THE UNIVERSITY
of EDINBURGH

THE UNIVERSITY
of EDINBURGH

TESTED TO DESTRUCTION

WILLIAM KEW

Advanced Spectroscopic, Spectrometric, and Chemometric Analysis of Scotch Whisky

Doctor of Philosophy
School of Chemistry
The University of Edinburgh

August 2018

DECLARATION

This declaration states:

a) that the thesis has been composed by myself, and

b) either that the work is my own, or, for work conducted as a member of a research group, that I have made a substantial contribution to the work, such contribution being clearly indicated, or

c) that the work has not been submitted for any other degree or professional qualification except as specified, and

d) that any included publications are my own work, except where indicated throughout the thesis and summarised and clearly identified on the declarations page of the thesis.

Edinburgh, August 2018

William Kew

LAY SUMMARY

Scotch Whisky is an important product - it is valuable both to the economy and culture of Scotland. Part of what makes whisky special is the way that it is produced. It is a result of several steps: fermentation of raw cereals, subsequent distillation to a higher alcohol strength, and 'maturation' – where the liquid spends at least three years in an oak barrel. Usually this barrel will have been used before for other alcoholic beverages. Despite this simple process, whisky is a mixture of many thousands of different chemical compounds. This thesis uses high-level analytical instrumentation and statistical methods to further the understanding of the structure, origins, and significance of the chemical composition of Scotch Whisky.

The most abundant compounds in whisky, water and ethanol (alcohol), present a challenge to one of the techniques used here. NMR – Nuclear Magnetic Resonance – uses a very strong magnetic field and electromagnetic (EM) pulses to probe the centre (or nucleus) of atoms. A positively charged nucleus acts as a 'bar magnet' and interacts with the strong magnetic field and EM pulses to reveal information on the structure, and quantity, of molecules in the sample. Being quantitative, NMR is overpowered by the high levels of ethanol and water, and so first a method was developed to suppress their signals. A sophisticated series of NMR experiments were then modified to include this 'solvent' suppression and revealed the identities of several dozen chemicals in Scotch Whisky. Statistical analysis identified chemical differences between malt and blended whisky.

Mass Spectrometry (MS) was also used to investigate the identities of chemical compounds found in whisky. Chemicals are first ionised, *i.e.* become charged and then placed within a strong magnetic field, where the speed of an ions resulting motion describes its mass accurately and precisely, allowing for determination of its molecular formula. Determination of thousands of molecular formulae - containing carbon, hydrogen, oxygen, and sulfur - was possible across dozens of samples. Whilst the complete structure and quantity of these compounds remains unknown, it was possible to use statistical methods to identify compounds that were more prevalent in Scotch Whisky matured in ex-Sherry barrels than in ex-Bourbon barrels. Structural aspects of some species were identified by means of fragmenting - or breaking - them up. Further work investigated other means of ionising the chemicals in Scotch Whisky. These alternative ionisation sources provided complementary information on the chemistry of Scotch Whisky.

ABSTRACT

Scotch Whisky is a complex mixture comprising thousands of chemical species at a diverse range of concentrations. The identities, origins, and significance of many of these compounds is largely unknown. Routine characterisation of Scotch Whisky mostly utilises techniques such as gas or liquid chromatography (GC, LC) coupled to a FID, UV, or low-resolution mass spectrometry (MS) detector. In this thesis, advanced spectroscopic and spectrometric techniques are investigated as potential complementary means to unravel the chemical complexity of Scotch Whisky. Chemometric methods are applied to decipher the significance or potential origin of many of these compounds.

Being predominantly (*ca.* 99 %) protonated 'solvent' – ethanol and water – ^{1}H and ^{13}C Nuclear Magnetic Resonance (NMR) required solvent suppression to be implemented into the acquisition of high resolution spectra. A 1D NOESY-presaturation sequence was modified and implemented in automation for this purpose. Furthermore, this solvent suppression was coupled with several 1D and 2D homo- and heterocorrelated NMR experiments for the analysis of Scotch Whisky. With limited sample preparation – only addition of buffer – an approximate limit of detection of 50 μM was achieved.

The developed NMR methodology was subsequently used for structural elucidation of dozens of compounds in Scotch Whisky. Quantification of these compounds was hindered by variable chemical shifts, signal overlap, and for some compounds the existence of equilibria of different forms. Quantification of ethanol concentrations in model solutions and genuine Scotch Whisky samples was successful. A large set of Scotch Whisky samples were analysed by the solvent suppressed 1D ^{1}H NMR methodology and various statistical techniques including statistical total correlation spectroscopy (STOCSY), independent and principal component analysis (ICA, PCA), and orthogonal partial least squares discriminant analysis (OPLS-DA). Various parameters were modelled, and discrimination of 'malt' or 'blend' status was achieved, whilst maturation wood type discrimination was less successful.

High resolution negative mode electrospray ionisation (ESI) Fourier transform ion cyclotron resonance (FTICR) MS was then used to examine a large set of Scotch Whisky samples. Thousands of unique molecular formulae were assigned within a 1 ppm threshold, representing an assignment rate of 72-88 % of the detected peaks in the spectra. Assignments were selectively confirmed by isotopic fine

structure (IFS) analysis, and structural information obtained by both quadrupolar isolation and fragmentation, and in-cell isolation and fragmentation. Similar chemometric methods as applied to NMR data were used to model sample parameters, and identification of potential maturation wood marker compounds was achieved.

Alternative ionisation sources – including atmospheric pressure chemical- and photo-ionisation (APCI, APPI), and laser desorption ionisation (LDI) – were compared for the analysis of Scotch Whisky by FTICR MS. The differing sources provide complementary compositional information on the sample set, with APCI and LDI being most different to ESI in terms of compounds ionised and spectral profiles. Positive mode ionisation was also successful, but molecular formula assignment was hindered by insufficient resolving power. Late experimentation pushed the achievable resolving power to 2.8 million at m/z 400, however the required approach is significantly more time consuming and prone to signal quality degradation. Formula assignment software, both commercial, open source, and in-house developed, were compared. The commercial and published open source software provided essentially identical results for Scotch Whisky, and whilst the in-house tools assigned fewer species (a subset of those assigned by the other tools), they did so with a smaller mean error of assignment. Various analysis and visualisation tools for MS data of complex mixtures were also developed.

PUBLICATIONS

The following is a list of peer-reviewed publications resulting from the work conducted during this Ph.D. Where this work is included in this thesis, this is declared at the start of each chapter.

(1) W. Kew, I. Goodall, D. Clarke and D. Uhrín, *Journal of The American Society for Mass Spectrometry*, 2017, **28**, 200–213.

(2) W. Kew, N. G. Bell, I. Goodall and D. Uhrín, *Magnetic Resonance in Chemistry*, 2017, **55**, 785–796.

(3) J. W. T. Blackburn, W. Kew, M. C. Graham and D. Uhrín, *Analytical Chemistry*, 2017, **89**, 4382–4386.

(4) W. Kew, J. W. Blackburn, D. J. Clarke and D. Uhrín, *Rapid Communications in Mass Spectrometry*, 2017, **31**, 658–662.

(5) W. Kew, J. W. T. Blackburn and D. Uhrín, *Analytical Chemistry*, 2018, **90**, 5968–5971.

(6) W. Kew, C. L. Mackay, I. Goodall, D. J. Clarke and D. Uhrín, *Analytical Chemistry*, 2018, **90**, 11265–11272.

ACKNOWLEDGMENTS

I know that love is ultimately the only answer to mankind's problems.
And I'm going to talk about it everywhere I go.

— Dr Martin Luther King Jr. (1967)

John Donne wrote that 'no man is an island', and despite the considerable individual effort and sacrifice required to complete a PhD, it could not be done alone.

First, I must thank my supervisor Dušan Uhrín for inviting me to work with him for these past four years, for his instruction, guidance, and patience. His faith in me allowed this project to blossom. Thanks also to David Clarke, who has been a substantial help as my co-supervisor, offering invaluable insights.

The Scotch Whisky Research Institute deserves special recognition for their superb support throughout this project. Ian Goodall has been an excellent industrial supervisor, providing me with great help, support, and freedom. John Conner is an encyclopaedia of maturation knowledge, which has been fundamental to this project. Thanks also to James Brosnan for his advice and support over the years. Also, I must thank the many people who made me very welcome, including Peter, Rebecca, Claire, Liz, John F, Barry, Mark, Shona, Mhairi, Pamela. Thanks also to the member companies for their support and time. Thanks also to the BBSRC for funding this project with the SWRI.

At the University of Edinburgh, I must give special thanks to the technical staff who made this entire project possible. Thanks to Juraj, Lorna, Logan, and Faye for their patience, instruction, support, troubleshooting, advice, and help over the past four years. Thanks also to my project students, Max and Ryan.

My group members, past and present. Thanks Harris, Zuzana, Nicholle, Natalia, John, Hannah, Adam, Ariana, Alan, Lorna, Yufan, and Justinas for the good times, for the cake, karaoke, and for the slivovice. John has been an excellent friend, sounding board, and gym buddy, and I have to give special thanks to him, Harris, and Natalia for making me feel so welcome when I started. I can't imagine having done this without all of you.

To my family, thank you for the support and love you have given me over the past 27 years. You have given so much, so freely, and I am eternally thankful to you all.

For Luke, thank you for your support, laughter, friendship, and love. You were my rock for a long time, and I will be forever grateful to have had you in my life.

Finally, I must thank the all the people whose names I may have forgotten but whose efforts I haven't. Admininstrators, technicians, conference organisers, conference buddies, students, and staff. Thanks especially to the chemistry community on Twitter, for `#RealTimeChem` and `#NMRChat` - a unique forum to bounce ideas off of. Thanks to the Nespresso machine in the office, the true hero of this story.

And to Scotland, a beautiful country and home to Scotch Whisky.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ABV Alcohol by Volume

ADC analog-to-digital converter

AI aromaticity index

ANOVA analysis of variance

APCI atmospheric pressure chemical ionisation

API atmospheric pressure ionisation

APPI atmospheric pressure photoionisation

BASHD band-selective homodecoupled

BURP band-selective uniform response pure-phase

CID collision induced dissocation

COSY Correlation Spectroscopy

CSSF chemical shift-selective filter

DBE double bond equivalence

DEPT distortionless enhancement of polarisation transfer

DHB 2,5-dihydroxybenzoic acid

DMS dimethyl sulfide

DMTS dimethyl trisulfide

DIPSI decoupling in the presence of scalar interactions

DOSY diffusion ordered spectroscopy

DQF double quantum filtered

DSS 4,4-dimethyl-4-silapentane-1-sulfonic acid

EC ethyl carbamate

ECD electron capture dissociation

EPR electron paramagnetic resonance

ESI electrospray ionisation

EXACT extended acquisition time

EXSY exchange spectroscopy

FID free induction decay

FTICR Fourier transform Ion Cyclotron Resonance

FT Fourier transform

FWHM full width half maximum

GC gas chromatography

HMBC Heteronuclear Multiple Bond Correlation

5-HMF 5-hydroxymethyl furfural

HPLC high performance liquid chromatography

HRMAS high resolution magic angle spinning

HSQC Heteronuclear Single Quantum Coherence

IAT ion accumulation time

ICA Independent Components Analysis

ICR Ion Cyclotron Resonance

IFS isotopic fine structure

INEPT Insensitive Nuclei Enhanced by Polarization Transfer

IRMPD infrared multiphoton dissociation

IST iterative soft thresholding

KMD Kendrick mass defect

LC liquid chromatography

LDI laser desorption ionisation

LED longtitudinal eddy delay

MALDI matrix assisted laser desorption ionisation

MMFDS methyl-2-methyl-3-furyl disulfide

MIR mid infrared

MIST modified iterative soft thresholding

MS Mass Spectrometry

NAS no age statement

NIR near infrared

NMR Nuclear Magnetic Resonance

NMS new make spirit

NOE nuclear Overhauser effect

NOESY Nuclear Overhauser Effect Spectroscopy

NOM natural organic matter

NUS non-uniform sampling

OPLS-DA orthogonal partial least squares discriminant analysis

RANSY ratio analysis NMR spectroscopy

PCA principal components analysis

PFG pulsed field gradient

PLS-DA partial least squares discriminant analysis

ppb parts-per-billion

ppm parts-per-million

PSYCHE pure-shift yielded by chirp excitation

PURGE presaturation using relaxation gradients and echoes

QTOF quadrupole time of flight

REST relaxation encoded spectroscopy

RF radio-frequency

RG receiver gain

SORI sustained off-resonance irradiation

SWA Scotch Whisky Association

SWRI Scotch Whisky Research Institute

SNR signal-to-noise ratio

STOCSY Statistical Total Correlation Spectroscopy

TIC total ion current

TOCSY Total Correlation Spectroscopy

TOF time of flight

UPLC ultra performance liquid chromatography

WET Water suppression Enhanced through T1 effects

ZQ zero quantum

Part I

INTRODUCTION

Best Pour Yourself a Double

1 INTRODUCTION

From the complex prima materia of barley and oak
through transformation, distillation and complex alchemy
over time, a liquid gold elixir
emerges sublime.

— Rachel Barrie, Master Blender (2018)

1.1 SCOTCH WHISKY

Whisky, from the Gaelic *Uisge Beatha* - 'Water of Life' - has been produced in Scotland for centuries. The result of a lengthy production process, the final product is revered around the world. Global exports exceed £4bn per year,[1] and its value as a cultural export is hard to overstate. But what is Scotch Whisky? What makes it special and unique? What made it the first, and main, global spirit? Considering the increasing modern competition from other countries, especially Japan and America, how does a traditionally produced spirit drink maintain such quality and presence on the international stage? How does it stand up to counterfeits and cheap imitations, and how can we be sure what is in the bottle is the premium product the consumer pays for?

There exists a great volume of research into the chemistry of distilled spirits - including Scotch Whisky. It is important to acknowledge the work which has come before, as it forms a foundation for the analysis and context of any future research. Research into mature spirit drinks, especially Scotch Whisky, has been conducted both in industrial contexts, such as at the Scotch Whisky Research Institute (SWRI), or in academic contexts. There has been research on all aspects of Scotch Whisky production, from raw materials, fermentation, distillation, maturation, all the way through to bottling. As it is beyond the scope of this text to review the entire breadth of whisky production, the interested reader is referred to the excellent, and thorough, book 'Whisky: Technology, Production and Marketing'.[2] However, to put this thesis in context, the key stages of Scotch Whisky production - and their known, or expected, chemical effects - will be discussed here.

1.1.1 *Production*

The production of Scotch Whisky is simultaneously remarkably simple and fascinatingly complex. Despite strict production regulations, enforced to ensure category protection against cheap imitation, as well to help define the general organoleptic attributes of the category, Scotch Whisky is a variable spirit drink category. Styles vary from the light and floral, to the mature and sweet, all the way through to the heavily peated drams which have very smoky aromas. Scotch Whisky is produced across Scotland in many styles, in many ways, and is far from homogeneous - and far from boring.

The regulations defining the category of Scotch Whisky are defined at national levels and protected internationally. The most recent definition comes from the Scotch Whisky Regulations 2009 (UK Statutory Instrument No. 2890), and the category of Scotch Whisky has international protected Geographical Indication status by the European Union Spirit Drinks Regulations (Regulation (EC) No. 110/2008), ensuring only Scotland can produce Scotch Whisky.

Section 3 of the Scotch Whisky Regulations 2009 (UK Statutory Instrument No. 2890) sets out the definition of Scotch Whisky. Whilst exhaustive for legal reasons, these regulations are summarised here thus:

- Scotch Whisky is made from four ingredients - water, malted barley and other whole grains, yeast, and caramel colouring E150a
- Malt whisky must be produced with 100 % malted barley
- Scotch Whisky must be distilled to no more than 94.8 % ethanol by volume
- Scotch Whisky must be matured in Scotland, for at least three years, in oak casks, of a volume of less than 700 L.
- The only whisky which can be produced, matured, or blended in Scotland is Scotch Whisky
- Only Scotland can produce Scotch Whisky

The categories of Scotch Whisky have a hierarchical nature, summarised in Figure 1.1. All whisky produced in Scotland is Scotch Whisky, but two main subcategories are produced - Single Malt and Single Grain whisky. The production of malt and grain whisky differs significantly in terms of raw materials, fermentation and distillation. The word 'Single' - as in Single Malt or Single Grain - indicates that the whisky is the product of only one distillery. When spirit is mixed from multiple distilleries, it is 'Blended'. If two (or more) Single Malts are blended, a Blended Malt is produced. If two (or more) Single Grains are mixed, a

Blended Grain is produced. The most common category of Scotch Whisky in terms of volume sold and consumed globally, however is Blended Scotch Whisky - a mix of Single Malt(s) and Single Grain(s).[3]

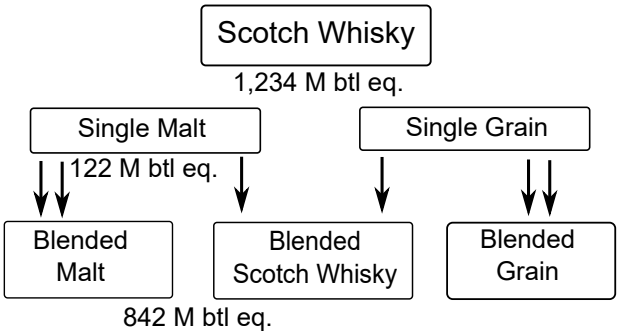


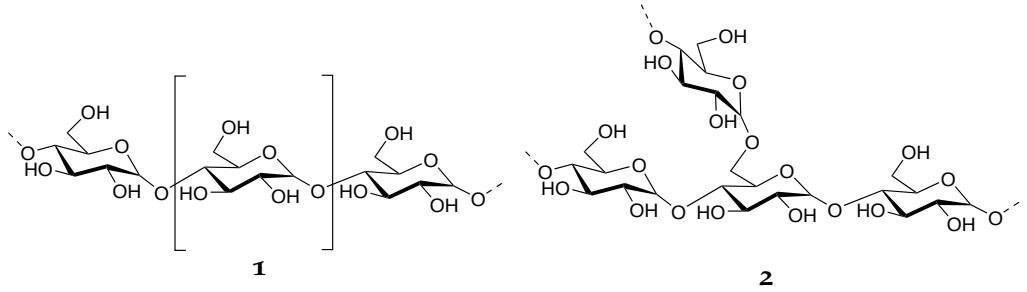
Figure 1.1: Hierarchy and export volumes of Scotch Whisky categories. Volumes in millions of of bottle equivalents. The 270 M blt eq. not shown includes Single Grain, Blended Grain, and bulk Scotch Whisky. Figures based on HMRC and Scotch Whisky Association (SWA).[3]

The two main categories of Scotch Whisky - malt and grain - differ significantly in their production (fermentation and distillation), though their maturation is much more similar. The overall process of Scotch Whisky production has 6 key steps:

- Raw cereals are processed prior to fermentation
- Fermentation to alcohol
- Distillation to higher alcohol strength
- Maturation - 3+ years
- Blending/Marrying/Vatting of Casks
- Bottling

1.1.1.1 *Raw Materials*

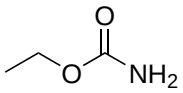
Malt Scotch Whisky begins with barley. Barley is an endosperm cereal containing starch in a protein (mostly gluten) matrix.[4,5] Starch, also known as amylum, is a glucose polysaccharide. Shorter (10^6 Da) linear starch molecules are amylose (**1**), whilst larger (10^8 Da) and branched ones are amylopectin (**2**). Example repeating units for both are shown in Figure 1.2. The linear connections are α-1,4 glycosidic linkages, whilst the branches are made with α-1,6 linkages.[6] Starch is the precursor to the fermentable sugars.



1 2

Figure 1.2: Structures of two major starch units, amylose (**1**) and amylopectin (**2**)

Aside from starch, barley contains enzymes, proteins, amino acids, and many small molecules including polyphenols such as flavanoids.[7] One source of concern for many years was the presence of glycosidic nitrile groups, such as present in epiheterodendrin. This is the major precursor to a compound of concern - ethyl carbamate (**3**). Ethyl carbamate is now controlled in many ways, with a key stage being selection of barley varieties low in glycosidic nitriles, as well as control of later processes - including fermentation and distillation.[5,8] Ethyl carbamate, at elevated levels, is reported to be toxic and a 'probable human carcinogen'.[8] The typical levels found in whisky these days are 40 parts-per-billion (ppb) ($40\,\mu g\,kg^{-1}$), well below the maximum allowed values of 1000 ppb for stone fruit spirits (brandies) allowed in the European Union (EU); note no general distilled spirits level is defined in the EU. For comparison, Canada sets a limit of 150 ppb for distilled spirits, and 400 ppb for stone fruit brandies.[9,10]



3

Figure 1.3: Structure of ethyl carbamate (**3**)

Grain whisky production uses other cereal (grain) sources, in addition to a small amount of malted barley. Predominantly, grain whisky uses wheat or corn, although not readily interchangeably due to process differences, though other cereals could be used, and have been used historically.[11] The advantage of grain sources are largely economical - corn and wheat yield more ethanol, $400\,L\,t^{-1}$, compared with $350\,L\,t^{-1}$ for barley. For bulk production, this difference is substantial.[11] The biochemistry of other cereals is similar to that of barley - a matrix containing proteins, starch, and small molecules. However, more complex subtle differences in their biology result in different properties relevant to production process and yields, a discussion beyond the scope of this thesis.[5,11]

1.1.1.2 *Malting*

The first stage in all alcohol production is preparing a source of fermentable sugars. In malt whisky production, the first stage is thus *malting*. This process involves taking the dried barley, wetting it, and allowing it to germinate partially. The germination process triggers the release of key enzymes for later saccharification, or enzymatic hydrolysis, of starch and oligosaccharides into smaller carbohydrates.[5] The key enzymes produced at this stage are α- and β-amylase. Only barley is malted for whisky production. Where other cereals are used in grain production, a small addition of malted barley is often made to introduce the required enzymes.

Additional enzymes are necessary during later fementation, namely endopeptidases for proteolysis. This breakdown of proteins into smaller peptides, amino acids, and nitrogen-containing species provides a necessary source of nutrition for yeast during fermentation. The selection of barley and other cereals with an optimal level of nitrogen is key in maximising alcohol yield.[5]

The malting process may last for several days, and is performed at a maltsers or maltings house. Most distilleries in Scotland source their malted barley externally and do not produce their own, though some still have traditional floor maltings. A floor malting involves laying the germinating barley out in a large room and regularly turning it, manually, to allow even air exposure and heat dissipation.

1.1.1.3 *Kilning and Peat*

Once the malting is complete, the malted barley is dried to arrest the germination process and to preserve it for later use. Most distilleries source externally produced malt, and even those that produce their own will need to store some for winter months. The malted barley is dried in a kiln using hot air, in a process known as *kilning*. To produce a 'peated' malt, the kilning is done with smoke and hot air from burning peat; this produces a peated whisky.[12,13] Contrary to popular misconception, peat in the water used at a distillery does not result in smoky aromas in the final whisky, however it does contribute to reduced alcohol production and a heavier (in sensory terms) new make spirit (NMS).[14] The peat smoke passes over the drying barley, infusing many phenolic compounds into the barley. Traditionally, most distilleries would have sourced their barley locally and had their own maltings, though this has changed in recent years.[5]

The level of peat infused is measured, via high performance liquid chromatography (HPLC), after the kilning stage, and not in the final product.[5] It is usually reported in parts-per-million (ppm) of total phenolic content. A traditional Islay whisky may use barley infused at around 50 ppm, though more re-

cent products have pushed the peat levels much higher. *Bruichladdich Octomore* series of whiskies contain the highest peat levels of any Scotch Whisky, with the *08.3* release containing a reported 308 ppm of phenolic content.[15] Given the sensory thresholds of phenols, more subtle production may call for a barley infused to 2-3 ppm.[5,16] It is worth noting that some phenolic compounds may be introduced into the whisky at a later stage, during maturation.

The chemistry of peat is itself complex. Peat is partially decomposed organic matter, produced in anaerobic, water-logged environments over thousands of years. Scottish peat is largely decomposed vegetation, including spaghnum mosses and heathers.[12] The slow and arrested degradation of this vegetation produces a complex chemistry. Vegetation, including both mosses here and oak used in maturation, comprise three main biopolymers - cellulose, hemicellulose, and lignin (discussed later). The ratio and composition of these biopolymers is species dependent, and the peat composition will vary both geographically and over time (depth).[12,13] Pyrolysis - gas chromatography (GC) - Mass Spectrometry (MS) was used by Harrison *et al.* to investigate the chemical composition of peat,[13] finding Islay and St. Fergus peats to be lignin-dominated, with Orkney and Tomintoul having more carbohydrates. Furthermore, Islay contained more nitrogen species and Orkney more aromatic hydrocarbons.[13]

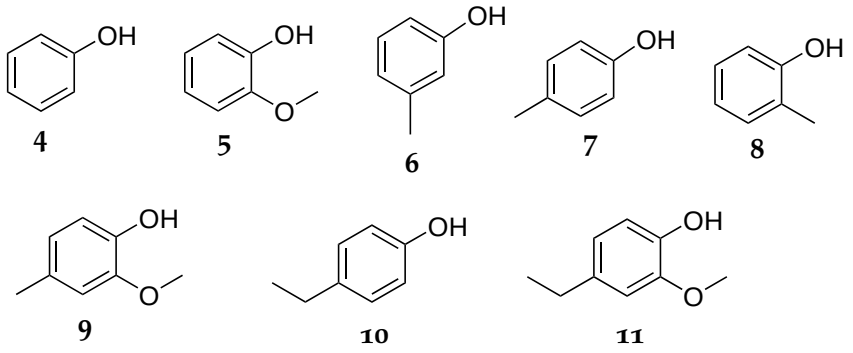


Figure 1.4: Structures of key phenolic compounds produced in peating malt. Phenol (**4**), guaiacol (**5**), *m*-cresol (**6**), *p*-cresol (**7**), *o*-cresol (**8**), 4-methylguaiacol (**9**), 4-ethylphenol (**10**), and 4-ethylguaiacol (**11**).[17]

The compounds introduced with peat smoke will vary depending on the source of the peat used, but the major known and flavour active species include phenols and guaiacols, key examples shown in Figure 1.4. Phenolic compounds have sensory thresholds, in water, of approximately ppb, equivalent to ($\mu g\,L^{-1}$), with phenol detectable at 9.5 $\mu g\,L^{-1}$.[18] In the ethanol : water matrix of Scotch whisky, these thresholds are higher and cresols and guaiacols are of greater significance. Nitrogen containing aromatic compounds - including pyridines and pyrazines - have also been reported in peated whiskies, by both GC-MS,[13] and more recently

p-H_2 hyperpolarised NMR.[19] The use of GC-MS is routine in analysis of phenolic compounds in Scotch Whisky, and work by Bendig *et al.* identified a number of bromophenols present in peated Islay Scotch Whisky samples.[20]

In addition to the peat-derivative compounds introduced during kilning, the heat can cause Maillard reactions within the barley. This process, reacting amino acids to form Strecker aldehydes, will affect even non-peated whiskies.[21] Amino acids such as valine, leucine, and methionine will react to produce aldehydes such as 2-methylpropanal, 3-methylbutanal, and methional respectively. These aldehydes may also have low sensory thresholds, with methional detectable at 0.26 ppb.[22]

1.1.1.4 *Mashing*

After malting and kilning, the barley is milled into a fine grist. This is achieved with roller mills, which crush the barley. The husk later acts as a natural filter. The grist is added to a mashtun with hot water, producing the *mash*. The water dissolves the enzymes and they begin the saccharification process. The enzymatic depolymerisation is performed by α- and β-amylase. α-amylase is most important is breaking the α-(1,4) linkages in starch, producing smaller carbohydrates such as fermentable sugars, including glucose, as well as non-fermentable dextrins. β-amylase is subsequently used for hydrolysis of dextrins, producing fermentable maltose (dimer) and maltotriose (trimer) sugars. A third enzyme, limit dextrinase, cleaves the α-(1,6) branching linkages in starch. Key products in this reaction are shown in Figure 1.5.

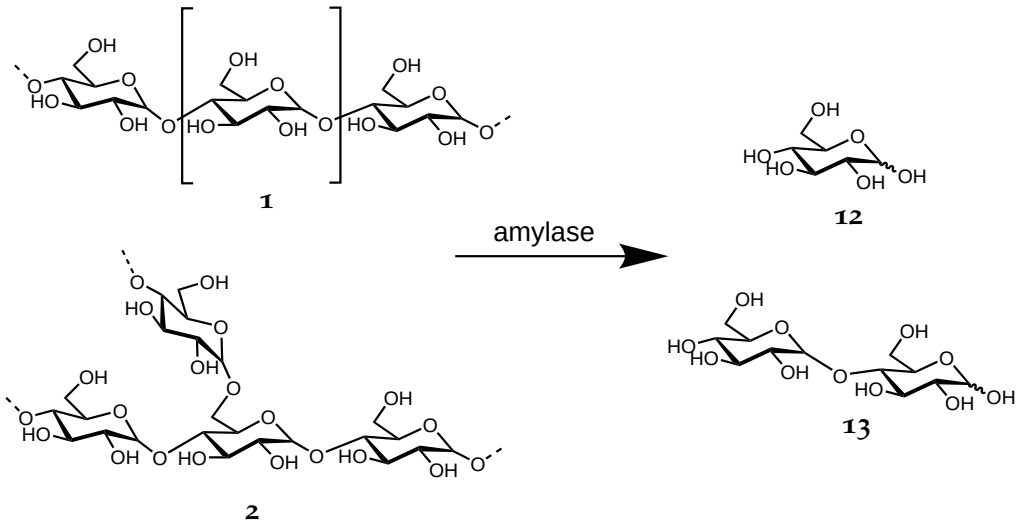


Figure 1.5: Key products from enzymatic depolymerisation of starch with α- and β-amylase. Starch units amylose (**1**) and amylopectin (**2**) are broken down to products including glucose (**12**) and maltose (**13**).

Whilst strict temperature control is not enforced in most distillery mashtuns, enzymes are most active at a specific temperature range (62 °C to 65 °C), and effective saccharification is established by control of the extraction water temperature.[5] At this stage, the mixture will be an inherently complex solution of proteins, sugars, and other constituents from the cereals. This process is typically repeated three times with increasingly hot water to ensure maximum sugar extraction. The resultant sugary mixture is known as the *wort*. The grind fineness and temperature of the mash impact the levels of wort lipids and wort gravity, which can have an impact on flavour.[5] Wort lipids can break down to aldehydes and ketones such as 1-octen-3-one and (*E*)-2-nonenal. Higher gravities may increase levels of esters in the new make spirit.[16] The wort is transferred to a washback vessel for fermentation.

1.1.1.5 *Fermentation*

Fermentation by the added yeast converts (primarily) mono- and di-saccharides into ethanol and carbon dioxide.[23] Additional metabolites are produced, such as esters, higher alcohols, glycerol, fusel oils, and more aldehydes and ketones, contributing to the complexity of the mixture.[21,23] Disaccharides, such as maltose and sucrose, the latter which may also exist within the cell structure, are first broken down to individual monosaccharides. Sucrose breaks down to glucose and fructose, whilst maltose breaks down into two molecules of glucose. Glucose is then further broken down into pyruvate via glycolysis.[23] Pyruvate decarboxylase converts pyruvate into acetaldehyde, and alcohol dehydrogenase finally produces ethanol. Within the brewing and distilling industry, a variety of yeast species are used from the genus *Saccharomyces*. The specific species used will influence not only alcoholic strength of the wash, but also the production of other flavour compounds, such as esters. *S. cerevisiae* is the most commonly used species in brewing, distilling, and baking - however each process uses different strains of this species, each providing different yields and flavour profiles. The fermentation takes between two and five days to complete. In the whisky industry, residual sugar is regarded as lost yield, and they will often monitor and optimise the fermentation to go to completion - that is, to ferment all available sugars to alcohol. During fermentation, other factors may influence the final spirit, such as microbial contamination from wild yeasts or lactic acid bacteria. If not controlled for, these may change the profile of the spirit.[23,24] The resultant product of fermentation - the *wash* - is not dissimilar to beer, and will have an Alcohol by Volume (ABV) of approximately 8 %.

During fermentation, one particularly significant negative by-product precursor is synthesised, isobutyraldehyde cyanohydrin. This is converted to ethyl carbamate

(EC) (**3**) - also known as urethane - during distillation.[8,9] EC is formed by the reaction of ethanol and urea or similar N-carbamyl compounds, excreted by the yeast. EC has been shown to occur in almost all yeast fermented products at levels around 10 ppb, though when distilled – especially in copper stills – the levels can increase to hundreds or even thousands of ppb.[8,9]

Fermentation produces many of the key volatile congeners present in Scotch Whisky. Many of these are summarised by Lee *et al.*[21] and Aylott and Mackenzie.[25] These include acetaldehyde, methanol, ethyl acetate, n-propanol, isobutanol, and 2- and 3-methyl butanol, and are present at levels of around $0.1\,\mathrm{g\,L^{-1}}$ to $3\,\mathrm{g\,L^{-1}}$ (absolute alcohol).[25]

1.1.1.6 *Distillation*

In malt whisky production, a batch process is used.[26] The wash is transferred to the first still, the wash still, and distilled to an alcoholic strength of approximately 25 %. This is then passed to the spirit still for the second distillation. This spirit is divided into foreshots, spirit cuts, and feints.[21] The foreshots and feints often contain less desirable aromas and flavours, as well as less ethanol, so are recycled to the wash still to be redistilled. The output of this is the new make spirit (NMS), at approximately 70 % ABV. The precise points at which to take the spirit cuts is a decision made by the distiller, and can influence greatly the final product. Most distilleries in Scotland use a traditional two pot still process.[21,25] The NMS is colourless but with a strong aroma, and is ready for maturation.

During distillation, a variety of chemical processes are occurring. Distillation carries over the more volatile components and increases the ethanol concentration. As the wash is heated for distillation, the additional heat may cause, or increase the rate of, a number of reactions including hydrolysis and esterification. There may also be more Maillard reactions in the wash still, producing cereal aromas.[21,26] Additionally, the copper in the still impacts the flavour by removing sulfur compounds from the spirit. Research at the SWRI demonstrated the importance of copper in key parts of the distillation apparatus at reducing the concentrations of a number of sulfury compounds in the spirit, including dimethyl sulfide (DMS) (**14**), dimethyl trisulfide (DMTS) (**15**), and methyl-2-methyl-3-furyl disulfide (MMFDS) (**16**).[27]

Compared to stainless steel stills, copper stills produce new make spirit with significantly lower sulfur compounds, confirmed by GC-MS and sensory analysis.[27] The research has not suggested a mechanism for the sulfur removal, however corrosion of the copper surface appears to be necessary. The group also noted that certain components of the still being copper or steel had a bigger impact than others, finding that the wash still condenser or spirit still pot were most

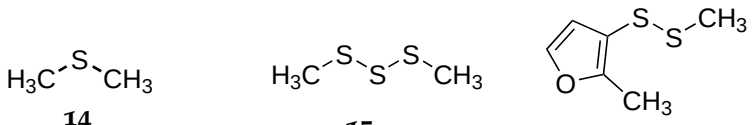


Figure 1.6: Structures of key sulfur compounds in Scotch Whisky: DMS(**14**), DMTS(**15**), and MMFDS(**16**)

significant. It should be noted this research was based on malt production; grain whisky stills often have copper sheets added to induce a similar effect.[11] Additionally, some degree of sulfury aroma can be a positive thing for the flavour of the final product, and some distillers may opt to produce a NMS with stainless steel in the distillation apparatus to help retain some sulfur. It should be noted that copper stills appear to increase levels of EC.[8] This may not be negative as EC is not particularly volatile. If EC is formed in the wash still, it is unlikely to carry over through the spirit still and thus to the final spirit.

1.1.1.7 *Grain Whisky Production*

Before discussing maturation, it is useful to compare malt NMS production (above) with that of grain NMS. Grain whisky production is largely similar - cereal sources are processed to produce fermentable sugars, fermented and distilled, and then matured - but on a more industrial scale. Whilst a normal malt distillery may produce hundreds of thousands of litres of spirit per year, a grain distillery will produce tens of millions of litres of spirit per year. This scale up is made possible by the use of a continuous distillation process, as opposed to malt whisky's batch process.[11]

Grain whisky production uses mostly raw, un-malted cereals - mostly wheat or corn - which need to be processed in such a way as to make the sugars more accessible for fermentation. This process is known as cooking, and involves adding hot water or steam to soften and solubilise the grain starch.[11] It is important not to cook the cereal at the wrong temperature, otherwise protein breakdown can cause gelatinisation, resulting in a porridge-like consistency.

The cooked grain is then transferred to a mashtun, with some quantity of malted barley and hot water. The malted barley provides a source of enzymes for saccharification and proteolysis. As with malt production, the starch is converted to smaller sugars. The wort is then fermented to produce a wash, before being distilled.

Grain distilleries use what is known as a Coffey or Patent still.[11,25] These tall columns consist of many metal plates which induce an element of reflux within the still. Wash is pumped in at a mid-point up the first still, the analyser, and the

components are separated with heat. The vapour is distilled off, and passed to the next column, the rectifier, whilst the spent wash is pumped out of the bottom of the still. The rectifier further purifies the spirit, recycling feints back to the first still, and eventually outputting the NMS to the spirit receiver. It is important to note that unlike malt stills, which are almost always made entirely of copper, Patent stills are usually stainless steel. To reduce the sulfury characteristic of the spirit, sacrificial copper may be added to the still.[11,27]

As a result of the production process, the grain spirit is much purer, cleaner, or more neutral, than malt spirit. After distillation and rectification, the grain NMS will be of a very high alcohol concentration, often around 94 %, but below the legal maximum of 94.8 %. The spirit is then typically diluted down to below 70 % for maturation. Unsurprisingly, many of the volatile congeners found in Scotch Whisky are far less abundant in grain whisky than malt whisky. Levels of the key volatile compounds can be used in authenticating unknown samples, such as in the case of blends to help determine the contribution of malt and grain whisky to their composition.[25]

1.1.1.8 *Maturation*

The longest and most significant stage in Scotch Whisky production is maturation. Maturation turns colourless NMS into legal Scotch Whisky. After the minimum of three years, the spirit has taken on colour and flavour from the wood, whilst simultaneously undergoing a myriad of reactions increasing and reducing levels of various congeners in the spirit.

The Scotch Whisky Regulations set out rules on maturation – the cask must be oak, and must have a capacity of no more than 700 L. Additionally, maturation must happen fully within Scotland. Beyond that, though, distillers have a large freedom to influence their product. Traditionally, Scotch Whisky will be matured in ex-Bourbon barrels (acquired with ease due to Bourbon regulations requiring all casks to be new and used only once).[25] These will be American White Oak (*Quercus alba*), and impart a specific flavour profile. The other mainly used cask type is an ex-Sherry butt. These casks are much larger, around 500 L compared to a 200 L bourbon barrel. Again, these are commonly made from American white or European oak (*Q. robur* or *Q. petrea*). Beyond these cask types, many more are also used, mostly made from European oaks.[25,28]

Casks are pre-treated in various ways. Beyond the intrinsic variability of the species of oak used, and the *terroir* of where the tree was grown, the provenance of the cask can include what part of the tree the wood came from (top or bottom, heart or edges), how old it is, how the wood was weathered, how the cask was coopered, and cask char or toast levels.[25,28] When preparing a Bourbon

barrel for first use, a cask is charred. This involves setting the inner surface of the barrel alight, to produce a surface layer of char, pyrolysing the oak cellulose, hemicellulose, and lignin. The extent of char will have an impact on the flavour. Alternatively, wood may be toasted - heated, but not burnt. This induces similar, but reduced, breakdown of the biopolymers without producing char. When a barrel is sourced for Scotch Whisky maturation, the producer has many options. The barrel could be simply refilled, with no modification. However, many producers may wish to regenerate the cask to increase the effects of maturation. This involves scraping off the inner surface, and re-charring or re-toasting the cask. Some producers may go further, break down a cask into its constituent staves and ends, and rebuild, or *cooper*, a different cask from the original. For example, *Laphroaig* finish some of their whiskies in a 'quarter cask', a rebuilt barrel a quarter the size of a normal barrel. This increases the relative surface area to whisky, inducing a faster and more intense maturation profile.

The length of maturation is also an important variable. Grain whiskies for blending may often be matured for as little as three years, however malts routinely age for 12 - 25 years. Some rarer whiskies may be aged for half a century or longer. This massive range of time induces significant differences between whiskies - even the same whisky from the same cask type aged for 10 or 20 years could be very different. Popular misconception is that older will always be 'more mature', as cask activity will ultimately dictate the maturation of a spirit. A new oak barrel will produce a more mature spirit much faster than a barrel which has been used many times, which may not produce an optimally mature spirit ever if the cask is 'exhausted'.[25,28,29]

A whisky may be matured in one type of cask, and then spend a shorter period in another barrel to impact additional effects. This process, known as finishing, adds to the complexity of the whisky. A distiller may source a cask previously used for maturation or fermentation of wine, or sweet or fortified wines, such as whiskies finished with Sauternes or Port casks.[29]

One final important aspect of maturation is known as 'the angels' share'. This term describes the volume of spirit lost to the atmosphere through evaporation through the porous oak casks. The porosity also allows oxygen to permeate the cask, a key part of maturation chemistry. Around 2 % of the contents of the cask may be lost each year to evaporation, though this varies with temperature and humidity.[25] In Scotland the relative humidity, generally, results in water being lost at a slower rate than ethanol, so the alcoholic strength gradually decreases over time.

The effect the previous maturation climate will have on the subsequent maturation of Scotch Whisky should not be understated. Bourbon maturation typic-

ally occurs in the US South, in a warm climate. Bourbon fill strength is similar to Scotch, at 62.5 % ABV. The combination of heat and high alcohol strength will extract a significant amount of compounds from the oak, resulting in reduced extractible content for subsequent barrel use, such as for Scotch Whisky. In contrast, Sherry wines are stored in oak in Jerez, Spain. Climatically, it is also a warm region, however the wine is of a far lower alcohol strength (12 % to 20 % ABV), and so will extract a different profile from the wood, and generally leave more residual extractible compounds for later re-use in Scotland. Furthermore, the Sherry will itself introduce a very different chemistry than Bourbon into the wood - Bourbon is a distilled spirit, similar to Scotch Whisky, whilst Sherry wine is non-distilled and produced from grapes. It will be far higher in proteins, amino acids, sugars, and other non-volatile compounds. As noted, Sherry casks for the Scotch Whisky industry can be made from European or American oak. European oak produces a higher level of extractable compounds. Bourbon is matured in American oak producting increased vanilla aromas and oak lactone (coconut aroma) levels.[21,25,29]

1.1.1.9 *Blending*

Scotch Whisky can be produced and sold in several categories. Often regarded as the premium category, Single Malt Whisky is produced from 100 % malted barley at a single distillery. However, the bottled product can be a mix of different casks, from different batches, so long as the whisky was produced at that one distillery. An age statement describes the youngest whisky component in the bottle. A Master Blender, responsible for producing the final product, will blend casks from a distillery to produce the specific flavour profile for a given brand, thus allowing consistency in an inherently variable product. This is designed to ensure that a customer's favourite brand tastes the same one month to the next, year after year.[25,30] However, it is not a scientific process and introduces another degree of complexity and variation to the whisky and the authentication of the brand.

Scotch Whisky can also be a Single Grain Whisky. This less common category describes a product of grain whisky distillation produced at a single distillery. Again, this will be a blending of a variety of casks to produce a consistent product. Alternatively, Scotch Whisky can be labelled as one of three forms of blend. A Blended Malt is a whisky which is made from 100 % malted barley, but the product is a blending of casks from different malt distilleries. This can be two, three, or dozens of distilleries.[30] Likewise, there exists Blended Grain whiskies – a product of grain whiskies produced at different distilleries and later blended together.[30]

Finally, there is the most common Scotch Whisky category, Blended Scotch Whisky. This consists of a blend of Single Malt whiskies and Single Grain whiskies, a minimum of one of each. Inherently, these source whiskies could come from one distillery which produces both grain and malt, but more commonly will come from multiple distilleries. This is due to the fact the malt and grain distillation processes are so different that most sites specialise in one or the other. Indeed, Scotland only has seven grain distilleries, supplying all the producers with spirit for their blends.[11]

This blending of casks - to produce anything from Single Malts to Blended Scotch Whisky - is known by a few terms. When casks are combined together into large tanks prior to bottling, they have been *vatted*. Should the blended spirits be allowed to further rest or mature (if returned to oak), they are described as being *married*.[30] Ultimately, all Scotch Whisky, unless it is sold as a 'Single Cask', is blended to produce the final product. This minor point is key in many of the challenges when designing experiments to analyse the effects of production on Scotch Whisky. For example, if a researcher wished to investigate the effect of maturation wood type and history on a final product, there is no realistic way of determining what the provenance of all the component whiskies and their maturation woods were. The exception to this is if the products were tracked from their original production and first fill of each cask, however the time-scales and commercial pressures involved hinder such experimental design.

1.1.1.10 *Chill Filtration*

The majority of Scotch Whisky undergoes a process called chill filtration prior to bottling. The whisky is chilled to approximately 4 °C, where some components, principally long chain fatty acid ethyl esters, precipitate out of solution and produce a haze. The whisky is passed through filter sheets made from cellulose, removing this haze.[21,25] The merits and downsides of this technique are strongly debated in the industry and amongst consumers, with many premium whiskies advertised as being 'non-chill filtered', whilst some markets demand a transparent product. Removal of these larger molecules does not impact sensory perception of the product, however if aldehydes and alcohols are caught up in these agglomerates and removed, this may impact the sensory preception.[21] Non-chill filtered whiskies may produce a haze when diluted from cask to bottle strength, around 60 % to 40 %.

Whisky, like wine, can precipitate mineral crystals over time. In whisky these tend to be calcium oxalate, and have no impact on flavour, but may be unwanted for aesthetic reasons. They are usually prevented by limiting the amount of cal-

cium introduced during production – such as through use of demineralised water for dilution.[31]

1.1.2 *Oak*

Oak is an angiosperm - flowering plant - of the genus *Quercus*, family *Fagaceae*, and consists of approximately 600 species, though in whisky maturation only a handful of these are commonly employed.[32] A complete discussion of the biochemistry of oak is beyond the scope of this thesis, and the enthusiastic reader is recommended to read Chen's chapter in 'Biotechnology of Lignocellulose'.[33] A brief summary of the relevant chemistry is presented here.

Like all wood, and indeed plant material, oak consists of mostly the polysaccharide cellulose and heteropolysaccharides known as hemicellulose. Cellulose (**17**) consists of a repeating chain of D-glucose linked at the β-(1-4) position, differentiating it from starch, whilst hemicellulose contains of a heterogenous mix of monosaccharides including xylose (**18**), mannose (**19**), rhamnose (**20**), glucose (**12**), and galactose (**21**). The ratio and linkages of these monosaccharides in hemicellulose varies between species, and sections of wood, making hemicellulose a highly heterogeneous biopolymer. Cellulose and hemicellulose exist in a matrix constituting around 55 % of the plant material.[33–35] Structures of cellulose and the individual sugars of hemicellulose are shown in Figure 1.7.

The third main component in wood is lignin, at around 20 % of dried weight.[35] Lignin is a highly complex polymer constructed of three main monomers, known as monolignols. These are *p*-coumaryl, coniferyl, and sinapyl alcohols, and are incorporated into lignin through perioxidase polymerisation.[34] The *p*-coumaryl (**22**), coniferyl (**23**), and sinapyl (**24**) alcohols form *p*-hydroxyphenol (H), guaiacyl (G), and syringyl (S) units respectively. The monolignol structures are shown in Figure 1.8. Angiosperms, such as oak, contain a mixture of G and S, with minor amounts of H.[33,35] As lignin is cross-linked, racemic, and can have molecular masses in excess of 10 kDa, the potential combinations of monomers into the final lignin can lead to very complex structures.[33]

In addition to cellulose, hemicellulose, and lignin, oak wood contains a number of other diverse polyphenolic compounds, as well as fatty acids, waxes, and terpenes.[33,35] In fact, there are over 8000 known phenolic compounds found in woods and plants.[35] The phenolic compounds found in wood can be divided into numerous classes, including phenolic acids, cinnamic acids, stilbenes, chalcones, flavones, flavonols, anthocyanins, tannins, and other polymers.[35]

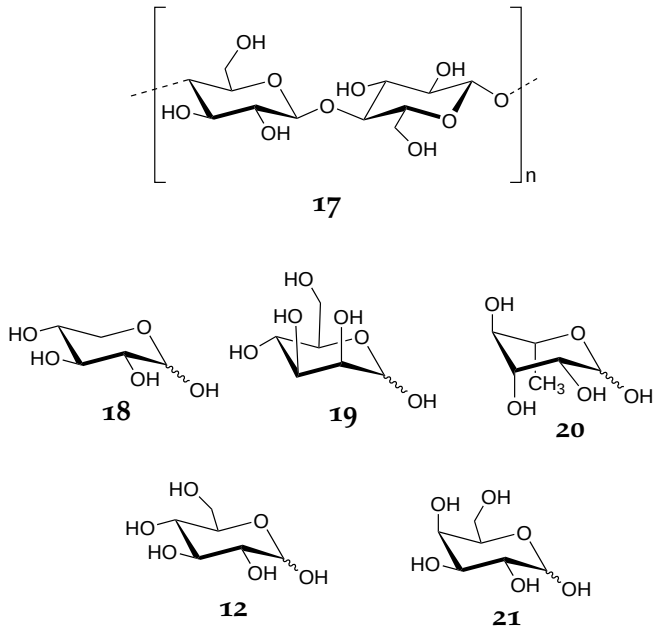


Figure 1.7: Structures of cellulose repeating unit (cellobiose, **17**) and hemicellulose components; xylose (**18**), mannose (**19**), rhamnose (**20**), glucose (**12**), and galactose (**21**). Monosaccharides shown in their pyranose (six-membered ring) form.

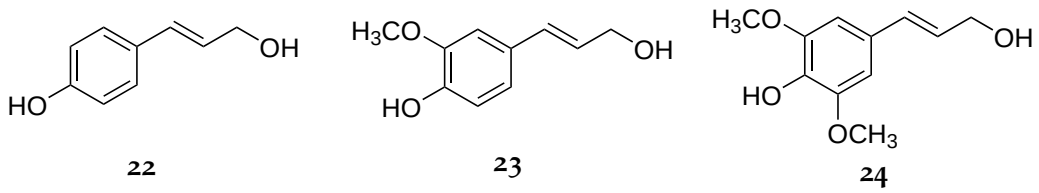


Figure 1.8: Structures of the three monolignols: *p*-coumaryl alcohol (**22**), coniferyl alcohol (**23**), and sinapyl alcohol (**24**).

Elementally, wood comprises approximately 50 % carbon, 42 % oxygen, 6 % hydrogen, 1 % nitrogen, and the remaining 1 % consists of ash content including trace metals.[33]

1.1.3 *E150a Caramel Colouring*

In addition to water, cereals, and yeast, the only other allowed component in Scotch Whisky is plain caramel colouring, also known as E150a or Class I caramel. This is added to a large number of Scotch Whisky products to regulate colour consistency across batch variation. The exact chemical composition of caramel colouring is itself an unknown, with previous work conducted in the field on that question.[36–38] Caramel is produced from heating of carbohydrates, some-

times in the presence of acids, bases, or salts. E150a must not be produced with any ammonium or sulfate compounds as catalysts. The exact recipes will be proprietary to the companies that produce and sell it to the distilleries and bottling companies, however most Scotch Whisky producers source it from a small number of companies, suggesting it may impart a chemical fingerprint specific to Scotch. The dominant components of caramel will be monosaccharides including glucose and fructose, and their pyrolysis products.[36–38] Addition of E150a does not have to be labelled, and very few producers will discuss levels of addition. As with chill filtration, caramel colouring is not designed to alter the flavour or aroma perception of the product, and is purely aesthetic. However, many consumers prefer their Scotch Whisky without addition, and so some premium products have a 'no caramel colouring' statement on their products.

1.1.4 *Chemical Analysis of Scotch Whisky*

Scotch Whisky is an important product. It is highly valuable economically and culturally. As such, there is great scientific interest in understanding the chemical processes which produce Scotch Whisky, and the chemistry of the final product. To fight adulterated and counterfeit Scotch Whisky, it is imperative to understand the chemical composition of Scotch Whisky.[25,38,39] Being able to identify markers or fingerprints, common to Scotch Whiskies, and absent in non-Scotch Whisky, or vice versa, would present a useful tool in the authenticity challenge.

Scotch Whisky, as a complex mixture, contains thousands of chemicals. Of these, only dozens are routinely analysed. From a food safety perspective, it is a vulnerable position to have a product that is not fully understood in terms of its composition. Whilst most of these compounds exist in low or trace amounts, it is important to study and characterise the mixture to rule out any potential harmful effects, or to modify the process such that they are effectively reduced, as has been practically applied in the case of ethyl carbamate.[8,9,16]

In the previous sections, the production of Scotch Whisky was discussed, introducing some aspects of the chemistry of Scotch Whisky. However, there exists a significant volume of prior research into the chemistry of Scotch Whisky and other matured distilled spirits. In this section, further research of interest is discussed.

The origins of flavour in Scotch Whisky from a biological and chemical perspective are discussed extensively in work by Lee *et al.* in 2001.[21] Aylott has published several papers over the past 20+ years discussing the chemistry of Scotch Whisky and authentication strategies based upon this.[25,40] These approaches have extended to the use of simple UV-Vis spectrophotometry for mobile brand au-

thentication.[41] UV-Vis and chemometric methods continue to be used for identification of whisky brands and counterfeits.[42]

A thorough review of the chemical composition, analysis, and authentication of whisky was presented by Wiśniewska *et al.* in 2014.[16] A review of the analysis of vodka was also published by Wiśniewska *et al.*[43] A thorough investigation, including sensory and chemical analysis, into a Scotch Whisky of unique historical significance - a whisky found at Sir Ernest Shackleton's 1907 Antarctic base camp - serves as an interesting review of the routine analyses in use today.[17] A multimethod approach for brand identification of distilled spirits was also presented in 1999.[44] A 2010 review discusses analytical and chemometric approaches to the characterisation of wine.[45] Riedl *et al.* published an extensive review of nontargetted approaches for food authentication, covering a large range of analytical and statistical techniques, and food products including alcoholic beveragess.[46]

Raman and near infrared (NIR) analysis of Scotch Whisky 'on-a-chip' performed well for a limited sample set (n=6 brands) for ethanol quantification and brand classification.[47] Kiefer and Cromwell extended the use of Raman spectroscopy for analysis of Single Malt Scotch Whisky in 2017, allowing for through-bottle analysis.[48] Chemometric analysis demonstrated potential for classification of whisky samples and authentication. Cognac was successfully discriminated from other distilled spirts by mid infrared (MIR) spectroscopy.[49] Total reflection X-ray fluoresence has also been used for Scotch Whisky analysis, distinguishing counterfeits from authentic samples.[50] Wiśniewska *et al.* performed a headspace MS, MIR and UV-Vis analysis for generic authentication of Scotch Whisky compared to American, Spanish, and Irish whiskies.[51] In this case, generic authentication was best achieved by the MS approach.

In 2005, Møller *et al.* coupled electrospray ionisation (ESI) with quadrupole time of flight (QTOF) for analysis of fifteen samples including American, Scotch, and 'presumably counterfeit whiskies' from Brazil.[52] This study used both positive and negative mode ESI and found ions of interest between m/z 80 and 550. Discriminant analysis successfully separated American and Brazilian whisky from Scotch Whisky, and there was some separation of Single Malt and Blended Scotch Whisky. Collins *et al.* used ultra performance liquid chromatography (UPLC)-QTOF MS in 2014 to profile non-volatile components in American whiskies, including Bourbon and rye whiskey.[53] In this study, young (<4 years) and old (>8 years) whiskies were distinguishable, as were Bourbon, rye, and Tennessee whiskies. The use of chromatographic separation and fragmentation allowed for identification of compounds of interest for distinguishing between American whisky types.[53]

In 2013, ESI was used with Fourier transform Ion Cyclotron Resonance (FTICR) MS for the analysis of whisky for the first time.[54] This work compared multiple samples of five authentic Scotch Whisky brands with counterfeit samples, and successfully discriminated between the real and fake samples, and showed differences between the brands, using either positive or negative mode ionisation. The high resolving power of FTICR allowed for the observation of hundreds of ions from m/z 300-700 (positive mode) and m/z 250-600 (negative mode). Exact masses were assigned chemical formula, although the confidence in these assignments is uncertain due to poor mass accuracies (assignment errors greater than 1 ppm were reported), and assignment rates were not reported. Structures were proposed for the most abundant species, however no structural elucidation was performed. Statistical analysis was based only on the 50 most abundant ions.[54]

The use of ESI FTICR MS for analysis of alcoholic drinks goes back to at least 2001, with Cooper and Marshall analysing the chemistry of wine.[55] Positive and negative mode ESI were both used, although negative mode presented greater diversity of spectra - a similar observation would be later made by whisky.[54] Formula assignments, within 1 ppm, were made for up to 94 % of species with >10 % relative abundance - a high threshold. More recent work by Roullier-Gall *et al.* has coupled these techniques for wine analysis in a metabolomics approach,[56] to understand more about the role of wood in wine production,[57,58] the *terroir* effect,[59] and even the combination of vintage and *terroir* in Burgundy wines.[60] Much more recently, the same group have presented analysis on whisky by similar techniques, investigating sample differences and wood extractions.[61] ESI has also been used for a fingerprint analysis of beers.[62]

Paper spray MS has recently been used for authentication of Blended Scotch Whisky,[63] This used chemometric techniques, including principal components analysis (PCA) and partial least squares discriminant analysis (PLS-DA), to compare the most commonly adulterated Scotch Whisky brands in Brazil with known counterfeits, and successfully discriminated between them. Subsequently, paper spray MS has also been used for quantification of the adulteration of whisky with sugar cane spirit.[64]

Other techniques have been also applied to whisky analysis. Isotope ratio analysis, based on a GC combustion approach, has been used to authenticate Scotch Whisky according to the $^{13}C/^{12}C$ ratios, specifically of volatile congeners including acetaldehyde, ethyl acetate, n-propanol, isobutanol, and amyl alcohols.[65] The reasoning behind this is the difference between barley (a C3 photosynthetic pathway) or corn or sugar cane (C4) which should allow for differentiation of organic compounds produced from fermentation of different cereals - thus discrimination between malt and blend or counterfeit whisky. C3 pathways result in

a depleted δ^{13}C‰. Unfortunately, wheat, and other common alcohol substrates such as grapes, also has a C3 pathway and this approach has limited applicability.[65] Counterfeits have also been identified by ^{2}H and ^{18}O stable isotope analysis, based on geographic origins of water having different isotopic ratios.[39] This work had promising results but would require an extensive database of isotope ratios for all water sources used in Scotch Whisky production across time, and knowledge of which water is used in each Scotch Whisky final product - a whisky may be distilled at one location, matured at another, and bottled (including dilution) at a third. A similar approach for analysis of ^{2}H levels in alcoholic spirits using quantitative Nuclear Magnetic Resonance (NMR) was presented in 1983,[66] and could be used to determine the botanical origin of ethanol. Increases in sensitivity in NMR may allow for analysis of the non-ethanol congeners. As with carbon ratios, the method applicability depends if the suspect sample has a Single Malt Scotch Whisky claim, which will have more specific isotopic ratios, rather than a blend, which may have much more natural variability and thus be harder to authenticate or to prove as counterfeit. ^{1}H NMR has been used to fingerprint Russian and Kenyan spirits for counterfeit identification via linear regression with some success.[67]

NMR has also been used to investigate hydrogen bonding between water and ethanol in alcoholic solutions,[68] and, along with Raman, to examine the effect of solutes in whisky on this hydrogen bonding.[69] This approach was also extended to Japanese sake.[70]

NMR, along with electron paramagnetic resonance (EPR) and UV-Vis, was used to study the antioxidant capacity of a Polish liquer, nalewka. The ^{1}H NMR method used presaturation of the water and ethanol, and the aromatic content of the sample was measured by integrating 6.2 - 8.2 ppm.[71] NMR has been used for analysis of solid lignin from rotting *Q. alba* by means of high resolution magic angle spinning (HRMAS). This identified a number of compounds previously observed in extracted lignins.[72] A novel NMR method was also developed to study intact wine bottles, yielding information on acetic acid levels as a measure of wine oxidation and quality.[73]

NMR, coupled with the presaturation using relaxation gradients and echoes (PURGE) solvent suppression,[74] was used for identification and quantification of drugs, specifically γ-hydroxybutryic acid (GHB, a 'date rape' drug), spiked into alcoholic and non-alcoholic beverages, including whisky.[75] This method was successful in all tested drinks matrices except wine, with variable observed detection limits and limits of quantification depending on the drug and matrix - in whisky mixed with diet coke, the detection limit was 0.01 % w/v although the limit of quantification was much poorer at 0.98 % w/v.[75]

Monakhova *et al.* published an automated NMR water and ethanol solvent suppression methodology in 2011.[76] This was based on the well established 1D NOESY presaturation sequence commonly used in metabolomics studies,[77] and achieved a reported limit of detection of 1 g hL^{-1}. This method was used for analysis of Greek grape marc spirits.[78] A similar approach was used for wine[79,80] and beer analysis.[81–83] Recently, an advanced NMR separation experiment - relaxation encoded spectroscopy (REST) - was applied to the analysis of beers, also.[84]

More focused analysis has included the use of GC-MS with chiral columns for enantiodifferentiation of whisky and cognac lactones.[85] It was known that maturation of spirits in oak produces optically pure (3S,4R) *trans*- and (3S,4S) *cis*-whisky lactones, and different oak species produce different ratios of *cis*/*trans*-whisky lactone. This method has significant potential for authentication and analysis of mature spirits.[85] The chemistry of oak lactone has been of interest for many years, with the precursor identified by means of a number of techniques, including MS and NMR, in 2000.[86] Likewise, much research has been conducted into ellagitannins and their derivatives in mature drinks.[87–90] Key polyphenols in Japanese whisky from ellagitannin oxidation were also identified.[91] This work identified - via column chromatography, UV, MS, and NMR - three new phenolic compounds, carboxyl ellagic acid and the rather unimaginatively named whisky tannins A and B. The latter two compounds are proposed to be produced from oxidation of a major oak ellagitannin castalagin, and are likely present in non-Japanese whisky, too. The stereochemistry of major ellagitannins is also an area of research interest.[92]

This section has demonstrated the significant complexity of Scotch Whisky in terms of its production and the resulting effects upon its chemistry. Furthermore, it shows how many researchers are investigating not only Scotch Whisky, but other alcoholic beverages by a large range of analytical methods, often coupling statistical analysis. The primary focus in much published work is in terms of authentication and detecting counterfeits, but there are other practical interests in understanding the origins of the compounds in Scotch Whisky, and their flavour impacts. As shown, a wide variety of analytical techniques are commonly employed for analysis of Scotch Whisky - especially sensory analysis, GC- and liquid chromatography (LC)-MS. Whilst there has been use of NMR for analysis of other alcoholic spirits, rigorous study of Scotch Whisky by NMR has not been conducted. Likewise, although FTICR MS has been used for a variety of alcoholic beverages, including whisky, a thorough analysis of Scotch Whisky using this technique had not been conducted at the onset of this PhD. The following sections introduce the major techniques used in this thesis - NMR, FTICR MS, and chemometric analysis.

1.2 NUCLEAR MAGNETIC RESONANCE

Nuclear Magnetic Resonance (NMR) is one of the most important and powerful spectroscopic techniques available, providing invaluable insight into chemical structure and dynamics across chemistry, biology, and materials science. It is a quantitative, non-destructive technique which has received significant attention and development since its discovery by Bloch and Purcell in 1945, who were later awarded the Nobel Prize in physics in 1952 for 'development of new methods for nuclear magnetic [precession] measurements and discoveries in connection therewith'.

NMR, specifically solution-state NMR in this thesis, serves routine use in many aspects of chemical research. It can provide enormous structural detail on molecules, allowing for their unambiguous structural elucidation. It is also quantitative, and can be used to measure relative and absolute concentrations of compounds in solution. As a solution state technique, it allows for the characterisation of many compounds in the liquid phase or dissolved in solution. In the contexts of research into Scotch Whisky, this is inherently advantageous.

1.2.1 *Basic Principles*

NMR relies on the magnetic properties of the nuclei within a compound or compounds in solution. At a basic level, these nuclei, possessing a magnetic moment, are essentially microscopic 'bar magnets'. These magnets, in the presence of a strong static magnetic field, align with or against the field (in the case of a spin-$\frac{1}{2}$ nuclei). These two orientations have near-identical energetics, and it is the small difference in energy levels which are probed by NMR. By application of radio-frequency (RF) pulses, the orientation and position of these bar magnets can be carefully manipulated. Their resultant motion induces an electrical current into the detector, which is recorded and processed giving rise to an NMR spectrum. This can provide quantitative and structural information on the sample. In practice, significant aspects of NMR spectroscopy - both data acquisition and processing - have been automated, making it increasingly a 'black-box' technique for the average user. However, whilst much of the full theoretical explanation involves advanced mathematics, it is possible to appreciate how NMR works with a few approximations and simplified models. It is far beyond the scope of this thesis to explain the full theory behind every aspect of NMR used here, and so the enthused reader is highly recommended to read the works of Hore,[93] Keeler,[94] Levitt,[95] and Claridge.[96] However, key concepts will be introduced here.

1.2.1.1 *Larmor Precession*

In the presence of a static, external magnetic field ($\mathbf{B}_0$), nuclei precess around the orientation of this field. This precession happens even at very low magnetic fields, such as the Earth's field ($\approx$50 μT). This motion is described in terms of its angular frequency, ω, with units of rad s^{-1} (value is nucleus and field strength dependent). This is also known as Larmor precession with an associated Larmor frequency, ν, typically of the order of MHz at high magnetic field strengths. The basic principles behind this and how this allows for NMR to be detected are described in the following sections.

1.2.1.2 *Spin*

The physical property of a nucleus which allows for NMR to exist is spin angular momentum. This quantum mechanical property has no direct macroscopic analogue, and can be difficult to visualise. As per Levitt,[95] it is not necessary to understand exactly what spin is. Spin, and the inherent positive charge of a nucleus, is why the nucleus has a magnetic property. The magnetic moment, μ, of a nucleus is related to the spin, I, via the magnetogyric ratio, γ, of the nucleus according to Equation 1.1.

$$\mu = \gamma I \tag{1.1}$$

The spin of a nucleus is a non-negative value extending from 0 in half-integer increments (*i.e.* 0, $\frac{1}{2}$, 1, $\frac{3}{2}$, 2, ...). The spin is, generally, determined by the number of protons and neutrons in a nucleus. If both values are even numbers, I = 0. If both are odd, I is an integer value (I = 1 or 2 or 3...). If one is odd and the other even, I is a half-integer value (I = $\frac{1}{2}$ or $\frac{3}{2}$ or $\frac{5}{2}$ or ...). Thus, hydrogen (^{1}H), has a spin-$\frac{1}{2}$ nucleus (often simply called proton or ^{1}H), whist its second isotope, deuterium, has a spin-1 nucleus (^{2}H), as it has both an odd number of protons and neutrons (1 of each).

Nuclei with a zero spin, I = 0, have no resulting magnetic moment (as per Equation 1.1) and thus are NMR inactive. Unfortunately, some of the most ubiquitous isotopes in chemistry - including ^{12}C and ^{16}O - have zero spin. Correspondingly, nuclei with I > 0 are described as NMR active. Spin-$\frac{1}{2}$ nuclei are the most commonly studied, and most straightforward nuclei to study in liquid state NMR, however nuclei with I > $\frac{1}{2}$ (quadrupolar nuclei) are also important NMR active nuclei. Again, unfortunately, some of the other most common nuclei in chemistry have quadrupolar nuclei, such as ^{14}N.

1.2.1.3 *Resonance*

Whilst precession always (and only) occurs in the presence of a magnetic field, it is resonance which results in signal detection by NMR. Spins in the external magnetic field adopt $2I+1$ orientations between $-I$ and $+I$, which for spin-$\frac{1}{2}$ gives two spin quantum numbers $+\frac{1}{2}$ and $-\frac{1}{2}$. These yield two energy states - α and β - that are not equivalent, and have a minor difference according to Equation 1.2:

$$E_\alpha = -\frac{1}{2}\hbar\gamma\mathbf{B}_0 \qquad E_\beta = +\frac{1}{2}\hbar\gamma\mathbf{B}_0 \tag{1.2}$$

For full derivation of these relationships, please refer to aforementioned texts including those by Levitt and Keeler.[94,95] As with other spectroscopic techniques, the key observation for NMR is not the energy levels themselves, but transitions between them. In this simple case, with only two levels, the energy difference between the two allowed states corresponds to $E_\beta - E_\alpha$.

$$\Delta E = \hbar\gamma\mathbf{B}_0 = \frac{h\gamma\mathbf{B}_0}{2\pi} \tag{1.3}$$

The energy required to transition these states may be delivered via a photon of frequency ν and thus energy $h\nu$, the energy level gap resonance frequency - or Larmor frequency - is simply Equation 1.3 divided by h (Equation 1.4).

$$\nu = \frac{-\gamma\mathbf{B}_0}{2\pi} \tag{1.4}$$

Thus, the Larmor frequency, ν, in Hz is defined by Equation 1.4 and the equivalent angular frequency, ω, is defined in rad s^{-1} as per Equation 1.5.

$$\omega = -\gamma\mathbf{B}_0 \tag{1.5}$$

NMR spectrometers are often defined by their ^{1}H Larmor frequency, and a proton in a 14.1 T magnetic field would expect to have a Larmor frequency, ν, of ≈600 MHz. A table summarising several key atomic nuclei and their NMR parameters is presented in Table 1.1.

1.2.1.4 *Energy Levels*

The two energy levels for a single spin-$\frac{1}{2}$ nuclei, α and β, will have populations determined by the Boltzmann distribution (Equation 1.6):[98]

$$\frac{n_\beta}{n_\alpha} = \exp\left(\frac{-\Delta E}{k_B T}\right) \tag{1.6}$$

Recalling Equation 1.3, for ^{1}H at 14.1 T, $\Delta E = 3.98 \times 10^{-25}$ J. This vanishingly small energy difference is why NMR is such a low sensitivity technique. The ratio

Table 1.1: Summary of key atomic nuclei and their NMR properties including spin, magnetogyric ratio (γ), Larmor frequency at 14.1 T (ν), and relative natural abundance.[97]

NUCLEUS	SPIN	γ 10^7 rad T^{-1} s^{-1}	ν MHz	NATURAL ABUNDANCE %
^{1}H	$\frac{1}{2}$	26.752	600.0	99.9885
^{2}H	1	4.107	92.2	0.0115
^{12}C	0	-	-	98.93
^{13}C	$\frac{1}{2}$	6.728	151.0	1.107
^{14}N	1	1.934	43.4	99.64
^{15}N	$\frac{1}{2}$	-2.713	60.9	0.364

of spin orientations - $\frac{n_\beta}{n_\alpha}$ - is thus just 0.999904 at 300 K - meaning that (on average) just 96 spins out of every 1,000,000 are preferentially aligned with, rather than against, the orientation of the magnetic field. It is important to note that these systems are dynamic and talking about alignment with or against the field for individual spins is largely an approximation. There is a temperature dependency in the Boltzmann distribution. However, at typical probe temperatures ($-50\,°C$ to $50\,°C$) this is of limited significance.[98]

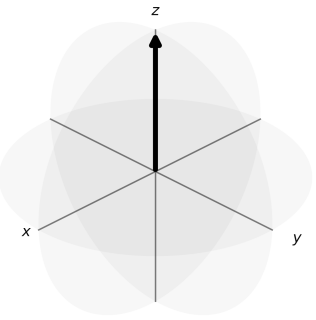


Figure 1.9: Vector model representation of bulk magnetisation.

It is inconvenient to consider every individual spin, and so instead the bulk magnetisation is considered. This is the sum of the vector spins, and is represented by the letter **M**. This is shown in Figure 1.9. By convention, the static magnetic field $\mathbf{B}_0$ is aligned along the z-axis, and at equilibrium the bulk magnetisation is $\mathbf{M}_0$ or $\mathbf{M}_{z0}$. The incidence RF pulses which are applied have a magnetic field component $\mathbf{B}_1$, and it is this which manipulates the orientations of **M**, for example pulls it into the x-y plane (also known as the transverse plane).[99]

The selection rule for observable NMR is that a spin can only change its spin quantum number by ± 1. For a single spin-$\frac{1}{2}$ system, there is only one possible transition α to β, whilst for increasing numbers of coupled spin-$\frac{1}{2}$ nuclei there are multiple transitions. These are known as quantum transitions. The energy levels for one, two, and three spin systems are shown in Figure 1.10. In the one spin system (^{1}H), the only transition is α to β, which is a single quantum transition. In the two spin system (^{1}H-^{1}H), a single α to β transition is a single quantum transition, whilst changing both ($\alpha\alpha$ to $\beta\beta$) is a double quantum transition. Flipping both without (practically) changing the overall energy - *i.e.* $\alpha\beta$ to $\beta\alpha$ - corresponds to a zero quantum transition. Transitions different from single quantum transitions are known as multiple quantum transitions. They are not directly observable in NMR, however in multidimensional experiments they can be indirectly observed after conversion to single quantum coherences, and may cause signal distortion. To remove this distortion, a number of pulse sequences may employ quantum filters, in particular zero quantum filtration as these cannot be removed by phase cycling.[100]

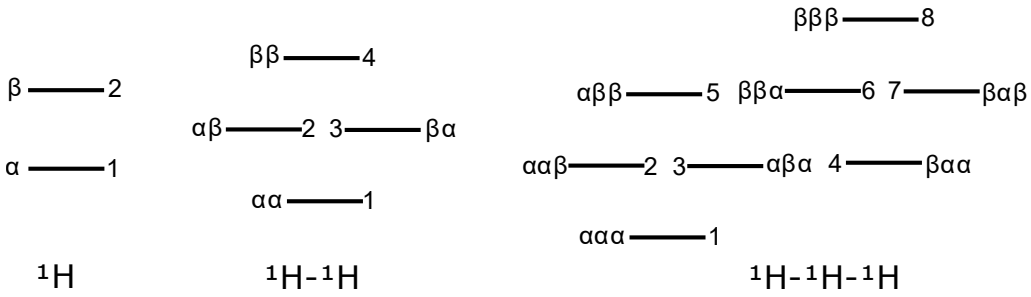


Figure 1.10: NMR energy levels for one, two, and three spin systems. Figure based on Keeler.[101]

1.2.1.5 *Chemical Shift*

NMR conveys information about molecules via a number of parameters, the most important being chemical shift and coupling. These dictate the position and multiplicity (or shape) of signals in the resulting spectrum. As previously mentioned, all nuclei will precess with a Larmor frequency dependent on the static magnetic field and their magnetogyric ratio. Proton frequencies may be at 600 MHz, for example. If no other effects were present, all protons within a sample would therefore precess at the same frequency, and the resulting spectrum would have only one signal.

In reality, the protons in a molecule may have very subtly different precession frequencies caused by local differences in the experienced magnetic field. The primary cause of these differences is the effect of orbiting electrons. Just as with

nuclei, electrons possess spin and have a charge, and thus have a magnetic moment. Their motion around a nucleus induces a secondary magnetic field which shields the nucleus from the external magnetic field, $\mathbf{B}_0$. The extent of shielding is dictated by the electron density around the nucleus, and the chemical environment.

Chemical shifts thus reflect the difference in the local environment and offset the actual precession frequency from the Larmor frequency produced by the external magnetic field. (De)shielding effects are small, and so the chemical shifts are orders of magnitude less than the Larmor frequency. At 600 MHz, a typical proton spectrum may have a range of chemical shifts covering just 6 kHz. Frequencies are magnetic field strength dependent, so chemical shifts are normalised to the the Larmor frequency by a simple division as per Equation 1.7, placing them in units of ppm.

$$\delta = 10^6 (\frac{\nu - \nu_{ref}}{\nu_{ref}}) \tag{1.7}$$

Chemical shift values for ^{1}H NMR typically range from 0 to 10 ppm. For example, a proton nucleus within a methyl group - not electron withdrawing - will be shielded and have a very low chemical shift around 1 ppm. A proton adjacent to an electron withdrawing group - for example, next to an oxygen - will be less shielded (more deshielded) and thus have a greater chemical shift, around 4 ppm. Aromatic protons may be around 6-8 ppm, and aldehydes up to 10 ppm.

Chemical shifts are affected by the solvent, temperature, sample concentration, pH, and the presence of other compounds in solution. There is no absolute correct chemical shift for individual protons, and instead spectra are often referenced to an internal standard such as 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). This is defined to be 0 ppm, which compensates for much chemical shift variability. However, in mixtures, there are compounding effects and post-processing for peak alignment may be required, as discussed later.

Chemical shifts cover a narrow range of frequencies, *e.g.* 10 ppm for ^{1}H and 200 ppm for ^{13}C. When conducting an NMR experiment, as opposed to measuring the absolute NMR frequencies - which are in MHz - and trying to resolve the very minor detail within a narrow band of the spectrum, the experiment is conducted in the 'rotating frame', whereby the spectrometer and receiver use a reference signal at a frequency very close to the Larmor frequencies in the sample (typically the centre of the spectrum). Subtraction of a reference signal from the observed signal reduces the signal frequency to just the elements of interest. This is in fact necessary for the limited sampling rates of the analog-to-digital converter (ADC) and to fulfil the Nyquist sampling theorem, discussed later.

1.2.1.6 *J-Coupling*

The second key detail in NMR spectra is information on couplings. In solution state NMR, the primary coupling to be concerned with is scalar or *J*-coupling. *J*-coupling is caused by interactions between nearby spins, and is also known as spin-spin coupling. This can be visualised by use of the energy levels from Figure 1.10.

For a single spin with no coupled partners (^{1}H), there is only a single transition, thus a single signal is observed. For the two spin system (^{1}H-^{1}H), there are four observable single quantum transitions. These are 1-2, 1-3, 2-4 and 3-4. These transitions are just two pairs of the same transitions, depending if it is spin A or B which changes state. Thus, each of A and B will have two signals, rather than just one. The energies are equivalent, and hence each nucleus will show only one signal. However, the presence of a coupling between the spins changes the absolute energy level depending on the state of the coupled spin. This causes tiny movement of the resonance frequency to the left or right (by amount of $\pm\frac{J}{2}$) and the two transitions are not equivalent any more. The result of this is a spectrum with two doublets. The doublets will be centred at the chemical shift of each nucleus, or spin, and placed around it by $\pm\frac{J}{2}$, where J is the coupling constant, typically between 1 Hz to 15 Hz for coupled protons. Heteronuclear one-bond coupling, *i.e.* between ^{1}H and ^{13}C is much larger, 125 Hz to 160 Hz. The size of coupling constants, generally, decrease as the number of bonds separating nuclei increases.

J-couplings tend to be smaller than chemical shift ranges in a spectrum, which makes practical aspects of NMR detection more straightforward. Couplings are important for multi-pulse correlated NMR experiments, as without couplings correlation is not possible. See subsection 1.2.2.

1.2.1.7 *NOE*

While *J*-couplings require spins to be coupled through bonds, the nuclear Overhauser effect (NOE) describes correlations through space. NOE is a result of cross relaxation through dipolar interactions of spins nearby in space. This cross relaxation rate is proportional to $\frac{1}{r^6}$, so NOE is a weak effect and for protons the distance limitation is approximately 5 Å. This allows the identification of spins which are nearby in space, even if they are many bonds apart - or not even within the same bonded system.

Thus, NOE can be used to determine which signals are correlated through space, such as with a Nuclear Overhauser Effect Spectroscopy (NOESY) experiment. This can be a selective 1D experiment, or more commonly a 2D experiment. In practical use, the NOE can be used for signal enhancement of one of

a coupled partner. A third, less obvious application, is an adaptation of the 1D NOESY experiment for presaturation of solvent signals. It was found that when water presaturation is combined with a three pulse NOESY sequence, excellent water suppression is achieved. The NOESY presaturation experiment is therefore common in metabolomics, and is used in this thesis for ethanol and water solvent suppression.

