GlaxoSmithKline

University of Strathclyde

Automated, Quantitative Process Analysis by NMR

PhD Thesis

Simon Andrew Watson

2019

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Date:

This thesis and the work within is dedicated to my wife, Kate, for all the support, advice and love that have made this study possible.

It's my turn to dedicate something to you for a change.

⋆

Acknowledgements

Andy Edwards and John Parkinson are thanked for their continued support, help and advice throughout the 5 years of study that this thesis represents.

Tran Pham is acknowledged for his work in fully deriving the Bruker sino real calculation and relating it to both internally and externally accepted SNR limits (see Appendix 1).

Andy Gibbs of Bruker UK is acknowledged for his help and advice with au programming and automation and for supplying the integration routine incorporated into the 35cl_quant programme (see eSuppl 1).

John Paul Cerrotti, Anna Gerdova, Bertram Manz and Jürgen Kolz of Magritek are acknowledged for supply and installing the Spinsolve NMR spectrometer. They are also thanked for supplying additional information on system hardware and software details, software updates and general advice throughout this project.

Denver Sheelan is thanked for the diligent preparation, acquisition and processing of samples and data for the deconvolution experiments.

Sam Genway and Benoit Mangili of Tessella are thanked for their support, training and advice that made the coding of the custom user interface possible.

Charles Goss and Neil Curtis are acknowledged as the source of the applications for the quantitative atline investigation and are thanked for support, advice and supplying samples.

Neil Curtis is acknowledged for supplying the HPLC assay data for Compound 2.

Jaclyn Dunn, Duncan Thompson and Alison Nordon are acknowledged for support and training in Matlab and PLS toolbox.

Audrey Zilliox and Michael Rothwell-Williams are acknowledged for supplying samples, testing the automated RAMS programme and offering positive suggestions that led to the development of many of its features.

Mark Hughes is acknowledged for supplying material and synthetic route information for the multivariate analysis studies.

Andy Roberts, Richard Upton, Sean Lynn and David Doyle are thanked for their time and effort in reviewing the experimental records that supported the work in this thesis.

Kate Kinnear, Nicholas Holmes, Caroline Duignan, Przemek Stasica and Denver Sheelan are thanked for reviewing the printed works.

Abstract

This thesis comprises a series of studies and investigations into applications of Nuclear Magnetic Resonance (NMR) to automated and quantitative process analysis. The work will cover the use of different nuclei, permanent and superconducting magnets, and quantitation by univariate and multivariate approaches. Customised programmes and scripts accompany the experimental work on each of these aspects.

Chapter 1 is an introduction to the history and status of NMR, as relevant to these works.

Chapter 2 describes the development and validation of a fully automated ^{35}Cl NMR quantitation method for chloride in solution. The method is embedded in existing walk-up software on a 400 MHz spectrometer and yields quantitative results with linearity $R^2 = 0.999$ and a detection limit of 0.1% w/w for a 10 mg sample.

Chapter 3 covers investigations into the capabilities and limitations of a 43 MHz bench top NMR spectrometer for samples in a tube and in flow. Aspects including long-term stability, characterisation of samples in flow, non-linearity of receiver and amplifier and synthesis of accurate shaped pulses are assessed.

Chapter 4 builds on this information and describes the development and programming of an automated custom interface designed for non-expert users that allows the bench top spectrometer to be used to rapidly supply quantitative results for atline process samples. Relative quantitation is achieved through a combination of sum integration and deconvolution.

Chapter 5 then extends the capabilities of the benchtop spectrometer through the development of a quantitative method for reaction mixtures of a pharmaceutical process using partial least squares modelling of ^{1}H spectra. The model is developed such that results were shown to be independent of typical spectral variations expected in real samples.

The thesis concludes with remarks on potential future extensions to the work discussed here.

Abbreviations

ADC	Analogue to Digital Converter
ALCOA	Attributable, Legible, Contemporaneous, Original, Accurate
API	Active Pharmaceutical Ingredient
AQ	AcQuisition time
ATMA	Automatic Tuning and Matching Accessory
BBFO	Broad-Band and Fluorine Observed
BFM	BromoFluoroMethane
BHT	ButylatedHydroxyToluene
CLS-MLR	Classical Least Squares – Multiple Linear Regression
CoA	Certificate of Analysis
COSY	COrrelation SpectroscopY
CPMG	Carr-Purcell-Meiboom-Gill
CSA	Chemical Shift Anisotropy
CST	Chemical Shift Tensor
DE	DEad time
DL	Detection Limit
DMAC	DiMethyl ACetamide
DNA	DeoxyriboNucleic Acid
DOSY	Diffusion Ordered SpectroscopY
DW	DWell time
EFNMR	Earth's Field Nuclear Magnetic Resonance
EM	ElectroMagnet
EPR	Electron Paramagnetic Resonance
FDA	Food and Drugs Administration
FID	Free Induction Decay
FMPI	FluoroMethylPhenyl Isocyanate
FP	Fluticasone Propionate
FT	Fourier Transform
FTIR	Fourier Transform InfraRed
GC	Gas Chromatography
GUI	Graphical User Interface
GSD	Global Spectral Deconvolution
HHLW	Half-Height Line Width
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High Performance Liquid Chromatography
HSQCed	Heteronuclear Single bond Quantum Correlation –multiplicity edited
IC	Ion Chromatography
ICH	International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use
ICP-OES	Inductively Coupled Plasma – Optical Emission Spectroscopy
ILS-MLR	Inverse Least Squares - Multiple Linear Regression
INEPT	Insensitive Nuclei Enhanced by Polarisation Transfer
IPA	IsoPropyl Alcohol
JRES	J-RESolved spectroscopy
LB	Line Broadening function
LC-MS	Liquid Chromatography-Mass Spectrometry

LV	Latent Variable
MeCN	Acetonitrile
MeOH	Methanol
MVA	MultiVariate Analysis
ND	Not Detected
NIR	Near-InfraRed
NMR	Nuclear Magnetic Resonance
NS	Number of Scans
PAT	Process Analytical Technology
PC	Principal Component
PCA	Principal Components Analysis
PLS	Partial Least Squares (Projection to Latent Structures)
PTFE	PolyTetraFluoroEthene
PULCON	PULse Length CONcentration determination
QC	Quality Control
QL	Quantitation Limit
RF	Radio Frequency
RG	Receiver Gain
RGA	Receiver Gain Adjust
RMS	Root Mean Square
RMSECV	Root Mean Squared Error of Cross Validation
RMSCV	Root Mean Square Coefficient of Variation
RNA	RiboNucleic Acid
% RSD	Percentage Relative Standard Deviation
ROESY	Rotating Overhauser Effect SpectroscopY
Rx	Receiver
SQUID	Superconducting QUantum Interference Device
SD	Standard Deviation
SNR	Signal-to-Noise Ratio
ssNMR	Solid State Nuclear Magnetic Resonance
T_1	Longitudinal (spin-lattice) relaxation time
T_2	Transverse (spin-spin) relaxation time
TAP	ThioAcid Propionate
TBME	TertiaryButylMethyl Ether
TCP-IP	Transmission Control Protocol – Internet Protocol
TD	Number of data points in the Time Domain
THF	TetraHydroFuran
TOCSY	Total COrrelation SpectroscopY
Tx	Transmitter
UV-Vis	Ultra-Violet/Visible light
VTEC	Volumetric Thermal Expansion Coefficient

Table of Contents

Chapter 1

Chapter 2

Chapter 5

Electronic Data

This thesis also contains electronic supplemental data, referenced in the text, for example, as “eSuppl 1”, etc.

These files are located in folders of the same name as the reference. The format of these files and the programmes required to open them are listed below:

eSuppl 1

Processing script in C source code in ".txt" file*

eSuppl 2

Pulse sequence for Spinsolve Expert (Magritek) in ".mac" file*

eSuppl 3

Processing script in Javascript for MestreNova (MestreLabs) in ".qs" file*

eSuppl 4

Processing script in Javascript for MestreNova (MestreLabs) in ".qs" file*

eSuppl 5

Complete RAMS control and user interface containing several file formats
".py" Python code for Python 3.4*
".ui" xml file generated by QT designer 5.6*
".png" image files**
".tff" truetype font file†

eSuppl 6

Data manipulation script in Python 3.4 in ".py" file*

* Human readable with any text viewer

** Readable with any image viewer

† Accessible by Microsoft Windows

Chapter 1

Introduction to NMR

1.1. Introduction

The phenomenon of nuclear precession in condensed matter and a method for its measurement were first described in two landmark papers, published a few weeks apart early in 1946 by two independent groups, working with almost no awareness of each other's research. These papers discussed the magnetic resonance of hydrogen nuclei, as observed by Edward Purcell *et al*. in a 1 kg block of wax [1] and by Felix Bloch *et al.* in a sample of water [2].

During the last 70 years of development, Nuclear Magnetic Resonance (NMR) spectroscopy has been successfully applied to all types of analyte analysis including identification, structure determination, quantitation and separation to name but a few. It's adoption across scientific disciplines now spans the physical, chemical, biological, medical and geological sciences and it is considered a prominent and highly-regarded analytical technique.

NMR spectroscopy is applied extensively in the pharmaceutical industry. It's specificity towards individual isotopes provides information on individual atoms in a drug molecule and, through probing inter and intra molecular interactions, NMR spectroscopy can show explicitly how the atoms are linked. Isotopic specificity is further augmented by inter- and intra-molecular interactions and dynamic behavior, resulting in high resolution spectra that can be used to identify individual chemical species in complex mixtures. Finally, a proportional relationship between the NMR signal and the number of nuclei responsible for that signal means that the phenomenon of NMR is inherently quantitative, and methods can be developed and validated to meet the highest standards set by regulatory agencies.

To develop a quantitative method for a given sample, the following aspects must be fully understood:

- Properties of the nuclei under analysis, including spin, gyromagnetic ratio, abundance, coupling mechanisms and relaxation pathways
- Capabilities of the spectrometer hardware

- Magnetic field strength
- Properties of the sample (solid or solution, static or flowing)
- Extraction of the required information from the raw data (processing)

There must then follow a process of assessment of the analytical method's capabilities by explicitly testing the method versus samples of known composition. In formal, regulated analytical environments, such as those found in the pharmaceutical industry, this is achieved through a process of method validation, whereby a developed method is tested against a series of pre-determined criteria. These standards are designed to test specific aspects of the method, *e.g.* accuracy, precision, detection limit, and so forth. Chapter 2 of this thesis explores this area in detail.

From a hardware perspective, NMR spectrometers require few moving parts, are constructed from solid-state components, and the transmitted and received signals pass through simple electrical conductors. Because of this, the physical day-to-day set up and running of the spectrometer requires little manual intervention by the analyst and so many steps in the acquisition and processing of data can be automated. A combination of automation and quantitation can lead to the development of fast, reliable quantitative methods that can deliver not just data, but the desired *information* to non-expert end users.

The following sections will expand on some of the critical aspects of NMR that influence the development of automated, quantitative methods.

1.2. Nuclear properties and interactions

One of the reasons that NMR spectroscopy has the wide-ranging utility previously noted is the large number of attributes and properties that it is sensitive to. These start with the structure of the atomic nucleus itself, reach out to the electronic and magnetic interactions within the overall atomic structure and extend beyond into the inter-nuclear, the inter-molecular and bulk macroscopic properties of the sample.

A second reason is the fact that NMR spectroscopy can be used, in principle, to analyse at least one isotope of almost all of the elements in the periodic table. Consequently, when considering if it is possible to investigate a sample, or a problem, by NMR spectroscopy the question asked is rarely, “Is it possible?”, but rather, “How long will it take?” and “What information will it yield?”. The questions can be answered by considering the properties and interaction of the nuclei under investigation.

1.2.1. Spin

The work preceding the lead up to the discovery of NMR was summarised by Bloch in his Nobel Prize lecture of 1952 and included the proposal by Pauli, in 1924, that atomic nuclei may possess an intrinsic angular momentum, accompanied by a magnetic moment. This intrinsic angular momentum is characterised by its spin quantum number denoted as *I* (in units of h/2π or $\hbar$).

Although multiple spin states are possible for each nucleus, for practical purposes it can be assumed that *I* corresponds to the ground state nuclear spin. As there are no simple rules for determining which of the many possible spin states equates to the ground state, it is common to regard *I* as an empirically determined property for each isotope. The value of *I* may be positive or negative (or zero) and is represented in multiples of 1/2.

Approximately 25% of all the naturally occurring NMR nuclei have spin = 1/2. A nucleus will have (2*I*+1) possible spin states and so spin 1/2 nuclei represent the simplest possible NMR condition with just 2 spin states.

Nuclei with spin = 1/2 also have the simplest magnetic moments, with a spherically symmetrical nuclear charge and dipole moment that can be visualised as a bar magnet with a north and south pole. For this reason, spin 1/2 nuclei have narrow lineshapes and hence spectra with good resolution. The lifetime of the excited state of spin 1/2 nuclei makes it possible to observe interactions between different nuclei *via* the property of spin-spin or "*J*" coupling. At a basic level this can be observed as splitting of the resonance lines due to homo- or heteronuclear coupling, giving information on the number of other nuclei in close proximity that are associated *via* a bonding framework. In more advanced 2D NMR methods, energy can be transferred between coupled spins and provide detailed and explicit structural information.

One downside is that spin 1/2 nuclei sometimes exhibit excessively long lifetimes in their excited states. The return from excited to equilibrium spin state is characterised by an exponential function with a time constant, T_1 (see later in this Chapter), which for spin 1/2 nuclei may have values of many hundreds of seconds or more. This in turn requires long delays between individual experiments to yield quantitative data.

Nuclei with higher spins (*i.e.* $I > 1/2$) are referred to as *quadrupolar nuclei*. These have non-spherically symmetrical charge distribution, the implications of which are discussed later.

1.2.2. Magnetogyric ratio

Nuclei with $I \neq 0$ posses angular momentum, from which arises the magnetic moment, μ as defined in equation (1.1) [3]:

$$\mu = \gamma I \qquad \textbf{(1.1)}$$

where γ is the *magnetogyric ratio*, a constant for a given nucleus. This determines both the frequency of the resonances and also their intensity, and has units of rad T^{-1} s^{-1}. In the absence of a magnetic field, a collection of nuclei will orientate themselves such that the angular momentum vectors will be aligned randomly across all directions in space (see Figure 1.1, left).

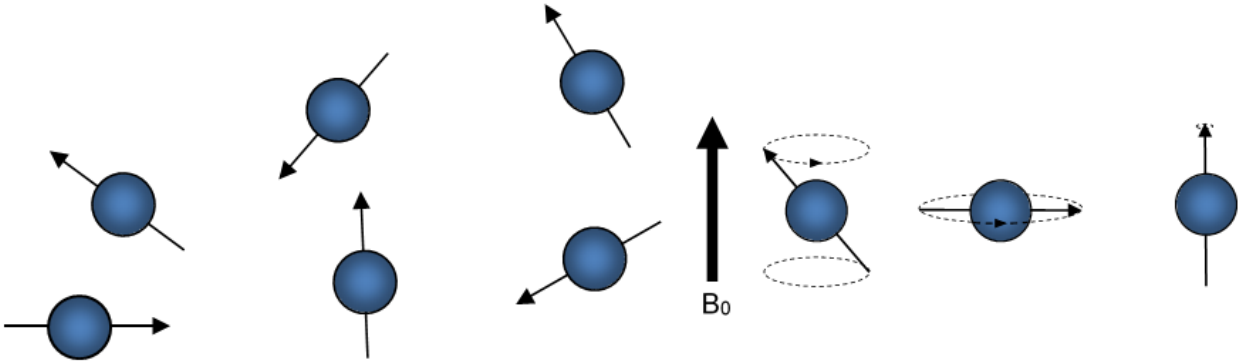


Figure 1.1 – Random orientation of spins in the absence of magnetic field (left), and nuclear pressession in the presence of a magnetic field (right)

When these nuclei are placed in a magnetic field, a torque is imparted on the magnetic moments and they trace circular paths, always staying at a constant angle to the applied magnetic field, B_0 (see Figure 1.1, right). The diagram shows examples of different circular paths that the magnetic moments could trace depending upon their orientation to B_0. This motion is known as *Larmor precession* and may be clockwise or counter-clockwise, depending on whether γ has a positive or negative value, respectively.

As time progresses, the random movement of molecules in solution leads the spins to be exposed to a range of slightly fluctuating fields, resulting in changes in the angle of precession. Since spins that are orientated parallel to the applied magnetic field are of lower energy than those orientated antiparallel to the field, a slight excess of spins aligning with B_0 will evolve. This leads to a small bulk magnetization vector aligned parallel to B_0, referred to as M_0.

The frequency of precession is the *Larmor precession frequency* and is defined either in rad s^{-1} by equation (1.2) or in Hz by equation (1.3) [4].

$$\omega = -\gamma B_0 \, rad \, s^{-1} \quad \textbf{(1.2)}$$

$$\nu = \frac{-\gamma B_0}{2\pi} \, Hz \quad \textbf{(1.3)}$$

^{3}H has the highest γ of any nucleus in the periodic table, closely followed by ^{1}H, then ^{19}F. It is of interest to visualise these values in a plot, arranged in order of increasing atomic number, with ν in MHz at 9.4 T on a $\log_{10}$ scale (see Figure 1.2). Note that some isotopes have been omitted for clarity.

Nuclei of a resonant frequency below that of ^{13}C are often referred to as *low-γ*. As well as giving a less intense NMR response (due to slower precession rates), these nuclei also cause problems with acoustic ringing [5], or probe ring-down. Ring-down manifests itself as additional in-phase intensity at the start of the oscillation NMR signal, referred to as the Free Induction decay (FID), resulting in baseline distortions of the Fourier Transform (FT) spectrum. It can usually be resolved by either an increase in the dead time between pulse and FID acquisition, back linear prediction of the FID, or post-FT baseline corrections.

A plot of the resonant frequencies of nuclei at 9.4 T against atomic number was created from tabulated literature data [6]. As can be seen from Figure 1.2, the spin 1/2 nuclei are distributed across the periodic table. Although these nuclei are often considered to be less difficult to observe by NMR, some have very low γ values, leading to poor receptivity, even when present at higher natural abundance. For example, the low γ of ^{187}Os (0.6161 rad T^{-1} s^{-1}) [7] results in 1000 $\times$ less sensitivity than for ^{13}C. Sensitivity is further diminished by the low natural abundance of 1.64%.

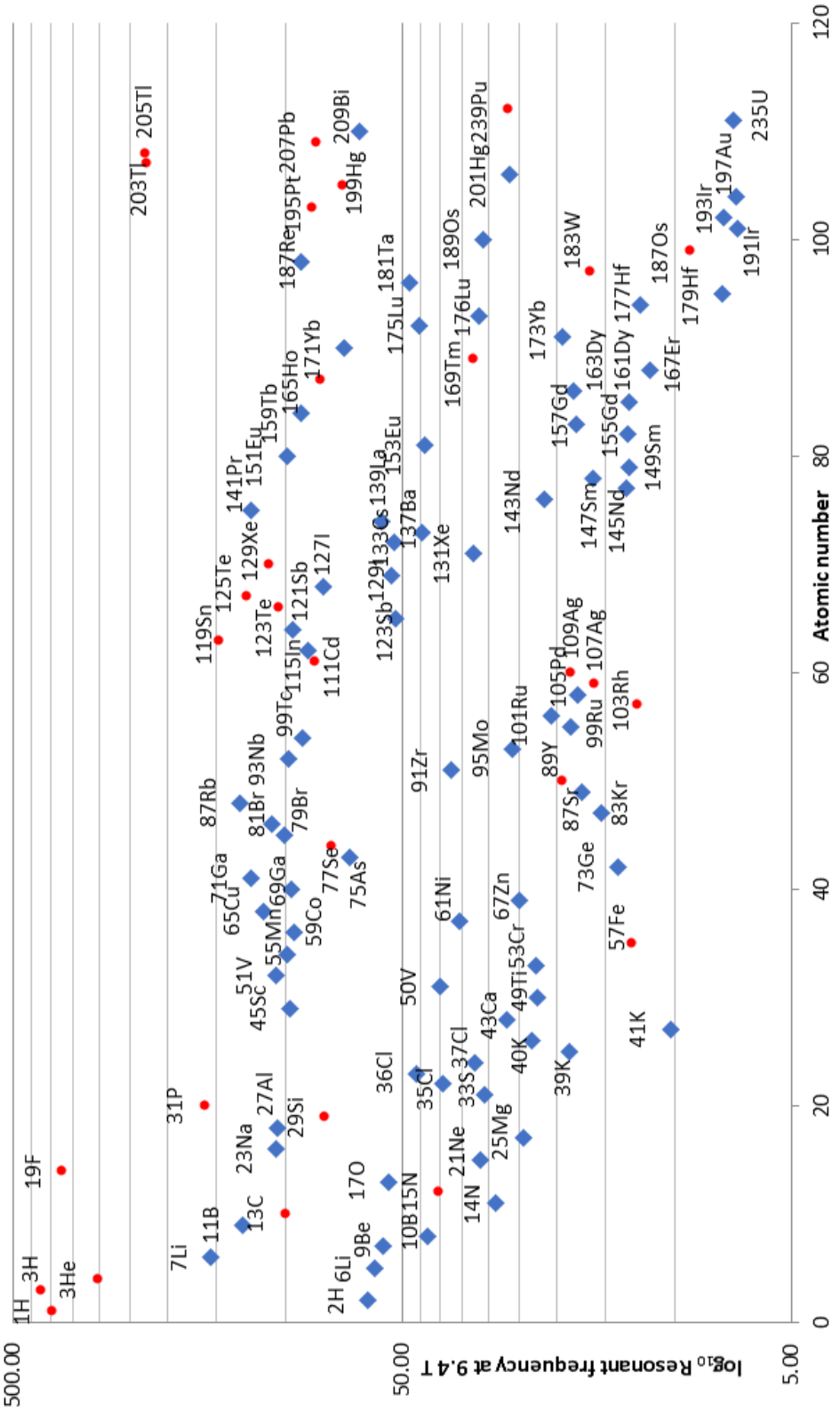


Figure 1.2 – Plot of log $_{10}$ NMR Resonant Frequencies at 9.4T (red circles denote spin ½, blue diamonds denote quadrupolar nuclei)

1.2.3. Natural abundance

The ratio of isotopes and their distribution in nature is intrinsically tied to their source and relative half-lives. For example, whilst ^{12}C and ^{13}C are stable and have abundancies of 98.9 and 1.1% abundancies, respectively, ^{14}C is radioactive with a half-life of 5,700 years and has an abundancy of ~ 1 x 10^{-12} %.

