holding water but also acting as a store of contaminants and source of nutrients. Soil, particularly peat, also stockpiles carbon preventing its release in the form of the infamous CO_2. Yet, due to climate changes and human activities, such as planting forests, SOM is disappearing, releasing their carbon. Thus if we can understand the roles of SOM in these processes, we could design safe ways to increase soil fertility, prevent contamination, avoid soil degradation and loss of carbon stores. To do this we must find a way to characterise individual components of soils. In other words, we have to identify forensically the chemical structures of individual molecules. This is a major challenge even for modern analytical chemistry. At present, we can use a powerful analytical technique called Nuclear Magnetic Resonance spectroscopy, or NMR for short, to resolve mixtures of tens of molecules. NMR identifies molecules by inspecting their individual atoms. By placing a sample in a magnetic field and subjecting it to radio waves we can make the carbons and hydrogens sing in their different languages. Even better, each molecule acts as a unique choir, which sings a song that

reveals its chemical structure. However, when listening to many similar molecules, present in complex mixtures, it is impossible to hear the voices from each individual choir; standard NMR methods do not work anymore. There are further issues. In soil mixtures there are many molecules present in very tiny amounts and therefore each song is sung very quietly. In addition, hydrogens are four times louder than carbons, but soil molecules have a carbon skeleton. To deal with these issues we attach tags to certain types of molecule. Their role is to bellow out ballads, which describe their immediate atomic surroundings. We can employ these tags as our forensic spies and by designing clever NMR experiments we hear only the choirs which have our spies amongst their members. In this work certain aromatic molecules called phenolics were tagged using a chemical reaction and each individual melody was separated with the help of specially designed multidimensional (three- or even four-dimensional) NMR experiments. Phenolics are claimed to play a major role in slowing down peat decomposition. Knowing their structure is essential in understanding their function at the molecular level in peatlands.

Towards this goal, the methodology was initially tested on a simple model mixture of 4 or 9 phenolic compounds before applying it to a SOM sample extracted from a Scottish peat bog. By interpreting the obtained lyrical fingerprints, the structures of individual phenolic molecules were identified. This methodology can be seen as the first steps towards unravelling the complex nature of SOM.

The principles behind this methodology are general. Using different tags and different NMR experiments will allow us to complete the molecular jigsaw puzzle of soils. Only then will we be able to explain and predict the responses of soil to our activities, making sure that we do not overexploit or degrade our beloved ‘dirt’.

Abbreviations

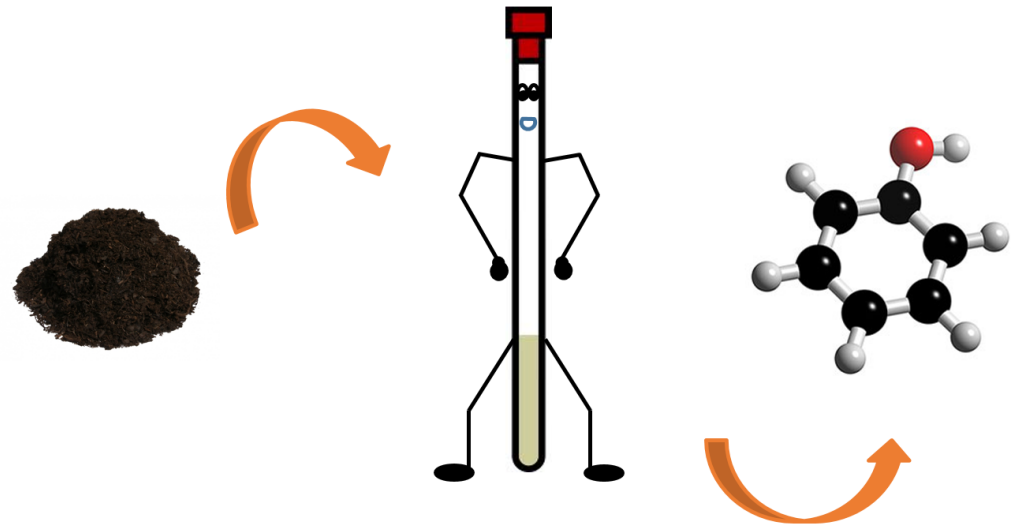
APT :	Attached Proton Test
APCI :	Atmospheric Pressure Chemical Ionisation
APPI :	Atmospheric Pressure PhotoIonisation
BPPLED :	Bipolar Pule Longitudinal Eddy Current Delay
B.S. :	Bloch Siegert
COSY :	COrrelation SpectroscopY
CONTIN :	CONTINuous distribution
CP-MAS :	Cross-Polarisation Magic Angle Spinning
CPMG :	Carr-Purcell-Meiboom-Gill
CRAM :	Carboxyl Rich Aliphatic Molecules
D_2O :	Deuterium Oxide
DANTE :	Delays Alternating with Nutations for Tailored Excitation
DBU :	1,8-Diazabicyclo[5.4.0]undec-7-ene
DEPT :	Distortionless Enhancement by Polarisation Transfer
DMC :	DiMethyl Carbonate
DMF :	DiMethylFormamide
DMS :	DiMethyl Sulphide
DMSO :	DiMethyl SulfOxide
DOC :	Dissolved Organic Carbon
DOM :	Dissolved Organic Matter
DOSY :	Diffusion-Ordered SpectroscopY
DPFGSE :	Double Pulsed Field Gradient Spin-Echo

DQ :	Double Quantum
DQF-COSY :	Double Quantum-Filtered COSY
ESI :	ElectroSpray Ionisation
FA :	Fulvic Acid
FID :	Free-Induction Decay
FT :	Fourier Transform
FT-ICR MS :	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
FT-IR :	Fourier Transform InfraRed spectroscopy
GC-MS :	Gas Chromatography Mass spectrometry
GW :	Ground Water
HA :	Humic Acid
HETCOR :	HETeronucleur CORrelation
HMBC :	Heteronuclear Multiple Bond Correlation
HMQC :	Heteronuclear Multiple Quantum Correlation
HPLC :	High-Performance Liquid Chromatography
HPSEC :	High-Pressure Size Exclusion Chromatography
HR-MAS :	High-Resolution Magic Angle Spinning
HS :	Humic Substances
HSQC :	Heteronuclear Single Quantum Correlation
IHSS :	International Humic Substance Society
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
INEPT :	Insensitive Nuclei Enhanced by Polarisation Transfer
IR :	InfraRed

KM :	Kendrick Mass
KMD :	Kendrick Mass Defect
LC :	Liquid Chromatography
LMW :	Low Molecular Weight
MDLT :	Material Derived from Linear Terpenoids
MS :	Mass Spectrometry
MW :	Molecular Weight
nD :	n-Dimensional
NMR :	Nuclear Magnetic Resonance
NOE :	Nuclear Overhauser Effect
NOESY :	Nuclear Overhauser Effect SpectroscopY
NOM :	Natural Organic Matter
NPK :	Nitrogen Phosphorus Potassium
NSHA :	Nordic Sea Humic Acid
PFG :	Pulsed Field Gradient
PGPB :	Plant Growth Promoting Bacteria
POM :	Particulate Organic Matter
PURGE :	Presaturation Utilising Relaxation Gradients and Echoes
QUAT :	QUATernary only
RF :	RadioFrequency
ROESY :	Rotating-frame Overhauser Effect SpectroscopY
SD :	Standard Deviation
SDBS :	Spectral DataBaSe for organic compounds
SEFT :	Spin-Echo Fourier Transform

SEMUT :	Spectral Editing with a Multiple-qUantum Trap
S/N :	Signal-to-Noise ratio
SOM :	Soil Organic Matter
SPARIA :	Substitution Patterns in Aromatic Rings by Increment Analysis
SPE :	Solid-Phase Extraction
SPFGSE :	Single Pulsed-Field Gradient Spin-Echo
SPR :	Shaped PResaturation
SRDOM :	Suwannee River Dissolved Organic Matter
SRFA :	Suwannee River Fulvic Acid
SRHA :	Suwannee River Humic Acid
T_1 :	Longitudinal relaxation time constant
T_2 :	Transverse relation time constant
TBAH :	TetraButylAmmonium Hydroxide
TD :	Time Domain point
TFF :	Tangential Flow Filtration
TMAH :	TetraMethylAmmonium Hydroxide
TOC :	Total Organic Carbon
TOCSY :	Total Correlation Spectroscopy
UV-vis :	UltraViolet-visible
WATERGATE :	WATER suppression by GrAdient Tailored Excitation
WET :	Water suppression Enhanced through T_1 effects

Chapter 1. Introduction



'Any intelligent fool can make things bigger and more complex….It takes a touch of genius –and a lot of courage -to move in the opposite direction'

Albert Einstein

1.1. Foreword

Soils represent open dynamic systems that are subjected to fluctuations in pH, redox potential, moisture availability, and temperature.[1] Each soil is a unique product of its geological setting, climatic conditions, biota, and duration of pedogenesis. Within soil matrices a series of complex chemical and biochemical reactions continuously act to convert plant and animal biomass into heterogeneous mixtures, termed soil organic matter (SOM). SOM is described as being composed of non-humic and humic substances. Non-humic substances consist of light-coloured molecules of known structure, such as amino acids, peptide/proteins, carbohydrates, lignin and lipids. Most of these substances do not persist in the environment and are either quickly re-mineralised or taken up by plants. Humic substances (HS), on the other hand, are described as a heterogeneous mixture of modified/highly modified organic molecules. These are referred to as refractory, to describe their greater resistance to further decomposition; some remaining in the soil for up to thousands of years before eventually being transformed into carbon dioxide and water as part of perpetual global recycling.[2, 3]

HS have numerous biogeochemical functions hence they have been subject of continued research over the last 150 years. Despite this, their exact molecular composition remains elusive causing a hindrance to the understanding of the structural-functional relationships of HS within different ecosystems.[3] Progress has been slow in this area due to controversies regarding the definition and nature of HS. Recent advances in high-resolution analytical techniques have allowed a better understanding of chemical classes and sizes of HS molecules. However, even today's powerful instrumentation has yet to provide conclusive, unambiguous structures of molecules contained in HS.

The overarching aim of this PhD project has thus been to develop methodologies that will unravel the molecular composition of HS. Towards this end, this study combines chemical modification of the prevailing functional groups of HS and purposely-designed Nuclear Magnetic Resonance (NMR) experiments.

This Chapter provides the necessary background information and review of the relevant literature. Firstly, some definitions are given followed by an overview of the current understanding of the composition of HS. The varied environmental roles of HS are described. Secondly, an examination of Scottish peatlands as well as peat phenolics and peatland plant

phytochemistry rationalises the direction of this study. Thirdly, an overview of existing applications of liquid-state NMR spectroscopy (and to a very small extent mass spectrometry) to HS provides the current status of the research field. This is followed by a summary of chemical modification strategies proposed for HS, before the concepts of NMR experiments are explained. Finally, the aims and objectives of this study are presented.

1.2. Organic matter in nature

The organic material present in terrestrial and aquatic systems is termed Natural Organic Matter (NOM). More specifically, the organic material contained within soil is termed Soil Organic Matter (SOM) or humus. The term humus (Latin for soil or earth) was originally used to describe the whole soil but became a descriptor of organic matter in 1761 and was propelled into common usage by Stevenson's 1994 book, where he defined humus as the decomposed organic material in soil.[2] Despite this, the terms NOM and SOM are used to denote both the decomposed and the non-decomposed organic components, the convention followed in this work. Organic matter found in aquatic systems is most commonly referred to as aquatic NOM and comprises of both Dissolved Organic Matter (DOM) and Particulate Organic Matter (POM). The former is defined as the aquatic NOM that passes through a 0.45 μm filter and includes both colloidal and truly dissolved components. Unsurprisingly, POM is specified as the aquatic NOM material that is retained by a 0.45 μm filter.

1.3. Humic substances (HS)

HS, the complex mixture of organic molecules produced by microbial and abiotic degradation of plant and animal residues, are the major components of organic matter in soil and natural waters. HS represent the main component of NOM, accounting for up to 80% and 60% of SOM and DOM respectively.[4] Due to their complexity, HS are typically not studied as a whole but instead are separated into operational fractions. In the following section the concept of operational fractions is explained followed by an overview of the importance of HS with special focus on their potential for increasing soil fertility. This section will close by presenting the current understanding of the nature of HS in terms of their molecular structure and properties.

1.3.1. Operational fractions of HS

HS were recognised as a complex mixture from the beginnings of scientific research into soil chemistry, which began in the mid-1800s.[5] All attempts to separate individual molecular species of HS have failed to this day. Thus, to simplify HS, at least partially, they have been historically subdivided into three operational fractions based on their solubility under different pH conditions.[6] The traditional operational fractions of HS are as follows:

- Humin - black coloured fraction insoluble at any pH.
- Humic Acid (HA) - brown or grey fraction soluble at high pH, but insoluble at low pH.
- Fulvic Acid (FA) – yellow-brown fraction soluble at any pH.

The procedure to obtain the above fractions from a sample of soil is summarised in Figure 1.1.

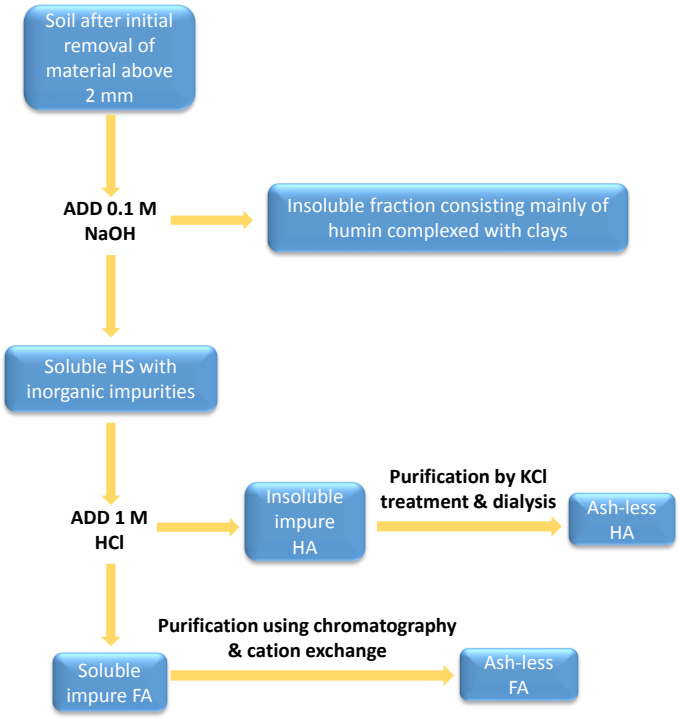


Figure 1.1 Outline of a protocol for the fractionation of soil HS into operational fractions, HA, FA and humin.

It must be reiterated that this division is purely operational and these fractions do not represent unique chemical species. Each fraction, instead, represents a heterogeneous mixture that will include partially degraded organic matter or known biomolecules.[7-9] There is also some doubt

as to whether humin is a separate fraction or just a proportion of HA which is insoluble in 0.1 M NaOH due to its interaction with other soil components, e.g. clay minerals.[10, 11]

It is important to note that the relative amount of each operational fraction of HS varies between soil types[2, 12] indicating that the processes generating HS are dependent on the local conditions. Similarly to SOM, DOM is routinely separated into HA and FA.

1.3.2. Importance of soil HS to Earth's ecosystems

HS account for a sizeable proportion of Earth's carbon pool: up to 80% of the 1550 Gt of the global carbon stored in SOM is estimated to exist in the form of HS.[4, 13] Understandably, HS are an important player in the carbon cycle, as well as other global biogeochemical cycles, where they contribute significantly to the fixation and release of CO_2.[1] Their ability to act as a long-term sink of carbon[14] has highlighted a potential role for HS in future measures to counteract the effects of global warming.[13, 15] Indeed, enhancing the carbon sequestration capacity of agricultural soils,[16] forest and grassland soils[15] has been proposed as a method for offsetting global CO_2 emissions by 5 to 15%. These propositions utilise the fact that HS are the relatively more stable components of SOM. For example, ^{14}C dating of HS extracted from different depths of a Taiwanese soil determined mean residence times of 143 to 1740 yr and 253 to 2200 yr for FA and HA, respectively (with increasing times correlating with increasing depth).[17] For comparison, the typical turnover rates of amino acids in soils are hours to days, while polysaccharides last for weeks to months, depending on the heterogeneity of the polymer. Thus, the prevalence of HS is key to soil's ability to maintain its carbon content.

HS are an important part of soil aggregates, functioning together with clay minerals to create unique and stable structures that are permeable to waters and, if not water logged, accessible to air. These aggregates give the soil its structural integrity[2] allowing it to act as a habitat for flora and fauna. The high surface area of HS enhances the water retention ability of soil,[10] while the acidity of HS provides soils with a high cation exchange capacity[2] e.g. 300-700 meq/100 g with Ca^{2+} at pH 7 (in absence of clay materials).[18] This also allows soils to buffer against sudden changes in pH.[2] The presence of carboxylic acid groups in HS compounds gives soil the ability to act as a weak acid exchanger[19] controlling the binding, storage and release of metals.[10] For example, HS can bind excess Al, thus decreasing its availability to plants.[10] HS also have the capacity to reduce toxic forms of metals such as Cr(VI) to the nontoxic Cr(III)[10, 19, 20] or to form

complexes with heavy metals/metalloids, such as arsenic,[21] or radionuclides, such as uranium,[22-24] and immobilise them.[25]

In addition to metals, other inorganic or organic molecules can form water soluble (or insoluble) associations with HS. Much interest in this area lies in the interaction of HS with xenobiotics. These compounds, which include pesticides or polyaromatic carbons, can be potentially removed from sewage or water systems using HS.[23, 26]

While it is evident that the interactions of HS with inorganic or organic pollutants are important, the detailed molecular characterisation of these binding events is absent, simply because the molecular composition of HS in unknown. It is often speculated that the binding mechanism involves hydrophobic/hydrophilic or cation exchange interactions, dispersive forces, hydrogen bonding, charge transfer or chelation.[23] Aliphatic/aromatic components of HS, in addition to their acidic functional groups, are said to be significant.[19] Nevertheless, the exact molecular nature of these interactions remains elusive and understanding of molecular make-up of soil will have major implications for assessment of contaminant risk and is a prerequisite for rational designing of remediation protocols.[27]

1.3.3. The role of HS in soil fertility

The role of HS in soil fertility has recently become very topical. As the quest for new farming technologies has been intensified by the need to respond to the projected global population growth to 11 billion by 2100,[28] crop production has to double in order to maintain worldwide food security.[29, 30] This can be achieved either by enhancing crop growth on the existing agricultural land, or by obtaining new land mainly through the reduction of the rainforest. The former solution is more ecologically sound although represents the more difficult challenge.[30, 31] Increasing the application of Nitrogen, Phosphorus, Potassium (NPK) fertilisers seems the easiest course of action to enhance crop growth on existing farmland but there are several issues to be considered. Firstly, the current overuse of these chemicals has already resulted in diminishing crop yields as well as increasing environmental problems, such as algae production in natural waters, augmented by fertiliser run-off from agricultural land.[32] Secondly, the accelerated depletion of global phosphorus reserves mean that the application of NPK fertilisers will become increasing difficult to support.[33] These facts strongly suggest that the use of NPK fertilisers as a

means of doubling crop production over the required timescale is not feasible and a more sustainable solution needs to be sought urgently.

Addition of HS to soil is considered as one possible solution for reducing the need for NPK fertilisers. This is not a new proposition, and much of the literature in this area dates from before the 1990s. HS are reported to increase soil fertility indirectly or directly. Indirect effects are related to the ability of HS to influence soil compaction, water retention or their interaction with nutrients, while the direct effects involve changes to plant respiration, photosynthesis, and root growth/architecture due to the interaction of plant cells with HS molecules. While the former effects are widely accepted, the latter are poorly understood.

There are numerous studies recording an increases in the amount of N, P, K, Ca, Mg and Fe in plants by applying small amounts of HA or FA to tobacco plants,[34] cucumbers,[35] gerbera[36] and tomatoes,[37] even in soils under saline stress,[38] to name but a few. For example, the treatment of corn seedlings with HS increased mobilisation of P and Fe[39] aiding their translocation throughout the plants.[40] While these studies promote the benefits of HS, the effects on nutrient bioavailability are reported to be dependent on the concentration of operational fractions used, with higher concentrations of HA and FA reported to decrease nutrient assimilation.[34, 36] It has also been suggested that FA is a better promoter of plant uptake due to its higher COOH content and thus binding capabilities.[41]

Historical accounts indicate that HS increases respiration and photosynthesis, with photosynthetic behaviour related to an increase in chlorophyll content.[42, 43] Contrastingly, Liu *et al.* reported increased photosynthesis only under certain concentrations of HA with the chlorophyll content remaining the same.[44] With no recent comprehensive studies proving direct or indirect influences,[45] the effect of HS on photosynthesis or respiration of plants remains elusive.

The promotion of root growth by HS, which can be classified as a direct effect, is most likely related to a hormonal effect, specifically the promotion of H^+-ATPase activity in plant cell membranes. A study of Canellas *et al.* demonstrated that HA both alone and supplemented with citrate increased the activity of plasma membrane H^+-ATPase in microsomal vesicles from maize roots.[46] The stimulation of H^+-ATPase induced by HA alone was around 50%, and it was increased to 80% in the presence of 0.005 mM exogenous citrate. The authors suggested that the modulation of H^+-ATPase activity may represent a mechanism by which the HA interacts with

plant roots, enhancing the electrochemical H^+ gradient and inducing the opening of citrate-permeable anion channels. Citrate exuded from the roots would then release auxin-like groups from HA, which in turn would further activate H^+-ATPase promoting root growth, according to the so called acid growth theory.[47] This theory was indirectly corroborated by NMR[48] and GC identification of the auxin molecules in HS samples.[49]

Despite the exact mechanism being unknown, the fact that a number of studies have reported an increase in plant growth upon addition of HS has sparked interest in industrial production of HS-containing products for agricultural use. A particular example includes the use of HA, extracted from Leonardite, an oxidation product of lignite, or peat,[50] as a fertiliser ingredient. While some studies demonstrate that these HS-based fertilisers stimulate plant growth and uptake of nutrients[37, 43] others find insignificant increases.[51] A possible reason for the lack of conclusive results could be down to the varying amount of HS in the commercial products or the nature of the HS itself; with Macarthy *et al*.[52] reporting that composition of Leonardite HS is very different to soil HS. Soil type is another important factor that may affect the success of humic-based fertilisers. For example, Kaplan *et al*.[53] tested liquid and solid forms of humic-based fertilisers and reported insignificant increase in nutrient uptake by strawberry plants grown on calcareous soil.

Another promising approach is the combination of HS with plant growth-promoting bacteria (PGPB). PGPB are a wide range of microorganisms that induce plant growth by several processes including fixation of N_2, increasing nutrient availability, enlarging root surface area, and enhancing beneficial symbioses for the host.[54] An example of this new biofertiliser concept was the combination of HS and endophytic diazotrophic bacteria proposed by Canellas *et al*.[47, 55] The results of their field studies, which showed increased maize bacterial colonisation upon addition of diluted HS. Understanding of these results on a molecular level could have far-reaching consequences.

The increasing focus on 'green' fertilisers has brought forward the idea of using HS isolated from vermicompost. Vermicompost is produced by the decomposition of organic waste by worms and has been shown to be an effective fertiliser for a number of crops, such as potato,[56] lettuce,[57] tomato[58] and strawberries.[59] The ability of vermicompost to act as fertiliser has been put down to the production of plant growth regulators, which are believed to be co-extracted with HS. This

is supported by the fact that HS extracted from vermicompost has been shown to increase root length, root weights, chlorophyll content, and importantly the yield (Table 1.1).[60]

Table 1.1 Reported enhancements after the addition of HS from vermicompost. Reproduced from ref.[60]

Crop type	Enhancements by HS extracted from vermicompost	References
Basil	Yield and chlorophyll content	Befrozfar *et al.*[61]
Pepper	Number of fruits, plant height and root weight	Arancon *et al.*[62]
Strawberry	Yield and number of fruits/plants	Arancon *et al.*[63]
Tomato	Yield and number of fruits	Atiyeh *et al.*[64]

The best results reported in terms of crop yield were from the combination of vermicompost-derived HA and PGPBs, which led to a 65% and 30% increase in maize and sugarcane yields, respectively.[47, 55, 65]

Although positive results have been reported in literature regarding the potential use of HS in enhancing crop yield, it should be noted that soil fertility depends on the complex interplay of mechanisms. While researchers believe that these are potentially controlled by HS, others have pointed out that progress in this area requires deciphering of the causative relationship, preferably at the molecular level.[45, 48] The lack of the molecular level information is thus hindering development of future HS-based fertilisers, a promising component of a sustainable solution to global food security.

1.3.4. Structural concepts of HS

The molecular structure of HS has been the subject of much controversy. The long-standing view is that they are polymeric systems with variable conformation depending upon the pH, concentration, and the ionic strength.[66, 67] HS at low pH, high concentrations, and high ionic strengths, were believed to exist in a random coil globular conformation,[68] while HS at high pH, low concentrations, and low ionic strengths, were believed to form flexible linear colloids.[2, 9, 67]

General acceptance of this 'polymer' theory came from the similarities in the measured physical properties of HS to those of biomolecules. For example, electrophoretic mobility, as well as sedimentation velocity studies were compatible with the existence of random coil

macromolecules with mass-weighted average molecular weights (MW) of up of to 50 kDa and radii of 4-10 nm.[9] The polymer theory was also consistent with the resistance of HS to further degradation and theories regarding their creation. The origin of the polymers was explained by well-known polymerisation reactions, such as the Maillard reaction.[69] This abiotic reaction occurs between amino acids and monosaccharides producing melanodins, which account for the observed dark colour of HS.[70] As a consequence, this theory has led to the creation of many pseudo-structures providing an ''average HS structure'' which many still use as descriptors of HS today (Figure. 1.2).

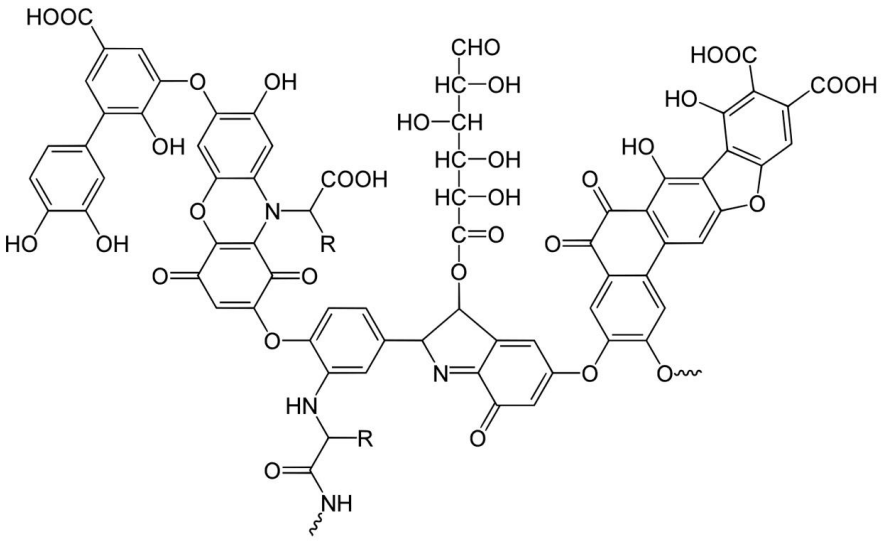


Figure 1.2 A molecular model of the 'average structure' of HS. Reproduced from ref.[2]

Despite the wide acceptance of the polymer theory of HS, there has yet been any scientific proof for the existence of polymers in HS fractions. In addition, this theory has a series of shortcomings. For example, different techniques gave widely different MW values. Equations at the heart of the interpretation of some measurements were used on seemingly polydisperse HS samples.[67] Moreover, the proposed HS formation reactions such as the Mallaird reaction utilise starting materials, monosaccharides and amino acids, that are rapidly lost from the soil and therefore unlikely to be available over the relatively long timescales required for the reactions to occur. Of course, there is always a fresh supply of these materials but not on the scale required to account for HS formation. In addition, this reaction requires an alkaline pH which would equate to acidic soils having less HS than alkaline soils, which is not the case.[70]

Applications of advanced spectroscopic, chromatographic and mass spectrometry (MS) techniques, which were not available during the early studies, have led to a new structural concept of HS. It is now proposed that HS exist as a heterogeneous mixture of molecules closely resembling the structures of their plant or microbial precursors.[71, 72] These molecules are believed to form supramolecular associations stabilised by hydrophobic dispersion forces and hydrogen bond formation.[67] These associations were first cited by Wershaw[67, 73] and ascertained by high pressure size-exclusion chromatography experiments (HPSEC).[9, 67, 74] These experiments revealed that the addition small organic acids to the HS samples decreased their observed molecular weights, consistent with break-up of molecular aggregates. Furthermore, the disaggregation was reversed by increasing the pH of the solution. 2D diffusion ordered NMR spectroscopy (DOSY) confirmed the presence of smaller molecules in HA. Diffusion coefficients measured after the addition of a well-known protein-denaturing agent, acetic acid, to samples of HA, corresponded to molecules with MWs in the range of 200-2500 Da.[75] Both the HPSEC and DOSY studies also unveiled that FA was less associated than HA. Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS) provided the ultimate proof that HS exists as a mixture of small and large compounds. Using soft-ionisation methods, incapable of inducing fragmentation of covalent bonds, the spectra obtained contained an envelope of ions, indicating the presence of many thousands of molecules in samples of FA/HA. The MW ranges reported were between 200-1000 Da with the average MW of 300-500 Da and 400-700 Da for FA and HA, respectively.[76, 77]

With the ever-increasing evidence for the concept of supramolecular association of smaller molecules, researchers have now formulated new definitions of the operational fractions of HS. HA is now described as consisting of associations of more hydrophobic molecules stabilised by dispersion forces at neutral or basic pH. In contrast, FA is thought to consist of hydrophilic molecules with less association.[78] Both HA and FA contain large and small molecules with FA having on average a greater proportion of smaller molecules than HA.

As part of the molecular association model, it is believed that HS can form micelles,[79] as first hinted by observations of lowered water surface tension in the presence of HA.[7, 80] Fluorescence studies of the fluorescent probe pyrene in an HS solution established that these HS micelles have inner hydrophobic cores surrounded by hydrophilic coatings. This was concluded based on the fact that fluorescence from pyrene was not quenched when magnesium bromide was added to pyrene/HS solution, meaning that the probe was encapsulated inside the hydrophobic core of HS

micelles.[81] The nature of the inner hydrophobic cores of micellar HS structures was further investigated by High-Resolution Magic Angle Spinning (HR-MAS) NMR. The aromatic proton signals of HS, without operational fractionation, extracted from forest soil were observed when the sample was suspended in DMSO-d_6 but not when suspended in D_2O. As DMSO-d_6 dissociates hydrogen bonds, the authors concluded that aromatic moieties were being protected in the hydrophobic cores of HS micelles in D_2O.[9, 82]

The formation of molecular aggregates or micelles in solutions of HS has important consequences for the interpretation of physical measurements. Depending on the solution conditions, the aggregates will change the shape and properties of HS producing 'unique' superhumic structures. Consequently, methods such as size exclusion or the use of the E4/E6 formula cannot correctly interpret the MW of HS samples.[67, 83] In order to analyse individual molecules these associations and micelles need to be broken.

If HS form molecular aggregates or micelles in solution it becomes questionable if there is anything unique about the molecules making up these aggregates, do HS actually exist? Although known biomolecules are not regarded as HS molecules,[2] their presence in HA/FA extracts is supported by MS and NMR data.[70] Biomolecules such as proteins decompose readily in soils, thus are not refractory enough to dominate the properties of HS. However, if HS molecules were to associate or form micelles with protective hydrophobic cores, it is then possible that biomolecules survive in the environment and hence contribute to the observed properties of HS.[67] Others advocate that HS are purely mixtures of plant/microbial carbohydrates, proteins, lipids, as well as degraded lignin, tannins, melanins and polyketides all produced in accordance with the prevailing ecological conditions.[70-72] While it is generally accepted that HS contains molecular associations of small molecules, polymers could still be present.[84] These polymers could be formed by a combination of Mallaird/polyphenol reactions.[69] A completely different view of HS structure was presented by Tan *et al*.[10] Using Scanning Electron Microscopy the authors propose that HA forms ordered nano-assemblies, suggesting that the associations while weak create ordered entities.[10]

From the above discussion, it is evident that much confusion still exists in defining what constitutes HS. The modern analytical techniques described therein allude to the presence of a range of molecules but the complexity of these mixtures keeps the composition of individual molecules hidden. What is clear is that there is no evidence to suggest that HS exist exclusively

as polymers or clear evidence to support that HS contain unique chemical compounds. Only by characterising structures of individual components can the field make conclusive arguments about the true molecular nature of HS.

1.3.5. Bulk properties of HA and FA

As discussed in the previous section, a comparison of HA and FA samples has established that HA is generally composed of higher MW compounds than FA. In terms of elemental composition, HA has a higher carbon and lower oxygen content than FA (Table 1.2).[2]

Table 1.2 Average elemental composition (expressed as percentage per mass with standard deviation (SD)) of HA and FA isolated from different sources. Reproduced from ref.[85]

	Source	C (% ± SD)	H (% ± SD)	O (% ± SD)	N (% ± SD)
Humic Acid	Soil	55.4 ± 3.8	4.8 ± 1.0	36 ± 3.7	3.6 ± 1.3
	Freshwater	51.2 ± 3.0	4.7 ± 0.6	40.4 ± 3.8	2.8 ± 1.6
	Marine	56.3 ± 6.6	5.8 ± 1.4	31.7 ± 7.8	3.8 ± 1.5
	Peat	57.1 ± 2.5	5.0 ± 0.8	35.2 ± 2.7	2.8 ± 1.0
Fulvic Acid	Soil	45.3 ± 5.4	5.0 ± 1.0	46.2 ± 5.2	2.6 ± 1.3
	Freshwater	46.7 ± 4.3	4.2 ± 0.7	45.9 ± 5.1	2.3 ± 2.1
	Marine	45.0 ± 4.0	5.9 ± 0.9	45.1 ± 6.0	4.1 ± 2.3
	Peat	54.2 ± 4.3	5.3 ± 0.5	37.8 ± 3.7	2.0 ± 0.5

The elemental compositions of HA and FA from different sources appear similar, indicating that bulk characteristics of HS are similar regardless of their origin (see Table 1.2). It should be noted that these figures are averaged from around 50 samples and the exact elemental composition may differ depending on local conditions, particularly for marine samples where levels of allochthonous and autochthonous inputs into water systems vary. Nitrogen content is low in all samples due to the assimilation of nitrogen substances such as amino acids. Unsurprisingly, the greatest nitrogen content is found in marine HA and FA, due to the high abundance of marine microorganisms in seawaters. Despite this, the elemental composition shows that in general HS are mostly made up of CHO compounds, a fact supported by MS studies.[77]

The differences in elemental composition can be related to the amount of functional groups present in individual fractions, of which carboxyl (COOH) and hydroxyl (OH) groups are considered the most abundant.[86] Since these groups are highly important for HS functions, such as intermolecular interactions, many studies are aimed at quantifying their proportion in FA and HA.[2, 87, 88] The content of COOH groups is typically measured indirectly using potentiometric titrations, e.g. the $Ca(OAc)_2$ method.[88] However it has been reported that the outcome of these indirect potentiometric studies depend on a number of factors, such as HS concentration and method of titration,[89] issues resolved by direct titration methods. An example of a direct titration study of several HS samples using 0.1 M NaCl reported an average COOH content of 6.6 ± 0.8 meq g^{-1} and 4.8 ± 0.4 meq g^{-1} for FA and HA, respectively. In addition, the content of these groups was higher in terrestrial FA than aquatic FA, with the opposite trend reported for HA. The phenolic contents were also measured as 1.4 ± 0.3 meq g^{-1} C^{-1} and 1.4 ± 0.4 meq g^{-1} C^{-1} for FA and HA, respectively.[86] Methylation in combination with gas chromatography has been tested as an alternative method for measuring COOH content, however only 80% of the COOH groups determined by titrations were observed.[90] Regardless of the procedure used, it is generally agreed throughout the literature that HA has lower COOH content than FA providing reasoning why the acidity of HA is generally lower than FA for soils, excluding spodosols, which are highly acidic.[10]

1.4. Importance of peat bogs and their phenolics

This section will provide a brief account of some important features of Scottish peatlands, discuss the role of peat bogs as a carbon sink, highlight the role of phenolics in the humification process and survey their occurrence in peatland flora. Importantly, it will articulate the reasons for studying peat phenolics.

1.4.1. Scottish peat bogs

The two most common types of peat bog in Scotland are the blanket bog, which cover most of the upper moorlands of Scotland and its isles, and the raised bog, commonly found in the central belt, Dumfries and Galloway, and Grampian plain with a few scattered locations in parts of the Western Highlands. Both types receive moisture from rainfall. One other type of peat, namely fen, occurs mostly in the Scottish Borders and on floodplains, hence receiving moisture also from groundwater. The subject of this study is a raised peat bog.

It is currently estimated that over 80% of Scotland's peat is in decline.[91] Reasons for this damage include:[92]

- Changes in land use whether commercial exploitation, wind farms, intensive farming or afforestation.
- Erosion by grazing animals, humans or weather.
- Drainage, which is man-made or climate-driven.
- Pollution from heavy industry.

There are growing incentives to restore peatlands throughout Scotland, highlighted in a number of consultation documents such as the Scottish Soil Framework[93] and strategies such as the 2020 Challenge for Scotland's Biodiversity.[94] These are mainly driven by the need to maintain the UK CO_2 emission targets; this could be widely helped by maintaining the health of peatlands. It is estimated that a 5% loss in UK peatland would equate to the entire annual UK CO_2 emissions.[95]

The other concern associated with degraded peatlands is water quality, with 70% of the UK drinking water originating from areas dominated by peatlands, and with the amount of DOC (and discolouration) in UK natural waters doubling in the last 30 years.[96] There are strict regulations on the amount of DOC in drinking water and thus these increased levels have to be reduced with more efficient and expensive removal technologies.

1.4.2. Peat bogs as a carbon sink

Northern peatlands cover an area of around 400×10^6 ha, and currently store around 547 Gt of organic carbon as waterlogged peat.[97] Scotland has the highest percentage of peatland cover in Europe, covering an estimated 1.8×10^6 ha. These peatlands store approximately 1.62 Gt of carbon, making them a very important carbon sink.[98, 99] Most of this carbon stored in the form of organic carbon. The organic content of peat in Scotland is high, ranging from 20% for a 'peaty' soil[91] to at least 60%-100% for peat.[99]

Peatlands are important sources of information on past and future climate changes because they respond strongly to changes in environmental conditions, in particular to hydrology.[100, 101] Historically, northern peatlands have removed CO_2 from the atmosphere faster than it has been rereleased and now contain 20–30% of the world's soil carbon stock,[102] which equates to over 60% of the atmospheric carbon pool.[103] Increasing temperatures, altered precipitation regimes

and elevated atmospheric CO_2 concentrations have complex effects on peatland carbon cycling.[104]

Peat contains a mixture of decomposed and partially decomposed material in a waterlogged anoxic environment. The rate of peat humification is dependent on the hydrology (water table) of a given ombrotrophic peat profile. The near surface layer above the water table, or held under temporary unsaturated conditions, is exposed to oxygen. This acrotelm region experiences the highest rates of decomposition supplying material for the anoxic permanently waterlogged catotelm.[105] Changes in hydrology and climate also change the predominant vegetation. Mosses and lichens prefer wetter conditions compared to vascular plants such as heather. The predominant vegetation then in turn controls the oxygen supply with mosses creating degradation resistant watertight blankets cutting the underlying peat off from a supply of nutrients as well as air. Drying of peat bogs leads to their destruction and release of the stored carbon. The reason for this is that once oxygen is allowed into the peat system, microbes, which are mainly dormant in anaerobic conditions, speed up the humification process.

1.4.3. The role of phenolics in the humification process in peat bogs

Phenolic compounds are ubiquitous to all wetlands, and have attracted intense interest as potent inhibitors of hydrolase enzymes,[106] which play a major role in humification. As a consequence of this inhibitory activity, the phenolics themselves, but also other organic molecules are degraded more slowly in the presence of phenolic compounds. An exception is phenol oxidase, an enzyme able to degrade these recalcitrant materials.[107] However, phenol oxidase requires bimolecular oxygen to function – a scarce resource in waterlogged peat.[108] It has been proposed that the low rate of biodegradation in peatlands is due to oxygen constraints on phenol oxidase, which allow phenolic materials to accumulate and inhibit pivotal hydrolase enzymes.[109] It is therefore important to identify what these compounds are and source where they are coming from.

1.4.4. Phenolics in peatland flora

It should be noted that previous studies have rebutted any link between vegetation and HS composition,[110] although a number of recent reports suggest an association between the HS molecular composition and the overlying plant types.[111] Given the fact that phenolic compounds

are widely distributed throughout the plant kingdom, constituting up to 60% of plant dry mass,[112] it is not surprising that they are found in soil.

Due to the loose definition (presence of at least one aromatic ring with a hydroxyl group) a large number of molecules belong to this group. Importantly, phenolic compounds constitute the building blocks of lignin - complex organic polymers forming the main classes of structural materials in the support tissues of vascular plants and some algae. The three common lignin monomers, monolignols, are paracoumaryl alcohol (Figure 1.3a), coniferyl alcohol (Figure 1.3b) and sinapyl alcohol (Figure 1.3c).

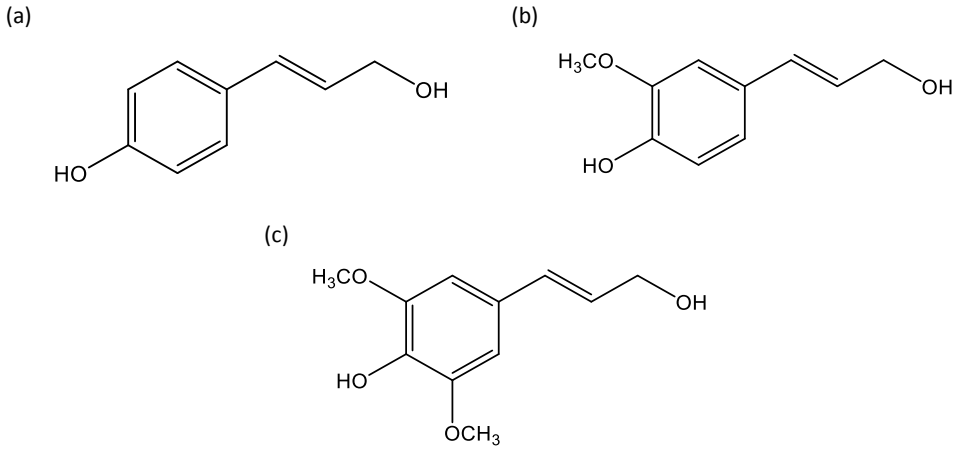


Figure 1.3 The three main monolignols (a) p-coumaryl alcohol; (b) coniferyl alcohol and (c) sinapyl alcohol.

Note the degradation products of lignin containing these building blocks are typically reported in the literature as p-hydroxy (or 4-hydroxy) phenols, vanillyl phenols and syringyl phenols.

The composition of lignin varies depending on plant groups such as angiosperms and gymnosperms. For example, angiosperms contain both vanillyl and syringyl phenols while gymnosperms contain mostly vanillyl phenols.[113]

Non-vascular plants do not contain lignin, however, they do contain aromatic compounds, of which phenolics constitute the bulk of the secondary metabolites.[114]

Of course, plants are not the only source of HS and compounds other than phenolics are found. Microbial decay products are another important source of HS.[71] The decay of microbes releases mostly proteins, nucleic acids, hetero- and homo- polysaccharides such as chitin, polyphenols, lipids and melanins. Some believe that the amount of compounds in soil derived from microbes is underestimated.[71, 115] NMR has identified proteinaceous and polysaccharide materials such as

peptidoglycan and lipoproteins, small short chain lipids and peptides of microbial origin.[71, 116, 117] Plants also release their own set of polysaccharides and large lipids like cutin.

The vegetation cover of peat plays an important role as a source of organic materials as well as a protection against the elements. Preeminent peat bog species of Scotland consist of non-vascular plants such as *Sphagnum* mosses (Figure 1.4a), *Lichen* and vascular plants such as *Calluna vulgaris* (Common Heather) (Figure 1.4b) and *Eriophorium vaginatum* (Cotton Grass) (Figure 1.4c). *Drosera rotundifolia*, the round-leaved sundew or common sundew (Figure 1.4d), is a carnivorous plant often found in bogs, marshes and fens. Out of all these species, the litter produced by the peat-forming moss, *Sphagnum*, is the dominant input of organic carbon into ombrogenous bogs and some geogenous fens.[118, 119]

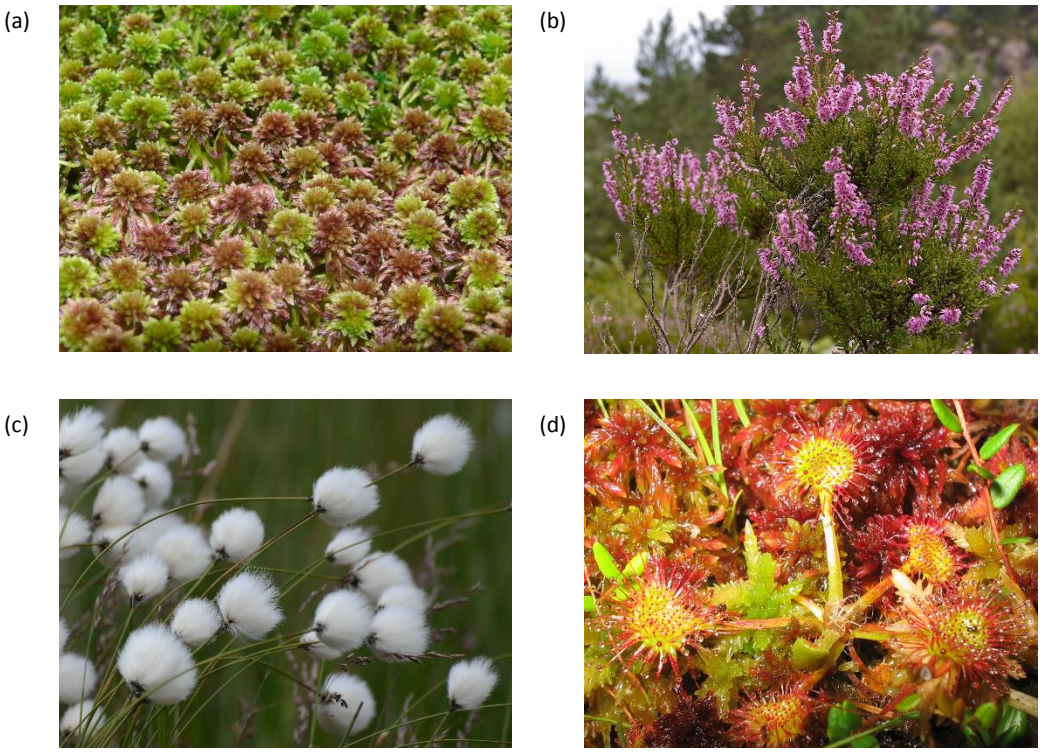


Figure 1.4 Examples of peat flora. (a) *Sphagnum capillifolium* (red peat moss),[120] (b) *Calluna vulgaris* (common heather),[121] (c) *Eriophorum vaginatum (tussock cotton grass),*[122] (d) *Drosera rotundifolia* (common sundew).[123]

Sphagnum does not contain lignin in its cell walls,[105] instead 4-hydroxy phenols such as sphagnum acid (3-(4-hydroxyphenyl)pent-2-enedioic acid) (Figure 1.5a), 4-hydroxybutenolide (Figure 1.5b), 1-(4-hydroxy-3-methoxyphenyl)ethanone (Figure 1.5c) and 4-hydroxybenzoic acid (Figure 1.5d) are the main aromatic components of the cell walls.

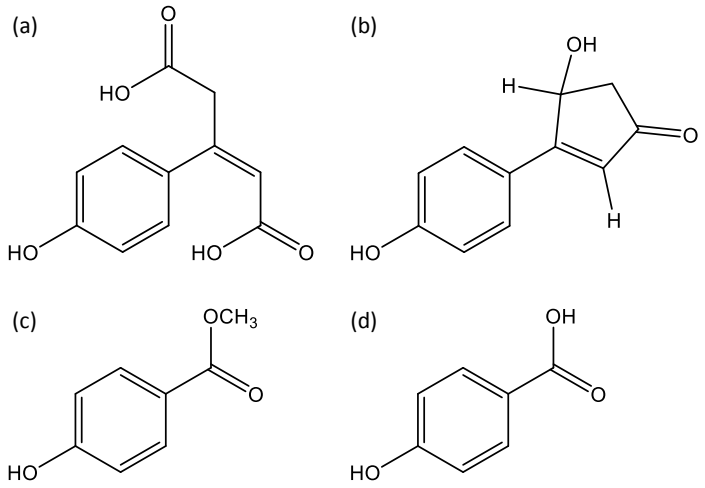


Figure 1.5 Compounds isolated from *Sphagnum* moss. (a) sphagnum acid; (b) 4-hydroxybutenolide;(c) 1-(4-hydroxy-3-methoxyphenyl)ethanone; (d) 4-hydroxybenzoic acid.

A Liquid Chromatography Mass Spectrometry (LC-MS) study of *Sphagnum papillosum* extracts also identified 1-(4-hydroxy-3-methoxyphenyl)ethanone and 4-hydroxybenzaldehyde.[124] These endogenous compounds exist either bound to other biopolymers or in water-soluble form as shown in the study of the compartmentalisation of phenolic constituents in *Sphagnum*.[125] The content of sphagnum acid can be monitored through its thermochemolysis products in a GC-MS experiments, making it a biomarker phenol specific to *Sphagnum*.[118] Other mosses have their own biomarkers, e.g. *Polytrichum commune* produced 3-(3,4-dimethoxyphenyl)-propanoic acid methyl ester on TetraMethylAmmonium Hydroxide (TMAH) treatment.[118] It has been suggested that sphagnum acid protects the sphagnum cell polysaccharides from microbial attack. Its removal under oxic conditions (periods of unsaturation) exposes the polysaccharides for attack once water logged conditions return.[118]

It should be noted that another line of thought suggests that cell-wall polysaccharides of *Sphagnum* play in their own right an important role in decay resistance of *Sphagnum* and were found to actively depress decomposition in vitro.[101, 126] Cell-wall holocellulose of *Sphagnum* mosses contains pectin-like rhamnogalacturonan I type polysaccharides together with xyloglucomannan and cellulose. When the cell wall undergoes autogenic acid hydrolysis, fragments of these pectin-like polymers are slowly released into the environment as so-called sphagnan. Uronic acids contained in the free or cell-wall bound sphagnan may either directly or indirectly inhibit microbial activity.[127]

Despite the fact that mosses do not contain lignin, a recent study using TMAH chemolysis detected lignin-derived phenols. As these phenols could not be formed by the moss, it was proposed that they were removed from the DOM that percolated into the hyaline cells (which make up about 80% of the plant's volume) and became clathrated and physically bound to the *Sphagnum* cell walls.[118] This makes the identification of biomarkers of *Sphagnum* and tracing their fate in the environment rather difficult.

Drosera rotundifolia contains phenolics,[128] although their detailed analysis is not available. Another species of sundew, *Drosera capensis*, was shown to contain a range of phenolics, gallic acid (3, 4, 5-trihydroxybenzoic acid) in particular.[129]

Another non-vascular species found commonly at peat bogs is lichen. In terms of specific lichen metabolites, 3-methoxy-5-methylphenol has been identified.[130]

Calluna vulgaris and *Eriophorium vaginatum* are vascular angiosperms (flowering plants) commonly found on the surface of peat. Both species flourish in drier spells compared to the water-loving mosses. *Culluna vulgaris* is a dwarf shrub dominating Scottish heath and moorlands. A very good review of the phytochemistry of *Culluna vulgaris* by Monschein *et al.*[131] lists flavonoids, chromones, catechins, procyanidins, athrocyadins, lipids, steroids, triterpenes as well as phenolic acids, phenols and phenol glycosides as heather metabolites. Due to its vascular nature heathers contains lignin.[132] The major aromatic compounds found in heather are listed in Table 1.3.

Table 1.3 List of known compounds isolated from heather. Reproduced from ref.[131]

Common Name	IUPAC Name
Caffeic acid	3-(3,4-Dihydroxyphenyl)-2-propenoic acid
Chlorgenic acid	(1S,3R,4R,5R)-3-([(2E)-3-(3,4-Dihydroxyphenyl) prop-2-enoyl]oxy)1,4,5-trihydroxycylohexanecarboxylic acid
p-Coumaric acid	(E)-3-(4-Hydroxyphenyl)-2-propenoic acid
p-Coumaroylquinic acid	(1R,3R,4S,5R)- 1,3,4- Trihydroxy- 5- ([3- (4-hydroxyphenyl)prop- 2-enoyl]oxy)cyclohexanecarboxylic acid
Dihydroxytoluene	3-Methylbenzene-1,2-diol
Ferulic acid	(E)-3-(4-Hydroxy-3-methoxy-phenyl)prop-2-enoic acid
Gentisic acid	2,5-Dihydroxybenzoic acid
Hydroquinone	Benzene-1,4-diol
Isochlorogenic acid	(1S,3R,4S,5R)-3-([3-(3,4-Dihydroxyphenyl)acryloyl]oxy)-1,4,5-trihydroxycyclohexanecarboxylic acid
Protocatechuic acid	3,4-Dihydroxybenzoic acid
Syringic acid	4-Hydroxy-3,5-dimethoxybenzoic acid
Vanillic acid	4-Hydroxy-3-methoxybenzoic acid

Table 1.3 shows a number of lignin-derived phenols such as syringic and vanillic acid as well as other lignin-related compounds. In common with *Sphagnum* (as well as cotton grass) is p-coumaric acid. Tannins are also more abundant in ericoids like heather providing sources of cathecols and aromatic acids such as gallic acid.[133, 134] Para-hydroxy phenols are produced by heather plants but the majority comes from mosses.[135]

Eriophorum augustifolium (common cotton grass) and *Eriophorum vaginatum* (tussock cotton grass) are two versatile sedges populating peat bogs. Not many studies have focused on the phytochemistry of the cotton grass. The most informative was the investigation of progressive stages of decay of *Eriophorum vaginatum* using alkaline extraction. Epoxyhydroxy acids,

dihydroxyacids and aromatic acids (specifically ferulic acid and p-courmaric acid) were identified and found to degrade quickly. Contrastingly, fatty acids and long chain hydroxyl acids were more resistant to decay.[136] A series of flavonoids were also isolated from the stem of cotton grass.[137] Schellekens *et al*.[133] claimed that ferulic acid dominates in low lignin containing species such as cotton grass and hence has been identified as a biomarker of this species. Upon pyrolysis they found that high ferulic acid, and also p-courmaric acid, content in cotton grass contributes to high amounts of 4-vinylguaicyl and p-hydroxyphenol, respectively. Ferulic acid was also identified in heather (Table 1.3) however, the higher lignin content in heather means that the other phenols overshadow its presence.

In general, increased amounts of dimethoxy phenols are taken as an indication for the presence of vascular plants typically associated with drier periods, which promote vascular plant growth. Considering that the major source of these phenols is lignin, it is important to understand how lignin decays in peat.[113] Lignin is preferentially lost over other more resistant compounds such as long chain aliphatics in oxic conditions when white rot is present.[101, 138] Under anaerobic conditions, i.e. in greater peat depths, lignin is not mineralised.[119] This does not mean that the lignin units are preserved under these conditions but instead undergo structural changes, with demethylation of methoxy groups being one of the important reactions.[139] Zaccone *et al.* noted that these changes occur with the accumulation of hydroxyl phenyls, increasing with peat depth.[119] The same author reported that HA samples extracted from peat followed the same trends, thus providing evidence that the diagenetic changes in the source materials are the direct input of humic material.[119]

It is therefore evident that the aromatic compounds in peat HS are plant-derived. This does not necessarily restrict their source to lignin, as phenols are also present in non-vascular plants. Considering that lignin content is used to represent aromatics in plant-soil interaction models, this assumption could be inaccurate and other sources should also be taken into account.[113] One complication that is inherent to these studies is the demethylation reactions that blur the distinction between the lignin- and non-lignin-derived phenols. This could be overcome by using a ^{13}C-labelled TMAH in GC-MS.[140, 141] Similarly, the NMR methods proposed here are capable of making a distinction between methylated and non-methylated aromatics.

The methods for characterisation of phenolics in peat could help to improve modelling (e.g., Ise *et al*.[142]) of the decomposition of peat and its links with changing climate and will certainly contribute to our understanding of this very important carbon store at the molecular level.

1.5. Application of high-resolution techniques to the molecular-scale characterisation of HS

The overwhelming complexity of HS prevents separation of individual molecules by currently available chromatographic techniques. Thus to unravel the structures of individual molecules in mixtures of thousands of molecules, it has become widely accepted that two high-resolution analytical techniques, NMR and MS techniques are the most promising tools for this task.[143, 144] While MS techniques, such as FT-ICR MS, can resolve numerous non-isobaric molecules and provide their molecular formulae, NMR can, in principle, provide information about the classes of molecules, their functionality, molecular structure and provide insight into their aggregation state and dynamics. However, in practice, NMR struggles to overcome the limitations of low sensitivity and resolution. Regardless of their weaknesses, both techniques have contributed significantly to molecular-scale characterisation of HS, the topic reviewed next.

1.5.1. Molecular-scale characterisation of HS by FT-ICR MS and Orbitrap MS

Two high-resolution mass spectrometry techniques, FT-ICR MS and Orbitrap MS, have made important contributions to our understanding of HS; these are briefly summarised in this section.

One of the main outcomes of MS experiments is the determination of molecular formulae from m/z ratios. For complex mixtures, containing thousands of compounds, the mass resolution has to be very high in order to obtain a molecular formula for each individual compound. Currently, FT-ICR MS has the highest resolution available, with the resolving power currently exceeding 200 000 and with a mass accuracy of < 0.5 ppm.[145] More detailed information about the underlying theory and instrumental set-up is given in the reviews by Marshall *et al*.[146] and Amster.[147] The key point relevant to this study is that the high resolution combined with soft ionisation techniques that avoid fragmentation of covalent bonds, such as electrospray, allow practically unambiguous interpretation of m/z values in terms of molecular formulae (considering sensible combinations of CHO atoms) in the range of 200-1000 Da.[77] Typical FT-ICR MS spectra

of HS contain a dense envelope of peaks containing thousands of ions with MW up to ~1000 Da,[148-150] as shown for Suwannee River Fulvic Acid (SRFA) in Figure 1.6a.[151]

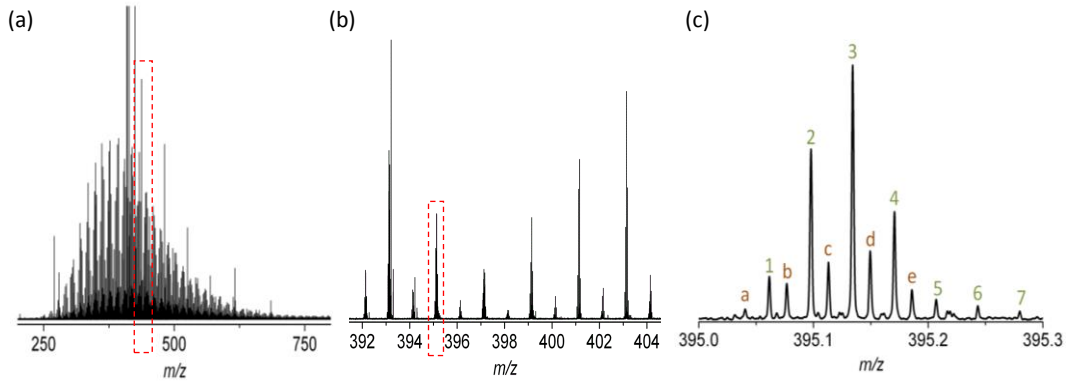


Figure 1.6 (a) FT-ICR MS spectrum of Suwannee River Fulvic Acid (SRFA) obtained in positive mode; (b) expansion of the region highlighted in (a) from 392 to 404 *m/z*; (c) further expansion of region highlighted in (b) illustrating two series of ions (1-7 and a-e) where each individual peak is separated from the next by 0.0364 Da, the mass difference between a CH_4 group and O. Image from ref.[151]

Analysis of HS MS spectra has revealed a number of distinct mass differences between individual packets of ions at each nominal mass, within these bunches and between prominent peaks (Figure 1.6b and c).[76, 148, 150, 152] Stenson *et al.*[77] examined these mass patterns, concluding that they represent homologous series of compounds differing in degree of saturation or a functional group change (e.g. CH_4 for O as shown in Figure 1.6c), similar patterns to that observed for petroleum compounds. They also observed that all ions in the MS spectra of HA and FA were singly charged and that these peak patterns were repeated throughout the spectra regardless of experiment conditions.

In order to inspect the large datasets obtained from HS MS spectra a number of visualisation methods have been tried. Visualisation of the mass patterns is best achieved by a Kendrick mass plot that shows Kendrick mass (Normalised IUPAC mass to the mass difference observed) vs Kendrick mass defect (KMD) (difference between IUPAC and KM).[153] The plots can be normalised to any particular combination of atoms, but the most common is a CH_2 group. Kendrick mass plots of HS samples with masses normalised to CH_2 have a diagonal shape, showing an increase in number of oxygens and double bond equivalents (DBE) with increasing mass. If plotted for all compounds present in the spectra these plots become dense and unintelligible, therefore the masses are typically divided into so-called z* series using the expression given in Figure 1.7.

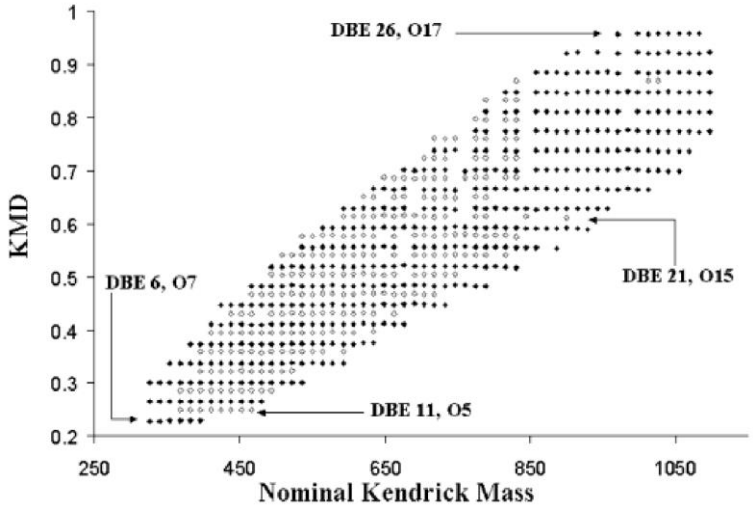


Figure 1.7 KMD vs Nominal Kendrick Mass for odd masses obtained from the FT-ICR-MS spectrum of SRFA, obtained in negative mode. The plot shows a subset of masses, z* series, (each represented by a dot) at every 14th *m/z* using the expression z* = modulus [NM/14]-14. Spacing between dots horizontally is 14 Da, which equates to CH_2 group, while vertical spacing is 0.0364 Da, which equates to the differences between a CH_4 and O. Reproduced from ref.[77]

These mass patterns are hypothesised to represent molecules related to one another by degree of degradation, for example demethylation (loss of CH_2) and aromatic ring opening (gain of O_2).[77] Enhanced understanding of these pathways has come from the comparison of the patterns obtained for a range of humic materials: FA or HA samples from different sources show the same general patterns, indicating similarities in the degradation mechanisms regardless of their source.[154, 155] KMD plots have also been produced by normalising the masses obtained with other combinations of atoms such as H_2, H_2O and CO_2.[156]

From the accurate masses obtained from HS FT-ICR MS spectra, molecular formulae can be assigned and this has been attempted manually[77, 153] and with the aid of computer programs by a number of researchers.[157] Whatever the assignment procedure used the results consisted entirely of $C_xH_yO_z$ compounds.

The most common way to inspect the obtained molecular formulae represent is by van Krevelen diagrams, plots of H/C vs O/C ratios derived from the molecular formulae. Originally designed for the study of the diagenetic evolution of coal or oil,[158] this type of plot has been widely accepted by the HS community as a way to categorise the molecular formulae and match them to particular compound classes. Regions of specific H/C vs O/C ratios are designated to different compound classes such as lignins, lipids or condensed aromatics. Based on these plots, the HS samples have been shown to contain many compounds in each class (Figure 1.8).

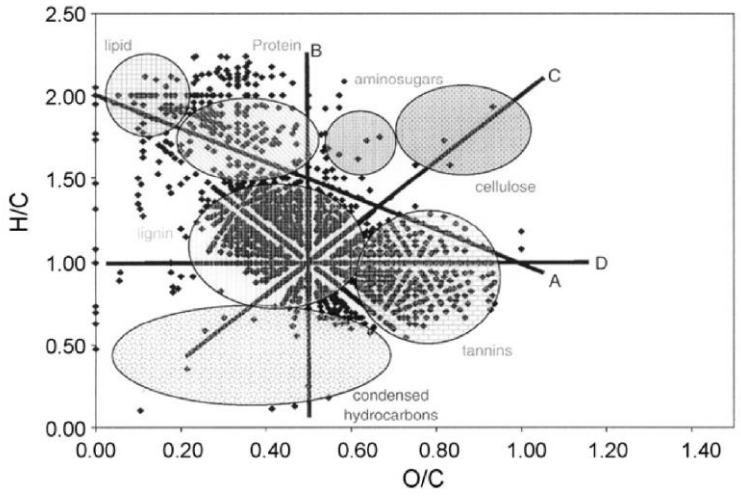


Figure 1.8 Van Krevelen diagram produced using the molecular formulae obtained from the FT-ICR MS spectrum of NOM isolated from swamp water. Superimposed on the plot are the designated compound class regions.[159]

As each peak on this diagram represents one chemical formula, van Krevelen plots have been used to compare different HS samples. These plots should, however be interpreted with caution as the presence of a peak in a particular region does not guarantee that the compound belongs to a particular compound class, as the same chemical formula can fit a number of different chemical structures.[160] This is exemplified by the fact that Hertkorn *et al.* observed that the major refractory component of DOM (especially marine DOM) is the Carboxylic Rich Acyclic Molecules (CRAM),[143] which appear in the same region of van Krevelen plots as 'lignin' molecules. Hatcher *et al*.[161] demonstrated this ambiguity further with a molecular formula $C_{26}H_{32}O_{11}$ that could be classed as a 'lignin' compound according to van Krevelen analysis but also a CRAM.

In an attempt to progress from molecular formulae to molecular structures, MS^n studies have been performed on HS samples. The most promising example by Witt *et al*.[162] used collision induced dissociation of several individual peaks isolated from a cluster with nominal mass of 365 Da. Observed mass losses, due to collision, matched the molecular masses of CO_2, H_2O, CH_4 and CO, indicating the prevalence of carboxyl and hydroxyl groups in these compounds. A number of parent structures were proposed, although the fragmentation was not conclusive enough to make definite assertions about their chemical structures.

There are a number of issues with FT-ICR MS, the most important to note is that obtained mass spectrum depends on the ionisation method used. ElectroSpray Ionisation (ESI) is the most commonly used ionisation technique for acidic compounds, but it is difficult to judge what proportion of the HS molecules are being ionised.[159] Other soft ionisation techniques such as

Atmospheric Pressure Chemical Ionisation (APCI) and Atmospheric Pressure Photoionisation (APPI) have also been applied to HS samples.[163, 164] Hertkorn *et al.* compared APCI, ESI and APPI and reported that out of the thousands of ions obtained, only 611 and 204 were common to all three ionisation techniques in negative and positive mode, respectively.[163] This could, for example, be caused by the preference of APPI/APCI to ionise smaller/less-polar molecules, while ESI preferentially ionises larger/more polar molecules. An alternative explanation is that the breakup of aggregates and thus ionisation of individual molecules was higher in one of these techniques. In addition, sample preparation, ESI conditions and instrumental parameters have to be carefully optimised as all these factors result in different ionisation efficiencies and different ion distributions.[148, 165] Further investigations are thus required to clarify how best to acquire and interpret MS spectra of HS.[166]

While the observed maximum *m*/*z* for FT-ICR MS spectra is limited only by instrumental factors,[76] the lowest *m*/*z* reported is ~250. This prevents detection of a potentially large number of very small molecules, a point proven by a recent MS study using an Orbitrap[167], the most recent addition to the high-resolution mass spectrometers as detailed by Markovo *et al*.[168] or Hu *et al*.[169] The Orbitrap has a maximum resolving power of 100,000 at *m*/*z* 200 and thus is capable of providing molecular formulae of HS for this mass range. The best example of the use of Orbitrap on HS was by Remucal *et al.* who compared the Orbitrap and FT-ICR MS spectra of SRFA. The main conclusion drawn was that ~430 molecules in SRFA were not detected by FT-ICR MS and these were all below 290 Da.[167] The authors also listed some molecules that match the molecular formula obtained, including aliphatic acids, phenols and aromatic acids.

FT-ICR MS and to a smaller extent Orbitrap MS are now routine tools to analyse the composition of NOM from various sources such as rivers,[77, 167] freshwater lakes,[170] peat pore water[171] and soil.[172] Researchers have compared marine with terrestrial (specifically mangrove pore water) DOM[154] and riverine with soil HA.[150] Most of the aforementioned studies presented their results in the form of van Krevelen diagrams. One of the most common observations from these diagrams is that a large proportion of the assigned molecules appear in the 'lignin' region, implying their terrestrial origin. For marine DOM samples this was a significant result as it indicated that terrestrial material can be traced using FT-ICR MS. However, as mentioned above caution should be taken when interpreting van Krevelen plots. Other techniques such as NMR need to be used alongside the high-resolution MS data to justify any conclusions regarding compound type.

1.5.2. Molecular-scale characterisation of HS by NMR spectroscopy

The application of NMR spectroscopy in the studies of HS has been extensive but the advancements have been limited by the capabilities of the spectrometers of the day, a limited selection of suitable experiments, and inherently by the overwhelming complexity of HS NMR spectra. In addition, a wide range of HS samples, e.g. from whole soils, NOM and DOM to HA and FA, from many different sources have been studied by NMR, making comparisons of results challenging. The aim of the following sections is to provide an overview of the research undertaken in this area. The focus will largely be on solution state NMR studies using ^{13}C and ^{1}H nuclei as this overlaps with the techniques used in this investigation. For information regarding $^{31}P/^{15}N$ NMR and solid–state methods applied to HS, the reader is referred to a review.[173] Solid-state and HR-MAS NMR are promising techniques for the study of the insoluble fractions of HS such as humin[11, 117] however at this stage neither technique has the resolving power to reveal individual molecular structures, which are at the heart of this project.

1.5.2.1. One-dimensional (1D) NMR studies of HS

The earliest NMR study of HS was that of a fractionated methylated soil FA.[174] The obtained ^{1}H spectrum contained mostly carbonyl and aliphatic signals with little aromatic or olefinic resonances. This came as a surprise as HS, especially from terrestrial sources, were believed to have a high lignin-like content. Variable aromatic content was shown in subsequently published spectra, indicating that such compounds were either removed during the extraction procedures or that the aromatics compounds contained within HS were heavily substituted, which would explain the absence of aromatic proton resonances in the earlier spectra.[175-179] Nevertheless, common to all solution state ^{1}H spectra was the lack of fine structure with overlapping signals and broad lines[180, 181] as exemplified by Figure 1.9a.[178]

^{1}H-decoupled ^{13}C spectra are usually less complex than ^{1}H spectra, with a greater spread of resonances. Thus, on realisation that ^{1}H spectra provide limited information, the focus moved to ^{13}C NMR spectroscopy. The first ^{13}C spectra obtained without an additional organic extraction step were of two soil HA and one FA from andisol, vertisol and podsol soils.[182] Many more examples from both aquatic and terrestrial sources flooded the literature[176-178, 182-185] demonstrating the increased interest in NMR as a technique for HS research. All spectra

contained carboxylic, aromatic and aliphatic carbons in relatively similar amounts as shown in Figure 1.9b.

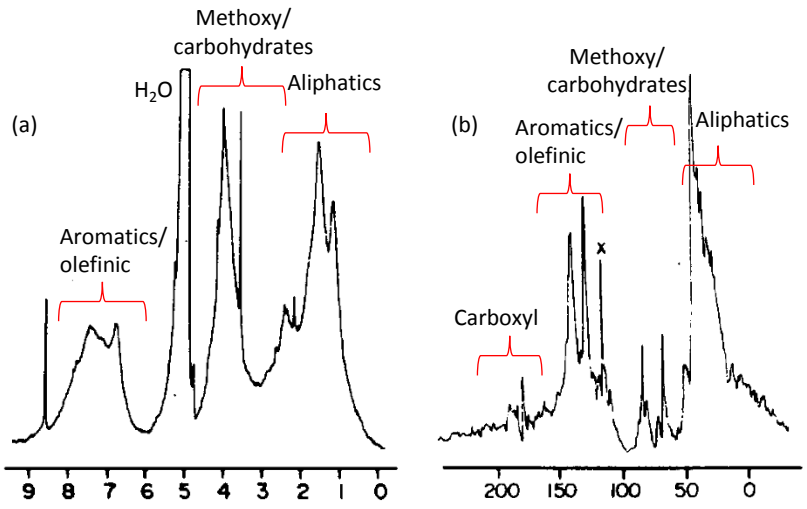


Figure 1.9 100 MHz (a) ^{1}H and (b) ^{13}C NMR spectra of peat HA in 99% D_2O/1% NaOD. Compound class regions are highlighted on the spectra. Adapted from ref.[178]

The ^{13}C spectra acquired like ^{1}H spectra, however, continued to show poor resolution with weak unresolved signals, showing that the inherent complexity is the major issue. This still is the case today even at the highest magnetic field of 800 MHz and above. Therefore all 1D spectra allow only the characterisation of the broad bands into four main regions representing aliphatic, carbohydrate, aromatic and carboxyl functional moieties.[181]

Thus, only rough comparisons of samples were possible based on ^{1}H and ^{13}C NMR spectra. For example, Stuermer and Payne[177] compared terrestrial and marine FA and HA samples, showing that terrestrial samples contained larger proportions of aromatic signals than the marine samples, which were interpreted as proof that lignin compounds are part of terrestrial HS samples. The aliphatic resonances, albeit more intense in the marine samples, showed similar chemical shifts to the terrestrial samples. Wilson *et al.* tried to explain this observation and suggested that while aquatic aliphatics are derived from algal or microbial lipids, similar compounds could be produced by microrganisms in the soil.[179]

Beyond a general description and comparison, it is clear that the complexity of HS prevents full separation of individual signals thus limiting the use of 1D NMR spectroscopy for the analysis of HS. Only few authors attempted to interpret the ^{1}H and ^{13}C spectra in more detail and make tentative assignments of some of the more resolved signals.[177, 178, 184, 186] The first comprehensive

study of both ^{1}H and ^{13}C of HA and FA was that of marine sediments (compared to a peat HA) by Hatcher *et al*.[178] This study, combined with the results of two other publications[179, 187] made the following observations:

- Unsubstituted aliphatic carbons and their associated protons resonate at 0-50 ppm and 0.6-1.5 ppm, respectively. Methyl protons and carbons, resonating at 0.9 ppm and 15 ppm, respectively, are in high abundance. Before the structural view of HS changed, these resonances were interpreted to indicate highly branched polymeric structures. Another strong resonance that appears at ~ 30 ppm was assigned to tertiary butyl units. Methylene protons and carbons in alkyl environments resonate at 1.0-1.4 ppm and 18-40 ppm, respectively.
- Protons and carbons α to carboxyl functional groups resonate at 1.8-2.2 ppm, while the carboxyl carbons appear at 175 ppm. Protons of methyl and methylene groups α to aromatic rings resonate at 2-3.3 ppm. Protons resonating at ~2.6 ppm are indicative of benzylic carbon attached protons.
- Protons resonating at around 3.7 ppm are indicative of O or N substituted compounds, mostly likely carbohydrates.
- Protons β and carbons α to oxygens resonate at 3-4.2 ppm and 60-110 ppm and are characteristic of carbohydrates. In addition, further evidence of the presence of carbohydrates is a peak at 65 ppm, characteristic for C6 of hexoses, and signals around 75 ppm, which are characteristic for C2-C5 of hexose ring carbons. In addition, signals at 100-105 ppm provide evidence of anomeric carbons. Aquatic HS samples were found to have greater intensity in this region than HS samples from terrestrial sources.
- Aromatic or conjugated olefinic carbon and protons resonate between are 105-170 and 6-8.5 ppm, respectively. Three broad peaks are typically observed at 7.2, 6.9 and 6.5 ppm that are typical of aromatic structures substituted with a weak electron donating group, a strong electron donating group and those with multiple substituents, respectively.

The above classification of signals is common to most published spectra. However, examination of spectra obtained from a number of studies identified additional resonances at:

- 115-117 ppm and 130-132 ppm, assigned to lignin-like phenols.[188]
- 156 ppm, assigned to phenolic or N-substituted aromatic carbons.[188]

- 170-180 ppm, assigned to carboxyl carbons. More specifically resonances at 167 ppm have been assigned to dihydroxybenzoic acids, and 172 ppm to either crotonic or dihydroxybenzoic acids.[189]
- 195 ppm (FA) and 192 ppm (HA), assigned to ketonic carbons.[2, 189, 190]
- 105 ppm, which appears as a shoulder on the signals of aromatics resonances, has been assigned to anomeric carbons.[191]

Another potential reason for the poor NMR spectra obtained in several studies is the solvent that was used. Hatcher *et al*.[178] noted that much of the earlier work[185, 192] used organic extraction which was only able to solubilise part of HS samples. The use of D_2O, with or without NaOD, is regarded to provide conditions closest to the natural environment of soil. Hence, this solvent was adopted for measuring NMR spectra of HA and FA.[178, 182, 184, 193] However, one should not disregard using more than one solvent for HS studies as this can highlight important differences in molecular composition. A favourite of many researchers is DMSO-d_6, which compared to D_2O provides additional key features in HS spectra (Figure 1.10). For example, due to chemical exchange with D_2O, amide signals will only appear in DMSO-d_6 spectra.[144]

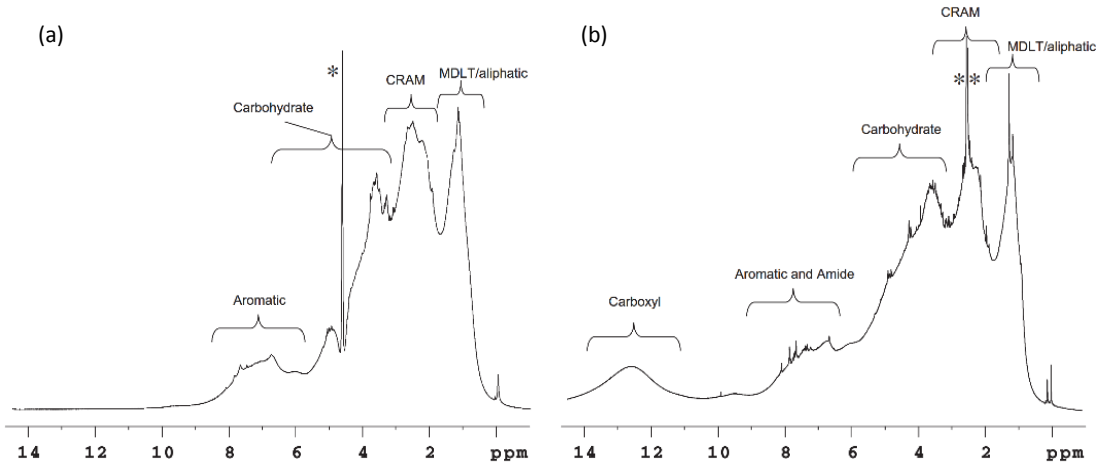


Figure 1.10 ^1H spectra of SRDOM in (a) D_2O and (b) DMSO-d_6. Residual HOD and DMSO-d_6 signals are highlighted by asterisks. Compound class regions are highlighted on both spectra. CRAM: carboxylic rich alicyclic molecules and MDLT: material derived from linear terpenoids. Reproduced from ref.[144]

The major issue for any solvent is the interference of the solvent signal. In the case of D_2O this signal arises, in ^1H spectra, from residual H_2O which can mask a large area of the spectrum and can also lead to artefacts in 2D spectra. Signals from solvents can be removed, with varying degree of success, by some form of signal suppression, as shown by the asterisk in Figure 1.10

where the water peak was suppressed to some extent by presaturation. Solvent suppression will be discussed in greater depth in Section 1.5.2.2.

When using D_2O another variable to consider is the solution pH, which can also affect the appearance of NMR spectra as shown in Figure 1.11.[194]

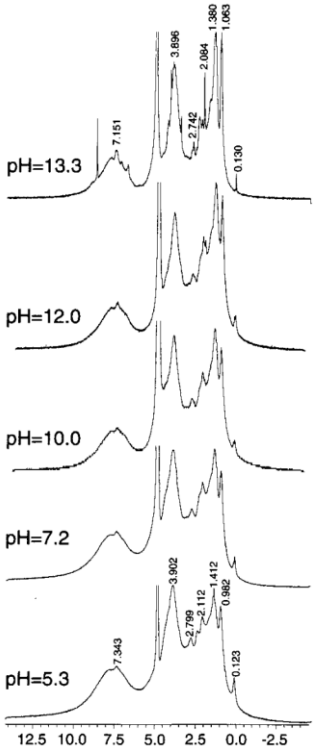


Figure 1.11 400 MHz ^{1}H spectra of soil HA at different pH values in 99% D_2O/1% NaOD.[194]

Figure 1.11 illustrates that changing the pH from 13.3 to 5.3 leads to broadening of resonances and loss of sharp signals. These changes could be explained by the extent of aggregation of the sample, related to changes in the protonation state of molecules at different pH values.

Despite the attempts of some authors, the resonance assignment of 1D spectra should be treated as tentative and should not, without further evidence, be fully trusted. The resonances do fall in to specific recognised regions, but their broad features hinder the separation of specific functional group types.[181] For example, aromatic signals can overlap with carbohydrate signals, as heteronuclei can shift signals into the anomeric region. Even relatively isolated signals could still be a superposition of many signals from similar moieties. In some instances, more resolved peaks can appear rising from a blanket of unresolved signals. One set of commonly reported peaks in ^{1}H spectra of NOM are those at 1.93, 2.43, 3.36 and 8.4 ppm (in NaOD), which were assigned to

acetate, succinate, methanol and formate by adding small amounts of these compounds to the HS samples.[195] However, this method of assignment is not applicable for the broad range of compounds in HS as their identities are masked by the overlapping humps of resonances.

Many researchers believe that increased magnetic field strength will resolve the signals of more compounds. Indeed increased resolution with the advent of stronger magnetic field instruments such as 600, 800 MHz or 1 GHz has allowed higher sensitivity, smaller sample sizes or shorter experimental times to be used.[196] However, some correctly argue[197] that the increasing magnetic field strength will never combat the inherent molecular complexity of HS samples, as increased broadening due to chemical shift anisotropy of loosely associated molecules will remain a problem.

Unlike MS spectra, ^{1}H NMR spectra can easily be acquired in a quantitative manner. Unfortunately, the quantification is hindered by the overlap inherent to ^{1}H spectrum of HS. As mentioned before, ^{13}C spectra are subject to less overlap and so a number of studies have attempted to quantify ^{13}C spectra of HS. However, obtaining quantitative ^{13}C spectra is not a simple task, as one must allow sufficient time for the relaxation of carbon spins (T_1) but also to suppress nuclear Overhauser enhancements, due to the non-universal increase in signal intensity upon ^{1}H decoupling.[198] Both these effects are taken into account by choosing an appropriate repetition time, pulse angle and using inverse-gated decoupling. Preston *et al.*[199] attempted to define these parameters to obtain relatively quantitative ^{13}C spectra of soil FA and HA samples using the shortest possible acquisition times. They measured T_1 values from 0.2 to 2 s from the HS samples, with carboxyl and methoxy groups accounting for the observed maximum. Thorn *et al.* concurred with these parameters to obtain relatively quantitative inverse-gated ^{1}H decoupled ^{13}C spectra of various IHSS standards.[189, 197]

^{13}C spectroscopy offers the possibility to extract information about the number of protons attached to each carbon through the use of a variety of spectral editing techniques. A number of different editing experiments have been applied to HS samples including Spin-Echo Fourier Transform (SEFT),[199] Attached Proton Test (APT),[200] Distortionless Enhancement by Polarisation Transfer (DEPT),[201-204] Sub-spectral Editing with a Multiple Quantum Trap (SEMUT)-90, QUATernary-only (QUAT)[202] or proton spin-echo DEPT (DEPT-Q).[205] A few examples are given below:

1. APT spectra of SRFA and SRHA[189] identified carbohydrates and secondary ethers in the region between 62-92 ppm. Anomerics and methoxy groups were observed at 92 and 56 ppm, respectively. It was also noted that the carboxyl region is split into two areas, 165-185 ppm (carboxylic acids) and 185-220 ppm (primarily ketones). The ketone region was also populated suggesting evidence of diaryl, alkyl diaryl and dialkyl ketones. APT was also carried out by Cook *et al.*[200] on a FA sample from a soil B horizon. They concluded that the carbohydrate and aliphatic carbons within the sample were highly functionalised, while the majority of aromatics were protonated.
2. Buddrus *et al.*[205] performed four experiments: two types of ^{1}H INADEQUATE, DEPT, and a standard spin-echo to determine the proportion of isolated methyl groups in a sample of Ground Water (GW) HS. They found that the most successful experiment for this task was the standard spin-echo that showed that 60% of GW HS contained methyl groups. From further analysis of the chemical shifts, they postulated that polycyclic isoprenoid structures were the likely source of the majority of these methyl groups. Both DEPT and ^{1}H-INADEQUATE overestimated the abundance of methyl groups due to interference from spins coupled to the methyl groups.
3. DEPT and QUAT experiments were also carried out on aquatic HS samples (Figure 1.12b-e) in addition to a standard carbon spectrum (Figure 1.12a).[202] The results showed a high proportion of quaternary carbons (78%) particularly in the aromatic region 100-160 ppm. Specifically, signals at 140 ppm and 150 ppm were assigned to aromatic carbons substituted with alkyl and hydroxyl groups, respectively. Using combined analysis of all spectra, they quantified the ratio of aliphatic to aromatic protons in their samples as 8.2: 1.8.
4. The J-SEFT[199] experiment was applied to soil FA and HA. The spectrum obtained was similar to DEPT. Additional assignments of the negative CH_3 peaks at 54.8 and 59 ppm was made to aromatic $COOCH_3$ and OCH_3 groups in HA spectra, respectively, while in FA peaks at 57.7 and 51 ppm were assigned to OCH_3 groups.

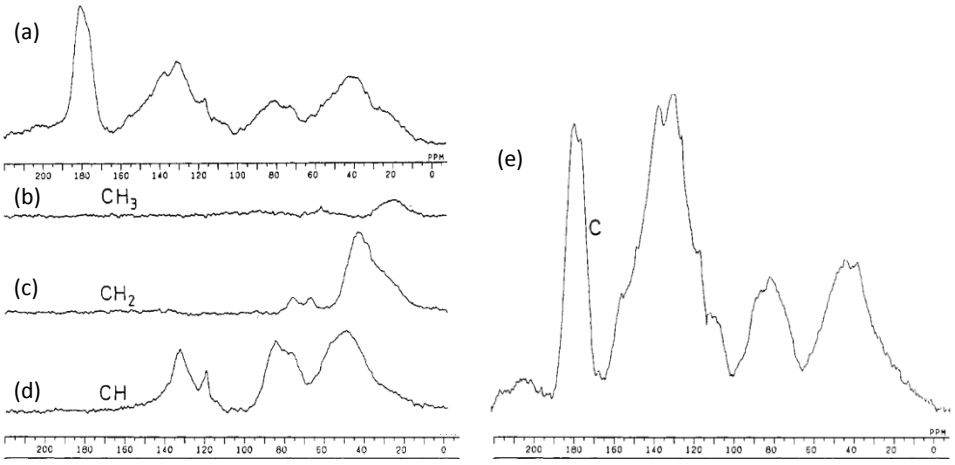


Figure 1.12 (a) ^{13}C spectrum; (b)-(d) DEPT sub-spectra; (e) QUAT spectrum showing only quaternary carbons of an aquatic HS sample.[202]

1.5.2.2. Solvent suppression techniques applied to HS

As mentioned in Section 1.5.2.1, one of the limitations of ^{1}H NMR spectroscopy of HS is the presence of broad solvent signals particularly from residual H_2O in D_2O samples. To alleviate this problem a number of water suppression techniques have been developed and applied to HS. For example, Wilson *et al.*[206] showed that irradiation of the water peak during the relaxation delay could be used to enhance the quality of the spectra. However, this technique also affects the signals surrounding the water peak. Improvements to water suppression included phase modulated irradiation,[207] composite pulses[208] and low flip angles.[209, 210] However, it was not until the introduction of pulsed field gradients (PFGs), see Section 1.7.6, in the form of WATERGATE that the solvent suppression was significantly improved. This sequence uses a combination of PFGs and tailored excitation and was first applied to HS samples by Lee *et al.*[211] to produce solvent free ^{1}H spectra for low concentration samples. Liu *et al.*[212] added a W5-DANTE element to WATERGATE creating the W5-WATERGATE that inverts all signals except water, which is dephased by the gradients in the sequence. Lam *et al.*[210] later added a train of water-selective shaped pre-saturation (SPR) irradiation pulses to the W5-WATERGATE sequence train to further improve water suppression (Figure 1.13). The authors reported that when compared to W5-WATERGATE a wider region around water (~1.1 ppm) is attenuated which is not particularly good for carbohydrate-rich samples that contain resonances close to water.

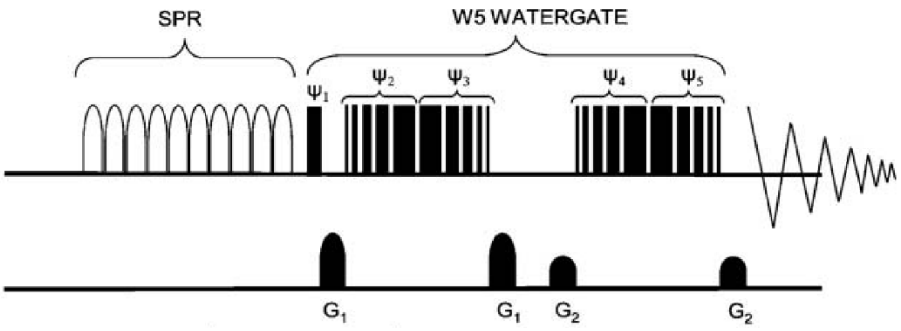


Figure 1.13 SPR-W5 WATERGATE solvent suppression pulse sequence. Selective pulses are depicted by an open 'shape', whereas hard pulses are indicated by solid blocks. Selective pulses were applied for 2 ms. 2000 pulses were applied each along the *x* direction, with a 4 ms delay separating each pulse. *G* represents gradient pulses while ψ the phases of the hard pulses used in W5-WATERGATE.[210]

Another method, that arguably produces even better results than WATERGATE, is the double-pulsed field gradient spin-echo (DPFGSE), which was successfully used on HS samples by Cook *et al.*[200] Simpson *et al.* argue that PURGE[213], which was developed for natural samples provides similar results to DPGSE but, due to the lower irradiation power used, leads to less attenuation of resonances around the water signal.[144] The best example of water suppression currently available is Perfect Spin Echo WATERGATE which limits attenuations of signals close to water (within 0.5 ppm) by preventing *J* modulation during the spin-echoes used to refocus chemical shift.[214] This provides superior performance to the above methods, though to date this method has not been applied to HS samples.