1.2.1.8 *Relaxation*

Alignment of the nuclear spins with the static field is not instantaneous, and it takes time for the system and bulk magnetisation to come to equilibrium. This is true for both initial insertion of the sample into the magnetic field, and after subsequent application of RF pulses ($\mathbf{B}_1$). This time is known as relaxation, and is characterised by two relaxation terms, T_1 and T_2. T_1 refers to longitudinal, or spin-lattice, relaxation, where excess energy is dissipated as heat. This relaxation describes the bulk magnetisation returning to $\mathbf{M}_0$ from a perturbed state. Typical values of T_1 for ^{1}H are 0.5 s to 5 s, and longer for ^{13}C, especially quarternary carbons. It has an exponential profile as shown in Equation 1.8.[99]

$$\mathbf{M}_z = \mathbf{M}_0(1 - exp^{\frac{-1}{T_1}}) \tag{1.8}$$

To observe a strong signal in multi-pulse NMR, it is important to allow the spins to relax. Typically, for quantitative application, delays of up to $5 \times T_1$ are used, equivalent to 99.3 % recovery after a 90° pulse. This can result in very long experiment times, and often, instead of 90° pulses, Ernst angle optimisation is used. This is calculated from Equation 1.9, and determines the optimal angle to use when fast repetition of pulses (t_r) are used. This is especially useful for very slowly relaxing species.[102]

$$\cos \alpha_e = exp^{\frac{-t_r}{T_1}} \tag{1.9}$$

T_2 is transverse, or spin-spin, relaxation. T_2 relaxation is a transfer of energy by spins mutually swapping between states - α and β - whereas T_1 relaxation was a loss of energy. It describes the effect of losing bulk magnetisation in the transverse plane ($\mathbf{M}_{xy}$) through fluctuation of the magnetic field homogenity along the z-axis as caused by molecular motion and magnetic field inhomogeneity. As a result each individual spin experiences a slightly different field, precessing with a different rate - and thus the spin vectors cancel out. The inhomogenity arises from two sources - $T_{2(\Delta\mathbf{B}_0)}$ static field inhomogenities - $\mathbf{B}_0$ will never be perfect, despite shimming - and T_2 local intra- and intermolecular interactions within

the sample. These effects are additive as per Equation 1.10, and the apparent T_2 becomes T_2^*.[99]

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{2(\Delta \mathbf{B}_0)}} \tag{1.10}$$

T_2^* causes an exponential decay in the recorded FID of an NMR experiment, and is thus directly related to the observed linewidth of the NMR spectrum as per Equation 1.11, where $\Delta\nu_{\frac{1}{2}}$ corresponds to the line full width half maximum (FWHM).

$$\Delta\nu_{\frac{1}{2}} = \frac{1}{\pi T_2^*} \tag{1.11}$$

T_2 can never be longer than T_1 (as T_1 returns all magnetisation to the z-axis), and for spin-$\frac{1}{2}$ nuclei in small molecules $T_1 \approx T_2$, whilst for larger species - *i.e.* proteins, T_2 can be much shorter.[99] Spin-$\frac{1}{2}$ have longer T_2 values than quadrupolar nuclei, and thus have much narrower lines typically.[99]

1.2.1.9 *Radiation Damping*

A significant phenomenon in concentrated samples in sensitive (high field, cryogenically equipped probe) spectrometers is radiation damping. This will be an issue in the case of raw samples of Scotch Whisky, which are predominantely protonated solvent. This effect is caused by a negative feedback of magnetisation between the sample and the probe coils. Specifically, it is caused by a coupling between transverse magnetisation of strong signals - *i.e.* solvent - and the RF coils. As the magnetisation precesses, it induces a current into the coil, which generates a magnetic field. This additional field accelerates the apparent T_1, and is known as T_{RD}.[103]

$$T_{RD} = \frac{2}{\gamma \mu_0 \eta Q M_0} \tag{1.12}$$

Equation 1.12 shows the relationships of radiation damping to physical parameters, including the magnetogyric ratio (γ), the magnetic permeability (μ_0), the 'filling factor' of the probe (η), the bulk magnetisation at equilibrium (M_0), and Q, the quality factor of the probe.[103]

$$Q = \frac{\omega L}{R} \tag{1.13}$$

Equation 1.13 shows the Q factor relates the Larmor frequency (ω), inductance (L), and resistance of the coil (R). Cryogenically cooled probes have a low

resistance, and are 'high Q factor' probes, and are therefore much more prone to radiation damping.[103]

1.2.1.10 *Saturation*

Related to equilibrium and relaxation is the concept of saturation. In an NMR experiment, signal is only observed when the net bulk magnetisation is perturbed from the z-axis and allowed to precess coherently. If the population of spins in the energy levels were equalised, there would be no net magnetisation, and thus no signal. At this point, the system is described as saturated. It will remain saturated until T_1 relaxation returns the system to equilibrium.

Saturation will occur if pulsed experiments are repeated faster than T_1 can recover the magnetisation, and is why there must be a relaxation delay between repetition of experiments. In this context, saturation is a negative factor, reducing signal intensity. However, there are scenarious where saturation - especially selective saturation - may be desirable.

Commonly in ^{1}H NMR for biological NMR, metabolomics, or whenever a protonated solvent is used in place of a deuterated one, the solvent signal can be orders of magnitude greater than the signals of interest for the solutes. For example, biological NMR and metabolomics routinely have samples in aqueous solution, and selective saturation of the protons in water allows for an increase in sensitivity. In this thesis, Scotch Whisky is studied by NMR and selective saturation of water and ethanol is necessary for acquisition of high quality NMR spectra.

Presaturation refers to the application of a long, low-power saturating pulse element during the recycle delay (d1) or other long delay within the pulse sequence, prior to acquisition. The purpose of presaturation is to saturate undesired signals so that they do not appear in the final spectrum. Presaturation equalises the numbers of nuclei at lower and higher energy levels, hence no signal is observed in principle. By presaturating the signals, they are not (net) affected by the rest of the pulse sequence, and it is relatively easy to ignore these spins in more complex pulse sequences. Presaturation, like other selective elements, is long and low powered, often lasting seconds.

Presaturation alone may induce a residual hump due to sample $\mathbf{B}_0$ and $\mathbf{B}_1$ inhomogeneity. To compensate for this, a non-selective 1D NOESY sequence may be combined with presaturation. This is routinely used in metabolomics studies, achieving excellent results for samples in aqueous solvents. The NOESY sequence used is essentially the first trace of a 2D sequence, with a short mixing time $\leq$10 ms to limit the NOE effect, taking advantage of phase cycling to suppress the distorted baseline hump.[77,104] This approach uses 90° pulses, arbitrary

angles can be implemented with FLIPSY.[105] Solvent suppression approaches, including alternatives to presaturation, are discussed later in this section.

1.2.1.11 *Pulsed Field Gradients*

Static field homogenity is important for high resolution NMR, to ensure all nuclei experience the same field and precess the same, regardless of position in the NMR tube. However, there are times when deliberate, precise magnetic field inhomogeneity can be used. This is achieved by the use of magnetic field gradients. Most commonly, these are available and applied only along the z-axis - use of x,y,z gradients allows for imaging such as in MRI. These gradients are applied for very short periods in the form of a pulse, and are commonly known as pulsed field gradient (PFG)s. When applied, $\mathbf{B}_0$ now has a dependency upon z-axis position - spins at the top of the active volume of the tube will experience a different field strength than those at the bottom.

Gradient strengths are weak compared to the static field, with those of the Bruker TCI cyroprobe capable of delivering a maximum $54\,\mathrm{G\,cm^{-1}}$, approximately 0.04 % of a 14.1 T magnet. This equates to a Larmor frequency for proton of approximately 0.24 MHz. However, this is sufficient to destroy the bulk magnetisation when it is in the transverse plane, and often this is used to filter out unwanted magnetisation, for example in the Trippleton z-filter for elimination of zero quantum interference.[100]

1.2.1.12 *Diffusion*

An interesting phenomenon in NMR is related to molecular motion within the sample. In an NMR spectrometer, within the probe, there exists a finite active volume of field homogeneity and coils for signal detection. NMR measures, directly, the precession of nuclear spins. However, by application of magnetic field gradients to the sample, it is possible to indirectly measure the *molecular* motion within the sample. Diffusion simply describes the translational motion of nuclei within the sample. In routine spectroscopy, the gradients are applied along the z-axis, and thus only vertical motion is detected - lateral motion is not detected.

The simple diffusion ordered spectroscopy (DOSY) experiment relies on a pair of PFGs with duration (δ or little delta) separated by a delay (Δ or big delta). The first PFG defocuses all magnetisation along the height of the sample. The subsequent delay allows molecules to diffuse within the sample, and the second PFG refocuses the spins by means of an identical but opposite phased pulse. Spins which did not diffuse (at least not vertically) will be perfectly refocused. Any

spins which have diffused will not be refocused, and the signals will attenuate accordingly.

1.2.1.13 *NMR Measurement*

The Nyquist sampling theorem states that the sampling rate of an analogue signal must be twice the highest frequency to be digitised. If an NMR experiment was conducted without subtraction of a reference signal, an ADC with a sampling frequency of 1.2 GHz or greater would be required. By using a rotating frame of reference, the required sampling frequency may be as low as 12 kHz. Given audio CDs sample at 44.1 kHz, hardware of this specification is readily available.

In Fourier transform (FT) NMR, the reference frequency is typically set to the centre of the spectrum. Commonly for samples in water, solvent suppression of the water may be performed. As water ($\approx$4.7 ppm) is close to the centre of the spectrum, the frequency may be set to this for presaturation. For a 10 ppm spectral width, this extends from approximately 0 to 10 ppm, covering a 6000 Hz frequency range. Because the reference frequency is subtracted, it is important to note that the measured frequencies are therefore from -3000 Hz to 3000 Hz. Therefore the highest frequency which needs to be measured is just two times 3000 Hz.

Three instrumental settings are connected to define the acquisition parameters of an NMR spectrum. These are the acquisition time, the spectral (or sweep) width, and the time domain. The acquisition time is how long the resulting signal is measured - typically from 10 ms to 10 s, and dictates the resolution of the spectrum.[106] The time domain describes how many points are measured within the acquisition time - *i.e.* how many digital points are used to describe the analogue signal. The distance between points in the time domain dictates the third parameter, the spectral width, as per the Nyquist sampling theorem. The time between points, or the frequency of the sampling, defines the spectral width. More time domain points will result in faster sampling, and a wider spectral width. Because of digitiser limitations with regards to sampling rates and memory buffers, all three parameters interact with each other.

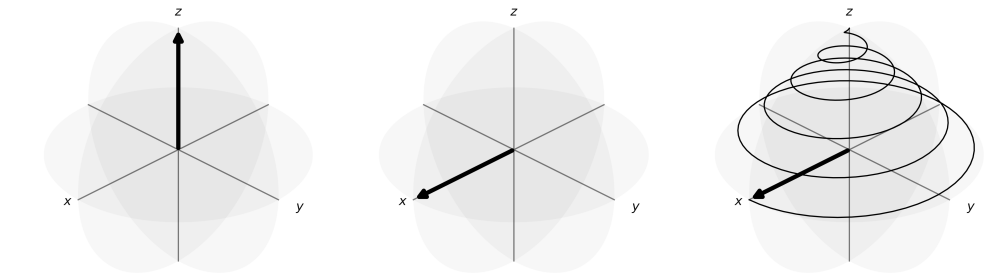
The Fourier transformation transforms time domain data into the frequency domain. Put simply, it de-constructs a time domain signal into a sum of sinusoidal waves from $-\infty$ to $+\infty$ frequencies. The mathematics and computer science behind the fast Fourier transform are beyond the scope of this thesis, but it is important to appreciate the impact the acquisition parameters have on the spectrum. Where the time domain data is long - the signal has been measured for a long time - it is possible to more precisely deconvolve that into the sum of its frequencies. Where it is short, there is more uncertainty. This is described by the

Gabor limit, that states it is impossible for a function to have both a narrow time domain and narrow frequency domain.[106] This is ubiquitous in signal processing, especially with regards to the Fourier transform, and is somewhat analogous to the Heisenberg uncertainty principle. Thus, for high resolution NMR spectra, acquisition times of many seconds are required. If the signal decays too fast, *e.g.* it has a short T_2, the Fourier transformed line will be correspondingly broad.

The NMR signal is measured as the nuclei, possesing a magnetic moment, induce a current into the receiver coils. The receivers are in the transverse plane (x,y-axis), and do not detect signal from nuclei at z-axis. The observed NMR signal follows an exponential decay characterised by a rate constant $R_2^* = \frac{1}{T_2}$. If the measurement is made on-resonance - *i.e.* there is one signal which is at the exact reference frequency - the resulting free induction decay (FID) follows a monoexponential decay. When the signal is off-resonance, the signal is an exponentially damped sine wave, as a result of precession during relaxation away from the transverse plane. Pictorially, this is represented in Figure 1.11. This figure shows the bulk magnetisation at equilibrium aligned with z-axis. In Figure 1.11b, the magnetisation has been rotated 90° into the transverse plane. Figure 1.11c shows the path taken by precession and relaxation back to z-axis. Figure 1.11d shows the measured FID signal by two orthogonal channels, which can be Fourier transformed to produce an NMR spectrum.

1.2.1.14 *NMR Dimensions*

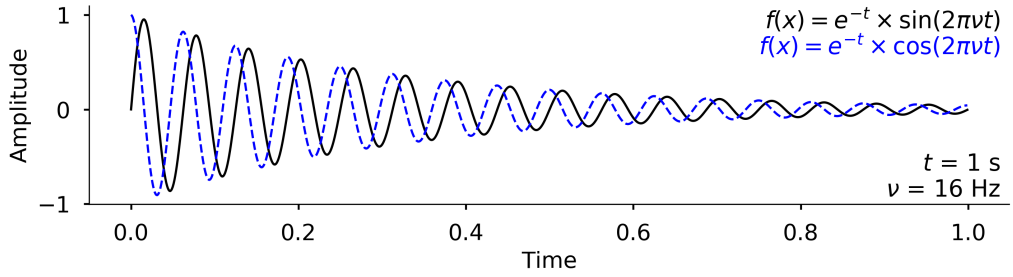
NMR experiments can have one time domain, and be one dimensional (1D), or have multiple time domains, and be multidimensional (2D, 3D, ..., nD). In a simple 1D ^{1}H NMR experiment, there is a delay, a 90° pulse, and measurement of the FID into one time domain (t_1). In a 2D experiment, for example ^{1}H-^{1}H COSY, there is a second time domain which is sampled incrementally during the pulse sequence in the form of a variable delay. This first time domain is labelled t_1, and is indirectly measured, whilst the directly measured FID becomes t_2. This convention continues for 3D experiments, where the two indirect dimensions are variable delays sampled across t_1 and t_2, whilst t_3 is the directly detected dimension. Accordingly, each additional dimension requires a significantly longer experimental time. Multi-dimensional NMR experiments are often acquired with far fewer time domain points than 1D NMR experiments in each indirect dimension, however the number of sampled points can still become very large - for example a 3D experiment with $256 \times 96 \times 96$ points (t_3, t_2, and t_1, respectively) will require 9216 repetitions of the experiment to sample all the indirect points. There are approaches around this, including the use of non-uniform sampling (NUS) whereby not every indirect point is sampled, and instead of a direct Fourier trans-



(a) Bulk magnetisation at equilibrium

(b) After 90° pulse into transverse plane

(c) Precession during relaxation



(d) Measured FID

Figure 1.11: Vector representation of simple 90° pulse, relaxation and precession, and measured FID. Note the FID samples both the x and y components (sine and cosine, respectively) of the signal, as detected with quadrature detection, allowing for phase detection.

form, the missing points are calculated using an alternative algorithm, for example iterative soft thresholding (IST). This can achieve nearly identical results in a fraction of the required experimental time, allowing for greater resolution in high dimension NMR than would normally be possible.

1.2.1.15 *Selective Pulses*

RF pulses in FT NMR take several forms. The most common is a hard, short pulse. This is used for broadband spin manipulation, for example a 30° or 90° rotation of all spins into the transverse plane. This pulse is high powered and very short (*ca.* 8 µs), and thus covers a large frequency range, easily 10 kHz or greater. The hard pulse power is often calibrated and not adjusted, and pulse angle is thus manipulated by changing pulse length - *e.g.* a 180° pulse lasts twice as long as a 90° pulse.

There are times when a more careful approach is necessary, and only specific spins (or frequencies) are intended to be manipulated. This is made possible through selective pulses. To achieve selectivity, *i.e.* a narrow band of frequencies excited, the pulse must be correspondingly longer, often in the order of 1 ms

to 100 ms or even more. In an opposite fashion to hard pulses, where angle is dictated by pulse length and power is not modulated, in a selective pulse the selectivity is dictated by the length, and the pulse angle is defined by the power level. Power levels in selective pulses are much lower, as they last much longer.

The selective pulse is often 'shaped', for example by a Gaussian envelope. This gives a band of excited frequencies according to a Gaussian curve. Other shapes achieve different excitation profiles. The reason for shaping the RF pulse is two-fold - to remove the 'sinc' type wiggles produced by selective rectangular pulses that affect the spins outside the intended excitation windows, and to produce a 'top-hat' profile. The latter reasons leads to longer pulses. In some cases, multiple frequencies may need to be excited simultaneously - *i.e.* for multiple solvent suppression - and in this case the shaped pulse is frequency modulated to allow this.

1.2.2 *NMR Experiments*

All FT NMR experiments rely on a specific series of RF pulses and delays to correctly manipulate the spins. These pulses and delays can be used in many different combinations, generating an entire suite or tool-box of NMR experiments, allowing for the characterisation of chemical structures. It is well beyond the scope of this thesis to explain in more detail how each experiment works. Indeed, this thesis does not develop any fundamentally new NMR pulse sequences. Instead, a very brief explanation of the purpose and interpretation of the most important NMR experiments used in this thesis is presented.

1.2.2.1 *1D ^1H*

The standard 1D ^{1}H NMR sequence involves a recycle delay (d1) which allows the system to return to equilibrium, followed by an excitation pulse (p1) of typically 30° or 90°, and measurement of the FID (t_1). As with all NMR experiments, this may be repeated a number of times; co-adding FIDs increases the observed signal-to-noise ratio (SNR). A ubiquitous misnomer is to refer to this as the 'number of scans' (ns), as FT spectroscopy does not 'scan', and the term refers back to continuous wave spectroscopy. Instead, number of transients is preferred, however even the author is guilty of referring to scans. This process works because signal increases linearly with the number of co-added transients, whilst noise is random and increases according to $\sqrt{ns}$. Thus, to double the SNR involves acquiring four-times as many transients (SNR $\propto \frac{ns}{\sqrt{ns}}$).

1.2.2.2 *1D ^{13}C*

Acquisition of 1D ^{13}C spectra follows essentially the same principles as acquisition of 1D ^{1}H spectra, except with the routine use of proton decoupling. This can take three forms - power gated decoupling, gated decoupling, and inverse gated decoupling. The difference in practice is when the decoupling is applied. In power gated decoupling, the decoupling element is applied throughout d1 and acquisition. This serves two purposes - decoupling during acquisition removes the ^{13}C-^{1}H coupling, turning each signal into a singlet (at natural abundance ^{13}C, there are statistically very few coupled ^{13}C-^{13}C nuclei).

The second purpose is to benefit from signal enhancement through the NOE. By continuously irradiating the sensitive nuclei (^{1}H), the system is inherently not in equilibrium. Relaxation is then only possible via dipole-dipole coupling, through space to nearby spins. The lower sensitivity nuclei is enhanced in sensitivity by a factor (ε_{NOE}) determined by the relative magnetogyric ratios of the nuclei. This is approximately calculated for fast motion (small molecules) as per Equation 1.14.[107]

$$\varepsilon_{NOE} = 1 + \frac{\gamma_I}{2\gamma_S} \tag{1.14}$$

In the case of the magnetogyric ratios of ^{1}H and ^{13}C, with γ_I (^{1}H) approximately four times greater than that of carbon (γ_S), the maximum net boost is approximately 3×. Molecular motion also plays a role in the NOE, and large and small molecules may have positive and negative NOEs. Unfortunately, this enhancement reduces the quantitative nature of the carbon spectrum - methyl carbons will get a greater boost than quaternary carbons, and is not always beneficial.

It is possible to gain this signal enhancement and retain coupling information where this is necessary, by decoupling only during the relaxation delay (d1). This is known as gated decoupling. Where the opposite is true - simplification of the spectrum is required, but NOE enhancement is not desired - inverse gated decoupling can be used, where the decoupling is applied only during acquisition. This may be for quantitative carbon experiments, or for spin of opposite sign magnetogyric ratio (*e.g.* ^{1}H and ^{15}N) which would result in a reduced signal via this mechanism ($\varepsilon_{NOE} = 1 + \frac{267}{2\times-27} = -4\times$).

1.2.2.3 *Homocorrelated NMR*

Homocorrelated 2D ^{1}H-^{1}H Correlation Spectroscopy (COSY) provides a spectrum showing correlations between *J*-coupled protons. Routinely it is used to de-

termine adjacent protons, but can be used for long range (LR-COSY) correlations or relayed correlations (RELAY-COSY).[108]

Homocorrelated 2D ^{1}H-^{1}H Total Correlation Spectroscopy (TOCSY) provides a spectrum showing correlations between *J*-coupled protons belonging to one spin system. It is very similar to that of COSY, except showing multiple correlations mediated by a chain of *J*-couplings. In a TOCSY experiment, the duration of the spin-lock (mixing) period dictates how far magnetisation can be transferred, and thus how many bonds can correlations be observed for - weaker correlations with smaller couplings will require a longer mixing time, up to 160 ms. This mixing time dependency can be used to determine relative connectivity of correlated spins, by acquiring a series of 2D TOCSY spectra with increasing mixing times, in a so-called 'Taylored' approach.[109]

TOCSY also finds use in selective 1D experiments, where excitation of a single signal, followed by the TOCSY spin-lock period, allows observations of correlations between spins in a much faster, higher resolution experiment. The drawbacks of this method are requiring a high-selectivity, which can be achieved by means of a chemical shift-selective filter (CSSF),[110] and the need to probe each signal of interest separately. However, where only a few signals are of interest, this can represent a substantial time saving compared with acquiring a full 2D spectrum. It is important to note that TOCSY transfers require an uninterrupted spin-system, that is, continuous network of *J*-coupled protons. If this is interrupted by a heteroatom (*i.e.* oxygen in an ether bond) or a quaternary carbon, the magnetisation will not transfer.

NOESY is commonly used in 1D and 2D experiments to determine the relationships between spins through space, via the NOE. Homocorrelated 2D NOESY spectra produce a strong autocorrelated diagonal through the spectrum, which can induce significant t_1 noise (or ridges, vertical artefacts through the spectrum from strong signals). Various efforts have been reported to reduce the magnitude of this noise.[111] The size of the NOE can be very small, and so NOE cross-peaks in the 2D spectra can be difficult to observe. Additionally, field strength and molecular size impact on the magnitude (and the sign) of the NOE, and at high fields, NOEs for medium sized molecules can approach zero.

1.2.2.4 *Heterocorrelated NMR*

In terms of heterocorrelated NMR experiments, one of the most important is 2D ^{1}H-^{13}C Heteronuclear Single Quantum Coherence (HSQC), which provides information on ^{1}H-^{13}C bonded pairs. It is a sensitive experiment, using the Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) element to boost sensitivity by transferring the polarisation from the more sensitive ^{1}H to ^{13}C prior to evolu-

tion and mixing, and then transferring the magnetisation back to ^{1}H for detection. As this experiment starts on proton and is detected on proton, only carbons directly bonded to a proton are detected. Specifically, only ^{1}H nuclei attached to ^{13}C nuclei are observed, which makes this experiment much less sensitive than the homocorrelated ones. Thus, quaternary carbons are not observed.

2D ^{1}H-^{13}C HSQC-TOCSY is a combination of the two sequences to provide correlations between protons and carbons multiple bonds apart, within the same spin system. As with the TOCSY experiment, choice of spin-lock mixing time is crucial. As the TOCSY element only allows correlations within spin-systems, again, quarternary carbons are not observed.

2D ^{1}H-^{13}C Heteronuclear Multiple Bond Correlation (HMBC) is a similar experiment to HSQC-TOCSY, allowing observation of multiple bond correlations. However, it does not require all protons to be in a spin system, and can observe quarternary carbons as well as 'go through them' and heteroatoms. Thus, it can be used to connect spin-systems identified by other experiments, for example bridging the glycosidic bond in a disaccharide. HMBC cross-peaks are mixed phased, and ^{3}J couplings are related to the dihedral angle between the proton and carbon according to the Karplus relationship (Equation 1.15). This means that for dihedral angles close to 90°, ^{3}J is close to zero, and the HMBC cross-peak may not be observed.

$$J(\phi) = C\cos 2\phi + B\cos\phi + A \tag{1.15}$$

1.2.2.5 *DOSY*

Diffusion was introduced previously, and DOSY is the most common application of diffusion to NMR spectroscopy. In a 2D experiment, a series of 1D spectra are acquired with increasing PFG strength. This causes a gradual decrease in signal intensity. The signal attenuation is related to the diffusion coefficient by the following, Equation 1.16.

$$I = I_0 \exp^{-D\gamma^2 g^2 \delta^2 (\Delta - \frac{\delta}{3})} \tag{1.16}$$

In Equation 1.16, I is the measured intensity and I_0 the unattenuated intensity, D is the diffusion coefficient, γ the magnetogyric ratio of the nucleus, δ the duration of the PFG, g the PFG strength, and Δ the duration of the diffusion delay. By fitting an exponential function to the decaying signals across the arrayed 1D spectra, it is possible to calculate D.

D can be related to the hydrodynamic radius (r_s) of the molecule as per the Stokes-Einstein equation, Equation 1.17.

$$D = \frac{k_B T}{6\pi\eta r_s} \tag{1.17}$$

Here, k_B is the Boltzmann constant, T is temperature, η is the viscosity of the solvent. This assumes the molecules are spherical, and that there is no convection.[112] It is possible, however, to approximate molecular size from this, and so DOSY finds application in the analysis of mixtures routinely. If two signals overlap from species with different diffusion coefficients, it will fail to fit an exponential accurately. Various means to compensate for this post-acquisition exist, including fitting multiple exponential functions. The best solution, however, is to avoid overlapping signals as far as possible. DOSY has been used for study of food supplements,[113] plant extracts,[114] biofluids,[115] and beverages,[116] including Port wine.[117]

1.2.3 *Solvent Suppression Techniques*

The acquisition of ^{1}H NMR spectra of aqueous or alcoholic samples is severely hindered by the high concentrations of solvent - water and / or ethanol. In the case of Scotch Whisky, the vast majority of the sample is water and ethanol, protonated solvents. ^{1}H and ^{13}C NMR spectra are highly sensitive to this. NMR signals are sampled within the dynamic range (typically 12-16 bits) of the ADC, which describes the range of signal amplitudes which can be simultaneously measured. The signal to be digitised should ideally use the full range of these bits to maximise integrity of digitisation. As the NMR signal is very weak, it first goes through a pre-amplifier which 'boosts' the signal according to a value set as the receiver gain (RG). This value must be chosen to ensure the signal going to the ADC uses the full range of bits; too weak a signal, not using the full range of bits, will have a greater noise level, whilst too much gain will cause a receiver 'overflow' and clipping of the FID. The dynamic range of an NMR spectrum is therefore finite, and when a sample contains a large range of concentrations of compounds, the gain and dynamic range will be limited by the most abundant species - in this case, the solvent signal.

One practical means to overcome this issue would be to remove the solvent, via rotary evaporation or lyophilisation. Subsequent re-dissolution of the sample into a deuterated solvent would allow spectra to be acquired without signal suppression. However, this has obvious drawbacks - volatile congeners would be lost, and drying aqueous samples can be a time consuming and laborious process. Beyond loss of volatiles, drying may induce other chemical changes within

the sample, too. It is preferable to be able to analyse a sample such as Scotch Whisky in its raw state.

Whilst the use of an ADC with greater dynamic range would somewhat resolve the problem of strong solvent signals, Delsuc and Lallemand reported a means to overcome this issue on existing hardware by improving the dynamic range of the NMR signal through oversampling.[118] Typically, sampling rates are chosen as the minimum of the Nyquist theorem - twice the highest frequency to be measured. Delsuc and Lallemand found that an oversampling factor of n yielded a gain in dynamic range of $\log_2(n)$ bits - *i.e.* for an oversampling factor of 8, there was a gain of 3 bits in dynamic range. Modern spectrometers use digital receivers and always sample at the maximum digitiser rate to benefit from this boost in dynamic range.

The most common means to reduce the problems associated with strong solvent signals in NMR spectroscopy is therefore to attenuate them prior to acquisition. Zheng and Price published an extensive review of solvent suppression in NMR in 2010, to which the interested reader is referred.[103] Claridge describes three main classes of solvent suppression techniques.[104] These are based on saturating the undesired signal, zero excitation of it, or PFG 'destruction' of the solvent signal.[104] Earlier in this section, presaturation was introduced, especially with regards to a method based on the presaturation of solvents at the beginning of the 1D NOESY pulse sequence.[77] This method is widely used in metabolomics studies, and an excellent review of its working principles was presented by Mckay in 2011.[77] Monokhova *et al.* adapted the sequence for automated suppression of water and ethanol signals in a single-channel approach, using a multiple-frequency selective saturation pulse. This pulse had encoded eight (the OH singlet, the ethanol triplet, and quartet) frequencies in a 'line-selective' approach. Despite this name, they are not perfectly line-selective methods. As demonstrated by Kupče *et al.*, when shaped pulses targeting very close frequencies are combined in a single selective pulse, they do not work in a line-selective manner; instead, they will saturate a broader region.[119] In practice this may not matter much for the frequency range within each multiplet, but the spread of irradiation between multiplets may attenuate signals of interest. A similar method was implemented by Ragone *et al.* to examine wine samples.[120]

Zero excitation schemes include so called 'jump-return', where the spectrometer is on-resonance for the solvent signal. A 90° pulse pulls all magnetisation into the transverse plane. Off-resonance spins (*i.e.* all but the solvent) precess and 'fan-out' over the course of a short delay. A subsequent opposite 90° pulse 'returns' the solvent magnetisation, which has not evolved within the rotating frame, back to the z-axis.[104] This method is imperfect and significant solvent sig-

nal remains, and more advanced versions have been developed, including binomial sequences.[103,104]

Finally, PFG methods include excitation sculpting,[121] which has proven to be a very efficient water suppression technique and been adopted for the suppression of multiple signals.[122] This is not perfect, and the selectivity of this approach is not optimal, attenuating a wider range of signals than for a presaturation based approach. Longer pulses, to increase selectivity, result in undesired *J*-modulation of multiplets. This latter problem can be addressed by the perfect-echo pulse sequence,[123] however this can induce its own problems with quantification as signals will spend longer in the transverse plane and may relax more prior to acquisition. Excitation sculpting has also been coupled with DOSY for water suppression in complex samples.[115] Water suppression Enhanced through T1 effects (WET)[124,125] suppresses water at the start of the sequence. It has been used successfully for acquisition of 1D and 2D spectra of grape pomace distilled spirits.[126] This study was performed using a room temperature probe, where WET performs well. Nevertheless, the efficiency of WET is known to deteriorate on cryoprobe instruments and a modification has been put forward that improves its performance.[127]

Note that certain sequences do not require solvent suppression. For example, a highly selective 1D TOCSY experiment might never excite the solvent resonances. Heterocorrelated experiments such as HSQC require magnetisation to transfer between ^{1}H and an X nucleus, such as ^{13}C. The ^{1}H-^{13}C magnetisation is then selected by PFGs or phase cycling, eliminating protons not attached to the ^{13}C, *i.e.* protons in water molecules. This will naturally filter out water (H_2O), but will not suppress organic solvents like ethanol (CH_3CH_2OH). Some 2D homocorrelated experiments filter out singlets (uncoupled spins, *i.e.* water), such as the double quantum filtered (DQF)-COSY, however this still will not suppress ethanol.[104]

1.3 FOURIER TRANSFORM MASS SPECTROMETRY

FTMS represents the highest possible resolving power and mass accuracy of any mass spectrometry technique.[128] There are three main classes of FTMS - Ion Cyclotron Resonance (ICR), Orbitrap, and linear ion traps.[128,129] The latter two are not used in this thesis, so will not be discussed further. Regardless, ICR has the highest possible mass accuracy and resolving power of any of these techniques. For a more detailed discussion on the theory and application of FTICR, there are excellent reviews by Marshall *et al.* from 1998,[130] 2002,[131] and a more recent one by Cho *et al.* from 2014.[132]

1.3.1 *Instrumentation*

An FTICR instrument contains many important parts, most significant of which is the ICR cell, discussed in the next section. However, there are additional required components. These include the superconducting magnet - FTICR can be performed at low fields, but higher fields yield higher performance. ICR performs best under ultra high vacuum, but routinely atmospheric pressure ionisation (API) sources are used - such as ESI. As such, a big part of the instrument consists of sections designed to direct ions and reduce pressure in stages. A simplified example schematic of an FTICR instrument is shown Figure 1.12.

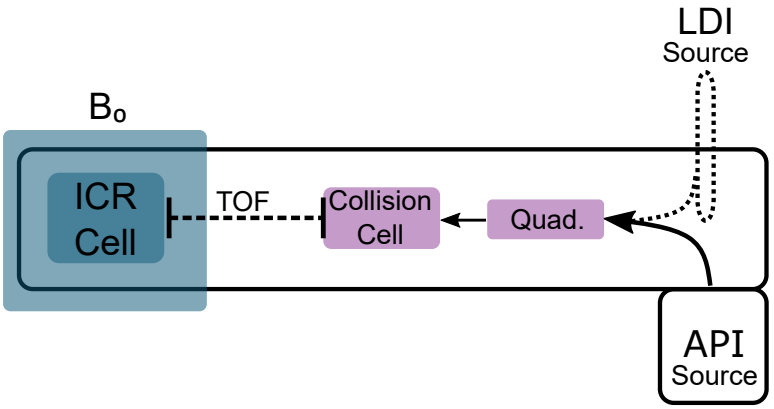


Figure 1.12: Simplified schematic of a FTICR instrument. From right to left shows the 'source', both the API and laser desorption ionisation (LDI) sources. Ions are directed through a funnel to the quadrupole (Quad.) followed by the collision cell. Packets of ions are then transmitted to the ICR cell with a predefined time of flight (TOF). The magnetic field ($\mathbf{B}_0$) is indicated.

The schematic shown in Figure 1.12 is based on the Bruker Solarix instrument, albeit highly simplified. The key components are shown, starting with the API source which is, obviously, at atmospheric pressure. This could be ESI, atmospheric

pressure photoionisation (APPI), or atmospheric pressure chemical ionisation (APCI), and generates ions from the solution state into the gas phase. Alternatively, a solid state source such as LDI can be used. This is used under vacuum, so the access chamber can be brought up to atmospheric pressure to change samples, and pumped down to vacuum again. The ions are funnelled into the first major component, the quadrupole. This quadrupole can act as a mass filter, removing ions below a threshold (Q1 mass) or within a specific range (isolating masses). At best, it has nominal mass resolution for filtering of ions.

After the quadrupole, ions are sent to the collision cell. This name is misleading, as it does not have to be operated in a collision manner. Instead, ions are accumulated here for a period of time (the ion accumulation time (IAT)). Ions can be fragmented here, if required, typically after isolation of single masses with the quadrupole. The IAT is an important parameter which, along with the original sample, dictates how many ions will be in the cell. Too many ions will lead to complications, discussed later, whilst too few may not yield optimal signal. Furthermore, ions are not stable indefinitely, and longer IAT can yield different spectra for complex mixtures, including natural organic matter (NOM).[133]

The accumulated ion packet is then sent into the ICR cell, which is within the centre of the static magnetic field ($\mathbf{B}_0$). A predefined TOF is set - only ions which travel the distance between cells within this TOF will stay within the ICR cell. This TOF parameter is thus another way to filter ion mass ranges - higher mass species will travel slower, and require a longer TOF. For small molecules, a TOF of 500 ms to 700 ms is typical here.

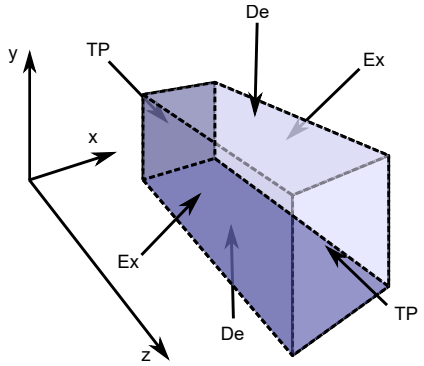


Figure 1.13: Simplified schematic of a cubic ICR cell showing the orientation of the trapping plates (TP), detection plates (De) and excitation plates (Ex)

A simple 3D schematic of a ICR cell is shown in Figure 1.13. Note that this shows a simple cubic cell, whilst more advanced cells exist in more exotic geometries, and the cell on the Solarix instrument is an 'infinity' cell, with a cylindrical shape. The infinity cell also has only two excitation (ex) and two detection (de) plates, so this approximation is sufficient. Ions are held within the ICR cell by

means of trapping plates (TP) which each have an applied voltage to repel ions. The magnetic field is oriented along the z-axis, and ions oscillate with cyclotron motion around this.

1.3.2 *ICR*

Superficially, FTICR MS is not dissimilar to FT NMR. Particles with charge (ions in MS, nuclei in NMR), have an oscillating motion within a magnetic field, which induces a current in a detector. The recorded sine wave is Fourier transformed to produce a frequency domain spectrum. Again, multiple transients (or the misnomer, 'scans') are co-added to increase the SNR. This approximation lends itself to a nice visualisation, however, there are several key differences, one being that after Fourier transformation, the mass spectrum has to be converted to another domain - mass (or more specifically mass-to-charge, m/z). The general workflow of an FTICR experiment is shown graphically in Figure 1.14, and explained in the following sections.

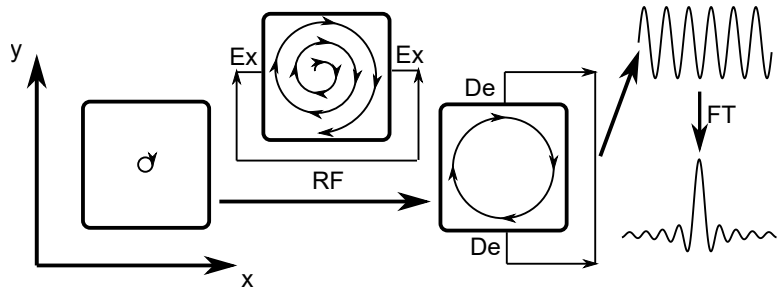


Figure 1.14: Workflow of the FTICR experiment showing the inital cyclotron motion of ground state ions, application of an RF pulse excites them to a coherent larger radius. Detection of this cyclotron motion results in a sine wave transient, which can be Fourier transformed to yield a frequency domain spectrum.

The key phenomenon here is ICR. Where NMR relied on precession, FTICR relies on cyclotron motion. A charged particle (ion) in a static magnetic field will oscillate with cyclotron motion. Essentially, it will spin around within the field. This motion is caused by the Lorentz force. The angular frequency (ω) of this motion is determined by Equation 1.18. For full derivation of the cyclotron motion, refer to Marshall *et al.*'s primer of 1998.[130]

$$\omega = 2\pi f = \frac{zeB}{m} \tag{1.18}$$

The angular cyclotron frequency (ω, rad s^{-1}) is related to the field strength (B), the mass of the ion (m), the elementary charge (e), and the number of charges

(z). The frequency (f, Hz) is then, at a fixed field strength, dictated by the mass-to-charge ratio (m/z) of the ion, as shown in Equation 1.19:

$$\frac{m}{z} = \frac{eB}{2\pi f} \tag{1.19}$$

As frequencies can be determined with high accuracy and precision, by means of Fourier transformation of time domain data, FTICR provides unparalleled mass resolution and accuracy. As with NMR, the magnetic field strength is not arbitrary, and high fields are required. The instrument in Edinburgh is equipped with a 12 T magnet - roughly equivalent to the field strength of a 500 MHz ^{1}H NMR spectrometer. The most powerful FTICR magnets in the world are now 21 T, equivalent to a roughly 900 MHz ^{1}H NMR spectrometer.

1.3.2.1 *ICR Detection*

One interesting point to note about FTICR measurement is the frequency range. In NMR, it is common to have Larmor precession frequencies in the millions of Hertz. In FTICR, the frequencies are smaller but span a larger range in a mass spectrum. For example, an ion at m/z 100 in a 12 T field (noting mass in kg in Equation 1.20 and a charge of 1) will oscillate with a frequency of 1.83 MHz. However, ions at m/z 400 and 800 will oscillate much slower, at just 458 kHz and 230 kHz, respectively.

$$f = \frac{eB}{2\pi m} \tag{1.20}$$

In NMR, measurements were made in a context of the rotating frame - the measured signal was subtracted from a reference frequency of the spectrum, giving small frequency ranges to be measured. In broadband FTICR, this is not feasible. As such, the absolute frequencies of FTICR signals are measured, which requires larger datasets and more advanced digitisers.

As mentioned in the NMR section, the Nyquist sampling theorem requires a sampling rate of at least twice that of the highest frequency to be measured. In FTICR, the highest frequency will be lowest mass to be measured in the spectrum. The same three parameters as in NMR are in play in FTICR - acquisition time or transient length, the time domain size, and the frequency range to be sampled (spectral width in NMR, low mass limit in FTICR). The acquisition times are comparable to high resolution NMR, lasting several seconds for ultra high resolution. The time domain size is reported in 'words' (W), a computer science term (for modern 64-bit hardware, a word contains 64-bits). Typical time domains (TD) are on the order of 4 MW, four million words is equivalent to 32 MBytes.

It is not possible with the Bruker Solarix instruments to manually define the acquisition time, and instead transient length - and thus resolution - is controlled by altering the time domain and the low mass limit. Increasing the time domain increases the number of points sampled, and for a fixed sampling rate, this increases the acquisition length. Increasing the low mass limit (say from m/z 100 to 200) will also increase the transient length. There is a finite amount of memory within the ADC - a faster sampling rate will fill this memory faster. To sample low masses, the transient length is accordingly reduced - or the time domain reduced. By increasing the low mass limit, the memory does not fill as fast, and so the transient can be recorded for longer. These hardware limitations are increasingly an issue with higher ICR frequencies - for example with higher fields (21 T), and with 2ω detection (see later). In fact, recent work at 21 T using 4ω detection could not sample ions with a m/z below 600 due to the frequency being too high.[134]

When broadband acquisition is not necessary, ultra high resolution can be achieved through narrowband acquisition using what is known in mass spectrometry as 'heterodyne' detection. This is essentially what is performed in NMR - the recorded signal is subtracted from a reference frequency, and thus lower frequencies need to be sampled, allowing for an increasing resolution. Heterodyne detection is not used routinely in complex mixture analysis, and has not been used in this thesis.

1.3.2.2 *ICR Complications*

Whilst ions do oscillate with cyclotron frequency within the ICR cell, they do so with a very small radius of motion. This makes it more challenging to measure the signal. As such, ions are excited to a higher radius orbit prior to acquisition. This is the origin of the resonance aspect of FTICR, and another analogous aspect to NMR. However, FTICR has far fewer pulse sequences and experiments. In routine broadband FTICR, the pulses and pulse sequences are seldom adjusted. The default pulse sequence is very similar to that of a single pulse NMR sequence - a delay at the start, followed by an excitation pulse and recording of the FID as the ions induce an image current within the detectors.[130] This was shown previously in Figure 1.14.

The equilibrium energy state has ions at a small orbit radius. Excitation can increase this to a higher radius. The larger radius is closer to the detector plates of the FTICR cell and induces a much stronger current. FTICR is performed under ultra high vacuum, and where the ions have a coherent movement (a so called 'packet' or 'ion cloud'), they can continue oscillating at this higher energy state for some time. Collisions allow for the energy to dissipate, which re-

duces the intensity of the FID. There is no equivalent to T_1 or T_2 relaxation in FTICR, which has the effect that an ideal, theoretical FTICR transient will be a sinusoidal wave with no exponential damping. This would allow for much longer transient lengths to be achieved, however instability with ion clouds and imperfect vacuums cause signal decay over time. Unfortunately, this signal decay both broadens and distorts the lineshapes. Lineshapes can be 'split' and severely degraded when signal decay is too rapid, and the data become useless. As such, an optimal transient length exists depending on ion cloud stability and vacuum quality. Furthermore, some ions are more stable than others - longer transients may decrease sensitivity if ions are not stable for the length of the transient, and varying profiles may be achieved if different ions have different stabilities. Magnetic field inhomogenity will negatively affect measurement, however the requirements for homogenity are not as strict as for solution state NMR. The new generation of dynamically harmonised ICR cells have electrodes which require 'shimming' to correct the field within the cell.[135] In the ICR cell used in this thesis, shimming is not an option.

The non-decaying FID requires substantial apodisation to avoid 'sinc' wiggles (also known as side lobes) - Fourier transform artefacts for a signal which has not decayed to 0 at the end. This effect is evident in Figure 1.14. Interestingly, Kanawati *et al.*,[136] observed that the side lobes observed with the Fourier transform of a non-fully decayed signal in ICR exhibit a drastically different resolution than those of genuine signals. They proposed a simple approach to filter out signals with resolving powers substantially greater than theoretically predicted as per Equation 1.22. This was only published in 2017, showing there are still plenty of simple improvements which can be made for ICR. Due to its late development, it is not implemented in this thesis. In NMR, apodisation and linear prediction are often combined to increase resolution and reduce 'sinc' wiggles. Linear prediction is not utilised routinely in FTICR, presumably for computational reasons. Linear prediction Cholesky decomposition was compared to Fourier transformation in ICR analysis in 1997 by Guan and Marshall, requiring hours of computational time to process 8k time domain points.[137] Computational power has grown significantly since 1997, and so this may be an area worth revisiting.

In NMR, broadband pulses are commonly hard, short pulses which excite a large frequency range. As FTICR covers much larger frequency ranges, a frequency swept excitation pulse must be used. This 'CHIRP' excitation is also known as an adiabatic pulse in NMR, and allows for irradiation over very large frequency ranges. As with selective pulses in NMR, it is the power of the pulse which is adjusted to change the degree of excitation - in this case, how much ion cyclotron radius is enlarged. Furthermore, CHIRP excitation excites all ions to

the same radius. This means that FTICR is quantitative with regards to ion populations within the cell. Frequency selective pulses can be used to manipulate the radius of selected ions as discussed later.

Unfortunately, additional factors can affect measurement of ICR, namely that the ions can have additional motions beyond simple cyclotron motion. These include the magnetron motion, and axial motion from the trapping plates at either end of the ICR cell. These lower frequency motions cause a slight modulation to the cyclotron frequency, and the net result is that it is the reduced cyclotron frequency (ω_+) which is routinely measured. On modern instrumentation, such as used here, magnetron motion and axial trapping motion are not a significant issue (these additional peaks are not observed, and mass calibration compensates for the reduced cyclotron frequency). Therefore, there will be no further discussion of this phenonemon. The interested reader is referred to the 1998 primer.[130]

Mathur and O'Connor presented a review of common artifacts in FTICR.[138] These include the presence of higher-order cyclotron frequencies, known as harmonics. These are essentially multiples of the base frequency. If an ion produces signals not only for its fundamental cyclotron frequency, but also even and odd multiples of this frequency, the resulting spectrum will be increasingly complex. As such, detection of just the fundamental frequency is desired.[130,138] As with magnetron motion, harmonic peaks are not observed on the instrumentation used in this work, and so their origins - theoretical and in practice - will not be discussed further. However, an interesting area of potential improvement in FTICR MS is that of using the harmonics to achieve higher resolving powers. This is possible by means of what is called 'quadrupolar detection', whereby the ICR cell detection and excitation plates are all used for detection, allowing signal detection along both x and y direction. This, coupled with deliberate methods to increase the presence of harmonic peaks, provides measurement of higher frequencies - thus resulting in higher resolving power. This is also known as 2ω or 4ω detection, and is achieving excellent results in complex mixture and biological mass spectrometry applications. New instrumentation (Solarix 2XR) will allow for routine 2ω, doubling the resolving power.[134,139] Application of 4ω detection at 21 T resolved isotopic peaks of ubiquitin (8.6 kDa) at m/z 714 with just a 12 ms transient length.[134]

In FTICR, additional concerns include space charge effects and peak coalesence. Space charge effects, simply, are a result of unfavourable electrostatic interactions between ions in a finite space, and induce errors with the measured frequency. Peak coalesence is a phenomenon where two (or more) closely separated frequencies can coalesce, producing a single peak at some value between the two correct ones. This can be difficult to confirm, however when measure-

ment mass accuracy is high, coalesced peaks should not be mistakenly assigned (and will remain unassigned).[130] Neither effect is observed to cause problems in the data presented in this thesis. Peak splitting is observed when the ICR vacuum is insufficient, or when the ICR cell is overfull of ions. However, adequate instrument maintenance and careful selection of acquisition parameters avoids this problem. Manual inspection of data during method 'tuning' and acquisition is necessary in every session to ensure acquisition of high fidelity data.

1.3.2.3 *Resolution*

In MS, two similar terms are used, often interchangeably, to describe spectrum resolution - resolution and resolving power. As NMR defines resolution as the linewidth (FWHM), to avoid ambiguity here resolution will also refer to the absolute linewidth. Resolving power, however, refers to the linewidth relative to the measured mass, as per Equation 1.21.

$$\text{Measured Resolving Power} = \frac{m}{\Delta m_{50\%}} \tag{1.21}$$

High resolution FT MS may achieve resolving powers (RP) in the order of 10^5 , to 10^6 , which is substantially greater than achieved by more common techniques like QTOF (10^4) or a single quadrupole mass detector such as in a GC-MS (10^1) to 10^2). The importance and choice of resolving power within the contexts of this research is discussed further in chapter 7.

The theoretical resolving powers can be calculated using a simplified equation from Cho *et al.*, shown in Equation 1.22.[132]

$$\text{Theoretical Resolving Power} = \frac{CBt}{m/z} \tag{1.22}$$

Where C is a constant dependent on processing parameters (apodisation and magnitude or absorption mode spectra) and instrument corrections, B is the field strength, t is the transient length. This shows that the resolving power is directly related to the field strength - higher field gives better performance - and indirectly proportional to the m/z - resolving power deteriorates across the mass spectrum.

Until recently, it was not possible to correctly phase broadband FTICR mass spectra. As such, the vast majority of data is still processed in magnitude mode, including data presented in this thesis. Phasing of mass spectra has been discussed in the literature extensively, including choice of apodisation functions.[140–143] Absorption mode processing, followed by phase correction, allows for a theoretical improvement in resolution of two-times from the magnitude mode. In reality, because of apodisation, the resolving power - in magnitude mode or absorption

mode - is never at its theoretical maximum. However, the improvements can still be substantial. In chapter 7, improvements of greater than two-times were possible by changing the apodisation and phasing the spectra, achieving a resolving power of 2.8×10^6. In this thesis, magnitude mode data are processed with full sine apodisation, whilst the example phased spectra in chapter 7 used the Kilgour apodisation.[142]

1.3.2.4 *Mass Calibration*

FTICR MS data must be calibrated before it can be interpreted. This involves two steps, the first converting the frequency to m/z, based on a fixed, known relationship. The common, two point calibration for the theoretical single ion in a uniform field is shown in Equation 1.23.[144]

$$m/z = \frac{A}{f^2} + \frac{B}{f} \tag{1.23}$$

Here, the mass-to-charge is determined by the frequency measured and two parameters - A and B. A and B correspond to the electric and magnetic field coefficients, and are empiricially determined against known standards - such as arginine or CsI clusters.[144,145] Bruker implements a three point calibration with an additional additive correction term C. This initial calibration should return a spectrum with a good mass accuracy ($<$10 ppm), however much better recalibration is both possible and necessary. Note also that the calibration will be optimal for the instrument and ICR cell parameters it was established for; subtle changes such as for trapping voltages or excitation pulses may degrade its performance, however instrumental recalibration is seldom necessary as spectra are recalibrated in post-processing, too.

The second step is post-processing (re)calibration, where each spectrum is manually recalibrated against the known observed compounds in the case of internal calibration. External calibration performs worse, achieving approximately three-times worse mass accuracy, and is best avoided where possible.[144] Internal calibration involves identifying a series of known species across the mass spectrum and fitting a correction function to their mass errors across the spectrum. For complex mixtures, a quadratic function is optimal and easily implemented; exotic piece-wise or iterative calibrations have shown improved performance but are not routinely available.[144,146] Calibration lists must contain a diverse range of species spanning the mass ranges of interest, and the identities must be confidently known - miscalibration will severely limit assignment, or worse, result in erroneous assignments.

Note that mass accuracy in FTICR is often measured in a relative unit, ppm or ppb. This relates the relative difference in theoretical mass (M_{theo}) with measured mass (M_{mass}), and multiplies by a fixed value (million, billion) to make the units more amenable, as shown in Equation 1.24.

$$error_{ppm} = \frac{M_{theo} - M_{mass}}{M_{mass}} \times 10^6 \tag{1.24}$$

Mass errors less than 1 ppm are commonly cited in the literature as a threshold for formula assignment, however for complex systems, much higher mass accuracies are required for unambiguous formula assignment based on exact mass alone. This is discussed further in chapter 7.

1.3.3 *Isolation and Fragmentation Methods*

FTICR offers a mass resolving power great enough to not require chromatographic prior separation. However, there are times that isolation of a narrow range of masses may be desired. This is achievable in two ways on an FTICR instrument - using the quadrupole, or using the ICR cell.

Quadrupole mass isolation is well established and well known, and is in fact a necessary function of quadrupole mass spectrometry. It works on a principle that applied electromagnetic fields deflect ions of given masses, allowing only desired masses to pass through. Deflected ions collide with the walls of the quadrupole, and are lost.

ICR isolation is a more advanced concept, and relies on a series of RF pulses to selectively excite or de-excite ions. This is discussed in more detail in the literature.[130] One such approach uses a stored waveform of the inverse Fourier transform of the frequencies desired to be excited.[147] Similar methods are utilised in 2D FTICR MS.[148] 2D FTICR MS couples selective excitation with fragmentation to achieve broadband fragmentation of a spectrum without chromatography, and shows tremendous promise with the increasing instrumentation and computational power.[148]

The isolation method implemented on the SolariX instrument is the 'sweep-and-shots' method, whereby, after broadband excitation, a notched 'sweep' pulse is applied. The notch around the frequencies of interest has zero amplitude, whilst all other ions are further excited, and 'ejected' from the cell. In reality, they are not ejected but collide with the cell wall and are lost. The remaining ions represent a fraction of the original spectrum, but some ions very close to those of interest may remain. 'Clean-up' RF pulses are applied to any residual unwanted ions, each 'shot' is a highly selective pulse which may last a relatively long time

(as with NMR, high frequency selectivity requires long pulse length). As the net result is far fewer ions in the cell than normal, it may be common to increase the IAT substantially. Likewise, it may be useful to couple in-cell isolation with pre-isolation using the quadrupole - *i.e.* filter a narrow 1 to 5 m/z range using the quadrupole, and then a 0.3 m/z notch with the sweeping pulse, followed by any required clean up shots.[130,149]

Once ions have been isolated, it is possible to perform fragmentation experiments and obtain potentially useful information on their structure. If chromatography has been coupled, and only individual compounds are being ionised at the source, it is possible to induce fragmentation at this stage. If ions are isolated sufficiently using the quadrupole, fragmentation within the collision cell is possible. Both cases use collision induced dissocation (CID), where neutral gas and a collision potential are applied to fragment species. If MS^3 is required, preliminary fragmentation within the source, and isolation of the fragment of interest in the quadrupole produces this tandem MS technique. Should the spectra be sufficiently complex that ICR cell isolation is required, fragmentation is still possible. There are several fragmentation methods that can be applied for this, including electron capture dissociation (ECD) and infrared multiphoton dissociation (IRMPD), both of which apply energy to the ions by means of electrons or photons, respectively. However, it is possible to perform the same form of fragmentation as in the collision cell - CID - within the ICR cell. This involves pulsing neutral gas molecules directly into the cell and thus increasing the number of collisions. Energy can be applied in the form of RF pulses, usually off-resonance, in what is known as sustained off-resonance irradiation (SORI).[130] This application of RF power within ≈500 Hz of the ion of interest introduces enough additional energy for CID. This method can work very well, but requires substantial parameter optimisation and instrument time.[130,149] SORI CID has been applied to wine,[55] polymers,[150] and in NOM studies.[149,151]

1.3.4 *Ionisation Methods*

Mass spectrometry measures the masses of ions. Therefore, one of the most important parts of a mass spectrometer is the ionisation source. In this section, the four ionisation sources used in this thesis are discussed. For all methods, both positive and negative polarities can be used.

1.3.4.1 *Electrospray Ionisation*

Electrospray Ionisation (ESI) is a soft-ionisation technique commonly employed as the ion source for FTICR and other MS instruments requiring an API source. In brief, a solution containing analytes of interest is sprayed through a nebuliser. A neutral nebuliser gas (nitrogen) at high pressure ($\approx$2 bar) helps to vaporise the solution producing a fine mist of micro-droplets. A high voltage is applied to either the nebuliser tip or the inlet for the instrument (the other is grounded), and an inert heated drying gas applied. The applied potential induces charge within each micro-droplet. As the micro-droplets are pulled into the vacuum, the dry gas evaporates the solvent. As this happens, charge is passed onto the analytes, which become ionised. The exact mechanism of this process is debated, but such discussion is ultimately beyond the scope of this thesis. What is important, however, is to note the requirements for ESI and the ion types it generates.

ESI produces singly and multiply charged ions - the latter increasingly common as the molecule grows in size (larger ions can stabilise more charges). For small molecules, ions are singly charged. ESI does not, technically, create ions, and relies on the molecule existing (at least in equilibrium) in ionic form in solution. In negative mode, ions are formed by loss of a H^+ ion, producing the M^- ion. ESI requires the analytes to be able to adopt this charge, which for (de)protonation requires the presence of labile protons - for example in carboxylic acid, alcohol, or amine functionality. ESI is not expected to ionise non-polar species. Commonly, ionisation efficiency can be improved or augmented by addition of a chemical modifier, such as an acid or base, to aid (de)protonation by changing the pH of solution.

Particularly in positive mode, though theoretically possible in the negative mode, ESI can also produce adducted ions. Here, instead (or as well) as loss/-gain of a proton, there is addition of another ion which carries the charge. For example, positive mode ions are commonly of the form $[M+Na]^+$ or $[M+K]^+$. Chloride adducts ($[M+Cl]^-$) are possible in negative mode, but rarely observed in complex mixtures. Radical ions are not typical of ESI, and are more likely if the analyte existed as a free radical in solution. The use of ESI for complex NOM samples was reviewed by Novotony *et al.*[152]

1.3.4.2 *Atmospheric Pressure Photo Ionisation*

Another API technique, atmospheric pressure photoionisation (APPI) involves spraying a solution through a nebuliser, formation of micro-droplets which are dried and pulled into the vacuum and instrument. Instead of relying on ionic species in solution and an electric potential, APPI utilises photons from a light source

to ionise species. Commonly, a Kr lamp is used, emitting ultraviolet photons with an energy of 10.6 eV. To ionise species successfully, the molecules must be able to absorb the UV radiation, so typically organic aromatic or conjugated systems ionise best. They must have an ionisation potential below 10.6 eV, and often a chemical dopant is added to solution to boost ionisation efficiency. For example, toluene (ionisation energy 8.8 eV) will ionise, and can charge transfer to less ionisable analytes of interest. APPI will produce (de)protonated species and radical ions readily. Adduct formation is much less likely than in ESI.

1.3.4.3 *Atmospheric Pressure Chemical Ionisation*

The final API source examined is atmospheric pressure chemical ionisation (APCI), which is very similar to APPI, except instead of a lamp emitting photons, a coronal discharge emits high energy electrons to induce ionisation. As APCI does not require polarity or conjugation, APCI will ionise a large variety of species in terms of structure, polarity and mass, including hydrocarbons. Again, (de)protonated ions are expected, along with radical ions. Adduct formation is again less likely than for ESI. APCI does not typically use a chemical modifier, and in fact its ability to ionise a diverse range of chemical structures can lead to problems with contamination - for example hydrocarbon impurities in the nebuliser gas.

1.3.4.4 *Laser Desorption Ionisation*

Finally, for dry, solid, or insoluble samples, laser desorption ionisation (LDI) can be used. This involves adhesing the sample to a plate which is inserted into the source, which is then pumped down to vacuum prior to ionisation. A laser beam is used to ablate (desorb) molecules from the plate, and also to ionise them. There is debate in the literature about exactly what happens during this process.[153,154] The laser (Smartbeam-II, Nd:YAG) used on this instrument produces photons at 355 nm, *i.e.* within the UV range. Therefore, compounds with conjugated systems (like with APPI) are likely to ionise well. Indeed, typically LDI is performed with the analyte co-crystallised in a matrix in matrix assisted laser desorption ionisation (MALDI). There are a number of MALDI matrices, but they typically are small aromatic molecules, for example 2,5-dihydroxybenzoic acid (DHB). These compounds act like toluene in APPI, are readily ionised, and can transfer charge to the analytes. MALDI has been used for complex mixture (NOM) analysis previously.[155] Interestingly, the compounds used in matrices are similar to known compounds in NOM, and recent work within the group demonstrated that matrix-free LDI outperformed MALDI for analysis of NOM whilst

also providing complementary information to ESI, and avoiding the issues of sample solubility typically associated with NOM analysis.[156]

The matrix sometimes also contains a salt, to provide an additional source of ions. Matrix-free LDI will have a shortage of protons, so positive mode ionisation is likely to be in the form of adducted ions and radical cations. Negative mode ionisation is likely to be deprotonated or radical anions. High laser powers may induce fragmentation, however this was not observed for the study on NOM.[156]

1.3.5 *Interpreting Mass Spectra*

The acquired FTICR mass spectra of complex mixtures can be complex and contain many signals. There are many aspects to consider when interpreting the spectra in the contexts of complex mixtures. The process of unambiguously assigning a molecular formula to a measured mass involves several factors.

Atoms have an exact mass. Each isotope of each element has a unique exact mass. By definition, ^{12}C has a mass of 12 Da exactly. Every other isotope has a non-integer exact mass, and for most stable isotopes, the exact masses are known to a very high precision and accuracy. As mentioned in the context of NMR, different isotopes have different natural abundances - ^{12}C is far more abundant than ^{13}C. As with masses, natural abundances are known to a high accuracy, too. The exact masses and natural abundances of key isotopes relevant to this work are summarised in Table 1.2.

The difference between the nominal (integer) mass and the exact mass is known as the mass defect. For many isotopes this is negative, however for the stable isotopes of hydrogen and nitrogen it is positive - see Table 1.2. The exact mass and mass defect are also specific to each isotope (at least amongst common isotopes), and therefore it is possible to calculate the elemental composition of a species from its measured exact mass by means of a brute force calculation - what combination of isotopes would give this mass? Of course, measurement error will make this imperfect, so a mass tolerance will be necessary. As discussed previously, commonly the relative mass error ppm is used. For a species at m/z 400, a 1 ppm error is a 0.4 mDa error. To put this in perspective, an electron mass (0.5486 mDa) at m/z 400 is 1.37 ppm, and the mass difference between $^{1}H_2$ and $^{2}H_1$ is 1.548 mDa, corresponding to 3.87 ppm at m/z 400. Therefore, high mass accuracy is able to discriminate very similar exact masses.

In complex mixtures, it is observed that most compounds are related and belong to homologous series. This means that the same base formula has an additional repeating unit, for example a series of fatty acids growing by a CH_2 group each time. The exact mass of CH_2 is 14.01565, and therefore - even without know-

Table 1.2: Exact masses[157] and natural abundances[158] of key isotopes and the electron

ISOTOPE	EXACT MASS	NAT. ABUN. %
e^-	0.0005486	-
^{1}H	1.007825	99.9885
^{2}H	2.014102	0.0115
^{12}C	12.00000	98.90
^{13}C	13.003355	1.10
^{14}N	14.003074	99.63
^{15}N	15.000109	0.37
^{16}O	15.994915	99.76
^{17}O	16.999131	0.04
^{18}O	17.999159	0.20
^{23}Na	22.989770	100.0
^{31}P	30.973763	100.0
^{32}S	31.972072	95.02
^{33}S	32.971459	0.75
^{34}S	33.967868	4.21
^{39}K	38.963708	93.2
^{41}K	40.961825	6.73

ing the elemental composition of a species - it will be possible to relate group ions together based on this mass difference.

To make visualisation and computation of these trends more straightforward, the Kendrick mass defect (KMD) was proposed.[159] This modifies the masses of the elements by redefining CH_2, not ^{12}C, to have a mass of exactly 14, not 14.01565. The calculation for this is performed as shown in Equation 1.25

$$KM = M_{meas} \times \frac{Base_{NM}}{Base_{EM}} \tag{1.25}$$

Here, the Kendrick mass (KM) for any ion can be calculated by multipling the measured mass (M_{meas}) by the ratio of the nominal mass (NM) of the new base unit to its exact mass (EM). CH_2 is the most common base unit, but any chemical group which could be additive in a series can be used, including CO, CO_2, H_2, and so on. The result of this process is that now all species within the same homologous series have the same numbers after the decimal place. The CH_2 Kendrick mass has been calculated for several species shown in Table 1.3.

Table 1.3: Nominal, exact, and CH_2 Kendrick masses for several compounds.

SPECIES	NOMINAL MASS	EXACT MASS	KENDRICK MASS
$C_5H_{10}O_6$	166	166.04774	165.8623296
$C_6H_{12}O_6$	180	180.06339	179.8623296
$C_7H_{14}O_6$	194	194.07904	193.8623296
$C_7H_{12}O_7$	208	208.05831	207.8259852

Thus Table 1.3 shows a series of compounds increasing by a CH_2 unit two times, followed by addition of a O and loss of H_2. Nominally, both changes are a mass change of 14 Da. It is not immediately obvious that the last species is unrelated to the other ones by looking at their exact masses. However, inspection of their Kendrick masses makes it clear that the first three species are related by a CH_2, whilst the last one is not.

isual inspection of the exact masses is not immediately obvious that the last species is not like the previous ones, whereas inspection of the Kendrick masses makes it clear that the first three compound are related by a CH_2 group, but not the final species.

This Kendrick mass analysis forms the basis of many formula assignment routines in complex mixture mass spectrometry. For example, if the mixture is full of homologous series, and sufficient Kendrick units have been defined, it will be pos-

sible to assign every formula from the manual assignment of just one, by connecting the rest in a network approach. Such an approach was implemented by Kujawinski and Behn in 2006,[160] and Kilgour *et al.* in 2012.[161] Manually, Kendrick mass analysis has been used historically widely in complex mixture mass spectrometry, although its direct use is in decline thanks to automated formula assignment tools. Graphically, Kendrick masses can be useful in what is known as a KMD plot, with KMD calculated as shown in Equation 1.26,[162] and plotted against nominal mass for a Scotch Whisky sample (data acquired in chapter 7) shown in Figure 1.15.

$$KMD = (NM - KM) \times 1000 \tag{1.26}$$

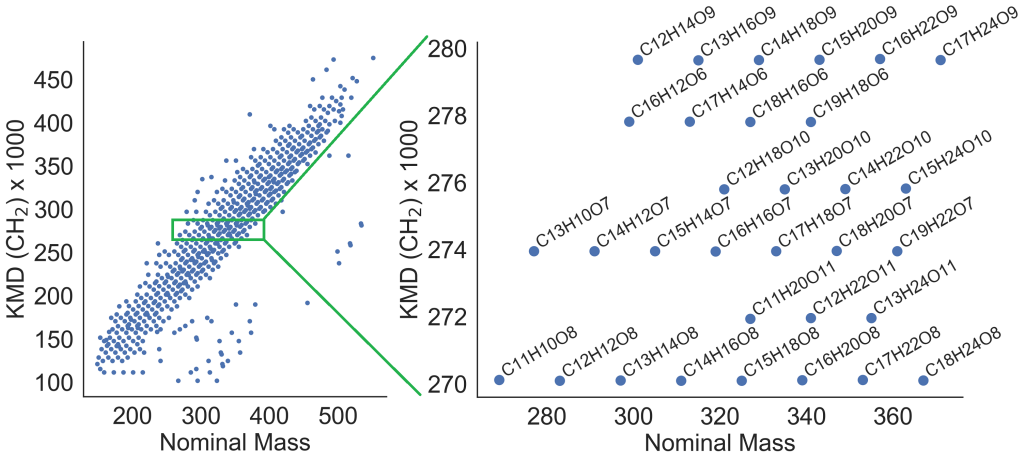


Figure 1.15: Example of plotting KMD against nominal mass for a Scotch Whisky ESI FTICR mass spectrum.

As KMDs can be similar, manual assignment commonly uses a z* approach to separate out species with similar - but different - KMDs.[163] The z*- for a CH_2 Kendrick mass unit - is calculated as shown in Equation 1.27. This will yield 14 groups (z*s), and each one should not have KMDs too close to each other.

$$z* = (modulo \frac{NM}{14}) - 14 \tag{1.27}$$

Formula assignment is an area of important development even today, and broadband formula assignment based on exact mass alone is not sufficient for CHNOSP species. This is discussed further in chapter 7.

Adding to the complexity of the spectra are isotopes. As shown in Table 1.3, almost every element has a number of isotopes. Whilst for the elements CHONS there is one predominant isotope, the minor isotopes are also observed in FTICR because of its sensitivity. Additional peaks in the spectra are thus caused by the

population of ions for each compound where the predominant isotope has been replaced with a minor one. These species are known as isotopologues - the same structure but with an isotopic substitution. The signal intensity for an isotopologue will correspond to its isotopic natural abundance and relative signal of the monoisotopic (major) species. The most common isotopologue in these sample types is the first ^{13}C isotopologue, where one ^{12}C has been replaced by a ^{13}C, at exactly + 1.003355 Da and at 1.1 % (× # carbons) relative abundance.

The fact that the isotopologues will have predictable abundances provides what is known as the isotopic fine structure (IFS). This is the position and intensities of all the observable isotopologues for any given species. This, coupled with exact mass accuracy, allows for unambiguous formula assignment. Unfortunately, the SNR has to be sufficient to observe many of the isotopologues. Some elements (*i.e.* P, Na, and K) do not have other isotopes.

Molecular formula does not equal structure, and one of the biggest weaknesses in FTICR MS is the lack of structural information yielded in broadband acquisition. Fragmentation was discussed earlier as a potential means to elucidate structures. However in a complex mixture, any ion peak may actually be multiple isomeric species with the same empirical formula but different structures.

1.3.6 *Visual Analysis of Mass Spectra*

Complex mixture mass spectra are often similar visually, and effort has been made in the literature to visualise them in more efficient ways. These include plotting the heteroatomic classes (*i.e.* bar plot of number of each O_n species), the double bond equivalence (DBE) (calculated as shown in Equation 1.28)[164] against carbon number, and van Krevelen diagrams.

$$DBE = 1 + \frac{1}{2} \times (2C - H + N + P) \tag{1.28}$$

DBE does not factor in oxygen, and many complex mixtures are dominated by CHO species. Thus, in 2006, Koch and Dittmar developed a new metric, the Aromaticity Index (AI).[164] This factors in oxygen, and better describes NOM chemistry than DBE alone. The equation for aromaticity index (AI) is shown in Equation 1.29.[164]

$$AI = \frac{1 + C - O - S - 0.5H}{C - O - S - N - P} \tag{1.29}$$

A modified version of AI was also recommended to account for σ-bound oxygen, whereby the number of oxygens is halved. This division is somewhat ar-

bitrary, and should be optimised if more is known about the chemistry of the sample.

Van Krevelen first published his eponymous diagram in 1950, discussing chemical changes in fuels.[165] He measured bulk elemental ratios, and plotted them on a scatter plot, connecting points based on chemical processes. The axes were elemental ratios, such as hydrogen-to-carbon (H/C) and oxygen-to-carbon (O/C). This plot was repurposed in 2003 by Kim *et al.* for 'graphical analysis' of high resolution mass spectral data of NOM.[166] Here, each assigned chemical species is plotted based on the molecular O/C and H/C ratios. The result is a scatter plot with many hundreds or thousands of points, many overlapping. In the case of petroleomic studies, where nitrogen may be common, one of the axis can be the N/C ratio. Three-dimensional plots have been published, but are inherently difficult to visualise on 2D paper.

Regions of the van Krevelen diagram are often assigned a chemical classification, for example lipids with $O/C < 0.2$ and $H/C \approx 2.0$, or carbohydrates with higher O/C ratios.[167] Recent work has aimed to push the van Krevelen diagram further, including basing molecular characterisation on more than just two elemental ratios.[168] During the course of this PhD, software to produce interactive van Krevelen diagrams was produced,[169] and later another open source tool was published by Brockman *et al.* to visualise metabolomics data with van Krevelen diagrams.[170]

1.4 STATISTICAL METHODS

The generation of significant volumes of data in modern spectroscopy and spectrometry has led to the development and implementation of numerous statistical methods to interpret and handle this data. In this section, some of the key statistical methods and backgrounds are detailed.

1.4.1 *Statistical Correlation Spectroscopy*

The basic premise behind statistical correlation spectroscopy relies upon pseudo-quantitative spectroscopy, *e.g.* NMR. In NMR, signals in a mixture, corresponding to the same compound, or to compounds with the same origin or biopathway, will increase and decrease proportionally to each other. In other words, these signals are correlated. If spectra of several samples of similar mixtures can be acquired, it may be possible to correlate signals within these spectra by looking at how they change across spectra. The requirements would be that the mixtures must contain, largely, the same components, and that the spectra are resolved enough to observe these correlations.

1.4.1.1 *STOCSY*

Statistical Total Correlation Spectroscopy (STOCSY) is an idea first implemented for ^{1}H NMR spectroscopy by Cloarec *et al.* in 2005.[171] It was based on work by Šašić *et al.* in 2000,[172] which generalised work originally performed by Noda in 1989 on two-dimensional infra-red spectroscopy.[173]

Mathematically, the required datasets would consist of two matrices, each containing n samples (or spectra) represented by v variables - data points or bins - such that each matrix is $\mathbf{X} = n \times v$. The correlation, therefore, would be between two matrices, $\mathbf{X_1}$ and $\mathbf{X_2}$. In the simple case of correlating one set of NMR data, the autocorrelation is calculated between $\mathbf{X_1}$ and the transpose of itself, as summarised in Equation 1.30:[171]

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X_1}^{\mathrm{t}}\mathbf{X_1} \tag{1.30}$$

In this case, $\mathbf{C}$ is a matrix of $v_1 \times v_2$, and each value represents the correlation coefficient between any given variables in the input datasets. Signals from the same compound should have $r = 1$, but in practice any noise or other signal distortion will make $r < 1$. The net result, if such a matrix were plotted as a heatmap, would be not dissimilar to a 2D TOCSY NMR spectrum. In practice, this can have some drawbacks. Typically, high-resolution 1D spectra are used as

inputs. Even if binned data are input, the resulting correlation matrix may be much higher resolution than a typical 2D NMR spectrum, and visualisation can be computationally slow to generate, and difficult to interpret.

More practically, it is helpful to view a series of selective, 1D STOCSY plots - analogously to the selective 1D TOCSY experiment. In this, the correlation is calculated between a given 'driver' peak (or narrow range of variables), and the entire data matrix. The result is now a vector C of v_1 correlation coefficients. These can be visualised in a 1D plot as either intensities (and thus a pseudo-NMR spectrum), or as the colour encoding of an averaged NMR spectrum from the dataset. Examples of such an approach are used in subsection 5.6.2 and shown in Figures 5.20, 5.21, and 5.22.

There are practical advantages to STOCSY. Firstly, acquiring 1D spectra is inherently faster than acquiring 2D spectra. Practically speaking, for a 2D homocorrelated NMR experiment, each increment is essentially acquiring another 1D spectrum, and thus to acquire 100 high resolution 1D NMR spectra may be more advantageous than to acquire a single 2D TOCSY spectrum with 100 t_1 increments. The resulting data will be of higher resolution, and represent 100 samples rather than simply one.

Adoption of STOCSY into the NMR community has been limited only to some metabolomics studies.[171] STOCSY has been used for NMR analysis in food science before, including for signal assignment in spectra of champagne base wines.[174] Further modifications to STOCSY include application to diffusion ordered NMR,[175] heteronuclear NMR,[176] and adaption for heterospectroscopy - coupling NMR and LC-MS[177]

1.4.1.2 *RANSY*

Developed six years after STOCSY was implemented for NMR, ratio analysis NMR spectroscopy (RANSY) is a similar statistical method to correlate signals within data. It is based on the principle that the *ratio* of the integrals of two or more NMR signals from the same compound will be equivalent across different samples and concentrations. RANSY uses a 'driving' peak to be the source for the ratio analysis; the ratio of this peak to all other peaks is calculated. The reciprocal of the coefficient of variation - or relative standard deviation - is then calculated by dividing the mean of each ratio by its standard deviation. The algorithm is provided in the original paper by Wei *et al.*[178] In this paper, RANSY was successfully applied to 1D and 2D NMR of biofluids including blood serum. RANSY, and STOCSY, have been implemented in the R program rDolphin for metabolomics analysis.[179] Both methods were implemented in Python for use in this thesis.

1.4.2 *Unsupervised Modelling*

Modern analytical instruments generate a lot of data. For studies involving many samples, this can become impractical to manually inspect to identify common features and differences. In this case, techniques which identify the maximum variation within and between samples, or which can pull out key information from the data are very useful. In preliminary studies, or in sample blinded studies, techniques which perform such tasks without predefined classification or labels are especially important. These techniques are termed 'unsupervised' because they use only the provided data to create a model, and no external information or factors.

1.4.2.1 *PCA*

The most commonly used unsupervised method in chemometric studies is principal components analysis (PCA). PCA is a data dimensionality technique which can also be used for visualisation. PCA takes high dimensional data - with many samples and many variables - and creates new axes to describe this data in fewer variables or dimensions. Put another way, it takes a number of correlated input variables (or axes), and calculates new uncorrelated principal components, or axes, which are linear combinations of the original variables. Bro and Smilde presented a tutorial review in a chemistry context in 2014, reference to which the reader is recommended.[180]

A matrix $\mathbf{X}$ ($m \times n$) of m rows (observations, samples, or spectra) and n columns (variables) is mean centred (subtraction of the mean from each variable), scaled, and the covariance matrix calculated. This covariance matrix is decomposed using either eigen analysis or singular value decomposition.[181,182] Without worrying about how this performed, all that is required for use is to know that PCA decomposes $\mathbf{X}$ as shown in Equation 1.31,[183] where $\mathbf{T}$ are the scores, $\mathbf{P}$ are the loadings, and $\mathbf{E}$ is the residual undescribed data.

$$\mathbf{X} = \mathbf{P}\mathbf{T}^T + \mathbf{E} \tag{1.31}$$

The scores ($\mathbf{T}$) describes the variation in the samples and the loadings ($\mathbf{P}$) the variation in the variables. Put another way, the loadings describes the contribution of each input variable to the new principal components.

PCA can be used for visualising the reduced dimensionality, also. The scores can be plotted as a two-dimensional scatter plot, placing each original observation in the new principal component space. Typically, the scores for principal component 1 are plotted on the x-axis, and scores for principal component 2 or

greater on the y-axis. When points appear close together, it indicates those observations are similar. Where points are far apart, they are dissimilar. Often, a scores plot will be described as containing clusters or groups of observations. This plot makes it possible to visualise the variation between the samples. Outliers of the model, those which are significantly different from the rest, are placed outwith the Hotelling's T^2 ellipse, which represents a generalisation of the t-test to multivariate data.[184] A 95 % confidence is commonly used.

The loadings relate the input variables to the new principal components. These can be plotted as a two-dimensional plot of the principal components, as with the scores, to visualise which variables are correlated and which are not. However, it can be more useful to plot the loadings of one principal component against the input variables, especially in the context of spectroscopic data. Here, a line plot will highlight input variables with large loadings, and thus importance in describing the variation in the data.

A biplot places both scores and loadings on one figure, further aiding visualisation of how the variation is explained by observations and variables. For data with many variables, the biplot can become crowded and difficult to interpret; biplots are more useful in models with few variables.

Pre-processing of data prior to PCA or other statistical analysis can be significant.[183] PCA typically involves mean centering of the data prior to decomposition, but also scaling of the data. This scaling can take the form of autoscaling (also known as unit variance, Equation 1.32), which sets the standard deviation of each variable to 1:[183]

$$\tilde{x}_{mn} = \frac{x_{mn} - \bar{x}_m}{s_i} \tag{1.32}$$

where x_{mn} is the original data point of variable m and sample n, $\tilde{x}_{mn}$ corresponds to the transformed data point, $\bar{x}_m$ is the mean of variable m, and s_i is the standard deviation. Alternatively, Pareto scaling, (Equation 1.33), may be used, and this retains some of the importance of the original scale whilst reducing the impact of very large ranges:[183]

$$\tilde{x}_{mn} = \frac{x_{mn} - \bar{x}_m}{\sqrt{s_i}} \tag{1.33}$$

In a simple case, the matrix $\mathbf{X}$ could be decomposed into as many principal components as original variables, at which point it would no longer be a data reduction method but simply have transposed the data into a new space. In reality, as the principal components are orthogonal, and because the first describes the most variation, the second the second most, etc., it is often the case that only a few principal components are necessary to describe most of the data.

The amount of variation described by each component, or the sum of components, is the R^2 value, sometimes called the goodness of fit. R^2 increases from 0 to 1. At R^2=0, no data is explained by the model, and at 1, all of the data (100 %) is described by the model. It approaches 1 as more components are added. Thus, the variation of the input data which is described by the model may be termed $R^2\mathbf{X}$.

The software used for calculating PCA and orthogonal partial least squares discriminant analysis (OPLS-DA) in this thesis is SIMCA (Sartorious, formerly Umetrics), which determines an optimal number of principal components based on a cross-validated prediction. Essentially, the model is calculated a number of times, incrementing the number of components, and each time leaving some of the original data out of the model. The model is then used to predict the values of those left-out data, and the sum of the squared difference between the observed and predicted values is calculated (also known as PRESS - predictive residual sum of squares). This goodness of prediction is termed Q^2, and varies from -1 to +1. A value of 0 indicates prediction as good as random change, negative values are worse than random, and positive values are good. Q^2 does not approach 1 as more components are added, and instead decreases as too many components are added. Thus, the maximum of Q^2 is used to determine the optimal number of components. PCA is not, however, used for predictive classification here, and so the Q^2 values are not necessary to report for PCA models.