The comparatively low abundance of ^{13}C and ^{15}N led to delayed NMR applications in these nuclei, until techniques such as Fourier Transform NMR signal averaging, INEPT [8] and inverse techniques were developed. There are some advantages of dilute spins, such as the simplification of 1D ^{13}C NMR spectra. The ^{1}H-^{13}C spin-spin coupling can be readily decoupled and the very low probability of two adjacent spin-active carbons being present in a molecule (0.0121%) means that there is no significant interference from ^{13}C-^{13}C couplings – one of the reasons why ^{13}C spectra exhibit such high resolution. In this case the low ^{13}C abundance also simplifies corresponding ^{1}H spectra as the ^{1}H-^{13}C coupled signals are present at *ca.* 1.1% of the signal area of each peak.

One obvious way in which unfavourable natural abundances are overcome is by isotopic labeling. This is usually used either to enhance the signal-to-noise ratio (SNR) of the spectra, or to take advantage of the properties of a normally low-level isotope *e.g.* ^{13}C and ^{15}N labeling in DNA and RNA studies [9]. In these examples, partial ^{2}H labeling can also be employed to reduce the strong ^{1}H-^{1}H dipolar coupling presence in these molecules as well as being a useful tool to modify the relaxation characteristics of particular atoms to enhance the lifetimes of certain spin states.

1.2.4. Relaxation pathways

When a sample is placed in a magnetic field, an equilibrium condition, as discussed above, is established over time resulting in a bulk magnetisation vector. This equilibrium magnetisation vector can be perturbed by the application of a pulsed electromagnetic field resonating at, or close to, the Larmor frequency, ω or ν (depending on units). In NMR spectroscopy, these fields correspond to radio frequencies.

In a simplified model, the equilibrium and perturbed states can be considered as states where the individual spins combine to give a bulk magnetisation vector parallel and anti-parallel to the external magnetic field, respectively. The energy gap between these states, ΔE, is then represented by equation (1.4).

$$\Delta E = h\nu \tag{1.4}$$

The transition between the parallel (low energy) and anti-parallel (high energy) states is then an excitation process where the bulk spin system absorbs energy, followed by relaxation back to the equilibrium state that emits energy. The relaxation process does not happen instantaneously as the energy must be transferred to another system, or systems, that can accept ΔE exactly. This process is called spin-lattice, or longitudinal, relaxation and is characterised by a time constant, T_1 as described in equation (1.5).

$$M_{z(\tau)} = M_{z(0)}\left(1 - e^{-\frac{\tau}{T_1}}\right) \tag{1.5}$$

Where $M_{z(\tau)}$ is the magnetisation in the z axis (see below) at time τ after an excitation pulse and $M_{z(0)}$ is the equilibrium magnetisation in the z axis, both of which are proportional to the NMR signal intensity at these times. As there is energy loss from the spin system to the surrounding environment, spin-lattice relaxation is an enthalpic process.

T_1 is an important factor to consider in the context of quantitation by NMR spectrometers as FT pulsed NMR spectrometers acquire and average the data from multiple transients to obtain the final NMR spectrum. T_1 dictates the length of the delay required between the transients and hence sets an upper limit on the speed at which quantitative data may be acquired. For example, 99.3% of equilibrium magnetisation is achieved when a delay of 5 × T_1 is left between transients (see Figure 1.3). T_1 may also be different for individual nuclei in a molecule. As it is not possible to predict T_1 values empirically, they must be determined for each signal arising from a sample.

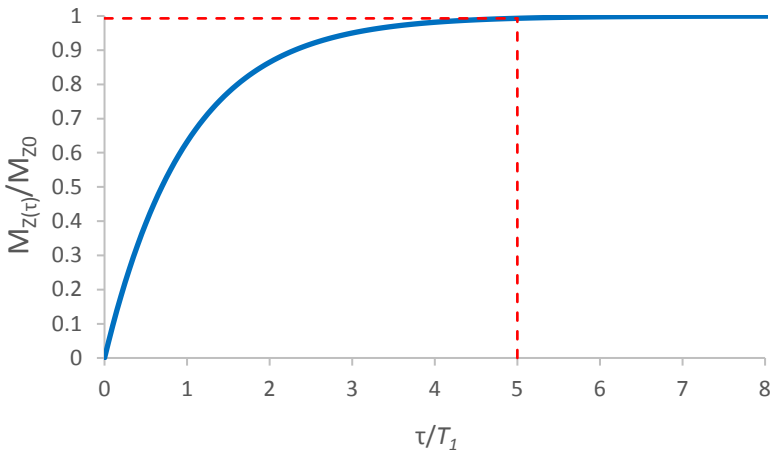


Figure 1.3 – Plot of $M_{Z(\tau)}/M_{Z0}$ *vs.* τ/T_1. Dotted line shows magnetisation at 99.3% of M_0 when $\tau = 5 \times T_1$.

The NMR experiment can also be considered using a vector model, where the vector of the applied magnetic field, B_0, is arbitrarily assigned as the z axis. Although at equilibrium there is a bulk magnetisation vector also aligned with the z axis, the individual precessing spins that make up this vector have random phase and so the field is effectively isotropic in the xy plane. As NMR spectrometers are only designed to detect rotating magnetic fields in the xy plane, no signal is detected at equilibrium.

The application of the RF excitation pulse to the sample has the effect of inducing spin-coherence: that is, directly after the pulse the precessing spins have the same phase. As time progresses, spins from different nuclei and molecules will experience slightly different magnetic fields due to the presence of other nearby spins. Energy will be exchanged between these spins, slowing or speeding up the precession so

that over time spin coherence is lost and the spins are once again randomly phased. This process is known as spin-spin, or transverse, relaxation and is characterised by a time constant, T_2, as described in equation (1.6).

$$M_{xy(\tau)} = M_{xy(0)}\left(e^{-\frac{\tau}{T_2}}\right) \quad \textbf{(1.6)}$$

Where $M_{xy(\tau)}$ is the magnetisation in the xy plane at time τ after an excitation pulse and $M_{xy(0)}$ is the equilibrium magnetisation in the xy axis. There is a loss of order in the form of phase coherence of the individual spins during spin-spin relaxation, but no net loss of energy, hence in contrast to spin-lattice relaxation this is an entropic process.

From the perspective of quantitation, T_2 affects the linewidth of signals in the NMR spectrum and also the observable lifetime of the NMR signal. T_2 is also affected by similar phenomena as those which affect T_1 and so the two are often considered together.

1.2.5. Coupling mechanisms

Originally used as a mechanical engineering term to describe the physical connection of two oscillating or rotating parts, coupling as applied to NMR spectroscopy can be considered as an interaction between two spin systems. The energy transfer between the spin systems can take place via a number of mechanisms, the principal ones being:

- Spin-Spin, scalar, or J-coupling
- Dipolar coupling
- Quadrupolar coupling

Spin-spin coupling is the transfer of energy between nuclear spins via electrons in chemical bonds and is almost exclusively intramolecular, whereas dipolar coupling involves the interaction of nuclear spins through space over a distance of a few

Ångstroms and hence may be intra- or intermolecular. Although both are of great importance to NMR spectroscopy [10] and form the basis of many NMR experiments, they are predominantly a consideration for structural work. Their impact on quantitation is mainly through the effect they have on peak shape and hence resolution and SNR, and so will not be discussed at length here.

Quadrupolar coupling arises as a result of the charge distribution of certain nuclei and can have a much greater influence on the appearance of spectra, T_1 and T_2. For this reason it is a significant consideration for some of the subsequent quantitation studies in this thesis.

The exact function that describes the shape or distribution of nuclear charge for a given atom is dependent only upon its spin quantum number, *I,* and not on the atomic weight or proton number. As these functions are mathematically proportional to the same spherical harmonics used to describe basic electron orbits, they can be thought of visually as taking the same appearance of the s-, p-, and d-orbitals. These so called *electropoles* are referred to as the total electric charge, dipole and quadrupolar moments respectively.

In the simplest case where $I = 1/2$, only the total electric charge moment exists. As in the case of the s-orbital, it is spherically symmetrical. This means that all spin 1/2 nuclei, regardless of size or number of nucleons, behave as point charges at the nuclear centre, with no dependence upon orientation or internal structure of the nucleus.

The dominant electropole to consider, when present, is the quadrupolar moment. This is present in all nuclei where $I > 1/2$, and for this reason these are referred to as *quadrupolar nuclei*. In these cases, the nuclear charge is no longer spherically symmetrical, and has an ellipsoid distribution. The quadrupolar moment is described by Q, and is measured in units of 10^{-24} cm^2, commonly referred to as the *barn* (10^{-27} cm^2 = 1 *millibarn*). Quadrupolar moments of the elements span over four orders of magnitude from < 1 mb for ^{6}Li to *ca.* 8000 mb for ^{176}Lu.

The distribution of charge from neighboring atoms in a molecule gives rise to an electric field gradient (*efg*) surrounding the nucleus. The quadrupolar moment will interact strongly with the *efg*, giving a resultant quadrupolar coupling; this in turn has several dominant effects upon the attributes of the nucleus for the purposes of acquiring NMR data.

If the molecule is in solution, then rapid isotropic tumbling will average out any orientation dependence of the quadrupolar coupling, and so there is no net effect on chemical shift. There is still a strong effect on the nuclear relaxation T_1 (equivalent to T_2 for small molecules in solution) as shown in equation (1.7) [11].

$$T_1^{-1} = T_2^{-1} = \frac{3\pi^2(2I+3)}{10I^2(2I-1)}\chi^2\left(1+\frac{\eta_Q^2}{3}\right)\tau_c \qquad \textbf{(1.7)}$$

where τ_c is the rotational correlation time – defined as the average (root-mean-squared) time taken for a molecule to tumble through an angle of *ca.* 1 radian; and χ is the quadrupolar coupling constant in equation (1.8).

$$\chi = \frac{e^2 q_{zz} Q}{h} \qquad \textbf{(1.8)}$$

in which e is the constant of fundamental electric charge and q_{zz} is the largest component of the *efg* tensor.

η_Q is the asymmetry parameter calculated from the principle components of the *efg* tensor as described in equation (1.9).

$$\eta_Q = \frac{q_{xx} - q_{yy}}{q_{zz}} \qquad \textbf{(1.9)}$$

From this expression it can be seen that the symmetry of a molecule or complex can have a significant effect upon T_2. Since linewidth is a function of T_2, signals from

asymmetric quadrupolar species in solution can be of the order of megahertz wide for nuclei with large *Q* values.

As discussed before, the resonant frequency is proportional to γ and low γ nuclei pose several problems in terms of acquiring data. Assuming no issues with isotopic abundance, nuclei with larger Q and low γ are therefore more challenging in terms of their data acquisition.

1.3. NMR spectrometer systems

1.3.1. Spectrometers

In contrast to early spectrometers, which were largely constructed by hand using a combination of custom electronics and off-the-shelf units designed for RF communications, modern spectrometers are industrially produced by vendors with electronic systems specifically designed and built for NMR.

Figure 1.4 is an example block diagram showing common spectrometer modules.

The user would interact with the spectrometer *via* a PC that in turn is connected to a spectrometer control unit (effectively an internal interface PC). From here, sub-modules either produce the RF pulse, receive and process the resulting NMR signal from the sample, or control other systems within the magnet or probe, such as shimming, lock and temperature. The prevalence of double-headed arrows in the diagram gives an indication of the level of feedback control between modules.

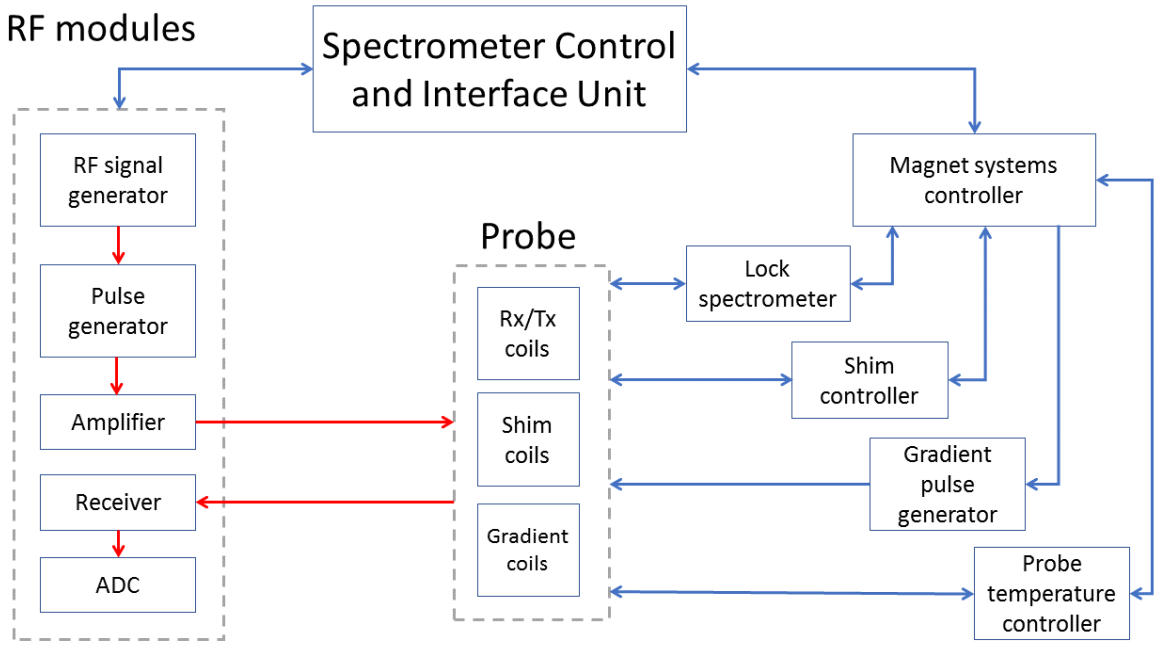


Figure 1.4 – Simplified block diagram of an NMR spectrometer (signal path in red, Rx = receiver, Tx = transmitter)

From a quantitative perspective, the critical components are those directly in the signal path, *i.e.* RF signal generators, pulse generators, amplifiers, receiver, Analogue to Digital Converter (ADC), and probes, as well as other associated cabling and filters (not shown here). The effects of some specific modules and components will be discussed further in the following Chapters.

1.3.2. Magnets

Alongside the electronics of the spectrometer, the other fundamental component required for NMR spectroscopy is a source of magnetic field.

Combining equations (1.3) and (1.4) to give equation (1.10):

$$\Delta E = h \cdot \frac{-\gamma B_0}{2\pi}\ Hz \qquad \textbf{(1.10)}$$

it can be seen that the energy gap, ΔE, between equilibrium and excited (or perturbed) states is proportional to the magnetic field strength, B_0, with the energy

levels being degenerate in the absence of a field. The larger the energy gap, the greater the population excess in the lower state (in accordance with the Boltzmann distribution) and hence the greater the SNR in the resulting NMR spectrum. Hence magnet design and strength is of great importance to NMR spectroscopy

In practice, there are other interconnected factors to consider, so the dependence of SNR on γ and B_0 is often approximated as in equation (1.11) [12] to:

$$\frac{signal}{noise} \propto \left|\gamma^{\frac{5}{2}}\right| B_0^{\frac{3}{2}} \qquad \textbf{(1.11)}$$

This could then be scaled appropriately by the natural abundance of the nuclei of interest.

There are four main sources of magnet field that are suitable for NMR spectrometers: permanent magnets, electromagnets, superconducting electromagnets and the earth itself.

Electromagnets (EM) require a stable, controlled source of electricity and efficient cooling systems to extract the heat generated by high currents in the solenoid coils. These factors have led to them being costly to build and run due to the demanding nature of the power handling [13]. EM spectrometers are still available from some suppliers [14], but are no longer in common use.

Permanent magnet spectrometers are more straightforward to maintain and, as long as the magnet is thermostatically controlled, they provide stable fields for a smaller running cost [15]. In the 1960s [16], the main alloy available was Alnico (composed of aluminium, nickel and cobalt), which gave field strengths up to 1.4 T (60 MHz).

Rare earth based materials such as SmCo, developed in the 1970s, and later NdFeB [17] in the 1980s, were able to produce the same field as Alnico magnets, for a much smaller size of magnet. The subsequent problem of maintaining magnetic field homogeneity over a usable volume with a smaller magnet was overcome following

the design and implementation of the Halbach array [18]. In its simplest form, a series of individual permanent magnets are arranged such that their magnetic fields augment on one side of the array whilst cancelling the magnet field to near-zero on the opposite side as shown in Figure 1.5, left.

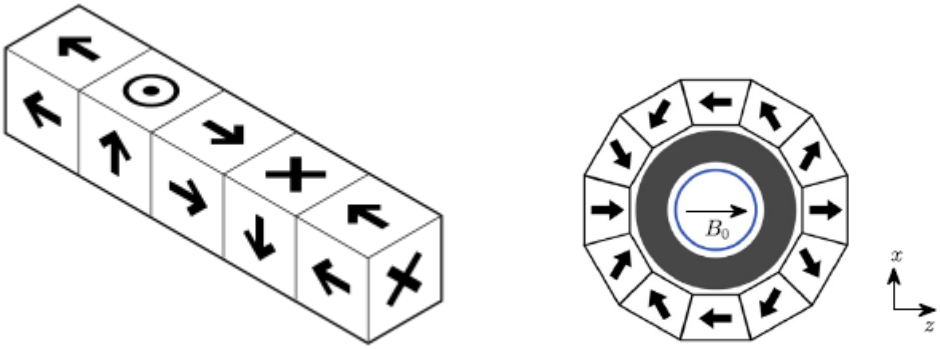


Figure 1.5 – A linear Halbach array (left) and cylindrical (right) Halbach array

In the array shown here, arrows denote the direction of field, with X pointing away and ⊙ pointing towards the viewer. The resultant magnetic field would give a strong field on the underside as represented and cancel out on the upper face. This concept can be extended to accommodate a cylindrical design as shown in Figure 1.5, right (taken from [19]) where B_0 denotes the direction of the magnetic field.

By stacking a series of such rings on top of one another, a homogenous field can be produced over a volume suitable to accept a 5 mm NMR tube as shown in Figure 1.6 (taken from [20]), with the field external to the cylinder being cancelled out. As an added advantage, the Z-axis of the field, as shown in these examples, is across the diameter of the cylinder and so a solenoid coil can be used to produce the orthogonal magnetic field in the xy plane that is required for excitation.

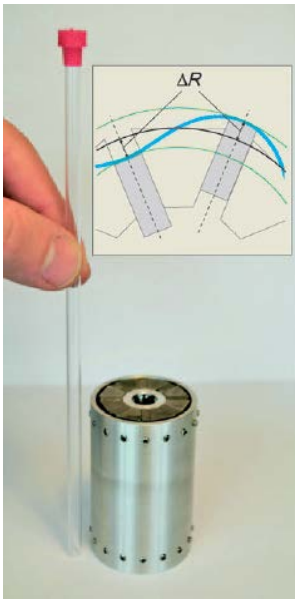


Figure 1.6 – A cylindrical Halbach array designed to accept a standard 5 mm NMR tube (inset shows adjustable magnets within the array to mechanically shim the overall field).

The phenomenon of superconductivity was discovered by Heike Kamerlingh in 1911 [21]. In summary, certain metals or alloys lose all resistance below a critical temperature, typically < 10 K [22]. Under these conditions, a current passing through the superconductor would encounter no resistance and so generate no heat. Solenoidal electromagnets created using these materials could also cope with higher current flows, hence higher field strength, and were persistent and self-stabilising as long as they were kept below their critical temperature with liquid helium.