1.5.2.3. Two-dimensional (2D) NMR studies of HS

Analogous to the 2D NMR spectroscopy of organic compounds or biomolecules, the addition of a second dimension, especially in heteronuclear experiments, increases the signal dispersion, uncovering information which is lost in 1D NMR spectra. The first 2D NMR spectrum was acquired on an HS sample using a heteronuclear *J*-resolved 2D experiment which gave the ^{1}H-^{13}C coupling constants vs ^{13}C chemical shifts along the F_1 and F_2 axis, respectively.[202] The authors used this experiment to reinforce the observations obtained by spectral editing techniques, and highlighted the benefits and applicability of 2D NMR to HS. At the time of Buddrus *et al.'s* experiments, 2D NMR was still a lengthy and not very sensitive endeavour. However today, with the advancements in spectrometer hardware, increased magnetic field strengths and introduction of cryoprobes, a wide range of 2D experiments can be used, each of which provide specific information. Table 1.4 summarises the 2D NMR experiments that have been applied to HS.

Table 1.4 Summary of 2D NMR experiments performed on HS samples.

NMR experiment	Acronym	Information	Interpretation
COrrelation SpectroscopY	COSY	Connectivity of protons on adjacent carbons.	Cross-peaks connect the chemical shifts of protons that are coupled. Cross-peaks appear around a central diagonal.
TOtal Correlation SpectroscopY	TOCSY	Connectivity of protons over multiple bonds.	Similar to COSY but multiple correlations are established by following horizontal or vertical lines drawn through the diagonal cross-peaks.
Heteronuclear Single Quantum Coherence/ Heteronuclear Multiple Quantum Coherence	HSQC/HMQC	^{1}H-^{13}C one bond correlations.	Cross-peaks represent carbon chemical shifts in one dimension and proton chemical shifts in the other dimension.
Heteronuclear Single Quantum Coherence/ Heteronuclear Multiple Quantum Coherence-Total Correlation SpectroscopY	HSQC/HMQC-TOCSY	^{1}H-^{13}C one bond and ^{1}H-^{1}H correlations.	Cross-peaks represent carbon chemical shift in one dimension and protons of individual spin systems in the other dimension.
Heteronuclear Multiple Bond Correlation	HMBC	^{1}H-^{13}C correlations over 2-4 bonds. Quaternary carbons also observed.	Cross-peaks represent carbon chemical shifts in one dimension and proton chemical shifts in the other dimension.
Nuclear Overhauser Effect SpectroscopY	NOESY	Interactions through space/chemical exchange.	Cross-peaks represent two protons that are close in space/chemical exchange.
Rotating-frame nuclear Overhauser Effect SpectroscopY	ROESY	Interactions through space/chemical exchange.	Cross-peaks represent two protons that are close in space/chemical exchange.
Diffusion Ordered SpectroscopY	DOSY	Separates species by diffusion coefficient.	Cross-peaks represent chemical shift in one dimension against diffusion coefficient in the other.

When applied to complex mixtures, the major expectation of 2D NMR is to reveal the peaks hidden in overlapped regions of 1D spectra.[22] This section will give a general overview of the application of 2D experiments to HS samples. In-depth information regarding the physics and mathematical treatment of each experiment will not be discussed here. This information can be found in more specialised literature.[215, 216]

Traditionally, 2D NMR experiments are separated into homonuclear experiments, investigating correlations between nuclei of the same kind and heteronuclear experiments, which correlate different nuclei. Homonuclear experiments performed on HS samples will be discussed first.

1.5.2.3.1. 2D homonuclear correlation experiments

There are two types of homonuclear experiments using either scalar or dipolar couplings. 2D COrrelation Spectroscopy (COSY) detects protons that are mutually coupled typically via 2-3 bonds (vicinal or geminal). There are a number of variations of the COSY experiment each having their own advantages and disadvantages depending on the studied system. The standard COSY-90 has the best sensitivity but for complex samples can provide unresolved spectra around the diagonal. COSY-45 simplifies the structure of the cross-peaks, narrows the diagonal and allows the two-bond and three-bond couplings to be distinguished. These benefits, however, come with a reduction in S/N and hence COSY-45 requires longer overall experimental time. Phase-sensitive COSY experiments, in principle, allow couplings to be measured. The resulting spectra have characteristic dispersive diagonal and anti-phase absorptive cross-peaks. However, as the dispersive tails extend far from the diagonal and can mask resonances, the Double Quantum-Filtered (DQF) COSY is more suitable. This experiment reduces the diagonal by strongly attenuating signals from non-coupled spins but importantly also produces absorptive diagonal peaks.[144, 217] However, the increased simplicity of the spectra again is associated with a reduction in S/N.

2D COSY spectra of HS have been acquired on a number of occasions. For example, Schmitt-Kopplin *et al.*[218] illustrated the use of gradient-selected COSY experiments to follow the effect of photo-irradiation on an HA sample under O_2 and N_2. Examination of the spectra obtained before and after photo-irradiation, indicated that the most photo-labile components are the O/N substituted aromatics, while functionalised aliphatic and carboxyl content increased under O_2. To simplify the 2D COSY spectra Haiber *et al.*[219] employed tangential flow multistage

ultrafiltration to separate fractions of an aquatic HS sample. A COSY spectrum was acquired for the < 1 kDa fraction that supported their HMQC experiments. While both these studies obtained some useful information, their COSY spectra were crowded or in the case of Haiber *et al.* contained only a few cross-peaks. One of the cleanest and detailed COSY spectra that can be found in the literature compares peat HA and FA(Figure 1.14).[188]

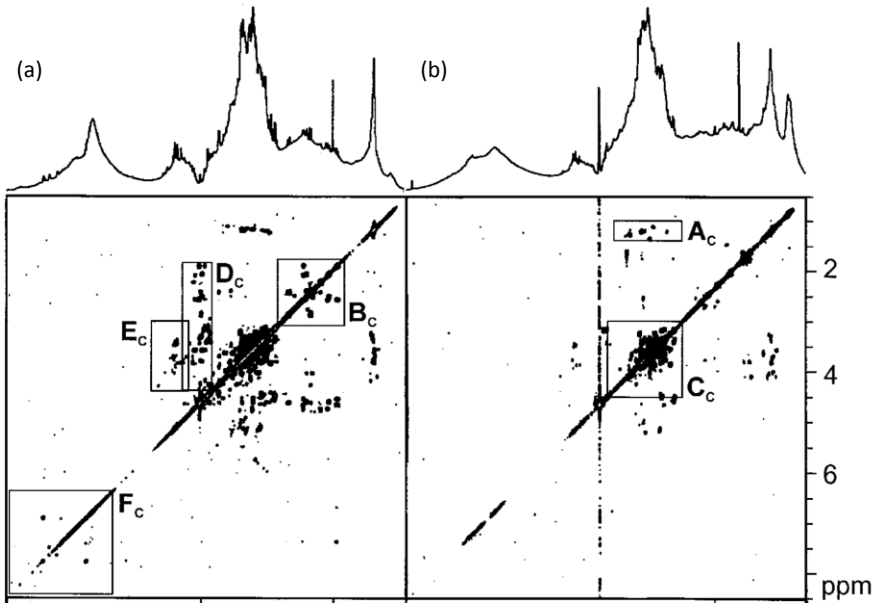


Figure 1.14 2D COSY spectra of peat (a) FA and (b) HA. Labelled regions donate: (A_C) deoxy sugars, ethers and esters; (B_C) functionalised aliphatics; (C_C) carbohydrates excluding anomerics; (D_C) functionalised aliphatics with one heteroatom; (E_C) carbohydrate with anomerics; (F_C) aromatics.[188]

These spectra present clear differences between HA and FA. The FA sample contains more correlations in all regions shown compared to the HA spectrum. Simpson *et al.*[196] used an array of 2D experiments to characterise a soil FA sample. A few tentative assignments obtained by COSY and backed by additional experiments revealed mono- and dicarboxylic acids, amino acids, carbohydrates and also 1, 2-, 1-4, 1, 3, 4- substituted benzenes. More recently, Woods *et al.*[220] measured a COSY spectrum from a DOM sample fractionated by 2D High-Performance Liquid Chromatography (HPLC). Using Bruker AMIX prediction software, small acids and diols were identified. The small acids found agree with the MS study by Remucal *et al.*[167] mentioned in Section 1.5.1 and the NMR study by Wilson *et al.*[195] in Section 1.5.2.1.

Phase sensitive COSY spectra of complex mixtures can suffer from self-cancellation of cross-peaks due to the overlap of anti-phase lines, therefore these spectra may not show all possible correlations. It is therefore argued that TOtal Correlation SpectroscopY (TOCSY) should be used in combination with COSY or as a stand-alone experiment for establishing homonuclear

correlations in HS samples. TOCSY uses multiple-step coherence transfer (instead of a single coherence transfer used in COSY) to map out the protons of each spin system. This is achieved by replacing the second 90° pulse in the standard COSY pulse sequence with an isotropic mixing sequence, a spin-lock. The duration of spin-lock, referred to as the mixing time, determines how far along a spin system the magnetisation travels. For example, short mixing times, 20-30 ms, give an almost COSY-like spectrum. Simpson *et al*.[144] observed that for NOM samples a 60 ms mixing time gives a great deal of long-range correlations even for fast relaxing systems. Cook *et al*.[200] used a similar mixing time of 50 ms to acquire a TOCSY spectrum of a FA sample and witnessed up to 7 cross-peaks for a given spin system. The nature of the spin-lock mixing sequence also affects the resulting spectra. Simpson *et al*.[144] tested three such schemes: MLEV, DIPSI and adiabatic mixing on a SRDOM sample. The results clearly indicated that for the DOM sample studied, adiabatic mixing sequences performed best as they are less susceptible to RF inhomogeneity.[221]

There are numerous examples of the use of TOCSY in the studies of HS and some provide high quality spectra, mostly applied in combination with COSY.[191, 196, 200, 217, 220, 222-224] In these studies, a number of correlations were identified as shown in the TOCSY spectrum of HA from an oak forest soil (Figure 1.15).[222]

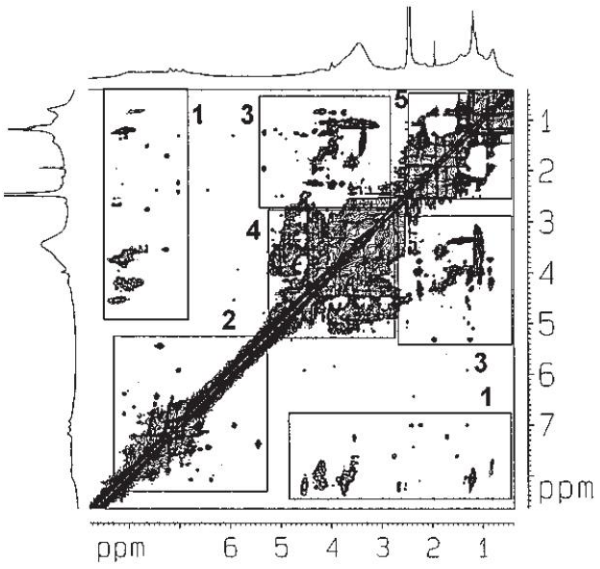


Figure 1.15 A 75 ms mixing time 2D TOCSY spectrum of oak forest soil HA. The labelled regions are assigned as follows: 1: amide side chains; 2: aromatics; 3: sugars, methane units bridging lignin aromatics or Hα-Hβ of amino acids; 4: methylene units adjacent to ethers, esters and hydroxyls in aliphatic chains, Hα-Hβ-Hγ of amino acid; 5: methylene in aliphatic chains, methyl units in amino acids and aliphatic chains.[222]

Spectra obtained from another soil in the same study showed that HA of different origins are very similar in the types of compounds presented and only differed slightly in their relative abundance in agreement with the elemental compositions seen in Table 1.2. Another study of soil HA and FA under comparable conditions[225] reported very similar spectra for HA, but the soil FA spectrum lacked many cross-peaks in sections labelled 1, 2, and 4 in Figure 1.15, indicating that in this case, HA and FA have slightly different compositions. This was also reported by Cook *et al.*[200] Hertkorn *et al.*[188], who carried out a detailed study of FA and HA from soil, noted a number of changes to the populations of cross-peaks in the spectral regions defined above. For example, region 4 in their spectrum was spilt into two, 4a: 4.4-3/1.4-1 ppm assigned to deoxy sugars ethers, esters, and 4b: 4.5-3.2/3-1.4 ppm assigned to functionalised aliphatics with one heteroatom. The authors also observed a reduction in cross-peaks for HA in section 2 and very little cross-peaks in region 4b of FA. It should also be noted that the same spectra contained no cross-peaks in section 1 for either HA or FA. This, however, does not rule out the presence of amino acids or peptides as different solvents were used. Hertkorn *et al.*[188] used D_2O while both Simpson *et al.*[222] and Cook *et al.*[200] used DMSO-d_6 thus the cross-peaks from peptide or amino acids would not be expected to be observed in D_2O spectra. Cook *et al.*[200] also acquired a ^{1}H-^{15}N HSQC spectrum, which clearly showed that amide moieties were present in the sample.

Selective TOCSY experiments have also been carried out on a SRDOM sample that was initially fractionated by 2D HPLC. By selectively exciting methyl proton resonances (1.06 ppm), Woods *et al.*[220] saw no TOCSY transfer, indicating that they are surrounded by quaternary carbons, while the excitation of the designated OR region (3.77 ppm) showed a wealth of resonances to the protons between 1.5-2.5 ppm. These results were interpreted as evidence of cyclic structures, however despite the connection between OR region and aliphatics region, this does not prove that these molecules are cyclic as the same connection could be made from acyclic or even lipid-like moieties.

In conclusion, TOCSY is a very powerful experiment and one can obtain acceptable resolution data from HS samples using this technique. However, cross-peak overlap is still a severe obstacle. Interpretation of TOCSY spectra can be more difficult than COSY spectra as it is not obvious if the observed cross-peaks are due to short or long-range correlations. Cross-peaks due to dipolar interactions can also occur in TOCSY spectra causing additional confusion.[144, 226]

Nuclear Overhauser Effect SpectroscopY, NOESY, is the most common experiment for mapping the through-space proximity of proton nuclei. It utilises dipolar interactions between spins that are less than 5 Å apart. The NOESY experiment is very important as it gives information regarding the 3D structure of molecules in solution and can provide tertiary structure information, which is not obtainable using through-bond experiments. For example, protons separated by heteronuclei like oxygen do not give through-bond cross-peaks in 2D COSY or 2D TOCSY spectra. NOESY spectra also show cross-peaks due to chemical exchange, which most likely are limited in HS to exchangeable protons in aprotic solvents.

A number of groups have applied 2D NOESY experiments to HS,[22, 194, 196, 217, 222, 223] however the majority of obtained spectra are quite poor, especially for HA samples. Chien and Bleam,[194] for example, investigated the changes in NOESY spectra obtained at different pH values (5.3 to 13.3), using a mixing time of 0.3 s and 0.05 M NaOD as the solvent. Their spectra contained negative NOE enhancements that, according to the authors, provides evidence for macromolecular structures, although an alternative explanation could involve aggregation of small molecules. The authors assigned the cross-peaks labelled A to E in their spectra obtained at pH 13.3 (Figure 1.16a) and 5.4 (Figure 1.16b). Cross-peak A situated at 5.2/4.1 ppm (Figure 1.16a) was assigned to interactions between oxygen-substituted structures and olefinic protons. With decreasing pH, broad and more intense cross-peaks appear, especially in the aliphatic and oxygen-substituted aliphatic regions (Figure 1.16b).

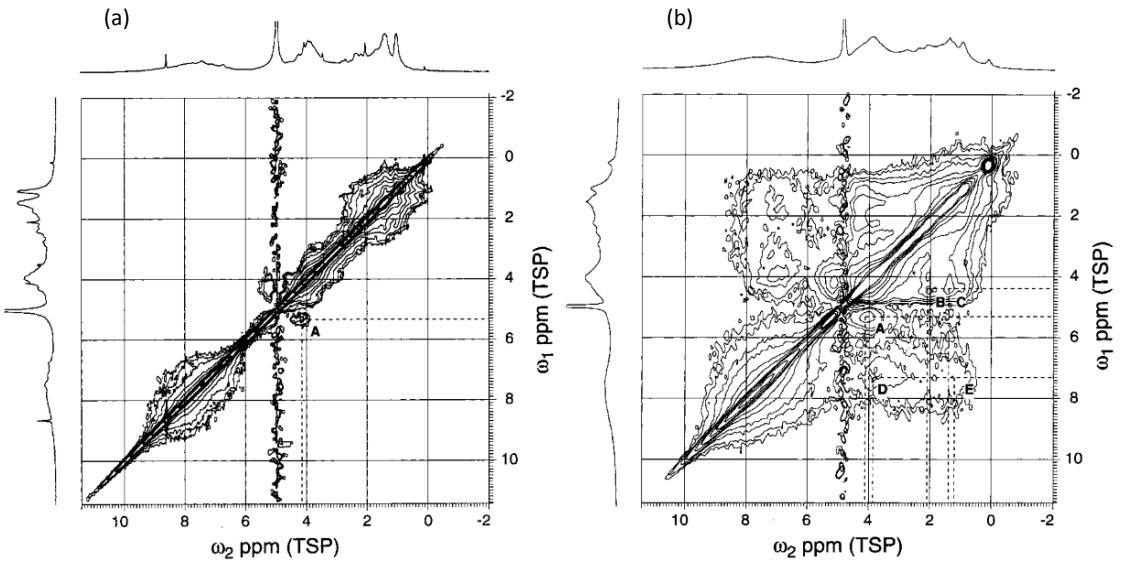


Figure 1.16 2D NOESY spectra of HA at (a) pH 13.3 and (b) pH 5.4.[194] Labelled regions denote correlations of (A) oxygen substituted structures with olefinic; (B) ethers with carbonyl methyls; (C) ether with methylenes; (D) aromatics with oxygen substituted aliphatics (phenol); (E) aromatics with methylenes.

The authors speculated that as pH decreases, the HS macromolecules compact bringing their flexible aliphatic chains closer in space. However, noting the results from the DOSY studies (see Section 1.5.2.3.2) and with the knowledge that their sample was HA at 10% w/v it is possible they are seeing the effects of aggregation, which increases with decreasing pH and not compacting of a macromolecular structure. It is also likely that the observed increased intensity is due to the protonation state of polysaccharides. Above pH 7 these molecules carry a negative charge and form more elongated molecules, while at lower pH they become neutral and start to compact, eventually precipitating as gels.

Fan *et al.*[223] acquired a 2D NOESY spectrum (Figure 1.17a) of a forest soil HA sample to back up their arguments based on the analysis of 2D TOCSY, as well as InfraRed (IR) spectra. They used a 200 ms mixing time, D_2O solvent and 2.8% w/v sample concentration at neutral pH.

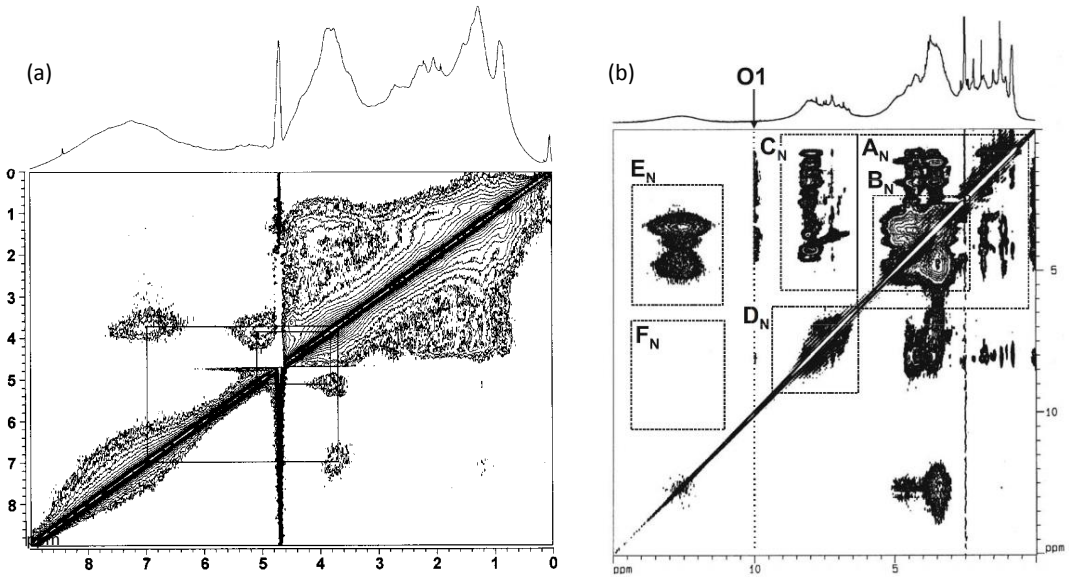


Figure 1.17 2D NOESY spectra of (a) forest soil HA in D_2O[223] and (b) soil HA in DMSO-d_6.[22] Labelled regions denote correlations between (A_N) aliphatic to aliphatic or aliphatic to hydroxyl; (B_N) aliphatic hydroxyl to aliphatic hydroxyl; (C_N) amide NH to aliphatic; (D_N) aromatic to aromatic; (E_N) carboxylic acid to aliphatic hydroxyl (chemical exchange); (F_N) carboxylic acid to aromatic hydroxyl (chemical exchange) resonances.

The spectrum collected appears very similar to the pH 7.3 2D NOESY spectrum collected by Chien and Bleam,[194] although the assignments differ. The same cross-peak labelled A in Figure 1.16 is not reported as an olefinic correlation but instead, and likely more correct, is assigned to the interaction between anomeric protons and the other ring protons of carbohydrates. The correlations from 7-7.3 ppm to 3.8-3.9 ppm, to 2.0 ppm, and then to 0.9 ppm were assigned to phenyl, methoxy, and propyl protons, respectively, speculated to be of lignin origin.

The spectra shown in Figure 1.16 and 1.17a show quite weak correlations. Part of the reason why could be the aggregation due to weak intermolecular forces. Changing the solvent to DMSO-d_6 is one way to reduce these interactions. This is demonstrated nicely by the 2D NOESY spectrum of HA (Figure 1.17b) acquired using the mixing time of 250 ms.[22] This spectrum shows NOEs of NH, aliphatics and aromatic protons, reported before, as well as chemical exchange cross-peaks between OH/COOH protons, only visible due to the change in solvent. In general, the above cases indicate the importance of sample conditions on the appearance of NOESY spectra of HA.

While HA is subject to aggregation, FA is less likely to aggregate and thus the 2D NOESY spectra appear more resolved (Figure 1.18).

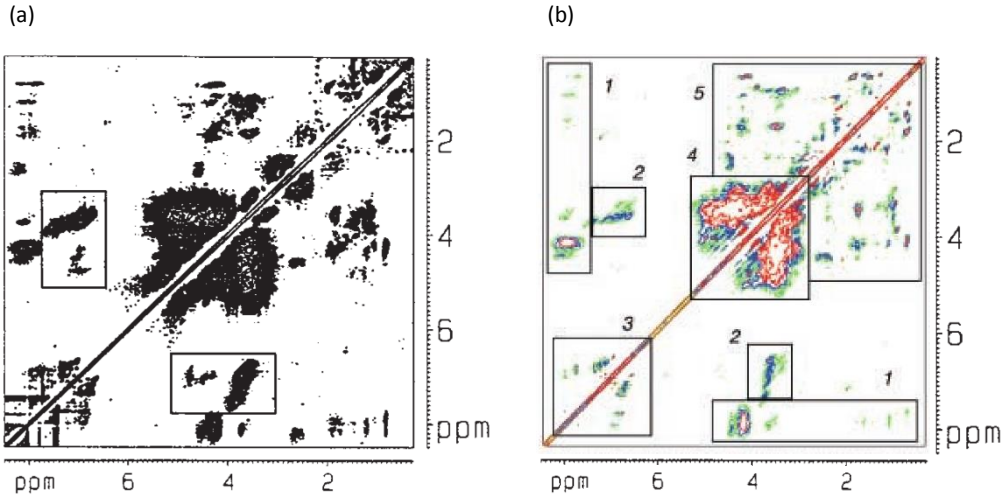


Figure 1.18 2D NOESY spectra of (a) pine forest soil FA[217] and (b) oak forest soil FA[222] in DMSO-d_6. Labels denote (1) amino acids; (2) lignin aromatics (methoxyl–aromatic interactions); (3) lignin aromatics (interactions between aromatic protons); (4) methylene and methyl units in amino acids and aliphatic structures.

The spectra shown in Figure 1.18 represent FA samples from a pine and oak soil, respectively, in DMSO-d_6.[217, 222] It is clear that both are very similar and show more defined cross-peaks than seen for the HA samples, particularly for the aromatic and aromatic methoxy/aliphatic regions.

Depending on the rotational correlation times, the NOE effect can have a positive or negative enhancement, or in some cases a zero enhancement. In the latter case, cross-peaks will be absent in the 2D NOESY. In such instances, Rotating-Frame Overhauser Effect SpectroscopY (ROESY) needs to be considered. While NOE peaks can be positive or negative, the ROE cross-peaks are always positive, irrespective of the size of the molecule. However, the ROESY pulse sequence uses a spin-lock in the same way as the TOCSY sequence, thus TOCSY cross-peaks can occur in

ROESY spectra, which have to be minimised. An example of a ROESY experiment performed on forest soil FA in DMSO-d_6 is shown in Figure 1.19.[217]

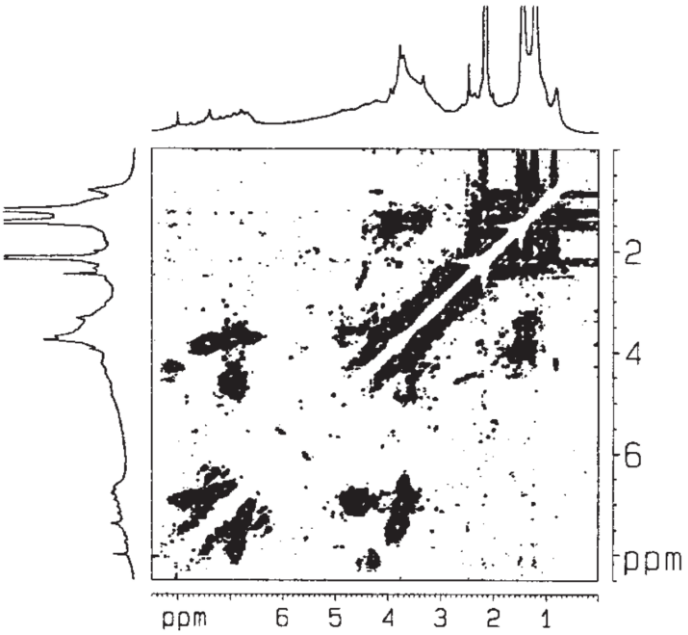


Figure 1.19 A 2D ROESY spectrum of pine forest soil FA in DMSO-d_6.[217]

The 2D ROESY spectrum was acquired to complement the 2D NOESY spectrum shown in Figure 1.18a. In this spectrum, it is interesting to note the disappearance of cross-peaks assigned to amino acids. This is likely due to T_2 effects, which may attenuate the signals of larger molecules.[217]

In summary, 2D NOESY and ROESY experiments are vital in providing information regarding remote protons in molecules, especially if the molecules are highy substituted and thus have limited through-bond proton-proton transfer capability. However, 2D NOESY and ROESY spectra obtained from HS samples contain weak negative cross-peaks, due to aggregation of samples. The cross-peaks overlap making specific correlations of individual molecules difficult to identify. In addition, many cross-peaks seen in the spectra are not unique as the same correlations can also be seen in *J*-correlated spectra. In general, the information obtainable from such spectra is limited.

1.5.2.3.2. Diffusion ordered NMR spectroscopy

Diffusion Ordered SpectroscopY or DOSY is an NMR technique that has had a major impact on HS studies. The DOSY experiment uses PFGs (see Section 1.7.6) to measure diffusion coefficients (*D*), which report on the hydrodynamic radii or size of molecules. In practice, this is achieved by

measuring a series of spectra with incremented PFG strength (*G*). From the resulting signal intensities (*I*), the diffusion coefficients are extracted using a least squares regression analysis of the linearised form of the following equation:

$$I = I_0\, exp\left[-D(G\gamma\delta)^2(\Delta - \frac{\delta}{3} - \frac{\tau}{2})\right], \quad [1]$$

where δ is the duration of the PFG, Δ is the diffusion delay time, τ is a short recovery delay and γ is the gyromagnetic ratio of the nucleus, typically ^{1}H.[227, 228] This procedure has been applied to HS samples; however, the interpretation of spectra is not straightforward due to the severe overlap of resonances in HS spectra. Better models for more complex systems such as HS were required and a few modifications of the basic procedure have been proposed to obtain better results.

For example, Dixon and Larive[228, 229] investigated the use of non-edited, relaxation-edited and spin-echo edited Bipolar Pulse Longitudinal Eddy current Delay (BPPLED) DOSY pulse sequences on a SRFA sample, to provide both selective diffusion information and to enhance spectral resolution of overlapped signals. The diffusion coefficients varied across the regions, demonstrating the complexity of the sample, but the obtained hydrodynamic radii agreed well with small angle X-ray scattering results, assuming spherical molecules. Despite the fact that the two edited DOSY experiments allowed clearer examination of overlapped regions compared to the non-edited BPPLED sequence, when applied to a model mixture, the benefits were less obvious when it was applied to FA. In the same study, a spin-echo edited approach did not show any improvement as the method can be tuned to only one value of ^{1}H-^{1}H coupling. Thus these experiments may work if there were two components, which can be isolated by the spin-echo or relaxation editing, but for multiple components there is too much overlap and the obtained diffusion coefficient must be classified as "apparent". For polydisperse mixtures such as FA and HA it would be expected that a continuum of diffusion coefficients exist. Thus, the use of a single exponential model to characterise the diffusion of individual molecules is bound to fail. This is illustrated in Figure 1.20 where linear regression using a single decay constant is presented for samples of SRFA and Nordic Sea HA (NSHA). It is evident, from Figure 1.20, that both samples deviate from a single exponential decay, a feature more pronounced for HA than FA. Thus to account for this inherent heterogeneity, DOSY NMR data need to be treated differently.

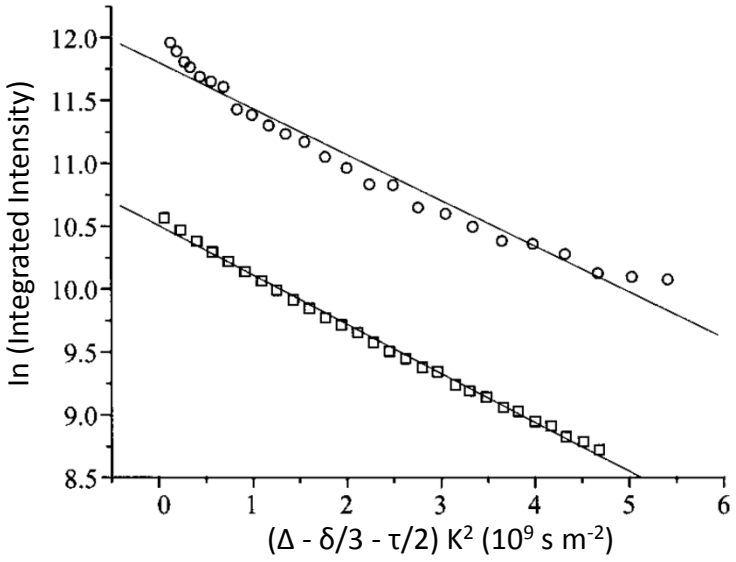


Figure 1.20 Representation of DOSY data of SRFA(□) and NSHA(○) in the form of echo intensity versus PFG strength.[229]

Analysis that goes some way towards accounting for the polydispersity of HS samples was initially developed by Morris *et al.*,[229] who produced the first 2D DOSY spectrum of an HS sample, a plot of ^{1}H chemical shift against diffusion coefficient. The advantage of this 2D representation is that it provides visual comparison of different components and their relative D.

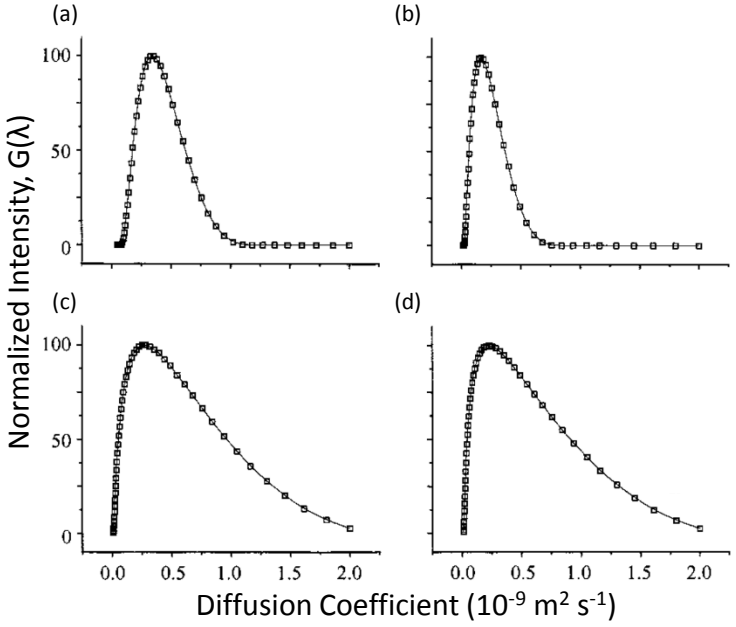


Figure 1.21 CONTIN diffusion coefficient distributions for the resonances between 1.7 to 3.3 ppm (assigned to protons α to aromatic or electronegative functional groups) of (a) SRFA, (b) SFHA, (c) NSHA and (d) peat HA.[229]

The same authors also used a procedure called CONTIN, a general purpose constrained regularisation for analysis of multi-exponential decay, which reports a continuous distribution of diffusion coefficients, thus better reflecting the polydispersity of the HA and FA samples (Figure 1.21). The CONTIN distributions of each chemical shift region indicate that all samples measured are mixtures of mostly small and some large molecules but there is no evidence of very large molecules. It should be noted here that to obtain accurate diffusion coefficient distributions, high-quality (i.e. high S/N ratio) data are required. Furthermore, to adequately describe the non-exponential decay of the NMR signal with gradient strength, at least 25 different gradients must be used in each PFG-NMR experiment.[229] Even though CONTIN analysis recognised the polydispersity, it can be at best semi-quantitative. It should not be used as a quantitative measure, which would require separation of individual components.[227]

Simpson *et al.*[227] used DMSO-d_6 as a solvent to disrupt the molecular aggregates of a soil FA sample and obtained a 2D DOSY spectrum using biexponential fitting (Figure 1.22).

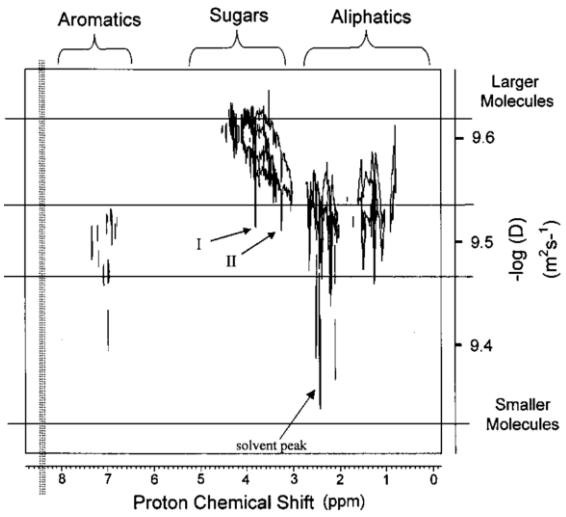


Figure 1.22 A 2D DOSY spectrum of soil FA in DMSO-d_6. Labels (I) and (II) denote CH_2 groups adjacent to esters and ethers, respectively. Reproduced from ref.[227]

The spectrum shows different diffusivities for different compound classes. Sugars exhibited the slowest diffusion, as would be expected for polysaccharides, compared to the much smaller, faster diffusing aromatics. This study provided clear, undeniable evidence that FA is a mixture and contains mostly small molecules. The same author carried out a more detailed study of aggregation by 2D DOSY.[75] The addition of acetic acid disrupted the aggregate formations in peat HA as seen by the separation of components in the diffusion dimension (Figure 1.23).

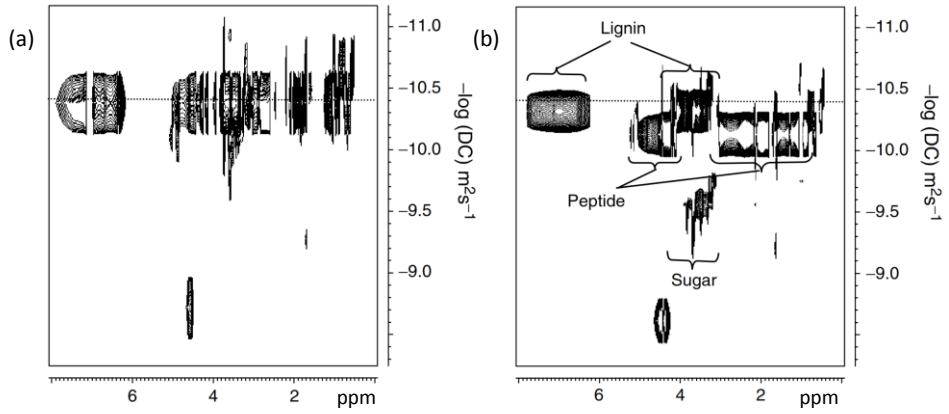


Figure 1.23 2D DOSY spectra of peat HA (a) before and (b) after the addition of acetic acid.[75]

Analysis of the 2D DOSY spectra, shown in Figure 1.23, showed that HA contains molecules such as lignin, peptides and oligosaccharides, which have diffusivities consistent with molecular weights of ~2500, ~1000 and 200–600 Da, respectively.

Additional experiments showed that the aggregation of HA was concentration dependent, while FA showed only the effects of viscosity (Figure 1.24a). The study also showed that while FA contains mostly small molecules some larger oligosaccharides were present (Figure 1.24b). A comparison with a number of maltodextrin standards (Figure 1.24a) determined the molecular weight of these oligosaccharides to be round ~1000 Da.

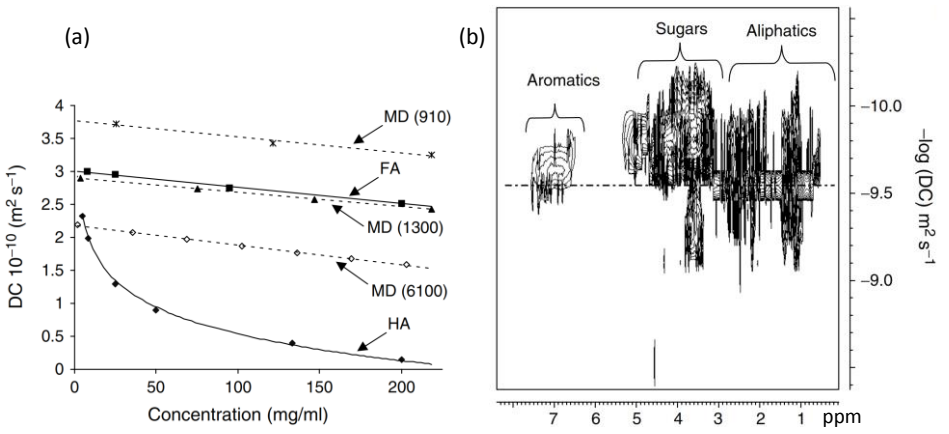


Figure 1.24 (a) Comparison of the diffusivities of a series of maltodextrins (MD), soil FA and peat HA. (b) 2D DOSY spectrum of soil FA. Reproduced from ref.[75]

A similar study was conducted by Piccolo *et al.*[230] who studied the diffusion of multiple HA and FA samples extracted from peat, a volcanic soil and coal. The study analysed the DOSY spectra

under different concentrations and upon addition of acetic acid. They reached the same general conclusions as Simpson *et al.*,[75] with the additional finding that the largest components observed depended on sample source. Specifically, the slowest diffusing components of HA from the volcanic, peat and coal sources were carbohydrates, alkyl groups or aromatics respectively.

1.5.2.3.3. 2D heteronuclear correlation experiments

Heteronuclear experiments have the advantage of using the increased frequency range of hetero-nuclei such as ^{13}C, which have their chemical shifts dispersed over a much wider area than protons thus potentially reducing overlap. Starting with one-bond correlations, one of the original experiments used to correlate ^{1}H with ^{13}C nuclei was HETeronuclear CORrelation spectroscopy or HETCOR.[231] However, this experiment is not commonly used anymore due to the fact it relies on ^{13}C detection, which is less sensitive than ^{1}H detection. Heteronuclear Multiple Quantum Correlation (HMQC)[232] and Heteronuclear Single Quantum Correlation (HSQC)[233] are the two heteronuclear experiments that use ^{1}H detection and are almost exclusively used today. While the HMQC experiment is less prone to problems associated with miscalibrated pulses, it suffers from lower resolution due to ^{1}H-^{1}H coupling modulation of cross-peaks in the ^{13}C dimension. The HSQC experiment does not have this problem and is the preferred experiment at present, albeit only recently for HS samples. The main issues with ^{1}H-detected experiments is that ^{12}C-attached protons must be suppressed. Although this used to be a problem in the past, PFGs used routinely these days deal with this issue very efficiently. However, as HSQC records proton-proton couplings in F_2, some argue that HETCOR, with an additional BIlinear Rotational Decoupling (BIRD) element, can resolve peaks with a 0.01 ppm resolution and that for complex mixtures this experiment should be used.[234] The recent appearance of pure-shift methods may settle this issue as they produce singlets in F_2.[235]

To follow up the COSY experiment discussed in Section 1.5.2.3.1, Haiber *et al.*[236] analysed their tangential flow multistage ultrafiltration fractions of aquatic SRFA and SRHA samples using ^{1}H, ^{13}C HETCOR. Initial inspection of 1D ^{1}H spectra acquired on all fractions (< 3 kDa to > 100 kDa) showed that the spectrum from the lowest MW cut-off (< 3 kDa) contained the most resolved peaks (small molecules), a first indication that fractionation by size had indeed taken place. Upon increasing MW, there were variations in the number of signals in certain regions in the SRFA sample spectra. For example, the aromatic and carbohydrate signals increased, while with increasing MW the aliphatic signals decreased. Further structural information came from a few

tentative assignments of the ^{1}H, ^{13}C HETCOR spectra, with an additional multiplicity edited HETCOR spectrum confirming them. The methoxy (58.8/3.94 ppm) and aromatic (117.5/7.07 ppm) signals were assigned to coumaryl, guaicyl or syringyl, the known monomers of lignin, all of which were more prevalent in LMW fractions, while aliphatic cross-peaks (e.g. 80.2/4.62 ppm) were assigned to phenylpropane chains. The authors hypothesised, albeit from very little information, that: (i) decomposition of lignin moieties involves loss of phenylpropane chains, followed by dealkylation of methoxy groups, (ii) aggregation of aromatics occurs during this process. The trends for SRHA sample were not very clear as the HETCOR spectra showed very few cross-peaks and ^{1}H spectra showed little differences between fractions. A possible explanation is that the SRHA sample was aggregated and thus ultrafiltration could not fractionate individual components.

Kingery *et al.*[191] attempted to show the power of a combination of heteronuclear and homonuclear 2D NMR experiments on soil HA. In this study 2D TOCSY and 2D HMQC were used together to make less ambiguous assignments. However, as witnessed for the aforementioned 2D HETCOR spectrum of SRHA,[236] the HMQC of the soil HA sample contained very few cross-peaks especially in the aromatic region. Nevertheless, using both these experiments, the authors identified correlations belonging to fatty acids and peptide amino acids.

Similar results were reported by Fan *et al.*[223] who acquired an HSQC spectrum, in addition to NOESY and TOCSY spectra, on a soil sample treated with a metal complexing agent, Tiron, used to remove exchangeable metal ions to increase the solubility of the HS sample. The spectra obtained were quite noisy but did display characteristic cross-peaks of amino acids and carbohydrates containing α- and β-pyranoses, long chain alkanes and methoxyphenylpropanyl structures. Furthermore, from the intensity of the cross-peaks the authors concluded that aromatic structures must be quite rigid (weak cross-peaks) unlike amino acids, which showed the strongest cross-peaks.

Simpson *et al.* examined a soil FA fraction at pH 12.6 using a series of 1D and 2D experiments including an 2D HSQC, an expansion of its aliphatic region is shown in Figure 1.25.[196] Using the 2D ^{1}H, ^{13}C HSQC, as well as the other 2D experiments, the authors identified mono and dicarboxylic acids, as well as esters, alcohols and ethers. The assigned dicarboxylic acid chemical shifts were similar to those reported by Fan *et al.*[223]

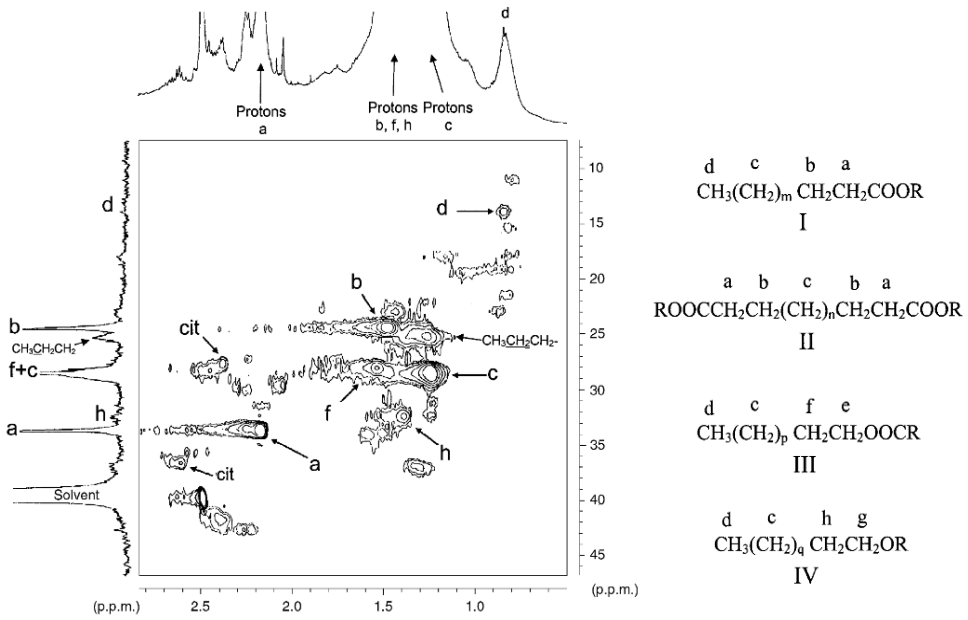


Figure 1.25 500 MHz 2D ^{1}H, ^{13}C HSQC spectrum of podzol FA in DMSO-d_6. Possible assignments of the labelled resonances are given on the aliphatic chains to the right of the spectrum. Note 'cit' indicates citrate.[196]

In other regions of the HSQC spectrum, they also tentatively reported carbohydrates, amino acid residues as well as substituted benzenes. However, despite the number of assignments, only small parts of the 2D spectrum were presented and showed considerable signal overlap.[196]

There is a number of ways one can enhance the quality of hetero-correlated spectra. For example, Simpson *et al.*[217] acquired 2D phase-cycled HMQC spectra with a BIRD pulse train preceding the HMQC pulses (Figure 1.26a), which reduced the cancellation artefacts dominating the spectrum acquired without the BIRD element (Figure 1.26b).

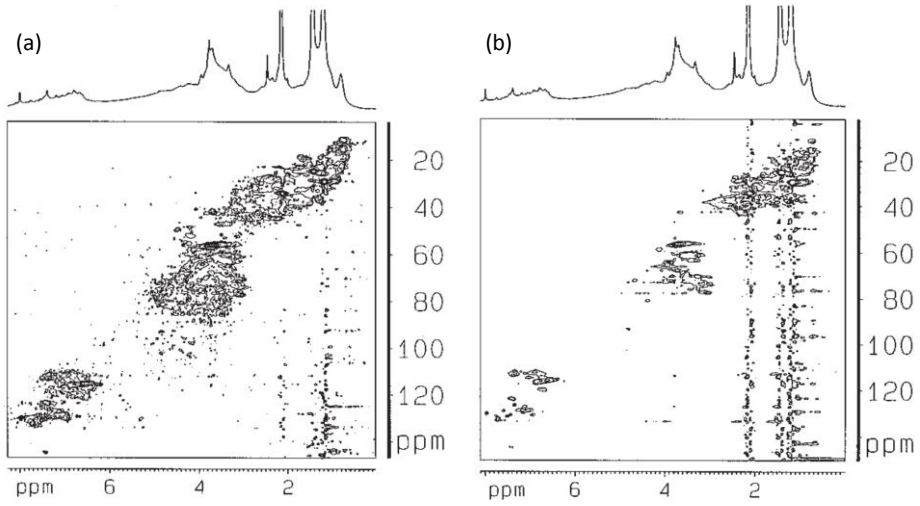


Figure 1.26 2D ^{1}H, ^{13}C HMQC spectra of pine forest FA (a) with and (b) without a BIRD pulse train.[217]

Simpson *et al.* remarked that this method cannot be applied to big molecules which have negative NOEs. It should be noted that this approach is now superseded by the use of PFGs offering superior reduction of cancellation artefacts.

Other modifications to the standard 2D experiments are only starting to be used in HS research. For example, the sensitivity-enhanced ^{1}H, ^{13}C HSQC was endorsed by Simpson *et al.* for giving the most information in the shortest time.[196, 217]

All experiments mentioned so far provide only one-bond correlations, which cannot provide information about connectivities of functionalised carbons. To get a real handle on the structures, especially those with carboxylic and hydroxyl functional groups, long-range ^{1}H-^{13}C couplings need to be used. The most powerful experiment for this task is the Heteronuclear Multiple Bond Correlation (HMBC) experiment, which reports on two and three bond heteronuclear couplings. Only three examples of the 2D ^{1}H, ^{13}C HMBC experiment acquired on HS samples have been reported in the literature to date.

The first example was provided by Simpson *et al.*[196] who used 2D HMBC to support their proposed assignments of phthalate and 1, 3, 4-trisubstiutued benzene structures of soil FA. The obtained spectrum showed very few cross-peaks but was still invaluable in the analysis.

Cook *et al.*[200] acquired a much richer 2D ^{1}H, ^{13}C HMBC spectrum of soil FA, with numerous cross-peaks in the aliphatic and single heteroatom substituted aliphatic region. Cross-peaks in the aromatic region were assigned to functionalised phenols. One major benefit of 2D HMBC is that quaternary phenolic carbons do not overlap with the carboxy carbons, which makes them easily distinguishable. For example, from the examination of the chemical shifts the authors were also able to discriminate between cyclic and acyclic ketonic carbons.

In addition to soil, 2D HMBC experiments have also been performed on lake DOM.[204] The focus of this study was on the major components of DOM, previously mentioned CRAM, Section 1.5.2.1, and Material Derived from Linear Terpenoids (MDLT). The resulting spectra showed very little aromatics or carbohydrates but high intensity in the functionalised aliphatic region, with assignments confirming previous findings of Hertkorn *et al.* regarding CRAM.[143] Using spectral prediction Lam *et al.*[204] suggested that a group of signals below 2 ppm in ^{1}H and between 20-80 ppm in ^{13}C spectra did not match cyclic structures nor could be due to fatty acids but in fact belonged to MDLT.

The only other example of the 2D HMBC experiment was a study focusing on carboxylic acid functionalities in soil HA/FA and SRFA/HA.[225] The 2D HMBC experiment was used to correlate carboxylic carbons with protons two or three bonds away. The authors used a coupling constant of 8 Hz to optimise the evolution intervals, and found that the majority of carboxyl groups were correlated to unsubstituted aliphatics and alicyclic acids. When they tried to optimise the experiment for a 5 Hz coupling, the intensity of aromatic carboxyls did not increase, only an overall decrease in intensity was observed due to longer evolution delays (calculated as $1/2\ {}^{n}J_{CH}$).

The main issue with 2D HMBC are the long evolution intervals, which lead to excessive relaxation during the pulse sequence and consequently poor sensitivity. Another issue is the large spread (0-8 Hz) of long-range coupling constants that makes it difficult to optimise the experiment fully. This can be partly solved by measuring more than one 2D HMBC spectrum, each with a different coupling constant, however this costs time. Finally, 2D HMBC experiments without refocusing periods produce mixed phased multiplets that can attenuate the cross-peak intensity due to the cancellation of opposite phase lines - a problem augmented by overlap of signals in complex mixtures.

1.5.2.3.4. Experiments combining homo- and heteronuclear polarisation transfers

The field of biomolecular NMR has produced a number of 3D experiments that combine different polarisation transfer pathways. Only recently, however, have such experiments found their way into other fields such as environmental NMR. Many of the 3D experiments also can be acquired in a 2D fashion. One of the most common variants is the 2D HSQC-TOCSY[22, 237] or 2D HMQC-TOCSY.[196] Both experiments correlate directly bonded proton-carbon pairs, but also extend the correlation to other protons of the same spin system. Such experiments are very useful for extending partial structures obtained from basic 2D methods. For example, Simpson *et al*.[217] used 2D HSQC-TOCSY to confirm their assignments from 2D COSY, HSQC. The same author used 2D HMQC-TOCSY[196] to correlate protons and carbons in the same spin system, piecing together information obtained from 2D COSY and 2D HMBC.

1.5.2.4. Three-dimensional (3D) NMR studies of HS

As illustrated above, even with the addition of a second dimension HS samples still provide complex 2D NMR spectra. Thus, higher dimensions are desirable to reduce the spectra overlap

as much as possible. There are not many examples in the literature of 3D NMR spectra of HS. One of the best examples was provided by Simpson *et al.*[72] who acquired a 3D HMQC-TOCSY spectrum of a pine forest soil FA sample. Despite still containing overlapping cross-peaks, extension into a 3rd dimension spread the peaks considerably (Figure 1.27a).

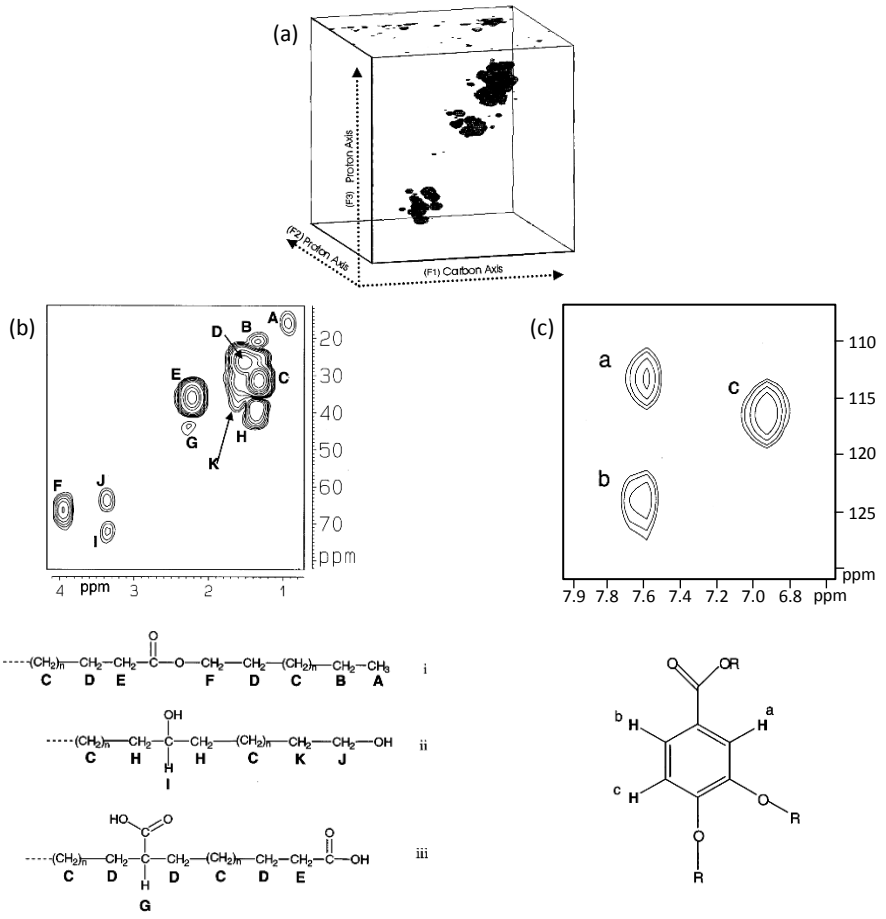


Figure 1.27 (a) 3D HMQC-TOCSY spectrum of pine forest soil FA; (b) F_1F_2 plane from the 3D spectrum in (a) at 1.3 ppm. Proposed structural fragments are shown below the spectrum with cross-peaks labelled accordingly; (c) F_1F_2 plane taken at 6.9 ppm. Proposed structural fragment is shown below the spectrum with cross-peaks labelled accordingly.[72]

A 2D plane extracted at the ^{1}H chemical shift of 1.3 ppm (Figure 1.27b) showed a number of HMQC cross-peaks, which could be tentatively assigned to aliphatic chains containing internal carboxylic, hydroxyl, ester groups and terminal carboxyl, hydroxyl or ester groups. The author speculated that due to the similarity between aliphatic cross-peaks of FA and that of tomato cuticles that these aliphatics originated from plant cuticles from forest floor vegetation. However, other lipid-rich plant components such as plant resins were not examined which could also explain the above cross-peaks.

Another 2D plane extracted at the ^{1}H chemical shift of 6.9 ppm (Figure 1.27c), shows aromatic HMQC cross-peaks, which can be tentatively assigned using database chemical shifts to the dimethoxy benzoate structure. The authors suggested that the assigned compound has a similar substitution pattern as lignin monolignols, although they could not derive the nature of the OR groups.

Despite this, 3D experiments present one way to separate out the individual spin systems for structural identification and as such 3D NMR spectroscopy promises to be a very powerful tool in HS studies.

1.5.2.5. The use of databases and predictive software in the analysis of NMR spectra of HS

In some NMR studies of HS, predictive software is used to aid interpretation of 2D spectra. This is done routinely by Simpson *et al.*[238, 239] using ACD labs prediction software. For example, this software was used to assign chemical shifts obtained from the pine forest soil FA sample COSY, TOCSY and HMQC spectra (see Sections 1.5.2.3.3 and 1.5.2.3.1). The assignments were not unambiguous as the programme gave multiple possible structures per chemical shift. The authors speculated that the most probable aromatic structures ware lignin-derived, however they did not observe lignin side chain resonances, which they suggested was due to degradation. In addition, 2D HMBC was not used to connect the resonances identified by 2D COSY and 2D HMQC experiments.

Perdue *et al.*[240] used a program called SPARIA (Substitution Patterns in Aromatic Rings by Increment Analysis) to predict the substitution patterns of aromatic rings by interpreting the 2D ^{1}H, ^{13}C HSQC spectra of SRNOM. The authors reported the predicted substitution patterns from the most intense cross-peaks analysed and concluded that these patterns are common to lignin but also non-lignin derived materials. The chemical shift matching was carried out using only the chemical shifts of CH pairs. Therefore, in order to prove that the interpretation is correct multiple correlated chemical shifts per molecule are required. This approach was illustrated by Simpson *et al.*[196] who combined data from 2D ^{1}H, ^{13}C HSQC, 2D ^{1}H, ^{13}C HMBC, 2D ^{1}H, ^{1}H COSY and 2D HMQC-TOCSY, thus increasing the number of correlated nuclei. This only works well of course if the cross-peaks are sufficiently separated.

1.5.2.6. Summary of the current molecular-scale understanding of soil and peat FA from NMR studies

This study focuses on the analysis of a peat FA sample, therefore the current understanding of the structural composition of soil and peat FA is presented here in two parts focusing on: non-aromatic and aromatic compounds.

Non-aromatic compounds:

Soil/peat FA contains low molecular weight compounds mixed with potentially large ones. DOSY experiments (Section 1.5.2.3.2) indicate that the large compounds are mainly polysaccharides, polypeptide or long chain aliphatics.[75] These long chain aliphatic compounds show intense signals at 1.8 ppm in ^{1}H spectra. In the 3D NMR study described in Section 1.5.2.4 these were found to be functionalised with carboxylic or hydroxyl functional groups.[72] It is likely that these long chain aliphatics are lipids sourced mainly from plants[72] and represent a significant proportion of peat FA.[188] Polysaccharides are generally present in greater amounts in soil/peat FA than in HA based on identified anomeric and ring carbon chemical shifts,[188, 203] while peptides appear in larger quantities in peat HA.[188]

Several studies propose that the aliphatic molecules in HS samples are similar in structure to plant and microbial constituents and their degradation products.[70-72] Kelleher *et al*.[71] overlaid the 2D HSQC ^{1}H, ^{13}C spectra of tomato cutin, amylopectin and bovine serum albumin, as well as kraft lignin, lignin generated in the pulp industry, representing the four classes of compounds found in plants (fatty acids, carbohydrates, proteins and lignin) and compared the result to a soil HA spectrum. The similarities were clear, indicating that aliphatic structures in HS resemble these biopolymers. This conclusion is supported by a comparison of the 2D HSQC plane extracted from the 3D spectrum of soil FA (Section 1.5.2.4), with a 2D ^{1}H, ^{13}C HSQC of tomato cutin which were practically identical.[72] This does not mean that these are the exact and only compounds present in HS, but instead implies that the major components have closely related compositions to these representative biopolymers.

Small molecules identified in FA samples include mono and dicarboxylic acids with an average chain length of ten carbons.[196] Very small molecules such as formate[241], acetate[195] and adipic acid[196] have also been identified in many soil/peat FA samples.

Aromatic compounds:

Most researchers point out that the aromatic structures in soil/peat FA are lignin-like. FA and HA NMR spectra from various studies highlight ^{13}C signals at 115-117 (ortho to OH/OMe groups as in softwood lignin) and 130-132 ppm (meta to OH/OMe groups as in grass lignin).[188, 203] A comparison of the 2D HSQC of soil HA and kraft lignin, mentioned above, showed very similar features.[71] Of course, this does not mean that unaltered lignin is a component of HS, but indicates that the degradation products must resemble the monolignols. However, based on the chemical shift analysis of isolated cross-peaks, the presence of several non-lignin structures including di- and tri- hydroxy/alkoxy substituted benzoic acids have been proposed.[72, 188, 196] As mentioned in Section 1.5.2.5, spectral prediction methods have also indicated the similarity of aromatic substitution patterns to lignin monolignols but also non-lignin aromatic molecules.

In summary, NMR spectroscopy thus far does indicate that the major compounds in HS resemble plant constituents, but at this stage the exact forms of these molecules remain unknown.

1.6. Chemical modification of HS for the purpose of structural characterisation

Carboxyl and hydroxyl groups are major functional moieties of HS (Section 1.3). These groups are important for the physiochemical properties and behaviour of HS in the environment (Section 1.3.2). Thus, it is unsurprising that a substantial research effort has focused on establishing the distribution of these functional groups in HS compounds.

NMR spectroscopy is well suited for probing the neighbourhood of such functional groups. However, to distinguish the signals originating from these molecules, and at the same time ensure high sensitivity of such experiments, chemical modifications of these groups by incorporating NMR active nuclei is required. The easiest method of chemical modification is the exchange of labile protons of carboxyl and hydroxyl groups with a group containing an NMR active element. The nuclei that could be considered for this task are ^{13}C, ^{29}Si, and ^{31}P. ^{31}P can be introduced into R-OH compounds using 2-chloro-4, 4, 5, 5-tetramethyldioxaphospholane[242] however, the resulting phosphitylated compounds contain a bulky five-membered ring and will therefore not be considered here further. The nuclei left to consider are ^{13}C and ^{29}Si. Before these two alternatives are examined further, it should be mentioned that ^{15}N and ^{19}F could also be

incorporated into HS. Indeed accounts of ^{15}N-tagged HS studies exist in the literature.[243] However, this involves different chemistries and not the replacement of hydrogen in OH groups. Fluorination, which has not been applied to HS yet, is a very promising technique, especially in the light of appearance of many fluorination protocols.[244] Nevertheless, the likelihood of side reactions and incomplete conversions are at the moment limiting factors.

1.6.1. Silylation of HS

Silylation of HS inserts a non-polar trimethylsilyl group in place of a labile OH or NH proton. This procedure creates a mixture of silyl esters, ethers and amides from carboxylic, hydroxyl and amide groups, respectively. The incorporated ^{29}Si nuclei can easily be monitored using 1D ^{29}Si NMR spectroscopy. This procedure has been applied to kraft lignin[245] but the partial overlap of the ^{29}Si chemical shifts of phenolic and alcoholic silyl ethers meant that only estimates of the proportions of each type were possible.

Silylation preceded by methylation was also tried on an HS sample in order to first react the most labile hydroxyl groups then silylate the remaining ones.[246] The authors acquired Insensitive Nuclei Enhancement by Polarisation Transfer (INEPT) spectra which showed the presence of silyl esters as well as silyl ethers, an indication that the methylating agent diazomethane did not fully methylate carboxylic acids, hindering distinction of COOH and OH groups. However, they did find that their INEPT spectrum was virtually identical to the quantitative ^{29}Si spectrum showing that INEPT is a possible method for quantitative studies of such functional groups. Herzog *et al.*[87] improved the methylation procedure by adding catalytic amounts of an organic acid, allowing better distinction of hydroxyl groups in a HA sample. It should be noted that the two-step method used only works if the HS fraction studied is soluble in the reagents used in both steps, which in this case was DMSO-d_6 for silylation and methanol for methylation.

The only 2D ^{1}H, ^{29}Si HSQC spectra acquired on silylated NOM sample was by Hertkorn *et al.*[22] The authors identified regions in the spectrum as corresponding to aliphatic and aromatic carboxylic acids, primary, secondary, and tertiary aliphatic alcohols, and phenols and nitrogen derivatives.

Although silylation is evidently a valuable derivatisation method, there are several drawbacks.

(i) The ^{29}Si-enriched silylating agents that would be required for this work are practically unavailable.

(ii) Silylated compounds are moisture sensitive and require special handling procedures.

(iii) Silyl groups are bulky and steric factors can cause incomplete substitution.

(iv) ^{29}Si-^{13}C two-bond (2·1-3·0 Hz) and three-bond (1-2·3 Hz) coupling constants are smaller[247] than the analogous coupling constants in ^{13}C-methylated compounds, hence the signal transfer to and from the aromatic ring carbons is more susceptible to relaxation losses in silylated compounds.

(v) Cryoprobes capable for performing triple resonance experiments, involving ^{1}H, ^{13}C, ^{29}Si, are not common place.

Thus, due to these reasons silylation is not at present a viable method for probing directly the nature of molecules in HS.

1.6.2. Methylation of HS

Methylation is a promising method for extracting information regarding the composition and structures of HS. In comparison to silylation, the methyl unit is smaller and hence its incorporation at less accessible sites will be more probable.[22] In addition, 100% ^{13}C-enriched methylation agents are readily available and the methylated compounds are stable. However, it has been well documented that, unlike silylation, side reactions can occur upon methylation such as C, N and S-methylation.[248] The potential for such side reactions depends on the methylating reagent and reaction conditions used. However, it is important to note that side reactions are not necessarily a negative outcome as their products can potentially aid characterisation of HS compounds.

Several methylation protocols have been attempted using HS in the past. These methods fall into two main categories; pyrolysis and chemical. Pyrolysis methods typically use TMAH as a methylating agent.[249] Upon application to HS, this method produced a variety of compounds as shown from subsequent GC analysis.[250] However, incomplete methylation can occur and even partial decarboxylation of benzoic acids, with hydroxyl groups ortho or para to a carboxylic acid group. Joll *et al.*[251] proposed that this decarboxylation depends on the mole ratio of TMAH used, which is calculated based on the number of carboxylic acid groups present, not known for HS

samples. In addition, working in the gas phase at temperatures of 300-500 °C can induce decomposition, bond cleavages of certain compounds and is claimed to have poor reproducibility.[252] Overall NMR spectra of HS samples modified by thermal chemolysis should be interpreted with caution.

Purely chemical methods, on the other hand, are performed in liquid phase. These methods use electrophilic reagents, stripping off the labile protons and inserting a CH_3 group from a methylating agent. The majority of studies report the use of diazomethane or dimethyl sulphide (DMS) as the methylating reagents.[253] Diazomethane has been used extensively on soil FA and HA[254, 255] and HS from aquatic environments.[90, 248] In most cases, methylation was successful as shown by the reduction of OH bands (3440 cm^{-1}) and increase in C-H (2960 cm^{-1}, 1380 cm^{-1}, 1365 cm^{-1}) stretches in IR spectra,[253] and by the appearance of OMe resonances (~52 ppm, ~56 ppm, ~60 ppm) in ^{13}C NMR spectra.[248]

The use of diazomethane has a number of disadvantages. Firstly, the reagent itself is explosive and dangerous. Secondly, HS have limited solubility in diethyl ether, which is used to produce diazomethane in most applications. It also has been reported that diazomethane adds nitrogen to HS samples during methylation and can initiate cyclisation reactions.[248] Diazomethane is also limited in its ability to methylate certain functional groups. Most studies showed that it only methylates carboxylic acids, enolic hydroxyls and phenolic hydroxyls but not carbohydrate or aliphatic hydroxyls.[256, 257] Nevertheless, it is a good reagent for determining the carboxyl or phenolic content only. Arsenie *et al.*[90] demonstrated this by performing methylation followed by saponification, where the methyl esters were hydrolysed.

DMS is another extensively used reagent for methylation of soil HA[253, 258, 259] and FA[253, 256] It is also toxic, but much easier to handle, in terms of safety and solubility, than diazomethane. However, in terms of its methylation capability it has been shown to increase the carbon content via C-methylation. In addition, it is noted in a number of studies that DMS only methylates phenolic and enolic groups and not carboxylic groups.[253] Contradicting these results, Briggs and Lawson,[259] reported that under milder alkaline conditions complete methylation is possible. This difference is potentially down to the solvent used. Kuŕaň *et al.*[258] methylated coal and peat HA samples using DMS in acetone and methanol. They reported that using aprotic acetone allowed the methylation of carboxylic acids, which did not methylate in protic methanol. This result

reflects not only the interaction of the protic solvents with the nucleophile but also the higher solubility of the base and hence alkalinity in such solvents.

It is evident that both reagents have good and bad points and in all instances their methylation efficiency has been questioned. To answer this, Schnitzer *et al.*[253] studied both reactions on a variety of HA samples. They found that the pH and nature of the HA sample had an effect on the performance of each reagent. In summary, the authors suggested that diazomethane is most suitable for methylation of HA of soils that tended towards a neutral pH (> pH 5.5), since little C- (or N-) methylation occurred. While for acidic HA samples DMS performed better in terms of side reactions. They also detailed that, for FA samples, both methods were poor and that methyl iodide (CH_3I) is a better alternative.

To improve the methylation efficiency some studies have used a combination of methods where a second methylation step follows the initial one. Wersaw *et al.*[257] devised one such method using diazomethane followed by CH_3I as the second methylating agent. In this procedure, the first step methylated the carboxylic acid, enolic and phenolic groups, while the second methylated the remaining aliphatic and carbohydrate hydroxyl groups. After both steps, the product had increased solubility in chloroform. In addition, a mild hydrolysis step was conducted to hydrolyse the esters, allowing differentiation between carboxylic groups and other groups. Thorn *et al.*[248] also used enriched diazomethane/CH_3I permethylation to study the differences in functional groups from different sources by ^{13}C NMR spectroscopy. To do this they used ^{13}C-enriched diazomethane and CH_3I in one reaction and non-enriched diazomethane but $^{13}CH_3I$ in another. They inspected the spectra after each step recording the observed resonances in the regions around 52, 56 and 60 ppm for methyl ester, phenolic/enolic and phenolic methoxy carbons, respectively. After the final step resonances were recorded at 55 and 59 ppm, which were linked to carbohydrates.

Mikita *et al.*[254] methylated both soil HA and aquifer FA, using $^{13}CH_3I$; the resulting ^{13}C NMR spectra showed only the methylated groups due to the incorporation of the fully abundant ^{13}C nucleus (Figure 1.28).

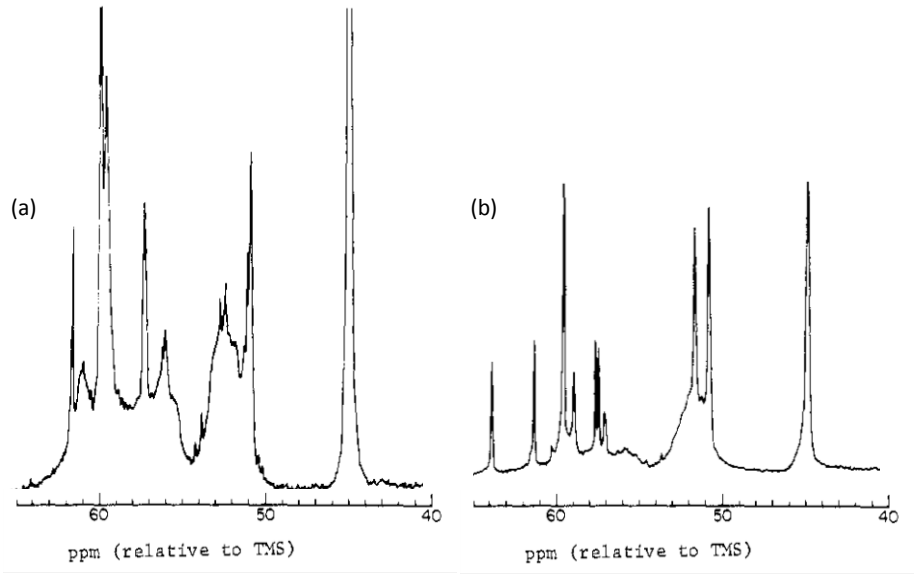


Figure 1.28 ^{13}C NMR spectra of ^{13}C-methylated (a) soil HA and (b) aquifer FA prepared using $^{13}CH_3I$ as a methylation agent.[254]

Combining the information from the spectra shown in Figure 1.28 with the work of Thorn *et al.*,[248] a summary of the assignments of chemical shifts made from the analysis of methylated HS samples is given in Table 1.5.

Table 1.5 Tentative assignment of ^{13}C methoxy resonances of methylated HS samples from two studies.

Assignment	Reported chemical shift (ppm)[a]	Reported chemical shift (ppm)[b]	Alternative assignment
R_2NCH_3	45.6	-	
Aliphatic $COOCH_3$	51.3	52	
Aromatic $COOCH_3$	52.3	52	
Aryl OCH_3	55.7-56.3	55	Carbohydrate OCH_3
		56	Enolic or phenolic OCH_3
$ROCH_2CH_2OCH_3$	58.9	-	
Carbohydrate OCH_3	58.9-60.8	59-60	Aliphatic or aryl OCH_3
R_2CHOCH_3	57.5	-	
Enolic OCH_3	61	61	Phenolic OCH_3 adjacent to two substituents

[a]Mikita *et al.*[254]; [b]Thorn *et al.*[248]

The assignments given in Table 1.5 were made using the chemical shifts of model compounds and therefore caution should be taken in extrapolating these to real samples. The spectra shown in Figure 1.28 compare both HA and FA albeit from two different sources. Nevertheless, it is evident that FA contains more $COOCH_3$ groups and a lower aromatic to aliphatic carboxyl ratio compared to HS, while HA has more phenolic and carbohydrate groups.

The major drawback of CH_3I is that it is prone to side reactions particularly carbon methylation. Thorn *et al*.[248] reported that following methylation by CH_3I the subsequent ^{13}C NMR spectra showed a broad hump of resonances around 24 ppm, which is the typical chemical shift region for C-CH_3 groups of activated aliphatic carbons. These side reactions can provide an additional source of information, but the unpredictable nature of such modifications may make their analysis by NMR more difficult. CH_3I itself is also very toxic and reactive. In addition, the reaction requires a strong base to remove the acidic protons, which creates copious amounts of inorganic salts as by-products. The base typically used is NaH, which is very reactive and requires anhydrous conditions. Thus, it is beneficial to look at milder reagents that provide full methylation of HS compounds with limited side reactions. One possibility is to use CH_3I in combination with the phase transfer catalyst, TBAH. This method was developed by Liotta *et al*.[260] to methylate coals. Phase transfer catalysts are soluble in both aqueous and organic phases. By forming a complex with a reagent, which is only soluble in one phase, it can transfer that reagent to the other phase, where the reagent itself is insoluble, to carry out a reaction.[261] In terms of HS, TBAH can replace NaH, dissociate the labile protons in aqueous phase, then transfer the humic anions to the organic phase to react with CH_3I. TBAH is also inexpensive and much safer to use compared to NaH.

Piccolo *et al*.[252] used TBAH with CH_3I to methylate an HA sample. The authors commented that the methylation was successful but residual reagent was difficult to remove, indicated by the increased amount of nitrogen present according to the elemental analysis. The same procedure was also tried on HA by Clemow *et al*.[262] who reported that by washing the filtrate with water many times, very little residual ammonium ions were left behind.

TBAH can act as a base in place of NaH, but the procedure still requires CH_3I to fully methylate HS. A review of methylation agents[263] highlighted some greener alternatives, including dimethyl carbonate (DMC), which is a non-toxic, eco-friendly compound.[264] This reagent in the presence of a nucleophile and at high temperatures has been found to methylate a range of organic

compounds, e.g. phenols[265] or flavonoids.[266] However, the requirement of high temperatures with a gas liquid phase-transfer catalyst or an autoclave[266, 267] is problematic, as in the case of complex mixtures their use can lead to random chemical reactions. To alleviate this problem a catalyst such as 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU), which allows the temperature to be lowered to 90 °C, was introduced.[266] A typical reaction scheme using DBU and a phenol is shown in Figure 1.29.

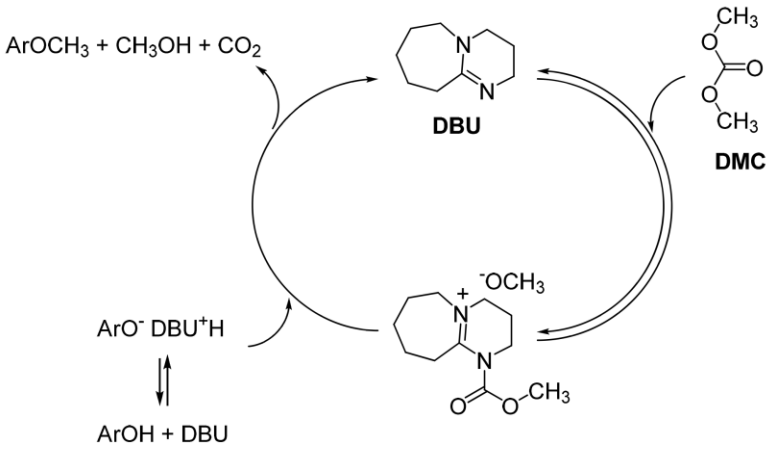


Figure 1.29 The reaction mechanism for the methylation of phenols using DMC with a DBU catalyst. Reproduced from ref.[266]

The drawback of DBU is that for some substrates like flavonoids, long reaction times are required (~72 h).[266] A possible solution is to use a microwave to heat the reaction mixture and shorten the reaction time significantly. For example, complete methylation of carboxylic acids or phenols has been achieved in 12 minutes.[268]

The use of DMC with or without a catalyst is attractive and has yet to be attempted on HS. The only potential concern is that DMC has also been used for N-methylation.[269] Nevertheless the nitrogen content of HS is low and NCH_3 resonances appear in a separate spectral region from OCH_3 signals. Carbonylation is also a possible outcome,[270] however, it is difficult to predict the outcome of DMC on HS as no work has been conducted on mixtures with multiple functionalities.

Evidently, there is still work to be done to develop eco-friendly, less hazardous methylation reactions that fully methylate HS without any side reactions. Considering the efficiency and safety of the above reactions, it was decided to explore two methylation procedures to prepare ^{13}C-labelled HS for spectroscopic studies, using CH_3I and NaH or CH_3I and the phase transfer catalyst, TBAH, as a base.

1.7. NMR theory and design of experiments

NMR spectroscopy is a powerful non-destructive analytical technique used for the structure determination of molecules. The modulus operandi of NMR is based on the intrinsic property of nuclei called spin, equipping them with angular and magnetic momenta; the latter interacts with and becomes ordered in an external magnetic field. As a spectroscopic technique, NMR probes the tiny energy level differences between the ordered spins states by application of radiofrequency (RF) radiation. The magnetic properties of different types of nuclei, characterised by their gyromagnetic constant, are different resulting in unique resonance frequencies. Nuclei also interact with each other, either directly or indirectly via electrons. These interactions empower NMR to provide structural information about molecules.

Since its discovery, 70 years ago, the arsenal of NMR experiments has expanded dramatically providing the spectroscopist with not only information about molecular compositions but also their diffusion properties and ability to interact or react with each other.

A long-standing weakness of NMR is its inability to deal, at the level of individual molecules, with complex mixtures. The lack of resolution and sensitivity reduces the potential of standard NMR experiments to reveal the structures of distinct molecules. To solve the structure of an unknown molecule, multiple correlations between nuclei need to be established to piece together individual atoms like pieces of a jigsaw.

As illustrated in Section 1.5.2, standard NMR experiments have been applied to HS but have provided limited information. This is because for chromatographically inseparable mixtures NMR methodology needs to be tailored to be able to separate out the individual molecules spectroscopically and establish multiple correlations between their nuclei to enable structural characterisation. One way how this could be achieved is illustrated in this study. NMR experiments can be described by pulse sequences; cascades of radiofrequency pulses separated by precisely timed delays that generate and select signals that carry specific information. In order to design, describe and analyse these sequences, mathematical and intuitive apparatus is required.

Simple NMR experiments can be described by a vector model. In this model, magnetic moments of spins are represented by vectors, which are summed to give a resultant magnetisation vector. The model then follows the motions of this vector under the influence of RF pulses and the

external magnetic field, allowing simple illustrations of some of the most fundamental NMR experiments for a spin-1/2 nucleus. However, usage of this model is limited as it is not able to describe complex NMR experiments, e.g. those that involve the evolution of multiple quantum coherences.

A full description of any spin manipulation is possible using the density operator theory.[271] However, with the in-depth mathematical description involved, it is a rather time consuming and complex approach, which lacks the ability to illustrate the physical nature of events. Spin product operator theory on the other hand is a middle road approach to NMR theory. Developed from density matrix theory by Ernst's group, it can be used to describe the inner workings of even the most complicated NMR experiments in a simple manner.[272] In this section, the basic spin product operator theory is described in order to give the reader the necessary background information to follow the workings of the proposed experiments.

1.7.1. Product spin operators

The information provided in this section is derived from a number of references.[215, 272, 273] Basic spin physics states that nuclei have a spin angular momentum, which is represented by a vector. This vector is described by its x, y and z components, I_x, I_y and I_z, which are associated with operators $\hat{I}_x$, $\hat{I}_y$ and $\hat{I}_z$. Operators like these are the machinery used in the spin product operator treatment of NMR experiments. Despite the name, an operator is used to describe states of existence before and after events in an NMR sequence, i.e. pulses or evolution intervals.

In a quantum mechanical description, the density operator,$\hat{p}$, describes mathematically the state of the system by summing the different amounts of each operator, as a function of time, t:

$$\hat{p}(t) = a(t)\hat{I}_x + b(t)\hat{I}_y + c(t)\hat{I}_z, \qquad [2]$$

where a, b and c represent the amounts of each operator which change during the course of a pulse sequence.

The density operator changes when the state changes and a general expression to describe this change can be summarised as:

$$\hat{p}(t) = \exp\left(-i\hat{H}t\right)\hat{p}(0)\exp\left(i\hat{H}t\right), \qquad [3]$$

where $\hat{p}(0)$ and $\hat{p}(t)$ represent the initial and final density operators, respectively. Eqn. 3 is equivalent to:

$$\exp\left(-i\hat{H}t\right)\{initial\ operator\ state\}\exp\left(i\hat{H}t\right) \equiv \cos\hat{H}\{initial\ operator\ state\} + \sin\hat{H}\{new\ operator\ state\} \quad [4]$$

Eqn. 4 is the fundamental equation of the product spin operator formalism representing the actions of a Hamiltonian describing a particular event in an NMR pulse sequence.[273]

1.7.1.1. Radiofrequency pulses

Pulses of oscillating radiofrequency waves are applied to manipulate spins, e.g. to flip the magnetisation away from the *z*-axis towards the *x*-*y* plane. Pulses change the state of the spin system in a very short period of time. To describe this motion mathematically one needs to consider the Hamiltonian of the event that can, for example, be written for a hard pulse applied from the *x*-axis as:

$$\hat{H}_{pulse} = \omega_{RF}\hat{I}_x, \quad [5]$$

where ω_{RF} is the RF frequency. Thus, the pulse can be thought to operate on the spin system. By adapting Eqn. 3 for a x-pulse of length t_p the following expression is obtained:

$$\hat{p}(t_p) = \exp\left(-i\omega_1 t_p\hat{I}_x\right)\hat{I}_z\exp\left(i\omega_1 t_p\hat{I}_x\right) \quad [6]$$

By applying Euler's trigonometric identity and the fact that $\omega_1 t_p$ is equivalent to flip angle β, Eqn. 6 can be solved as:

$$\hat{p}\left(t_p\right) = cos\beta\hat{I}_z - sin\beta\hat{I}_y, \quad [7]$$

which is a particular form of Eqn. 4. What this treatment shows is that the application of a *x*-pulse changes the state of the system from $\hat{I}_z$ to $-\hat{I}_y$ by an amount dependent on the flip angle. This is exactly the same result as provided by the vector model, where, e.g. for $\beta = 90°$, the pulse flips the magnetisation from the *z*-axis toward the –*y*-axis. Similarly, if 180° *x*-pulse was applied then the final state of the spin system would be $-\hat{I}_z$, the magnetisation vector was inverted.

1.7.1.2. Evolution under chemical shift

After a 90° pulse, the magnetisation vector is tipped into the *x*-*y* plane. Spins with a non-zero resonance offset Ω will start to precess in this transverse plane until relaxation returns them to

their original state. In terms of spin product operators, the rotation around the z-axis can be described by the chemical shift Hamiltonian ($\hat{H}_{chemical\ shift}$):

$$\hat{H}_{chemical\ shift} = \Omega \hat{I}_z, \qquad [8]$$

where $\hat{I}_z$ is the operator identifying the axis of rotation, and Ω is the resonance offset (or strictly speaking the residual angular velocity). After a period t of free precession, at resonance offset Ω, the transformation is summarised as:

$$-\hat{\boldsymbol{I}}_{\boldsymbol{y}} \xrightarrow{\Omega t \hat{I}_z} -\hat{\boldsymbol{I}}_{\boldsymbol{y}} \cos \Omega t \ + \ \hat{\boldsymbol{I}}_{\boldsymbol{x}} \sin \Omega t$$

To determine the new operator state after pulses or free precession, the diagrams shown in Figure 1.30 can be used as a visual aid.

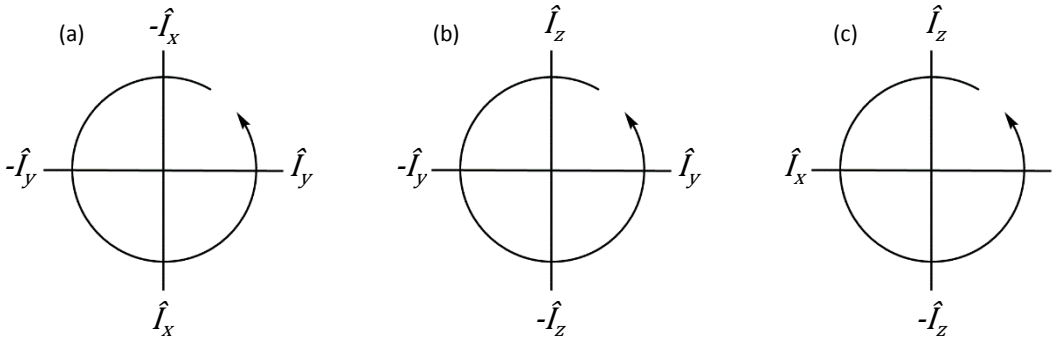


Figure 1.30 The effect of (a) free precession and (b) *x* or (c) *y* pulses on the state of product spin operators of a spin ½ nucleus. Reproduced from ref.[273]

1.7.1.3. Evolution under coupling

Scalar or J coupling is the through bond interaction between spins. The J coupling arises when spins polarise the bond electrons, which in turn polarise the spins of the other bonded nuclei. The result observed in a spectrum, between two coupled spins, is the splitting of peaks with half of the signal representing the coupled partners spin in a lower energy α state, and the other half representing the coupling to the higher energy β state. Such coupling is independent of the magnetic field and depends only on the nature of the nuclei, number of bonds and geometry.

Spin product theory can be used to describe coupling even where the vector model struggles. However, to use spin product formalism for coupling interactions new operators have to be introduced. The first four, $2\hat{I}_z\hat{S}_x$, $2\hat{I}_x\hat{S}_z$, $2\hat{I}_z\hat{S}_y$, $2\hat{I}_y\hat{S}_z$, all describe anti-phase magnetisation, which is only detectable during its conversion to in-phase magnetisation during the acquisition.

Compared to the in-phase Lorentzian lines observed for single spin operators ($\hat{I}_x$, $\hat{I}_y$ and $\hat{I}_z$), the anti-phase magnetisation yield opposite phase Lorentzian lines. In addition, another four operators, $2\hat{I}_y\hat{S}_x$, $2\hat{I}_x\hat{S}_y$, $2\hat{I}_y\hat{S}_y$, $2\hat{I}_x\hat{S}_x$, describe multiple quantum (MQ) coherences, which represent invisible NMR states. They can be observed indirectly and evolve under chemical shift but not under mutual *J* coupling of the spins involved in this coherence.

Coupled spins evolve under the influence of coupling once the magnetisation is transferred to the transverse plane. Thus to describe the transformation in spin product operators the Hamiltonian for coupling evolution ($\hat{H}_J$) applied to two coupled spins *I* and *S* is:

$$\hat{H}_J = 2\pi J_{IS}\hat{I}_z\hat{S}_z, \qquad [9]$$

where J_{IS} is the coupling constant in Hz.