1.4.2.2 *ICA*

Related to PCA is Independent Components Analysis (ICA), which is used to decompose data into individual components. Where PCA calculated components of maximum variation, ICA attempts to determine the underlying components in the data. An application, and analogy, for the use of ICA is to decompose multiple overlapping voices in a recording. With one measurement, it will be impossible. As the number of independent recordings of this conversation increases, it is possible to decompose the data into the individual voices, or underlying signals. This is known as blind source separation.[185]

The general principle is to determine the independent components from multiple measurements of the same mixture. This can be modelled as Equation 1.34, where $\mathbf{x}$ is the measured data, $\mathbf{A}$ is the so-called mixing matrix, and $\mathbf{s}$ the independent components, or source signals.

$$\mathbf{x} = \mathbf{As} \tag{1.34}$$

There are many algorithms to solve this relationship, and thus perform ICA, but that is beyond the scope of this thesis. The key points to consider are that to completely and correctly solve the relationship, there must be as many measurements as underlying signals. For the case of NMR, if a mixture class contained three compounds, at least three spectra would be required to resolve the compounds. Each measurement should be independent - *i.e.* the samples should be different, and have differing ratios of the compounds.

ICA has been applied to NMR of mixtures,[186] and to solve the problem of overlapping signals in DOSY.[185] In mixtures, the number of components is often not known, and the number of measurements limited by the number of different samples. It is necessary to define the number of components prior to calculation of ICA, and it is not possible to objectively and automatically determine the optimal number of components. Therefore, a degree of subjectivity creeps into ICA.

1.4.3 *Supervised Modelling*

Statistical analysis is also used to model specific parameters, for example in metabolomics studies between test and control subjects. Where models are created, or trained, with both measured data and a parameter to model, they are known as 'supervised models'. Here, the input data consists again of a matrix $\mathbf{X}$ $m \times n$ of m rows (observations, samples, or spectra) and n columns (variables), but also a second matrix or vector, $\mathbf{Y}$, containing the parameter to be modelled.

If the model is to predict a continuous numerical variable - such as solubility - it is a regression problem. If the model is to predict categorical variables - such as test and control - it is a classification problem. Computationally, classification here involves regression against binary values (in the case of two-class classification), where $\mathbf{Y}$ contains either 1s or 0s (or -1 and +1).[187] For multiple class classification, $\mathbf{Y}$ can be a binary matrix.

1.4.3.1 *OPLS-DA*

A commonly used supervised method is orthogonal partial least squares discriminant analysis (OPLS-DA). OPLS-DA first finds the optimal separation between the classes, and then determines the maximal variation within the classes. It is an extension of PLS-DA, with the addition that it determines orthogonal components, unrelated but interfering with the classification to be modelled.[188] An example of this may be in a drugs trial where there is significant variation between male and female participants, independent of their test or control group status.

OPLS-DA will separate out orthogonal variation (sex) from the data necessary to model the parameter of interest (group).

Brereton and Lloyd presented a tutorial review of PLS-DA in 2014,[187] which describes in detail the application, implementation, use, and limitations of this method. The mathematics of PLS-DA and OPLS-DA are more complex than those of PCA, and the reader is refered to the literature.[187,188]

OPLS-DA models generated by SIMCA also have an R^2 value (for both **X** and **Y**), and are produced in a cross-validated manner, both to calculate the goodness of prediction statistic Q^2, but also to try to generate a model which is valid. In predictive modelling, it is possible to over fit a model. In OPLS-DA, this can happen because there are too many classes, too many variables, or too many components calculated. The Q^2 statistic is a good measure of the genuine predictive ability of the model because it is generated in a cross-validated method. Values above 0.5 are largely considered acceptable. However, additional validation strategies can be employed to increase confidence in the model. Worley and Powers evaluated means to combine PCA and OPLS-DA analysis to avoid overfitting of the data and spurious models.[189] However, more routine means to validate the model take two forms.[190]

Permutation testing involves scrambling **Y** prior to model generation, and calculating the R^2 and Q^2 values. This is repeated a number of times - 20 or more. A randomised **Y** vector should have poor predictive ability, and the Q^2 values should be <0, or at least less than for the original model.[190]

Additionally, analysis of variance (ANOVA) can be used to calculate test statistics to determine the goodness of the model, including the notorious p-value. ANOVA determines if two distributions (or their variance) are statistically significantly different (the t-test determines if the means are different). The resulting p-statistic is between 0 and 1, with values of $p<0.05$ or 0.01 typically regarded to indicate statistical significance.[190] CV-ANOVA is performed in SIMCA and performs a permuted cross validation of the data to determine the significance.[191]

2

AIMS

My goal is simple.
It is a complete understanding of the universe,
why it is as it is, and why it exists at all.

— Stephen Hawking (1985)

Due to its economic, cultural, and demonstrated scientific interest, Scotch Whisky is a product worth investigation with the highest level analytical techniques available. It is a complex mixture, containing thousands of compounds, and many of them are unknown. Their significance for flavour or quality, their origins, and their quantities, are all important properties to determine. Furthering the understanding of Scotch Whisky in terms of its chemical composition will benefit the industry and the consumer, and development of new techniques for this type of analysis will benefit the wider research community.

The aims of this PhD thesis are thus as follows:

- Investigate and develop an NMR methodology for the acquisition of 1D ^{1}H spectra of Scotch Whisky
- Develop a solvent suppression methodology to allow NMR spectra to be acquired for native samples
- Automate this method for high throughput experiments, and implement the solvent suppression method into other 1D and 2D NMR experiments for structural elucidation
- Evaluate the quantification of ethanol and minor compounds in Scotch Whisky by NMR
- Develop and optimise FTICR MS protocols for the analysis of Scotch Whisky as a complex mixture
- Develop tools for FTICR MS data processing, analysis, and visualisation
- Investigate alternative FTICR MS ionisation methods for the analysis of Scotch Whisky
- For both NMR and FTICR MS, acquire spectra for a large range of Scotch Whisky samples for metabolomics type study

3

METHODOLOGY

Well like my grandma used to say, "If at first you don't succeed..."
'...try a larger thermonuclear reaction?'

— Lt. Cols. Mitchell and Carter, Stargate SG-1 (2006)

This chapter is provided for reference of the key methodologies and technical details used in this thesis. Discussion is provided in the subsequent results chapters.

3.1 NMR

The following details sample information, instrumental parameters, and data processing related to chapter 4 and chapter 5.

3.1.1 *Sample Preparation*

For NMR analysis of Scotch Whisky samples, a standardised sample preparation was implemented as follows. Samples were analysed in 5 mm Wilmad 535-PP-7 Precision NMR tubes rated for 600 MHz. D_2O (99 %), acetic acid-d_4 (99.5 %), and sodium acetate-d_3 (99 %) were acquired from Sigma-Aldrich Co. Samples (500 µL) of Scotch Whisky, provided by the SWRI, were mixed with D_2O/sodium acetate/acetic acid buffer solution (100 µL) containing DSS-d_6 as an internal standard. The final concentrations of buffer and internal standard were 25 mM and 1 mM, respectively. Samples of very high ethanol strength (>80 %) were diluted using ultra high quality (UHQ) water (1:1) before mixing with the buffer solution.

Reference standards of compounds were prepared for NMR by dilution in a 40 % ethanol : water solution at approximate concentrations of 1 mM to 10 mM, followed by preparation as above for a Scotch Whisky sample.

Spike-in samples were prepared by spiking in a small amount of a reference compound into a sample of Scotch Whisky at two levels of addition. These were approximate levels at or below the expected pre-existing level of compound in solution (0.1 mM to 1 mM). Preparation continued as for a normal Scotch Whisky sample. Where the reference standard was a powder, no dilution compensation was made. Where it was liquid, equivalent volume of 40 % ethanol : water solu-

tion was added to the comparison samples. For spike-in analysis, a control with no added compounds (only dilution as required) was prepared.

A series of ethanol : water solutions between 35 % to 70 % ABV were prepared using volumetric glassware. These stock solutions were further prepared by addition of the D_2O buffer as for normal Scotch Whisky samples as above.

A solution of acetaldehyde in ethanol : water was prepared for analysis of the equilibrium hemiacetals formed. One drop of acetaldehyde was added to 100 µL of ethanol : water solution (40 % ABV). Accurate measurement of acetaldehyde was not possible due to high volatility of acetaldehyde, and was not necessary for the nature of experiments performed. This solution was made up to 600 µL using D_2O prior to addition to an NMR tube for analysis.

3.1.2 *Experimental Parameters*

All spectra were acquired at 300 K, unless stated otherwise. NMR experiments were developed and implemented on 4-channel Avance III HD 600 and 800 MHz Bruker spectrometers equipped with TCI cryogenically cooled probes with automatic matching and tuning and 16-slot (600 MHz) or 24-slot (800 MHz) sample changerx. All spectra were acquired at 600 MHz unless otherwise noted. Acquisition of a 1D ^{1}H spectrum is preceded by the acquisition of three preliminary spectra in automation using Bruker IconNMR and Python scripting. The 1D ^{1}H spectra were acquired in full automation, whereas subsequent 2D experiments were acquired in manual mode.

3.1.2.1 *Solvent Suppression*

This section details the precursor experiments required for solvent suppression. The experiments are performed sequentially in automation, starting with experiment number 10.

Experiment 10 uses the AU program *efindwater2.wk* for automated acquisition. This AU program is based on *Findwater* by Peter Gierth and Andrew Gibbs, included with TopSpin. This routine performs pulse calibrations for ^{1}H on both channel one and three. Having determined the correct 90° pulse for ^{1}H, it determines the approximate water resonance via acquisition of a 1D ^{1}H spectrum with a 360° pulse, suppressing most of the water signal. The first 4,096 points are linear back predicted - 512 in the case of high (>50 %) ABV samples - to remove the fast-decaying broad component, and the residual sharp component peak picked (within pre-defined limits of 5.2 to 4.2 ppm) and set as the water frequency. These experiments are stored in experiment numbers 99999-99997. The AU program re-

turns to experiment number 10, executes *getprosol* with the determined 90° pulse, sets p1 to 0.4 µs and RG to 1 prior to acquisition of the NMR spectrum with on-resonance presaturation of water, leaving the ethanol CH_2 and CH_3 unattenuated. Acquisition time was 3.93 s and relaxation delay of 4 s. The time domain consisted of 64k points covering a sweep width 13.8 ppm. One dummy scan was performed prior to co-addition of two transients.

Experiment 11 acquires a reverse INEPT spectrum using acquisition and relaxation times of 6.06 and 4 s, four dummy scans and 16 scans with ^{1}H and ^{13}C carrier frequencies of 2.2 and 40.0 ppm, respectively. The FID is zero filled once and Fourier transformed using exponential broadening of 0.2 Hz. The antiphase satellite ethanol signals are peak picked, and the chemical shifts of CH_2 and CH_3 signals of ^{12}C isotopomers are calculated by adding 1.53 and 1.11 Hz to the corresponding chemical shifts of ^{13}C isotopomers, respectively, to compensate for the ^{13}C isotope shift. A frequency-modulated rectangular-shaped pulse, *EthanolShape*, is generated starting from a square pulse (1000 points). This pulse inverts the on-resonance CH_2 protons and the off-resonance CH_3 protons. Its exact length, to be around 50 ms, is calculated to allow n 2πrotations for the off-resonance CH_3 protons. This experiment also stores the centre position of the CH_2 signal as the carrier frequency for the final experiment. The power used for this pulse corresponds to $\gamma B_1/2\pi$ equal to 28 Hz. In automation, it uses the Python script *createEthanolShape_ wk.py*.

Experiment 12 acquires a single scan 1D ^{13}C spectrum with inverse-gated ^{1}H decoupling. The spectrum is acquired using an acquisition time of 2.04 s with ^{1}H spectrum and ^{13}C spectrum carrier frequencies set to 2.20 and 40 ppm, respectively. The spectrum is peak picked and the positions of the ethanol carbon signals are used to calculate a shaped pulse, *CarbonShape*, for decoupling of the ethanol satellites in the final experiment. The ^{13}C spectrum carrier is placed exactly in the middle of the CH_2 and CH_3 ^{13}C spectrum chemical shifts. A 180° rectangular cosine modulated shaped pulse consisting of around 500 points and length approximately 2.5 ms is generated. The exact number of points and length is optimised per sample. Its exact length is calculated to allow n 2πrotations for both resonances. The power level was set to $\gamma B_1/2\pi$ equal to 333 Hz, and the pulses were phase cycled using m4p5 scheme, where m4 and p5 are 0°, 0°, 180°, 180° and 0°, 150°, 60°, 150°, 0°, respectively.[192,193]

Experiment 13 acquires the final, ethanol and water suppressed spectrum. It imports the various offsets, calculated pulse shapes, and power levels to acquire a 1D ^{1}H spectrum. The experiment uses the NOESY presaturation pulse sequence, with a nominally short NOE mixing period (1 ms to 10 ms). For water pre-saturation, the power level was set to $\gamma B_1/2\pi$ equal to 20 Hz. The receiver gain is set to

45.2, and spectra are obtained using four dummy scans and 32 scans. The FID is acquired with digitised sampling of 128k time domain points over 16 ppm, yielding the acquisition time of 6.82 s at 600 MHz or 5.14 s at 800 MHz; a 4.05 s relaxation/pre-saturation delay was used. The FID was zero filled once and Fourier transformed using an exponential line broadening of 0.2 Hz.

For fuller discussion of the methodology of these experiments, please refer to chapter 4.

3.1.2.2 *Implementation Into Other Experiments*

The developed solvent suppression was implemented into other NMR experiments by addition of a presaturation block to existing pulse sequences. Details are provided below.

The pulse sequence of a magnitude mode COSY experiment (Bruker pulse sequence *cosygpppqf*) was modified to include solvent signal presaturation during a 3 s relaxation delay as outlined for the acquisition of 1D ^{1}H NMR spectra. The 2D COSY spectra were acquired using 1k and 4k complex points in F1 and F2 using spectral widths of 10 and 13 ppm in F1 and F2 yielding t1 and t2 acquisition times of 85.2 and 262 ms, respectively. Four scans were accumulated per increment resulting in the overall acquisition time of 3 hr 52 min. A forward linear prediction to 2k points was used in F1 and zero filling to 8k was applied in F2. A sine square window function was used for apodization prior to Fourier transformation in both dimensions.

Similar modifications to those described for the COSY experiments were implemented into a phase-sensitive TOCSY experiment using a z-axis decoupling in the presence of scalar interactions (DIPSI)-2 spin-lock[194] starting from the Bruker pulse sequence *dipsi2esgpph*. Note that the excitation sculpting (*es*) element of the sequence was disabled. A 40 ms spin-lock was applied at $\gamma B_1/2\pi$ equal to 10 kHz. The PFGs surrounding the DIPSI spinlock were applied at 1% and 3%. The spectrum was acquired using 768 and 4k complex points in F1 and F2 using spectral widths of 10 and 13 ppm in F1 and F2 yielding t1 and t2 acquisition times of 63.9 and 262 ms, respectively. Sixteen scans were accumulated per increment resulting in the overall acquisition time of 8 hr 8 min. A forward linear prediction to 2k points was used in F1 and zero filling to 8k was applied in F2. A cosine square window function was used for apodization prior to Fourier transformation in both dimensions.

A band-selective 2D TOCSY experiment[195] was designed starting from the Bruker pulse sequence *dipsi2gpphzs*. Two 2D band-selective TOCSY spectra were acquired focusing on methyl and methylene regions centred at 0.741 and 1.289 ppm, respectively, using the relaxation time of 2 s, mixing time of 80 ms, and

eight scans per increment. As these regions do not include solvent signals, no signal suppression was necessary. A zero-quantum suppression was achieved using 20 ms CHIRP pulse with a simultaneous PFG (11%) before and after mixing according to Thrippleton *et al.*[100]

The methyl focussed band-selective TOCSY experiment was acquired using the following parameters: 128 and 4k complex points in F1 and F2, respectively, spectral widths of 0.083 and 7.0 ppm in F1 and F2, yielding t1 and t2 acquisition times of 487 ms and 1.28 s, respectively. The overall acquisition time was 1 hr and 16 min. A forward linear prediction to 256 points was used in F1 followed by zero filling to 512 points. Zero filling to 8,192 was applied in F2. A cosine square window function was used for apodization prior to Fourier transformation in both dimensions.

The methylene focussed band-selective TOCSY experiment was acquired using the following parameters: 576 and 4k complex points in F1 and F2, respectively, spectral widths of 0.832 and 8.4 ppm in F1 and F2, yielding t1 and t2 acquisition times of 576 and 405 ms, respectively. The overall acquisition time was 4 hr and 14 min. A forward linear prediction to 1k points was used in F1. A zero filling to 8k points was applied in F2. A cosine square window function was used for apodization prior to Fourier transformation in both dimensions.

A magnitude mode *J*-resolved experiment (Bruker pulse sequence *jresqf*) was modified to include the presaturation elements as above. Spectra were acquired using 96 and 4k complex points in F1 and F2, respectively. Sixteen scans were acquired per increment using spectral widths of 40 Hz and 13 ppm in F1 and F2, yielding t1 and t2 acquisition times of 0.26 and 1.2 s, respectively. The overall experimental time was 2 hr and 20 min. A forward linear prediction to 256 points was used in F1, followed by zero filling to 512 points. A zero filling to 8,192 was applied in F2. A sine square window function was used for apodization prior to Fourier transformation in both dimensions.

A 1D CSSF TOCSY experiment[110] using DIPSI-2 spin-lock[194] and zero-quantum suppression[100] was modified to include solvent signal presaturation during the relaxation delay. The spin-lock at $\gamma B_1/2\pi$ equal to 10 kHz was applied from 10 to 160 ms. Selective inversion of a selected proton was achieved using a 40 ms Gaussian pulse applied on channel 1 after switching to the desired frequency from the frequency of the CH_2 ethanol signal used during the presaturation. The spectra were obtained with 32 increments (2.6 ms) of the CSSF (total length 83.2 ms) with two scans each. The 1D TOCSY spectra were acquired using 32k time domain points and a spectral width of 14 ppm, yielding an acquisition time of 1.46 s; a 2.0 s relaxation delay was used. The FID was zero filled once and Fourier transformed using exponential line broadening of 0.3 Hz.

An echo/anti-echo distortionless enhancement of polarisation transfer (DEPT) edited ^{1}H, ^{13}C HSQC experiment, Bruker pulse sequence *hsqcedetgpsisp2.3* was modified to include explicit adiabatic bilevel decoupling[196–201] and the solvent suppression as described for the acquisition of 1D ^{1}H NMR spectra. Standard Bruker broadband CHIRP ^{13}C pulses were used including 500 μs and 2 ms long inversion and refocusing pulses, respectively.

Analogous 2D ^{1}H, ^{13}C HSQC-TOCSY with DIPSI-2 spin lock but without the multiplicity selection was also modified (standard Bruker pulse program *hsqcdietgpsisp.2*) to include signal suppression. The heterocorrelated spectra were acquired using 1k and 3k complex points in F1 and F2, respectively. Eight scans (HSQC) or 16 scans (HSQC-TOCSY) were acquired per increment using spectral widths of 15 and 165 ppm in ^{1}H and ^{13}C dimensions yielding t1 and t2 acquisition times of 21 and 176 ms, respectively. The ^{13}C carrier frequency was set to 80 ppm. The spectrum was zero filled to 8k complex points in F2, and linear predicted with 512 coefficients to 2k complex points in F1. The cosine square window function was used before Fourier transformation; a 40 ms spin-lock was applied at $\gamma B_1/2\pi$ equal to 10 kHz in the HSQC-TOCSY experiment.

The same HSQC sequence was implemented at 800 MHz with the following modifications: 768 and 4k complex points in F1 and F2, with sweep widths of 165 and 16 ppm, yielding acquisition times of 11.6 ms and 160 ms, respectively. Four scans were acquired per increment. For processing of this spectrum, resolution was boosted by means of forward linear prediction in F2 from 4k to 8k complex points, and modified iterative soft thresholding (MIST) forward prediction from 768 to 2k complex points and subsequent zero filling to 4k in F1 in *MestreNova*.

A constant-time phase-sensitive gradient-selected ^{1}H, ^{13}C HMBC experiment with twofold low-pass *J*-filter to suppress one-bond correlations and no decoupling during acquisition,[202,203] Bruker standard pulse program *hmbcctetgpl2nd*, was modified to include signal presaturation during the relaxation delay, as described above. Additional ethanol CH_2 and water signal saturation ($\gamma B_1/2\pi$ equal to 28 Hz) were applied during the long-range evolution delay optimised for nJCH of 6 Hz. A two-step low pass J-filter was applied optimised for the $^{1}J_{CH}$ couplings of ethanol (125.7 and 142.6 Hz). The 90° ^{13}C pulses of the filter were implemented as composite 90°x- 90°y pulses. The 2D ^{1}H, ^{13}C HMBC spectra were acquired using 910 and 4k complex points in F1 and F2, respectively. Sixteen scans were acquired using spectral widths of 14 and 200 ppm in ^{1}H and ^{13}C dimensions yielding t1 and t2 acquisition times of 15 and 243 ms, respectively. The ^{13}C carrier frequency was set to 100 ppm. The spectrum was zero filled to 8k complex points in F2, and linear predicted with 512 coefficients to 2k complex points in F1. The cosine square window function was used before Fourier transformation.

3.1.3 *Other NMR Experiments*

For the experiments on acetaldehyde in subsection 5.4.3, the following experiments were used.

A ^{1}H DOSY experiment was performed at 600 MHz with the a time domain of 32k points and 8 diffusion points in a linear ramp from 5 % to 95 % gradient strength. The pulse sequence was a convection compensated Bruker sequence *ledbpgp2s.compensated*. Diffusion delays were big delta of 100 ms and little delta of 1.5 ms. 16 scans were acquired per increment with 4 dummy scans and a receiver gain of 16. Sweep width was 20 ppm giving acquisition time of 1.32 s.

A similar ^{1}H DOSY experiment was performed with compensation for chemical exchange using the PROJECTED sequence.[204] This was acquired also at 600 MHz with a time domain of 32k points and 16 diffusion points in a linear ramp from 5 % to 95 % gradient strength. Diffusion delays were little delta of 2 ms and 10 echoes each consisting 20 ms yielding a big delta of 200 ms. 16 scans were acquired per increment with 4 dummy scans and a receiver gain of 8. Sweep width was 11 ppm giving acquisition time of 2.47 s.

The ^{13}C DOSY experiment was acquired on a 500 MHz spectrometer equipped with He-cooled carbon-optimised probe and AVANCE III HD console, using the stimulated echo INEPT (Bruker sequence *stebpgpin1s*). After 4 dummy scans, 128 scans were acquired per increment, with 10 diffusion points linearly ramped from 5 % to 95 % gradient strength. Sweep width was 220 ppm yielding an acquisition time of 0.59 s. The receiver gain was set to maximum. The diffusion delays were big delta of 100 ms and little delta of 1.5 ms.

DOSY spectra were processed with GNAT.[205] This included zero filling to 64k with 1 Hz Lorentzian line broadening. Diffusion dimension was processed with 1k digital points. In ^{1}H spectra peak picking was performed above the level of the acetaldehyde CH3 ^{13}C satellites. The diffusion fit equation was a monoexponential.

A 1D ^{13}C spectrum was acquired for this acetaldehyde solution (Bruker sequence *zgpg30*). Spectrum was acquired on the 800 MHz spectrometer at a ^{13}C frequency of 201 MHz. Receiver gain was set to maximum. After 4 dummy scans, 1024 transient were co-added with 64k time domain points prior to zero filling once to 128k points, application of 1 Hz exponential line broadening and Fourier transformation.

3.2 FTICR MS

The following details sample information, instrumental parameters, and data processing related to chapter 6 and chapter 7.

3.2.1 *Sample Preparation*

Scotch Whisky samples were, again, provided by the SWRI. A total of 85 authentic Scotch Whisky samples were analysed, consisting of a mixture of malts and blends. The main sample set was the SWRI standard sample set from 2014 ($n_{malts} = 24$, $n_{blends} = 28$). Further samples were sourced from subsets of previous years' standard sample sets, *i.e.* 2010 ($n_{malts} = 23$, $n_{blends} = 1$) and 2012 ($n_{malts} = 8$, $n_{blends} = 1$). A malt new-make spirit sample was also analysed. Samples are referred to by an anonymised reference label of the format SYY-XXXX, where YY refers to the sample set year and XXXX is a unique identifier for the sample in that year.

In chapter 7, four samples from the 2014 standard sample set were used. These were a Sherry cask matured Single Malt (S14-1941), Blended Scotch Whisky (S14-1944), peated Single Malt (S14-1962), and a Bourbon cask matured Single Malt (S14-2196).

Methanol and water, both LC-MS grade, were purchased from Fisher Chemical. Nitrogen gas (99.998%) for APCI was purchased from BOC.

3.2.2 *Instrumental Parameters*

All mass spectra in this thesis was acquired on a SolariX FTICR MS (Bruker Daltonics) equipped with a 12 T actively shielded superconducting magnet and Infinity ICR cell.

3.2.2.1 *ESI*

For ESI analysis in chapter 6, the following settings were used. Nebuliser gas flow was set to 1.8 bar, drying gas was 6 L min^{-1} at 180 °C. The capillary voltage was set at 4 kV and the end plate offset at 500 V. Samples were infused at a flow rate of 120 µL h^{-1}. Broadband spectra were acquired with 200 co-added transients between m/z 98.3 and 1000 into a 4 MWord time domain with transient length of 1.1185 s. The Q1 mass was set to m/z 100. Time of flight was set to 0.6 ms with an ion accumulation time (IAT) of 150 ms. For acquisition of narrow windows for isotopic pattern analysis, ions were filtered using the quadrupole, and

the ion accumulation time was increased to 2000 ms. For fragmentation experiments, specific ions were isolated using the quadrupole and fragmented using CID in the collision cell (MS2) or source and collision cell (MS3). Solvent blanks were run after every 12 samples and at the start and end of each experimental session. Three samples were analysed in replicate across different days to validate the instrumental reproducibility. Spectra were acquired in negative mode. Samples were prepared 1:10 dilution into 50:50 methanol/water, using 30 μL of raw sample per analysis.

In chapter 7, ESI spectra were acquired in a similar fashion as described previously. Samples were infused at a flow rate of 200 $\mu L\,h^{-1}$ and ions were accumulated for 150 ms. Spectra were acquired between m/z 98.3 and 2000 using a 4 MWord time domain with transient length of 1.1185 s.

3.2.2.2 *APCI*

For APCI analysis, spectra were acquired with a corona needle discharge of 4 μA. The capillary voltage set at 4.5 kV and end plate offset at 500 V. The nebuliser pressure was 2 bar at 370 °C, with dry gas flow of 3.5 $L\,min^{-1}$ at 220 °C. Samples were infused at a flow rate of 400 $\mu L\,h^{-1}$ and ions were accumulated for 300 ms. APCI nebuliser gas was supplied from a nitrogen gas cylinder (99.998%, BOC). Samples were prepared 1:1 dilution into 50:50 methanol/water. Spectra were acquired in negative mode.

3.2.2.3 *APPI*

For APPI analysis, spectra were acquired with a source equipped with a krypton lamp at 10.6 eV. The capillary voltage was set at 1.5 kV and end plate offset at 500 V. The nebuliser pressure was 2.5 bar at 400 °C, with dry gas flow of 4 $L\,min^{-1}$ at 220 °C. Samples were infused at a flow rate of 2 $mL\,h^{-1}$ and ions were accumulated for 600 ms. Samples were infused neat, with no addition of solvents. Spectra were acquired in negative mode.

3.2.2.4 *LDI*

For LDI analysis, spectra were acquired with a MALDI source fitted with a solid-state 1 kHz SmartBeam TMII laser. Laser focus was set to minimum, with 1000 shots fired at a rate of 1000 Hz per scan. Laser power was tuned to the minimum required to observe signal, however this was still high (up to 100%). Spectra were acquired with a selective accumulation of transients based upon a minimum and maximum total ion current (TIC) threshold. Samples were spotted (1 μL) dir-

ectly onto a MALDI plate and allowed to air dry. This was repeated one or more times to ensure sufficient spot thickness. Spectra were acquired in negative mode.

3.2.3 *FTICR MS Analysis*

All spectra were acquired using ftmsControl 2.1 (Bruker Daltonics) and processed according to the default parameters, including zero-filling once and full sine apodization prior to Fourier transformation, unless otherwise stated. Spectra were internally re-calibrated using a quadratic function and peak picked in Data Analysis 4.4 (Bruker Daltonics). Spectra were peak picked with a SNR threshold of 4 and a minimum absolute intensity of 2×10^6, a method suggested by Riedel and Dittmar.[206]

For formula assignment in chapter 6, the peak lists were exported as text files to PetroOrg S-10.2 (The Florida State University). In chapter 7, formula assignment was also performed with Formularity[207] and in-house developed scripts. Further discussion of formula assignment is made in chapter 7.

3.3 CHEMOMETRICS

All PCA and OPLS-DA analysis was performed using SIMCA 14.1 (Sartorius, formerly Umetrics). All other data analysis was performed using Python and in-house written scripts or publicly available packages.

For NMR chemometric analysis, data were collated into a table of spectra using MestreNova to bin the data. The binning was uniform. Residual solvent regions were excluded or zeroed prior to analysis. This is discussed further in chapter 5.

For FTICR MS chemometric analysis, a data table was constructed for n samples (observations) against m variables (assigned molecular formulae) using peak intensities, where NA values (peaks not found in a given sample) were filled with a random value at the level of the noise. Thus, this results in no formulae being excluded from the model. Data were then normalised, per spectra, to a sum total intensity of 1. Data were then mean centred and Pareto or auto scaled prior to PCA or OPLS-DA model construction.

Part II

RESULTS AND DISCUSSION

Served Straight Up

4

NMR METHOD DEVELOPMENT

This difficult science is formed slowly,
but it preserves every principle which it has once acquired.

— Joseph Fourier (1878)

4.1 DECLARATION

Most of the work in this chapter was previously published in the following paper:

> W. Kew, N.G.A. Bell, I. Goodall, and D. Uhrín, *Magnetic Resonance in Chemistry*, 2017, **55**, 785-796

The research and experiments described in this chapter were performed in collaboration with Prof. Dušan Uhrín and Dr Nicholle Bell. Custom scripts for automating acquisition were initially developed by Boris Mitrovich, and customised by myself as noted in the text.

4.2 INTRODUCTION

To be able to acquire NMR spectra of many Scotch Whisky samples over the course of this PhD, it was first necessary to develop an NMR methodology which would allow for the automated acquisition of high-resolution spectra - the challenge, of course, being the very high levels of water and ethanol present in solution.

NMR is an inherently quantitative technique, and the highly abundant 'solvents' in whisky (ethanol and water) dominate the spectra and severely limit its interpretability for two main reasons. Firstly, the intense solvent signals overlap with signals of minor components and obscure them. Secondly, to avoid a receiver overflow from these intense signals, the receiver gain must be reduced. The ADC has a limited dynamic range, and with a low receiver gain, compounds with low abundance will not be sampled properly and their intensity will be at the noise level. As such, a solvent suppression method was required. This chapter details the development of an advanced solvent suppression technique which is used throughout later chapters for the acquisition of NMR spectra of hundreds of samples. For further background on the theory, and other examples, of solvent suppression please refer to section 1.2.

4.3 SAMPLE PREPARATION

A minimal sample preparation routine was developed. Whilst it would be technically possible to acquire NMR spectra of Scotch Whisky with no sample preparation, for a variety of technical reasons this is not optimal. Without the addition of deuterium it would not be possible to 'lock' the spectrometer frequency. This would result in a degradation in spectrum quality, especially for longer acquisitions such as in 2D experiments where magnetic field drift would negatively affect the quality of the spectra. Additionally, chemical shifts are concentration, solvent, and pH dependent. Every sample of Scotch Whisky will have a unique combination of congener concentrations, exact ethanol strength, and pH. The addition of a sodium acetate/acetic acid buffer somewhat compensates for the pH variation between samples. This could be, and indeed is, added at the same time as the D_2O, required for lock. Ethanol strength standardisation, however, would require accurate measurement of the ethanol strength in the first place. The decision was made, therefore, to not standardise this aspect of the samples. Finally, the addition of DSS-d_6 allows for a consistent chemical shift reference as the default spectrum reference to the water signal - *ca.* 4.7 ppm - would be variable according to ethanol strength and pH. Any residual chemical shift differences between samples are addressed post-acquisition and dealt with using peak-alignment techniques as discussed in subsubsection 5.6.1.1.

4.4 SOLVENT SUPPRESSION

Whilst there exist a variety of solvent suppression techniques already, it was deemed necessary to develop our own. Solvent suppression is most commonly applied to water samples in biological NMR contexts, *e.g.* proteins, DNA, or metabolomics. In the case of Scotch Whisky, there is the additional complication of ethanol at a minimum 40 % ABV. At the onset of this PhD, there were a couple of published schemes for the automated suppression of ethanol and water (see section 1.2). However, these either did not decouple the carbon satellites of ethanol,[76] or where they did decouple, they were developed for room temperature probes only.[78] On a cryogenically cooled probe, it is undesirable to decouple carbon for an extended period of time due to heating of the sample and probe. Heating of the sample can cause convection within the tube and affect magnetic field homogenity, whilst excessive power applied for a long time is potentially damaging to the hardware. Typically, experiments are limited to *ca.* 180 ms for continuous broadband carbon decoupling on a cryoprobe - as per manufacturer recommendations. In the case of high-resolution 1D ^{1}H NMR, acquisition time

may last for several seconds. Thus, it was necessary to selectively decouple carbon signals with a low power pulse, which avoids the issue of heating. Overall, therefore, the criteria was to suppress three signals - the OH from water and ethanol, and the CH_2 and CH_3 of ethanol, as well as to decouple the ethanol signals. As exact ratios of solvents, and thus chemical shifts, will vary sample-to-sample, it is necessary to optimise the suppression parameters per-sample, and automatically if the methodology is to be suited to high-throughput application.

Normally, determination of signal frequencies is made simply by acquiring the relevant NMR spectra and peak picking. Unfortunately, a typical single-pulse 90° or 30° 1D ^{1}H NMR spectrum would exhibit receiver ADC overflow, even with a minimal RG. Furthermore, due to the high Q factor of cryogenically cooled probes, even an extremely short RF pulse, such as 0.4 µs, produces radiation damping. Therefore, different initial experiments are a requirement to locate the exact frequencies of ethanol and water.

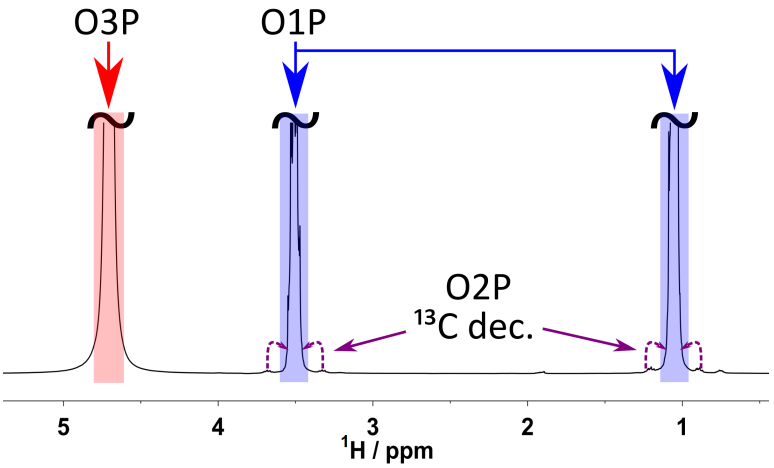


Figure 4.1: 1D ^{1}H NMR spectrum of Scotch Whisky showing the schematic of solvent suppresion. O1P sits on resonance of the ethanol CH_2 signal and presaturates CH_2 and CH_3 signals. Simultaneously, the carbon channel (on carrier frequency O2P, at the midpoint between the ethanol carbon signals) decouples both ^{13}C signals from the protons of ethanol, collapsing their signals to the midpoint of each multiplet. They are then presaturated by the O1P suppression. The third channel is also on ^{1}H and presaturates the OH signal of water and ethanol. This spectrum was acquired at 400 MHz on a room temperature probe.

The basis for the solvent suppression experiment is a modified 1D NOESY type sequence with presaturation during the recycle delay. This experiment has been widely adopted in the metabolomics community, and as this research is not dissimilar from that area, it was an appropriate choice. The final acquisition involves presaturation of the OH signal on the third channel of the spectrometer with a continuous, low power pulse. The second channel is used for a selective carbon

decoupling pulse during this period, and the first channel applies a selective presaturation pulse to the ethanol CH_2 and CH_3 signals. The decoupling of the carbon signals allows for their proton resonances to be saturated by the first channel ethanol suppression. A schematic of this methodology is shown in Figure 4.1. To determine the optimal power levels, pulse durations, and shaped pulse modulations, three precursor experiments are required.

4.4.1 *Pulse Calibration and Water Suppression*

The first preliminary experiment (Expno 10, AU program *efindwater2.wk*) serves to locate the exact resonance frequency of water and to calibrate 90° ^{1}H pulses on the first and third channels. The water signal is located using a 360° pulse in a single scan experiment. The backwards linear prediction of the first 4,096 points used during the processing removes the fast decaying broad component of the water signal, allowing for the accurate determination of the frequency of the sharp component. The determined value is then used to irradiate the water signal on the third channel. Interestingly, it was found for high alcohol strength samples (ABV $>$ 50 %), back prediction of the first 4,096 points removed all of the water signal. In these cases, only the first 512 points are back predicted. The exact cause of this faster relaxing OH signal is not known, but may be related to chemical exchange between water and ethanol OH protons, and differing hydrogen bond strengths. The peak picking of the water resonance frequency is performed between 4.2 and 5.2 ppm, and the result is routinely around 4.7 ppm.

For the final acquisition, this is all that is required of this first precursor stage as the water signal will be suppressed by a continuous, low power pulse on resonance. However, for a quality check, an additional spectrum is acquired in a simple pre-saturation experiment (Bruker pulse sequence *zgpr*). This acquires a 1D ^{1}H spectrum with water suppressed and a minimal flip angle pulse (0.4 µs, approximately 4.5°), resulting in a spectrum containing the OH signal significantly attenuated, with the full intensity ethanol CH_2 and CH_3 signals. An example of the spectrum acquired from this experiment is provided in Figure 4.2.

Should the OH resonance have been mislocated, suppression will be unsuccessful and immediately apparent. In manual acquisition mode, the operator could resolve this issue at this point. In automation, it serves as a debugging point. Additionally, radiation damping distorts the ethanol CH_2 and CH_3 signals, however their ^{13}C satellites can be inspected for a measure of shimming quality.

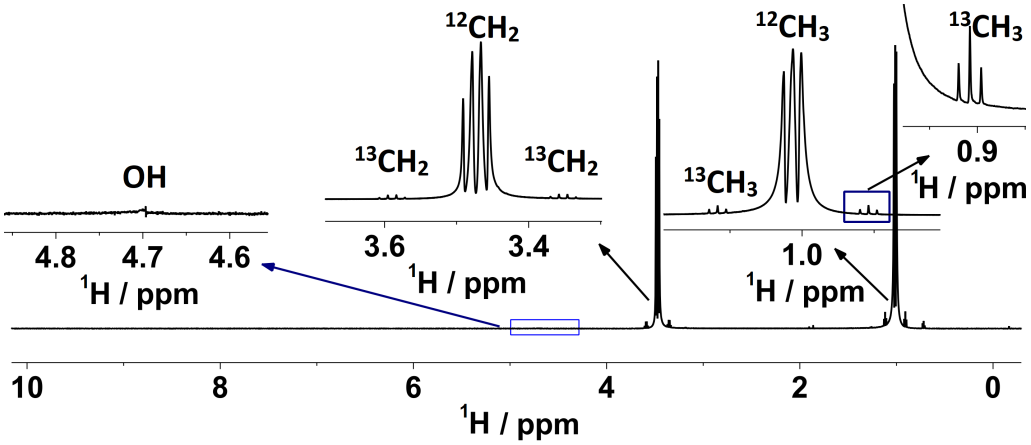


Figure 4.2: 1D ^{1}H NMR spectrum with OH signal of water and ethanol suppressed. Expansions inset show key regions of spectrum (left-to-right): residual OH signal, CH_2 quartet and satellites, CH_3 triplet and satellites, rightmost CH_3 satellite.

4.4.2 *Ethanol Shape Generation*

The second experiment (Expno 11, Python script *createEthanolShape_ wk.py*) determines the exact chemical shifts of CH_2 and CH_3 signals of ^{12}C isotopomers of ethanol by first determining the equivalent chemical shifts in ^{13}C isotopomers by means of a non-refocused reverse INEPT experiment.[208,209] This approach was chosen as radiation damping prevents accurate determination of the chemical shifts of ethanol's protons in the ^{12}C isotopomers. Radiation damping is not an issue for their much weaker ^{13}C satellite signals, which can be acquired as the first trace of a non-refocused gradient-selected HSQC[210] or, in this case, a non-refocused reverse INEPT[211] experiment.

Both of these methods transfer magnetisation from the naturally low abundant ^{13}C nuclei to directly bonded ^{1}H, and thus suppress the major ^{12}C isotopomer signals which are not NMR active. In the case of the HSQC experiment, the magnetisation starts on ^{1}H, whilst in the reverse INEPT it starts on the much lower abundant ^{13}C nuclei. Refocusing of the magnetisation was not necessary in these experiments.

The HSQC method produced distorted multiplets affected by the evolution of proton-proton couplings, whereas the reverse INEPT yielded pure anti-phase multiplets with respect to $^{1}J_{CH}$. A comparison of the HSQC and INEPT-based methods is shown in Figure 4.3.

Despite the inherently lower sensitivity of the INEPT method, a SNR > 4,000 was routinely obtained, which is more than sufficient for accurate frequency determination by peak picking. The main signals were suppressed significantly

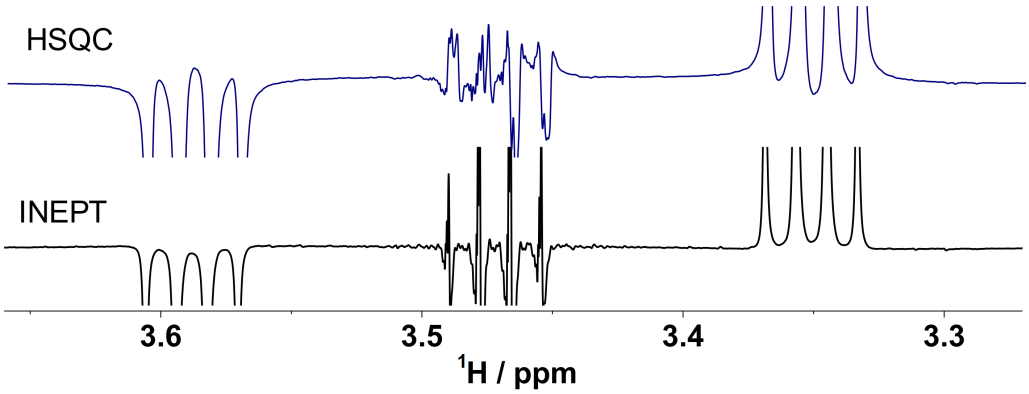


Figure 4.3: Partial 1D ^{1}H reverse INEPT (bottom) and HSQC (top) NMR spectra showing the CH_2 signals of ethanol. ^{12}C bound ^{1}H signals are suppresed whilst ^{13}C bound ^{1}H signals are anti-phase. The HSQC spectrum shows a distorted baseline and multiplets affected by J evolution, whilst the INEPT has pure phase signals.

(<25 % of the height of the satellite signals), which was sufficient to allow an increased RG allowing for adequate digitisation. An example of the spectrum acquired from this experiment is provided in Figure 4.4.

The chemical shifts of ^{12}C isotopomers of ethanol were therefore obtained on the basis of the frequency of ^{13}C isotopomers and considering the ^{13}C/^{12}C isotope shift. The values of these isotope shifts were experimentally determined to be −1.53 Hz and −1.11 Hz for CH_2 and CH_3, respectively. In practice, this is performed with a simple Python script. In short, the Python routine executes TopSpin's built-in peak picking (`XCMD("PPL")`) and then loads this list into memory as an array. The peaks are filtered to include only the 10 most intense peaks, and then the centre of each sub-multiplet (quartet or triplet) is determined, followed by the centre of each overall multiplet (CH_2 or CH_3), isotopic shift corrections applied, and finally the distance between the centre of these multiplets is determined. The resulting distance, in Hz, is then multiplied by a factor for n 2π rotations. At 600 MHz, n = 36, and at 800 MHz, n = 50, were used. Note that, whilst there should be 14 signals peak picked (6 (2 x triplet) + 8 (2 x quartet)), only the top 10 intense peaks are used. This is because the four least intense peaks (the outer parts of the quartets) may approach the level of the residual ^{12}C isotopomer and result in erroneous peak picking. For debugging, a non-fatal error message is produced if too many or too few peaks are picked.

Peak picking of the ^{13}C isotopomers was performed in automation, with only specific chemical shift ranges allowed to ensure that the signals of the ^{12}C isotopomers were not peak picked. For variable temperature experiments, these ranges required modification. The location of these ethanol CH_2 and CH_3 signals allows for the generation of the required doubly-selective shaped pulse (*Ethan-*

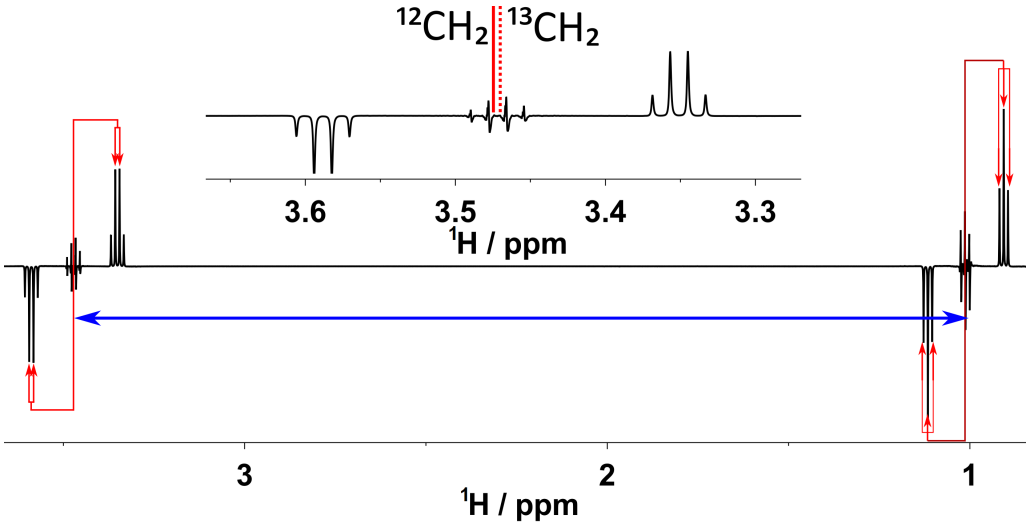


Figure 4.4: 1D ^{1}H reverse INEPT NMR spectrum. ^{12}C bound ^{1}H signals are suppresed whilst ^{13}C bound ^{1}H signals are anti-phase. Annotations show the peak picking locations and subsequent frequency and distance determinations as described in the text. Blue line shows the distance between the midpoint of the two multiplicities. Inset expansion of the CH_2 region shows the quality and resolution of the spectrum, as well as visualising the effect of isotopic shift (not to scale).

olShape) for their presaturation. Its exact length is calculated for each sample, taking into account the exact chemical shift difference between the CH_2 and CH_3 signals and allowing the off-resonance signal (CH_3 here) to make exactly a multiple of 2π rotations. For more technical detail on the generation of this shaped pulse, see section 4.6.

This precision is essential for high quality suppression of the off-resonance signal as illustrated in Figure 4.5. Here, at 600 MHz, for a difference of 1.470 Hz between the CH_2 and CH_3 signals, a 50 094 µs pulse produces 36 2π rotations of the off-resonance CH_3 signal (Figure 4.5 far left). When the pulse length was recalculated to add additional rotation in 20° increments, the suppression of the CH_3 signal became progressively worse. Lengthening the pulse by as little as 38 µs, or 0.076 % of the original pulse length, which corresponds to the time needed to generate such a 20° phase shift, has a noticeable effect.

The exact rotation of only one off-resonance signal can be achieved in this way; therefore, another channel was used to suppress the water signal as described in subsection 4.4.1. However, for instruments with fewer channels, it is possible to frequency modulate a shaped pulse to suppress more than two resonances simultaneously. Indeed, this is the approach of pre-existing experiments including that by Monakhova *et al.*[76]

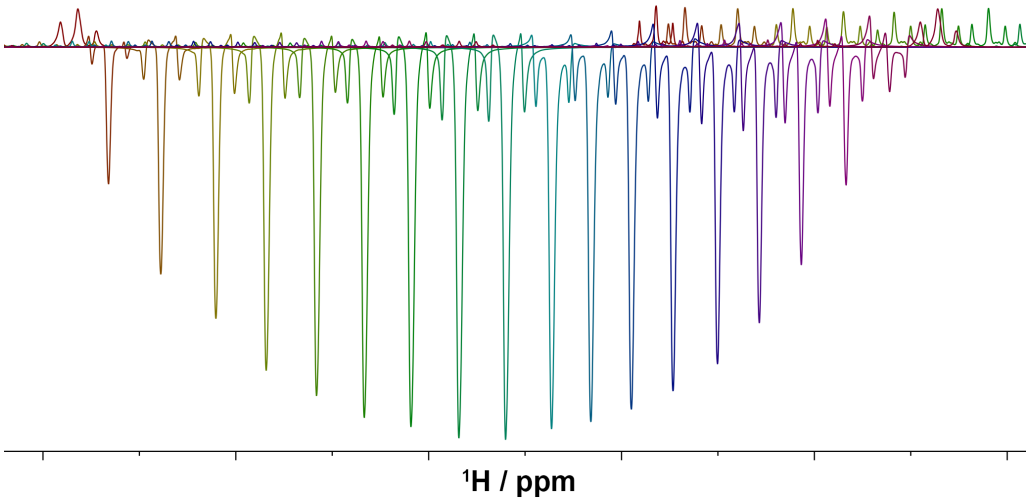


Figure 4.5: Residual signal of CH_3 protons of ethanol obtained using the pulse sequence of Figure 4.8, as a function of the pulse length of the presaturation pulse. On the left, the pulse length, 50 094 µs, was calculated to allow 36*2π rotations, whereas the subsequent experiments increased this value by 38 µs. Each recalculated shape allowed for an additional phase shift increment of 20°. The range of 360° is shown in 17 increments. Note that the central line of the CH_3 triplet is disproportionately taller.

4.4.3 *Ethanol Decoupling*

The third and final preliminary experiment (Expno 12, Python script *createCarbonShape.py*) was a 1D ^{13}C NMR spectrum acquired using inverse-gated decoupling (Bruker pulse sequence *zgig*), which applies decoupling only during acquisition. Typically, power-gated decoupling is used for acquisition of ^{13}C NMR to increase signal strength by means of the NOE. However, this effect can impact on quantification. Furthermore, it was not necessary to boost sensitivity as the ethanol is at least 6.85 M in solution. The major reason for acquiring 1D ^{13}C NMR spectra is to determine the exact ^{13}C chemical shifts of ethanol resonances. The pulse sequence is performed in a single scan and the two carbon signals are well resolved with a SNR > 740:1 at 43 % ABV. An example spectrum is shown in Figure 4.6.

This information is required for setting up the parameters for the decoupling of ^{13}C satellites of ethanol in the ^{1}H spectrum of whisky. As previously discussed, it is not possible to apply broadband ^{13}C-decoupling during a multi-second relaxation delay or acquisition of ^{1}H spectra on a cryogenically cooled probe. Therefore, decoupling of ethanol signals was performed in a selective manner. For this purpose, a cosine modulated rectangular pulse, *CarbonShape*, of approximately 2.5 ms was generated at the frequency midway through the two ethanol signals.

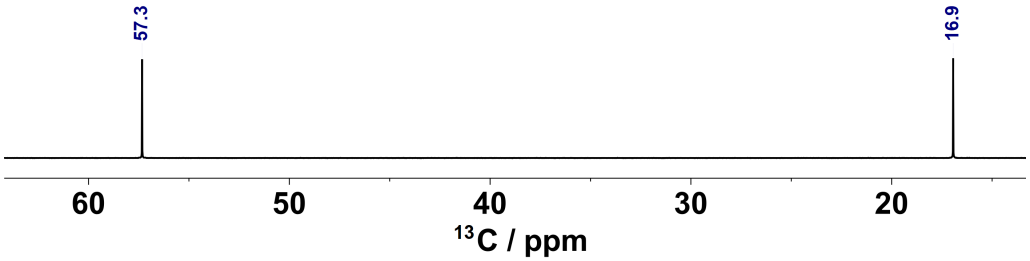


Figure 4.6: 1D ^{13}C NMR spectrum showing the two carbon signals of ethanol.

As with the ^{1}H presaturation pulse, its exact length was calculated to allow n*2*π rotations of both signals. Likewise, the shape was generated using a Python script, which peak picked the spectrum, measured the distance between these two signals in Hertz, and calculated the required pulse length. At 600 MHz, n = 6, and at 800 MHz, n = 8. Technical details on the calculation of this shaped pulse are discussed further in section 4.6. This selective decoupling resulted in a residual signal <1 % of the original ^{13}C satellite signal height (Figure 4.7). It was not necessary to apply decoupling during the acquisition.

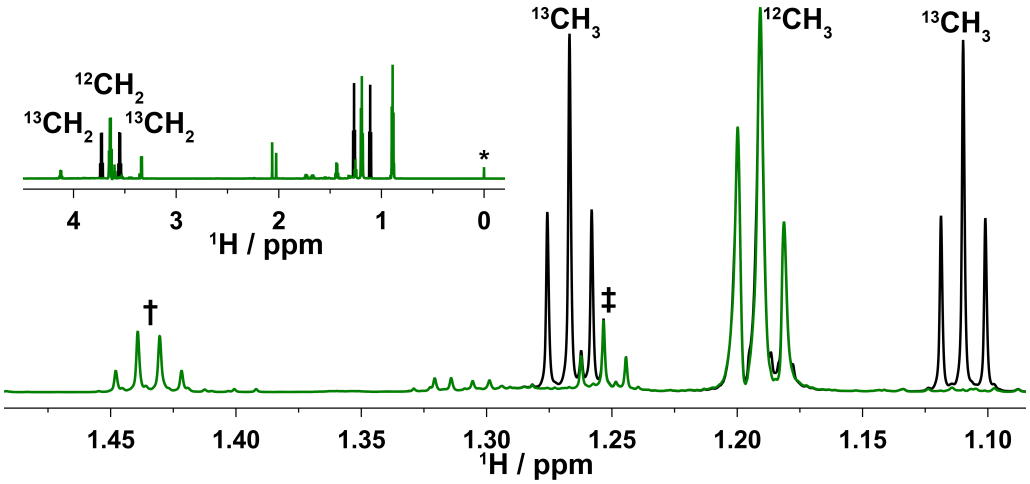


Figure 4.7: Partial 1D ^{1}H NMR spectrum of Scotch Whisky (S15-3896) acquired at 800 MHz showing the effect of ^{13}C decoupling on (green) and off (black) of the ethanol CH_3 signal. These spectra are acquired with ethanol and water suppression. Single dagger (†, 1.43 ppm) indicates the signals of 3-methylbutan-1-ol and double dagger (‡, 1.255 ppm) indicates a triplet from ethyl acetate. Both pairs of signals overlay perfectly in the two spectra, indicating a lack of heating from the decoupling. Also, the triplet ‡ becomes clearer when decoupling is activated.

This spectrum also serves a dual purpose. These high ethanol concentrations yield single scan spectra with a high SNR that are suitable for the quantification of the ethanol concentration. Quantification of ethanol, including by means of the reverse INEPT spectrum, is discussed further in subsection 5.5.1.

4.4.4 *Final Acquisition*

The combination of parameters and shaped pulses generated in the three preliminary experiments allows for optimal acquisition of the final ^{1}H NMR spectrum. This experiment (experiment number 13, Python script *AcqWhisky.py*) is acquired with the pulse sequence as shown in Figure 4.8.

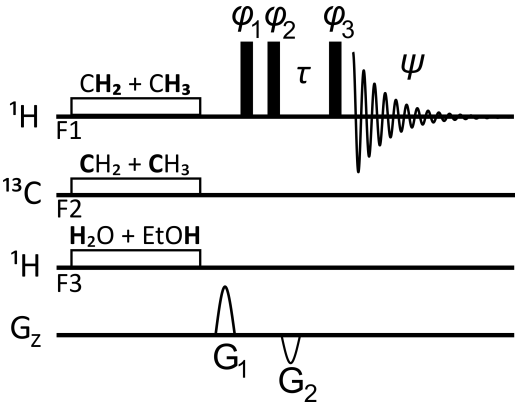


Figure 4.8: Pulse sequence for simultaneous suppression of water, ethanol and ^{13}C ethanol satellites in whisky samples. Narrow-filled rectangles represent 90° pulses. The 1 ms CHIRP shaped PFG were applied at G1 = 50% and G2 = -11% followed by a 200 μs gradient recovery delay resulting in a τ period of 1.2 ms. The following phase cycle was used ϕ 1 = x, -x; ϕ 2 = 8(x), 8(-x); ϕ 3 = 2x,2(-x), 2y,2(-y), and ψ = x, 2(-x), x, y, 2(-y), y, (-x), 2x, (-x), (-y), 2y, (-y).

The spectrum is acquired with presaturation and decoupling during the relaxation delay. The carrier frequency of channel 1 is set to that of the CH_2 signal of ethanol, and the *EthanolShape* is pulsed here. Channel 2 performs the selective ethanol decoupling with *CarbonShape*, and channel 3 performs a low power continuous irradiation for presaturation of the OH signal. Full parameters are provided in subsubsection 3.1.2.1. Note that the relaxation delay is 4.05 s and the acquisition time is 6.82 s for a spectral width of 16 ppm in 128k time domain points at 600 MHz. As the presaturation pulse element lasts only 50 ms, the pulse sequence loops over the suppression element continuously for the length of the relaxation delay. For acquisition, the RG is set to 45.2 to allow for direct comparison of spectra. This is a belts-and-braces approach to quantification, as all samples have been prepared identically, and have an internal standard present. A typical resulting acquired spectrum is shown in Figure 4.9.

This 1D ^{1}H NMR spectrum Scotch Whisky shows remarkable suppression of water/ethanol signals with residual 'solvent' signals smaller than the overlapping CH_3 signals of higher alcohols at 0.88 ppm (Figure 4.9). The residual CH_2 signals of ethanol are smaller than the signal of nine CH_3 protons of 1 mM DSS used as internal standard. The level of the ethanol signals suppression can be

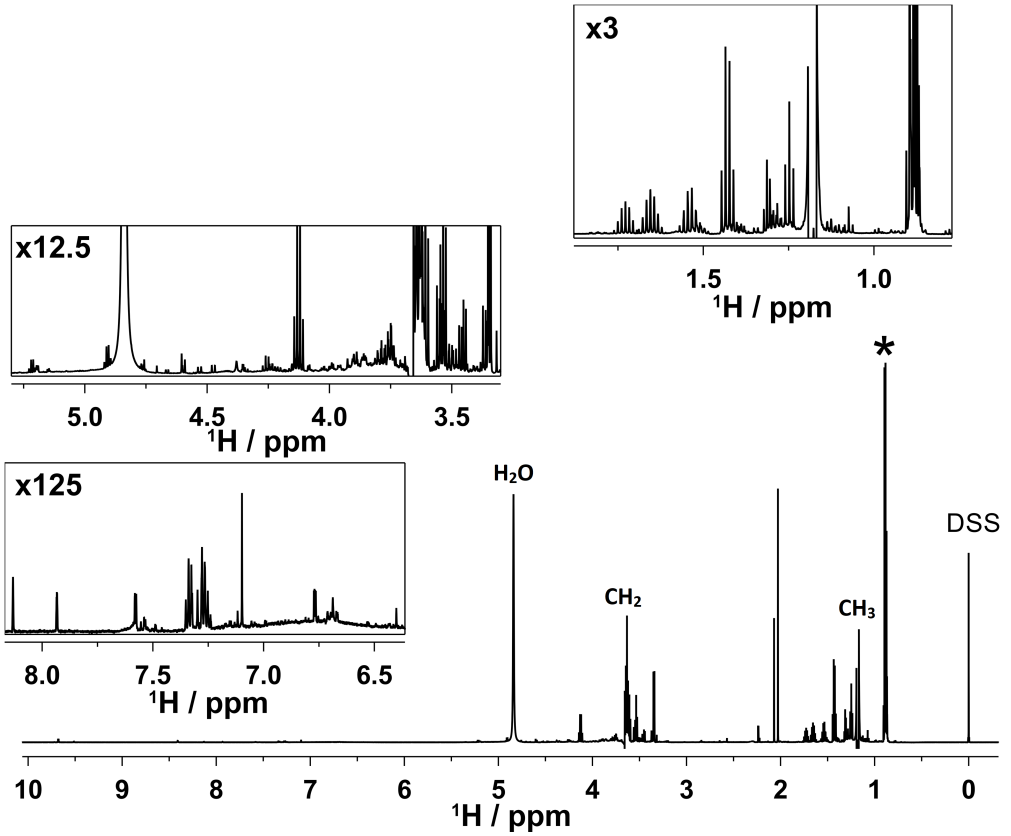


Figure 4.9: Typical 1D ^{1}H NMR spectrum of Scotch Whisky (S14-1941) with OH signal and CH_2 and CH_3 signals of ethanol suppressed. Internal standard DSS (1 mM, 0 ppm) is annotated. Asterisk indicates overlapping methyl signals from several major congeners. Insets show regions of interest including higher alcohols (0.8-1.8 ppm), carbohydrates (3-5.5 ppm) and aromatic compounds (6.5-8.5ppm)

quantified by recording a spectrum without the carbon decoupling (Figure 4.7). Here, the residual ^{12}C isotopomer ethanol signals have approximately the same intensity as their satellites, representing suppression of ≈99.5 % of the signal. The CH_2 signals is typically better suppressed as its chemical shift coincides with the carrier frequency, whereas the CH_3 is off-resonance. The water signal is typically suppressed to that of the level of the DSS signal.

The final spectrum shows the signals of many different compounds clearly resolved and potentially identifiable and quantifiable, displaying pure phase multiplets as shown in Figure 4.9. The limit of detection was determined to be approximately 50 µM. This could be improved by increasing the number of scans or using a higher field instrument. In reality, however, both higher field and more scans offer only marginal improvements that would be difficult to justify economically, and thus this detection threshold is as low as practically possible for routine NMR of Scotch Whisky today.

The developed method exceeds the reported performance of similar approaches in the literature, for example the work of Monakhova *et al.*[76] One metric of performance is the improvement in receiver gain achieved by means of solvent suppression - in this case the method presented here achieved a receiver gain of 45.2, compared to the value of 28.5 reported by Monakhova *et al.*[76] Whilst different spectrometers were used, they were both Bruker Avance-era consoles, so the receiver gain values are comparable. An alternative means to show the improvement is to compare the SNR of a minor compound in the non-suppressed and suppressed experiments. Monakhova *et al.* showed a ≈10× improvement in SNR between their initial *zgpr* experiment (RG of 1, ns =8) and their fully suppressed one.[76] This was measured for the methanol content in their samples, however methanol is at a low level in Scotch Whisky, and not routinely observed in the initial water presaturation experiment here. As such, the SNR for DSS (0 ppm) was chosen to compare. For sample S14-1941, DSS had a SNR of 38 in the water presaturation experiment, and 6435 in the final acquisition, representing a 170-fold improvement in SNR. However, this should be normalised against the different number of co-added transients (2 and 32). The theoretical improvement in SNR from just the increase in number of scans is 4× ($= \frac{\sqrt{32}}{\sqrt{2}}$), and thus the solvent suppression yields a true 42.5× improvement - four times greater than that achieved by Monkahova *et al.*[76] Additionally, the method presented by Monkahova *et al.*[76] does not suppress the ^{13}C satellites of ethanol, was not implemented for 2D NMR experiments, and its performance on cryogenically cooled probes is unknown.

4.5 EXTENSION TO OTHER NMR EXPERIMENTS

Despite the obvious usefulness of the solvent suppressed 1D ^{1}H NMR spectra of Scotch Whisky, confident structure elucidation requires more sophisticated NMR experiments to be utilised. Although these experiments may be more time consuming and less sensitive, they can be used for a selection of Scotch Whisky samples to provide more detailed information about the compounds present. In this section, the adaptation of several commonly used experiments to include whisky solvent suppression is detailed.

4.5.1 *Homocorrelated NMR*

Most commonly and accessibly, in terms of sensitivity and often simplicity, homocorrelated NMR experiments provide a crucial means towards the structural elucidation of compounds. This involves first establishing proton-proton chemical

shift correlations - *i.e.* which signals in the ^{1}H NMR spectra are part of the same molecule or spin system. This is typically achieved by using *J*-coupling constants to transfer the magnetisation between nearby protons in COSY or TOCSY type experiments. Alternatively, through-space correlations could be used to inform structural elucidation; however, the usefulness of this for small molecular weight congeners is limited, and NOESY-type experiments were therefore not included here.

Acquisition of COSY spectra in aqueous solutions is challenging,[212] and therefore, a basic gradient-selected 2D COSY method (Bruker pulse sequence *cosygpppqf*) was adapted here. This allows for the simple acquisition of spectra with sufficient sensitivity. The use of gradient-selection reduces the need for complex phase cycling to remove undesirable signals. However, this experiment results in data which cannot be phased when processed, and requires the use of severe window functions which reduces the sensitivity.

For acquisition of 2D TOCSY spectra a phase-sensitive method (adapted from Bruker sequence *dipsi2esgpph*), using DIPSI-2,[194] was chosen. Unlike the COSY experiment, this TOCSY experiment results in pure-phase absorptive line shapes, allowing for a better theoretical resolution.

Excellent results were achieved for both methods as seen in Figure 4.10. The quality and level of the solvent suppression can be judged by inspecting the vertical projections positioned at the sides of the 2D spectra to be similar to that achieved in the 1D ^{1}H NMR spectrum (see Figure 4.9). The methyl signals of higher alcohols resonating at 0.88 ppm are more intense than the residual signal of the methyl protons of ethanol. The residual solvent signals do generate t1-noise in their three locations in the 2D spectra; however, this is at an acceptable level, and does not prevent spectral assignment.

In both the COSY and TOCSY spectra, several correlations, between aliphatic protons resonating in a 1 ppm region, between 0.8 – 1.8 ppm, and the methyl protons around 0.88 ppm or other protons up to 5.2 ppm, can be seen. These protons belong to compounds including higher alcohols, such as 2- and 3-methylbutan-1-ol, n-propanol, iso-butanol, and ethyl acetate. For more details on compound assignment, see section 5.4.

Despite adding a dimension to the NMR spectra, by going from 1D to 2D, not all signals are fully resolved. In complex mixtures, molecules containing very similar fragments are present at a range of concentrations. This can lead to strong signals concealing weaker ones and preventing their identification, especially when wide proton multiplets are present. In the NMR spectra presented thus far, none of the proton-proton couplings have been removed. However, their decoupling in one dimension would be highly desirable - increasing resolution in

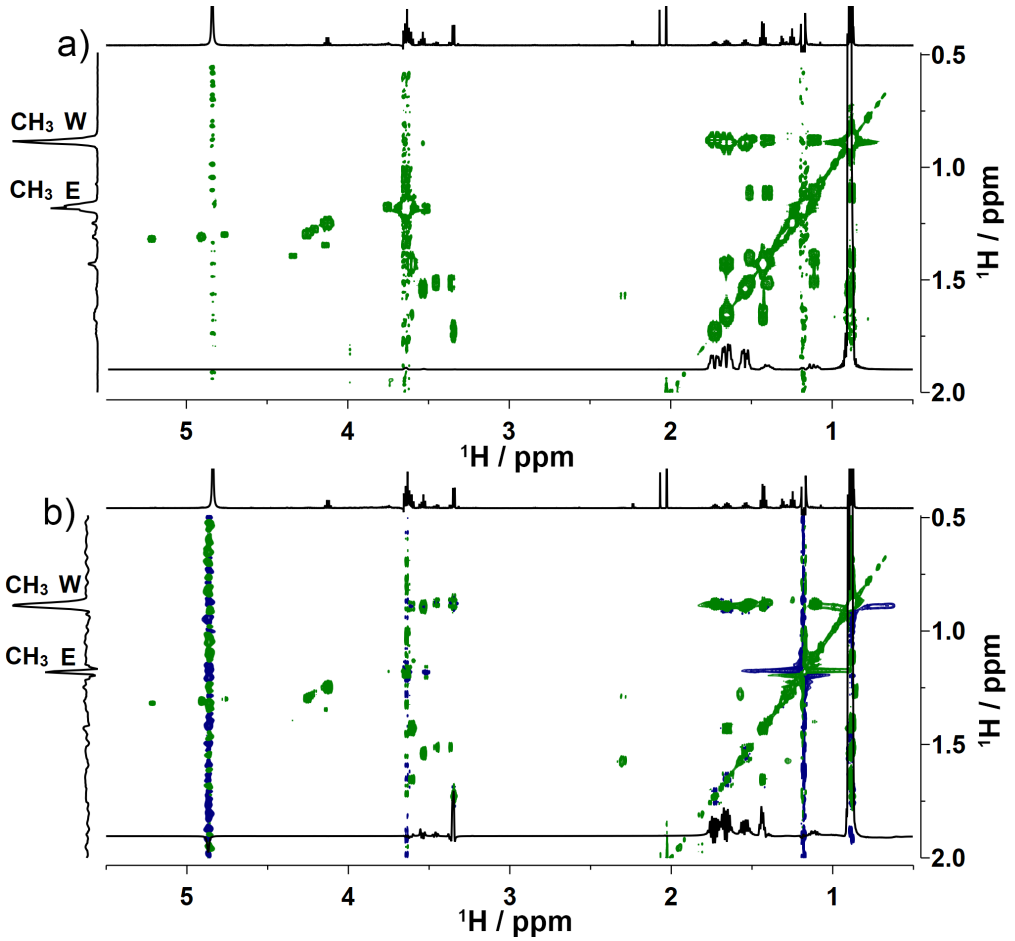


Figure 4.10: Partial (a) 2D COSY and (b) 40 ms mixing time 2D TOCSY spectra of Scotch Whisky (S14-1941) acquired using the developed solvent suppression scheme of Figure 4.8. A 1D spectrum is shown at the top, scaled to show appropriate features, whereas vertical projections are shown on the sides. Hortizonal traces at 0.88 ppm are overlayed to show the quality of the spectrum. Methyls of ethanol (E) and other compounds in whisky (W) have been labeled on the vertical projections.

that dimension whilst preserving their multiplicity in the other, thus retaining the important information provided by *J* couplings. Particularly suitable for low concentration samples are band-selective experiments that increase the sensitivity through resolution improvement. Towards this end, we have implemented a solvent suppression scheme into a band-selective homodecoupled (BASHD) 2D TOCSY experiment,[195] achieving excellent resolution as illustrated in Figure 4.11.

Band-selective experiments involve, as the name suggests, selective irradiation - in this case inversion for *J* refocusing. By selectively inverting spins within a narrow band, both they and their coupled partners can be decoupled in the indirect dimension (F1). This, coupled with high digital resolution in F1, allows for significantly improved resolution of cross-peaks. The band is necessarily nar-

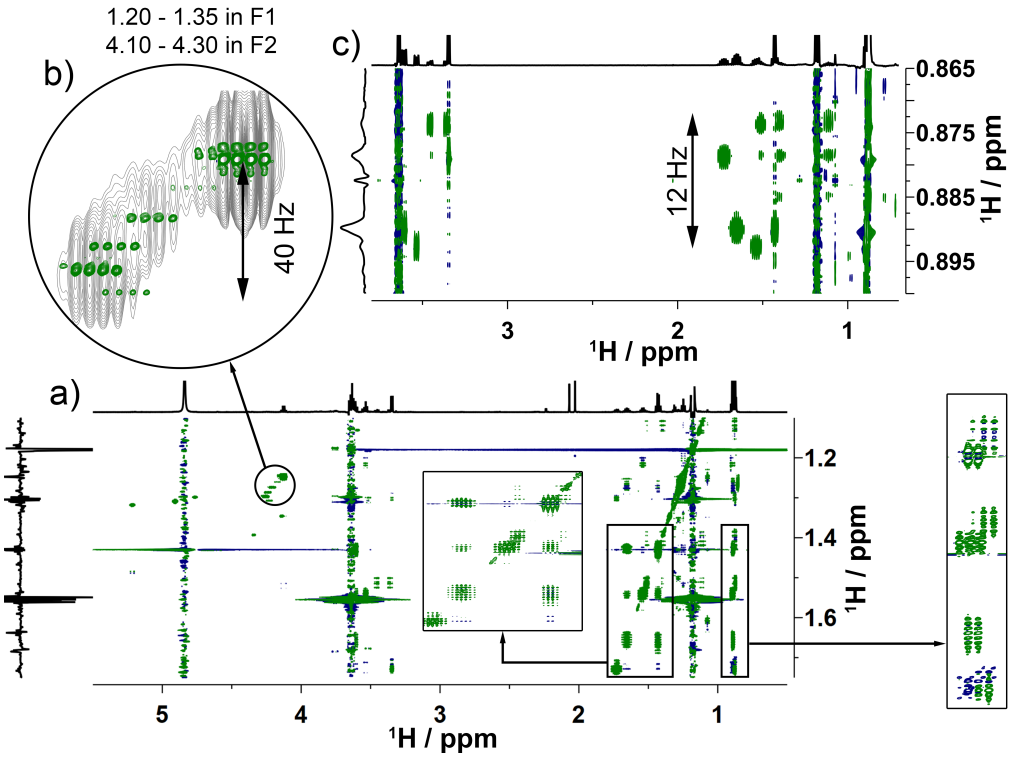


Figure 4.11: Partial BASHD 2D TOCSY spectra of Scotch Whisky (S14-1941) acquired with a 80 ms mixing time. Separately, acquired 1D spectra are shown as external horizontal projections, whereas internal vertical projections are shown. (a) A region of 500 Hz centred on 1.43 ppm was inverted by a 10-ms Iband-selective uniform response pure-phase (BURP)2 pulse. Two expansions are shown as indicated by arrows; (b) expansion of a circled region from (a) in green overlaid with the same region of 2D TOCSY spectrum shown in Figure 4.10b. (c) In a separate experiment, a region of 50 Hz centred on 0.88 ppm was inverted by a 20 ms IBURP2 pulse, again acquired with a 80 ms mixing time.

row to avoid inverting coupled partners. If this happens, the signals are not decoupled. However, this cannot always be avoided. Additionally, the use of a zero quantum (ZQ) suppression scheme here,[100] applied before and after the mixing time, removes unwanted anti-phase signals, further improving the spectral quality.

In Figure 4.11a, a 500 Hz region centred at 1.43 ppm was selectively inverted. This region does contain some mutually coupled protons which prevents the refocusing of some *J* couplings. Still, even this region benefits from excellent resolution afforded by a high digital resolution, resulting in a long t1 acquisition time (576 ms) as shown in the inset of Figure 4.11a. Protons from within this region resonating in a small window of 40 Hz centred on 1.13 ppm do not have coupled partners in the inverted region but are *J* coupled to protons resonating at around

4.05 ppm. As such, they appear as singlets in F1 as shown in Figure 4.11b. At least 10 resonances were identified in this narrow region due to removal of the couplings in F1.

In a separate experiment (Figure 4.11c), the complex overlapping methyl signals at 0.88 ppm - resonating within 15 Hz - could also be resolved in F1 by applying a 20 ms inversion pulse that does not invert any additional resonances. Several cross peaks in two regions around 1.6 and 3.4 ppm were clearly resolved connecting the protons resonating here with the methyl groups. This spectrum was acquired with a t1 acquisition time of 165 ms. The cross-peaks are all singlets in F1, achieving a resolution dictated by the digital resolution and natural linewidth.

The aforementioned homocorrelated experiments provide invaluable information on the structure of the compounds within Scotch Whisky. However, they are not without any shortcomings. Whilst the solvent-suppressed 1D ^{1}H NMR spectrum of Scotch Whisky has an acquisition time of approximately 6.5 min, two dimensional experiments have inherently longer acquisition times to achieve the required digital resolution in the indirect dimension. For example, the 2D TOCSY above had an acquisition time of nearly 8 h, and even the band-selective TOCSYs, covering much narrower spectral widths, required between 1 h and 3 h. The amount of spectrometer time required becomes even more significant when multiple TOCSY experiments are performed with varying mixing times to correlate more signals within a spin-system.[109]

This issue could be addressed by converting these experiments into their 1D analogs. This is possible provided there are resolved signals in the 1D ^{1}H spectrum. These fully or partially resolved signals can be selectively excited and *J* correlation experiments performed starting from these signals. Limited signal overlap in certain regions of ^{1}H NMR spectra of Scotch Whisky allows for such separation of signals in 1D experiments by means of a highly selective 1D CSSF.[110] This filter allows for the frequency, or chemical shift, selective excitation of resonances separated by as little as 1 Hz to 2 Hz.[110]

The original CSSF-TOCSY pulse sequence was modified to include the solvent suppression element (see subsubsection 3.1.2.2).[110] An example of the spectra acquired from this experiment is shown in Figure 4.12. Here, the anomeric proton of β-D-glucopyranose at 4.52 ppm was selected and then subjected to varying TOCSY spin-lock mixing times of 20 ms to 160 ms. For comparison, the normal, solvent-suppressed 1D 1H NMR spectrum is shown at the bottom of Figure 4.12. The power of the CSSF is noted by observing that a partially overlapping signal, just 2.9 Hz to the left of H-1 of β-D-glucopyranose was not excited. From this series of experiments, each taking approximately 3.5 min, the protons of the gluc-

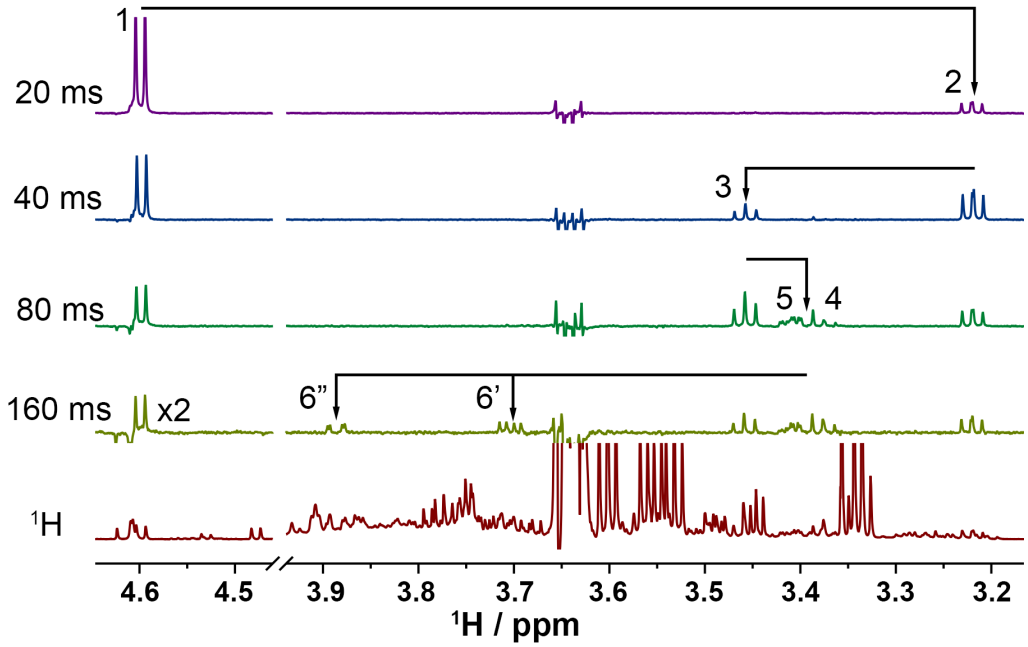


Figure 4.12: 800 MHz 1D ^{1}H CSSF-TOCSY NMR spectra of Scotch Whisky. A selective excitation is applied to the anomeric proton of β-D-glucopyranose at 4.62 ppm with TOCSY mixing times of 20 ms to 160 ms, from top to bottom. 160 ms spectrum y-axis was scaled up by a factor of 2. TOCSY transfers through spin system are annotated. Bottom NMR spectrum shows the normal ^{1}H NMR spectrum for this sample with solvent suppression. Residual ethanol CH_2 signal is visible at 3.64 ppm.

ose ring were assigned in just a few minutes simply by connecting the signals in order of appearance with increased mixing time. Also noticeable is the excellent suppression of the CH_2 signals of ethanol, even better than in the routine solvent-suppressed experiment. The longest mixing time (160 ms) is scaled up by a factor of 2 in the figure - the reduced sensitivity is due to relaxation and magnetisation dilution during the longer mixing time.

Whilst historically the first 2D NMR experiment, *J*-resolved spectra are largely neglected these days in favour of chemical shift correlation experiments. However, *J*-resolved NMR remains a powerful tool in complex mixture analysis, allowing for the separation of overlapping multiplets and the ability to more accurately measure their coupling constants, especially very small ones, in a relatively fast and sensitive experiment.[213,214] Thus, the last homocorrelated experiment presented here is a 2D *J*-resolved experiment. An example spectrum is shown in Figure 4.13.