Superconducting magnets have now become the standard for NMR spectroscopy and manufacturers continue to produce successively higher field spectrometers. In 2010, the worlds' first 1 GHz NMR spectrometer was installed at the European Centre for High Field NMR in Lyon by Bruker [23], representing a step up in magnetic field strength to 23.5 T. This magnet was subsequently superseded in design by a shielded version at the same field strength installed at the University of Bayreuth, Germany, exemplifying that developments in magnet technology are not just restricted to increasing field strength, but also extending to improving physical size limits and stray field reduction. At the time of writing the highest commercially

available magnet is a 1.2 GHz magnet under construction (starting in 2017) by Bruker [24].

The final source of magnetic field to discuss here is that of the earth itself. The earth's magnet field overall approximates to a dipole, and has the advantage that it is highly homogenous over short distances. The strength of this field varies from around 35 μT at the equator to 64 μT at the poles. This is of the order of 10^5 – 10^6 fold less than the fields produced by other sources discussed above and corresponds to Lamor frequencies of ~2 kHz. The SNR obtained at these frequencies is very low and requires either pre-polarisation [25], dynamic nuclear polarisation or the use of superconducting quantum interference devices (SQUIDs) [26] to detect the signal. For this reason, earth's field NMR (EFNMR) is not a mainstream technique, although spectrometers are commercially available for educational use [27] or for specific areas of research [28].

1.3.3. SNR enhancement

Although not utilised or explored in this thesis there are approaches available other than increasing field strength to either increase signal or decrease noise.

It has already been stated that signal from a sample is dependent upon the difference in population between the ground and excited spin states, which is in turn dependent upon the energy gap, ΔE (see equation (1.10)). Equation (1.12) gives the ratio of the population of upper and lower spin states (N_{upper} and N_{lower} respectively), in terms of this energy gap.

$$\frac{N_{upper}}{N_{lower}} = e^{-\frac{\Delta E}{kT}} \tag{1.12}$$

In practice, the population difference is very small. For example, when observing proton on a 9.4 T magnetic (ν = 400 MHz) at a temperature of 300K, at equilibrium there would be a population of 1,000,000 in the upper spin state and 1,000,0064 in

the lower spin state. Polarising the nuclei beyond this equilibrium is known as *hyperpolarization*, and can be achieved by several means including dynamic nuclear polarisation (DNP)[29], parahydrogen induced polarisation (PHIP)[30] and signal amplification by reversible exchange (SABRE)[31]. DNP relies on transferring the spin polarisation of from unpaired electrons to the analyte nucleus at a temperature of a few K, then warming the sample up and rapidly acquiring NMR data. This often requires long polarisation times of many hours. In contrast, PHIP chemically reacts hydrogen enriched in the *para* spin isomer with a substrate molecule. As only one spin state is populated in parahydrogen, the resultant signal after excitation is much greater and signal enhancements of the order of 10^4 can be achieved. SABRE represents a subsequent advancement in allowing the spin population of parahydrogen to be transferred to an analyte molecule via temporary association with a metal complex. As this does not require the direct chemical addition of H_2, the analyte remains unchanged.

Noise has been reduced in spectrometer consoles by gradual improvements in RF electronics including use of higher integrated circuitry and multi-layered circuit boards that reduce the signal path. A much greater step-change in noise reduction was made with the introduction of cryogenically cooled NMR probes, or “cryoprobes” [32]. By pumping compressed helium around the excitation coils and pre-amplifiers at temperatures of *ca.* 20K the thermal noise is reduced such that a SNR increase of around ×4 is achieved. These systems are widely available and common on high-end spectrometers in industry and academia.

1.4. Batch reaction and flow process monitoring

NMR is generally used as a tool for analysing isolated samples that have either been extracted or synthesised. However, many of the strengths of NMR discussed so far also make it highly desirable as a real time method for obtaining kinetic and structural information during a chemical reaction and for monitoring an ongoing industrial flow process [33-35]. Other factors in favor of the technique include fast acquisition times and flexibility in the setup and acquisition of different experiments in a sequence.

In the case of batch reaction monitoring, it is expected that the concentration of starting material, intermediate and product formation will change. This requires quantitative methods capable of monitoring these species over a wide concentration range for a defined period of time. In the case of flow process monitoring, once steady state has been achieved, it is expected that the species concentrations will be maintained at a constant level [36]. In this instance, long-term stability of the instrumentation becomes critical, as is the ability to distinguish small levels of concentration change within a narrow range.

For either case, monitoring can be carried out by removing a sample from a reaction vessel or flow process, quenching if necessary to halt the reaction, and then analysing by, for example, HPLC, GC or NMR. It can be of greater advantage to observe the reaction as it proceeds with as little interference as possible by monitoring the process in real time.

There are multiple approaches to real time reaction and process monitoring. The main methods are atline, *in situ*, online and inline [37].

1.4.1. Atline reaction monitoring

Traditionally, analysis by NMR (and other analytical techniques) takes place offline. That is, a process or chemical reaction is sampled, and the sample material taken to another location for analysis – this may require preparation (*e.g.* weighing, dilution), queueing up alongside prepared standards, manual post processing of the data to yield the results and the accompanying process of recording these details into an electronic laboratory notebook system. There are many steps in this process and analysis may take many hours, or even days.

Atline analysis seeks to reduce the turnaround time from sampling to results by developing instruments, sample introduction accessories and automated processing such that the analytics instruments can be co-located in the same laboratory or process area as the chemistry that requires the analysis [38, 39]. Sample

preparation and introduction needs to be trivial, and acquisition and processing should be automated.

The application of NMR to atline analysis will be explored further in subsequent Chapters of this thesis.

1.4.2. *In situ* reaction monitoring

With *in situ* reaction monitoring, the reaction under investigation is scaled down and run in an NMR tube, effectively bringing the reaction to the instrument. This may be to gain basic kinetic information for reactions as exemplified by kinetic studies into free-radical co-polymerisation by Mahdi *et al*. [40]. This approach also offers the opportunity to determine additional information. For example, in work by Mao *et al*. [41], an aldol condensation pathway was proposed for a series of functionalised pyrrolizidine producing reactions based on *in situ* NMR reaction monitoring. In the same works, *in situ* NMR was used to rapidly screen the effect of different additives on the reaction. In other work by Codina *et al.* [42], mechanistic understanding was gained for a chemical process under development in the pharmaceutical industry by reaction monitoring in an NMR tube. The work presented also highlights the advantage of being able to calculate immediate mass balance on the integrals of the species present without need for calibration.

Problems with resolution of complex mixtures can be overcome by acquisition of 2D correlation experiments that can offer additional dispersion of the signals for the species of interest in a second dimension. For example, Duarte-Ruiz *et al*. [43] studied the species formed in reaction between DMSO-d6 and iodomethane utilising HSQC and HMBC experiments over a series of many days. In the case of work by Tauler *et al.* [44], analysis of a reaction between ^{15}N labelled cisplatin and an amino acid-nucleotide hybrid was followed with a series of (^{1}H, ^{15}N) HSQC experiments. Multivariate analysis was then used to interpret the data, producing reaction profiles and allowing kinetics to be calculated.

In both of these examples, 2D data acquisition was only possible because the reactions were very slow, raising the issue that 2D experiments have relatively long acquisition times compared to 1D experiments. The additional time required to acquire 2D data can be partially alleviated by so-called ultrafast 2D experiments [45], which take advantage of the fact that the majority of a typical 2D spectra is “empty space” containing no useful data. Prior knowledge of the spectral appearance is then used to only acquire regions of interest in a process referred to as non-uniform sampling. An example by Queriroz *et al.* [46] utilises ultrafast 2D methods to acquired HSQC experiments in less than 1 s to monitor acetal hydrolysis *in situ*.

In a second example, Giraudeau *et al.* [47] demonstrated that it was also possible to fit an electrochemical cell into an NMR tube in a high field magnet and use ultrafast 2D experiments to follow redox reactions *in situ*.

Analysis by heteronuclear NMR offers the advantage of being able to interrogate the reaction with respect to the reagents, and the resulting spectra are often simpler with less interference from other species [48-52]. The ability to interleave experiments that analyse specific nuclei such as ^{13}C, ^{27}Al, ^{11}B, ^{29}Si, ^{31}P, ^{35}Cl, ^{79}Br and ^{15}N as the reaction proceeds allows a high level of discrimination to be made between the species present. A recent work that exemplifies the heteronuclear approach is given by Lippens *et al.* [53], where ^{1}H and ^{31}P experiments were acquired simultaneously to monitor the phosphorylation of glucose by hexokinase. A calibration curve was generated to relate chemical shift changes to the pH of the solution allowing an additional parameter to be monitored during the ongoing reaction. Another recent study from Sandford et *al.* [54] uses the relatively high sensitivity of the ^{19}F nucleus to understand the mechanism and side reactions of arylsulfonate electrophiles with tetramethyl ammonium fluoride. ^{19}F spectra also benefit from a wide chemical shift range of several hundred ppm that provides advantages in resolution, even at lower field strengths.

Some additional disadvantages to running reactions in an NMR tube, other than those already discussed, include:

- Absence of stirring
- Lack of stepwise addition of reagents
- Formation of precipitate leading to poor magnetic field homogeneity which in turn adversely affects the lineshape of the resulting data
- A narrow glass tube is not necessarily a good model for a typical reaction vessel

A further disadvantage is that the manual mixing of reagents and manual transfer to the magnet causes an initial delay before acquisition can begin, limiting the reaction rates which can be observed by this technique. Bruker has now provided an automated approach to *in-situ* analysis [55], providing a unit with a series of syringe pumps that sits directly above the NMR magnet. These syringes inject reagents into an integrated mixing chamber and deliver the solution directly into the NMR probe in a flowcell in around 130 ms, allowing data to be acquired on reactions with half-lives as low as 1 s by use of a stopped-flow approach.

1.4.3. Online and inline reaction monitoring

By contrast, with online reaction monitoring, the reaction is carried out in a larger vessel in close proximity to the NMR spectrometer, which monitors the contents of the vessel, preferably with minimum interference. Online monitoring enables parameters such as temperature, stirring rate and reagent addition rate to be investigated as well as the basic chemical mechanism. An example of monitoring in typical reaction vessels is the work carried out by Bernstein *et al.* [51]. Figure 1.7 (taken from this reference) shows the 100 mL glass vessel set up with typical heating, sample introduction and stirring capabilities.

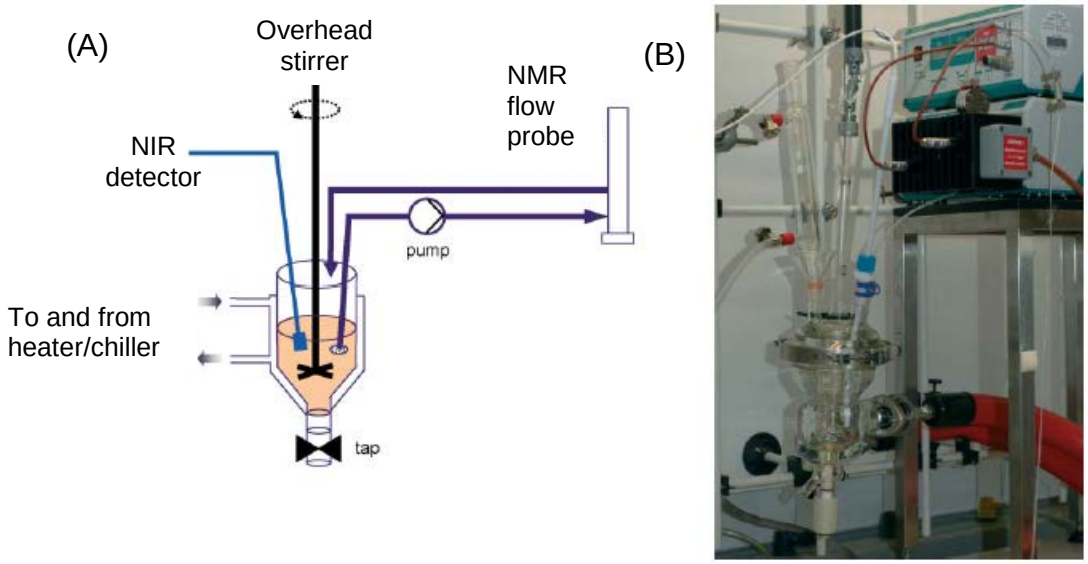


Figure 1.7 – Flow reactor apparatus shown in (A) schematic and (B) photograph

In this set up, the reaction is sampled by means of a pump, and drawn through heated narrow bore lines to a flow probe located in a nearby NMR spectrometer.

Complete off-the-shelf solutions are now available to achieve online NMR monitoring at high field. For example Bruker's InsightMR system includes temperature controlled lines to take the reaction solution to and from the reaction vessel via an integrated flowcell that fits down a standard narrow bore magnet using existing 5 mm probes [56].

As will be discussed in more detail in chapter 3, benchtop NMR is particularly attractive for online reaction monitoring as the spectrometers are cheaper compared to high field instruments and can be located much closer to chemical reaction setups. In an example by Warren *et al.* [57], the kinetics of a polymerisation reaction are ascertained for the same reaction in both batch and flow chemistry modes. This would have required no more additional hardware than a flowcell insert and a small pump, both of which can be supplied by the vendor along with the spectrometer at minimal additional expense. Singh and Blümich [58] also use a similar setup to monitor the effect of different isotopes on the hydrolysis of acetic anhydride and ethyl trifluoroacetate, in this case including a cold water bath in the flow path to cool the reagents before entering the magnet.

In both of the examples above separate experiments were set up and run and the data for each processed manually. In contrast, a chemistry platform build by Kappe *et al*. [59] is designed around an automated modular microreactor and is capable of queuing and running multiple experiments with integrated analysis by NMR, IR and UPLC. Such a platform would allow rapid optimisation of a chemical reaction with lower resources and time spend in the lab performing manual chemistry and analytical tasks.

The small form factor of benchtop instruments may also make them easier to install in an industrial plant environment. The work by Maiwald et al. [60] uses benchtop NMR to acquired and automatically processed data for a chemical transformation in a continuous reactor. Although this was performed at a laboratory scale, the intention of the work was to demonstrate the applicability of this approach to the control of larger scale manufacturing processes. Later works also led by Maiwald [35] demonstrate explicitly how a benchtop spectrometer can be installed in an explosion-proof housing in an industrial plant environment. In this example automated data acquisition and processing was used to optimise plant performance of the process and also provide reference data to calibrate a NIR also attached to the process.

Although the term inline and online and sometimes used interchangeable, including in the literature references given here, they are different sampling regimes. Whereas online monitoring takes a diverted sample which then returns to the main flow, online monitoring places the instrument directly in the main flow. Due to the constraints of magnet/probe diameter NMR would only be used inline in case where the main flow could pass through a relatively small diameter tubing. The platform by Kappe *et al.* referenced above is an example of this.

1.5. Data Processing

The concepts introduced thus far have primarily covered the acquisition of NMR spectroscopic data. Once raw data is acquired from the spectrometer, it may undergo multiple levels of processing in several steps to yield quantitative numerical results.

The first level of processing converts the FID time domain data to a spectrum in the frequency domain by Fourier Transform. This is usually performed automatically by most spectrometer software, with opportunity for the user to set or adjust critical parameters. Some of the typical parameters and functions applied during this step, such as zero-filling and apodisation, are discussed further in Chapter 2.

At the next level, the FT spectrum is manipulated and corrected to remove or adjust features that affect the consistency of the spectral appearance. For example, phase correction to adjust for the effects of phase difference between transmitter and receiver components in the spectrometer, baseline correction to remove artefacts, chemical shift referencing to correct for field drift, derivatisation to change the emphasis of the observed spectral features, and so on. The implications of these are discussed in Chapters 2, 4 and 5.

The final level of processing extracts numerical data, representing the quantity of analytes present in the sample, from the spectrum. These numerical data are then used in the calculations and algorithms that yield the quantitative results. Two of the main methods for extracting the numerical data are integration and multivariate analysis (MVA). Integration calculates the area of specific analyte peaks and is the simplest and most common method applied when peaks are baseline resolved. If there is poor resolution of analyte peaks or interference from other baseline artefacts, spectral deconvolution, where mathematical functions are fitted to match the peak lineshapes in the spectrum, may be used instead. As spectra become more complex and peak overlap increases it may not even be possible to identify clearly which peaks belong to which analytes. In such cases, a process of MVA may be used, where whole spectra are analysed mathematically for differences and these

are then correlated to known concentrations in calibration solutions to produce a predictive model for future samples. These different approaches are compared and discussed in depth in Chapter 4.

1.6. Automation

As the development of NMR hardware and software has increased over time, the degree of user interaction required to generate NMR data has subsequently decreased. Set up processes such as tuning, matching, locking and shimming are routinely automated in walk-up systems, and most basic 1 and 2 dimensional spectra can be acquired quickly by non-expert users *via* a graphical user interface (GUI).

Vendors such as Gilson provide simple modular automated sample preparation platforms for solid powder and liquid handling and Bruker have produced platforms specific for dilution and dispensing of samples into NMR tubes, such as the SamplePro Tube [61].

Ideally, sample preparation should be minimised by developing methods that can analyse, for example, neat reaction or process solutions (see Chapter 4 and 5), or sample directly from the process.

At the highest level of automation it is necessary to step back from the individual unit steps of preparation, acquisition and processing and view the analysis as part of a cycle of information/data flow. This starts with the attributes, history and reason for creation of an analytical sample. A sub-set of this information needs to be transferred to an analytical platform by a suitably trained user, and data acquisition initiated. Following data acquisition, the data requires processing to render new information in the form of results. Finally, all the output from the platform (raw data, processed data, processing details, results) needs to be archived and disseminated such that the relevant new information is made available where it is required, and all other associated information is retrievable in the future. Additional considerations around long-term archive maintenance and (if required), deletion of information are also part of the lifecycle aspects of the flow.

Within industry, this high-level view of information flow is referred to as knowledge management . The aspects of information creation and storage are also referred to in terms of Data Integrity (DI). For the pharmaceutical industry, there exist formal requirements from agencies such as the U.S. Food and Drug Administration (FDA). An example of this are the “ALCOA” principles [62], which require that experimental records and associated data are:

Attributable	-Is it clear who carried out and recorded an experiment
Legible	-The information is readable and permanent
Contemporaneous	-The records/data were created at the same time as the experiment
Original	-Raw data should be retained as well as derived data
Accurate	-The records should reflect the work that actually took place

This acronym was originally quoted by a US FDA official in the 1990s, but since then additional principles have been added including Complete, Consistent, Enduring, and Available.

Having listed some of the aspects of analysis that can be automated, it becomes obvious that the challenge of developing a highly automated analytical platform becomes one of integration. Although vendors of most analytical instruments will offer automated solutions to fit a variety of scenarios, it is still often the case that an analysis will require a specific instrument, with specific data processing, reporting and storage requirements that cannot be met by a single piece of software. The developing analyst then needs to be able to integrate several pieces of software such that the end user does not need to access multiple software packages and transfer data between them. This can be achieved in several ways, for example by writing scripts within a software package that can enhance its capability to perform additional specific functions such as data input, processing, archiving and reporting (see Chapter 2). Another approach is to write new software to act as a “glue” to coordinate several other packages and provide a single GUI for the end user (see Chapter 4).

These approaches are all reliant on vendors providing scripting and/or external control functions in their software. The presence, or lack, of such functionality should therefore be a deciding factor in the selection and purchase of new instrumentation.

1.7. Scope of report

The potential to utilise NMR spectroscopy as a powerful quantitative analytical tool has been discussed, and some examples of current applications and technological developments have been shown. The basic parameters and properties of the NMR phenomena and spectrometer, magnets and other associated hardware have also been introduced with respect to their impact on quantitation. Finally, the advantages of automation to the end user, and its challenges to the developing analyst, were presented.

The work presented in this thesis will expand on the themes of quantitation and automation and show that NMR spectroscopy is not exclusively a technique that requires large, immobile instruments run by NMR experts, but one where methods can be developed for the non-expert user, using small, portable instruments.

It will be shown that a fundamental understanding of the spectrometer and experimental parameters is required to control the analysis to the point where accurate, reproducible data can be produced with minimum user intervention. Utilization and integration of existing software, along with the coding of custom processing routines and user interfaces can be used to pull together different aspects of an NMR method and turn them into a unified analytical tool.

The work discussed here will start with the development of an automated chloride analysis method by walk-up high field NMR. The work will then move to a portable low field bench top NMR spectrometer, investigating key acquisition and processing aspects. This will lead on to the development of an automated user interface and basic quantitative methodology for atline analysis of simple mixtures. Finally, the application of multivariate analysis to the quantitation of more complex mixtures at low field will be explored.

For each chapter, a separate project aim introducing the overall subject of the studies and key results will be given. This will be followed by a more detailed background to the subject area, including relevant literature references. The experimental details will then be listed, grouped by study type, and followed by the results and discussion, also broken down in the same study type and order. Finally, the work will be summarised in a conclusion.

Overall concluding remarks to the report will be given at the end.

References

[1] E. M. Purcell, H. C. Torrey and R. V. Pound, Resonance Absorption by Nuclear Magnetic Moments in a Solid. Physical Review, 69 (1946) 37-38.