The states α and β of the coupled spin will precess at $\Omega \pm \pi J$. The end result follows the same format as the chemical shift evolution transformation (see Section 1.7.1.2), but depends on whether the initial state of magnetisation was in-phase or anti-phase. For example, the in-phase $\hat{I}_x$ operator will transform under $\hat{H}_J$ coupling as follows:

$$\hat{\boldsymbol{I}}_{\boldsymbol{x}} \xrightarrow{2\pi J_{IS} t \hat{I}_z \hat{S}_z} \hat{\boldsymbol{I}}_{\boldsymbol{x}} \cos \pi J_{IS} t + \boldsymbol{2\hat{I}_y\hat{S}_z} \sin \pi J_{IS} t$$

While the anti-phase state $2\hat{I}_x\hat{S}_z$ will transform as follows:

$$\boldsymbol{2\hat{I}_x\hat{S}_z} \xrightarrow{2\pi J_{IS} t \hat{I}_z \hat{S}_z} \boldsymbol{2\hat{I}_x\hat{S}_z} \cos \pi J_{IS} t + \hat{\boldsymbol{I}}_{\boldsymbol{y}} \sin \pi J_{IS} t$$

The following representation can be used to visualise coupling transformations (Figure 1.31).

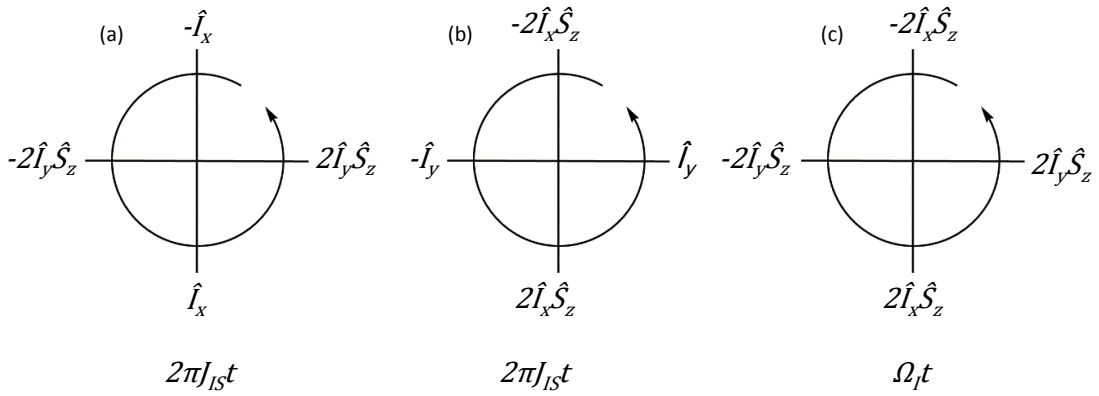


Figure 1.31 Summary of product spin operator transformations of spin *I* in a two-spin *IS* system by (a) and (b) scalar coupling ($2\pi J_{IS}$) and (c) chemical shift evolution (Ω_I) during period *t*. Adapted from ref.[273]

1.7.2. Building blocks of NMR pulse sequences

Many pulse sequence elements have been developed over the history of NMR spectroscopy that can be used to generate a specific spin condition. This section will highlight the principles behind the pulse sequence elements used throughout this work.

1.7.2.1. The spin-echo

The most basic of elements, used often to illustrate the link between product spin operator formalism and the vector model, is the spin-echo. Using the commonly used pulse sequence symbols, this element is drawn in Figure 1.32.

Figure 1.32 Spin-echo pulse sequence element. Thin and thick bars represent 90° and 180° pulses, respectively. The delays τ are arbitrary.

Considering this pulse sequence, the vector model can be used to deduce the outcome of chemical shift evolution (Figure 1.33).

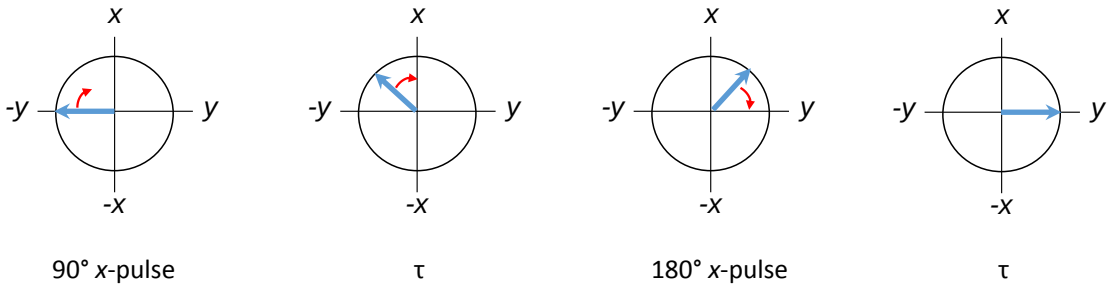


Figure 1.33 Fate of the magnetisation (blue arrow) during the spin-echo pulse sequence of Figure 1.32.Adapted from ref.[274]

The initial 90° *x*-pulse puts the spins in the transverse plane where during time τ they precess. The 180° *x*-pulse flips the spins 180° around the *x*-axis to their mirror image position causing the chemical shift to refocus. At this point, the result will differ depending on the type of coupling. If the spins are of a different kind (e.g. ^{1}H-^{13}C) then the coupling will, at the end of the 2τ period, be refocused (Figure 1.34a). However, if the spins are homonuclear (e.g. ^{1}H-^{1}H) the coupling is

not refocused because the α and β states of both spins are changed when the 180° x-pulse occurs (Figure 1.34b).[215, 216]

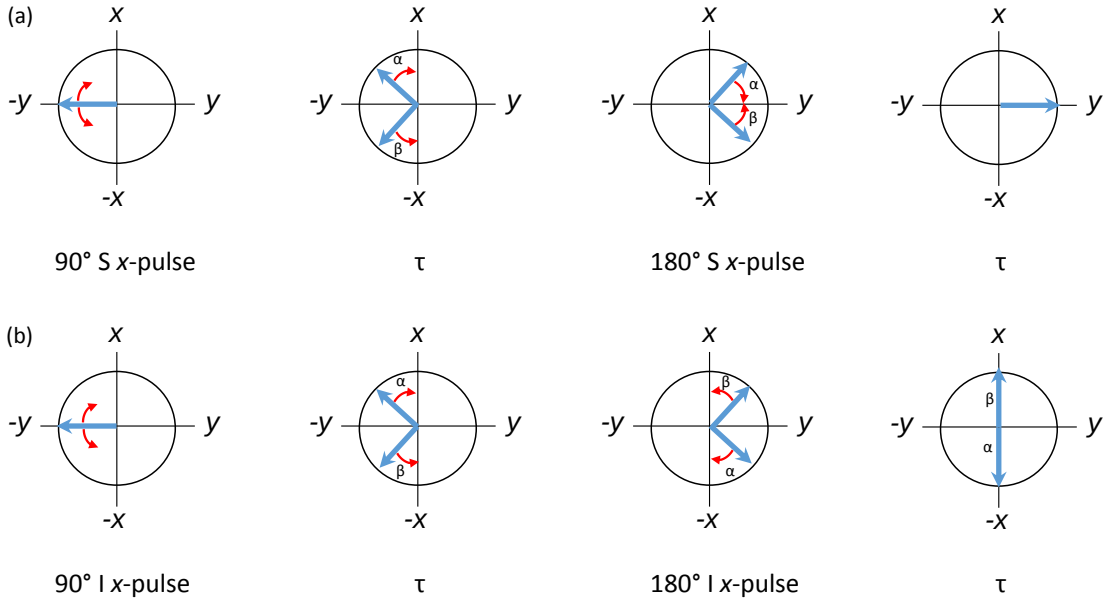


Figure 1.34 Vector model showing the outcome of (a) heteronuclear (b) homonuclear spin-echoes considering the spin-spin coupling evolution of two ½ spins. Adapted from ref.[274]

The same result can be obtained using spin product formalism, where the appropriate operator is applied to describe the action of individual pulses and time periods.

For two heteronuclear coupled spins this is formulated as:[215]

1. 90° x-pulse followed by free chemical shift precession during time τ:

$$\hat{\boldsymbol{I}}_{\boldsymbol{z}} \xrightarrow{(\pi/2)\hat{I}_x} -\hat{\boldsymbol{I}}_{\boldsymbol{y}} \xrightarrow{\Omega\tau\hat{I}_z} -\hat{\boldsymbol{I}}_{\boldsymbol{y}}\cos(\Omega\tau) + \hat{\boldsymbol{I}}_{\boldsymbol{x}}\sin(\Omega\tau)$$

2. Free J coupling precession during the first τ period:

$$\xrightarrow{2\pi J_{IS}\tau\hat{I}_z\hat{S}_z} -\hat{\boldsymbol{I}}_{\boldsymbol{y}}\cos(\Omega\tau)\cos(\pi J_{IS}\tau) + \hat{\boldsymbol{I}}_{\boldsymbol{x}}\sin(\Omega\tau)\cos(\pi J_{IS}\tau)$$
$$+ 2\hat{\boldsymbol{I}}_{\boldsymbol{x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}\cos(\Omega\tau)\sin(\pi J_{IS}\tau) + 2\hat{\boldsymbol{I}}_{\boldsymbol{y}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}\sin(\Omega\tau)\sin(\pi J_{IS}\tau)$$

3. 180° I pulse from the x direction:

$$\xrightarrow{\pi\hat{I}_x} +\hat{\boldsymbol{I}}_{\boldsymbol{y}}\cos(\Omega\tau)\cos(\pi J_{IS}\tau) + \hat{\boldsymbol{I}}_{\boldsymbol{x}}\sin(\Omega\tau)\cos(\pi J_{IS}\tau)$$
$$+ 2\hat{\boldsymbol{I}}_{\boldsymbol{x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}\cos(\Omega\tau)\sin(\pi J_{IS}\tau) - 2\hat{\boldsymbol{I}}_{\boldsymbol{y}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}\sin(\Omega\tau)\sin(\pi J_{IS}\tau)$$

4. After the free chemical shift precession during time τ:

$$\xrightarrow{\Omega\tau\hat{I}_z} -\hat{\boldsymbol{I}}_{\boldsymbol{y}}\cos(\pi J_{IS}\tau) - \mathbf{2}\hat{\boldsymbol{I}}_{\boldsymbol{x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}\sin(\pi J_{IS}\tau)$$

5. Free J coupling precession during the second τ period:

$$\xrightarrow{2\pi J_{IS}\tau\hat{I}_z\hat{S}_z} -\hat{\boldsymbol{I}}_{\boldsymbol{y}}\cos^2(\pi J_{IS}\tau) + \mathbf{2}\hat{\boldsymbol{I}}_{\boldsymbol{x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}\sin(\pi J_{IS}\tau) - \mathbf{2}\hat{\boldsymbol{I}}_{\boldsymbol{x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}\sin(\pi J_{IS}\tau)\cos(\pi J_{IS}\tau) - \hat{\boldsymbol{I}}_{\boldsymbol{y}}\sin^2(\pi J_{IS}\tau)$$

6. Using a simple trigonometric identity ($cos^2A + sin^2A = 1$) this can be simplified to:

$$\rightarrow -\hat{\boldsymbol{I}}_{\boldsymbol{y}}$$

This result shows that the magnetisation has been refocused, which is the same outcome as obtained by the vector model. In the above example of heteronuclear spin-echo, the 180° pulse was applied to I spin, while the S spin was untouched (Figure 1.35a). Of course two other cases are possible where the 180° pulse is placed on S instead of I (Figure 1.35b), or where two 180° pulses are applied to both spins (Figure 1.35c).

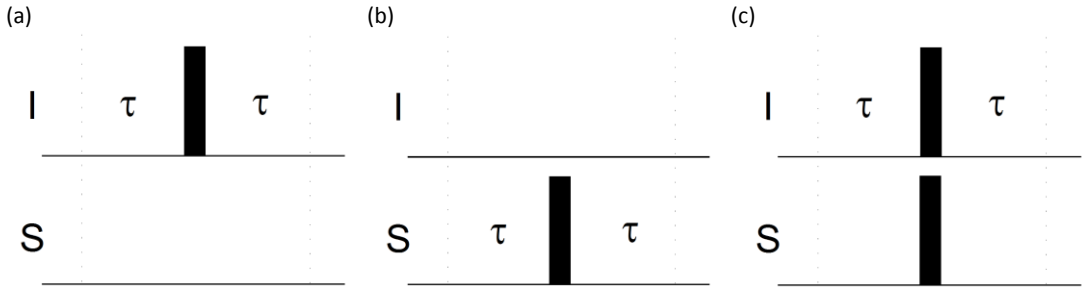


Figure 1.35 Heteronuclear spin-echoes. (a) refocuses δ_I and J_{IS}; (b) refocuses δ_S and J_{IS}; (c) refocuses δ_I and δ_S but not J_{IS}. Bars represent 180° pulses. τ is an arbitrary delay.

In a) and b) the chemical shift (δ) of the spin receiving the 180° pulse is refocused, while that of the other spin is not. In both cases, the coupling between I and S is refocused. In c) the effect of two 180° degree pulses refocuses the chemical shifts but not the coupling between I and S.

1.7.3. Insensitive Nuclei Enhanced by Polarisation Transfer (INEPT)

The spin product formalism may not be as intuitive for the spin-echo as the vector model, however it comes to the fore with more complicated pulse sequences. One pulse sequence element frequently used in this study is the INEPT sequence.[275] INEPT is a polarisation transfer experiment that allows transfer of the magnetisation from a more sensitive, high gyromagnetic ratio nucleus, to a less sensitive, low gyromagnetic ratio nucleus, e.g. proton to carbon. This sequence is crucial in many NMR experiments, not only to increase their sensitivity, but also for correlating spins. The pulse sequence of INEPT is shown in Figure 1.36.

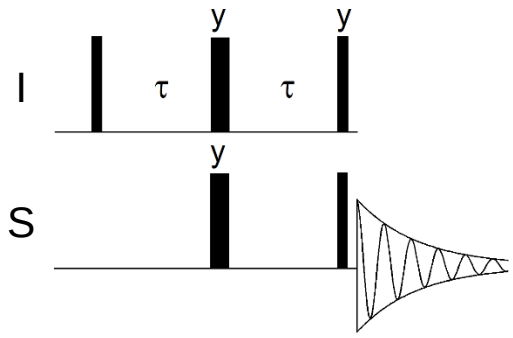


Figure 1.36 INEPT pulse sequence. Thin and thick bars represent 90° and 180° pulses, respectively. τ is set to $1/4J_{IS}$. Unless stated otherwise the pulses were applied from the *x*-direction.

This sequence can be described as a heteronuclear spin-echo (Figure 1.35c) with the addition of two 90° pulses on both channels. Returning to the beginning, the first 90° *x*-pulse creates transverse magnetisation of ^{1}H (*I*) spins. This is followed by a spin-echo with two 180° pulses applied to both ^{13}C (*S*) and ^{1}H (*I*) spins. As shown in Section 1.7.2., simultaneous application of these pulses has the effect of refocusing ^{1}H (*I*) chemical shifts but allowing the heteronuclear coupling to evolve. Setting the τ period to $1/4J$ results in the ^{1}H (*I*) spins adopting pure anti-phase states. The last two 90° pulses transfer the magnetisation from the ^{1}H (*I*) spins to the ^{13}C (*S*) spins. Thus, in a typical ^{1}H coupled ^{13}C INEPT spectrum one would see magnified anti-phase multiplets for each CH_n group.

In terms of product spin operators the INEPT sequence can be illustrated by the following steps:[215]

1. 90° *x*-pulse on spin *I* followed by a spin-echo. Only the coupling needs to be taken into account as the spin-echo will refocus the chemical shift:

$$\boldsymbol{a\hat{I}_z + b\hat{S}_z} \xrightarrow{(\pi/2)\hat{I}_x} \boldsymbol{-a\hat{I}_y + b\hat{S}_z} \xrightarrow{2\pi J_{IS}\tau \hat{I}_z \hat{S}_z} \boldsymbol{a}(-\boldsymbol{\hat{I}_y}\cos\pi J_{IS}\tau + \boldsymbol{2\hat{I}_x\hat{S}_z}\sin\pi J_{IS}\tau) + \boldsymbol{b\hat{S}_z}$$

where *a* represents the ratio of gyromagnetic constants of spins *I* and *S*. In the case of ^{1}H and ^{13}C, $a = {}^{\gamma_H}/_{\gamma_C}\, b = 4b$.

2. Two 180° pulses on *I* and *S* can be treated sequentially:

$$\xrightarrow{\pi\hat{I}_y} \boldsymbol{a}(-\boldsymbol{\hat{I}_y}\cos\pi J_{IS}\tau - \boldsymbol{2\hat{I}_x\hat{S}_z}\sin\pi J_{IS}\tau) + \boldsymbol{b\hat{S}_z}$$

$$\xrightarrow{\pi\hat{S}_y} \boldsymbol{a}(-\boldsymbol{\hat{I}_y}\cos\pi J_{IS}\tau + \boldsymbol{2\hat{I}_x\hat{S}_z}\sin\pi J_{IS}\tau) - \boldsymbol{b\hat{S}_z}$$

3. *J* coupling evolution follows and the result can be simplified using a trigonometric identify:

$$\xrightarrow{2\pi J_{IS}\tau \hat{I}_z \hat{S}_z} \boldsymbol{a}(-\hat{\boldsymbol{I}}_{\boldsymbol{y}} \cos^2 \pi J_{IS}\tau + \boldsymbol{2\hat{I}_x\hat{S}_z} \cos \pi J_{IS}\tau \sin \pi J_{IS}\tau + \boldsymbol{2\hat{I}_x\hat{S}_z} \sin \pi J_{IS}\tau \cos \pi J_{IS}\tau + \hat{\boldsymbol{I}}_{\boldsymbol{y}} \sin^2 \pi J_{IS}\tau) - \hat{\boldsymbol{S}}_{\boldsymbol{z}}$$

$$\equiv -\boldsymbol{a}(\hat{\boldsymbol{I}}_{\boldsymbol{y}} \cos 2\pi J_{IS}\tau - \boldsymbol{2\hat{I}_x\hat{S}_z} \sin 2\pi J_{IS}\tau) - \boldsymbol{b\hat{S}_z}$$

4. The expression is further reduced by substituting $1/4J$ for τ:

$$\equiv \boldsymbol{2a\hat{I}_x\hat{S}_z - b\hat{S}_z}$$

5. The final step is a 90° y-pulse on I and x-pulse on S:

$$\xrightarrow{(\pi/2)\hat{I}_y} \boldsymbol{-2a\hat{I}_z\hat{S}_z - b\hat{S}_z} \xrightarrow{(\pi/2)\hat{S}_x} \boldsymbol{2a\hat{I}_z\hat{S}_y + b\hat{S}_y}$$

The last two terms represent polarised and natural I magnetisation of spin S. If I is ^{1}H and S is ^{13}C, the first term denotes the anti-phase carbon magnetisation, which will be 4 times that represented by the in-phase magnetisation term, S_y. When these two terms are added the result for a CH group is an anti-phase doublet of increased intensity, as shown in Figure 1.37.

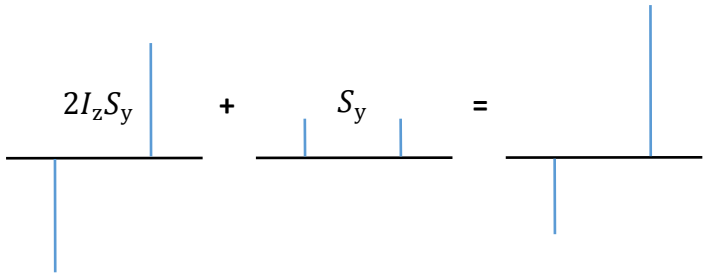


Figure 1.37 The signals of the *IS* spin system generated by the INEPT pulse sequence shown in Figure 1.36.

In practice the natural carbon magnetisation is removed by phase cycling or PFGs. As can be seen in the above example, product spin operators can clearly demonstrate the evolution of spins in more complex NMR experiments. This treatment will be used to describe the pulse sequences developed in this work. Other pulse sequence elements used in this study are attended to next.

1.7.4. Constant-time chemical shift labelling

One of the consequences of increasing the dimensionality of NMR experiments is the increase in the overall length of the pulse sequence, which in turn can lead to relaxation induced signal losses. One way of reducing these losses is to combine several evolution events into one interval as done, e.g. by the so-called constant-time chemical shift labelling intervals, a common feature of biomolecular pulse sequences. A constant-time interval can be used for chemical shift labelling of one resonance type, while concurrently creating the anti-phase state required for the ensuing

polarisation transfer and potentially refocusing the couplings used in the previous polarisation transfer step. An example of a constant-time period is shown in Figure 1.38.

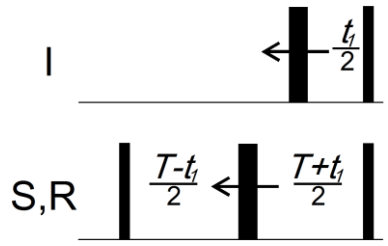


Figure 1.38 An example of chemical shift labelling of spin *S* during a constant-time period in a *ISR* spin system. T = $1/2J_{RS}$. All pulses are applied from *x*.

The effect of this pulse sequence on spin *S*, starting from its anti-phase state with respect to spin *R* and in-phase with respect to spin *I*, can be described by product spin operators as follows:

$$-4\hat{\boldsymbol{I}}_z\hat{\boldsymbol{S}}_x\hat{\boldsymbol{R}}_y \xrightarrow{(\pi/2)\hat{S}_x,\hat{R}_x} -4\hat{\boldsymbol{I}}_z\hat{\boldsymbol{S}}_x\hat{\boldsymbol{R}}_z \xrightarrow{2\pi J_{RS}T} -2\hat{\boldsymbol{I}}_z\hat{\boldsymbol{S}}_y \xrightarrow{\pi\hat{I}_x;\ \pi\hat{S}_x} -2\hat{\boldsymbol{I}}_z\hat{\boldsymbol{S}}_y$$

$$\xrightarrow{\Omega_S t_1\hat{S}_z} -2\hat{\boldsymbol{I}}_z\hat{\boldsymbol{S}}_y\cos\Omega_S t_1 + 2\hat{\boldsymbol{I}}_z\hat{\boldsymbol{S}}_x\sin\Omega t_1 \xrightarrow{(\pi/2)\hat{I}_x;(\pi/2)\hat{S}_x} 2\hat{\boldsymbol{I}}_y\hat{\boldsymbol{S}}_z\cos\Omega_S t_1$$

The segment has effectively two spin-echoes. By moving the pulses as t_1 increases, the overall T period is held constant. Thus, the J coupling evolution is avoided while at the same time t_1 is incremented for chemical shift labelling of *S*, in this instance. Note that the situation described above matches the constant-time segment used after the DQ labelling period in the INEPT-INADEQUATE-HSQC pulse sequence shown in Figure 4.3.

1.7.5. Selective and broadband radiofrequency pulses

Short 90° or 180° degree pulses will excite or invert spins over a large frequency range around the selected carrier frequency. The precession frequency (ω, $rads^{-1}$) of a pulse of flip angle β is determined by the pulse length (τ_p):

$$\omega = \frac{\beta}{\tau_p} \qquad [10]$$

In order to have all signals in a given spectrum excited, the maximum frequency offset has to be lower than the precession frequency (or RF field strength) of the pulse. In this way, all spins effectively feel on-resonance excitation. If the offset is comparable to the RF field applied then the signals further from the carrier frequency will feel less than the desired flip angle leading to so called off-resonance effects. The signals may still appear in the spectrum due to the fact that the

excitation profile of the pulse is to the first approximation the Fourier transform of the shape. The Fourier transform of a rectangular pulse is a sinc function where the *x*-axis is frequency in Hz. This means that signals away from the maximum excitation will feel an attenuated excitation or even no excitation if they happen to fall in a null point of the sinc function. Being able to change the excitation window leads to selective excitation/inversion. In this way increasing the pulse length narrows the range over which the pulse is effective, meaning we can choose to excite/invert certain resonances in a given spectrum. Of course this could be done with a rectangular pulse of low power, however the sinc 'wiggles', that are inherent to such pulses, lead to undesirable excitation. These 'wiggles' can be removed if low power pulses are shaped. Different shape pulses, also known as selective or 'soft' pulses, are available. Depending on the way the RF power is modulated or shaped, the excitation/inversion profile will have a more or less 'top hat' appearance. The parameters of each selective pulse in terms of duration, maximum power and offset need to be set specifically in order to obtain the desired inversion/excitation profile. The most popular is the Gaussian shape (Figure 1.39a) which under FT appears as Gaussian with no 'wiggles' (Figure 1.39c). Another example is the r-SNOB pulse used in a number of the proposed experiments. Its shape and FT profile are depicted in Figure 1.39b and Figure 1.39d, respectively.

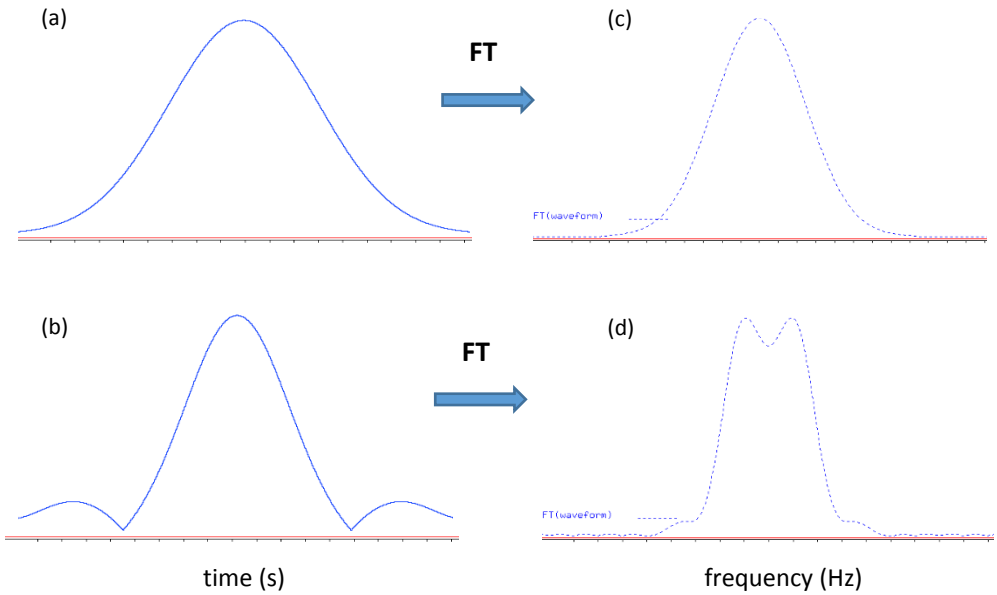


Figure 1.39 (a) Gaussian and (b) r-SNOB shape pulses with their excitation profiles (Fourier transform (FT)) shown in (c) and (d), respectively.

Going in the opposite direction, a simple rectangular 180° pulse does not cover the frequency range of ^{13}C resonances on > 600 MHz NMR instruments. This issue can be addressed by using

broadband adiabatic, or phase and amplitude modulated, pulses. The adiabatic condition states that in situations where the B_1 field of the pulse is not aligned on resonance the field felt by the nuclei is the effective field B_{eff}. If this B_{eff} is strong enough, a slow rate sweep of the B_1 field will be followed by B_{eff}.[216] In other words different spins will be affected at different points in this adiabatic passage. Excitation or inversion is then achieved without using high power, making it a better choice for carbon inversion or decoupling. In addition, these pulses have better resilience to the inhomogeneity of the B_1 field.

1.7.6. Pulsed field gradients (PFGs)

An inherent part of NMR experiments is the selection of some signals and destruction of others. There are two ways to achieve this: (i) phase cycling and (ii) PFGs. Phase cycling requires repeating the same pulse sequence while changing phases of specific pulses, typically one at the time. This increases the overall experimental time but also leaves room for cancellation artefacts as the undesired signals are subtracted in subsequent scans. Any short term instability of the spectrometer will therefore result in imperfect signal cancellation and the appearance of artefacts in the spectra, e.g. increased t_1 noise in 2D experiments.

PFGs operate within one scan by destroying unwanted signal outright. This has the advantage of a reduced number of scans and removal of cancellation artefacts.[276] Another application of PFGs is in quadrature detection of chemical shifts in the indirectly detected dimensions of nD experiments often with increased sensitivity.

PFGs generate a magnetic field gradient along one of the sample tube axes, usually z, causing molecules at different positions in the NMR tube to experience a different applied field. During the short duration of PFGs (typically 0.5-2 ms) magnetisation is dephased. After the gradient has been applied for time Δ the phase at any position in the sample, ϕ_z is given by:

$$\phi_z = \gamma(B_o + G_z)\Delta \qquad [11]$$

As B_o is constant across the sample, it can be ignored, while the remaining $\gamma G_z \Delta$ is the spatially dependent phase induced by the gradient. Using product spin operators the effect of a gradient on spin I in the transverse plane can be expressed as:

$$\hat{\boldsymbol{I}}_{\boldsymbol{x}} \xrightarrow{\gamma G_z \Delta I_z} \hat{\boldsymbol{I}}_{\boldsymbol{x}} \cos(\gamma G_z \Delta) + \hat{\boldsymbol{I}}_{\boldsymbol{y}} \sin(\gamma G_z \Delta)$$

The net result is elimination of signal as the spins precessing with identical Larmor frequencies will be placed in different orientations due to the accrued phase. This can be reversed by applying another PFG with opposite polarity.

It follows on that using PFGs allows selective dephasing of signals when combined with certain pulse sequence elements thus providing the experimentalist with a powerful tool for spin manipulation. This is illustrated here on an example of a single pulsed-field gradient spin-echo (SPFGSE) shown in Figure 1.40a. Using product spin operators the action of this pulse sequence element can be described (ignoring any J or Ω evolution) as follows:

$$\hat{\boldsymbol{I}}_{\boldsymbol{z}} \xrightarrow{\pi/2\hat{I}_x} -\hat{\boldsymbol{I}}_{\boldsymbol{y}} \xrightarrow{\gamma G_1 \Delta \hat{I}_z} -\hat{\boldsymbol{I}}_{\boldsymbol{y}} \cos(\gamma G_1 \Delta) + \hat{\boldsymbol{I}}_{\boldsymbol{x}} \sin(\gamma G_1 \Delta) \xrightarrow{\pi \hat{I}_x} \hat{\boldsymbol{I}}_{\boldsymbol{y}} \cos(\gamma G_1 \Delta)$$
$$+ \hat{\boldsymbol{I}}_{\boldsymbol{x}} \sin(\gamma G_1 \Delta) \xrightarrow{\gamma G_1 \Delta \hat{I}_z} \hat{\boldsymbol{I}}_{\boldsymbol{y}} \cos^2(\gamma G_1 \Delta) + \hat{\boldsymbol{I}}_{\boldsymbol{y}} \sin^2(\gamma G_1 \Delta) \equiv \hat{\boldsymbol{I}}_{\boldsymbol{y}}$$

The product spin operators show that the part of the magnetisation that received a perfect 180° pulse is restored.

The part of the magnetisation that did not receive a 180° pulse continues to be dephased by the second PFG and acquires a gradient induced spatially dependent phase, ϕ_z and thus the magnetisation is not detected. The gradient echo thus removes the necessity to phase-cycle the 180° pulse while eliminating signals that could potentially be the source of artefacts.

This gradient spin-echo also forms a platform for selective excitation of spins. If the non-selective 180° pulse (Figure 1.40a) is replaced by a frequency selective pulse, only the magnetisation of the spins within the inversion window of that pulse will be refocused in a so called selective single pulse-field gradient spin-echo (selective SPFGSE).

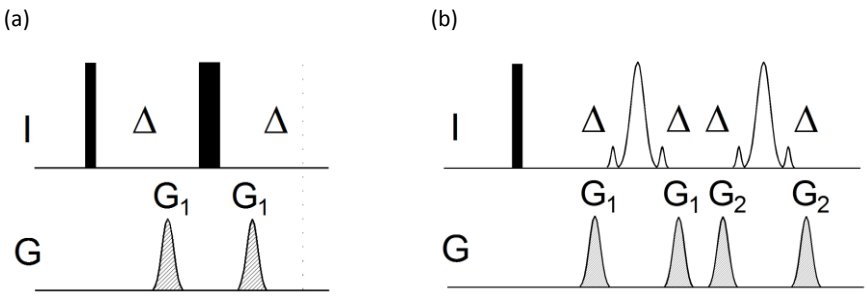


Figure 1.40 (a) SPFGSE and (b) selective DPFGSE where G_1 and G_2 denote gradient strengths. Δ is an arbitrary delay typically the sum of the gradient duration and the recovery time. Thin and thick bars represent 90° and 180° pulses respectively. Sequence (b) contains selective 180° pulses.

However a much more robust selective excitation element is the selective DPFGSE (Figure 1.40b).[277] This segment reinforces the action of the SPFGSE in that only the spins that experience selective inversion twice escape dephasing by the gradients. Additional advantages of the DPFGSE is that a non-uniform phase induced by the first selective 180° pulse is eliminated by the second. Selective excitation by a DPFGSE is therefore ideally suited to be used at the beginning of a t_1 period, as it generates pure phase of signals across the whole excited frequency range.

1.8. Aims and objectives

Knowledge of the molecular composition of HS is an absolute requirement for building the structure-functional relationships of HS molecules. Such information will enable the origin of SOM molecules to be traced, improve modelling of plant-soil interactions and explain at the molecular level the interactions of soil with heavy metals and radionuclides. It will also underpin explanation of the role of soil in nutrient mediation, and enable evidence-based applications of HS in crop production. The understanding of the decomposition of peat SOM and the effects of a changing climate will be advanced through molecular level investigation of these processes.

Due to the complexity of HS, a combination of chromatography and standard NMR and MS techniques is inadequate to the problem at hand. To unravel the compounds contained in this intractable mixture, 'spectroscopic separation' and subsequent structural analysis is required. The former can be achieved by chemical modification of HS with NMR active tags, while the latter requires specifically designed nD NMR experiments. As COOH and OH groups are fundamental to the function of HS this investigation focuses on molecules carrying these groups, specifically the phenolic compounds that are seen to play a role in preserving peat SOM.

Thus the aim of this study is to design a methodology, which combines ^{13}C methylation with isotope-filtered nD NMR spectroscopy, and apply it to a HS sample to obtain structural information about the phenolic molecules.

The objectives to achieve this aim are to:

- extract HA, FA and DOM samples from peat samples.
- investigate an effective protocol for ^{13}C methylation of a phenolic molecules, prepare methylated model mixtures and a methylated HS sample.
- develop new NMR experiments that exploit polarisation transfer pathways involving the $^{13}CH_3$ tags.
- test the methodology on methylated model mixtures and acquire a complete set of experiments on a methylated HS sample.
- analyse nD NMR spectra and propose structures of phenolic moieties/molecules of the studied HS sample.
- compare the identified structures with those of known metabolites and degradation products of plants present on the sampling site.

Chapter 2. Materials and Methods

Red Moss of Balerno raised peat bog, Midlothian, Scotland

Image from www.currie-scc.gov.uk

Acknowledgments of work described in this section

Extraction and processing of the Red Moss of Balerno peat FA sample was conducted by a summer student Adam Michalchuk, School of Chemistry, University of Edinburgh.

Methylation reactions were conducted by an undergraduate group (Victoria Camus, Alison Dickson, Asha Hoque), Dr Lorna Murray and John Blackburn, School of Chemistry, University of Edinburgh.

2.1. Samples used in this study

This study used a series of HS samples and two model mixtures, referred to as model mixture **I** and model mixture **II**. The source and composition of the samples will be outlined in this section.

2.1.1. HS samples

The HS samples used in this study were i) Suwannee River Fulvic Acid (SRFA) and Humic Acid (SRHA), which were purchased from the International Humic Substances Society (IHSS); ii) FA and HA extracted from peat collected from Needle's Eye natural analogue site, Dumfries and Galloway, Scotland (NEFA and NEHA); iii) HA and FA extracted from peat collected from Red Moss of Balerno reserve, Midlothian, Scotland (RMFA and RMHA); iv) DOM extracted from peat pore water collected at Needle's Eye natural analogue site, Dumfries and Galloway, Scotland (NEDOM). Samples ii) and iii) were prepared following a modified IHSS method outlined by Thurman and Malcolm.[278] Sample iv) was processed using tangential flow ultrafiltration (TFF).

2.1.2. Model mixture I

As explained in Section 1.4, this study aims to identify phenolic molecules in HS samples. A mixture of four compounds (Figure 2.1) was therefore used to test the methylation protocols and design the NMR experiments. The unmethylated molecules in model mixture **I** have been assigned letters **A'** to **D'** while the methylated molecules are referred to as **A** to **D**.

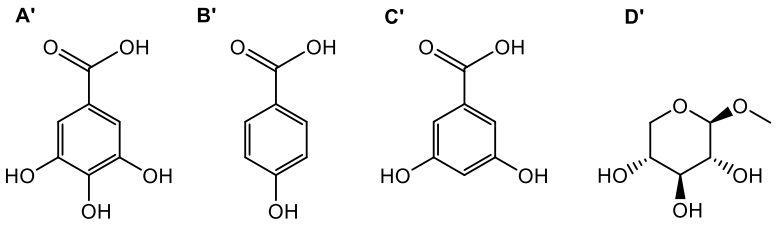


Figure 2.1 Model mixture **I** (**A'**) 3, 4, 5-trihydroxybenzoic acid; (**B'**) 4-hydroxybenzoic acid; (**C'**) 3, 5-dihydroxybenzoic acid; (**D'**) Methyl β-D-xylopyranoside.

In order to tailor the NMR experiments to phenols, model mixture **I** contains organic aromatic acids, namely 3, 4, 5-trihydroxybenzoic acid, 4-hydroxybenzoic acid, 3, 5-dihydroxybenzoic acid and a monosaccharide methyl β-D-xylopyranoside. The phenolic acids were chosen as they contain the most prevalent functional groups in HS samples; carboxyl and hydroxyl groups. Compound **A'** has been proposed to be part of HA from GC-MS[140] and NMR[196] studies, while

B' has been identified in GC-MS studies of HA.[140] Compound **C'** was chosen to complement **A'** and **B'** by having two hydroxyl groups, whilst the carbohydrate **D'** was used to test whether non-aromatic molecules interfere with the NMR methods.

2.1.3. Model mixture II

A more complex mixture of 9 compounds (Figure 2.2) was also prepared.

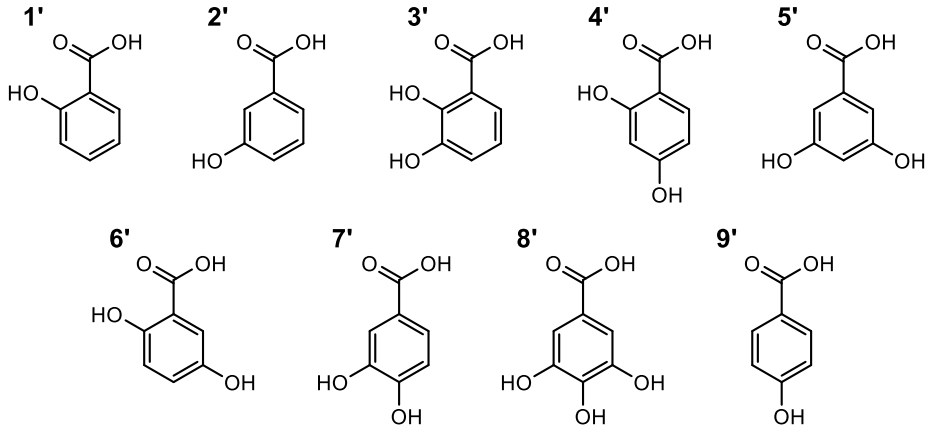


Figure 2.2 Model mixture **II** (**1'**) 2-hydroxybenzoic acid; (**2'**) 3-hydroxybenzoic acid; (**3'**) 2, 3-dihydroxybenzoic acid; (**4'**) 2, 4-dihydroxybenzoic acid; (**5'**) 3, 5-dihydroxybenzoic acid; (**6'**) 2, 5-dihydroxybenzoic acid; (**7'**) 3, 4-dihydroxybenzoic acid; (**8'**) 3, 4, 5-trihydroxybenzoic acid; (**9'**) 4-hydroxybenzoic acid.

Again this mixture contained a number of phenolic acids. The unmethylated molecules are labelled **1'**-**9'**, while the methylated molecules are labelled **1**-**9**. Compounds **2'**, **7'**, **8'** and **9'** have been proposed to be part of HA from GC-MS studies,[140] while the others were chosen to have a variety of substitution patterns. Methylated model mixture **II** was used to test (i) the efficiency of methylation when applied to a more diverse set of structures and (ii) compare the performance of the developed NMR experiments.

2.2. Materials required for isolation and purification of HS

The following chemicals were used for the extraction of HA and FA from peat: HCl (37%, analytical grade, VWR International), NaOH (≥ 98%, analytical grade, Fisher Scientific). Purification of the HA samples required a Visking Dialysis membrane (12-14000 Da, Medicell International), KCl (≥ 99%, analytical grade, Sigma Aldrich), KOH (≥ 85%, analytical grade, Sigma Aldrich) and $AgNO_3$ (≥ 99%, analytical grade, Sigma Aldrich).

Purification of RMFA required a Superlite™ DAX-8™ resin (Supelco, Sigma Aldrich), a DOWEX™ 50WX8-200 resin (Supelco, Sigma Aldrich), as well as acetonitrile (> 99.99%, HiPerSolv(R), VWR International), methanol (> 99.9%, analytical grade, VWR International), HCl (37%, analytical grade, VWR International), NaOH (≥ 99%, analytical grade, Sigma Aldrich) and Milli-Q(R) water (produced in house by a Millipore filtration system (EMD Millipore Corporation, Billerica, MA, USA)). Purification of the Needle's Eye DOM sample required 8 micron (Whatman™, 540, 1540090) and 2.7 micron (Whatman™, 542, 1542110) hardened ashless filter paper and 0.45 micron bottle-top vacuum filter systems (Corning®, Sigma Aldrich). Concentration of the NEDOM sample required Milli-Q(R) water (produced in house by a Millipore filtration system (EMD Millipore Corporation, Billerica, MA, USA)), and Masterflex™ Norprene tubing to withstand flow pressures up to 60 psi.

2.3. Materials required for the methylation of the model mixtures and HS

Chemicals required for the methylation reactions were methyl iodide (^{12}C and ^{13}C) (> 99%, CK Gas Products), sodium hydride (NaH) (60% (w/w) in mineral oil, Sigma Aldrich), anhydrous dimethylformamide (DMF) (> 99.9%, SeccoSolv(R), Merck), tetrahydrofuran (THF) (> 99.9%, analytical grade, VWR International), HCl (37%, analytical grade, VWR International), chloroform (> 99.9%, analytical grade, Sigma Aldrich) and TBAH (40% in water, Sigma Aldrich), tri-sodium citrate (> 99.9%, analytical grade, Sigma Aldrich) and hexane (> 95%, HiPerSolv(R). VWR International).

2.4. Isolation of HA collected from Needle's Eye peat (NEHA)

A sample of peat (3 kg) was collected from Needle's Eye natural analogue site, on the Solway Coast of Dumfries and Galloway, Scotland (Figure 2.3), 54°52'N, 3°45'W. This site is close to a mineralised uranium vein exposed at the outcrop of the cliff in form of a Criffel granite complex.[279] Uranium from the vein is dissolved and transported in surface water from the cliff towards Southwick Water (see Figure 2.3).

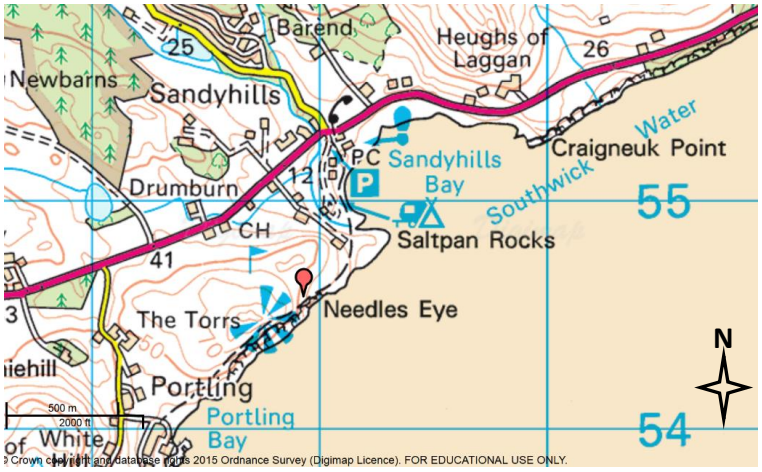


Figure 2.3 Location of the sampling site at Needle's Eye, Dumfries and Galloway highlighted by a red marker. Map produced using EDINA Digimap.

An area of 1 m deep anoxic peat runs 30 cm from the cliff edge. This peat has high organic carbon content, ranging from 35 to 90%[280] and it is reported that 80-90% of the dissolved uranium is complexed by this organic matter,[281] preventing transport to the Southwick Water. The organic matter acts as a long-term sink for uranium, however there is concern that climatic conditions could dry out the peat and release the uranium in the form of colloids.[280] The cliff area and peat are conserved by the Scottish Wildlife Trust. As this site is in a rural location away from anthropogenic organic matter sources, this site is of considerable importance for metal binding studies particularly far field radionuclide transport and retardation. The understanding of these interactions with HS contained within the organic matter is key to understanding this process. Thus, permission was obtained from the Scottish Wildlife Trust to collect a sample of peat from this site for molecular analysis.

HA were isolated from the peat sample following a modified version of the IHSS certified procedure outlined by Thurman and Malcolm.[278] The main modification was that the absence of a N_2 atmosphere in order to make the procedure simpler and faster. Justification for this choice can be found in Tan *et al.*[10, 282] who reported that omission of the N_2 atmosphere does not result in oxidation as long as the procedure is carried out within a few hours. In brief, the peat sample (3 kg wet weight) was processed by initially removing any visible plant material and stones before passing it through a 2 mm metal sieve. The sieved peat was separated into twenty-eight 80 g portions. Each proportion was acidified using HCl (400 ml, 1 M) and shaken for one hour. The samples were left to settle, upon which two layers formed consisting of a dark solid residue and a pale brown solution. The supernatant was separated by centrifugation (8873 x g, 5 min) of

layered mixture and was designated as FA extract 1. The crumbly dark residue was neutralised with NaOH (250 ml, 1 M), and then the volume topped up with less concentrated NaOH (0.1 M), to meet the recommended 10:1 (v/w) ratio of solution: peat set by the IHSS. The neutralised samples were shaken for 3 hours before centrifugation (8873 x g, 5 min). The resulting dark brown supernatant was isolated and the dark brown pellet was discarded. To separate FA and HA, the supernatant was acidifed with a minimum amount of HCl (6 M) required to bring the pH to 1 or 2, shaken and left for several hours. The resulting dark brown insoluble component represented HA, while the yellow solution was designated as FA extract 2. These fractions were isolated by centrifugation (8873 x g, 5 min) and stored for further processing.

2.5. Isolation of HA and FA from Red Moss peat (RMHA and RMFA)

Peat samples were extracted from a lowland raised bog, Red Moss, located at Balerno, a suburb of Edinburgh (See Figure 2.4), coordinates 55°51'N, 3°20'W.[283]

Figure 2.4 Red Moss of Balerno raised peat bog Site, 55°51'N, 3°20'W. Sampling location highlighted by a red marker. Map produced using EDINA Digimap.

The site is designated as a Site of Specific Scientific Interest as it is one of twenty remaining raised bogs in Edinburgh and the Lothians,[284] and has been conserved as a nature reserve by the Scottish Wildlife Trust since 1986.

The bog is a habitat for various plants (Table 2.1) and animal species, such as the dragonfly, adder and lizard.[285] Under the living vegetation, the bog itself is composed of an initial layer (within the first 10 cm) of decaying vegetation followed by well decomposed peat, which can be found at depths of 6 metres in places,[286] that has accumulated over thousands of years since the end of the last ice age. Concerns have been raised regarding the drying of raised bogs and the resulting

effect on carbon storage and biodiversity (Section 1.4.2).[284] Preservation of this and other raised bogs would benefit from a better understanding of the molecular composition particularly that of phenols as outlined in Section 1.4.3.

Table 2.1. Plant species common to Red Moss of Balerno raised peat bog.[287]

Latin name	Common name	Classification
Calluna vulgaris	Heather	angiosperm
Hypnum jutlandicum	Heath plate moss	bryophyte
Sphagnum papillosum	Sphagnum moss	bryophyte
Sphagnum cuspidatum	Feathery bog moss	bryophyte
Sphagnum capillifolium	Red peat moss	bryophyte
Sphagnum magellanicum	Magellan's peat moss	bryophyte
Sphagnum recurvum	Flat-topped bog moss	bryophyte
Polytrichum commune	Haircap moss	bryophyte
Hypnum cupressiforme	Cypress-leaved plait moss	bryophyte
Pleurozium schreberi	Feather moss	bryophyte
Aulacomnium palustre	Ribbed bog moss	bryophyte
Eriophorum angustifolium	Cotton grass	angiosperm
Eriophorum vaginatum	Tussock cotton grass	angiosperm
Empetrum nigrum	Crowberry	angiosperm
Vaccinium myrtillus	Blueberry	angiosperm
Cladonia arbuscula	Lichen	fungi
Drosera rotundifolia	Common sundew	angiosperm
Erica tetralix	Cross-leaved heath	angiosperm

Samples of peat were taken at 3 different depths, 0-10 cm (layer 1, 1.4 kg), 10-20 cm (layer 2, 3 kg) and 20-30 cm (layer 3, 3.1 kg). Another sample of peat (3 kg) was taken from below the initial layer of vegetation. From previous elemental analysis, these sections have similar carbon content (~40% relative to sectional 105 °C dried peat weights), although the deeper samples contain more highly decomposed peat.[286] The same procedure as outlined in Section 2.4 was used to extract both HA and FA from each sample.

2.6. Purification of RMHA and NEHA

After the acidification step (Section 2.4) both the RMHA and NEHA samples still contained inorganic impurities. In order to remove these a procedure for purification was carried out. Note this was only carried out on the second sample of RMHA and the NEHA sample. The purification procedure was as follows.

The HA residue was redissolved in a minimal amount of KOH (0.1 M), before solid KCl was added. The mixture was then centrifuged (8873 x g, 5 min) to separate out solid impurities. The HA solution was precipitated using a minimal amount of HCl (6 M). To remove Na^+ cations and obtain an H^+-saturated HA sample, the solid residue after acidification was dialysed against Milli-Q(R) until the water gave a negative Cl^- reading in an $AgNO_3$ test. The H^+-saturated slurry was then freeze-dried, sealed and refrigerated.

2.7. Purification and concentration of RMFA

In order to isolate RMFA from any inorganics the IHSS protocol contains a two stage purification/concentration procedure. The first stage involves affinity chromatography using a DAX™-8 resin, while the second stage uses DOWEX™ 50WX8 cation exchange to convert the FA sample into an H^+ saturated form**.** In this section, a brief description of each stage is given, including the details of the preparation, sample use and regeneration of each resin.

2.7.1. Monitoring column capacity using UV-vis absorbance

The IHSS procedure for isolation of FA requires the use of affinity chromatography to concentrate and purify the sample. An estimate of the amount of Dissolved Organic Carbon (DOC) in the eluent during this step is required to avoid overloading the column (column breakthrough). In order to monitor DOC, the UV-vis absorbance at 254 nm of the column eluent was measured. This utilises the linear relationship between the DOC concentration and UV-vis absorbance of the aromatics[288] contained within HS samples.

To justify this method, the organic content of the rinse of the conditioned column and RMFA sample were measured as DOC using a Shimadzu TOC-V analyser. This analyser measures organic carbon via a combustion technique. The sample is split in the analyser, where half of it is mixed with phosphoric acid to convert the inorganic carbon into CO_2. The other half is combusted

at 680 °C with the aid of a platinum catalyst. The resulting CO_2 from both sections is cooled, dehumidified and analysed by a non-dispersive IR detector. The DOC is then calculated as total carbon minus the inorganic carbon.

RMFA sample layer 1 was diluted using Milli-Q(R) water 5, 10, 15 and 20 times. DOC was measured using a 10 ml sample of each dilution. The UV-vis absorbance was then measured for diluted calibrant sample at 254 nm, using a Lightwave II UV-vis spectophotometer, Biochrom with Milli-Q(R) water as a reference. Only the FA samples diluted 5 and 10 times gave UV-vis absorbance (the other samples were too dilute). The DOC of the FA dilutions and their mean UV-vis absorbance of 3 replicates are given in Table 2.2.

Table 2.2 DOC and associated UV-vis absorbance at 254 nm for series of diluted RMFA samples

Dilution factor	DOC (mg l^{-1})	Mean UV-vis absorbance at 254 nm of 3 replicates
0	9.005	0.377
5	1.801	0.083
10	0.930	0.038

The TOC-analyser system blank had a DOC measurement of 0.045 mg l^{-1}, while the rinse from the conditioned column gave a DOC measurement of 0.415 mg l^{-1} and UV-vis absorbance of 0.007. A calibration curve was not drawn as only 3 points were obtained. Nevertheless, from the table, there is an increase in absorbance vs DOC confirming the aforementioned relationship. Thus, taking into account that the DOC of the rinse is much lower than the undiluted RMFA sample oversaturation of the column can be clearly judged based on a step change in UV-vis absorbance of the eluent.

2.7.2. DAX™-8 affinity chromatography

To purify and concentrate the RMFA samples a DAX™-8 (Superlite™) column was used. DAX™-8 is a polymethyl methacrylate non-ionic macroporous resin with a mean pore diameter of 225 Å, particle size of 0.3 mm and surface area of 160 $m^2 g^{-1}$. It is well-established that this resin is a suitable replacement of the Amberlite™ XAD resin outlined in the IHSS method, but subsequently discontinued.[289] The technical specifications are very similar, and some authors even highlight that the organic impurities, or fines content, of DAX™-8 is lower than XAD

resin.[290] Both these resins absorb and desorb organic matter easily by weak intermolecular interactions and capture a large fraction of organic matter. It has been reported that this method does not bind all FA molecules,[291] however, DAX™-8 is the most widely used resin in HS research as it conforms to the IHSS standards. According to literature, the extraction efficiency of NOM on the DAX-8™ resin is approximately 90% for a typical marsh water sample.[292] It should also be noted that DAX™-8 resin represents a hydrophobic layer, thus the obtained sample will be the hydrophobic component of FA. A study by Hanninen *et al*.[293] demonstrated that the hydrophobic component of peat FA contained the aromatic molecules of FA and therefore there is no requirement during this study to retain the hydrophilic eluent.

2.7.2.1. Preparation of the DAX-8™ resin

In order to prepare the column, the DAX-8™ resin was first dissolved in NaOH (0.1 M) to remove fines, which were decanted off. This process was repeated until the organic content of the supernatant was below the detection limits of UV-vis spectrophotometer at 254 nm. The DAX™-8 resin was then poured as a methanol slurry into a 25 mm (outer diameter) x 584 mm glass-fritted chromatography column with 250 ml solvent reservoir (Aldrich(R) calibrated, Z560375) so that approximately 50% of the column length was filled with the resin. The column was then packed using methanol. To condition the column for RMFA application, the resin was rinsed as follows:

i. 1 column volume of Milli-Q(R) water to remove methanol.
ii. 1 column volume of acetonitrile.
iii. 5-10 column volumes of methanol.
iv. 2 column volumes of Milli-Q(R) water.
v. 3 x 1 column volume of 0.1 M NaOH followed by 0.1 M HCl.
vi. Milli-Q(R) water until the conductivity was below 10 μS cm^{-1} measured using a Jenway 4200 conductivity meter to ensure the removal of acid and base.

The pH of the eluent was checked at each stage using pH indicator paper (Sigma Aldrich). The pH of the water before application of RMFA was neutral.

Before applying the RMFA sample, the pH was adjusted to 2 using NaOH (0.1 M). The RMFA samples were pre-filtered using 8 micron hardened ashless filter paper then 4 micron hardened ashless filter paper in order to remove any small particles. Once filtered, the samples were loaded

onto the DAX™-8 column. The pH of the effluent was measured and the organic content was checked via UV-vis spectrophotometry. The capacity of the column was reached when organic content, indicated by the absorbance at 254 nm, increased suddenly. Once the capacity was reached the RMFA was eluted using NaOH (0.1 M) and refrigerated.

2.7.2.2. Regeneration of the DAX™-8 column

Once the RMFA had been fully eluted, the DAX™-8 column was rinsed as follows to regenerate the column for the next batch of RMFA:

1. NaOH 0.1 M until DOC decreased below 9 mg l^{-1} as indicated via UV-Vis absorbance.
2. 1 column volume of methanol.
3. 1 column volume of acetonitrile.
4. 1 column volume of methanol.
5. Milli-Q® water until free of methanol.

Lastly, the resin was rinsed sequentially NaOH (0.1 M), HCl (0.1 M) and Milli-Q® water to remove remaining impurities and condition the column for the next batch of RMFA. The conductivity of the eluent of final Milli-Q® water rinsing step was below 10 μS cm^{-1}.

2.7.2.3. Issues with the DAX™-8 column

A number of issues were encountered when using the DAX™-8 column. The first was the blocking of the column frit. This initially occurred after the first batch of RMFA was eluted off, during regeneration, when the flow rate slowed to a standstill. Variables such as temperature and pressure were increased with no success. Methanol was applied to the resin to release the blockage on the frit, however, once the NaOH or HCl were applied, the flow stopped. The frit is made of glass fibre with a porosity between 15-25 μm, thus should not be blocked by the material which was filtered through 4 μm paper. However, on contact with the manufacturers, the blocking mechanism was suggested to involve a complex interplay of very small particles or big molecules fouling the frit. Only sonication the frit and subsequent application of methanol was found to clean the frit and return the flow rate to normal. However, upon repacking of the column, the same problem occurred. Thus a new resin was used and sample was passed through as quickly as possible and not left in contact with sample overnight. In addition, the subsequent

samples were filtered under vacuum using a 0.2 μm PES filtration system (VWR(R) vacuum filtration system, Standard line, VWR International) to try to ensure blockage of the frit from the sample was not possible.

2.7.3. DOWEX™ 50WX8 cation exchange

The cation exchange resin used in the IHSS procedure is the AG-MP-50. However, this resin is now discontinued and the widely used alternative is DOWEX™-50WX8 (Supelco, Sigma-Aldrich Co. LLC). DOWEX™-50WX8 is a strong acid cation exchange resin whose base matrix is a copolymer of styrene and divinylbenzene (8% cross linkage). This base is functionalised with SO_3H groups which provide the exchangeable H^+ ions. The resin pore size range used in this study was 100-200 mesh providing a capacity of 4.8 meq g^{-1}.

2.7.3.1. Preparation and regeneration of the DOWEX™ 50WX8 cation exchange columns

The DOWEX™-50WX8 resin was packed as a methanol slurry into five 10 ml solid phase extraction (SPE) columns (Bio-Rad laboratories Inc.). The resin was then washed with 5 bed volumes of methanol to remove impurities, followed by Milli-Q(R) water to remove the methanol. The following procedure was conducted before and after sample application to condition and re-condition the columns:

1) Flush column with water to remove remaining sample.
2) Add 1 M NaOH until the eluent from the column is alkaline (monitored with pH paper).
3) Add 4 bed volumes 2 M HCl – (colour changes from golden brown to darker brown on application of HCl).
4) Flush with Milli-Q(R) water until the eluent conductivity is below the IHSS recommended 5 μS cm^{-1}.

2.7.3.2. DOWEX™ 50WX8 cation exchange of RMFA

Prior to loading sample, the water level above the resin was minimised. RMFA was then applied drop wise until the sample height above the resin was 2 cm. At this point, RMFA was poured into the column in batches of 100 ml. The sample was then allowed to flow under gravity and the first 2-3 ml of eluent were discarded. Then the remaining eluent was collected under positive pressure

with use of balloons. On application of RMFA, saturation was detected by the colour change of the resin from light to olive brown as well as an increase in pH of the eluent. After cation exchange the eluent was lyophilised in 50 ml portions, collected and refrigerated until further use.

2.8. Collection of DOM from Needle's Eye (NEDOM)

For comparison with the other two HS extracts, samples of peat pore water were taken from a bog section of the Needle's Eye site near the base of the cliff. A 1 m pit was dug, revealing dark water. The hole quickly filled with water, which was then collected using a bucket inserted into the pit. Three 10 l containers were filled, with aid of a funnel and designated as samples 1, 2 and 3. Sample 1 contained water mostly from surface flow and was considered to be rainwater.

All samples were passed initially through 2 mm, 64 μm and 20 μm stainless steel sieves to remove any solid material. Subsequently the samples were filtered under vacuum through 8 μm and 2.7 μm hardened ashless filter paper. The liquid was then centrifuged to remove any further solid material before passing the supernatant through 0.45 μm Corning® bottle-top vacuum filter systems.

2.8.1. Ultrafiltration of NEDOM

To concentrate the DOM in the collected peat water samples 1-3, ultrafiltration through a 1 kDa Ultrasette™ lab Tangential flow filtration (TFF) device with Omega™ membrane with suspended screen was conducted. TFF involves passing a solution horizontally through the centre of two membrane plates under pressure. The material smaller than 1 kDa flows through the membrane into the filtrate tank, while the large molecules recirculate and concentrate in the feed chamber.

This filtration process has the advantage over downwards flow filtration processes by preventing fouling of the membrane as the horizontal crossflow prevents larger molecules blocking the filtrate, allowing the flux to remain at full rate as the filtered volume increases.[294] The horizontal flow requires pressure that is provided by a peristaltic pump. The set-up of the ultrafiltration system is shown in Figure 2.5.

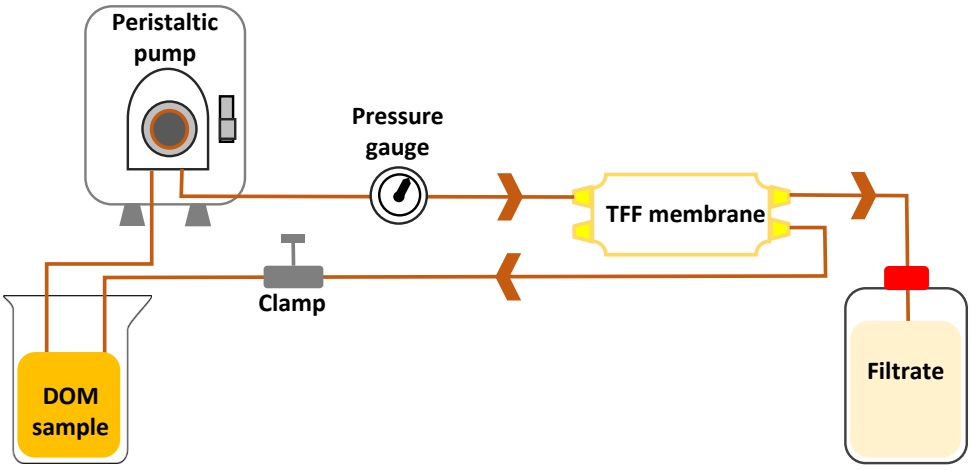


Figure 2.5 Schematic of the ultrafiltration set up used to concentrate NEDOM. Adapted from ref.[294]

A peristaltic pump is connected to the sample and membrane via MasterFlex® tubing. The sample passes through a cross flow membrane under constant pressure (40 psi). The filtrate was collected in a 10 l container while the retentate flows back to the sample vessel. Upon starting the above procedure, the system is equilibrated by adjusting the speed of the rotor until a constant flow rate is attained. Once the filtrate container was full, the contents were discarded, while the DOM sample after concentration was collected and stored in a 4 °C cold room to prevent any bacterial growth. Between samples 1-3, the unit was cleaned with 0.1 M NaOH then Milli-Q® water and the tubes checked for wear and tear. Using this procedure the volume of sample was reduced from 10 l to 400 ml, changing in colour from light yellow to dark orange.

After ultrafiltration, 50 ml aliquots of the retentate samples were freeze-dried and samples 2 and 3 were combined for NMR analysis.

2.9. Methylation procedures

Two methylation methods were tested on the model mixture **I**. Both methods used methyl iodide (CH_3I) as the methylating agent. CH_3I reacts via a basic S_N2 nucleophilic substitution reaction with carboxylate and alkoxide groups to give methyl esters and ethers, respectively. A base is required to activate the acidic functional groups. It is the base used which constitutes the difference between the two methods tested, the first uses NaH, denoted as Method A, while the second uses TBAH as the base, denoted as Method B. As mentioned in Section 1.6.2, NaH is a strong base but also flammable in air. TBAH on the other hand is a medium strength base that is much safer to use.

The methylation scheme of method A using CH_3I and NaH with an alcohol (ROH) is shown in Figure 2.6.

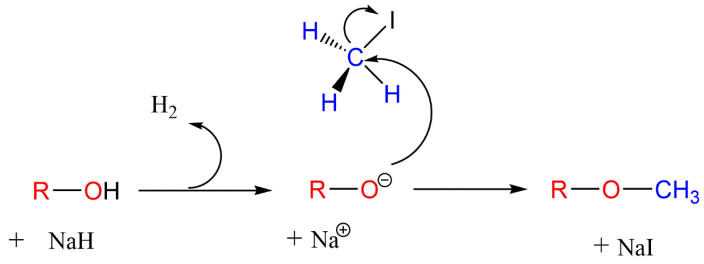


Figure 2.6 Methylation of ROH using method A.

As stated in Section 1.6.2, the benefit of TBAH is that in addition to being the base in the reaction it is also a phase transfer catalyst, which is soluble in the organic as well as aqueous phase. Phase transfer catalysts facilitate the migration of a reactant from one phase (aqueous) to another (organic), in which the reaction is taking place.[261] In the case of methylation, TBAH dissolves the hydroxyl containing compound ROH, and transfers it to the organic phase (THF) where the reaction with CH_3I takes place (Figure 2.7).

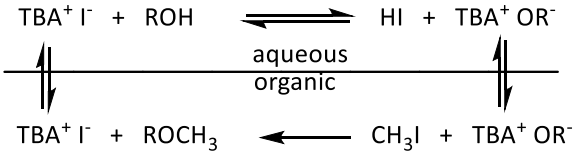


Figure 2.7 Methylation of ROH using method B with methyl iodide (MeI) and the aid of the phase transfer reagent TBAH. Adapted from ref.[261]

Method A was conducted in this study using modifications to the methylation procedures (which has been already successfully applied to HS samples) used by Thorn *et al.* and Mikita *et al.*,[248, 254] while the literature methylation procedure[252, 262] was used in method B.

To obtain the desired ^{13}C-labelled compounds $^{13}CH_3I$ was used. Both methods were first attempted on the model mixture **I** using inexpensive unlabelled CH_3I. The chosen method was then applied to model mixture **II** and the HS samples.

2.9.1. Methylation of model mixture I using method A

Additional steps were incorporated into the literature method following the analysis of the NMR spectra recorded after the initial methylation attempt, namely:

1. Pre-removal of the oil used to store NaH by washing the amount required with hexane.
2. Removal of DMF from the reaction product by rotary evaporation (18 mbar at 60 °C).

Further modifications were made throughout the course of testing method A leading to a final method using $^{13}CH_3I$ as detailed below.

NaH, 60% in oil, (0.754 g, 18.85 mmol) was washed with hexane (5 cm^3, 37.99 mmol) before DMF (10 cm^3, 129.68 mmol) was added and cooled in an ice bath to 5 °C, under N_2. Compounds **A** (0.247 g, 1.45 mmol), **B** (0.200 g, 1.45 mmol), **C** (0.224 g, 1.45 mmol) and **D** (0.200 g, 1.22 mmol) were dissolved in DMF (10 cm^3, 129.68 mmol) and the NaH solution was added while stirring. The mixture was then warmed to room temperature and stirred for 2 h. CH_3I (1.24 cm^3, 18.85 mmol) was added and stirred for a further 16 h before adding chloroform (30 cm^3). After stirring for a further 5 h, the mixture was neutralised using 1 M HCl. Milli-Q® water (40 cm^3) was added creating a separate layer from chloroform. The water layer was separated and washed with chloroform (40 cm^3). The organic fractions were combined, washed with water (3 x 40 cm^3) and dried over $MgSO_3$. Chloroform was evaporated off and DMF was removed at 18 mbar at 60 °C to give 1.087 g of dark yellow oil. ^1H NMR spectra was measured in $CDCl_3$ to check if DMF was removed. If DMF still remained in large quantities it was evaporated under vacuum again. Once DMF was reduced to a minimum, the sample was kept for further NMR analysis.

2.9.2. Methylation of model mixture I using method B

A mixture of the model compounds was prepared using the sample amounts as Section 2.9.1 dissolved in THF (5 cm^3, 61.66 mmol). The resulting solution was slowly stirred, while adding TBAH (0.5010 g, 1.93 mmol). After stirring for another 1.5 h CH_3I (0.6 cm^3, 9.64 mmol) was added forming a brown solid. To this, THF (4 cm^3, 49.33 mmol) was added to dissolve the solid before neutralising the solution with HCl (50 µl, 0.1 M). THF was then removed under reduced pressure at 50-70 °C. The mixture was then extracted using chloroform (50 cm^3) and Milli-Q® water (3 x 50 cm^3) using the same procedure as outlined for method A. The resulting mixture was washed with

an aqueous solution of NaCl or trisodium citrate to remove unreacted TBAH and dried over $MgSO_4$. The remaining chloroform was removed by evaporation to leave a brown oil.

2.9.3. Methylation of mixture II, RMFA and NEHA using method A

The methylation scheme used on model mixture **II** is shown in Figure 2.8.

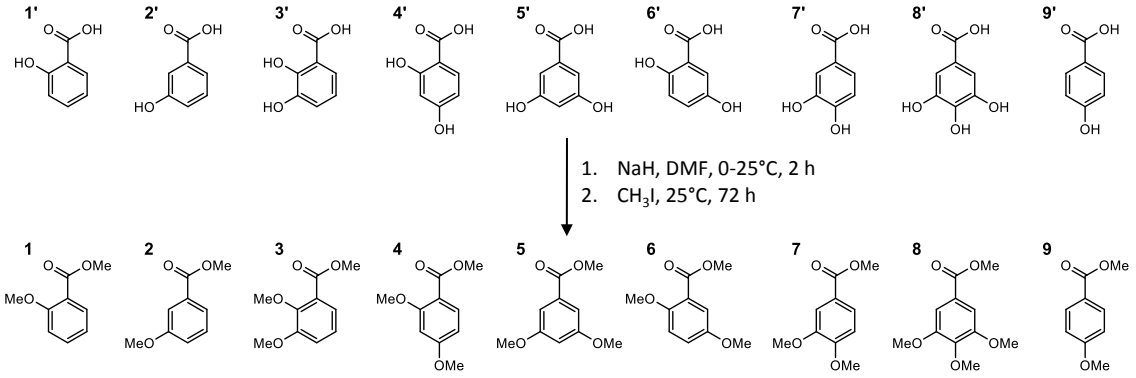


Figure 2.8 Methylation of model mixture **II** using method A.

Under nitrogen, NaH 60% in oil (0.142 g, 3.55 mmol) was washed with hexane (5 cm^3, 0.04 mmol) and DMF (10 cm^3, 129.98 mmol) added. The stirring suspension was cooled in an ice bath to 5 °C. A mixture of 2-hydroxybenzoic acid **1′** (0.014 g, 0.10 mmol), 3-hydroxybenzoic acid **2′** (0.014 g, 0.10 mmol), 2,3-dihydroxybenzoic acid **3′** (0.015 g, 0.10 mmol), 2,4-dihydroxybenzoic acid **4′** (0.015 g, 0.10 mmol), 3,5-dihydroxybenzoic acid **5′** (0.015 g, 0.10 mmol), 2,5-dihydroxybenzoic acid **6′** (0.015 g, 0.10 mmol), 3,4-dihydroxybenzoic acid **7′** (0.015 g, 0.10 mmol), 3,4,5-trihydroxybenzoic acid **8′** (0.017 g, 0.10 mmol) and 4-hydroxybenzoic acid **9′** (0.014 g, 0.10 mmol) was dissolved in DMF (5 cm^3, 64.99 mmol) and the solution was added slowly to the stirring suspension of NaH/DMF.

The mixture was warmed to room temperature and stirred for 2 h. ^{13}C-labelled CH_3I (0.263 cm^3, 4.23 mmol) was added and stirred for a further 72 h under reflux. The mixture was neutralised using 0.5 M HCl and Milli-Q® water (30 cm^3) added. Chloroform (30 cm^3) was added, the Milli-Q® water layer was separated and the organic fractions were combined. The combined organic fractions were washed with water (2 x 30 cm^3), dried over $MgSO_4$ and evaporated (after removing chloroform, DMF was removed at 9 mbar @ 60 °C) to give a mixture of methyl (2-methoxy)benzoate **1**, methyl 3-methoxybenzoate **2**, methyl 2, 3-dimethoxybenzoate **3**, methyl 2, 4-dimethoxybenzoate **4**, methyl 3, 5-dimethyoxybenzoate **5**, methyl 2, 5-dimethoxybenzoate **6**,

methyl 3, 4-dimethoxybenzoate **7**, methyl 3, 4, 5-trimethoxybenzoate **8** and methyl 4-methoxybenzoate **9**.

The same reaction procedure was used to methylated RMFA (0.6 g) and NEHA (0.5 g) using ^{13}C-labelled and unlabelled methyl iodide, respectively.

2.10. Hydrolysis of mixture II

A fully methylated sample of model mixture **II** was subjected to hydrolysis by adding a few droplets of HCl (0.1 M) to the mixture in chloroform in order to convert the esters back to carboxylic acid groups. The sample was subjected to evaporation to remove chloroform and then re-dissolved in $CDCl_3$. This provided a sample containing only methoxy ethers. The hydrolysis reaction was conducted to investigate the effects on chemical shift when converting ester groups back to carboxylic acids. This could add another layer of information for the molecular characterisation of phenolic molecules in methylated mixtures.

2.11. FT-IR spectrometry of methylated NEHA

In order to investigate the results of methylation of the NEHA, FT-IR spectra were acquired before and after the methylation procedure. FT-IR spectroscopy was performed using a few drops of methylated sample on a Spectrum 100 Perkin Elmer ATR-FT-IR spectrometer. Note this was not conducted for the RMFA sample due to the small amount of material collected after methylation.

2.12. NMR parameters used for the initial inspection of the model mixture I, model mixture II RMFA, RMHA, NEHA, and NEDOM.

All spectra were acquired on a 800 MHz AVANCE III Bruker spectrometer equipped with a TCI ^{1}H-optimised 5 mm cryoprobe with z-gradients. The temperature was set to 15 °C unless otherwise specified. The reason for this was to slow down the diffusion of molecules and thereby reduce diffusion-limited losses in the INADEQUATE experiments that have long delays between the de- and re-focusing PFGs. The NMR spectra of unmodified samples were all acquired at 18 °C. The pulse programs developed in this work are presented in Appendix A (see external web link).

For NMR analysis, methylated model mixture **I** obtained from method A (Section 2.9.1) and B (Section 2.9.2) were each dissolved in $CDCl_3$ (500 µl). A ^{13}C-methylated model mixture **I** obtained from method A was also dissolved in $CDCl_3$ (550 µl).

For NMR analysis, ^{13}C-methylated model mixture **II** (Section 2.9.3) was dissolved in $CDCl_3$ (800 µl) by taking the appropriate amount of the concentrated solution to prepare a sample with an average concentration of 1.4 mM of each compound **1-9** in $CDCl_3$ (550 µl).

For NMR analysis, unlabelled methylated NEHA (Section 2.9.3) was dissolved in $CDCl_3$ (550 µl). ^{13}C-methylated RMFA (10 mg) was also dissolved in $CDCl_3$ (550 µl).

All NMR tubes were sealed to avoid evaporation of $CDCl_3$.

Non-methylated samples of NEHA, RMHA and NEDOM (10 mg) samples were dissolved in NaOD/D_2O (0.01%/550 µl), while RMFA (10 mg) were dissolved in D_2O (550 µl) samples. 1D ^1H spectra were acquired for initial characterisation using low-power HOD presaturation. 8 scans were collected with 4 dummy scans. The spectral width was set between 20-40 ppm depending on the sample. The acquisition and relaxation delay was maintained at 1 and 8 s respectively.

2.13. NMR parameters used for the characterisation of methylated model mixture I

All 1D NMR spectra were acquired using standard Bruker pulse sequences and parameters. A **1D ^1H** spectra were acquired using a relaxation delay and acquisition times of 2 and 2.92 s respectively, 4 scans were accumulated in 39 s. The FIDs were zero filled once and a 0.3 Hz line broadening was applied prior to FT. **1D ^{13}C** spectra were acquired using a relaxation and acquisition times of 2 and 1.36 s; the number of scan was adjusted to provide adequate S/N ratio. The FIDs were zero filled once and a 1 Hz exponential line broadening was applied prior to FT.

A **2D ^1H, ^{13}C HSQC** spectrum was recorded using the Bruker *hsqcetgpsisp2.2* sequence. 1024 x 512 time domain (TD) points were acquired during the t_2 and t_1 acquisition times of 1 s and 116 ms, respectively, covering 0.6 and 11 ppm. The delays were optimised for a $^1J_{CH}$ = 145 Hz. The ^1H and ^{13}C carrier frequencies were set to 3.9 and 56.7 ppm, respectively. The relaxation time was set to 0.5 s and 2 scans were acquired per t_1 increment. The spectrum was acquired without the ^{13}C decoupling.

A **2D INEPT-INADEQUATE** spectrum was acquired using the pulse sequence shown in Figure 4.2 and pulse program named *inept_inadequate_2D*. The t_1 and t_2 acquisition times were 14.1 and 63.8 ms, respectively; 4 scans were accumulated into each of the 1536 complex data points sampling a spectra width of 270 ppm in F_1, making a total experimental time of 3 h. The spectral width in F_2 was 10 ppm. The relaxation delay was set to 1.5 s. The carrier frequency was set to 5 and 115 ppm for ^{1}H and ^{13}C, respectively. The experiment was optimised for an $^nJ_{CC}$ = 4 Hz and $^1J_{CH}$ = 145 Hz. 2 ms and 500 μs CHIRP adiabatic pulses were used for refocusing and inversion of ^{13}C resonances, respectively. The acquisition in t_1 was delayed by half of the dwell time, with the corresponding phases in this DQ dimension set to 90 and -180. Bi-level adiabatic decoupling was used during t_2.

A **4D HCCH$_3$** spectrum was acquired using the pulse sequence shown in Figure 4.8 and pulse program called *HCCH$_3$_4D*. The t_1, t_2, t_3 and t_4 acquisition times were set to 11, 11, 17 and 200 ms, respectively and spectral widths were 1.1, 14, 3 and 1.6 ppm, respectively. 32 scans were accumulated into each of the 20 complex data points in F_1, making a total experimental time of 1 d 16 h. The relaxation delay was set to 1 s. The ^{13}C frequency offsets were changed from 55 to 110 to 95 then back to 55 ppm, while the ^{1}H offsets were changed from 7.3 to 5.55 to 3.8 ppm during the pulse sequence. The experiment was optimised for an $^nJ_{CC}$ = 5 Hz, $^1J_{CHarom}$ = 155 Hz and $^1J_{CH_3}$ = 145 Hz. A 1 ms r-SNOB pulse was used in the DPFGSE and 2 ms composite adiabatic CHIRP pulses were used to refocus the ^{13}C resonances. For model mixture **I**, the trim pulse was not used. The acquisition in t_1, t_2 and t_3 was delayed by half of the dwell time, with the corresponding phases set to 90 and -180 in these dimensions. Xy32 decoupling was used during t_4 at γB_1= 3571 Hz. Non-uniform sampling at 25% was employed for this experiment.

A 2D **X-filtered HMBC** spectrum was acquired using the pulse sequence shown in Figure 4.10 and pulse program called *xfilter_hmbc_2D*. The following parameters were used: t_1 and t_2 acquisition times of 27 and 213 ms, respectively with spectral widths 140 and 12 ppm in F_2 and F_1; 4 scans were accumulated into each of the 1536 complex data points making the total experimental time of 4 h. The relaxation delay was set to 2 s while the long-range coupling evolution interval was set to $28/2^*{}^1J_{CH_3}$, where $^1J_{CH_3}$ = 144.5 Hz. This minimized the one-bond correlations. The carrier frequency for ^{1}H was set to 5.5 ppm, while the ^{13}C frequency was changed from 55 to 115 then back to 55 ppm during the pulse sequence. 2 ms composite adiabatic CHIRP

pulses was used to refocus the ^{13}C resonances. A refocused version of HMBC using echo-anti-echo for sign-discrimination in F_1 was used. Bi-level adiabatic decoupling was used during t_2.

A **2D X-filtered NOESY** spectrum was acquired using the pulse sequence shown in Figure 4.13 and pulse program called *xfilter_noesy_2D*. t_1 and t_2 acquisition times were 53.3 and 213 ms, respectively and spectral widths were 3 and 12 ppm; 32 scans were acquired into each of the 256 F_1 complex data points making a total experimental time of 5 h. The relaxation delay was 1.35 s. The mixing time was set to 600 ms. The X-filter was optimised for a $^1J_{CH_3}$ = 145 Hz. The acquisition in t_1 was delayed by half of the dwell time and the ϕ_1 phase was incremented by 90 degrees simultaneously with t_1 incrementation. This resulted in phases of 45, -180 in F_1.

A **2D X-filtered NOESY-TOCSY** spectrum was acquired using the pulse sequence shown in Figure 4.14 and pulse program called *xfilter_noesy_tocsy_2D*. The majority of the parameters were the same as for the 2D X-filtered NOESY experiment. The spectral widths were 7.4 and 12 ppm. A DIPSI2 mixing sequence was applied for 40 ms and the t_1 and t_2 acquisition times were 21.6 ms and 53 ms, respectively. The total experimental time was 5 h and 3 min. The carrier frequency was placed at 4.5 and 55 ppm for ^1H and ^{13}C channels, respectively. In order to remove zero quantum coherences a 20 ms adiabatic CHIRP pulse was applied. 2 ms r-SNOB pulses were used in the DPFGSE. Only one non-selective hard 180° pulse was used during the mixing time. The acquisition in t_1 was delayed by half of the dwell time, with the corresponding phases set to 90 and -180 in F_1.

A **4D HMQC-NOESY-TOCSY** spectrum was acquired using the pulse sequence shown in Figure 4.17 and program called HMQC_NOESY_TOCSY_4D. The delays were optimised for a $^1J_{CH_3}$ = 145 Hz. 3072 x 32 x 24 x 48 TD points were acquired during the t_4, t_3, t_2 and t_1 acquisition times of 192, 12, 8 and 13 ms, respectively, making a total experimental time of 2 d 21 h. The ^1H, ^1H, ^{13}C and ^1H dimensions covered 10, 1.7, 0.8 and 10 ppm, respectively. The NOESY and TOCSY mixing times were 1 s and 40 ms, respectively. The relaxation delay was set to 1 s, while the number of scans per increment was 8. The ^{13}C carrier frequency was set to 56.2 ppm while the ^1H carrier frequency was changed from 3.9 to 5.5 and 7.3 during the NOESY mixing time. The 1 and 0.6 ms PFGs were applied at strengths specified in Figure 4.17. A 1.5 ms r-SNOB 180° pulse was used for the DPFGSE. Two non-standard selective 180° pulses were placed at specified places during the NOESY mixing time. These pulses were created using two 2800 μs IBURP2 pulses, by time inverting the second pulse and phase shifting it to invert the aromatic protons. The acquisition in

t_1, t_2 and t_3 was delayed by half of the dwell time, with the corresponding phases set to 90 and -180 in these dimensions. Non-uniform sampling at 25% was employed for this experiment.

2.14. NMR parameters used for the characterisation of ^{13}C-methylated model mixture II

All 1D NMR spectra were acquired using standard Bruker pulse sequences and parameters. **1D ^{1}H** spectra were acquired using a relaxation delay and acquisition time of 1.5 and 2.72 s respectively. 1 scan was accumulated using a spectral width of 15 ppm. A **1D ^{13}C** spectrum was acquired using a relaxation and acquisition times of 2 and 1 s, respectively. 1 scan was acquired using spectral width of 20 ppm. The FIDs were zero filled once and a 1 Hz exponential line broadening was applied prior to FT.

Methyl and aromatic regions of the **1D ^{13}C** spectrum were also acquired separately. The aromatic region was acquired using 5120 scans with the acquisition time and relaxation delay of 0.4 and 2 s, respectively. The spectral width was 100 ppm, and carrier frequency set to 130 ppm. The methoxy of ^{13}C spectrum was acquired using 1 scan, with the acquisition time and relaxation delay set to 2 s. The spectral width was 20 ppm and the carrier frequency was set to 55 ppm.

A **2D ^{1}H, ^{13}C HSQC** spectrum (methoxy region shown in Figure 6.1) was recorded using the Bruker *hsqcetgpsisp2.2* sequence. 1024 x 512 TD points were acquired during the t_2 and t_1 acquisition times of 1 s and 116 ms, respectively, covering 0.6 and 11 ppm. The delays were optimised for a $^{1}J_{CH}$ = 145 Hz. The ^{1}H and ^{13}C carrier frequencies were set to 3.9 and 56.7 ppm, respectively. The relaxation time was set to 0.5 s and 2 scans were acquired per t_1 increment. The spectrum was acquired without ^{13}C decoupling.

A **2D ^{1}H, ^{13}C HSQC** spectrum was recorded using the Bruker *hsqcetgpiajcsp* sequence. The delays were optimised for a $^{1}J_{CH}$ = 145 Hz. 128 x 2048 TD points for each of the CLIP and CLAP experiments acquired in interleaved mode with t_2 and t_1 acquisition times of 640 and 26.5 ms, respectively, covering 2 and 12 ppm. The overall experimental time was 23 min. The relaxation time was set to 2 s and 2 scans were acquired per t_1 increment. The carrier frequency was set to 3.9 and 59 ppm in ^{1}H and ^{13}C channels, respectively. The spectrum was acquired without ^{13}C decoupling.

A **2D DOSY** spectrum (shown in Figure 6.3) was acquired using the Bruker sequence *dstebpgp3s*. 48 spectra with a diffusion gradient increasing in a linear fashion from 0.963 to 47.187% of the maximum strength of the gradient. δ was 1.5 ms, while Δ was 97.4 ms. The G value was 4257.7 HzG^{-1}. Bipolar gradients of 0.5 ms were applied. The relaxation delay 1.5 s, while acquisition time was 2.5 s. 16 scans were acquired in to each spectrum over a spectral width of 8 ppm, making the total experimental time 37 min. The 2D DOSY spectrum was processed using triexponential fitting of signal intensities.

A **3D INEPT-INADEQUATE-HSQC** spectrum was acquired using the pulse sequence shown in Figure 4.3 and pulse program called *INEPT-INADEQUATE_HSQC_3D*. 3072 x 128 x 256 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 181, 13 and 8 ms, respectively. The ^{1}H, ^{13}C and ^{13}C dimensions covered 75, 24 and 10 ppm, respectively. The relaxation delay was set to 1.2 s while the number of scans per increment was 4. The total experimental time was 14 h 47 min. The experiment was optimised for a $^{n}J_{CC}$ = 6 Hz and $^{1}J_{CH}$ = 153 Hz. The ^{1}H and ^{13}C carrier frequencies were set to 5.5 and 110 ppm, respectively. 2 ms PFGs were applied at strengths listed in Figure 4.3. A 2 ms composite adiabatic CHIRP ^{13}C pulse was utilised for refocusing carbon resonances, while a 500 μs CHIRP adiabatic pulse was used for inverting carbon resonances. The acquisition in t_1 and t_2 was delayed by half of the dwell time and the ϕ_1 phase was incremented by 60 degrees simultaneously with t_1 incrementation. This resulted in phases of 38, -180 in F_1. Bi-level adiabatic decoupling was used during t_3. Non-uniform sampling at 25% was employed for this experiment.

A **3D IPAP-INEPT-INADEQUATE-HSQC** spectrum was acquired using pulse sequence shown in Figure 4.4 and pulse program called *IPAP_INEPT_INADEQUATE_HSQC*_3D. 4096 x 384 x 320 TD points were acquired during the t_3, t_2, and t_1 acquisition times of 511, 39.8 and 10.6 ms, respectively. The ^{1}H, ^{13}C and DQ dimensions covered 5, 24 and 75 ppm. The relaxation delay was set to 1.4 s, while the number of scans per increment was 4. The overall duration of the experiment was 2 d, 10 h. The experiment was optimised for a $^{n}J_{CC}$ = 6 Hz and a $^{1}J_{CH}$ = 153 Hz. The ^{1}H and ^{13}C carrier frequencies were set to 5.5 and 110 ppm, respectively. 2 ms PFGs were applied at strengths listed in Figure 4.3. A 2 ms composite adiabatic CHIRP ^{13}C pulse was utilised for refocusing of carbon resonances, while a 500 μs CHIRP adiabatic pulse was used for inverting carbon resonances. The acquisition in t_1 and t_2 was delayed by half of the dwell time and the ϕ_1 phase was incremented by 80 degrees simultaneously with t_1 incrementation. This resulted in phases of 12, -180 in F_1. Non-uniform sampling at 10% was employed for this experiment.

A **3D HMQC-HMBC** spectrum was acquired using the pulse sequence shown in Figure 4.11 and pulse program called *hmqc_hmbc_3D*. 384 x 160 x 160 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 240, 33.1, and 11.4 ms, respectively. The ^{1}H, ^{13}C and ^{13}C dimensions covered 1, 12 and 35 ppm. The relaxation delay was set to 1.3 s, while the number of scans per increment was 4. The overall duration of the experiment was 12 h 36 min. The delays were optimised for a $^{1}J_{CH}$ = 145 Hz and a $^{n}J_{CH}$ = 6 Hz. The ^{1}H and ^{13}C carrier frequencies were set to 3.9 and 115 ppm, respectively. The acquisition in t_1 and t_2 was delayed by half of the dwell time and the ϕ_1 phase was incremented by -115 degrees simultaneously with t_1 incrementation. This resulted in phases of 120, -180 in F_1. A 2 ms composite adiabatic CHIRP ^{13}C pulse was utilised for refocusing of carbon resonances while a non-standard 150 μs r-SNOB 180° pulse was used to refocus the $^{1}J_{CH}$ evolution during the HMBC step. Xy32 decoupling was used during t_3 at γB_1 = 3571 Hz. Non-uniform sampling at 25% was employed for this experiment.

A **3D HMQC-NOESY** spectrum was acquired using the pulse sequence shown in Figure 4.15 and pulse program called *HMQC_NOESY_3D*. The delays were optimised for a $^{1}J_{CH_3}$ = 144.5 Hz. 8192 x 224 x 24 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 512, 53 and 4 ms, respectively. The ^{1}H, ^{13}C and ^{1}H dimensions covered 10, 12 and 0.25 ppm. The NOESY mixing time was 600 ms. The relaxation delay was set to 2 s, while the number of scans per increment was 4. The carrier frequencies in ^{1}H and ^{13}C dimensions were 3.9 and 56.6 ppm respectively. The acquisition in t_1 and t_2 was delayed by half of the dwell time, with the corresponding phases set to 90 and -180 in these dimensions. A 2 ms r-SNOB 180° pulse was used in the DPFGSE. Two non-standard selective 180° pulses were placed at specified places during the NOESY mixing time. These pulses were created using two 2800 μs IBURP2 pulses, by time inverting the second pulse and phase shifting it to invert the aromatic protons. Non-uniform sampling at 25% was employed for this experiment.