Due to its simplicity, a magnitude mode 2D *J*-resolved experiment[215] was selected as the basic pulse sequence to be modified with the developed signal suppression technique. More advanced *J*-resolved sequences were investigated, in-

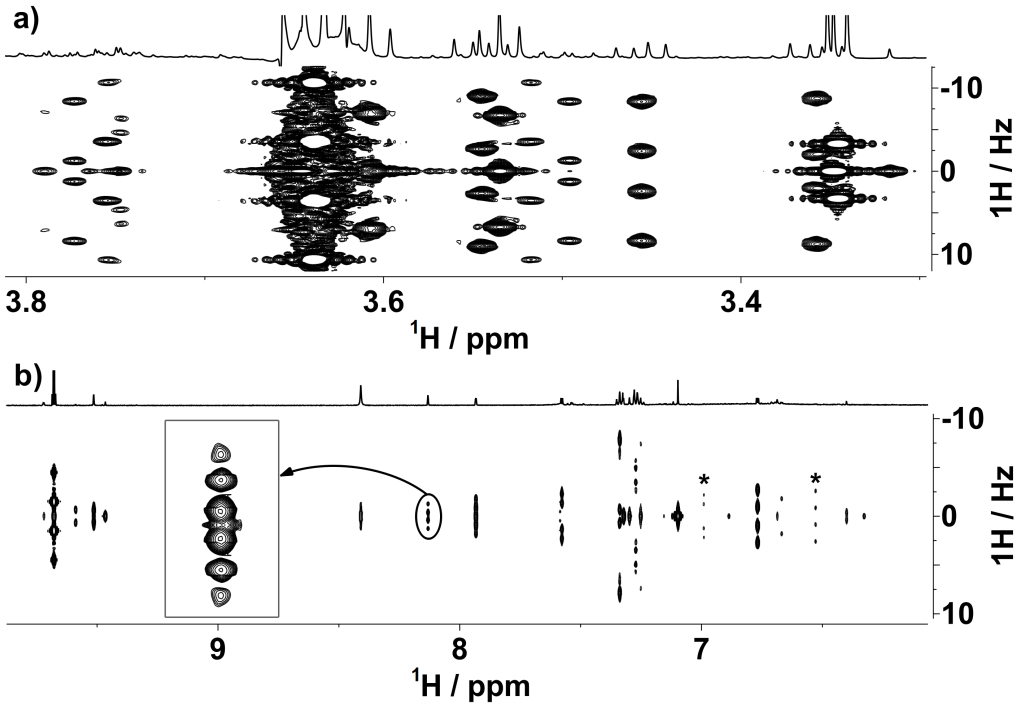


Figure 4.13: Magnitude mode 2D *J*-resolved spectrum of Scotch Whisky (S14-1941) (a) region around the CH_2 signal of ethanol (b) aromatic region. The spectrum was tilted and symmetrised in F1. The multiplets labelled with an asterisk in (b) belong to signals that are below the noise level in the 1D spectrum. The inset in (b) shows the multiplet of a signal at 8.13 ppm that is unresolved in the 1D spectrum.

cluding pure-shift methods, however these proved to be more difficult to achieve acceptable results with the solvent suppression technique. The resulting spectra, (*e.g.* see Figure 4.13), are acquired in less than 90 min, and yet provide excellent results.

High-quality spectra were obtained resolving multiplets of many compounds at a variety of concentrations. The solvent suppression was, again, very efficient, as illustrated in Figure 4.13a, where a region around the CH_2 resonances of ethanol is shown. Here, many minor signals are clearly resolved with only a 40 Hz wide strip lost to the t1 noise of ethanol signal. As an example, note the partially overlapping triplet to the right of the CH_2 signal of ethanol which is well resolved in the *J*-resolved spectrum. Additionally, note the well resolved multiplicities within a 0.1 ppm region at 3.5 ppm, which would be difficult or impossible to accurately deconvolute in the 1D spectrum without *a priori* knowledge.

In Figure 4.13b, the aromatic region of the spectrum reveals valuable details of signals, especially of those which were at the noise level in the 1D NMR spectrum. For example, two multiplets are resolved in the 2D spectrum which are not even observed in the 1D spectrum (annotated with asterisks in Figure 4.13b).

This increase in sensitivity is due, largely, to the increased total number of scans through the indirect dimension. Furthermore, *J*-resolved spectra have capacity to compensate for magnetic field inhomogeneity and reveal splittings of signals that appear as singlets in a 1D spectrum; this is very valuable in establishing the identity of these signals (see the inset in Figure 4.13b).

4.5.2 *Heterocorrelated 2D NMR*

Whilst the above sequences allow for the observation and interrogation of a number of proton-proton correlations, full structural characterisation requires information on other nuclei within each molecule. Most commonly, this corresponds to ^{13}C. In this subsection, three of the most common and useful heterocorrelated NMR experiments were adapted to incorporate solvent suppression for Scotch Whisky. These are 2D ^{1}H, ^{13}C HSQC, HSQC-TOCSY, and HMBC experiments.

Compared to ^{1}H, ^{13}C NMR will be substantially less sensitive due to both the lower gyromagnetic ratio and, more significantly, the much lower natural abundance of ^{13}C - only 1.1 % compared to ^{1}H at 99.9 %. As discussed in the introduction, there are ways to alleviate some of the sensitivity problems with carbon NMR, most importantly by taking advantage of the increased sensitivity of proton for both detection and establishing the initial polarisation transfer. To that end, the three experiments detailed here rely on both of these strategies. Direct detection of ^{13}C is not performed on Scotch Whisky beyond that of the third precursor experiment as previously discussed in subsection 4.4.3.

The reduced natural abundance, however, affords some advantages. Most significantly, the ^{13}C heterocorrelated spectra of Scotch Whisky will only show its most abundant compounds. This simplification, along with access to the chemical shifts and heteronuclear connectivity of ^{13}C, makes these experiments very powerful, producing spectra which assist significantly with unambiguous compound identification.

The major challenge associated with these experiments, along with sensitivity, is the solvent suppression. This time, however, both ^{12}C and ^{13}C isotopomers of ethanol must be suppressed. Insufficient suppression of the former can lead to excessive t1 noise, whereas failure to suppress the latter would produce very strong ethanol cross peaks obscuring the signals of interest.

Solvent suppression during the relaxation delay was added to standard Bruker pulse sequences (*hsqcedetgpsisp2.3*, *hsqcdietgpsisp.2*, and *hmbcctetgpl2nd* - see subsubsection 3.1.2.2), providing excellent results. This was sufficient for the HSQC and HSQC TOCSY sequences, while additional measures had to be taken for HMBC. A partial HSQC spectrum is shown in Figure 4.14.

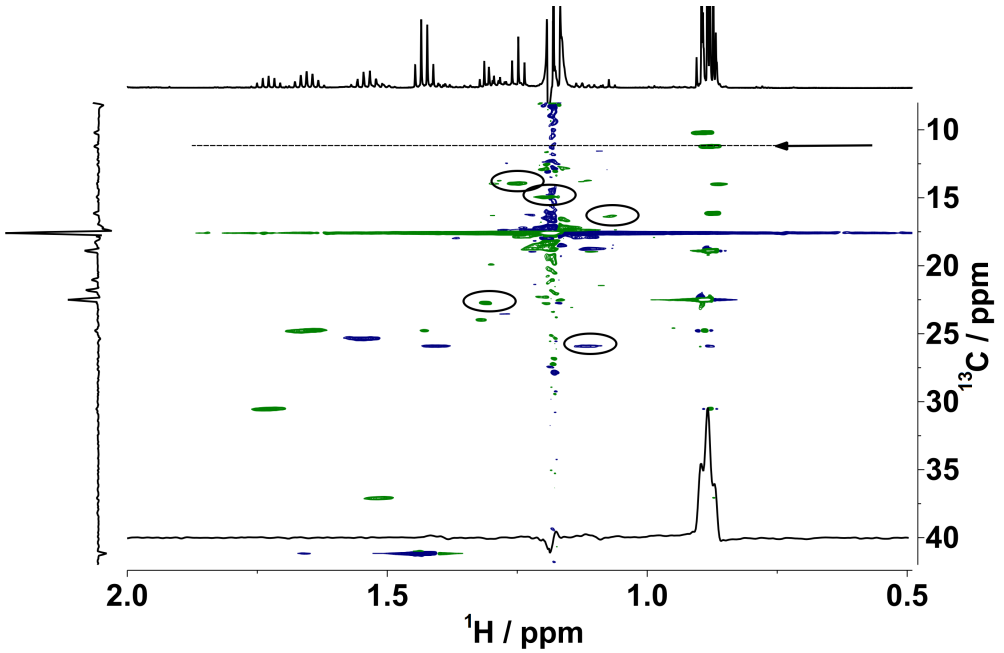


Figure 4.14: Partial 2D ^{1}H, ^{13}C HSQC spectrum of Scotch Whisky (S14-1941) showing the low ppm aliphatic region. The 1D spectrum is shown at the top, whereas an internal maximum projection is presented on the left-hand side of the spectrum. Signals close to CH_3 protons of ethanol in both the ^{12}C and ^{13}C isotopomers are circled. A row taken at $\delta(^{13}C) = 11.22$ ppm highlighted by an arrow is presented at the bottom of the spectrum.

The HSQC experiment is phase sensitive, allowing for multiplicity editing to discriminate the CH_3/CH and CH_2 cross-peaks. Decoupling during acquisition further increases the spectral quality. This bilevel decoupling element had to be explicitly coded into the pulse sequence to resolve technical difficulties with the spectrometer hardware. Specifically, due to the complexity of the pulse sequence and the numerous pulses and shapes - required for all elements including the solvent suppression - the console returned a FIFO error if the decoupling was not explicit. As discussed previously, cryogenically cooled probes do not allow broadband carbon decoupling for extended periods of time. In the 1D ^{1}H NMR experiment, and in the presaturation element, we avoid this by selectively decoupling ethanols carbon signals. This is sufficient as in the ^{1}H spectrum, only ethanol has carbon satellites of any appreciable intensity. Likewise, it is sufficient for presaturation as only the ethanol signal needs to be suppressed. The purpose of the decoupling during acquisition, however, is to provide a much simplified spectrum free of proton-carbon coupling, and must be broadband to encompass all cross-peaks. This ^{13}C decoupling during acquisition limits the maximum direct dimension acquisition time t2 to approximately 180 ms. Across a 15 ppm spec-

tral width in ^{1}H, this corresponds to 3,072 time domain points and a FID resolution of 5.8 Hz at 600 MHz.

The limited F2 resolution is not a significant problem as proton-proton couplings make cross-peaks of multiplets relatively wide anyway. However, if the spectrum was acquired pure-shift in F2, *i.e.* proton-proton couplings were removed, higher digital resolution would be required to take advantage of this. Two main strategies exist to achieve higher resolutions in F2. The first is the extended acquisition time (EXACT) method - 'chunking' the acquisition of the FID so that decoupling can be paused when the FID is not being detected.[216] During the off-chunks, when the FID is not being detected, simple refocusing pulses effectively decouple carbon, using much less power and generating much less heat. The gaps result in recorded zeroes in the FID. These missing sections of the FID are then filled in post-acquisition via NUS methods. This modern method would require modification to the pulse sequence but is theoretically possible. However, neither the pure-shift or EXACT acquisitions have been implemented for Scotch Whisky yet. The second strategy to improve the digital resolution is during processing, by linear prediction or application of NUS algorithms to computationally increase the length of the FID.

As shown in Figure 4.14, the acquired HSQC spectrum is of a high quality, with numerous cross-peaks visible. The spectrum was acquired with 1,024 increments in F1, (t1 = 20 ms), providing a high resolution in both dimensions. The residual ethanol CH_3 signal is the dominant feature of this region of the spectrum, showing the t1 noise. However, many genuine cross-peaks are visible even very close to this signal and the t1 ridges. A horizontal trace at $\delta(^{13}\mathrm{C}) = 11.22$ ppm shows the quality, resolution, and sensitivity of the spectrum at a given congener methyl cross-peak. The overlap of several methyl signals at $\delta(^{1}\mathrm{H}) = 0.88$ ppm is resolved in the carbon dimension showing signals for at least nine different methyls from different molecules. Acquisition of this spectrum took approximately 5 h, substantially longer than the 1D ^{1}H NMR spectrum, however providing invaluable information necessary for structural assignment.

Complementing the HSQC data, the HSQC-TOCSY experiment allows for the correlation of protons and carbons more than one bond away by means of a TOCSY transfer. A partial example spectrum is shown in Figure 4.15.

Just as with the homocorrelated TOCSY, a DIPSI-2 spin-lock sequence was used for the mixing period, allowing for straightforward comparison with the ^{1}H,^{1}H TOCSY data. Likewise, the HSQC sequence elements are very similar - using the same sensitivity improvements, phase sensitive detection and explicit decoupling during acquisition. Note the spectrum is not multiplicity edited here. These similarities allow the HSQC and HSQC-TOCSY spectra to be superim-

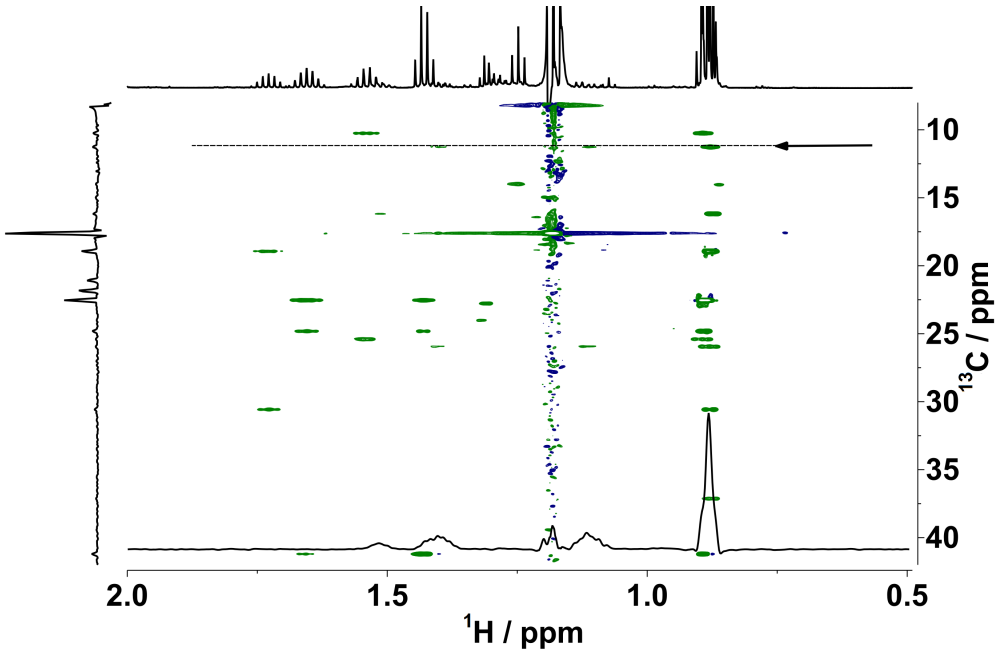


Figure 4.15: Partial 2D ^{1}H, ^{13}C HSQC-TOCSY spectrum of Scotch Whisky (S14-1941) showing the low ppm aliphatic region. The 1D spectrum is shown at the top, whereas an internal maximum projection is presented on the left-hand side of the spectrum. A row taken at $\delta(^{13}C) = 11.22$ ppm highlighted by an arrow is presented at the bottom of the spectrum.

posed and directly compared for aiding interpretation. Furthermore, they were acquired with identical spectral widths and number of time domain points, and processed in the same way. The TOCSY mixing period was 40 ms, corresponding to transfers across 2 to 3 (occasionally 4) bonds, depending on the coupling constants. These spectra were acquired with twice the number of scans per increment compared to the HSQC spectrum, and thus took approximately 10 h to acquire.

The acquired spectrum (Figure 4.15) shows many more cross peaks in the aliphatic region than seen in the HSQC spectrum, corresponding to the TOCSY-cross peaks. The ethanol CH_3 signal is still the most abundant and induces substantial t1 ridges, approximately 15 Hz wide, however many cross-peaks of minor compounds in their vicinity are still resolved and aid assignment of compounds in Scotch Whisky. Again, a horizontal trace of $\delta(^{13}C) = 11.22$ ppm shows the quality of the data in terms of resolution and sensitivity, and the t1 ridges are of a similar intensity to genuine cross-peaks, but are easily distinguished by their noisy, mixed phase character.

As mentioned above, the HSQC and HSQC-TOCSY sequences required no additional solvent suppression after the presaturation block. However, for the HMBC sequence, additional solvent suppression was implemented during the

long-range proton-carbon coupling ($^{n}J_{CH}$) evolution period. This was possible as this delay, set to $\frac{1}{2}J$ for a 6 Hz coupling constant, is relatively long 83 ms. During this period, a continuous wave irradiation was applied on the third channel for water suppression and on the first channel for CH_2 suppression. Theoretically, this block could use a shaped pulse for doubly-selective saturation as in the presaturation block; however, this was not implemented and the 50 ms pulse may not be compatible with this 83 ms delay. Regardless, the CH_3 protons are suppressed sufficiently. An example of the acquired HMBC spectrum is shown in Figure 4.16 for the same region as the HSQC and HSQC-TOCSY spectra.

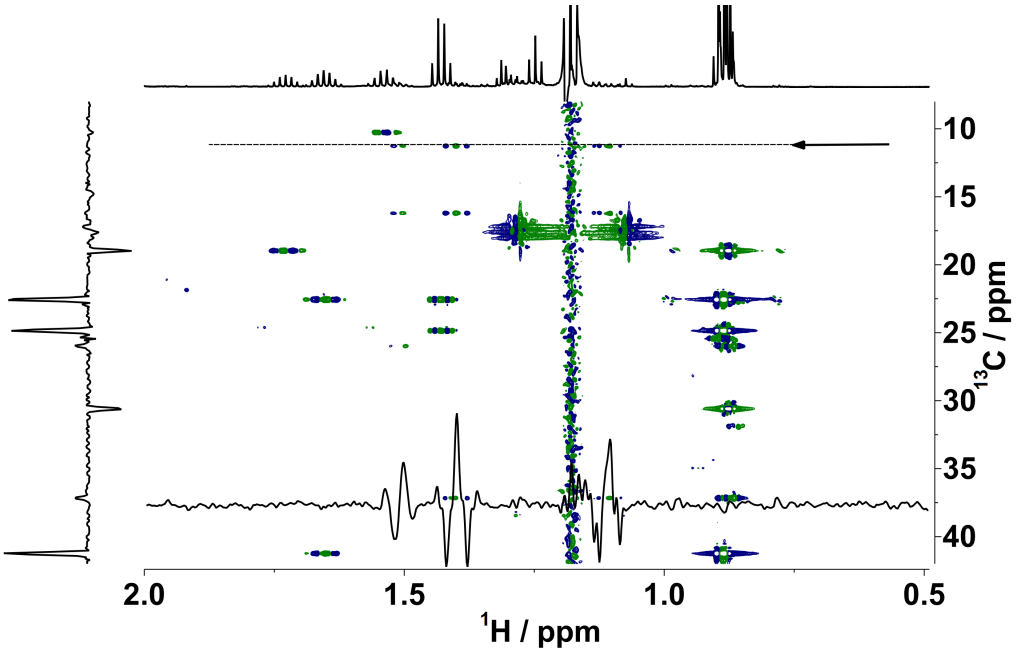


Figure 4.16: Partial 2D ^{1}H, ^{13}C HMBC spectrum of Scotch Whisky (S14-1941) showing the low ppm aliphatic region. The 1D spectrum is shown at the top, whereas an internal maximum projection is presented on the left-hand side of the spectrum. A row taken at $\delta(^{13}C)$ = 11.22 ppm highlighted by an arrow is presented at the bottom of the spectrum.

The HMBC sequence differs from the HSQC sequence in several ways. HMBC allows observation of multiple bond correlations (hence the name). However, unlike in the HSQC-TOCSY sequence, these multiple bond correlations can go through heteroatoms or involve quarternary carbons - ^{1}H TOCSY transfers correlate only protonated carbons within the same spin system. Furthermore, although the residual CH_2 cross peak of ethanol is a singlet in the HSQC and HSQC-TOCSY spectra (see Figure 4.14 and Figure 4.15, respectively), it appears as a $^{1}J_{CH}$ doublet in Figure 4.16 as ^{13}C decoupling is not applied during acquisition in this HMBC sequence. This is despite a two-fold low-pass *J*-filter applied in

this experiment and set to 125 Hz and 142 Hz, corresponding to $^{1}J_{CH}$ of ethanol. Correspondingly, there is weaker t1 noise through the mid-point of this doublet.

The cross-peaks observed in the HMBC spectra are well resolved, and the experiment is phase sensitive. However, the spectrum cannot be phased as the ^{1}H signals are of inherently mixed phase. From this, locating the exact midpoint of a cross-peak can be more challenging, and so HSQC cross-peaks are used for determination of ^{13}C chemical shifts where possible. The spectral quality is improved by using a constant-time sequence. Here, the total length of the sequence remains constant which eliminates the *J*-modulation in F1, improving the resolution in the indirect dimension.

The power of the HMBC spectra is in its ability to determine the chemical shifts of quarternary carbons whilst gaining the sensitivity boost of a polarisation transfer and proton-detected experiment. Additionally, the ability to connect spin systems which are not correlated in TOCSY spectra is another use of the HMBC data. This spectrum was again high resolution. Without the acquisition length limits in t2, a larger time domain and narrow spectral width could be used resulting in a t2 of 243 ms corresponding to an FID resolution of 4.1 Hz. The indirect dimension resolution is comparable to the HSQC sequences, although slightly poorer due to fewer increments (910 instead of 1024) and a larger spectral width in F1. Again, a horizontal trace at $\delta(^{13}\mathrm{C}) = 11.22$ ppm shows the quality of signal obtained, with the residual t1 noise not hindering the interpretation at all. The HMBC spectrum took about 9 h to acquire.

The HSQC based spectra were acquired with carbon spectral widths of 165 ppm centred at 80 ppm, whilst the HMBC used 200 ppm centred at 100 ppm. This modification for HMBC allows the observation of most carbon signals, including quarternary carbons, without folding or aliasing of the spectrum. The HSQC and HSQC-TOCSY - along with some of the homocorrelated 2D spectra - exhibited folding of cross-peaks with very high chemical shifts, however their corresponding chemical shift in the directly detected dimension allows for their deconvolution.

Overall, the presented 1D and 2D NMR experiments represent a powerful toolkit for the structural analysis of compounds in Scotch Whisky. High resolution and sensitivity enhanced experiments, performed at high magnetic field with cryogenically cooled probes, represent the limits of detection with conventional NMR for the analysis of Scotch Whisky without a separation or prior concentration step. The developed NMR solvent suppression technique is readily combined with other pulse sequences to expand this toolkit, and future work could include the incorporation of this suppression into other experiments, including DOSY,[117] REST,[84] and broadband pure-shift experiments.[217]

4.6 AUTOMATING ACQUISITION

Automating the acquisition of 1D ^{1}H NMR spectra was required because of the large numbers of samples and the sophisticated solvent suppression routine implemented. For full details of the theory and implementation, see section 4.4 and subsubsection 3.1.2.1. Briefly, for each sample, the suppression parameters are optimised. This includes shaped pulse lengths, power levels, frequencies, etc. Whilst much of the effort in automating some of these routines was already implemented by Boris Mitrovich, a former summer student, it was necessary to refine and optimise his work. Note that the more advanced NMR experiments were not required to be implemented in automation, though such effort would be relatively trivial.

As discussed earlier, for the automated suppression of the ^{1}H signals of ethanol, a 1D reverse INEPT NMR spectrum is acquired to determine the exact frequencies of ethanols CH_2 and CH_3 signals. For two signals, such as in this case, it is possible to be, in effect, 'on resonance' by application of a pulse which performs n 2πrotations. At 600 MHz, this corresponds to rotating 36 times 2π, and results in a pulse of approximately 50 ms.

Inherently, this pulse is generated digitally, and must comprise a number of discrete points. For simple shapes, as few as 100 points are sufficient. However, to more accurately describe more complex functions, 1000 points (or more) are used. These points have a uniform length, dictated by the overall length of the shaped pulse and the number of points used. The final caveat in this is that the spectrometer hardware in the console - specifically the frequency controller and signal generation unit - have technical limitations. For the Avance III HD consoles coupled to our 600 MHz and 800 MHz spectrometers, these limitations are that the minimum 'event time', or length of a pulse element, must be 25 ns. This event time is a limit for any digital setting of frequency, phase or amplitude.[218]

For a pulse of 50 000 µs (50 ms) and 1000 points, the corresponding length of each point is simply 50 µs (50000/1000). The problem is in the calculated versus required precision. For two signals separated by exactly 1475.6087 Hz, the pulse length would be calculated to be 50 148.7953 µs. When this is split into 1000 points, each point now has a length of 50.1488 µs. Obviously, this length does not fulfil the criteria of minimum unit length time, and so a programmatic means to 'round' this time was required. *N.B.* For simplicity, numbers have been truncated to four decimal places, however floating point numbers in TopSpin are calculated with up to seven points of precision. Note that TopSpin only uses two decimal places of precision when executing pulses, however.

Algorithm 4.1 Rounding Pulse Length Function

```
def pulselopt(time):
    npoints = 1000
    timeA = float(time)/float(npoints)
    timeB = timeA/0.025
    timeA = int(timeB)*0.025*1000
return timeA
```

Algorithm 4.1 demonstrates a simple, but effective means to 'correct' a pulse length so it fulfils this criteria. Simply, the pulse length is divided by 1000, and this length then divided by the minimum unit of time. This value is then rounded to an integer, and multiplied back up by the minimum unit length and number of points. The result, in this example, is 50 150 μs. The overall pulse length has been increased by 1.2047 μs, equivalent to an increase of 0.0024 % or an offset from n 2πrotations of 0.63°. This increase is relatively insignificant and has no adverse effects on the signal suppression.

A similar issue was observed for the second component of ethanol suppression, the decoupling of the ^{13}C satellites in the ^{1}H NMR spectrum. As explained in subsection 4.4.3, a shaped pulse is calculated to last approximately 2.5 ms for n 2πrotations to selectively decouple the ^{13}C satellites of ethanol.

There is an additional layer of difficulty here. Whereas suppressing ^{1}H-^{12}C protons of ethanol uses a long 50 ms pulse, this decoupling pulse is 20-times shorter (approximately 2.5 ms). Therefore there is greater likelihood of creating a pulse with an incompatible length. A similar solution was implemented as shown in Algorithm 4.1. To calculate an optimal, but also technically feasible, pulse of such a short length with so many points becomes more challenging. Instead, fewer points (n=500), and a variable number of them, can be used as the decoupling pulse need not be as exact as the suppression pulse. Simultaneous optimisation of two parameters - length and number of points - required further modifications as shown in Algorithm 4.2.

Algorithm 4.2 Optimising pulse length and number of points

```
minlength = 0.025
time=float(int(time))
low, high = 450, 550
def pointcalc(time,low,high):
    for i in range(low,high):
        z = (time / i) / minlength
        if z % 1 == 0:
            points = i
            return points
        elif i == (high−1):
            return False
        else:
            i + 1
    return
points = pointcalc(time,low,high)
if points != False:
    time = int(time)
elif points == False:
    time = time + 1.0
    points = pointcalc(time,low,high)
    if points != False:
        time = int(time)
    else:
        time = time − 2.0
        points = pointcalc(time,low,high)
        if points != False:
            time = int(time)
        else:
            points = 500
            time = int(time)
```

Algorithm 4.2 details the more complex routine required to optimise both a shaped pulse length and number of points. The algorithm determines an optimal pulse length $\pm 1\,\mu s$ the rounded (integer) value of this original length. It simultaneously determines the appropriate number of points between 450 and 550 points to describe this shape whilst operating within the minimum unit length of time. If such optimisation were to be impossible, the final part of the algorithm sets the number of points to 500 and the length to the integer length. However, this fail-safe has not been used.

Here, a measured distance of 6056.0045 Hz at 150 MHz (^{13}C) corresponds to an original pulse length of 2476.8806 μs. This algorithm will calculate an optimal

pulse length of 2478 μs in 472 points. This difference from the optimally calculated value is a ≈1.1 μs, ≈0.05 %, again an insignificant change.

It may be considered that this approach is excessive when most Scotch Whisky products are sold at 40 % ABV. However, samples of, and relating to Scotch Whisky, may be much stronger - up to 94.8 % ABV. For a diverse range of samples, and to be able to automate acquisition of their spectra, encoding flexibility and redundancy is more than programmer's indulgence; it is necessary.

The bulk of the remaining acquisition scripting was written by Boris Mitrovich with only minor modifications by myself, and are not worth further discussion. My other modifications include adding debugging output, updating elements for newer versions of TopSpin, and adding in compensations for variable temperature experiments.

4.7 CONCLUSIONS

This chapter detailed the development of an NMR methodology for the acquisition of high quality spectra of Scotch Whisky. Specifically, the ability to acquire spectra of Scotch Whisky with minimal sample preparation, in its raw state, by means of an advanced solvent suppression method. The solvent suppression technique was built upon the pre-existing 1D NOESY-presat type sequences, and this block was implemented into a number of widely used and important NMR pulse sequences to create a tool-kit of sequences which could be used to characterise the minor compounds in Scotch Whisky.

Automating the acquisition of the 1D ^{1}H NMR spectra of Scotch Whisky was a necessary step towards implementing this as a routine method on the spectrometer. This involved relatively straightforward scripting and Bruker's IconNMR interface, however there was not-insignificant troubleshooting required to mitigate hardware limitations. Hardware limitations were also a concern in the pulse sequence development itself - in terms of carbon decoupling and power levels, as well as in terms of pulse timings and explicitly encoding decoupling into the HSQC sequences.

The characterisation of the NMR detected compounds in Scotch Whisky is presented in section 5.4. The resulting spectra are of high quality, high resolution, represent the limits of NMR sensitivity today, and will be utilised more fully in the following chapters. The quantitative nature of the 1D ^{1}H NMR experiment, along with the precursor experiments, will allow for the quantification of many compounds in Scotch Whisky by NMR. Simultaneously, the fast and automated means of acquiring the 1D spectra will allow for high-throughput, metabolomics-type studies of Scotch Whisky, detailed in later chapters.

5

NMR ANALYSIS OF SCOTCH WHISKY

Science and everyday life cannot and should not be separated.

— Rosalind Franklin (1940)

5.1 DECLARATION

The reference ethanol-water solutions were prepared and NMR spectra acquired by two undergraduate RME project students - Ryan Glenwright and Klaudia Murinova. All other work is my own.

5.2 INTRODUCTION

A significant advantage of NMR, in comparison to other techniques such as MS, is that it allows for the simultaneous *de novo* identification and quantification of a range of compounds with diverse chemistries and quantities. Whilst this is, and will prove to be, a necessary quality of NMR, Scotch Whisky has been - at least partially - previously characterised by other techniques such as GC or LC-MS. Using these techniques, the identities and typical concentrations of many compounds have already been identified in Scotch Whisky.

As discussed in the introduction, many of the more abundant, flavour and aroma active compounds in Scotch Whisky are known in the literature. Broadly speaking, the classes of compounds in Scotch Whisky can be described as follows:[16,25]

- Higher Alcohols
- Sugars
- Cask Extractives
- Fatty Acids/Ethyl Esters
- Sulfides
- Phenols

Beyond these classes of compounds, two other key compounds are known to exist in Scotch Whisky - EC and whisky lactone. Combined, all these compounds

cover a range of concentrations from nM to mM. NMR has, of course, a higher limit of detection than an optimised GC-MS method for food analysis,[219] so the lowest abundant species - such as EC and the sulfur species - are not expected to be observed in the NMR spectra of native Scotch Whisky. EC has a median concentration of 40 $\mu g\,kg^{-1}$, in European whisky. However, in other spirits, especially those derived from stone fruits, the levels are much higher and are even quantifiable by NMR techniques.[9,10]

The major known sulfur species - including DMS(**14**), DMTS(**15**), and MMFDS(**16**) - are all highly odour active at low concentrations, and only the lowest concentrations provide desirable characters to the spirit.[27] Above these levels, they introduce unpleasant sulfur aromas - such as rotten egg or meat. A study from the SWRI reported typical concentrations of these sulfur species at approximately 20 $\mu g\,kg^{-1}$ to 40 $\mu g\,kg^{-1}$ - equivalent to approximately 80 nM to 400 nM.[27] The low expected concentrations of EC and the sulfur species therefore precludes them from being observed in the NMR spectra of untreated whisky samples.

This chapter details analysis of the NMR spectra of Scotch Whisky, including the identification of whisky congeners and quantification of ethanol and other compounds. NMR is also used to study classification and product consistency in a large sample set of Scotch Whiskies using statistical analysis and chemometric techniques.

5.3 SAMPLE TYPES & CONSIDERATIONS

Scotch Whisky samples were provided by the SWRI. The primary sample set for this study consisted of its 'standard sample set' of Scotch Whisky. This set is curated biennially from final products produced by SWRI member companies. The SWRI utilise market data to determine which samples to curate, aiming to design a representative set of Scotch Whisky within a limited number of samples - typically around 50 samples per set. The samples provided here were curated in 2010, 2012, and 2014, and represent the majority of Scotch Whisky sold as total UK and export case sales; 2014 (73 %), 2012 (71 %), and 2010 (67 %).[220–222] In total, these three years of samples comprised of 148 samples - 2010 (n=47), 2012 (n=49), and 2014 (n=52). Sample IDs are based on their year of addition to the SWRI library (*e.g.* S10 for 2010, S12 for 2012) followed by a four-digit unique ID. *N.B.* Some samples from the 2012 set were provided in 2013 and have S13 codes. As with all samples in this thesis, they have been anonymised. See Appendix B for a list of samples used in this chapter.

Whilst the provenance of these samples is excellent - they came direct from the producer to the SWRI - not every production parameter is known about every

sample. That said, at least some detail is known as this has to be provided on the label by law. For example, if it is a malt, grain, or blended whisky, and if it is the product of a single distillery or a blend of distilleries. The sample set contains no single or blended grain whisky, and thus comprises Single Malts, Blended Malts, and Blended Scotch Whisky. For simplicity, this is reduced to malts (both single and blended, n=71) and blends (blended whisky, n=77).

Geographical designation is another legally defined parameter, however it is optional, and it can only be designated if the product is exclusively from that region. The sample set consists of Highland (n=29), Islay (n=17), Lowland (n=6), and Speyside whisky (n=17). A further 79 have no origin claim - the blends and blended malts. There were no samples from the other SWA legally defined geographical region, Campbeltown.

Age statements, like geographical origin, are optional, but if defined they represent the youngest whisky in the bottle (as all whisky - bar single cask bottlings - will comprise a blend of casks of varying ages). The sample ages were 8 (n=1), 10 (n=21), 12 (n=41), 16 (n=4), to 21 (n=2) years old. Again, a further 79 have no age claim. However, these are a mix of no age statement (NAS) blends and single malts.

Maturation wood types are most commonly ex-Bourbon or ex-Sherry, and whilst there is no requirement to detail these on the final product, the maturation woods are known for some of the sample set. These woods are either the main maturation or finishing wood. For simplicity, this detail is streamlined - if a whisky was matured or finished in a type of wood, that wood is a 'maturation wood type'. These include ale (shorthand 'A', n=3), bourbon (shorthand 'B', n=26), bourbon and port (shorthand 'BP', n=1), bourbon and sherry (shorthand 'BS', n=21), and sherry (shorthand 'S', n=10). A total of 87 samples have an unknown maturation wood type.

Finally, some 18 whiskies were known to contain peated malt, with 130 either unpeated or not known to contain peated malt.

These characteristics of the sample set allow for some supervised chemometric investigation - see section 5.6. However, there are many confounding variables and unknowns which compromise the ability to draw too many conclusions about specific whisky production characteristics with this sample set and how they are reflected in NMR spectra.

5.4 STRUCTURE ELUCIDATION

As with other analytical techniques, the use of reference standards accelerates compound identification and method development - especially for complex matrices.

Such approaches have been therefore applied here where possible. Thus, the first stage in the identification of compounds present in Scotch Whisky involved acquiring reference spectra for compounds expected or known to be present, as reported by the literature or in consultation with the SWRI. The reference spectra must be acquired under similar conditions, *i.e.* in a 40 % ABV solution and of a typical concentration to how they may be present in whisky; this minimises the chemical shift differences that can otherwise complicate compound identification.

5.4.1 *Higher Alcohols*

Many of the higher alcohols present in whisky have distinctive ^{1}H NMR spectra and a simple visual comparison of the spectra allows for a degree of confidence in asserting whether these compounds are present in whisky or not. An example of this approach is shown in Figure 5.1.

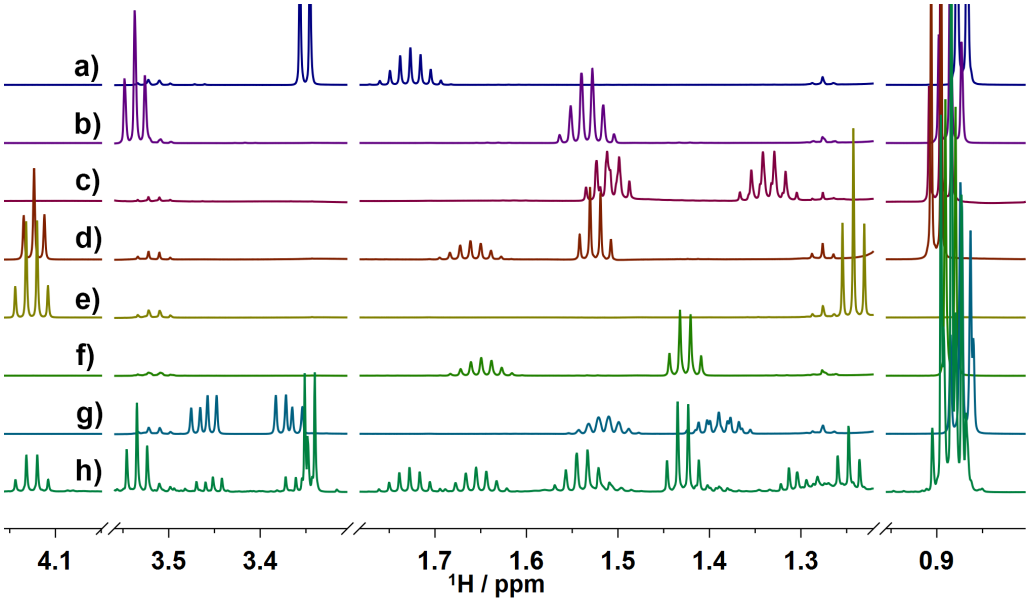


Figure 5.1: Stacked partial 1D ^{1}H NMR spectra of h) Scotch Whisky (S14-2858) and a-g) reference compounds in 40 % ethanol solution. Reference compounds are a) isobutanol (**25**), b) n-propanol (**26**), c) n-butanol (**27**), d) iso-amyl acetate, e) ethyl acetate (**28**), f) 3-methylbutanol (**29**), g) 2-methylbutanol (**30**). Numbered compounds have structures shown in Appendix A. Spectra have been scaled independently.

Here, seven major volatile compounds known to exist in whisky are compared to a typical whisky spectrum. Compounds such as ethyl acetate (**28**) (Figure 5.1e) and 3-methylbutanol (**29**) (Figure 5.1f) have distinctive and well resolved multiplicities which allow for their confident identification. Of the compounds in

Figure 5.1, two compounds cannot be confirmed simply by comparison of the spectra. These compounds, n-butanol (**27**) and iso-amyl acetate (Figure 5.1c,d), may be below the limits of detection and/or obscured by overlapping signals from more abundant species. Potential means to resolve this may involve the acquisition of pure-shift NMR spectra.[217,223]

A similar strategy was employed for other expected compounds, however those with few ^{1}H NMR signals, especially those which mostly present singlets, are difficult to confirm by this strategy. Furthermore, chemical shifts will vary between the reference solution and each whisky sample due to subtle changes in concentration, pH, and alcohol strength. Comparison of reference spectra to whisky relies on a degree of subjective interpretation for shifted peaks, which is why signal multiplicity analysis is vital. On the other hand, resolving singlets in an overlapped region and compensating for an unknown chemical shift difference is inherently difficult.

5.4.2 *Carbohydrates*

Another major class of compounds expected in Scotch Whisky are sugars or carbohydrates. These give complex NMR spectra, and structurally similar carbohydrates may yield similar spectra.[224] Therefore, the next stage in identifying compounds in whisky was to perform 'spike-in' experiments. This involves spiking in a small amount of a pure compound into a sample of whisky which may contain this compound. By comparing spectra at two levels of addition and a control, it is not only possible to determine if a compound is present in the mixture, but what its chemical shifts are in a specific matrix. Provided the level of addition is low enough, the chemical shifts should not change between the three spectra. An example of this approach on the addition of glucose is shown in Figure 5.2.

Here, it is abundantly clear that glucose is present in solution as all of its peaks overlap perfectly with the peaks resulting from the two levels of spiked-in addition. The typical signals of a sugar in NMR are the anomeric protons, which are well resolved at 5.2 and 4.6 ppm, corresponding to the α and β-pyranose configurations, respectively.

When successful, spike-in experiments do provide certainty that a given compound is present in the mixture, and make identification of compounds far more straightforward. Another advantage of the spike-in method over comparison with reference spectra is that compounds with fewer distinctive signals can be confirmed, too. As chemical shifts do not need to be adjusted, even singlets can be identified and associated with a compound. It is also important to select an appropriate whisky sample for the compound of interest – not all whiskies will

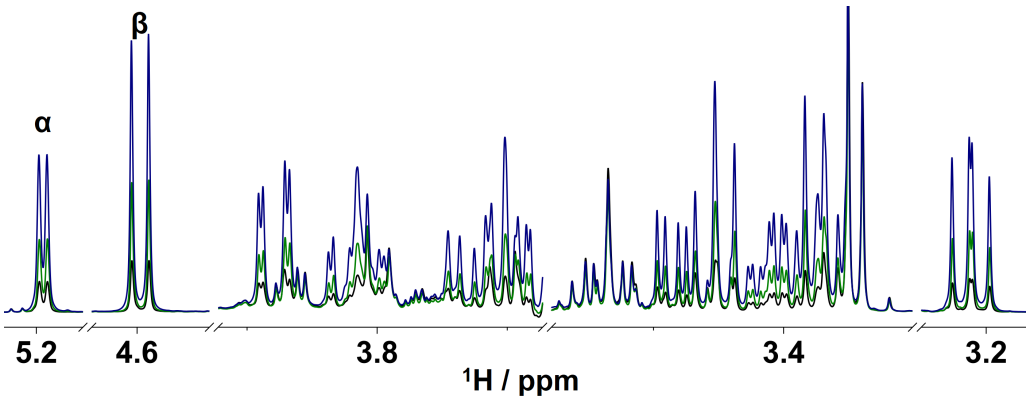


Figure 5.2: Superimposed partial 1D ^{1}H NMR spectra of Scotch Whisky (S10-1313) with glucose spiked-in at two levels of addition. Black shows the normal, unadulterated whisky spectrum, whilst blue and green represent the addition. The α and β anomeric protons have been annotated. Signals increase proportionally where glucose signals are, whilst signals unrelated to glucose are unaffected - *e.g.* 3.53 ppm shows no change, whilst the anomeric protons (5.2 and 4.6 ppm) or 3.2 ppm show substantial changes.

contain the same compounds at the same concentrations. In this case, a whisky suspected to contain a reasonable, NMR related, concentration of glucose was chosen. Unfortunately, it is not possible to assign all peaks – or even resolve the more complex overlapping signals – by spike-in experiments alone.

Another convenient 1D ^{1}H NMR experiment to assign compounds in whisky is a selective 1D TOCSY, in this case using a CSSF allowing for highly selective excitation.[110] An example of this technique was shown in the previous chapter for the assignment of β-D-glucopyranose (**12a**) (see Figure 4.12). The acquisition of several 1D TOCSY spectra with increasing mixing times allow for the assignment of all of the protons within a spin-system, and thus positive identification. This approach works very well for compounds with extended spin-systems, *i.e.* carbohydrates, but is not applicable to compounds where TOCSY transfer is not possible, such as at quaternary carbons or heteroatoms.

Another approach relies on the use of homo- and heterocorrelated 2D NMR experiments with appropriate solvent suppression implemented. By applying this modification to the standard tool-kit of small molecule structural elucidation experiments - including COSY, TOCSY, HSQC and HMBC - it was possible to elucidate the structures of many of the compounds present in Scotch Whisky.

An example of the assignment procedure, using HSQC and HSQC-TOCSY spectra, is shown in Figure 5.3 for both the α and β-pyranose forms of glucose. In general, analysis of HMBC spectra was also necessary for the identification of quaternary carbons.

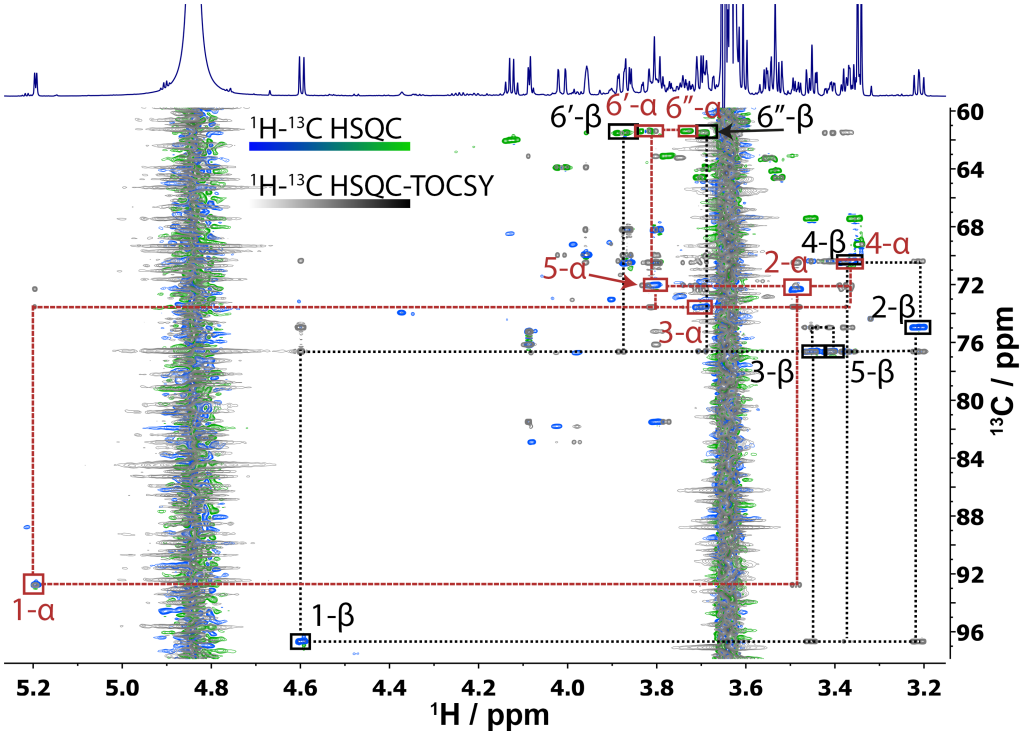


Figure 5.3: Superimposed partial 2D ^{1}H-^{13}C HSQC (blue/green) and HSQC-TOCSY (grey) NMR spectra of Scotch Whisky (S10-1313) showing the assignment of α and β-glucopyranose (**12b** and **12a**). Blue cross-peaks correspond to CH_2 signals whilst green are CH.

The elucidation of glucose and fructose in solution in whisky is an interesting challenge. Despite the experimental toolkit available, the problem is confounded by the fact these compounds are known to exist in multiple tautomers in solution. Fructose, for example, is reported to have five distinct forms in aqueous solution, with the β-pyranose (**31a**) being the predominant one (≈70 % at 20 °C) with the β-furanose (**31b**) and α-furanose (**31c**) comprising the majority of the remaining forms (≈20 % and ≈5 % at 20 °C, respectively),[225] while the α-pyranose and open-chain keto tautomers make up the remaining 5 %.

Tautomeric ratios of fructose and glucose are temperature dependent, but interestingly there is a solvent depenency, too. For glucose in water, six possible forms are reported including two acyclic forms (aldehyde and hydrate).[226] However, the pyranose forms dominate (99.6 %). The α:β-pyranose ratio is reportedly approximately 1:2.1 in water at 30 °C.[226] The spectra here were acquired at 25 °C and in whisky - an ethanol:water solution. By measuring the ratio of the anomeric protons in the 1D ^{1}H spectrum, in Scotch Whisky (S10-1313) 40 % ABV, the ratio is closer to 1:1.3. The proximity of the signals of anomeric protons, especially the β-tautomer, to the suppressed water signal may account for some signal attenuation and differences in this ratio. However, it is likely that pH and ethanol are

also altering it. For fructose, which does not have such well resolved anomeric signals (in the major conformers),[225] it is more challenging to quantify the ratio of signals. This will present a substantial challenge for absolute quantification of these compounds.

With the high-resolution data acquired, it was possible to identify most of the cross peaks present in the HSQC spectra of whisky. An example expansion of one such spectrum is shown in Figure 5.4. In this figure, the carbohydrate region of a Scotch Whisky (S10-1313) is shown and 33 cross-peaks are assigned out of approximately 40 in this region.

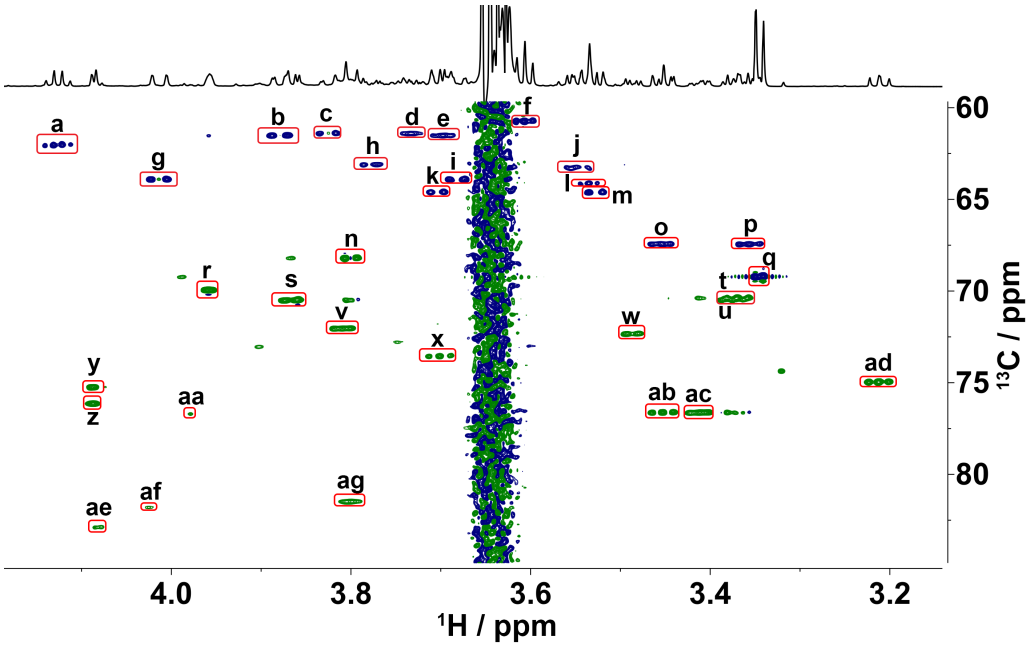


Figure 5.4: Partial 2D ^{1}H-^{13}C HSQC NMR spectrum of Scotch Whisky (S10-1313) with multiple cross-peaks in the carbohydrate region and assigned cross-peaks labelled. Annotations are detailed in Table 5.1. Multiplicity editing results in blue cross-peaks for CH_2 signals and green for CH.

For processing of the spectrum shown in Figure 5.4, in addition to routine linear prediction in F2, an NUS algorithm (MIST) was utilised for forward prediction in F1 in MestreNova. This allowed for an increase in digital resolution with reduced artefacts. The use of MIST in this case reduced F1 wiggles, which helped to resolve two of the signals from fructose labelled 'y' and 'z' in Figure 5.4. Normal linear prediction does not resolve these signals well from each other. Some linear prediction artefacts are noticeable in F2 at the cross-peak labelled 'q', for example. However, the spectrum resolution in both F1 and F2 is excellent.

There were some signals which could not be identified in compounds which were still determined to be present in whisky. For example, protons with a similar chemical shift to that of water or ethanol may be attenuated or entirely sup-

Table 5.1: Assignments of HSQC cross peaks in Figure 5.4 for Scotch Whisky.

SYMBOL	ASSIGNMENT	SYMBOL	ASSIGNMENT
a	ethyl acetate	r	5, β-fructopyranose
b	6′, β-glucopyranose	s	4, β-fructopyranose
c	6′, α-glucopyranose	t	4, β-glucopyranose
d	6, α-glucopyranose	u	5, α-glucopyranose
e	6, β-glucopyranose	v	4, α-glucopyranose
f	3-methylbutanol	w	3, α-glucopyranose
g	6′, β-fructopyranose	x	2, α-glucopyranose
h	6, β-fructofuranose	y	4, β-fructofuranose
i	6, β-fructopyranose	z	3, β-fructofuranose
j	1/1′, β-fructofuranose	aa	4, α-fructofuranose
k	1′, β-fructopyranose	ab	2, β-glucopyranose
l	n-propanol	ac	3, β-glucopyranose
m	1, β-fructopyranose	ad	5, β-glucopyranose
n	3, β-fructopyranose	ae	3, α-fructofuranose
o	2-methylbutanol	af	5, α-fructofuranose
p	2-methylbutanol	ag	5, β-fructofuranose
q	iso-butanol		

pressed, making their identification difficult or impossible. Similarly, proton signals which are overlapped with more abundant compounds may be impossible to resolve or assign in this way.

Assignment of the carbons in each molecule is non-trivial due to the low abundances of the compounds and the low natural abundance of ^{13}C. Even at a high magnetic field with a cryogenically cooled probe, not every expected carbon signal could be observed for some compounds, especially those approaching the limits of detection. For example, going back to higher alcohols, none of the carbons in n-butanol could be identified, and some of those in 5-HMF were not found either.

Interestingly, whilst many of the compounds in the ^{1}H NMR spectra of Scotch Whisky are present in all or most samples, some are only present in a small number of samples. These tend to be the more unusual sample types. For example, the whiskies finished in an ale cask (S10-1511, S12-1241, and S14-2087 - the same product from three years of production) contain an unidentified set of signals at 5.30 ppm (Figure 5.5).

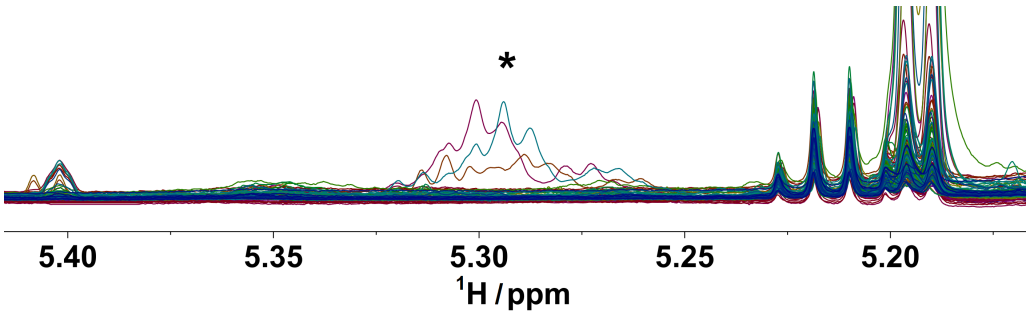


Figure 5.5: Partial 1D ^{1}H NMR spectra for 148 Scotch Whisky samples showing unidentified signals at 5.30 ppm (*) unique to the three samples finished in ale casks. Visible at $\delta(^{1}H) = 5.19$ ppm are the α anomeric protons of glucose, overlapping with the quartet of an acetaldehyde-related compound.

This chemical shift range is typical for anomeric protons of carbohydrates, and it seems probable that these signals correspond to carbohydrates found in beer, such as maltoses and dextrins.[82] That it is not a well resolved signal indicates the presence of multiple, structurally similar species, including different length oligomers.[227]

To further investigate the identities of these carbohydrates, additional NMR experiments were performed. Whilst complete structural assignment was not possible, the COSY and HSQC spectra (shown in Figure 5.6) were able to confirm that these are carbohydrate signals likely of dextrins or maltoses.

Figure 5.6a shows the carbohydrate region of the COSY spectrum, highlighting the proximity of these ale sugar COSY cross-peaks to those of the α anomeric pro-

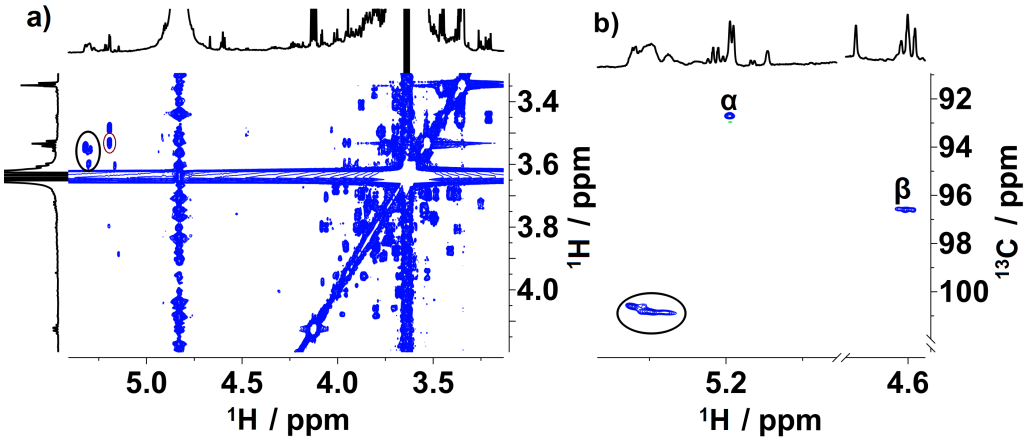


Figure 5.6: 2D a) ^{1}H-^{1}H COSY and b) ^{1}H-^{13}C HSQC NMR spectra for unknown carbohydrates in ale-cask matured Scotch Whisky (S12-1241). Black ellipse indicates cross-peaks corresponding to the ale sugars, and in a) red ellipse indicates cross-peak of α-glucose. Spectra acquired at 800 MHz (^{1}H) and 200 MHz (^{13}C).

ton of glucose, indicating structural similarities with dextrins. In both the COSY and HSQC spectra, as well as the 1D ^{1}H spectrum, the ale sugar signals are not resolved indicating, again, they are a number of different structurally species. In the HSQC spectrum, the ale sugar cross-peaks have similar carbon chemical shifts to those of α and β-glucose, too. However, they are more deshielded and, again, this is consistent with anomeric signals of maltoses and dextrins.[227]

Overall, using all of the aforementioned techniques, over twenty-two different structures have been identified in Scotch Whisky and partially or completely assigned. These structures are shown, with chemical shifts where determined, in Appendix A along with tabulated data and coupling constants. These comprise a range of compounds such as higher alcohols, sugars, and cask extractives including gallic and syringic acids.

5.4.3 *Acetaldehyde*

Carbohydrates are not alone in existing in multiple forms in solution. Interconversion of carbohydrate configuration involves ring-opening and closing, and the anomeric position in the closed ring form is an intramolecular hemiacetal. Another very important chemical in whisky is acetaldehyde. Despite its structural simplicity, acetaldehyde presented an interesting challenge to characterise and quantify in Scotch Whisky. In this section, the chemistry of acetaldehyde in ethanol water solutions, in the context of Scotch Whisky, is investigated.

Acetaldehyde will react with water, forming hemiacetals and poly(oxymethylmethylene) glycol oligmers - essentially hemiacetal oligomers.[228] This reaction will be under equilibrium. Additionally, Jaubert *et al.*[229] proposed that acetaldehyde (**32**) will react with ethanol (**33**), to form the hemiacetal (**34**), and oligomers of this hemiacetal, (*e.g.* **35**). These structures and reactions are summarised in Figure 5.7.

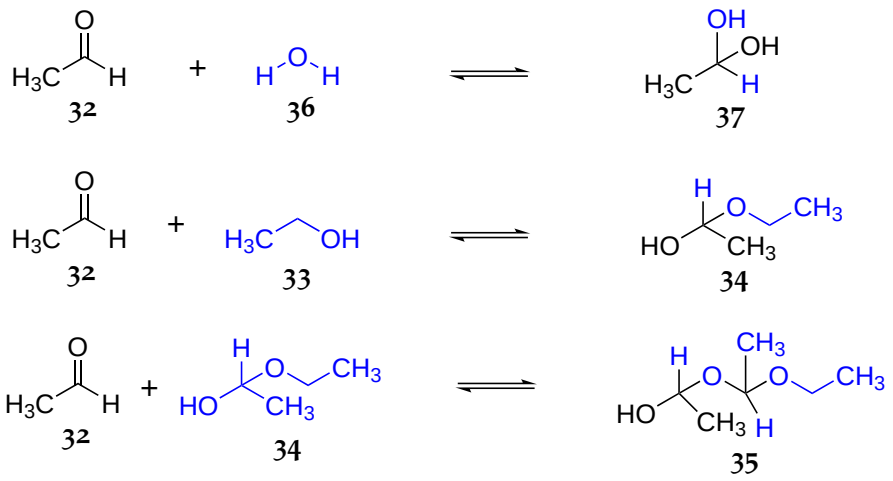


Figure 5.7: Scheme of hemiacetal formation from acetaldehyde (**32**) and water (**36**) or ethanol (**33**), producing the water hemiacetal (**37**), ethanol hemiacetal (**34**) , and ethanol oligomeric hemiacetal (**35**).

Studies on the equilibria of acetaldehyde in binary solvent systems of ethanol and water - with trace amounts of other alcohols and acids - are lacking, although Jaubert *et al.* noted that in the presence of acid, the reactions can also form acetals (*i.e.* 'acetal' (**38**) as it is known in the Scotch Whisky literature, shown in Figure 5.8), potentially irreversibly.[229] Similar reactions may happen for other alcohols in solution - however as ethanol is the dominant alcohol in whisky, these other reactions are not of significance here.

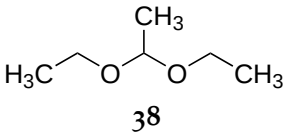


Figure 5.8: Structure of acetal (**38**).

To resolve the composition of acetaldehyde in aqueous ethanol solution, a reference acetaldehyde spectrum was obtained (Figure 5.9). This involved dissolving a drop of pure acetaldehyde in a 40 % ethanol:water solution (100 µL) and addition of the D_2O buffer used for the normal whisky experiments (500 µL). The substantially more diluted levels of protonated water and ethanol in this model

mixture remove the need to use the developed solvent suppression routines, affording faster and more sensitive acquisition.

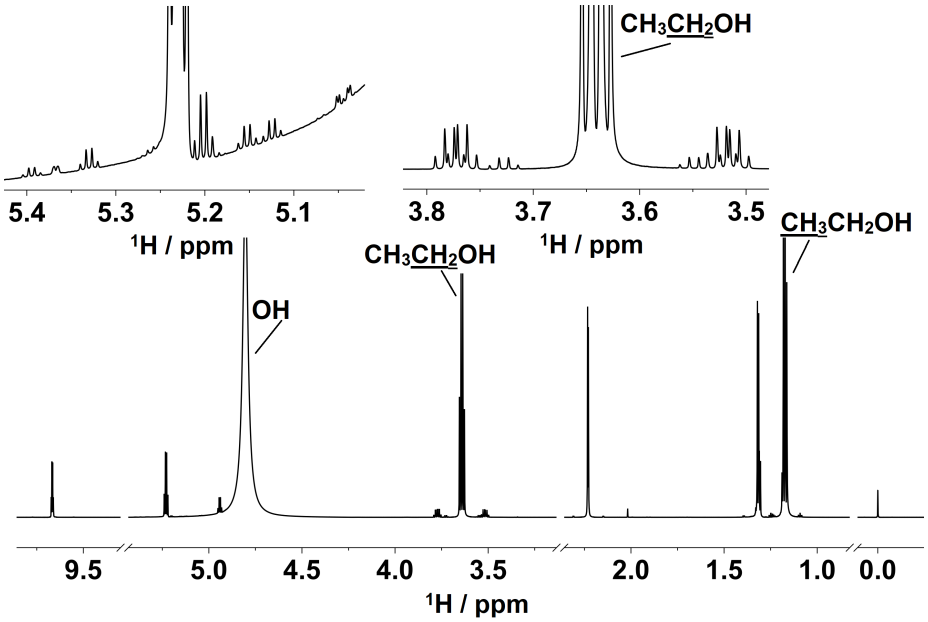


Figure 5.9: 1D ^{1}H NMR spectrum of acetaldehyde in a 40 % ethanol water solution. Acetaldehydes aldehyde signal is at 9.6 ppm. Top left expansion shows region around 5.2 ppm and several minor multiplets here, residual water signal is distorting the baseline. Top right expansion shows multiplets surrounding ethanol CH_2 signals.

The expected methyl signal ($\delta(^{1}H) = 2.23$ ppm, $\delta(^{13}C) = 30.8$ ppm) and aldehyde signal ($\delta(^{1}H) = 9.88$ ppm, $\delta(^{13}C) = 208.9$ ppm) (chemical shifts as measured in whisky) were observed. However, additional signals were also observed. These include significant quartets at 5.22 ppm (see also Figure 5.5) and 4.92 ppm (both with $^{3}J_{HH} = 5.2$ Hz). The top left expansion in Figure 5.9 reveals the presence of many minor signals - quartets and doublets of doublets - adjacent to these more prominent quartets. There are also complex signals adjacent to the ^{13}C satellites of ethanol CH_2, shown in the top right expansion, and overlapping signals at 1.3 ppm (overlapping doublets coupled to the aforementioned quartets), along with a signal overlapping with the ethanol triplet for CH_3. Beyond a standard 1D ^{1}H NMR spectrum, further spectra were acquired to understand the nature of these signals.

A ^{1}H-^{1}H TOCSY NMR spectrum was also acquired, with a long mixing time of 120 ms to ensure multiple-bond TOCSY-transfer. Most correlations only connected two signals - *i.e.* the spin systems consisted of only two proton environments. One set of signals did have three correlations - between geminal protons of non-equivalent CH_2 protons (later confirmed by HSQC) and their neighbouring CH_3

protons. These belong to an ethyl group bonded to a chiral carbon through an ether bond, such as seen in a hemiacetal.

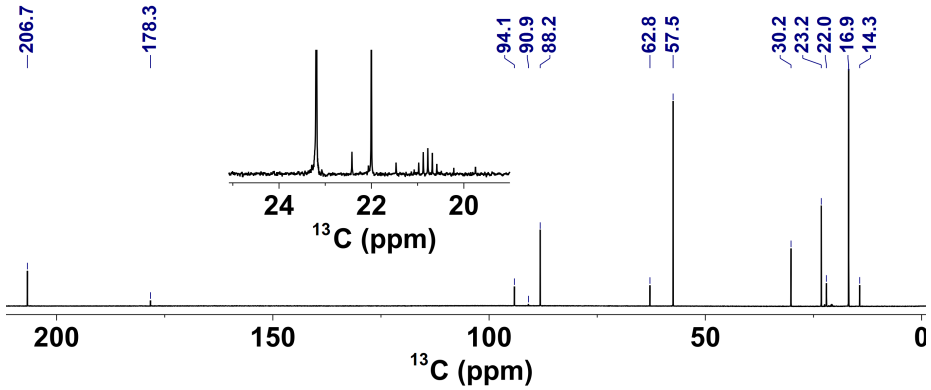


Figure 5.10: 1D ^{13}C NMR spectrum of acetaldehyde in a 40 % ethanol water solution. Peaks have been picked and annotated with their chemical shifts for clarity. Expansion shows minor signals around 22 ppm. Spectrum acquired at 200 MHz

The acquired ^{13}C NMR spectrum, Figure 5.10, shows over 12 singlets corresponding to acetaldehyde equilibria products. Bearing in mind that both acetaldehyde and ethanol have only two carbon environments each, the presence of 12 signals indicates that a range of species are present. The expansion in this figure shows some minor peaks which have not been labelled, including a multiplet at 20.8 ppm. This experiment was acquired with the power gated ^{1}H decoupling, so this multiplicity must correspond to a deuterated species. The D_2O buffer solution contains deuterated acetic acid, and this accounts for the signals at $\delta(^{13}C)$ = 178 ppm (singlet) and 20.8 ppm (pentet). The carbon spectra were cross-referenced with the work of Jaubert *et al.*[229] confirming that these signals do indeed correspond to hemiacetals. Identified peaks also include acetaldehyde ($\delta(^{13}C)$ = 206.7 ppm and 30.2 ppm) and ethanol ($\delta(^{13}C)$ = 57.5 ppm and 16.9 ppm).

It is possible that the species that give rise to these signals are involved in chemical exchange. To investigate if exchange was present, NOESY spectra were acquired with two mixing times (200 ms and 700 ms) to try to observe exchange signals and identify signals from protons close in space. The results were of limited use, showing no NOESY cross-peaks, only one exchange spectroscopy (EXSY) cross-peak and several COSY-like artefacts. The EXSY cross-peak is shown in Figure 5.11 and confirms that acetaldehyde is in exchange with a methyl at approximately 1.27 ppm.

Diffusion measurements can be used to infer the relative size of species in solution. Initially, a 2D ^{1}H DOSY spectrum was acquired using a standard parameter

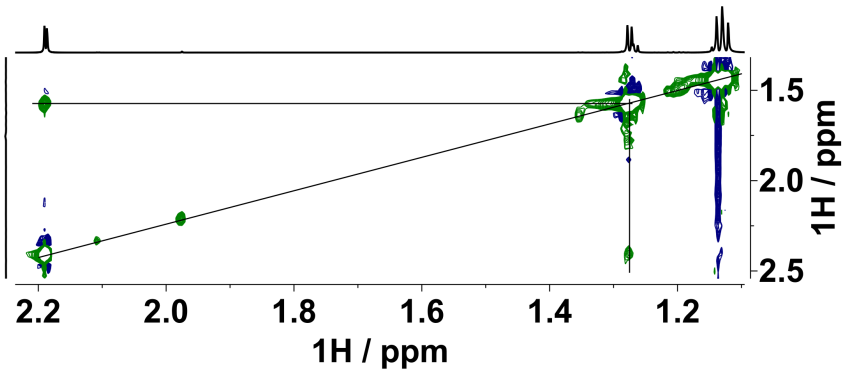


Figure 5.11: Partial 2D ^{1}H NOESY NMR spectrum of acetaldehyde in ethanol:water. Region shown includes three diagonal cross-peaks and one EXSY cross-peak showing exchange between acetaldehyde and an unknown compound methyl. Mixing time was 700 ms

set and pulse sequence. The acquired spectrum is shown in Figure 5.12a. In these DOSY spectra, cross peaks at the bottom belong to molecules that diffuse fastest and have the smallest hydrodynamic radius - *i.e.* at a simple approximation, are the smallest. In Figure 5.12a, the ^{1}H water (4.7 ppm, 18 Da) and DSS$-d_6$ (0 ppm, 224 Da) signals are well resolved, and therefore the mass range of species in solution can be approximated; all the major species present must have a mass between 18 Da to 224 Da. Similarly, acetaldehyde - especially the aldehyde signal (9.6 ppm) - is well resolved, confirming the unknown compounds must be larger than acetaldehyde.

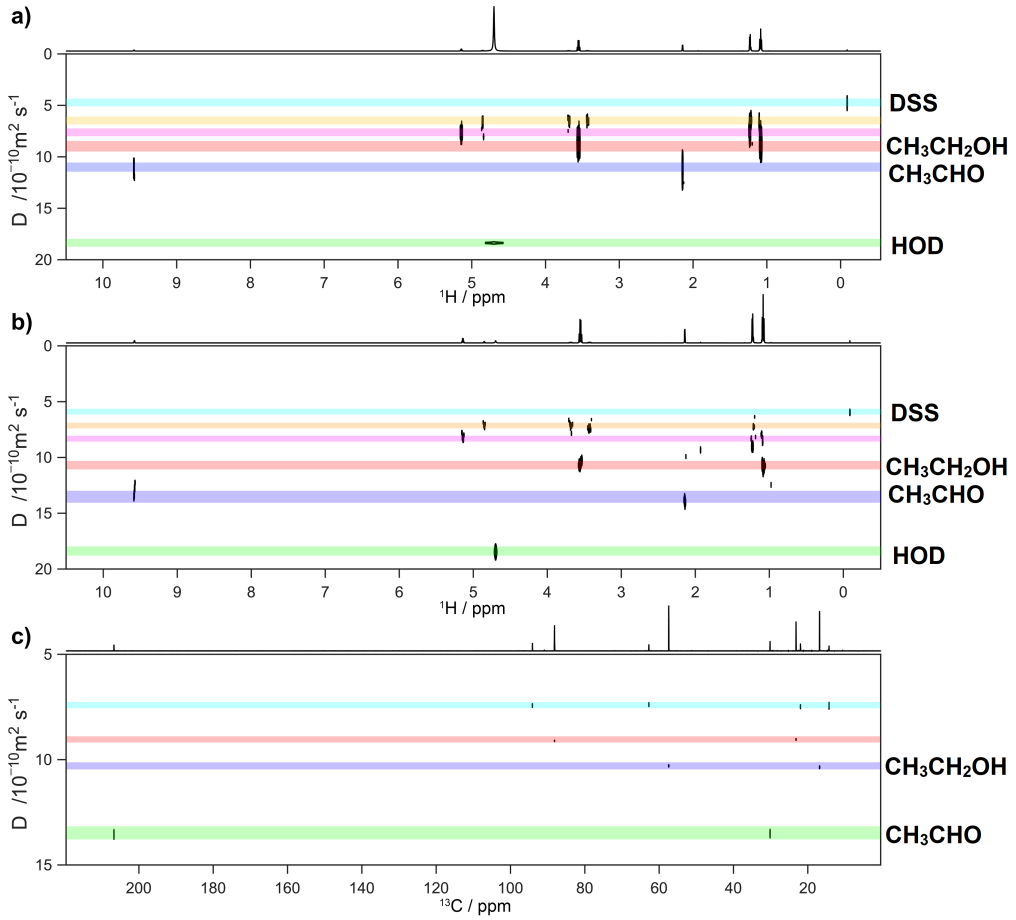


Figure 5.12: 2D DOSY NMR spectra of acetaldehyde in ethanol:water. a) 600 MHz ^{1}H DOSY spectrum acquired with stimulate echo, longtitudinal eddy delay (LED) and convection compensation (Bruker sequence *ledbpgp2s.compensated*), b) 600 MHz ^{1}H DOSY spectrum acquired with chemical exchange suppression (PROJECTED sequence),[204] c) 125 MHz (500 MHz spectrometer) ^{13}C DOSY spectrum acquired with stimulated echo and INEPT (Bruker sequence *stebgpin1s*). Coloured bars added for clarity.

Unfortunately, many of the cross peaks are not well resolved in the diffusion dimension - they are spread or 'smeared' vertically. This can be due to several reasons, however in this case it is likely due to two compounding issues. Firstly, routine DOSY data processing involves fitting a monoexponential decay to a series of attenuated peaks. If a peak is the superposition of signals from multiple, different sized molecules, the monoexponential fit may have a large error associated with it. Overlap, either partial or complete, of signals will distort the DOSY spectrum.

A simple means to resolve overlapping peaks, and thus determine if they are present, is by 'pure-shift' NMR, here obtained by acquisition of a 1D ^{1}H pure-shift yielded by chirp excitation (PSYCHE) NMR spectrum acquired for the acetaldehyde sample, and shown in Figure 5.13a,c. It can be seen that there are overlapping signals in several regions of the spectra, including the ethanol CH_3 signal at 1.13 ppm and other methyls at 1.28 ppm. The ethanol CH_2 signal at 3.6 ppm appears to have an underlying secondary signal too, however the resolution of the PSYCHE spectrum is not sufficient to be sure of this. However, the complex multiplets either side at 3.5 and 3.7 ppm are seen to be clearly just one signal each. The PSYCHE method has a significantly reduced sensitivity and longer experimental time than the normal 1D ^{1}H experiment.[217] A pure-shift DOSY experiment does exist, but it is a pseudo-3D experiment and takes a significant amount of time to acquire.[230]

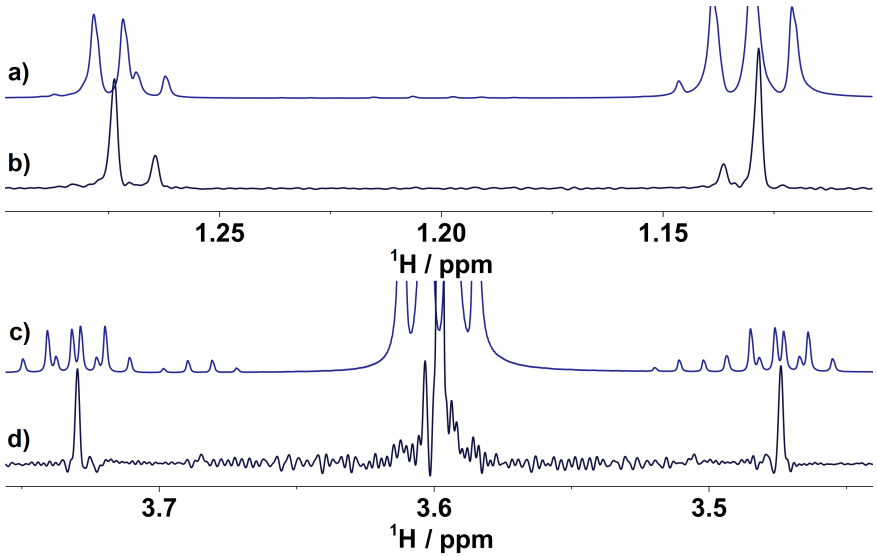


Figure 5.13: Normal (a,c) and pure-shift (b, d) 1D ^{1}H NMR spectrum of acetaldehyde in ethanol water solution. Spectrum (b,d) acquired with PSYCHE-TSE pure-shift method with 16 scans per increment and 64 increments giving a reconstructed TD of 12800 points, zero filled to 32k points with 0.3 Hz line broadening applied prior to Fourier transform. Expansions show a,b) 1.1 to 1.2 ppm and c,d) 3.45 to 3.75 ppm.

The second cause of broad signals will be chemical exchange. If the species are exchanging on the timescale of the diffusion measurement, their decay will be a weighted average of the decays of exchanging species. Whilst the DOSY cross-peaks of ethanol may be broad due to signal overlap, the aldehyde signal was not observed to have significant overlap. Thus, its unresolved diffusion must be due to exchange. To compensate for this effect, Aguilar *et al.* developed the PROJECTED pulse sequence for measuring diffusion whilst suppressing chemical exchange.[204]

The resulting spectrum acquired with this sequence is shown in Figure 5.12b, and shows marked improvement over the normal DOSY sequence. The cross peaks are significantly narrower in the diffusion dimension, increasing the precision of the measurement. The accuracy is unknown. For example, the DSS$-d_6$ signal has a measured difussion coefficient (D) of $5.96(20) \times 10^{-10}\,m^2\,s^{-1}$, whilst in the normal spectrum it was $4.78(60) \times 10^{-10}\,m^2\,s^{-1}$. This corresponds to a mass - as calculated according to Evans *et al.*[231] - of between 172 Da to 154 Da in the PROJECTED measurement and 206 Da to 359 Da for the normal sequence. Thus, the PROJECTED sequence is more precise but less accurate than the normal experiment.

Bearing in mind that there is a significant error in the masses calculated from both of the ^1H diffusion measurements, it is still possible to infer relative size and number of different species present based on the DOSY spectra. As the PROJECTED data is better resolved, it is used in the following discussion. As can be seen in Figure 5.12b, from bottom-to-top, there is water, acetaldehyde, ethanol, two major bands of signals, and DSS$-d_6$. These two bands likely correspond to hemiacetals.

Despite the suppression of exchange in the DOSY spectrum, there is still signal overlap which impedes analysis. To resolve this, a ^{13}C DOSY spectrum was acquired and is shown in Figure 5.12c. This used an INEPT sequence to diffuse on ^1H prior to magnetisation transfer to ^{13}C for detection. As such, this spectrum only shows carbons directly bonded to a proton, and lacks the acetic acid quarternary carbon observed in the 1D ^{13}C spectrum (Figure 5.10).

The ^{13}C DOSY spectrum (Figure 5.12c) shows excellent resolution in both chemical shift and diffusion dimensions, allowing for precise measurements. It has fewer signals than the ^1H due to the significantly reduced sensitivity - and therefore took longer to acquire than a ^1H DOSY spectrum (1 h compared to 15 min), and shows no signal for DSS or water, giving fewer reference points. This reduced sensitivity focuses the problem on only the major compounds present. It clearly shows acetaldehyde, ethanol, and two other, larger, species and is consistent with the ^1H DOSY data. The measured chemical shifts, diffusion rates, and calculated

masses are summarised in Table 5.2 for the ^{13}C DOSY. Diffusion values were calculated using GNAT,[205] and masses calculated from a spreadsheet published by Evans *et al.*[231]

Table 5.2: Measured diffusion values and calculated masses for major compounds in ^{13}C DOSY spectrum. Errors not shown for calculated masses - all were much less than 1 Da.