[2] F. Bloch, W. W. Hansen and M. Packard, Nuclear Induction. Physical Review, 69 (1946) 127-127.

[3] R. J. Abraham (1971), Introduction. Analysis of high resolution NMR spectra, Elsevier Publishing Company, Amsterdam, 1-4.

[4] R. E. Ernst, G. Bodenhausen and A. Wokaun (1987), Nuclear Spin Hamiltonian. Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Oxford University Press, Oxford, 44-49.

[5] E. Fukushima and S. B. W. Roeder, Spurious ringing in pulse NMR. Journal of Magnetic Resonance, 33 (1979) 199-203.

[6] D. E. Lide (2000), Nuclear spins, moments and other data related to NMR spectroscopy. CRC Handbook of Chemistry and Physics, 80th Edition, CRC Press, New York, 9-92 - 99-94.

[7] A. G. Bell, W. Koz´min´ski, A. Linden and W. von Philipsborn, ^{187}Os NMR Study of (η^6-Arene)osmium(II) Complexes: Separation of Electronic and Steric Ligand Effects. Organometallics, 15 (1996) 3124-3135.

[8] G. A. Morris and R. Freeman, Enhancement of nuclear magnetic resonance signals by polarization transfer. Journal of the American Chemical Society, 101 (1979) 760-762.

[9] K. Lu, Y. Miyazaki and M. Summers, Isotope labeling strategies for NMR studies of RNA. Journal of Biomolecular NMR, 46 (2010) 113-125.

[10] P. J. Hore (2004), Spin-spin coupling. Nuclear Magnetic Resonance, Oxford University Press, Oxford, 22-43.

[11] R. K. Harris and B. E. Mann (1978), Introduction. NMR and the Periodic Table, Academic Press Inc., London, 1-19.

[12] F. A. Nelson and H. E. Weaver, Nuclear Magnetic Resonance Spectroscopy in Superconducting Magnetic Fields. Science, 146 (1964) 223-232.

[13] E. Fukushima and B. W. Roeder (1981), NMR Hardware - Magnets. Experimental Pulse NMR - A nuts and bolts approach, Addison Wesley, Massachusetts, 312-318.

[14] The Eft-60 & Eft-90 NMR Spectrometers, https://aiinmr.com/high-quality-durable-nmr-spectrometers-pmnmr-instruments/, Retrieved 12 Dec 2018.

[15] B. A. Evans and R. E. Richards, Temperature-controlled permanent magnet for high-resolution nuclear magnetic resonance. Journal of Scientific Instruments, 37 (1960) 353.

[16] J. W. Emsley, J. Feeney and L. H. Sutcliffe (1967), NMR Spectrometers and their Accesories - The Magnet. High Resolution NMR Spectroscopy, Pergamon, Oxford, 201-202.

[17] M. Sagawa, S. Fujimura, H. Yamamoto, Y. Matsuura and K. Hiraga, Permanent magnet materials based on the rare earth-iron-boron tetragonal compounds. IEEE Transactions on Magnetics, 20 (1984) 1584-1589.

[18] K. Halbach, Design of permanent multipole magnets with oriented rare earth cobalt material. Nuclear Instruments and Methods, 169 (1980) 1-10.

[19] J. Mitchell, L. F. Gladden, T. C. Chandrasekera and E. J. Fordham, Low-field permanent magnets for industrial process and quality control. Progress in Nuclear Magnetic Resonance Spectroscopy, 76 (2014) 1-60.

[20] E. Danieli, J. Perlo, B. Blümich and F. Casanova, Small Magnets for Portable NMR Spectrometers. Angewandte Chemie International Edition, 49 (2010) 4133-4135.
[21] H. Kamerlingh, "For his investigations on the properties of matter at low temperatures which led, *inter alia*, to the production of liquid helium.". Nobel Prize in Physics (1913)
[22] H. Kamerlingh, Investigations into the Properties of Substances at Low Temperatures, which Have Led, amongst Other Things, to the Preparation of Liquid Helium. Nobel Lecture (1913)
[23] Bruker, Chemistry: Breaking the billion-hertz barrier. Nature, 463 (2010) 605-606.
[24] H. Schwalbe, Editorial: New 1.2 GHz NMR Spectrometers— New Horizons? Angewandte Chemie International Edition, 56 (2017) 10252-10253.
[25] A. Mohoric and J. Stepísnik, NMR in the Earths magnetic field. Progress in Nuclear Magnetic Resonance Spectroscopy, 54 (2009) 166-182.
[26] L. Q., Y. Zhang, H. J. Krause, A. I. Braginski, M. Burghoff and L. Trahms, Nuclear magnetic resonance in the earth's magnetic field using a nitrogen-cooled superconducting quantum interference device. Applied Physics Letters, 91 (2014) 072505.
[27] Earths field NMR, http://www.teachspin.com/earth-s-field-nmr-gradient-field-coil-system.html, Retrieved 02 Nov 2018.
[28] M. W. Hunter, R. Dykstra, M. H. Lim, T. G. Haskell and P. T. Callaghan, Using Earths Field NMR to Study Brine Content in Antarctic Sea Ice: Comparison with Salinity and Temperature Estimates. Applied Magnetic Resonance, 36 (2009) 1-8.
[29] C. R. Bowers and D. P. Weitekamp, Parahydrogen and synthesis allow dramatically enhanced nuclear alignment. Journal of the American Chemical Society, 109 (1987) 5541-5542.
[30] T. C. Eisenschmid, R. U. Kirss, P. P. Deutsch, S. I. Hommeltoft, R. Eisenberg, J. Bargon, R. G. Lawler and A. L. Balch, Para hydrogen induced polarization in hydrogenation reactions. Journal of the American Chemical Society, 109 (1987) 8089-8091.
[31] R. W. Adams, J. A. Aguilar, K. D. Atkinson, M. J. Cowley, P. I. P. Elliott, S. B. Duckett, G. G. R. Green, I. G. Khazal, J. López-Serrano and D. C. Williamson, Reversible Interactions with para-Hydrogen Enhance NMR Sensitivity by Polarization Transfer. Science, 323 (2009) 1708-1711.
[32] P. Styles, N. F. Soffe and C. A. Scott, The first cryoprobe – Some recollections. Journal of Magnetic Resonance, 213 (2011) 355-356.
[33] M. V. Gomez and A. de la Hoz, NMR reaction monitoring in flow synthesis. Beilstein Journal of Organic Chemistry, 13 (2017) 285-300.
[34] P. Giraudeau and F.X. Felpin, Flow reactors integrated with in-line monitoring using benchtop NMR spectroscopy. Reaction Chemistry & Engineering, 3 (2018) 399-413.
[35] S. Kern, L. Wander, K. Meyer, S. Guhl, A. R. G. Mukkula, M. Holtkamp, M. Salge, C. Fleischer, N. Weber, R. King, S. Engell, A. Paul, M. P. Remelhe and M. Maiwald, Flexible automation with compact NMR spectroscopy for continuous production of pharmaceuticals. Analytical and Bioanalytical Chemistry, 411 (2019) 3037-3046.

[36] A. R. Bogdan and A. W. Dombrowski, Emerging Trends in Flow Chemistry and Applications to the Pharmaceutical Industry. Journal of Medicinal Chemistry, 62 (2019) 6422-6468.
[37] B. G. Lipták (2017), Analyzer selection and application. Analysis and Analyzers, CRC Press, Florida, 26 - 27.
[38] N. Schmitz, A. Friebel, E. von Harbou, J. Burger and H. Hasse, Liquid-liquid equilibrium in binary and ternary mixtures containing formaldehyde, water, methanol, methylal, and poly(oxymethylene) dimethyl ethers. Fluid Phase Equilibria, 425 (2016) 127-135.
[39] N. N. A. Ling, A. Haber, E. F. May, E. O. Fridjonsson and M. L. Johns, By-line NMR emulsion droplet sizing. Chemical Engineering Science, 160 (2017) 362-369.
[40] M. Abdollahi, S. Mehdipour-Ataei and F. Ziaee, Using 1H-NMR spectroscopy for the kinetic study of the in situ solution free-radical copolymerization of styrene and ethyl acrylate. Journal of Applied Polymer Science, 105 (2007) 2588-2597.
[41] Z.-Y. Mao, Y.-W. Liu, P. Han, H.-Q. Dong, C.-M. Si, B.-G. Wei and G.-Q. Lin, Regio- and Stereoselective Cascades via Aldol Condensation and 1,3-Dipolar Cycloaddition for Construction of Functional Pyrrolizidine Derivatives. Organic Letters, 20 (2018) 1090-1093.
[42] I. M. Clegg, C. M. Gordon, D. S. Smith, R. Alzaga and A. Codina, NMR reaction monitoring during the development of an active pharmaceutical ingredient. Analytical Methods, 4 (2012) 1498-1506.
[43] E. Avella-Moreno, N. Nuñez-Dallos, L. Garzón-Tovar and A. Duarte-Ruiz, Reactions and products revealed by NMR spectra of deuterated dimethylsulfoxide with iodomethane in neutral and basic media. Journal of Sulfur Chemistry, 36 (2015) 535-543.
[44] J. Jaumot, V. Marchán, R. Gargallo, A. Grandas and R. Tauler, Multivariate Curve Resolution Applied to the Analysis and Resolution of Two-Dimensional [1H,15N] NMR Reaction Spectra. Analytical Chemistry, 76 (2004) 7094-7101.
[45] A. Tal and L. Frydman, Single-scan multidimensional magnetic resonance. Progress in Nuclear Magnetic Resonance Spectroscopy, 57 (2010) 241-292.
[46] L. H. K. Queiroz Jr., P. Giraudeau, F. A. B. dos Santos, K. T. de Oliveira and A. G. Ferreira, Real-time mechanistic monitoring of an acetal hydrolysis using ultrafast 2D NMR. Magnetic Resonance in Chemistry, 50 (2012) 496-501.
[47] R. Boisseau, U. Bussy, P. Giraudeau and M. Boujtita, In Situ Ultrafast 2D NMR Spectroelectrochemistry for Real-Time Monitoring of Redox Reactions. Analytical Chemistry, 87 (2015) 372-375.
[48] T. Bartik, Comments on the synthesis of trisulfonated triphenylphosphine: reaction monitoring by NMR spectroscopy. Inorganic Chemistry, 31 (1992) 2667–2670.
[49] R. K. Grover, B. S. Joshi, S. Batra, R. Roy and A. P. Bhaduri, Translactonization of erythromycin A during oximation: mixture analysis and reaction monitoring by NMR. Magnetic Resonance in Chemistry, 39 (2001) 355-360.
[50] B. Y. W. Man, Substitution reactions in ionic liquids. A kinetic study. Tetrahedron Letters, 46 (2005) 7641-7645.

[51] M. A. Bernstein, M. Stefinovic´ and C. J. Sleigh, Optimising reaction performance in the pharmaceutical industry by monitoring with NMR. Magnetic Resonance in Chemistry, 45 (2007) 564-571.
[52] S. J. Chan, A. G. Howe, J. M. Hook and J. B. Harper, ^{79}Br NMR spectroscopy as a practical tool for kinetic analysis. Magnetic Resonance in Chemistry, 47 (2009) 342-347.
[53] N. Cox, R. Kuemmerle, P. Millard, E. Cahoreau, J.-M. François, J.-L. Parrou and G. Lippens, Integrated pH Measurement during Reaction Monitoring with Dual-Reception 1H–31P NMR Spectroscopy. Analytical Chemistry, 91 (2019) 3959-3963.
[54] S. D. Schimler, R. D. J. Froese, D. C. Bland and M. S. Sanford, Reactions of Arylsulfonate Electrophiles with NMe4F: Mechanistic Insight, Reactivity, and Scope. The Journal of Organic Chemistry, 83 (2018) 11178-11190.
[55] InsightXpress, https://www.bruker.com/products/mr/mr-in-pharma/insightxpress/overview.html, Retrieved 02 Nov 2018.
[56] New Approaches to On-line and In-Situ NMR Reaction Monitoring: Fast and Furious Data, https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/MagneticResonance/NMR/InsightMR_white_paper.pdf, Retrieved 02 Nov 2018.
[57] S. T. Knox, S. Parkinson, R. Stone and N. J. Warren, Benchtop flow-NMR for rapid online monitoring of RAFT and free radical polymerisation in batch and continuous reactors. Polymer Chemistry, (2019)
[58] K. Singh and B. Blümich, Online monitoring of the kinetic isotope effect in chemical reactions with 1H and 19F low-field NMR spectroscopy. Analyst, 143 (2018) 4408-4421.
[59] P. Sagmeister, J. D. Williams, C. A. Hone and C. O. Kappe, Laboratory of the future: a modular flow platform with multiple integrated PAT tools for multistep reactions. Reaction Chemistry & Engineering, 4 (2019) 1571-1578.
[60] S. Kern, K. Meyer, S. Guhl, P. Gräßer, A. Paul, R. King and M. Maiwald, Online low-field NMR spectroscopy for process control of an industrial lithiation reaction—automated data analysis. Analytical and Bioanalytical Chemistry, 410 (2018) 3349-3360.
[61] SamplePro Tube, https://www.bruker.com/products/mr/nmr/automation/samplepro-tube/overview.html, Retrieved 28 Sep 2018.
[62] Current Expectations and Guidance, including Data Integrity and Compliance With CGMP, https://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/UCM518522.pdf, Retrieved 02 Nov 2018.

Chapter 2

Standardless, Automated Quantitation of Chloride by NMR

2.1. Introduction

2.1.1. Project aim

In this Chapter, the use of heteronuclear open access NMR technology to support synthetic chemistry development will be explored. Alternative uses of open access spectrometers will be sought and quantitative methods developed that run under the same automation software as other routine experiments to provide new tools for chemical analysis.

In the process of developing a synthetic chemical route such that it is suitable to become a manufacturing scale process, individual synthetic steps are investigated in detail. Parameters investigated can include temperature, solvent, stirring speed, reagent concentration, reagent batch/manufacturer and so on. This may require running hundreds of reactions, which in turn require fast, efficient analysis of species of interest to support the work. Within industry, there is therefore a strong driver to continue to look for ways to reduce the resource and time required for the analysis of common analytes.

The aim of the following investigation is to develop and validate a quantitative, fully automated method for chloride analysis, ideally not requiring standards to be run along with the samples, and with the lowest possible run time and best possible detection/quantitation limits.

A summary of this work has been published in a peer reviewed journal [1], a copy of which is included in Appendix 3.

2.1.2. Background

2.1.2.1. Chlorine and NMR

Chlorine is one of the more commonly encountered halogens in synthetic chemistry. The strong electronegativity of the atom combined with an electron deficient $3p^5$ outer electron shell configuration means that it will readily form compounds with almost all other elements in the periodic table. Chlorine is the 20th most abundant element on the earth. The seas contain trillions of tons of chlorine, with world production of salt from mining exceeding 160 million tons per annum [2]. Thus it is relatively easy and cheap to source chlorine from sodium chloride.

Chlorine has an electronegativity of 3.16 on the Pauling scale (the 3rd highest after fluorine and oxygen) and so forms reactive centres on molecules. This reactivity also stems from the weakness of the C-Cl bond when compared to the C-C bond and the relatively high stability of the resulting chloride anion formed when a chlorine atom is lost from organic chlorides. For this reason it is used in a wide variety of substitution and addition reactions, as well as being present in catalysts, reagents and as a native counterion.

Chlorine is typically encountered in the pharmaceutical industry in the form of chloride (Cl^-) either intentionally present as a counterion for API salts or a contaminant left over as a reaction by-product or from brine washes. For this reason it is important to be able to confirm accurately the stoichiometry of a chloride salt, or the removal of chloride contamination from a process, and so a quantitative analytical method is required.

As with any nucleus, the utility of NMR to chlorine analysis can be discussed in terms of the nuclear properties of the isotope. Chlorine is present in nature as two stable isotopes, ^{35}Cl and ^{37}Cl (see Table 2.1 for properties [3]).

Table 2.1 – Nuclear properties of chlorine isotopes

Isotope	^{35}Cl	^{37}Cl
Natural abundance (%)	75.53	24.47
γ (rad T^{-1} s^{-1})	2.6212	2.1815
Larmor frequency at 9.4 T (MHz)	39.19	32.62
Nuclear spin	3/2	3/2
Quadropolar moment (mB)	-100	-79
Overall sensitivity (^{13}C = 1)	20.15	3.76

^{37}Cl has a smaller quadrupolar moment, but as ^{35}Cl has 3 × greater abundance and a higher overall sensitivity it tends to be the preferred isotope for analysis by NMR. Although 20 × more sensitive than ^{13}C, ^{35}Cl is still over 200 × less sensitive than ^{1}H and so every effort is made in the following works to maximise the SNR of the data by optimising all relevant parameters.

For the purposes of context, the quadrupolar moment of ^{35}Cl is close to that of the more commonly encountered nucleus, ^{23}Na. Rapid isotropic tumbling in solution removes any net effect of the quadrupolar moment on chemical shift, but still leaves a strong effect on T_1, T_2 and hence a large linewidth. This effect is minimised for situations where the nucleus is in an environment of spherical symmetry and thus has a low electric field gradient. As an example, the linewidth for aqueous NaCl is quoted in literature as *ca.* 10 Hz, whereas the linewidth for CCl_4 is *ca.* 10 kHz [4]. This has the additional advantage that for the types of sample that will be typically encountered, the only resonances likely to be observed in the ^{35}Cl NMR spectrum will be either Cl^- or the less common ClO_4^- as any covalently bonded chlorides will have linewidth of the same order as, or greater than, the full spectral width.

2.1.2.2. Automation and analysis

Historically, chemists would have determined chloride content by using Mohr's method [5], where aqueous silver nitrate is titrated into the solution for analysis, with potassium chromate added as an indicator to visualise the endpoint. Modern laboratories are more likely to use either automated titration apparatus, in which the endpoint is determined potentiometrically, or a technique such as ion chromatography (IC) [6].

Analogous to HPLC, IC is a highly industrialised technique that can be hyphenated to advanced automation systems capable of weighing, dissolving and injecting samples into the chromatograph, processing data and then reporting results. However, most systems require human intervention in the form of analytical experts to prepare samples, queue them on an autosampler, integrate the chromatography and report results back to the customers. All liquid chromatography techniques also need a constant flow of mobile phase to operate, requiring further time on the part of the analytical expert to prepare solutions, flush columns and then dispose of waste.

Application notes from a method developed on a Thermoscientific ICS-300 IC [7] quote a limit of quantitation for chloride in drinking water of 500 μg/L, which equates to 500 ppb. The limit of detection would be expected to be lower still (see later in this chapter). Application notes for even their most basic Aquion IC [8] shows the ability to achieve an intensive peak for chloride at around 100 mg/L, or 100 ppm, with a 5 min run time. By visual inspection, the limit of detection would be expected to be at least an order of magnitude or more lower. These instruments are design primarily for trace analysis and quantitation of routine samples from a chemical process would therefore likely require significant dilution to prevent overloading both the chromatographic column and the detector.

NMR spectrometers are ideally suited for automation and walk up analysis. Apart from regular fills of liquid nitrogen and helium to maintain the superconducting coils at temperature, the systems require minimum or no daily intervention, except to empty autosamplers if required and to check on the overall performance of the

instrument. Although more complex data can require many hours of manual processing, for routine processes and calculations it can generally be assumed that if a task or series of tasks can be performed by an analytical expert using a computer, then it can be automated.

Only one example of quantitative chloride analysis by NMR was found in the literature, by Lim and Lee [9]. The published works state that chloride was detected at a concentration of 20 μg/mL, or 20 ppm, with a 5 min run time on a 700 MHz NMR spectrometer, but by visual inspection this peak was only just discernable. This indicates that the inherent sensitivity is likely to be appropriate for process development samples without the need for repeated dilution steps or lengthy run times. The authors also note that improved signal to noise can be achieved with longer run times.

2.1.2.3. Validation

As noted in Chapter 1, analytical methodology can be rigorously tested *via* a process of validation, as defined in accepted industry standards such as ICH [10] or USP.

With reference to ICH standards for the development of a quantitative method [11], the main validation aspects that require consideration are:

- Accuracy
- Precision
    - Repeatability
    - Intermediate precision
- Specificity
- Detection limit
- Quantitation limit
- Linearity
- Range

The suite of validation aspects to be applied to a given method depends upon the degree of confidence required for the results the method will yield. For example, if a purity figure is to be determined for a batch of material destined for a human clinical trial, then the highest level of confidence is required and all validation aspects listed must be covered. For the purposes of this method, which will be to support synthetic route development, intermediate precision will not be required. Range is not typically assessed by isolated experiments, but instead represents the range of results over which the method has demonstrated acceptable linearity, accuracy and precision. Accuracy will not be formally assessed as this requires the analysis of certified reference standards. Instead, the recovery of prepared samples of chloride containing different species will be assessed. All other aspects will be further discussed in the following sections.

2.2. Experimental

All data were acquired on a 9.4 T Ultrashield 400 MHz Bruker Avance-III NanoBay NMR spectrometer equipped with a BCU-05 temperature controller and using a 5 mm BBFO probe with a broad band channel tunable from ^{31}P to ^{15}N (including ^{19}F) with automatic tuning and matching (ATMA). ^{35}Cl data were acquired at a frequency of 39.19 MHz.