A **3D HMQC-NOESY-TOCSY** spectrum was acquired using the pulse sequence shown in Figure 4.16 and pulse program called *HMQC_NOESY_TOCSY_3D*. The delays were optimised for a $^{1}J_{CH_3}$ = 144.5 Hz. 8192 x 256 x 24 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 512, 53 and 4 ms. The ^{1}H, ^{13}C and ^{1}H dimensions covered 10, 12 and 0.25 ppm. The NOESY and TOCSY mixing times were 600 and 80 ms, respectively. The relaxation delay was set to 2 s, while the number of scans per increment was 8. The overall experimental time of 11 h. The carrier frequencies in ^{1}H and ^{13}C dimensions were 3.9 and 56.6 ppm respectively. The acquisition in t_1

and t_2 was delayed by half of the dwell time, with the corresponding phases set to 90 and -180 in these dimensions. Two hard 180° pulse were used during the mixing time, while a 2 ms r-SNOB 180° pulse was used for the DPFGSE. In order to remove zero quantum coherences a 20 ms adiabatic CHIRP pulse was applied simultaneously with 8% gradient after the DIPSI2 spin-lock. Non-uniform sampling at 25% was employed for this experiment.

A **3D HcCH$_3$** spectrum was acquired using a modified version of the pulse sequence shown in Figure 4.8 and the pulse program named *HcCH$_3$_3D*. 384 x 32 x 64 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 240, 53 and 29 ms, respectively. The delays were optimised for a $^1J_{CHarom}$ = 155 Hz, $^3J_{CC}$ = 6 Hz and $^1J_{CH_3}$ = 145 Hz. The ^{1}H, ^{13}C and ^{1}H dimensions covered 1, 1.5 and 1.4 ppm. The relaxation delay was set to 1.5 s, while the number of scans per increment was 4. The overall experimental time was 1 h and 11 min. To reduce off resonance effects the carrier frequencies were changed at specific places. The ^{13}C frequency offsets were changed from 55 to 110 to 95 then back to 55 ppm, while the ^{1}H offsets were changed from 7.1 to 5.5 to 3.9 ppm during the pulse sequence. The acquisition in t_1 and t_2 was delayed by half of the dwell time and ϕ_5 was incremented by -140 degrees simultaneously with t_2 incrementation. This resulted in phases 90, -180 in F_1 and 154, -180 in F_2. 2 ms composite adiabatic CHIRP ^{13}C pulses were utilised for refocusing carbon resonances. A 1 ms trim pulse is applied after the DPFGSE on full power while 1 ms r-SNOB 180° pulses were used for the DPFGSE. Xy32 decoupling was used during t_3 at γB_1= 3571 Hz. Non-uniform sampling at 25% was employed for this experiment.

A **3D hCCH$_3$** was acquired using a modified version of the pulse sequence shown in Figure 4.8 and the pulse program named *hCCH$_3$_3D*. 384 x 32 x 256 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 240, 53 and 27 ms, respectively. The ^{1}H, ^{13}C and ^{13}C dimensions covered 1, 1.5 and 24 ppm. The relaxation delay was set to 1.5 s, while the number of scans per increment was 4, making the total experimental time 4.5 h. The delays were optimised for a $^1J_{CHarom}$ = 155 Hz, $^3J_{CC}$ = 6 Hz and $^1J_{CH_3}$ = 145 Hz. To reduce off resonance effects the carrier frequencies were changed at specific places. The ^{13}C frequency offsets were changed from 57 to 110 to 96 then back to 55 ppm, while the ^{1}H offsets were changed from 7.1 to 5.5 to 3.9 ppm during the pulse sequence. The acquisition in t_1 and t_2 was delayed by half of the dwell time and the ϕ_5 phase was incremented by 165 degrees, while the ϕ_6 phase was incremented by -140 degrees simultaneously with t_1 and t_2 incrementation. This resulted in phases 28, -180 in F_1 and 157, -180 in F_2. A 2 ms composite adiabatic CHIRP ^{13}C pulses were utilised for refocusing carbon resonances. 1 ms trim

pulse is applied after the DPFGSE on full power, while 1 ms r-SNOB 180° pulses were used for the DPFGSE. Xy32 decoupling was used during t_3 at γB_1= 3571 Hz. Non-uniform sampling at 25% was employed for this experiment.

A **4D HCCH$_3$** spectrum was acquired using the pulse sequence shown in Figure 4.8 and the pulse program called HCCH$_3$_4D. 384 x 16 x 128 x 32 TD points were acquired during the t_4, t_3, t_2 and t_1 acquisition times of 240, 27, 13, and 14 ms, respectively. The ^{1}H, ^{13}C, ^{13}C and ^{1}H dimensions covered 1, 1.5, 24 and 1.4 ppm. The delays were optimised for a $^1J_{CHarom}$ = 155 Hz, $^3J_{CC}$ = 6 Hz and $^1J_{CH_3}$ = 145 Hz. The relaxation delay was set to 1.5 s, while the number of scans per increment was 8, making the overall experimental time 14 h and 20 min. To reduce off resonance effects the carrier frequencies were changed at specific places. The ^{13}C frequency offsets were changed from 57 to 110 to 96 then back to 55 ppm, while the ^{1}H offsets were changed from 7.1 to 5.5 to 3.9 ppm during the pulse sequence. The acquisition in t_1, t_2 and t_3 was delayed by half of the dwell time and the ϕ_5 phase was incremented by 165 degrees while the ϕ_6 phase was incremented by -140 degrees simultaneously with t_2 and t_3 incrementation, respectively. This resulted in phases of 90, -180 in F_1, 188, -180 in F_2 and 154, -180 in F_3. 2 ms composite adiabatic CHIRP ^{13}C pulses were utilised for refocusing carbon resonances. A 1 ms trim pulse is applied after the DPFGSE with full power, while 1 ms r-SNOB 180° pulses were used for the DPFGSE. Xy32 decoupling was used during t_4 at γB_1= 3571 Hz. Non-uniform sampling at 25% was employed for this experiment.

2.15. NMR parameters used for the characterisation of the hydrolysed ^{13}C-methylated model mixture II

Initial characterisation was carried out using the same experiment parameters outlined in Section 2.14.

A **2D ^{1}H, ^{13}C HSQC** spectrum was recorded using the Bruker *hsqcetgpiajcsp* sequence. The delays were optimised for a $^1J_{CH}$ = 145 Hz. 128 x 2048 TD points for each of the CLIP and CLAP experiments acquired in interleaved mode with t_2 and t_1 acquisition times of 640 and 26.5 ms, respectively, covering 2 and 12 ppm. The overall experimental time was 23 min. The relaxation time was set to 2 s and 2 scans were acquired per t_1 increment. The carrier frequency was set to 3.9 and 59 ppm in ^{1}H and ^{13}C channels, respectively. The spectrum was acquired without ^{13}C decoupling.

A **3D IPAP-INEPT-INADEQUATE-HSQC** spectrum was acquired using the same parameters as the fully ^{13}C-methylated model mixture **II** (see Section 2.14).

A **3D HMQC-NOESY-TOCSY** spectrum was acquired using the pulse sequence shown in Figure 4.17 and pulse program called *HMQC_NOESY_TOCSY_3D*. The delays were optimised for $^{1}J_{CH_3}$ = 144.5 Hz. 8192 x 64 x 32 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 512, 79.6 and 66.7 ms. The ^{1}H, ^{13}C and ^{1}H dimensions covered 10, 2 and 0.3 ppm. The NOESY and TOCSY mixing times were 600 and 80 ms, respectively. The relaxation delay was set to 2 s, while the number of scans per increment was 8. The overall experimental time of 11 h. The carrier frequencies in ^{1}H and ^{13}C dimensions were 3.98 and 56.3 ppm, respectively. The acquisition in t_1 and t_2 was delayed by half of the dwell time, with the corresponding phases set to 90 and -180 in these dimensions. Two hard 180° pulse were used during the mixing time, while a 2 ms r-SNOB 180° pulse was used for the DPFGSE. In order to remove zero quantum coherences a 20 ms adiabatic CHIRP pulse was applied simultaneously with 8% gradient after the DIPSI2 spin-lock. Non-uniform sampling at 25% was employed for this experiment.

2.16. NMR parameters used for the characterisation of methylated NEHA

A **1D ^{1}H** spectrum was recorded using the standard Bruker *zg* pulse sequence. The spectral width, relaxation delay and number of scans were set to 20 ppm, 8 s and 8 scans, respectively.

2.17. NMR parameters used for the characterisation of ^{13}C-methylated RMFA

A **1D ^{1}H** spectrum was recorded using the Bruker *zg* sequence. The spectral width, relaxation delay and number of scans were set to 20 ppm, 8 s and 8 scans, respectively. A **1D ^{13}C** spectrum for ^{13}C-methylated RMFA focusing on methoxy resonances was measured using the standard Bruker *zgpg* sequence. The spectral width was set to 260 ppm and the carrier frequency was set to 55 ppm. 512 scans were acquired with relaxation delay set to 2 s and a acquisition time of 630 ms, making the total experimental time 24 min.

A **2D ^{1}H, ^{13}C HSQC** spectrum (methoxy region shown in Figure 8.1) was recorded using the standard Bruker *hsqcetgp* sequence. 6144 x 1024 TD points were acquired during the t_2 and t_1

acquisition times of 320 and 127 ms, respectively, covering 12 and 20 ppm, making the total acquisition 1 h and 32 min. The delays were optimised for a $^{1}J_{CH}$ = 145 Hz. The ^{1}H and ^{13}C carrier frequencies were set to 3.8 and 55 ppm. The relaxation time was set to 2 s and 2 scans were acquired per t_1 increment. Xy32 decoupling was used during t_2 at γB_1= 3571 Hz.

A **2D ^{1}H, ^{13}C HSQC** spectrum (the aromatic region is shown in Figure 8.2) was recorded using the Bruker *hsqcedetgpsisp2.3* sequence. 4096 x 1536 TD points were acquired during the t_2 and t_1 acquisition times of 128 and 18.2 ms, respectively, covering 20 and 210 ppm, making the total acquisition 3 h. The delays were optimised for a $^{1}J_{CH}$ = 145 Hz. The ^{1}H and ^{13}C carrier frequencies were set to 4.7 and 100 ppm. The relaxation time was set to 1.6 s and 4 scans were acquired per t_1 increment. Bi-level adiabatic decoupling was used during t_2.

A **2D DOSY** spectrum (shown in Figure 8.15) was acquired using the standard Bruker sequence *dstebpgp3s*. 64 spectra with a diffusion gradient increasing in a linear fashion from 0.963 to 47.187% of the maximum gradient strength. δ was set to 1.5 ms, while Δ was set to 97.4 ms. The G value was 4257.7 Hz G^{-1}. Bipolar gradients of 0.5 ms were applied. The relaxation delay was 1.5 s, while the acquisition time was 2.5 s. 16 scans were acquired in to each spectrum over a spectral width of 8 ppm making the total experimental time 37 min. The 2D DOSY spectrum was processed using triexponential fitting of the signal intensities.

A **3D INEPT-INADEQUATE-HSQC** spectrum was acquired using the pulse sequence shown in Figure 4.3 and pulse program called *INEPT-INADEQUATE_HSQC_3D*. 3072 x 128 x 256 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 240, 13 and 8 ms, respectively. The ^{1}H, ^{13}C and ^{13}C dimensions covered 8, 24 and 75 ppm, respectively. The relaxation delay was set to 1.2 s while the number of scans per increment was 4, making the total experimental time 61 h. The experiment was optimised for a $^{n}J_{CC}$ = 6 Hz and $^{1}J_{CH}$ = 153 Hz. The ^{1}H and ^{13}C carrier frequencies were set to 5.5 and 110 ppm, respectively. A 2 ms composite adiabatic CHIRP ^{13}C pulse was utilised for refocusing carbon resonances, while another 500 μs CHIRP adiabatic pulse was used for inverting carbon resonances. The acquisition in t_1 and t_2 was delayed by half of the dwell time and the ϕ_1 phase was incremented by 60 degrees simultaneously with t_1 incrementation. This resulted in phases of 24, -180 in F_1.

A **3D HMQC-HMBC** spectrum was acquired using the pulse sequence shown in Figure 4.11 and pulse program called *hmqc_hmbc_3D*. 3072 x 96 x 320 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 240, 20, and 10 ms, respectively. The ^{1}H, ^{13}C and ^{13}C dimensions covered

8, 12 and 80 ppm. The relaxation delay was set to 1.3 s, while the number of scans per increment was 4. The overall duration of the experiment was 2 d 12 h 47 min. The delays were optimised for a $^{1}J_{CH}$ = 145 Hz and $^{n}J_{CH}$ = 6 Hz. The ^{1}H and ^{13}C carrier frequencies were set to 3.8 and 115 ppm, respectively. The acquisition in t_1 and t_2 was delayed by half of the dwell time and the ϕ_1 phase was incremented by -50 degrees simultaneously with t_1 incrementation. This resulted in phases of 153, -180 in F_1. A 2 ms composite adiabatic CHIRP ^{13}C pulse was utilised for refocusing of carbon resonances while a non-standard 150 μs r-SNOB 180° pulse was used to refocus the $^{1}J_{CH}$ evolution during the HMBC step. Xy32 decoupling was used during t_3 at γB_1 = 3571 Hz.

A **4D HCCH$_3$** was spectrum was acquired using the pulse sequence shown in Figure 4.8 and the pulse program called *HCCH$_3$_4D*. 575 x 24 x 56 x 16 TD points were acquired during the t_4, t_3, t_2 and t_1 acquisition times of 240, 15, 7, and 8 ms, respectively making the total experimental time 64 h. The ^{1}H, ^{13}C, ^{13}C and ^{1}H dimensions covered 1.5, 4, 20 and 1.2 ppm. The delays were optimised for a $^{1}J_{CHarom}$ = 155 Hz, $^{3}J_{CC}$ = 6 Hz and $^{1}J_{CH_3}$ = 145 Hz. The relaxation delay was set to 0.9 s, while the number of scans per increment was 8. To reduce off resonance effects the carrier frequencies were changed at specific places. The ^{13}C frequency offsets were changed from 56 to 110 to 96 then back to 56 ppm, while the ^{1}H offsets were changed from 7.5 to 5.52 to 3.8 ppm during the pulse sequence. The acquisition in t_1, t_2 and t_3 was delayed by half of the dwell time and the ϕ_5 phase was incremented by 135 degrees, while the ϕ_6 phase was incremented by -20 degrees simultaneously with t_2 and t_1 incrementation, respectively. This resulted in phases of 90, -180 in F_1 and 188, -180 in F_2 and 154, -180 in F_3. 2 ms composite adiabatic CHIRP ^{13}C pulses were utilised for refocusing carbon resonances. A 1 ms trim pulse is applied after the DPFGSE with full power, while 1 ms r-SNOB 180° pulses were used for the DPFGSE. Xy32 decoupling was used during t_4 at γB_1= 3571 Hz.

A **3D HcCH$_3$** was aquired using a modified version of the pulse sequence shown in Figure 4.8 and the pulse program named *HcCH$_3$_3D*. 512 x 72 x 80 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 213, 45 and 42 ms, respectively. The delays were optimised for a $^{1}J_{CHarom}$ = 155 Hz, $^{3}J_{CC}$ = 6 Hz and $^{1}J_{CH_3}$ = 145 Hz. The ^{1}H, ^{13}C and ^{1}H dimensions covered 1.5, 4 and 1.2 ppm, respectively. The relaxation delay was set to 1.1 s, while the number of scans per increment was 24, making the total experimental time 60 h. To reduce off resonance effects the carrier frequencies were changed at specific places. The ^{13}C frequency offsets were changed from 55 to 110 to 95 then back to 55 ppm, while the ^{1}H offsets were changed from 7.1 to 5.5 to 3.9 ppm during the pulse

sequence. The acquisition in t_1 and t_2 was delayed by half of the dwell time and ϕ_1 to ϕ_4 were incremented by -20 degrees simultaneously with t_1 incrementation. This resulted in phases of 83, -180 in F_1 and 90, -180 in F_2. 2 ms composite adiabatic CHIRP ^{13}C pulses were utilised for refocusing carbon resonances. A 1 ms trim pulse is applied after the DPFGSE on full power while 1 ms r-SNOB 180° pulses were used for the DPFGSE. Xy32 decoupling was used during t_3 at γB_1= 3571 Hz.

A **3D hCCH$_3$** was acquired using a modified version of the pulse sequence shown in Figure 4.8 and the pulse program named *hCCH$_3$_3D*. 512 x 72 x 96 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 213, 45 and 12 ms, respectively. The ^{1}H, ^{13}C and ^{13}C dimensions covered 1.5, 4 and 20 ppm. The relaxation delay was set to 1.1 s, while the number of scans per increment was 16 making the total experimental time 47 h. The delays were optimised for a $^1J_{CHarom}$ = 155 Hz, $^3J_{CC}$ = 6 Hz and $^1J_{CH_3}$ = 145 Hz. To reduce off resonance effects the carrier frequencies were changed at specific places. The ^{13}C frequency offsets were changed from 56 to 110 to 96 then back to 56 ppm, while the ^{1}H offsets were changed from 7.15 to 5.52 to 3.8 ppm during the pulse sequence. The acquisition in t_1 and t_2 was delayed by half of the dwell time and ϕ_5 was incremented by +135, while ϕ_6 was incremented by -90 degrees simultaneously with t_1 and t_2 incrementation, respectively. This resulted in phases of 28, -180 in F_1 and 129, -180 in F_2. 2 ms composite adiabatic CHIRP ^{13}C pulses were utilised for refocusing carbon resonances. A 1 ms trim pulse is applied after the DPFGSE on full power while 1 ms r-SNOB 180° pulses were used for the DPFGSE. Xy32 decoupling was used during t_3 at γB_1= 3571 Hz.

A **3D HMQC-NOESY** spectrum was acquired using the pulse sequence shown in Figure 4.15 and pulse program called *HMQC_NOESY_3D*. 8192 x 224 x 64 time domain points were acquired during the t_3, t_2 and t_1 acquisition times of 512, 43 and 4 ms, respectively. The ^{1}H, ^{13}C and ^{1}H dimensions covered 9, 13 and 1 ppm. The delays were optimised for a $^1J_{CH_3}$ = 144.5 Hz. The NOESY mixing time was 600 ms. The relaxation delay was set to 2 s, while the number of scans per increment was 4, making the total experimental time 51 h. The carrier frequencies in ^{1}H and ^{13}C dimensions were 3.9 and 56.6 ppm respectively. The acquisition in t_1 and t_2 was delayed by half of the dwell time, with the corresponding phases set to 90 and -180 in these dimensions. A 2 ms r-SNOB 180° pulse was used in the DPFGSE. Two non-standard selective 180° pulses were placed at specified places during the NOESY mixing time. These pulses were created using two 2800 μs

IBURP2 pulses, by time inverting the second pulse and phase shifting it to invert the aromatic protons.

A **3D HMQC-NOESY-TOCSY** spectrum was acquired using the pulse sequence shown in Figure 4.16 and pulse program called *HMQC_NOESY_TOCSY_3D*. 8192 x 224 x 64 TD points were acquired during the t_3, t_2 and t_1 acquisition times of 465, 43 and 4 ms, respectively, making a total experimental time of 47 h. The ^{1}H, ^{13}C and ^{1}H dimensions covered 9, 13 and 1 ppm, respectively. The delays were optimised for a $^{1}J_{CH_3}$ = 144.5 Hz. The NOESY and TOCSY mixing times were 600 and 80 ms, respectively. The relaxation delay was set to 2 s, while the number of scans per increment was 8. The carrier frequencies in ^{1}H and ^{13}C dimensions were 3.7 and 57 ppm, respectively. The acquisition in t_1 and t_2 was delayed by half of the dwell time, with the corresponding phases set to 90 and -180 in these dimensions. Two hard 180° pulse were used during the mixing time, while a 2 ms r-SNOB 180° pulse was used for the DPFGSE. In order to remove zero quantum coherences a 20 ms adiabatic CHIRP pulse was applied simultaneously with 8% gradient after the DIPSI2 spin-lock.

2.18. NOESY transfer efficiency using ^{13}C-methylated model mixture II

The NOESY transfer was assessed by measuring the standard NOESY sequence *noesyphpr* using a 600 ms mixing time. The NOE step can occur from aromatic to methoxy protons or vice versa. In order to measure the efficiency of the NOE transfer to and from the methoxy groups, the diagonal peaks for the aromatic protons and methoxy protons were integrated as well as the cross-peaks. For a given transfer step, the cross-peak integral was divided by the diagonal integral. As the no ^{13}C decoupling was used, the methoxy diagonal peak are split by $^{1}J_{CH_3}$ couplings. To consider this, the transfer efficiency from the methyl methoxy protons was multiplied by a factor of 2. Using this method, the estimated Me to Arom and Arom to Me transfer percentages were 3.8 and 0.6% respectively.

Chapter 3. Results and Discussion. Characterisation of obtained HS samples and methylated model mixture I, model mixture II, RMFA and NEHA

HS samples always appear like fluffy Yorkshire puddings after lyophilisation

As part of this study several samples of NOM were prepared as outlined in *Materials and Methods, Chapter 2*. This section presents the initial NMR characterisation of these samples as well as the methylation of selected ones. In addition, the results of the methylation procedures on model mixture **I** and **II,** outlined in Section 2.9, are described.

3.1. HS samples

Samples were isolated, purified and concentrated using the IHSS protocols detailed in *Materials and Methods, Chapter 2*. After the last lyophilisation step powdered samples of NEHA (14.410 g), RMHA (4.511 g) and RMFA layer 1 (0.045 g), RMFA layer 2 (0.2130 g) and RMFA layer 3 (0.1770 g) were obtained. The NEDOM samples were concentrated using ultrafiltration as outlined in Section 2.8.1. After lyophilisation NEDOM samples 2 and 3 were combined to give one sample (0.543 g), while sample 1 gave less material (0.166 g). Sample 1 was not combined with samples 2 and 3 because it was believed to be mostly rainwater and not pore water.

3.2. Initial NMR characterisation and comparison of the obtained HS samples

To inspect the molecular characteristics of prepared NOM samples their ^{1}H spectra were measured, using the parameters given in Section 2.12, and shown in Figure 3.1. Due to the fact that NEDOM sample 1 was mostly rainwater this sample was not subjected to further analysis. In terms of the RMFA layers, the layer collected between the depths of 20 and 30 cm was chosen due to the fact it was considered to be the oldest and most humified of the three layers.[286]

The ^{1}H spectra, shown in Figure 3.1 are typical of NOM samples and reflect the presence of aromatic (8.5 to 6 ppm), carbohydrate/methoxy (6-4 ppm) and aliphatic (< 4 ppm) molecules. The signal around 4.8 ppm is the HOD signal. The differences between these samples lies in the relative amount of the different compound classes. It is immediately evident that RMFA, RMHA and NEHA have large amount of carbohydrates or methoxy group containing molecules, based on the intense signals between 3-5 ppm. In addition, these samples also show two distinct aliphatic signals centred around 1.48 and 1.08 ppm. These are characteristic of long chain aliphatics as reported in Section 1.5.2. Both SRFA and SRHA show intense aliphatic signals between 2-3 ppm, which are absent in the NEDOM/NEHA samples and less prominent in the RMFA and RMHA peat samples. These signals are likely due to the CRAM molecules, which

show resonances between 1.6 and 3.2 ppm. These molecules are clearly specific to riverine/marine samples.[220]

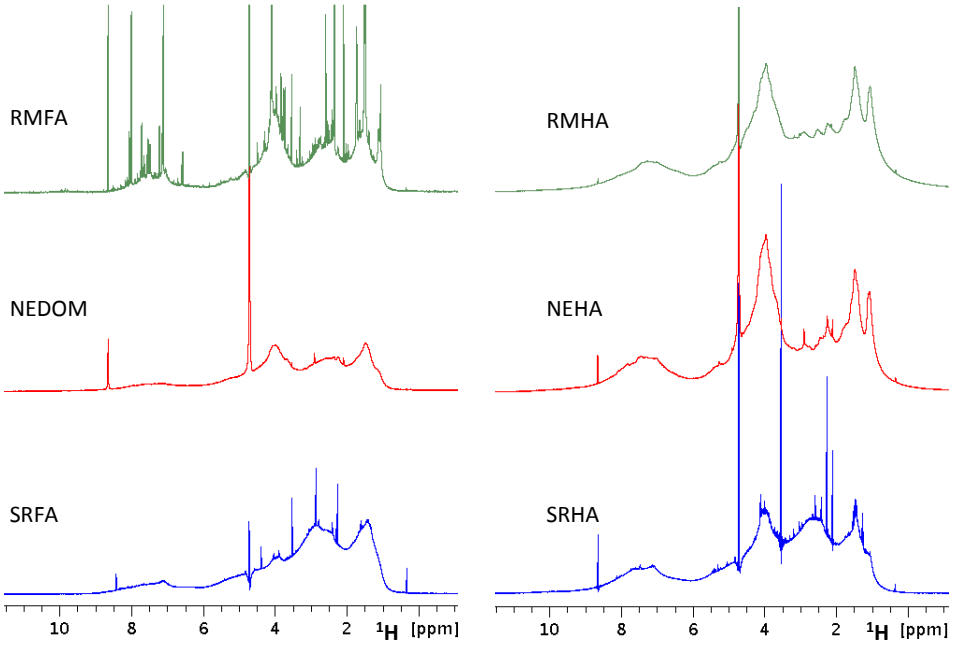


Figure 3.1 800 MHz ^{1}H spectra of the HS samples, Red Moss Fulvic Acid (RMFA), Needle's Eye Dissolved Organic Matter (NEDOM), Suwannee River Fulvic Acid (SRFA), Red Moss Humic Acid (RMHA), Needle's Eye Humic Acid (NEHA), Suwannee River Humic Acid (SRHA).

The RMHA and NEHA ^{1}H spectra look very similar to that obtained from the highland peat HA (see 1D projection in Figure 1.14) sample by Hertkon *et al.*[188] indicating that the relative amounts of compound classes are similar in different peat samples. The NEDOM sample have similar features to the SRDOM sample (see Figure 1.10). The aromatics are pronounced in all HA samples, as well as RMFA. An intense signal, most likely attributed to formate, appears around 8.4-8.6 ppm in varying amounts in all samples.[195] As this is unlikely an artefact of the extraction procedure, it is likely formate is contained in each HS sample encapsulated in the molecular aggregate structures but released when the sample is dissociated at higher pH, and thus this compound is not lost during dialysis. Comparing all the samples, the ^{1}H spectrum of RMFA shows a larger number of resolved signals of individual compounds. As this sample went through additional concentration by affinity chromatography, these compounds were likely enriched. As phenolic compounds are the focus of this study, RMFA, RMHA and NEHA samples were taken for further consideration due to their higher aromatic content in their ^{1}H spectra. The next property to consider was the solubility as the methylation reactions require the HS sample

to dissolve in the aqueous phase. RMFA dissolved easily in D_2O, while both RMHA and NEHA required base to be added. Thus RMFA was the chosen candidate for further study, however, given the small amount of sample, the initial methylation studies were performed on the NEHA sample which gave the most material (Section 3.1.).

3.3. Methylation of model mixture I, model mixture II, NEHA and RMFA

Methylation methods A and B were tested on the model mixture **I**. Using the most successful method, model mixture **I** was methylated using ^{13}C-labelled methyl iodide. Model mixture **II** was also methylated using the preferred method with ^{13}C-labelled methyl iodide. As outlined in Section 3.2, the NEHA and RMFA were selected for methylation. The NEHA sample was used to test the selected methylation procedure with unlabelled methyl iodide, while the RMFA was methylated using ^{13}C-labelled methyl iodide. This section will discuss the methylation results and initial characterisation, using 1D ^{1}H and ^{13}C NMR, as well as FT-IR spectroscopy.

3.3.1. Methylation of model mixture I using method A with unlabelled CH_3I

Methylation was carried out using method A as detailed in Section 2.9.1 on model mixture **I** (see Figure 2.1). As outlined in Section 2.9.1, two issues had to be addressed, DMF and oil removal. Two ^{1}H NMR spectra after two separate methylation attempts on the model mixture **I** using method A are shown in Figure 3.2. Both spectra show methoxy proton signals (circled), which proves that methylation has indeed taken place. In addition, there was no indication of C-methylation as observed by Thorn *et al.*[248] However, spectrum (Figure 3.2a) contains large solvent signals due to DMF. The amount of DMF in the sample was reduced substantially by evaporation under vacuum during the second work-up (Figure 3.2b). The other issue that had to be addressed was the presence of the oil from the NaH suspension. To remove oil, NaH was washed with dry hexane prior to use which was thorough in the first attempt but not during the second.

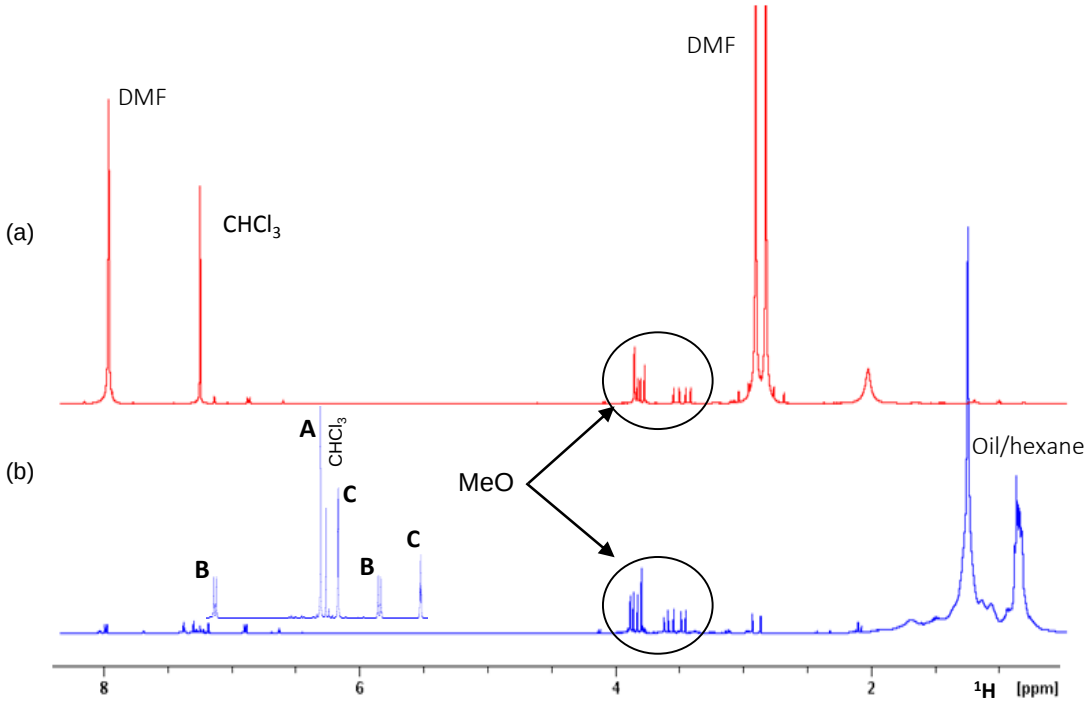


Figure 3.2 ^{1}H NMR spectra of two methylation attempts on the model mixture **I** (a) without DMF removal and oil removal (b) with DMF removal and unsuccessful oil removal. The residual solvent peaks and that of oil are labelled. The aromatic signals from model mixture **I** are labelled A-C on the inset. The methoxy signals are circled.

3.3.2. Methylation of the model mixture I using method B with unlabelled CH_3I

The methylation procedure outlined in Section 2.9.2 was performed on model mixture **I** (Figure 2.1). The methoxy region of the ^{1}H spectrum of the product is shown in Figure 3.3 compared with the result from method A. As can be seen from the spectra, method **B** (blue signals) produces additional signals in the methoxy sugar region. Seven rather than four signals were observed for the MeO groups of the Methyl β-D-xylopyranoside implying that 50% of the compound had only three hydroxyl groups methylated (signals 1, 2, 3 in Figure 3.3). This partial methylation is likely due to the weaker base used by this method. Furthermore, partial methylation is also indicated by the weaker signals of the aromatic methoxy groups of compound **A**. This molecule has three methoxy groups positioned ortho to one another. It is possible that the weaker base could not achieve full methylation of these groups.

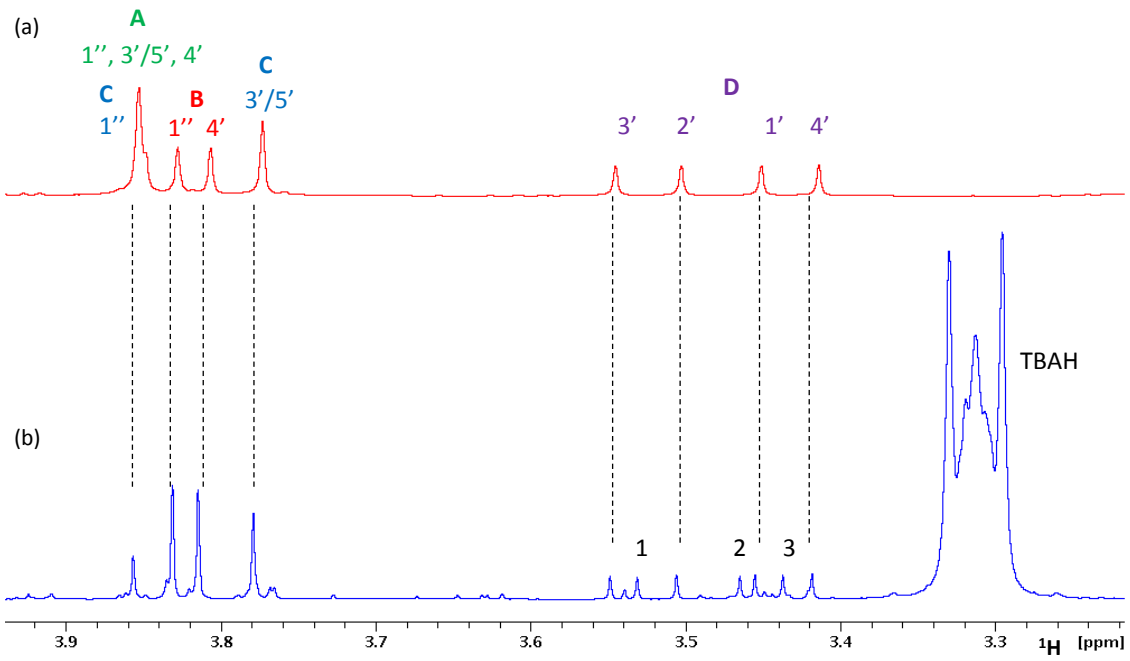


Figure 3.3 The methoxy region of the ^{1}H NMR spectra of the methylated model mixture **I** using (a) method A and (b) method B. The protons of molecules **A**, **B**, **C** and **D** are labelled in green, red, blue and purple, respectively. These colours as well as the resonance assignment follow that given in Figure 3.4. Additional signals of partially methylated molecule **D** are labelled 1-3. The signal of residual TBAH is also labelled.

In addition, complete removal of TBAH was not achieved, as indicated by the large signals of the base in the ^{1}H spectrum. A washing procedure used by Piccolo *et al.*[252] to remove the residual TBAH was also not successful. Considering the incomplete methylation as well as difficulties removing the phase catalyst this procedure was not considered further for the methylation reaction of HS samples.

3.3.3. Methylation of model mixture I using method A with ^{13}C-labelled CH_3I

It was evident from the ^{1}H spectra that method A was more successful than method B. Thus, this procedure was chosen to methylate the model mixture before using it on the HS samples.

The Method A procedure was finalised and repeated to ensure reproducibility and to produce a clean sample of model mixture **I** (Section 2.9.1) for the development of pulse sequences. The resulting ^{13}C decoupled and coupled ^{1}H NMR spectra, with the methylated molecules **A-D,** are shown in Figure 3.4. Note the colour scheme shown will be used throughout this section.

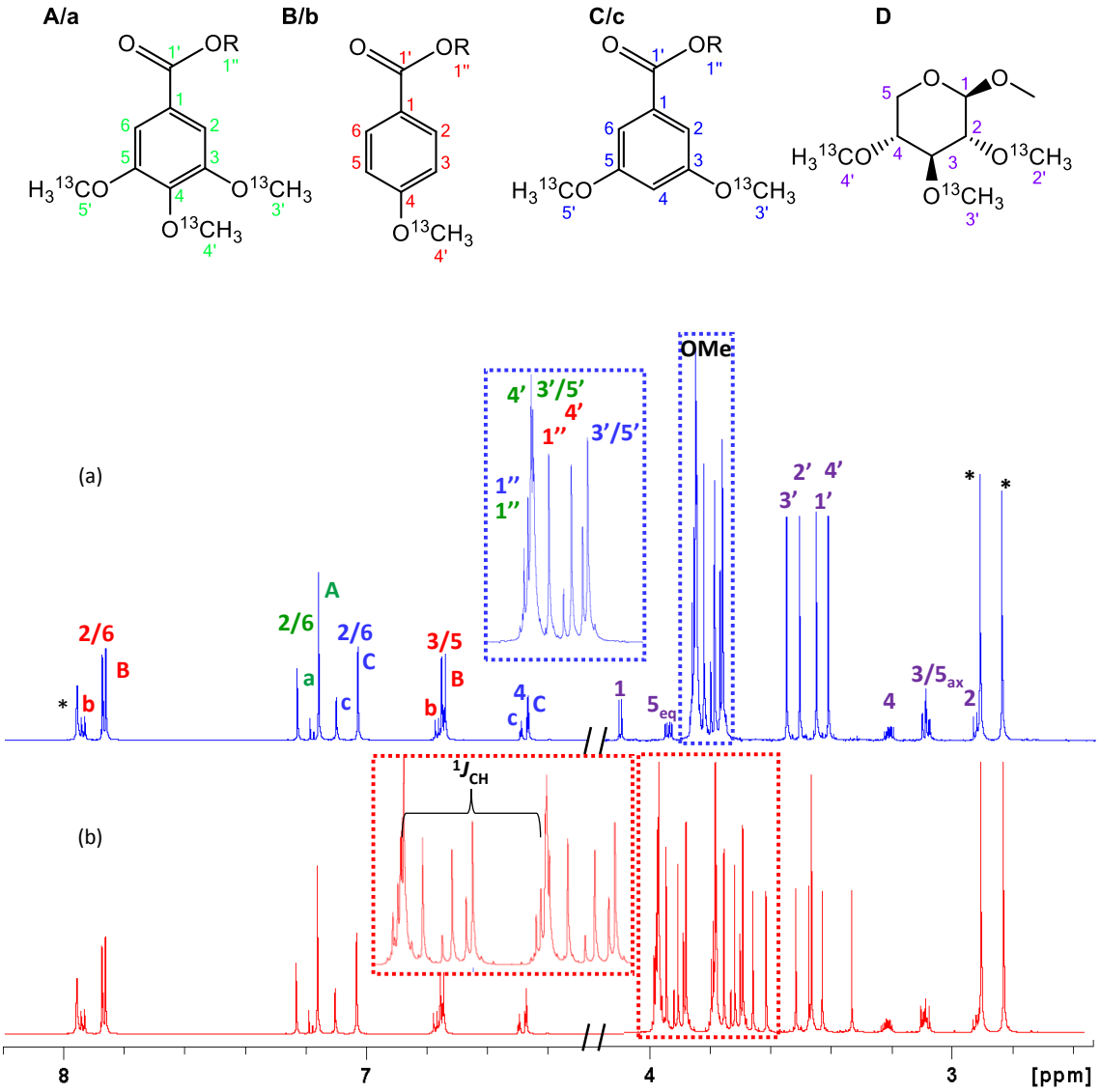


Figure 3.4 800 MHz (a) ^{13}C-decoupled and (b) ^{13}C-coupled ^{1}H spectra of ^{13}C-methylated model mixture **I**. Residual solvents DMF signals are labelled with an asterisk. The insets enclosed in dashed boxes show expanded regions of the methoxy protons which are split by their $^1J_{CH}$ coupling in (b). The fully methylated molecules (**A**-**D**) and partially methylated molecules (**a**-**c**) of mixture **I** are shown above the spectra. R indicates $O^{13}CH_3$ for **A**-**C** and OH for **a**-**c**. The colours used for the resonance labels will be used throughout this section to help differentiate signals from the different molecules of mixture **I**.

It is evident that this methylation attempt produced a more complicated mixture than observed previously. On closer inspection it was concluded that the heterogeneity is caused by partial hydrolysis of ester groups. Considering the methylation procedure as outlined in Section 2.9.1, this hydrolysis has occurred due to excess acid used during the neutralisation step. This result was not taken as a failure but instead the resulting mixture offered a more complex sample for the development of the nD NMR experiments. Therefore the fully methylated molecules in model mixture **I** are labelled **A**-**D**, while those partially methylated are labelled **a**-**c**. The methyl region

in the ^{1}H spectrum is overlapped; however the methoxy carbon region shows a better dispersion of signals (Figure 3.5).

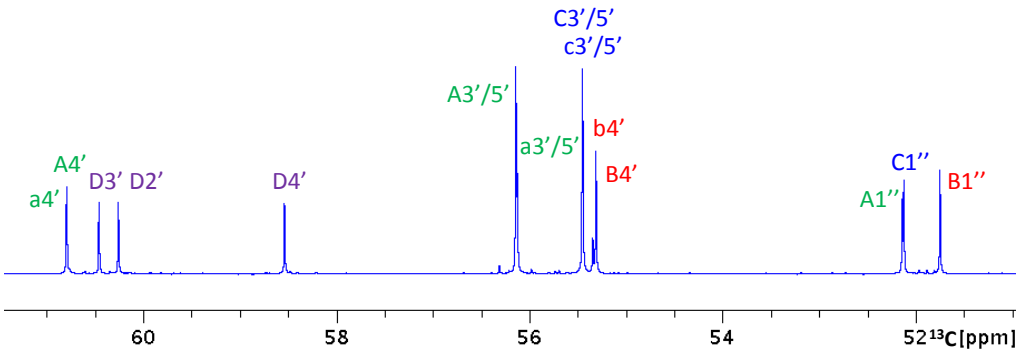


Figure 3.5 Methoxy region of the ^{13}C spectrum of the model mixture **I**. **A**-**D** and **a**-**c** represent the fully and partially methylated molecules. For labelling of the carbon resonances see Figure 3.4.

3.3.4. Methylation of model mixture II using method A and ^{13}C-labelled CH_3I

Model mixture **II** (see Figure 2.2) was methylated using method A with ^{13}C-labelled methyl iodide. To inspect the results of the methylation procedure, a ^{1}H spectrum was acquired on the methylated model mixture **II**. For clarity the aromatic (8 to 6.5 ppm) and methoxy (4.1 to 3.7 ppm) regions of the ^{1}H spectrum are shown separately (Figure 3.6a and b). The assignment of the ^{1}H resonances was only possible with the aid of 2D NMR spectra (see Section 6.2).

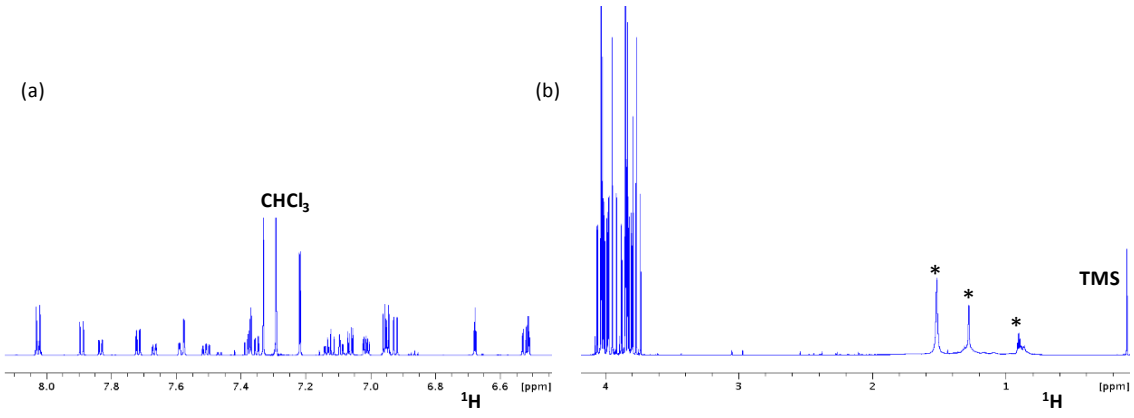


Figure 3.6 1D ^{1}H spectrum showing (a) aromatic region and (b) methoxy region of ^{13}C-methylated model mixture **II**. Impurities are labelled by an asterisk. Signals of the solvent $CHCl_3$ and an internal standard TMS are labelled.

Some impurities (marked with an asterisk) from the methylation reaction/extraction procedure are also visible (1.6 to 0.8 ppm). Some of these signals can be attributed to oil or residual hexane.

As these peaks appear away from the aromatic and methoxy signals of interest, they can be ignored.

In addition to the ^{1}H spectrum, a ^{13}C spectrum was also acquired (Figure 3.7).

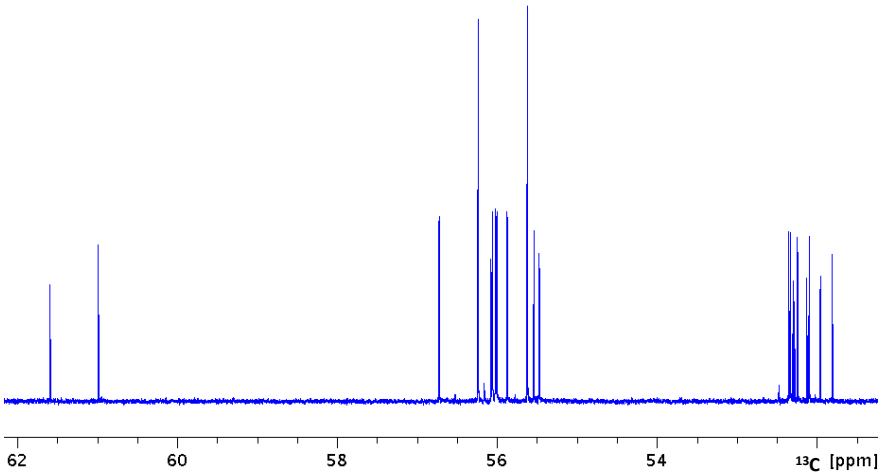


Figure 3.7 Methoxy region of the ^{13}C spectrum of ^{13}C-methylated model mixture **II**.

Focusing on its methoxy region, three distinct regions are clearly visible. Above 60 ppm there are two signals, which are due to methoxy groups that are sandwiched between two other methoxy groups. The signals between 57-55 ppm belong to the other phenolic methoxy carbons, while the signals of the ester methoxy carbons appear below 54 ppm.

3.3.5. Methylation of NEHA and RMFA using method A

After a series of experiments on model compounds, methylation using method A was carried out on two HS samples, NEHA and RMFA with unlabelled and ^{13}C labelled CH_3I, respectively (Section 2.9.3). This section will describe the results.

3.3.5.1. Methylation of NEHA using method A with unlabelled CH_3I

The outcome of the methylation was initially assessed by comparing FT-IR and ^{1}H NMR spectra before and after the methylation procedure (see Section 2.11 and 2.12). The FT-IR spectra before and after methylation of NEHA are shown in Figure 3.8a and b, respectively.

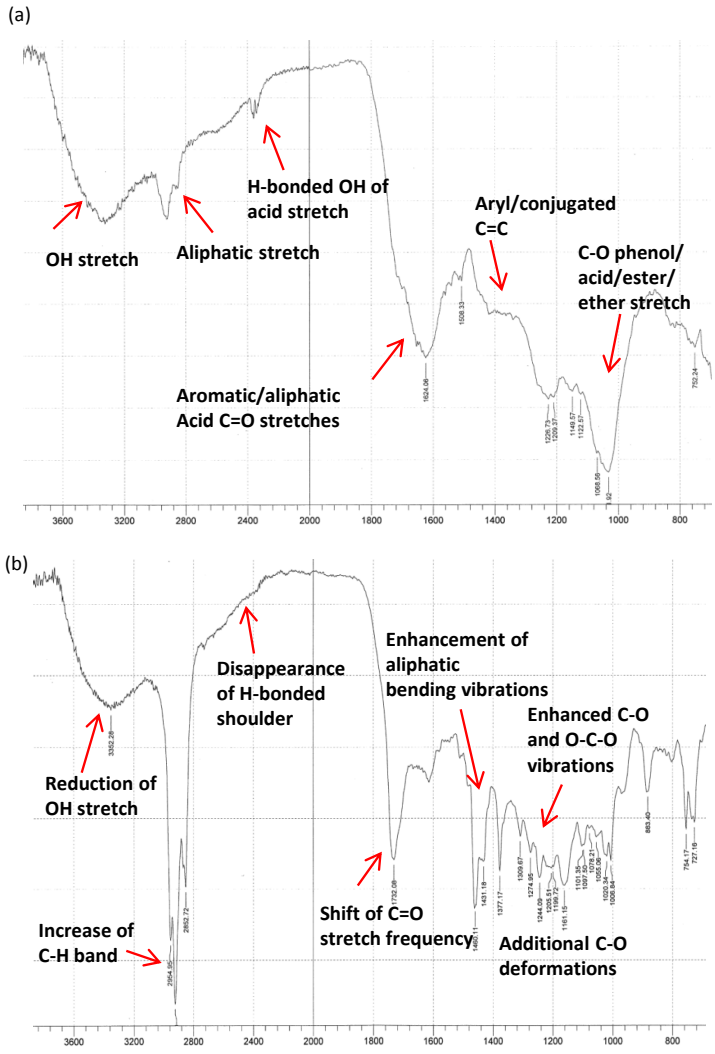


Figure 3.8 FT-IR spectra of NEHA (a) before and (b) after methylation using method A with unlabelled CH_3I. Spectra are annotated showing the typical functional groups and indicate changes induced by methylation.

The spectrum of methylated NEHA features broad peaks of ester, aromatic, phenol acids, aliphatic and alcohol functionalities.[252, 295] Changes indicated in Figure 3.8b, are typical for the conversion of carboxylic acid and alcohol groups into methyl esters and ethers, respectively. This can be seen by the increase in the intensity of the CH stretch (2954 cm^{-1}) due to the incorporation of CH_3 groups but also the shift of the C=O stretch from 1624 cm^{-1} to 1732 cm^{-1}. It should be noted that this band is not in the textbook 1735-1750 cm^{-1} range indicating that there may be conjugation occurring which lowers the expected carbonyl frequency or that the carbonyl groups are attached to aromatic systems e.g. p-methoxy benzoate C=O stretch is around 1714 cm^{-1}. There is also a marked decrease in carbohydrate associated bands (~1040 cm^{-1}), which could be a result of their removal during the chloroform extraction step. Of course the remaining signals in this region

could also be assigned to C-O deformations. FT-IR cannot reveal the exact nature of the groups present as bands can move depending on the nature of the molecule itself and thus may be misinterpreted. However, this method is a very useful and fast way to determine if methylation has taken place. The ^{1}H spectra of peat HA before and after methylation are shown in Figure 3.9.

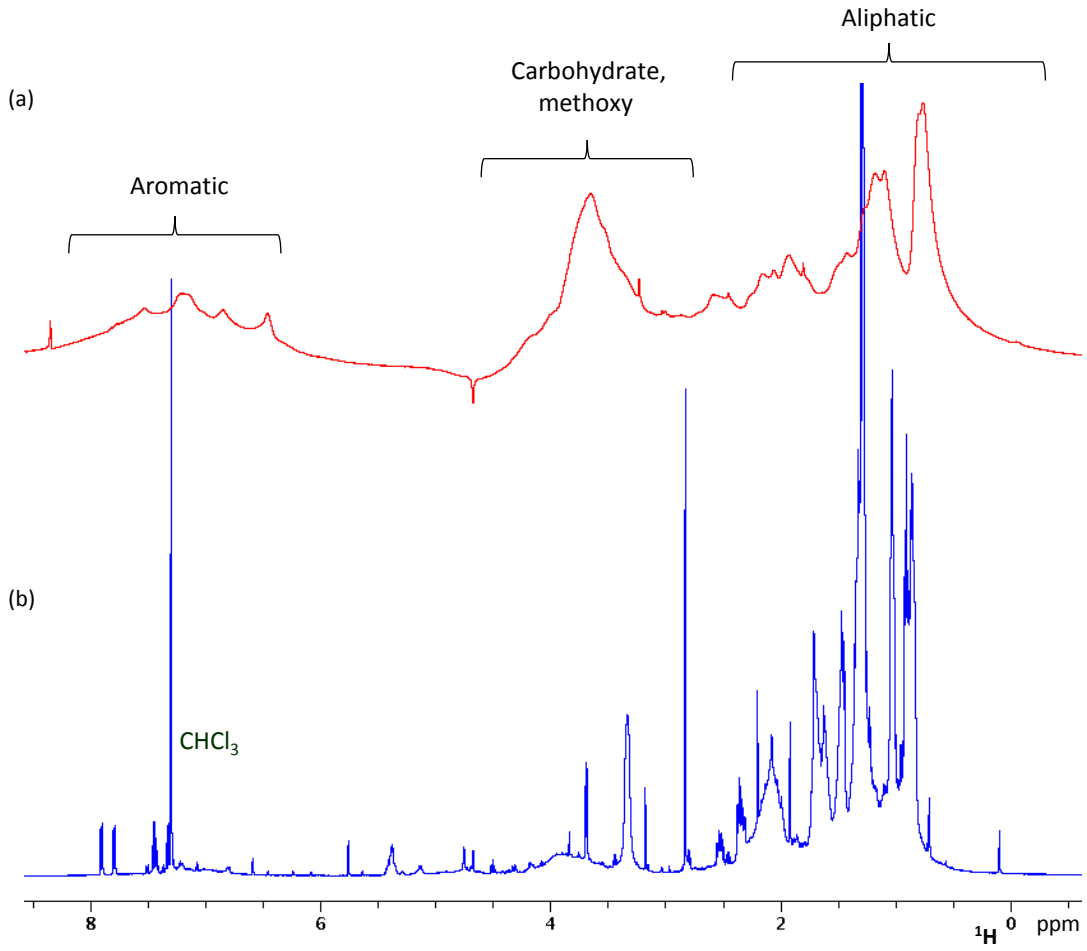


Figure 3.9 ^{1}H spectra of NEHA (a) before and (b) after methylation using method A with unlabelled CH_3I. Typical spectral regions are annotated on the top spectrum.

The ^{1}H spectrum before methylation (Figure 3.9a) shows typical broad bands assignable to aromatics, carbohydrates/methoxy and aliphatic compounds. The methylated spectrum (Figure 3.9b) is clearly different showing a number of more resolved resonances. Methylation is indicated by the appearance of unresolved signals around 3.7 ppm, which belong to methoxy protons. The spectrum obtained is similar but not exactly the same as that obtained from methylated lignite HA (not shown) by Piccolo *et al.*[252] This may be due to the different methylation procedure used or due to the fact that the sample used in the other study was from lignite. The spectrum contains

intense signals in the aliphatic region, in agreement with the increased intensity in the FT-IR spectrum. These presumably belong to nonpolar compounds that were extracted into chloroform. The amount of carbohydrates is clearly reduced in agreement with the FT-IR spectrum. The aromatic region of the methylated sample contains few resolved resonances that were assigned to phthalates. These have been previously identified in a soil FA sample by Simpson *et al.*[196] Phthalates are known plasticisers and common organic pollutants. These could originate from laboratory equipment, e.g. pipette tips or organic solvents. However, as phthalates were not observed in the ^{1}H spectra obtained from model mixture **I** or **II**, which were subjected to the same procedure as the HA sample, they must be part of the HA sample**.** In general, the low intensity of aromatic signals in this spectrum indicates that either (i) methylation of the phenols did not occur (ii) methylation did occur but only for a small number of the phenols or (iii) the sample did not dissolve properly in the aqueous phase. Repetition of the methylation procedure resulted in the same result, thus this sample was not deemed suitable for characterisation of the phenolics.

3.3.5.2. Methylation of RMFA using method A with ^{13}C-labelled CH_3I

To examine the extent of the ^{13}C-methylation of RMFA, a 1D ^{1}H NMR spectrum was acquired before (Figure 3.10a) and after methylation using method A with ^{13}C-labelled CH_3I (Figure 3.10b).

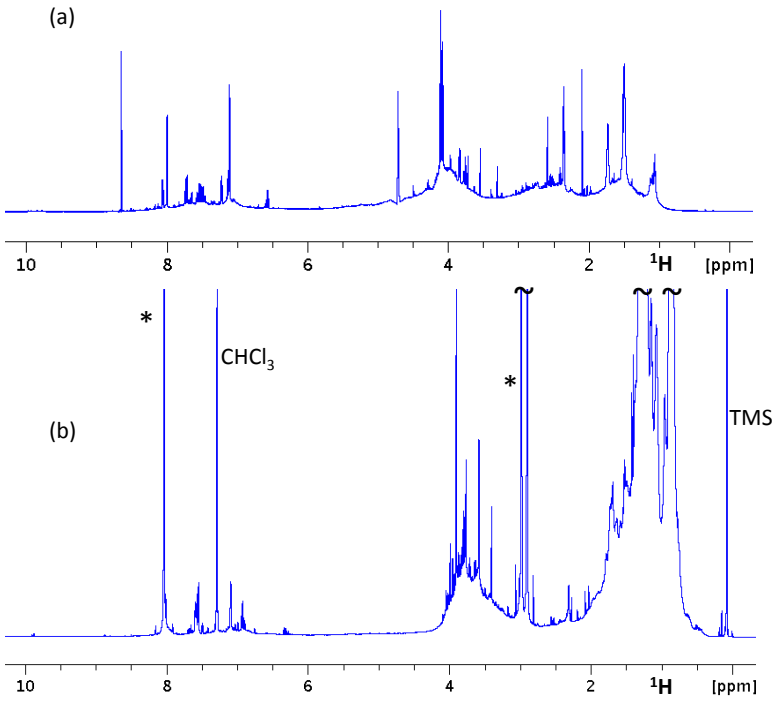


Figure 3.10 ^{1}H spectrum of RMFA (a) before and (b) after ^{13}C methylation. Signals of residual DMF are highlighted with an asterisk. Signals of the solvent $CHCl_3$ and internal standard TMS are labelled.

The spectrum of the ^{13}C-methylated sample shows increased intensity in the methoxy region compared with the non-methylated sample. It is evident that there is a series of more abundant molecules superimposed on a less abundant 'hill' in this region. There also appears to still be a lot of signals of aliphatic protons below 2 ppm that belong to other methylated molecules. Some of these could be due to C- and N-methylation. In addition, the methylation and extraction protocol has also removed the majority of the signals in the region between 3 and 2 ppm, which were likely due to polysaccharides.

The methoxy region of the ^{13}C NMR spectrum is shown in Figure 3.11.

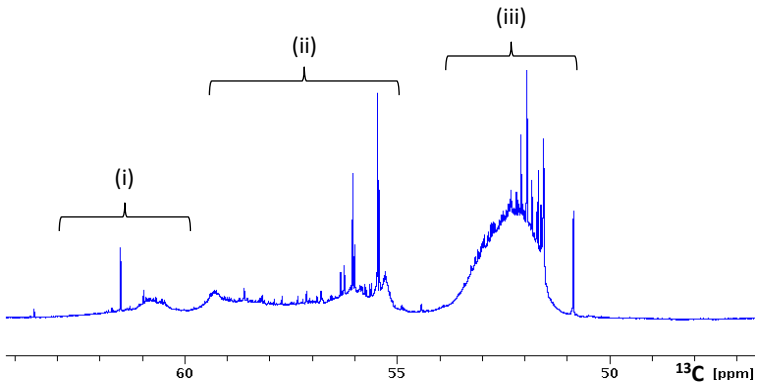


Figure 3.11 Methoxy region of a ^{13}C NMR spectrum of ^{13}C-methylated RMFA showing three regions: (i) aromatic methoxy groups sandwiched between to bulky substituents, (ii) aliphatic and aromatic methoxy ethers and (iii) aliphatic and aromatic methoxy esters.

This spectrum shows a large number of resonances that can be divided into three regions. The first, spanning from 50-54 ppm, represents aliphatic or aromatic methoxy esters. Sharp signals superimposed on top of a hump show that the sample contains a small number of major compounds and many minor compounds. The region spanning 55 to 60 ppm belongs to the aromatic, carbohydrate and aliphatic methoxy ethers. There is a much wider variation in these groups compared to the esters, again with a mixture of major and minor compounds. The 1D ^{13}C spectrum cannot distinguish between methoxy groups in various structural environments. In addition, there is a less abundant group of resonances above 60 ppm which represent aromatic or aliphatic methoxy groups that are sandwiched between two other functional groups. For example in tri-methoxybenzene the para methoxy carbon would show such a high methoxy chemical shift. Overall the initial assessment of the ^{1}H and ^{13}C NMR spectra indicated that this sample is suitable for further NMR investigation.

3.4. Chapter Conclusions

Two methylation procedures were tried and tested on model mixture **I**. While the method utilising the phase transfer catalyst TMAH is safer, it led to incomplete methylation and difficulties in removing signals of the base. Methylation involving CH_3I and NaH was therefore selected as the preferred method. After successful methylation of model mixture **I** both model mixture **I** and **II** were methylated using ^{13}C enriched CH_3I to provide samples for testing the NMR experiments. HS samples were successfully isolated from peat and peat pore water sources. These samples were inspected by ^1H NMR and from analysis of the phenolic region as well as the initial amount of sample extracted, NEHA and RMFA were chosen for further analysis. The methylation of NEHA using unlabelled CH_3I did not show persuasive evidence for high concentration of methylated phenolic compounds. RMFA was methylated using ^{13}C-labelled CH_3I producing a range of methoxy groups and a clear presence of methylated phenols. Thus, it was decided that this sample is suitable for further NMR analysis.

Chapter 4. Results and Discussion. NMR methodology development

800 MHz spectrometer, School of Chemistry University of Edinburgh

4.1. Design of NMR pulse sequences

A number of novel and existing pulse sequences were used in this study. These will be shown pictorially as well as described by spin product operators. These experiments utilise various couplings of ^{13}C-methylated phenolic compounds (Figure 4.1) taking advantage of the ^{13}C isotope labelling, which increases the sensitivity of the carbon nucleus hundred fold compared to natural abundance carbons.

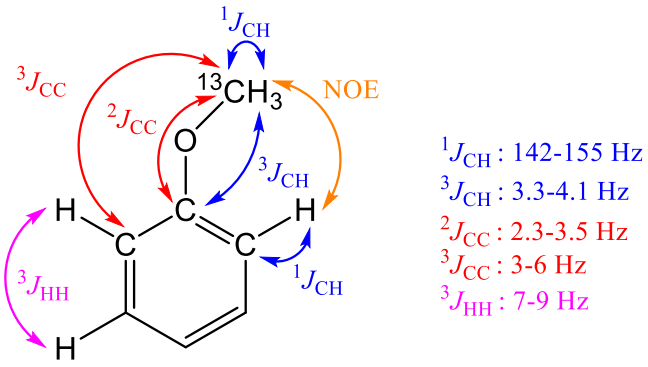


Figure 4.1 Polarisation transfer pathways using heteronuclear and homonuclear couplings from a ^{13}C-labelled methoxy group to the nuclei of the aromatic ring. The *J* couplings measured in this work are displayed on the right.

Importantly, the NMR experiments only pick up the molecules that carry the $^{13}CH_3$ groups. These groups hence become 'spies', probing the surrounding nuclei. The following sections describes the NMR experiments designed to utilise these 'spies'.

4.1.1. INEPT-INADEQUATE based experiments

2D INEPT-INADEQUATE was originally proposed by Otting *et al.*[296] Its purpose is to trace out pairs of ^{13}C coupled nuclei. This is normally quite an insensitive experiment due to the low natural abundance of ^{13}C. However, if one of the carbons is ^{13}C-enriched, the sensitivity of the INEPT-INADEQUATE can be compared to that of HSQC or HMBC experiments. Compared to the regular ^{13}C-detected INADEQUATE, the INEPT step boosts sensitivity due to the large ^{1}H polarisation and higher sensitivity of proton detection. If the signal is detected on methyl singlets then the major drawback of ^{1}H detection, splitting of ^{1}H multiplets, is absent. The modified version of the experiment used in this study is shown in Figure 4.2.

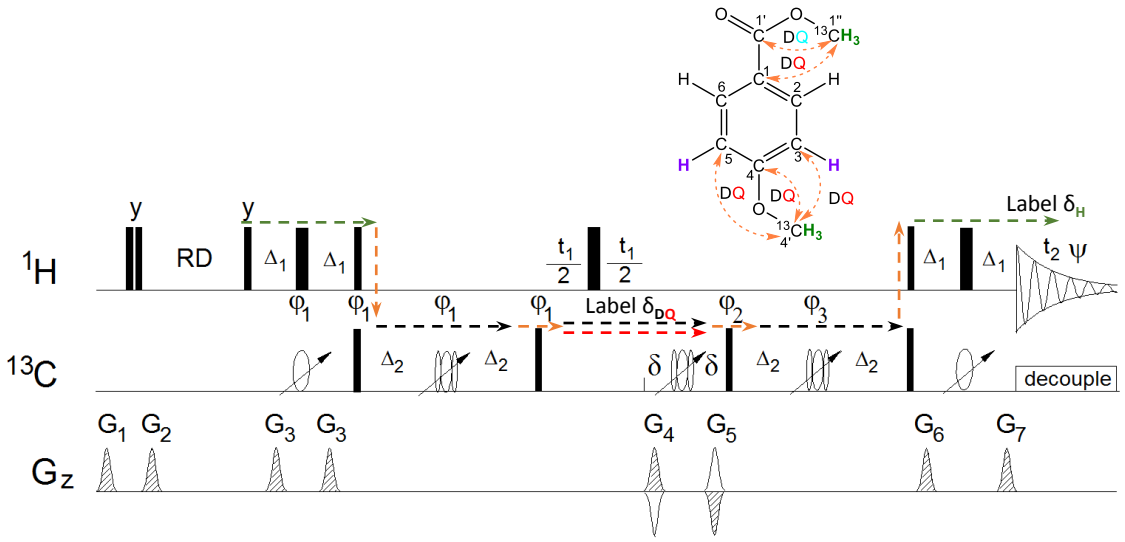


Figure 4.2 2D INEPT-INADEQUATE showing one polarisation transfer pathway. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. $\Delta_1 = 1/4^1J_{CH_3}$, $\Delta_2 = 1/4^nJ_{CC}$, δ = 1.2 ms. t_1 and t_2 indicate incremental periods during which chemical shift labelling occurs. The following phase cycle was used: φ_1 = x, y, -x, -y; φ_2 = 8x, 8(-x); φ_3 = 4x, 4(-x); Ψ = 2(x, -x, x, -x), 2(-x, x, -x, x). 1 ms PFGs were applied with the following strengths: G_4 = 40%, G_5 = -40%, G_6 = 3% and G_7 = 43.16% were used for N-type signal selection, while G_4 = -40%, G_5 = 40%, G_6 = 3% and G_7 = 43.16% were used for P-type signal selection. Spectra were processed using the echo-anti-echo protocol. 180° adiabatic inversion and refocusing CHIRP pulses were used on the carbon channel. The inset highlights the nuclei of which the chemical shifts are obtained. The green and purple boldfaced text represents methyl and aromatic protons, respectively. The green, black and black red arrows represent methyl proton, methyl carbon and DQ carbon magnetisation. Note DQ stands for Double Quantum, which is black/red or black/turquoise to indicate that the DQ correlation occurs between the methyl carbon and aromatic or carbonyl carbon, respectively.

The initial PFG-90°$_x$90°$_y$-PFG pulse sequence element at the end of acquisition ensures purging of ^{1}H magnetisation, which can survive from one scan to the next. The ensuing relaxation delay is then followed by an INEPT step, which creates anti-phase carbon magnetisation, $2\hat{I}_z\hat{S}_y$. This evolves during the subsequent carbon spin-echo under the influence of J_{CC} coupling between the carbon spins *S* and *R*, while the ^{13}C chemical shift is refocused creating $4\hat{I}_z\hat{S}_x\hat{R}_z$ coherences. These further evolve as summarised using product spin operators, considering only terms giving rise to signal. Intervals Δ_1 and Δ_2 were optimised as shown in Figure 4.2.

$$\mathbf{4\hat{I}_z\hat{S}_x\hat{R}_z} \xrightarrow{(\pi/2)\hat{S}_x,\hat{R}_x} \mathbf{-4\hat{I}_z\hat{S}_x\hat{R}_y} \xrightarrow{\Omega_{DQ}t_1\hat{S}_z,\hat{R}_z;\,\pi\hat{I}_x\,\pi\hat{S}_x,\hat{R}_x} [\![labels\ DQ]\!] \xrightarrow{(\pi/2)\hat{S}_x,\hat{R}_x} \mathbf{4\hat{I}_z\hat{S}_x\hat{R}_z}$$
$$\xrightarrow{2\pi J_{SR}(2\Delta_2)\hat{S}_z,\hat{R}_z;\,\pi\hat{S}_x,\hat{R}_x} \mathbf{\hat{I}_z\hat{S}_y} \xrightarrow{(\pi/2)\hat{S}_x;\,(\pi/2)\hat{I}_x} \mathbf{\hat{I}_y\hat{S}_z} \xrightarrow{2\pi J_{IS}(2\Delta_1)\hat{I}_z\hat{S}_z;\,\pi\hat{I}_x;\,\pi\hat{S}_x} \mathbf{\hat{I}_x}$$

This pulse sequence works for both $^1J_{CC}$ and $^nJ_{CC}$ simply by setting the Δ intervals to an appropriate value. In the ^{13}C-methylated sample the magnetisation starts on the ^{13}C labelled methyl protons. The INEPT step transfers the magnetisation to the methyl carbons. The magnetisation is then left to evolve under $^{2,3}J_{CC}$ of coupled carbons followed by the creation of DQ

coherences, which are labelled during t_1. The DQ coherences are selected by PFGs and then brought back into observable proton magnetisation for detection during t_2. A parallel pathway starts on protonated aromatic carbons if an $^nJ_{CC}$ exists between these carbons and the methoxy carbons.

4.1.1.1. 3D INEPT-INADEQUATE-HSQC

In order to analyse an INEPT-INADEQUATE spectrum, one has to know at least one of the SQ chemical shifts per pair of ^{13}C coupled spins. This normally means acquiring a separate ^1H, ^{13}C correlating experiment such as an HMQC or HSQC. For one molecule or a simple mixture this is adequate, however for complex mixtures this is not possible as unambiguous assignments cannot be made. Thus a 3D NMR experiment was created to combine both the DQ and SQ labelling in one experiment to resolve this issue.[297]

This involved a simple modification of the 2D INEPT-INADEQUATE. The SQ ^{13}C coherences are labelled in a constant-time manner during the second ^{13}C spin-echo. To prevent the modulation of the carbon magnetisation by $^1J_{CH}$ couplings caused by a moving ^{13}C 180° pulse, a 180° ^1H pulse is applied which follows the movement of the ^{13}C pulse. A similar approach has thus far been applied only to one-bond carbon-carbon correlations and has appeared under various names.[297] It was decided to refer to this experiment here as 3D INEPT-INADEQUATE-HSQC.

The pulse sequence, labelled nuclei and polarisation transfer pathway of this experiment are shown in Figure 4.3.

Figure 4.3 illustrates the pathway from methyl proton, INEPT transfer to methyl carbon, creation of a DQ coherence by the 90° carbon pulse. The DQ coherence is then labelled during t_1. The SQ methyl carbon coherence is restored by the 90° carbon pulse and labelled during a constant-time period $2\Delta_2$ during t_2. Here also the carbon-carbon coupling is refocused while the $^1J_{CH}$ is not evolving, keeping the anti-phase proton-carbon magnetisation. This allows the reverse INEPT step to transfer the magnetisation back to the methyl protons, which are labelled during t_3. Figure 4.3 indicates the DQ correlation is created between the methyl proton and aromatic carbon but it also occurs between the methyl proton and carbonyl group of the methyl esters. The resulting spectrum is a 3D cuboid with proton, carbon and DQ dimensions. However as there is no frequency selection, a pathway also exists which starts and ends on aromatic proton. This experiment can thus be thought of as a pseudo 4D experiment.

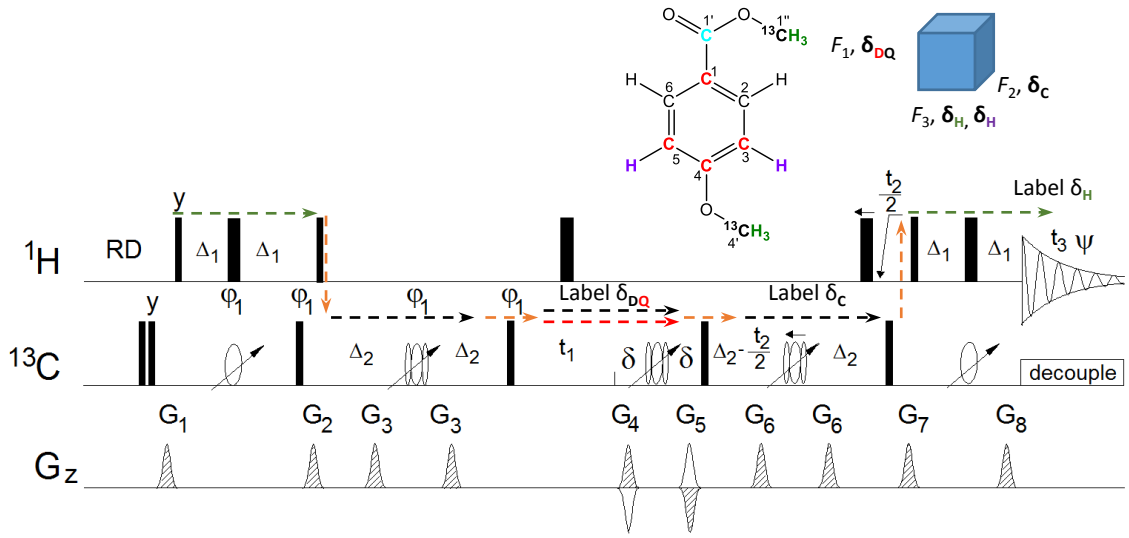


Figure 4.3 3D INEPT-INADEQUATE-HSQC showing one possible polarisation transfer pathway. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. $\Delta_1 = 1/4^1J_{CH_3}$, $\Delta_2 = 1/4^nJ_{CC}$, $\delta = 1.2$ ms. t_1 and t_2 indicate incremental periods during which chemical shift labelling occurs. The following phase cycle was used: φ_1 = x, y, -x, -y; Ψ = x, -x, x, -x, -x, x, -x, x. 1 ms PFGs were applied with the following strengths: G_1 = 18%, G_2 = 37%, G_3 = 9%, G_6 = 5%, G_7 = 70% and G_8 = 35.14%. G_4 = 35% and G_5 = -35% were used for N-type signal selection, while G_4 = -35% and G_5 = 35% were used for P-type signal selection. Spectra were processed using the echo-anti-echo protocol. 180° adiabatic inversion and refocusing CHIRP pulses were used on the carbon channel. The inset highlights the nuclei of which the chemical shifts are obtained. The green, black, purple, red and turquoise boldfaced text and arrows represents the methyl proton, methyl carbon, aromatic proton, aromatic carbon and carbonyl carbons, respectively. Note DQ stands for Double Quantum. In the figure, DQ is black/red or black/turquoise to indicate the DQ correlation occurs between methyl carbons and aromatic or carbonyl carbons, respectively. The inset shows the 3D cuboid obtained from this experiment.

Several modifications were made to the 3D INEPT-INADEQUATE-HSQC experiment in order to deal with issues specific to the samples used in this study. The first change was the inclusion of an IPAP module during the final refocusing spin-echo (Figure 4.4).

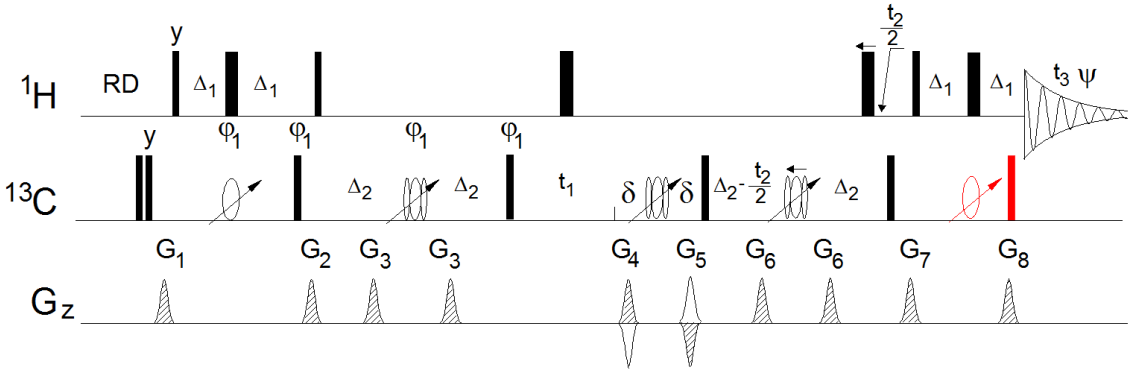


Figure 4.4 3D IPAP-INEPT-INADEQUATE-HSQC. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. Parameters used are as in Figure 4.3. The segment coloured in red is present only when acquiring the in-phase (IP) spectrum.

In this modified sequence the ^{13}C decoupling is removed and instead In-Phase or Anti-Phase (or IPAP) CH doublets are acquired. The AP spectrum is acquired by omitting the ^{13}C 180° pulse during the second $2\Delta_1$ period, while the IP spectrum requires a purging 90° ^{13}C pulse to be added at the end of a regular INEPT spin-echo to remove any residual anti-phase magnetisation. This modification was explored in order to increase the resolution in the directly detected dimension. As the required lengthening of the acquisition time is not allowed on cryoprobe instruments in the presence of ^{13}C decoupling, two IPAP spectra yielding in-phase and anti-phase doublets are acquired. The two spectra are then added, subtracted and recombined by moving the signals into their chemical shift positions. This is only possible if the $^1J_{CH}$ coupling constants are uniform. This condition is fulfilled separately for phenolic methoxy groups and methoxy esters and therefore this method can be utilised here. The spread of the $^1J_{CH}$ aromatic couplings prevents it to be used for this region of the spectrum. The signal with 50% intensity is thus contained in two spectra, which gives the possibility to decide between the signal and the noise by comparing the two spectra.

The second modification concerned optimisation of the spectral widths in the DQ and SQ dimensions. It was decided to set the ^{13}C *RF* carrier frequency to the middle of the relevant aromatic ^{13}C resonances (~110 ppm) and fold the $^{13}CH_3O$ signals. This meant that the signals in the DQ dimension would require spectral width of 150 ppm to prevent signal aliasing, even though the signal only appears in about half of this frequency range. In order to avoid sampling empty space it is possible to effectively move the DQ frequencies to the centre of a much narrower spectral window. This can be done if the pulses preceding the incrementable period t_1 are shifted in phase every time t_1 is incremented. In the case of DQ coherences, the relationship between the correct and shifted frequencies takes the following form:

$$\mathrm{DQ\ (final)} = \mathrm{DQ\ (orig)} + \mathrm{sw}\ (F_1\mathrm{final})/(\tfrac{180}{x}) \qquad [12]$$

Where x is an arbitrary phase shift in degrees. Using this equation it is possible to determine the original DQ frequencies. This procedure is illustrated in Figure 4.5.

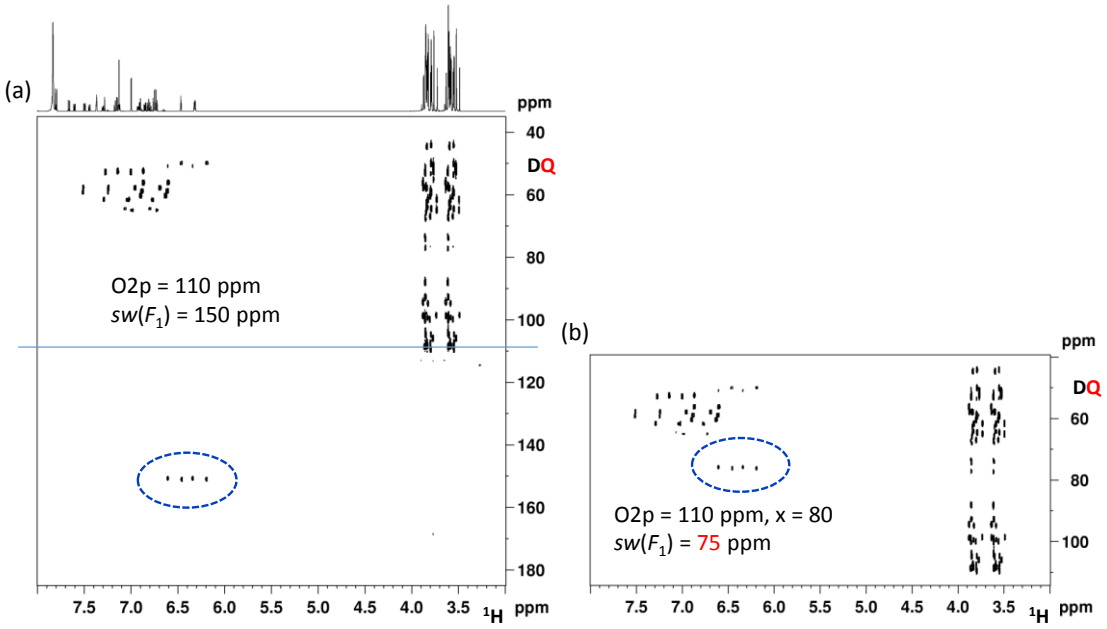


Figure 4.5 Optimising the spectral width of the DQ dimension. The first F_1F_3 DQ plane of the 3D IP INEPT-INADEQUATE-HSQC spectrum of the model mixture **II** acquired without the ^{13}C decoupling. The ^{13}C carrier was set at 110 ppm (blue line); this value represents zero DQ frequency. The t_1 acquisition time was 2.8 ms. (a) Spectral width was set to 150 ppm; no incrementation of the φ_1 phase was applied. (b) Spectral width was set to 75 ppm and φ_1 was incremented by 80° every time the t_1 interval was increased. This represents the optimal sampling of the DQ dimension without the need for signal aliasing. The exception are the circled signals, which will be discussed in Section 6.2.1.

The third modification involved the suppression of signals from methylated non-aromatic molecules. This is achieved by modifying the defocusing period $2\Delta_2$ of the 3D INEPT-INADEQUATE-HSQC and inverting selectively frequencies outside of the aromatic and methoxy regions of the carbon spectra, by pulses A and B, respectively (Figure 4.6a). Pulses A and B were chosen to cover the 75 to 85 and 25 to 35 ppm ranges as shown in Figure 4.6b.

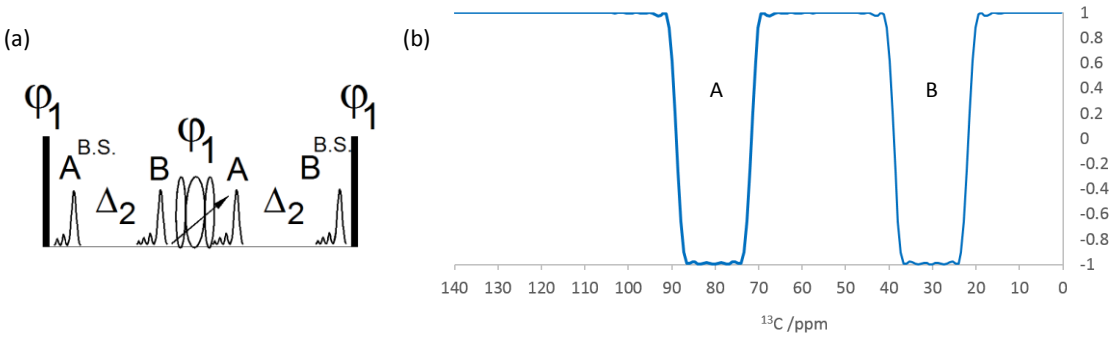


Figure 4.6 The defocusing $2\Delta_2$ interval of the 3D INEPT-INADEQUATE-HSQC with 180° pulses inverting the regions shown in (b). B.S signifies the Bloch-Siegert shift compensating pulses. A and B denote the inverted regions.

The pulses labelled B.S. remove the Bloch-Siegert shifts. Since there is also a composite 180° pulse in the middle of this spin-echo, the J_{CC} couplings of the selectively inverted carbons with the

methoxy carbons will not evolve, hence their signals will be suppressed (Figure 4.6b). This modification mainly supresses the signals of methylated carbohydrates.

The result of this suppression is shown in Figure 4.7.

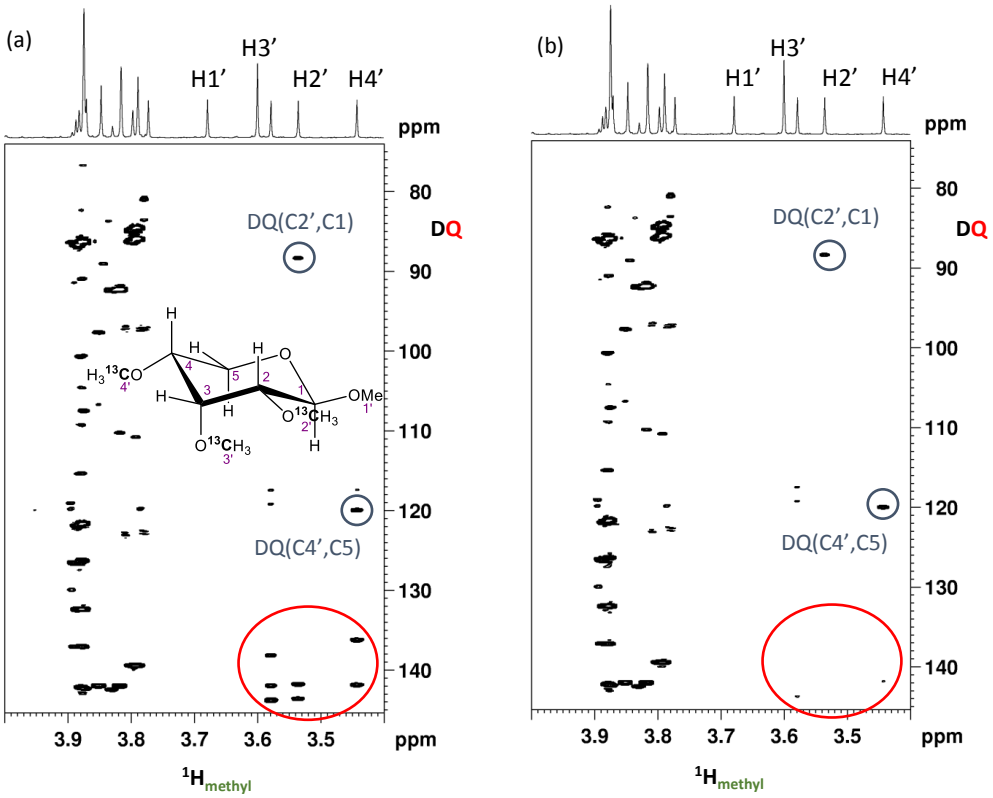


Figure 4.7 Illustration of the signal suppression of non-aromatic methylated compounds using a mixture **I** including a ^{13}C-methylated methyl β-D-xylopyranoside (Compound **D**). The carbon chemical shifts of **D** are 104.6 (C1), 84.9 (C3), 83.1 (C2), 79.3 (C4) and 63.0 (C5). (a) and (b) show the methyl region of the first DQ plane of the 3D IPAP INEPT-INADEQUATE-HSQC experiment with no suppression and with the suppression of aliphatic resonances, respectively. The red circle highlights the signals that were suppressed by the procedure.

The residual signal intensity in the suppressed spectrum was 6 ± 4% of the original signals. The C2'C1 and C4'C5 DQ coherences (highlighted in blue in Figure 4.7) were not suppressed at all, as C1 and C5 chemical shifts are outside of the inversion band of the 1.5 ms IBURP-2 pulses applied at 80 and 30 ppm. The intensity of the DQ coherences involving the aromatic OMe carbons (cross-peaks on the left side of the spectra) were not affected at all.

4.1.2. $HCCH_3$ experiments

The 3D INEPT-INADEQUATE-HSQC experiment provides chemical shifts of the atoms in the ipso and ortho positions to the methoxy groups. However as aromatic protons are generally multiplets, while the methoxy protons are always singlets with the intensity of three protons, the

superior S/N of methoxy proton signals was noticeable in this experiment. Therefore a set of experiments was designed to provide the chemical shifts of the aromatic CH pairs utilising a different polarisation transfer pathway and the detection of the signal on methyl protons. This approach naturally leads to a 4D experiment that was named according to the polarisation transfer pathway used, as 4D HCCH$_3$. In the following, this experiment is discussed.

The 4D HCCH$_3$ correlates aromatic protons adjacent to methoxy groups with their directly bonded carbon and the nuclei of the methoxy groups. The pulse sequence of the 4D HCCH$_3$ including the polarisation transfer pathway and the sampled nuclei are shown in Figure 4.8.