COMPOUND	CHEM. SHIFT (^{13}C ppm)	DIFFUSION ($10^{-10}\,m^2\,s^{-1}$)	CALC. MASS (Da)
Acetaldehyde CHO	206.7	13.55 ± 0.17	31
Acetaldehyde CH_3	30.2	13.51 ± 0.13	31
Ethanol CH_2	57.4	10.30 ± 0.02	52
Ethanol CH_3	16.8	10.36 ± 0.03	51
X1 - Signal A	88.2	9.11 ± 0.01	67
X1 - Signal B	23.2	9.04 ± 0.02	68
X2 - Signal A	94.1	7.44 ± 0.11	101
X2 - Signal B	62.8	7.49 ± 0.05	100
X2 - Signal C	22.0	7.39 ± 0.05	103
X2 - Signal D	14.3	7.43 ± 0.05	101

As can be seen in Table 5.2, the calculated masses for ethanol and acetaldehyde are inaccurate - ethanol is overestimated by ≈6 Da whilst acetaldehyde is underestimated by ≈13 Da. However, they are of the correct magnitude, allowing for approximation. The two other species - X1 and X2 - are larger with calculated masses of ≈67 Da and ≈101 Da.

Finally, 2D ^{1}H-^{13}C HSQC and HMBC spectra were acquired. As the proposed hemiacetals contain ether bonds, only HMBC, and not TOCSY or other homonuclear *J*-correlated experiments, will allow the unambiguous correlation between signals to be determined. Whilst the ^{13}C DOSY was able to resolve the carbon signals from different compounds, it could not correlate this to proton chemical shifts. Based on their chemical shifts, relative diffusion rates, and number of signals, it is possible to assign the two major unknown species as the hemiacetal formed between acetaldehyde and water (X1, (**37**)) and the hemiacetal formed between acetaldehyde and ethanol (X2, (**34**)).

This focused investigation on acetaldehyde in a model ethanol water solution has identified the two major hemiacetals formed, both of which are routinely visible in Scotch Whisky NMR spectra. The equilibria between these species will be pH and solvent concentration dependent, and thus variable between samples.

Quantification of acetaldehyde by integration of signals from only its native structure, and not its hemiacetals, by NMR is therefore likely to be inaccurate. Simultaneous integration of acetaldehye and its major hemiacetals would be required, but the proximity of signals to the residual water and ethanol solvent signals will further impede analysis.

5.5 QUANTIFICATION

One of the most significant advantages of NMR is that it is a quantitative technique when performed correctly. In this section, the quantification of both a major component of Scotch Whisky (ethanol) and minor compounds is investigated.

5.5.1 *Ethanol Quantification*

Despite the chemical complexity of Scotch Whisky, one of the most important aspects of the product is the alcoholic strength. By law, it must be a minimum of 40 % ABV, and so-called 'cask strength' products may be significantly stronger. It is therefore routinely necessary to quantify the ethanol strength of a sample. In the developed solvent suppression methodology, there are two spectra acquired for ethanol suppression which may serve a dual purpose and may be used for ethanol quantification. These are the reverse INEPT experiment (see subsection 4.4.2) and the single scan 1D ^{13}C NMR experiment (see subsection 4.4.3).

To test both methods, a series of model ethanol water solutions were produced between 35 % and 70 % ABV, covering the expected ranges of ethanol in a Scotch Whisky sample. Quantification below this level is not required for authentication as any sample below the legal minimum of 40 % ABV would not be classed as Scotch Whisky. Likewise, samples above 70 % ABV are very unlikely to be encountered in final products, although new make spirits may exceed this level - especially for grain spirits. To make the calibration standards consistent with the Scotch Whisky NMR method, the D_2O buffer containing DSS$-d_6$ and acetic acid/sodium acetate was used in the sample preparation.

In the ^{1}H coupled reverse INEPT experiment (see subsection 4.4.2) a 1D ^{1}H NMR spectrum with ^{13}C bound protons of ethanol anti-phase is obtained, whilst the ^{12}C bound protons are suppressed. For quantification of ethanol, the ^{13}C satellite signals could theoretically be integrated. Beyond acquisition parameters, accurate quantification relies upon good phasing and baseline. However, the anti-phase signals in this spectrum present a challenge. In most cases it was possible to achieve acceptable, but imperfect phasing, of these signals. There were also baseline distortions affecting the integration of these signals caused by imperfect

suppression of ^{12}C-bound protons, which even relatively aggressive algorithms, like the Whittaker smoother,[232,233] could not completely compensate for. Finally, as ethanol strength varies, so does the chemical shift of its protons. Integrating each signal area consistently is non-trivial in this instance. As such, this method is not expected to produce highly accurate results.

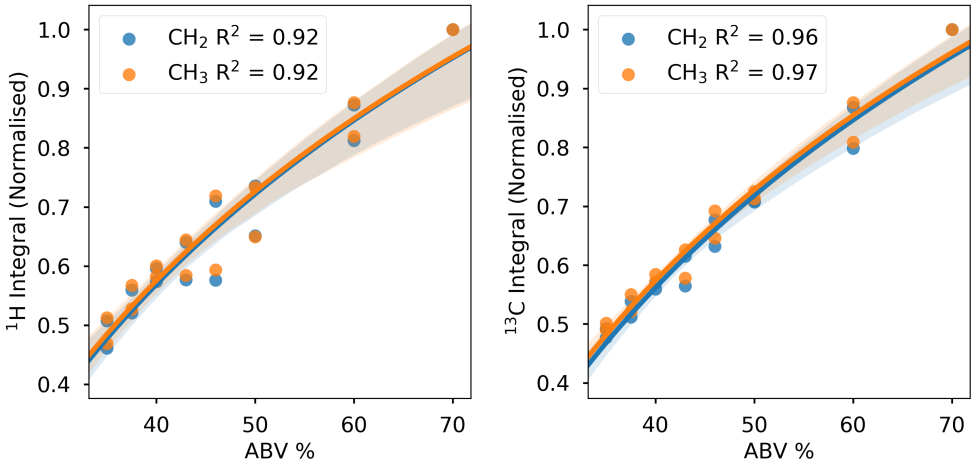


Figure 5.14: Quantification of ethanol by integration of a) ^{1}H NMR and b) ^{13}C NMR as described in the text. Note that the integrals were normalised to the maximum value to aid visualisation. Linear fits, and correlation coefficients, were calculated for log(ABV) v. integral, whilst the plot is of the original variables. Log fits achieved superior results. R^2 values shown are the square of the Pearson correlation coefficient. The light coloured areas indicate a 95 % confidence interval.

The spectra were manually phased prior to integration with the Data Analysis tool of MestreNova, which allows for variable integration regions across a series of stacked spectra - thus reducing the effect of chemical shift variation. The asbolute values of the two integrals corresponding to the two parts of the anti-phase doublets for each of the CH_2 and CH_3 signals were summed and then normalised to their maximum values to allow visualisation on one plot. The results are shown in Figure 5.14a, with R^2 values of 0.92 for both integrals. The confidence interval is at its narrowest between 35 % and 45 % ABV.

Ethanol could also be quantified from the inverse-gated 1D ^{13}C NMR spectrum (see subsection 4.4.3). Despite being a single scan experiment on a ^{13}C - which has lower natural abundance and sensitivity than proton - because ethanol in Scotch Whisky is at the molar level, the resulting SNR is more than sufficient for quantification. Spectra were integrated and normalised as above, and the results are shown in Figure 5.14b. The ^{13}C method performs better than for the ^{1}H reverse INEPT method, with R^2 values exceeding 0.96. The confidence interval is significantly narrower across the whole range of alcohol strengths.

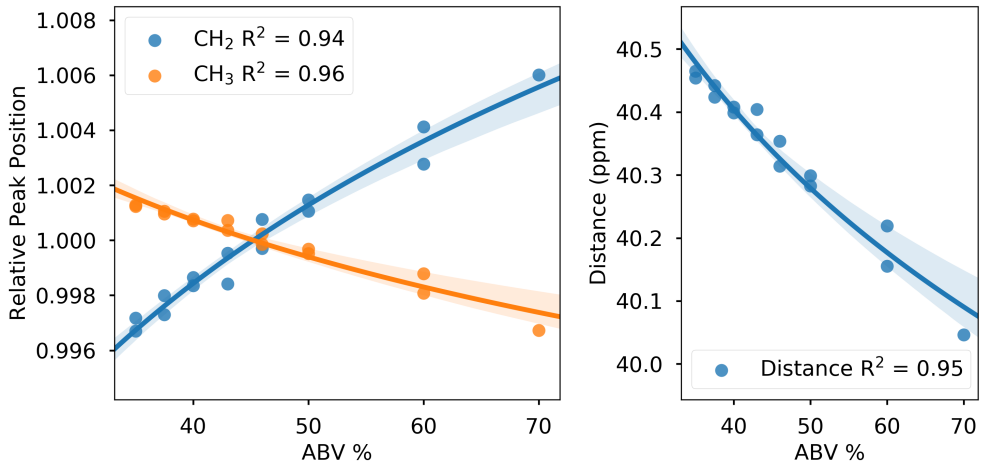


Figure 5.15: Quantification of ethanol by ^{13}C NMR peak positions. (a) shows the relative peak position for both carbon signals, (b) shows the distance between peaks in ppm. Note that the peak positions in a) were normalised to the mean value to aid visualisation. Linear fits, and correlation coefficients, were calculated for log(ABV) versus peak position/distance, whilst the plot is of the original variables. R^2 values shown are the square of the Pearson correlation coefficient. The light coloured areas indicate a 95 % confidence interval.

As already mentioned, chemical shift varies with concentration. It was noted that the ^{13}C peaks of ethanol shift closer together as alcohol strength increases. This effect was more noticeable and consistent than for the ^{1}H spectra. This represents a potentially powerful means to quantify ethanol strength where the usual concerns about phase, baseline, and integration region selection do not matter. Furthermore, determining the peak position does not require full polarisation of the carbon nuclei to be stabilised as per quantitative NMR, and the experiment could be performed faster. Additionally, the number of scans and receiver gain can vary, provided the SNR is sufficient for peak picking. In this example, and in Scotch Whisky at 40 %, these points are rather academic as quantative results are achievable in a single scan. However, measuring peak positions is inherently more straightforward than accurately and consistently integrating Lorentzian peak areas. To measure the accuracy of using peak position as a means of quantification, the peak positions in the carbon spectrum were measured and normalised. The results are shown in Figure 5.15a. The correlation coefficients are comparable to those of the integration method at around $R^2 = 0.95$, with narrow confidence intervals.

The distance between the peaks represents a more robust measurement which may be used for quantification. Measuring the distance between the two peaks eliminates potential error in determining the resonance frequency of ethanol signals. Shown in Figure 5.15b, this is indeed the case, with a R^2 value of 0.95. This

observation has potential value when you consider that the automated solvent suppression already determines the distance between the carbon peaks for selective decoupling (subsection 4.4.3). The pulse length calculated is directly proportional to this distance, and could therefore be used for a means of automatic ethanol strength quantification without the need for user intervention or manual data processing.

Whilst the above data confirms that ethanol can be quantified on model ethanol:water solutions, the important test is on real samples of Scotch Whisky. In this chapter, 148 samples of Scotch Whisky were analysed by NMR. All these samples had a narrow range of reported alcohol strengths of 40 %, 43 %, or 46 % ABV. As the ^{1}H INEPT method yielded poorer results for our model solutions, and requires increased processing effort, only the ^{13}C integrals and peak positions were used here.

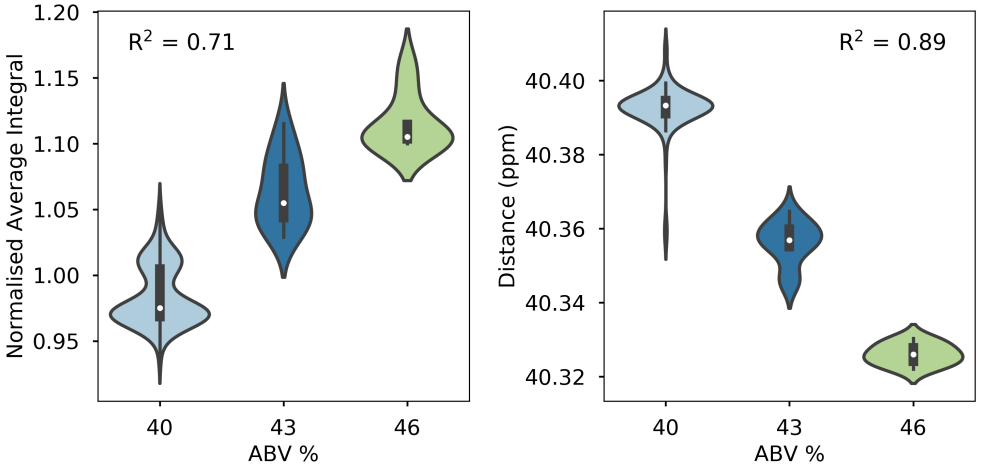


Figure 5.16: Violin plots for the a) mean ^{13}C integrals of ethanol signals, b) distance (ppm) between ethanol CH_2 and CH_3 signals. A violin plot shows the distribution of data in a similar means to a box plot, though the shape of the 'violin' represents a kernel density estimate of the distribution of values within the data. The white dots are the mean, whilst the black bars are essentially a box plot. R^2 values (square of Pearson correlation coefficient) are shown.

First, the carbon integrals were compared to the label strength. Given the narrow range of alcohol strengths investigated, a purely linear regression was used for these data. As expected, there was a correlation, however the R^2 value was only 0.71. (Figure 5.16a). Interestingly, when viewed separately, the methyl carbon gave a greater R^2 than the methylene (0.78 v 0.62). (Figure 5.17a). The cause of this is unknown, as in the model solutions both carbon integrals performed equally well. These R^2 values are significantly poorer than for the model solutions.

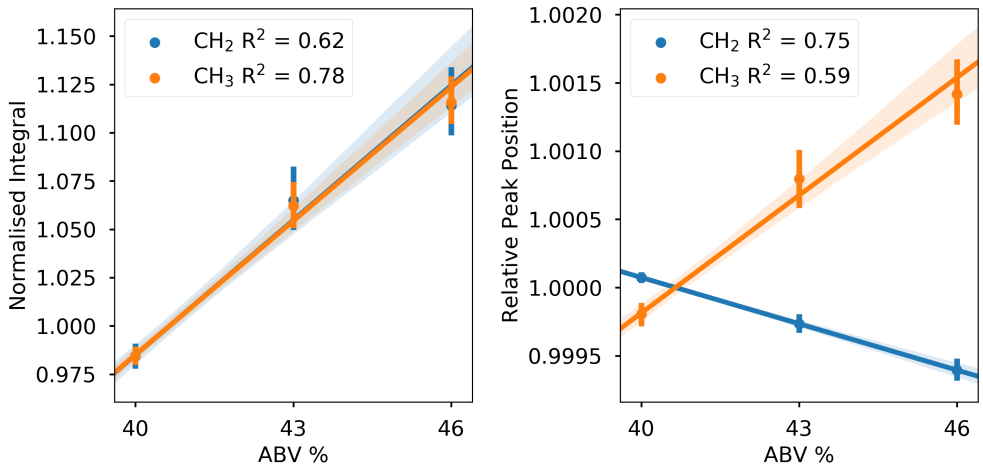


Figure 5.17: Linear regression plots for ^{13}C mean normalised integrals of ethanols a) CH_2 and CH_3 signal and b) relative chemical shifts of CH_2 and CH_3. Relative chemical shifts calculated by mean normalisation. R^2 (square of Pearson correlation coefficient) values are shown along with 95 % confidence intervals.

Secondly, the chemical shifts of the ethanol carbon signals was investigated. The correlation between the absolute chemical shifts of these signals and the label ABV was limited, with R^2 values of 0.75 and 0.59 for CH_2 and CH_3, respectively. See Figure 5.17b. The relative distance between these carbons was used to infer ethanol strength (a smaller distance indicates higher strength). The correlation of these parameters showed a much stronger correlation, with an R^2 of 0.89, as shown in Figure 5.16b. This supports the quantification of ethanol by this means.

The violin plots show a kernel density estimate of the distribution of measurements of integrals or distances. As can be seen in Figure 5.16, there is substantial variation in the distribution of values for each ABV. Label ABV need not be exactly precise - a 0.3 % tolerance is allowed (EU Regulation 2011/1169) - and this may explain some of the variance seen and poorer R^2 values when compared to the model ethanol solutions. Another potential source of error include evaporation - samples had been opened and subsampled both at SWRI and here, and ABV may have changed as a result over time. This issue is impossible to avoid without sampling directly at the bottling hall and storing frozen and sealed until this analysis - an obviously impractical sampling scheme for this experiment.

5.5.2 *Quantification of Other Compounds*

Quantification of individual compounds in the full 1D ^1H NMR spectrum is possible. With a long recycle time - with the acquisition time and relaxation delay totalling nearly 11 seconds - the proton spectrum should be quantitative. How-

ever, several of the compounds present may adopt multiple forms in solution - with unknown ratios. This has been discussed previously with regards to carbohydrates and aldehydes. Similarly, if any higher alcohols are involved in hemiacetal formation, their quantification may be affected. Peak overlap also hinders quantification, requiring sophisticated line fitting approaches. As chemical shifts may alter between samples, it is also possible for the ability to resolve overlapping species to falter between samples. Various commercial and open source software have been developed in an attempt to quantify compounds in a mixture, including – Bruker Assure, MestreNova SMA, ACD/Labs Quanalyst, Bayesil,[234] BATMAN,[235] and rDolphin.[179] These tools were tested, excluding Bayesil and rDolphin, and all struggled with the Scotch Whisky spectra and failed to produce acceptable quantification results when benchmarked against SWRI quantification of the same samples. They also require significant user input. Based on these difficulties, it was decided not to focus further effort on quantification of other analytes in Scotch Whisky in this work. The potential to do so exists and is an area worth future investigation.

5.6 CHEMOMETRIC ANALYSIS

Statistical analysis is best performed on a reduced set of input variables, where the variation between them depends on the sample composition more than inconvenient variation –- chemical shift changes, overlapping peaks, etc. One approach to reduce the number of variables in each NMR spectrum is to quantify the compounds present. This would be the gold standard of input data as it allows for easy verification of the input data and intuitive understanding of the resulting models. However, it would require accurate quantification of most, if not all, of the signals in the spectrum - any compounds not quantified would be data wasted. In the absence of quantitative information about the compounds derived from ^{1}H NMR, statistical analysis is performed on binned data.

5.6.1 *Data Preprocessing*

For the statistical analysis of spectroscopic data, the first step is to ensure spectra are processed in as similar fashion as possible. In other words, to properly examine the variation between samples, there must be minimal variation in acquisition and processing of data. To this end, spectra are acquired in automation and processed with automatic zero order phase corrections, pivoted around DSS. Automatic baseline corrections are applied, and signals are referenced to DSS at 0 ppm. Spectra are then superimposed and visually inspected. Spectra

with baselines significantly distorted are manually reprocessed with appropriate phase and baseline corrections.

5.6.1.1 *Peak Alignment*

After phase and baseline corrections, residual chemical shift variations between spectra must be resolved. Spectra are peak aligned interactively using icoshift.[236] This alignment step is performed in TopSpin passing the data through Octave for processing using in-house written scripts. The icoshift process is repeated up to twice to generate optimal results. It can, however introduce slight artefacts at alignment-region boundaries. An example is shown for signals of isobutanol and 3-methylbutanol in Figure 5.18 a) before and b) after peak alignment.

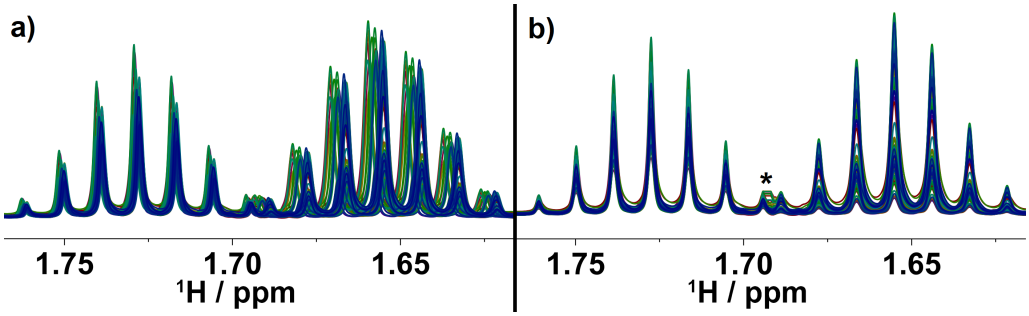


Figure 5.18: Partial 1D ^{1}H NMR spectra of all 148 Scotch Whisky spectra between 1.78 and 1.60 ppm showing a) unaligned and b) aligned peaks. Asterisk indicates minor artefact produced with manual icoshift region selection.

In Figure 5.18b, the small distortion to the signal introduced by alignment is annotated. It is clear, however, that the alignment procedure has, overall, improved the comparability of the spectra. To quantify this peak alignment improvement, the peak positions for the isobutanol multiplet between 1.78 and 1.68 ppm were measured on the binned data for the aligned and unaligned dataset.

Within this region of the spectra, the median standard deviation in peak positions across 148 spectra for the unaligned data was five-times greater than the bin width (0.001 ppm), whilst for the aligned data it was effectively zero. At an even finer bin width (0.0001 ppm), it was four-times greater than the bin width for the unaligned data, and again effectively zero for the aligned data. Therefore, the alignment procedure allows us to use a high resolution binning.

5.6.1.2 *Spectral Binning*

Binning of the NMR data serves two purposes. Firstly, it acts to reduce subtle differences in peak positions, which left unchecked could affect statistical interpretation. Secondly, it reduces the size of the data, or number of variables, to

process. In the sample set of 148 samples, each spectrum was acquired with 128k time domain points and zero filled to 256k points. Performing statistical analysis on a 148 x 256k matrix - totalling 37 million data points - would be computationally taxing, although not impossible. The distance between points in the Fourier transformed spectrum is 0.00006 ppm. By binning at, for example, 0.001 ppm - and excluding the left and right extremes of the spectra which contain only noise - the data can be reduced to only 10 000 points per spectrum, giving a total of only ≈1.4 million points, a 26-fold reduction in data. At 600 MHz, 0.001 ppm is only 0.6 Hz, less than the linewidth of a signal, and so much of the important detail in the spectrum is retained. A comparison of bin sizes is shown in Figure 5.19.

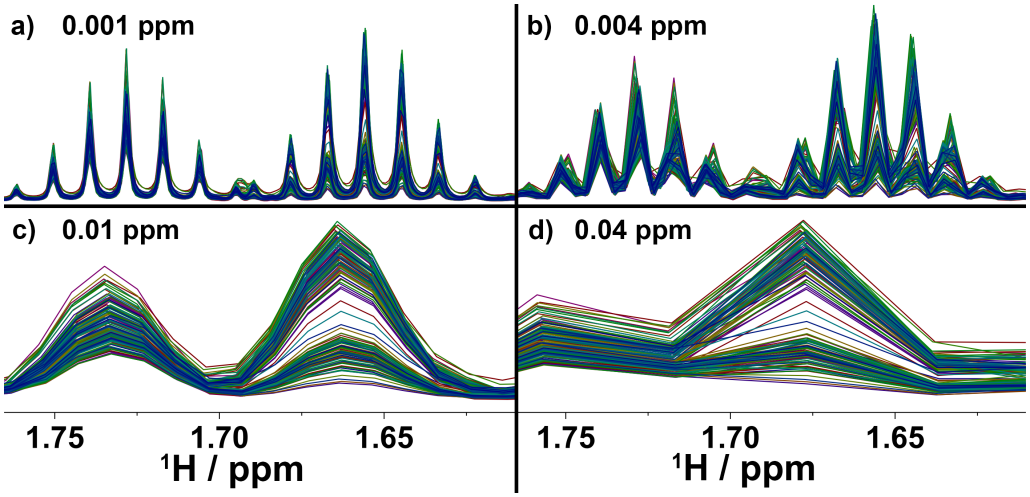


Figure 5.19: Partial 1D ^{1}H NMR spectra of all 148 Scotch Whisky spectra between 1.78 and 1.60 ppm showing variable bin sizes from 0.001 ppm to 0.04 ppm.

The bin size comparisons in Figure 5.19 demonstrate how a fine bin size (0.001 ppm) retains most of the data whilst simultaneously reducing the number of variables. At coarser bin sizes, the multiplicities are first distorted (Figure 5.19b), then peak positions shift (Figure 5.19c), and then the baseline is severely distorted (Figure 5.19d). It is abundantly clear that coarse bin sizes come with severe penalties to the quality of the data, and any models constructed from such reduced data may suffer from lack of interpretability or validity against the real data.

The binning procedure used here is a simple uniform binning method where every x ppm region is integrated and reduced from many points (variables) to just one value. Alternative methods do exist for binning. At the other extreme from the simplistic uniform binning method exists fitting each lineshape in the spectrum and returning an area for these. This is closely related to 'simply' quantifying each compound in the sample. Various methods to address line fitting have been implemented, including the GSD method in MestreNova,[237] which op-

erates *ab initio*, or methods which fit lineshapes based on *a priori* knowledge, such as the BATMAN routine.[235] The former method is prone to error and does not provide any chemical information, whilst the latter requires substantial spectral database development and fine tuning of the parameters, including those of the initial Bayesian model. However, even if this line fitting approach was successful per spectrum, to tabulate this across all spectra for statistical analysis would require the same lineshapes (compounds) to be present - and resolvable - in each spectrum. This is inherently not the case in Scotch Whisky where some samples may contain unique species. Any approach to resolve this - such as excluding peaks found only in some samples and not in others - would be a compromise not associated with uniform binning.

Between these two extremes exist a variety of sophisticated binning routines which do not require spectral deconvolution or *a priori* knowledge of the composition of each sample. These include manually defining integral regions, kernel based binning (*e.g.* gaussian bins), and 'intelligent' binning. Many of these routines are available in the MVAPack suite of tools.[238] Whilst these methods may be preferable for techniques such as PCA, the uniformly binned data are preferred in spectral reconstruction for model interpretation from loadings, or for application with other statistical methods including STOCSY and ICA.

5.6.2 *STOCSY*

In an NMR spectrum, signals corresponding to the same chemical species will increase and decrease proportionally as concentrations of those compounds increase or decrease. STOCSY takes advantage of this relationship and can reveal correlations in 1D NMR spectra between signals from the same compounds, but also from compounds which occur together.

When successful, STOCSY has the power to reveal correlations equivalent to those observed through space (like NOESY), from *J*-coupling (like COSY or TOCSY), between spin systems (like HMBC), between exchanging systems (like EXSY), and even from chemically distinct compounds with similar origins. The potential information gained from such an approach is significant, and it is investigated here.

1D STOCSY has been selected for use here due to its more straightforward interpretation. An example is shown in Figure 5.20.

Figure 5.20a shows the median NMR spectrum across the 148 samples in this dataset. Here, the 148 aligned and binned spectra were averaged by taking the median value for signal intensity/bin integral at each point. For calculation of the 1D STOCSY, a 'driver' peak must be selected. In this example, it is the region

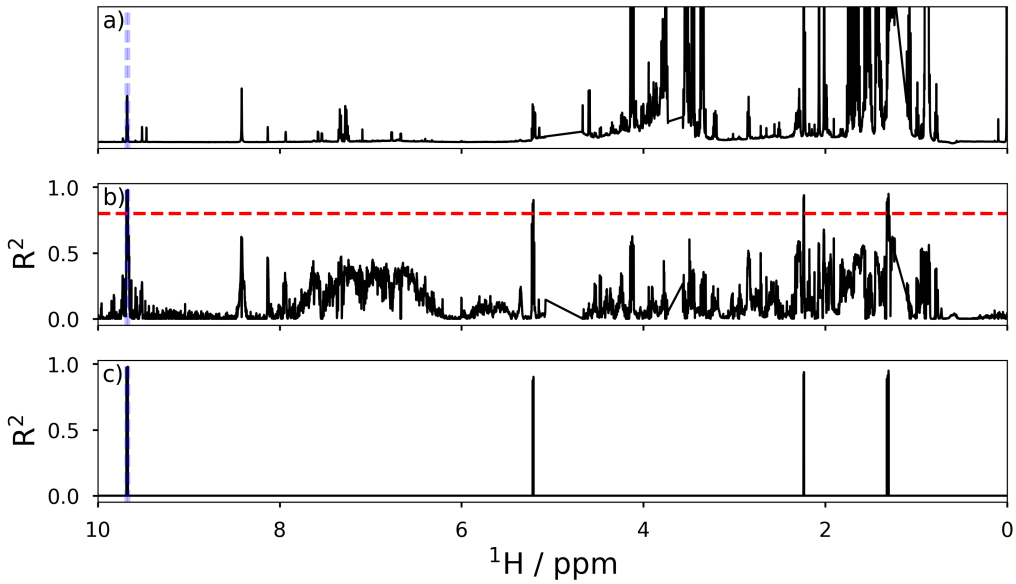


Figure 5.20: Example of STOCSY NMR analysis for acetaldehyde. a) shows the median NMR spectrum for 148 Scotch Whisky samples, b) shows the resulting 1D STOCSY spectrum with R^2 values for y-axis, and c) shows the peaks from b) where R^2 >0.8. Vertical dashed blue lines have been added to highlight the driver peak of acetaldehyde at 9.68 ppm, and a red horizontal dashed line in b) added to show $R^2 = 0.8$ threshold.

around 9.69 ppm corresponding to the aldehyde signal of acetaldehyde (**32**). In Figure 5.20b, the R^2 values (the square of the Pearson correlation coefficient) is shown as the y-axis against chemical shift, highlighting signals which correlate with those of the selected driver peak - acetaldehyde. It is clear that there are a handful of strongly correlating signals, and a range of low or medium correlating species in the spectra. To aid visualisation, a threshold R^2 value of 0.8 was set, below which all values were zeroed. The result of this is shown in Figure 5.20c, and highlights there are four highly correlated signals to the aldehyde of acetaldehyde.

Acetaldehyde was extensively discussed in subsection 5.4.3, and it is no surprise that the aldehyde correlates highly with both the methyl of acetaldehyde (2.2 ppm) but also two of the ethanol hemiacetal (**34**) signals. The other hemiacetals signals which would be expected based on previous assignment but which are not shown were either overlapping with the residual solvent signals (removed for statistical analysis), or overlapping with uncorrelated signals - dampening their correlation. It is evident that STOCSY is a technique with potential, but it is not infallible.

An alternative means to visualise the 1D STOCSY data is to colour code the median NMR spectrum with the associated R^2 values. This approach is shown in Figure 5.21.

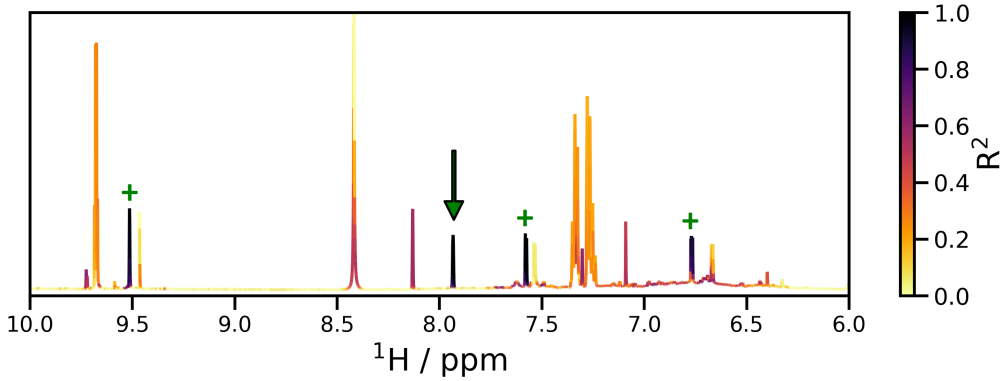


Figure 5.21: 1D STOCSY NMR spectrum for furfural. Spectrum calculated as the median NMR spectrum for 148 Scotch Whisky samples, with colour coding according to R^2 values for STOCSY driven by furfural at 7.93 ppm, as indicated by a green arrow; the strongest correlations are annotated with a '+' symbol.

This 1D NMR spectrum (Figure 5.21) has intensities representing the median values of the sample set, but is colour coded for correlation between the driver peaks and other peaks. With this colour map, the strongest correlations are nearly black, whilst the weakest are very pale. It is easily seen that in this example, using one of the aromatic protons on furfural (**39**) at 7.93 ppm as the driver signal, that it is strongly correlated with the other aromatic signals of furfural at 7.58 ppm and 6.77 ppm, and the aldehyde signal at 9.51 ppm, indicated with '+' symbols on the spectrum. Thus, the strongest correlations are with other protons in the same molecule. Weaker correlations are observed with other compounds, formic acid (**40**) at 8.5 ppm, ethyl formate (**41**) at 8.1 ppm, and unidentified signals at 7.1 and 9.7 ppm.

However, the advantage of colour coding correlation over using a R^2 threshold is that it is easier to visualise moderate correlations. Furfural is mainly associated with malt whisky production - including malting and the use of pot stills.[25] In this STOCSY spectrum, it is also seen to be correlated with signals of other cask extractives - including syringaldehyde (**42**) at 7.31 ppm, and also with 2-phenylethanol (**43**) at 7.26 and 7.34 ppm. 2-Phenylethanol is a higher alcohol produced during fermentation and associated more with malt whisky than grain, again due to the use of pot stills and distillation to a lower final ABV.[239]

Similarly, the colour coding allows easier visualisation of the weakest correlations. In Figure 5.21, one of the strongest correlations is to 7.58 ppm, to another aromatic proton of furfural. A very similar signal, also a doublet, is commonly

observed adjacent to this signal at 7.54 ppm, corresponding to 5-hydroxymethyl furfural (5-HMF) (**44**). Despite the structural similarity, 5-HMF concentration is highly dependent on the addition of caramel colouring.[25] As the addition of caramel will vary whisky to whisky, and be (largely) independent of malt production, it is reasonable that there is no correlation between the signals of furfural and 5-HMF.

STOCSY was also used for investigation of the higher alcohol signals. Refer to Figure 5.1 for example 1D ^{1}H spectra of several common higher alcohols. In Figure 5.22, the multiplet of 3-methylbutanol (**29**) at 1.66 ppm was used as the driver signal. Again, the strongest correlations are to signals from the same molecule - in this case the quartet at 1.43 ppm. The methyl signals in 3-methylbutanol are strongly correlating, but superimposed with other signals, at 0.88 ppm. The other higher alcohols in this region - such as iso-butanol (**25**) at 1.73 ppm - have medium correlations, which makes sense given their co-production during fermentation.

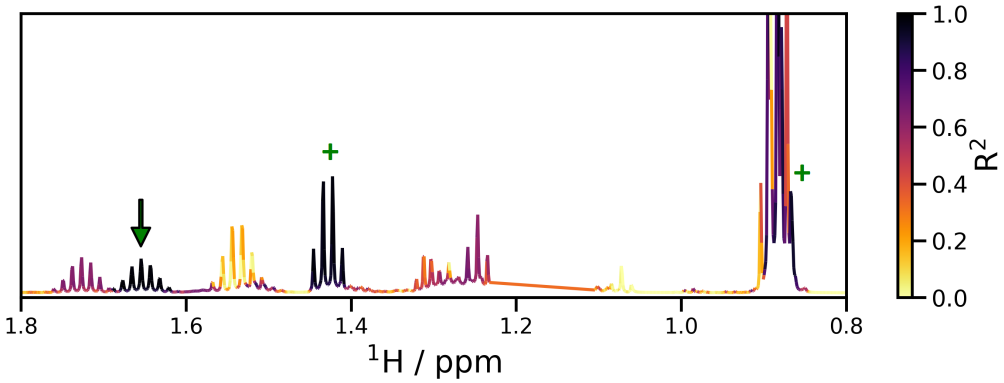


Figure 5.22: 1D STOCSY NMR spectrum for 3-methylbutanol. Spectrum calculated as the median NMR spectrum for 148 Scotch Whisky samples, with colour coding according to R^2 values for STOCSY driven by 3-methylbutanol at 1.66 ppm, as indicated by a green arrow; the strongest correlation is annotated with a '+' symbol.

Interestingly, however, Figure 5.22 shows that one of the higher alcohol signals has a very weak correlation with 3-methylbutanol; n-propanol (**26**), at 1.54 ppm, has an R^2 value close to 0. Whilst n-propanol is also a fermentation byproduct,[240] this lack of correlation is explainable. As grain whisky production results in lower levels of methyl butanols than malt production, correlation of these compounds in this mixed sample set is disturbed.[25] STOCSY analysis of only malt whisky samples may yield different results. Beneath the signals of n-propanol is a minor signal corresponding to 2-methylbutanol (**30**). This signal is largely obscured by n-propanol in the ^{1}H NMR spectra of Scotch Whisky.

An alternative technique to STOCSY is ratio analysis NMR spectroscopy (RANSY).[178] As the name suggests, RANSY uses the ratio of signal intensities to determine correlation, as opposed to STOCSY which just uses correlation coefficients. RANSY was investigated, but did not show any more information, or a higher quality of information than obtained from STOCSY, and is not shown here.

5.6.3 *ICA*

Statistical methods exist to decompose a mixture of signals based on multiple measurements of different mixtures. ICA is a method to separate out additive signals in a multivariate spectrum, and so is well suited to the problem of mixtures in NMR, and has previously been shown to have potential application in separating out compounds in mixtures by analysing NMR spectra.[186,241] The primary advantage of ICA is in separating out signals in an intuitive fashion. ICA is explained in the introduction. An example of ICA for Scotch Whisky NMR data is shown in Figure 5.23.

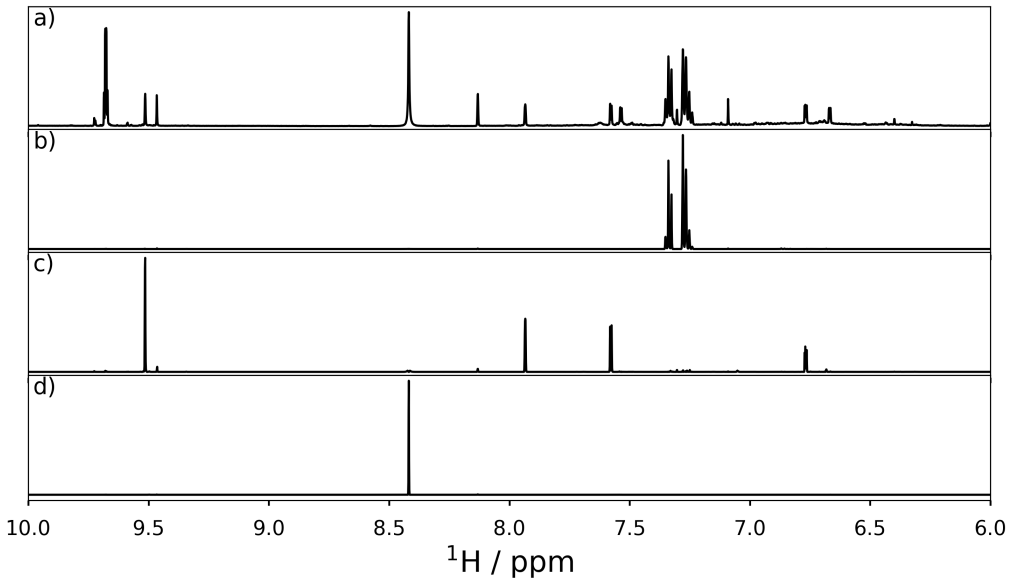


Figure 5.23: ICA applied to NMR spectra of Scotch Whisky, constructed with six components. a) shows median NMR spectrum for 148 samples, and ICA components b) 1, c) 3, and d) 4. Intensity values for components have been squared to aid visualisation.

In Figure 5.23, an ICA model has been constructed for six components using the binned spectra of 148 samples between 10 and 6 ppm. The median NMR spectrum is again shown in Figure 5.23a, whilst Figure 5.23b-d show ICA components 1, 3 and 4. These components were selected to highlight the more successful outcome of the ICA. The number of components is chosen before the model is con-

structed, and affects the outcome. Trial-and-error determined six components provided acceptable results, but further experimentation is necessary. ICA can theoretically decompose a signal into n-1 components, where n is the number of measurements (spectra, here).

Component 1, shown in Figure 5.23b, has managed to separate out, very cleanly, the signals from 2-phenylethanol (**43**) (a higher alcohol) from all the other signals. Similarly, Figure 5.23c shows component 3 separating out the signals of furfural (**39**) from other compounds, and Figure 5.23d shows that component 4 has resolved formic acid (**40**) from all other signals in this region.

ICA is shown to be reasonably successful - provided the appropriate number of components is selected. Experimentation with other spectral regions and differing numbers of components provided mixed results. ICA is biased towards parameter selection, and therefore will be potentially difficult to use when the compounds in the mixture are completely unknown. At this point, most of the major compounds in the ^{1}H NMR spectra of Scotch Whisky have been identified - or at least correlated - and ICA serves limited further benefit here.

5.6.4 *Classification Analysis*

Chemometrics is an area of statistics which is concerned with applying statistical methods to chemical data to better understand the large volumes of data which can readily be generated these days. It is widely used in areas including metabolomics, and most commonly involves methods such as PCA and OPLS-DA, explained in the introduction.

As discussed in section 5.3, the primary sample set for chemometric investigation were the 'standard sample sets' of Scotch Whisky from 2010, 2012, and 2014. Whilst the samples are anonymised, it is known if they are a blend (Blended Whisky) or a malt (both Single Malt and Blended Malt), and their alcohol strength, along with, for some samples: age statements, geographical origins, maturation wood types, and if the malt was peated or not.

5.6.4.1 *PCA*

Initial classification analysis involved applying PCA to the ^{1}H NMR spectra acquired of the standard sample set (n=148). PCA performed well with R^2 values of 0.53 (PC1), 0.16 (PC2), and 0.05 (PC3). A total of 17 components were calculated by the software SIMCA, using its own cross-validated optimisation methodology, with a cumulative R^2 of 0.95. An advantage of PCA over ICA is that the number of components is less important. That is, in PCA the first compon-

ent describes the maximum variation in the data, the second is orthogonal and describes the next largest variation, etc. PCA components are deterministic in a sense. ICA, however, decomposes data into a pre-defined number of components - therefore component 1 will be different if there are three total components or ten total components. Additionally, an optimal number of principal components can be determined for PCA - *i.e.* via a scree plot - whereas for ICA no equivalent metric exists. Statistics of all PCA and OPLS-DA models are summarised in Table 5.3.

PCA is an unsupervised technique, using just the measured data (NMR spectra) as a blind input. Strictly speaking, it is not a classification technique. However, visualisation of the scores - typically by a scatter plot of two components - can be aided by encoding a known classification with scatter plot marker colours or shapes.

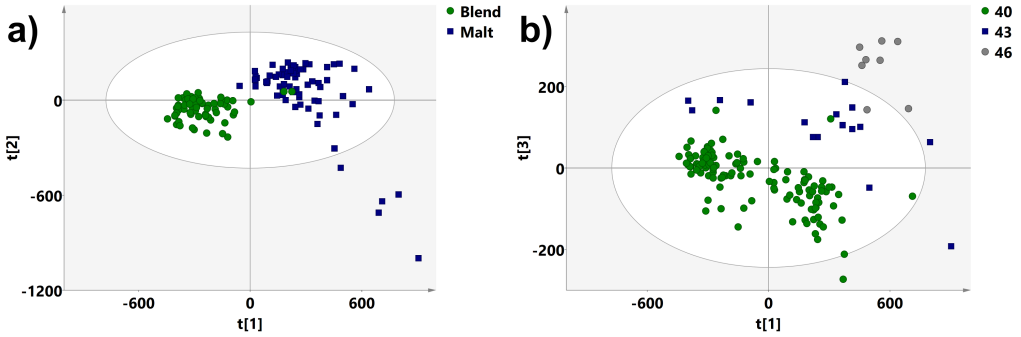


Figure 5.24: Chemometric results for modeling the aligned NMR data PCA scores plots for PC1 v PC2 shown in a) colour coded for blend or malt whisky, b) PC1 v PC3 colour coded for label alcohol strength (ABV %).

Colour coding of sample type - blend or malt - is shown in the scores plot of Figure 5.24a. Here, a clear separation emerged in PC1 with only a couple of 'misplaced' samples. These two blends, overlapping with malt samples, (S14-2858 and S13-0110) correspond to premium Blended Scotch Whisky samples with a high proportion of malt in their blend. The outliers to the bottom right are whiskies matured in Bourbon and Sherry casks (one was matured in both Bourbon casks and Port pipes - BP), and are - S10-1313, S12-2388, S13-0090, S14-1964, S14-1972(BP). As Sherry and Port casks can introduce significantly more maturation related character than Bourbon alone, these samples may be outliers for this reason. Inspection of their spectra revealed no obvious deficits in spectra attributable to poor shimming or solvent suppression, hence technical reasons could be excluded.

The loadings plot for PC1 is shown in Figure 5.25a and has a global positive trend and a similar appearance to that of the typical NMR spectrum of a

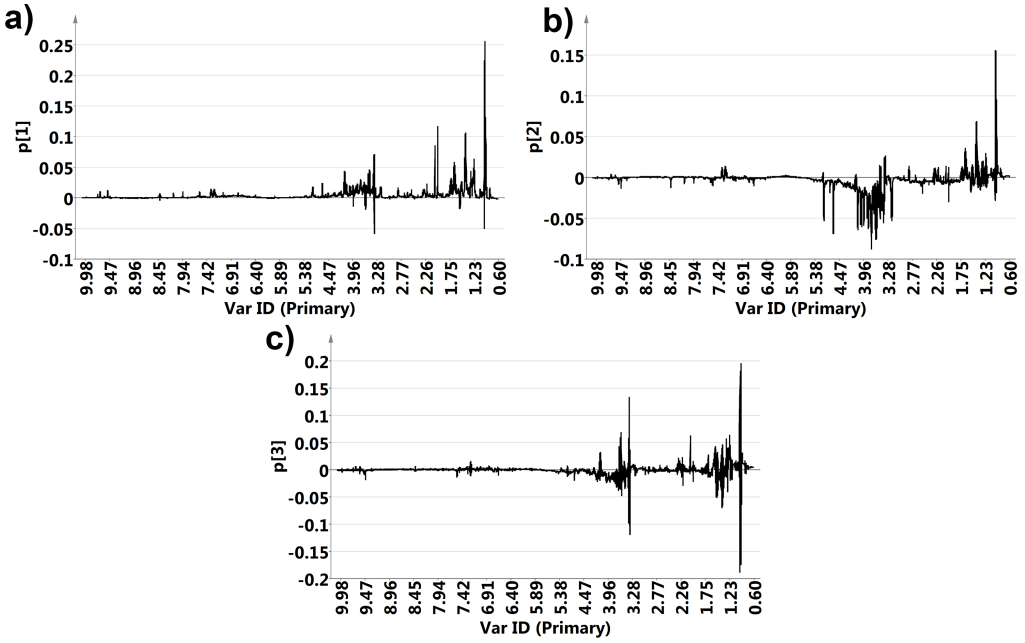


Figure 5.25: Loadings plots for PCA of NMR data based on the standard set of 148 sampleshown are principal components a) 1, b) 2, and c) 3.

Scotch Whisky. Based on this plot, it is likely that the more complex, more 'mature' whisky samples are placed further to the right on PC1. Given that malts have a more complex new make spirit, and are often more premium products with more maturation related character, this fits with expectations. The loadings for PC2 in Figure 5.25b show a negative contribution around the carbohydrate region. Analysis of the spectra confirmed that the outliers in Figure 5.25a are outliers because of their higher carbohydrate content.

Interestingly, PC3 can largely separate out samples based on their alcohol strength (40 % to 46 %) (Figure 5.24c). These bottle strengths are of a very limited range, so it is unclear how this model would translate to cask strength samples (>50 %). As peak alignment and binning of spectra was performed, the discrimination may be due to relative changes in concentrations of compounds which exist in equilibrium of forms - *i.e.* α- or β-D-glucopyranose, or acetaldehyde in its hemiacetal forms (see subsection 5.4.2 and subsection 5.4.3). The largest loadings in PC3 correspond to chemical shifts near the residual ethanol solvent signals. Despite the removal of the ethanol signals from the processed spectrum, clearly some residual distortion may remain.

5.6.4.2 *OPLS-DA*

Supervised modelling techniques - such as OPLS-DA - allow for more targeted investigation of specific questions. Blend versus malt classification was examined

more closely with OPLS-DA using the same dataset. Again, SIMCA calculates an optimal number of components via a cross-validated method. This model was constructed with 1 predictive and 3 orthogonal components providing an $R^2\mathbf{X}$ of 0.76, $R^2\mathbf{Y}$ of 0.94. The scores plot for this model is shown in Figure 5.26.

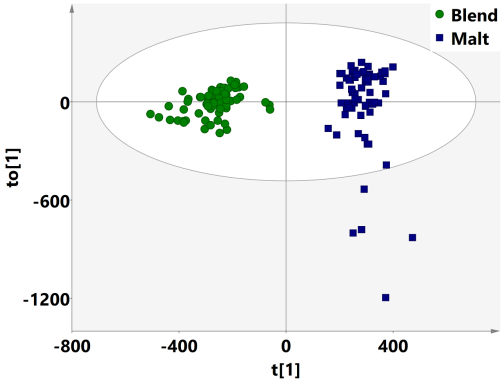


Figure 5.26: NMR based OPLS-DA scores plot for blend versus malt.

Figure 5.26 shows apparent perfect separation of the two classes along component 1, and the high Q^2 value of 0.92 indicates a model with a high degree of accuracy. Model statistics are summarised in Table 5.3. The outliers in OPLS-DA scores plot are the same as in the first PCA scores plot, as the y-axis here represents the intra-class variability (*i.e.* principal components of the residual data).

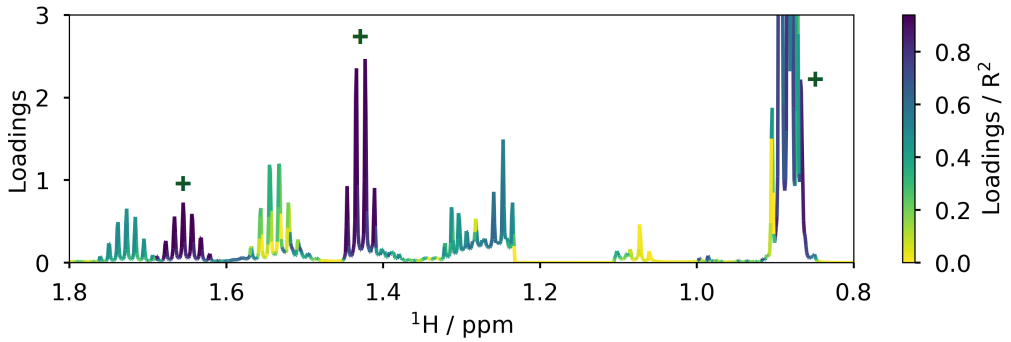


Figure 5.27: S-line plot for blend versus malt OPLS-DA model from NMR data between 1.8 and 0.8 ppm. Y-axis shows pseudo 1D NMR spectrum, line colour represent R^2 of variable to class discrimination - darker signals are more prevalent in malt whiskies. 3-methylbutanol signals are labelled with a '+'.

Like PCA, OPLS-DA can be used to visualise which variables are key in generating the model, *i.e.* which input variables are most important for class separation. Here, the S-line plot (expansion around key region shown in Figure 5.27) indicates which signals are key in discriminating blended whiskies from malt whiskies. This analysis showed that the signals belonging to the higher alcohols

are most significant as indicated by the correlation coefficient, with key importance ascribed to the signals of 3-methylbutanol.

OPLS-DA was also used to examine maturation wood type. By including only those samples with known maturation and finish wood types, and samples where more than one example of the style were present, the sample set was reduced to three classes - Bourbon (B) only (n=26), Sherry (S) only (n=10), and Bourbon and Sherry (BS) maturation (n=21) for a total of 57 samples. The OPLS-DA model generated is shown in Figure 5.28a.

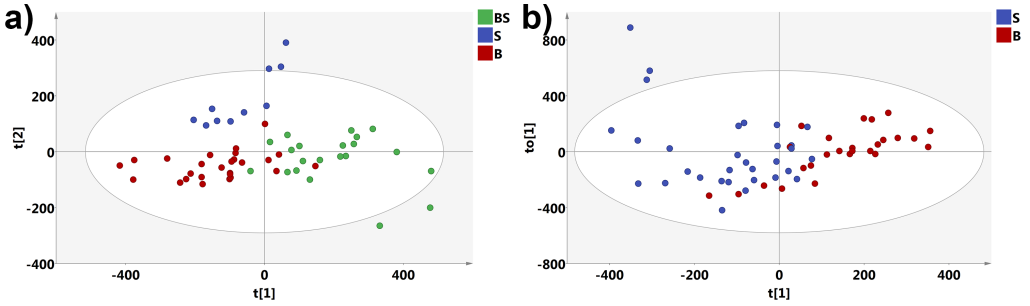


Figure 5.28: OPLS-DA Scores plots for maturation wood type with a) three class (B, S and BS), and b) two classes (B and S, where BS is now labelled S).

The constructed model had an $R^2\mathbf{X}$ of 0.71, $R^2\mathbf{Y}$ of 0.59. The scores plot, shown in Figure 5.28a, shows some degree of separation between the three classes, however, the model is not perfect in separating them out. OPLS-DA is known to perform less well for more than 2 classes, so validation was especially important here, especially as the Q^2 value was low at only 0.36. Permutation testing of classification of the BS class shows that the model is valid, as shown in Figure 5.29.

The two criteria for validity are that a) the Q^2 values for the permutations (left of each plot) are lower than the original (right of each plot), and/or b) that the Q^2 y-axis intersection is below zero. Both of these criteria are met as shown in Figure 5.29, and therefore the model shows a good predictability. Furthermore, CV-ANOVA can be used to calculate the statistical significance of the model. Here, the *p*-value is <0.05, giving a high confidence in the model.

Not only do OPLS-DA models perform poorly for greater than two classes, S-line plots cannot be generated for more than two class models. In this case, the interpretation is further hindered as the three classes are not entirely distinct – BS should show similarity and overlap with B and S, in varying amounts depending of ratio of Bourbon to Sherry character in the spirit. As such, an alternative model was generated to compare Bourbon and Sherry maturation. As only 10 samples were exclusively matured in Sherry casks, and the question of interest is about Sherry or Bourbon, those whiskies matured in both types of cask (BS) were relabelled as Sherry maturation (S). As such, the model becomes an indicator for

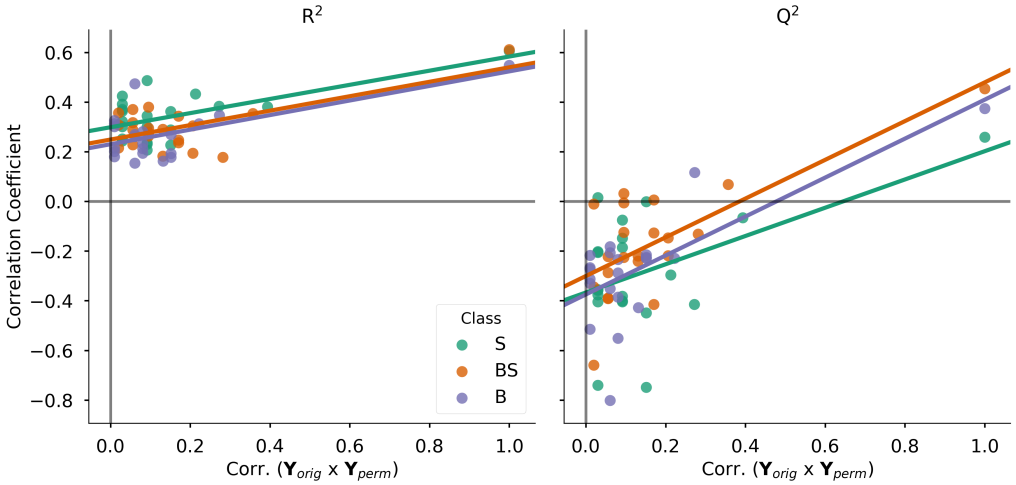


Figure 5.29: OPLS-DA permutation testing for NMR maturation wood type model. **Y**-vector permutation repeated 20-times with calculated R^2 values in the left panel and Q^2 values in the right panel. The x-axis shows the correlation between the original **Y**-vector ($\mathbf{Y}_{orig}$), and permuted **Y**-vector ($\mathbf{Y}_{perm}$). Calculations shown for all three **Y** variables - B = bourbon, S = sherry, BS = bourbon and sherry.

Sherry character, and should separate out Bourbon features. The resulting model had an $R^2\mathbf{X}$ of 0.52, $R^2\mathbf{Y}$ of 0.50, and Q^2 of 0.39, representing some loss in R^2 but a slight improvement in Q^2 compared to the three-class model. The scores plot is shown in Figure 5.28b, with general separation, but also some overlapping samples.

Examination of the loadings indicated a number of moderately correlated signals relating to this class separation, including many of the signals between 6 and 10 ppm, with the noticeable exception of 2-phenylethanol, which is uncorrelated to the class separation. 2-Phenylethanol is a fermentation by-product rather than maturation related congener, and it is reasonable for it not to be correlated. The other compounds in this region are primarily maturation related - including syringaldehyde and acetaldehyde – and it is logical that they would be more abundant in sherry cask matured whiskies. These are often deliberately styled to have a more mature characteristic. Additionally, there is moderate correlation between the carbohydrate region and sherry maturation – this may come from the wood, or more likely from carry over of the sweeter sherry wines compared to the dry (no, or very little, residual sugar) bourbon barrels.

Beyond maturation wood types and blend or malt classification, several other parameters were modelled with varying degrees of success. The parameters and statistics of all models are summarised in Table 5.3. In this table, 'PCA' is the original PCA model, whilst the rest are OPLS-DA models. 'Blend/Malt' corresponds to the blend versus malt model, discussed previously. 'Wood (4)' is a model for

classification based on four maturation wood types - B, S, BS, and the ale cask finished whiskies. 'Wood (3)' and 'Wood (2)' correspond to the maturation wood type models for B, S, and BS (Wood (3)), as discussed above. Of the wood models, the fewest classes (Wood 2) performs best when evaluated by CV-ANOVA in terms of F-statistic and *p*-value, although all three had *p*-values below 0.05, indicating statistical significance.

Table 5.3: Summary of statistics for chemometric models built using 1D ^{1}H NMR data. All models (bar PCA) were built with OPLS-DA. Components describes the number of principal components (PCA) or predictive and orthogonal components (p+o, OPLS-DA). *n* is the number of observations (samples). F (F-statistic) and *p* (*p*-value) were calculated from CV-ANOVA.

MODEL	COMPONENTS	*N*	$R^2\mathbf{X}$	$R^2\mathbf{Y}$	Q^2	F	*P*
PCA	17	148	0.95				
PCA	3	148	0.74				
Blend/Malt	1+3	148	0.76	0.94	0.92	197	<0.05
Peat	1+4	148	0.72	0.85	0.74	38.7	<0.05
Wood (4)	3+6	60	0.87	0.81	0.54	2.30	<0.05
Wood (3)	2+2	57	0.71	0.59	0.36	3.30	<0.05
Wood (2)	1+1	57	0.52	0.50	0.39	8.38	<0.05
Age	3+0	69	0.49	0.26	0.10	1.20	0.244
Origin	2+4	69	0.78	0.48	0.25	1.16	0.263

In Table 5.3, the 'peat' model compared the known peated samples and all other samples. Originally this was modelled using the whole spectrum, but this model was improved (Q^2 increased from 0.66 to 0.74, F-statistic increased from 15.9 to 38.7) by only including the regions 6-10 ppm of the spectra in the model. Both the original and revised model were deemed statistically significant ($p < 0.05$). Investigation of the S-line plot highlighted regions of high correlation and thus potential interest. These were primarily around 6.85 and 7.1 ppm, and the most significant region is presented in Figure 5.30.

Interestingly, the important features highlighted in Figure 5.30 do not correspond to the signals of compounds identified in this investigation thus far. Peated whiskies are known to be distinguished by their phenolic aromas, however the typical concentration of these phenol compounds is in the low ppm (or $\mu g\, L^{-1}$), which is at, or below, the NMR limit of detection here. However, reference NMR spectra have been acquired for a number of typical peated whisky phenolic compounds. Relevant regions of their ^{1}H NMR spectra are shown in Figure 5.31. It

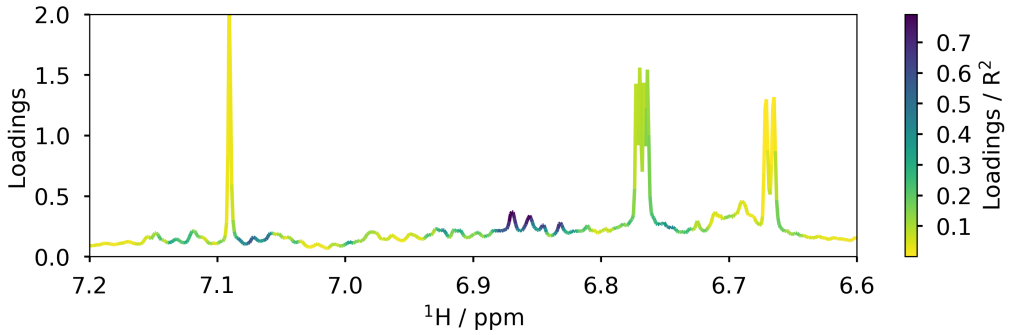


Figure 5.30: S-line plot for peated OPLS-DA model from NMR data between 7.2 and 6.6 ppm. Y-axis shows pseudo 1D NMR spectrum, line colour represent R^2 of variable to class discrimination - darker signals are more prevalent in peated whiskies.

is not possible to assign a specific phenolic compound, or compounds, to the signals in the S-line plot which discriminate peated whiskies. However, the similarities in chemical shifts between the reference compounds and the S-line plot are highlighted by the emulation of signals shown in Figure 5.31. This indicates that the OPLS-DA model discriminates between peated and non-peated whiskies because of phenolic compounds despite the NMR sensitivity limitations.

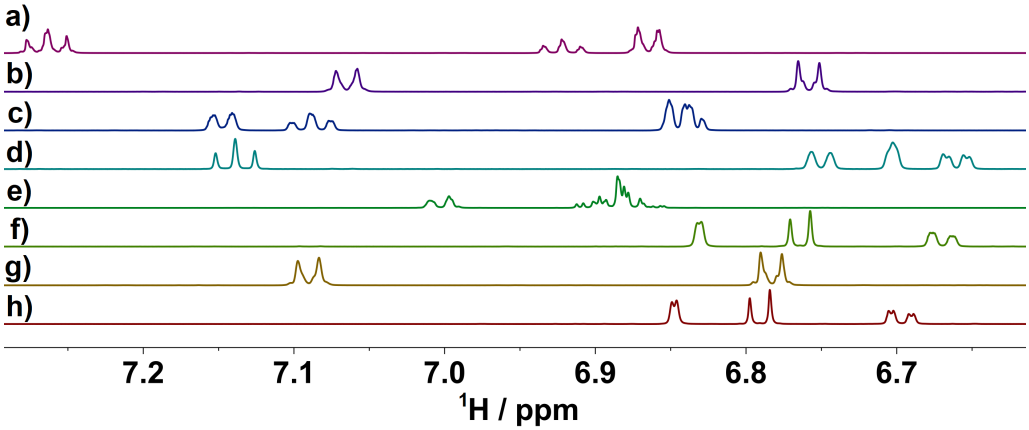


Figure 5.31: Partial stacked 1D ^{1}H NMR spectra of reference phenolic compounds: a) 4-ethylguaiacol (**11**), b) 4-ethylphenol (**10**), c) 4-methylguaiacol (**9**), d) guaiacol (**4**), e) *m*-cresol (**6**), f) *o*-cresol (**8**), g) *p*-cresol (**7**), h) phenol (**4**). Reference samples prepared in 40 % ABV solution and NMR spectra acquired as per routine Scotch Whisky method at 600 MHz.

Two further OPLS-DA models are summarised (Table 5.3) which were not deemed to be statistically significant. These are 'age' and geographical 'origin'. Both had low Q^2 values and CV-ANOVA *p*-values which were much greater than the significance threshold (0.05), indicating neither model was accurate. This is not overly surprising. Age statements simply indicate the youngest whisky in the

bottle - older components may be blended in with younger for product consistency. Furthermore, age is but one variable and cask activity is far more significant - three years in a first fill cask will produce a more mature whisky than ten years in an exhausted cask. Regionality in Scotch Whisky simply indicates where it is produced. However, Scotch Whisky does not have *terroir* in the same sense as wine - the barley does not have to be grown locally to the distillery (it rarely is), neither does the peat (if used) have to be sourced locally. Most significantly, perhaps, the barrels used for maturation have, almost exclusively, been previously used in another country - and these are used across Scotland without a regional specialisation. Put simply, there is no reason for it to be possible to model geographic origin of Scotch Whisky by NMR.

As discussed before, the chemometric analysis could, in principle, have been performed on quantified compounds. However, there is a non-trivial difficulty in accurately, and ideally automatically, quantifying of a wide range of compounds from which to build a data matrix. Furthermore, the handful of well resolved and identified multiplets which may be fitted and quantified most reliably are for compounds which are already well known, and characterisable by routine methods (such as GC-MS). They are also the easiest signals to bin and interpret in the NMR spectra. Quantifying compounds would also lose all the minor signals which may be significant, but in the quantification regime would be either unresolved or unidentified/unidentifiable. Nevertheless, quantification of these compounds in Scotch Whisky by NMR is an important potential area of future work. It is also possible that a mixed approach based on both quantified contribution of identified compounds and non-specific binning of other regions of spectra could be successful.

5.7 CONCLUSIONS

NMR spectroscopy provides an unparalleled, quantitative, insight into the diverse chemistry of Scotch Whisky. Acquisition of the solvent suppressed 1D ^{1}H NMR spectra is routine and fast, and the suite of complementary 1D and 2D NMR experiments allows for the structural characterisation of dozens of compounds in Scotch Whisky.

Many compounds in Scotch Whisky were identifiable by comparison to reference and spike-in NMR spectra, supported by homo- and heterocorrelated NMR experiments where possible. For some chemicals, which exist in multiple forms in solution, more investigation was required. In this chapter, the characterisation of the α and β tautomers of glucopyranose, fructopyranose, and fructofuranose was successful. Two hemiacetal forms of acetaldehyde were identified by means

of several NMR experiments, including DOSY, whilst other, less significant, acetaldehyde equilibrium compounds were observed but remain uncharacterised.

Quantification of ethanol was successful, especially by use of ^{13}C NMR. This quantification yielded high accuracy based either on integration of signals or regression analysis of chemical shifts. The accuracy was lower for real samples than on model solutions. Interestingly, the chemical shift difference between the two carbon signals proved to be the most accurate means of quantification in these samples. Quantification of other compounds in Scotch Whisky was hindered by variable signal overlap and peak position, current software limitations, and the unknown ratios of structural isomers in solution.

Statistical analysis of the NMR spectra of Scotch Whisky provided insight into chemical identification, origins, and sample discrimination. Binning of the spectra was performed uniformly with a fine bin width, allowing for retention of ^{1}H multiplicities whilst reducing the data matrix size considerably. STOCSY was utilised to correlate signals from the same compound or compounds with similar origins, whilst ICA decomposed the NMR spectra into compound spectra with some success. Classification of the standard sample set of Scotch Whiskies was performed with PCA and OPLS-DA in terms of: blend versus malt, maturation wood types, alcohol strength, age statement, geographical origin, and peated status. The statistical models were interrogated for validity and to understand the chemical reasoning behind the outcomes of these models.

6

ESI FTICR MS ANALYSIS OF SCOTCH WHISKY

There are known knowns...
there are known unknowns...
But there are also unknown unknowns...

— Donald Rumsfeld (2002)

6.1 DECLARATION

Most of the work in this chapter was previously published in the following paper:

> W. Kew, I. Goodall, D. J. Clarke, and D. Uhrín, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 200-213

6.2 INTRODUCTION

Where NMR provides quantitative structural information of components of a mixture, it suffers from a poor resolution and sensitivity when compared to FTICR MS. The resolving power and sensitivity of FTICR MS is unparalled, and it is the golden standard when it comes to molecular level characterisation of complex mixtures. Whilst chromatography can be coupled to FTICR MS, FTICR MS is capable of such high resolving power that often this is not necessary and direct infusion can be sufficient. Two ionisation sources are most commonly coupled to FTICR MS for complex mixture analysis, ESI and APPI, though the latter is almost exclusively utilised in petroleomics studies. ESI has been the *de facto* standard ionisation technique for complex mixture analysis by FTICR MS, including NOM, biofluids, and food products, for nearly twenty years. ESI has previously been used for analysis of whisky by Q-TOF in 2005,[52] and FTICR MS in 2013.[54] These studies examined questions of authenticity and complexity, but suffered from small and irregular sample sets, and lack of data analysis tools - for example, chemometrics in Garcia *et al.* was performed on just the 50 most abundant ions, excluding the vast majority of the data acquired. More recently, and after the work in this chapter was originally published, further work has been conducted analysing the complexity of 'barrel-aged' whisky by ESI FTICR MS.[61]

This chapter investigates the chemical complexity of Scotch Whisky, including new make spirits and mature final products, by ESI-(-) FTICR MS. The mass

spectra and their complexity were examined with formula assignment rates of 72 % to 88 % across 85 mature Scotch Whisky samples. Select molecular formula assignments were confirmed by IFS, and structural information was elucidated by means of tandem-MS experiments both in the source and quadrupole, and by means of in-cell isolation and SORI-CID. Further visualisation of the data by means of van Krevelen and DBE plots revealed chemical space regions of interest, and chemometric analysis highlighted potential compounds of interest for discrimination of maturation wood types.

6.3 SAMPLE SET

As with chapter 5, this chapter deals with the untargetted analysis of a diverse sample set of final product Scotch Whiskies. A subset of the same sample set was used, however, due to experimental time and manual data acquisition, overall 85 samples were used. These encompass all of the 2014 samples (n=52), and the Single Malts from 2010 and 2012, as well as one blend from 2010 and 2012 to examine product consistency. As in chapter 5, Single Malt and Blended Malt whiskies are classified here as 'malt', whilst blended whiskies are 'blend'. There remain no Single or Blended grain whiskies in this sample set. A malt NMS was also examined for comparison. See Appendix B for a list of samples used in this chapter.

6.4 FORMULA ASSIGNMENT

In this chapter, exact masses were assigned chemical formula using PetroOrg. Formulae were limited to $C_{0\text{-}100}$ $H_{0\text{-}200}$ $O_{1\text{-}20}$ $S_{0\text{-}1}$ and a 1 ppm error threshold. Only singly charged deprotonated formulae were assigned, as consistent with similar work using ESI in the literature.[54,242]

Elemental limits for assignment were tested iteratively by including higher O and S limits and with broader mass error thresholds. However, no confident assignments could be made for a species with a heteroatomic class larger than O_{19} or $O_{12}S$ for any sample in this study. See section 7.8 for further discussion on formula assignment and software.

6.5 MASS SPECTRA

For this large study, it was feasible to use only one ionisation source and polarity. Negative mode ESI was used primarily for its ubiquity in the literature and, ulti-

mately, because it was the only methodology established within the department at that time. Naturally, these ESI spectra do not represent the full chemical composition of each sample. In chapter 7, four different ionisation sources are utilised to examine a reduced sample set of Scotch Whiskies, and their differences discussed further. Similarly, the use of positive mode ionisation is also discussed there.

Technical details of acquisition are provided in section 3.2. In short, samples were diluted 1:10 in a 50:50 methanol water solution prior to direct infusion. Spectra were acquired into a 4 MW time domain between m/z 98.3 and 1000 with a transient length of 1.1185 s; 200 transients were co-added prior to zero filling, apodisation and Fourier transformation. This yielded a resolving power of approximately 300,000 at m/z 400. Spectra were internally calibrated against known and expected species in Scotch Whisky, along with homologous series of compounds. Calibration lists are included in Appendix C.

6.5.1 *One Sample*

Figure 6.1 shows the ESI spectra for a typical sample, S14-2373 - a 10-year-old Sherry cask matured Highland Single Malt Scotch Whisky. This sample is a complex mixture, with 3325 peaks picked (SNR > 4). Internal calibration based on known and confirmed chemical species achieved a standard deviation of 217 ppb up to m/z 577. Figure 6.1a shows the full intensity broadband spectrum from m/z 100 to 700, whilst Figure 6.1b scales the y-axis to reveal the significant underlying complexity.

As with all complex mixtures, any given nominal mass contains many distinct signals. This spectrum is complex and at m/z 397 there are at least a dozen signals observed (Figure 6.1c). Here, 12 peaks were detected and assigned unambiguous monoisotopic formulae within a 1 ppm error threshold. The assignments in this panel are summarised in Table 6.1.

Of the 3325 peaks observed in this spectrum, 1902 were assigned monoisotopic formulae with a median mass error of 35 ppb. A further 571 peaks were assigned as $^{13}C_1$ isotopologues, median error 67 ppb. As stated above, in this work, formula assignment allowed for up to one sulfur atom per molecule. The presence of sulfur in some of these species was confirmed by IFS later in subsection 6.5.3. Despite inclusion, sulfur remains a minor component of this sample - in this sample, 1813 of the 1902 assigned monoisotopic peaks were CHO species, with only 89 being CHOS.

In total, over 74 % of the peaks could be assigned in this sample. This was a significant result when contemporary efforts on complex mixture analysis often

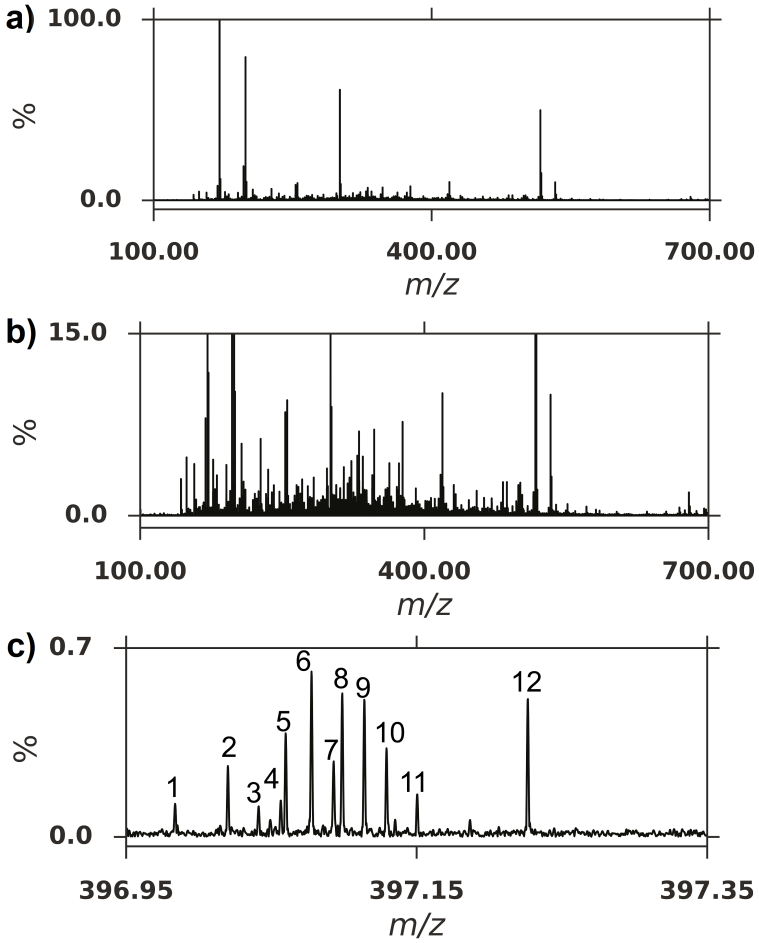


Figure 6.1: ESI mass spectrum of sample S14-2373 showing a) broadband mass spectrum between m/z 100 and 700, b) broadband mass spectrum between m/z 100 and 700 zoomed in to 15 % relative abundance, c) 12 peaks of the zoomed in region of m/z 396.95 - 397.35; labelled assignments are detailed in Table 6.1

struggled with molecular formula assignment. Garcia *et al.* presented only 20 assignments per spectrum,[54] and the ability here to confidently assign thousands of molecular formulae across not just one, but dozens of mass spectra is a strong result.

6.5.2 *Sample Set*

Across the entire sample set studied (n=85), 4271 unique monoisotopic formulae were assigned; 82 % were CHO and 18 % CHOS. The number of unique chemical formulae assigned increases approximately logarithmically with the number of new samples, as shown in Figure 6.2.

Table 6.1: Select ESI MS formula assignments of Scotch Whisky m/z 397. See Figure 6.1 for reference. Measured m/z values shown for the $[M\text{-}H]^-$ ion. *Neutral molecular formula are given.

#	m/z	FORMULA*	ERROR (ppb)
1	396.9837595	$C_{18}H_6O_{11}$	63
2	397.0201095	$C_{19}H_{10}O_{10}$	27
3	397.0412325	$C_{16}H_{14}O_{12}$	43
4	397.0564975	$C_{20}H_{14}O_9$	20
5	397.0598919	$C_{17}H_{18}O_9S$	38
6	397.0776630	$C_{17}H_{18}O_{11}$	71
7	397.0928843	$C_{21}H_{18}O_8$	17
8	397.0987888	$C_{14}H_{22}O_{13}$	62
9	397.1140314	$C_{18}H_{22}O_{10}$	28
10	397.1292975	$C_{22}H_{22}O_7$	53
11	397.1504005	$C_{19}H_{26}O_9$	14
12	397.2265544	$C_{18}H_{38}O_7S$	16

Here, the number of unique formulae for each additional spectrum was calculated (with 100 permutations of sample order to remove potential bias), and the average number of unique formulae identified against sample number plotted. A log linear regression was fit to the data, demonstrating the diminishing returns with increasingly large sample sets. Similar observations have been made for DOM in lakes,[243] where the number of new formulae decreased significantly above approximately 50 samples. In that study, all the samples were Swedish Lakes - here, all the samples are Scotch Whisky. It is highly probable that the di-

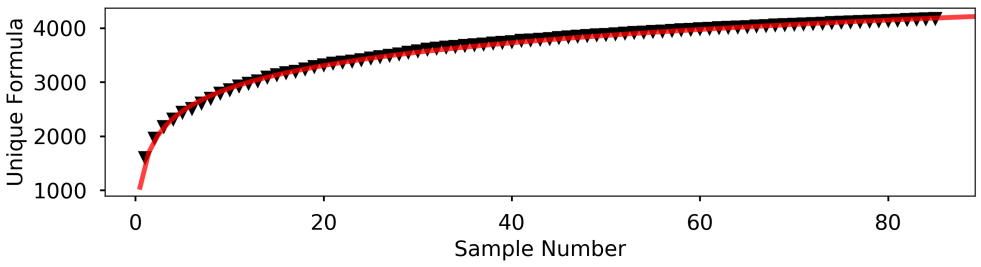


Figure 6.2: Number of unique molecular formulae assigned against increasing sample set size. Values plotted represent the average of 100 permutations of sample set order. Log fit of the data is plotted in red.

versity of formulae would be greater for a more diverse sample set - *i.e.* whiskies from around the world.

This demonstrates the need for a larger sample set to reflect the chemical diversity of Scotch Whisky, and at the same time suggests that the sample size used here, 85 samples, was sufficiently large. On average, each sample had 1647 monoisotopic formula assignments. 90 % of the unique assignments could be obtained from ≈50 % of the sample set.

Only 407 formula were common to all samples, and 1201 formula were present in over 75 % of samples. In total, 44 samples identified sulfur containing compounds. Whilst CHO compounds and sulfur containing low molecular weight congeners have been identified by other techniques in Scotch Whisky,[16,21,27] a larger CHOS molecule, such as that assigned to peak 12 in Table 6.1 ($C_{18}H_{38}O_7S$) has not been identified previously.

6.5.3 *Isotopic Fine Structure*

Formula assignment based solely on exact mass accuracy runs the risk of error, especially as more heteroatoms are allowed. To confirm the assignment of not only CHOS compounds, but also several CHO compounds - and thus validate the assignments generated - isotopic fine structure (IFS) analysis was utilised. As FTICR MS has such high sensitivity and resolution, it is quite possible to readily observe isotopologues of major ions in a broadband spectrum, especially the first ^{13}C isotopologue. However, true fine structure, including much less abundant isotopes such as ^{18}O, ^{34}S, and multiple ^{13}C substitutions requires greater sensitivity and resolution than routine broadband acquisition affords.

Targetted IFS is thus much more amenable when select regions of the spectrum are isolated - using the quadrupole mass filter and increasing the IAT. Additionally, the concentration of the sample could be increased - but this may leave residual contamination in the instrument and is best avoided where possible. By filtering out all but a narrow mass range, and increasing the IAT, the effective concentration of only a specific mass range of the sample is increased in the analyser. This allows for increased sensitivity. Furthermore, selective mass acquisition appears to allow for a more stable ion cloud in the ICR cell, allowing transient length to be increased substantially without signal degradation, such as peak splitting. This increases the resolution of the acquired spectrum.

IFS has previously been used to confirm formula assignments,[244–246] including the use of selectively isolating ions of interest.[247] The assignment of $[C_{13}H_{19}O_8S]^-$ (m/z 335.08062, -0.02 ppm) is confirmed by IFS as shown in Figure 6.3 in the S14-2373 sample.