All samples were weighed using a 5-place Mettler MX5 microbalance. Dilutions were made using either a Rainin 1000 µL electronic displacement pipette or a Gilson 25 µL positive displacement pipette. A volume of 600-700 µL of each solution was transferred to Norell Standard Series™ 5 mm diameter NMR tubes for analysis. Data were acquired using a pulse and acquire sequence (Bruker zg pulse sequence) and the final parameters following this investigation are summarised in Table 2.2.

Table 2.2 – Method parameters

Parameter	Value
90° Pulse width (µs)	13.4
Pulse power (W)	100
Relaxation delay (ms)	1
Dead time (µs)	17
Acquisition time (s)	0.174
Number of data points	2048
Dwell time (µs)	42.4
Line broadening (Hz)	10
Sample temperature (K)	300

2.2.1. Sodium chloride stock solutions

Stock solutions were prepared by dissolving sodium chloride in deuterium oxide (99.9 atom % D), both supplied by Sigma Aldrich.

FID artefact

Data were acquired on the 10 mg/mL chloride linearity sample as prepared below. ^{35}Cl data were acquired with a pulse width of 14 µs at 1, 5, 10, 20, 40, 60, 80, 100 W.

T_1, T_2 and temperature

Data were acquired on the 10 mg/mL chloride linearity sample as prepared below. T_1 ^{35}Cl data were acquired using an inversion recovery pulse sequence, with recovery delays of 0, 10, 20, 30, 40, 50, 60, 80, 100, 150 ms for experiments at 273, 283, 293, 303 K; 0, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200 ms at 313 K; 0, 20, 30, 40, 50, 65, 80, 100, 150, 200, 300 ms at 323, 333 K; 0, 20, 50, 65, 80, 100, 125, 150, 200, 300 ms at 343, 353 K.

T_2 ^{35}Cl data were acquired using a CPMG sequence, with an echo delay of 2 ms and loop counts of 2, 4, 6, 8, 10, 12, 14 at 273, 283 K; 2, 4, 6, 8, 10, 12, 14, 16, 18 at 293 K; 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 at 303, 313 K; 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 at 323, 333 K; 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 at 343, 353 K.

Receiver gain

Data were acquired on the 10 mg/mL chloride linearity sample as prepared below. ^{35}Cl data were acquired with receiver gains of 0.1, 1, 2, 4, 8, 16, 32, 64, 128, 203.

Linearity

Solutions of sodium chloride in D_2O were prepared at chloride concentrations of 0.01, 0.02, 0.06, 0.1, 0.2, 0.4, 0.6, 1.0, 1.5, 2.0, 2.5, 5.0 and 10.0 mg/mL. For each concentration one sample was prepared and dispensed into a 5 mm NMR tube, a single measurement taken and all data were acquired under identical conditions.

It will be assumed that this method will be developed such that it can determine the % w/w chloride content of a solid sample prepared in solution with a nominal concentration of 10 mg/mL. In this context, the solutions of sodium chloride prepared for linearity are therefore equivalent to samples that would produce final processed results from this method of 0.1, 0.2, 0.6, 1, 2, 4, 6, 10, 15, 20, 25, 50 and 100% w/w chloride, respectively. Although in practice it would not be possible for a sample to contain 100% w/w chloride, this concentration was run for purposes of investigating the upper linear range of response for chloride on this spectrometer.

Repeatability

Data were acquired on the 0.2 mg/mL chloride linearity sample as prepared above using the parameters in Table 2.2. Data were acquired over the following time periods in the following sequence:

1) Once per hour for 24 hours
2) Once per day for 7 days
3) Once per hour for 24 hours
4) Periodically over 1 month, typically in blocks of 12 experiments or more per 24 hours

Effects of concentration on tuning and matching

The 0.01 mg/mL chloride linearity sample as prepared above was placed in the magnet and automatically tuned and matched. The sample was then removed from the magnet and the 0.02, 0.1, 1, 10 mg/mL samples were placed in the magnet in turn without further tuning and matching and the appearance of the wobb curve recorded. This was repeated with no sample in the magnet.

Effects of concentration on pulse width

Additional samples to the linearity set were prepared using the same process. The ^{35}Cl p_{90} was determined manually by observation of the null point with a p_{360} pulse using samples at 1, 1.5, 2, 2.5, 5, 10, 20 mg/mL chloride.

Interaction with other species and recovery

Samples of different compounds were prepared such that each contained a fixed chloride content. This was achieved by dissolving a weighed amount of each material in a measured volume of a pre-prepared stock solution of sodium chloride.

Approximately 10 mg of each of the following compounds was dissolved in 1 mL of a solution of 0.2 mg/mL chloride in acetonitrile-d_3/D_2O 50/50 v/v: sodium sulphate, sodium bromide, paracetamol, glucose, succinic acid, maleic acid.

Approximately 10 mg of each of the following compounds were dissolved in 1 mL of a solution of 0.2 mg/mL chloride in D_2O: acesulfame K, quinine, succinimide, paracetamol, benzoic acid, butylatedhydroxytoluene (BHT), ascorbic acid, vanillin, 2-

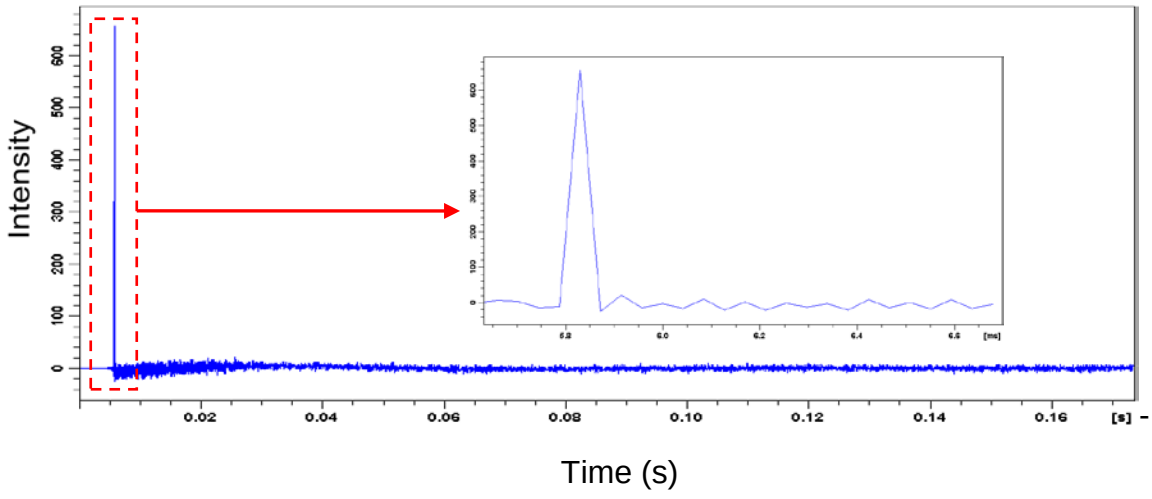
hydroxy-5-methylbenzaldehyde, benzophenone, 3-methyl-4-nitrobenzoic acid, benzenesulphonic acid, glucose, phenylboronic acid, glycine, sodium dodecyl sulphate (SDS), potassium carbonate, di-sodium hydrogen phosphate, sodium acetate.

Sodium chloride was prepared to a nominal concentration of 0.1 mg/mL chloride in each of the following diluents: methanol-d_4/D_2O 50/50 v/v, acetonitrile-d_3/D_2O 50/50 v/v

2.3. Results and discussion

2.3.1. FID artefact

Initial ^{35}Cl data acquired on a 10 mg/mL chloride solution showed the presence of a large artefact occupying the first few data points of the FID as shown in Figure 2.1. Several problems may arise from this phenomenon, but the main issue for the purposes of acquiring data for this study is one of dynamic range. The receiver gain (RG) should be set such that the highest point in the FID is not clipped. If the FID contains one or more artefact points of greater intensity than the analyte signal, then the RG will be set artificially high and the actual data from the chloride will not fill the dynamic range optimally resulting in reduced SNR in the FT spectrum.



Time (s)

Figure 2.1 – Artefact in ^{35}Cl NMR FID

This artefact could arise from a number of sources; including chloride background resonances from the probe, breakthrough of RF frequencies from one part of the signal path within the console to another, or probe ring-down (see Chapter 1). Further investigation showed that the artefact is present, even when there is no sample present in the probe. It does, however, require the whole signal path from console to probe to preamps to be connected, and the intensity and phase of the signal is not affected by pulse width, but it is affected by pulse power (see Figure 2.2) and also tuning and matching.

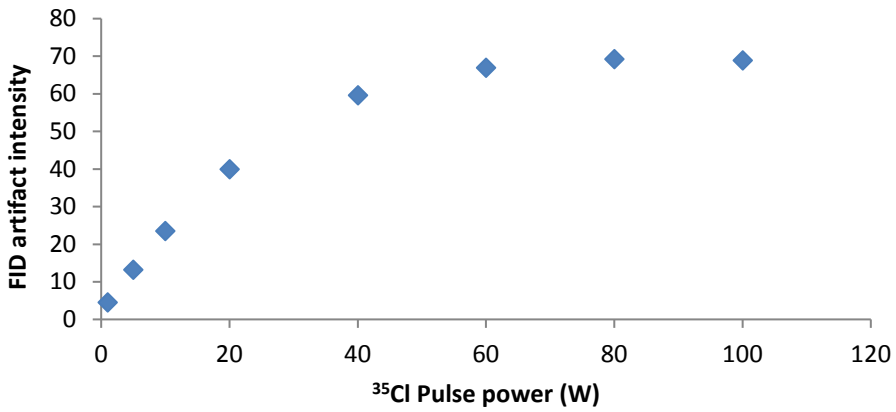


Figure 2.2 – Plot of artefact intensity with pulse power

Since the whole signal path is required to produce the artefact it is not due to RF breakthrough within the console, and the lack of correlation between pulse width and intensity suggests it is not from a contaminant in the probe.

The most likely explanation is probe ring-down. This is the presence of residual energy from the RF pulse still resonating in the tuned circuit of the probe coil during acquisition, which is then picked up by the receiver along with the start of the FID. As it is due to residual energy left after the RF pulse is turned off, the overall pulse width won't have a significant effect, but the power of the pulse will.

One solution is to increase the delay between the pulse turning off and the receiver acquiring data – known as the dead time (DE). Increasing DE from 6.4 to 17 µs

results in almost total removal of the artefact from the FID (see Figure 2.3), and no significant change in the SNR of the chloride resonance in the FT spectrum.

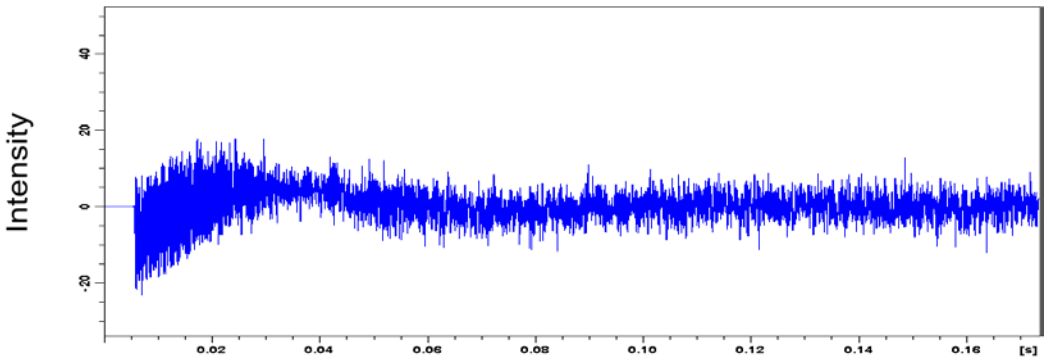


Figure 2.3 – FID arising from the removal of the ring-down artefact following optimisation of the dead time (DE)

A DE value of 17 µs was therefore selected. This value may need to be revisited for samples of lower concentration, where the artefact may still be of greater intensity than the chloride signal.

2.3.2. T_1, T_2 and temperature

Inversion recovery and CPMG experiments at 300 K yielded a T_1 value of 32.3 ms and a T_2 value of 31.4 ms, respectively. As T_1 determines the minimum ideal relaxation delay (> 5 × T_1 for > 99% relaxation), the effect of temperature on relaxation was also investigated.

Proportional, linear relationships between temperature (K) and T_1 (ms) and T_2 (ms) were observed with linear regression equations of y = 0.6409x – 159.9 and y = 0.5527x – 134.46, respectively, and R^2 values of 0.9873 and 0.9932, respectively.

This is expected for rapidly rotating molecules in the extreme line narrowing regime [12]. However, as the sample will be temperature controlled, it is not expected that temperature will be a significant issue for accurate quantitation.

The apparent T_2 as determined by the half-height line width (HHLW, $\Delta\nu_{1/2}$) of the peak is referred to as T_2^* and is calculated by equation (2.1)[13, 14].

$$T_2^* = \frac{1}{\pi\Delta\nu_{1/2}} \quad \textbf{(2.1)}$$

which is related to T_2 by the relationship in equation (2.2) [13].

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{2\Delta B_0}} \quad \textbf{(2.2)}$$

here, $T_{2\Delta B_0}$ is the spin-spin relaxation due to inhomogeneity of the static magnetic field. Since the observed HHLW of the chloride peak is *ca*. 10 Hz, then T_2^* is approximately 32 ms and since $T_2^* \approx T_2$, $T_{2\Delta B_0} \approx 0$. That is, the observed T_2^* has a dominant contribution from the quadrupolar relaxation pathway.

The above equations also demonstrate that an increase in temperature will give rise to a decrease in the observed HHLW as shown in equation (2.3).

$$\Delta\nu_{1/2} \propto \frac{1}{T} \quad \textbf{(2.3)}$$

This was experimentally observed in the data.

Data from Hertz and Weingärtner [15] showing calculated T_1 values for infinite dilutions of chloride in different solvents gave a value of 40.0 ms for water. This value is comparable to the value for D_2O determined here. Hertz and Weingärtner also suggest in their work that water will have the longest T_1 of all solvents, and so the relaxation parameters as determined and presented in this thesis are still applicable to solutions in other solvents.

2.3.3. Cycle time optimisation

The pulse-and-acquire sequence is one of the most basic in NMR and the key components are shown in Figure 2.4.

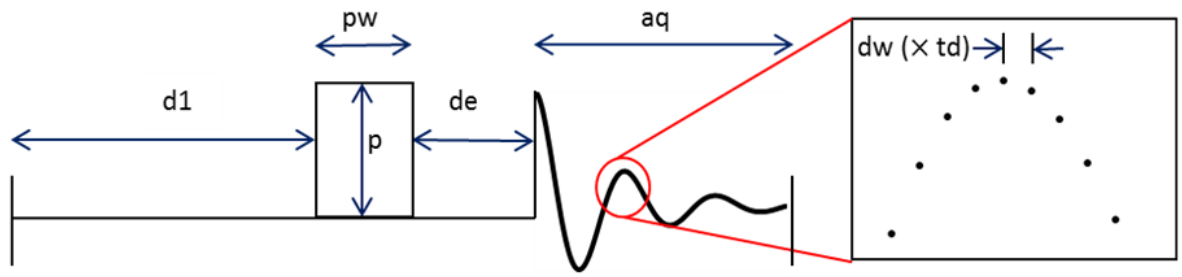


Figure 2.4 – Details of pulse and acquire sequence

To fully optimise SNR per unit time, each of the components contributing to the overall cycle time was considered, as even a small time saving could have a significant effect over many thousands of scans.

The dead time has already been discussed in 2.3.1 and was set at 17 µs to minimise the ring-down artefact in the FID.

Sufficient data points must be acquired to prevent truncation of the FID. A truncated FID approximates to a square wave in the time domain, which Fourier transforms to a sinc function, resulting in baseline "wiggles" around the base of the resonance in the frequency domain. As T_2 is very short for chloride in D_2O, it was possible to acquire only 2048 points with dw = 42.4 µs, for which aq = 0.174 s, without observing any truncation of the FID.

For nuclei such as ^{1}H or ^{13}C, where T_1 in the solution state is usually in the order of 1 - 10 s, the relaxation delay is typically set at around $5 \times T_1$ if quantitative data are required. In the example of these solutions of chloride in D_2O, $5 \times T_1$ is less than the acquisition time, aq, *i.e.* the system is almost completely relaxed by the time the acquisition has ended, so d1 can be set with an arbitrarily short delay of 1 ms.

The 90° pulse width as measured from the 360° null point is 13.4 µs for a pulse power of 100 W. Even with the short values of d1 and aq used, the duty cycle (*i.e.* proportion of time that RF power is applied to the probe) for the sequence is < 0.01% and so should not cause any damage to either probe or amplifier.

The parameters as determined above should give an overall cycle time of approximately 175 ms. However, there is an additional delay of 30 ms before the relaxation delay to allow tasks within the spectrometer console to take place, so the cycle time is closer to 205 ms and 6144 scans can be acquired in just over 21 minutes of experiment time.

2.3.4. Receiver gain

Within the NMR spectrometer console itself, acquisition of repeatable and quantitative data is dependent upon the performance of the transmitter and the receiver components of the signal path (the impact of the transmitter is discussed in 2.3.9).

The linearity of the receiver was tested by measuring the absolute integral of the chloride resonance over the full range of receiver gain values permitted by the spectrometer. Linear regression yielded a correlation coefficient (R^2) of 0.9987 and hence was shown to be acceptable for this method. These data also show that the absolute integrals of the chloride resonance can be directly compared between samples if the area of the resonance in each spectrum is divided by the RG.

It was also noted that although the 10 mg/mL chloride sample is not of high enough concentration to cause clipping at the highest RG set here, there is a rapid loss of SNR observed at RG values below 32 (Figure 2.5). This proves the importance of removing artefacts from the FID (as discussed in 2.3.1) that may lead to an artificially low RG resulting in a significant decrease in the SNR.

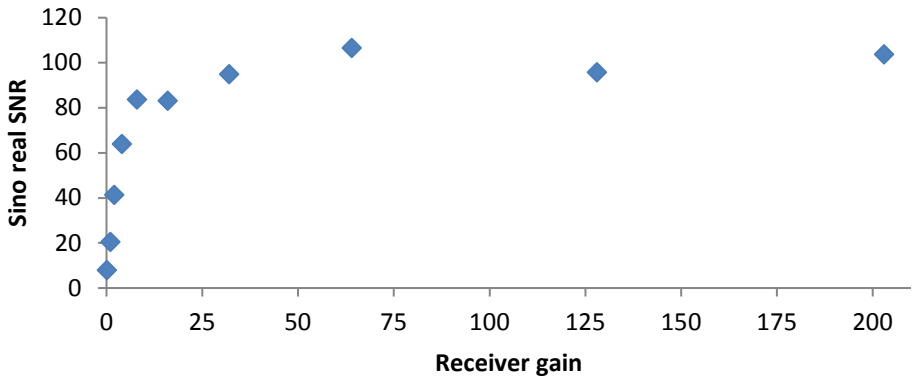


Figure 2.5 – Relationship between receiver gain and SNR

2.3.5. Line broadening

After the point at which all resonances have decayed in the FID, all subsequent data collected by the receiver is noise. This will add to the noise in the FT spectrum without contributing signal and therefore decrease the SNR of the resonances of interest. One way to alleviate this problem is to multiply each data point in the FID by an exponential (or apodisation) function, reducing the contribution of the noise at the end of the FID compared to the more intense signals at the start of the FID.

Manipulating the data in this way will increase the apparent decay of the resonances in the FID and hence broaden peaks in the FT spectrum according to equation (2.1). In doing so the peak height also decreases, and so there is an optimum degree of apodisation for a given spectrum, which is the point at which the function doubles the HHLW in the FT spectrum [16]. For example, when LB of approximately 10 Hz is applied to the 10 mg/mL chloride solution ^{35}Cl NMR spectrum, the HHLW is doubled from 10 to 20 Hz.

In a later section (2.3.13), spectra of samples dissolved in acetonitrile-d_3/D_2O 50/50 are introduced, as this diluent offers better solubility for typical small organic molecules. In this instance, the chloride peak HHLW is now of the order of *ca.* 70 Hz. Whilst it would appear ideal to optimise the LB value specific to the HHLW of each spectrum, it must be remembered that the aim of this investigation is to directly compare the area of peaks in different spectra to allow quantitation, and this requires

the integration process to be consistent. Although applying an exponential apodisation function with different LB values would not change the peak integrals in the FT spectrum, it would have an effect on peak width and the appearance of noise. This, in turn, may change the degree of accuracy to which the software can integrate the peaks. For reasons of consistency, LB values were fixed at 10 Hz for all subsequent data acquired by this method.

2.3.6. Specificity

ICH guideline Q2 defines specificity as:

"...the ability to assess unequivocally the analyte in the presence of components which may be expected to be present. Typically, these might include impurities, degradants, matrix, etc."

This aspect has already been discussed in 2.1.2.1. and it has been concluded that only the Cl^- anion will be observed in these spectra with no interference from species other than the ClO_4^- anion. Therefore, no further experimental work is required.