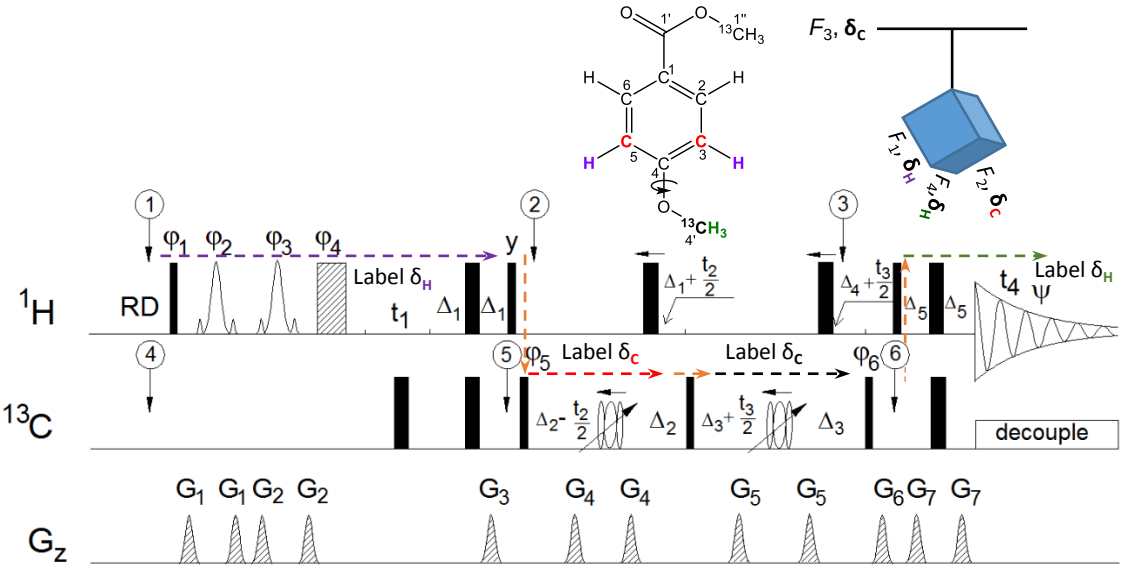


Figure 4.8 4D HCCH$_3$ with polarisation transfer used to obtain the highlighted nuclei on the molecule shown. The green, black, purple and red boldfaced text and arrows represent the methyl protons, methyl carbons, aromatic protons and aromatic carbons, respectively. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. 1 ms ^{1}H r-SNOB pulses applied during the initial DPFGSE select aromatic ^{1}H resonances (labelled during t_1). The aromatic carbons, methyl carbons and methyl protons are labelled in t_2, t_3 and t_4, respectively. The following parameters were used: $\Delta_1 = 1/4^1J_{CHarom}$, $\Delta_2 = 1/4^3J_{CC}$, $\Delta_3 = 1/4^3J_{CC}$, $\Delta_4 = 0.0915/^1J_{CH3}$, $\Delta_5 = 1/4^1J_{CH3}$. The following phase cycle was used: φ_3 = 2(x), 2(y), φ_5 = x, -x, ψ =(x, -x, -x, x). Phases φ_1 and φ_5 were incremented according to the States-TPPI protocol. If phase shifting applied in t_1, t_2, t_3 the phases φ_1 *to* φ_4 and φ_5 and φ_6 were phase shifted, respectively. 1 and 0.6 ms gradient pulses were applied at the following strengths: G_1 = 11%, G_2 = 50%, G_3 = 23%, G_4 = 8%, G_5 = 17%, G_6 = 33% and G_7 = 10%. In order to reduce the off-resonance effects the carrier frequencies were changed at the specified places, indicated by balloon topped arrows in the pulse sequence. ①= 7.15 ppm, ②= 5.52 ppm, ③= 3.8 ppm, ④= 110 ppm, ⑤= 96 ppm, ⑥= 56 ppm. Composite CHIRP 180° pulses used for refocusing periods. Trim pulse were used artefact suppression. The inset shows the resulting 4D spectrum as cuboids on a string.

Using product spin operators the polarisation transfer pathway of the 4D HCCH$_3$ experiment after the initial DPFGSE can be summarised as shown below. Intervals Δ_1-Δ_4 were optimised as shown in Figure 4.8. Only the terms leading to useful signal are shown.

$$-\hat{\boldsymbol{I}}_{\boldsymbol{y}}(arom) \xrightarrow{\Omega_I t_1 \hat{I}_z;\, \pi\hat{S}_x} [\![labels\ \delta_I]\!] \xrightarrow{INEPT\ (2\Delta_1)} \mathbf{2}\hat{\boldsymbol{I}}_{\boldsymbol{z}}\hat{\boldsymbol{S}}_{\boldsymbol{y}}(arom)$$

$$\xrightarrow{2\pi J_{SR}(CT2\Delta_2)\hat{S}_z\hat{R}_z;\, 2\pi J_{IS}(CT2\Delta_2)\hat{I}_z\hat{S}_z;\, \pi\hat{I}_x;\, \pi\hat{S}_x} [\![evolve\ \ {}^3J_{CC} + labels\ \delta_S]\!] \rightarrow \mathbf{2}\hat{\boldsymbol{S}}_{\boldsymbol{y}}\hat{\boldsymbol{R}}_{\boldsymbol{z}} \xrightarrow{(\pi/2)\hat{S}_y,\hat{R}_y} -\mathbf{2}\hat{\boldsymbol{S}}_{\boldsymbol{z}}\hat{\boldsymbol{R}}_{\boldsymbol{y}}$$

$$\xrightarrow{2\pi J_{SR}(CT2\Delta_3)\hat{S}_z\hat{R}_z;\, 2\pi J_{IR}(CT2\Delta_3)\hat{I}_z\hat{R}_z;\, \pi\hat{I}_x;\, \pi\hat{R}_x} [\![evolve\ \ {}^3J_{CC} + labels\ \delta_R]\!] \rightarrow \mathbf{2}\hat{\boldsymbol{I}}_{\boldsymbol{z}}\hat{\boldsymbol{R}}_{\boldsymbol{y}} \xrightarrow{Reverse\ INEPT\ (2\Delta_5)} \hat{\boldsymbol{I}}_{\boldsymbol{x}}(methyl)$$

The polarisation transfer of the 4D $HCCH_3$ experiment starts on aromatic protons whose chemical shift is labelled during t_1. The magnetisation is then transferred to the directly bonded aromatic carbons which are labelled in t_2. The magnetisation passes to the methoxy carbon for chemical shift labelling during t_3 and finally to the methyl protons for detection. Carbon decoupling over a narrow range of methyl carbons allows low decoupling power, hence longer acquisition times and better resolution. Chemical shifts of all four of these nuclei are sampled, leading to a 4D experiment. The selection of aromatic protons is achieved via a DPFGSE, which removes all but the aromatic protons with high efficiency. It also eliminates phase errors across the inverted frequency range while allowing any shape pulse to be used for inversion. It is therefore preferable to a shorter SPFGSE sequence. The r-SNOB pulses were chosen due to their near 'top-hat' profile.

The 3D INEPT-INADEQUATE-HSQC can be classified as an out-and-back experiment, while in the 4D $HCCH_3$ the flow of the magnetisation is unidirectional. Some signal is lost in the 4D $HCCH_3$ experiment during the preparation of the anti-phase carbon magnetisation of the CH_3 groups which can only be quantitative for CH groups. The 3D INEPT-INADEQUATE-HSQC experiment however also suffers signal losses due to DQ filtration and leakage of the signal into carbon- carbon ZQ coherences.[298] Hence, the sensitivity of the two techniques is comparable as is confirmed Section 6.4. In order to maximise the digital resolution of the 4D spectrum, folding and shifting of the resonances as described in Section 4.1.1 was invoked.

By fixing one of the ^{13}C chemical shift labelling periods, two 3D versions of this 4D experiment can be created: 3D $HcCH_3$ and 3D $hCCH_3$, which only sample the chemical shifts of the aromatic protons or aromatic carbons, respectively. These experiments have the potential advantage of higher sensitivity and achievable resolution but suffer from reduced information content. The benefit of the 4D version is that it correlates chemical shifts from four nuclei, which is not easily done from two separate 3D spectra in cases of severe overlap.

4.1.3. HMBC-based experiments

The HMBC experiment maps the long-range correlations, typically via $^{2}J_{CH}$ and $^{3}J_{CH}$ coupling constants. This section introduces the basic HMBC experiment, followed by the ^{13}C-filtered HMBC experiment designed for use on ^{13}C-methylated mixtures. This experiment can easily be converted into a 3D HMQC-HMBC, the pulse sequence discussed last.

The basic 2D HMBC experiment takes the form shown in Figure 4.9 where the first 90° carbon pulse removes the one-bond correlations, while the second converts the anti-phase magnetisation of the long-range coupled protons into DQ heteronuclear coherences.

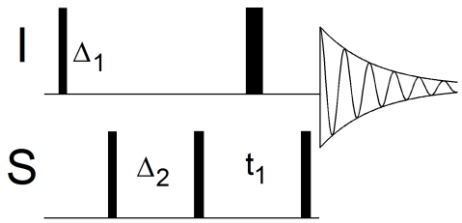


Figure 4.9 HMBC pulse sequence with low pass *J* filter. The thin and thick rectangles represent 90° and 180° pulses, respectively. t_1 is the incremental period during which chemical shift labelling occurs. The Δ_1 and Δ_2 delays are optimised for $^{1}J_{CH}$ and $^{n}J_{CH,}$ respectively.

The ^{13}C chemical shifts are labelled during t_1, followed by a conversion of DQ coherences into SQ proton magnetisation, which is typically acquired without ^{13}C decoupling**.**

4.1.3.1. ^{13}C-filtered HMBC

Considering the methylated mixtures, the HMBC is an excellent method for mapping the correlations between the methoxy protons and aromatic carbons at ipso positions. In order to select magnetisation of only protons directly attached to ^{13}C, an X-filter was placed at the beginning of the HMBC pulse sequence and the low pass filtered was removed. The mixed-phase nature of the HMBC cross-peaks is not desirable for complex mixtures as it degrades the resolution and can lead to cancellations of partially overlapped signals. This was avoided by designing a refocused HMBC experiment. Because the methoxy protons are singlets, no J_{HH} modulation is taking place and pure in-phase signals are recorded. Refocusing of the $^{n}J_{CH}$ couplings also allow the use of ^{13}C decoupling to enhance the sensitivity of the experiment. This experiment also yields excellent resolution. The pulse sequence for the ^{13}C-filtered HMBC is shown in Figure 4.10.

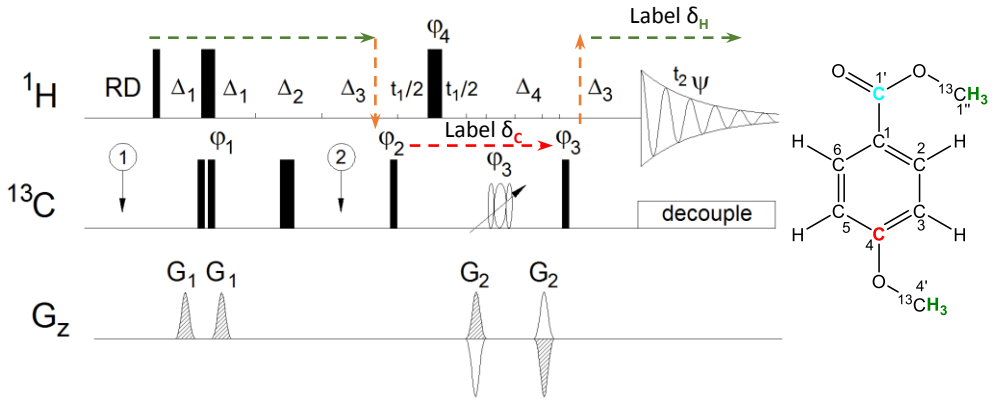


Figure 4.10 2D ^{13}C-filtered refocused HMBC. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. The following parameters were used: $\Delta_1 = 1/2{}^1J_{CH_3}$, $\Delta_2 = (2\delta_G + t_{180}(C_{adia}) + 2\,t_1{}^0 + t_{180}(^1H))$, $\Delta_3 = 1/2{}^nJ_{CH}$ (even multiple of Δ_1), $\Delta_4 = t_{180}(^1H) + 2t_1{}^0$. t_1 and t_2 indicates the incremental period during which chemical shift labelling occurs. The following phase cycle was used: φ_1 = 2(x), 2(-x), φ_2 = x,-x, φ_3 = 4(x), 4(-x), φ_4 = 8(x), 8(-x), ψ = x, -x, -x, x, -x, x, x, -x. Gradient pulses were applied at the following strengths: G_1 = 30% and G_2 = ± 80%. Quadrature detection in t_1 was achieved using echo, anti-echo protocol. 180° adiabatic CHIRP pulses were used to invert the wide bandwidth of ^{13}C resonances. To reduce the off-resonance effects the carrier frequencies were changed at specified places, indicated by balloon topped arrows: ①= middle of methyl carbons and ②= middle of aromatic and methyl carbons. Decoupling was applied during acquisition. The inset highlights the nuclei of which the chemical shifts are obtained in this experiment. The green, purple and red boldfaced text and arrows represent the methyl protons, aromatic protons and aromatic carbons, respectively.

An extra Δ_2 interval with a 180° ^{13}C pulse applied mid-way was required to ensure proper timing. A complication arose from the fact that the protons of interest are also attached to ^{13}C hence the effective evolution of $^1J_{CH_3}$ couplings during the long-range evolution intervals Δ_3 was minimised by setting this interval to an even multiple of 1/2 $^1J_{CH_3}$. This was possible because of nearly uniform values of $^1J_{CH_3}$ couplings within functional group types. For 9 model compounds the mean aromatic methoxy $^1J_{CH_3}$ was measured as 144.59 ± 0.34 Hz, while the mean ester methoxy $^1J_{CH_3}$ was 147.05 ± 0.21 Hz.

The ^{13}C-filtered refocused HMBC can be described by spin-product operators using the Δ intervals given in Figure 4.10 and ignoring the time keeping delays Δ_2 and Δ_4:

$$-\hat{\boldsymbol{I}}_y(methoxy) \xrightarrow{2\pi J_{IS}\Delta_1\hat{I}_z\hat{S}_z;\ \pi\hat{I}_x;\ \pi S_x or\ 0\hat{S}_x} \pm 2\hat{\boldsymbol{I}}_x\hat{\boldsymbol{S}}_z \xrightarrow{2\pi J_{IS}\Delta_1\hat{I}_z\hat{S}_z} \pm\hat{\boldsymbol{I}}_y \xrightarrow{2\pi J_{IR}\Delta_3\hat{I}_z\hat{R}_z} \pm 2\hat{\boldsymbol{I}}_x\hat{\boldsymbol{R}}_z \xrightarrow{(\pi/2)\hat{R}_x}$$

$$\pm 2\hat{\boldsymbol{I}}_x\hat{\boldsymbol{R}}_y \xrightarrow{\Omega_I t_1 R_z;\ \pi\hat{I}_x;\ \pi R_x} [\![labels\ \delta_R]\!] \xrightarrow{(\pi/2)R_x} \pm 2\hat{\boldsymbol{I}}_x\hat{\boldsymbol{R}}_z \xrightarrow{2\pi J_{IR}\Delta_3\hat{I}_z R_z} \pm\hat{\boldsymbol{I}}_y(methoxy)$$

The ^{13}C-filtered HMBC selects the ^{13}C-labelled protons, transfers the magnetisation to the $^nJ_{CH}$ coupled carbons and labels their chemical shifts in F_1. The magnetisation is then passed back to

the protons of the methyl groups, which are detected under ^{13}C decoupling in the F_2 dimension as pure-phase singlets.

4.1.3.2. 3D HMQC-HMBC

In order to complement the 3D INEPT-INADEQUATE-HSQC experiment a 3D HMQC-HMBC experiment was designed to correlate the methoxy group nuclei with the ipso aromatic carbons. Due to its simplicity and relatively uniform and large $^3J_{CH_3, C_i}$ coupling constants (3.3 - 4.1 Hz) this experiment is expected to be more sensitive than the 3D INEPT-INADEQUATE-HSQC. The pulse sequence of the 3D HMQC-HMBC is shown in Figure 4.11.

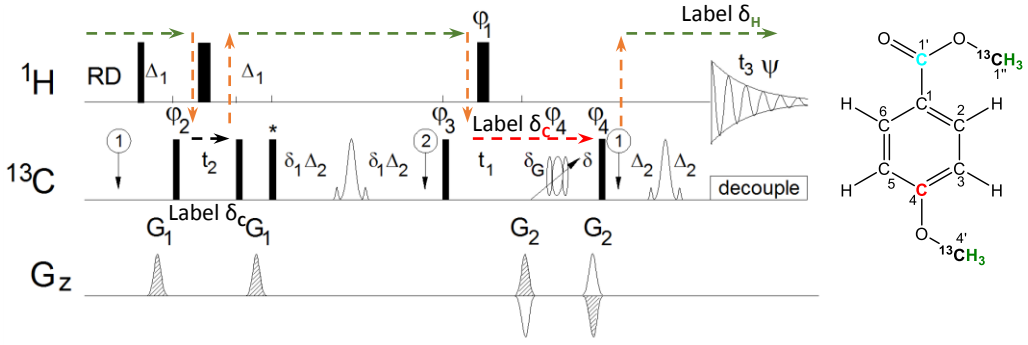


Figure 4.11 3D HMQC-HMBC sequence showing the polarisation transfer pathway. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. The following parameters were used: $\Delta_1 = 1/2^1J_{CH_3}$, $\delta = \delta_G + t_{180}(^1H) + 2t_1^0$, $\Delta_2 = 1/4^nJ_{CH} - t_{180}(^{13}C_{SNOB})/2$, $\delta_1 = (\delta_G + 0.5*t_{180}(^{13}C_{adia}) + 0.5*t_{180}(^1H)$, where δ_G is the gradient duration plus recovery time. t_1 and t_2 indicates the incremental period during which chemical shift labelling occurs. The following phase cycle was used φ_1 = 8(x), 8(-x), φ_2 = 2(x), 2(-x), φ_3 = x, -x, φ_4 = 4(x), 4(-x), ψ = x, -x, -x, x, -x, x, x, -x. 2 ms gradient pulses were applied at the following strengths: G_1 = 10% and G_2 = 80%. Quadrature detection in t_1 was achieved using echo, anti-echo protocol. In order to reduce the off-resonance effects the carrier frequencies were changed at specified places, indicated by the balloon topped arrows: ①= middle of methyl carbons ②= middle of aromatic and methyl carbons. 2 ms CHIRP refocusing pulses and 150 us r-SNOB pulses were used. The 90° ^{13}C pulse labelled with an asterisk converts any non-refocused methyl proton magnetisation into MQ coherences, thus eliminating it fully. ^{13}C decoupling was applied during acquisition. The molecule shown highlights the nuclei whose chemical shifts are obtained in this experiment. The green, black, red and turquoise boldfaced text and arrows represent the methyl protons, methyl carbons, aromatic carbons and carbonyl carbons, respectively.

The pulse sequence starts with a HMQC part, which labels the ^{13}C chemical shifts of the methoxy carbons (t_1) followed by creation of anti-phase proton magnetisation with respect to quaternary ipso carbons. Once converted into MQ coherences the chemical shift of the quaternary carbons are labelled in the HMBC part of the sequence. The conversion to SQ proton coherences is followed by the refocusing of $^3J_{CH_3, C}$ couplings and detection under ^{13}C decoupling. The coherence transfer of the 3D HMQC-HMBC is expressed using the product spin operators as follows:

$$-\hat{\boldsymbol{I}}_{\boldsymbol{y}}\,(methoxy) \xrightarrow{2\pi J_{IS}\Delta_1 \hat{I}_z} \boldsymbol{2\hat{I}_x\hat{S}_z} \xrightarrow{(\pi/2)\hat{S}_x} \boldsymbol{-2\hat{I}_x\hat{S}_y} \xrightarrow{\Omega_S t_2 \hat{S}_z;\, \pi\hat{I}_x} [\![Labels\ \delta_S]\!] \xrightarrow{(\pi/2)\hat{S}_x} \boldsymbol{-2\hat{I}_x\hat{S}_z} \xrightarrow{2\pi J_{IS}\Delta_1 \hat{I}_z\hat{S}_z;\, (\pi/2)\hat{S}_x}$$
$$\boldsymbol{-\hat{I}_y} \xrightarrow{2\pi J_{IR}(2\Delta_2+2\delta_1)\hat{I}_z\hat{R}_z;\, \pi\hat{R}_{x(methoxy)}} \boldsymbol{2\hat{I}_x\hat{R}_z} \xrightarrow{(\pi/2)\hat{R}_x} \boldsymbol{-2\hat{I}_x\hat{R}_y} \xrightarrow{\Omega_R t_1 \hat{R}_z;\, \pi\hat{I}_x;\, \pi\hat{R}_x} [\![Labels\ \delta_R]\!]$$
$$\xrightarrow{(\pi/2)\hat{R}_x} \boldsymbol{2\hat{I}_x\hat{R}_z} \xrightarrow{2\pi J_{IR}(2\Delta_2)\hat{I}_z\hat{R}_z;\, \pi\hat{S}_{x(methoxy)}} \boldsymbol{\hat{I}_y}\,(methoxy)$$

As with the ^{13}C-filtered HMBC precautions had to be taken to prevent the evolution of $^1J_{CH_3}$ coupling of fully labelled methoxy groups. This was achieved by setting the Δ_1 interval to the mean 1/2 $^1J_{CH_3}$ value. Any imperfectly refocused signals at the end of the HMQC segment were converted into MQ coherences by a 90° ^{13}C pulse. Evolution of $^1J_{CH_3}$ during the defocusing and refocusing delays was prevented by the application of methoxy carbon selective inversion pulses mid-way through these intervals. This is a more effective measure for stopping the evolution due to $^1J_{CH_3}$ couplings than setting the long-range evolution intervals to an even multiple of 1/2 $^1J_{CH_3}$.

4.1.4. NOESY-based experiments

In the NMR experiments developed so far, the magnetisation was transferred using scalar couplings involving natural abundance ^{13}C nuclei. This limits the sensitivity of the experiments to the point that concatenating another polarisation transfer with low efficiency such as a TOCSY or NOESY step is hardly feasible. However, the NOESY experiment uses a more efficient transfer from the methoxy group to the aromatic ring protons, without involving a natural abundance ^{13}C nucleus, thus transfer of the magnetisation further using proton-proton coupling constants is possible. The NOE transfer efficiency between a methoxy group and an aromatic ring (and vice versa) was investigated using 1D NOESY experiments. Details of how this was achieved are given in Section 2.18. The end result showed that transfer from methoxy to aromatic protons showed a NOE transfer efficiency between 3.8%, while the transfer efficiency from aromatic to methoxy protons was only 0.6%. The reason for this is down to the fact that the transfer is more efficient from three protons to one proton than the other way around, reflecting the cross relaxation. This efficiency increase compared to previous experiments was deemed sufficient to proceed with designing experiments that utilise the NOESY polarisation transfers. Towards this end, several experiments have been designed/tested to ascertain the feasibility of this approach. These experiments are 2D $^{13}CH_3$-filtered NOESY, 2D NOESY-TOCSY, 3D HMQC-NOESY and 4D HMQC-NOESY-TOCSY. Of course, each of these experiments can be easily modified to reduce their dimensionality e.g. 2D HMQC-NOESY or 2D/3D HMQC-NOESY-TOCSY. To understand

the nature of these experiments an initial description of the 2D NOESY pulse sequence is given first.

The NOESY experiment maps the NOEs due to through space interactions between protons, typically separated by less than 5 Å. The proton-proton NOE is observed as an intensity change of a proton signal, caused by dipolar coupling with another proton, in 1D spectra or as a cross-peak in a 2D NOESY spectrum. The basic 2D NOESY experiment is shown in Figure 4.12.

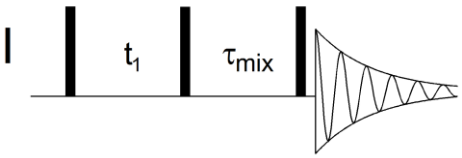


Figure 4.12 Basic 2D NOESY pulse sequence. The thin bars represent 90 pulses. The t_1 delay is the incremental period, while the τ_{mix} delay represents the mixing time.

Using the product spin operators and according to Hore *et al.*[215], the NOESY experiment can be described as follows:

The first 90°_{x} pulse creates transverse magnetisation of spin I_1:

$$\hat{\boldsymbol{I}}_{\boldsymbol{1z}} \xrightarrow{(\pi/2)\hat{I}_x} -\hat{\boldsymbol{I}}_{\boldsymbol{1y}}$$

Its chemical shift is labelled during t_1 and the magnetisation is transferred back to $\hat{I}_z$ by the second 90°_{x} pulse, where NOE transfer to the other close-by spins (I_2) occurs. The final 90° brings the magnetisation into observable form:[215]

$$\xrightarrow{\Omega t_1 \hat{I}_z;\, 2\pi J_{I_1 I_2} t_1 \hat{I}_z} \xrightarrow{(\pi/2)\hat{I}_x} \xrightarrow{\tau_{mix}} \xrightarrow{(\pi/2)\hat{I}_x} -\hat{\boldsymbol{I}}_{\boldsymbol{1y}} \cos\Omega_{I_1} t_1 \cos\pi J_{I_1 I_2} t_1 + 2\hat{\boldsymbol{I}}_{\boldsymbol{1x}}\hat{\boldsymbol{I}}_{\boldsymbol{2z}} \cos\Omega_{I_1} t_1 \sin\pi J_{I_1 I_2} t_1 + \hat{\boldsymbol{I}}_{\boldsymbol{1x}} \sin\Omega_{I_1} t_1 \cos\pi J_{I_1 I_2} t_1 + 2\hat{\boldsymbol{I}}_{\boldsymbol{1y}}\hat{\boldsymbol{I}}_{\boldsymbol{2z}} \sin\Omega_{I_1} t_1 \sin\pi J_{I_1 I_2} t_1$$

In addition to the desired $\hat{I}_z$ magnetisation the second 90°_{x} degree pulse creates some MQ and SQ terms:

$$-\hat{\boldsymbol{I}}_{\boldsymbol{1z}} \cos\Omega_{I_1} t_1 \cos\pi J_{I_1 I_2} t_1 - 2\hat{\boldsymbol{I}}_{\boldsymbol{2x}}\hat{\boldsymbol{I}}_{\boldsymbol{2y}} \cos\Omega_{I_1} t_1 \sin\pi J_{I_1 I_2} t_1 + \hat{\boldsymbol{I}}_{\boldsymbol{1x}} \sin\Omega_{I_1} t_1 \cos\pi J_{I_1 I_2} t_1 - 2\hat{\boldsymbol{I}}_{\boldsymbol{1z}}\hat{\boldsymbol{I}}_{\boldsymbol{2y}} \sin\Omega_{I_1} t_1 \sin\pi J_{I_1 I_2} t_1$$

The phase cycling simplifies this by eliminating the DQ and SQ coherences, which leaves:

$$-\hat{\boldsymbol{I}}_{\boldsymbol{z}} \cos\Omega_{I_1} t_1 \cos\pi J_{I_1 I_2} t_1 + \boldsymbol{ZQ} \sin\Omega_{I_1} t_1 \sin\pi J_{I_1 I_2} t_1$$

ZQ appears due to a MQ coherence $\hat{I}_{1x}\hat{I}_{2y}$, which is a mixture of ZQ and DQ terms, ($2\hat{I}_x\hat{S}_y$ = DQ − ZQ).

During mixing time, τ_{mix}, the magnetisation is transferred to the other close by spins by cross relaxation which are made observable diagonal and off diagonal peaks by the final 90°_x pulse:

$$-\hat{\boldsymbol{I}}_{\boldsymbol{1z}} \xrightarrow{\tau_{mix}} -a\hat{\boldsymbol{I}}_{\boldsymbol{1z}} + b\hat{\boldsymbol{I}}_{\boldsymbol{2z}} \xrightarrow{(\pi/2)\hat{I}_z} -a\hat{\boldsymbol{I}}_{\boldsymbol{1y}} + b\hat{\boldsymbol{I}}_{\boldsymbol{2y}}$$

Where $\hat{\boldsymbol{I}}_{\boldsymbol{1y}}$ is the diagonal and $\hat{\boldsymbol{I}}_{\boldsymbol{2y}}$ the off diagonal cross-peak. *a* and *b* are coefficients that depend on factors such as the transfer efficiency and the length of the mixing time.

4.1.4.1. ^{13}C-filtered NOESY

As can be seen from the 2D NOESY spectra of HS samples presented in Section 1.5.2.3.1 these are complex and difficult to interpret. In the ^{13}C-methylated HS sample only the NOEs between the methoxy protons and protons of the parent compounds are of interest. These can be selected using a X- or more specifically a ^{13}C-filter. The ^{13}C-filtered NOESY experiment developed by Otting and Wüthrich[299] uses the evolution of $^1J_{CH_3}$ couplings to select only magnetisation either of ^{13}C-H or ^{12}C-H protons, while the other is effectively removed by appropriate phase cycling. In this study the ^{13}C-filter selects the ^{13}C labelled protons and replaces the initial 90°_x pulse of the NOESY pulse sequence. The pulse sequence of the 2D ^{13}C-filtered NOESY is shown in Figure 4.13.

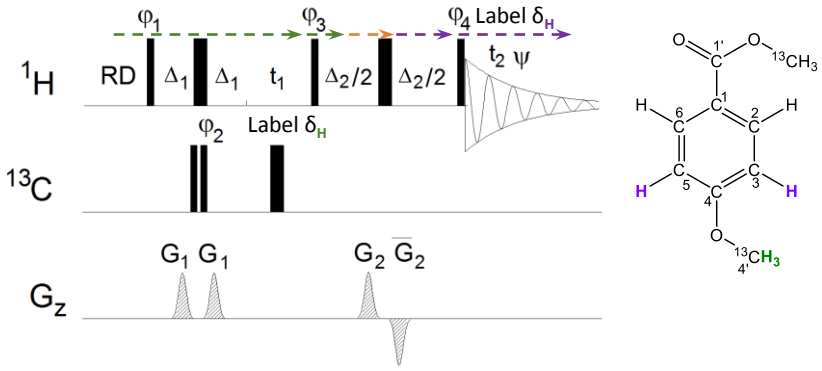


Figure 4.13 2D ^{13}C-filtered NOESY pulse sequence with the polarisation transfer pathway. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. The following parameters were used: $\Delta_1 = 1/2^1J_{CH_3}$, $\Delta_2 = \tau_{mix}$. t_1 indicates the incremental period during which chemical shift labelling occurs. The following phase cycle was used: φ_1 = 2(x), 2(-x), φ_2 = x,-x, φ_3 = 16(x), 16(-x), φ_4 = 4(x), 4(-x), 4(y), 4(-y); ψ = x, -x, -x, x, -x, x, x, -x, y, -y, -y, y, -y, y, y, -y, -x, x, x, -x, x, -x, -x, x, -y, y, y, -y, y, -y, -y, y. Gradient strengths were: G_1 = 21%, G_2 = 15% and -15%. The molecule shown highlights the nuclei whose chemical shifts are obtained in this experiment. The green and purple boldfaced text and arrows represent the methyl and aromatic protons, respectively.

In short, this pulse sequence selects the ^{13}C methoxy groups and labels the associated proton chemical shifts in the F_1 dimension. The magnetisation is then transferred via NOE to nearby protons whose chemical shifts are read by the last 90° pulse and displayed along the F_2 dimension. In terms of simplified product spin operators, the outcome of the ^{13}C-filtered NOESY experiment on the ^{13}C-methylated compounds can be summarised as follows, where $\hat{I}_1$= methoxy proton, $\hat{I}_2$= ortho aromatic protons carbons and $\hat{S}$ = methoxy carbon.

$$-\hat{\boldsymbol{I}}_{\mathbf{1y}} \xrightarrow{2\pi J_{IS}\Delta_1 \hat{I}_z\hat{S}_z;\, \pi\hat{I}_x} \mathbf{2}\hat{\boldsymbol{I}}_{\mathbf{1}x}\hat{\boldsymbol{S}}_{\boldsymbol{z}} \xrightarrow{\pi\hat{S}_x \text{ or } 0\hat{S}_x} \pm\mathbf{2}\hat{\boldsymbol{I}}_{\mathbf{1}x}\hat{\boldsymbol{S}}_{\boldsymbol{z}} \xrightarrow{2\pi J_{IS}\Delta_1 \hat{I}_z\hat{S}_z} \pm\hat{\boldsymbol{I}}_{\mathbf{1y}} \xrightarrow{\Omega_I t_1 \hat{I}_z;\, \pi\hat{S}_x} [\![labels\ \delta_H\ (methoxy)]\!]$$
$$\xrightarrow{(\pi/2)\hat{I}_x} \pm\hat{\boldsymbol{I}}_{\mathbf{1z}} \xrightarrow{\Delta_2;\, \pi\hat{I}_x} \pm\hat{\boldsymbol{I}}_{\mathbf{1z,2z}} \xrightarrow{(\pi/2)\hat{I}_x} \pm\hat{\boldsymbol{I}}_{\mathbf{1y,2y}}$$

The pulsed field gradients of opposite polarity surrounding the central 180° pulse of the mixing time de-phase DQ, but not ZQ coherences. As the $^{13}CH_3O$ carbons do not have a J coupling with aromatic protons, ^{13}C decoupling is not needed during t_2. Long acquisition times can therefore be used, resulting in better resolution and revealing the multiplicity of the proton signals, thus adding to the information obtainable from the experiment.

4.1.4.2. ^{13}C-filtered NOESY-TOCSY

Going one step further, TOCSY transfer can extend the ^{13}C-filter NOESY experiment to further transfer the magnetisation, after the initial NOESY step, using the J_{HH} coupling of the aromatic protons. To formally extend the ^{13}C-filtered NOESY experiment a TOCSY step is added after the t_1 period. However, crucially, the chemical shifts of the methoxy protons are not labelled. Instead the chemical shifts of protons nearby the methoxy groups are labelled after the NOESY step before returning their transverse magnetisation to the z-axis. During the following TOCSY period magnetisation is transferred to nearby protons via J_{HH} couplings. The final read pulse flips these protons into the transverse plane where they are detected during t_2. Another difference between this experiment and the ^{13}C-filtered NOESY (Figure 4.13) is that it uses DPFGSE for the selection of methoxy protons in place of the first 90° pulse of the ^{13}C filter. This provides cleaner spectra as the aromatic protons are dephased out right and spectra therefore show reduced cancellation artefacts. The pulse sequence of such an experiment is shown in Figure 4.14.

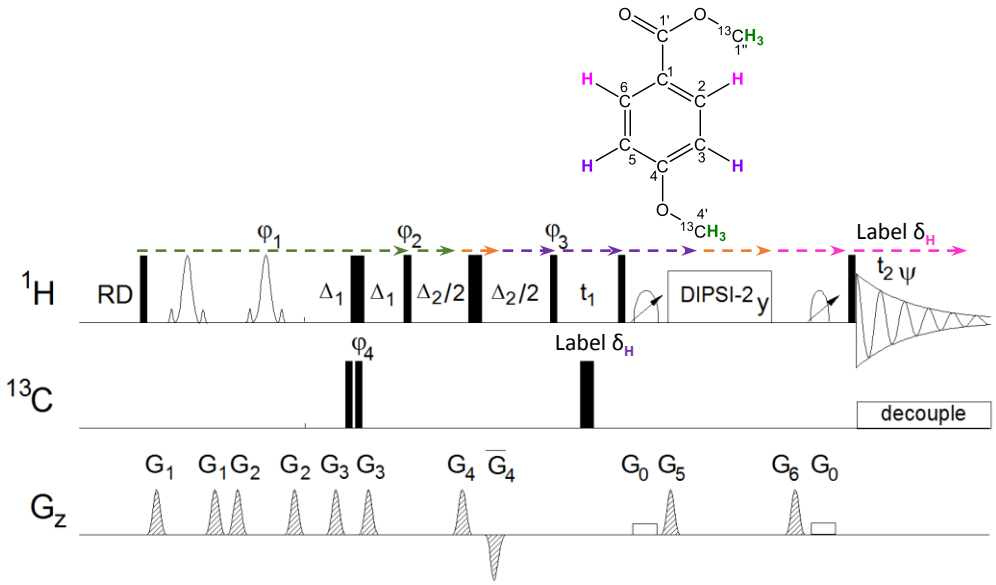


Figure 4.14 2D ^{13}C-filtered NOESY-TOCSY with the polarisation transfer pathway. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. The following parameters were used: $\Delta_1 = 1/2^1J_{CH3}$, $\Delta_2 = \tau_{mix}$. t_1 indicates the incremental period during which chemical shifts labelling occurs. The following phase cycle was used: φ_1 = 8(y), 8(x), φ_2 = 4(x), 4(-x), φ_3 = 2(x), 2(-x), φ_4 = x, -x, ψ = x, -x, -x, x, -x, x, x, -x, -x, x, x, -x, x, -x, -x, x. Gradient strengths were: G_0 = 8%, G_1 = 1 1%, G_2 = 17%, G_3 = 21%, G_4 = 15%, G_5 = 7%, G_6 = 31%. The MQ coherences were purged by two adiabatic pulses during which low level gradients were applied, before and after the TOCSY step. r-SNOB 180° shaped pulses were used in the initial DPFGSE and CHIRP pulses were used simultaneously with G_0 gradients. ^{13}C decoupling was applied during acquisition. The molecule shown highlights the nuclei whose chemical shifts are obtained in this experiment. The green, purple and pink boldfaced text and arrows represent the methyl protons, aromatic protons (ortho to the methoxy group) and aromatic protons (meta to the methoxy group), respectively.

Neglecting the differences in position of the labelling period, the spin product operators describing this pulse sequence are identical to those of the ^{13}C-filtered NOESY until after the NOE mixing time, where a TOCSY transfer takes place.

$$\pm\hat{\boldsymbol{I}}_{2z} \xrightarrow{DIPSI\ spin-lock} \pm\hat{\boldsymbol{I}}_{2z,3z} \xrightarrow{(\pi/2)\hat{I}_x} \pm\hat{\boldsymbol{I}}_{2y,3y}$$

The addition of two adiabatic pulses during which low strength PFGs are applied before and after the TOCSY spin-lock, dephases the DQ and ZQ coherences generating pure-phase multiplets. DIPSI-2, a *z*-filtered spin-lock sequence suitable for the narrow band of aromatic proton resonances, was used.

When applied to the ^{13}C-methylated samples, the 2D ^{13}C-filtered NOESY-TOCSY selects the protons of ^{13}C-labelled methoxy groups and transfers the magnetisation to nearby protons via NOEs. The TOCSY step then transfers the magnetisation from these protons to their *J* coupled partners, i.e. aromatic ring protons.

4.1.4.3. 3D HMQC-NOESY

This experiment starts with a DPFGSE selecting only the methoxy protons followed by chemical shift labelling of methoxy carbons during the t_2 period of the HMQC. What follows is a t_1 period that labels the chemical shift of the methoxy protons while refocusing the evolution of $^1J_{CH_3}$ couplings. The magnetisation of methoxy protons is then prepared for a NOE transfer along the z-axis by a 90° ^{1}H pulse.

The pulse sequence of the 3D HMQC-NOESY is shown in Figure 4.15.

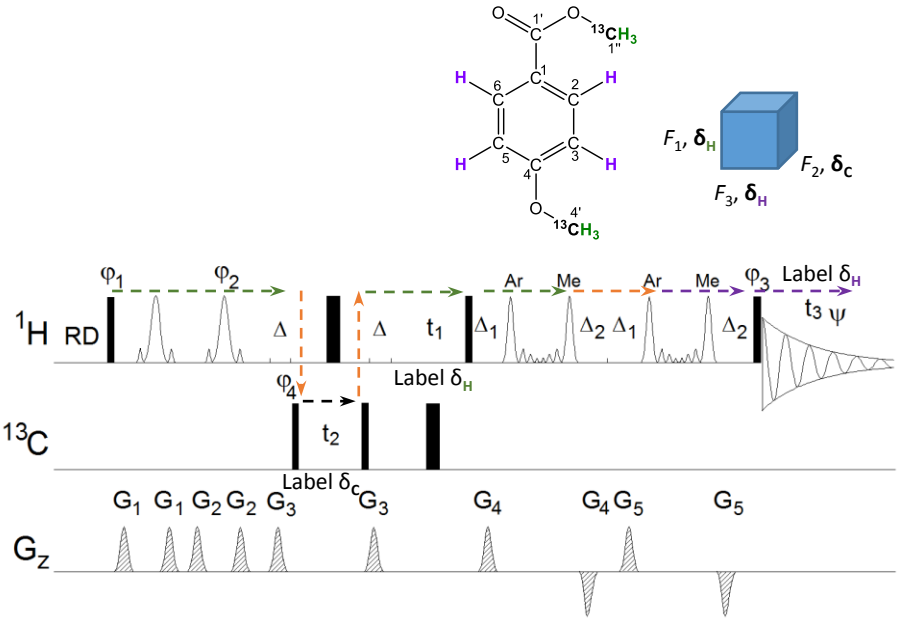


Figure 4.15 3D DPFGSE-HMQC-NOESY showing the polarisation transfer pathway. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. The following parameters were used: $\Delta = 1/2\ J_{CH_3}$, $\Delta_1 = \tau_{mix}*\alpha$, where α = 0.25-0.4, $\Delta_2 = \tau_{mix}*(0.5-\alpha)$. t_1 and t_2 indicate incremental periods during which chemical shift labelling occurs. 1 and 0.6 ms gradient pulses were applied at the following strengths: G_1 = 33%, G_2 = 17%, G_3 = 12%, G_4 = 9%, G_5 = 25%. The following phase cycle was used: φ_1 = 2(x), 2(-x), φ_2 = y, φ_3 = 4(x), 4(-x), φ_4 = x, -x, ψ = x, -x, -x, x, -x, x, x, -x. r-SNOB pulses were used for DPFGSE, while IBURP inversion pulses were applied at 3.9 and 7.4 ppm during the mixing time. The molecule shown highlights the nuclei whose chemical shifts are obtained in this experiment. The green, black and purple boldfaced text and arrows represent the methyl protons, methyl carbons and aromatic protons, respectively.

Starting after the DPFGSE the polarisation transfer pathway using the product spin operators is expressed as follows, where $\hat{I}_1$= methoxy proton, $\hat{I}_2$= ortho aromatic protons and $\hat{S}$ = methoxy carbons:

$$\hat{\boldsymbol{I}}_{\boldsymbol{1y}} \xrightarrow{2\pi J_{IS}\Delta \hat{I}_z \hat{S}_z} -2\hat{\boldsymbol{I}}_{\boldsymbol{1x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}} \xrightarrow{(\pi/2)\hat{S}_x} 2\hat{\boldsymbol{I}}_{\boldsymbol{1x}}\hat{\boldsymbol{S}}_{\boldsymbol{y}} \xrightarrow{\Omega_S t_2 \hat{S}_z;\, \pi \hat{I}_x} [\![labels\ \delta_S]\!] \xrightarrow{(\pi/2)\hat{S}_x} 2\hat{\boldsymbol{I}}_{\boldsymbol{1x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}$$

$$\xrightarrow{2\pi J_{IS}\Delta \hat{I}_z \hat{S}_z} \hat{\boldsymbol{I}}_{\boldsymbol{1y}} \xrightarrow{\Omega_I t_1 \hat{I}_z;\, \pi \hat{S}_x} [\![Labels\ \delta_I]\!] \xrightarrow{(\pi/2)\hat{I}_x} \hat{\boldsymbol{I}}_{\boldsymbol{1z}} \xrightarrow{\Delta_1;\, \Delta_2;\, 2(\pi \hat{I}_{x(methoxy)}, \pi \hat{I}_{x(arom)})} \hat{\boldsymbol{I}}_{\boldsymbol{1z,2z}} \xrightarrow{(\pi/2)\hat{I}_x} \pm \hat{\boldsymbol{I}}_{\boldsymbol{1y,2y}}$$

The HMQC segment was used in this experiment instead of the HSQC because i) HSQC has more pulses and thus is more susceptible to miscalibrations; and ii) the $2\hat{S}_x\hat{I}_x$ coherences generated by the HMQC do not show undesirable J_{HH} coupling evolution in F_1 because the methyls are singlets, thus there is no compromise to the resolution and sensitivity.

An efficient NOE transfer for these small molecules requires long NOE build up times, during which considerable auto relaxation occurs. This leads to appearance of axial peaks at zero F_1 frequency and increased t_1 noise. To maintain high quality spectra the aromatic and methoxy protons were inverted twice during the NOE mixing time by band selective pulses. In this way, the NOE transfer was limited between these groups. In practice these selective pulses may not be necessary and could be replaced by non-selective ones if no NOE is expected to occur between the methoxy protons and aliphatic protons. When using the selective pulses these can in principle be applied simultaneously or consecutively. After testing several options, the IBURP2 pulse was found to be the most effective. Two phase shifted pulses centred on the aromatic and methoxy regions were added, with the second pule time reversed. However, on examining the final inversion profile, this procedure was deemed unsatisfactory due to an effect reported by Kupce and Freeman.[300] The so-called 'close encounter' effects between two soft pulses are due to one inverted event affecting the other, even when one of the regions was inverted using a time reversed pulse. In order to prevent these effects, two shaped pulses inverting the methoxy and aromatic resonances, consecutively, were used.

As shown by Stott *et al.*[301] placing a 180° pulse in the middle of the mixing time interval is less effective compared to two inversion events spaced appropriately. It is best to use two 180° pulses, one at ~0.25 *τ_{mix} and the other at ~0.75 *τ_{mix} to get more efficient suppression of artefacts.

4.1.4.4. 3D HMQC-NOESY-TOCSY

It is straight forward to extent the 3D HMQC-NOESY pulse sequence into a 3D HMQC-NOESY-TOCSY. Magnetisation is kept along the z-axis after the NOESY mixing time and a DIPSI-2 spin-lock is applied surrounded by ZQ removal elements.[302] The pulse sequence, polarisation transfer pathway and sampled nuclei are shown in Figure 4.16.

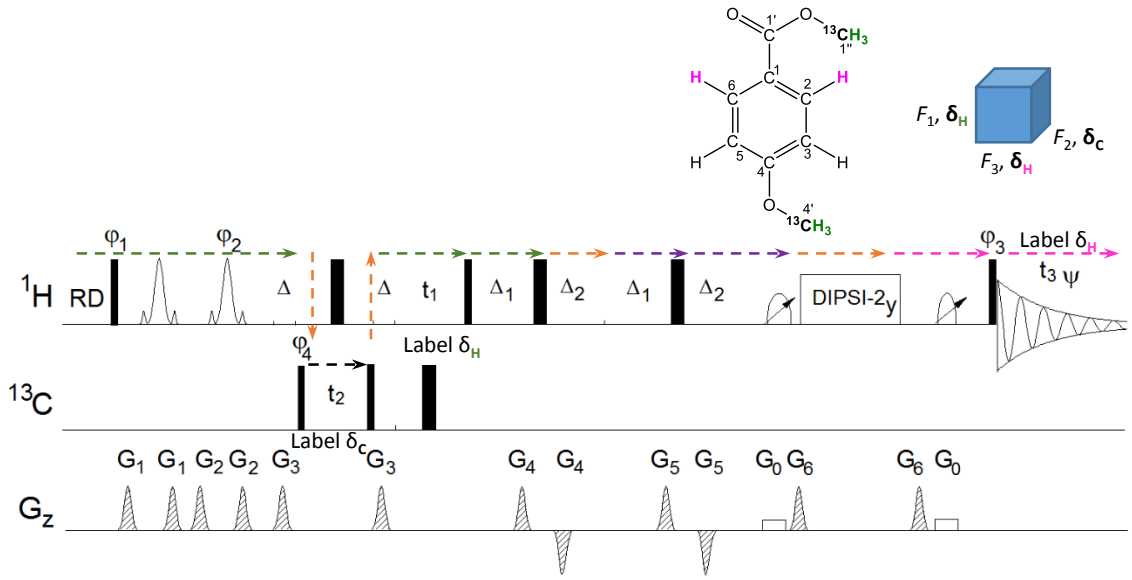


Figure 4.16 3D DPFGSE-HMQC-NOESY-TOCSY pulse sequence with the polarisation transfer pathway. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. The following parameters were used: $\Delta = 1/2\ ^1J_{CH}$, $\Delta_1 = \tau_{mix}$* α, where α = 0.25-0.45, $\Delta_2 = \tau_{mix}$*(0.5- α). t_1 and t_2 indicate incremental periods during which chemical shift labelling occurs. 1 and 0.6 ms gradient pulses were applied at the following strengths: G_0 = 4%, G_1 = 33%, G_2 = 17%, G_3 = 12%, G_4 = ± 9%, G_5 = ± 25%, G_6 = 7%. The following phase cycle was used: φ_1 = 2(x), 2(-x), φ_2 = y, φ_3 = 4(x), 4(-x), φ_4 = x, -x, ψ = x, -x, -x, x, -x, x, x, -x. r-SNOB pulses were used for DPFGSE and CHIRP pulses were applied before and after the DIPSI spin-lock. The molecule shown highlights the nuclei whose chemical shifts are obtained in this experiment. The green, black and pink boldfaced text and arrows represent the methyl protons, methyl carbons and aromatic protons (meta to the methoxy groups), respectively. The purple arrows represent aromatic protons, through which the polarisation is passed but their chemical shift is not labelled.

Starting after the DPFGSE, the polarisation transfer pathway in product spin operators can be described as follows, where $\hat{I}_1$= methoxy proton, $\hat{I}_2$= ortho aromatic protons carbons, $\hat{I}_3$= meta or para aromatic protons and $\hat{S}$ = methoxy carbon.

$$\hat{\boldsymbol{I}}_{\mathbf{1y}} \xrightarrow{2\pi J_{IS}\Delta\hat{I}_z\hat{S}_z} -2\hat{\boldsymbol{I}}_{\mathbf{1x}}\hat{\boldsymbol{S}}_{\mathbf{z}} \xrightarrow{(\pi/2)\hat{S}_x} 2\hat{\boldsymbol{I}}_{\mathbf{1x}}\hat{\boldsymbol{S}}_{\mathbf{y}} \xrightarrow{\Omega_S t_2\hat{S}_z;\ \pi\hat{I}_x} [\![labels\ \delta_S]\!] \xrightarrow{(\pi/2)\hat{S}_x} 2\hat{\boldsymbol{I}}_{\mathbf{1x}}\hat{\boldsymbol{S}}_{\mathbf{z}} \xrightarrow{2\pi J_{IS}\Delta\hat{I}_z\hat{S}_z}$$

$$\hat{\boldsymbol{I}}_{\mathbf{1y}} \xrightarrow{\Omega_I t_1\hat{I}_z;\ \pi\hat{S}_x} [\![Labels\ \delta_I]\!] \xrightarrow{(\pi/2)\hat{I}_x} \hat{\boldsymbol{I}}_{\mathbf{1z}} \xrightarrow{2\Delta_1;\ 2\Delta_2;\ 2\pi\hat{I}_x,} \hat{\boldsymbol{I}}_{\mathbf{1z,2z}} \xrightarrow{DIPSI} -\hat{\boldsymbol{I}}_{\mathbf{1z,2z,3z}} \xrightarrow{(\pi/2)\hat{I}_x} -\hat{\boldsymbol{I}}_{\mathbf{1y,2y,3y}}$$

The sequence allows transfer from the aromatic protons generated by the NOESY step as described for the HMQC-NOESY experiment to other *J*-coupled aromatic protons.

4.1.4.5. 4D HMQC-NOESY-TOCSY

The 3D HMQC-NOESY-TOCSY experiment can be easily extended to a 4D experiment by labelling the chemical shift of aromatic protons after the NOESY transfer. The magnetisation is

then brought back to the *z*-axis for a DIPSI spin-lock and detected during the direct acquisition period t_4.

The pulse sequence, polarisation transfer pathway and sampled nuclei are shown for the 4D HMQC-NOESY-TOCSY in Figure 4.17.

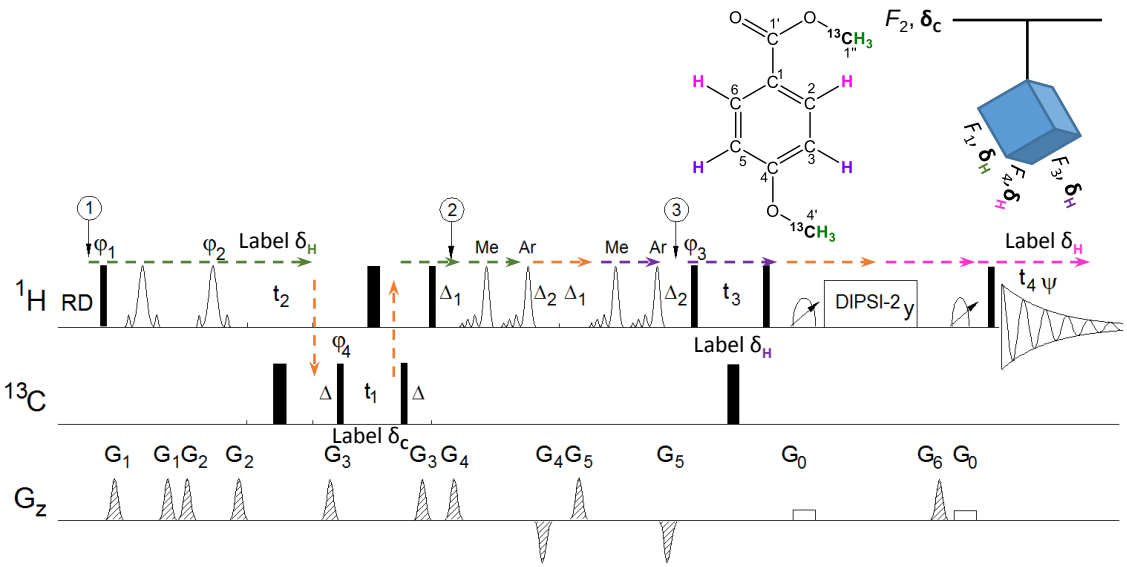


Figure 4.17 4D HMQC-NOESY-TOCSY with polarisation transfer pathway. The thin and thick rectangles represent 90° and 180° pulses, respectively. Unless stated otherwise the pulses were applied from the *x*-direction. The following parameters were used: $\Delta = 1/2\,{}^1J_{CH}$, $\Delta_1 = \tau_{mix}*\alpha$, where $\alpha = 0.25\text{-}0.45$, $\Delta_2 = \tau_{mix}*(0.5\text{-}\alpha)$. t_1, t_2 and t_3 indicate incremental periods during which chemical shift labelling occurs. 1 and 0.6 ms gradient pulses were applied at the following strengths: $G_0 = 4\%$, $G_1 = 33\%$, $G_2 = 17\%$, $G_3 = 12\%$, $G_4 = \pm 9\%$, $G_5 = \pm 25\%$, $G_6 = 7\%$. The following phase cycle was used: φ_1 = 2(x), 2(-x), φ_2 = y, φ_3 = 4(x), 4(-x), φ_4 = x, -x, ψ = x, -x, -x, x, -x, x, x, -x. In order to reduce the off- resonance effects and easily optimise the spectral windows in individual dimensions the carrier frequencies were changed at specified places, indicated by the balloon topped arrows: ①= middle of methyl protons ②= middle of aromatic and methyl protons and ③= middle of aromatic protons. r-SNOB pulses were used for DPFGSE, non-standard IBURP pulses were used during the mixing time and CHIRP pulses were applied before and after the DIPSI spin-lock. The molecule shown highlights the nuclei whose chemical shifts are obtained in this experiment. The green, black, purple and pink boldfaced text and arrows represent the methyl protons, methyl carbons, aromatic protons (ortho to the methoxy groups) and aromatic protons (meta to the methoxy groups), respectively.

Starting after the DPFGSE the polarisation transfer pathway expressed using product spin operators is as follows, where $\hat{I}_1$= methoxy proton, $\hat{I}_2$= ortho aromatic protons carbons, $\hat{I}_3$= meta or para aromatic protons and $\hat{S}$ = methoxy carbon.

$$\hat{\boldsymbol{I}}_{\boldsymbol{1y}} \xrightarrow{\Omega_I t_2 \hat{I}_z;\, \pi \hat{S}_x} [\![labels\ \delta_I]\!] \xrightarrow{2\pi J_{IS} \Delta \hat{I}_z \hat{S}_z} -2\hat{\boldsymbol{I}}_{\boldsymbol{1x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}} \xrightarrow{(\pi/2)\hat{S}_x} 2\hat{\boldsymbol{I}}_{\boldsymbol{1x}}\hat{\boldsymbol{S}}_{\boldsymbol{y}} \xrightarrow{\Omega_S t_1 \hat{S}_z;\, \pi \hat{I}_x} [\![Labels\ \delta_S]\!] \xrightarrow{(\pi/2)\hat{S}_x} 2\hat{\boldsymbol{I}}_{\boldsymbol{1x}}\hat{\boldsymbol{S}}_{\boldsymbol{z}}$$

$$\xrightarrow{2\pi J_{IS} \Delta \hat{I}_z \hat{S}_z} \hat{\boldsymbol{I}}_{\boldsymbol{1y}} \xrightarrow{(\pi/2)\hat{I}_x} \hat{\boldsymbol{I}}_{\boldsymbol{1z}} \xrightarrow{2\Delta_1;2\Delta_2;\, 2\left(\pi \hat{I}_{x(methoxy)}, \pi \hat{I}_{x(arom)}\right)} \hat{\boldsymbol{I}}_{\boldsymbol{1z,2z}} \xrightarrow{(\pi/2)\hat{I}_x} -\hat{\boldsymbol{I}}_{\boldsymbol{1y,2y}}$$

$$\xrightarrow{\Omega_I t_3 \hat{I}_z;\, \pi S_x} [\![Labels\ \delta_I]\!] \xrightarrow{(\pi/2)\hat{I}_x} -\hat{\boldsymbol{I}}_{\boldsymbol{1z,2z}} \xrightarrow{\boldsymbol{DIPSI}} -\hat{\boldsymbol{I}}_{\boldsymbol{1z,2z,3z}} \xrightarrow{(\pi/2)\hat{I}_x} \hat{\boldsymbol{I}}_{\boldsymbol{1y,2y,3y}}$$

4.2. Chapter Conclusions

From a number of possible polarisation transfer pathways, outlined in Figure 4.1, a series of experiments utilising the labelled ^{13}C methoxy group have been designed. These experiments are summarised in Table 4.1.

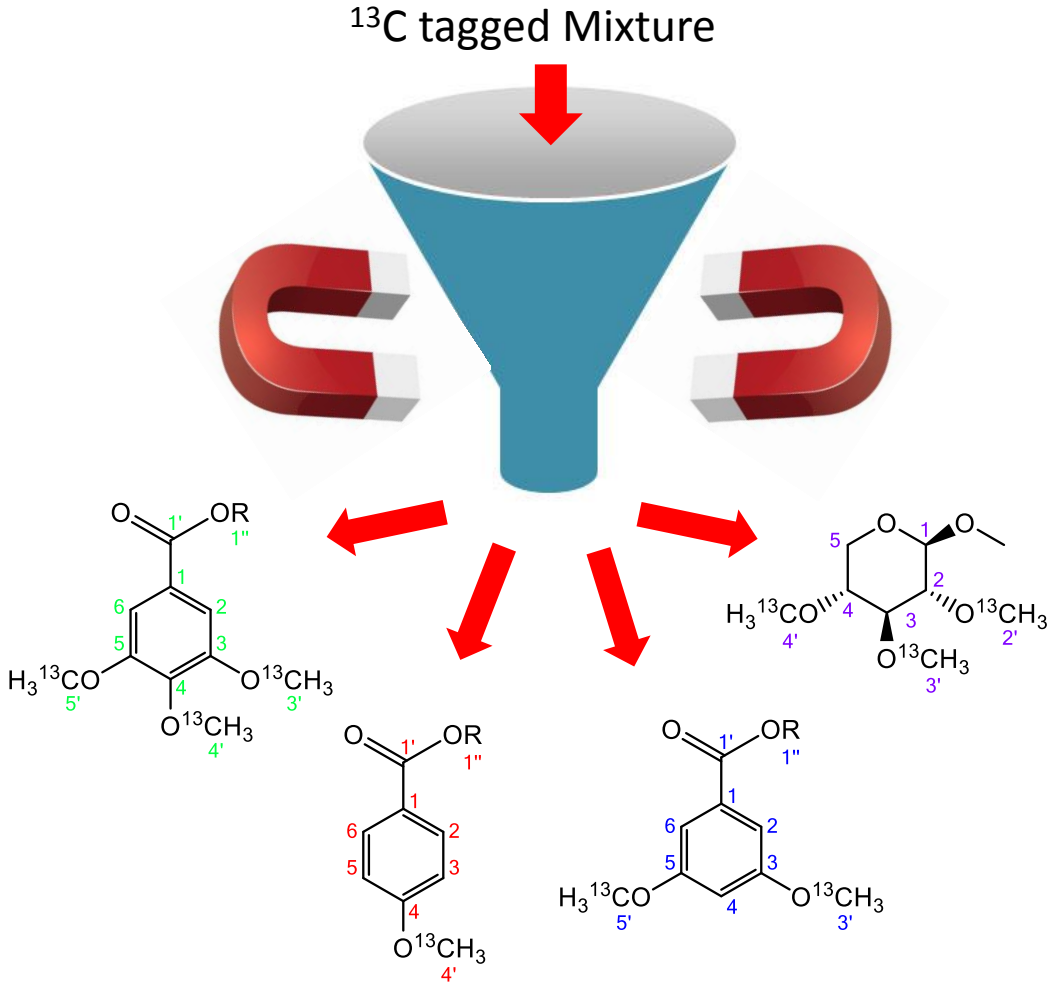
Table 4.1. Summary of the ^{13}C-filtered nD NMR experiments and the chemical shifts they provide.[a]

Experiment/Dimension	F_1	F_2	F_3	F_4
2D INEPT-INADEQUATE	$[C_{Arom}C(H_3)]_{DQ}$	$(C)H_3/H_{Arom}$	-	-
3D INEPT-INADEQUATE-HSQC	$[C_{Arom}C(H_3)]_{DQ}$	$C(H_3)/C_{Arom}$	$(C)H_3/H_{Arom}$	-
4D $HCCH_3$	H_{Arom}	C_{Arom}	$C(H_3)$	$(C)H_3$
3D $HcCH_3$	H_{Arom}	$C(H_3)$	$(C)H_3$	-
3D $hCCH_3$	C_{Arom}	$C(H_3)$	$(C)H_3$	-
2D ^{13}C-filtered HMBC	$Cipso_{Arom}$	$(C)H_3$	-	-
3D HMQC-HMBC	$C(H_3)$	C_{Arom}	$(C)H_3$	-
2D ^{13}C-filtered NOESY	$(C)H_3$	H_{Arom}	-	-
2D ^{13}C-filtered NOESY-TOCSY	$(C)H_3$	H_{Arom}	-	-
3D HMQC-NOESY	$(C)H_3$	$C(H_3)$	$H_{Arom\text{-}ortho}$	-
3D HMQC-NOESY-TOCSY	$(C)H_3$	$C(H_3)$	$H_{Arom\text{-}meta}$	-
4D HMQC-NOESY-TOCSY	$C(H_3)$	$(C)H_3$	$H_{Arom\text{-}ortho}$	$H_{Arom\text{-}meta}$

[a] The chemical shifts of nuclei enclosed in parentheses are not labelled.

The experiments summarised in Table 4.1 were tested on model mixtures **I** and **II** as discussed in Chapters 5 and 6.

Chapter 5. Results and Discussion. Application of ^{13}C-filtered NMR experiments to model mixture I



The pulse sequences introduced in Chapter 4 were tested and optimised on the ^{13}C-methylated model mixtures **I** and **II**. The experimental details of the presented spectra are given in Chapter 2. ^{13}C-methylated model mixture **I** was used to test the following experiments 2D INEPT-INADEQUATE, 2D ^{13}C-filtered HMBC, 4D HCCH_3, 2D ^{13}C-filtered NOESY, 2D ^{13}C-filtered NOESY-TOCSY and 4D HMQC-NOESY-TOCSY. The obtained results are briefly presented next. A more thorough analysis of a complementary set of experiments is presented using model ^{13}C-methylated mixture **II** in Chapter 6.

5.1. 2D INEPT-INADEQUATE

The 2D INEPT-INADEQUATE spectrum of ^{13}C-methylated model mixture **I**, acquired using the pulse sequence of Figure 4.2, with expansions of particular regions, is shown in Figure 5.1. Exemplar polarisation pathways are shown on molecule **B**. CH_3-Cq correlations (magenta and purple boxes) show only one cross-peak at a particular ^{13}C DQ frequency, while CH_3-CH correlations show two (orange boxes) cross-peaks. The 2D INEPT-INADEQUATE spectrum shows the proton chemical shifts along the F_2 dimension while the F_1 dimension represents the DQ frequency. As can be seen from the assigned cross-peaks, the experiment provides the desired information. For example, molecule **B** shows the CH_3-Cq cross-peaks representing correlations from the methoxy protons to the quaternary aromatic carbons, and also CH_3-CH cross-peaks representing the correlations from the methoxy groups to the aromatic protons. In addition, this experiment also allows access to the chemical shifts of the ester groups and the correlations from the ester methoxy groups to the quaternary aromatic carbons (purple box, expansion not shown). There is also expected to be one two-bond and one or two three-bond CH_3-CH correlations from the sugar **D** as well (grey box, expansion not shown). Thus, this experiment is very useful in providing multiple ^{13}C chemical shifts of the same molecule. It should be noted that the DQ frequencies obtained in this experiment can only be interpreted if an additional ^{1}H, ^{13}C HSQC spectrum is acquired. This is a feasible approach for the simple mixture used here. However, in complicated mixtures ambiguities will mount quickly. Therefore, this experiment needs to be modified by incorporating an HSQC step and forming a 3D experiment as illustrated on model mixture **II** in the next Chapter.

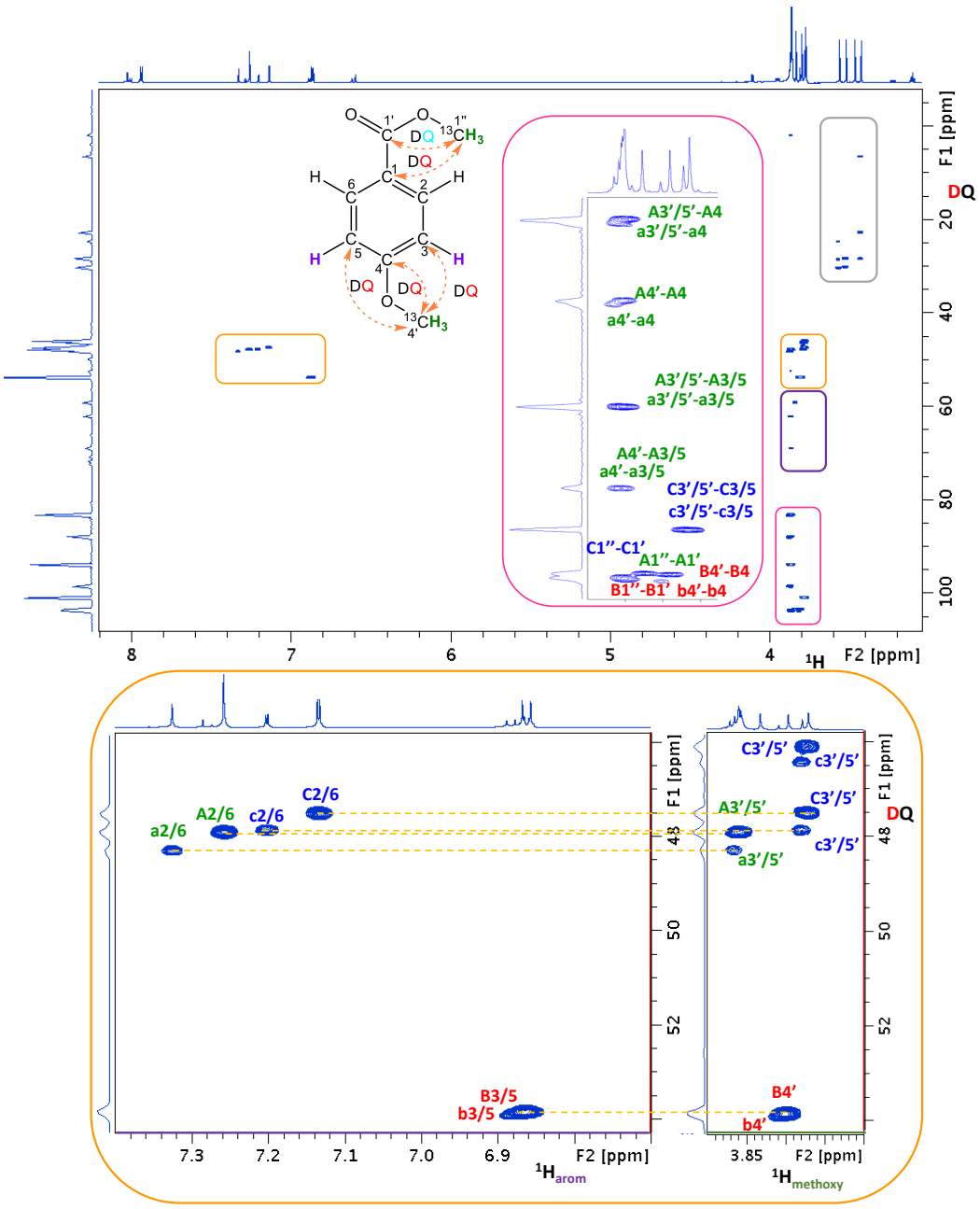


Figure 5.1 2D INEPT-INADEQUATE spectrum of ^{13}C-methylated model mixture **I**, acquired using the pulse sequence shown in Figure 4.2. The magenta box and expansion shows the region containing the aromatic CH_3-C_q correlations, while orange box and expansion shows the aromatic CH_3-CH correlations. The purple and grey boxes indicate the regions containing ester CH_3-C_q and sugar CH_3-CH correlations, respectively. The correlations to CH sugar protons not visible due to their low intensity. **A**-**C** and **a**-**c** represent the fully and partially methylated molecules. Exemplar polarisation transfer pathways are shown on molecule **B**. The cross-peak labels refer to molecules shown in Figure 3.4.

5.2. 2D ^{13}C-filtered HMBC

The methoxy proton to aromatic or ester ipso carbon region of the ^{13}C-filtered HMBC spectrum of ^{13}C-methylated model mixture **I**, acquired using the pulse sequence of Figure 4.10 is shown in Figure 5.2. This experiment selects the methoxy protons and transfers the magnetisation via $^{3}J_{CH_3,Ci}$ coupling constants to ipso carbons.

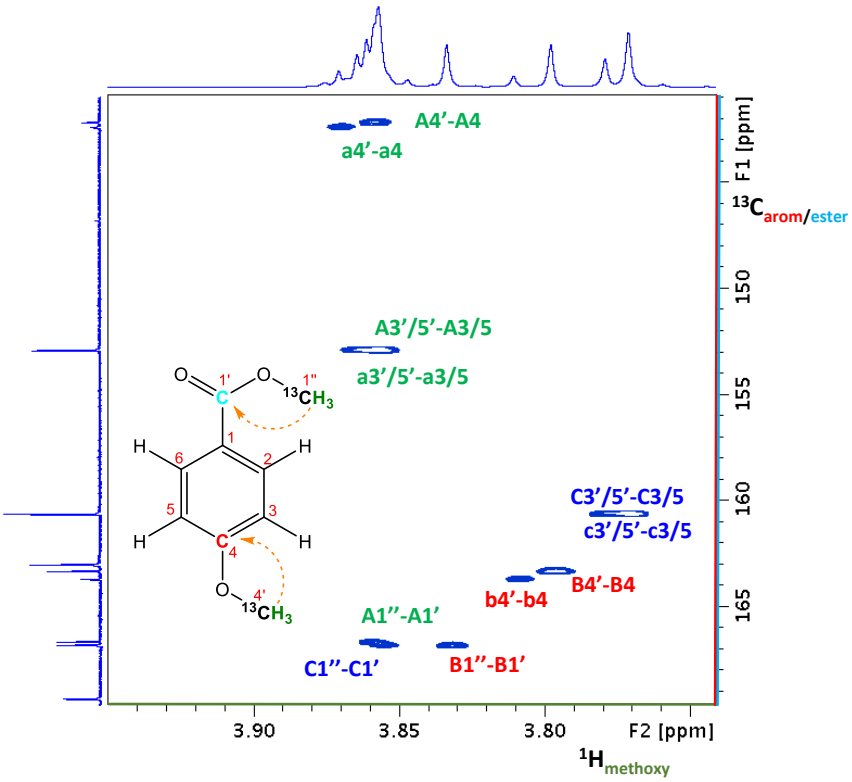


Figure 5.2 2D ^{13}C-filtered HMBC spectrum of ^{13}C-methylated model mixture **I**, acquired using the pulse sequence shown in Figure 4.10. **A**-**C** and **a**-**c** represent the fully and partially methylated molecules, respectively. Exemplar polarisation pathways are illustrated on molecule **B**. The cross-peak labels refer to molecules shown in Figure 3.4.

The ^{13}C-filtered HMBC spectrum shows the ^{13}C-labelled methoxy protons along the F_2 dimension and the $^{3}J_{CH_3,Ci}$ coupled carbons along the F_1 dimension. From the number of cross-peaks and their assignments, it is clear that the experiment shows the desired correlations. For example, molecule **B** shows two long-range correlations from the methoxy and the ester protons to the quaternary aromatic carbon (4'-4) and carboxyl carbon (1''-1'), respectively. Thus the ^{13}C-filtered HMBC is capable of correlating ^{13}C-labelled methoxy protons with ipso carbons.

5.3. 2D ^{13}C-filtered NOESY

The methoxy to aromatic proton region of the 2D ^{13}C-filtered NOESY spectrum obtained for the ^{13}C-methylated model mixture **I** using the pulse sequence of Figure 4.13 is shown in Figure 5.3.

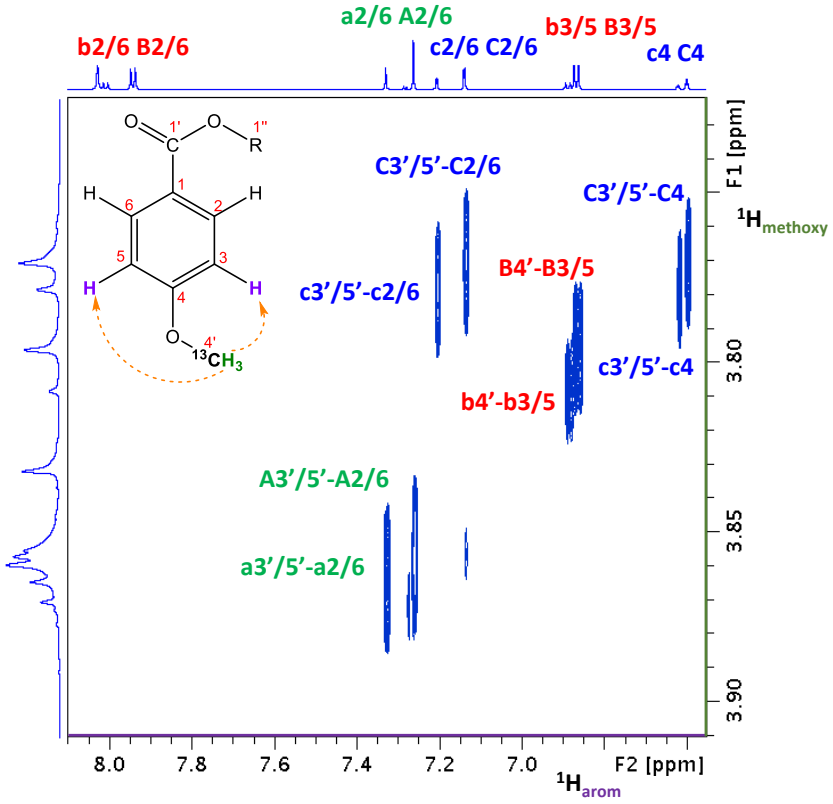


Figure 5.3 Methoxy to aromatic proton region of the 2D ^{13}C-filtered NOESY spectrum of ^{13}C-methylated model mixture **I** obtained using the pulse sequence shown in Figure 4.13, with a 600 ms mixing time. **A**-**C** and **a**-**c** represent the fully and partially methylated molecules. The polarisation transfer pathways are illustrated on molecule **B**. Cross-peak labels refer to molecules shown in Figure 3.4.

The spectrum correlates the chemical shifts of ^{13}C-labelled methoxy protons in the F_1 dimension with the aromatic protons along the F_2 dimension. Cross-peak analysis clearly shows that the X-filter successfully selected the NOE from the ^{13}C-labelled methoxy groups. As exemplified on molecule **B**, it is also clear that the NOESY transfer took place from the methoxy protons to the ortho protons on the aromatic rings. The NOESY cross-peaks were positive as expected for molecules of this size. Transfer did not extend to the protons meta to the methoxy group in this 600 ms mixing time NOESY experiment. These results indicate that the ^{13}C-filtered NOESY is effective in transferring the magnetisation from labelled methoxy groups to ortho aromatic protons and could be valuable for the analysis of HS samples.

5.4. 2D ^{13}C-filtered NOESY-TOCSY

The aromatic region of the 2D ^{13}C-filtered NOESY-TOCSY spectrum of the ^{13}C-methylated model mixture **I** acquired using the pulse sequences of Figure 4.14 is shown in Figure 5.4.

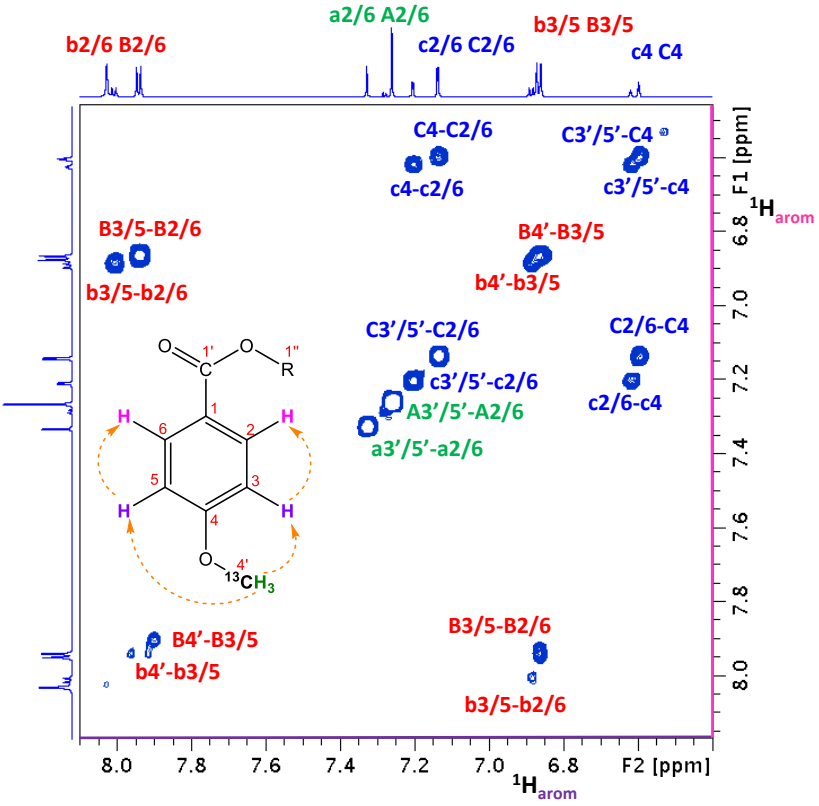


Figure 5.4 2D ^{13}C-filtered NOESY-TOCSY spectrum of ^{13}C-methylated model mixture **I** obtained using pulse sequence shown in Figure 4.14. **A**-**C** and **a**-**c** represent the fully and partially methylated molecules, respectively. Exemplar polarisation transfer pathways are shown on molecule **B**. Cross-peak labels refer to molecules shown in Figure 3.4.

The 2D ^{13}C-filtered NOESY–TOCSY spectrum shows cross-peaks generated from the NOE between methoxy protons and aromatic protons ortho to these groups on the diagonal and the TOCSY transfer between the aromatic protons along the F_2 dimension. The spectrum shows a number of cross-peaks, which have been assigned. For example using molecule **B**, the diagonal cross-peak labelled **b/B** 4'-3/5, originating from the NOESY transfer from the H4' methoxy protons to the aromatic proton H3/5, shows a TOCSY transfer to the H2/6 aromatic protons. Thus, the NOESY-TOCSY experiment is effective in extending the correlations between methoxy and aromatic protons of the same ring via J_{HH} couplings.

5.5. 4D NMR methods

4D NMR spectra cannot be visualised directly and have to be inspected as 3D cuboids on a string (Figure 5.5).

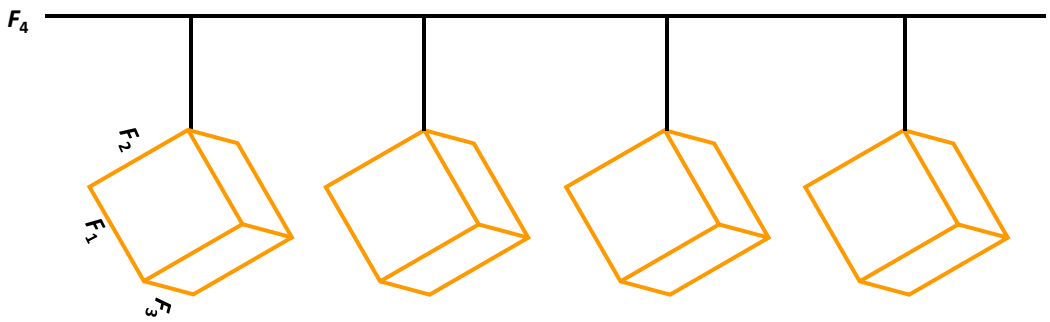


Figure 5.5 Visualisation of 4D NMR spectra as cuboids on a string where one chemical shift axis represents a string and the remaining three correlated chemical shifts are obtained from cuboids hanging from the string.

Any of the four correlated chemical shifts can be chosen to be the string. In this study the methoxy carbons were chosen to be represented by the string. In this chapter two 4D experiments are presented: 4D HCCH$_3$ and 4D HMQC-NOESY-TOCSY.

The 4D HCCH$_3$ experiment correlates the proton and carbon chemical shifts of the methoxy groups with those of the aromatic CH groups in ortho positions to these methoxy groups. The 4D spectrum of ^{13}C-methylated model mixture **I** is shown in Figure 5.6. The string, represented here by the methoxy region of the 1D ^{13}C spectrum of ^{13}C-methylated model mixture **I**, identifies three cuboids for the pairs of methoxy group in the mixture. Due to the limited resolution of the 4D spectrum each cuboid contains signals from both the fully and partially methylated molecules. The cross-peaks located in these cuboids provide the chemical shifts of the methoxy protons, aromatic protons and carbons ortho to these methoxy groups. As illustrated in Figure 5.6, molecules **B/b** show two cross-peaks in the 3D cuboid one for the ester (**B**) and one for the acid (**b**) between the methoxy group at position 4 and the aromatic protons/carbons at 3/5. Similar information is obtained for molecules **A/a** which shows correlations between the methoxy groups at 3/5 to aromatic protons/carbons at 2/6. Molecule **C** however shows two sets of cross-peaks: one set from the correlations from the methoxy group at 3/5 with the aromatic protons at 2/6 and another set from the same methoxy groups with the aromatic proton/carbon at position 4. The latter set of cross-peaks are negative because the carbon at position 4 couples to two fully ^{13}C-labelled methoxy carbons resulting in a change of phase.

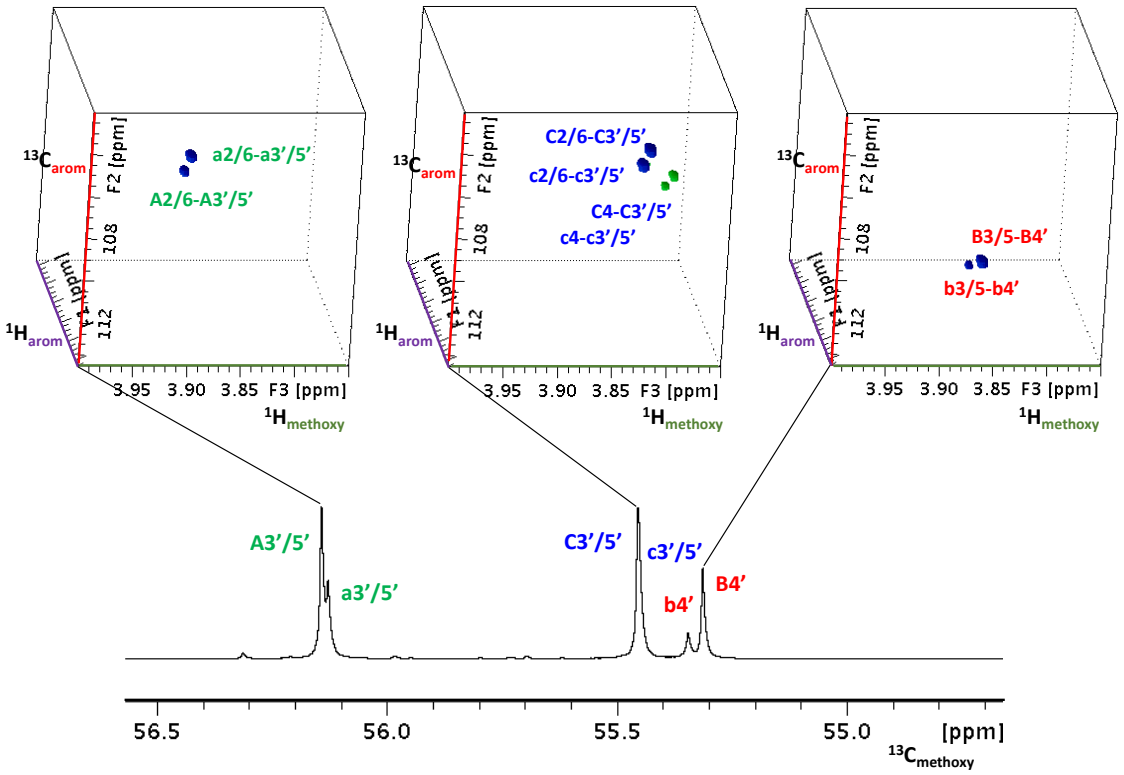


Figure 5.6 4D $HCCH_3$ spectrum of ^{13}C-methylated model mixture **I,** acquired using the pulse sequence shown in Figure 4.8, depicted as cuboids on a string. The cuboids show the correlations between methoxy protons and aromatic protons and carbons. The correlations within each cuboid are connected to a methoxy carbon on the string. **A**-**C** and **a**-**c** represent the fully and partially methylated molecules, respectively. The labels refer to the molecules shown in Figure 3.4.

The 4D HMQC-NOESY-TOCSY experiment extends the information provided by the 2D ^{13}C-filtered NOESY-TOCSY. It separates the individual NOESY-TOCSY planes in a 4D space using the proton and carbon chemical shifts of the methoxy groups. Partial analysis of the 4D HMQC-NOESY-TOCSY spectrum obtained for ^{13}C-methylated model mixture **I** is shown in Figure 5.7. The string represents the methoxy carbon dimension as was the case for the 4D $HCCH_3$ spectrum. Only molecule **B** has aromatic protons ortho to each other and thus will show TOCSY correlations for the short DIPSI-2 mixing time used. Therefore only this cuboid extracted through the methoxy carbon of molecule **B** is shown in Figure 5.7. The cuboid shows NOESY cross-peaks on the diagonal, while the TOCSY cross-peaks are located off the diagonal. It is evident from the 2D NOESY-TOCSY projection of this cuboid that there is a series of NOESY cross-peaks on the diagonal representing correlations between the methoxy protons and aromatic protons of both molecules **B** and **C**. This is because the digital resolution of the ^{13}C methoxy dimension used was low. Molecule **B** has weaker NOESY peaks showing that transfer to a TOCSY partner had occurred. The TOCSY peaks are clearly visible on the dotted horizontal line drawn from the

NOESY peaks. In this particular spectrum decoupling was used during the directly detected period. However if decoupling is removed one can obtain the J_{HH} couplings from the projection.

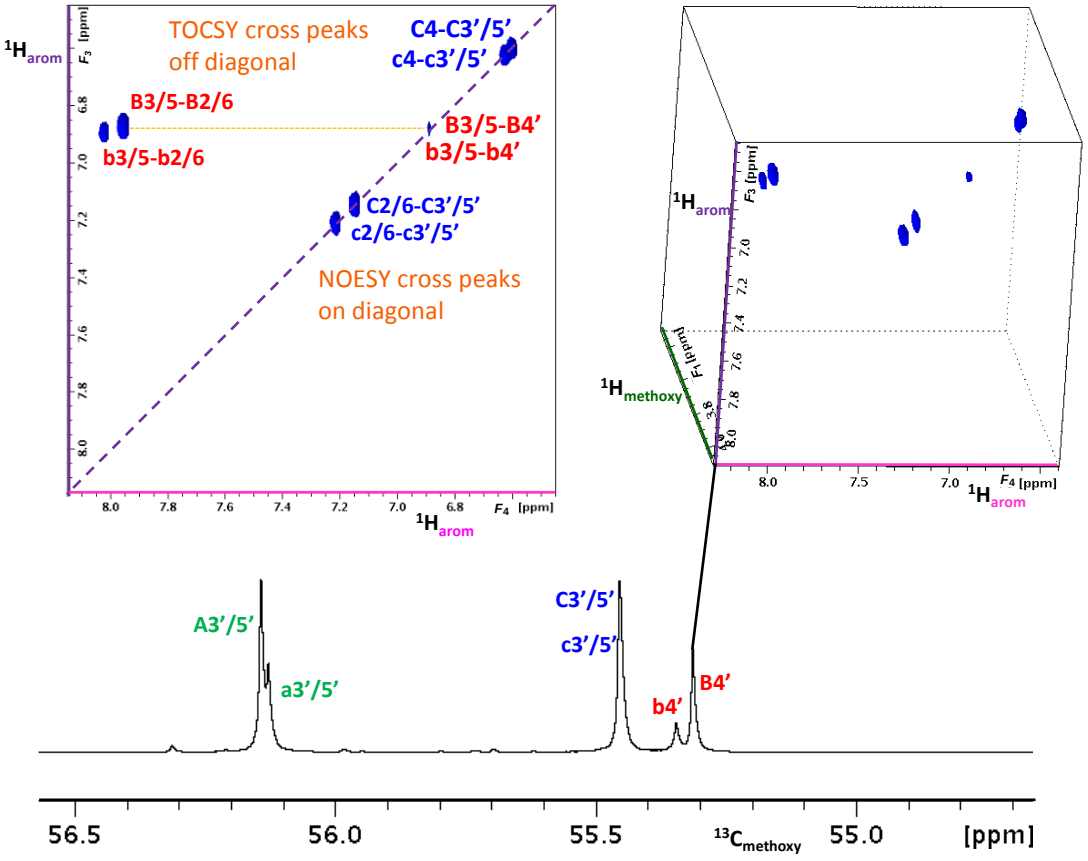


Figure 5.7 4D HMQC-NOESY-TOCSY spectrum of ^{13}C-methylated model mixture **I**, acquired using the pulse sequence shown in Figure 4.17. **A**-**C** and **a**-**c** represent the fully and partially methylated molecules, respectively. The insets shows a 3D cuboid and projection of the cuboid extracted from the ^{13}C methoxy chemical shift of molecules **B/b**. The low resolution in the methoxy carbon dimension resulted in cross-peaks from **C/c** molecule appearing. However due to the nature of this molecule there is no TOCSY transfer expected. The labels refer to the molecules shown in Figure 3.4.

The 4D HMQC-NOESY-TOCSY clearly correlates the methoxy groups with the aromatic protons ortho and meta to these groups. This information will be vital in the NMR analysis of complex mixtures such as HS.