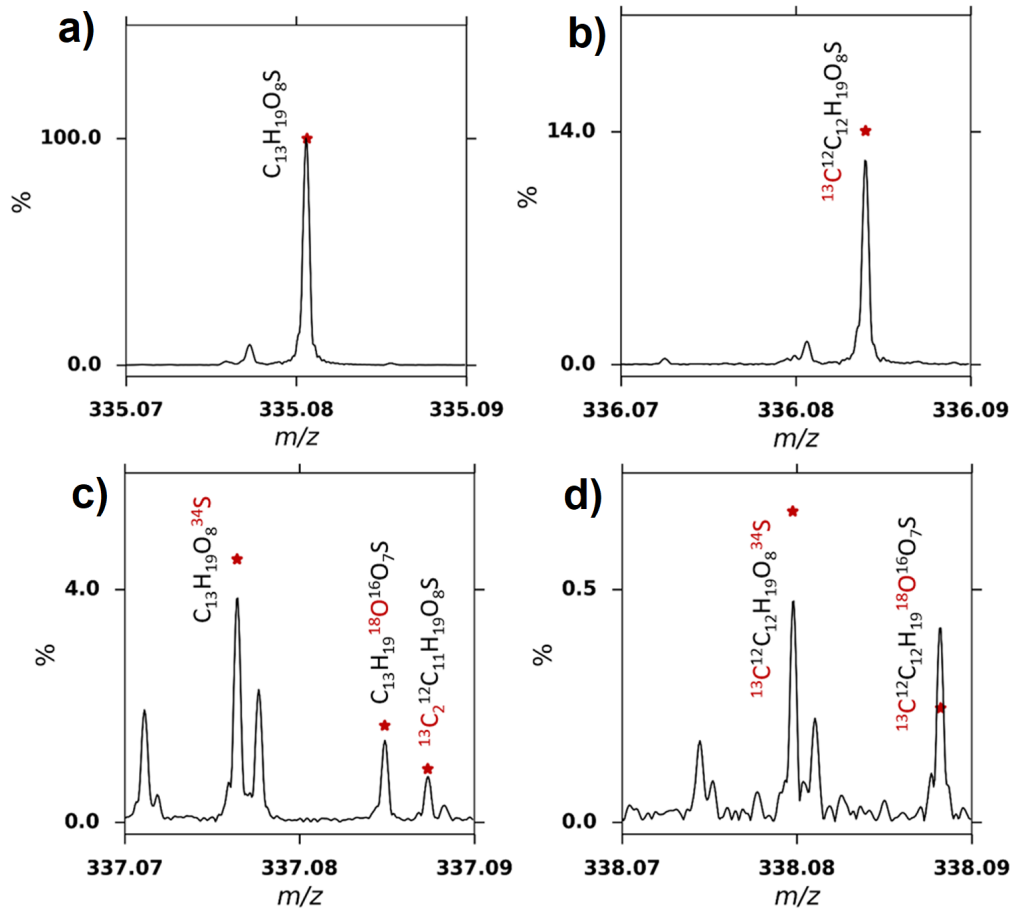


Figure 6.3: IFS analysis of $[C_{13}H_{19}O_8S]^-$ in Scotch Whisky (S14-2373) showing the a) monisotopic peak, b) first, c) second, and d) third isotopologue regions. The recorded spectrum is shown in black, whilst theoretical masses and relative abundances are plotted as red stars. Chemical formulae are annotated with isotopic substitution in red.

Here, a window of 7 mass units was isolated, a mass spectrum acquired and calibrated against previously assigned CHO species from the broadband spectrum. The predicted isotopic pattern ($[C_{13}H_{19}O_8S]^-$) was calculated and is annotated as red stars on the spectral data (Figure 6.3). The first isotope (Figure 6.3a) had an assignment error of 89 ppb, resolution of 760,500 and SNR of 13,935. The original broadband spectrum was not sufficient for IFS with equivalent values of 119 ppb (error), 375,000 (resolution) and 744 (SNR). Such a substantial increase in SNR allowed observation of isotopic ions at the second, third, and fourth isotope nominal mass.

For example, the final observed isotopologue ($^{13}C^{12}C_{12}H_{19}^{18}O_{16}O_7^{32}S$), with a theoretical natural abundance of only 0.23 % of the monoisotopic form, was still observed with a resolution of 795,000 and SNR of 57. In contrast in the broadband spectrum this peak was below the intensity threshold for peak picking. The in-

creased resolution and SNR achieved using this technique allowed isotopologues with the same nominal mass to be distinguished and to confirm the elemental composition of this molecule. This is demonstrated clearly in the third isotope signal (Figure 6.3c) – where the intensity ratios of the O/S/C isotope peaks were compared to the predicted isotope pattern and confirm the overall elemental composition of this ion as $[C_{13}H_{19}O_8S]^-$. This illustrates the power and importance of windowing spectra for isotope fine structure assignments. This technique was repeated as part of regular analysis when calibrating or assigning spectra. Further examples are presented for $[C_{30}H_{45}O_7]^-$, and two other formulae, $[C_{18}H_{29}O_3S]^-$, and $[C_{12}H_{19}O_{10}]^-$, in a different sample (S14-1908), in Figure 6.4.

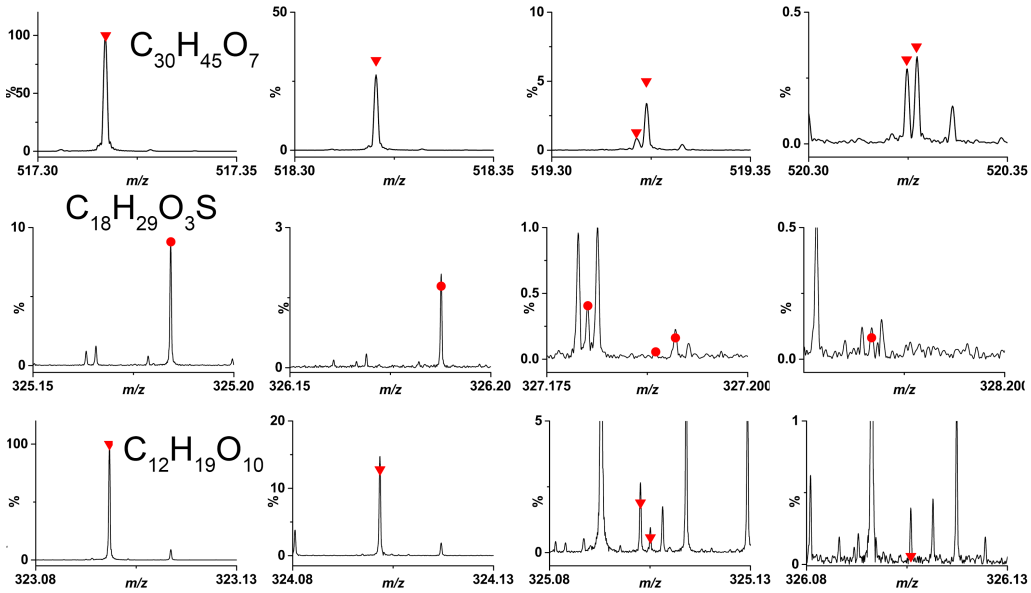


Figure 6.4: IFS analysis of $[C_{30}H_{45}O_7]^-$ (sample S14-2373), $[C_{18}H_{29}O_3S]^-$ and $[C_{12}H_{19}O_{10}]^-$ (both sample S14-1908) in Scotch Whisky showing the (left to right) monisotopic peak, first, second, and third isotopologue regions. The recorded spectrum is shown in black, whilst theoeretical masses and relative abundances are plotted as red circles and arrows.

The combination of exact mass accuracy and IFS allows for the confident assignment of molecular formulae in a sample. However, there remains no structural information obtained from these methods as the translation of molecular formulae alone into chemical structures is not possible. Any given formula can correspond to many possible structures. Using the S14-2373 sample as an example, the mass spectrum had 1902 unique formulae assigned. These formulas were cross-referenced against the *ChemSpider* (The Royal Society of Chemistry) database, and 1735 formulae had at least one known structure in the database. Many had more; in total over 320,000 possible known structures were found for just 1902 input formulae. Based on simple combinatorics, it is expected that many

Table 6.2: Subset of molecular formulae (neutral) assigned in Scotch Whisky (S14-2373) by ESI along with associated absolute errors of assignment in ppb and possible compound identification based on known chemistry of Scotch Whisky. Asterisks indicate assignments also observed by Garcia *et al.*[54]

M/Z	ERROR ppb	FORMULA	PROPOSED ID
151.0400608	46	$C_8H_8O_3$	Vanillin
155.1077472	39	$C_9H_{16}O_2$	Whisky Lactone
167.0349732	54	$C_8H_8O_4$	Vanillic Acid
169.0142383	50	$C_7H_6O_5$	Gallic Acid
171.1390419	67	$C_{10}H_{20}O_2$	Ethyl octanoate / decanoic acid*
177.0557077	57	$C_{10}H_{10}O_3$	Coniferaldehyde
179.0561001	65	$C_6H_{12}O_6$	Monosaccharide
181.0506241	46	$C_9H_{10}O_4$	Syringaldehyde
191.0349739	44	$C_{10}H_8O_4$	Scopeletin
197.0455386	43	$C_9H_{10}O_5$	Syringic Acid*
199.1703420	58	$C_{12}H_{24}O_2$	Ethyl decanoate / dodecanoic acid*
207.0662751	35	$C_{11}H_{12}O_4$	Sinapaldehyde
227.2016473	28	$C_{14}H_{28}O_2$	Ethyl Dodecanoate*
281.2486228	67	$C_{18}H_{34}O_2$	Ethyl-9-Hexadecenoate
283.2642638	35	$C_{18}H_{36}O_2$	Ethyl Hexadecanoate
300.9989723	61	$C_{14}H_6O_8$	Ellagic Acid*
341.1089493	42	$C_{12}H_{22}O_{11}$	Disaccharide*

of the higher mass compounds will have many more possible structures, this number is likely a significant underestimate.[248] That said, based on *a priori* knowledge of compounds present in Scotch Whisky, it is possible to propose identities for some peaks in the spectrum; for example, a large peak at m/z 300.99897 corresponds to $[C_{14}H_5O_8]^-$, which is likely to be ellagic acid. Further identifications are proposed in Table 6.2, with molecular formulae assignments also made by Garcia *et al.* highlighted with an asterisk.[54] Note that all but one of the CHO-only ESI-(-) assignments reported by Garcia *et al.*[54] could be found here in a single sample (S14-2373). The N-containing and S_2 species were not assigned.

6.5.4 *Tandem MS*

Whilst it is not possible to determine structure from formula alone, mass spectrometry is commonly used in structural elucidation, and though the overall focus in this thesis is application of direct infusion broadband molecular ion analysis, tandem MS can be of use. In this subsection, two means to achieve this without prior chromatographic separation are presented - use of the quadrupole, or in-cell isolation and fragmentation.

6.5.4.1 *In Source and Quadrupole*

The compound $C_{30}H_{46}O_7$ was observed in all but one of the Scotch Whisky samples in this study, and whilst the relative abundance of the molecular ion varied significantly (relative standard deviation of 161 %), it was deemed to be a compound of interest. Furthermore, in some spectra it was a dominant ion, such as seen in Figure 6.1. It is the most abundant ion not easily given a chemical identity based solely on exact mass and literature knowledge.

$C_{30}H_{46}O_7$ has been reported previously, including during an LC-MS study of Irish Whiskey,[249] with the chromatography suggesting that two structural isomers were present. MS/MS analysis highlighted several major fragments, however, no structure was proposed. Due to the evident interest in this compound, its fragmentation was repeated here.

The quadrupole mass filter was used to isolate a very narrow region of just 1 m/z around m/z 517 in sample S14-1941. As this ion was so abundant within the spectrum, and the most abundant ion by a significant amount within this mass range, quadrupole filtering was sufficient for isolation. The less abundant species within the same m/z may still be present, and may fragment, but naturally with much lower abundances, and thus will not impede interpretation. Additionally, the high mass accuracy will allow confident identification of the fragments of

interest. Fragmentation was achieved through CID within the collision cell, producing a number of fragments. MS^3 fragmentation requires prior fragmentation within the source, and quadrupole filtering of the subsequent ion to be further fragmented. The major fragmentation peaks observed (Figure 6.5) are summarised in Table 6.3 and are consistent with those previously reported.[249]

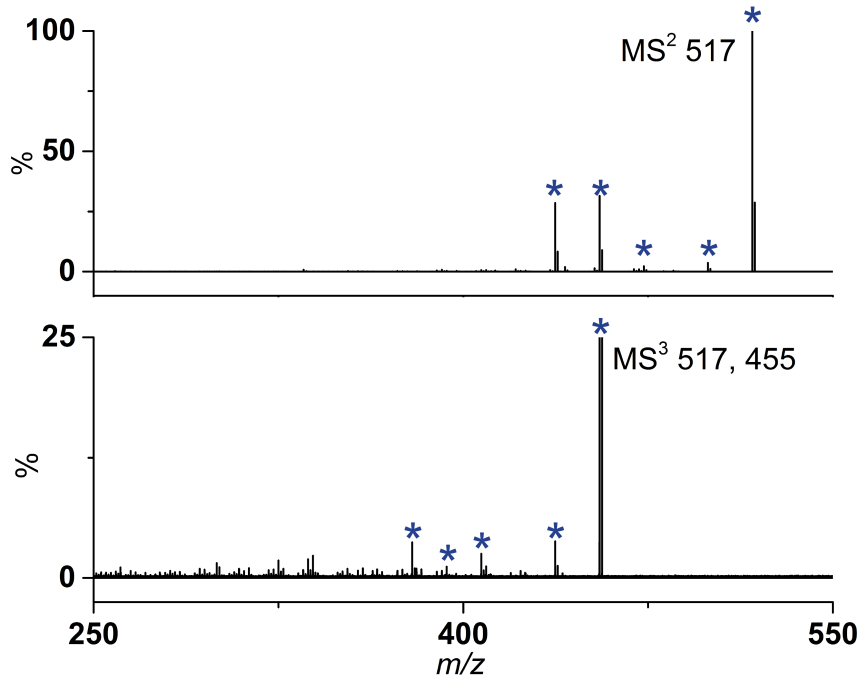


Figure 6.5: Fragmentation spectra for compound $C_{30}H_{46}O_7$ in Scotch Whisky (S14-1941); (top) MS^2 517 and (bottom) MS^3 517-455. Key peaks which are listed in Table 6.3 are annotated. Fragmentation performed as described in the text.

Fragmentation was predominantly by loss of H_2O and CO_2, with some loss of CH_2 in the MS^3 experiment. This suggests a number of cleavable hydroxyl and carboxyl groups, along with terminal methyl functionality. Full structural elucidation was not possible; however, cross-referencing the *ChemSpider* database suggests several triterpenoid derivatives. Based on the tandem MS data of this species, and structures in the literature for the same or similar formula,[89,250,251] this compound is very likely an oleanane-type triterpenoid. Minor fragment peaks observed in Figure 6.5 (bottom) are due to the reduced selectivity of performing the initial fragmentation in the source - the fragment ion subsequently filtered with the quadrupole may not be as distinct, in terms of abundance, from other ions within that mass range.

In the broadband mass specta of Scotch Whisky, additional compounds including $C_{30}H_{46}O_8$, a hydroxylated equivalent, and $C_{36}H_{56}O_{12}$, a glucosylated equivalent, were observed. Compounds such as glucosylated bartogenic acid (Glu-BA, $C_{36}H_{56}O_{12}$) have been shown to exist in *Q. robur* (a white oak species)[251] and

Table 6.3: Fragmentation peaks observed for molecular ion $[C_{30}H_{45}O_7]^-$ using quadrupole and collision cell. Larger errors are observed as spectra not calibrated post acquisition. Both MS^2 and MS^3 performed on named masses (m/z 517 and 455)

MS^N	M/Z	ION $[M-H]^-$	FRAGMENTATION	ERROR ppm
MS	517.31708	$C_{30}H_{45}O_7$	M-H	1.43
MS^2	499.30591	$C_{30}H_{43}O_6$	M-H-H_2O	1.21
(517)	473.32671	$C_{29}H_{45}O_5$	M-H-CO_2	1.14
	455.31610	$C_{29}H_{43}O_4$	M-H-CO_2-H_2O	1.28
	437.30559	$C_{29}H_{41}O_3$	M-H-CO_2-$2H_2O$	1.21
MS^3	455.31608	$C_{29}H_{41}O_3$	M-H-CO_2-H_2O	1.32
(517, 455)	437.30612	$C_{29}H_{41}O_3$	M-H-CO_2-$2H_2O$	1.25
	407.29504	$C_{28}H_{39}O_2$	M-H-CO_2-$3H_2O$	1.27
	393.27939	$C_{27}H_{37}O_2$	M-H-CO_2-$3H_2O$-CH_2	1.31
	379.30014	$C_{27}H_{39}O$	M-H-CO_2-$4H_2O$-CH_2	1.32

bartogenic acid (BA) itself has been found in oak wood extracts. Therefore, the identity of this unknown compound is likely bartogenic acid or a structurally similar isomer.

6.5.4.2 *In Cell Isolation and Fragmentation*

Isolation and fragmentation of species using the source, quadrupole, and collision cell was possible for $C_{30}H_{46}O_7$ without prior chromatographic separation simply because it was so abundant relative to other signals. Selective isolation and fragmentation of less abundant species, especially 'grass' ions, is not possible this way in a complex mixture. However, FTICR MS lends itself to a far more sophisticated means of ion isolation and fragmentation.

Within the ICR cell in a routine broadband experiment, a frequency-swept pulse excites all ions to the same radius for detection. However, selective excitation can be used to manipulate the cyclotron radius of specific ions - similar to how selective RF pulses can be applied in NMR. With this technique, it is possible to increase - or decrease - the radius of select ions. At an extreme, it can be used to 'eject' ions - in reality, excite them to the radius of the cell, where they collide with the walls and are adsorbed and lost. This approach is used here to isolate individual ions, or exact masses, for further investigation.

Once the ion of interest has been isolated, it is possible to fragment the ion within the ICR cell. Whilst advanced methods, specifically in proteomics, may involve use of ECD or IRMPD, a simpler approach is possible in the form of SORI CID. SORI uses a long, low-power RF irradiation to manipulate the position and coherence of the ion cloud, in this case off-resonance from the ion of interest (typically within $\pm$ 500 Hz). Simultaneously, the pressure within the cell is increased by injection of neutral gas to trigger fragmentation via CID. This mechanism is essentially the same as how ions are fragmented within the collision cell.

The first target for this method was another significant and ubiquitous molecular formula assignment within the Scotch Whisky sample set - $C_{14}H_6O_8$. This molecular formula has a low hydrogen-to-carbon ratio, and indicates a conjugated unsaturated system. The formula very likely corresponds to a well known maturation related congener in whisky - ellagic acid. However, cross referencing against the *ChemSpider* database indicates eight reported compounds, although it is ellagic acid which has the most associated data sources and references.[252]

The work flow, and power, of in-cell isolation is demonstrated in Figure 6.6. Here, a complex mass spectrum of a Scotch Whisky sample (S14-1911, a Blended Scotch Whisky) (Figure 6.6a-c) is simplified to a single peak. In the original broadband spectrum, there were 2426 peaks detected above the SNR threshold. By application of a quadrupole mass filter, centred at m/z 301 $\pm$ 4 (Figure 6.6d-f), this is reduced to a narrow mass range containing several dozen ions. In-cell isolation by selective excitation ('sweep') removes all remaining unwanted signals (Figure 6.6g-i). Naturally, the IAT is increased correspondingly, and as such the SNR increases from the broadband spectrum to the isolated ion spectrum. This boost in SNR is necessary to ensure sufficient signal for fragmentation.

Once an individual ion has been isolated, SORI CID can be used to fragment it. The application of SORI CID to $C_{14}H_6O_8$ is shown in Figure 6.7. Careful optimisation of the collision energy must be performed - too low a power will produce no fragmentation, whilst too high may destroy all the signal. This process is time consuming, and iterative - the IAT and other instrument settings must be tuned to compensate for the signal attenuation caused by isolating a single species and then fragmenting it. Once fragmentation is observed, many transients must still be co-added to ensure sufficient SNR. Also, spectra must be acquired without CID energy applied to act as a control for auto-fragmentation or noise peaks, as shown in Figure 6.7a.

Figure 6.7 demonstrates a highly successful isolation of a peak of interest - $C_{14}H_6O_8$ - in Scotch Whisky without any chromatographic separation. In-cell isolation and subsequent fragmentation resulted in a number of fragment ions of high and low abundances. Confirmation of the identity of this species as ellagic

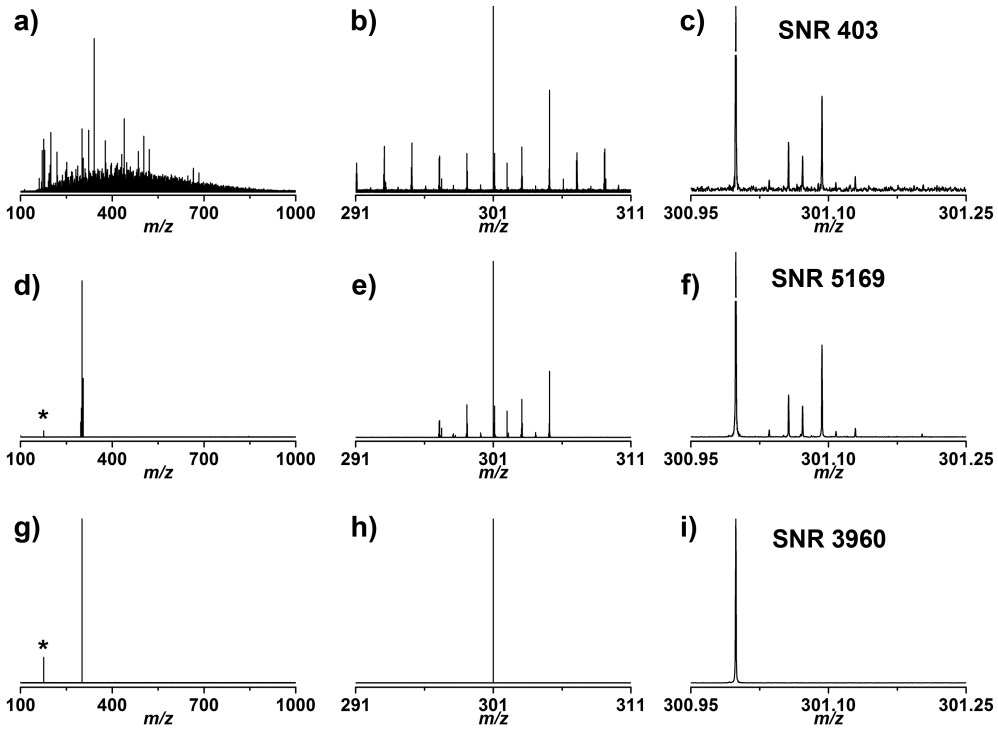


Figure 6.6: In-cell isolation of $C_{14}H_6O_8$ in Scotch Whisky (S14-1911). a-c) progressive expansions of a regular broadband ESI FTICR mass spectrum, d-f) progressive expansions of the mass spectrum with quadrupole mass filter applied at m/z 301 $\pm$ 4, g-i) progressive expansions of application of in-cell isolation after quadrupole mass filtering. Expansions shown for full spectrum (a,d,g), m/z 291 - 311 (b,e,h), and m/z 300.95 - 301.25. SNR annotated for molecular ion signal. * indicates turbo noise signal.

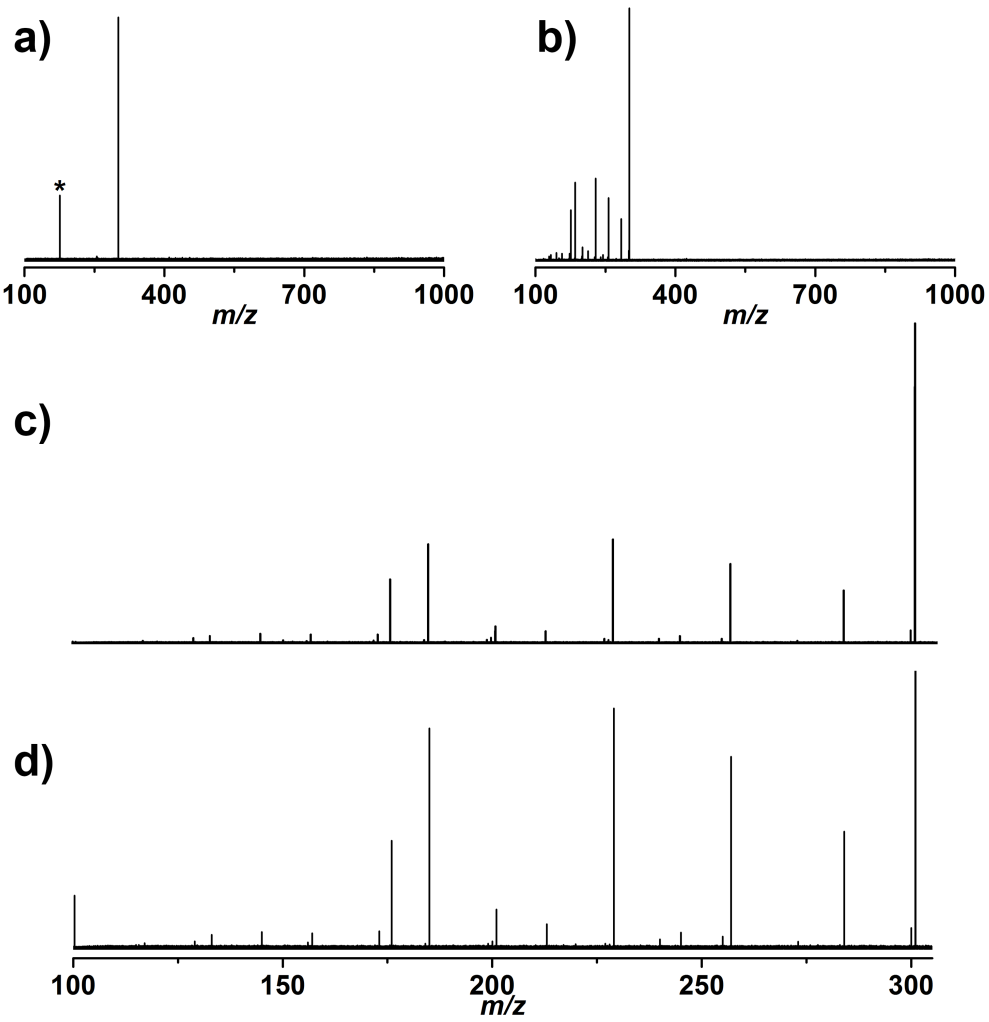


Figure 6.7: In-cell isolation and SORI CID of $C_{14}H_6O_8$ in Scotch Whisky (S14-1911); a) shows the full mass range with the ion isolated but no CID energy applied - the peak labelled with a asterisk is a noise peak, b) shows the effect of applying SORI CID, c) is an expansion of b), and d) shows the same isolation and fragmentation process applied to a standard of ellagic acid in methanol:water.

acid was achieved by acquiring the fragmentation pattern for a reference standard of ellagic acid. This compound was dissolved in methanol:water and isolated and fragmented in the same manner as the Scotch Whisky sample. The fragmentation pattern for $C_{14}H_6O_8$ in Scotch Whisky is shown in Figure 6.7c, whilst the reference standard is in Figure 6.7d. They are, essentially, identical with only subtle abundance differences (*e.g.* the species around m/z 200 and 245). The peak at m/z 100.33395 is one third of the mass of the ellagic acid molecular ion and likely a harmonic peak.

The high resolving power, mass accuracy, and sensitivity of FTICR MS applies also to fragmentation spectra. This is shown in Figure 6.8 where 24 distinct fragment ions of the ellagic acid reference standard were detected and confidently identified based on exact mass and mass differences. The identities of these fragments are presented in Table 6.4.

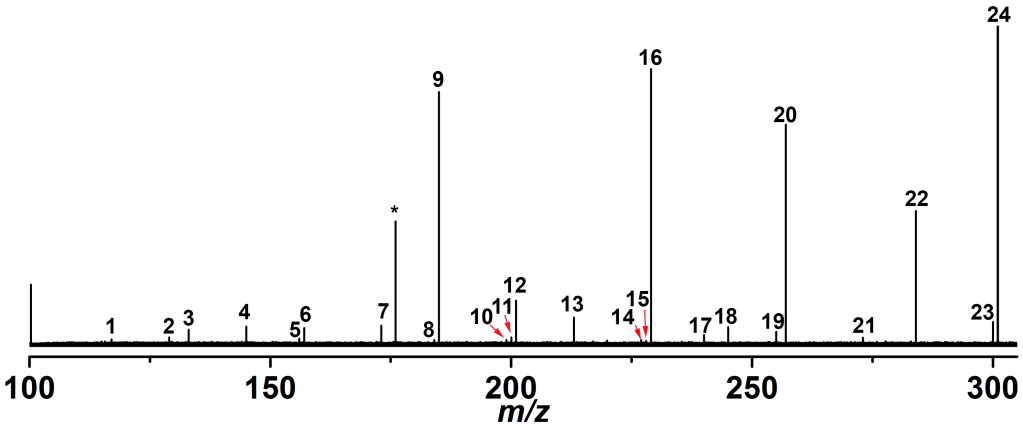


Figure 6.8: Annotated in-cell isolation and SORI CID of $C_{14}H_6O_8$. Ellagic acid reference tandem MS spectrum is shown.* is a turbo noise peak. Peak annotations are summarised in Table 6.4

Ellagic acid represented a relatively straightforward application of this process - despite being in a complex mixture, it was a relatively abundant ion. It is possible that a sample of Scotch Whisky may have a greater amount of this compound and simple quadrupole isolation may be sufficient - as for $C_{30}H_{46}O_7$ previously. However, in-cell isolation and fragmentation is required to study the ions immediately adjacent to ellagic acid, or to study other minor compounds in solution.

The ability to isolate ions as close as 15 mDa is demonstrated in Figure 6.9 for the region containing ellagic acid. Ions shown in Figure 6.9a,d were successfully isolated and fragmented, whilst those shown in Figure 6.9b,c could only be isolated and not fragmented. This may be due to the much lower achieved SNR - these ions are less abundant, and may be less stable. No fragmentation

Table 6.4: Fragmentation peaks observed for molecular ion $[C_{14}H_5O_8]^-$ in a reference standard of ellagic acid using in cell isolation and SORI CID. Spectrum was single point calibrated to the molecular ion. See text for details and Figure 6.8 for spectrum.

PEAK	ION $[M-H]^-$	ERROR ppm	PEAK	ION $[M-H]^-$	ERROR ppm
24	$C_{14}H_5O_8$	0.00	12	$C_{11}H_5O_4$	0.70
23	$C_{14}H_4O_8^{\bullet}$	0.03	11	$C_{11}H_4O_4^{\bullet}$	0.80
22	$C_{14}H_4O_7^{\bullet}$	0.18	10	$C_{11}H_3O_4$	0.60
21	$C_{13}H_5O_7$	0.29	9	$C_{11}H_5O_3$	0.84
20	$C_{13}H_5O_6$	0.34	8	$C_{11}H_4O_3^{\bullet}$	0.77
19	$C_{13}H_3O_6$	0.16	7	$C_{10}H_5O_3$	1.00
18	$C_{12}H_5O_6$	0.35	6	$C_{10}H_5O_2$	1.16
17	$C_{13}H_4O_5^{\bullet}$	0.51	5	$C_{10}H_4O_2^{\bullet}$	1.18
16	$C_{12}H_5O_5$	0.55	4	$C_9H_5O_2$	1.25
15	$C_{12}H_4O_5^{\bullet}$	0.48	3	$C_8H_5O_2$	1.29
14	$C_{12}H_3O_5$	0.63	2	C_9H_5O	1.35
13	$C_{12}H_5O_4$	0.65	1	C_8H_5O	1.55

was observed for these species at any applied CID energy. A higher concentration sample may be necessary to isolate these ions with acceptable sensitivity for fragmentation. However, of course, it may be that they do fragment but lose the charge carrying part of the molecule and are no longer observable.

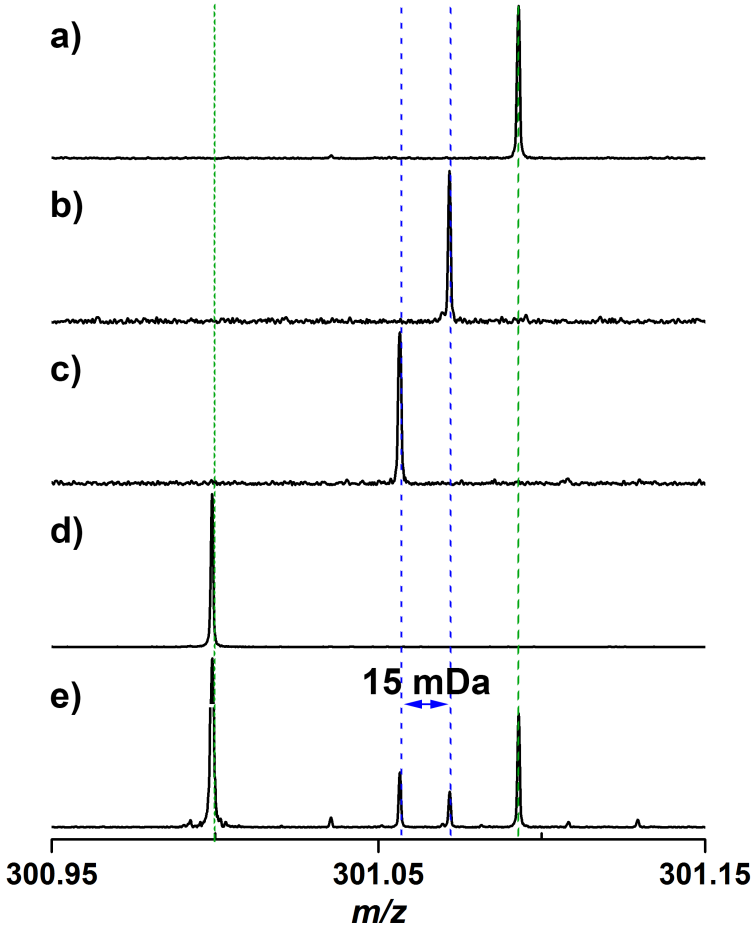
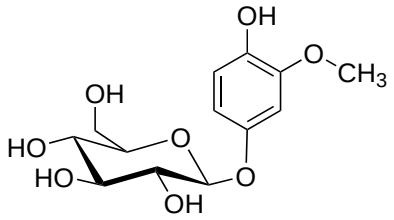


Figure 6.9: In-cell isolation of several ions within a single nominal mass in Scotch Whisky (S14-1911). a-d) shows successful isolation of individual peaks. a) and d) were fragmented successfully whilst b) and c) were not. The mass difference between b) and c) was only 15 mDa

To achieve this highly selective isolation, the in-cell isolation method used both 'sweep' and 'shots'. Sweep excites all ions except for a narrow band around the ion of interest to the cell wall, thus removing them. Shots are subsequent highly selective irradiation of residual unwanted ions. These selective pulses are inherently longer, and further attenuate signal. The need for 'shots' is demonstrated in Figure 6.11a,b where (a) the use of sweep excitation removed all but the nearest ions, and (b) shots eliminated those residual unwanted species.

The ion shown in Figure 6.9a corresponds to the molecular formula $C_{13}H_{18}O_8$. Cross referencing ChemSpider, and considering the expected chemistry of Scotch

Whisky, the species is likely a glycosylated aromatic compound, perhaps a phenol, such as 4-hydroxy-3-methoxyphenyl-β-D-glucose. The isolated ion was fragmented via SORI CID, with the fragmentation pattern shown in Figure 6.11c. A reference standard could not be readily obtained for comparison. However, the presence of a series of fragments with formulae indicating unsaturated systems, culminating with $[C_7H_7O_3]^-$ at m/z 139.04027 (1.47 ppm), indicates this structure likely does contain a phenolic group. Subtracting the fragment formulae from the molecular ion results in $C_6H_{10}O_5$, which would be indicative of a glycoside. An example possible structure is shown in Figure 6.10, corresponding to tachioside (**45**), a known lignan glycoside.[253]



45

Figure 6.10: Possible structure of compound with molecular formula $C_{13}H_{18}O_8$ (**45**).

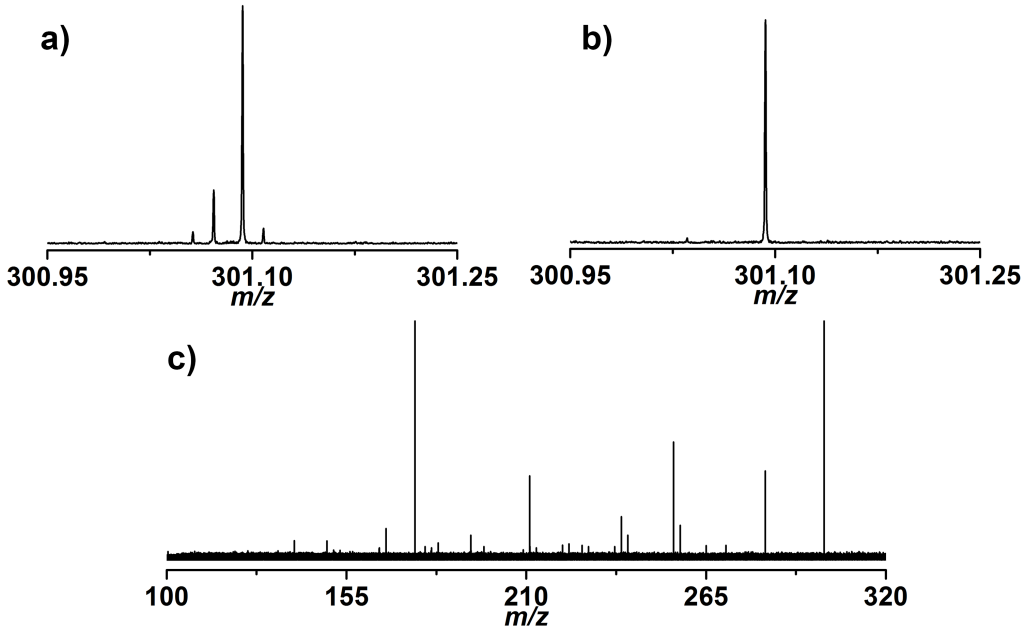


Figure 6.11: In-cell isolation and SORI CID of $C_{13}H_{18}O_8$ in Scotch Whisky. a) shows the nominal mass with 'sweep' only in cell isolation, b) shows the subsequent application of cleanup 'shots', and c) shows the expanded SORI CID fragmentation spectrum.

In addition to ellagic acid and the glycoside, several other ions across the mass spectrum were selectively isolated and fragmented. A species at m/z 341.10893, molecular formula $C_{12}H_{22}O_{11}$ - a disaccharide - fragmented. As typical for carbohydrate fragmentation, the mass spectra were complex with many peaks observed and identifiable. Comparison to reference standards of other disaccharides - including maltose, lactose, and sucrose - yielded no conclusive results. A peak at m/z 503.16176 ($C_{18}H_{32}O_{16}$) - fragmented predominately to same mass as the disaccharide, and it is likely this species is a similar trisaccharide - *i.e.* maltotriose fragmenting to maltose, or cellotriose fragmenting to cellobiose.

This methodology has significant potential, but with so many thousands of compounds in Scotch Whisky, it will be necessary to determine which species are priority targets for future work.

6.5.5 *Data Visualisation*

The comprehensive assignment of chemical formulae in mass spectra of the studied whisky sample set (74 % for sample S14-2373 and 72 % to 88 % across the entire sample set) allows the data to be visualised based on the atomic composition of the assigned species. In this subsection, various means are utilised to visualise the chemical diversity across the sample set.

6.5.5.1 *Heteroatomic Class Distribution*

Visualisation of the complex mixture as a whole, for example by inspecting the heteroatomic class distribution within a sample or set of samples, allows for further understanding of the nature of the complex mixture. Figure 6.12 shows the heteroatomic class distribution calculated for the entire set of Scotch Whiskies, comparing the chemistry of malt and blend whisky. Here the data are represented as a violin plot, which shows the distribution of the counts of heteroatomic classes across the type of samples, *e.g.* the number of compounds with the heteroatomic class O_2 within the malt whisky type. The distributions were calculated as a kernel density estimate.

Overall, the general trend is a unimodal distribution centred at around O_9 for blends, with the malts' distribution centred around O_{10}. Generally the malt whiskies have a higher count at each heteroatomic class than the blend whiskies. Additionally, there is a larger spread of counts around the middle heteroatom classes, with the upper and lower extremes having narrower distributions.

Within each heteroatomic class, the blends generally have a narrower unimodal distribution, whereas the malts have a larger distribution. In several cases,

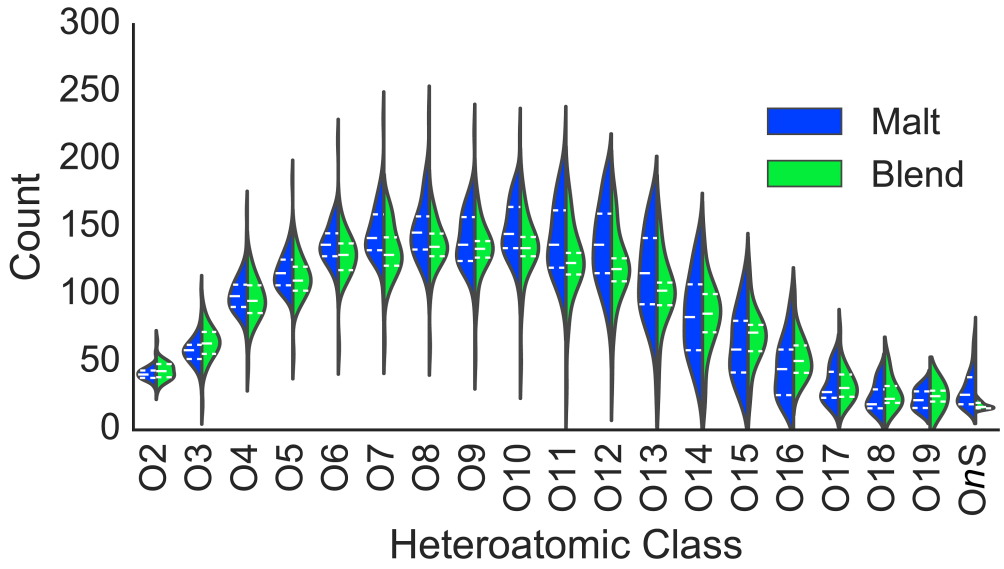


Figure 6.12: Violin plot for heteroatomic class distributions for the entire sample set. Malts are in blue and blends are in green. The shapes of the plots represent the kernel density estimate, calculated using the Silverman kernel bandwidth rule of thumb, of the distribution of counts across the sample set, quartiles are represented with white dotted lines within the plots. Each violin has been scaled to the same width. Sulfur containing heteroatomic classes are summarised in the single O_nS class.

this distribution is (at least) bimodal, for example for malts at O_{17}. Interestingly, the O_nS class is distinctly different between malts and blends, with a much narrower and smaller distribution in blends than in malts. The total counts for this class are also low relative to other heteroatomic classes. The distribution of the individual CHOS classes across the sample set is shown in Figure 6.13, showing that malts contain far more sulfur species, with a greater range of occurrences at higher oxygen count.

This heteroatomic class distribution violin plot shows that it may be possible to discriminate between types of Scotch Whisky by their mass spectral profiles via chemometric methods. This is investigated further in section 6.6.

6.5.5.2 *Van Krevelen and DBE plots*

Two other commonly used visualisations, van Krevelen and DBE versus carbon number plots, are shown for four samples in Figure 6.14. These samples are three mature Scotch Whisky samples including a single malt matured in ex-Sherry casks (S14-2373), a single malt matured in ex-Bourbon casks (S14-1948), and a Blended Scotch Whisky (S14-1916). A malt NMS was also shown for comparison (S11-0034). Using the molecular formulae, the van Krevelen diagram (Fig-

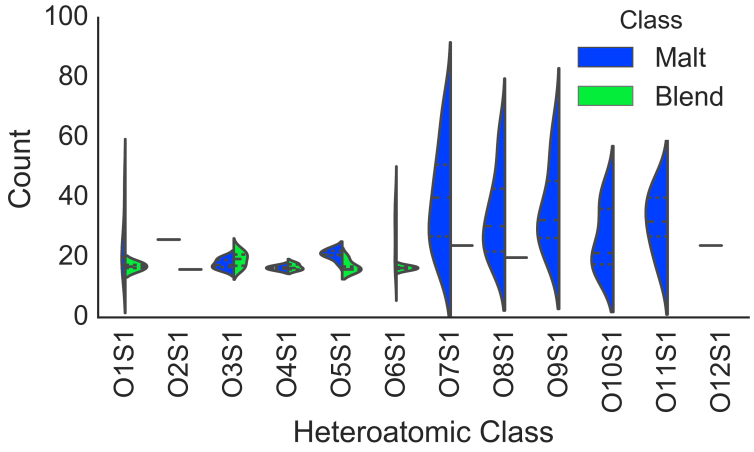


Figure 6.13: Violin plot for OS heteroatomic class distributions for the entire sample set. Malts are in blue and blends are in green. The shapes of the plots represent the kernel density estimate, calculated using the Silverman kernel bandwidth rule of thumb, of the distribution of counts across the sample set, quartiles are represented with dotted lines within the plots. Each violin has been scaled to the same width.

ure 6.14 left column) sorts assigned species based on the hydrogen-to-carbon (H/C) versus the oxygen-to-carbon ratios (O/C). Extra information is encoded; the marker size reflects signal abundance and the colour reflects the mass. ESI FTICR MS is not quantitative, and signal abundance is only an approximation for compound abundance in the sample. In the resulting plot, major biogeochemical classes of compounds (such as lipids, carbohydrates, lignins, ellagitannins, etc.) cluster in specific characteristic regions.

The van Krevelen plots (Figure 6.14 left column) reveal several features which are common to all Scotch Whisky samples analysed. These include large signals for fatty acids or esters at $O/C < 0.3$ and $H/C = 2$, representing formulae including $C_{10}H_{20}O_2$ and $C_{12}H_{24}O_2$, compounds which may be formed during fermentation, or via esterification during maturation. There is also a significant central star region, extending outwards from O/C of 0.5 and H/C of 1 corresponding to aromatic molecules; lignin and ellagitanin type compounds. At the centre of the star are formulae such as $C_8H_8O_4$, for example vanillic acid, a known maturation related compound. There is also a common region at O/C of 1 and H/C of 2 corresponding to carbohydrates. These peaks will represent compounds such as glucose and fructose, as well as related compounds extracted through maturation from cask heat treatment or carry over from previous fills.

Despite the similarities between the whisky samples as shown by the van Krevelen diagrams, it is clear that there is substantial chemical diversity amongst the samples. The van Krevelen plot for sample S14-2373 (Figure 6.14a), a 12-

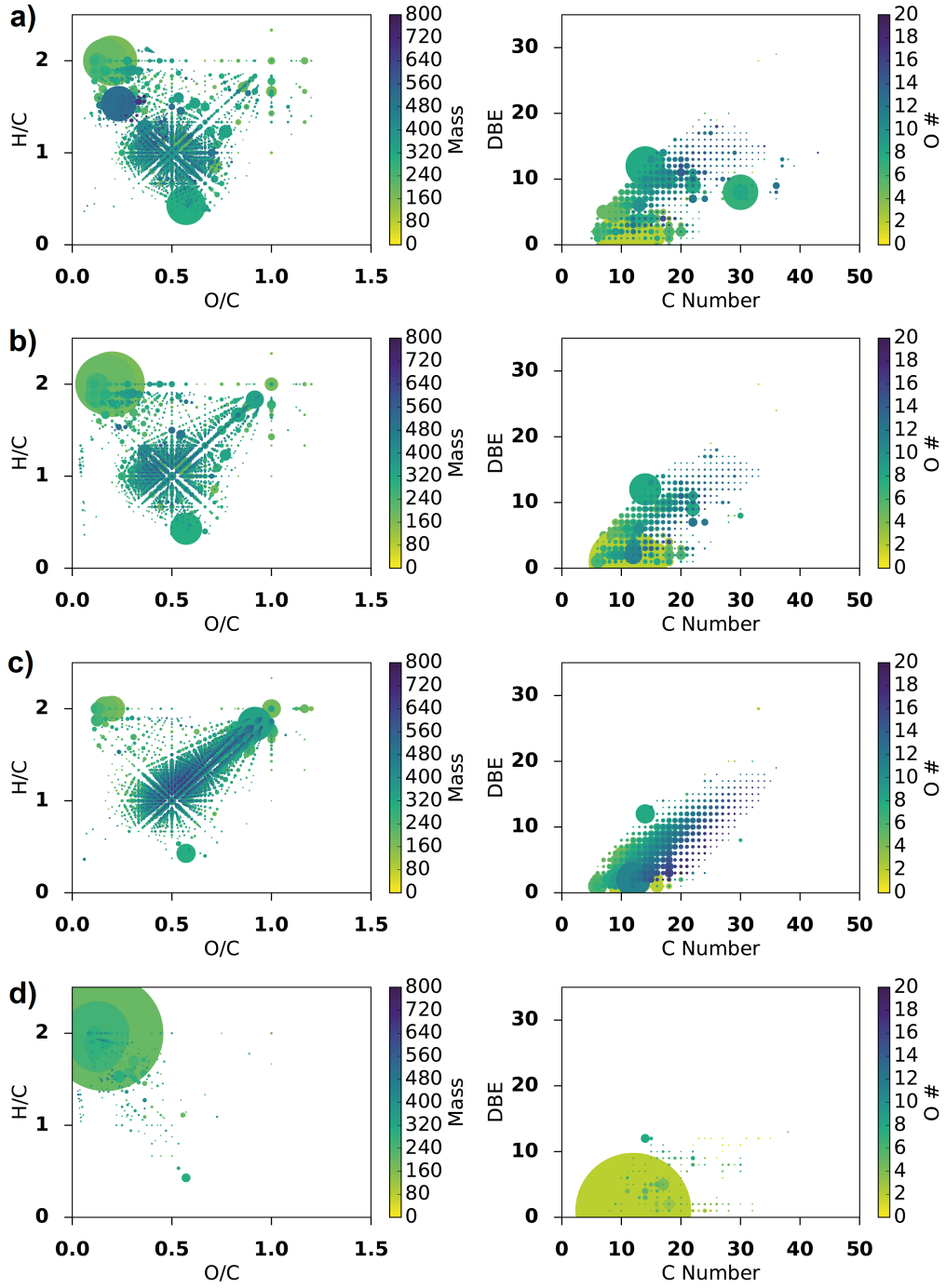


Figure 6.14: Van Krevelen (left) and DBE versus carbon number plots (right) for four different samples; (top to bottom) a) S14-2373 an ex-Sherry cask matured Highland single malt, b) S14-1948 an ex-Bourbon cask matured Highland single malt, c) S14-1916 a blended whisky, d) S11-0034, a new make malt spirit from a Speyside distillery. In the van Krevelen diagrams, points are coloured according to mass and sized according to relative signal intensity. In the DBE versus carbon plots, points are coloured according to oxygen number and sized according to relative signal intensity

year old Sherry cask matured Highland single malt, shows a significant complexity within the central star region around O/C 0.5 and H/C 1.0. The higher abundance species towards the bottom at O/C of 0.6 and H/C of 0.4 corresponds to $C_{14}H_6O_8$, with the elemental formula of ellagic acid, a known polyphenolic tannin-hydrolysis product present in Scotch Whisky.[28,40] There is also a high abundance species at O/C of 0.2 and H/C of 1.5 that corresponds to $C_{30}H_{46}O_7$, likely bartogenic acid as discussed in subsection 6.5.4. The second sample, S14-1948, (Figure 6.14b) a 10-year old ex-Bourbon cask matured Highland single malt, appears slightly less complex than the ex-Sherry cask matured sample. Interestingly, $C_{30}H_{46}O_7$ is of significantly lower relative abundance in this sample. The third sample, S14-1916, (Figure 6.14c) a blended whisky, has a complex central star region. However, in this case, the complexity spreads towards H/C 2 and O/C 1. This stark difference compared to the malt samples shown may be due to the differences in production methods. Some of the major species in this region correspond to (from the top right to the centre) $C_6H_{12}O_6$, $C_{12}H_{22}O_{11}$, $C_{12}H_{20}O_{10}$, $C_{12}H_{18}O_9$. These are likely carbohydrates or carbohydrate derivatives, possibly 'caramel-type' compounds.[36]

For comparison, a malt NMS (S11-0034) was also analysed using this methodology (Figure 6.14d). This sample is non-peated and from a Speyside distillery. This spectrum contained far fewer peaks than the spectra of the mature samples, with many nominal m/z not represented. As a result, there were fewer assigned compounds in this spectrum (only 252) indicating the lack of complexity. The major compounds present are fatty acid or ester type compounds, including $C_{10}H_{20}O_2$. These compounds will correspond to volatile congeners produced during fermentation and carried over through distillation. When compared to the three mature Scotch Whisky samples, it is abundantly clear that the chemical complexity of Scotch Whisky reflected in their mass spectra arises through maturation.

Furthermore, as trend lines through data points on the van Krevelen plot represent the characteristic loss or gain of repeating chemical units, such as apparent methylation or oxidation, these plots have been used to visualise potential chemical transformations occurring in complex samples.[165,166] The trend lines extending from the central star region represent different chemical transformations; for example, vertical lines correspond to (de)hydrogenation, diagonal lines (de)methylation, and horizontal lines oxidation/reduction. With one sample representing just a single time point, it is impossible to infer the direction of these trend lines with these samples - if a direction even exists. However here comparisons of matured whisky to the new-make spirit allow a general trend to be inferred. The trend lines correspond to an oxidative process, pointing towards

the central star region, where the types of compounds present are increasingly aromatic, unsaturated, and oxygen rich. This change in chemistry may be due to chemical transformations within the new-make spirit, however it is likely that the majority of changes are due to extraction of additional compounds from the cask. Investigation of this would require samples of the same spirit as it matures over time, however this process takes years and such sample sets are challenging to curate.

The chemical diversity can also be represented by assigning to each molecular formulae its DBE. The carbon number of each species is plotted against DBE as a scatter plot; the points are again sized by relative abundance, but coloured by oxygen number. The DBE versus carbon number plots (Figure 6.14 right column) demonstrate similar trends in the samples as observed with the van Krevelen plots. The ex-Sherry cask matured sample (Figure 6.14a right) highlights the same major species, including the relatively high abundance compound $C_{30}H_{46}O_7$ (DBE=8). The most significant difference between the first and second sample's DBE versus carbon number plot is this high abundance compound, which is of significantly lower abundance in the ex-Bourbon matured sample (Figure 6.14b). The blended whisky (Figure 6.14c) has a distinctly different DBE plot. The distribution of species in this plot appears far more linear, with fewer species of significantly higher relative abundance present. The DBE plot of the new-make spirit is sparse, with the major components having a very low DBE and carbon number around 10 to 15.

6.6 CHEMOMETRIC ANALYSIS

The wealth of information provided by high resolution mass spectrometry, including the confident assignment of thousands of molecular formula, lends itself to chemometric analysis. The input data for chemometric analysis is a matrix $\mathbf{X}$ of $m \times n$, with m variables (normalised abundance of each assigned molecular formula) and n observations (each spectrum or sample). Most statistical methods do not support missing data well, and as only $\approx$10 % of formula were common to all spectra, a solution was required. The two main options are a) discard variables (formula) not found in, *e.g.* 50 % of samples, or b) fill missing values with some random value. The former option would discard significant volumes of data - and important data, the formula which are different between samples. Filling missing values was thus chosen. Here, if a formula was not assigned in a given sample (but was in others), it was given a random abundance value equivalent to the level of the noise. This, in effect, changes the missing value from

being 'not present' to 'not detected above the peak picking and error thresholds'. Similar approaches have been used by other groups.[60,183]

The sample set is of good provenance. However, being real final products, not every manufacturing parameter or variable is known. The first step is again unsupervised analysis.

6.6.1 *Principal Components Analysis*

A PCA model (Figure 6.15a) was constructed from the entire sample set using SIMCA, as detailed in the methodology. Here, 9 principal components described 85 % of the data, with the first two representing 50 % of the variance in the data. As with the ^{1}H NMR data in chapter 5, the scores plot has been colour coded by blend or malt status. The malt samples are more separated than the blends, which are more clustered. This reflects greater diversity between the malt samples than the blends, and is consistent with the heteroatomic class plots. There are a few samples outwith the confidence intervals. These outliers are malts - the three in the top right are peated Islay single malts, two of which were the same product from different years of production (2012, 2014), and the third was also peated, from an unknown distillery on Islay. The three outliers on the bottom right were all the same product from different years, specifically a Highland single malt matured in Sherry casks. The overall spread of samples in the scores plot reflects the diverse and numerate variables causing the chemical complexity in Scotch Whisky. The loadings for this model (not shown) indicated a number of carbohydrate-type formula were more common in blends, and cask extractive-type formula more common for malts. The cause for this may be potentially greater use of caramel colouring in blends and natural colour in malt whiskies, however this is speculation as it is unknown which samples do or do not contain caramel colouring. The relative abundance of the compounds highlighted by the loadings may be a means of whisky classification.

6.6.2 *OPLS-DA*

Supervised techniques, such as OPLS-DA, allow for specific chemometric interrogation of the data. Whilst the ^{1}H NMR data in chapter 5 allowed for near perfect separation of blends and malts by PCA alone, the mass spectra did not. Thus, OPLS-DA was used to better distinguish between the malts and blends. This model was built with four orthogonal components with an $R^2\mathbf{X}$ of 0.64, $R^2\mathbf{Y}$ of 0.83, with a cumulative Q^2 of 0.61. The supervised model successfully distinguished these two types; the scores plot for the predictive and first orthogonal

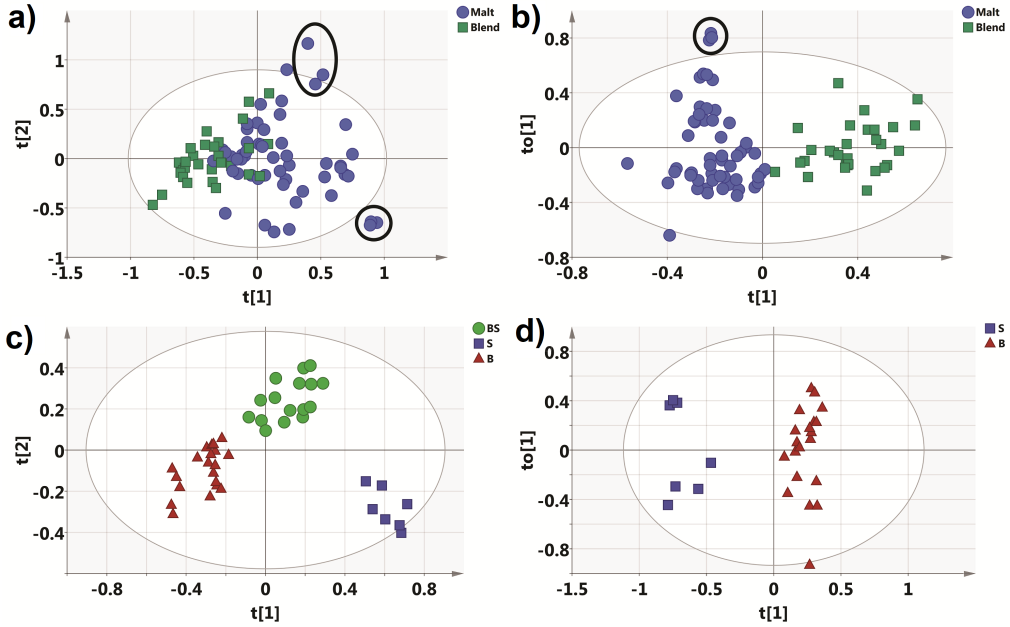


Figure 6.15: Scores plots for statistical models built around samples. a) PCA scores plot for all Scotch Whisky samples analysed (n=85) coloured according to malt (blue) or blend (green). Two examples of samples of the same whisky product from differing production years are circled. The bottom circle represents samples discussed in Figure 6.17 b) OPLS-DA Scores plot for the Malt (blue) versus Blend (green) (n=85) model. Three samples of the same whisky product from different production years are circled. c) OPLS-DA scores plot for a model based on malts with known maturation wood type (n=43). BS (green) = ex-Bourbon and ex-Sherry, S (red) = ex-Sherry, B (blue) = ex-Bourbon. d) OPLS-DA Model for malts matured in ex-Sherry only or ex-Bourbon only casks (n=27), S (red) = ex-Sherry, B (blue) = ex-Bourbon.

component is shown in Figure 6.15b. The three closely positioned single malt outliers (Figure 6.15b, top left, circled) represent the same samples grouped as the bottom right outliers in the PCA plot (Figure 6.15a).

The model identified several key formulae for discriminating between malt and blend whiskies (listed in Table 6.5). As with the PCA model, the blends are discriminated by having a higher relative abundance of a number of carbohydrate-type species, whereas the malts have a higher relative abundance of cask extractive compounds. The underlying differences between blends and malts are multiple; grain whisky is distilled to a higher alcohol strength, producing a lighter spirit; malt whiskies are a premium product and may be matured for longer in more active casks. Interestingly, this model placed one blended whisky (S14-2858) very near the malts. Upon further analysis, this blended sample was found to represent a premium blended whisky with a high malt content.

Table 6.5: Key variables (formulae) discriminating between malts and blends for the OPLS-DA model. Possible identities are proposed based on formula and expected chemistries. "Carbohydrate" includes derivative or related compounds.

VARIABLE ID	POSSIBLE IDENTITY	DISCRIMINATES TYPE
$C_9H_{10}O_5$	Syringic acid	Malt
$C_{14}H_6O_8$	Ellagic acid	Malt
$C_{30}H_{46}O_7$	Bartogenic acid	Malt
$C_7H_6O_5$	Gallic acid	Malt
$C_{12}H_{20}O_{10}$	Disaccharide	Blend
$C_{12}H_{22}O_{11}$	Disaccharide	Blend
$C_9H_{16}O_8$	Carbohydrate	Blend
$C_{12}H_{18}O_9$	Carbohydrate	Blend

These results echo those observed for the ^{1}H NMR data, with a larger sample set, in chapter 5. Given the major differences between blends and malts are presence of grain spirit in the blends, and grains contain far fewer compounds (they are distilled to higher ABV), it is reasonable that NMR would outperform ESI FTICR MS for this classification. That is to say, ESI FTICR MS is less sensitive to low molecular weight, volatile congeners which make malt spirit distinct from grain.

Again, as with the ^{1}H NMR data in chapter 5, several other variables could potentially be modelled by OPLS-DA. An OPLS-DA model was then constructed for samples of known geographical region ($n = 54$), however there was limited predictive power with a Q^2 of only 0.20. This is similar to the NMR data (section 5.6), with Q^2 0.25. As stated there, geographical origin should not be predictable as there is no *terroir* in Scotch Whisky production. Similarly, age is reflective of only the youngest spirit in the blend - and cask activity varies significantly - thus no success was achieved modelling age by OPLS-DA, either.

Next, an OPLS-DA model was constructed for peated and non-peated samples from the 2014 standard sample set ($n_{peated} = 13$, $n_{non\text{-}peated} = 18$). The predictive power of this model was also limited, with a Q^2 of 0.14. Note that peat is introduced at the malting stage, prior to distillation, and so the compounds related to peated characteristic will have been produced through pyrolysis and must be volatile enough to pass through distillation. As such, they are likely to be low molecular weight volatile compounds, and thus, as with the blend and malt classification, 'peat' characterisation is likely not suitable for ESI FTICR MS. As such,

peat character is better determined by NMR, although the sensitivity limitations present a challenge.

Analysis of the new make spirit sample (Figure 6.14d) demonstrated that the complexity in Scotch Whisky arises through maturation. Therefore, OPLS-DA modelling maturation wood types was the next target. An OPLS-DA model was built for the malt samples across all years with known maturation wood type (n = 43) (Figure 6.15c). Three classes were considered – whiskies which had been matured in ex-Bourbon casks only (B), ex-Sherry casks only (S), or a mixture of ex-Bourbon and ex-Sherry casks (BS). This model had 4 orthogonal components with an $R^2\mathbf{X}$ of 0.73, $R^2\mathbf{Y}$ of 0.86, and Q^2 of 0.61. The first predictive component (x-axis) separates class S from B, and the second predictive component (y-axis) separates the S and BS classes. The separation of B and BS classes is smaller, possibly due to some of BS samples only being finished in an ex-Sherry cask, rather than matured more evenly (over time) between wood types.

To understand the discrimination between ex-Sherry and ex-Bourbon cask maturation, a further model was built using just the B and S classes only (n = 27) (Figure 6.15d). Unlike in the equivalent NMR model, where BS and S classes were combined, here the BS samples were excluded. The resulting model had 2 orthogonal components with an $R^2\mathbf{X}$ of 0.65, $R^2\mathbf{Y}$ of 0.96, and a Q^2 of 0.80. Despite the small S class sample size, the model clearly discriminated between the two classes. Additionally, the ex-Sherry cask matured samples are clustered into two groups on the first orthogonal component (y-axis); the top cluster represents the same product from different years, the bottom four are two different products from two years. The ex-Bourbon cask matured samples are more spread across this component, representing larger variation in the larger sample size. The major variables contributing to this separation, as identified by the OPLS-DA model, are summarised in Table 6.6.

A bar plot for the relative abundances of the formulae summarised in Table 6.6 is shown in Figure 6.16, classified per maturation wood type, including the BS class. From this bar plot, it is seen that the samples matured in both ex-Bourbon and ex-Sherry casks generally show abundances of these compounds in between that of ex-Bourbon or ex-Sherry cask only matured whiskies. Based on known chemistry of Scotch Whisky, possible identities of some of the major discriminating formulae include mono- and disaccharides ($C_6H_{10}O_6$, $C_{12}H_{22}O_{11}$, $C_{12}H_{20}O_{10}$), and fatty acids or esters ($C_{10}H_{20}O_2$, $C_{12}H_{24}O_2$). Identities of two sulfur containing species, $C_{13}H_{20}O_8S$ and $C_{15}H_{24}O_8S$, are not known. However, they are discriminating factors as they are not found in any ex-Bourbon cask only matured whiskies. In contrast, $C_{12}H_{22}O_{11}$, a disaccharide, is almost absent in ex-Sherry cask matured whiskies, whilst noticeable in ex-Bourbon cask matured samples. B class

Table 6.6: Key variables (formulae) discriminating between Sherry and Bourbon barrel matured Scotch Whisky based on the OPLS-DA model. Possible identities are proposed based on formula and expected chemistries. "Carbohydrate" includes derivative or related compounds.

VARIABLE ID	POSSIBLE IDENTITY	DISCRIMINATES TYPE
$C_{30}H_{46}O_7$	Bartogenic acid	Sherry
$C_{14}H_6O_8$	Ellagic acid	Sherry
$C_{13}H_{20}O_8S$	Unknown	Sherry
$C_6H_{10}O_6$	Glucono delta-lactone	Sherry
$C_{15}H_{24}O_8S$	Unknown	Sherry
$C_{30}H_{46}O_8$	Unknown	Sherry
$C_7H_6O_5$	Gallic Acid	Sherry
$C_9H_{10}O_5$	Syringic Acid	Sherry
$C_{10}H_{20}O_2$	Decanoic Acid	Bourbon
$C_{12}H_{24}O_2$	Dodecanoic Acid	Bourbon
$C_{12}H_{22}O_{11}$	Disaccharide	Bourbon
$C_{16}H_{32}O_2$	Hexadecanoic Acid	Bourbon
$C_{16}H_{30}O_2$	Hexadec-9-enoic acid	Bourbon
$C_{12}H_{20}O_{10}$	Carbohydrate	Bourbon
$C_{14}H_{28}O_2$	Tetradecanoic acid	Bourbon

samples have a higher relative abundance of several O_2 class compounds, likely fatty acids or esters. $C_{30}H_{46}O_7$ is the highest relative abundance compound that clearly discriminates ex-Sherry from ex-Bourbon cask matured whiskies. Based on earlier tandem MS work (subsection 6.5.4), its identity is proposed to be bartogenic acid. Interestingly, its presence has not been reported in Scotch Whisky before, although it has in other oak matured spirits and wine.[89,249–251] These compounds, especially those with unknown structure, are worth future investigation for both structural elucidation and quantification, and for determination of their origins.

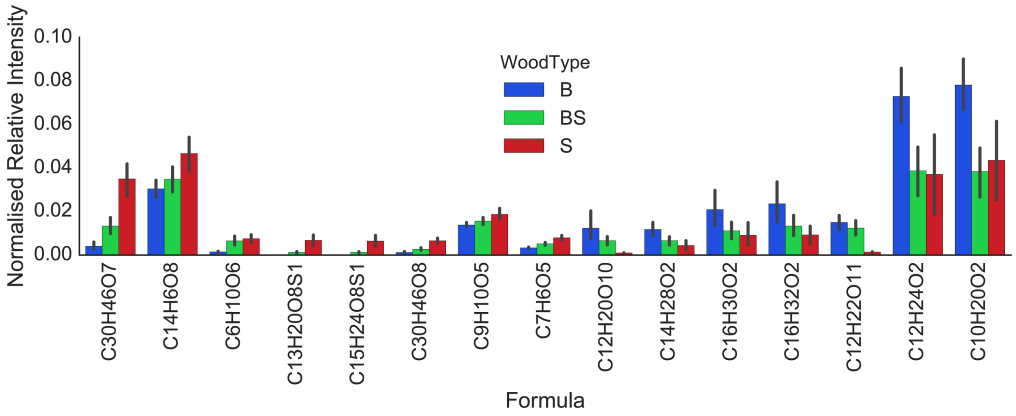


Figure 6.16: Bar plot of discriminating formulae against normalised relative intensity within three classes of samples – ex-Bourbon cask only matured (B, blue), ex-Bourbon and ex-Sherry cask matured (BS, green), and ex-Sherry cask only matured (S, red). Black lines indicated 95% confidence intervals. Formulae are ordered according to their location on the OPLS-DA S-Plot for B versus S class.

6.6.2.1 *Product Consistency*

Based on the fact three samples of the same product - but from different years of production - clustered so tightly in the PCA and OPLS-DA scores plots (Figure 6.15a,b), closer inspection of these samples was performed. Figure 6.17 shows the mass spectra, van Krevelen and DBE plots for the three samples.

In Figure 6.17 the mass spectral data are strikingly similar, with the same dominant peaks corresponding to $C_{14}H_6O_8$ and $C_{30}H_{46}O_7$. Additionally, two species $C_{10}H_{20}O_2$ and $C_{12}H_{24}O_2$ occur with similar relative intensity and above average intensity. Van Krevelen and DBE versus carbon number plots further reflect the similarity of these samples. The consistency of these samples was quantified by calculating the median relative standard deviation for the normalised relative abundance of each formula. This coefficient of variation was only 6.3 %. In comparison, the same value for the entire sample set was 25 %. Furthermore, the three

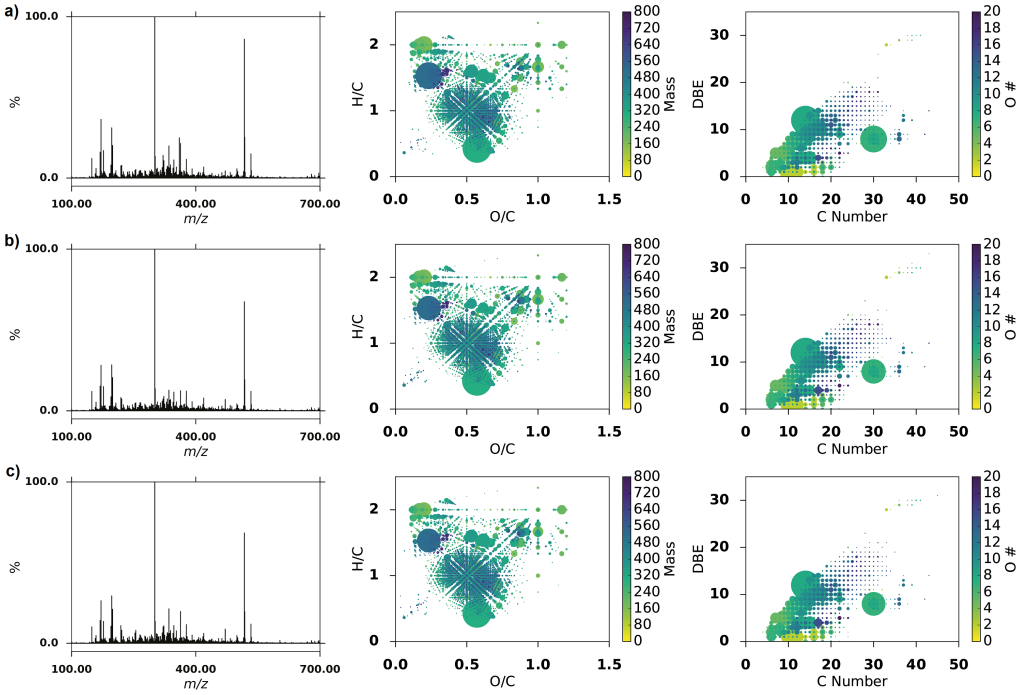


Figure 6.17: Samples of the same Scotch Whisky product from three years, a) S10-1218, b) S12-1293, c) S14-1941, a 12-year old ex-Sherry cask matured Highland single malt. The left column shows the broadband mass spectra for each sample, the middle column shows a van Krevelen diagram coloured according to mass and sized according to normalised relative abundance, the right column shows a plot of carbon number versus double bond equivalent coloured according to oxygen number and sized according to normalised relative abundance.

products had 2826 unique formula across all three samples with 2208 (78 %) of those common to all three - recall the entire sample set had only 407 (9.5 %) common across 4271 unique formula. Therefore, the ESI FTICR MS profile of this product remains consistent across three years of production years. The sample set analysed here contained two more products produced across three years. Again, these samples also displayed high consistency across years.

It is somewhat remarkable that the chemical complexity is consistent between production years, even though Scotch Whisky is made from naturally variable raw materials (barley, yeast, water, oak) across different seasons. This level of consistency is achieved through blending of casks to produce the desired product. This work, undertaken by the Master Blender, produces a chemical consistency across thousands of compounds using just sensory perceptions - primarily smell and taste. The sensory profile is inherently based in chemistry, however the key aromas of whisky were thought to be related to around 80 volatile aroma compounds.[254,255] Here, the observed similarities cross thousands of compounds,

suggesting a correlation between sensory profile and a far broader range of compounds than previously known.

6.7 CONCLUSIONS

The use of negative mode ESI coupled to high resolution FTICR MS has revealed the chemical complexity of Scotch Whisky. Here, thousands of unique molecular formula were confidently assigned across dozens of samples. The mass spectra were assigned CHOS formula based on exact mass with parts-per-billion mass accuracy, and isotopic fine structure successfully confirmed many assignments, including those of sulfur containing species not previously reported in the literature.

Comparison of mature whisky samples to a malt NMS confirms that the complexity of Scotch Whisky, as observed by FTICR MS, arises through maturation. Several ions of interest, including $C_{30}H_{46}O_7$ and $C_{14}H_6O_8$, were isolated in this complex mixture by means of quadrupole isolation and in cell isolation, and CID was used to fragment ions to identify structures. Based on comparisons with literature and reference standards, identites of bartogenic acid ($C_{30}H_{46}O_7$) and ellagic acid ($C_{14}H_6O_8$) were proposed. Di- and trisaccharides were also investigated, but their fragmentation patterns were inconclusive.

Visualisation of the data by means of heteroatomic class distributions, van Krevelen diagrams and DBE versus carbon number plots allowed for more straightforward comparison of samples. Chemometrics, PCA and OPLS-DA, identified potential marker compounds for discriminating whisky types (blend and malt), as well as maturation wood types (Bourbon or Sherry casks). The combination of both MS and NMR datasets for chemometric study holds promise, and is worth future investigation. Algorithms for multi-block (use of multiple input datasets) PCA and PLS-DA were presented by Worley and Powers,[238,256] and ESI-MS and NMR metabolomics data have been combined for statistical analysis.[257]

The structures, quantities, and specific origin of many of the compounds in Scotch Whisky remain unknown. Determination of structures and quantification will likely require extensive purification via chromatography and NMR characterisation. This time consuming process will need specific compounds of interest identified, likely through further experimentation with specific sample sets and questions. The origins of maturation related compounds - which are the major source of complexity - will require analysis of wood extractions and time series sample sets; however, these sample sets are challenging to acquire due to the time scales involved in Scotch Whisky production.

7

FTICR MS METHOD DEVELOPMENT

Mais apud me omnia fiunt Mathematicè in Natura
but in my opinion, all things in nature occur mathematically

— René Descartes (1640)

7.1 DECLARATION

Most of work in this chapter was previously published in the following papers:

W. Kew, J. W. T. Blackburn, D. J. Clarke, and D. Uhrín, *Rapid Communi. Mass Spectrom.*, 2017, **31**, 658-662

W. Kew, C. L. Mackay, I. Goodall, D. J. Clarke, and D. Uhrín, *Anal. Chem.*, 2018, **90**, 11265-11272

7.2 INTRODUCTION

With the successful use of negative mode electrospray ionisation (ESI) FTICR MS for the analysis of a large sample set of Scotch Whisky in chapter 6, a number of possible follow up questions arise. For example, how important was the use of ESI, and would other ionisation sources such as atmospheric pressure photoionisation (APPI) and atmospheric pressure chemical ionisation (APCI) have provided substantially different information? Contemporaneous work within the group investigated matrix free laser desorption ionisation (LDI) as an ionisation source for NOM,[156] using just the natural organic complexity as a 'matrix' - would this work for Scotch Whisky, too? Negative mode ionisation was used, but could positive mode ionisation provide complementary information? Trial-and-error, and visual inspection of the data, determined an optimal sample concentration for ESI analysis - but how important would sample concentration be for the final spectral quality?

Equally, as understanding and experience with the instrumentation developed, questions were raised about how to best optimise acquisition and processing parameters. Magnitude mode spectra are ubiquitous in FTICR MS, but phasing spectra would provide substantial resolution improvements. Likewise, with better understanding of how to improve spectral acquisition conditions, how far can

the resolution be pushed? What are the limitations associated with ultra high resolution data acquisition? And ultimately, how high a resolution is required for analysis of Scotch Whisky?

This chapter deals with the above topics and other questions associated with improving FTICR methodology for the acquisition of complex mixture, specifically Scotch Whisky, mass spectra. This chapter also discusses formula assignment tools and software developed as part of this PhD for analysis of FTICR MS data.

7.3 SAMPLES

Four Scotch Whisky samples were selected from the 2014 standard sample set for analysis in this chapter. The four samples represent a diverse range of Scotch Whisky types - including a Sherry cask matured Single Malt (S14-1941), a Blended Scotch Whisky (S14-1944), a peated Single Malt of unknown maturation wood (S14-1962), and a non-peated Bourbon cask matured Single Malt (S14-2196). Based on their provenance and their ESI spectra (see chapter 6), this sample set is expected to contain a diverse, but common, range of chemistries typically observed in Scotch Whisky. In chapter 6, 2955 unique monoisotopic formulae were assigned for these four samples in total, with only around a third common to all four samples.

7.4 NEGATIVE MODE IONISATION

Initially, both positive and negative ionisation modes were investigated. For both polarities, the same ionisation sources and equivalent analyser conditions were used, yielding data of comparable quality. However, the acquired positive mode data contained not just $[M+H]^+$ ions but also adducts of the form $[M+Na]^+$ and $[M+K]^+$. Negative mode spectra were exclusively assigned as $[M-H]^-$ or $[M-e]^{-\bullet}$. In positive ion mode, the multiple ionisation mechanisms resulted in increased spectral complexity, which made unambiguous formula assignment impossible at this resolving power. For a fuller discussion of this issue, see section 7.5. Based on this, only negative mode data are analysed and presented further. Calibration lists are included in Appendix C.

7.4.1 *Ionisation Source Optimisation*

Each of the four ionisation sources investigated required a degree of optimisation. However, to keep the experiment from becoming excessively complex, acquisi-

tion was kept as simple as possible. That is, no chemical modifiers, dopants, or matrices were added to any samples, and (for solution state modes), only simple dilution with high quality solvents - where necessary - was used. Further optimisation of each technique could be investigated. For example, Ruddy *et al.* recently published efforts to investigate the effects of ion suppression in positive ESI of petroleum complex mixtures.[258] In another publication, the use of a strong base (tetramethyl ammonium hydroxide) was investigated to aid negative mode ESI of complex mixtures.[259] However, to investigate all possible optimisations and aspects of each ionisation source in both polarities for Scotch Whisky would be a substantial undertaking. The purpose of this experiment, therefore, was to investigate the information yielded for a select set of Scotch Whisky samples by four different ionisation techniques using a simple approach, avoiding the use of chemical modifiers or dopants.

7.4.1.1 *ESI*

As presented in chapter 6, ESI proved to be a sensitive technique, necessitating sample dilution. Initial experimention with a 1:100 dilution was somewhat successful, requiring only 3 μL of sample per analysis. However, residual instrument or solvent contamination, especially by fatty acids, was observed at this low level of sample addition. Consequently, increasing concentration to a 1:10 dilution increased genuine signal whilst relatively reducing contamination signal. Similarly, initial experimentation had investigated use of NH_4OH as a modifier for ionisation. This yielded no significant improvement, whilst increasing risk of sample contamination. Consequently, samples for ESI require minimal volumes (10 μL to 20 μL); a factor that may be important for limited samples, such as historical or rare materials.[17] Final sample preparation entailed 1:10 dilution into a 50:50 methanol water solution.