2.3.7. Linearity

Linearity is defined in ICH guideline Q2 as follows:

"The linearity of an analytical method is its ability to obtain results which are directly proportional to the concentration of the analyte in the sample."

A plot of absolute area versus chloride content for the linearity samples is shown in Figure 2.6. Note that the RG was set as part of the automation process using the RGA function, but since the values obtained for all samples were equal, then absolute area can be taken without further compensation.

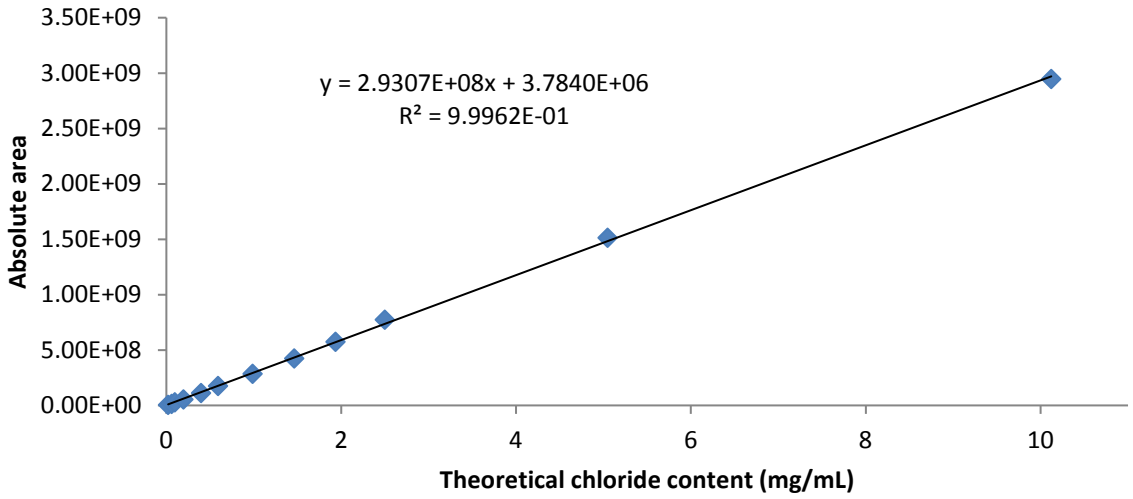


Figure 2.6 – Linearity of Chloride content *vs.* absolute peak area over full range

A good degree of linearity is observed across the full range of content with a correlation coefficient of > 0.999, which is acceptable for use in a quantitative method. This demonstrates that for sample data acquired over a period of less than an hour, with the sample solvent and solution properties here, the response is linear. It also can be assumed that any difference in RG values could be compensated for with minimum additional error added.

These data were further interrogated by calculating residuals (predicted chloride content from regression subtracted from actual chloride content) and plotting these values against the theoretical chloride values (see Figure 2.7).

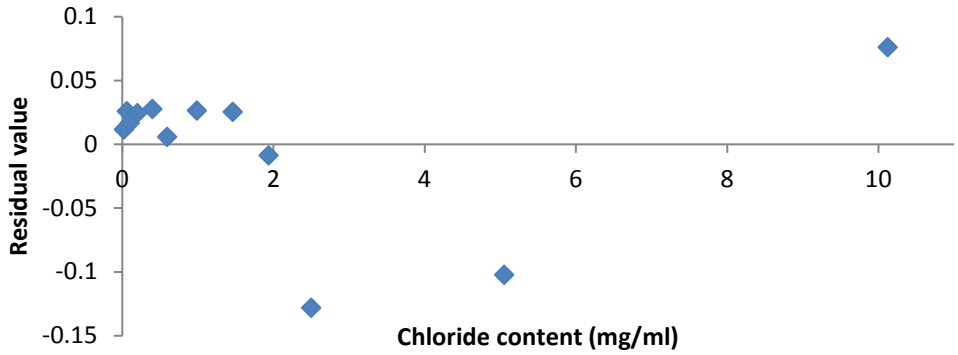


Figure 2.7 – Plot of residuals for chloride content

The values do not appear randomly distributed around zero. Fitting higher order polynomials (rather than a linear equation) to the linearity data gives a higher correlation coefficient, but this does not conclusively confirm that these relationships are any more valid. It should be noted that the spectra for the linearity data were processed by manual phasing, and by manually setting the integration region for each spectrum so as to minimise the amount of noise included. For this reason, and given that each point on the calibration curve is only prepared in singlicate, it is equally as likely that the unequally distributed deviations from the linear curve could be explained by experimental error in integration and/or manual phasing.

In conclusion, the degree of linearity shown here is considered acceptable for the purposes of this method. The validity of this decision will be tested by analysing samples spiked with known quantities of chloride as part of the interaction and recovery assessment (see 2.3.13).

As a comparison, Lim and Lee obtained correlation coefficients of > 0.999 by use of internal standards placed coaxially within the NMR tube over the range 20 to 5000 µg/mL (or 0.02 to 5 mg/mL) [9] .

2.3.8. Detection and quantitation limits

The ICH guideline Q2 defines the detection limit (DL) and quantitation limit (QL) as:

Detection limit – "*The detection limit of an individual analytical procedure is the lowest amount of analyte in a sample which can be detected but not necessarily quantitated as an exact value."*

Quantitation limit – "*The quantitation limit of an individual analytical procedure is the lowest amount of analyte in a sample which can be quantitatively determined with suitable precision and accuracy."*

These limits are critical in determining how data are reported from an analytical method. If a result falls below the DL, then there is no confidence that the result is

distinguishable from noise and is reported as “not detected” (ND). If a result is between the DL and QL, then there is confidence that the analyte signal has been observed, but the accuracy of the figure cannot be guaranteed and so is reported as “<QL” and the QL quoted. If the result is above QL then a numerical figure can be reported with confidence.

The minimum SNR acceptable for DL and QL are usually quoted as 3:1 and 10:1 respectively. However, these are for situations where noise is calculated as “peak-to-peak”, *i.e*. the distance between the highest and lowest points of noise over a given range. Topspin software, as used to process these data, calculates noise based on a root-mean-squared (RMS) calculation. It is therefore necessary to derive the minimum figures for DL and QL using the “sino real” command, which are 3.75:1 and 12.5:1 respectively (see Appendix 1).

The 0.01 mg/mL chloride solution has a sino real SNR figure of 3.92:1. This is greater than the required 3.75:1 and so is acceptable for a DL. The 0.06 mg/mL chloride solution has a sino real SNR figure of 16.04:1. This is above the limit of 12.5:1, and so is acceptable for a QL. It is possible to estimate the mg/mL chloride content that a peak at 12.5:1 would represent and thus estimate the theoretical QL as shown in equation (2.4).

$$Estimated\ QL = \frac{12.5}{16.04} \times 0.06\ mg/mL\ = 0.0468\ mg/ml\ chloride \quad \textbf{(2.4)}$$

Thus, if a sample of nominal concentration 10 mg/mL was analysed using the current method, then the DL and QL for chloride content would be 0.1 and 0.5% w/w respectively. This will hold true for any solutions with the same HHLW. As the actual samples are more likely to be dissolved in acetonitrile-d_3/D_2O 50/50 (see 2.3.13), the SNR and DL/QL will be calculated for each spectrum and used as limits to report data against in the processing programme (see 2.3.14)

Work carried out by Lim and Lee [9] does not include the SNR of the lowest concentration samples and so a direct comparison cannot be made.

2.3.9. Repeatability

Repeatability is defined in ICH guideline Q2 as follows:

“Repeatability expresses the precision under the same operating conditions over a short interval of time.”

Although extending this to analyses over different days is strictly classed as an aspect of intermediate precision, this would also include analyses on different instruments by different analysts and so the study here will be considered as repeatability only.

As discussed in 2.3.4, the quality and reliability of quantitative data are directly dependent upon the performance of the transmitter. Within the NMR spectrometer, the main components to consider here are the amplifiers. As with receivers, the most modern amplifiers are described as linear, *i.e.* changes made in the software to the power of pulses are proportional to the actual power output of the amplifier. Of equal importance is the stability of the amplifiers over time. NMR consoles are temperature controlled internally with fans, so that any change in the laboratory conditions should not affect the components in the amplifiers and hence change the output.

Replicates of the 0.2 mg/mL chloride sample over the initial 24 hour time period return an average result of 0.157 ± 0.006 mg/mL chloride, where the mg/mL chloride was calculated from the linearity regression equation from Figure 2.6 and the error is quoted as ± 1 standard deviation (SD). Whilst this level of variability would be considered acceptable for a quantitative method, Figure 2.8 shows some possible evidence of a decreasing trend in response over time.

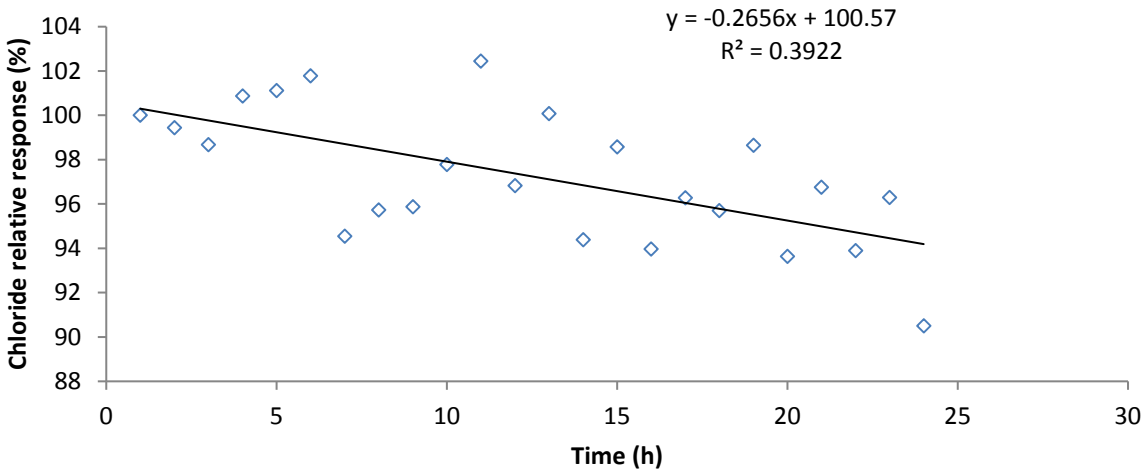


Figure 2.8 – Change in chloride response over 24 h

Figure 2.9 shows data acquired over a total of 7 days (including the initial 24 hour period).

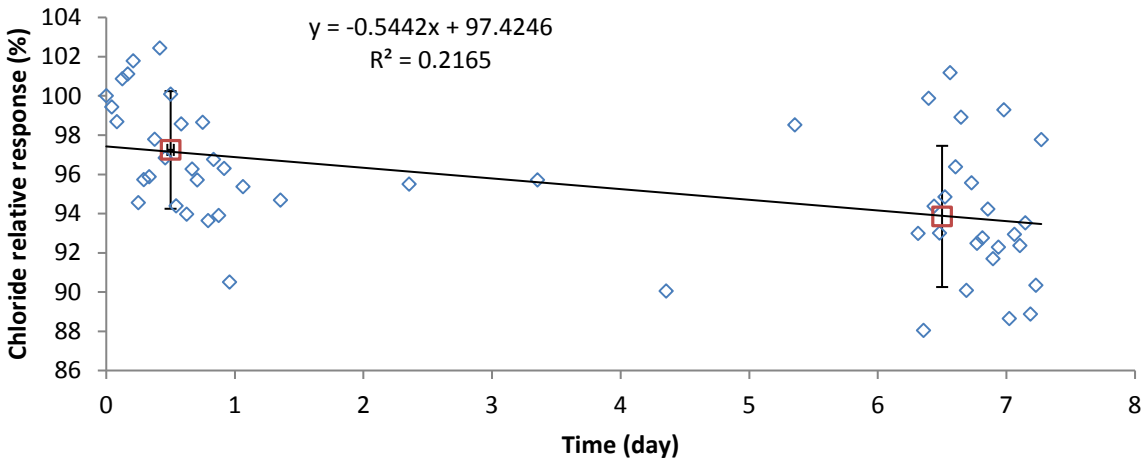


Figure 2.9 – Change in chloride response over 1 week

The red boxes show the average response for the first and second 24 hour periods and the error bars show ± 1 × SD. The error bars for these overlap and so although there is a drop in the average response over 1 week, the change is not significant compared to the degree of variability of response over 1 day.

Figure 2.10 shows data acquired over a period of several months, with the results averaged for each 24 hour period (with ≥ 12 experiments per 24 hour period). There appears to be less of a trend apparent over this duration. There is a variation in the average values for a given day, but there is no overall trend over time. Although it is more typical to use ± 2 × SD when comparing data (*i.e.* 95% confidence that the true value lies within this region) there is already overlap between all data at ± 1 × SD (*i.e.* within the region of *ca.* 68% confidence). Furthermore, the standard deviation of the averaged values across the days is 2.03, whereas the standard deviations for the values for each day range from 2.0 to 3.8. This shows that there is more intra-day variation than inter-day variation in the data.

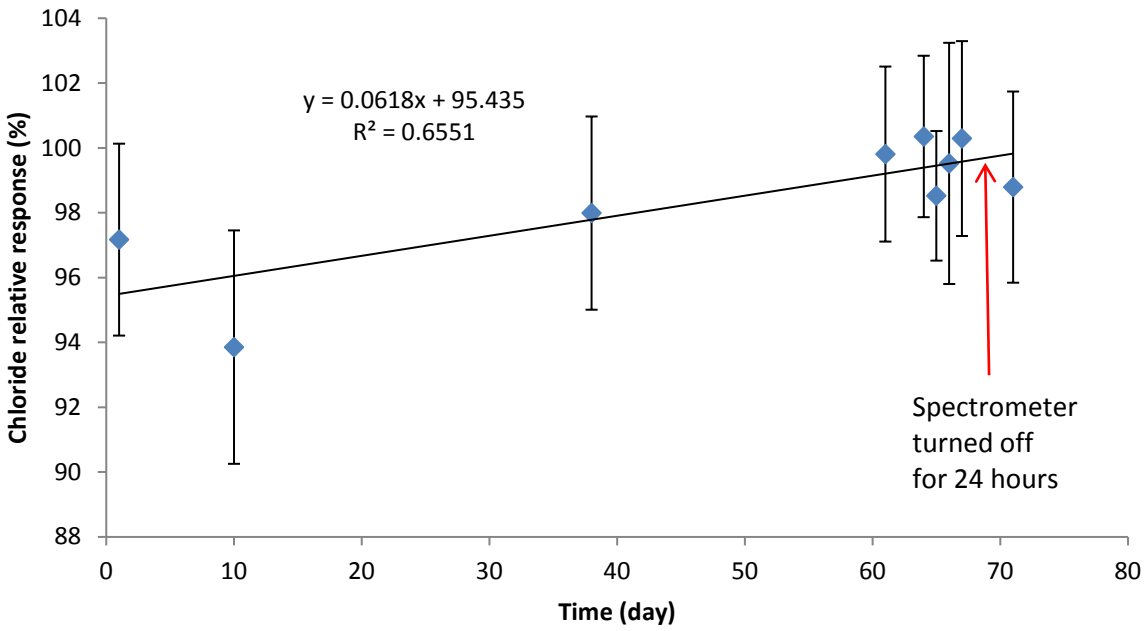


Figure 2.10 – Change in chloride response over 10 weeks, with error bars of 1 × standard deviation

As shown in Figure 2.10, the spectrometer was turned off for a period of approximately 24 hours and no significant differences are observed in the data acquired before and after this time.

To further exemplify the effect of the level of variation upon an actual measurement, the day with the greatest level of variation in results would yield a chloride content of

0.150 ± 0.008 mg/mL (quoting ± 1 × SD) or 0.150 ± 0.150 mg/mL (quoting ± 2 × SD), equating to 5.0% RSD. This would be deemed an acceptable level of variation for a typical analytical method for most purposes. For comparison purposes, the average of all data over this period yielded a chloride content of 0.158 mg/mL.

It can be concluded from these data that there is no obvious trend in spectrometer response over a period of 10 weeks and that turning the amplifier and receiver off for a period of time also has no significant effect upon response.

As all data discussed here were acquired with the same NMR tube and tuning and matching were not adjusted between consecutive samples on the same day, then the sources of these variations must lie in either the instrument response or the data processing. Further investigations into the processing method may be required to determine if these errors can be minimised.

2.3.10. Coil loading and reciprocity

The discussions so far have centered either on the optimisation of parameters directly under the analyst's control or assessing the general capabilities of the method. Another factor to consider is the effect that a given sample will have on the performance of the spectrometer, and specifically the efficiency of energy transfer to and from the probe coil.

An NMR probe contains one or more coils, each of which forms part of a tuned circuit. Figure 2.11 shows an example of a basic parallel tuned circuit consisting of a single inductor coil and two variable capacitors.

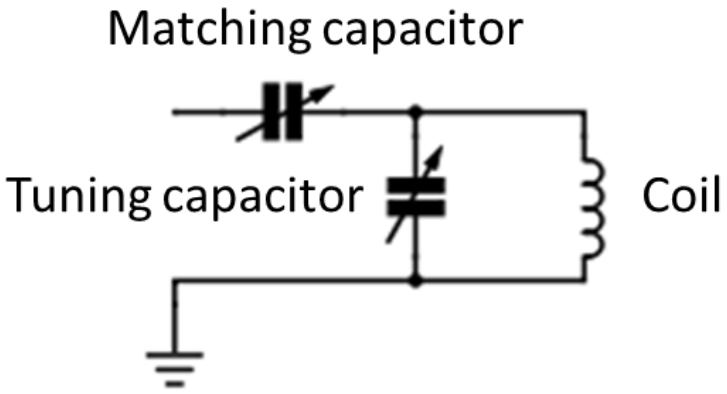


Figure 2.11 – A simplified NMR probe circuit

In practice, the coil would be large enough to fit around the NMR tube, and may be one of several geometries (*e.g.* saddle, solenoid, *etc*.). Before acquisition, a sample would be placed in the coil and the circuit tuned to the Larmor frequency of the nuclei of interest, ω_0, in accordance with (2.5) [17].

$$\omega_0 = \frac{1}{\sqrt{LC}} \tag{2.5}$$

where L is the inductance of the coil and C the capacitance of the overall circuit.

At the same time, the impedance (that is, the overall opposition to current flow taking into account resistive, capacitive and inductive components) is matched to that of the rest of the signal pathway, typically 50 Ω. The condition for this is described by (2.6).

$$\frac{Z - R_0}{Z + R_0} \rightarrow 0 \tag{2.6}$$

where Z is the impedance of the circuit and R_0 the reference impedance (here 50 Ω). In practice, this is achieved by adjusting the tuning and matching capacitors, the visualisation of which will be discussed later.

An RF pulse, with defined pulsewidth, phase, amplitude and distribution, would then be introduced into the coil at the Larmor frequency, thus irradiating the sample. Once

the pulse is turned off, the weak RF signal coming back from the resonating nuclei in the sample is detected by the same coil and is transferred, *via* the pre-amplifiers, to the receiver.

The magnitude of the resulting RF signal output is, besides other factors, dependent upon the efficiency of the energy transfer between coil and sample and back again. This is described in RF electrical engineering terms as the 'Q', or 'quality' factor which, for the example of the parallel LC circuit [17], is defined in equation (2.7).

$$Q = \frac{\omega L}{r} \tag{2.7}$$

Although not fully discussed here, Q [18] is related to SNR of a signal by the relationship shown in (2.8).

$$\frac{s}{n} \propto \sqrt{Q} \tag{2.8}$$

The conductivity and dielectric constant of the sample will change the Q factor and hence, in the specific example of this method, it can be seen that the absolute area of the chloride resonance in the FT NMR spectrum will change depending on the properties of the sample.

Fortunately, it is not necessary to explicitly determine the Q factor for each sample, as the principle of reciprocity can be applied. Work by Hoult and Richards [18, 19] (later exemplified by Burton *et al.* [18, 19]) proves that in the case where excitation and detection take place in the same coil, signal intensity has a constant proportionality to analyte concentration for a constant flip angle.

Any effect that changing the sample has on the Q of the coil will have an equal effect on the transmission of the irradiating pulse into the sample and the reception of the resonating signal back from the sample. This effect can be measured and

compensated for by determining the p_{90} for the sample and relating it to the p_{90} for a standard.

Changes in NaCl concentration, different solvents and the presence of other analytes can all potentially change Q. The change was investigated indirectly, by looking at the difference in the tuning/matching “wobb” curve and directly by measuring the change in p_{90} for different samples.

2.3.11. Effect of concentration on tuning and matching

The parallel circuit shown in Figure 2.11 indicates only the most basic components of a tuned circuit. In practice, probes have much more complex circuits to allow consistent performance over a wide range of nuclei and to minimise interactions where several tuned circuits are coupled to the same coil.

Tuning and matching is therefore a process where these two components are adjusted for the nuclei and the sample until an optimum is reached. This is commonly performed by sweeping over a narrow frequency range of a few MHz around the frequency of interest and measuring the response compared to an internal 50 Ω reference load. This graphical display is known as a “wobb” curve in Bruker terminology, as shown in Figure 2.12.