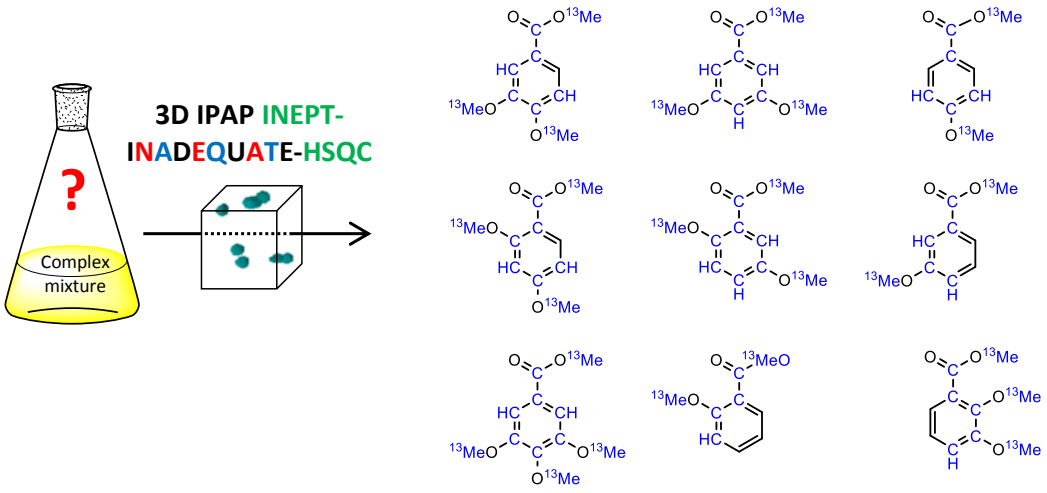
5.6. Chapter conclusions

A series of ^{13}C-filtered NMR experiments were successfully applied to ^{13}C-methylated model mixture **I**. Using ^{13}C-filtered NOESY/NOESY-TOCSY allowed access to chemical shifts of the methoxy protons as well as the aromatic protons ortho and also meta to these groups, while the

4D HMQC-NOESY-TOCSY provides the same chemical shifts with the addition of the methoxy carbon chemical shift. The ^{13}C-filtered HMBC easily provides chemical shifts for each quaternary carbon of the aromatic and carboxyl groups attached to methoxy groups.

While the 4D $HCCH_3$ provides the chemical shifts of the methoxy protons and carbons as well as the aromatic CH pairs ortho to the methoxy groups, the 2D INEPT-INADEQUATE also provides the chemical shifts of aromatic quaternary carbons. However, an additional HSQC spectrum is required to decode the DQ frequencies in the 2D INEPT-INADEQUATE.

Chapter 6. Results and Discussion. Application of the ^{13}C-filtered NMR experiments to model mixture II



N. G. A. Bell, L. Murray, M.C. Graham, D. Uhrín, *Chem. Commun.* **2014**, 50, 1694-1697.

After successful testing of the basic principles of ^{13}C-filtered experiments on ^{13}C-methylated model mixture **I**, additional experiments were developed using a more complex mixture of nine compounds (**1-9**), ^{13}C-methylated model mixture **II** (Figure 6.1). Using this sample, the following experiments were optimised: 3D INEPT-INADEQUATE-HSQC, 3D IPAP-INEPT-INADEQUATE-HSQC, 3D HcCH$_3$, 3D hCCH$_3$, 4D HCCH$_3$, 3D HMQC-HMBC, 3D HMQC-NOESY and 3D HMQC-NOESY-TOCSY. Before the results of these experiments are discussed, a more detailed spectral characterisation of ^{13}C-methylated model mixture **II** is presented.

6.1. 2D NMR characterisation of ^{13}C-methylated model mixture II

The methoxy region of the 2D ^{1}H, ^{13}C HSQC spectrum of ^{13}C-methylated model mixture **II**, acquired using the parameters described in Section 2.14 is shown in Figure 6.1 with the ^{13}C-methylated molecules.

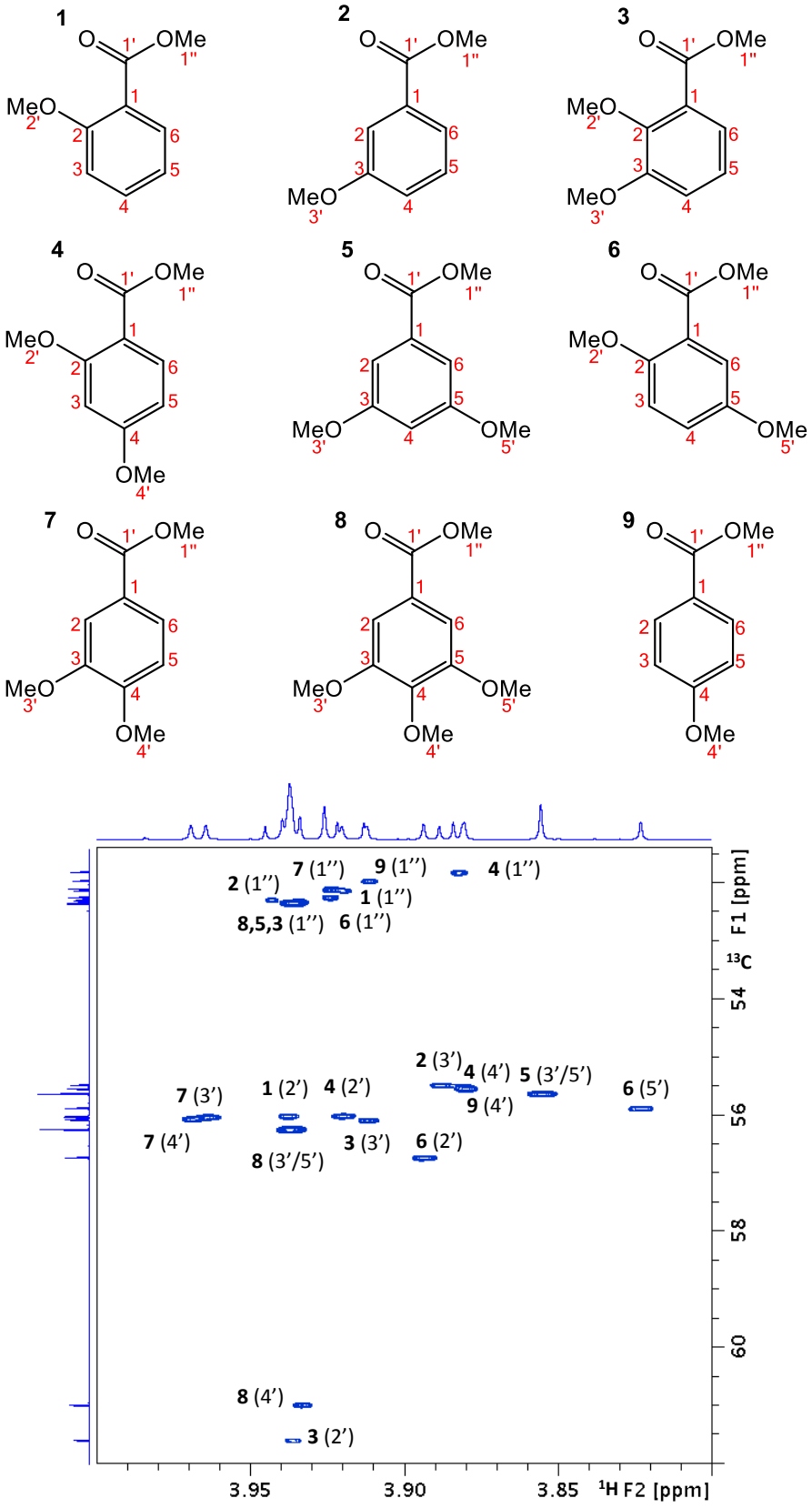


Figure 6.1 Methoxy region of 2D ^{1}H, ^{13}C HSQC spectrum of ^{13}C-methylated model mixture **II**. Labels refer to the molecules shown in the inset above the spectrum.

As can be seen from this spectrum, carbon chemical shifts are separated into three distinct regions: methoxy esters (< 54 ppm), methoxy ethers (54-57 ppm) and methoxy groups sandwiched between two bulky substituents (> 60 ppm). The aromatic region the ^{1}H, ^{13}C 2D HSQC spectrum of ^{13}C-methylated model mixture **II** is shown in Figure 6.2.

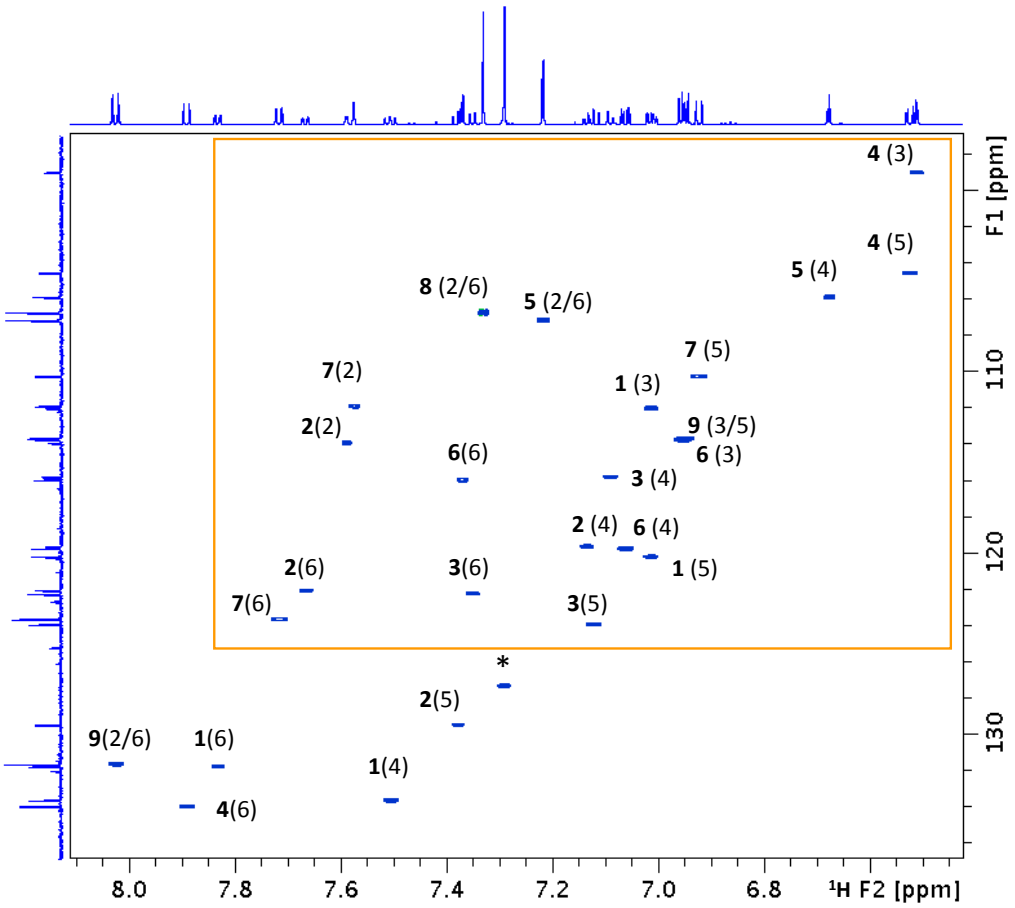


Figure 6.2 The aromatic region of the 2D ^{1}H-^{13}C HSQC of ^{13}C-methylated model mixture **II**. Labels refer to the molecules shown in Figure 6.1. The highlighted section contains resonances belonging to nuclei ortho or para to methoxy groups.

The spectrum has a diagonal appearance due to the influence of correlated ortho and para effects of the ester and ether groups on both types of nuclei. The orange rectangle in Figure 6.2 shows the spectral region where the nuclei ortho and para to methoxy groups are found.

The protonated and quaternary carbons of ^{13}C-methylated model mixture **II** were assigned by analysing the ^{1}H and ^{13}C spectra (Figure 3.6 and 3.7) as well as the ^{1}H, ^{13}C HSQC (Figure 6.1 and

6.2) and ^{1}H, ^{13}C HMBC spectra (data not shown). All spectra indicated that the mixture contained only methylated compounds with little indication of the starting material.

In order to characterise the sample further, a 2D DOSY spectrum was obtained (Figure 6.3).

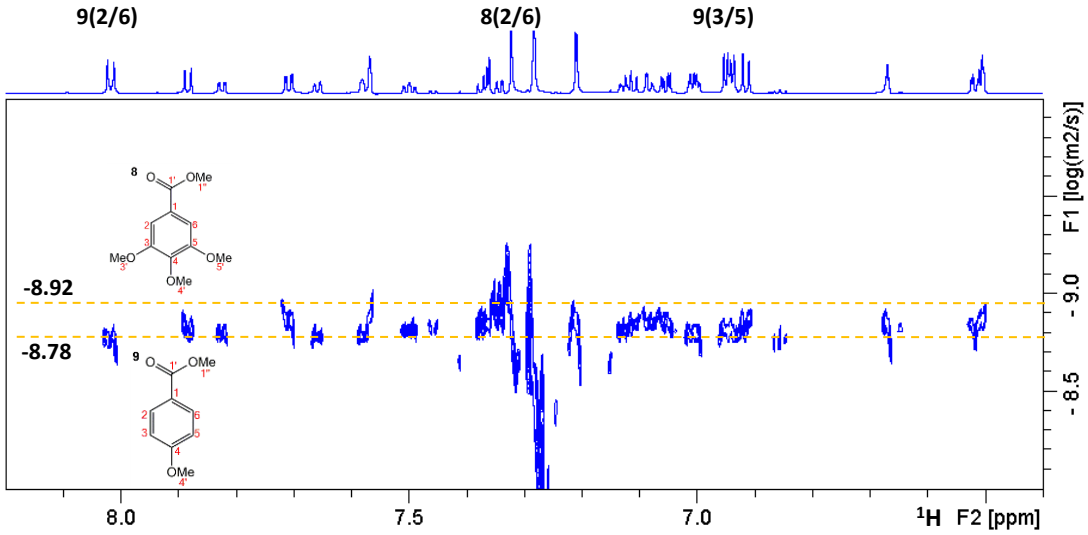


Figure 6.3 2D DOSY spectrum of ^{13}C-methylated model mixture **II.** The structures of the fastest (molecules **9**) and slowest (molecules **8**) diffusing molecules are shown together with corresponding logD values. All other molecules fall between these values indicated by the dashed lines.

The 2D DOSY spectrum shows the decadic logarithm of the diffusion coefficient along the F_1 dimension and the ^{1}H chemical shift along F_2. Signals from the same molecule should have the same diffusional coefficient. The 2D DOSY spectrum of the ^{13}C-methylated model mixture **II** was acquired in order to (i) obtain a reference for the ^{13}C-methylated RMFA sample and (ii) to inspect the variations of diffusion coefficients between the similar molecules of ^{13}C-methylated model mixture **II**. The spectrum shown in Figure 6.3 shows the diffusion values fall between logD of -8.78 and -8.92. In line with expectations, based on their molecular size and the relative position of the carboxylic and phenolic groups, these extreme values correspond to molecules **9** and **8**.

6.2. Application of ^{13}C-filtered NMR experiments on ^{13}C-methylated model mixture II

The 3D INEPT-INADEQUATE-HSQC, 3D IPAP-INEPT-INADEQUATE-HSQC, 3D HMQC-HMBC, 3D HMQC-NOESY, 3D HMQC-NOESY-TOCSY, 3D $HcCH_3$, 3D $hCCH_3$ and 4D $HCCH_3$ spectra were acquired on ^{13}C-methylated model mixture **II** to investigate and compare their performance before applying them to the RMFA sample. This section will demonstrate how these

experiments provide chemical shifts and coupling constants, using molecule **6** of ^{13}C-methylated model mixture **II** to illustrate the process of structure determination. The chemical shifts/coupling constants obtained from all other molecule are listed in Appendix B (see external web link).

6.2.1. 3D (IPAP) INEPT-INADEQUATE-HSQC

The 2D INEPT-INADEQUATE was illustrated on ^{13}C-methylated model mixture **I** in Chapter 5.1. Two 3D versions of the experiment were applied to ^{13}C-methylated model mixture **II**: the refocused/decoupled and the IPAP version introduced in Sections 4.1.1.1. The benefits of the IPAP version can be seen on the methoxy region of a ^{1}H-^{13}C HSQC, which illustrates the increased resolution obtainable with IPAP implemented (Figure 6.4).

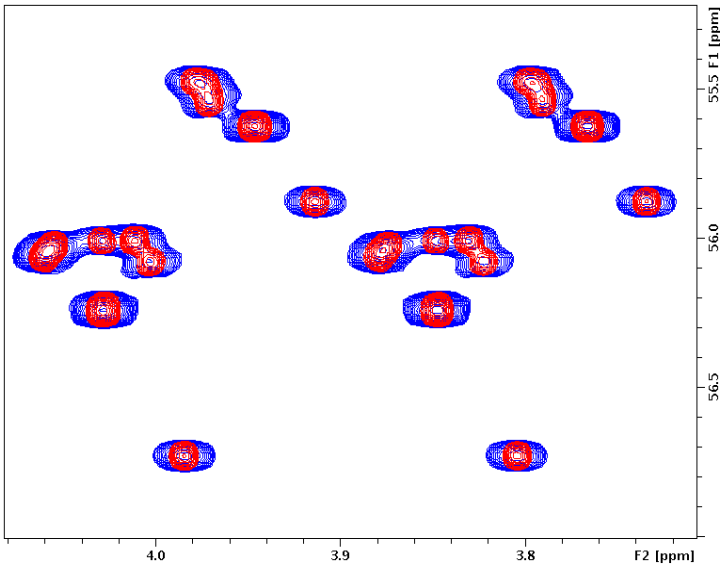


Figure 6.4 Overlay of two IP 2D ^{1}H, ^{13}C HSQC spectra showing the methoxy region of ^{13}C-methylated model mixture **II**. Both spectra were acquired without ^{13}C decoupling. The acquisition time used for the blue spectrum was 150 ms, i.e., the typical time for ^{13}C decoupled spectrum, while the red spectrum was acquired over 250 ms, the value used for the IPAP procedure.

The IPAP experiment does not use ^{13}C decoupling and thus the CH correlations are split by their $^{1}J_{CH}$ coupling constant. Two spectra are acquired in an interleaved mode, showing in-phase or anti-phase $^{1}J_{CH}$ doublets. As high-resolution 3D spectra produce large files and the cross-peaks in the 3D INEPT-INADEQUATE-HSQC are confined to specific spectral regions, the spectra were processed in a strip mode, to produce separate IP and AP 3D spectra for the methoxy and aromatic regions. The corresponding IP and AP spectra are then added and subtracted producing simplified spectra. These are now similar to the regular spectra obtained using ^{13}C decoupling, except for the position of the cross-peaks, which are now offset by $\pm\ ^{1}J_{CH}/2$ from their ^{1}H chemical

shifts. If the $^{1}J_{CH}$ splitting is uniform for a particular group of CH resonances these two spectra can be recombined by shifting them left or right along the F_3 dimension by $^{1}J_{CH}/2$. In this way, the sensitivity of the experiment is restored with an additional benefit of better resolution in the ^{1}H dimension. The 3D IP-INEPT-INADEQUATE-HSQC spectrum obtained from ^{13}C-methylated model mixture **II** shows an abundance of cross-peaks in both the aromatic and methyl regions (Figure 6.5).

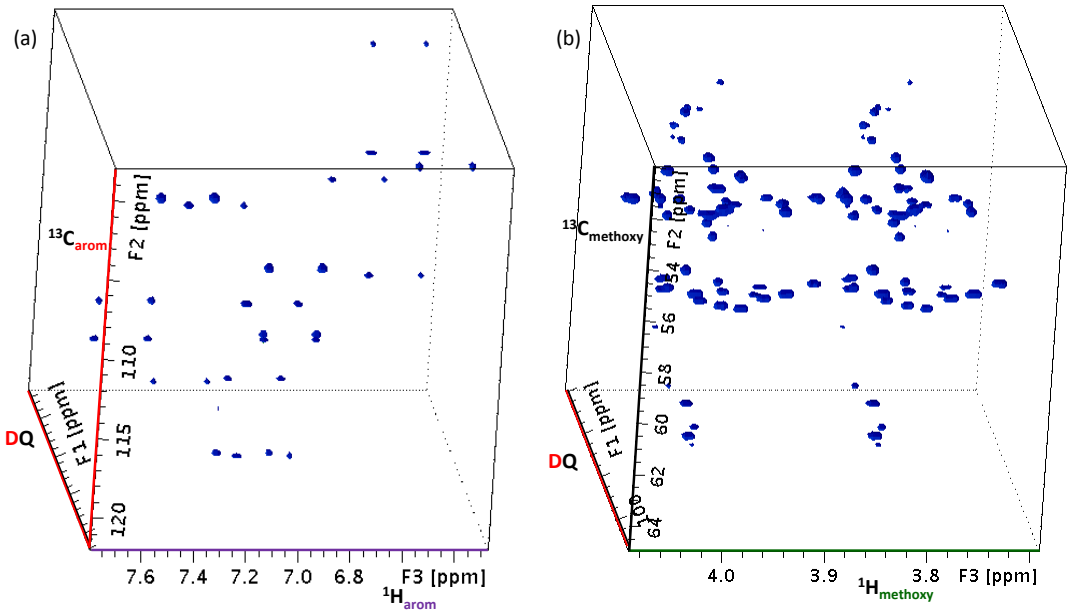


Figure 6.5 The (a) aromatic and (b) methoxy regions of the 3D IP-INEPT-INADEQUATE-HSQC spectrum, acquired using the pulse sequence shown in Figure 4.4, of ^{13}C-methylated model mixture **II**. The cross-peaks are split by the $^{1}J_{CH}$ coupling in F_3.

The F_1F_3 projections of both spectral regions, generated after combining the IP and AP spectra, are shown in Figure 6.6. This projection shows cross-peaks originating from the DQ correlations between methoxy and aromatic carbons. Correlations involving protonated aromatic carbons have two cross-peaks, one each in the aromatic and methoxy regions sharing the same DQ frequency, while those belonging to quaternary carbons only have one cross-peak in the methoxy region.

As explained in Section 4.1.1.1, the spectral width in the DQ dimension was set to 75 ppm and φ_1 was incremented by 80° every time the t_1 interval was increased. This represents the optimal sampling of the DQ dimension without the need for signal aliasing.

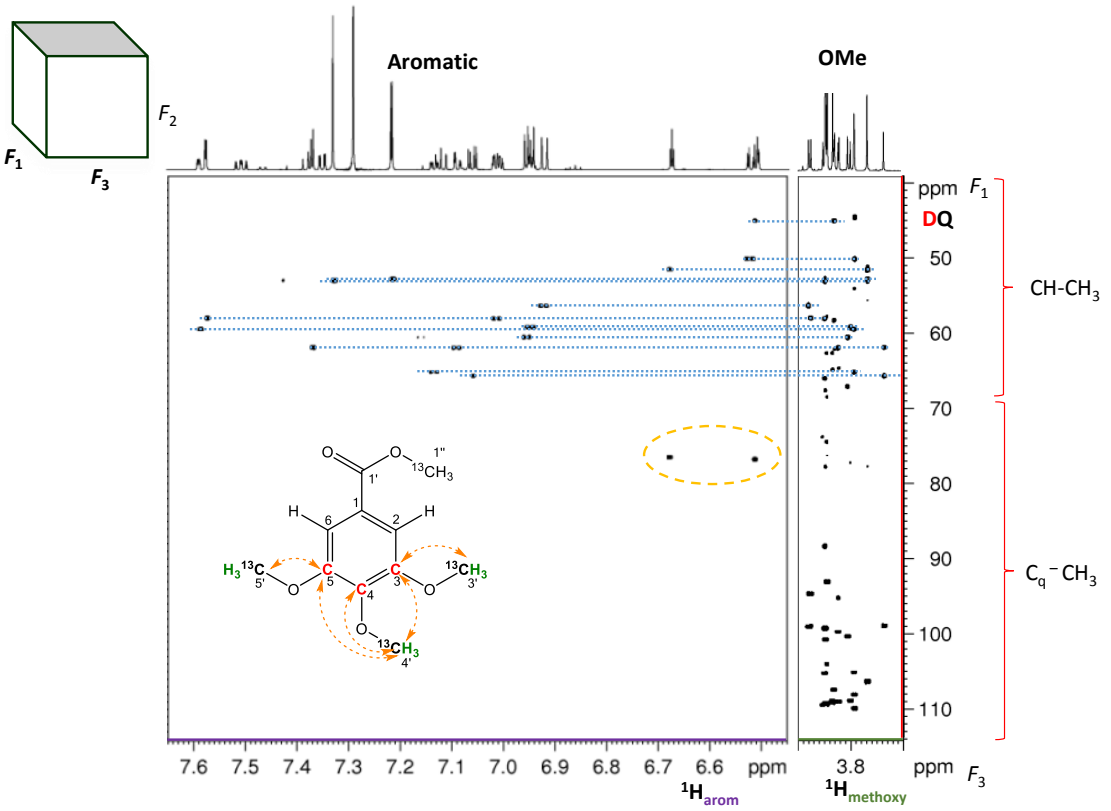


Figure 6.6 2D F_1F_3 projections of the 3D IPAP-INEPT-INADEQUATE-HSQC spectra showing the aromatic (left) and methoxy (right) regions of ^{13}C-methylated model mixture **II**. The 3D cube displayed at the top left indicates which face of this projection is displayed. The dotted lines connect the methoxy and aromatic resonances sharing the same DQ frequency. These only appear if a protonated carbon is next to a methoxy group. The ^{13}C carrier frequency (110 ppm) was not re-referenced to 0 hence this needs to be taken into account when calculating the SQ chemical shifts. The inset shows the observed correlations on an exemplar structure. The circled cross-peaks are examined further below.

An exception were the circled signals also shown in Figure 6.6, which belong to DQ coherences created between the carbons of two -O^{13}CH$_3$ groups coupled to the same aromatic carbons as in molecules **4** and **5**. Their true DQ frequencies are ~ 1 ppm (2 x 55.5 – 110 ppm), hence they are aliased here. Their origin can be explained by following the evolution of magnetisation using the product spin operators with reference to the pulse sequence of Figure 4.4. The heteronuclei S and R_1 and R_2 refer to the aromatic carbon ortho to two methoxy groups and the two methoxy carbons, respectively.

$$\hat{I}_z \xrightarrow{(\pi/2)\hat{I}_y} \hat{I}_x \xrightarrow{\pi\hat{I}_x,\pi\hat{S}_x(2\Delta_1)} 2\hat{I}_y\hat{S}_z$$

$$\xrightarrow{(\pi/2)\hat{I}_x,(\pi/2)\hat{S}_x} -2\hat{I}_z\hat{S}_y \xrightarrow{\pi\hat{S}_x,\hat{R}_{1x},\hat{R}_{2x}(2\Delta_2)} -8\hat{I}_z\hat{S}_y\hat{R}_{1z}\hat{R}_{2z}\sin(\pi J_{R1,S}2\Delta_2)\sin(\pi J_{R2,S}2\Delta_2) \xrightarrow{(\pi/2)\hat{S}_x} 8\hat{I}_z\hat{S}_z\hat{R}_{1y}\hat{R}_{2y}$$

Followed by the back conversion to SQ ^{13}C coherences, refocusing of $^{n}J_{CC}$ couplings and a reverse INEPT step, the $2R_{1y}R_{2y}$ DQ coherences between the two methoxy carbons are detected on the

aromatic proton *I*. These signals are therefore expected to appear when a CH pair is sandwiched between two methoxy groups, as is the case for molecules **4** and **5**.

To determine the chemical shifts of carbons involved in the DQ coherences one requires the SQ chemical shift of one carbon. The strength of the 3D INEPT-INADEQUATE-HSQC is that this information is contained in the same spectrum as seen in the F_2F_3 projection (Figure 6.7), which represents the 2D HSQC spectrum.

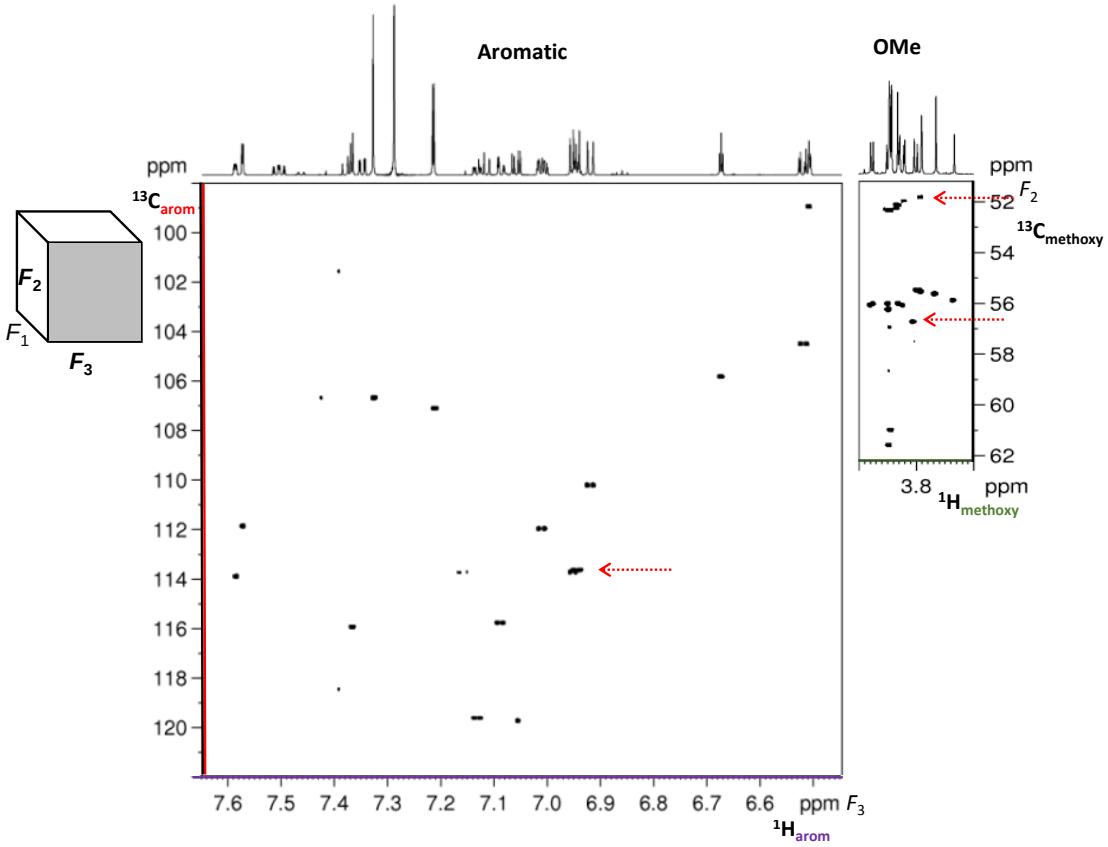


Figure 6.7 2D F_2F_3 projection of the aromatic and methoxy regions of the 3D IPAP-INEPT-INADEQUATE-HSQC spectrum of ^{13}C-methylated model mixture **II**. The 3D cube displayed at the top left indicates which face of this projection is displayed. The red arrows indicate cross-peaks analysed further in Figure 6.8.

The F_2F_3 projections are good indicators of the level of information obtainable from this sample. By comparing the methoxy region of the 2D ^{1}H, ^{13}C HSQC with the HSQC projection the number of molecules which yielded signal can be assessed. In the case of ^{13}C-methylated model mixture **II** signals of all molecules were visible. Assessing the aromatic region is more difficult, as the number of protonated carbons in non-ortho positions is generally not known. However, for ^{13}C-methylated model mixture **II** all ortho CH pairs were present in the spectrum.

Analysis of the 3D IPAP-INEPT-INADEQUATE-HSQC spectra starts with the identification of the cross-peaks in the methoxy part of the HSQC projection and extraction of 2D planes through either the methoxy proton or methoxy carbon chemical shifts. In this and all other spectra presented in this work, the 2D planes are obtained at specific methoxy carbon chemical shifts. An example of the 2D planes obtained for the methoxy groups at 52.21 ppm and 56.67 ppm are shown in Figure 6.8b and Figure 6.8c.

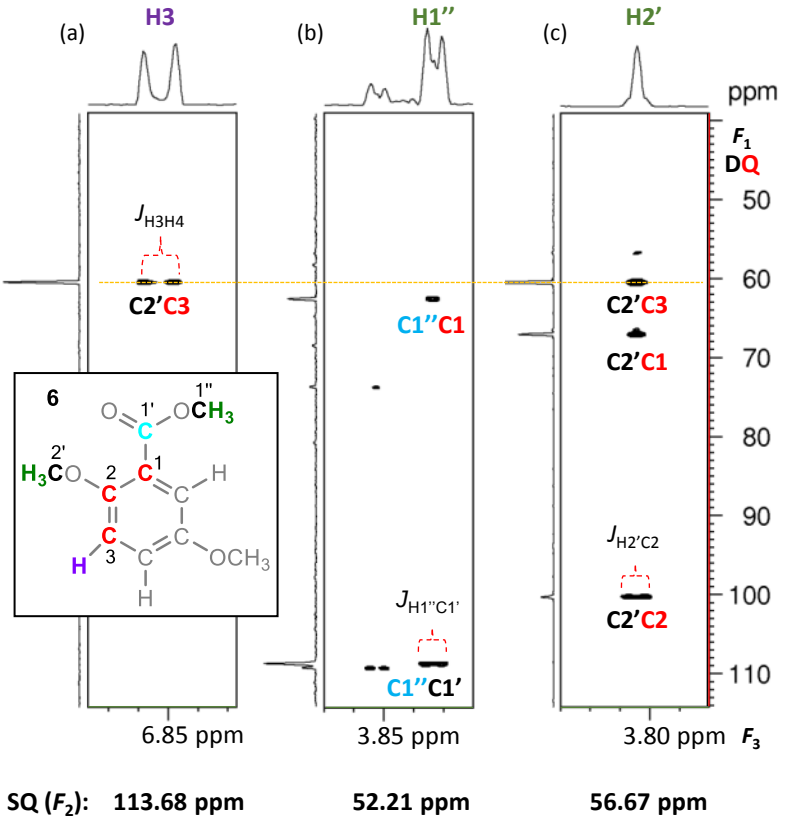


Figure 6.8 2D F_1F_3 IPAP-INEPT-INADEQUATE-HSQC planes extracted via the (a) aromatic (b) methoxy ester (c) methoxy ether carbons of molecule **6** of ^{13}C-methylated model mixture **II**. The inset highlights the nuclei whose chemical shifts are obtainable from analysis of these planes.

The most informative planes are the F_1F_3 planes which show correlations of the methoxy protons and the DQ coherences involving methoxy and aromatic carbons. In the case of molecule **6**, the plane at 56.67 ppm (Figure 6.8c) shows three cross-peaks. Using the DQ frequency, the carbon chemical shifts of the observed cross-peaks (SQ_2) on each plane can be calculated as:

$$SQ_2 = DQ + O2P - SQ_1 \quad [13]$$

Where SQ_1 is the chemical shift of the carbon in the F_2 dimension, while O2P is the carrier frequency. Using this relationship, the chemical shifts of the three carbons involved in a DQ

correlation with the methoxy carbon at 56.67 ppm can be calculated as 120.24, 153.47 and 113.68 ppm. The carbon at 153.47 ppm (DQ frequency of 100.17 ppm) yields a cross-peak due to $^{2}J_{CC}$ coupling between the methoxy carbon and the quaternary ipso carbon. This cross-peak is split by the $^{3}J_{CH_3, Ci}$ coupling involving the methoxy protons and the ipso carbon. The cross-peaks at 113.68 and 120.24 ppm (DQ frequencies of 60.38 and 66.94 ppm) are due to the correlations with aromatic carbons ortho to the methoxy group. Only one of these shows a cross-peak at the same DQ frequency in the aromatic region of the spectrum (Figure 6.8a), implying that one ortho carbon is protonated while the other is quaternary. Once a protonated aromatic carbon has been identified the shape of the DQ cross-peak in the aromatic region of the spectrum is inspected. If this proton is next to another proton, as was the case for molecule **6**, the cross-peak will be split by a large ortho coupling. Next, the SQ carbon chemical shift of the quaternary carbon (120.24 ppm) is inspected. If a $COO^{13}CH_3$ group is attached to this carbon, it should appear in a correlation involving a methoxy carbon of an ester group. Methoxy ester protons typically show two correlations, one with the quaternary ipso carbon and the other with the carbonyl groups, mediated by $^{3}J_{CC}$ and $^{2}J_{CC}$ couplings, respectively. In the case of molecule **6**, the SQ chemical shift of the quaternary carbon (120.24 ppm) was found to yield a cross-peak with a methoxy ester groups at 52.21 ppm. Thus a F_1F_3 plane extracted at 52.21 ppm (Figure 6.8b) shows the expected pattern: two cross-peaks, one with the quaternary ipso carbon and the other with the carbonyl carbon. The latter cross-peak is split by a $^{3}J_{CC}$ coupling constant. This information led to the conclusion that molecule **6** has one methoxy group ortho to an ester group.

Similarly, the methoxy group at C5 of molecule **6** was analysed. This methoxy group yielded assignment of the remaining C4H4 and C6H6 pairs. Thus the chemical shifts of all carbons of molecule **6** could be completed from the analysis of the 3D IPAP-INEPT-INADEQUATE-HSQC spectrum alone. However, in complex mixtures ambiguity could arise and additional experiments, in particular the 3D HMQC-NOESY-TOCSY, would likely be required to connect the two fragments together.

The results shown in this section focused on the IPAP version of the INEPT-INADEQUATE-HSQC. The ^{13}C decoupled non-IPAP spectrum was also acquired. This spectrum provided identical information as presented above, therefore this data is not discussed here.

6.2.2. 4D HCCH$_3$

As illustrated already for ^{13}C-methylated model mixture **I**, the 4D HCCH$_3$ experiment is a powerful method for obtaining the assignment of aromatic CH resonances ortho to the methoxy groups.

Using the pulse sequence shown in Figure 4.8, a 4D HCCH$_3$ spectrum was also obtained for ^{13}C-methylated model mixture **II**. To illustrate the results, the 3D spectra extracted for the methoxy groups of molecule **6** are shown in Figure 6.9.

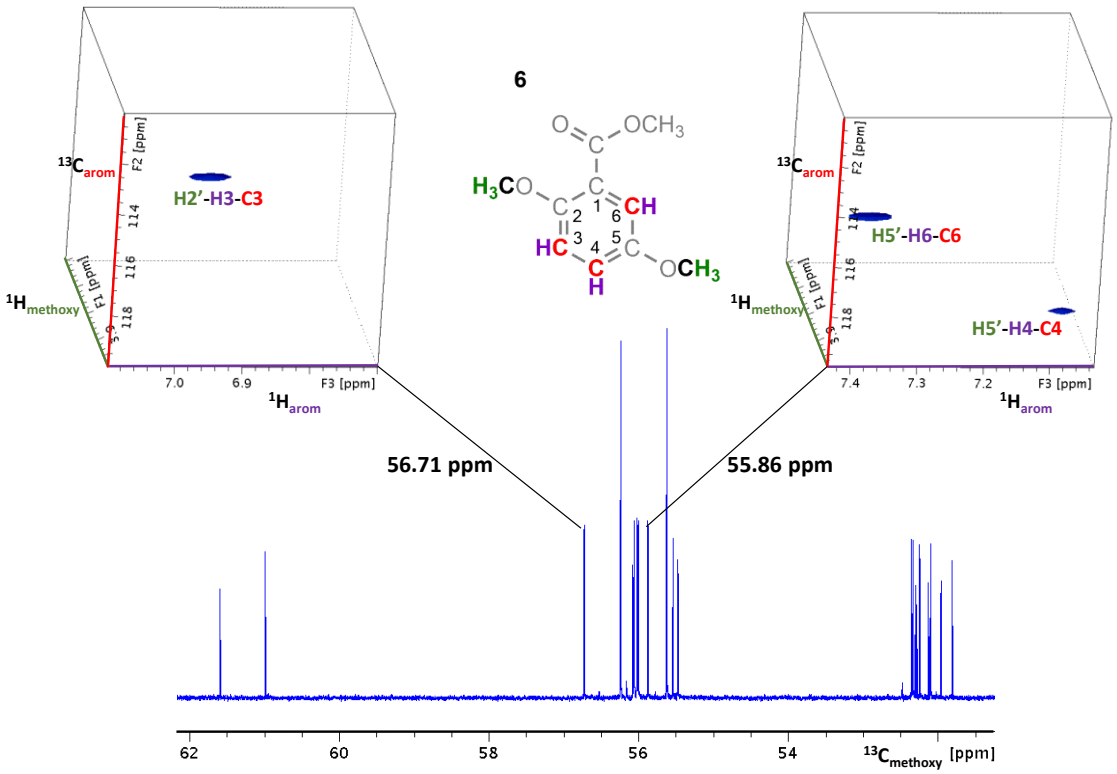


Figure 6.9 4D HCCH$_3$ spectrum of ^{13}C-methylated model mixture **II** obtained using the pulse sequence shown in figure 4.8. Cuboids extracted at methoxy carbons at 55.86 ppm and 56.71 ppm indicate the correlations from the methoxy carbons of molecule **6**. Inset shows molecule **6**, with nuclei whose chemical shifts are obtained highlighted.

These show one cross-peak arising from the correlations to the C2 methoxy group and two cross-peaks arising from the C5 methoxy group. As can be seen in Figure 6.9 the 4D HCCH$_3$ allows separation of individual resonances associated with each methoxy group and thus is very useful for cases where there is overlap in more than one dimension. However, as 4D experiments are generally less sensitive than 3D experiments and their digital resolution is limited, two 3D versions of these experiments were also tested on ^{13}C-methylated model mixture **II**.

6.2.3. 3D hCCH$_3$ and HcCH$_3$

3D hCCH$_3$ and 3D HcCH$_3$ experiments label either the carbon or proton chemical shift, respectively, of a CH moiety next to a methoxy group. The corresponding 3D spectra of ^{13}C-methylated model mixture **II** acquired using the 3D versions of the pulse sequence shown in Figure 4.8, are shown in Figure 6.10.

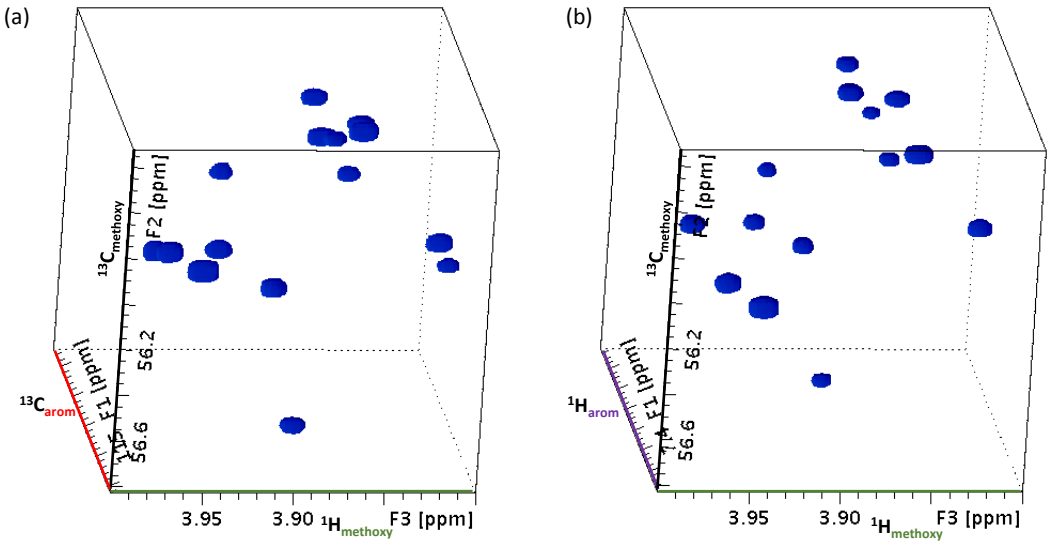


Figure 6.10 3D (a) hCCH$_3$ and (b) HcCH$_3$ spectra of ^{13}C-methylated model mixture **II** acquired using the 3D version of the pulse sequence shown in Figure 4.8.

To examine these spectra further, their 2D F_1F_3 projections were obtained (Figure 6.11).

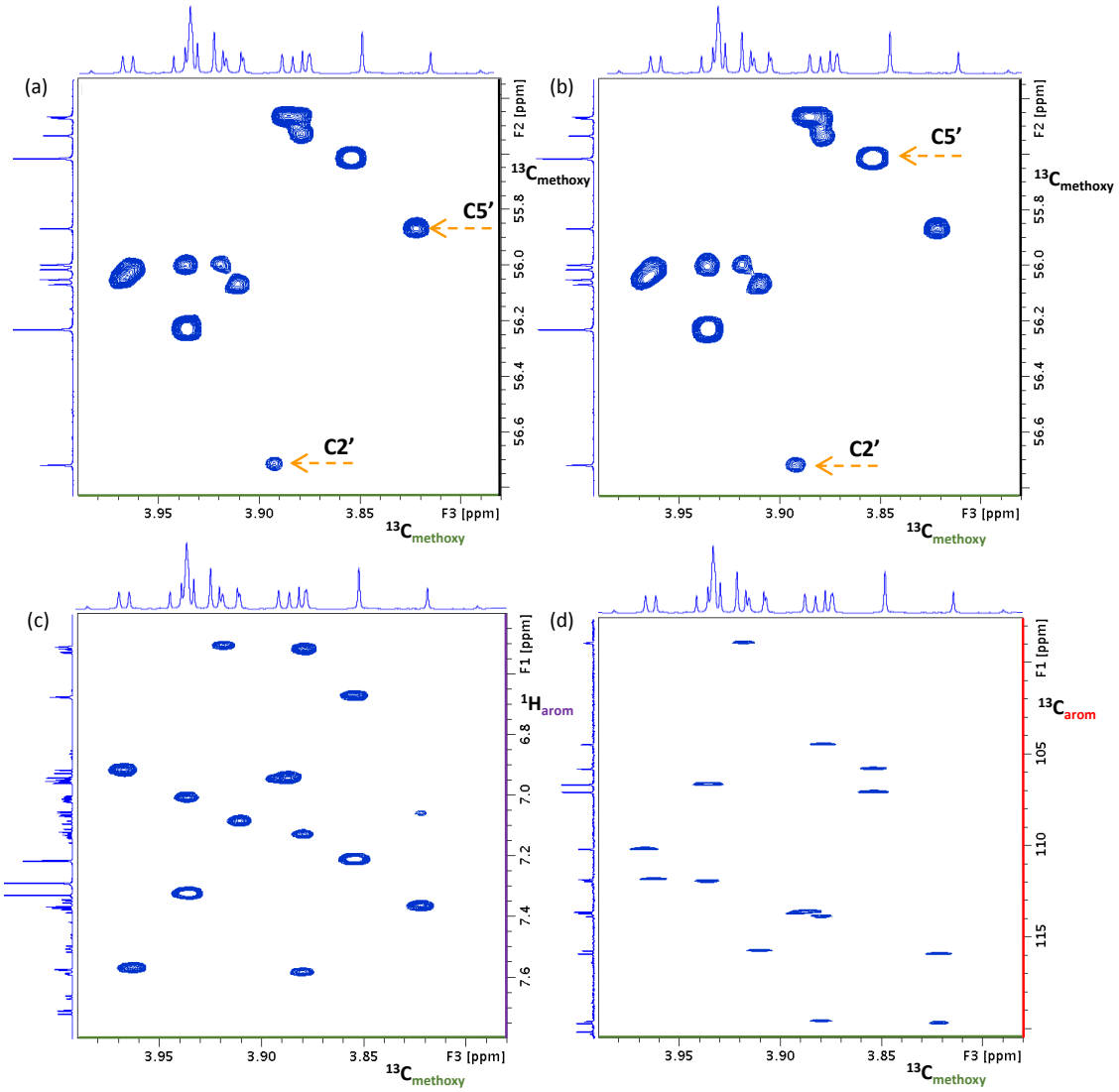


Figure 6.11 F_2F_3 projections of the (a) 3D $HcCH_3$ and (b) 3D $hCCH_3$ and F_1F_3 projections of the (c) 3D $HcCH_3$ (d) and 3D $hCCH_3$ spectra shown in Figure 6.10**.** The arrows indicate the methoxy groups of molecule **6**.

There is some overlap in the cross-peaks in the F_2F_3 projections but the F_1F_3 projections show the expected number of cross-peaks (15) for ^{13}C-methylated model mixture **II** representing all possible correlations between methoxy group and aromatic CH moieties ortho to these groups. Only two methoxy groups in the mixture do not show correlations in these spectra due to the absence of an ortho aromatic CH moiety. The methoxy resonances belonging to molecule **6** are indicated with arrows in Figure 6.11 a and b. The 2D F_1F_3 planes extracted through these chemical shifts are shown in Figure 6.12.

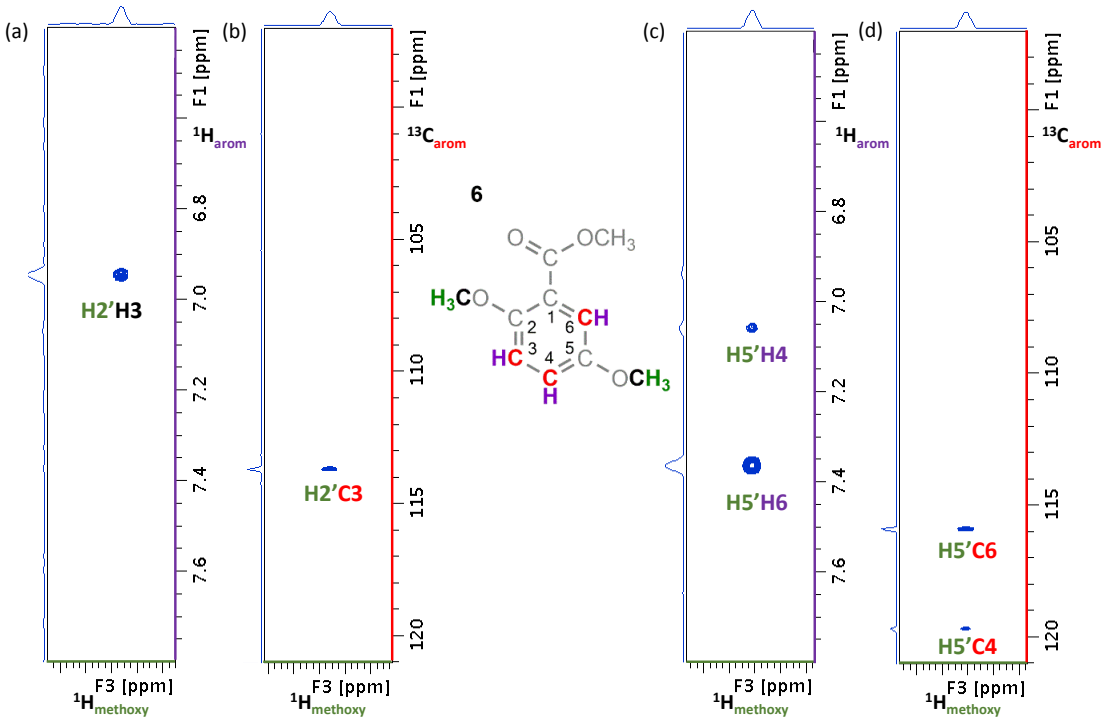


Figure 6.12 (a) and (b) show 2D F_1F_3 planes of the 3D $HcCH_3$ and 3D $hCCH_3$ spectra, respectively extracted at the methoxy carbon at 56.71 ppm (C2'), while (c) and (d) show the same planes for the methoxy carbon at 55.86 ppm (C5'). The inset shows molecule 6 with the nuclei whose chemical shifts were obtained highlighted.

The chemical shifts obtained from the 3D experiments were identical to those obtained from the 4D version. The sensitivity of the 3D and 4D spectra is compared later in this chapter.

6.2.4. 3D HMQC-HMBC

A 3D HMQC-HMBC spectrum of ^{13}C-methylated model mixture **II** obtained by the pulse sequence of Figure 4.11 in shown in Figure 6.13.

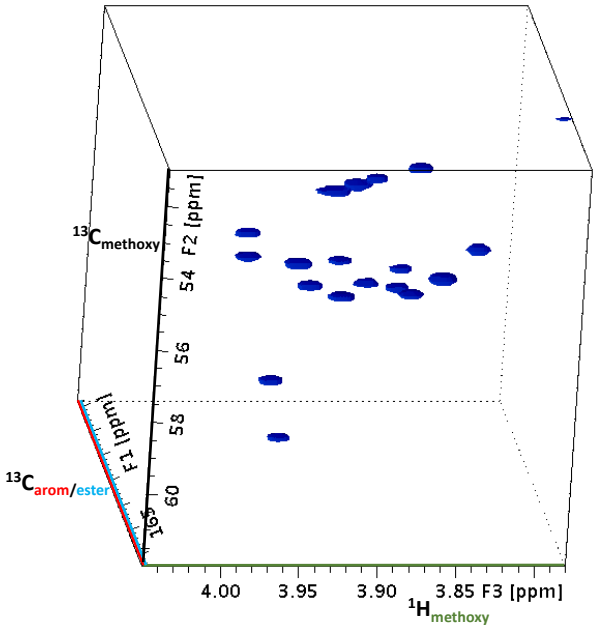


Figure 6.13 3D HMQC-HMBC spectrum of ^{13}C-methylated model mixture **II** acquired using the pulse sequence shown in Figure 4.11.

All 23 cross-peaks expected for ^{13}C-methylated model mixture **II** are present, although some overlap was observed for the ester cross-peaks. As these cross-peaks carry little structural information, the focus is on the ipso carbons carrying the methoxy groups. To analyse the region of the spectrum where these cross-peaks appear the 2D F_1F_3 and F_2F_3 projections were obtained (Figure 6.14).

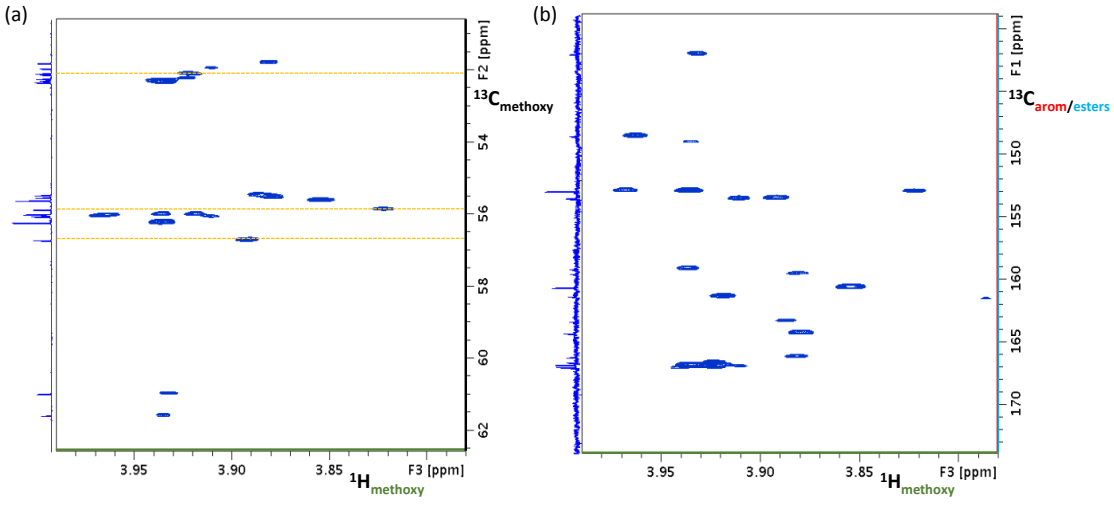


Figure 6.14 (a) F_2F_3 and (b) F_1F_3 projections of the 3D HMQC-HMBC spectrum of ^{13}C-methylated model mixture **II.** The dotted lines indicate the chemical shifts of methoxy carbon of molecule **6**.

The 2D F_2F_3 projection is equivalent to the methoxy ^{1}H, ^{13}C HSQC spectrum, while the F_1F_3 projection shows the correlations between the methoxy proton and quaternary aromatic carbons. Three F_1F_3 2D planes extracted at the methoxy carbon chemical shifts indicated by the dotted lines in Figure 6.14, are shown in Figure 6.15.

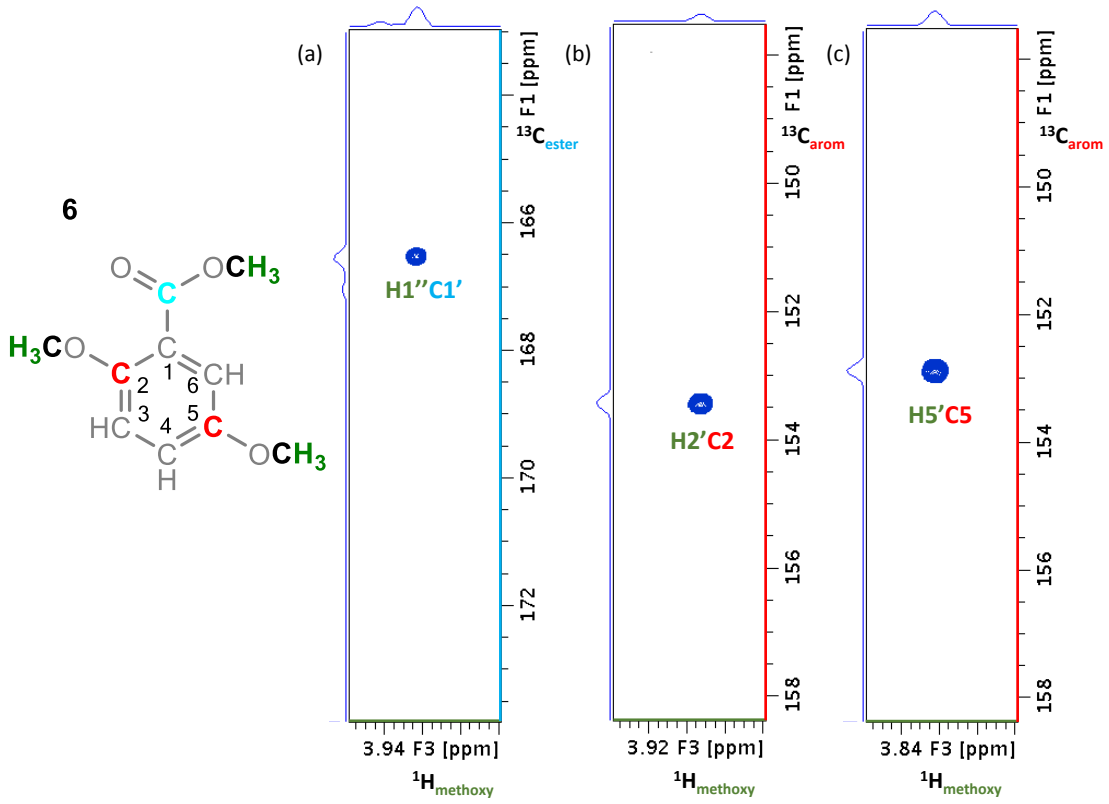


Figure 6.15 2D F_1F_3 planes obtained from the 3D HMQC-HMBC spectrum extracted from the F_2 methoxy carbon chemical shifts for molecule **6** at (a) 52.22 ppm; (b) 56.70 ppm; (c) 55.86 ppm. The inset shows molecule **6** highlighting the nuclei whose chemical shifts are obtained from this experiment.

Each F_1F_3 plane shows only one cross-peak belonging to C1', C2 and C5, respectively. Outside of the context of the other experiments, the data obtained from the 3D HMQC-HMBC is of limited value. However, this rather sensitive experiment is useful for confirming the assignments obtained from the 3D INEPT-INADEQUATE-HSQC spectra, especially if both carbons ortho to the methoxy group are quaternary.

6.2.5. 3D HMQC-NOESY

A 4D HMQC-NOESY-TOCSY (Section 5.5) experiment was tested on ^{13}C-methylated model mixture **I**, which gave information about protons ortho and meta away from the methoxy groups in one spectrum. To build the molecular information bit by bit, it was decided to perform two

3D experiments on ^{13}C-methylated model mixture **II**, the first being the 3D HMQC-NOESY. This experiment provides correlations between the aromatic protons that are close in space to the methoxy protons (aromatic protons ortho to the methoxy groups) and the nuclei of the methoxy groups. The 3D spectrum obtained from ^{13}C-methylated model mixture **II** contained intense negative auto correlation peaks of methoxy protons (not shown) while the cross-peaks of interest were contained only in the aromatic region of the spectrum. 2D F_1F_3 and F_2F_3 projections were thus obtained focusing on the aromatic protons (Figure 6.16).

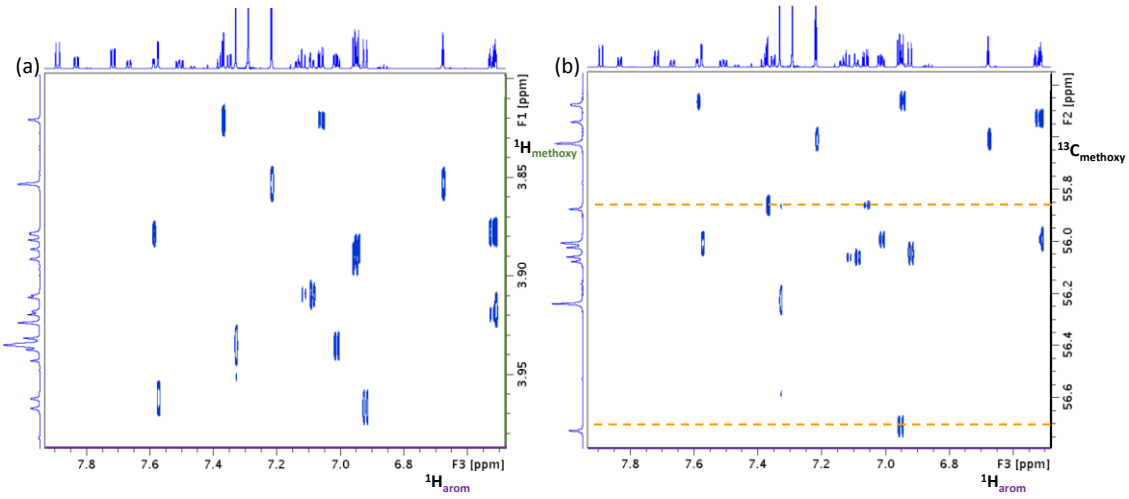


Figure 6.16 Aromatic region of the 2D (a) F_1F_3 and (b) F_2F_3 projections of the 3D HMQC-NOESY spectrum of ^{13}C-methylated model mixture **II**, acquired using the pulse sequence shown in Figure 4.15. The dotted lines indicate the methoxy carbon chemical shifts of molecules **6**.

The F_1F_3 projection shows the methoxy proton to aromatic proton correlations, while the F_2F_3 projection shows the methoxy carbon to aromatic proton correlations. Both projections show 1 or 2 cross-peaks in rows representing the methoxy carbon or proton. The projections could be used to extract the chemical shifts and coupling constants of the aromatic protons. However, it is clearer, especially in areas of overlap, to extract rows for each methoxy chemicals shift. Therefore using the 2D F_2F_3 projection, two planes extracted through the two methoxy carbons of molecule **6,** indicated by the dotted lines in Figure 6.16, are shown in Figure 6.17.

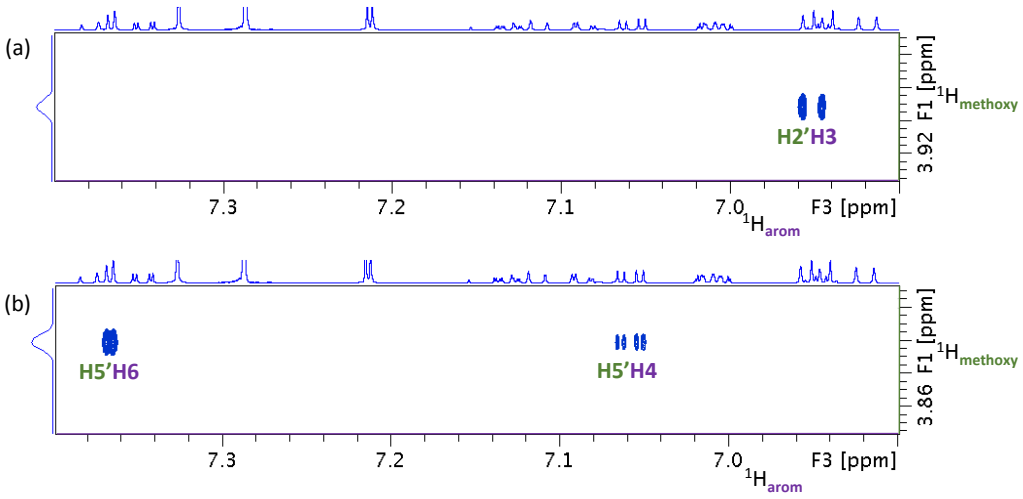


Figure 6.17 2D F_1F_3 planes of the 3D HMQC-NOESY spectrum shown in Figure 6.16 extracted through methoxy carbons of molecule **6** at (a) 56.71 (C2') and (b) 55.84 (C5').

The 2D plane extracted through the methoxy chemical shift of 55.84 ppm shows two cross-peaks, one intense doublet (3.1 Hz) at 7.36 ppm and one doublet of doublets at 7.06 (9.1, 3.3 Hz) ppm. The other plane extracted at 56.71 ppm shows one doublet (9.0 Hz) at 6.95 ppm. This information implies that the first methoxy group has two protons in ortho positions, while only one proton is next to the second methoxy group. The sizes of the coupling constants suggest that these three protons are part of the same spin system.

6.2.6. 3D HMQC-NOESY-TOCSY

Although it is possible to label all four chemical shifts of the nuclei involved in the polarisation transfer pathway of the 4D HMQC-NOESY-TOCSY experiment, as was shown for ^{13}C-methylated model mixture **I** in Section 5.5, it is preferable to reduce its dimension and not label the chemical shift after the NOESY step. Instead, the identity of the NOESY peaks is established through comparison with the 3D HMQC-NOESY spectrum.

Using the pulse sequence outlined in Section 4.1.4.4, a 3D HMQC-NOESY-TOCSY spectrum was obtained from ^{13}C-methylated model mixture **II**. As with the 3D HMQC-NOESY spectrum, the 3D HMQC-NOESY-TOCSY spectrum has intense auto correlation peaks of methoxy groups (data not shown). 2D F_1F_3 and F_2F_3 projections of the aromatic region of the spectra are used for initial analysis (Figure 6.18).

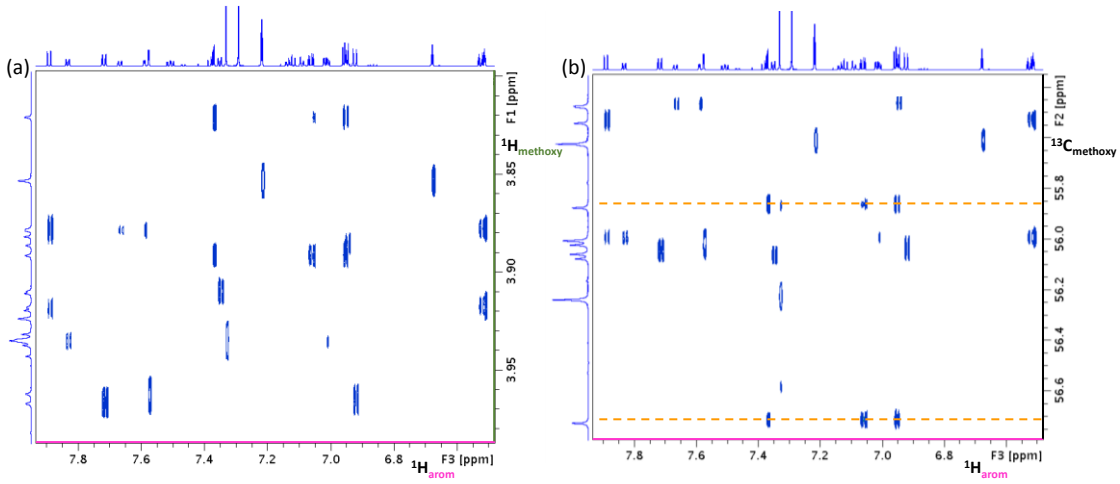


Figure 6.18 Aromatic region of the 2D (a) F_1F_3 and (b) F_2F_3 projections of the 3D HMQC-NOESY-TOCSY of ^{13}C-methylated model mixture **II** acquired using the pulse sequence shown in Figure 4.16. The dotted lines indicate the methoxy carbon chemical shifts of molecule **6**.

A comparison with the corresponding region of the 3D HMQC-NOESY spectra in (Figure 6.16) shows that the TOCSY spectrum contains more cross-peaks. These additional TOCSY cross-peaks are extremely valuable as they provide proton chemical shifts of protons remote from the methoxy groups. As with the 3D HMQC-NOESY, 2D F_1F_3 planes were extracted at specific methoxy carbon frequencies shown in Figure 6.18b. The two planes extracted through the methoxy carbons of molecule **6** are shown in Figure 6.19, with the NOESY cross-peaks circled.

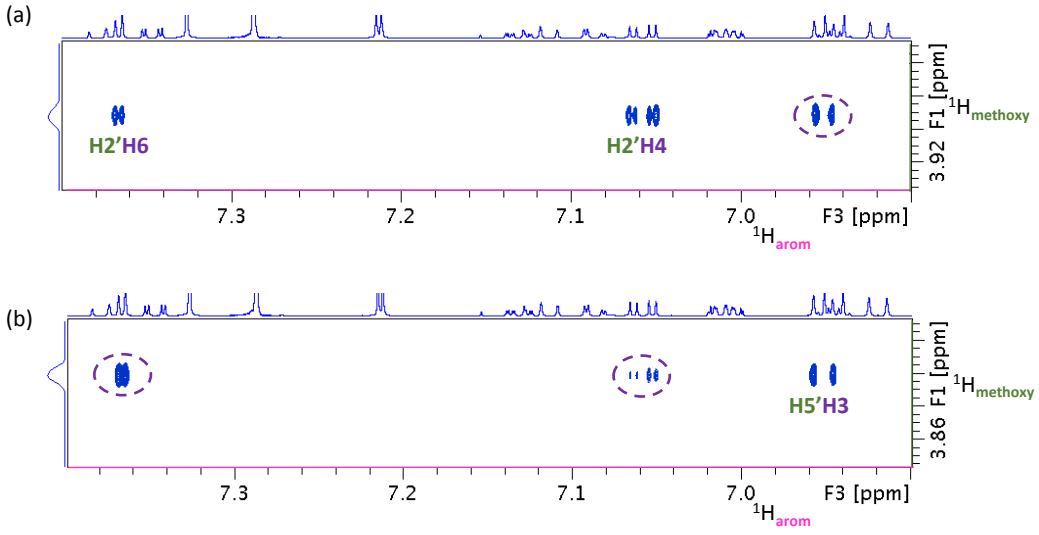


Figure 6.19 2D F_1F_3 planes of the 3D HMQC-NOESY-TOCSY extracted at methoxy carbons of molecule **6** at (a) 56.71 (C2') and (b) 55.84 (C5') ppm in F_2. The NOESY cross-peaks are circled.

A 70 ms TOCSY mixing time was used to obtain this spectrum, hence the magnetisation was transferred via three- and four-bond proton-proton couplings. It can be seen that the same set of cross-peaks was obtained in both spectra indicating that indeed the identified protons are part of the same molecule. The protons at 6.95, 7.06 and 7.36 ppm can therefore be assigned to positions H3, H4 and H6 respectively (Figure 6.19). The 3D HMQC-NOESY-TOCSY is a very powerful addition to the arsenal of experiments for the assignment of resonances from ^{13}C-methylated compounds.

6.3. Relative merits of the ^{13}C-filtered NMR experiments

The combined information obtained from the experiments presented in the previous section allowed full characterisation of the nine compounds of ^{13}C-methylated model mixture **II**. The ^{13}C and ^{1}H chemical shifts and coupling constants obtained from this analysis are listed in Appendix B (see external web link). Table 6.1 summarises the results for compound **6** only.

Table 6.1 ^{1}H and ^{13}C chemical shifts and *J* couplings obtained from the individual NMR experiments for molecule **6**.

Experiment	3D HMQC-HMBC	3D HMQC-NOESY	3D HMQC-NOESY-TOCSY	3D HcCH$_3$ 3D hCCH$_3$	4D HCCH$_3$	3D INEPT-INADEQUATE-HSQC
Position	Chemical shift **^{1}H**/^{13}C (ppm)/(*J* coupling Hz)					
1	-	-	-	-	-	120.24
1'	166.52	-	-	-	-	166.64
1''	**3.92**	**-**	**-**	**-**	**-**	**3.92**
	52.22	-	-	-	-	52.21
2	153.43	-	-	-	-	153.47
2'	**3.90**	**3.89**	**3.89**	**3.89**	**3.89**	**3.89**
	56.70	56.71	56.71	56.71	56.69	56.67
3	**-**	**6.95**	**6.95**	**6.94**	**6.94**	**6.95**
	-	(9.1)	(8.1)	-	-	(8.9)
	-	-	-	113.67	113.65	113.68
4	-	**7.06**	**7.06**	**7.06**	**7.05**	**7.05**
	-	(9.1,3.3)	(9.7, 3.2)	-	-	(9.0)
	-	-	-	119.67	119.64	119.67
5	152.88	-	-	-	-	152.89
5'	**3.82**	**3.82**	**3.82**	**3.82**	**3.82**	**3.82**
	55.85	55.86	55.86	55.87	55.86	55.81
6	**-**	**7.37**	**7.37**	**7.37**	**7.35**	**7.37**
	-	(3.1)	(3.4)	-	-	-
	-	-	-	115.88	115.85	115.87

The results summarised in Table 6.1 show that all experiments provide the chemical shifts of methoxy groups. Only the 3D INEPT-INADEQUATE-HSQC and 3D HMQC-HMBC provide information about the ester groups and chemical shifts of quaternary carbons to which the methoxy groups are linked. The $HCCH_3$ type experiments provide complementary information regarding the aromatic CH pairs ortho to the methoxy groups. Experiments containing a NOESY and TOCSY step yield chemical shifts of ortho and meta (in principle also para) protons, respectively.

The 3D INEPT-INADEQUATE-HSQC could be regarded as the only experiment required for the analysis of this mixture, as it identifies all accessible carbon chemical shifts. However as can be seen in the example of molecule **6**, for unknown molecules the 3D HMQC-NOESY and 3D HMQC-NOESY-TOCSY are required to piece the structural fragments together. Furthermore, in the case of more complex mixtures, the 3D/4D $HCCH_3$ and 3D HMQC-HMBC experiments may become useful to separate out fragments overlapped in up to three dimensions, especially in the analysis of HS samples.

The issue not addressed so far is the relative sensitivity of the proposed experiments. One of the reasons for acquiring all experiments on ^{13}C-methylated mixture **II** was to compare their performance and S/N they can provide. The results of this investigation are presented next.

6.4. Relative sensitivity of the ^{13}C-filtered NMR experiments

The S/N ratio achievable in the individual experiments were compared by acquiring 3D/4D NMR spectra using similar experimental conditions, including comparable digital resolution in the indirectly detected dimensions. In order to achieve this and to accommodate different spectral widths required by individual experiments, different overall acquisitions times had to be used. These differences were eliminated by scaling the measured S/N ratios and producing values reflecting the S/N achievable in a 14 hour long 3D INEPT-INADEQUATE-HSQC experiment. As all experiments have in common the ^{1}H, ^{13}C correlations of the methoxy groups, 1D traces were extracted from the HSQC/HMQC planes at individual methoxy carbon chemical shifts. The S/N of methoxy protons was measured, except for the 3D HMQC-NOESY and 3D HMQC-NOESY-TOCSY, where the S/N was determined for aromatic protons. For the TOCSY experiment, the NOESY peaks were not included. For the 3D INEPT-INADEQUATE-HSQC spectra the S/N ratio was determined in both aromatic and methoxy regions. The S/N of the 4D $HCCH_3$ spectrum was

obtained by extracting 3D cubes at individual methoxy ^{13}C chemical shifts, followed by the extraction of 1D slices from the 2D HSQC planes, as done for the 3D experiments. The obtained data are presented for clarity in two graphs (Figure 6.20 and 6.21) in the form of bar charts, where the count represents the number of cross-peaks with a particular S/N ratio within the 0-3000 range, increasing in steps of 300.

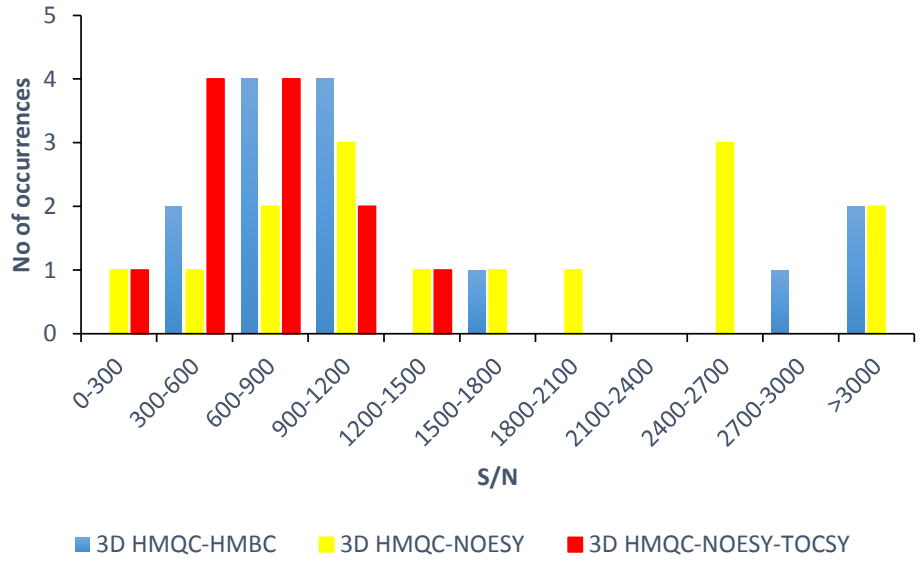


Figure 6.20 Comparison of the S/N measured in the 3D HMQC-HMBC, 3D HMQC-NOESY and 3D HMQC-NOESY-TOCSY spectra. See text for details.

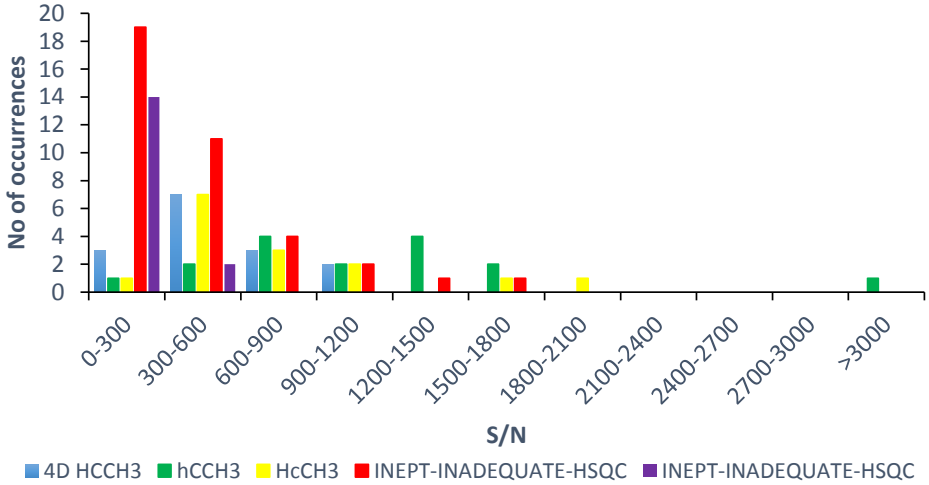


Figure 6.21 Comparison of the S/N measured in the 4D $HCCH_3$, 3D $hCCH_3$, 3D $HcCH_3$ and 3D INEPT-INADEQUATE-HSQC experiments. For the INEPT-INADEQUATE-HSQC the S/N was measured for methoxy signals (red) and aromatic signals (purple). See text for details.

The sensitivity of the 3D HMQC-HMBC experiment (Figure 6.20) is similar to that observed in the 3D $HCCH_3$ spectra. The analysis of the 3D HMQC-HMBC included correlations of the carboxyl carbons, which increased the number of reported observations. Figure 6.20 also contains the analysis of the 3D HMQC-NOESY and the 3D HMQC-NOESY-TOCSY cross-peaks. Here the achieved S/N ratios are comparable to those observed in the 3D HMQC-HMBC spectra. One has to bear in mind that in the 3D HMQC-HMBC the magnetisation passes through natural abundance carbons and the S/N is measured on a singlet containing three protons. The results for the NOESY and TOCSY spectra on the other hand report on the S/N ratio of aromatic protons, usually split by proton-proton couplings. Qualitatively, the relative sensitivity of the NOESY (TOCSY) experiments normalised to three protons of a singlet requires a scale up factor of 3-6, which is in line with the excepted NOE efficiency of a few percent, the limiting factor in the sensitivity of these experiments.

The sensitivity of the NOESY experiment is only slightly better than that of the TOCSY experiment because the 70 ms DIPSI transfer was very efficient in transferring the magnetisation to *J*-coupled spins, where the S/N was measured.

Figure 6.21 summarises the S/N ratios achieved in the 3D/4D $HCCH_3$ family of experiments and the 3D INEPT-INADEQUATE-HSQC. The sensitivity of the two 3D $HCCH_3$ experiments is comparable as is that of the 4D $HCCH_3$ and 3D INEPT-INADEQUATE-HSQC. A high number of analysed cross-peaks in the latter spectrum reflects the fact that both protonated and quaternary ipso carbons were inspected. The decreased S/N of the 4D spectrum relative to the 3D $HCCH_3$ experiments is caused by the increased dimensionality of this experiment. Losses due to DQ filtration in the 3D INEPT-INADEQUATE-HSQC experiment were the main cause of the lower S/N obtained for this experiment.

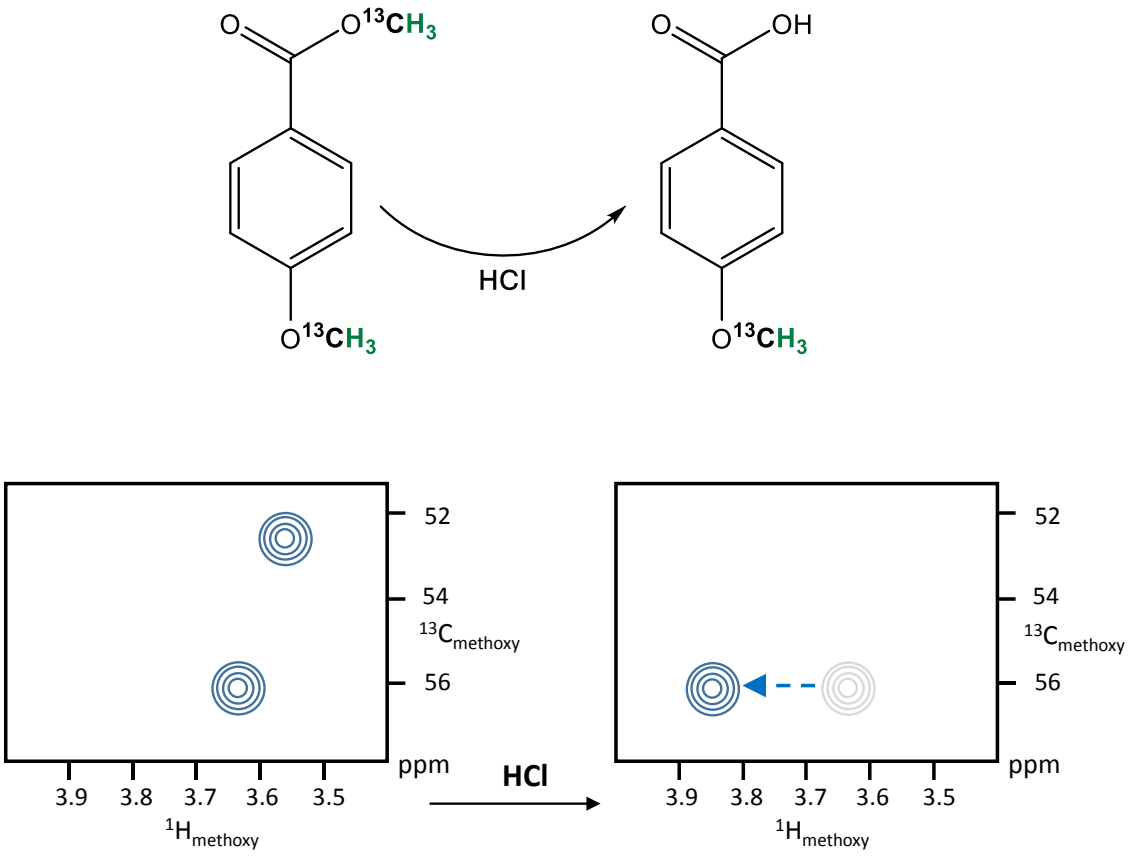
In summary, the S/N analysis presented here suggests that the sensitivity of all proposed experiments is roughly comparable. The only exception is the aromatic protons of the 3D INEPT-INADEQUATE-HSQC experiment, which are the limiting step in obtaining all accessible chemical shifts of nuclei in the immediate vicinity of the methoxy groups. In an ideal situation where a sufficient S/N is obtained for these protons, no other experiment is likely needed to characterise these nuclei. The chemical shifts of more remote protons, which are very useful pieces of information, are accessible via the 3D HMQC-NOESY-TOCSY experiment. Therefore in an ideal case these two experiments should be sufficient to obtain a complete set of correlated

chemical shifts. If however the ortho protons are not obtainable from the INEPT-INADEQUATE-HSQC experiment, the $HCCH_3$ experiment(s) need to be acquired. Similarly, if the ipso carbons do not appear in the 3D INEPT-INADEQUATE-HSQC, the more sensitive 3D HMQC-HMBC can be employed to provide this data. However as the $HCCH_3$ type experiments use the SQ rather than the DQ frequency axis the individual signal separation may be the argument for acquiring these types of spectra.

6.5. Chapter conclusions

A series of ^{13}C-filtered NMR experiments were successfully applied to ^{13}C-methylated model mixture **II**. The combination of the experiments allowed the full characterisation of the mixture. The only issue arising was that of the ester group correlations. Weak correlations via $^4J_{CC}$ located the positions of these groups, however their weak strength indicates that these are unlikely to be obtained for the HS samples. Analysis of the experiments shows that there is sufficient S/N in the 3D INEPT-INADEQUATE-HSQC and 3D HMQC-NOESY-TOCSY spectra. These experiments should therefore be adequate for complex mixtures as they provide the maximum number of correlations.

Chapter 7. Results and Discussion. Application of ^{13}C-filtered NMR experiments to hydrolysed model mixture II



Methyl iodide methylated both the carboxylic and hydroxyl groups in ^{13}C-methylated model mixture **II** to give methyl esters and methoxy ethers, respectively. However, the esters can easily be hydrolysed to yield carboxylic acids. This will simplify the spectra, but will also induce chemical shift changes to the parent molecules. It was therefore decided to hydrolyse part of the sample (as detailed in Section 2.10) and reassign the resonances by acquiring a limited set of experiments. This exercise was conducted to see if the combined analysis of the fully and partially methylated samples provides additional information. The results of this exercise are presented here.

7.1. Initial NMR characterisation of hydrolysed model mixture II

To verify that the ester hydrolysis was successful the ^{13}C spectra before and after the hydrolysis procedure were acquired (Figure 7.1).

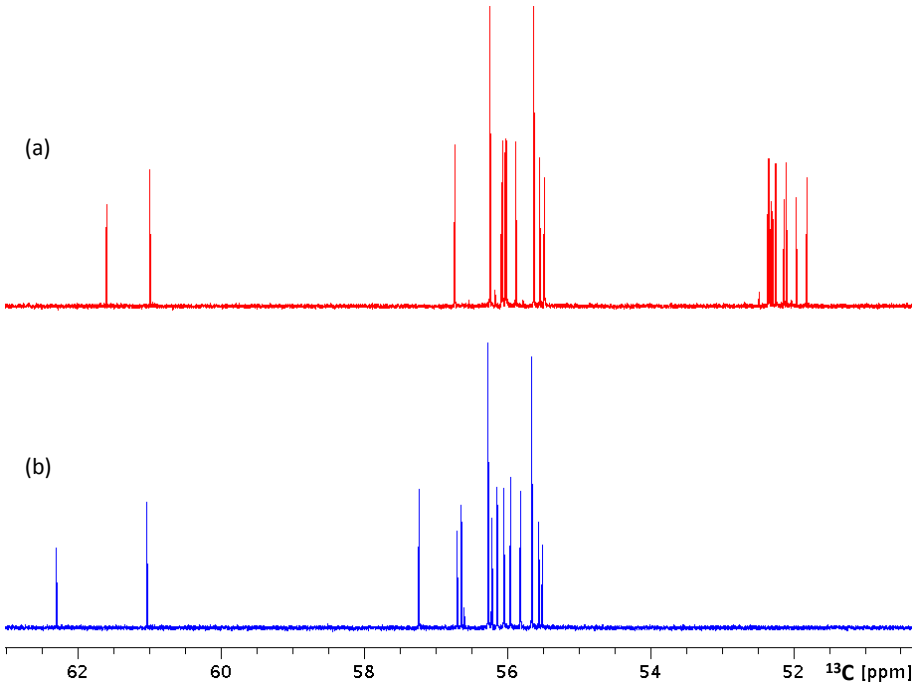


Figure 7.1 ^{13}C spectra of ^{13}C-methylated model mixture **II** (a) before and (b) after hydrolysis of ester groups.

Inspection of the methoxy region of the ^{13}C spectra in Figure 7.1 shows the disappearance of the ester methoxy groups. There are also differences in the ^{13}C chemical shifts of methoxy carbons. The change from ester to acid therefore has an effect on the electron density not only of the

aromatic carbons but also that of the methoxy carbons. These changes are clearly visible in an overlay of the 2D ^{1}H, ^{13}C HSQC spectra obtained before and after hydrolysis (Figure 7.2).

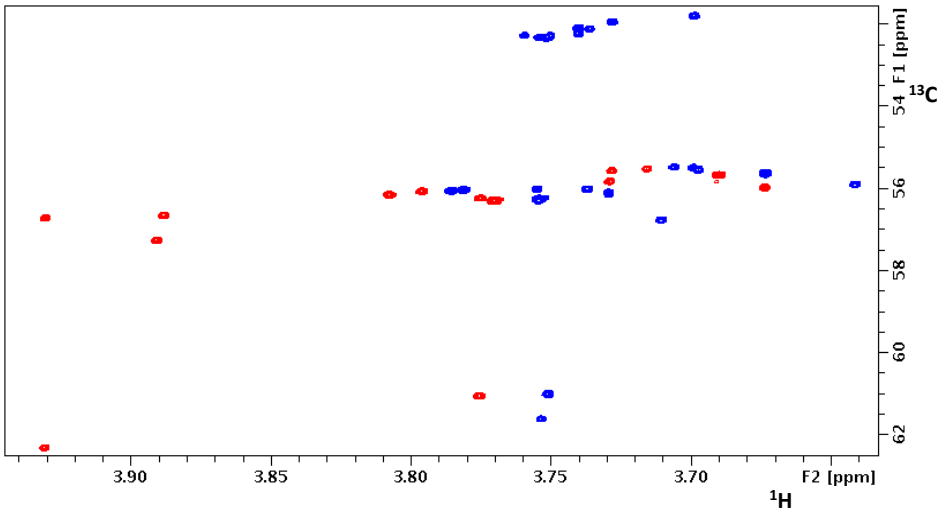


Figure 7.2 Methoxy ether region of 2D ^{1}H, ^{13}C HSQC showing the methoxy cross-peaks of ^{13}C-methylated model mixture **II** before (blue) and after (red) hydrolysis.