N.B. contamination - of samples, solvents, or instrument - is an important factor to be aware of when using such highly sensitive equipment. Careful selection and testing of solvents, pipette tips, glassware, etc. was performed to ensure the spectra represent the sample. Unfortunately, instrument contamination, and variability, is more difficult to protect against. Inter-day (week, or month) reproducibility has not been addressed for complex mixture analysis by FTICR MS. As such, all samples to be compared are run on the same day, or consecutive days. Quality controls and solvent blanks are also run to provide a measure of instrument state. Note that solvent blanks in ESI are never devoid of signals, however the magnitude and number of peaks in a solvent blank can be used to infer contamination levels. The complexity of ESI spectra, 'fact or artifact', and their interpretation in NOM studies was discussed by Novotny *et al.* in 2014.[152]

7.4.1.2 *APPI*

APPI required significantly greater infusion rates and concentrations than ESI; a typical analysis required 150 μL to 250 μL of sample. APPI requires a component of the sample to be ionisable below 10.6 eV, which is the power of the source lamp. To achieve this, toluene (ionisation potential 8.8 eV)[260] is typically added as a modifier. Water and ethanol (most of a Scotch Whisky sample by volume) have ionisation potentials much higher than toluene (12.7 and 10.47 eV respectively).[261,262] Nevertheless, APPI was successfully applied to ionise samples of mature Scotch Whisky with no prior sample dilution and without addition of toluene, though requiring higher flow rates (≈33 μL/min).

It should be noted that below a sample-specific flow rate threshold, no signal was obtained. In this limited sample set, it was observed that the 'darkest' (in appearance) samples performed best - yielding the most signal and ions, whilst the lighter coloured samples required the highest flow rates or even manual injection into the source to acquire any signal. Despite the ionisation energy of ethanol being approximately equal to the power of the APPI lamp, it is perhaps more probable that there is a critical concentration of a compound, or compounds, in Scotch Whisky which ionise directly in APPI, and then charge transfer to ionise other species. This trend of darkest samples ionising more readily mirrored the findings with LDI. Experimentation with new make spirit samples, which are inherently colourless, failed to yield any signal for APPI. This, combined with the flow rate and concentration requirement, further supports the theory that it is minor compounds in Scotch Whisky which ionise in APPI, and it is not reliant on the solvent. Note that, similarly to colourless new make spirit, APPI solvent blanks do provide blank spectra. Addition of toluene may allow use of APPI with colourless distilled spirits, however the selectivity of ionisation would no longer exist.

7.4.1.3 *APCI*

APCI required flow rates greater than those used for ESI (6 μL/min), but lower than needed for APPI. Additionally, more concentrated samples were required for APCI than ESI; diluted 1:1 with methanol water, some 50 μL to 150 μL of sample was required per APCI analysis. Interestingly, APCI requires higher purity nebuliser gas as it transpired APCI ionises any residual hydrocarbon impurities present in the in-house generated nitrogen. A negative control (*i.e.* pure solvent) with in-house generated nitrogen gas confirmed the need to use a dedicated high purity nitrogen cylinder to acquire APCI spectra. Similar tests revealed no such requirement for any of the other ionisation sources. The dry-gas, sur-

rounding the glass capillary, remained the in-house nitrogen, but this did not appear to present the same issues.

7.4.1.4 *LDI*

Finally, LDI mass spectra were readily obtained after some method optimisation. Sample volumes required were at their lowest for LDI: 1 μL samples (neat) were spotted onto the MALDI plate and allowed to air dry for several minutes. Depending on the sample, it was often necessary to re-spot the sample several times to increase concentration. Analogously with APPI, it was found that samples lighter in colour required more repeat spots to build up sufficient concentration. Due to unstable scan-to-scan ion count, the use of a TIC threshold filter prior to summing transients was required. Without this filter the SNR suffered, and peak positions varied, causing peak splitting and broadening. The use of this filter typically added a two-fold time penalty to acquisition, further increasing when more scans were dropped. This time penalty appeared to be due to dead time as the acquisition computer Fourier transformed and processed each transient to determine if it should be co-added. This process took approximately as long as each scan, and the process was serialised, impeding acquisition speed. To acquire signal, a high laser power (up to 100 %) and narrow beam focus were required. No ions were observed below such laser settings. There is a risk of fragmentation associated with the higher laser powers utilised. However, previously published work from the group, comparing MALDI - a 'soft' ionisation technique - with matrix-free LDI of complex mixtures found no evidence for fragmentation.[156]

7.4.2 *Spectrum Quality*

As in chapter 6, all acquired spectra were of high quality, with a resolving power of *ca.* 300,000 at m/z 400 and a median SNR at m/z 401 of between 16 (LDI) and 35 (APPI), with ESI and APCI between these values. As acquisition settings were similar and digitizer settings identical, all spectra have comparable resolution and noise levels. Across the four whisky samples, there were on average between 895 (APPI) and 4111 (LDI) peaks picked. Of these, between 75 % (LDI) and 95 % (APPI) were assigned chemical formula within the error thresholds. Formula assignment are further discussed in section 7.8 (methodology) and subsection 7.4.3 (results). The numbers are summarised in Table 7.1.

For all ionisation modes the spectra are dominated by a small number of abundant peaks, with most of the complexity at much lower relative abundance. For example, in APCI the most abundant peak corresponds to m/z 255.23295 ($[C_{16}H_{31}O_2]^-$,

Table 7.1: Statistics for mass spectra across four samples and four ionisation sources in negative mode. Columns show the mode, total number of peaks detected (# Peaks), number of monoisotopic assignments (# Mono.), percentage of monoisotopic radical anions assigned (% Rad.), total percentage of formula assigned, including isotopologues (% Assign.), and the mean error of assignment in ppb (Error (ppb)). *Mean values across four samples. **Mean values across all four samples and ionisation sources

MODE / SAMPLE	# PEAKS	# MONO.	% RAD.	% ASSIGN.	ERROR (ppm)
APCI*	2795	2070	30.9 %	89.9 %	0.06
S14-1941	3980	2805	34.7 %	84.5 %	0.07
S14-1944	1549	1342	27.9 %	96.0 %	0.05
S14-1962	4347	3000	36.0 %	83.4 %	0.06
S14-2196	1305	1131	24.8 %	95.9 %	0.05
APPI*	895	777	10.2 %	95.3 %	0.04
S14-1941	1295	1080	18.4 %	93.7 %	0.05
S14-1944	531	499	2.4 %	96.8 %	0.03
S14-1962	717	654	6.0 %	96.5 %	0.04
S14-2196	1038	875	13.9 %	94.0 %	0.04
ESI*	1254	875	0.8 %	82.9 %	0.04
S14-1941	1714	1068	0.9 %	76.0 %	0.05
S14-1944	757	626	0.5 %	88.4 %	0.03
S14-1962	1581	1066	0.8 %	81.3 %	0.04
S14-2196	963	740	1.1 %	85.9 %	0.03
LDI*	4111	2664	19.8 %	75.4 %	0.08
S14-1941	3696	2678	22.6 %	82.6 %	0.08
S14-1944	4626	2604	18.6 %	66.1 %	0.09
S14-1962	4254	2866	19.1 %	77.3 %	0.08
S14-2196	3869	2506	18.9 %	75.4 %	0.08
Total**	2264	1596	15.4 %	85.9 %	0.05

0.01 ppm), likely a fatty acid such as palmitic acid. As can be seen in the Figure 7.1b, there are a multitude of peaks at each odd m/z, with much smaller peaks at each even m/z corresponding to CHO species and their ^{13}C-containing isotopologues, respectively. This trend is repeated for all ionisation modes, nevertheless, the peak abundance at even m/z varies, *e.g.* more abundant peaks were seen in the APCI and LDI spectra than in the spectra acquired by the other two methods. Many of the signals at even m/z were assigned as radical anions, especially in APCI and LDI. Furthermore, the lack of peaks between each nominal mass confirms that all species are singly charged. These observations agree with previous findings for negative mode ESI spectra of Scotch Whisky and other complex mixtures,[54,61,156] including those in the last chapter (chapter 6).

In contrast to the recent work of Roullier-Gall *et al.*,[61] nitrogen containing species were neither detected nor assigned in the ESI, nor any other ionisation source, mass spectra of Scotch Whisky. A selection of their reported nitrogen compounds were searched for in the spectra of Scotch Whisky and the vast majority were either not detected, present but below the noise threshold for peak picking, and / or had the chemical formula of a surfactant, and thus were likely contaminants. Low molecular weight nitrogen species have also been reported in Scotch Whisky by *p*-H_2 hyperpolarised NMR,[19] but have not been observed here.

Example mass spectra for one whisky sample at three levels of expansion (m/z 100-1000, 325-335, 333-333.2) are shown in Figure 7.1.

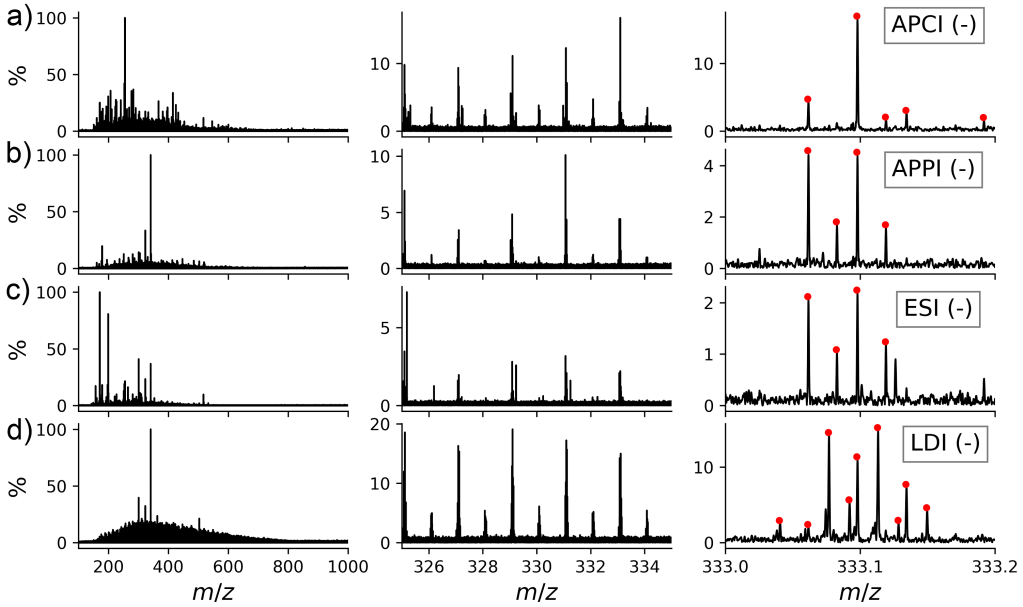


Figure 7.1: Negative FTICR mass spectra at three expansions (left-to-right) m/z 100-1000, m/z 325-335, m/z 333-333.2, for one Scotch Whisky sample (S14-1944), acquired with four ionisation sources (top-to-bottom) a) APCI, b) APPI, c) ESI, d) LDI. Assignments summarised in Table 7.2.

The spectrum acquired using APPI is dominated by a species at m/z 341.10893 ($[C_{12}H_{21}O_{11}]^-$, 0.03 ppm), which corresponds to the molecular formula of a disaccharide. This compound also dominates the LDI spectrum. Neither APPI nor LDI contain fatty acid-type signals of any significant abundance, suggesting that these ionisation sources may ionise different classes of compounds, such as polyphenols and lignin derivatives.

The spectrum obtained using ESI is consistent with previously published data,[54,61] and with the analysis of a larger sample set by ESI presented in chapter 6. Here, it is dominated by two large signals at m/z 171.13905 and 199.17035 ($[C_{10}H_{19}O_2]^-$, 0.01 ppm and $[C_{12}H_{24}O_2]^-$, 0.01 ppm), corresponding to two fatty acids. The next most abundant signals are at m/z 300.99900 and 341.10893 ($[C_{14}H_5O_8]^-$, 0.03 ppm and $[C_{12}H_{21}O_{11}]^-$, 0.03 ppm), corresponding to ellagic acid (a known cask extractive), and a disaccharide (as observed in the APPI data).

The LDI spectrum was the most distinct amongst the four ionisation sources. Aside from the previously mentioned disaccharide (m/z 341.10893), and ellagic acid (m/z 300.99900), the spectrum has a skewed normal distribution of signal intensities. This type of mass spectrum profile is more typically observed for NOM.[156,263] Given the compounds present in Scotch Whisky are known to exist at a wide range of concentrations, this relatively uniform distribution of abundances suggests an even less quantitative ionisation mechanism for LDI than ESI. Of all the ionisation techniques used, the LDI spectrum had the highest relative signal intensity and the most peaks at this single nominal mass (see Figure 7.1). This was due to LDI presenting a more normally distributed profile of ion intensities; the other ionisation sources were not normally distributed.

The final expansions of the spectra within 0.3 mass units at m/z 333 are shown in Figure 7.1c; Table 7.2 contains molecular formulae identified by all ionisation techniques in this region. In the APCI spectrum all five peaks have been assigned monoisotopic molecular formulae. The APPI expansion has a similar appearance to the APCI spectrum with three out of four assigned peaks common with the APCI data. The ESI spectrum displays six peaks, however only four were assigned – these were the same species as assigned in the APPI spectrum. LDI showed the largest number - nine peaks were observed and assigned. Only two of these were common with the other three spectra, and a third was common with a formula identified in the APCI spectrum.

Table 7.2: Select formula assignments for a Scotch Whisky (S14-1944) and mass errors for APCI, APPI, ESI, and LDI at m/z 333. Crosses in the final four columns indicate a peak that was assigned in the given spectrum. See Figure 7.1

m/z	MOLECULAR FORMULA*	ERROR (ppm)	APCI	APPI	ESI	LDI
333.04049	$C_{19}H_{10}O_6$	-0.09				X
333.06160	$C_{16}H_{14}O_8$	-0.02	X	X	X	X
333.07685	$C_{20}H_{14}O_5$	-0.02				X
333.08273	$C_{13}H_{18}O_{10}$	-0.01		X	X	
333.09215	$C_{24}H_{14}O_2$	-0.13				X
333.09798	$C_{17}H_{18}O_7$	-0.02	X	X	X	X
333.11327	$C_{21}H_{18}O_4$	-0.11				X
333.11911	$C_{14}H_{22}O_9$	-0.02	X	X	X	
333.12848	$C_{25}H_{18}O$	0.04				X
333.13436	$C_{18}H_{22}O_6$	0.00	X			X
333.14966	$C_{22}H_{22}O_3$	-0.13				X
333.19187	$C_{16}H_{30}O_7$	0.01	X			

7.4.3 *Formula Assignments*

All 16 spectra were assigned simultaneously using *Formularity*,[207] with peak alignment prior to formula assignment. Again, Table 7.1 summarises the number of peaks detected and formula assignments for each spectrum.

All ionisation sources achieved greater than 75 % molecular formula assigned across the average of four whisky samples, which represent excellent assignment rate for MS analysis of complex mixtures. All assignments were made with a maximum error threshold of $\pm$0.25 ppm, and the average for all ionisation modes was considerably lower than that threshold (0.07 ppm). The average distributions of errors achieved using each ionisation technique are shown in Figure 7.2.

LDI has the broadest error distribution, with APPI and ESI having the narrowest error distributions overall. However, the distributions of errors appear largely normally distributed around 0 ppm for all ionisation modes, indicative of correctly calibrated data.[144] Overall, an average of 86 % of peaks were assigned a formula with a mean error of assignment of just 0.05 ppm. These statistics give high confidence in the assignment of data.

Across all 16 spectra acquired – including four samples and four ionisation modes – there were a total of 20809 monoisotopic molecular formula assigned. Accounting for the duplication of radical and deprotonated ions, 3993 were unique

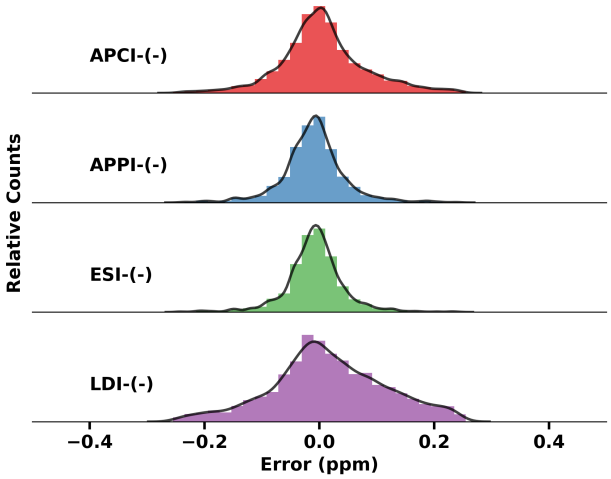


Figure 7.2: Histograms of formula assignment mass errors for each of the ionisation sources in negative mode. Y-axes scaled independently.

molecular formula. LDI and APCI had the greatest number of peaks and assignments, however APPI and ESI had better rates of assignment than LDI. APPI had a very high assignment success (95 %) with an average error of only 0.04 ppm.

The assignment of radical anions was necessary to comprehensively interpret the acquired data and to achieve high formula assignment rates, especially for APCI and LDI. As shown in Table 7.1, radical anions were assigned in each ionisation mode. Radical anions and the first isotopic peak of the same compounds were well resolved; an example is shown in Figure 7.3.

APCI had an average of 31 % of monoisotopic peaks assigned as radicals, with 20 % for LDI and 10 % for APPI. Fewer than 1 % of monoisotopic peaks were assigned as radicals for ESI. Across all four ionisation sources, these radical anions do not constitute a significantly different chemistry to those assigned as deprotonated ions. Table 7.3 shows the numbers of formula that were uniquely identified from radical anion assignments.

This was most significant for APCI, where 6 % of the unique molecular formula were only identified as radical anions, but even this is a relatively small number. A majority of compounds found as radical anions were also found as deprotonated ions. Across all assigned formula, radical ions had a median ion abundance of 64 % of that of their equivalent deprotonated ions. For completeness, all subsequent analysis is based on de-duplicated formula lists and reported ion abundances are the mean of the radical and deprotonated form when both forms were assigned. Clearly, however, radical assignments are not crucial to covering the chemical diversity of Scotch Whisky.

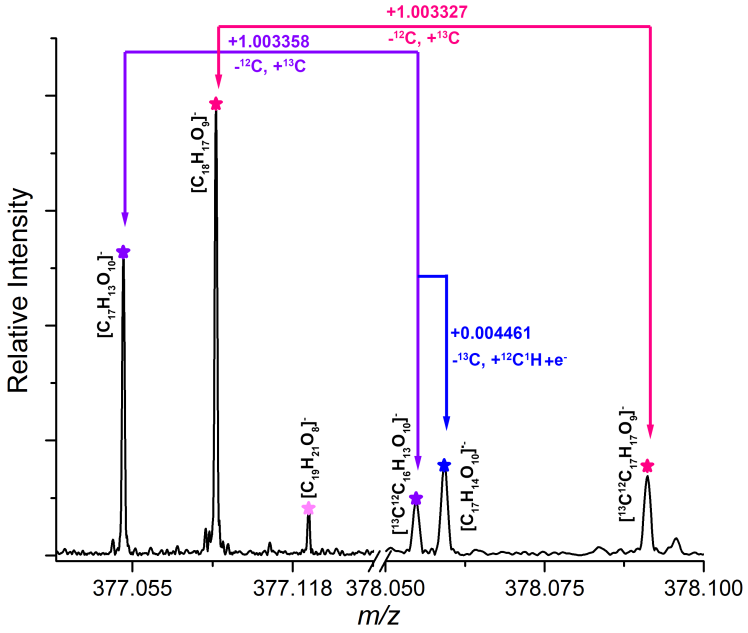


Figure 7.3: Region of APPI (-) mass spectrum for sample S14-1941 at m/z 377-378 showing several assignments, including one compound ($C_{17}H_{14}O_{10}$) ionised both as deprotonated and radical anion form. Mass differences have been annotated to aid inspection. Theoretical peak positions and intensities are plotted as coloured scatter marks with star glyphs.

Table 7.3: Number of new formula identified from assignment of radicals, total number of formula after de-duplication of radical anion and deprotonated ion assignments for the same molecule (total # de-duplicated), and the resulting % new formula from assignment of radicals. Numbers shown as mean for each ionisation mode across four Scotch Whisky samples, and as mean across all ionisation sources.

SOURCE	# NEW FROM RADICALS	TOTAL # DE-DUPLICATED.	% NEW FROM RADICALS
APCI	91	1483	6.1 %
APPI	2	686	0.3 %
ESI	3	871	0.3 %
LDI	27	2163	1.3 %
Mean	31	1301	2.4 %

7.4.4 *Formula Intersections*

To investigate the intersections of formula assigned between each ionisation mode, *UpSet* plots were utilised.[264] It is generally not possible to accurately plot scaled three set Venn diagrams using circles, and impossible to even approximate this

for four sets.[264,265] Instead, the UpSet plots are an attractive alternative to represent the size of intersections across many sets, as they display the exclusive intersections of sets of data. The UpSet plot shown in Figure 7.4, presents a cumulative account of individual ionisation methods across all four samples.

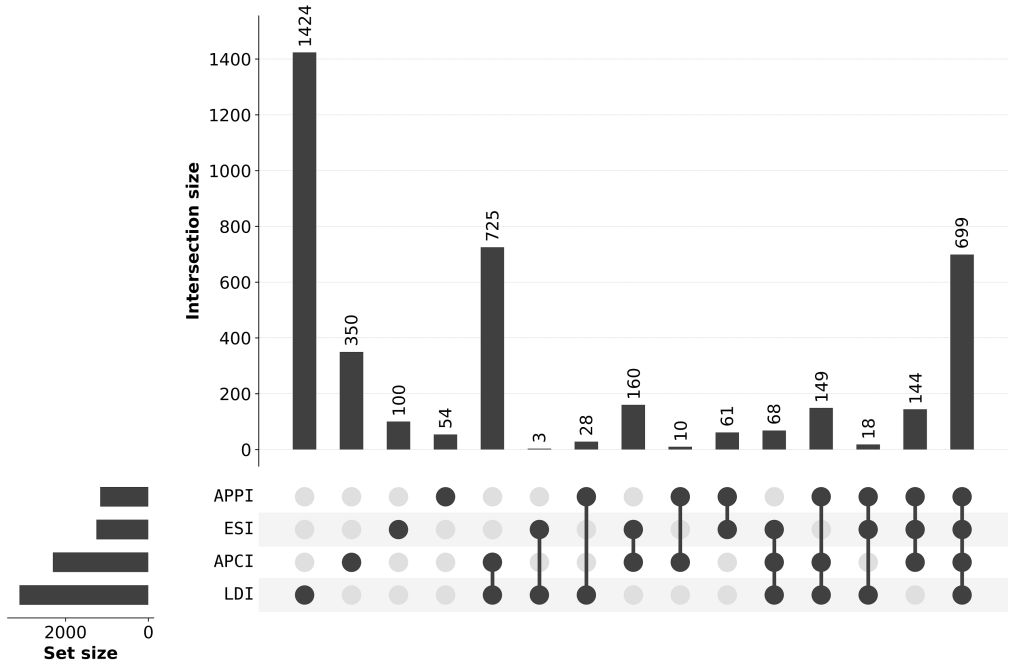


Figure 7.4: UpSet plots showing intersections of each ionisation source. Set size on the left indicates number of unique formulae for each source across the four whisky samples. Vertical bar plots indicate the number of formulae found in each exclusive intersection.

LDI has the most unique formulae assigned (Figure 7.4), whereas APPI provides the least (1424 compared to 54 formulae). APCI and LDI had the greatest number of common formula in their pairwise intersection (725), and overall there were 699 formulae common to all negative mode ionisation sources. Pairwise, ESI and LDI are least similar, which is to be expected due to their fundamentally different ionisation mechanisms. Similar results were previously obtained for SRFA.[156] The presented UpSet plot clearly highlight the complementary information provided by the four ionisation sources in negative mode.

Additional comparison can be made on the level of individual samples. The distribution of the number of times formula were assigned across the 16 spectra is shown in Figure 7.5a. There were 663 unique formulae assigned in only one spectrum - over 83 % of formula were identified in at least two spectra. Only 178 of the unique formulae were identified in all 16 spectra.

The distribution of unique formula across spectra for each sample is shown in Figure 7.5b. This figure shows that for each whisky sample, most formula ap-

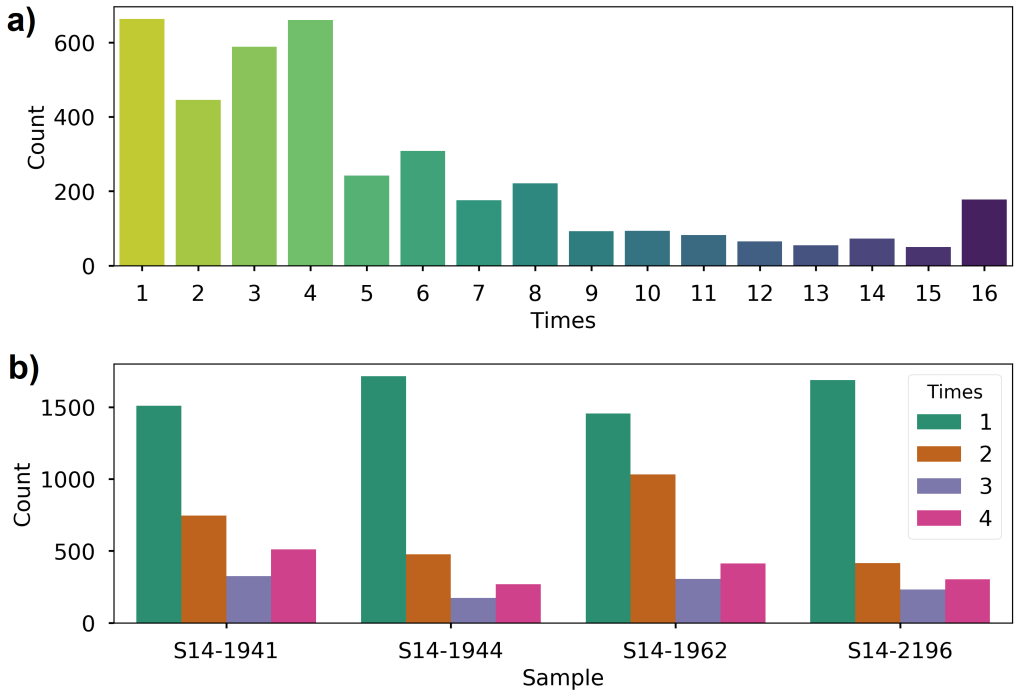


Figure 7.5: Formula occurence charts across samples and ionisation sources. a) Number of times each unique formula (n=3993) appears across the 16 acquired spectra, b) number of times each unique formula appears across four ionisation modes for individual whisky samples.

pear in only one ionisation source. Interestingly, samples S14-1941 and S14-1962 appear to have far more formula common to two or more sources than the other two samples. The exact cause for this is unknown, as there are many variables about the samples and class of compounds ionised. These results highlight that there is significant chemical diversity between the whisky samples and between the ionisation sources.

7.4.5 *Chemical Diversity*

Whilst formula intersections give some information about chemical diversity of the selected four Scotch Whisky samples as revealed by the different ionisation modes, complex mixtures are rarely analysed on an individual formula level. Instead, it is more typical to investigate the distribution of heteroatomic classes or to visualise the data through van Krevelen diagrams and DBE versus carbon number plots.

The heteroatomic, or oxygen, class distributions for the four ionisation modes extend to O_{19}, as shown in Figure 7.6. Formula were assigned above this class, up to O_{25}, however only 34 formulae out of 20809 were assigned above O_{19} and

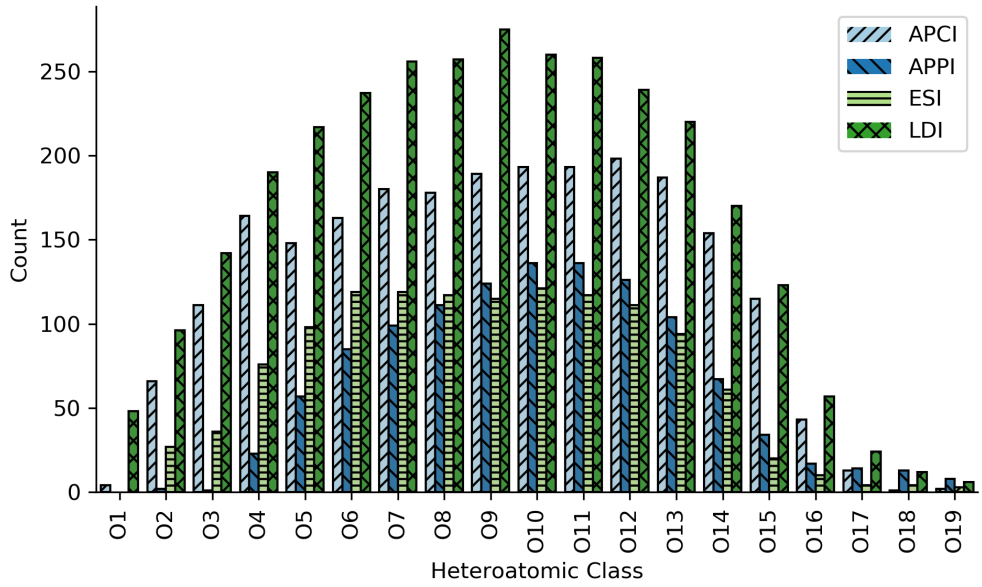


Figure 7.6: Heteroatomic class histogram for ionisation modes. Y-Axis is the count of formula containing the given number of oxygens. Each unique formula within each ionisation mode across all samples is counted only once, *e.g.* if the same formula appears in more than one sample in a given ionisation source, it only counts once.

the figure was truncated accordingly. Similarly, seven CH only formulae were assigned, however these were also excluded from this figure.

LDI clearly contains the most formula, and their class distribution extends from O_1 upwards. LDI, along with APPI, present a normal distribution of oxygen classes centred around O_{10} and O_{11}, respectively. APPI has very few formulae identified below O_4. It appears to be preferably ionising oxygen rich aromatic compounds, *e.g.* lignin derived compounds. APCI, like LDI, has formula extending from O_1. APCI presents a bimodal distribution around O_4 and O_{12}. Finally, ESI presents an almost bimodal distribution around O_6 and O_{11}, perhaps indicative of mono- and disaccharides, and polyphenolic glycosides. Counts for oxygen classes above O_{15} drop off significantly for all ionisation modes. This is in part due to instrumental settings, particularly the optimised mass range during analysis. It is interesting that the distribution of oxygen classes is similar across different ionisation modes (ESI, LDI and APPI), and that APCI is the most unusual, with a local maximum at O_4. LDI also shows a higher abundance of compounds below the O_4 class, as seen previously in the LDI spectra of SRFA.[156]

The van Krevelen diagrams provide a simple representation of the CHO chemical space as reported by the mass spectra. This is shown for all ionisation sources in Figure 7.7 as a cumulative plot for all four whiskies allowing their conveni-

ent visual comparison. As with those presented in chapter 6, the van Krevelen diagrams are a scatter plot, where each glyph (scatter marker) represents an assigned formula in terms of its oxygen-to-carbon ratio (O/C, x-axis) and hydrogen-to-carbon ratio (H/C, y-axis). Here, the van Krevelen (and DBE plots) show the median signal abundance across the four samples for each ionisation mode. For further discussion on the generation of these plots, see section 7.7.

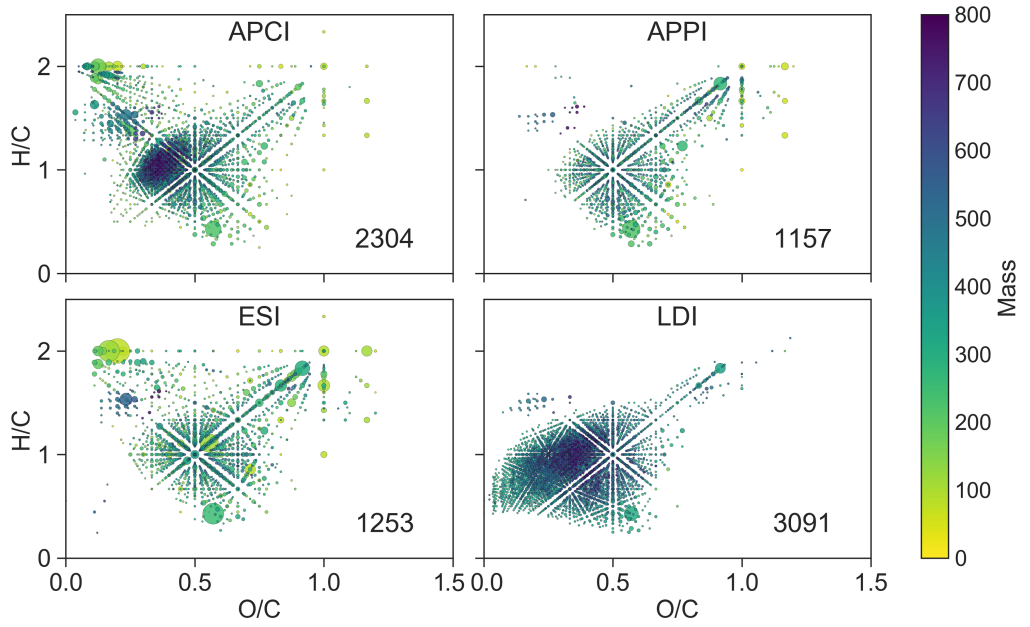


Figure 7.7: Van Krevelen diagrams for each ionisation source. Each assignment from individual ionisation modes was plotted with the size of the glyph representing the median abundance across the four whisky samples. The colour represents the mass of the peak. The number in the bottom right corner shows the number of unique formula identified for each ionisation source across four samples.

The van Krevelen representation of the ESI data contains several lower mass species at high H/C ratio and low O/C ratio corresponding to fatty acids or alcohols. The top right region, high H/C and high O/C corresponds to carbohydrates, and the diagonal towards the centre likely indicates carbohydrate breakdown products or glycosidic lignin derivatives. The 'central star' region, at H/C 1 and O/C 0.5, is occupied by cask extractives, likely tannin and lignin derived compounds. The ESI spectra and van Krevelens for a much larger sample set are discussed in chapter 6.

ESI shares some similarities with APCI and APPI, with the APCI containing more of the low O/C species, while APPI is dominated by the central star-to-carbohydrate region. The central star region is likely to contain UV active compounds, which are preferentially ionised using APPI. As colour is an area of interest to the mature spirits industry, this is potentially a useful finding. APCI has

nearly twice as many assignments than ESI and APPI, however it does not appear to be covering a significantly larger chemical space. There are more species identified around H/C 1.4 and O/C 0.2 for the APCI, exemplified by the molecular formula $C_{30}H_{46}O_7$ and related compounds. These compounds are likely to be triterpenoids as detailed in chapter 6. Their increased number in the APCI spectra will be useful for the investigation of the wood extracts chemistry. LDI shows the greatest difference from ESI; it contains nearly three times as many formula, has no fatty acid region, and only a nominal amount of carbohydrate and trend line assignments. It is heavily dominated by species with O/C ratios below 0.5 and H/C ratios below 1.2. A similar trend was observed in comparisons of ESI and LDI spectra of SRFA.[156]

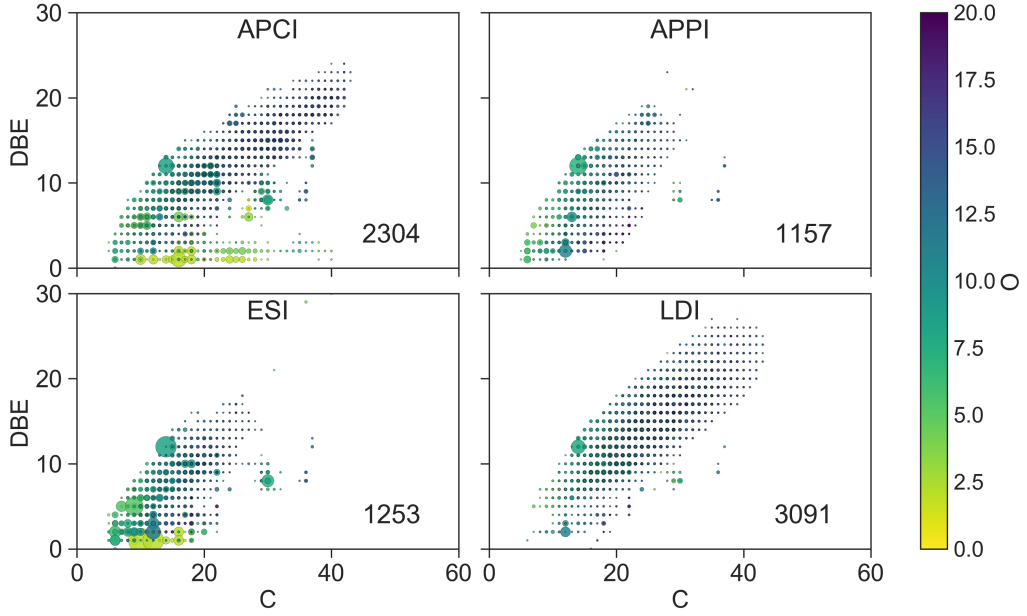


Figure 7.8: DBE versus C number plots for each ionisation source. Each assignment from individual ionisation modes was plotted with the size of the glyph representing the median abundance across the four whisky samples. Colour represents the oxygen count of the peak. The number in the bottom right corner shows the number of unique formula identified for each ionisation source across four samples.

The DBE versus carbon number plots are shown in Figure 7.8. These are styled similarly to the van Krevelens, except colour now reflects oxygen count as the DBE calculation does not factor in oxygen (see section 7.7). ESI plots share a similar profile to APPI, however the APPI data do not contain many of the low DBE and low oxygen species, *i.e.* the fatty acid compounds. Both APPI and ESI do not extend much beyond C_{30} and DBE of 18. There is some similarity between the APCI and ESI data, however APCI shows a large DBE and C number range, extending up to C_{40} and a DBE of 25. It includes higher carbon number compounds

with low DBE and low oxygen number, which may be longer chain unsaturated fatty acid-type species, as well as higher DBE species with higher oxygen number, which may be complex aromatic compounds. LDI, again, is the most different. Like APPI, it lacks the fatty acid species, however, it extends beyond C_{40} and up to DBE 28 in a largely normal distribution, reflecting observations made of the raw mass spectra. Again, the DBE analysis shows that there is an overlap between the ionisation techniques, but complementary data are to be obtained by each technique.

Further visualisations of the unique and common formula assignments were made to compare between the ionisation sources. A van Krevelen diagram for the 699 formulae common to all ionisation modes is shown in Figure 7.9. These are predominately wood extractive compounds, lignin derivatives and glycosides.

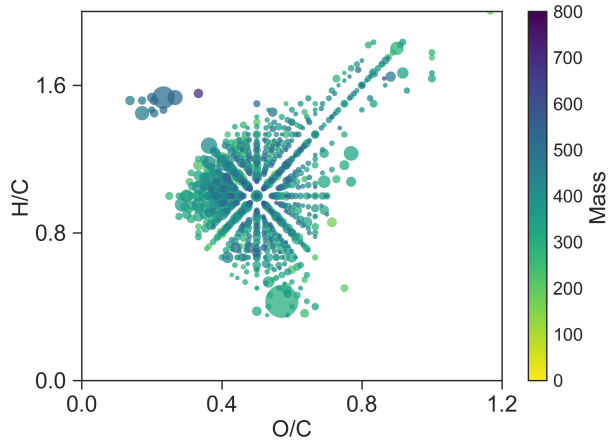
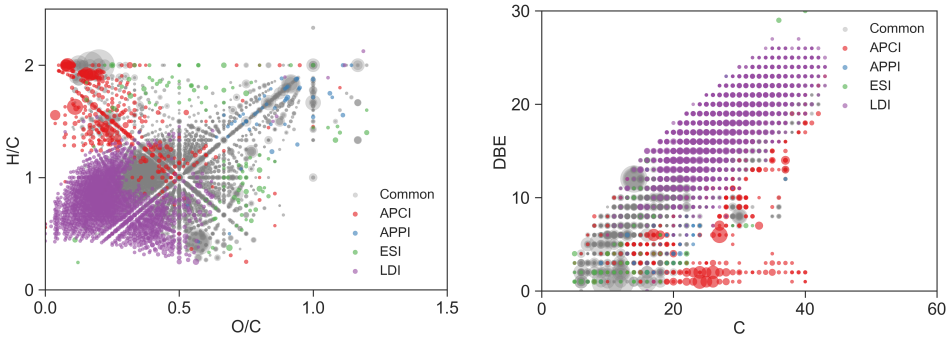


Figure 7.9: Van Krevelen diagram for 699 formulae common to all four ionisation techniques including all four samples.

Van Krevelen and DBE versus C number plots showing the unique formula for each ionisation mode are shown in Figure 7.10a) and b), respectively. These diagrams reaffirm the analysis presented above and show, *e.g.* that LDI and APCI produced most of the unique formulae, typically with a low O/C ratio (both techniques) and a H/C ratio below 1 (LDI) or between 1 and 2 (APCI). They also illustrate nicely the formula intersections.

Another metric that can be used to classify the compounds based on their molecular formulae is the aromaticity index (AI). AI categorises the compounds as non-aromatic, aromatic, and condensed aromatics. The modified AI was calculated for all assigned formulae. The modified AI is calculated with the number of oxygens halved. This estimates the number of σ-bound oxygen atoms and produces a more realistic assessment of the aromaticity index.[164] The relative percentages of formulae corresponding to these compound classes are shown in Figure 7.11.



(a) van Krevelen (b) DBE versus carbon number

Figure 7.10: Visualisation of the formula unique to each ionisation source and common to all sources

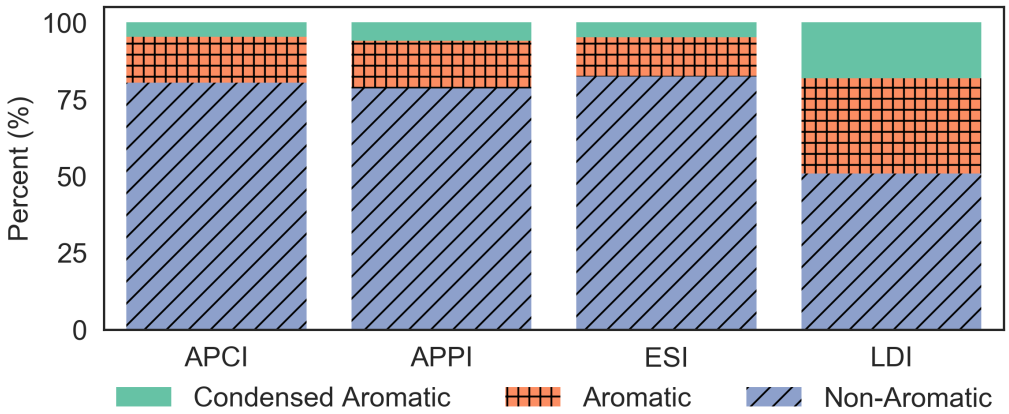


Figure 7.11: Modified Aromaticity Index (AImod) bar plot. Counts of formula have been normalised to 100 % individually for each ionisation technique to compensate for significantly more formulae identified in LDI and APCI than APPI and ESI.

AI calculates a numerical value between 0 and 1 for a chemical formula, taking into account numbers of C, H, O, and other elements (where appropriate).[164] Values of AI between 0 and 0.5 are classified as 'non-aromatic', 0.5 to 0.67 as 'aromatic', and above 0.67 as 'condensed aromatic'. It is important to note this describes the averaged overall molecule, which may contain both aromatic and aliphatic moieties. The relative AI distributions amongst the four ionisation sources confirmed the van Krevelen analysis - ESI is most similar to APCI and APPI in terms of relative numbers of aromatic and non-aromatic species identified. LDI is different, with nearly 50 % of assigned formula described as aromatic or condensed aromatic. This prevalence of aromatic or condensed aromatic compounds observed by LDI was even more pronounced (≈92 %) in the LDI spectra of SFRA.[156] A comparison of the corresponding van Krevelen diagrams showed,

that unlike in the SRFA sample, compounds with H/C $\geq$ 1 were present in significant numbers in the whisky samples. These compounds reflect the higher content of non-aromatic moieties in whisky samples which are ionisable by LDI.

7.5 POSITIVE MODE IONISATION

Positive mode data requires higher resolving powers for unambiguous and confident assignment. This is because in positive mode, adducted ions are more prevalent and probable. In provisional positive mode experimentation on Scotch Whisky, all four ionisation methods yielded protonated ions, $[M+H]^+$. However, ESI also yielded confirmed sodium adducts, $[M+Na]^+$, and LDI also generated potassium adducts, $[M+K]^+$. All four ionisation sources thus yielded many peaks, and whilst it is possible to confirm a subset of assignments via IFS, it was not possible to acquire broadband data with high enough data quality across multiple samples to yield IFS for formula assignment. Furthermore, there does not yet exist software to perform broadband formula assignment using IFS.

To confirm the limitations of the instrumentation presently in use, calculations were performed to determine what possible mass differences could occur in complex mixtures of CHO species. All possible, chemically logical, monoisotopic formulae were generated using in-house developed scripts. See section 7.7 for more detail. Possible formulae within 1 mass unit at m/z 401 and 799 in both positive and negative mode were investigated as an example. Positive species were calculated as the protonated form, or as the sodium or potassium adducts. Negative mode species were calculated as purely deprotonated, as no adducted formula have been observed in the mass spectra of Scotch Whisky.

Table 7.4: Theoretical complexity of CHO spectra in positive and negative ionisation modes, showing - for two values of m/z - the number of possible species (# CHO), minimum distances between them, and formula change required for such a mass difference. Smallest gaps (min. δmass) are shown in both absolute units (mDa) and relative units (ppm).

POLARITY	MASS (m/z)	# CHO	MIN. δ mDa (ppm)	FORMULA CHANGE
Positive	401	164	0.152 (0.38)	$-C_3$ -Na $+H_4$ +O +K
Positive	799	621	0.152 (0.19)	$-C_3$ -Na $+H_4$ +O +K
Negative	401	60	3.5 (8.8)	$-H_8$ $-O_{13}$ $+C_{18}$
Negative	799	211	2.37 (2.97)	$-C_{25}$ $+H_{12}$ $+O_{18}$

Note that radical anions have been confirmed in negative mode, but these are, obviously, significantly heavier than the deprotonated form - mass difference of $^{1}H^{+}$ and e^{-} is 1.007 276 Da. Therefore, the monoisotopic ions (and thus most significant peaks) would not overlap or reduce the mass gaps indicated in Table 7.4.

It is clear, and inherently obvious, that as mass increases the number of chemically logical, monoisotopic formula increases as well. The relative (ppm) gaps between species decreases as mass increases. The absolute, mDa, smallest gap does not change between the two nominal masses in the positive mode as the smallest gap is due to the same molecular change ($-C_3Na$ $+OH_4K$). In negative mode, the smallest gap is much larger, as it is due to much more significant compositional changes (*e.g.* changes including $-C_{25}$ or $+C_{18}$, see Table 7.4). In this case, it may be possible to assign formula based on other criteria than just mass accuracy, for example the types of compounds expected in a sample, or classes of compounds to be ionised by a given source - *i.e.* aromatics in APPI, fatty acids in ESI - which may be determined from DBE or AI calculations, but this is a subjective strategy.

To determine how significant these mass differences are, resolving power calculations were performed based on Equation 7.1.

$$\text{Measured Resolving Power} = \frac{m}{\Delta m_{50\%}} \tag{7.1}$$

The measured resolving power is defined as the mass of a peak divided by its full width at half maximum ($\Delta m_{50\%}$) as shown in Equation 7.1. To calculate how to achieve these resolving powers, in the context of how long a transient must be (recalling that in FT spectroscopy, transient length $\propto$ spectral resolution), further calculations were performed using Equation 7.2.[132] The theoretical resolving power can be calculated from the magnetic field strength (B, in Tesla), the length of the recorded transient (t, in seconds), the mass of interest (m/z) and a constant (C) as shown in Equation 7.2.[132]

$$\text{Theoretical Resolving Power} = \frac{CBt}{m/z} \tag{7.2}$$

For data acquired on the instrument used here, C was empirically determined to be approximately 1×10^{7}, accounting for processing factors such as apodisation and magnitude mode processing. C was not determined for phased spectra. The results for these calculations are summarised in Table 7.5.

Here we have calculated resolving power required to differentiate between two peaks on their mass difference. This is also based on the peaks being of equal intensity, and thus not requiring baseline resolution. For a large difference in ion intensities, higher resolving powers would be needed.

Table 7.5: Theoretical resolving powers and transient lengths required for positive and negative ionisation modes at two values of m/z. Values calculated for an magnetic field strength of 12 T and spectra processed in magnitude mode.

POLARITY	MASS (m/z)	MIN. δ mDa (ppm)	RESOLUTION REQUIRED	TRANSIENT LENGTH (S)
Positive	401	0.152 (0.38)	2,638,223	8.8
Positive	799	0.152 (0.19)	5,257,105	35
Negative	401	3.5 (8.8)	115,000	0.38
Negative	799	2.37 (2.97)	337,168	2.2

As can be seen from the data in Table 7.5, there is a stark difference in experiment required for negative and positive ionisation of a complex mixture like Scotch Whisky. Thus, if in positive mode, both sodium and potassium adducted ions are possible, it is impossible to discriminate between these assignments without a very high resolving power. For ionisation sources such as ESI, which is known to produce sodiated and protonated ions, and theoretically could produce potassiated ones also, it is not possible to resolve and assign species confidently without resolving power in the millions. At 12 T, this equates to long acquisition times of 8.8 s at m/z 401, and 35 s at m/z 799. These transient lengths are difficult or impossible to achieve on this instrument in broadband acquisition mode, due to limits with vacuum quality and ion cloud instability. Higher resolving powers are investigated in practice in section 7.6. For negative mode ionisation, with only deprotonated species, the resolving powers required are in the low hundreds of thousands. These are entirely feasible with short transient lengths at 12 T in magnitude mode.

These results confirm that without higher resolution data, requiring a higher resolution instrument or method, it is impossible to confidently assign positive mode data of complex mixtures from broadband mass spectra. There are potential means to work towards this limit. By use of phased, absorption mode spectra, the resolution could be increased between $\sqrt{2}$ and 2, depending on apodisation. With 2ω quadrupolar detection, the resolution could be doubled again (or transient length halved).[139,266,267] At a higher magnetic field, resolution would also increase, as per Equation 7.2. However, the improvement at 21 T compared to 12 T is only 1.75-times ($= \frac{21}{12}$, Equation 7.2), and there are currently only two 21 T FTICR instruments in the world. Combination of these three improvements results in a net ≈6-times improvement in resolution for the same length of transient. This alone is obviously still insufficient for short transients, but longer transients

(6 seconds or longer) are possible with improved ICR cell design and control, vacuum levels, and ion counts.

The calculations made here were for a relatively simple case of CHO only formulae, singly charged as deprotonated (negative) or protonated (positive) or sodium or potassium adducts (positive). As discussed, radical anions would not impede monoisotopic peak discrimination. Other ionisation states were not considered. Addition of more heteroatoms, *e.g.* NSP, would increase difficulty further again. Also, these calculations are for monoisotopic formula only; inclusion of isotopologue peaks would further increase the complexity of the theoretical number of peaks. Finally, these calculations are based on ideal spectra of equal ion counts. Large differences in ion counts, unstable ion-clouds, space charge effects, peak coalescence, and thermal noise will aggravate the situation. Of course, no sample is going to contain every possible formula. Overall, though, it is easy to see that negative mode data with simply deprotonated formula are much more straightforward to acquire with sufficient resolution, and to assign, than positive mode data.

From a practical perspective, positive mode data were acquired without any substantial complications compared to the negative mode data. If one could be confident of only producing one type of ion (*e.g.* all protonated, no adducts), it may be possible to assign data at this resolution, as the minimum gaps would be comparable to those for negative mode (where only one type of ion is expected).

7.6 RESOLVING POWER

In the previous sections, the ionisation sources were compared and optimised. However, analyser settings were kept essentially the same. When it came to positive mode ionisation, the achieved resolving power was insufficient for unambiguous formula assignment. In this section, increasing the achievable resolving powers in broadband acquisition on the current instrumentation are tested. Sample S14-1941 was used exclusively in this section.

FTICR represents the gold standard of high resolution mass spectrometry. The instrument used throughout this thesis was equipped with a high field, 12 T magnet and modern electronics and ICR cell. However, newer hardware now exists. In this and the previous chapter, the resolving power achieved at m/z 400 was approximately 300,000. However, FTICR is theoretically capable of much greater resolving powers. In section 7.5, the required resolving powers were calculated for positive and negative ionisation modes. Without access to higher field strengths or 2ω detection, the question arises - how far can the resolving power

be pushed on the available equipment, without compromising spectral quality, sensitivity, or time required?

The possible resolving power is dependent on the length of the recorded transient (as with NMR). In FTICR MS, the transient length is controlled by balancing the low mass (highest frequency) limit and time domain size in megawords. Typically, 4 MW is now routinely used in the literature. The balance is related to the Nyquist theorem, as explained further in the introduction. In short, longer transients can be achieved by increasing the time domain (4 MW to 8 MW will double transient length), or by increasing the low mass (m/z 100 to 200 will double transient length). The consideration for the user is where the regions of interest in the spectrum are.

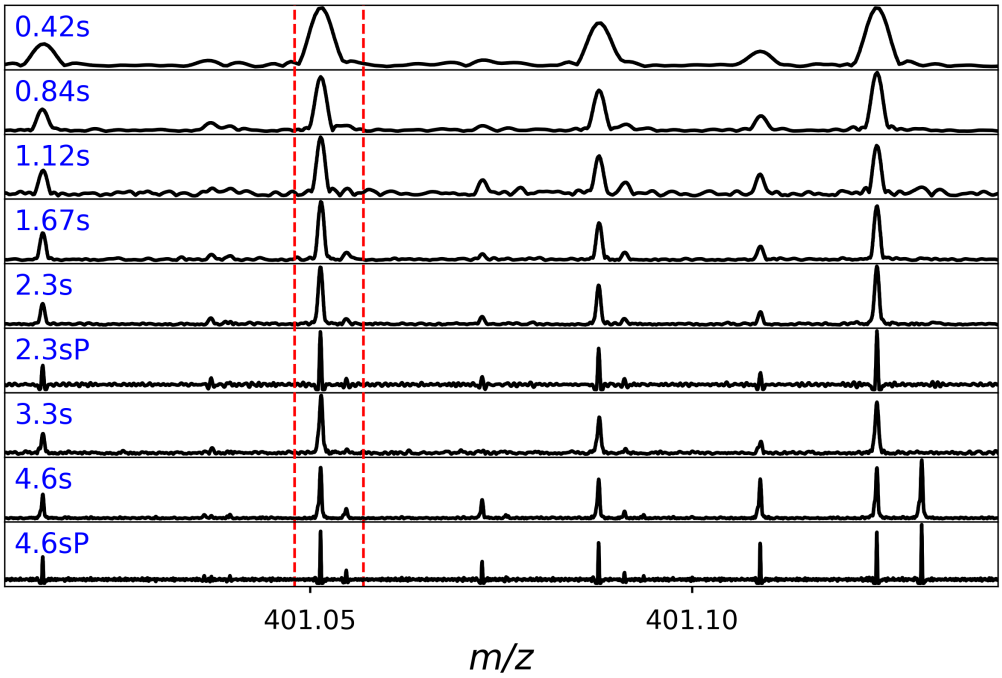


Figure 7.12: Expansions around m/z 401 for the negative ESI FTICR mass spectrum of Scotch Whisky (S14-1941) acquired with different transient lengths. Transient lengths are annotated on each subplot, 'P' indicates the spectrum was phased post-acquisition, where the other spectra were presented in magnitude mode. Note that 4.6s and 4.6sP were acquired with a longer IAT, causing different relative signal intensities. Red dashed lines indicate region expanded in Figure 7.13.

To determine both the limits of, and optimal, transient lengths possible, a series of spectra were acquired with different parameters. An expansion around m/z 401 is shown for all acquired spectra in Figure 7.12. These spectra were acquired consecutively, altering only the low mass limit or the time domain size to achieve required change in transient length. For all but the 4.6 s spectra, 100 transients

were co-added in a normal acquisition. The 4.6 s transient data were acquired using segmented acquisition - the Q1 mass filter and isolation window (50 m/z) were applied and sequentially moved through the mass range, summing 50 transients at each position. As quadrupole filtering reduces the number of ions which make it to the ICR cell, to achieve greater SNR, the IAT was increased to 1 s from 0.1 s, though this yields different relative signal abundance - exemplified by the furthest most right peak in Figure 7.12. IAT is known to affect the spectrum profile in complex mixture MS.[133] Theoretically, the increased SNR could be used for broadband IFS analysis, however variable ion counts in different spectral windows mean that a fixed IAT resulted in areas of split peaks due to overfilling of the ICR cell.

The increasing resolving powers shown in Figure 7.12 show diminishing returns in Scotch Whisky. Going from 0.42 s to 0.84 s and 1.12 s yields much narrower lines, and resolves signals previously hidden in the noise or merged with nearby signals. An example of this is shown in Figure 7.13. At 1.12 s a minor signal at approximately m/z 401.055 begins to be resolved, though it isn't a particularly sharp feature until the phased 2.3 s data. Based on the highest resolution spectrum (4.6sP), this is $[C_{16}H_{17}O_{10}S]^-$ (0.10 ppm), and likely a real species, although no entry exists in the *ChemSpider* database for this molecular formula.

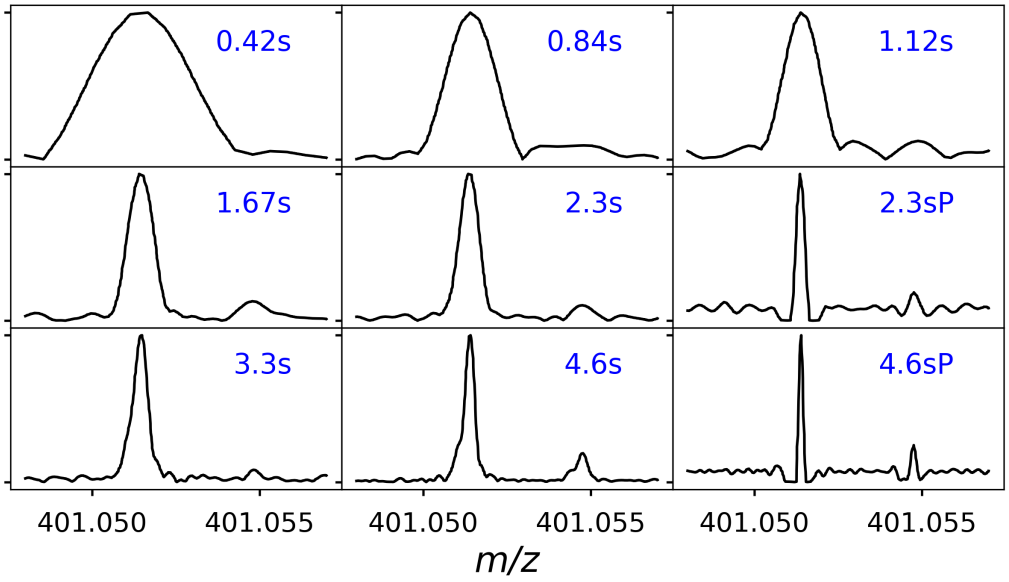


Figure 7.13: Narrow expansions around m/z 401.05 acquired with different transient lengths. Transient lengths are annotated on each subplot, 'P' indicates the spectrum was phased post-acquisition.

Figure 7.13 also highlights another significant part of acquiring high resolution data - unlike in NMR, where transients decay according to T_2 relaxation,

and provide simple Lorentzian line shapes, FTICR transients decay because of ion cloud instability. Substantial apodisation is routinely applied in FTICR, to limit so-called 'wiggles' or side lobes by decaying the transient to 0 at its end. The ideal transient in FTICR decays only minimally, as ion-cloud instability decay distorts the resulting line shape. This effect is noticeable in the 3.3 s and 4.6 s data in Figure 7.13, where the large signal does not have a symmetrical line shape, but in fact shows distortion on the left side. This reduces resolving power and mass accuracy, and is an undesirable effect. Interestingly, phasing (with different apodisation) appears to resolve this problem in 4.6sP. This is similar to NMR, where line shape distortions are caused by magnetic field inhomogenity and apodisation can narrow the lines. This apodisation and phasing can also distort the line shapes, introducing partial negative 'dips' either side of the peaks.

Table 7.6: Summary of resolving powers achieved using different time domain and low masses. Aquisition time factors shown relative to the method used throughout this thesis. Resolving powers (RP) measured at m/z, and calculated for ideal system with no apodisation. *Spectra processed using Kilgour apodisation[142] and phased.

FID (S)	LOW MASS	TD (MW)	IAT	EXPT. RP	THEOR. RP	EXPT. TIME EQ.
0.42	147.4	1	0.1	120 k	160 k	0.4
0.84	147.4	2	0.1	240 k	320 k	0.8
1.12	98.3	4	0.1	300 k	430 k	1.0
1.67	147.4	4	0.1	430 k	640 k	1.5
2.3	202.7	4	0.1	600 k	880 k	2.0
2.3*	202.7	4	0.1	1.5 M	1.6 M	2.0
3.3	147.4	8	0.1	800 k	1.3 M	2.8
4.6	202.7	8	1.0	1.2 M	1.8 M	34.4
4.6*	202.7	8	1.0	2.8 M	3.6 M	34.4

The resolving powers achieved are summarised in Table 7.6, and show that magnitude mode data consistently achieve poorer resolution than theoretically possible - due to the heavy apodisation - whilst the phased data are much closer to theoretical values. Whilst the phased 4.6 s data achieves an excellent 2.8 million resolving power - approaching the levels required for positive mode data acquisition - it required significantly longer to acquire, some 34.4 times as long as the normal broadband method presented earlier and in the previous chapter. Furthermore, whilst the IAT could be increased for some regions of the segmen-

ted acquisition, in others (not shown) it caused substantial peak splitting and other signal distortions when the ion count was too high within the ICR cell.

Ultimately, instrumental parameter selection is always going to be a compromise between time and quality. In the case of Scotch Whisky, it does not appear to be so complex as to require ultra-high resolving powers for negative mode ESI. Based on all of the above, it appears transient lengths of 1.12 s in magnitude mode are probably too short to resolve the most minor features, and ideally 1.67 s or, better yet, 2.3 s, and phased data, should be routinely used in future.

7.7 DEVELOPMENT OF NEW SOFTWARE

To process, analyse, and visualise the data generated during this PhD, particularly the FTICR MS data, a number of software tools were developed using *Python* and assorted packages. One of these tools was for formula assignment, and visualisation tools were written - some of which produced figures for this thesis. The visualisation tools generated both static plots (using *matplotlib*),[268] including van Krevelen and DBE plots, as well as interactive plots (using the *Bokeh* package)[269] viewable online. An initial release of the scripts is freely available online(`https://github.com/wkew/FTMSVisualization`) along with a publication.[169] Furthermore, an online interface (`https://wkew.co.uk/ftmsvisonline/`), built using *Wooey* (`https://github.com/wooey/Wooey`), allowing use of some of these scripts has also been made available. Complete discussion of these tools is not included here; they are already published, and there are space limitations in this thesis.

7.7.1 *Formula Generation and Assignment*

Early on, one of the key missing aspects in the processing work flow was a means to automate the assignment of high resolution mass spectrometry data. As discussed in section 7.8, tools were available at later stages, and that section compares the performance of those tools. However, the tools developed in-house, despite not having as high assignment rates as Formularity, do serve several purposes, including calculation of possible formulae as utilised in section 7.5. That formula generation capability is actually a necessary stage in the formula assignment strategy.

The formula assignment methodology implemented here essentially cross references exact masses from an input peak list with a pre-compiled database of possible chemical formula. There are a few additional sophistications added to reduce risk of false assignments. The most significant of these relies on homologous chemical series and Kendrick mass defects. This characteristic aspect of

complex mixtures is well known, and is the key to many formula assignment strategies, including those developed in parallel for Formularity. In the in-house tools, the following Kendrick units are searched for: CH_2, OH_2, CO_2, and H_2. Kendrick masses were discussed in the introduction.

The script loads in a peak list, and for each Kendrick mass unit calculates the Kendrick masses of each peak. Peaks are then grouped and sorted according to their z*, checked against several criteria, and assignments are then made. After all the monoisotopic assignments have been made, the script iterates through each monoisotopic assignment searching for ^{13}C isotopologues within a pre-defined error threshold.

The criteria which are checked against are designed to reduce the risk of false assignments. For example, whilst Kendrick mass defects should be exactly the same for species in the same series, there will be a small error due to mass accuracy. A tolerance value (0.3 mDa) is allowed for to compensate for this. The exact value of this may need alteration depending on spectral quality. Additionally, this script requires multiple hits (typically 3) within a series of compounds for assignment, based on the assumption that there will not be individual species unrelated to any others within the sample. Along with a relative mass error threshold ($\pm$0.5 ppm), there is an absolute mass accuracy tolerance (2 mDa) for assignment. Finally, for isotopologue matching, the script allows for a maximum 1 mDa error for the m/z +1.003355 theoretical position of the first ^{13}C isotopologue. Most of these values have been empirically optimised.

Future work could incorporate isotopic peaks to confirm assignment. Whilst ideally this would use the full IFS - when sensitivity and resolution allow for it - in a simpler sense, the relative intensity of the first ^{13}C isotopologue should provide enough confirmation for a formula assignment. For example, as discussed in section 7.5, the smallest mass difference in a negative mode spectrum at m/z 401 (8.8 ppm) is exchanging -H_8O_{13} for +C_{18}. That large difference in carbon numbers will induce a substantial change in the theoretical signal intensity of the ^{13}C isotopolgue. For a C_{12} compound with 100 % monoisotopic relative intensity, the first isotopologue will have a (1.1 % × 12) 13.2 % relative intensity. In contrast, an equivalent ion with +C_{18} (thus C_{30}) would have a 33 % abundant isotopologue. This significant difference will be observable for species with high SNR and natural isotopic abundances. This is a similar strategy to how Bruker DataAnalysis scores formula assignments with its manual assignment tool.

Initial formula generation is based on criteria developed by Kind and Fiehn for 'Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry'.[270] Essentially, the user defines a polarity to generate ions for (negative or positive), if adducts are to be considered (sodium and / or

potassium), and which elements they wish to include. By default, the elemental limits, in terms of numbers of CHO, etc., are taken from Kind and Fiehn,[270] however these can be overruled to allow for specific elemental amounts - *i.e.* max one sulfur. The tool has been used for elements including CHONSP, as well as sodium and potassium adducts, and deuterium.

The formula generator iterates through every combination of every allowed element. By use of condensed efficient Python, it is possible to iterate through all the criteria (rules) and combinations in a small amount of code and relatively fast. Most of the conditions can be condensed into one `if` statement within one function returning a simple `True` or `False`. This is shown in Algorithm 7.1. These conditions check the relative ratios of elements. Where elements are not in use, they have a value of 0 and thus return 0 in any of these calculations. As such, C and H must always be included - however, this is obvious within organic mixtures. These limits could be compared to the limits of a van Krevelen diagram and where chemistries may expect to fall, however they are based on the work of Kind and Fiehn.[270] Another limit (not shown) is that the maximum number of heteroatoms must not exceed 1.3x the number of carbons. These criteria may be a cause in why this tool assigns fewer species than Formularity or PetroOrg, however unassigned (or non-generated) species will represent fringe chemistries.

Algorithm 7.1 Formula generation conditions

```
def conditions(c,h,n,o,s,p,na,k,d):
    if (      0.2 < float(h)/float(c) <3    #hc ratio
    and     float(o)/float(c) < 1.2         #oc ratio
    and     float(n)/float(c) < 1.3         #nc ratio
    and     float(s)/float(c) < 0.8         #sc ratio
    and     float(p)/float(c) < 0.8         #pc ratio
    and     float(d)/float(c) < 3
    ):
        return True
    else:
        return False
```

The calculated dictionaries of all possible formulae are stored in a series of text files, each covering m/z 100 ranges. This is useful for formula assignment, where each dictionary can be read in separately and stored as a separate array. Thus, when iterating through peak lists and formula databases, the searches can be limited to narrower mass ranges to accelerate performance. Furthermore, pre-compilation of possible formula means they only have to be generated once, rather than each time the assignment script is run.

7.8 FORMULA ASSIGNMENT SOFTWARE

Assignment of chemical formulae from exact mass is not a trivial problem. Beyond assumptions about elemental and adduct compositions, as discussed earlier, there are a number of challenges when trying to assign many formulae across multiple spectra. With access to instrument time, IFS can be used to confirm essentially any small molecule composition. But for broadband analysis of dozens of samples, as presented in this thesis, automated strategies are required.

In this chapter, exact masses were assigned chemical formulae using the software Formularity, whereas in the previous chapter, formula were assigned using PetroOrg. Additionally, a formula assignment program was developed in house as part of this research (discussed further in section 7.7). The reasons for these three developments are essentially pragmatic. At the onset of this PhD, there was no tool available within the department for automated assignment of molecular formula. During the course of my PhD, I wrote a tool for formula assignment - however, this development was interrupted by acquisition of a licence for PetroOrg, a commercial piece of software explicitly designed for molecular formula assignment in complex mixtures. In 2017, Formularity was published.[207] It was based on work published much earlier,[160] but had been re-written and made much more accessible to use. Formularity, like the in-house tool, cross references masses against a pre-compiled database of possible chemical formula and extends assignment lists using Kendrick mass defects. The information of how PetroOrg works is not publicly available.

In this section, these three tools are compared in terms of speed, consistency, and practical aspects. Data acquired for sample S14-1941, acquired as part of chapter 6, will be used throughout. This spectrum has 4694 peaks detected above the peak picking threshold. To ensure comparable results, assignment limits are kept consistent. Specifically, a maximum error of 1 ppm and formula limited to $C_{0\text{-}100}$ $H_{0\text{-}200}$ $O_{1\text{-}20}$ $S_{0\text{-}1}$. This is a lowest common denominator approach based on PetroOrg's parameters, though Formularity and in-house assignment tools have more sophisticated rules. Furthermore, the assignments above m/z 700 are discarded to exclude potentially spurious results.

The number, and errors, of assignments for each of the three methods are summarised in Table 7.7. This table shows that Formularity and PetroOrg provide comparable rates of assignment (73 % to 76 %) and comparable mass errors of assignment. In contrast, the in-house developed assignment tool assigns fewer formula, only 58 %. However, the mass errors are half the size as for Formularity or PetroOrg. It would appear that the in-house method is stricter with assign-

ment than the other two programs, and whilst it assigns fewer formula, it is more confident with those assignments it does make on average.

Table 7.7: Summary of assignment statistics for sample S14-1941 by three different programs - PetroOrg, Formularity, and in-house scripts.

	PETROORG	FORMULARITY	IN-HOUSE
# Total Peaks	4694	4694	4694
# Assigned	3430	3549	2707
# Monoisotopic	2509	2588	2140
Mean Error (ppm)	0.093	0.095	0.049
StDev	0.093	0.096	0.051

To see if the assignments are consistent between programs, visualisation of the intersections of these assignments is presented. Figure 7.14 shows the UpSet plot for the assignments across the three programs. It is clear that the vast majority of assignments are consistent to all three programs. As PetroOrg and Formularity assigned more, they have a shared group of assignments not found with the in-house tools. Only a handful (17) assignments were unique to Formularity, and only 15 were unique to the in-house tools. No assignments were unique to PetroOrg. These results suggest Formularity is providing the greatest coverage of assignments.

From a practical point of view, there are other considerations for the three programs. PetroOrg is commercial and thus is available to fewer groups, limiting reproducibility. In contrast, Formularity and the in-house tools are free and open source tools. The open-source aspect provides additional advantages, most importantly that it is possible to determine what the algorithm for assignment is. PetroOrg, in contrast, is a black-box approach with no public information on how it works. Furthermore, software which is open source allows the end user to modify these tools for their own specific applications.

Anecdotally, PetroOrg is the slowest (computationally) tool, whilst Formularity is the fastest. Batch processing is possible in all three tools, although only Formularity can perform it in parallel and execute peak-alignment prior to formula assignment. However, none of these methods are particularly time consuming, and all will assign thousands of CHO peaks in under a minute on a modern desktop computer.

All tools can have additional elements or ionisation modes, *i.e.* adducts or radicals, considered. With the in-house scripts, this involves creating additional formula databases, and potentially significant user modification. Formularity provides

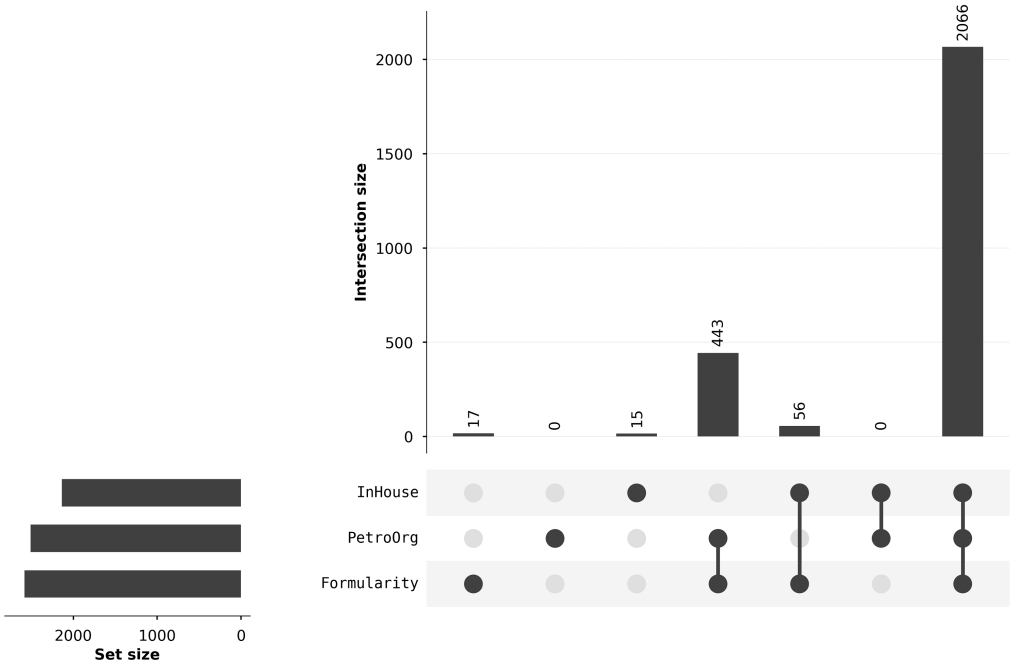


Figure 7.14: UpSet plot showing intersections for formula assignments made for three different programs - PetroOrg, Formularity, and in-house scripts.

a more straightforward means to achieve this, however it can only assign one ionisation mode at a time - so analysis must be repeated for radical anions after deprotonated species, for example. The risk with a sequential approach is assigning the same species to multiple forms - especially when considering adducts, such as sodium and potassium, as explained in section 7.5. PetroOrg provides means to consider radicals and adducts, though how this is calculated is unclear. Furthermore, PetroOrg becomes much slower the more elements or adducts are allowed.

Based on all the above considerations, Formularity is recommended as the best tool for routine formula assignment of complex mixture mass spectra. However, all researchers should of course validate assignment strategies for their own complex mixtures and data.

7.9 CONCLUSIONS

In chapter 6, negative mode ESI was utilised to examine dozens of Scotch Whisky samples. In this chapter, alternative ionisation sources and positive mode ionisation were investigated as a complementary means for the analysis of Scotch Whisky. Solution state atmospheric techniques, specifically APCI and APPI, were found to produce complementary information to ESI. APCI presented a chal-

lenge with its high sensitivity for any contamination in the nebuliser gas, whilst APPI only produced signal when the samples were sprayed neat and at a high flow rate. Therefore, APCI and APPI are not best suited to volume limited samples, or particularly light and uncomplicated samples, *i.e.* NMS. A solid-state technique, LDI was found to provide a very different profile of Scotch Whisky, in both compounds ionised and relative signal intensities. LDI suffered from variable signal scan-to-scan, and was most time consuming technique for which to acquire high quality spectra. All techniques, APCI, APPI, ESI and LDI, provide useful information, and no one ionisation source will comprehensively characterise a complex mixture such as Scotch Whisky. Furthermore, this is the first reported attempt to use APCI, APPI, or LDI for analysis of Scotch Whisky or other mature spirit drinks.

Positive mode ionisation was investigated. Due to adduct formation, it generates more complex spectra than negative mode. At the resolving powers achieved in the routine methodology, it was not possible to unambiguously assign the positive mode data. Following this, it was investigated what resolving powers were required - and what could be achieved on this instrument. With segmented acquisition and phasing of the spectra, a maximum resolving power of 2.8 million at m/z 401 was achieved, however at the expense of significant instrumental time.

Finally, three formula assignment tools were compared, including commercial (PetroOrg), open source (Formularity), and in-house tools. Whilst the in-house tools performed objectively well, the Formularity and PetroOrg had higher assignment rates. All three tools assigned essentially the same species, with very little conflicting assignments. Ultimately, Formularity performed best in terms of numbers of assignments, speed of assignment, and the fact it is open source allowing for free (gratis) use and free (libre) understanding of the underlying algorithms. The in-house tools I developed come with superior data visualisation scripts included - Formularity does not produce graphics, and PetroOrg has limited visualisation tools. Furthermore, the scripts I developed are open source Python and therefore could easily be adapted by other groups into their own workflows, or coupled with other open source Python software for FT MS analysis, for example the 'spike' Python package from Chiron *et al.*[271]

8

CONCLUSIONS

Science and art belong to the whole world,
and before them vanish the barriers of nationality.

— Johann Wolfgang von Goethe (1813)

8.1 SUMMARY

In this work, two complementary, powerful, high-resolution techniques have been used to analyse Scotch Whisky. Scotch Whisky represents a complex mixture - and comprehensive analysis of it is no small feat. Whilst many aspects of the chemistry remain unknown, this thesis represents a step forward toward the molecular level characterisation of Scotch Whisky.

To conduct analysis, two techniques were chosen - NMR and FTICR. These divergent techniques have similar instrumentation - superconducting high strength magnets, RF pulse sequences and rely on Fourier transformation of time domain data to yield spectra. However, the physics behind both techniques are quite divergent, with NMR examining the spins of atomic nuclei, whilst FTICR probes the mass-to-charge of molecular ions. The former thus provides information on chemical structure and dynamics, whist the latter describes the mass of each species.

NMR analysis of Scotch Whisky was conducted on native samples, without removal of the protonated solvents. Therefore, the first step was to develop and evaluate a solvent suppression methodology. A modified 1D NOESY presaturation method was developed, automated, and implemented into a variety of 1D and 2D, homo- and heterocorrelated NMR experiments, performed at high field 600 MHz and 800 MHz. The results were very good, offering high resolution and high sensitivity experiments. Using these tools, dozens of chemical structures were elucidated in Scotch Whisky. Additionally, dynamic processes were observed, including chemical exchange and equilibrium between acetaldehyde and its hemiacetals of water and ethanol. For these measurements additional NMR experiments - including DOSY - were utilised. As NMR is quantitative, quantification of ethanol in Scotch Whisky was performed successfully, although quantification of minor congeners was more difficult due to peak overlap, chemical shift changes, and equilibria of forms.

A metabolomics approach was taken to analysise a large sample set of mature Scotch Whisky products, the 1D ^{1}H NMR spectroscopy, covering a diverse range of styles. Samples from multiple years of production were also included. Application of chemometric methods, including STOCSY, unsupervised (ICA and PCA) and supervised (OPLS-DA) methods revealed relationships in the data between correlated species and samples. Malt and blend Scotch Whisky were successfully discriminated based on their higher alcohol content, specifically 3-methylbutanol, echoing results from GC-MS and analysis by Aylott and Mackenzie.[25] Peated samples exhibited characterisatic phenolic peaks, however these were at the limits of detection. Maturation wood types could be discriminated, though no NMR-identifiable chemical markers could be elucidated.

FTICR MS was performed initially using ESI in the negative mode, as typical for complex mixture analysis, yielding high quality spectra for dozens of Scotch Whisky samples. Modern analysis software achieved high (>70 % rates of formula assignments with small mass accuracy errors ($\ll$1 ppm). Chemical formula assignments were confirmed using IFS analysis, confirming the presence of higher molecular weight sulfur containing species in some Scotch Whisky samples for the first time. Across the sample set, heteroatomic class distributions could be visualised, and malt and blended Scotch Whisky samples were found to have differing chemistries. Examining this further, chemometric methods were applied in unsupervised (PCA) and supervised (OPLS-DA) ways to further model the Scotch Whisky mass spectra. PCA clustered, loosely, samples that were malts or blends, whilst supervised methods were able to accurately classify Scotch Whisky according to maturation wood type, amongst other things. Key potential marker compounds were identified for maturation wood type. Isolation and fragmentation of chemical species was achieved using both the quadrupole for highly abundant species, and in-cell isolation and fragmentation for minor species. Fragmentation patterns helped inform structural elucidation.