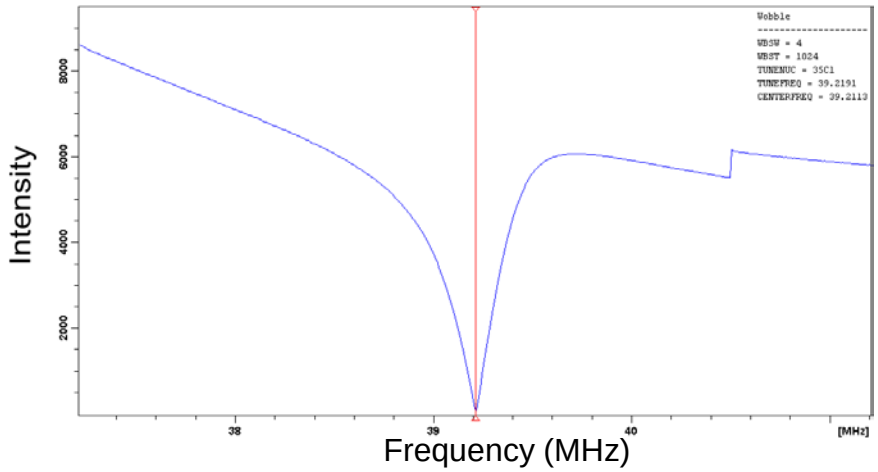


Figure 2.12 – Wobb curve used to optimise probe tuning and matching for a sample of 0.01 mg/mL chloride

The horizontal position of the dip in the curve is adjusted with the tuning capacitor and the vertical position with the matching capacitor, although the two components are interactive and so an iterative process is required. The probe is optimally tuned and matched when the dip reaches a minimum, at a frequency equal to that of the transmitter frequency for the nuclei of interest [14]. For older probes, and currently all solid state NMR probes, this task is performed manually with rotating adjustment rods that protrude from the bottom of the probe. On newer probes, such as the BBFO ATMA probe used here, these rods are connected to electrical motors and so the process can be carried out automatically by software alone.

The Wobb curve was recorded for samples at higher concentration, but with the same tuning and matching settings. For example, the curve for 10 mg/mL chloride is shown in Figure 2.13.

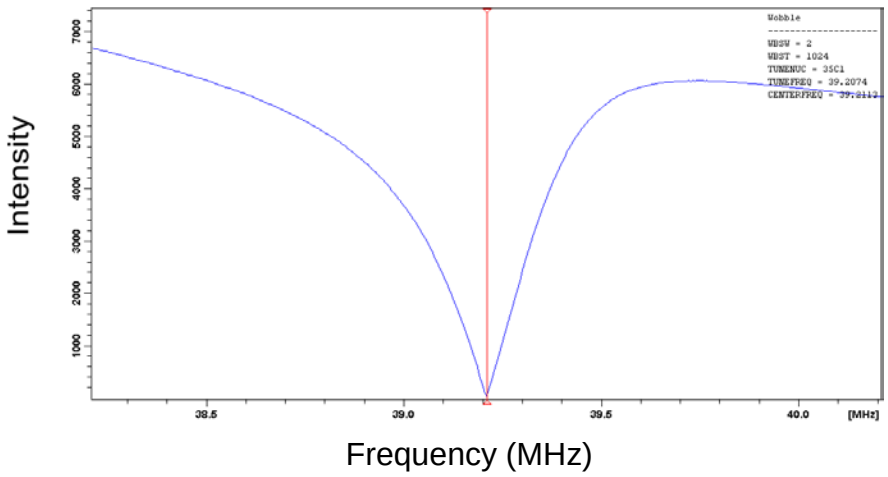


Figure 2.13 – Wobb curve of sample at 10 mg/mL Chloride

No significant differences are observed in the appearance of the curves shown above, so it is assumed that the effect on tuning and matching, and hence Q, is negligible for the concentration range shown here. For comparison**,**
Figure 2.14 shows the wobb curve when no sample is placed in the magnet.

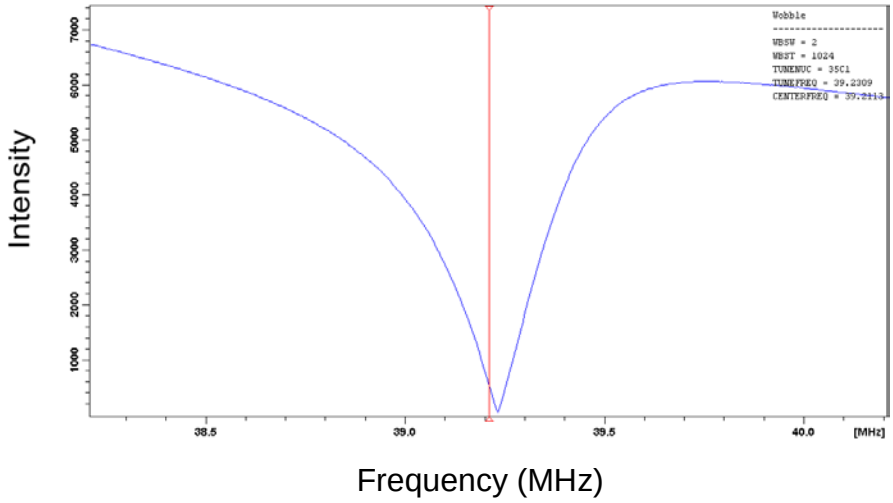


Figure 2.14 – Wobb curve with no sample in magnet

2.3.12. Effect of concentration on pulse width

As discussed above, the pulse width required for a 90° flip angle (p_{90}) determined for a given sample proves a direct measure of the correction factors required to compensate for changes in Q, given that certain criteria are met.

Table 2.3 shows observed ^{35}Cl pulse width required for a 360° flip angle (p_{360}) and calculated p_{90} and relative responses at increasing chloride concentration.

Table 2.3 – Effects of chloride concentration on ^{35}Cl pulse width

Chloride concentration (mg/mL)	P_{360}/μs	P_{90}/μs	% Relative response
1	55.10	13.7750	100.0000
1.5	55.40	13.8500	99.9963
2	55.45	13.8625	99.9950
2.5	55.55	13.8875	99.9918
5	55.65	13.9125	99.9877
10	55.75	13.9375	99.9828
20	55.90	13.9750	99.9740

A non-linear relationship is observed between pulse width and concentration (see Figure 2.15) consistent with expectation that higher concentrations will result in longer pulse widths, as the Q factor decreases. The overall change in p_{90} over this range is 0.2 μs. It can be readily shown that if the flip angle deviates from p_{90} by θ, then the intensity of the signal will decrease from the maximum value by a factor of cos θ [19]. From here, the relative response for the values of the p_{90} measured can be calculated as shown in Table 2.3.

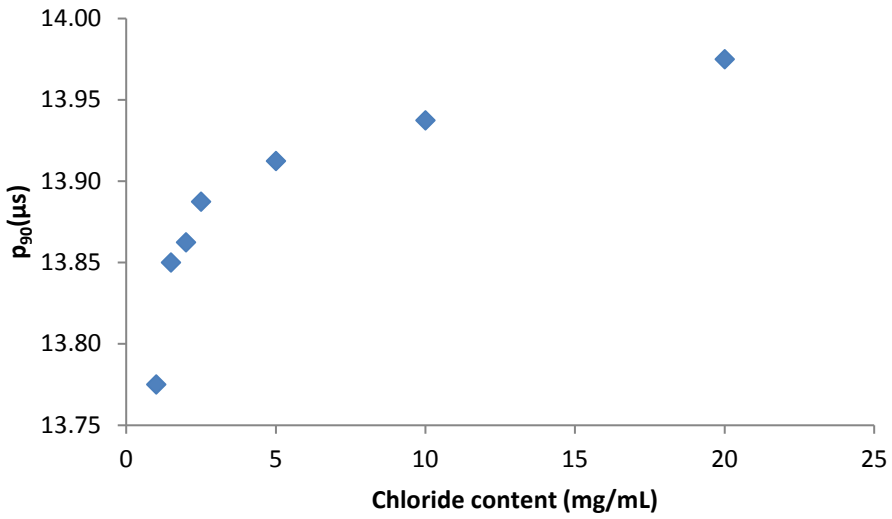


Figure 2.15 – Effects of chloride concentration on ^{35}Cl pulse width

The changes in signal intensity are minimal compared to the error observed from repeated analysis of the same sample (see 2.3.9), due to the cos θ relationship. For example, ± 1 µs deviation in a 14 µs pulse would equate to a decrease in the signal of 2.51% relative to the maximum. For a sample containing 0.158 mg/mL chloride (from the average value determined in 2.3.9) this would be a decrease of 0.004 mg/mL. This is less than the error of ± 0.008 mg/mL obtained from repeat analysis of the same sample using the current parameters in 2.2.

Due to low SNR, it was not possible to accurately measure the p_{90} for samples of concentration lower than 1 mg/mL chloride. However, no significant changes in the wobb curve were observed at these concentrations (see 2.3.11) and a plot of response *vs*. concentration shows a linear relationship down to a concentration of 0.01 mg/mL chloride.

In summary, these data show that the concentration of sodium chloride has no significant effect upon Q, and thus no compensation of the response is required over the concentration range discussed here.

2.3.13. Interaction with other species and recovery

The preceding work in this study has determined only the effects of NaCl in D_2O on the quantitation of chloride content. To determine if this method will be accurate for the quantification of chloride in actual samples, the effect of different species in the solution on the quantification must also be investigated.

Table 2.4 shows the theoretical chloride contents for a series of compounds dissolved in D_2O and spiked with a fixed amount of chloride.

Table 2.4 – Recovery values for analyte solutions in D_2O containing chloride

Analyte	Theoretical chloride content (mg/mL)	Experimental chloride content (mg/mL)	% Recovery
Sodium sulphate	0.185	0.179	96.8
Sodium bromide	0.192	0.196	102.2
Paracetamol	0.195	0.197	100.7
Glucose	0.185	0.169	91.3
Succinic acid	0.187	0.190	101.4
Maleic acid	0.185	0.192	104.0

ICH guidelines do not mandate specific limits for recovery, but GSK internal documents require recovery of 90 – 110%, criteria which have been met for the species determined here.

The number of species in Table 2.4 is small, but the data shows that any interactions from other anions such as SO_4^{2-} and Br^-, or from small organic molecules with aromatic, olefinic, phenolic, amide, hydroxyl and carboxylic acid functionalities, do not affect the method within the criteria specified. Only the sample containing glucose shows a difference between the actual and measured content of >± 0.007 mg/mL (as obtained from repeat analysis of the same sample with the current parameters in 2.2). It may be that this is indicative of an interaction affecting the chloride response.

Only a minor percentage of compounds likely to be encountered in the development of pharmaceutical molecules are soluble in D_2O. For analysis by HPLC, a mixture of equal volumes of acetonitrile and water, or methanol and water, are known to dissolve many compounds of differing functionality. Table 2.5 shows data for samples of sodium chloride in typical HPLC solvents.

Table 2.5 – Effect of diluent solvent on quantitation

Diluent	Ratio	Theoretical chloride content (mg/mL)	Experimental chloride content (mg/mL)	% Recovery
Methanol-d_4/D_2O	1:1	0.104	0.112	107.8
Acetonitrile-d_3/D_2O	1:1	0.106	0.112	106.0

The recovery values are within the specified acceptance criteria. Although the presence of methanol and acetonitrile has no significant effect on the signal response, there is a difference in the appearance of the spectra as shown in Figure 2.16.

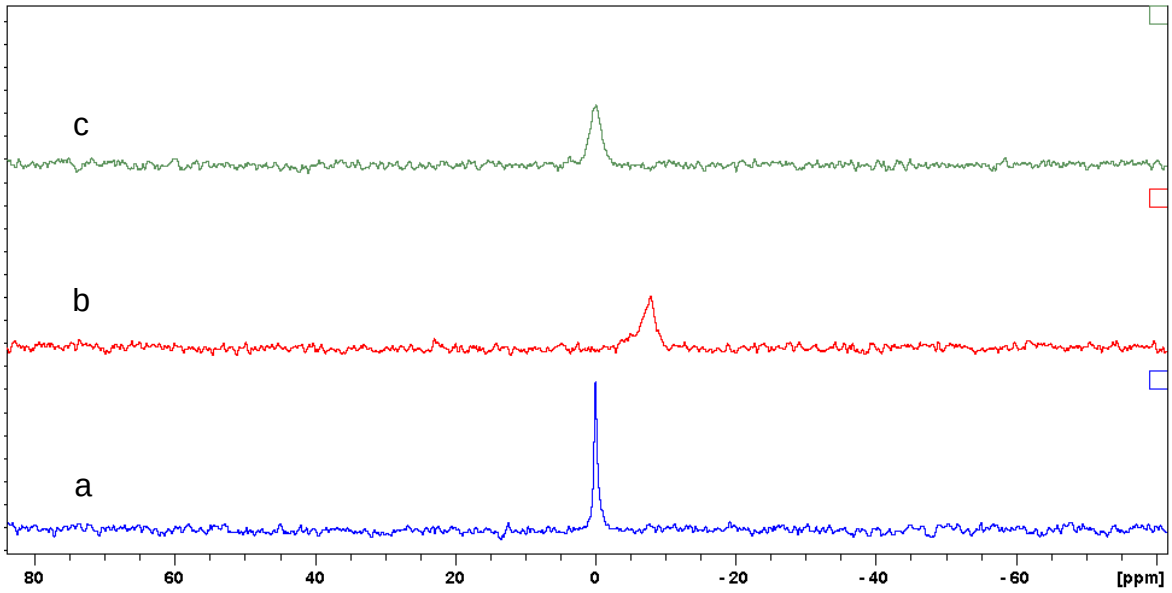


Figure 2.16 – ^{35}Cl spectra at 0.1 mg/mL chloride concentration with diluents a) D_2O, b) Methanol-d_4/D_2O, c) Acetonitrile-d_3/D_2O

The spectrum acquired in D_2O shows a typical profile with a single resonance at *ca.* 0 ppm, a HHLW of 20 Hz and SNR of 27:1. In contrast, the spectrum acquired in methanol-d_4/D_2O shows that the chloride peak has shifted to *ca*. $\delta\ ^{35}Cl$ = -7.7 ppm, with a HHLW of *ca.* 79 Hz and SNR of 10:1. Furthermore, this peak appears to contain more than a single resonance and is no longer a simple Lorentzian lineshape. The spectrum acquired in acetonitrile-d_3/D_2O has a shift of *ca.* $\delta\ ^{35}Cl$ = 3.8 ppm a HHLW of 73 Hz and SNR of 12:1.

These differences in appearance between the solvent systems are likely due to changes in the solvation sphere around the chloride ion. In D_2O solutions the species present is a 6-coodinate octahedral complex, $[Cl(D_2O)]^-$, whereas the presence of MeCN-d_3 is likely to disrupt this symmetry. This will lead to an increase in the electric field gradient at the chloride nucleus, resulting in broader lines. A change was made to the automation programme to adjust the processing depending upon the linewidth to compensate for different solutions (see 2.3.14). The linewidth is first assessed and then the integration region set based on this. This ensures that the integration region is wide enough to include the whole peak, but not so wide as to include excessive noise, which will introduce inaccuracies, especially at lower chloride concentrations.
Table 2.6 shows the recovery values of solutions of 20 different compounds spiked with chloride.

Table 2.6 – Recovery values for analyte solutions in acetonitrile-d_3/D_2O containing chloride

Analyte	Theoretical chloride content (mg/mL)	Experimental chloride content (mg/mL)	% Recovery
Acesulfame K	0.193	0.183	94.8
Quinine	0.193	0.187	97.0
Succinimide	0.180	0.178	99.0
Paracetamol	0.191	0.189	99.1
Benzoic acid	0.181	0.173	95.5
Succinic acid	0.183	0.180	98.4
BHT	0.184	0.165	89.7
Ascorbic acid	0.213	0.209	98.3
Vanillin	0.183	0.187	102.3
2-Hydroxy-5-methylbenzaldehyde	0.195	0.184	94.6
Benzophenone	0.188	0.182	96.8
3-Methyl-4-nitrobenzoic acid	0.187	0.189	101.1
Benzenesulphonic acid	0.194	0.192	98.8
Glucose	0.194	0.186	96.0
Phenylboronic acid	0.188	0.171	90.9
Glycine	0.190	0.181	95.5
Sodium dodecyl sulphate	0.184	0.178	96.7
Potassium carbonate	0.192	0.179	93.4
Di-Sodium hydrogen phosphate	0.190	0.183	96.5
Sodium acetate	0.190	0.185	97.3

Only one value falls just outside the range of 90 – 110% recovery, that of BHT. For the purposes of the method, it can be shown here that the presence of typically encountered organic moieties and inorganic species do not significantly affect the quantitation of chloride in this diluent.

In future work, accuracy could more rigorously assessed by the analysis of reference samples with certified levels of chloride. Accuracy is formally defined in ICH guideline Q2 as follows:

"The accuracy of an analytical procedure expresses the closeness of agreement between the value which is accepted either as a conventionally true value or an accepted reference value and the value found."

Hence analysis of samples of well-characterised references would be required.

2.3.14. Automation

The ultimate aim of this study was to develop a fully automated ^{35}Cl NMR quantitative chloride experiment for a walk-up NMR spectrometer. To achieve this, the method must not only be developed with respect to experimental conditions and parameters, but it must also be capable of data acquisition, processing and reporting with minimal user interaction.

The proceeding work has been carried out using Bruker Topspin 3.0 software under MS Windows 7. Automation is controlled by the Bruker IconNMR software that integrates with Topspin to provide a walk up front-end for the user. For most experiments, the user will input only the dataset name, diluent, experiment and title information such as batch and/or sample preparation details. For this experiment, the user also needs to input sample weight and solution volume. The calculation to produce the result will then need to include other information from acquisition parameters, such as receiver gain and number of scans, and also the regression parameters determined from the linearity experiments in 2.3.7. Finally, a result must be reported back in the correct format. A specific computer programme is therefore needed to handle the different user inputs, process the data and calculate and report the output.

In this instance, an automation (au) programme can be written as a list of commands designed to carry out a series of tasks in Topspin based on the programming language C. C [20] was developed by Dennis Richie at the Bell Labs between 1969 and 1973 as a general purpose, imperative language and is considered as a middle level programming language.

An au programme may contain a combination of standard C language statements and additional pre-defined macros for the purposes of handling datasets and parameters in Topspin. These are then translated into C statements by Topspin and

subsequently compiled into executable binary files using a C compiler. A summary flow diagram for running and processing this experiment is given in Figure 2.17.

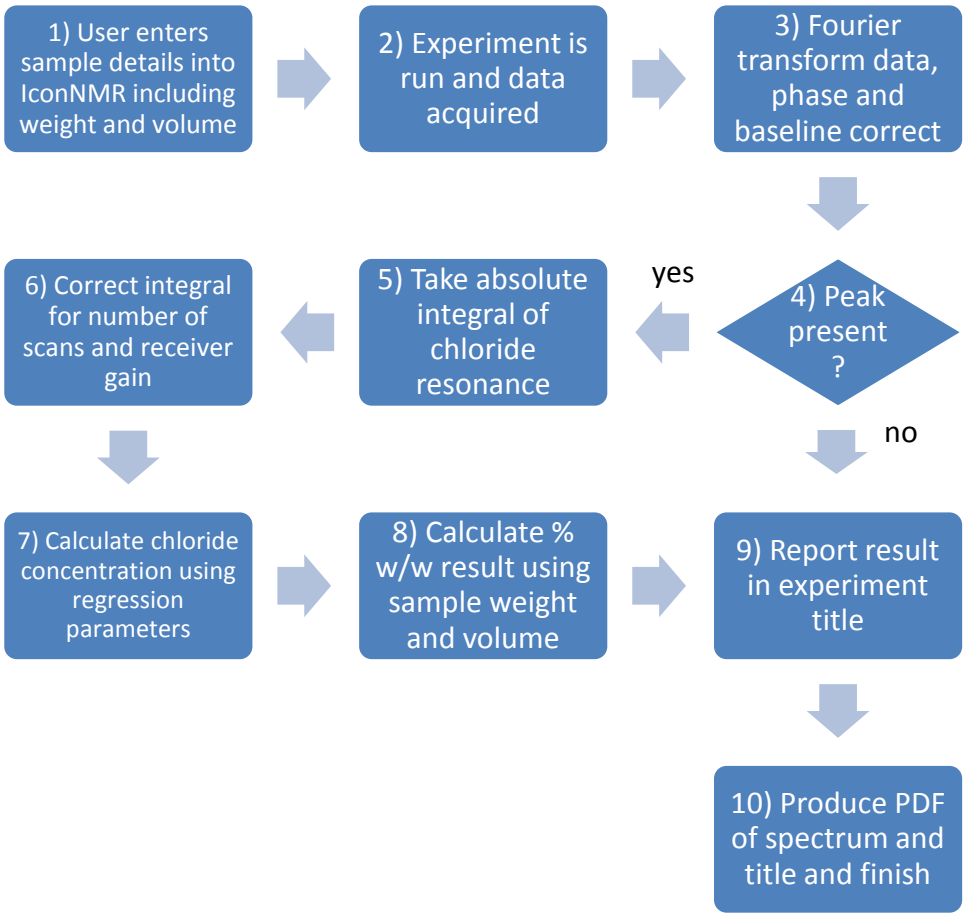


Figure 2.17 – Diagram showing work flow for ^{35}Cl quantitation experiment

The programme written to cover steps 3 to 9 of this work flow is called 35cl_quant (see eSuppl 1). The programme includes comments that explain each section of code in more detail.