The spectrum shows that the ^{1}H resonances in addition to the ^{13}C resonances have moved. These chemical shift changes have occurred despite the fact the chemical modification has occurred 5 or more bonds away from any methoxy group. It was hoped that changes would also occur to the chemical shifts of the aromatic nuclei, which could be interpreted to obtain structural information about the methylated molecules. With this in mind, two ^{13}C-filtered NMR experiments were applied to the hydrolysed model mixture **II,** as detailed in the next section.

7.2. Application of ^{13}C-filtered NMR experiments to the hydrolysed methylated model mixture II

In order to determine the chemical shift changes of aromatic protons and carbons, which occur upon hydrolysis, two of the ^{13}C-filtered experiments were acquired; 3D INEPT-INADEQUATE-HSQC and the 3D HMQC-NOESY-TOCSY. From the resonance assignment of fully ^{13}C-methylated model mixture **II**, presented in Chapter 6, these two experiments were sufficient to fully characterise the molecules in this mixture. The chemical shifts obtained for the hydrolysed molecule **6** as well as the chemical shift differences (Δ) between hydrolysed (δ_{acid}) and fully methylated (δ_{ester}) molecule **6** are summarised in Table 7.1. The obtained differences varied between -2.3 to +2.6 ppm.

Table 7.1 Chemical shifts obtained from the 3D HMQC-NOESY-TOCSY and the 3D INEPT-INADEQUATE-HSQC spectra of hydrolysed methylated molecule **6**. The differences in chemical shift between the benzoate and benzoic acid versions of molecule **6** are also given.

Experiment/ Molecular position	3D HMQC-NOESY-TOCSY δ_{Acid} (ppm)	Chemical shift $\Delta/(\delta_{Acid}-\delta_{Ester})$ (ppm)	3D INEPT-INADEQUATE-HSQC δ_{acid} (ppm)	Chemical shift $\Delta/(\delta_{Acid}-\delta_{Ester})$ (ppm)
1	-	-	117.97	-2.27
2	-	-	152.35	-1.12
2'	4.07	+0.18	4.07	+0.18
	57.22	+0.51	57.27	+0.54
3	7.03	+0.08	7.04	+0.08
	-	-	113.26	-0.42
4	7.16	+0.10	7.04	-0.01
	-	-	122.28	+2.61
5	-	-	154.44	+1.55
5'	3.85	+0.03	3.85	+0.03
	55.96	+0.1	55.98	+0.17
6	7.72	+0.35	7.72	+0.35
	-	-	116.23	+0.36

These chemical shifts changes reflect the change in electron withdrawing capacity between an ester and an acid. To quantify the effects of de-esterification, all molecules were analysed using the 3D INEPT-INADEQUATE-HSQC to map the effect on carbon (Table 7.2) and both the INEPT-INADEQUATE-HSQC and HMQC-NOESY-TOCSY on proton chemical shifts (Table 7.3). As a point of reference, the "molecule 0" entry in both tables represents the chemical shift changes between benzoic acid and methyl benzoate.[303]

Table 7.2 ^{13}C chemical shift changes (δ_{acid}-δ_{ester}) determined from the 3D INEPT-INADEQUATE-HSQC spectra of the fully and partially ^{13}C-methylated model mixture **II**. Carbon positions are labelled relative to the carboxylic acid group.

	^{13}C CHEMICAL SHIFT CHANGES (ppm)					
Molecule	IPSO	ORTHO	ORTHO	META	META	PARA
"0"	-0.79	0.52	0.52	0.12	0.12	0.93
1	-2.33	-1.01	-	-0.3	-	1.72
2	-	0.56	-	0.19	0.1	0.99
3	-3.89	-0.85	-	-1.44	-	1.76
4	-1.79	-1.75	-	2.04	0.33	0.96
5	-1.01	0.59	0.59	0.12	0.12	0.86
6	-2.35	-1.12	0.35	-0.42	1.54	1.6
7	-1.1	0.39	1.16	0.1	0.19	0.8
8	-1.02	0.69	0.69	0.15	0.15	0.86
9	-	0.83	0.83	0.18	0.18	0.73

- Positions not accessible by the 3D INEPT-INADEQUATE-HSQC

Changes in ^{13}C chemical shifts after de-esterification ranged from -3.8 to +2.04 ppm. As expected the largest changes were observed on the ipso carbon. Four compounds stood out from the rest in terms of the magnitude and location of the changes; these were compounds **1**, **3**, **4** and **6**. While all other compounds showed small differences (~ 1 ppm) that were very similar to those observed for molecule "0", the four compounds showed large differences in all positions, ipso to para. By in large, this was the order of the observed changes but there were a few exceptions, e.g. the meta position of compound **4**. These four significant movers had the methoxy groups in ortho position, which is clearly the reason for the difference observed between the two classes of compounds.

Table 7.3 ^{1}H chemical shift changes (δ_{acid}-δ_{ester}) determined from the 3D INEPT-INADEQUATE-HSQC and 3D HMQC-NOESY-TOCSY spectra of the fully and partially ^{13}C-methylated samples. The proton positions are labelled relative to the carboxylic acid group

	^{1}H CHEMICAL SHIFT CHANGES (ppm)*				
Molecule	ORTHO	ORTHO	META	META	PARA
"0"	0.14	0.14	0.07	0.07	0.06
1	-	-	0.09	-	-
	0.4	-	0.17		
2	0.06	-	-	-	0.05
	0.06	0.08	0.04	-	0.05
3	-	-	-	-	0.1
	0.41	-	-		
4	-	-	0.11	0.15	-
	0.27	-	0.13	0.17	-
5	0.06	0.06	-	-	0.05
	0.07	0.07	-	-	0.05
6	0.35	-	0.09	-	-
	0.35	-	0.08	-	0.1
7	0.05	-	0.04	-	-
	0.05	0.09	0.03	-	-
8	0.07	0.07	-	-	-
	0.07	0.07	-	-	-
9	-	-	0.04	0.04	-
	0.07	0.07	0.03	0.03	-

*Red denotes chemical shift changes from 3D HMQC-NOESY-TOCSY; - denotes positions not accessible by these experiments.

The biggest changes in ^{1}H chemical shifts (+ 0.27 to + 0.41 ppm) occur at the ortho position for molecules **1**, **3**, **4** and **6**. These are the only molecules which have a methoxy group next to the carboxylic group and thus have only one ortho proton. The chemical shift of the ortho protons in the other molecules generally increased by up to 0.07 ppm. The changes in the meta and para

positions are small and in line with those for molecule "0". Overall, the ^{1}H chemical shift changes between the fully and partially methylated compounds are not very significant.

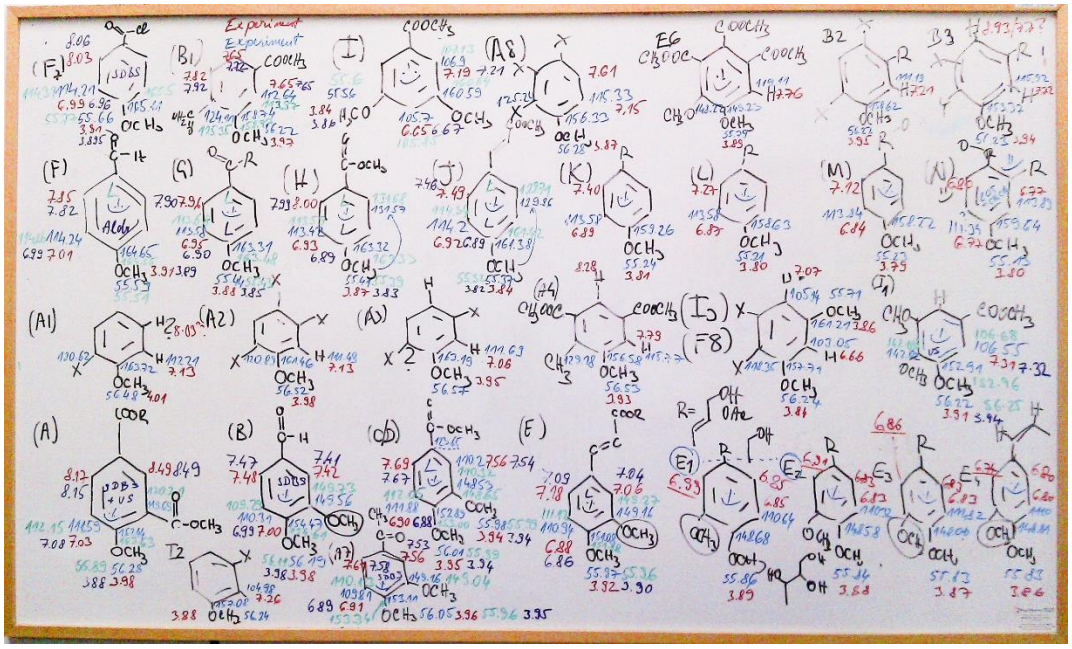
7.3. Chapter conclusions

Hydrolysis was conducted on the ^{13}C-methylated model mixture **II**. Initial characterisation using ^{13}C 1D NMR spectra and ^{1}H, ^{13}C HSQC showed that the hydrolysis procedure was successful and led to removal of the ester groups. Two ^{13}C-filtered NMR spectra, 3D INEPT-INADEQUATE-HSQC and 3D HMQC-NOESY-TOCSY, were acquired on the sample and the observed chemical shift differences were quantified with reference to the fully ^{13}C-methylated sample.

Based on this analysis, it was judged that the changes in chemical shift after the hydrolysis would be difficult to interpret in structurally heterogeneous complex mixtures containing unknown compounds.

The information provided by the hydrolysis of the sample would likely not add much to what was already established. Hydrolysis was therefore not performed on the ^{13}C-methylated RMFA sample.

Chapter 8. Results and Discussion. Molecular-scale characterisation of RMFA by ^{13}C-filtered NMR spectroscopy



The office white board came in useful when analysing the nD NMR spectra.

As detailed in Chapter 3, the HS sample chosen for further NMR analysis is RMFA. This chapter provides a detailed characterisation of the methoxy groups contained in ^{13}C-methylated RMFA. This is followed by the results of the ^{13}C-filtered NMR experiments that were previously tested on model mixture **II** (Section 6.2). This includes analysis of 3D INEPT-INADEQUATE-HSQC, 4D HCCH$_3$, 3D HcCH$_3$, 3D hCCH$_3$, 3D HMQC-HMBC, 3D HMQC-NOESY and 3D HMQC-NOESY-TOCSY spectra. The experimental parameters for each experiment are listed in Section 2.17. The procedure to obtain molecular structures from spectra of an unknown mixture is presented. The molecules are classified in terms of compound type and size. The chapter concludes with a comparison of the identified molecules with literature data and plant metabolites.

8.1. Characterisation of ^{13}C-methylated RMFA sample by 2D NMR Spectroscopy

The 1D ^{1}H and ^{13}C spectra shown in Figure 3.10 and 3.11 of ^{13}C-methylated RMFA showed significant overlap, therefore to obtain clarification of the methoxy group types present in the sample a 2D ^{1}H, ^{13}C HSQC spectrum was acquired (Figure 8.1).

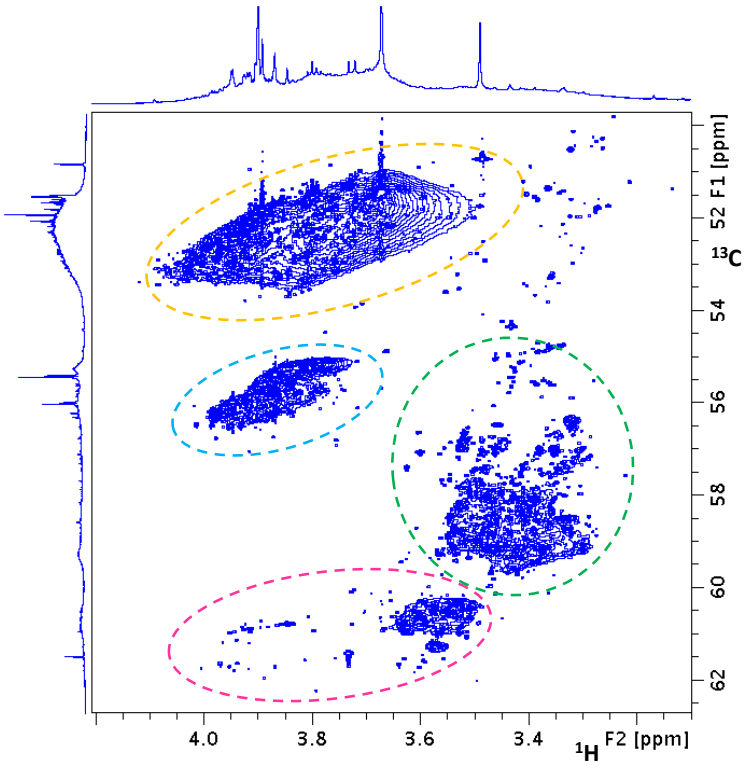


Figure 8.1 2D ^{1}H, ^{13}C HSQC spectrum of ^{13}C-methylated RMFA. The dashed regions indicate distinct methoxy regions present in the sample as aliphatic or aromatic ester (----), aromatic (----), carbohydrate (----) and sandwiched aliphatic or aromatic (----) methoxy groups.

The 2D ^{1}H, ^{13}C HSQC spectrum allows the distinction between aliphatic and aromatic methoxy groups, which have similar methoxy proton chemical shifts but differ in methoxy carbon chemical shift, as highlighted in Figure 8.1. Subsequent analysis of this sample focuses on the aromatic methoxy ether region investigating the surrounding environments of methylated hydroxyl groups in phenolic molecules.

In order to further characterise the compounds present in the ^{13}C-methylated RMFA sample the aromatic and aldehydic region of a ^{1}H, ^{13}C HSQC was also acquired (Figure 8.2).

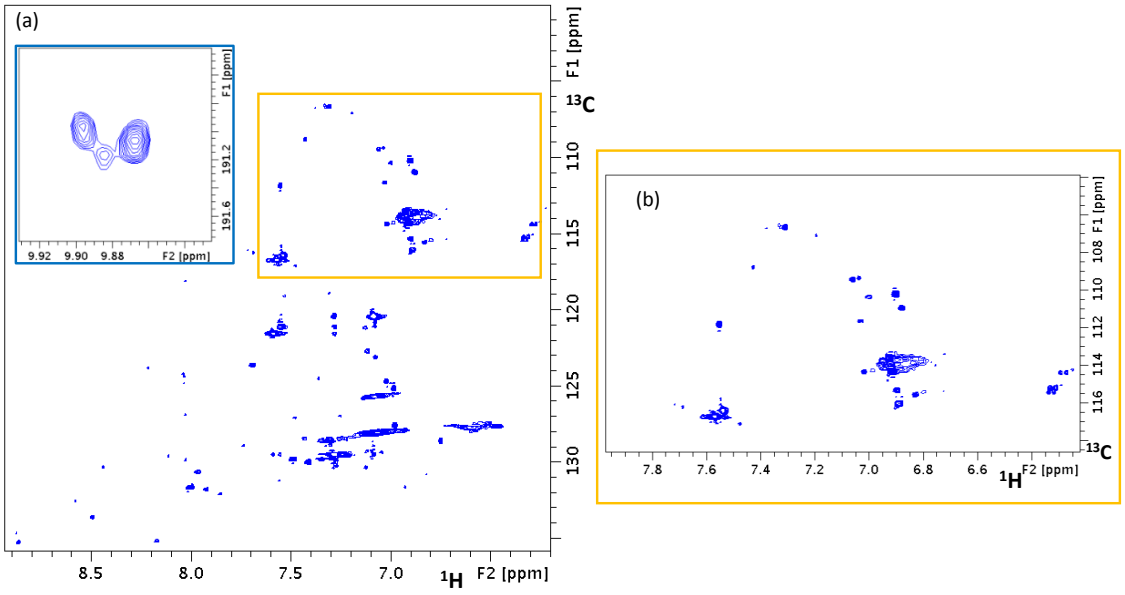


Figure 8.2 (a) Aromatic region of the 2D ^{1}H, ^{13}C HSQC spectrum of ^{13}C-methylated RMFA with the (b) expansion of the region indicated by the orange box in (a). This region contains cross-peaks belonging to nuclei ortho to the methoxy groups. The aldehydic region, shown as an inset (blue box) in (a), contains two intense cross-peaks and one weak cross-peak.

This spectrum shows numerous aromatic proton-carbon correlations. Resonances above 120 ppm represent CH pairs that are not ortho, or para, to electron donating groups. The CH pairs with ^{1}H chemical shifts above 7.3 ppm are potentially next to carbonyl containing groups, which have a strong deshielding effect on ortho protons. Figure 8.2b indicates that there are several aromatic CH pairs ortho and para to methoxy groups. Outwith the aromatic and aliphatic regions of the ^{1}H, ^{13}C HSQC three peaks were identified in the aldehyde region (See inset in Figure 8.2a).

8.2. Procedure for the analysis of the ^{13}C-filtered NMR spectra

In order to build molecular fragments from ^{13}C-methylated mixtures using the ^{13}C-filtered NMR methodology, the 2D ^{1}H, ^{13}C HSQC spectrum was used as a starting point for inspection of all 3D and 4D NMR spectra. If the molecule does not show a methoxy cross-peak in the HSQC spectrum, it will not show any resonances in the ^{13}C-filtered experiments. Thus building molecular fragments starts by identifying a particular methoxy cross-peak and using its ^{13}C chemical shifts to extract 2D planes or 3D cuboids from individual 3D/4D spectra. These are then inspected to obtain chemical shifts of aromatic nuclei. A particular order in which the spectra are inspected was developed that proved most efficient. The procedure to extract chemical shifts (or if possible coupling constants) from each ^{13}C-filtered experiments is summarised as follows:

1. Identify one methoxy group and record its ^{1}H and ^{13}C chemical shifts from the 2D ^{1}H, ^{13}C HSQC spectrum.
2. Identify the quaternary ipso carbon in the 3D HMBC-HMQC spectrum and record its ^{13}C chemical shift.
3. Inspect the 3D HcCH$_3$ and 3D hCCH$_3$ spectra to identify the chemical shifts of aromatic ^{1}H and ^{13}C nuclei next to the methoxy groups. If there are no peaks in either of the spectra, the methoxy group is either surrounded by two quaternary carbons or is attached to non-aromatic/olefinic moiety. If the 2D planes extracted at the same ^{1}H/^{13}C methoxy chemical shift contain more than one cross-peak, then the 4D HCCH$_3$ experiment can be used to remove this ambiguity.
4. Use the 3D INEPT-INADEQUATE-HSQC spectrum to verify the assignment obtained in steps 2 and 3 and inspect for the presence of quaternary carbons, additional to the ipso carbon chemical shifts. If the molecule contains aromatic protons, inspect the aromatic proton vs DQ plane to extract the J_{HH} couplings if present.
5. Use the 3D HMQC-NOESY spectrum to confirm the assignment of the ortho protons. By examining their J_{HH} couplings suggest the substitution pattern of the aromatic ring.
6. Finally, using the 3D HMQC-NOESY-TOCSY spectrum record the chemical shifts and coupling constants of any additional protons of the aromatic ring that can be reached using TOCSY transfer from the protons ortho to the methoxy groups.

Using the above procedure, a number of fragments of aromatic molecules were identified in the ^{13}C-methylated RMFA sample starting from the methoxy cross-peaks labelled in the 2D ^{1}H, ^{13}C

HSQC spectrum (Figure 8.3). The numbers displayed on Figure 8.3 will be used throughout this chapter as identifiers for the proposed molecules in Section 8.3.

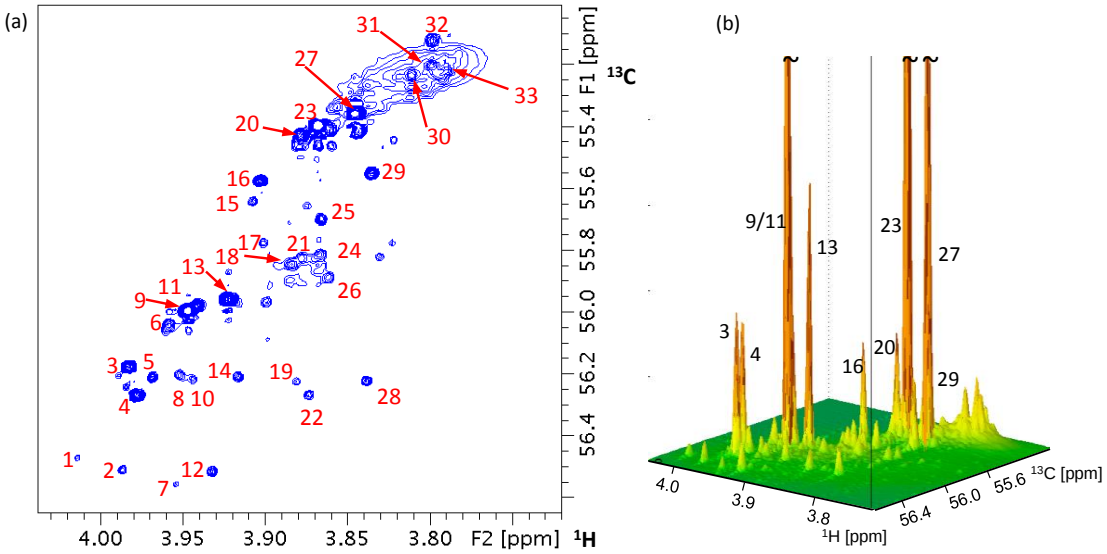


Figure 8.3 (a) Aromatic methoxy region of the 2D ^{1}H, ^{13}C HSQC spectrum of ^{13}C-methylated RMFA. Each cross-peak above a threshold is labelled. The numbers are used throughout the analysis as compound identifiers. (b) Intensity stack plot of the same region with peaks of the major compounds sitting on top of a 'carpet' of minor compounds.

8.3. From correlated chemical shifts to molecular structures

Once a set of chemical shifts associated with each investigated methoxy group was obtained, initial molecular fragments were drawn. The chemical shifts were analysed for the compatibility with different functional groups. In the first instance, it was assumed that the molecules were substituted benzene molecules. If a match was not found, other classes of aromatic molecules were considered such as flavonoids which containing multiple aromatic/aliphatic rings. Initially, the chemical shifts for each fragment obtained were compared to those of methoxy benzene. The differences observed were then compared with the substituent effects tabulated for mono-substituted benzene rings (Table 8.1). When multiple substituents are present the chemical shift of the aromatic ring nuclei can be calculated using the following equations:

$$\delta_{13_C} = 128 + \sum \delta_i \qquad [14]$$

$$\delta_{1_H} = 7.28 + \sum \delta_i \qquad [15]$$

Where δ_i are the chemical shift effects tabulated in Table 8.1.

Table 8.1 ^{1}H and ^{13}C chemical shift substituent effects calculated using mono-substituted benzene derivatives[304]

Substituent	Chemical shift effect (δ_i) ($^{13}C/^{1}H$)			
	Ipso	Ortho	Meta	Para
CH_3	9.3	0.8 -0.2	0 -0.12	-2.9 -0.22
CH_3CH_2	15.6	-0.4 -0.14	0 -0.06	-2.6 -0.17
$(CH_3)_2CH$	20.6	-2.5 -0.13	0.1 -0.08	-2.4 -0.18
$(CH_3)_3C$	22.4	-3.1 0.02	-0.1 -0.08	-2.9 -0.21
CF_3	-9.0	-2.2	0.3	3.2
C_6H_5	13	-1 0.37	0.4 0.20	-1 0.10
CH_2=CH	9.5	-2 0.06	0.2 -0.03	-0.5 -0.10
CH≡C	-6.1	3.8 0.15	0.4 -0.02	-0.2 -0.01
CH_2OH[(a)]	12	-1 0.07	0 0.07	-1 0.07
COOH	2.1	1.5 0.85	0 0.18	5.1 0.27
$COOCH_3$	2.1	1.1 0.71	0.1 0.11	4.5 0.21
COCl	5	3 0.84	1 0.22	7 0.36
CHO	8.6	1.3 0.56	0.6 0.22	5.5 0.29
$COCH_3$	9.1	0.1 0.62	0 0.14	4.2 0.21
$COCF_3$	-5.6	1.8 -	0.7 -	6.7 -
COC_6H_5	9.4	1.7 0.47	-0.2 0.13	3.6 0.22
CN	-15.4	3.6 0.36	0.6 0.18	3.9 0.28
OH	26.9	-12.7 -0.56	1.4 -0.12	-7.3 -0.45
OCH_3	31.4	-14.4 -0.48	1.0 -0.09	-7.7 -0.44
$OCOCH_3$	23	-6 -0.25	1 -0.03	-2 -0.13
OC_6H_5	29	-9 -0.29	2 -0.05	-5 -0.23
NH_2	18.0	-13.3 -0.75	0.9 -0.25	-9.8 -0.65
$N(CH_3)_2$	23	-16 -0.66	1 -0.18	-12 -0.67
$N(C_6H_5)_2$	19	-4 -	1 -	-6 -

$NHCOCH_3$	11	-10	0	-6
		0.12	-0.07	-0.28
NO_2	20	-4.8	0.9	5.8
		0.95	0.26	0.38
NCO	5.7	-3.6	1.2	-2.8
		-	-	-
F	34.8	-12.9	1.4	-4.5
		-0.26	0	-0.2
Cl	6.2	0.4	1.3	-1.9
		0.03	-0.02	-0.09
Br	-5.5	3.4	1.7	-1.6
		0.18	-0.08	-0.04
I	-32	10	3	1
		0.39	-0.21	0
CH_3OCH_2[(a)]	9.8	-0.1	-0.3	-0.9
		-	-	-
$CH_3CH_2OCH_2$	10.6	-0.3	0.4	-0.4
		-	-	-
$CONH_2$	5.9	-0.9	-0.1	2.9
		0.61	0.10	0.17
$CONHCH_3$	6.6	-1.1	0.4	3.2
		0.51	0.10	0.17
$CON(CH_3)_2$	8.4	-1	0.3	1.5
		-	-	-
CH_3CH_2COO	2.6	1.5	0.3	4.8
CH_3COOCH_2	6.3	1.3	0.4	-1
CH_3OCOCH_2	6.1	1.29	0.6	-0.9
$CH_3CO(CH_2)_2$	13	0.5	0.3	-1.9
$(CH_3)_2CHCO$	8.4	0.6	0.3	4.8
		0.67	0.39-0.03	0.39-0.03
CH_3CH_2CO	9.0	0.6	0.01	4.8
		-	-	-
$CH_3CH_2CH_2CO$	9.2	0.6	0.03	4.8
		0.67	0.27	0.14
$CH_3OPhHC=CHCO$	10.6	0.4	0.6	2.2
		0.73	0.22	0.29
$CH_3OPhCOCH=CH$	7.2	0.3	0.9	2.3
		0.35	0.13	0.11
COOPh	1.7	2.1	0.5	5.5
		0.90	0.17	0.27
OCOPh[(b)]	1.8	1	0.3	5.3
		-0.09	0.09	-0.08
PH:flavonoid[(b)]	10.9	-1.8	0.9	0.7
		-0.23	0.15	0.11
SH	2.8	1	1.3	-2.5
		-0.08	-0.16	-0.22

(a) ref[305] (b) ref[303]

It is well known that the substituent effects, listed in Table 8.1, are only accurate for mono-substituted benzenes. However these effects can be used in a qualitative manner, as the trends they reflect are valid, e.g. a methoxy group will have a high negative ortho and para effect. To provide some error boundaries, Figure 8.4 shows the chemical shifts obtained for di-substituted methoxy benzenes using the Spectral DataBaSe (SDBS)[303] database compared to those calculated based on placing an additional methoxy group to a mono-substituted methoxy benzene at different positions.

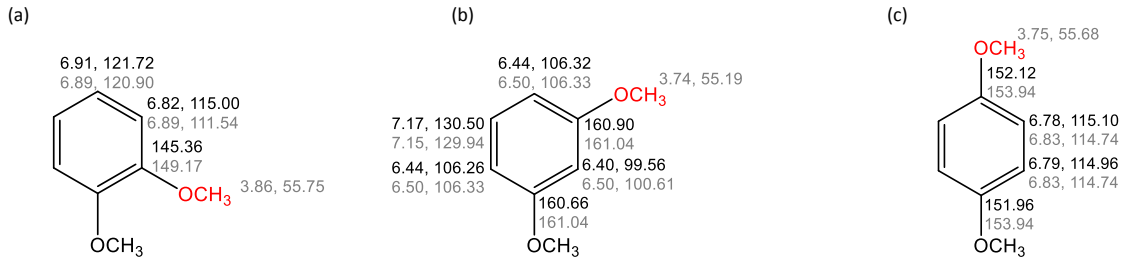


Figure 8.4 Comparison of the experimental (black[303]) and calculated (grey) chemical shifts for di-substituted methoxy benzene. The calculated chemical shifts were obtained by placing a second methoxy group (red) in the (a) ortho (b) meta or (c) para position of the methoxy benzene.

Although the agreement between the calculated and experimental data is not absolute, the trends can be easily discerned. The prediction worked well for the meta and para substituted molecules (^{13}C chemical shift differences of 0.1 and 0.4 ppm, respectively) while differences of up to 4 ppm were observed for the ortho di-substituted molecule. ^{1}H chemical shifts predicted differ from the database only by ~ 0.1 ppm. Despite these shortcomings the position of a methoxy group can be easily deduced.

As illustrated here, the obtained chemical shifts need to be treated with caution, especially for ortho substituted molecules. It is also good practice to find a partially substituted molecule in a chemical shift database and consider only the effect of the last substituent, rather than starting these calculations from a benzene ring using Eqn. 14 and 15. Two chemical shift databases were utilised in this study, SDBS[303] and the NMR Database of Lignin and Cell Wall Model Compounds.[306]

8.4. Example of the analysis of ^{13}C-filtered NMR spectra. Identification of compound 23

This section demonstrates the combination of procedure outlined in Section 8.2 for the analysis of ^{13}C-filtered NMR spectra with the chemical shift effects described in Section 8.3. Cross-peak number **23** of Figure 8.3 (δ ^{13}C/ ^{1}H of 55.41/3.87 ppm) was analysed in this Section. A 2D F_1F_3 plane extracted from the 3D HMQC-HMBC spectrum at the methoxy carbon chemical shift of 55.41ppm shows two cross-peaks (Figure 8.5).

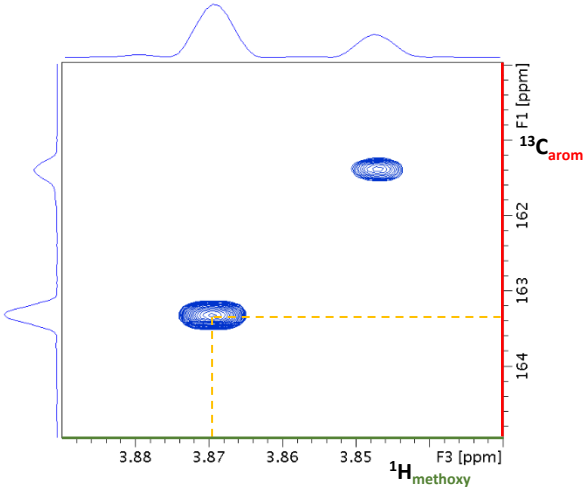


Figure 8.5 The 2D F_1F_3 (aromatic carbon vs methoxy proton) plane extracted from the 3D HMQC-HMBC spectrum of ^{13}C-methylated RMFA at the methoxy carbon chemical shift of 55.41 ppm (Molecule **23**).

The quaternary carbon at 163.32 ppm has the correct ^{1}H methoxy chemical shift (3.87 ppm) hence is identified as the ipso carbon carrying the investigated methoxy group.

To identify whether this methoxy group has protonated or quaternary aromatic carbons in the ortho positions, the 4D HCCH$_3$ spectrum was examined next. A 3D cube extracted at the 55.41 ppm methoxy carbon chemical shift is shown in Figure 8.6. This cuboid shows two cross-peaks with different ^{1}H methoxy chemical shifts indicating that there are two methoxy groups with very similar methoxy carbon chemical shifts, a fact already seen in the analysis of the 3D HMQC-HMBC spectrum. However the methoxy proton chemical shift (3.87 ppm) uniquely identifies the investigated methoxy group. The 2D projections of this cuboid (Figure 8.6) allow easy identification of the aromatic proton and carbon chemical shifts (6.93 and 113.42 ppm) of the aromatic CH moiety ortho to this methoxy group.

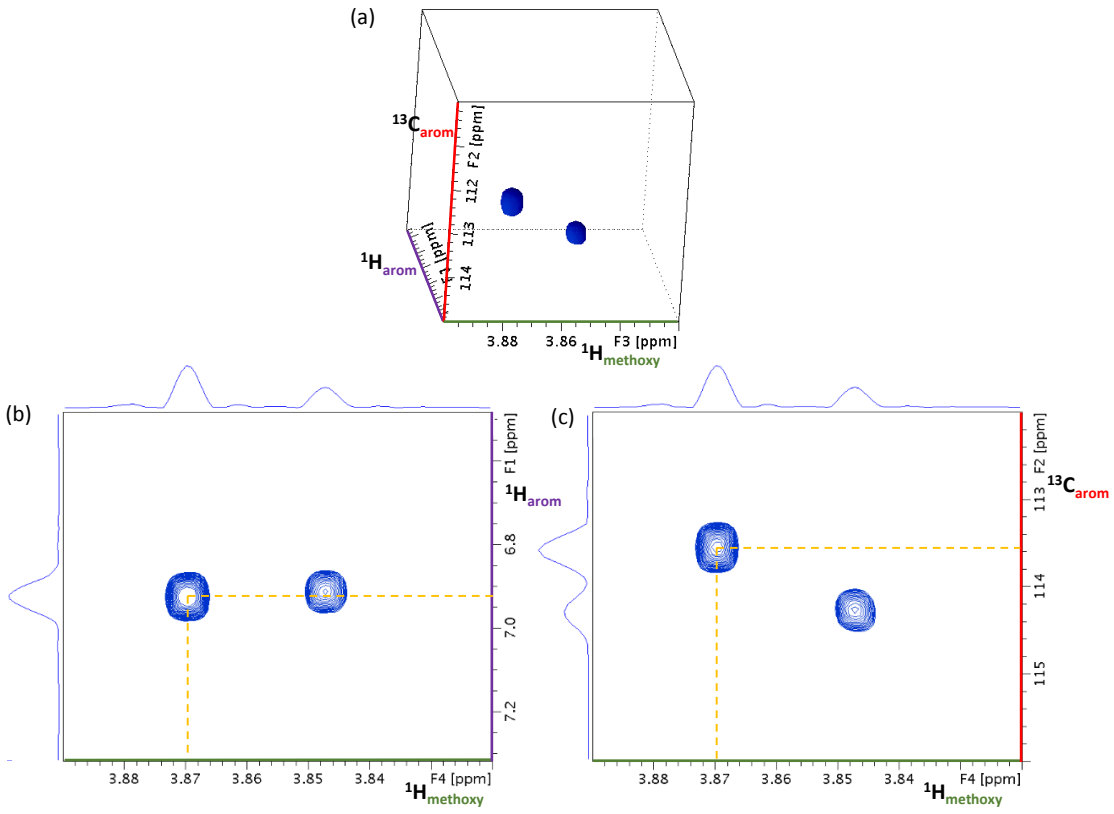


Figure 8.6 (a) A 3D cube extracted from the 4D $HCCH_3$ spectrum of ^{13}C-methylated RMFA at the methoxy carbon chemical shift of 55.41 ppm. (b) 2D F_1F_4 projection of (a), (c) 2D F_2F_4 projections of (a). The dashed orange lines correlate the chemical shifts of the methoxy proton with aromatic proton (b) and aromatic carbon (c) of molecule **23**.

Identical information was obtained from the 3D $hCCH_3$ and $HcCH_3$ experiments (data not shown). To confirm the correlations found and to look for additional quaternary carbons the F_1F_3 (DQ vs methoxy or aromatic proton) planes (Figure 8.7) were extracted from the 3D INEPT-INADEQUATE-HSQC spectrum. Two regions were inspected containing the aromatic (Figure 8.7a) and methoxy (Figure 8.7b) correlated resonances. The methoxy region is expected to contain a maximum of three cross-peaks. There are only two intense and one weak signals on the line of the methoxy chemical shift of 3.87 ppm. The two intense signals at 113.42 and 163.42 ppm have already been identified from the analysis of the previous experiments. This indicates that the fragment is symmetrical around the methoxy group. The plane extracted at the chemical shift of the aromatic carbon (113.42 ppm) shows one cross-peak at 6.93 ppm, the chemical shift already identified. This signal is split into a doublet by a large ortho coupling indicating that there is another proton at a meta position relative to the methoxy group. If the weak signal at 131.68 ppm is from molecule **23** it originates from a $^{4}J_{CH_3,C}$ coupling between the methoxy carbon and the aromatic carbon in a meta position. At this point it is unclear whether it is protonated or not.

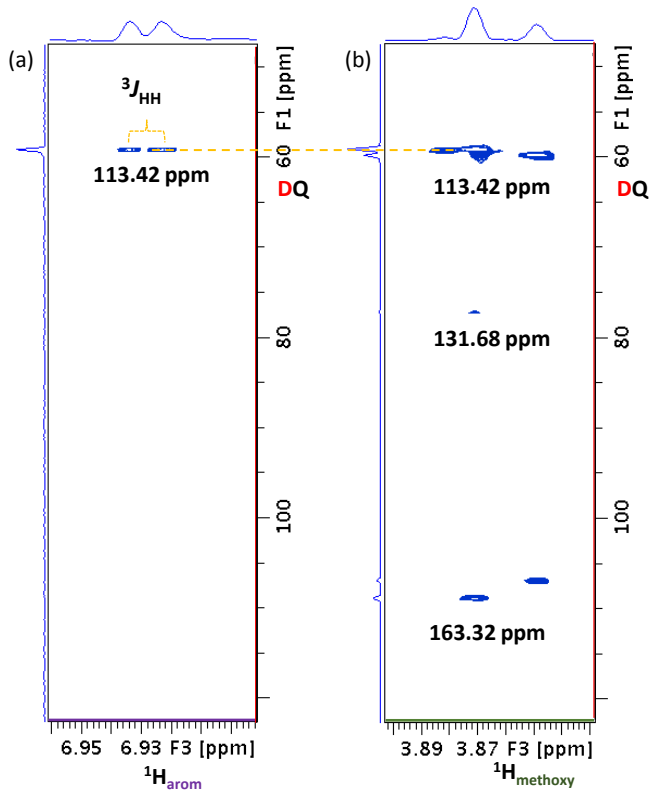


Figure 8.7 2D F_1F_3 planes extracted from the 3D INEPT-INADEQUATE-HSQC spectrum of ^{13}C-methylated RMFA via the (a) aromatic carbon at 113.42 ppm and (b) methoxy carbon at 55.41 ppm. The SQ chemical shifts of the aromatic carbons are given. The aromatic cross-peak in (a) is split by a large $^3J_{HH}$ coupling. Note that the F_1 dimension is a DQ axis and thus chemical shifts are calculated using Eqn 13.

The final spectra to be examined are the 3D HMQC-NOESY and 3D HMQC-NOESY-TOCSY. Figure 8.8 shows an overlay of 2D F_1F_3 planes extracted from the 3D HMQC-NOESY and 3D HMQC-NOESY-TOCSY spectra at the methoxy carbon chemical shift of 55.41 ppm.

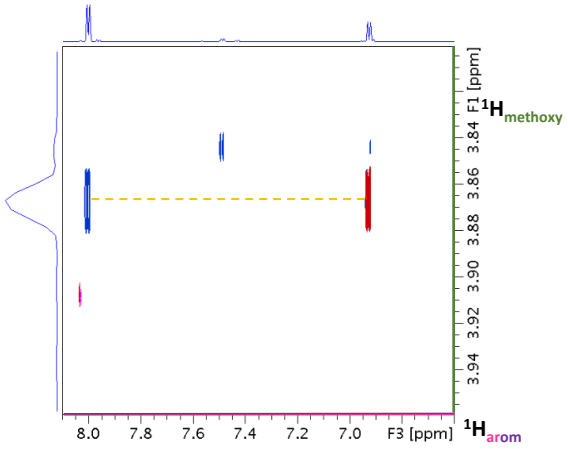


Figure 8.8 Overlay of the 2D F_1F_3 planes of 3D HMQC-NOESY (red) and 3D HMQC-NOESY-TOCSY (blue) of ^{13}C-methylated RMFA, extracted at the methoxy carbon chemical shift of 55.41 ppm.

The 3D HMQC-NOESY spectrum shows the expected correlations between the methoxy protons and the aromatic protons in the ortho position. The F_1 row at 3.87 ppm, representing the correlations from the methoxy proton, contains one cross-peak at 6.93 ppm split into a doublet by the ortho $^3J_{HH}$ splitting already identified from Figure 8.7. The other coupled aromatic protons can then be identified from the 3D HMQC-NOESY-TOCSY spectrum. The only correlation found along the F_1 row at 3.87 ppm is at 7.99 ppm. This cross-peak has a large ortho $^3J_{HH}$ splitting indicating the molecule is a para di-substituted aromatic ring. This confirms that the cross-peak at 131.68 ppm in Figure 8.7 represents a protonated aromatic carbon. The information obtained from the above experiments is compatible with the fragment shown in Figure 8.9.

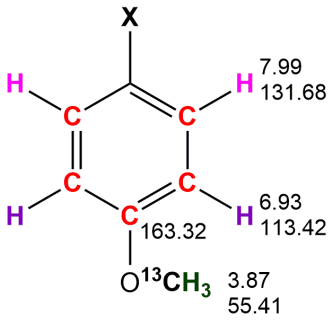


Figure 8.9 Molecular fragment identified by the analysis of a suite of ^{13}C-filtered NMR experiments using the methoxy group cross-peak at 3.87/55.41 ppm (Molecule 23).

A comparison of the fragment in Figure 8.9 with methoxy benzene indicated that the group in position C1 must have a large positive proton chemical shift effect and a carbon chemical shift effect of around +4 ppm. Examination of Table 8.1 indicates that this is typical for an electron-withdrawing group such as a carboxylic acid or ester group. As the methylation procedure was carried out with excess methylating reagent it is unlikely that the group is a carboxylic acid. To examine if the methyl ester was part of the original mixture or was created during methylation, the carboxylic region of the 3D HMQC-HMBC was examined (Figure 8.10). If the latter was the case this spectrum should show a correlation that matches the database information for methyl p-methoxy benzoate. The methoxy proton *vs* carbonyl carbon F_1F_3 projection of the 3D HMBC-HMQC spectrum shows several intense signals of methylated ester groups. For the ester group in methyl p-methoxy benzoate the lignin database of chemical shifts[306] gives chemical shifts of 166.82, 51.82, 3.86 ppm. Considering these chemical shifts as well as the large relative intensity expected for the cross-peak of this compound, Figure 8.10 indicates the possibility that molecule **23** contains a ^{13}C-methylated ester group, the potential cross-peak is indicated by an asterisk.

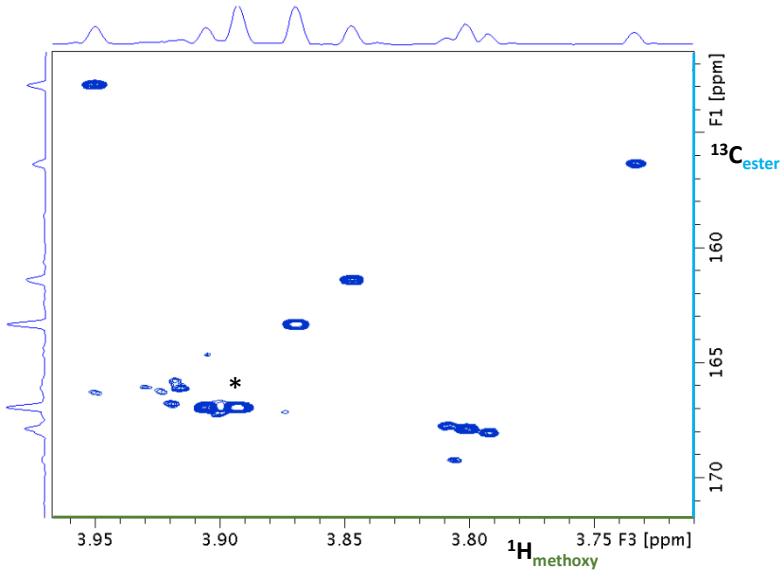


Figure 8.10 Methyl ester region of the 2D F_1F_3 projection of the 3D HMBC-HMQC spectrum of ^{13}C-methylated RMFA. The ester methoxy peak matching the chemical shift of methyl p-methoxy benzoate is labelled by an asterisk.

The 2D F_1F_3 plane extracted from the 3D INEPT-INADEQUATE-HSQC spectrum (Figure 8.11) at the methyl ester methoxy carbon (51.89 ppm) shows methyl proton vs DQ ^{13}C correlations.

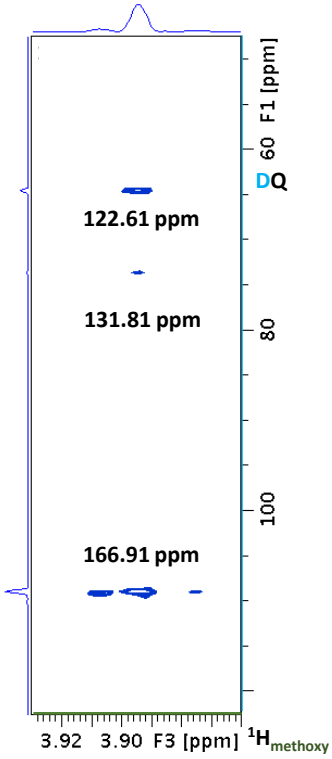


Figure 8.11 2D F_1F_3 plane extracted at 51.89 ppm from the 3D INEPT-INADEQUATE-HSQC of ^{13}C-methylated RMFA, showing methyl ester proton vs ^{13}C DQ correlations. The SQ chemical shifts are given, calculated from the DQ frequencies using Eqn 13.

This plane contains two intense and one weak cross-peaks in line with the methoxy ^{1}H chemical shift of 3.88 ppm. The SQ ^{13}C chemical shifts of these cross-peaks are 166.91, 122.61 and 131.81 ppm, respectively. The most intense correlations occur between the carbons of the ester methoxy group and the ester carbonyl group and the ipso quaternary carbon of the aromatic ring, respectively. The weak correlation is between the methoxy group carbon and the carbons ortho to the methyl ester group. This correlation was also identified in the previous 2D plane of the 3D INEPT-INADEQUATE-HSQC spectrum (Figure 8.7) and proves this methyl ester group belongs to molecule **23**. Thus, the assigned chemical shifts agree very well with the structure of methyl p-methoxybenzoate (Figure 8.12).

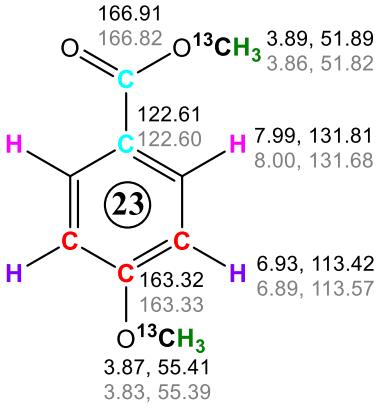


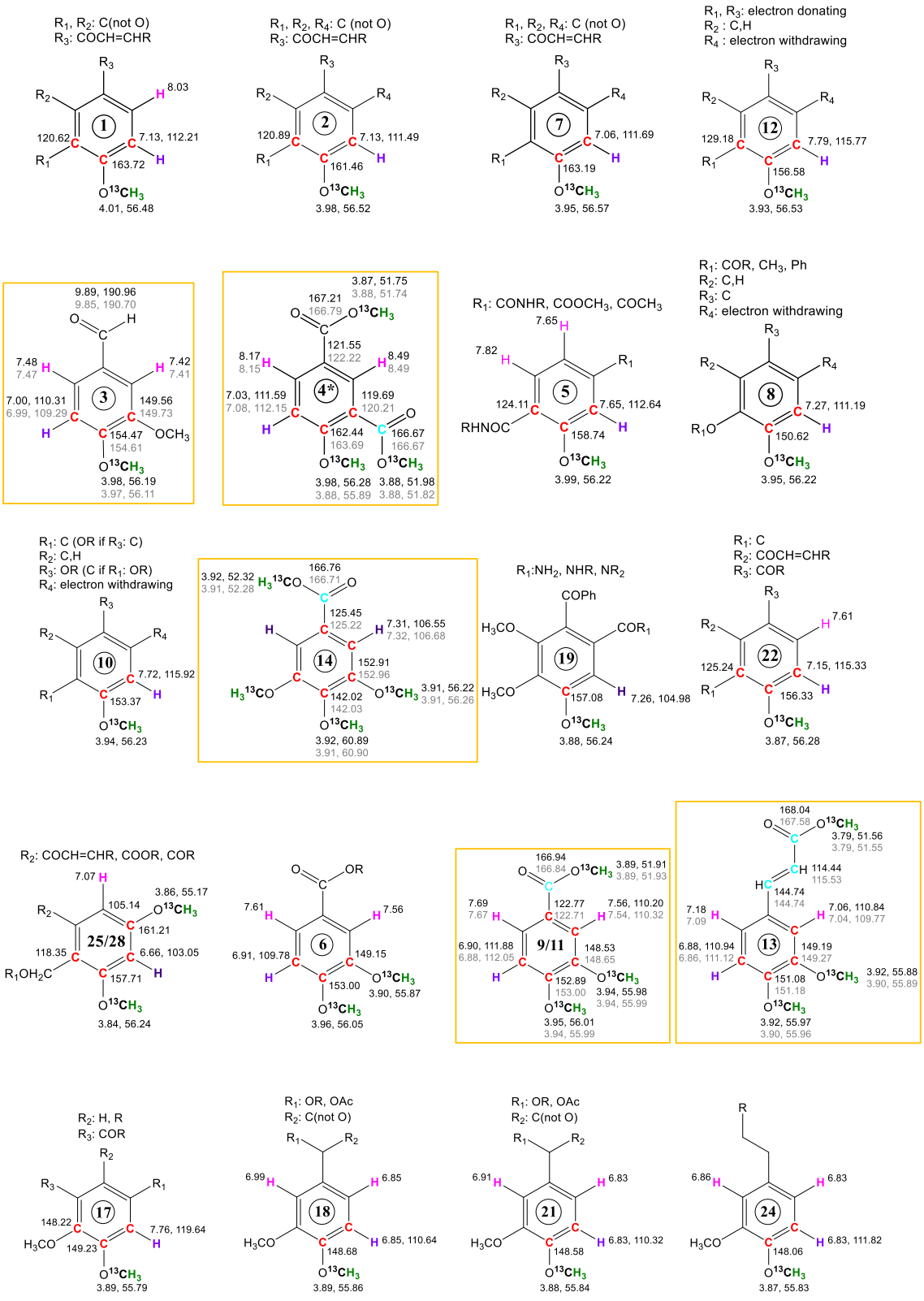
Figure 8.12 Comparison of the experimental (black) and database[303, 306] (grey) ^{1}H and ^{13}C chemical shifts obtained for molecule **23**.

It should be noted that the $^{4}J_{CH_3,C}$ couplings are small and correlations mediated by them will only be observed for the most abundant compounds. Nevertheless, even in the absence of this correlation the structure of compound **23** could be implied from the obtained chemical shifts. The procedure illustrated in this section using molecule **23** was repeated for all the methoxy cross-peaks labelled in the 2D ^{1}H, ^{13}C HSQC spectrum (Figure 8.3). The results are presented next.

8.5. Categorisation of the phenolic molecules identified in the ^{13}C-methylated RMFA

The same procedure outlined in the previous section was used to assign molecular structures to the molecular fragments obtained from the analysis of the NMR experiments from the ^{13}C-methylated RMFA sample. The detailed arguments leading to the structures obtained are given

in the Appendix C (see external weblink). Here a summary of the molecules is given (Figure 8.13). The order in which the molecules are presented follows the increasing ^{13}C methoxy chemical shift.



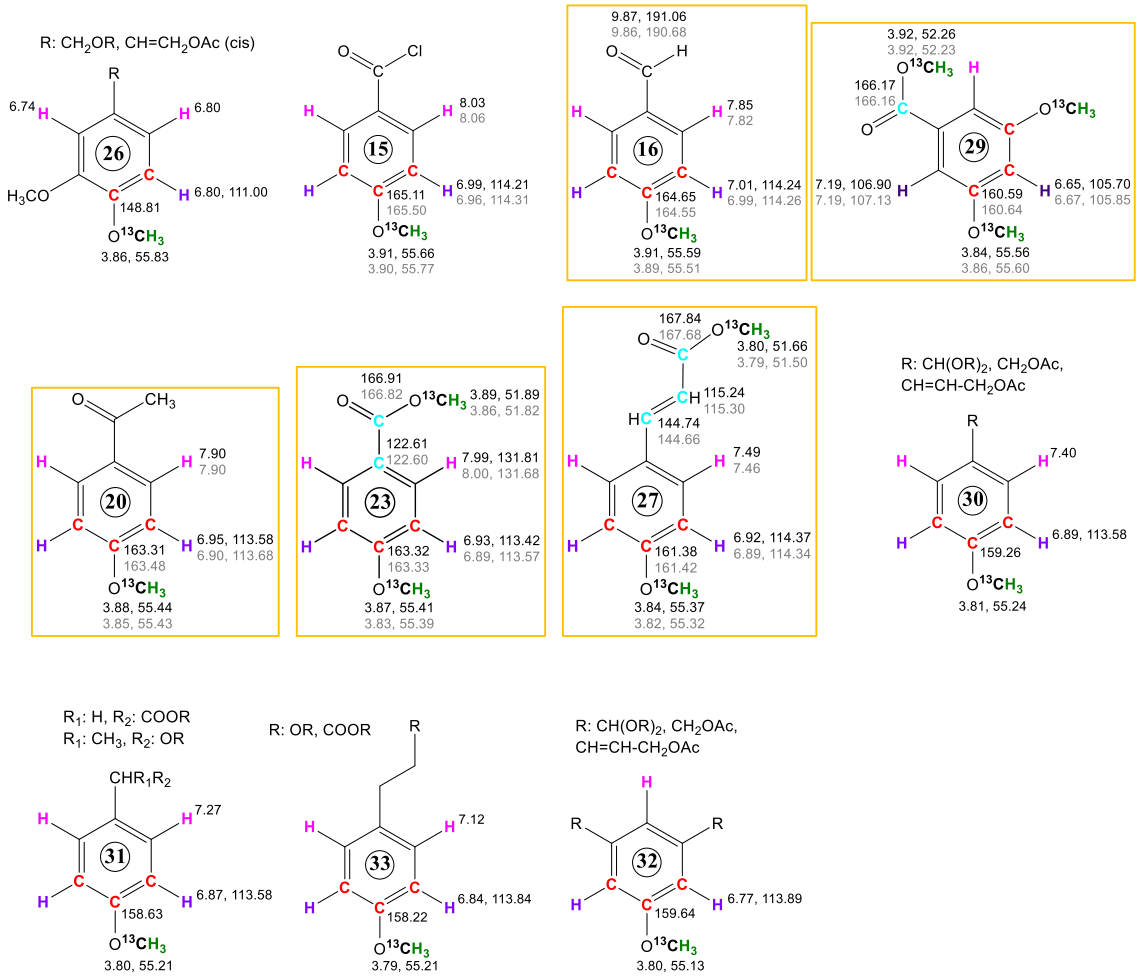


Figure 8.13 Phenolic molecule identified from the analysis of the ^{13}C-filtered NMR spectra of RMFA. The nuclei whose chemical shifts were obtained from the ^{13}C-filtered NMR experiments are coloured and boldfaced in red, purple, pink, green, black and turquoise. A comparison of the experimental (black) and database[303, 306]/calculated[304] (grey) ^{1}H and ^{13}C chemical shifts are reported on each molecule. The major compounds identified are highlighted with orange boxes. These molecules are analysed in detail in Section 8.7 and 8.8.

To determine if there is any correlation between the molecules identified and the position of the methoxy cross-peaks, the 2D ^{1}H, ^{13}C HSQC spectrum is inspected again. This spectrum showed that the signal overlap seen in 1D ^{1}H and ^{13}C spectra is reduced substantially by spreading the signals into two dimensions. A typical correlation between the proton and carbon chemical shifts results in signals appearing along the diagonal of this 2D matrix. This spread of resonances enabled the spectrum to be divided into four regions labelled (a) to (d) as shown in Figure 8.14.

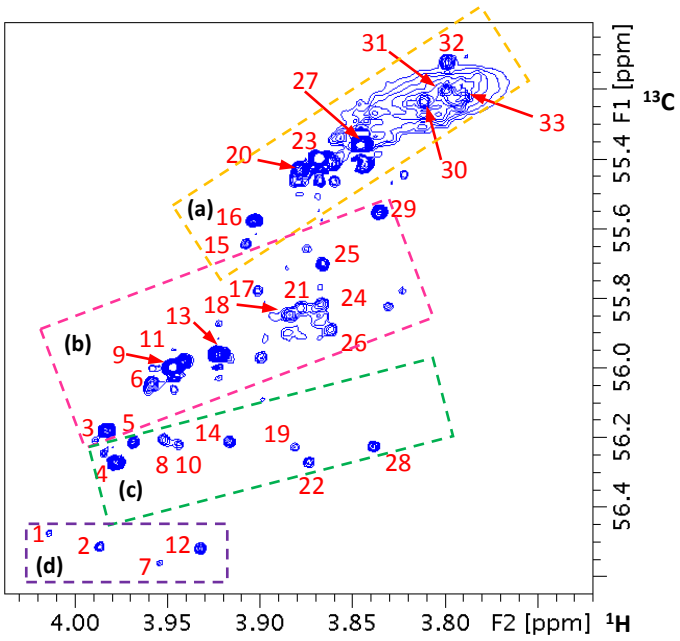


Figure 8.14 Methoxy region of the ^{1}H, ^{13}C HSQC (shown again) split into regions indicated by dashed rectangles. For further explanation, see text below.

Region (a) contains resonances of methoxy groups of p-substituted benzene rings, while region (b) contains methoxy groups of 3,4-disubstituted rings, with an exception of cross-peak **25** that belongs to one of the two 3,5 methoxy on 3,5 substituted molecule **25/28**. Region (c) is the most heterogeneous in terms of the structures of corresponding molecules. Cross-peak **28** belongs to the other methoxy group of the molecule **25/28**, while cross-peaks **14** and **19** belongs to compounds containing three methoxy groups. Molecule **19** only contains one ^{13}C-labelled and two unlabelled methoxy groups, while in molecule **14** all OH groups at positions 3, 4, 5 were ^{13}C-methylated (the C4 OMe group is not shown in Figure 8.14, as its carbon resonates at 60.89 ppm). The remaining molecules of this group had only one methoxy group located next to an unsubstituted CH site on otherwise highly substituted rings containing 2 to 3 additional substituents. Region (d) contains signals of methoxy groups located next to an unsubstituted CH site on otherwise highly substituted rings containing 3 to 4 additional non-oxygen substituents, i.e. similar to some molecules of the previous group, albeit with a different character of substituents causing more deshielding of methoxy resonances. The classification of different regions of the ^{1}H, ^{13}C HSQC spectrum shown in Figure 8.14 will enable characterisation of molecules of ^{13}C-methylated HS just through simple recording of 2D ^{1}H, ^{13}C HSQC spectra. As illustrated here, caution has to be exercised, as this classification is not absolute. Further samples need to be analysed by the developed methodology in order to verify the accuracy of this method.

It would appear from analysis of the obtained compounds, that the majority are multiple substituted benzene rings. In order to compare the relative sizes (hydrodynamic radii) of the phenolic molecules contained within ^{13}C-methylated RMFA with those in ^{13}C-methylated model mixture **II** a 2D DOSY spectrum was acquired. The spectral comparison is provided next.

8.6. Diffusion-ordered spectroscopy (DOSY) of ^{13}C-methylated RMFA

The 2D DOSY spectrum obtained from ^{13}C-methylated RMFA is overlaid with that of methylated model mixture **II** in Figure 8.15a and b.

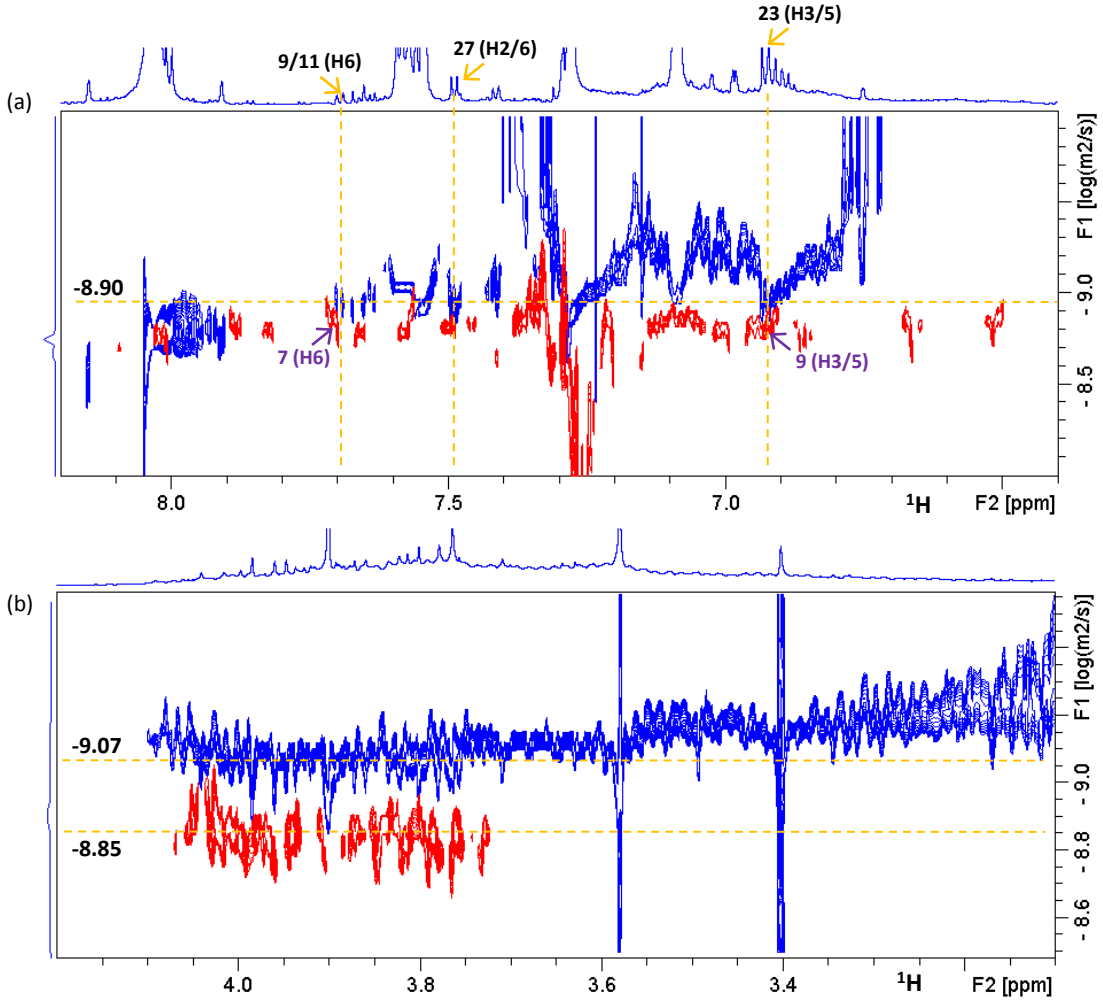


Figure 8.15 Overlay of 2D DOSY spectra of ^{13}C-methylated RMFA (blue) and ^{13}C-methylated model mixture **II** (red). (a) Shows the aromatic region. The orange horizontal dashed line indicates the approximate position of DOSY peaks for compounds **9/11**, **23** and **27** at logD of -8.90. The purple arrows indicate the position of DOSY peaks for compounds **7** and **9** of ^{13}C-methylated model mixture **II** (identical compounds to those identified in 9/11 and 23 of ^{13}C-methylated RMFA) (b) shows the methoxy region with dashed orange lines indicating the average logD values for the methoxy protons associated with aromatic rings for both samples.

A direct comparison of the two spectra is enabled by the fact that two out of the three major methylated compounds identified in the ^{13}C-methylated RMFA sample, compounds **9/11** and **23**, are compounds **7** and **9** of the ^{13}C-methylated model mixture **II**. It can be seen at the top of Figure 8.15 that the aromatic region of the ^{1}H spectrum of ^{13}C-methylated RMFA sample contains signals of varied intensity; nevertheless, it was possible to identify protons H6 of **9/11** and H3/5 of **23** (see the top projection of the DOSY spectrum on Fig. 8.15a). The corresponding DOSY signals lie approximately on the line of logD at -8.90. This value is very close to the DOSY cross-peaks of corresponding compounds in ^{13}C-methylated mixture **II** (purple arrows in Figure 8.15a), indicating that the diffusivity of these molecules in both samples is very similar. The resonance signal of the H2/6 protons of compound **27,** another major compound of the RMFA sample, at 7.49 ppm is also identifiable from the rest. The DOSY signals from these protons also lie on the line at logD of -8.90. The signals of the remaining 30 compounds identified in the ^{13}C-methylated RMFA sample are all hidden by other ^{1}H resonances.

The methoxy region of these 2D DOSY spectra can also be inspected (Figure 8.15b). The mean position of the aromatic methoxy signals at logD of -8.85 and -9.07 for the ^{13}C-methylated model mixture **II** and RMFA sample, respectively, are indicated by the dashed line on Figure 8.15b. The difference between these two values translates into a factor of 1.7 between the mean diffusion coefficients. While the logD values for the ^{13}C-methylated model mixture **II** obtained from the aromatic and methoxy protons are in perfect agreement. This cannot be said for the ^{13}C-methylated RMFA sample, where the methoxy protons appear to be associated with larger molecules. This discrepancy can be explained by the fact that the methoxy protons in the RMFA sample are completely overlapping with protons of all types of ester and sugar aliphatics (see Figure 8.1). The measured diffusion coefficient cannot therefore be interpreted rigorously. Nevertheless, the molecules carrying phenolic methoxy groups resonating between 3.7 and 4.2 ppm still show the largest apparent diffusion coefficients (smaller molecular size) amongst the other types of methoxy containing molecules.

The results of 2D DOSY analysis of ^{13}C-methylated RMFA confirm that the molecules fully identified in the RMFA sample are of similar size to that of the ^{13}C-methylated model mixture **II**. This indicates that the major phenolic molecules in RMFA are small molecules. Further investigation through the acquisition of a pseudo 3D DOSY-HSQC spectrum focusing on methoxy resonances is envisaged in order to analyse the relative sizes of more ^{13}C-methylated compounds.

8.7. Comparison of the phenolic molecules identified with literature data

As discussed in Section 1.5.2.6, there are only a few NMR-based studies that have attempted to identify phenolic molecules in HS. Some of the aromatic ring substitution patterns identified in this work were previously suggested based on the bulk matching of experimental and theoretical 2D ^{1}H, ^{13}C HSQC spectra by considering the effects of -OR and -COOR groups on the ^{1}H and ^{13}C chemical shift of aromatic rings.[238, 240] These approaches relied on matching a limited number of descriptors to numerous possibilities. As a consequence, and due to the reliance on the ortho couplings of aromatic protons, mostly para disubstituted benzene derivatives were identified and only a rudimentary elaboration of the nature of the substituents was possible.[238] The 3D HSQC-TOCSY experiment applied to plant derived HS[72] did not provide qualitatively higher information than the discussed 2D experiments, even though it was able to identify the presence of 3,4 dialkoxysubstituted molecules. In contrast, the methodology presented here provides a multitude of correlated chemical shifts, which allows convergence to a single structure/or a defined structural motif, thus yielding unprecedented structural details for phenolic compounds in HS samples.

As mentioned in Section 1.6.2, pyrolysis is a chromatography based technique for the compositional analysis of HS.[307] In this method, fragmented products, obtained by applying thermal energy, are analysed by GC/MS[6] or field ionization mass spectrometry.[308] The advantages of this method are the easy pre-treatment and the use of a small amount of sample; however, the basic method suffers from a number of undesired side reactions such as decarboxylation, cyclisation and dehydrogenation.[140] These reactions were reported to be minimised by means of simultaneous pyrolysis and methylation using TMAH.[249] Such protection of undesired modification of thermally or chemically labile functional groups gave way to a one-step reaction referred to as 'thermochemolysis'.[309] TMAH GC-MS analysis of 18 HA samples from different soils identified a number of phenolics:[140] amongst these, six of the ten major compounds (see Table 8.2) identified in this work (**3**, **9/11**, **14**, **16**, **20** and **27**). It is not possible to directly compare these results as a study using an identical sample is needed to evaluate the merits of both techniques. However, there are several issues with the TMAH GC-MS analysis that should be noted.

(i) Despite reducing the occurrence of potentially modifying side reactions through simultaneous methylation, there is no guarantee, that this problem is fully eliminated. This could lead to compounds being missed or falsely identified.

(ii) In order to positively identify the eluted compounds, standards are required. Indeed, 13 phenol derivatives were used in the study by Ikeya,[140] including all six compounds identified by both methods. This means that *de novo* structure determination by thermochemolysis is problematic.

(iii) The basic thermochemolysis cannot distinguish OH and OCH_3 substituents. This makes it impossible to ascertain relative inputs of fresh lignin, demethylated lignin and polyphenols into HS. To overcome this limitation, ^{13}C-TMAH thermochemolysis that methylates using a ^{13}C-labelled methyl group permits the differentiation of original and added methyl groups by MS and therefore the potential source of many 'lignin' phenols.[134] As illustrated above, this issue can be addressed by NMR provided the resonances of $^{13}CH_3O$ groups from the same molecule can be identified.

Quantification by both methods is problematic as the efficiency of methylation cannot be fully ascertained; thermochemolysis is, in addition, affected by evaporation efficiency and the extent of chromatographic separation, while NMR requires signal separation to integrate peaks as the basis for quantification. As mentioned above a comparative study of both methods on the same sample is required to quantitatively assess their performance.

8.8. Tracing the origin of the phenolic molecules identified in ^{13}C-methylated RMFA

The application of ^{13}C-filtered nD NMR spectroscopy to the ^{13}C-methylated RMFA sample has identified 31 phenolic compounds/structural motifs. From this set, the structures of 11 compounds were fully characterised (Figure 8.13) and these will be examined further in this section. Except for compounds **4**, **15** and **29,** the remaining eight compounds have been found by others in the plants present on the Red Moss sampling site (Table 8.2). Compound **15** (4-methoxybenzoyl chloride) is unlikely to be of plant origin and could be a contaminant. Acyl chlorides react with water readily, indicating that if this functional group is found in a soil compound, it must be protected from contact with water, potentially encapsulated in HS

aggregates or micelles. It is also possible that the limited number of obtained chemical shifts for this compound, despite an excellent match to 4-methoxybenzoyl chloride, could belong to another compound of unknown structure. The structure of this molecule therefore requires further verification. While compound **29** (3,5 dihydroxybenzoic acid), has been found in the hill raspberry, this plant has not been reported as common to the Red Moss Balerno raised bog. A search of the literature did not link this compound to other plant species present, such as crowberry and huckleberry. This could be interpreted as either compound **29** (i) has yet to be identified in the plant species present or (ii) its source is unknown. Finally, compound **4** was not found in the plants present in the Red Moss site, but has been described in *Decalepis hamiltonii*,[310] a species of plant in the family *Apocynaceae*, which is endemic to Peninsular India but the origin of this compound in the ^{13}C-methylated RMFA sample is unknown.

With these three exceptions, it was possible to link the remaining eight compounds to the major plants present on the RMFA collection site (see Table 8.2). The first four compounds listed in Table 8.2, **23**, **27**, **20** and **16** are derived from p-hydroxybenzoic acid, p-coumaric acid, p-hydroxyacetophenone and p-hydroxybenzaldehyde, respectively. 4-hydroxybenzoic acid is the most abundant phenolic compound in the RMFA sample and the quantity of all other major compounds is stated in Table 8.2 relative to this compound. All four compounds are amongst the major compounds identified by Mellegard[124] in *Sphagnum papillosum* albeit in different ratios than found in the RMFA sample. Compounds **23** and **27** were also identified in *Sphagnum fallax* and *Sphagnum cuspidatum*.[125] while compound **20** was also identified in *Sphagnum Magellanicum*.[125]

Sphagnum acid and its derivatives (Figure 1.5) are the only other *Sphagnum* p-hydroxyphenols reported in the literature. When the chemical shifts of the 31 compounds were compared with the chemical shifts of the aromatic moiety of sphagnum acid, reported in acetone-d_6,[125] those of compound **30** matched most closely. The methoxy cross-peak of compound **30** is located in the middle of unresolved cross-peaks in the p-hydroxy phenol region of the ^{1}H, ^{13}C HSQC spectrum of the ^{13}C-methylated RMFA (Figure 8.14). It is very likely that at least some of the signal intensity seen in this part of the spectrum originates from *Sphagnum* acid and its derivatives. This includes compounds **31**, **32** and **33**. Their positive identification would require measuring spectra of ^{13}C-methylated *Sphagnum* acid and its derivatives in $CDCl_3$.

Table 8.2 Major compounds identified in ^{13}C-methylated RMFA, their relative abundance and potential sources.

Number	Compound	Relative abundance[a] /%	Source
23	Methyl 4-methoxybenzoate	100	*Sphagnum* moss: 4-hydroxybenzoic acid.[124, 125]
27	Methyl 4-methoxycinnamate/ Methyl 3-(4-methoxyphenyl) acrylate*	57	*Sphagnum* moss/cotton grass/heather: p-coumaric acid (4-hydroxycinnamic acid).[124, 125]
20	4-Methoxyacetophenone	14	*Sphagnum* moss: 4-hydroxyacetophenone.[124, 125]
16	4-Methoxybenzaldehyde	10	*Sphagnum* moss: 4-hydroxybenzaldehyde.[124]
9/11	Methyl 3, 4-dimethoxy benzoate	55/11[b]	Heather: vanillic acid (4-hydroxy-3-methoxybenzoic acid).[131, 311] *Sphagnum moss*: 3, 4-dihydroxybenzoic acid and other non-vascular plants.[118, 312]
13	3, 4-Dimethoxycinnamate/ Methyl 3-(3, 4-dimethoxyphenyl) acrylate*	26[c]	Heather: caffeic acid (3, 4-dihydroxycinnamate).[131, 311] Heather/Cotton grass: ferulic acid (3-methoxy-4-hydroxycinnamic acid).[131, 133, 311]
3	3, 4-Dimethoxybenzaldehyde	14	*Sphagnum* moss: 3, 4-dimethoxybenzaldehyde,[118] vanillin (4-hydroxy-3-methoxybenzaldehyde).[124]
4	Dimethyl 4-methoxy isophthalate	14	4-Hydroxy isophthalic acid.[310]
29	Methyl 3, 5-dimethoxybenzoate	11	Hill raspberry: 3, 5-dihydroxybenzoic acid.[313]
14	Methyl 3, 4, 5-trimethoxybenzoate	9/3[d]	Heather: syringic acid (4-hydroxy-3, 5-dimethoxy benzoic acid).[131]

[a] The relative abundance was determined by integrating the methoxy cross-peaks of the 2D ^{1}H, ^{13}C HSQC spectrum shown in Figure 8.3, [b] Integrals of 4-OMe/3-OMe groups, respectively. [c] The 3-OMe and 4-OMe groups of compound **13** overlap. [d] Integrals of 4-OMe/3,5-OMe groups. For interpretation of these results see text below. * Indicates IUPAC name.

The next three compounds (**9/11**, **13** and **3**) are 3, 4 –OR substituted. Compound **3** is vanillin derived, as only the C4-hydroxyl group was found to be ^{13}C-methylated, indicating that C3 carries a $^{12}CH_3O$- group. This molecule has been detected in several *Sphagnum* species,[118, 124] but could also be classified as a lignin compound. For the other two compounds, signals of both C3 and C4

$^{13}CH_3O$- groups were found indicating that these compounds have hydroxyl substituents amenable to methylation. If only the dihydroxy molecules were present, integrals of both methoxy signals would be the same. However, the measured ratio of integrals of the C4 and C3 attached $^{13}CH_3O$ groups of compounds **9/11** is 5:1 and indicates that some of the **9/11** parent molecules carry a $^{12}CH_3O$ group. Because the intensity of the C4-$^{13}CH_3O$ signal was higher, it can be concluded that a coniferyl alcohol type monolignols (Figure 1.3) with a $^{12}CH_3O$ group at C3 is the major compound. The observed integral ratio translates into the 4:1 ratio of 4-OH, 3-OCH_3 and 3, 4-OH carrying molecules. With the rest of the chemical shifts being compatible with a COOH group at C1, compound **9/11** represents lignin-associated vanillic acid[131] (3-methoxy, 4-hydroxy benzoic acid) and protocatechuic acid (3, 4-dihydroxy benzoic acid) which is found in nonvascular plants[312] in a 4:1 ratio.

Similar analysis for molecule **13,** whose origin could be either ferulic acid (3-methoxy, 4-hydroxy substitution) or a caffeic acid (3, 4-dihydroxy substitution), could not be performed, as the signals of the $^{13}CH_3O$ groups attached at positions C3 and C4 are fully overlapped. Both acids have been reported to be present in heather and a cotton grass (see Table 8.2).[131, 133, 311]

It should be noted that several C3-$^{12}CH_3O$/C4-$^{13}CH_3O$ containing compounds (**17**, **18**, **21**, **24** and **26**) are amongst the 31 identified structures. However, the low intensity of their methoxy cross-peaks did not allow similar analysis to be performed.

Compound **14** is axially symmetrical and contains chemically equivalent C3 and C5 methoxy groups and a unique C4 methoxy group. Once again, the origin of this compound could be a gallic acid (3, 4, 5-trihydroxy substitution) or syringic acid (3, 5-dimethoxy, 4-hydroxy substitution) (see Table 8.2). The observed ratio of integrals (9:3) for C4-$^{13}CH_3O$ and 3, 5-$^{13}CH_3O$ groups indicates that syringic acid and gallic acid are present at the 5:1 ratio. The former was found in heather,[131] while the latter is present in tannins found in ericoids.[134]

Overall, it would therefore appear that for both compound **9/11** and **14**, the lignin-based sources prevail. Nevertheless, as demethylation can occur in an anaerobic environment, and the peat extracted in this study was under anaerobic contains, further investigation is required to relate these ratios to the original compounds.

In addition to already discussed *Sphagnum* acid like molecules **30-33,** several structures with highly substituted benzene rings were identified. Nevertheless, in accord with the literature, para

substituted phenols are prevalent in nature[133] of which this study identified eight compounds (**16**, **29**, **20**, **23**, **27**, **30-33**). These are more likely to originate from non-vascular plants. Another large group of identified compounds were several partially characterised 3, 4 di-OR substituted benzenes (**6**, **17**, **18**, **21**, **24**, **26**), in addition to fully characterised compounds **3**, **9/11**, **13** discussed above. These compounds, differing in the nature of the C1 attached sidechain, are more likely to originate from vascular plants.

In conclusion, this study indicates that the majority of compounds identified by ^{13}C-filtered NMR experiments in the ^{13}C-methylated RMFA sample are directly linked to the plants that cover the bog from which the sample was taken. Establishing such a link is a significant result. Under anaerobic conditions the decay of plants is slow; some of the compounds released during the bursting of plant cells at the end of their cycle, are transported to the anoxic soil environments, where they are prevented from further demineralisation. This is supported by the two aldehyde groups found in Figure 8.13, which belong to molecules **3** and **16**. As an aldehyde is a relatively reactive group, these compounds must have been preserved in anoxic conditions and not oxidised during the sample preparation.

8.9. Chapter conclusions

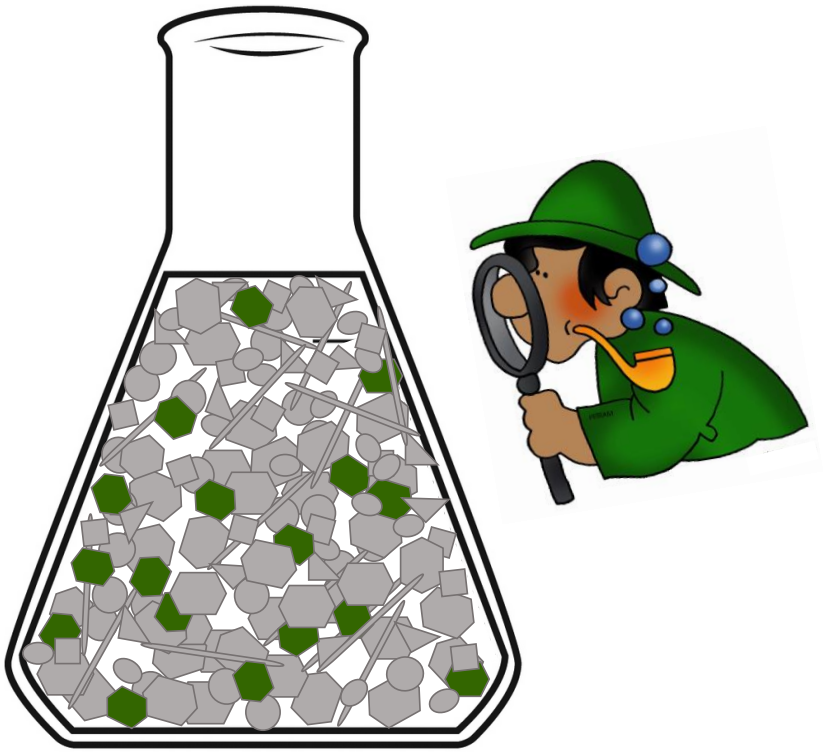
From the combined analysis of the data collected from the ^{13}C-filtered NMR experiments 31 molecules/moieties were identified in the methylated RMFA sample. The compounds fully identified consist of a number of mono and di-substituted benzoates, ketones and aldehydes. For some of the benzoates it was possible, using the 3D INEPT-INADEQUATE-HSQC spectrum, to link the ester group to the aromatic ring. Furthermore, several compounds containing an alkene substituent akin to the most common monolignols were identified. The remaining compounds contain a number of substituents that, at this time, could not be fully structurally characterised. It is also possible that a number of incomplete aromatic rings investigated are part of larger structures containing more than one aromatic ring.

It is encouraging that regions of the 2D ^{1}H, ^{13}C HSQC spectrum of ^{13}C-methylated RMFA sample could be associated with certain compound classes. A quick acquisition of such spectra of ^{13}C-methylated HS samples could provide reasonable indication of the compounds present.

A comparison of the 2D DOSY spectra obtained for ^{13}C-methylated model mixture II and the ^{13}C-methylated RMFA sample confirmed that the RMFA sample contains very small molecules.

From the 31 molecules/moieties identified, 11 of the compounds were fully characterised. By comparing these with metabolites found in the flora present at the Red Moss sampling site, 8 compounds could be directly linked to species of heather, *Sphagnum* moss or cotton grass. This is a significant result as it indicates that molecules present in HS are clearly related to the source plants and protected by the peat bog conditions from further breakdown. This is particularly exemplified through the identification of aldehydes, which are a very reactive compound class.

Chapter 9. Conclusions and recommendations for future work



(Image of the detective courtesy of Clipart, phillipmartin.info)

One alone cannot change the world but one can throw a stone across the waters and create many ripples.

Mother Theresa

Chapter 8 demonstrated the combined application of the methylation protocol and the ^{13}C-filtered nD NMR experiments to the identification of phenolic molecules in the RMFA sample and traced a number of the major compounds identified to flora present at the sampling site. This chapter provides the final concluding remarks as well as suggested future directions aimed at extending the work conducted in this study.

9.1. Concluding Remarks

Structure determination of molecules contained within complex mixtures is at the frontier of analytical chemistry. When chromatography fails to separate individual compounds, even NMR spectroscopy, the most powerful solution-based structure elucidation technique, is unable to elucidate individual molecular structures. In this work a novel "spectroscopic separation" technique has been proposed, which relies on the introduction of an NMR active nucleus (by modifying specific functional groups of unseparated molecules) to provide a way of reducing the overwhelming amount of resonances detected by NMR when studying complex mixtures. The signals from the tagged molecules are spread into several dimensions of nD NMR spectra providing multiple correlated chemical shifts and coupling constants of the parent molecules. This is possible due to the scalar and dipolar couplings that exists between the nuclei of the tag and the nuclei of the parent molecule. The data obtained was interpreted using spectroscopic knowledge and database information to suggest the molecular structure(s) compatible with the observations. This approach represents a paradigm shift in the application of isotopically labelled tags, away from relying only on the signal of the tag, towards using the tag to report on the parent molecule in a more direct way.

The focus of this study were the phenolic compounds contained in peat. These were studied as part of an operational fraction of HS, so-called FA, prepared from Red Moss raised peat bog, Balerno, Scotland. The hydroxyl groups and carboxyl groups of phenolic compounds are the prevailing functional groups of HS and are important to many functions of soil. It is therefore very important to understand their individual molecular structures. Such knowledge will eventually be able to explain their roles e.g. in soil fertility or the longevity of peat bogs. Phenolics have been shown to interfere with the degradation processes in anoxic waterlogged conditions and hence play an important role in modulating the carbon storage capabilities of peat. Knowledge of their structures will enable the study of their interactions with degrading enzymes

at a molecular level, contributing to efforts in designing novel technologies for carbon storage or aiding peat restoration attempts.

The structures/structural motifs of 31 phenolic compounds in RMFA were identified in this study by attaching $^{13}CH_3O$- groups to their skeletons and designing novel NMR experiments and analysing the obtained spectra. The identified compounds were small molecules, containing only a single aromatic ring, which can be classified as of both lignin and non-lignin origin. It was possible to link 8 of the 11 major compounds identified in the RMFA sample to the flora of the collection site, in particular *Sphagnum* moss and *Calluna vulgaris*. It is therefore evident that these compounds were sourced from the decaying plants covering the peat bogs. Throughout this work, NMR has been shown to be capable of determining the structure of individual molecules in complex mixtures.

It is a basic premise of this research that in order to get a true handle on the function of soil, it has to be understood at a molecular-scale. The results achieved in this study provide the first steps in a long journey towards this goal.

9.2. Recommendations for future work

The application of ^{13}C-filtered NMR spectroscopy successfully solved the structures of major phenolic molecules in an FA sample extracted from a raised peat bog. This represents the initial steps in the full-scale molecular characterisation of HS by NMR. The principles behind this approach can be explored in many different directions.

Firstly, the NMR experiments were carried out on a 10 mg ^{13}C-methylated FA sample. This small amount limited the sensitivity of the measurements. Increasing the amount of methylated sample and repeating the outlined experiments is expected to yield insight into the structures of less abundant phenolic compounds present in this sample. This goal can also be achieved by increasing the magnetic field strength of the spectrometer and re-acquiring a few crucial experiments on a 1 GHz NMR spectrometer.

The RMFA sample analysed in this study was collected from the 20-30 cm layer. Samples were collected from another two layers, 0-10 and 10-20 cm. These samples could be methylated and, at least, their 2D ^{1}H, ^{13}C HSQC characterised and the results compared across the different layers. This could give an insight into the fate of the phenolic compounds with peat depth.

A general issue that needs to be addressed by the whole soil science community is the treatment of soil and HS sample preparation. There is an increasing perception in the scientific community that the protocols developed by the International Humic Substances Society are outdated and potentially sample altering. One particular problem with the method is the use of strong acids (pH 1) that have unknown effects on the molecules of HS. The use of strong acids could be avoided by not separating fulvic acid from humic acid and doing a base extraction only. In this case one would obtain a sample of NOM. However, in general, new protocols, when they become available, need to be compared and their effects on the composition of phenolic compounds tested.

Organic matter in other environmental matrices, such as dissolved organic matter (DOM) in lochs, rivers and seas, which may be easier to study from the point of view of sample preparation are equally amenable to the NMR studies outlined here. Deciphering their structural composition will undoubtedly contribute to our understating of global carbon cycling.

Despite the focus on methylated phenolic compounds, the procedure also methylated the aliphatic carboxyl and hydroxyl functional groups. In order to study this compound class further,

the NMR experiments have to be tailored to their analysis. Particular attention should be paid to the carboxyl groups that are abundant in HS. This applies both to aromatic and aliphatic molecules. These investigations may also require further methylation protocols to be developed for, for example, methylation of carboxyl groups only.

The tagging strategy used in this study was methylation. This involved placing a ^{13}C nucleus a minimum of 2 bond lengths from the nearest carbon atom of an HS molecule, imposing a limit to the extent of polarisation transfer. Another possibility is to place ^{19}F on the molecules. For example, the carboxyl groups could be replaced with ^{19}F by means of a decarboxylation reaction. Due to the sensitivity, a large dispersion of chemical shifts and sizeable coupling constants, ^{19}F would be an excellent reporter group in investigating the structures of HS molecules. Tagging strategies and further NMR experiments need to be developed to take full advantage of this tag.

The work presented here characterised an FA sample extracted from a raised peat bog. The other operational fraction HA requires similar characterisation. This would allow a comparison of the types of phenolic compounds in these samples. It could answer the question of whether the same compounds are in HA and if there is a diagenetic link between the two fractions. This requires resolution of issues with solubility of HA samples, which generally are much more difficult to work with.

TMAH GC-MS analysis or thermochemolysis is the only other technique used for the molecular-scale characterisation of HS compounds. It would be of interest to conduct a parallel study that would characterise identical samples.

Once the NMR procedures of structural characterisation of HA and FA have been further developed using multiple tagging strategies, the methodology should be employed to compare the composition of different peat samples as well as organic layers in other soil types. By examining the molecular composition from different sources, researchers may finally be able to make links between cause and effect and start explaining the structural functional relationship of soil at the molecular-scale.

References

[1] J. I. Hedges, J. M. Oades, *Organic Geochemistry* **1997**, *27*, 319-361.

[2] F. J. Stevenson, *Humus Chemistry: Genesis, Composition, Reactions*, John Wiley and Sons, Hoboken, New Jersey, USA, **1994**.

[3] M. W. I. Schmidt, M. S. Torn, S. Abiven, T. Dittmar, G. Guggenberger, I. A. Janssens, M. Kleber, I. Koegel-Knabner, J. Lehmann, D. A. C. Manning, P. Nannipieri, D. P. Rasse, S. Weiner, S. E. Trumbore, *Nature* **2011**, *478*, 49-56.

[4] E. M. Peña-Méndez, J. Havel, J. Patočka, *Journal of Applied Biomedicine* **2005**, *3*, 13-24.

[5] D. L. Sparks, *Geoderma* **2001**, *100*, 303-319.

[6] M. H. B. Hayes, P. MacCarthy, R. L. Malcolm, R. S. Swift, *Humic substances II. In search of structure*, John Wiley and Sons, Chichester, UK, **1989**.

[7] R. L. Wershaw, *Soil Science* **1999**, *164*, 803-813.

[8] C. M. Cooke, N. J. Bailey, G. Shaw, J. N. Lester, C. D. Collins, *Bulletin of Environmental Contamination and Toxicology* **2003**, *70*, 761-768.