Following from the success of negative mode ESI, a variety of other ionisation sources - APCI, APPI, and LDI - were used with FTICR for the analysis of Scotch Whisky for the first time. Positive mode ionisation across all four sources was investigated, but insufficient resolving powers hindered unambiguous formula assignment due to the presence of potassium or sodium ion adducts. The four ionisation techniques provided complementary information across a set of Scotch Whisky samples, revealing both a diversity of chemistry across samples and across ionisation sources. The use of just one ionisation source provides a very limited perspective of the chemistry of a complex mixture. Each method had its own advantages and disadvantages. ESI and APCI were prone to fatty acid and hydrocarbon contamination, respectively. LDI and APPI - both used

without ionisation efficiency boosters like a matrix or dopant (*e.g.* toluene) - required concentrated, mature Scotch Whisky samples; NMS would not ionise using LDI or APPI. Thus, selection of an ionisation source, and optimisation of such methods, does rely upon careful consideration of the sample types.

FTICR provides the highest resolution mass spectra possible, although for the work in this thesis the technique was not pushed to its limits. Indeed, for negative mode analysis of Scotch Whisky, ultra high resolution (resolving power $> 1 \times 10^6$) does not appear to be necessary for monoisotopic formula assignment. However, the limitations of the current instrumentation were pushed by means of increased transient length in broadband acquisition, by means of phasing the absorption mode spectra, and by acquiring the data in a segmented approach with both long transients and absorption mode processing. These achieved very high resolving powers of 2.8×10^6 at m/z 400 - approaching the resolutions necessary for positive mode analysis and more complex mixtures. However, such methods require the instrument to be in an optimal condition, and a significantly longer acquisition time - up to 34-times longer for the best results.

New software tools were developed for the analysis of complex mixture MS data, including formula generation and assignment, data processing and analysis, and visualisation. These Python tools were published open source, and, partially, implemented in a free online tool to allow the wider community to use and adapt them. The ubiquitous van Krevelen diagram was given a modern refresh by creation of interactive van Krevelen (and other) diagrams, allowing inspection of the data in a dynamic way. This links data and different visualisations - including the spectra - to allow a more thorough inspection of the data than achievable in a static plot.

8.2 FUTURE STEPS

This work has been successful in both method development and application. NMR, FTICR MS and data analysis methods have been developed, complementing the existing tools within the department and in the wider analytical community. Application of these techniques to mature Scotch Whisky samples has yielded new information on maturation chemistry and differences in sample types. This work opens up many possible future avenues of research, to further comprehend the chemistry of Scotch Whisky, and to solve related analytical challenges.

One of the primary areas of interest in Scotch Whisky, and wider food and drink research, is product protection. This includes authentication - both brand and category. Based on the work conducted here, it is highly probable that both FTICR and NMR will be successful in generic, categorical authentication of Scotch

Whisky; NMR will be able to quantify compounds in a similar way that GC/LC-MS are used now, and FTICR will provide a complex chemical fingerprint of samples. Based on the work in chapter 6, where the same product from multiple production years yielded nearly identical mass spectra, it may even be possible to use FTICR for brand authentication. Significant validation with large, well curated sample sets - both authentic and counterfeit - will be required, and an extensive database built up for either or both of these methods to work for such a purpose. Unfortunately, reproducibility concerns with FTICR, and the economic and technical restrictions associated with both techniques will limit their adoption by industry.

Product protection also includes identification of compounds of concern and their origins. Origin determination will require specific experimental design, or large sample sets. Identification of unknown compounds, especially low abundance species, will require some form of chromatographic purification and concentration.

Identification of the origins of the chemical complexity in Scotch Whisky, as observed by FTICR MS, will require acquisition of specific sample sets. Scotch Whisky contains compounds produced during fermentation, distillation, and maturation, and through addition of E150a caramel colouring. Three main avenues are possible to deconvolve the complexity based on these potential origins. The most obvious is a full process, time series analysis - a sample set consisting of samples from every stage of the production process, including the raw cereals and yeasts, the wort, mash, wash and new make spirits, and then a time series sample set of the maturing spirit. Given the inherent variability in each stage of each process, multiple experiments should be conducted simultaneously. Scotch Whisky requires three years of maturation at a minimum, but the chemistry will continue to develop over five, ten, twenty years and more. Unfortunately, no such sample set yet exists.

Time series analysis can, somewhat, be overcome by means of model systems. Maturation modelling could be performed by use of miniature, small volume casks, increasing the relative surface area and then increasing extraction of wood related congeners. Unfortunately, such an approach will not replicate real maturation. However, accepting the limitations of model systems, further work could include examining wood extractions by using sticks or shavings of wood in ethanol:water solutions. Variables in this could include pH, alcohol strength, temperature, humidity, and also wood species, origin, provenance and history, pretreatment (char, toast), and repeat extractions to model the diminishing returns associated with old, exhausted, and inactive wood. To purely characterise the

chemistry of wood and wood extractable compounds, alternative solvents could be considered.

Fast production processes, including fermentation and distillation, could be examined on real samples without significant difficulty in obtaining them. This may allow an insight into the biochemical processes of fermentation, the impact of differing yeasts and microflora, and further research onto the impact of copper in distillation, or wood versus steel washbacks. Likewise, comparisons of cereals could be made at these stages, separating out the effects of maturation and blending.

Blending - for production of Single and Blended Scotch Whisky - is believed to differ if the spirits are allowed to 'marry' for a period of time, or further mature together. This process is worth potential investigation, as is chill filtration, which, whilst previously studied, has not been characterised by a technique as sensitive as FTICR.

Finally, E150a caramel colouring is added to many Scotch Whisky products. It is added for aesthetic colour reasons, and not to impact on flavour. Only one type of E150 is added (a, and not b, c, or d), and not all producers use it. Therefore, analytical strategies to detect, identify, and quantify caramel colouring are of significant interest in the contexts of product protection. Due to its poorer limits of detection, NMR methodology may need further development to detect E150a, whilst FTICR has significant promise to qualitatively confirm the presence of E150a. Discrimination between caramel types, and between samples with addition and without, will be important. Identification of marker compounds in caramel which are not naturally in Scotch Whisky will require large, well curated sample sets. Previous research in the literature suggests caramel colouring - and wood extractions, also - will themselves be complex mixtures. Each aspect could itself be a substantial area of research. Colour in mature spirits, both wood-derived and caramel-derived, is not a well understand property on a molecular level, but it is certainly worthy of future research.

BIBLIOGRAPHY

(1) S. W. A. (SWA), *2017 Export Figures*, http://www.scotch-whisky.org.uk/news-publications/news/2017-export-figures/, Accessed 21/06/2018, 2018.

(2) *Whisky*, ed. I. Russell and G. Stewart, Elsevier, 2nd edn., 2014.

(3) S. W. A. (SWA), *Scotch Whisky Exports Analysis 2017*, http://www.scotch-whisky.org.uk/news-publications/news/scotch-whisky-exports-analysis-2017/, Accessed 22/06/2018, 2018.

(4) D. E. Briggs, *Journal of the Institute of Brewing*, 1963, **69**, 13–19.

(5) T. A. Bringhurst and J. Brosnan, in *Whisky*, Elsevier, 2014, pp. 49–122.

(6) A. M. Smith, *Biomacromolecules*, 2001, **2**, 335–341.

(7) W. Friedrich and R. Galensa, *European Food Research and Technology*, 2002, **214**, 388–393.

(8) H. L. Riffkin, R. Wilson, D. Howie and S. B. Muller, *Journal of the Institute of Brewing*, 1989, **95**, 115–119.

(9) Z. Jiao, Y. Dong and Q. Chen, *Comprehensive Reviews in Food Science and Food Safety*, 2014, **13**, 611–626.

(10) Y. B. Monakhova, T. Kuballa and D. W. Lachenmeier, *ISRN Analytical Chemistry*, 2012, **2012**, 1–5.

(11) D. Murray, in *Whisky*, Elsevier, 2014, pp. 179–198.

(12) B. Harrison, J. Ellis, D. Broadhurst, K. Reid, R. Goodacre and F. G. Priest, *Journal of the Institute of Brewing*, 2006, **112**, 333–339.

(13) B. M. Harrison and F. G. Priest, *Journal of Agricultural and Food Chemistry*, 2009, **57**, 2385–2391.

(14) C. A. Wilson, Ph.D. Thesis, Heriot-Watt University, 2008.

(15) Bruichladdich, *Octomore 08.3*, https://www.bruichladdich.com/octomore/octomore-masterclass-edition-083, Accessed: 21/06/2018.

(16) P. Wiśniewska, T. Dymerski, W. Wardencki and J. Namieśnik, *Journal of the Science of Food and Agriculture*, 2014, **95**, 2159–2166.

(17) J. Pryde, J. Conner, F. Jack, M. Lancaster, L. Meek, C. Owen, R. Paterson, G. Steele, F. Strang and J. Woods, *Journal of the Institute of Brewing*, 2011, **117**, 156–165.

(18) W. Young, H. Horth, R. Crane, T. Ogden and M. Arnott, *Water Research*, 1996, **30**, 331–340.

(19) N. K. Hermkens, R. L. Aspers, M. C. Feiters, F. P. Rutjes and M. Tessari, *Magnetic Resonance in Chemistry*, 2017, **56**, 633–640.

(20) P. Bendig, K. Lehnert and W. Vetter, *Journal of Agricultural and Food Chemistry*, 2014, **62**, 2767–2771.

(21) K.-Y. M. Lee, A. Paterson, J. R. Piggott and G. D. Richardson, *Journal of the Institute of Brewing*, 2001, **107**, 287–313.

(22) I. Blank, in *ACS Symposium Series*, American Chemical Society, 2002, pp. 25–53.

(23) I. Russell and G. Stewart, in *Whisky*, Elsevier, 2014, pp. 123–146.

(24) A. Wanikawa, K. Hosoi and T. Kato, *Journal of the American Society of Brewing Chemists*, 2000, **58**, 51–56.

(25) R. Aylott and W. Mackenzie, *Journal of the Institute of Brewing*, 2010, **116**, 215–229.

(26) D. A. Nicol, in *Whisky*, Elsevier, 2014, pp. 155–178.

(27) B. Harrison, O. Fagnen, F. Jack and J. Brosnan, *Journal of the Institute of Brewing*, 2011, **117**, 106–112.

(28) J. R. Piggot, J. M. Conner, A. Paterson and J. Clyne, *International Journal of Food Science & Technology*, 2007, **28**, 303–318.

(29) J. Conner, in *Whisky*, Elsevier, 2014, pp. 199–220.

(30) S. J. Macleod, in *Whisky*, Elsevier, 2014, pp. 221–227.

(31) R. Aylott, in *Whisky*, Elsevier, 2014, pp. 243–270.

(32) D. J. Mabberley, *The Plant-Book: A Portable Dictionary of the Vascular Plants*, Cambridge University Press, 1997.

(33) H. Chen, in *Biotechnology of Lignocellulose*, Springer Netherlands, 2014, pp. 25–71.

(34) J. H. Grabber, J. Ralph, R. D. Hatfield and S. Quideau, *Journal of Agricultural and Food Chemistry*, 1997, **45**, 2530–2532.

(35) B. Zhang, J. Cai, C.-Q. Duan, M. Reeves and F. He, *International Journal of Molecular Sciences*, 2015, **16**, 6978–7014.

(36) A. Golon and N. Kuhnert, *Food & Function*, 2013, **4**, 1040.

(37) M. Boscolo, L. G. Andrade-Sobrinho, B. S. Lima-Neto, D. W. Franco and M. M. C. Ferreira, *Journal of AOAC International*, 2002, **85**, 744–750.

(38) A. C. McIntyre, M. L. Bilyk, A. Nordon, G. Colquhoun and D. Littlejohn, *Analytica Chimica Acta*, 2011, **690**, 228–233.

(39) W. Meier-Augenstein, H. Kemp and S. Hardie, *Food Chemistry*, 2012, **133**, 1070–1074.

(40) R. I. Aylott, A. H. Clyne, A. P. Fox and D. A. Walker, *The Analyst*, 1994, **119**, 1741.

(41) W. M. MacKenzie and R. I. Aylott, *The Analyst*, 2004, **129**, 607.

(42) A. R. Martins, M. Talhavini, M. L. Vieira, J. J. Zacca and J. W. B. Braga, *Food Chemistry*, 2017, **229**, 142–151.

(43) P. Wiśniewska, M. Śliwińska, T. Dymerski, W. Wardencki and J. Namieśnik, *Food Analytical Methods*, 2015, **8**, 2000–2010.

(44) P. J. Lehtonen, L. A. Keller and E. T. Ali-Mattila, *Zeitschrift für Lebensmitteluntersuchung und -Forschung A*, 1999, **208**, 413–417.

(45) J. Saurina, *TrAC Trends in Analytical Chemistry*, 2010, **29**, 234–245.

(46) J. Riedl, S. Esslinger and C. Fauhl-Hassek, *Analytica Chimica Acta*, 2015, **885**, 17–32.

(47) P. C. Ashok, B. B. Praveen and K. Dholakia, *Optics Express*, 2011, **19**, 22982.

(48) J. Kiefer and A. L. Cromwell, *Analytical Methods*, 2017, **9**, 511–518.

(49) D. Picque, P. Lieben, G. Corrieu, R. Cantagrel, O. Lablanquie and G. Snakkers, *Journal of Agricultural and Food Chemistry*, 2006, **54**, 5220–5226.

(50) C. A. Shand, R. Wendler, L. Dawson, K. Yates and H. Stephenson, *Analytica Chimica Acta*, 2017, **976**, 14–24.

(51) P. Wiśniewska, R. Boqué, E. Borràs, O. Busto, W. Wardencki, J. Namieśnik and T. Dymerski, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2017, **173**, 849–853.

(52) J. K. S. Møller, R. R. Catharino and M. N. Eberlin, *The Analyst*, 2005, **130**, 890.

(53) T. S. Collins, J. Zweigenbaum and S. E. Ebeler, *Food Chemistry*, 2014, **163**, 186–196.

(54) J. S. Garcia, B. G. Vaz, Y. E. Corilo, C. F. Ramires, S. A. Saraiva, G. B. Sanvido, E. M. Schmidt, D. R. Maia, R. G. Cosso, J. J. Zacca and M. N. Eberlin, *Food Research International*, 2013, **51**, 98–106.

(55) H. J. Cooper and A. G. Marshall, *Journal of Agricultural and Food Chemistry*, 2001, **49**, 5710–5718.

(56) C. Roullier-Gall, M. Witting, R. D. Gougeon and P. Schmitt-Kopplin, *Frontiers in Chemistry*, 2014, **2**, DOI: 10.3389/fchem.2014.00102.

(57) R. D. Gougeon, M. Lucio, M. Frommberger, D. Peyron, D. Chassagne, H. Alexandre, F. Feuillat, A. Voilley, P. Cayot, I. Gebefugi, N. Hertkorn and P. Schmitt-Kopplin, *Proceedings of the National Academy of Sciences*, 2009, **106**, 9174–9179.

(58) R. Gougeon, M. Lucio, A. De Boel, M. Frommberger, N. Hertkorn, D. Peyron, D. Chassagne, F. Feuillat, P. Cayot, A. Voilley, I. Gebefügi and P. Schmitt-Kopplin, *Chemistry - A European Journal*, 2009, **15**, 600–611.

(59) C. Roullier-Gall, M. Lucio, L. Noret, P. Schmitt-Kopplin and R. D. Gougeon, *PLoS ONE*, 2014, **9**, ed. A. C. Gill, e97615.

(60) C. Roullier-Gall, L. Boutegrabet, R. D. Gougeon and P. Schmitt-Kopplin, *Food Chemistry*, 2014, **152**, 100–107.

(61) C. Roullier-Gall, J. Signoret, D. Hemmler, M. A. Witting, B. Kanawati, B. Schafer, R. D. Gougeon and P. Schmitt-Kopplin, *Frontiers in Chemistry*, 2018, **6**, DOI: 10.3389/fchem.2018.00029.

(62) A. S. Araújo, L. L. da Rocha, D. M. Tomazela, A. C. H. F. Sawaya, R. R. Almeida, R. R. Catharino and M. N. Eberlin, *The Analyst*, 2005, **130**, 884.

(63) J. A. R. Teodoro, H. V. Pereira, M. M. Sena, E. Piccin, J. J. Zacca and R. Augusti, *Food Chemistry*, 2017, **237**, 1058–1064.

(64) F. Tosato, R. M. Correia, B. G. Oliveira, A. M. Fontes, H. S. França, W. K. T. Coltro, P. R. Filgueiras and W. Romão, *Analytical Methods*, 2018, **10**, 1952–1960.

(65) I. Parker, S. Kelly, M. Sharman, M. Dennis and D. Howie, *Food Chemistry*, 1998, **63**, 423–428.

(66) G. J. Martin, M. L. Martin, F. Mabon and M. J. Michon, *Journal of Agricultural and Food Chemistry*, 1983, **31**, 311–315.

(67) T. Kuballa, T. Hausler, A. O. Okaru, M. Neufeld, K. O. Abuga, I. O. Kibwage, J. Rehm, B. Luy, S. G. Walch and D. W. Lachenmeier, *Food Chemistry*, 2018, **245**, 112–118.

(68) A. Nose and M. Hojo, *Journal of Bioscience and Bioengineering*, 2006, **102**, 269–280.

(69) A. Nose, M. Hojo, M. Suzuki and T. Ueda, *Journal of Agricultural and Food Chemistry*, 2004, **52**, 5359–5365.

(70) A. Nose, M. Myojin, M. Hojo, T. Ueda and T. Okuda, *Journal of Bioscience and Bioengineering*, 2005, **99**, 493–501.

(71) J. Polak and M. Bartoszek, *Journal of Food Composition and Analysis*, 2015, **40**, 114–119.

(72) A. B. Koenig, R. L. Sleighter, E. Salmon and P. G. Hatcher, *Journal of Wood Chemistry and Technology*, 2010, **30**, 61–85.

(73) A. Weekley, P. Bruins, M. Sisto and M. Augustine, *Journal of Magnetic Resonance*, 2003, **161**, 91–98.

(74) A. J. Simpson and S. A. Brown, *Journal of Magnetic Resonance*, 2005, **175**, 340–346.

(75) C. T. Lesar, J. Decatur, E. Lukasiewicz and E. Champeil, *Forensic Science International*, 2011, **212**, e40–e45.

(76) Y. B. Monakhova, H. Schäfer, E. Humpfer, M. Spraul, T. Kuballa and D. W. Lachenmeier, *Magnetic Resonance in Chemistry*, 2011, **49**, 734–739.

(77) R. T. Mckay, *Concepts in Magnetic Resonance Part A*, 2011, **38A**, 197–220.

(78) C. Fotakis, D. Christodouleas, K. Kokkotou, M. Zervou, P. Zoumpoulakis, P. Moulos, M. Liouni and A. Calokerinos, *Food Chemistry*, 2013, **138**, 1837–1846.

(79) R. Godelmann, F. Fang, E. Humpfer, B. Schütz, M. Bansbach, H. Schäfer and M. Spraul, *Journal of Agricultural and Food Chemistry*, 2013, **61**, 5610–5619.

(80) M. Martin-Pastor, E. Guitian and R. Riguera, *Analytical Chemistry*, 2016, **88**, 6239–6246.

(81) I. Duarte, A. Barros, P. S. Belton, R. Righelato, M. Spraul, E. Humpfer and A. M. Gil, *Journal of Agricultural and Food Chemistry*, 2002, **50**, 2475–2481.

(82) L. Mannina, F. Marini, R. Antiochia, S. Cesa, A. Magri, D. Capitani and A. P. Sobolev, *ELECTROPHORESIS*, 2016, **37**, 2710–2719.

(83) C. Sánchez-Estébanez, S. Ferrero, C. M. Alvarez, F. Villafañe, I. Caballero and C. A. Blanco, *Food Analytical Methods*, 2017, **11**, 11–22.

(84) G. D. Poggetto, L. Castañar, R. W. Adams, G. A. Morris and M. Nilsson, *Chemical Communications*, 2017, **53**, 7461–7464.

(85) H.-G. Schmarr, M. Mathes, K. Wall, F. Metzner and M. Fraefel, *Journal of Chromatography A*, 2017, **1516**, 135–141.

(86) E. Masson, R. Baumes, C. L. Guernevé and J.-L. Puech, *Journal of Agricultural and Food Chemistry*, 2000, **48**, 4306–4309.

(87) J.-L. Puech, C. Mertz, V. Michon, C. L. Guernevé, T. Doco and C. H. du Penhoat, *Journal of Agricultural and Food Chemistry*, 1999, **47**, 2060–2066.

(88) A. Glabasnia and T. Hofmann, *Journal of Agricultural and Food Chemistry*, 2007, **55**, 4109–4118.

(89) A. Marchal, P. Waffo-Téguo, E. Génin, J.-M. Mérillon and D. Dubourdieu, *Analytical Chemistry*, 2011, **83**, 9629–9637.

(90) L. Sindt, M. Gammacurta, P. Waffo-Teguo, D. Dubourdieu and A. Marchal, *Journal of Natural Products*, 2016, **79**, 2432–2438.

(91) M. Fujieda, T. Tanaka, Y. Suwa, S. Koshimizu and I. Kouno, *Journal of Agricultural and Food Chemistry*, 2008, **56**, 7305–7310.

(92) Y. Matsuo, H. Wakamatsu, M. Omar and T. Tanaka, *Organic Letters*, 2014, **17**, 46–49.

(93) P. Hore, *Nuclear Magnetic Resonance*, Oxford University Press, 2nd edn., 2015, 128 pp.

(94) Keeler, *Understanding NMR Spectroscopy*, John Wiley & Sons, 2nd edn., 2010, 526 pp.

(95) M. H. Levitt, *Spin Dynamics*, Wiley John + Sons, 2001, 744 pp.

(96) T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier LTD, Oxford, 2016.

(97) M. Berglund and M. E. Wieser, *Pure and Applied Chemistry*, 2011, **83**, 397–410.

(98) P. Hore, in, Oxford University Press, 2nd edn., 2015, ch. Introduction, pp. 1–9.

(99) T. D. Claridge, in *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier, 2016, pp. 11–59.

(100) M. J. Thrippleton and J. Keeler, *Angewandte Chemie International Edition*, 2003, **42**, 3938–3941.

(101) J. Keeler, in, Wiley, 2nd edn., 2010, ch. Energy levels and NMR spectra, pp. 23–46.

(102) T. D. Claridge, in *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier, 2016, pp. 133–169.

(103) G. Zheng and W. S. Price, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2010, **56**, 267–288.

(104) T. D. Claridge, in *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier, 2016, pp. 457–498.

(105) D. Neuhaus, I. M. Ismail and C.-W. Chung, *Journal of Magnetic Resonance, Series A*, 1996, **118**, 256–263.

(106) M. Hall, *First Break*, 2006, **24**, DOI: 10.3997/1365-2397.2006027.

(107) M. Levitt, in, Wiley John + Sons, 2001, ch. Relaxation, pp. 513–570.

(108) T. D. Claridge, in *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier, 2016, pp. 203–241.

(109) C. Chung, S. Wimperis and J. Keeler, *Journal of Magnetic Resonance, Series A*, 1995, **114**, 188–200.

(110) P. T. Robinson, T. N. Pham and D. Uhrín, *Journal of Magnetic Resonance*, 2004, **170**, 97–103.

(111) C. Brissac, T. E. Malliavin and M. A. Delsuc, *Journal of Biomolecular NMR*, 1995, **6**, DOI: 10.1007/bf00197635.

(112) N. Esturau, F. Sánchez-Ferrando, J. A. Gavin, C. Roumestand, M.-A. Delsuc and T. Parella, *Journal of Magnetic Resonance*, 2001, **153**, 48–55.

(113) S. Balayssac, S. Trefi, V. Gilard, M. Malet-Martino, R. Martino and M.-A. Delsuc, *Journal of Pharmaceutical and Biomedical Analysis*, 2009, **50**, 602–612.

(114) A. Cherni, E. Chouzenoux and M.-A. Delsuc, *The Analyst*, 2017, **142**, 772–779.

(115) S. Balayssac, M.-A. Delsuc, V. Gilard, Y. Prigent and M. Malet-Martino, *Journal of Magnetic Resonance*, 2009, **196**, 78–83.

(116) R. Cao, A. Nonaka, F. Komura and T. Matsui, *Food Chemistry*, 2015, **171**, 8–12.

(117) M. Nilsson, I. F. Duarte, C. Almeida, I. Delgadillo, B. J. Goodfellow, A. M. Gil and G. A. Morris, *Journal of Agricultural and Food Chemistry*, 2004, **52**, 3736–3743.

(118) M. Delsuc and J. Lallemand, *Journal of Magnetic Resonance (1969)*, 1986, **69**, 504–507.

(119) E. Kupce and R. Freeman, *Journal of Magnetic Resonance, Series A*, 1995, **112**, 261–264.

(120) R. Ragone, P. Crupi, S. Piccinonna, C. Bergamini, F. Mazzone, F. P. Fanizzi, F. P. Schena and D. Antonacci, *Food Science and Biotechnology*, 2015, **24**, 817–826.

(121) T. Hwang and A. Shaka, *Journal of Magnetic Resonance, Series A*, 1995, **112**, 275–279.

(122) C. Dalvit, *Journal of Biomolecular NMR*, 1998, **11**, 437–444.

(123) R. W. Adams, C. M. Holroyd, J. A. Aguilar, M. Nilsson and G. A. Morris, *Chem. Commun.*, 2013, **49**, 358–360.

(124) R. Ogg, R. Kingsley and J. Taylor, *Journal of Magnetic Resonance, Series B*, 1994, **104**, 1–10.

(125) S. H. Smallcombe, S. L. Patt and P. A. Keifer, *Journal of Magnetic Resonance, Series A*, 1995, **117**, 295–303.

(126) C. Fotakis and M. Zervou, *Food Chemistry*, 2016, **196**, 760–768.

(127) P. S. C. Wu and G. Otting, *Journal of Biomolecular NMR*, 2005, **32**, 243–250.

(128) L. Lim, F. Yan, S. Bach, K. Pihakari and D. Klein, *International Journal of Molecular Sciences*, 2016, **17**, 104.

(129) R. T. Hilger, P. J. Wyss, R. E. Santini and S. A. McLuckey, *Analytical Chemistry*, 2013, **85**, 8075–8079.

(130) A. G. Marshall, C. L. Hendrickson and G. S. Jackson, *Mass Spectrometry Reviews*, 1998, **17**, 1–35.

(131) A. G. Marshall and C. L. Hendrickson, *International Journal of Mass Spectrometry*, 2002, **215**, 59–75.

(132) Y. Cho, A. Ahmed, A. Islam and S. Kim, *Mass Spectrometry Reviews*, 2014, **34**, 248–263.

(133) D. Cao, J. Lv, F. Geng, Z. Rao, H. Niu, Y. Shi, Y. Cai and Y. Kang, *Analytical Chemistry*, 2016, **88**, 12210–12218.

(134) J. B. Shaw, M. V. Gorshkov, Q. Wu and L. Paša-Tolić, *Analytical Chemistry*, 2018, **90**, 5557–5562.

(135) D. Nolting, R. Malek and A. Makarov, *Mass Spectrometry Reviews*, 2017, DOI: `10.1002/mas.21549`.

(136) B. Kanawati, T. M. Bader, K.-P. Wanczek, Y. Li and P. Schmitt-Kopplin, *Rapid Communications in Mass Spectrometry*, 2017, **31**, 1607–1615.

(137) S. Guan and A. G. Marshall, *Analytical Chemistry*, 1997, **69**, 1156–1162.

(138) R. Mathur and P. B. O'Connor, *Rapid Communications in Mass Spectrometry*, 2009, **23**, 523–529.

(139) E. Cho, M. Witt, M. Hur, M.-J. Jung and S. Kim, *Analytical Chemistry*, 2017, **89**, 12101–12107.

(140) S. C. Beu, G. T. Blakney, J. P. Quinn, C. L. Hendrickson and A. G. Marshall, *Analytical Chemistry*, 2004, **76**, 5756–5761.

(141) D. P. A. Kilgour, M. J. Neal, A. J. Soulby and P. B. O'Connor, *Rapid Communications in Mass Spectrometry*, 2013, **27**, 1977–1982.

(142) D. P. A. Kilgour and S. L. V. Orden, *Rapid Communications in Mass Spectrometry*, 2015, **29**, 1009–1018.

(143) D. P. A. Kilgour, K. O. Nagornov, A. N. Kozhinov, K. O. Zhurov and Y. O. Tsybin, *Rapid Communications in Mass Spectrometry*, 2015, **29**, 1087–1093.

(144) D. F. Smith, A. Kharchenko, M. Konijnenburg, I. Klinkert, L. Paša-Tolić and R. M. A. Heeren, *Journal of The American Society for Mass Spectrometry*, 2012, **23**, 1865–1872.

(145) X. Lou, J. L. J. Dongen and E. W. Meijer, *Journal of the American Society for Mass Spectrometry*, 2010, **21**, 1223–1226.

(146) A. N. Kozhinov, K. O. Zhurov and Y. O. Tsybin, *Analytical Chemistry*, 2013, **85**, 6437–6445.

(147) S. Guan and A. G. Marshall, *Analytical Chemistry*, 1993, **65**, 1288–1294.

(148) M. A. van Agthoven, M.-A. Delsuc, G. Bodenhausen and C. Rolando, *Analytical and Bioanalytical Chemistry*, 2012, **405**, 51–61.

(149) M. Witt, J. Fuchser and B. P. Koch, *Analytical Chemistry*, 2009, **81**, 2688–2694.

(150) T. Gruendling, S. Weidner, J. Falkenhagen and C. Barner-Kowollik, *Polymer Chemistry*, 2010, **1**, 599.

(151) A. Fievre, T. Solouki, A. G. Marshall and W. T. Cooper, *Energy & Fuels*, 1997, **11**, 554–560.

(152) N. R. Novotny, E. N. Capley and A. C. Stenson, *Journal of Mass Spectrometry*, 2014, **49**, 316–326.

(153) M. Glückmann, A. Pfenninger, R. Krüger, M. Thierolf, M. Karasa, V. Horneffer, F. Hillenkamp and K. Strupat, *International Journal of Mass Spectrometry*, 2001, **210-211**, 121–132.

(154) M. Karas and R. Krüger, *Chemical Reviews*, 2003, **103**, 427–440.

(155) D. Cao, H. Huang, M. Hu, L. Cui, F. Geng, Z. Rao, H. Niu, Y. Cai and Y. Kang, *Analytica Chimica Acta*, 2015, **866**, 48–58.

(156) J. W. T. Blackburn, W. Kew, M. C. Graham and D. Uhrín, *Analytical Chemistry*, 2017, **89**, 4382–4386.

(157) J. Meija, T. B. Coplen, M. Berglund, W. A. Brand, P. D. Bièvre, M. Gröning, N. E. Holden, J. Irrgeher, R. D. Loss, T. Walczyk and T. Prohaska, *Pure and Applied Chemistry*, 2016, **88**, DOI: 10.1515/pac-2015-0305.

(158) R. K. Harris, E. D. Becker, S. M. C. de Menezes, R. Goodfellow and P. Granger, *Pure and Applied Chemistry*, 2001, **73**, 1795–1818.

(159) C. A. Hughey, C. L. Hendrickson, R. P. Rodgers, A. G. Marshall and K. Qian, *Analytical Chemistry*, 2001, **73**, 4676–4681.

(160) E. B. Kujawinski and M. D. Behn, *Analytical Chemistry*, 2006, **78**, 4363–4373.

(161) D. P. A. Kilgour, C. L. Mackay, P. R. R. Langridge-Smith and P. B. O'Connor, *Analytical Chemistry*, 2012, **84**, 7431–7435.

(162) R. P. Rodgers and A. G. Marshall, in *Asphaltenes, Heavy Oils, and Petroleomics*, Springer New York, 2007, pp. 63–93.

(163) A. C. Stenson, A. G. Marshall and W. T. Cooper, *Analytical Chemistry*, 2003, **75**, 1275–1284.

(164) B. P. Koch and T. Dittmar, *Rapid Communications in Mass Spectrometry*, 2006, **20**, 926–932.

(165) D. Van Krevelen, *Fuel*, 1950, **29**, 269–284.

(166) S. Kim, R. W. Kramer and P. G. Hatcher, *Analytical Chemistry*, 2003, **75**, 5336–5344.

(167) R. L. and P. G., in *Fourier Transforms - Approach to Scientific Principles*, InTech, 2011.

(168) A. Rivas-Ubach, Y. Liu, T. S. Bianchi, N. Tolić, C. Jansson and L. Paša-Tolić, *Analytical Chemistry*, 2018, **90**, 6152–6160.

(169) W. Kew, J. W. Blackburn, D. J. Clarke and D. Uhrín, *Rapid Communications in Mass Spectrometry*, 2017, **31**, 658–662.

(170) S. A. Brockman, E. V. Roden and A. D. Hegeman, *Metabolomics*, 2018, **14**, DOI: 10.1007/s11306-018-1343-y.

(171) O. Cloarec, M.-E. Dumas, A. Craig, R. H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes and J. Nicholson, *Analytical Chemistry*, 2005, **77**, 1282–1289.

(172) S. Šašić, A. Muszynski and Y. Ozaki, *The Journal of Physical Chemistry A*, 2000, **104**, 6380–6387.

(173) I. Noda, *Journal of the American Chemical Society*, 1989, **111**, 8116–8118.

(174) Y.-S. Hong, C. Cilindre, G. Liger-Belair, P. Jeandet, N. Hertkorn and P. Schmitt-Kopplin, *Journal of Agricultural and Food Chemistry*, 2011, **59**, 7237–7245.

(175) L. M. Smith, A. D. Maher, O. Cloarec, M. Rantalainen, H. Tang, P. Elliott, J. Stamler, J. C. Lindon, E. Holmes and J. K. Nicholson, *Analytical Chemistry*, 2007, **79**, 5682–5689.

(176) H. C. Keun, T. J. Athersuch, O. Beckonert, Y. Wang, J. Saric, J. P. Shockcor, J. C. Lindon, I. D. Wilson, E. Holmes and J. K. Nicholson, *Analytical Chemistry*, 2008, **80**, 1073–1079.

(177) D. J. Crockford, E. Holmes, J. C. Lindon, R. S. Plumb, S. Zirah, S. J. Bruce, P. Rainville, C. L. Stumpf and J. K. Nicholson, *Analytical Chemistry*, 2006, **78**, 363–371.

(178) S. Wei, J. Zhang, L. Liu, T. Ye, G. A. N. Gowda, F. Tayyari and D. Raftery, *Analytical Chemistry*, 2011, **83**, 7616–7623.

(179) D. Cañueto, J. Gómez, R. M. Salek, X. Correig and N. Cañellas, *Metabolomics*, 2018, **14**, DOI: 10.1007/s11306-018-1319-y.

(180) R. Bro and A. K. Smilde, *Anal. Methods*, 2014, **6**, 2812–2831.

(181) H. Abdi and L. J. Williams, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, **2**, 433–459.

(182) J. Shlens, *arXiv:1404.1100* [*cs.LG*], 2014.

(183) R. A. van den Berg, H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde and M. J. van der Werf, *BMC Genomics*, 2006, **7**, 142.

(184) J. E. Jackson, *A User's Guide to Principal Components*, Wiley-Interscience, 1991.

(185) J. Zhong, N. DiDonato and P. G. Hatcher, *Journal of Chemometrics*, 2012, **26**, 150–157.

(186) Y. B. Monakhova, R. Godelmann, T. Kuballa, S. P. Mushtakova and D. N. Rutledge, *Talanta*, 2015, **141**, 60–65.

(187) R. G. Brereton and G. R. Lloyd, *Journal of Chemometrics*, 2014, **28**, 213–225.

(188) J. Trygg and S. Wold, *Journal of Chemometrics*, 2002, **16**, 119–128.

(189) B. Worley and R. Powers, *Current Metabolomics*, 2016, **4**, 97–103.

(190) M. N. Triba, L. L. Moyec, R. Amathieu, C. Goossens, N. Bouchemal, P. Nahon, D. N. Rutledge and P. Savarin, *Molecular BioSystems*, 2015, **11**, 13–19.

(191) L. Eriksson, J. Trygg and S. Wold, *Journal of Chemometrics*, 2008, **22**, 594–600.

(192) M. H. Levitt, R. Freeman and T. Frenkiel, in *Advances in Magnetic Resonance*, ACADEMIC PRESS, INC., 1983, vol. 11, pp. 47–110.

(193) R. Tycko, A. Pines and J. Guckenheimer, *The Journal of Chemical Physics*, 1985, **83**, 2775–2802.

(194) A. J. Shaka, C. J. Lee and A. Pines, *Journal of Magnetic Resonance (1969)*, 1988, **77**, 274–293.

(195) V. V. Krishnamurthy, *Magnetic Resonance in Chemistry*, 1997, **35**, 9–12.

(196) A. G. Palmer, J. Cavanagh, P. E. Wright and M. Rance, *Journal of Magnetic Resonance (1969)*, 1991, **93**, 151–170.

(197) L. E. Kay, P. Keifer and T. Saarinen, *Journal of the American Chemical Society*, 1992, 10663–10665.

(198) J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S. Glaser, O. Sørensen and C. Griesinger, *Journal of Biomolecular NMR*, 1994, **4**, 301–306.

(199) W. Willker, D. Leibfritz, R. Kerssebaum and W. Bermel, *Magnetic Resonance in Chemistry*, 1993, **31**, 287–292.

(200) C. Zwahlen, P. Legault, S. J. F. Vincent, J. Greenblatt, R. Konrat and L. E. Kay, *Journal of the American Chemical Society*, 1997, **119**, 6711–6721.

(201) R. D. Boyer, R. Johnson and K. Krishnamurthy, *Journal of Magnetic Resonance*, 2003, **165**, 253–259.

(202) T. D. W. Claridge and I. Pérez-Victoria, *Org. Biomol. Chem.*, 2003, **1**, 3632–3634.

(203) D. O. Cicero, G. Barbato and R. Bazzo, *Journal of Magnetic Resonance*, 2001, **148**, 209–213.

(204) J. A. Aguilar, R. W. Adams, M. Nilsson and G. A. Morris, *Journal of Magnetic Resonance*, 2014, **238**, 16–19.

(205) L. Castañar, G. D. Poggetto, A. A. Colbourne, G. A. Morris and M. Nilsson, *Magnetic Resonance in Chemistry*, 2018, **56**, 546–558.

(206) T. Riedel and T. Dittmar, *Analytical Chemistry*, 2014, **86**, 8376–8382.

(207) N. Tolić, Y. Liu, A. Liyu, Y. Shen, M. M. Tfaily, E. B. Kujawinski, K. Longnecker, L.-J. Kuo, E. W. Robinson, L. Paša-Tolić and N. J. Hess, *Analytical Chemistry*, 2017, **89**, 12659–12665.

(208) R. Freeman, T. H. Mareci and G. A. Morris, *Journal of Magnetic Resonance (1969)*, 1981, **42**, 341–345.

(209) A. Shaka and R. Freeman, *Journal of Magnetic Resonance (1969)*, 1982, **50**, 502–507.

(210) T. Parella and J. Belloc, *Magnetic Resonance in Chemistry*, 2001, **39**, 311–315.

(211) G. A. Morris and R. Freeman, *Journal of the American Chemical Society*, 1979, **101**, 760–762.

(212) D. L. Mattiello, Warren, L. Mueller and B. T. Farmer, *Journal of the American Chemical Society*, 1996, **118**, 3253–3261.

(213) J. Kikuchi, Y. Tsuboi, K. Komatsu, M. Gomi, E. Chikayama and Y. Date, *Analytical Chemistry*, 2016, **88**, 659–665.

(214) R. Novoa-Carballal, E. Fernandez-Megia, C. Jimenez and R. Riguera, *Nat. Prod. Rep.*, 2011, **28**, 78–98.

(215) A. Kumar, *Journal of Magnetic Resonance (1969)*, 1978, **30**, 227–249.

(216) I. E. Ndukwe, A. Shchukina, K. Kazimierczuk, C. Cobas and C. P. Butts, *ChemPhysChem*, 2016, **17**, 2799–2803.

(217) M. Foroozandeh, R. W. Adams, N. J. Meharry, D. Jeannerat, M. Nilsson and G. A. Morris, *Angewandte Chemie International Edition*, 2014, **53**, 6990–6992.

(218) *Specifications for AVANCE III HD*, Bruker BioSpin, 2011.

(219) S. Grosch, Y. B. Monakhova, T. Kuballa, W. Ruge, R. Kimmich and D. W. Lachenmeier, *European Food Research and Technology*, 2012, **236**, 267–275.

(220) A. S. Gray, *The Scotch Whisky Industry Review*, Pagoda Scotland Ltd, 34th, 2009.

(221) A. S. Gray, *The Scotch Whisky Industry Review*, Pagoda Scotland Ltd, 36th, 2011.

(222) A. S. Gray, *The Scotch Whisky Industry Review*, Pagoda Scotland Ltd, 36th, 2013.

(223) G. D. Poggetto, L. Castañar, G. A. Morris and M. Nilsson, *RSC Advances*, 2016, **6**, 100063–100066.

(224) W. A. Bubb, *Concepts in Magnetic Resonance*, 2003, **19A**, 1–19.

(225) T. Barclay, M. Ginic-Markovic, M. R. Johnston, P. Cooper and N. Petrovsky, *Carbohydrate Research*, 2012, **347**, 136–141.

(226) Y. Zhu, J. Zajicek and A. S. Serianni, *The Journal of Organic Chemistry*, 2001, **66**, 6244–6251.

(227) B. O. Petersen, M. S. Motawie, B. L. Møller, O. Hindsgaul and S. Meier, *Carbohydrate Research*, 2015, **403**, 149–156.

(228) A. Scheithauer, E. von Harbou, H. Hasse, T. Grützner, C. Rijksen, D. Zollinger and W. R. Thiel, *AIChE Journal*, 2014, **61**, 177–187.

(229) S. Jaubert and G. Maurer, *The Journal of Chemical Thermodynamics*, 2014, **68**, 332–342.

(230) M. Foroozandeh, L. Castañar, L. G. Martins, D. Sinnaeve, G. D. Poggetto, C. F. Tormena, R. W. Adams, G. A. Morris and M. Nilsson, *Angewandte Chemie International Edition*, 2016, **55**, 15579–15582.

(231) R. Evans, G. D. Poggetto, M. Nilsson and G. A. Morris, *Analytical Chemistry*, 2018, **90**, 3987–3994.

(232) P. H. C. Eilers, *Analytical Chemistry*, 2003, **75**, 3631–3636.

(233) C. Cobas, *Magnetic Resonance in Chemistry*, 2018, DOI: 10.1002/mrc.4747.

(234) S. Ravanbakhsh, P. Liu, T. C. Bjordahl, R. Mandal, J. R. Grant, M. Wilson, R. Eisner, I. Sinelnikov, X. Hu, C. Luchinat, R. Greiner and D. S. Wishart, *PLOS ONE*, 2015, **10**, ed. D. Monleon, e0124219.

(235) J. Hao, W. Astle, M. D. Iorio and T. M. D. Ebbels, *Bioinformatics*, 2012, **28**, 2088–2090.

(236) F. Savorani, G. Tomasi and S. Engelsen, *Journal of Magnetic Resonance*, 2010, **202**, 190–202.

(237) C. Cobas, F. Seoane and S. Sykora, *Global Spectral Deconvolution (GSD) of 1D-NMR spectra*, Poster at SMASH Conference, Santa Fe (NM, USA), 2008.

(238) B. Worley and R. Powers, *ACS Chemical Biology*, 2014, **9**, 1138–1144.

(239) G. Walker and A. Hill, *Beverages*, 2016, **2**, 38.

(240) W. M. Ingledew, in *The Alcohol Textbook: Ethanol Production by Fermentation and Distillation*, ed. T. Lyons, D. Kelsall and J. Murtagh, Nottingham University Press, 1995, pp. 55–79.

(241) Y. B. Monakhova, A. M. Tsikin, T. Kuballa, D. W. Lachenmeier and S. P. Mushtakova, *Magnetic Resonance in Chemistry*, 2014, **52**, 231–240.

(242) A. C. Stenson, W. M. Landing, A. G. Marshall and W. T. Cooper, *Analytical Chemistry*, 2002, **74**, 4397–4409.

(243) A. M. Kellerman, T. Dittmar, D. N. Kothawala and L. J. Tranvik, *Nature Communications*, 2014, **5**, DOI: 10.1038/ncomms4804.

(244) D. Miura, Y. Tsuji, K. Takahashi, H. Wariishi and K. Saito, *Analytical Chemistry*, 2010, **82**, 5887–5891.

(245) K. Park, M. Kim, S. Baek, I. Bae, S.-W. Seo, J. Kim, Y. Shin, Y.-M. Lee and H. Kim, *Plant Methods*, 2013, **9**, 15.

(246) T. Nagao, D. Yukihira, Y. Fujimura, K. Saito, K. Takahashi, D. Miura and H. Wariishi, *Analytica Chimica Acta*, 2014, **813**, 70–76.

(247) S. M. Miladinović, A. N. Kozhinov, M. V. Gorshkov and Y. O. Tsybin, *Analytical Chemistry*, 2012, **84**, 4042–4051.

(248) J.-L. Reymond, *Accounts of Chemical Research*, 2015, **48**, 722–730.

(249) N. Helle, M. Baden, K. MacNamara and D. Dabrowska, *LCGC Europe*, 2011, **24**, 448–467.

(250) A. Marchal, E. Génin, P. Waffo-Téguo, A. Bibès, G. D. Costa, J.-M. Mérillon and D. Dubourdieu, *Analytica Chimica Acta*, 2015, **853**, 425–434.

(251) A. Marchal, A. Prida and D. Dubourdieu, *Journal of Agricultural and Food Chemistry*, 2016, **64**, 618–626.

(252) ChemSpider, *Ellagic Acid*, `http://www.chemspider.com/Chemical-Structure.4445149.html`, Accessed: 16/06/2018.

(253) W.-H. Cai, K. Matsunami, H. Otsuka, T. Shinzato and Y. Takeda, *Journal of Natural Medicines*, 2009, **63**, 408–414.

(254) G. Steele, R. Fotheringam and F. Jack, in *Distilled Spirits: Tradition and Innovation*, ed. J. H. Bryce and G. G. Stewart, Nottingham University Press, 2004, pp. 161–168.

(255) L. Poisson and P. Schieberle, *Journal of Agricultural and Food Chemistry*, 2008, **56**, 5820–5826.

(256) B. Worley and R. Powers, *Chemometrics and Intelligent Laboratory Systems*, 2015, **149**, 33–39.

(257) D. D. Marshall, S. Lei, B. Worley, Y. Huang, A. Garcia-Garcia, R. Franco, E. D. Dodds and R. Powers, *Metabolomics*, 2014, **11**, 391–402.

(258) B. M. Ruddy, C. L. Hendrickson, R. P. Rodgers and A. G. Marshall, *Energy & Fuels*, 2018, **32**, 2901–2907.

(259) V. V. Lobodin, P. Juyal, A. M. McKenna, R. P. Rodgers and A. G. Marshall, *Analytical Chemistry*, 2013, **85**, 7803–7808.

(260) K. T. Lu, G. C. Eiden and J. C. Weisshaar, *The Journal of Physical Chemistry*, 1992, **96**, 9742–9748.

(261) J. W. Rabalais, T. P. Debies, J. L. Berkosky, J.-T. J. Huang and F. O. Ellison, *The Journal of Chemical Physics*, 1974, **61**, 516–528.

(262) K. M. A. Refaey and W. A. Chupka, *The Journal of Chemical Physics*, 1968, **48**, 5205–5219.

(263) W. Kew, J. W. T. Blackburn and D. Uhrín, *Analytical Chemistry*, 2018, **90**, 5968–5971.

(264) A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot and H. Pfister, *IEEE Transactions on Visualization and Computer Graphics*, 2014, **20**, 1983–1992.

(265) H. Chen and P. C. Boutros, *BMC Bioinformatics*, 2011, **12**, 35.

(266) L. Schweihard, M. Lindinger and H.-J. Kluge, *International Journal of Mass Spectrometry and Ion Processes*, 1990, **98**, 25–33.

(267) L. Schweikhard, *International Journal of Mass Spectrometry and Ion Processes*, 1991, **107**, 281–292.

(268) J. D. Hunter, *Computing In Science & Engineering*, 2007, **9**, 90–95.

(269) Bokeh Development Team, *Bokeh: Python library for interactive visualization*, 2014.

(270) T. Kind and O. Fiehn, *BMC Bioinformatics*, 2007, **8**, 105.

(271) L. Chiron, M.-A. Coutouly, J.-P. Starck, C. Rolando and M.-A. Delsuc, *arXiv:1608.06777* [*physics.comp-ph*], 2016.

Part III

APPENDIX

The Angels' Share

A

APPENDIX: NMR ASSIGNMENTS

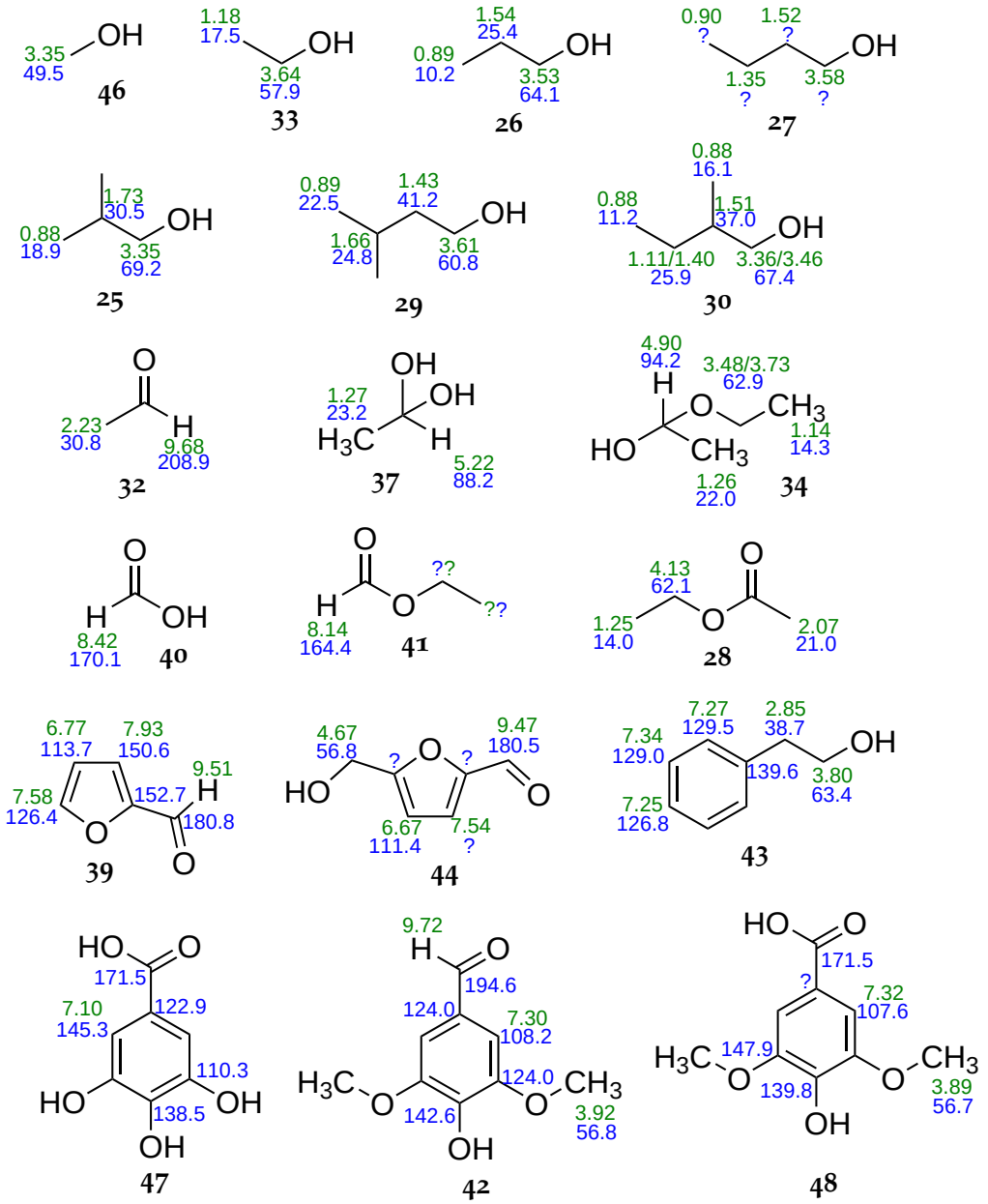


Figure A.1: NMR assigned structures excluding carbohydrates. Tabulated data, including names and coupling constants, is available in Table A.1. Green indicates ^{1}H chemical shifts, blue indicates ^{13}C chemical shifts. Unknown values are indicated with a question mark.

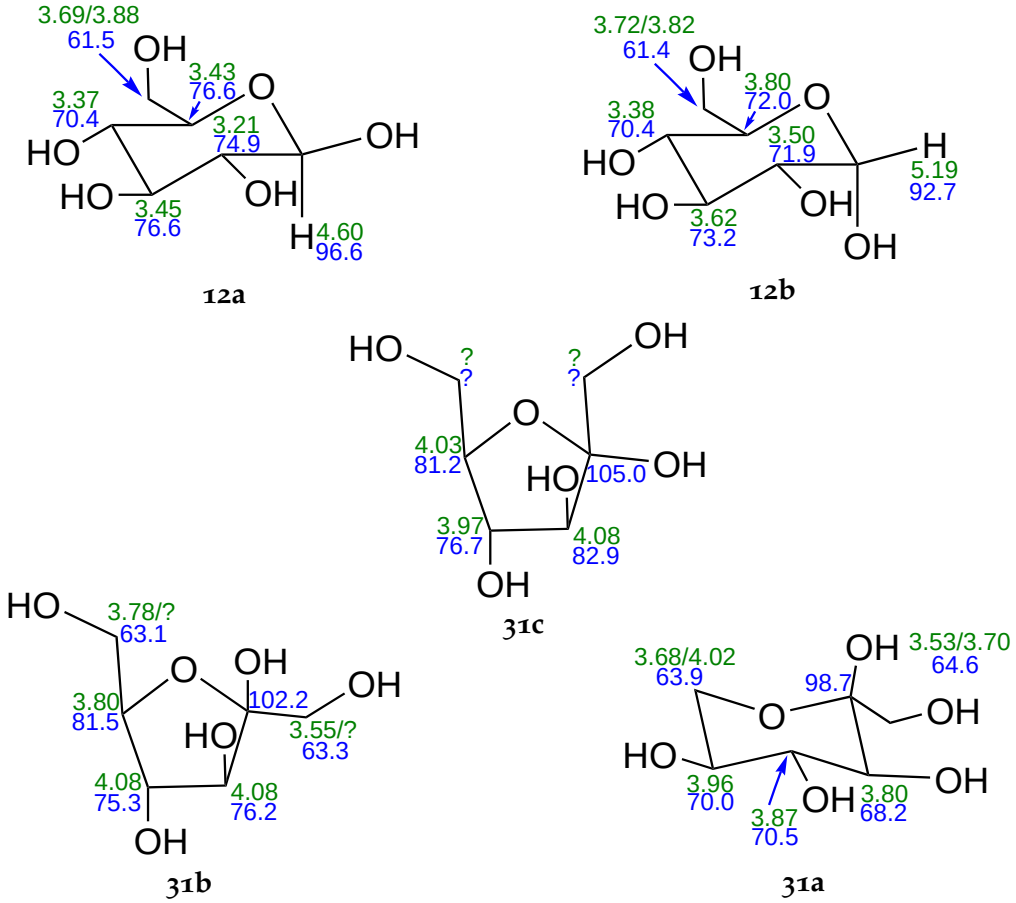


Figure A.2: NMR assigned carbohydrate structures. Tabulated data, including names and coupling constants, is available in Table A.1. Green indicates ^{1}H chemical shifts, blue indicates ^{13}C chemical shifts. Unknown values are indicated with a question mark.

Table A.1: NMR assignments including ^{13}C and ^{1}H chemical shifts and coupling constants. Not all values were measured or measurable, due to low SNR or overlapping signals.

COMPOUND	^{13}C / PPM	^{1}H / PPM	J / HZ	M	GROUP	H
water (**36**)	-	4.70	-	s	OH	1
methanol (**46**)	49.5	3.35	-	s	CH_3	3
ethanol (**33**)	16.9	1.19	7.1	t	CH_3	3
	57.5	3.64	7.1	q	CH_2	2
n-propanol (**26**)	10.2	0.89	7.5	t	CH_3	3
	25.4	1.54	-	m	CH_2	2
	64.1	3.53	6.7	t	CH_2	2
n-butanol (**27**)	?	0.90	7.4	t	CH_3	3
	?	1.35	-	m	CH_2	2
	?	1.52	-	m	CH_2	2
	?	3.58	6.7	t	CH_2	2
iso-butanol (**25**)	18.9	0.88	6.7	d	CH_3	6
	30.5	1.73	-	m	CH	1
	69.2	3.35	6.6	d	CH_2	2
2-methylbutanol (**30**)	11.2	0.88	6.5	t	CH_3	3
	16.1	0.88	7.5	d	CH_3	3
	25.9	1.11/1.40	-	m	CH_2	2
	37.0	1.51	-	m	CH	1
	67.4	3.36	10.8,6.6	dd	CH_2	1
	67.4	3.46	10.7,5.9	dd	CH_2	1
3-methylbutanol (**29**)	22.5	0.89	6.7	d	CH_3	6
	24.8	1.66	-	m	CH	1
	41.2	1.43	6.7	q	CH_2	2
	60.8	3.61	?	t	CH_2	2
acetaldehyde (**32**)	30.8	2.23	3.0	d	CH_3	3

Continued on next page

Table A.1 – continued from previous page

COMPOUND	^{13}C / PPM	^{1}H / PPM	J / HZ	M	GROUP	H
	208.9	9.68	3.0	q	CH	1
water hemiacetal	23.2	1.27	5.2	d	CH_3	3
of acetaldehyde (**37**)	88.2	5.22	5.2	q	CH	1
ethanol hemiacetal	14.3	1.14	-	?	CH_3	3
of acetaldehyde (**34**)	22.0	1.26	5.2	d	CH_3	3
	62.9	3.48/3.73	9.7,7.1	dq	CH_2	2
	94.2	4.90	5.2	q	CH	1
ethyl acetate (**28**)	14.0	1.25	7.2	t	CH_3	3
	21.0	2.08	-	s	CH_3	3
	62.1	4.13	7.2	q	CH_2	2
formic acid (**40**)	170.1	8.42	-	s	CH	1
ethyl formate (**41**)	164.4	8.14	-	s	CH	1
	?	?	-	-	CH_2	2
	?	?	-	-	CH_3	3
furfural (**39**)	113.7	6.77	3.7,1.7	dd	CH	1
	126.4	7.58	3.7,0.8	dd	CH	1
	150.6	7.93	1.7,0.8	dt	CH	1
	152.7	-	-	-	C	0
	180.8	9.51	0.9	d	CH	1
5-HMF (**44**)	56.8	4.67	-	s	CH_2	2
	114.4	6.67	3.6	d	CH	1
	?	7.54	3.6	d	CH	1
	180.5	9.47	-	s	CH	1
	?	-	-	-	C	0
	?	-	-	-	C	0
2-phenylethanol (**43**)	129.0	7.33	-	m	CH	2
	129.5	7.26	-	m	CH	2

Continued on next page

Table A.1 – continued from previous page

COMPOUND	^{13}C / PPM	^{1}H / PPM	J / HZ	M	GROUP	H
	126.8	7.25	-	m	CH	1
	38.7	2.85	7.1	t	CH_2	2
	63.4	3.78	7.1	t	CH_2	2
	139.6	-	-	-	C	0
gallic acid (**47**)	110.3	-	-	-	C	0
	122.9	-	-	-	C	0
	138.5	-	-	-	C	0
	143.3	7.10	-	s	CH	2
	171.5	-	-	-	C	0
syringic acid (**48**)	56.7	3.89	-	s	CH_3	6
	107.6	7.32	-	s	CH	2
	139.8	-	-	-	C	0
	147.9	-	-	-	C	0
	171.5	-	-	-	C	0
	?	-	-	-	C	0
syringaldehyde (**42**)	56.8	3.92	-	s	CH_3	6
	108.2	7.30	-	s	CH	2
	124.0	-	-	-	C	0
	124.0	-	-	-	C	0
	142.6	-	-	-	C	0
	194.6	9.72	-	s	CH	1
α-glucopyranose (**12b**)	92.7	5.19	3.7	d	CH	1
	71.9	3.50	9.8,3.7	dd	CH	1
	73.2	3.62	-	m	CH	1
	70.4	3.38	-	m	CH	1
	72.0	3.80	-	m	CH	1
	61.4	3.72/3.82	?	dd	CH_2	2
β-glucopyranose (**12a**)	96.6	4.60	7.9	d	CH	1

Continued on next page

Table A.1 – continued from previous page

COMPOUND	^{13}C / PPM	^{1}H / PPM	J / HZ	M	GROUP	H
	74.9	3.21	9.3,7.9	dd	CH	1
	76.6	3.45	-	m	CH	1
	70.4	3.37	-	m	CH	1
	76.6	3.43	-	m	CH	1
	61.5	3.69	12.1,5.7	dd	CH_2	1
	61.5	3.88	12.1,2.1	dd	CH_2	1
β-fructopyranose (**31d**)	98.7	-	-	-	C	0
	64.6	3.53/3.70	?	?	CH_2	2
	68.2	3.80	?	?	CH	1
	70.5	3.87	?	?	CH	1
	70.0	3.96	?	?	CH	1
	63.9	3.68/4.02	?	?	CH_2	2
β-fructofuranose (**31b**)	102.2	-	-	-	C	0
	63.3	3.55/?	?	?	CH_2	2
	76.2	4.06	?	?	CH	1
	75.3	4.08	?	?	CH	1
	81.5	3.80	?	?	CH	1
	63.1	3.78/?	?	?	CH_2	2
α-fructofuranose (**31c**)	105.0	-	-	-	C	0
	?	?/?	?	?	CH_2	2
	82.9	4.08	?	?	CH	1
	76.7	3.97	?	?	CH	1
	81.2	4.03	?	?	CH	1
	?	?/?	?	?	CH_2	2

B

APPENDIX: SAMPLE INFORMATION

Table B.1: List of samples used in chapter 5 (NMR) and chapter 6 (FTMS). Details include only known information on samples. 'Region' indicates SWA defined geographical region of origin. 'Wood' represents the maturation, and finish if applicable, wood types - B = Bourbon, S = Sherry, A = Ale, P = Port.

ID	TYPE	REGION	AGE	PEAT	WOOD	ABV	NMR	FTMS
S10-0462	Blend					40	Y	
S10-0463	Blend					40	Y	
S10-0464	Blend					40	Y	
S10-1011	Blend					40	Y	
S10-1012	Blend					40	Y	
S10-1013	Blend					40	Y	
S10-1014	Blend					40	Y	
S10-1015	Blend		12			40	Y	
S10-1016	Malt	Speyside	12			40	Y	Y
S10-1017	Malt	Islay	16	Y	BS	43	Y	Y
S10-1018	Blend					43	Y	
S10-1019	Malt	Highland	10		S	40	Y	Y
S10-1020	Malt	Islay		Y		43	Y	Y
S10-1021	Blend					40	Y	
S10-1022	Malt	Highland	10			40	Y	Y
S10-1023	Malt	Highland	12		B	40	Y	Y
S10-1024	Blend					40	Y	
S10-1127	Malt	Islay	10	Y	B	40	Y	Y
S10-1131	Malt	Highland	12	Y	S	40	Y	Y
S10-1132	Blend		12			40	Y	
S10-1133	Blend					40	Y	Y
S10-1134	Blend					40	Y	
S10-1180	Malt	Highland	10		B	40	Y	Y

Continued on next page

Table B.1 – continued from previous page

ID	TYPE	REGION	AGE	PEAT	WOOD	ABV	NMR	FTMS
S10-1181	Blend					43	Y	
S10-1182	Blend					40	Y	
S10-1183	Malt	Highland	8			40	Y	Y
S10-1218	Malt	Highland	12		S	40	Y	Y
S10-1306	Malt	Highland	10			40	Y	Y
S10-1313	Malt	Lowland			BS	43	Y	Y
S10-1314	Malt	Lowland			B	40	Y	Y
S10-1315	Malt	Islay	12	Y		40	Y	Y
S10-1407	Blend					40	Y	
S10-1408	Malt	Highland	12		B	40	Y	Y
S10-1509	Malt	Speyside	12		BS	40	Y	Y
S10-1510	Malt	Speyside	12		BS	40	Y	Y
S10-1511	Blend				A	40	Y	
S10-1512	Blend					40	Y	
S10-1849	Malt	Islay	10	Y	B	46	Y	Y
S10-1850	Malt	Highland	10		B	40	Y	Y
S10-1851	Malt	Highland			BS	46	Y	Y
S10-2015	Blend					40	Y	
S10-2054	Blend		12			40	Y	
S10-2055	Malt	Speyside	12		B	40	Y	Y
S10-2056	Blend					40	Y	
S10-2057	Blend					40	Y	
S10-2058	Malt	Speyside	10		BS	43	Y	Y
S10-2059	Blend					40	Y	
S12-0259	Blend					40	Y	
S12-0275	Malt	Speyside	12			40	Y	Y
S12-0418	Malt	Islay	16	Y	BS	43	Y	
S12-1077	Blend					40	Y	
S12-1145	Blend					40	Y	
S12-1146	Blend					40	Y	

Continued on next page

Table B.1 – continued from previous page

ID	TYPE	REGION	AGE	PEAT	WOOD	ABV	NMR	FTMS
S12-1147	Malt	Islay	12		BS	46	Y	Y
S12-1236	Malt	Islay	10	Y	B	40	Y	
S12-1237	Blend					40	Y	
S12-1238	Malt					40	Y	
S12-1239	Malt	Speyside	12		BS	40	Y	
S12-1240	Malt	Speyside	12		BS	40	Y	Y
S12-1241	Blend				A	40	Y	
S12-1242	Blend					40	Y	
S12-1243	Blend				S	40	Y	
S12-1244	Blend					40	Y	
S12-1245	Malt	Highland	10		S	40	Y	
S12-1246	Malt	Islay		Y		43	Y	
S12-1292	Blend					40	Y	Y
S12-1293	Malt	Highland	12		S	40	Y	Y
S12-1294	Blend					40	Y	
S12-1295	Blend		12			40	Y	
S12-1296	Malt	Highland	12	Y	S	40	Y	
S12-1484	Malt	Highland	10		B	40	Y	
S12-1485	Malt	Islay	10	Y	B	46	Y	Y
S12-2106	Blend		12			40	Y	
S12-2388	Malt	Highland			BS	46	Y	
S12-2410	Malt	Highland	12		B	40	Y	
S12-2411	Blend					40	Y	
S12-2513	Blend					40	Y	
S12-2514	Malt	Highland	10			40	Y	Y
S12-2515	Malt	Highland	12		B	40	Y	
S12-2516	Blend					40	Y	
S12-3431	Blend					40	Y	
S12-3432	Blend					40	Y	
S12-3433	Malt	Highland	12		BS	40	Y	

Continued on next page

Table B.1 – continued from previous page

ID	TYPE	REGION	AGE	PEAT	WOOD	ABV	NMR	FTMS
S12-3434	Malt	Highland	10		B	40	Y	
S12-3570	Blend					40	Y	
S12-3571	Blend		12			40	Y	
S12-3572	Malt	Speyside	12		B	40	Y	
S13-0013	Blend					43	Y	
S13-0026	Blend					40	Y	
S13-0041	Blend					40	Y	
S13-0090	Malt	Lowland			BS	43	Y	Y
S13-0091	Malt	Lowland			B	40	Y	Y
S13-0092	Malt	Islay	12	Y		40	Y	
S13-0110	Blend		21			40	Y	
S13-0111	Malt	Speyside	10		BS	43	Y	
S13-0112	Blend					40	Y	
S14-1906	Malt	Islay	16	Y	BS	43	Y	Y
S14-1907	Blend					40	Y	Y
S14-1908	Malt	Speyside	12			40	Y	Y
S14-1909	Blend		12			40	Y	Y
S14-1911	Blend					40	Y	Y
S14-1913	Malt	Islay	12		BS	46	Y	Y
S14-1914	Blend					40	Y	Y
S14-1915	Blend					40	Y	Y
S14-1916	Blend					40	Y	Y
S14-1919	Blend					40	Y	Y
S14-1920	Malt	Speyside	16		B	43	Y	Y
S14-1939	Blend		12			40	Y	Y
S14-1940	Blend					40	Y	Y
S14-1941	Malt	Highland	12		S	40	Y	Y
S14-1942	Malt	Highland	12	Y	S	40	Y	Y
S14-1943	Blend					40	Y	Y
S14-1944	Blend					40	Y	Y

Continued on next page

Table B.1 – continued from previous page

ID	TYPE	REGION	AGE	PEAT	WOOD	ABV	NMR	FTMS
S14-1947	Malt	Islay	10	Y	B	46	Y	Y
S14-1948	Malt	Highland	10		B	40	Y	Y
S14-1962	Malt	Islay	12	Y		40	Y	Y
S14-1963	Malt	Lowland			B	40	Y	Y
S14-1964	Malt	Lowland			BS	40	Y	Y
S14-1972	Malt	Highland	12		BP	46	Y	Y
S14-2079	Blend					40	Y	Y
S14-2080	Blend					40	Y	Y
S14-2081	Malt	Highland	10		B	40	Y	Y
S14-2082	Blend					40	Y	Y
S14-2083	Malt	Islay	10	Y	B	40	Y	Y
S14-2085	Malt					40	Y	Y
S14-2086	Blend		12		B	40	Y	Y
S14-2087	Blend				A	40	Y	Y
S14-2088	Blend					40	Y	Y
S14-2089	Malt	Speyside	12		BS	40	Y	Y
S14-2090	Malt	Speyside	12		BS	40	Y	Y
S14-2195	Blend					40	Y	Y
S14-2196	Malt	Highland	12		B	40	Y	Y
S14-2319	Malt	Speyside			BS	43	Y	Y
S14-2335	Blend					40	Y	Y
S14-2336	Blend					40	Y	Y
S14-2337	Malt	Highland	12		B	40	Y	Y
S14-2338	Blend					40	Y	Y
S14-2372	Blend					40	Y	Y
S14-2373	Malt	Highland	10		S	40	Y	Y
S14-2374	Malt	Islay		Y		43	Y	Y
S14-2375	Blend					43	Y	Y
S14-2815	Malt	Speyside	12		B	40	Y	Y
S14-2816	Blend					40	Y	Y

Continued on next page

Table B.1 – continued from previous page

ID	TYPE	REGION	AGE	PEAT	WOOD	ABV	NMR	FTMS
S14-2817	Blend		12			40	Y	Y
S14-2818	Blend					40	Y	Y
S14-2856	Malt	Speyside	10		BS	40	Y	Y
S14-2857	Blend					40	Y	Y
S14-2858	Blend		21			40	Y	Y

C

APPENDIX: CALIBRATION LISTS

Table C.1: Calibration lists for negative APCI and APPI FTICR mass spectra of Scotch Whisky. Formula shown as $[M\text{-}H]^-$ ion formula.

APCI		APPI	
FORMULA	M/Z	FORMULA	M/Z
$C_8H_7O_3$	151.040068	$C_6H_9O_5$	161.046644
$C_6H_9O_5$	161.045547	$C_7H_5O_5$	169.014247
$C_8H_7O_4$	167.034982	$C_6H_{11}O_6$	179.056111
$C_9H_9O_4$	181.050632	$C_6H_{11}O_7$	195.051026
$C_{10}H_{19}O_3$	187.133968	$C_8H_{11}O_7$	219.051026
$C_{12}H_{23}O_2$	199.170353	$C_8H_{13}O_8$	237.061591
$C_{11}H_{11}O_4$	207.066282	$C_9H_{15}O_8$	251.077241
$C_{12}H_{11}O_4$	219.066282	$C_9H_{15}O_9$	267.072155
$C_{11}H_{13}O_5$	225.076847	$C_{10}H_{17}O_9$	281.087806
$C_{12}H_9O_5$	233.045547	$C_{11}H_{17}O_9$	293.087806
$C_{15}H_{29}O_2$	241.217304	$C_{14}H_5O_8$	300.998991
$C_{13}H_{13}O_5$	249.076847	$C_{12}H_{17}O_9$	305.087806
$C_{16}H_{31}O_2$	255.232954	$C_{11}H_{19}O_{10}$	311.098370
$C_{13}H_{13}O_6$	265.071762	$C_{12}H_{19}O_{10}$	323.098370
$C_{15}H_{13}O_5$	273.076847	$C_{13}H_{15}O_{10}$	331.067070
$C_{18}H_{35}O_2$	283.264254	$C_{12}H_{21}O_{11}$	341.108935
$C_{17}H_{23}O_4$	291.160183	$C_{16}H_{17}O_9$	353.087806
$C_{17}H_{15}O_5$	299.092497	$C_{17}H_{19}O_9$	367.103456
$C_{14}H_5O_8$	300.998991	$C_{18}H_{17}O_9$	377.087806
$C_{12}H_{21}O_9$	309.119106	$C_{19}H_{21}O_9$	393.119106
$C_{17}H_{15}O_6$	315.087412	$C_{20}H_{21}O_9$	405.119106
$C_{18}H_{15}O_6$	327.087412	$C_{15}H_{25}O_{13}$	413.130064
$C_{19}H_{17}O_6$	341.103062	$C_{15}H_{25}O_{14}$	429.124979

Continued on next page

Table C.1 – continued from previous page

APCI		APPI	
FORMULA	M/Z	FORMULA	M/Z
$C_{18}H_{15}O_7$	343.082326	$C_{15}H_{27}O_{15}$	447.135544
$C_{21}H_{19}O_6$	367.118712	$C_{16}H_{27}O_{15}$	459.135544
$C_{19}H_{21}O_8$	377.124191	$C_{17}H_{27}O_{15}$	471.135544
$C_{21}H_{19}O_8$	399.108541	$C_{18}H_{29}O_{15}$	485.151194
$C_{22}H_{21}O_8$	413.124191	$C_{18}H_{29}O_{16}$	501.146108
$C_{22}H_{23}O_9$	431.134756	$C_{30}H_{45}O_7$	517.317077
$C_{22}H_{25}O_{10}$	449.145320	$C_{18}H_{31}O_{17}$	519.156673
$C_{24}H_{15}O_{10}$	463.067070	$C_{19}H_{33}O_{17}$	533.172323
$C_{25}H_{17}O_{11}$	493.077635	$C_{20}H_{33}O_{17}$	545.172323
$C_{25}H_{17}O_{12}$	509.072549	$C_{21}H_{33}O_{17}$	557.172323
$C_{30}H_{45}O_7$	517.317077	$C_{20}H_{35}O_{18}$	563.182888
$C_{28}H_{29}O_{10}$	525.176621	$C_{22}H_{33}O_{17}$	569.172323
$C_{28}H_{29}O_{11}$	541.171535	$C_{21}H_{33}O_{18}$	573.167238
$C_{29}H_{29}O_{11}$	553.171535	$C_{20}H_{35}O_{19}$	579.177802
$C_{30}H_{31}O_{11}$	567.187185	$C_{22}H_{35}O_{18}$	587.182888
$C_{29}H_{33}O_{12}$	573.197750	$C_{21}H_{35}O_{19}$	591.177802
$C_{30}H_{33}O_{12}$	585.197750	$C_{22}H_{37}O_{19}$	605.193452
$C_{32}H_{33}O_{11}$	593.202835	$C_{21}H_{37}O_{20}$	609.188367
$C_{32}H_{31}O_{12}$	607.182100	$C_{24}H_{37}O_{19}$	629.193452
$C_{32}H_{35}O_{13}$	627.208315	$C_{24}H_{37}O_{20}$	645.188367
$C_{33}H_{33}O_{13}$	637.192665	$C_{24}H_{39}O_{20}$	647.204017
$C_{33}H_{39}O_{14}$	659.234529	$C_{36}H_{57}O_{12}$	681.385551
$C_{33}H_{41}O_{15}$	677.245094	$C_{37}H_{51}O_{12}$	687.338600
$C_{38}H_{39}O_{13}$	703.239615	$C_{36}H_{57}O_{13}$	697.380465
$C_{39}H_{39}O_{13}$	715.239615		
$C_{38}H_{41}O_{14}$	721.250179		
$C_{39}H_{39}O_{14}$	731.234529		
$C_{39}H_{39}O_{15}$	747.229444		

Table C.2: Calibration lists for negative ESI and LDI FTICR mass spectra of Scotch Whisky. Formula shown as $[M\text{-}H]^-$ ion formula.

ESI		LDI	
FORMULA	M/Z	FORMULA	M/Z
$C_4H_5O_6$	149.009161	$C_{10}H_7O_3$	175.040068
$C_9H_{17}O_2$	157.123403	$C_{11}H_9O_3$	189.055718
$C_7H_5O_5$	169.014247	$C_{12}H_{11}O_3$	203.071368
$C_6H_9O_6$	177.040462	$C_{13}H_{13}O_3$	217.087018
$C_6H_{11}O_7$	195.051026	$C_{12}H_7O_5$	231.029897
$C_{12}H_{23}O_2$	199.170354	$C_{13}H_9O_5$	245.045547
$C_{11}H_{11}O_4$	207.066282	$C_{14}H_{11}O_5$	259.061197
$C_8H_{13}O_7$	221.066676	$C_{15}H_{13}O_5$	273.076847
$C_8H_{13}O_8$	237.061591	$C_{16}H_{15}O_5$	287.092497
$C_9H_{13}O_8$	249.061591	$C_{14}H_5O_8$	300.998991
$C_{12}H_5O_7$	261.004076	$C_{17}H_{17}O_5$	301.108147
$C_{11}H_{15}O_8$	275.077241	$C_{18}H_{19}O_5$	315.123797
$C_{12}H_{15}O_8$	287.077241	$C_{19}H_{21}O_5$	329.139447
$C_{13}H_7O_8$	291.014641	$C_{20}H_{23}O_5$	343.155097
$C_{14}H_5O_8$	300.998991	$C_{18}H_{13}O_8$	357.061591
$C_{13}H_{15}O_9$	315.072156	$C_{19}H_{15}O_8$	371.077241
$C_{13}H_{15}O_{10}$	331.067070	$C_{20}H_{17}O_8$	385.092891
$C_{12}H_{21}O_{11}$	341.108935	$C_{21}H_{19}O_8$	399.108541
$C_{15}H_{19}O_9$	343.103456	$C_{22}H_{21}O_8$	413.124191
$C_{16}H_{19}O_9$	355.103456	$C_{23}H_{23}O_8$	427.139841
$C_{17}H_{21}O_9$	369.119106	$C_{23}H_{21}O_9$	441.119106
$C_{19}H_{17}O_9$	389.087806	$C_{24}H_{23}O_9$	455.134756
$C_{20}H_{19}O_9$	403.103456	$C_{25}H_{25}O_9$	469.150406
$C_{20}H_{21}O_{10}$	421.114020	$C_{26}H_{27}O_9$	483.166056
$C_{20}H_{23}O_{11}$	439.124585	$C_{26}H_{25}O_{10}$	497.145321
$C_{19}H_{25}O_{13}$	461.130064	$C_{27}H_{27}O_{10}$	511.160971
$C_{17}H_{27}O_{15}$	471.135544	$C_{28}H_{29}O_{10}$	525.176621

Continued on next page

Table C.2 – continued from previous page

ESI		LDI	
FORMULA	M/Z	FORMULA	M/Z
$C_{22}H_{25}O_{12}$	481.135150	$C_{29}H_{31}O_{10}$	539.192271
$C_{22}H_{25}O_{13}$	497.130064	$C_{30}H_{33}O_{10}$	553.207921
$C_{18}H_{31}O_{16}$	503.161758	$C_{31}H_{35}O_{10}$	567.223571
$C_{24}H_{27}O_{12}$	507.150800	$C_{30}H_{29}O_{12}$	581.166450
$C_{30}H_{45}O_{7}$	517.317077	$C_{31}H_{31}O_{12}$	595.182100
$C_{30}H_{45}O_{8}$	533.311992	$C_{32}H_{33}O_{12}$	609.197750
$C_{24}H_{29}O_{14}$	541.156279	$C_{33}H_{35}O_{12}$	623.213400
$C_{25}H_{27}O_{14}$	551.140629	$C_{34}H_{37}O_{12}$	637.229050
$C_{29}H_{31}O_{12}$	571.182100	$C_{36}H_{27}O_{12}$	651.150800
$C_{28}H_{37}O_{13}$	581.223965	$C_{37}H_{29}O_{12}$	665.166450
$C_{22}H_{35}O_{19}$	603.177802	$C_{38}H_{31}O_{12}$	679.182100
$C_{30}H_{47}O_{11}S$	615.284457	$C_{39}H_{33}O_{12}$	693.197750
$C_{24}H_{39}O_{19}$	631.209102	$C_{40}H_{35}O_{12}$	707.213400
$C_{27}H_{23}O_{18}$	635.088987	$C_{41}H_{37}O_{12}$	721.229050
$C_{36}H_{55}O_{12}$	679.369901		
$C_{36}H_{57}O_{13}$	697.380465		
$C_{27}H_{43}O_{23}$	735.220061		
$C_{44}H_{51}O_{15}$	819.323345		
$C_{42}H_{67}O_{18}$	859.433289		
$C_{45}H_{33}O_{29}$	1037.111299		
$C_{47}H_{37}O_{29}$	1065.142599		

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

`http://postcards.miede.de/`

Final Version as of 15th October 2018 (`classicthesis`).