[9] R. Sutton, G. Sposito, *Environmental Science & Technology* **2005**, *39*, 9009-9015.

[10] K. H. Tan, *Humic Matter in Soil and the Environment: Principles and Controversies, Second Edition*, CRC Press, Taylor and Francis Group, Boca Raton, Florida, USA, **2014**.

[11] G. Song, E. H. Novotny, A. J. Simpson, C. E. Clapp, M. H. B. Hayes, *European Journal of Soil Science* **2008**, *59*, 505-516.

[12] N. Fasurova, L. Pospisilova, *Soil and Water Research* **2011**, *6*, 147-152.

[13] R. Lal, *Science* **2004**, *304*, 1623-1627.

[14] A. Piccolo, *Humic Substances in Terrestrial Ecosystems*, Elsevier Science, Amsterdam, Netherlands, **1996**.

[15] R. Lal, *Geoderma* **2004**, *123*, 1-22.

[16] R. Spaccini, A. Piccolo, P. Conte, G. Haberhauer, M. H. Gerzabek, *Soil Biology & Biochemistry* **2002**, *34*, 1839-1851; R. Lal, R. F. Follett, J. Kimble, C. V. Cole, *Journal of Soil and Water Conservation* **1999**, *54*, 374-381.

[17] M. Wang, S. Chang, *Canadian Journal of Soil Science* **2001**, *81*, 299-307.

[18] Y. Harada, A. Inoko, *Soil science and plant nutrition* **1975**, *21*, 361-369.

[19] G. Davies, E. A. Ghabbour, C. Steelink, *Journal of Chemical Education* **2001**, *78*, 1609-1614.

[20] P. Janos, V. Hula, P. Bradnova, V. Pilarova, J. Sedlbauer, *Chemosphere* **2009**, *75*, 732-738.

[21] S. Wang, C. N. Mulligan, *Chemosphere* **2009**, *74*, 274-279.

[22] I. V. Perminova, K. Hatfield, N. Hertkorn, *Use of Humic Substances to Remediate Polluted Environments: From Theory to Practice: Proceedings of the NATO Adanced Research Workshop on Use of Humates to Remediate Polluted Environments: From Theory to Practice, held in Zvenigorod, Russia, 23-29 September 2002, Vol. 52*, Springer, Dordrecht, Netherlands, **2005**.

[23] A. Helal, D. Imam, S. Khalifa, H. Aly, *Radiochemistry* **2006**, *48*, 419-425.

[24] R. S. Ray, S. Krueger, N. Roesch, *Dalton Transactions* **2009**, 3590-3598.
[25] J. F. McCarthy, J. M. Zachara, *Environmental Science & Technology* **1989**, *23*, 496-502.
[26] N. Senesi, *Science of the Total Environment* **1992**, *123*, 63-76; S. Stefaniak, I. Twardowska, H. E. Allen, M. M. Häggblom, *Viable Methods of Soil and Water Pollution Monitoring, Protection and Remediation*, Springer, Dordrecht, Netherlands, **2007**.
[27] P. M. Bertsch, J. C. Seaman, *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 3350-3357.
[28] P. Gerland, A. E. Raftery, H. Sevcikova, N. Li, D. Gu, T. Spoorenberg, L. Alkema, B. K. Fosdick, J. Chunn, N. Lalic, G. Bay, T. Buettner, G. K. Heilig, J. Wilmoth, *Science* **2014**, *346*, 234-237.
[29] D. Tilman, C. Balzer, J. Hill, B. L. Befort, *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 20260-20264.
[30] J. A. Foley, N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. O'Connell, D. K. Ray, P. C. West, C. Balzer, E. M. Bennett, S. R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockstrom, J. Sheehan, S. Siebert, D. Tilman, D. P. M. Zaks, *Nature* **2011**, *478*, 337-342.
[31] D. Tilman, K. G. Cassman, P. A. Matson, R. Naylor, S. Polasky, *Nature* **2002**, *418*, 671-677.
[32] J. Penuelas, B. Poulter, J. Sardans, P. Ciais, M. van der Velde, L. Bopp, O. Boucher, Y. Godderis, P. Hinsinger, J. Llusia, E. Nardin, S. Vicca, M. Obersteiner, I. A. Janssens, *Nature Communications* **2013**, *4*.
[33] J. Elser, E. Bennett, *Nature* **2011**, *478*, 29-31.
[34] V. A. Mylonas, C. B. McCants, *Journal of Plant Nutrition* **1980**, *2*, 377-393.
[35] B. S. Rauthan, M. Schnitzer, *Plant and Soil* **1981**, *63*, 491-495.
[36] A. Nikbakht, M. Kafi, M. Babalar, Y. P. Xia, A. Luo, N.-a. Etemadi, *Journal of Plant Nutrition* **2008**, *31*, 2155-2167.
[37] F. Adani, P. Genevini, P. Zaccheo, G. Zocchi, *Journal of Plant Nutrition* **1998**, *21*, 561-575.
[38] O. Turkmen, A. Dursun, M. Turan, C. Erdinc, *Acta Agriculturae Scandinavica Section B-Soil and Plant Science* **2004**, *54*, 168-174.
[39] Y. S. Lee, R. J. Bartlett, *Soil Science Society of America Journal* **1976**, *40*, 876-879.
[40] P. C. De Kock, *Science* **1955**, *121*, 473-474.
[41] A. Muscolo, M. Sidari, S. Nardi, *Journal of Geochemical Exploration* **2013**, *129*, 57-63.
[42] Z. Sladky, *Biologia Plantarum* **1959**, *1*, 142-150.
[43] Y. Chen, T. Aviad, *Humic substances in soil and crop sciences: Selected readings* **1990**, 161-186.
[44] C. H. Liu, R. J. Cooper, D. C. Bowman, *Hortscience* **1998**, *33*, 1023-1025.
[45] S. Nardi, D. Pizzeghello, A. Muscolo, A. Vianello, *Soil Biology & Biochemistry* **2002**, *34*, 1527-1536.
[46] L. P. Canellas, L. R. L. Teixeira Junior, L. B. Dobbss, C. A. Silva, L. O. Medici, D. B. Zandonadi, A. R. Facanha, *Annals of Applied Biology* **2008**, *153*, 157-166.

[47] L. Canellas, F. Olivares, *Chemical and Biological Technologies in Agriculture C7 - 3* **2014**, *1*, 1-11.
[48] S. Trevisan, O. Francioso, S. Quaggiotti, S. Nardi, *Plant Signaling & Behavior* **2010**, *5*, 635-643.
[49] L. P. Canellas, F. L. Olivares, A. L. Okorokova-Facanha, A. R. Facanha, *Plant Physiology* **2002**, *130*, 1951-1957.
[50] Y. Chen, M. De Nobili, T. Aviad, in *Soil organic matter in sustainable agriculture* (Eds.: F. Magdoff, R. R. Weil), CRC Press, **2004**, p. 103.
[51] T. K. Hartz, T. G. Bottoms, *Hortscience* **2010**, *45*, 906-910.
[52] R. L. Malcolm, P. Maccarthy, *Environmental Science & Technology* **1986**, *20*, 904-911.
[53] N. Pilanali, M. Kaplan, *Journal of Plant Nutrition* **2003**, *26*, 835-843.
[54] J. K. Vessey, *Plant and Soil* **2003**, *255*, 571-586.
[55] L. P. Canellas, D. M. Balmori, L. O. Medici, N. O. Aguiar, E. Campostrini, R. C. C. Rosa, A. R. Facanha, F. L. Olivares, *Plant and Soil* **2013**, *366*, 119-132.
[56] S. S. Bhattacharya, W. Iftikar, B. Sahariah, G. N. Chattopadhyay, *Resources Conservation and Recycling* **2012**, *65*, 100-106.
[57] F. Papathanasiou, I. Papadopoulos, I. Tsakiris, E. Tamoutsidis, *Journal of Food Agriculture & Environment* **2012**, *10*, 677-682.
[58] R. M. Atiyeh, N. Arancon, C. A. Edwards, J. D. Metzger, *Bioresource Technology* **2000**, *75*, 175-180.
[59] R. Singh, R. R. Sharma, S. Kumar, R. K. Gupta, R. T. Patil, *Bioresource Technology* **2008**, *99*, 8507-8511.
[60] R. Joshi, J. Singh, A. P. Vig, *Reviews in Environmental Science and Bio-Technology* **2015**, *14*, 137-159.
[61] M. Befrozfar, D. Habib, A. Asgharzadeh, M. Sadeghi-Shoae, M. Tookalloo, *Annals of Biological Research* **2013**, *24*, 8-12.
[62] N. Q. Arancon, C. A. Edwards, S. Lee, R. Byrne, *European Journal of Soil Biology* **2006**, *42*, S65-S69.
[63] N. Q. Arancon, C. A. Edwards, P. Bierman, C. Welch, J. D. Metzger, *Bioresource Technology* **2004**, *93*, 145-153.
[64] R. M. Atiyeh, S. Lee, C. A. Edwards, N. Q. Arancon, J. D. Metzger, *Bioresource Technology* **2002**, *84*, 7-14.
[65] D. Martinez-Balmori, F. L. Olivares, R. Spaccini, K. P. Aguiar, M. F. Araujo, N. O. Aguiar, F. Guridi, L. P. Canellas, *Journal of Analytical and Applied Pyrolysis* **2013**, *104*, 540-550.
[66] M. M. Kononova, *Soil organic matter: Its nature, its role in soil formation and in soil fertility*, Pergamon Press Ltd., Oxford, UK, **2013**.
[67] A. Piccolo, *Soil Science* **2001**, *166*, 810-832.
[68] K. Ghosh, M. Schnitzer, *Soil Science* **1980**, *129*, 266-276.
[69] L. C. Maillard, *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* **1912**, *155*, 1554-1556.
[70] J. Burdon, *Soil Science* **2001**, *166*, 752-769.

[71] B. P. Kelleher, A. J. Simpson, *Environmental Science & Technology* **2006**, *40*, 4605-4611.
[72] A. J. Simpson, W. L. Kingery, P. G. Hatcher, *Environmental Science & Technology* **2003**, *37*, 337-342.
[73] R. L. Wershaw, *Environmental Science & Technology* **1993**, *27*, 814-816.
[74] A. Piccolo, P. Conte, *Advances in Environmental Research* **2000**, *3*, 508-521.
[75] A. J. Simpson, *Magnetic Resonance in Chemistry* **2002**, *40*, 72-82.
[76] A. C. Stenson, W. M. Landing, A. G. Marshall, W. T. Cooper, *Analytical Chemistry* **2002**, *74*, 4397-4409.
[77] A. C. Stenson, A. G. Marshall, W. T. Cooper, *Analytical Chemistry* **2003**, *75*, 1275-1284.
[78] A. N. Fernandes, M. Giovanela, V. I. Esteves, M. M. de Souza Sierra, *Journal of Molecular Structure* **2010**, *971*, 33-38.
[79] M. Klavins, O. Purmalis, *Environmental Chemistry Letters* **2010**, *8*, 349-354.
[80] M. Tschapek, C. Wasowski, R. M. T. Sanchez, *Plant and Soil* **1981**, *63*, 261-271.
[81] R. R. Engebretson, R. Von Wandruszka, *Environmental Science & Technology* **1999**, *33*, 4299-4303; R. R. Engebretson, R. Vonwandruszka, *Environmental Science & Technology* **1994**, *28*, 1934-1941.
[82] A. J. Simpson, W. L. Kingery, D. R. Shaw, M. Spraul, E. Humpfer, P. Dvortsak, *Environmental Science & Technology* **2001**, *35*, 3321-3325.
[83] A. Piccolo, in *Advances in Agronomy, Vol. 75* (Ed.: D. L. Sparks), **2002**, pp. 57-134.
[84] R. Baigorri, M. Fuentes, G. González-Gaitano, J. M. García-Mina, *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2007**, *302*, 301-306.
[85] J. A. Rice, P. Maccarthy, *Organic Geochemistry* **1991**, *17*, 635-648.
[86] J. D. Ritchie, E. M. Perdue, *Geochimica Et Cosmochimica Acta* **2003**, *67*, 85-96.
[87] H. Herzog, P. Burba, J. Buddrus, *Fresenius Journal of Analytical Chemistry* **1996**, *354*, 375-377.
[88] T. Andjelkovic, J. Perovic, M. Purenovic, S. Blagojevic, R. Nikolic, D. Andjelkovic, A. Bojic, *Analytical Sciences* **2006**, *22*, 1553-1558.
[89] S. L. Khil'ko, A. I. Kovtun, V. I. Rybachenko, *Solid Fuel Chemistry* **2011**, *45*, 337-348.
[90] I. Arsenie, H. Boren, B. Allard, *Science of the Total Environment* **1992**, *116*, 213-220.
[91] P. M. C. Bruneau, S. M. Johnson, *Scotland's peatland - definitions & information resource, Scottish Natural Heritage Commissioned Report, http://www.snh.org.uk/pdfs/publications/commissioned_reports/701.pdf*, **2014**.
[92] K. Marsden, S. Ebmeier, *SPICe Briefing, Peatlands and Climate Change*, Scottish Parliament Information Centre, http://www.scottish.parliament.uk/ResearchBriefingsAndFactsheets/S4/SB_12-28.pdf, **2012**.
[93] *The Scottish Soil Framework*, Scottish Government, http://www.gov.scot/resource/doc/273170/0081576.pdf, **2009**.
[94] *2020 Challenge for Scotland's Biodiversity*, Scottish Government, http://www.gov.scot/Publications/2013/06/5538, **2013**.

[95] C. G. Bain, A. Bonn, R. Stoneman, S. Chapman, A. Coupar, M. Evans, B. Gearey, M. Howat, H. Joosten, C. Keenleyside, J. Labadz, R. Lindsay, N. Littlewood, P. Lunt, C. J. Miller, A. Moxey, H. Orr, M. Reed, P. Smith, V. Swales, D. B. A. Thompson, P. S. Thompson, R. Van de Noort, J. D. Wilson, F. Worrall, *IUCN UK Commission of Inquiry on Peatlands, http://www.iucn-uk-peatlandprogramme.org*, Edinburgh, **2011**.

[96] J. Labadz, T. Allott, M. Evans, D. Butcher, M. Billett, S. Stainer, A. Yallop, P. Jones, M. Innerdale, N. Harmon, K. Maher, R. Bradbury, D. Mount, H. O'Brien, R. Hart, *Peatland Hydrology: Draft Scientific Review to IUCN Peatland Programme Commission of Inquiry on Peatlands, http://www.uplandhydrology.org.uk/wp-content/uploads/2013/12/Review-6-Peatland-Hydrology_0.pdf*, **2010**.

[97] Z. Yu, J. Loisel, D. P. Brosseau, D. W. Beilman, S. J. Hunt, *Geophysical Research Letters* **2010**, *37*.

[98] *Towards an assessment of the state of UK Peatlands, JNCC report*, Joint Nature Conservation Committee, http://jncc.defra.gov.uk/pdf/jncc445_web.pdf, **2011**.

[99] S. J. Chapman, R. E. E. Artz, J. U. Smith, P. Smith, *Summary of the findings of an expert workshop on peatlands*, Expert workshop to establish the current state of knowledge of and future evidence needs for the extent and condition of carbon stocks in Scottish peatlands, Scottish Government, http://www.gov.scot/Publications/2010/02/19145611/0, **2009**.

[100] M. Philben, K. Kaiser, R. Benner, *Journal of Geophysical Research-Biogeosciences* **2014**, *119*, 897-909.

[101] J. Schellekens, R. Bindler, A. Martinez-Cortizas, E. L. McClymont, G. D. Abbott, H. Biester, X. Pontevedra-Pombal, P. Buurman, *Geochimica Et Cosmochimica Acta* **2015**, *150*, 74-89.

[102] E. Gorham, *Ecological Applications* **1991**, *1*, 182-195.

[103] W. C. Oechel, S. J. Hastings, G. Vourlitis, M. Jenkins, G. Riechers, N. Grulke, *Nature* **1993**, *361*, 520-523.

[104] L. R. Belyea, *Carbon Cycling in Northern Peatlands* **2009**, *184*, 5-18.

[105] H. Biester, K. H. Knorr, J. Schellekens, A. Basler, Y. M. Hermanns, *Biogeosciences* **2014**, *11*, 2691-2707.

[106] R. G. Wetzel, *Hydrobiologia* **1992**, *229*, 181-198.

[107] G. P. McLatchey, K. R. Reddy, *Journal of Environmental Quality* **1998**, *27*, 1268-1274.

[108] A. Pind, C. Freeman, M. A. Lock, *Plant and Soil* **1994**, *159*, 227-231.

[109] C. Freeman, N. Ostle, H. Kang, *Nature* **2001**, *409*, 149-149.

[110] C. Amalfitano, R. A. Quezada, M. A. Wilson, J. V. Hanna, *Soil Science* **1995**, *159*, 391-401.

[111] A. Otto, C. Shunthirasingham, M. J. Simpson, *Organic Geochemistry* **2005**, *36*, 425-448; J. S. Clemente, M. J. Simpson, A. J. Simpson, S. F. Yanni, J. K. Whalen, *Geoderma* **2013**, *192*, 86-96; O. Pisani, K. M. Hills, D. Courtier-Murias, A. J. Simpson, N. J. Mellor, E. A. Paul, S. J. Morris, M. J. Simpson, *Organic Geochemistry* **2013**, *62*, 7-16.

[112] R. G. Cates, D. F. Rhoades, *Biochemical Systematics and Ecology* **1977**, *5*, 185-193.

[113] M. Thevenot, M.-F. Dignac, C. Rumpel, *Soil Biology & Biochemistry* **2010**, *42*, 1200-1211.
[114] J. Dai, R. J. Mumper, *Molecules* **2010**, *15*, 7313-7352.
[115] J. S. Clemente, E. G. Gregorich, A. J. Simpson, R. Kumar, D. Courtier-Murias, M. J. Simpson, *Environmental Chemistry* **2012**, *9*, 97-107.
[116] A. Otto, M. J. Simpson, *Biogeochemistry* **2005**, *74*, 377-409.
[117] A. J. Simpson, G. Song, E. Smith, B. Lam, E. H. Novotny, M. H. B. Hayes, *Environmental Science & Technology* **2007**, *41*, 876-883.
[118] G. D. Abbott, E. Y. Swain, A. B. Muhammad, K. Allton, L. R. Belyea, C. G. Laing, G. L. Cowie, *Geochimica Et Cosmochimica Acta* **2013**, *106*, 177-191.
[119] C. Zaccone, D. Said-Pullicino, G. Gigliotti, T. M. Miano, *Organic Geochemistry* **2008**, *39*, 830-838.
[120] B. Haynold, "Sphagnum capillifolium", Licensed under CC BY-SA 3.0 via Commons, https://commons.wikimedia.org/wiki/File:Sphagnum_capillifolium_131007.jpg#/media/File:Sphagnum_capillifolium_131007.jpg, Wikipedia, **2007**.
[121] Aqwis, "Calluna Vulgaris", Licensed under CC BY-SA 3.0 via Commons, https://commons.wikimedia.org/wiki/File:CallunaVulgaris.jpg#/media/File:CallunaVulgaris.jpg, Wikipedia, **2006**.
[122] B. Bendall, "Eriophorum Cotton Grass", Licensed under Attribution via Commons, https://commons.wikimedia.org/wiki/File:Eriophorum_Cotton_Grass.JPG#/media/File:Eriophorum_Cotton_Grass.JPG, Wikipedia, **2013**.
[123] Migas, "Drosera-rotundifolia", Licensed under CC BY-SA 3.0 via Commons, https://commons.wikimedia.org/wiki/File:Drosera-rotundifolia.jpg#/media/File:Drosera-rotundifolia.jpg, Wikipedia, **2005**.
[124] H. Mellegard, T. Stalheim, V. Hormazabal, P. E. Granum, S. P. Hardy, *Letters in Applied Microbiology* **2009**, *49*, 85-90.
[125] S. Rasmussen, C. Wolff, H. Rudolph, *Phytochemistry* **1995**, *38*, 35-39.
[126] T. Hajek, S. Ballance, J. Limpens, M. Zijlstra, J. T. A. Verhoeven, *Biogeochemistry* **2011**, *103*, 45-57.
[127] T. Stalheim, S. Ballance, B. E. Christensen, P. E. Granum, *Journal of Applied Microbiology* **2009**, *106*, 967-976.
[128] M. Tkalec, M. Dobos, M. Babic, E. Jurak, *Acta Physiologiae Plantarum* **2015**, *37*, 1-9.
[129] J. Kovacik, B. Klejdus, K. Repcakova, *Plant Physiology and Biochemistry* **2012**, *52*, 21-27.
[130] J. Schellekens, P. Buurman, I. Fraga, A. Martinez-Cortizas, *Palaeogeography Palaeoclimatology Palaeoecology* **2011**, *299*, 56-69.
[131] M. Monschein, J. I. Neira, O. Kunert, F. Bucar, *Phytochemistry Reviews* **2010**, *9*, 205-215.
[132] M. A. Goni, J. I. Hedges, *Geochimica Et Cosmochimica Acta* **1992**, *56*, 4025-4043.
[133] J. Schellekens, P. Buurman, T. W. Kuyper, *Soil Biology & Biochemistry* **2012**, *53*, 32-42.

[134] T. R. Filley, K. G. J. Nierop, Y. Wang, *Organic Geochemistry* **2006**, *37*, 711-727.
[135] I. Turcotte, S. A. Quideau, *Canadian Journal of Soil Science* **2012**, *92*, 153-164.
[136] K. Lehtonen, M. Ketola, *Organic Geochemistry* **1993**, *20*, 363-380.
[137] M. Maver, E. F. Queiroz, J. L. Wolfender, K. Hostettmann, *Journal of Natural Products* **2005**, *68*, 1094-1098.
[138] I. Kogel-Knabner, *Soil Biology & Biochemistry* **2002**, *34*, 139-162.
[139] T. R. Filley, G. D. Cody, B. Goodell, J. Jellison, C. Noser, A. Ostrofsky, *Organic Geochemistry* **2002**, *33*, 111-124.
[140] K. Ikeya, S. Yamamoto, A. Watanabe, *Organic Geochemistry* **2004**, *35*, 583-594.
[141] T. R. Filley, R. D. Minard, P. G. Hatcher, *Organic Geochemistry* **1999**, *30*, 607-621.
[142] T. Ise, A. L. Dunn, S. C. Wofsy, P. R. Moorcroft, *Nature Geoscience* **2008**, *1*, 763-766.
[143] N. Hertkorn, R. Benner, M. Frommberger, P. Schmitt-Kopplin, M. Witt, K. Kaiser, A. Kettrup, J. I. Hedges, *Geochimica Et Cosmochimica Acta* **2006**, *70*, 2990-3010.
[144] A. J. Simpson, D. J. McNally, M. J. Simpson, *Progress in Nuclear Magnetic Resonance Spectroscopy* **2011**, *58*, 97-175.
[145] B. P. Koch, T. Dittmar, *Rapid Communications in Mass Spectrometry* **2006**, *20*, 926-932.
[146] A. G. Marshall, C. L. Hendrickson, G. S. Jackson, *Mass Spectrometry Reviews* **1998**, *17*, 1-35.
[147] I. J. Amster, *Journal of Mass Spectrometry* **1996**, *31*, 1325-1337.
[148] T. L. Brown, J. A. Rice, *Analytical Chemistry* **2000**, *72*, 384-390.
[149] A. Fievre, T. Solouki, A. G. Marshall, W. T. Cooper, *Energy & Fuels* **1997**, *11*, 554-560.
[150] E. B. Kujawinski, P. G. Hatcher, M. A. Freitas, *Analytical Chemistry* **2002**, *74*, 413-419.
[151] J. P. Elliot, C. S. Idle, F. Tzakoniati, *Anaylsis of Suwannee river fulvic acid by electrospray ionisation fourier transform ion cyclotron resonance mass spectrometry, Research methods exercise undergraduate degree poster, University of Edinburgh*, **2011**.
[152] G. Plancque, B. Amekraz, V. Moulin, P. Toulhoat, C. Moulin, *Rapid Communications in Mass Spectrometry* **2001**, *15*, 827-835; J. A. Leenheer, C. E. Rostad, P. M. Gates, E. T. Furlong, I. Ferrer, *Analytical Chemistry* **2001**, *73*, 1461-1471.
[153] C. A. Hughey, C. L. Hendrickson, R. P. Rodgers, A. G. Marshall, K. N. Qian, *Analytical Chemistry* **2001**, *73*, 4676-4681.
[154] B. P. Koch, M. R. Witt, R. Engbrodt, T. Dittmar, G. Kattner, *Geochimica Et Cosmochimica Acta* **2005**, *69*, 3299-3308.
[155] T. Reemtsma, A. These, A. Springer, M. Linscheid, *Environmental Science & Technology* **2006**, *40*, 5839-5845.
[156] S. Kim, R. W. Kramer, P. G. Hatcher, *Analytical Chemistry* **2003**, *75*, 5336-5344.
[157] T. Grinhut, D. Lansky, A. Gaspar, N. Hertkorn, P. Schmitt-Kopplin, Y. Hadar, Y. Chen, *Rapid Communications in Mass Spectrometry* **2010**, *24*, 2831-2837; E. B. Kujawinski, M. D. Behn, *Analytical Chemistry* **2006**, *78*, 4363-4373.
[158] D. W. Van Krevelen, *Fuel* **1950**, *29*, 269-289.

[159] T. Reemtsma, *Journal of Chromatography A* **2009**, *1216*, 3687-3701.
[160] T. Reemtsma, *Journal of Mass Spectrometry* **2010**, *45*, 382-390.
[161] R. L. Sleighter, P. G. Hatcher, *Marine Chemistry* **2008**, *110*, 140-152.
[162] M. Witt, J. Fuchser, B. P. Koch, *Analytical Chemistry* **2009**, *81*, 2688-2694; E. N. Capley, J. D. Tipton, A. G. Marshall, A. C. Stenson, *Analytical Chemistry* **2010**, *82*, 8194-8202.
[163] N. Hertkorn, M. Frommberger, M. Witt, B. P. Koch, P. Schmitt-Kopplin, E. M. Perdue, *Analytical Chemistry* **2008**, *80*, 8908-8919.
[164] A. Piccolo, M. Spiteller, A. Nebbioso, *Analytical and Bioanalytical Chemistry* **2010**, *397*, 3071-3078; T. Pfeifer, U. Klaus, R. Hoffmann, M. Spiteller, *Journal of Chromatography A* **2001**, *926*, 151-159.
[165] K. Ikeda, R. Arimura, S. Echigo, Y. Shimizu, R. A. Minear, S. Matsui, *Water Science and Technology* **2000**, *42*, 383-390; C. E. Rostad, J. A. Leenheer, *Analytica Chimica Acta* **2004**, *523*, 269-278; R. Flerus, B. P. Koch, P. Schmitt-Kopplin, M. Witt, G. Kattner, *Marine Chemistry* **2011**, *124*, 100-107; C. McIntyre, C. McRae, *Organic Geochemistry* **2005**, *36*, 543-553.
[166] R. L. Sleighter, P. G. Hatcher, *Journal of Mass Spectrometry* **2007**, *42*, 559-574.
[167] C. K. Remucal, R. M. Cory, M. Sander, K. McNeill, *Environmental Science & Technology* **2012**, *46*, 9350-9359.
[168] M. Hardman, A. A. Makarov, *Analytical Chemistry* **2003**, *75*, 1699-1705.
[169] Q. Z. Hu, R. J. Noll, H. Y. Li, A. Makarov, M. Hardman, R. G. Cooks, *Journal of Mass Spectrometry* **2005**, *40*, 430-443.
[170] E. C. Minor, C. J. Steinbring, K. Longnecker, E. B. Kujawinski, *Organic Geochemistry* **2012**, *43*, 1-11; A. M. Kellerman, T. Dittmar, D. N. Kothawala, L. J. Tranvik, *Nature communications* **2014**, *5*; J. A. Leenheer, T. I. Noyes, C. E. Rostad, M. L. Davisson, *Biogeochemistry* **2004**, *69*, 125-141.
[171] M. M. Tfaily, R. Hamdan, J. E. Corbett, J. P. Chanton, P. H. Glaser, W. T. Cooper, *Geochimica Et Cosmochimica Acta* **2013**, *112*, 116-129.
[172] M. M. Tfaily, R. K. Chu, N. Tolic, K. M. Roscioli, C. R. Anderton, L. Pasa-Tolic, E. W. Robinson, N. J. Hess, *Analytical Chemistry* **2015**, *87*, 5206-5215.
[173] R. L. Wershaw, M. A. Mikita, *NMR of humic substances and coal, Vol. 3*, Lewis Publishing, Chelsea, Michigan, USA, **1987**.
[174] D. H. R. Barton, M. Schnitzer, *Nature* **1963**, *198*, 217-218.
[175] R. Ishiwatari, *Geochemical Journal* **1968**, *2*, 175-184.
[176] H. Lentz, H. D. Ludemann, W. Ziechmann, *Geoderma* **1977**, *18*, 325-328.
[177] D. H. Stuermer, J. R. Payne, *Geochimica Et Cosmochimica Acta* **1976**, *40*, 1109-1114.
[178] P. G. Hatcher, R. Rowan, M. A. Mattingly, *Organic Geochemistry* **1980**, *2*, 77-85.
[179] M. A. Wilson, *Journal of Soil Science* **1981**, *32*, 167-186.
[180] G. T. Felbeck, in *Advances in Agronomy, Vol. 17* (Ed.: A. G. Normam), Academic Press, **1965**, pp. 327-368.
[181] R. L. Wershaw, M. A. Mikita, C. Steelink, *Environmental Science & Technology* **1981**, *15*, 1461-1463.
[182] F. J. Vila, H. Lentz, H. D. Ludemann, *Biochemical and biophysical research communications* **1976**, *72*, 1063-1070.

[183] G. Sposito, G. D. Schaumberg, T. G. Perkins, K. M. Holtzclaw, *Environmental Science & Technology* **1978**, *12*, 931-934; H. B. Hayes, W. S. Wilson, R. S. o. C. I. Services, *Humic Substances in Soils, Peats and Waters: Health and Environmental Aspects*, Royal Society of Chemistry, Information Services, Cambridge, UK, **1997**; M. Schnitzer, C. M. Preston, *Plant and Soil* **1983**, *75*, 201-211.
[184] M. A. Wilson, A. J. Jones, B. Williamson, *Nature* **1978**, *276*, 487-489.
[185] M. A. Wilson, K. M. Goh, *Plant and Soil* **1977**, *46*, 287-289.
[186] P. G. Hatcher, I. A. Breger, M. A. Mattingly, *Nature* **1980**, *285*, 560-562.
[187] K. Thorn, D. W. Folan, P. MacCarthy, US Geological Survey Water-Resources Investigation Report, **1989**.
[188] N. Hertkorn, A. Permin, I. Perminova, D. Kovalevskii, M. Yudov, V. Petrosyan, A. Kettrup, *Journal of Environmental Quality* **2002**, *31*, 375-387.
[189] K. A. Thorn, *Science of the Total Environment* **1987**, *62*, 175-183.
[190] C. M. Preston, R. Hempfling, H. R. Schulten, M. Schnitzer, J. A. Trofymow, D. E. Axelson, *Plant Soil* **1994**, *158*, 69-82; A. Piccolo, L. Campanella, B. M. Petronio, *Soil Science Society of America Journal* **1990**, *54*, 750-756.
[191] W. L. Kingery, A. J. Simpson, M. H. B. Hayes, M. A. Locke, R. P. Hicks, *Soil Science* **2000**, *165*, 483-494.
[192] D. Grant, *Nature* **1977**, *270*, 709-710; Ishiwata.R, *Chemical Geology* **1973**, *12*, 113-126.
[193] P. G. Hatcher, G. E. Maciel, L. W. Dennis, *Organic Geochemistry* **1981**, *3*, 43-48.
[194] Y. Y. Chien, W. F. Bleam, *Environmental Science & Technology* **1998**, *32*, 3653-3658.
[195] M. A. Wilson, P. J. Collin, R. L. Malcolm, E. M. Perdue, P. Cresswell, *Organic Geochemistry* **1988**, *12*, 7-12.
[196] A. J. Simpson, J. Burdon, C. L. Graham, M. H. B. Hayes, N. Spencer, W. L. Kingery, *European Journal of Soil Science* **2001**, *52*, 495-509.
[197] R. Newman, K. Tate, P. Barron, M. Wilson, *Journal of Soil Science* **2006**, *31*, 623-631.
[198] C. M. Preston, *Soil Science* **1996**, *161*, 144-166.
[199] C. M. Preston, B. A. Blackwell, *Soil Science* **1985**, *139*, 88-96.
[200] R. L. Cook, D. D. McIntyre, C. H. Langford, H. J. Vogel, *Environmental Science & Technology* **2003**, *37*, 3935-3944.
[201] J. Lambert, P. Burba, J. Buddrus, *Magnetic Resonance in Chemistry* **1992**, *30*, 221-227.
[202] J. Buddrus, P. Burba, H. Herzog, J. Lambert, *Analytical Chemistry* **1989**, *61*, 628-631.
[203] G. I. Ivanova, E. W. Randall, *Central European Journal of Chemistry* **2003**, *1*, 10-27.
[204] B. Lam, A. Baer, M. Alaee, B. Lefebvre, A. Moser, A. Williams, A. J. Simpson, *Environmental Science & Technology* **2007**, *41*, 8240-8247.
[205] J. Lambert, J. Buddrus, *Magnetic Resonance in Chemistry* **1996**, *34*, 276-282.
[206] M. Wilson, P. Collin, K. Tate, *Journal of Soil Science* **1983**, *34*, 297-304.
[207] X. A. Mao, C. H. Ye, *Chemical Physics Letters* **1994**, *227*, 645-650.
[208] A. Bax, *Journal of Magnetic Resonance* **1985**, *65*, 142-145.

[209] D. Neuhaus, I. M. Ismail, C. W. Chung, *Journal of Magnetic Resonance Series A* **1996**, *118*, 256-263.
[210] B. Lam, A. J. Simpson, *Analyst* **2008**, *133*, 263-269.
[211] G. S. H. Lee, M. A. Wilson, B. R. Young, *Organic Geochemistry* **1998**, *28*, 549-559.
[212] M. L. Liu, X. A. Mao, C. H. Ye, H. Huang, J. K. Nicholson, J. C. Lindon, *Journal of Magnetic Resonance* **1998**, *132*, 125-129.
[213] A. J. Simpson, S. A. Brown, *Journal of Magnetic Resonance* **2005**, *175*, 340-346.
[214] R. W. Adams, C. M. Holroyd, J. A. Aguilar, M. Nilsson, G. A. Morris, *Chemical Communications* **2013**, *49*, 358-360.
[215] P. J. Hore, J. A. Jones, S. Wimperis, *Nmr: The Toolkit*, Oxford University Press Inc., New York, USA, **2000**.
[216] R. Freeman, *Spin choreography: basic steps in high resolution NMR*, Spektrum, **1997**.
[217] A. J. Simpson, *Soil Science* **2001**, *166*, 795-809.
[218] P. Schmitt-Kopplin, N. Hertkorn, H. R. Schulten, A. Kettrup, *Environmental Science & Technology* **1998**, *32*, 2531-2541.
[219] S. Haiber, P. Burba, H. Herzog, J. Lambert, *Fresenius Journal of Analytical Chemistry* **1999**, *364*, 215-218.
[220] G. C. Woods, M. J. Simpson, A. J. Simpson, *Water Research* **2012**, *46*, 3398-3408.
[221] E. Kupce, P. Schmidt, M. Rance, G. Wagner, *Journal of Magnetic Resonance* **1998**, *135*, 361-367.
[222] A. J. Simpson, W. L. Kingery, M. H. B. Hayes, M. Spraul, E. Humpfer, P. Dvortsak, R. Kerssebaum, M. Godejohann, M. Hofmann, *Naturwissenschaften* **2002**, *89*, 84-88.
[223] T. W. M. Fan, R. M. Higashi, A. N. Lane, *Environmental Science & Technology* **2000**, *34*, 1636-1646.
[224] E. Kaiser, A. J. Simpson, K. J. Dria, B. Sulzberger, P. G. Hatcher, *Environmental Science & Technology* **2003**, *37*, 2929-2935.
[225] A. P. Deshmukh, C. Pacheco, M. B. Hay, S. C. B. Myneni, *Geochimica Et Cosmochimica Acta* **2007**, *71*, 3533-3544.
[226] R. L. Cook, *Analytical and Bioanalytical Chemistry* **2004**, *378*, 1484-1503.
[227] A. J. Simpson, W. L. Kingery, M. Spraul, E. Humpfer, P. Dvortsak, R. Kerssebaum, *Environmental Science & Technology* **2001**, *35*, 4421-4425.
[228] A. M. Dixon, C. K. Larive, *Analytical Chemistry* **1997**, *69*, 2122-2128.
[229] K. F. Morris, B. J. Cutak, A. M. Dixon, C. K. Larive, *Analytical Chemistry* **1999**, *71*, 5315-5321.
[230] D. Smejkalova, A. Piccolo, *Environmental Science & Technology* **2008**, *42*, 699-706.
[231] R. Freeman, G. A. Morris, *Journal of the Chemical Society-Chemical Communications* **1978**, 684-686; A. Bax, G. A. Morris, *Journal of Magnetic Resonance* **1981**, *42*, 501-505; A. Bax, R. H. Griffey, B. L. Hawkins, *Journal of Magnetic Resonance* **1983**, *55*, 301-315.
[232] L. Muller, *Journal of the American Chemical Society* **1979**, *101*, 4481-4484.
[233] G. Bodenhausen, D. J. Ruben, *Chemical Physics Letters* **1980**, *69*, 185-189.
[234] A. Bax, *Journal of Magnetic Resonance* **1983**, *53*, 517-520.
[235] K. Zangger, *Progress in Nuclear Magnetic Resonance Spectroscopy* **2015**, *86-87*, 1-20.

[236] S. Haiber, H. Herzog, P. Burba, B. Gosciniak, J. Lambert, *Environmental Science & Technology* **2001**, *35*, 4289-4294.
[237] A. J. Simpson, M. J. Simpson, *Nuclear magnetic resonance analysis of natural organic matter in Biophysico-Chemical Processes Involving Natural Nonliving Organic Matter in Environmental Systems*, Wiley-Blackwell, **2009**.
[238] A. J. Simpson, B. Lefebvre, A. Moser, A. Williams, N. Larin, M. Kvasha, W. L. Kingery, B. Kelleher, *Magnetic Resonance in Chemistry* **2004**, *42*, 14-22.
[239] G. C. Woods, M. J. Simpson, P. J. Koerner, A. Napoli, A. J. Simpson, *Environmental Science & Technology* **2011**, *45*, 3880-3886; B. G. Pautler, P. T. Sanborn, A. J. Simpson, M. J. Simpson, *Organic Geochemistry* **2013**, *63*, 122-138.
[240] E. M. Perdue, N. Hertkorn, A. Kettrup, *Analytical Chemistry* **2007**, *79*, 1010-1021.
[241] O. Francioso, S. Sanchez-Cortes, V. Tugnoli, C. Ciavatta, C. Gessa, *Applied Spectroscopy* **1998**, *52*, 270-277.
[242] P. Dais, A. Spyros, *Magnetic Resonance in Chemistry* **2007**, *45*, 367-377.
[243] K. A. Thorn, J. B. Arterburn, M. A. Mikita, *Environmental Science & Technology* **1992**, *26*, 107-116; K. A. Thorn, L. G. Cox, *Organic Geochemistry* **2009**, *40*, 484-499.
[244] T. Liang, C. N. Neumann, T. Ritter, *Angewandte Chemie-International Edition* **2013**, *52*, 8214-8264.
[245] M. O. J. Nieminen, E. Pulkkinen, E. Rahkamaa, *Holzforschung* **1989**, *43*, 303-307.
[246] K. A. Thorn, D. W. Folan, J. B. Arterburn, M. A. Mikita, P. Maccarthy, *Science of the Total Environment* **1989**, *81-2*, 209-218.
[247] J. Schraml, J. Past, J. Puskar, T. Pehk, E. Lippmaa, R. Brezny, *Collection of Czechoslovak Chemical Communications* **1987**, *52*, 1985-1991.
[248] K. A. Thorn, C. Steelink, R. L. Wershaw, *Organic Geochemistry* **1987**, *11*, 123-137.
[249] J. M. Challinor, *Journal of analytical and applied pyrolysis* **1991**, *20*, 15-24.
[250] C. Saizjimenez, B. Hermosin, J. J. Ortega-Calvo, *Water Research* **1993**, *27*, 1693-1696.
[251] C. A. Joll, T. Huynh, A. Heitz, *Journal of Analytical and Applied Pyrolysis* **2003**, *70*, 151-167.
[252] A. Piccolo, P. Conte, A. F. Patti, *Analytical and Bioanalytical Chemistry* **2006**, *384*, 994-1001.
[253] Schnitzer.M, *Soil Science* **1974**, *117*, 94-102.
[254] M. A. Mikita, C. Steelink, R. L. Wershaw, *Analytical Chemistry* **1981**, *53*, 1715-1717.
[255] F. J. Gonzalezvila, H. D. Ludemann, F. Martin, *Geoderma* **1983**, *31*, 3-15; S. U. Khan, Schnitzer.M, *Canadian Journal of Chemistry* **1971**, *49*, 2302-2309.
[256] M. Schnitzer, S. U. Khan, *Humic Substances in the Environment*, Books on Demand, Atlanta, Georgia, USA, **1972**.
[257] R. L. Wershaw, D. J. Pinckney, *Science* **1978**, *199*, 906-907.
[258] P. Kuran, P. Janos, L. Madronova, J. Novak, J. Kozler, *Talanta* **2008**, *76*, 960-963.
[259] G. C. Briggs, G. J. Lawson, *Fuel* **1970**, *49*, 39-48.
[260] R. Liotta, *Fuel* **1979**, *58*, 724-728; R. Liotta, K. Rose, E. Hippo, *Journal of Organic Chemistry* **1981**, *46*, 277-283.
[261] M. Rabinovitz, Y. Cohen, M. Halpern, *Angewandte Chemie-International Edition in English* **1986**, *25*, 960-970.

[262] L. M. Clemow, G. Favas, W. R. Jackson, M. Marshall, A. F. Patti, P. J. Redlich, *Fuel* **1999**, *78*, 567-572.
[263] G. Lamoureux, C. Agueero, *Arkivoc* **2009**, 251-264.
[264] S. Sharma, S. Ameta, V. Sharma, in *Proceedings of the World Congress on Engineering and Computer Science, Vol. 2*, **2010**.
[265] S. Ouk, S. Thiebaud, E. Borredon, P. Le Gars, *Applied Catalysis a-General* **2003**, *241*, 227-233.
[266] R. Bernini, F. Crisante, M. C. Ginnasi, *Molecules* **2011**, *16*, 1418-1425.
[267] W. C. Shieh, S. Dell, O. Repic, *Journal of Organic Chemistry* **2002**, *67*, 2188-2191.
[268] W. C. Shieh, S. Dell, O. Repic, *Organic Letters* **2001**, *3*, 4279-4281.
[269] S. Ouk, S. Thiebaud, E. Borredon, B. Chabaud, *Synthetic Communications* **2005**, *35*, 3021-3026.
[270] S. Memoli, M. Selva, P. Tundo, *Chemosphere* **2001**, *43*, 115-121.
[271] T. C. Farrar, J. E. Harriman, *Density Matrix Theory and Its Applications in Nmr Spectroscopy*, Farragut Press, Madison, Wisconsin, USA, **1991**.
[272] O. W. Sorensen, G. W. Eich, M. H. Levitt, G. Bodenhausen, R. R. Ernst, *Progress in Nuclear Magnetic Resonance Spectroscopy* **1983**, *16*, 163-192.
[273] J. Keeler, *Lecture notes, Chapter 6: Product Spin Operators, www-keeler.ch.cam.ac.uk/lectures/understanding/chapter_6.pdf*, **2004**.
[274] T. D. W. Claridge, *High-resolution NMR Techniques in Organic Chemistry*, Elsevier, Oxford, UK, **2009**.
[275] G. A. Morris, R. Freeman, *Journal of the American Chemical Society* **1979**, *101*, 760-762.
[276] J. Keeler, *Understanding NMR spectroscopy*, John Wiley and Sons, Chichester, UK, **2011**.
[277] K. Stott, J. Stonehouse, J. Keeler, T. L. Hwang, A. J. Shaka, *Journal of the American Chemical Society* **1995**, *117*, 4199-4200.
[278] E. M. Thurman, R. L. Malcolm, *Environmental Science & Technology* **1981**, *15*, 463-466.
[279] A. B. MacKenzie, A. N. Whitton, T. M. Shimmield, C. M. Houston, R. D. Scott, P. T. Hooker, *Natural decay series at the Needles's Eye natural analogue site, II, 1989-1991*, British Geological Survey, **1991**.
[280] X. Xu, *PhD thesis: Uranium associations and migration behaviour at the Needle's Eye natural analogue site in SW Scotland, Vol. PhD*, University of Edinburgh, **2013**.
[281] P. Jamet, P. J. Hooker, J. M. Schmitt, E. Ledoux, P. E. D. Orres, *Mineralium Deposita* **1993**, *28*, 66-76.
[282] K. H. Tan, J. C. Lobartini, D. S. Himmelsbach, L. E. Asmussen, *Communications in Soil Science and Plant Analysis* **1991**, *22*, 861-877.
[283] J. G. Farmer, P. Anderson, J. M. Cloy, M. C. Graham, A. B. MacKenzie, G. T. Cook, *Science of the Total Environment* **2009**, *407*, 5578-5588.
[284] T. Duffy, *Warning of threat to wildlife as lothians peat bogs dry up*, Edinburgh Evening News, http://www.edinburghnews.scotsman.com/news/warning-of-threat-to-wildlife-as-lothian-s-peat-bogs-dry-up-1-2718670, **2013**.

[285] *Red Moss of Balerno - Peatland Alliance Project*, IUCN UK, http://www.iucn-uk-peatlandprogramme.org/peatland-gateway/uk/peatland-alliance/projects/red-moss-balerno, **2015**.
[286] J. Cloy, *PhD thesis: Chemical records of environmental pollution in ombrotrophic peat bogs, Vol. PhD*, University of Edinburgh, **2006**.
[287] C. J. Hardacre, M. R. Heal, *Journal of Geophysical Research-Atmospheres* **2013**, *118*, 977-991; F. Everingham, I. Tanner, *Raised Bog Inventory of Red Moss Balerno*, Scottish National Heritage, http://www.snh.org.uk/Peatlands/RMB002.htm, **1994**.
[288] S. Khan, W. Yaoguo, Z. Xiaoyan, X. Youning, Z. Jianghua, H. Sihai, *International Journal of Environmental Pollution and Remediatio* **2014**, *2*, 18-23.
[289] H. Z. Ma, H. E. Allen, Y. J. Yin, *Water Research* **2001**, *35*, 985-996; J. Peuravuori, T. Lehtonen, K. Pihlaja, *Analytica Chimica Acta* **2002**, *471*, 219-226.
[290] J. J. Farnworth, *Superlite DAX-8: A replacement for Amberlite XAD-8 resin for the isolation of Humic Substances in IHSS Newsletter, http://www.humicsubstances.org/documents/Old%20IHSS%20Newsletters/NL_12_1994_edited.pdf*, IHSS, **1994**.
[291] S. Hiradate, T. Yonezawa, H. Takesako, *Geoderma* **2006**, *132*, 196-205.
[292] M. Thurman, *Organic Geochemistry of Natural Waters in Developments in Biogeochemistry, Vol. 2*, M. Nijhoff and Dr W. Junk Publishers, Dordrecht, Netherlands, **1985**.
[293] K. Hanninen, J. Knuutinen, P. Mannila, *Chemosphere* **1993**, *27*, 747-755.
[294] L. Schwartz, *Introduction to Tangential Flow Filtration for laboratory and process development applications*, Pall Corporation, http://www.pall.com/main/laboratory/literature-library-details.page?id=34212, **2015**.
[295] G. Ricca, F. Severini, G. Di Silvestro, C. M. Yuan, F. Adani, *Geoderma* **2000**, *98*, 115-125.
[296] J. Weigelt, G. Otting, *Journal of Magnetic Resonance Series A* **1995**, *113*, 128-130.
[297] J. Chung, J. R. Tolman, K. P. Howard, J. H. Prestegard, *Journal of Magnetic Resonance Series B* **1993**, *102*, 137-147; T. Saito, R. L. Rinaldi, *Journal of Magnetic Resonance Series A* **1996**, *118*, 136-139.
[298] A. Meissner, D. Moskau, N. C. Nielsen, O. W. Sorensen, *Journal of Magnetic Resonance* **1997**, *124*, 245-249.
[299] G. Otting, K. Wuthrich, *Quarterly Reviews of Biophysics* **1990**, *23*, 39-96.
[300] E. Kupce, R. Freeman, *Journal of Magnetic Resonance Series A* **1995**, *112*, 261-264.
[301] K. Stott, J. Keeler, Q. N. Van, A. J. Shaka, *Journal of Magnetic Resonance* **1997**, *125*, 302-324.
[302] M. J. Thrippleton, J. Keeler, *Angewandte Chemie-International Edition* **2003**, *42*, 3938-3941.
[303] *Spectral Database for Organic Compounds, SDBSWeb*, National Institute of Advanced Industrial Science and Technology, http://sdbs.db.aist.go.jp, **2015**.
[304] T. Liptaj, *Tables of ^{1}H and ^{13}C NMR parameters*, Slovak Technical University, Bratislava, Slovakia, **1981**.

[305] P. L. Corio, B. P. Dailey, *Journal of the American Chemical Society* **1956**, *78*, 3043-3048.
[306] J. Ralph, S. Ralph, L. Landucci, *NMR Database of Lignin and Cell Wall Model Compounds*, United States Forest Product Laboratory, https://www.glbrc.org/databases_and_software/nmrdatabase/NMR_DataBase_2009_Complete.pdf, **2004**.
[307] P. Leinweber, H. R. Schulten, *Journal of Analytical and Applied Pyrolysis* **1999**, *49*, 359-383.
[308] H. R. Schulten, *Journal of Analytical and Applied Pyrolysis* **1987**, *12*, 149-186.
[309] D. J. Clifford, D. M. Carson, D. E. McKinney, J. M. Bortiatynski, P. G. Hatcher, *Organic Geochemistry* **1995**, *23*, 169-175.
[310] A. Srivastava, L. J. M. Rao, T. Shivanandappa, *Food Chemistry* **2012**, *132*, 1959-1965.
[311] M. A. F. Jalal, D. J. Read, E. Haslam, *Phytochemistry* **1982**, *21*, 1397-1401.
[312] C. Dai, P. Liu, C. Liu, B. Wang, R.-y. Chen, *China journal of Chinese materia medica* **2006**, *31*, 1080-1082; S. Kakkar, S. Bais, *ISRN pharmacology* **2014**, *2014*, 952943-952943; R. J. Horvat, S. D. Senter, *Journal of Agricultural and Food Chemistry* **1980**, *28*, 1292-1295.
[313] S. Khadem, R. J. Marles, *Molecules* **2010**, *15*, 7985-8005.

Appendices

Appendices referred to throughout the thesis are located on the Edinburgh Research Archive at http://hdl.handle.net/1842/14212.

The following are included:

Appendix A. NMR pulse programs.

Appendix B. Chemical shifts obtained for model mixture II.

Appendix C. Detailed discussion on the structural determination of RMFA molecules based on the analysis of ^{13}C-filtered nD NMR experiments.

Publications

The following publications resulted from this work:

1. N.G.A. Bell, L. Murray, M.C. Graham, D. Uhrín, NMR methodology for complex mixture 'separation', *Chem. Commun.* 50, 1694-1697, **2014**.

2. N.G.A. Bell, A.A.L. Michalchuk, J.W.T. Blackburn, M.C. Graham, D. Uhrín, Isotope-Filtered 4D NMR Spectroscopy for Structure Determination of Humic Substances, *Angew. Chem. Int.* 29, 8382-8385, **2015**.

These publications are included below.

ChemComm

COMMUNICATION

NMR methodology for complex mixture 'separation'†

Cite this: *Chem. Commun.*, 2014, **50**, 1694

Received 21st November 2013,
Accepted 24th December 2013

DOI: 10.1039/c3cc48907h

www.rsc.org/chemcomm

Nicholle G. A. Bell,[a] Lorna Murray,[a] Margaret C. Graham[b] and Dušan Uhrín*[a]

Mixture 'separation' by NMR is demonstrated through the development of a pseudo 4D NMR experiment, 3D IPAP INEPT-INADEQUATE-HSQC, designed for the structural elucidation of ^{13}C tagged compounds.

The structural elucidation of compounds contained within complex mixtures is a challenging task. Despite the advances in chromatography, some mixtures cannot be separated. Humic substances (HS), produced by the biodegradation of plant and animal residues,[1] are the best known example of an 'inseparable' mixture consisting of thousands of organic compounds. Ubiquitous in nature, they make up a considerable proportion of the Earth's carbon pool and are key players in many biogeochemical processes.

In order to comprehend the functional roles of HS on a molecular level, their structural composition needs to be deciphered. So far HS have been characterized only on the level of compound classes and presence of individual functional groups.[2–4] NMR spectroscopy and mass spectrometry are widely regarded as the two most promising analytical techniques for revealing the structure of individual HS molecules.[5–7] Nevertheless, HS pose a considerable analytical challenge to both techniques, *e.g.*, the signal overlap in NMR spectra of complex mixtures prevents separation of resonances belonging to individual molecules and hence their identification using standard NMR techniques. Increasing the dimensionality of NMR experiments, selective excitation, or DOSY spectroscopy[8,9] are the three most common approaches attempted to circumvent these problems.

Unfortunately, these quickly fail when the complexity of mixtures increases. Inevitably, for chromatographically inseparable mixtures the "separation" must be done spectroscopically. One way of achieving this is by tagging the molecules with isotopically labelled nuclei. Once in place, the polarization transfer pathways are directed through these tags, reducing the complexity of spectra significantly. As illustrated here, this approach has a potential to elucidate molecular fragments of compounds contained in complex mixtures.

The bulk of HS is composed of oxygen rich $H_xC_yO_z$ compounds with the M_w range of $\sim$200–1000 g mol^{-1}.[10] Particularly prevalent functionalities decorating aromatic and aliphatic skeletons of HS are the hydroxyl and carboxyl groups. These groups are crucial to the interaction of HS with species such as heavy metals as well as the self-association of HS molecules. The methodology presented here aims at characterizing the aromatic moieties of HS carrying OH and COOH groups and relies on introducing ^{13}C-enriched –O^{13}CH$_3$ and –COO^{13}CH$_3$ groups into HS compounds. HS have been methylated in the past and the inspection of –O^{13}CH$_3$ resonances yielded some rudimentary information about the nature of their COOH and OH groups.[11] The novelty of our approach is that it uses labels to spy on their neighbourhood, obtaining the ^{1}H and ^{13}C chemical shifts and ^{1}H–^{1}H and ^{1}H–^{13}C coupling constants of the nuclei in their vicinity.

To achieve this aim we have designed a novel 3D NMR experiments, referred to here as 3D IPAP INEPT-INADEQUATE-HSQC (Fig. 1).

This NMR experiment can be viewed as a 3D extension of a 2D INEPT-INADEQUATE.[12] It is also related to 3D HCCH experiments.[13,14] 3D IPAP INEPT-INADEQUATE-HSQC exploits polarization transfer pathways shown in Fig. 1, and correlates the double-quantum (DQ) coherencies of long-range coupled carbons (F_1) with corresponding single-quantum ^{13}C chemical shifts (F_2) and the ^{1}H chemical shifts (F_3). The polarizations transfer is tuned for $^{2,3}J_{CC}$ (Fig. S1, ESI†) and correlates the chemical shifts of nuclei in $^{13}CH_3\cdots{}^{13}CH_y$ (y = 0, 1) fragments. It starts and ends concurrently on methoxy and aromatic protons located next to the methoxy groups. For $CH_3\cdots CH$ fragments it provides chemical shifts of all four nuclei and can therefore be regarded as a pseudo 4D experiment; all three chemical shifts are obtained for the $CH_3\cdots C_q$ moieties. Since the acquired NMR signal is filtered *via* isotopically enriched $^{13}CH_3$ groups, the resulting spectra are significantly simplified.

The limited ^{1}H and ^{13}C chemical shifts range of methoxy groups (Fig. S2, ESI†) necessitates the use of high digital resolution in the

[a] *EastChem, School of Chemistry, University of Edinburgh, King's Buildings, West Mains Rd, Edinburgh, Scotland, EH9 3JJ, UK. E-mail: dusan.uhrin@ed.ac.uk*
[b] *School of Geosciences, University of Edinburgh, King's Buildings, West Mains Rd, Edinburgh, Scotland, EH9 3JN, UK*

† Electronic supplementary information (ESI) available: Materials and methods, details of the optimization of NMR parameters, values of coupling constants, S/N ratios, examples of processing of IPAP spectra and a suppression of signals from aliphatic OMe groups. See DOI: 10.1039/c3cc48907h

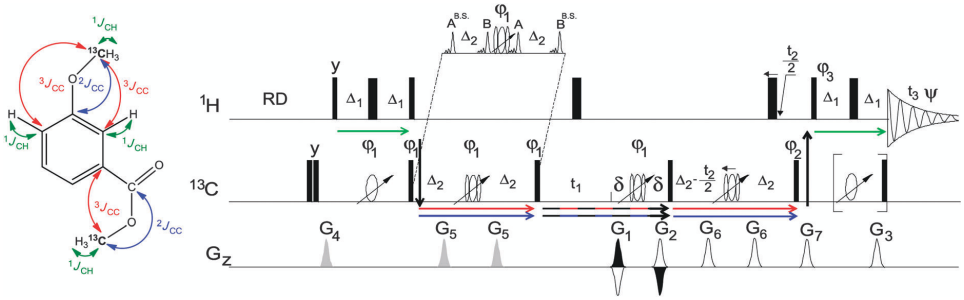


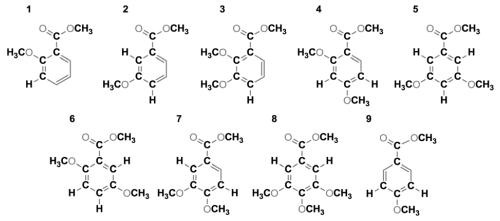
Fig. 1 The pulse sequence of the 3D IPAP INEPT-INADEQUATE-HSQC and its polarisation transfer pathways utilising various coupling constants. The details of the pulse sequence are given in the ESI.†

F_2 and F_3 dimensions of the 3D spectra. In order to achieve this in a realistic time, we have (i) optimized ^{1}H and ^{13}C carrier frequencies and spectral widths in all dimensions, (ii) frequency shifted DQ coherences during t_1, (iii) aliased signals in the F_2 dimension, (iv) used non-uniform sampling and (v) acquired pure in phase and antiphase (IPAP)[15] ^{13}C-coupled multiplets during t_3, as detailed in the ESI,† Fig. S3–S5.

The proposed method is illustrated on a model mixture of nine benzoic acid derivatives, **1a–9a**, which were fully methylated using $^{13}CH_3I$, yielding compounds **1–9** (Scheme 1, average concentration of 1.4 mM).

2D F_2F_3 (HSQC) (Fig. 2a and b) and 2D F_1F_3 (DQ) (Fig. 2c and d) planes of the 3D IPAP INEPT-INADEQUATE-HSQC spectrum of the model mixture illustrate the resolving power of this experiment. The spectrum was analysed as outlined in Fig. 2e–g. Shown here are the 2D DQ (F_1F_3) planes extracted at the ^{13}C shifts of the aromatic and Me carbons (F_2) indicated by red arrows in Fig. 2a and b. The obtained ^{13}C and ^{1}H chemical shifts allowed unambiguous identification of the fragment highlighted in the inset. The fragment was assigned to the correct molecule by considering the electronic effects of OMe and COOMe groups on the ^{13}C and ^{1}H chemical shifts of benzene and by analysing the proton–proton and long-range proton–carbon coupling constants determined in F_3. The inspection of the complete 3D spectrum lead to the identification of the fragments highlighted in Scheme 1 and their assignment to individual molecules.

When applied to HS, the signals from methylated aliphatic hydroxyl and carboxyl groups can potentially interfere with the interpretation of the 3D IPAP INEPT-INADEQUATE-HSQC spectra. These can be eliminated by band-selective inversion of spins resonating within the 65–95 and 15–45 ppm regions of the ^{13}C spectra applied during the initial carbon spin-echo, $2\Delta_2$, (see the inset of Fig. 1). Spins resonating in these regions effectively receive a 360° pulse, thus refocusing their $^{2,3}J_{CC}$ couplings with the $^{13}CH_3O$ carbons and eliminating their signals (see Fig. S6, ESI†).



Scheme 1 Nine methylated compounds. The molecular fragments identified by the analysis of the 3D IPAP INEPT-INADEQUATE-HSQC spectrum based on the chemical shift and coupling constants are highlighted in bold.

How sensitive is the 3D IPAP INEPT-INADEQUATE-HSQC experiment? Because the DQ coherences are created between the fully labelled $^{13}CH_3$ carbons and natural abundance ^{13}C spins, the theoretical sensitivity of this experiment is comparable to that of a refocused gradient-selected ^{1}H, ^{13}C HMBC optimized for $^nJ_{CH}$ couplings. We have determined the S/N ratios in 1D F_3 traces extracted from F_1F_3 planes of the 3D spectrum obtained by the addition of the 3D IP and AP INEPT-INADEQUATE-HSQC spectra. The values were normalized (Fig. S7, ESI†) to the average concentration of the methylated compounds (1.4 mM). Due to their singlet character, the S/N is higher for $-O^{13}CH_3$ (60–1900 : 1) than for the aromatic resonances (57–600 : 1). Somewhat lower values were obtained for $-COO^{13}CH_3$ signals (40–240 : 1). These values clearly reflect the sizes of $^{2,3}J_{CC}$ couplings (Fig. S1, ESI†) involved in the polarisation transfer and can be increased for small coupling constants by optimizing the Δ_2 intervals for smaller J_{CC} couplings (6 Hz at present). This will have to be balanced against faster relaxation in larger molecules. We note that the relaxation effects can be reduced by using $^{13}CD_2H$ in place of $^{13}CH_3$ groups. This modification opens a route toward a design of more efficient 3D INADEQUATE-based experiments thus compensating partially for the loss of protons in $^3CD_2H$ groups.

Considering the lowest S/N observed (40 : 1), a 10-fold reduction of the concentration would still yield spectra of adequate quality. Thus assuming a 140 μM concentration of a 500 g mol^{-1} M_w compound, 38.5 μg dissolved in 550 μl of $CDCl_3$ are required per compound. A 19.3 mg strong mixture of compounds containing 500 unique substitution patterns by –OH and –COOH groups on aromatic rings will therefore provide 3D spectra with sufficient S/N ratios to be analyzed. As the methylation procedure includes extraction into the organic phase, and thus selects a subset of HS compounds, simplification of these complex mixtures is inherent to this process. This could be further improved by employing partial fractionation of HS prior to methylation.

Obtaining ^{1}H and ^{13}C chemical shifts of the nuclei in the vicinity OH and COOH by itself does not lead to a structure, and

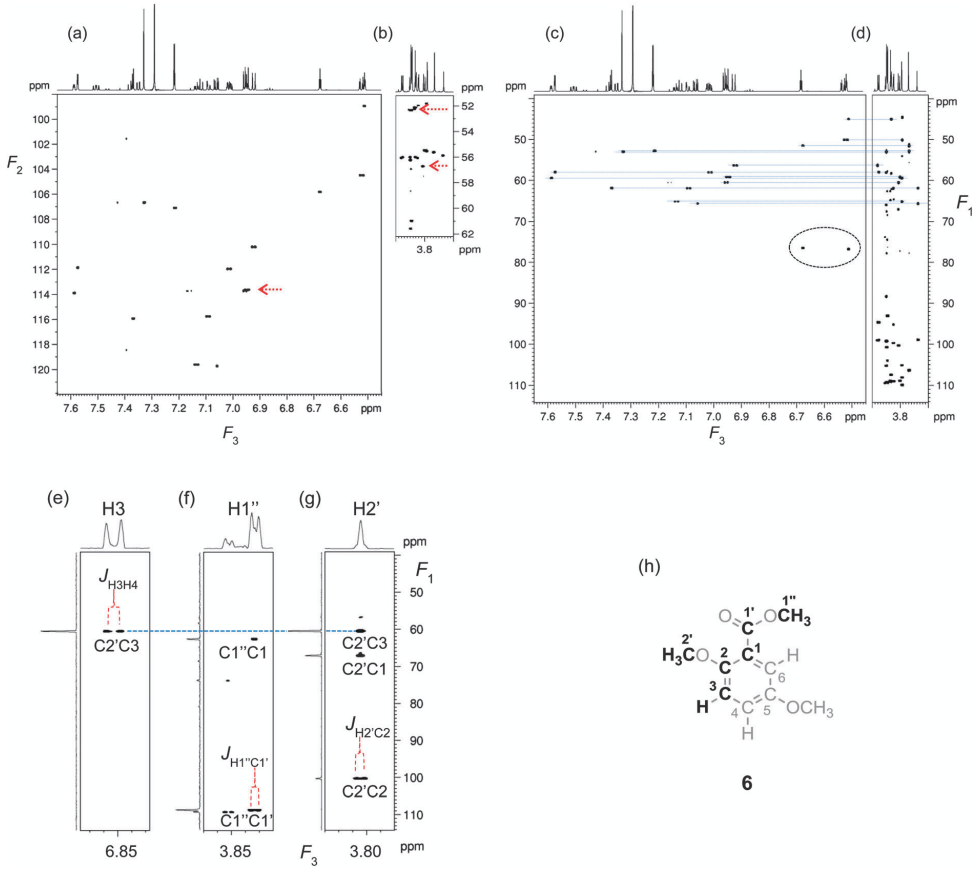


Fig. 2 2D projections of the 3D IPAP INEPT-INADEQUATE-HSQC spectrum obtained by the addition of the inphase and anti-phase 3D spectra. (a, b) F_2F_3 HSQC and (c, d) F_1F_3 DQ projections of the aromatic (a, c) and methyl (b, d) regions of the spectrum. The circled signals belong to the aliased DQ coherences of OMe carbons in compounds **4** and **5**. See Fig. S4 (ESI†) for the explanation of their origin. Blue lines connect DQ coherences of $CH_3\cdots CH$ fragments detected simultaneously on the aromatic and methyl protons. The red arrow indicates the F_2 coordinates of planes shown in (e–g). 2D DQ (F_1F_3) planes extracted at the CH and OMe ^{13}C chemical shifts indicated in (a and b) showing (e) aromatic, (f) ester and (g) methoxy proton regions of the spectrum. The blue line indicates a shared DQ frequency. Observed splittings are annotated. (h) The identified molecular fragment.

such data needs to be interpreted by considering the precursors of HS molecules and their chemical shifts. However, it is clear that this methodology supersedes information content of simple ^{1}H, ^{13}C HSQC spectra of HS.[16] The key is the increased number of nuclei that are identified as belonging to the same molecule. Additional valuable information can be obtained by hydrolysing the $-COO^{13}CH_3$ esters, while keeping the $-O^{13}CH_3$ groups untouched. The resulting site-specific changes of chemical shifts of aromatic spins can be quantified through recording the 3D IPAP INEPT-INADEQUATE-HSQC of a partially hydrolysed mixture (data not shown). Such information is very valuable for the identification of fragments carrying both OH and COOH groups.

In conclusion, we conjecture that tagging by ^{13}C labelled methyl groups in combination with 3D NMR spectroscopy is a very promising approach to the analysis of complex mixtures such as HS. To the best of our knowledge, this work is the first example of a sophisticated NMR experiment designed to explore the chemical environment around the tags, rather than just the signals from the tags themselves. This approach is not limited to methylation, other tags, containing NMR active, fully abundant nuclei such as ^{15}N or ^{19}F can also be explored. Finally, many other mixtures can benefit from this approach, including those found in food, beverages, natural products, or biological samples, to name a few.

This work was supported by the NERC grant NE/L00044X/1 to MCG and DU. NGAB would like to acknowledge the support of the University of Edinburgh Principal's Career Development Scholarship and NERC. We thank J. Bella for maintaining the NMR spectrometers.

Notes and references

1 F. J. Stevenson, *Humus Chemistry: Genesis, Composition, Reactions*, John Wiley & Sons, Ltd, New York, 2nd edn, 1994.
2 B. Lam, A. Baer, M. Alaee, B. Lefebvre, A. Moser, A. Williams and A. J. Simpson, *Environ. Sci. Technol.*, 2007, **41**, 8240–8247.
3 N. Hertkorn, R. Benner, M. Frommberger, P. Schmitt-Kopplin, M. Witt, K. Kaiser, A. Kettrup and J. I. Hedges, *Geochim. Cosmochim. Acta*, 2006, **70**, 2990–3010.
4 R. Benner, J. D. Pakulski, M. McCarthy, J. I. Hedges and P. G. Hatcher, *Science*, 1992, **255**, 1561–1564.
5 L. A. Cardoza, A. K. Korir, W. H. Otto, C. J. Wurrey and C. K. Larive, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2004, **45**, 209–238.
6 N. Hertkorn, C. Ruecker, M. Meringer, R. Gugisch, M. Frommberger, E. M. Perdue, M. Witt and P. Schmitt-Kopplin, *Anal. Bioanal. Chem.*, 2007, **389**, 1311–1327.
7 A. J. Simpson, D. J. McNally and M. J. Simpson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2011, **58**, 97–175.
8 A. J. Simpson, W. L. Kingery and P. G. Hatcher, *Environ. Sci. Technol.*, 2003, **37**, 337–342.
9 R. Novoa-Carballal, E. Fernandez-Megia, C. Jimenez and R. Riguera, *Nat. Prod. Rep.*, 2011, **28**, 78–98.
10 A. C. Stenson, W. M. Landing, A. G. Marshall and W. T. Cooper, *Anal. Chem.*, 2002, **74**, 4397–4409.
11 M. A. Mikita, C. Steelink and R. L. Wershaw, *Anal. Chem.*, 1981, **53**, 1715–1717.
12 J. Weigelt and G. Otting, *J. Magn. Reson., Ser. A*, 1995, **113**, 128–130.
13 J. Chung, J. R. Tolman, K. P. Howard and J. H. Prestegard, *J. Magn. Reson., Ser. B*, 1993, **102**, 137–147.
14 T. Saito and R. L. Rinaldi, *J. Magn. Reson., Ser. A*, 1996, **118**, 136–139.
15 L. S. Yao, J. F. Ying and A. Bax, *J. Biomol. NMR*, 2009, **43**, 161–170.
16 E. M. Perdue, N. Hertkorn and A. Kettrup, *Anal. Chem.*, 2007, **79**, 1010–1021.

Angewandte
Communications

NMR Spectroscopy

International Edition: DOI: 10.1002/anie.201503321
German Edition: DOI: 10.1002/ange.201503321

Isotope-Filtered 4D NMR Spectroscopy for Structure Determination of Humic Substances**

*Nicholle G. A. Bell, Adam A. L. Michalchuk, John W. T. Blackburn, Margaret C. Graham, and Dušan Uhrín**

Abstract: *Humic substances, the main component of soil organic matter, could form an integral part of green and sustainable solutions to the soil fertility problem. However, their global-scale application is hindered from both scientific and regulatory perspectives by the lack of understanding of the molecular make-up of these chromatographically inseparable mixtures containing thousands of molecules. Here we show how multidimensional NMR spectroscopy of isotopically tagged molecules enables structure characterization of humic compounds. We illustrate this approach by identifying major substitution patterns of phenolic aromatic moieties of a peat soil fulvic acid, an operational fraction of humic substances. Our methodology represents a paradigm shift in the use of NMR active tags in structure determination of small molecules in complex mixtures. Unlike previous tagging methodologies that focused on the signals of the tags, we utilize tags to directly probe the identity of the molecules they are attached to.*

The world's population is estimated to rise to 11 billion by 2100,[1] putting unparalleled pressure on agricultural food production.[2] New sustainable means of increasing land fertility whilst avoiding the current overuse of NPK fertilizers are needed.[2c,3] Humic substances, which exist in soil as complex mixtures of thousands of organic compounds derived from animal and plant remains, could form an integral part of green and sustainable solutions to the soil fertility problem. Humic substances play a vital role in physical and biochemical soil functions,[4] stimulate plant growth,[5] and ameliorate the effect of contaminant metals and chemical residues, which may build up over time and hinder plant growth.[6] It is unclear, however, if all fractions of humic substances are beneficial to plant growth,[5c] and if certain humic molecules may indeed be harmful to humans.[7] Thus, to address the potential role of humic substances in restoring the fertility of intensively farmed agricultural soils, it is essential to determine their molecular composition. Such an achievement will allow development of a molecular rather than a phenomenological description of soil—a fundamental step toward defining the structure–function relationships of humic substances. Nevertheless, the two high-resolution analytical techniques, Fourier transform ion cyclotron resonance mass spectrometry (FT ICR MS)[8] and NMR spectroscopy,[6b,9] which are essential in this endeavor, have serious shortcomings: MS only provides molecular formulae, and standard 2D or 3D NMR experiments cannot deliver the structures of individual compounds contained in complex mixtures.

The development of more powerful multidimensional (*n*D) NMR experiments is essential in overcoming these limitations.[10] However, *n*D NMR alone cannot address the inherent complexity of chromatographically inseparable mixtures,[10a] and some form of "spectroscopic separation" is still necessary. To achieve this, we have recently developed an isotope-filtered *n*D NMR methodology[11]—a combination of isotopic tagging and *n*D NMR. Unlike previous tagging approaches that focused on the signals of the tags,[6b,12] we utilize them to probe directly the identity of the tagged molecules. The power of this approach is in the ability to provide multiple correlated chemical shifts of individual molecules.

Here we report how isotope-filtered *n*D NMR enables characterization of phenolic moieties of humic molecules and illustrate its principles using a 4D $^{13}CH_3O$-filtered NMR experiment, 4D $HCCH_3$, which correlates chemical shifts of four nuclei—the aromatic CH atoms *ortho* to methoxy groups and those of $^{13}CH_3O$ atoms. The polarization transfer pathway of this experiment starts on aromatic protons and finishes on methyl protons utilizing $^1J_{CH_{ar}}$, $^3J_{CC}$, and $^1J_{CH_3}$ couplings (Figure 1 a). The pulse sequence of 4D $HCCH_3$ is shown in Figure S1 in the Supporting Information.

The workings of this experiment are illustrated using a peat soil fulvic acid (FA) sample that was methylated using $^{13}CH_3I$.[11] The sample was first characterized by acquiring a 2D ^1H,^{13}C HSQC spectrum (Figure 1 b). Its methoxy region shows a spread of $^{13}CH_3$ cross peaks reflecting different chemical environments of the methoxy groups. As the 4D $HCCH_3$ experiment was designed to investigate phenolic compounds, an expansion of the phenolic methoxy region of

[*] N. G. A. Bell, A. A. L. Michalchuk, J. W. T. Blackburn, Dr. D. Uhrín
EastChem School of Chemistry, University of Edinburgh
Joseph Black Building
David Brewster Rd, Edinburgh EH9 3FJ (UK)
E-mail: dusan.uhrin@ed.ac.uk
Dr. M. C. Graham
School of Geosciences, University of Edinburgh, Grant Institute
James Hutton Road, Edinburgh EH9 3FE (UK)

[**] This work was supported by the NERC grant NE/L00044X/1 to M.C.G. and D.U. N.G.A.B. would like to acknowledge the support of the University of Edinburgh Principal's Career Development Scholarship and NERC. We thank J. Bella and Dr. L. Murray for maintaining NMR spectrometers. The peat sample was collected with the permission of the Scottish Wildlife Trust.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201503321.

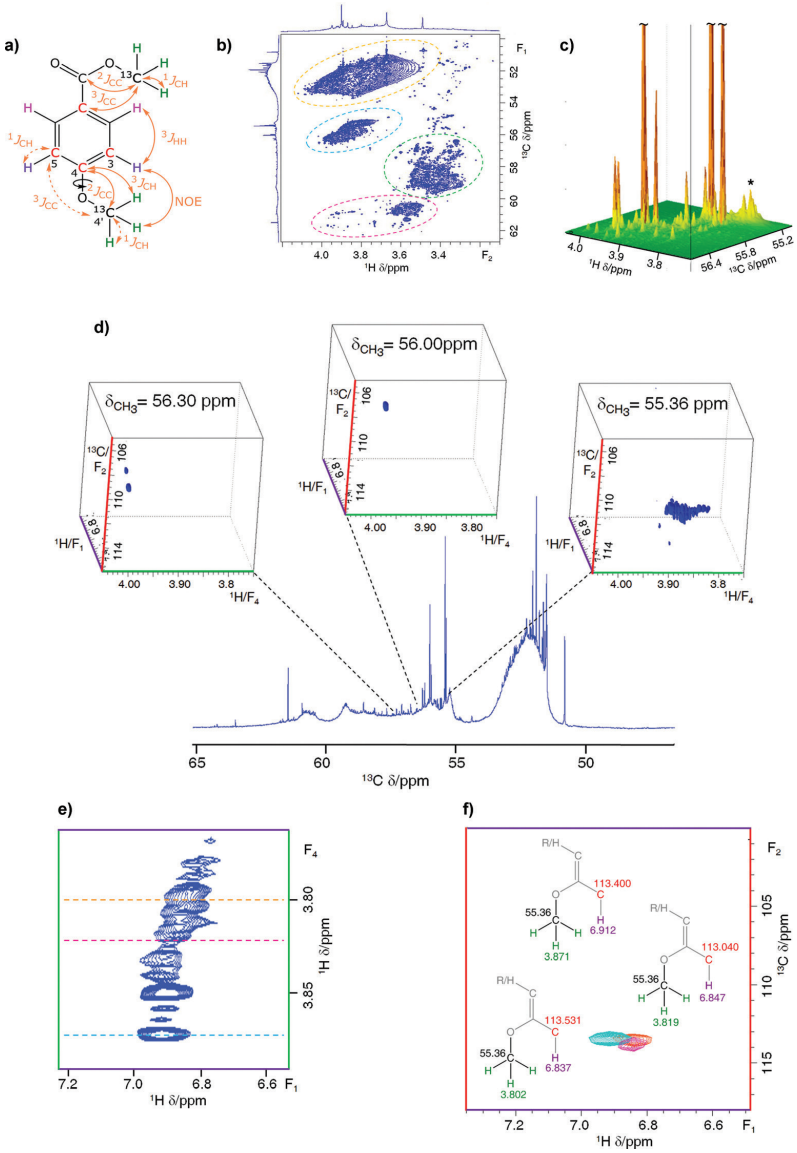


Figure 1. a) An exemplar aromatic compound highlighting nuclei accessible by $^{13}CH_3O$-filtered *n*D NMR experiments. The interactions mediating the polarization transfers in the 4D $HCCH_3$ are shown as dashed double-headed arrows. The color coding of nuclei is the same as used for the chemical shift axes of *n*D NMR spectra; b) the methoxy region of the 800 MHz 2D $^1H,^{13}C$ HSQC spectrum of $^{13}CH_3O$-methylated FA. Circled areas identify subregions as ester (orange), phenolic (cyan), carbohydrate (green), and aromatic/aliphatic sterically hindered methoxy groups (magenta); c) A stack plot of the phenolic methoxy resonances corresponding to the cyan-circled cross peaks in (b). The region of the spectrum used to illustrate the 4D $HCCH_3$ experiment below is labeled with an asterisk. d) Exemplar 3D cuboids extracted from a 800 MHz 4D $HCCH_3$ spectrum of ^{13}C-methylated FA at $^{13}C(H_3)$ chemical shifts indicated by the dashed lines. e) An F_1F_4 ($H_{ar}(C)H_3$) projection of a cuboid extracted at 55.36 ppm; f) An overlay of three 2D F_1F_2 ($H_{ar}C_{ar}$) planes extracted from this cuboid at proton methoxy chemical shifts indicated by the colored dashed lines in (e). Insets show the identified structural fragments, which belong to compounds **23**, **32**, and **33** (Figure S2).

the 2D $^1H,^{13}C$ HSQC spectrum is shown as a stacked plot in Figure 1c. This presentation reveals the presence of several major and numerous minor phenolic compounds in this FA sample.

A partial 4D $HCCH_3$ spectrum of ^{13}C-methylated FA shown in Figure 1d illustrates how isotope filtration combined with the dispersion of signals in a 4D space leads to comprehensive structural information. A cuboid extracted at 55.36 ppm is inspected. This region of the spectrum contains severely overlapping medium intensity signals labeled with an asterisk on Figure 1c. An F_1F_4 (or H_{ar}-($C_{ar}COC)H_3$) projection of the 55.36 ppm cuboid shows an adequate spread of signals in the directly detected F_4 dimension (Figure 1e). An overlay of three 2D F_1F_2 (or $H_{ar}C_{ar}$) planes taken from this cuboid at methyl proton chemical shifts indicated by the dashed lines in Figure 1e reveals similar, but non-identical, H_{ar}/C_{ar} chemical shifts (Figure 1f). The inspected region of the 4D $HCCH_3$ spectrum thus provides unprecedented chemical shift separation of aromatic compounds which have identical aromatic ring substitution patterns, whereas their substituents show structural variations further away from the aromatic ring.

Additional 3D $^{13}CH_3O$-filtered experiments were designed to complement the 4D experiment. 3D variants of the 4D $HCCH_3$, 3D $H(C)CH_3$, and 3D $(H)CCH_3$, do not label aromatic carbon or proton resonances, respectively, and hence provide similar, but more ambiguous informa-

tion than the 4D spectra albeit with higher digital resolution. Chemical shifts of nuclei other than those *ortho* to the $^{13}CH_3O$ groups (Table S1), were obtained by 3D INEPT-INADEQUATE-HSQC[11]/3D C_qCH_3 (quaternary aromatic carbons) and 3D CH_3-NOESY/3D CH_3-NOESY-TOCSY (*meta* and *para* protons relative to $^{13}CH_3O$ groups). These NMR experiments utilize polarization transfer through various couplings shown in Figure 1 a. The NOESY- and TOCSY- based experiments also show the splitting of aromatic proton resonances due to $^3J_{HH}$ or $^4J_{HH}$ couplings, hence providing additional valuable structural information.

In summary, multiple chemical shifts and coupling constants were obtained for individual molecules by analyzing the 3D/4D $^{13}CH_3O$-filtered spectra. When cross-referenced against database information,[13] this lead to the identification of nine major phenolic compounds of this FA sample (Figure 2). These comprise 1,3,4-, 1,3,5-trisubstituted as well

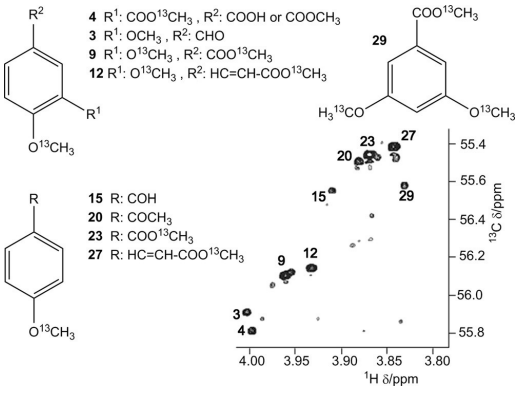


Figure 2. Major phenolic compounds identified in the methylated FA sample; their methoxy cross peaks are labeled on the 2D ^1H,^{13}C HSQC spectrum.

as 1,4-disubstituted hydroxybenzenes that differ by the nature of the *para* substituent relative to the $^{13}CH_3O$ group.

Twenty-one additional structures/structural motifs were also identified (Figure S2). Some of these contained highly substituted aromatic rings with four or five substituents. These molecules mainly relate to plant-derived lignin precursors such as vanillin, coumaryl, syringyl, coniferyl monolignols, but also flavonoids. The presence of several of these molecules in humic substances has been postulated based on the results of tetramethyl ammonium hydroxide (TMAH)-assisted pyrolysis GC-MS analysis.[14] Aromatic ring substitution patterns identified in our work were previously suggested based on the bulk matching of experimental and theoretical 2D ^1H,^{13}C HSQC spectra; the latter were calculated by considering the effects of -OR and -COOR groups on the ^1H and ^{13}C chemical shift of CH atoms.[15] These approaches relied on matching a limited number of descriptors to numerous possibilities. In contrast, our methodology provides a multitude of correlated chemical shifts, which allows convergence to a single structure/structural motif, thus yielding unprecedented structural details for phenolic compounds in samples of humic substances.

Further advances of this methodology are not limited to methylation. In addition, other tags containing NMR-active nuclei such as ^{15}N, ^{19}F, and ^{31}P, represent promising candidates for tagging various functional groups. Once fully developed, this methodology will lay ground for the structure–function investigations of humic compounds thus enabling exploration of their roles in improving the soil fertility and sustainable food production. Isotope-filtered *n*D NMR spectroscopy is also applicable to the investigation of natural organic matter in general, but also complex mixtures other than those found in the environment, for example, small-molecular metabolites, plant extracts, oil, food, and beverages.

Experimental Section

Peaty soil was collected from the Red Moss, an ombrotrophic peat bog near Balerno, central Scotland. The FA sample (60 mg) was extracted using the International Humic Substance Society (IHSS) protocols (http://www.humicsubstances.org/isolation.html) and methylated as described previously[11] yielding ^{13}C-methylated FA (10 mg). The sample was dissolved in 550 μL of $CDCl_3$ for NMR analysis at 15 °C. The pulse sequence of the 4D $HCCH_3$ is shown in Figure S1. The parameters of NMR experiments are detailed in the Supporting Information.

Keywords: complex mixtures · humic substances · isotopic labeling · NMR spectroscopy

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 8382–8385
Angew. Chem. **2015**, *127*, 8502–8505

[1] P. Gerland, A. E. Raftery, H. Sevcikova, N. Li, D. Gu, T. Spoorenberg, L. Alkema, B. K. Fosdick, J. Chunn, N. Lalic, G. Bay, T. Buettner, G. K. Heilig, J. Wilmoth, *Science* **2014**, *346*, 234–237.

[2] a) D. Tilman, C. Balzer, J. Hill, B. L. Befort, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20260–20264; b) J. A. Foley, N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. O'Connell, D. K. Ray, P. C. West, C. Balzer, E. M. Bennett, S. R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockstrom, J. Sheehan, S. Siebert, D. Tilman, D. P. M. Zaks, *Nature* **2011**, *478*, 337–342; c) D. Tilman, K. G. Cassman, P. A. Matson, R. Naylor, S. Polasky, *Nature* **2002**, *418*, 671–677.

[3] a) J. Elser, E. Bennett, *Nature* **2011**, *478*, 29–31; b) N. Gilbert, *Nature* **2009**, *461*, 716–718; c) P. M. Vitousek, H. A. Mooney, J. Lubchenco, J. M. Melillo, *Science* **1997**, *277*, 494–499; d) J. Penuelas, B. Poulter, J. Sardans, P. Ciais, M. van der Velde, L. Bopp, O. Boucher, Y. Godderis, P. Hinsinger, J. Llusia, E. Nardin, S. Vicca, M. Obersteiner, I. A. Janssens, *Nat. Commun.* **2013**, *4*, 0.

[4] a) S. Nardi, D. Pizzeghello, A. Muscolo, A. Vianello, *Soil Biol. Biochem.* **2002**, *34*, 1527–1536; b) H. L. Bohn, B. L. McNeal, G. A. O'Connor, *Soil chemistry*, 2nd ed., Wiley, New York, **1985**.

[5] a) S. Trevisan, D. Pizzeghello, B. Ruperti, O. Francioso, A. Sassi, K. Palme, S. Quaggiotti, S. Nardi, *Plant Biol.* **2010**, *12*, 604–614; b) A. Muscolo, M. Sidari, S. Nardi, *J. Geochem. Explor.* **2013**, *129*, 57–63; c) L. P. Canellas, F. L. Olivares, *Chem. Biol. Technol. Agriculture* **2014**, *1*, 1–11.

[6] a) J. F. McCarthy, J. M. Zachara, *Environ. Sci. Technol.* **1989**, *23*, 496–502; b) N. Hertkorn, A. Kettrup in *Use of Humic Substances to Remediate Polluted Environments: From Theory to*

Practice, *Vol. 52* (Eds.: I. V. Perminova, K. Hatfield, N. Hertkorn), Springer, **2005**, pp. 391 – 435.
[7] Y. C. Hseu, H. W. Huang, S. Y. Wang, H. Y. Chen, F. J. Lu, R. J. Gau, H. L. Yang, *Toxicol. Appl. Pharmacol.* **2002**, *182*, 34 – 43.
[8] A. C. Stenson, A. G. Marshall, W. T. Cooper, *Anal. Chem.* **2003**, *75*, 1275 – 1284.
[9] a) A. J. Simpson, W. L. Kingery, M. H. B. Hayes, M. Spraul, E. Humpfer, P. Dvortsak, R. Kerssebaum, M. Godejohann, M. Hofmann, *Naturwissenschaften* **2002**, *89*, 84 – 88; b) B. P. Koch, M. R. Witt, R. Engbrodt, T. Dittmar, G. Kattner, *Geochim. Cosmochim. Acta* **2005**, *69*, 3299 – 3308; c) N. Hertkorn, R. Benner, M. Frommberger, P. Schmitt-Kopplin, M. Witt, K. Kaiser, A. Kettrup, J. I. Hedges, *Geochim. Cosmochim. Acta* **2006**, *70*, 2990 – 3010.
[10] a) N. Hertkorn, C. Ruecker, M. Meringer, R. Gugisch, M. Frommberger, E. M. Perdue, M. Witt, P. Schmitt-Kopplin, *Anal. Bioanal. Chem.* **2007**, *389*, 1311 – 1327; b) K. Mopper, A. Stubbins, J. D. Ritchie, H. M. Bialk, P. G. Hatcher, *Chem. Rev.* **2007**, *107*, 419 – 442.
[11] N. G. A. Bell, L. Murray, M. C. Graham, D. Uhrin, *Chem. Commun.* **2014**, *50*, 1694 – 1697.
[12] a) M. A. Mikita, C. Steelink, R. L. Wershaw, *Anal. Chem.* **1981**, *53*, 1715 – 1717; b) K. A. Thorn, C. Steelink, R. L. Wershaw, *Org. Geochem.* **1987**, *11*, 123 – 137.
[13] a) IAST, National Institute of Advanced Industrial Science and Technology, **2015**; b) S. A. Ralph, J. Ralph, L. L. Landucci in http://ars.usda.gov/services/docs.htm?docid = 1328, **2004**.
[14] B. Chefetz, Y. Chen, C. E. Clapp, P. G. Hatcher, *Soil Sci. Soc. Am. J.* **2000**, *64*, 583 – 589.
[15] a) A. J. Simpson, B. Lefebvre, A. Moser, A. Williams, N. Larin, M. Kvasha, W. L. Kingery, B. Kelleher, *Magn. Reson. Chem.* **2004**, *42*, 14 – 22; b) E. M. Perdue, N. Hertkorn, A. Kettrup, *Anal. Chem.* **2007**, *79*, 1010 – 1021.

Received: April 13, 2015
Published online: June 2, 2015