Below is a summary of the tasks carried out by the code:

Step 1, there are a number of options that allow a user to input information into the walk-up software. In this case, the acquisition user parameters (USERA1 and USERA2) were utilised. These can be added to the list of variables that a walk-up

user can directly change prior to running a sample. When the user submits the sample, step 2 is initiated.

Step 2 is handled by IconNMR automation and step 3 is subsequently initiated.
Step 3 Fourier transforms the data and performs a series of phase corrections and a baseline correction that should produce an acceptable spectrum for most samples.

Step 4 checks if a peak was picked. If no peak is present, the programme jumps to Step 9. If a peak is present then the programme continues to Step 5.

Step 5 consists of a routine that takes the integer values of each data point within a region (see later) and adds them up to produce an absolute integral. This routine also takes into account scaling factors used by Topspin to store real data.

Step 6 takes the number of scans and receiver gain and scales the absolute integral relative to those parameters from the linearity experiments.

Step 7 takes the regression parameters (as stored in the parameter set as user processing parameters USERP1 and USERP2) and calculates the chloride concentration.

Step 8, the % w/w is then calculated from the sample weight and solution volume.

Step 9, the results are appended to the title file (taking into account the presence/absence of a chloride peak).

Step 10, the spectrum is plotted as a PDF file with the chloride results appearing in the title (see Appendix 2 for example PDF output).

The programme as shown enables the experiment to be run and the data processed entirely under automation. The only user intervention required is the sample preparation and in entering the weight and volumes into IconNMR. Two specific subsequent modifications to the programme are discussed below.

2.3.14.1. Determining optimal integration region

In practice, it was found that the peak width of the chloride resonance would vary considerably depending on the other species present. For example, solutions of NaCl in D_2O would give HHLW of *ca.* 20 Hz, whereas for some analytes dissolved in D_2O/MeCN 50/50 this would be up to 80 Hz. Setting an arbitrarily broad region for integration increases the variability of the results by increasing the noise region included in the integral for narrower resonances.

To resolve this issue, an additional function was written into the programme to determine the approximate linewidth of the chloride resonance (named "half_height_width" in the code in eSuppl 1). This function is called between steps 3 and 4 in Figure 2.17.

This function works by looking at intensity values of individual data points in the processed data file for the sample. Bruker Topspin stores real and imaginary FID data in 1r and 1i files, respectively. These are binary files containing a series of 32 bit signed integer values (or mantissa) for each data point, which each are related to their actual absolute value, abs, by the relationship in (2.9).

$$abs = mantissa \times 2^{NC\_Proc} \quad \textbf{(2.9)}$$

where NC_Proc is a single exponent stored as an acquisition parameter to be applied to all points in the file.

In the function half_height_width, the absolute intensity of the highest peak (apex) is first read in from the 1r file. Then, the next two data points to the right (*i.e*. lower ppm) of the apex are read in. If both of these are greater than half of the apex intensity, then each point moves along one, descending down the right-hand side of the peak (see Figure 2.18). When point 2 is less than the half-height intensity, the programme determines whether peak 1 or 2 are closest to the half-height and then stores the index of this point. The programme moves back to the apex and repeats

the process descending down the left-hand side of the peak, stopping again when the point nearest to half height intensity is found.

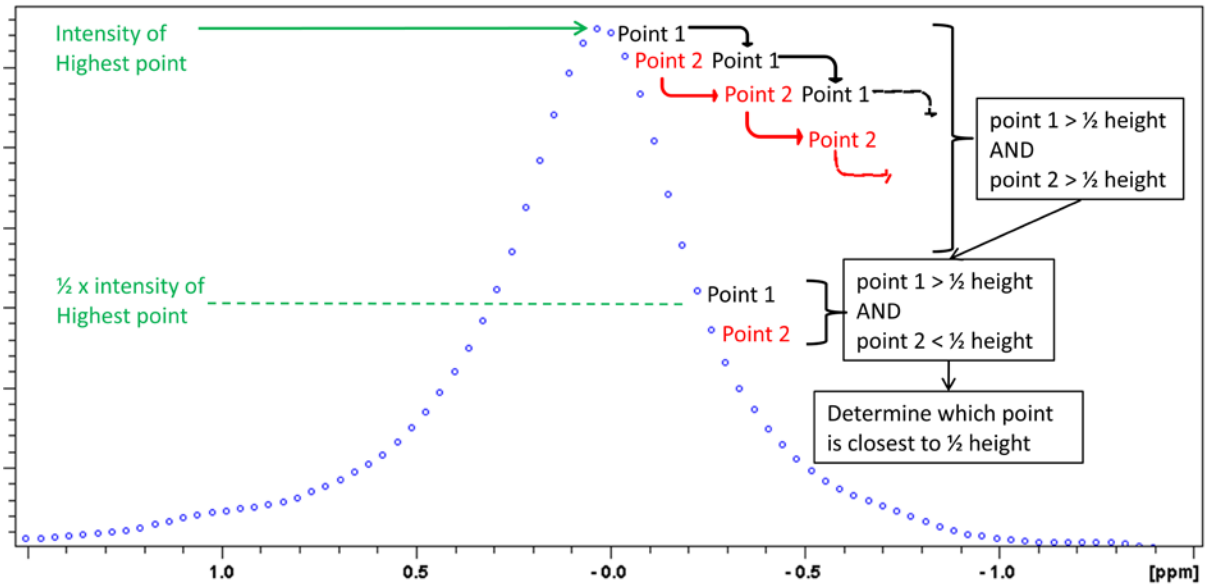


Figure 2.18 – Schematic showing the principles of the function half_height_width

By subtracting the index value from the half-height point on each side of the peak, the approximate peak width can be determined in units of data points. Finally, multiplying this number by the spectral width in Hz and dividing by the number of points in the real spectrum yields the approximate peak width in Hz.

The integral region can now be set such that minimum noise is included in the integration. Note that integration is achieved not by using the built-in functions in Topspin but by a similar process of reading in the intensity of the data points over the region of interest and adding these values up to produce a total absolute integral (see the function “intregion” in eSuppl 1).

2.3.14.2. Automatic phase correction

NMR pulses and the resulting FID can be described as electromagnetic waves, having the properties of frequency, amplitude and phase. For example, in the pulse sequence shown in Figure 2.4, the pulse will have amplitude of “p”, a frequency set close to the Larmor frequency of the nucleus of interest, and also a phase set

relative to the receiver. Any offset in the phase of the applied pulse and the receiver will manifest itself as a deviation from a true absorbance peak shape (see later) in the FT spectrum. Since the factors that cause this offset cannot be readily predicted, a constant phase correction has to be applied to the whole spectrum post-FT. This is referred to as a *zero order phase correction*. More predictable factors such as pulse length, dead time, pre-acquisition delays *etc*. require the application of a further phase correction, this time frequency dependent. This is referred to as a *first order phase correction*, but since the spectra discussed here contain only one resonance that is close to the transmitter frequency, this does not have a significant effect and will not be discussed further.

In the ^{35}Cl NMR spectra discussed here, with the current experimental set up, the zero order phase correction required to produce a resonance in the FT spectrum of pure absorption phase is close to 180°. However, slight changes in the relative phase difference between the pulse and the receiver from one sample to the next will require the zero order phase correction to be adjusted for each spectrum.

This can be done by eye, which can be highly subjective, or by use of automated phase correction routines. The automated routines available in Bruker Topspin were not capable of achieving a consistent absorption phase peak in this case and so a custom routine was written into the processing programme.

One measure of pure absorption mode is the degree to which the resulting peak can be modeled by an ideal mathematical lineshape, in this case a Lorentzian function. Bruker Topspin has a basic fitting module, Sola, and other programmes, such as DMfit [21] can be used to fit Lorentzian functions to peaks. By incrementally changing the phase correction and observing the goodness of fit, optimal phase correction could be obtained. In practice, it was not trivial to interface the fitting programmes and the main AU programme and hence an alternative approach was used.

A simplified measure of ideal peak shape is the symmetry of the peak as assessed at peak half-height, an approach for which is described here. First, an imaginary line

is drawn down from the apex of a peak and a point is taken on this line at half the height of the peak apex. The distances between this point and the nearest data points of the peak directly left and right of the line are then taken. Finally, the difference between these two distances is calculated as a measure of the peak symmetry, with maximum symmetry being achieved at the point of minimum difference. Figure 2.19 shows this pictorially for partially dispersive (asymmetric) and purely absorptive (symmetric) spectra.

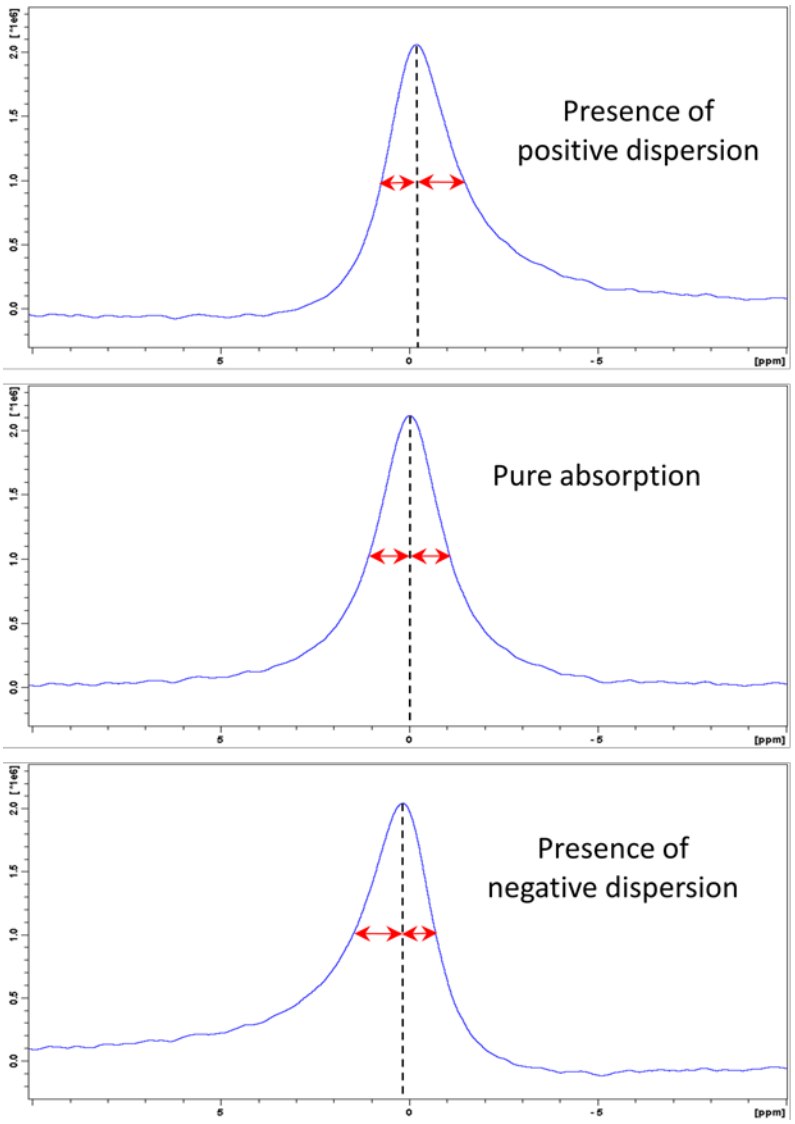


Figure 2.19 – Examples of peak shapes with varying phase errors

This approach has the advantage that it can be completely controlled and run *via* an au programme directly manipulating the processing of the spectrum and then

determining the relevant values by accessing the 1r data file (see 2.3.14.1). Figure 2.20 shows a plot of results as acquired by a subsequent modification to the au programme 35cl_quant.

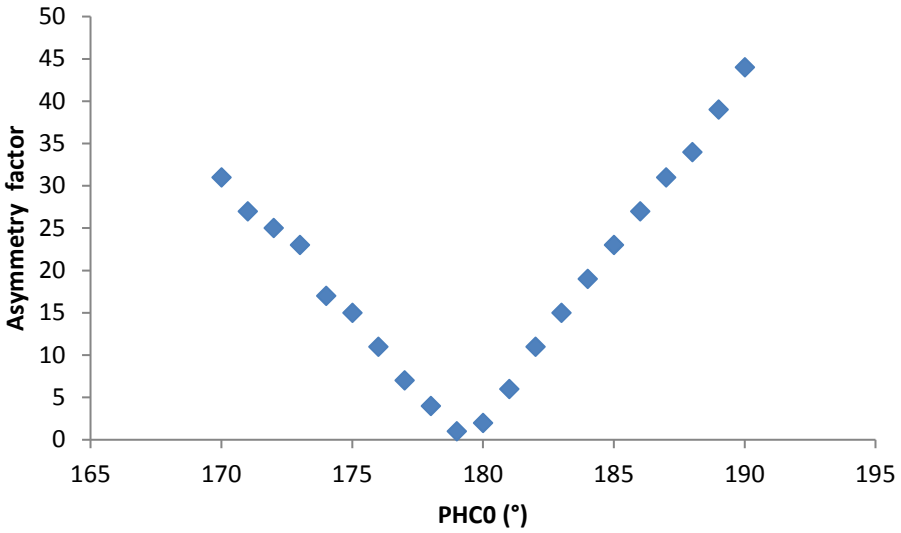


Figure 2.20 – plot of asymmetry factor *vs*. PHC0 as determined by 35cl_quant au programme

The asymmetry factor is calculated by the absolute difference between the apex line at half height and each edge of the peak. The function handling this optimisation is called “phase_opt” in the code in eSuppl 1, and is called as part of step 3 in the flow diagram in Figure 2.17.

2.4. Conclusions

A method for chloride analysis has been developed, with approximately 1 hour run time per sample and with no standard preparation required. The acquisition and processing steps are completely automated and incorporated into the existing walk up user interface. Reports are automatically produced and stored for the user to access. Assuming a nominal sample concentration of 10 mg/mL in D_2O, a DL and QL of 0.1 and 0.5% w/w was achieved respectively for D_2O solutions and the method was shown to be linear over the full range of 0.02 to 10 mg/mL chloride (equivalent to processed results from this method in the range of 0.2 to 100% w/w chloride – although, as previously noted, a result of 100% w/w would not be possible from a real sample).

The validity of this method in the general case has been assessed here and found to be acceptable compared to industry standard guidelines. If this method was to be used for a specific purpose then validation would be repeated, using the analyte in question and linearity samples in replicate covering the range of expected values for the application.

Compared to the only other literature example of chloride analysis by NMR found, the work presented here has produced a much greater level of knowledge in terms of the applicability of this approach to real samples from the pharmaceutical industry. It has been demonstrated that results are independent of the presence of many common chemical species in the sample, and explicitly shown that modern spectrometers can yield quantitative data based on absolute signal response with a stability over many weeks. Finally, it was shown that it is possible to seamlessly integrate the method into existing systems and make the quantitation accessible to non-expert users.

References

[1] S. A. Watson, A. J. Edwards and J. A. Parkinson, Standardless, Automated Determination of Chlorine-35 by ^{35}Cl Nuclear Magnetic Resonance. Analytical Letters, 50 (2017) 161-172.

[2] J. Emsley (2003), Chlorine. Nature's Bulding Blocks, Oxford, Oxford, 105-110.

[3] G. H. Fuller, Nuclear Spins and Moments. Journal of Physical and Chemical Reference Data, 5 (1976) 835-1092.

[4] R. K. M. Harris, B.E. (1978), The Halogens - Chlorine, Bromine and Iodine. NMR and the Periodic Table, Academic Press Inc., London, 421-438.

[5] K. F. Mohr, Neue Massanalytische Bestimmung des Chlors in Verbindungen. Justus Liebigs Annalen Der Chemie, 07 (1856) 335-338.

[6] G. D. Christian (2004), Liquid Chromatography. Analytical Chemistry, J. Wiley and sons, Chichester, UK, 625-627.

[7] Quantitation of Trace and Major Anions in Water by Ion Chromatograpy in a High-Throughput Laboratory, https://assets.thermofisher.com/TFS-Assets/CMD/Application-Notes/CAN-114-High-Throughput-IC-Anions-Bromate-LPN3023-EN.pdf, Retrieved 05 Sep 2019.

[8] Municipal drinking water analysis by fact IC, https://assets.thermofisher.com/TFS-Assets/CMD/Application-Notes/AB-120-IC-Municipal-Drinking-Water-Fast-IC-AB71940-EN.pdf, Retrieved 05 Sep 2019.

[9] H. Lim and S.-G. Lee, Quantitative analysis of chloride by chlorine-35 NMR spectroscopy. Bulletin of the Korean Chemical Society, 27 (2006) 972-973.

[10] International Conference on the Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (2013)

[11] Validation of Analytical Procedures: Text and Methodology Q2(R1). International Conference on the Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1994)

[12] R. K. Harris and B. E. Mann (1978), Introduction. NMR and the Periodic Table, Academic Press Inc., London, 1-19.

[13] T. D. W. Claridge (1999), Practical aspects of high-resolution NMR. High-Resolution NMR Techniques in Organic Chemistry, Pergamon, Oxford, 45-110.

[14] T. D. W. Claridge (1999), Tuning the Probe. High-Resolution Techniques in Organic Chemistry, Pergamon, Oxford, 83-85.

[15] H. Weingartner and H. G. Hertz, Magnetic Relaxation by Quadrupolar Interactions of Ionic Nuclei in Non-aqueous Electrolyte Solutions. Berichte der Bunsengesellschaft fur Physikalische Chemie, 81 (1977) 1204-1221.

[16] A. E. Derome (1987), Why Bother with pulse NMR. Modern NMR Techniques for Research Chemists, Pergamon Press, Oxford, 9-29.

[17] E. Fukushima (1981), Tank Circuits, Impedance Matching, and all that. Addison Weasley, Reading, Pennsylvania, 407-415.

[18] D. I. Hoult and R. E. Richards, The signal-to-noise ratio of the nuclear magnetic resonance experiment. Journal of Magnetic Resonance, 24 (1976) 71-85.

[19] I. W. Burton, M. A. Quilliam and J. A. Walter, Quantitative ^{1}H NMR with External Standards:Use in Preparation of Calibration Solutions for Algal

Toxins and Other Natural Products. Analytical Chemistry, 77 (2005) 3123-3131.
[20] D. E. K. Richie, B.W. (1988), The C Programming Language, Prentice Hall, New Jersey,
[21] D. Massiot, F. Fayon, M. Capron, I. King, S. Le Calvé, B. Alonso, J.-O. Durand, B. Bujoli, Z. Gan and G. Hoatson, Modelling one- and two-dimensional solid-state NMR spectra. Magnetic Resonance in Chemistry, 40 (2002) 70-76.

Chapter 3

Acquiring and Processing Data with Low Field Bench Top NMR

3.1. Introduction

3.1.1. Project aim

High field NMR spectroscopy is considered a mature technology with decades of development in magnet design, component modules (*e.g.* for locking, shimming, pulse generation and amplification, *etc.*) and overall system architecture. Hardware and software have been designed together to complement the specific needs of customer applications by a small number of vendors supplying a large number of instruments over a long period of time. This has led to a wealth of in-depth information about all aspects of the spectrometers being available in manuals, books and general literature, providing a firm bank of knowledge to work from when developing new quantitative and automated applications.

By comparison, there is now a relatively recent range of low field, bench top NMR spectrometers with a large number of vendors for a smaller (although growing) number of systems and users. Literature for these instruments is being published regularly and, where appropriate, references will be included in the introduction to this chapter, but it will take time to achieve the coverage and breadth of knowledge that high field NMR users have come to expect. Consequently, at the time of writing, there are many gaps in the available literature with respect to the “nuts-and-bolts” operation of specific spectrometers and their capabilities. Bench top NMR spectrometers offer advantages to industry in terms of cost, portability, robustness and low maintenance in a number of applications in online, atline and offline analysis. However, without the fundamental practical knowledge discussed above, it is not possible to assess what level of quantitation and automation is feasible from bench top NMR spectrometers.

With this in mind, the aims of the investigations reported in this chapter are to gain a basic, practical understanding of the capabilities of the Magritek Spinsolve NMR spectrometer. These investigations will cover aspects such as characterisation of

flowcell zones in flow mode, long-term stability of field homogeneity and measurement of the non-linear properties of the spectrometer components, including correction strategies to allow accurate synthesis of shaped pulses.

This introduction will identify gaps in the current literature for operational procedures of the Spinsolve. Subsequent experimentation will then seek to fill gaps in the current and lay the foundations for future quantitative and automated applications. Examples of these applications are covered in later chapters.

3.1.2. Background

3.1.2.1. Magritek Spinsolve spectrometer

At the time of writing, there are seven main vendors producing bench top NMR spectrometers commercially [1-8]. The Spinsolve, from Magritek, will be the spectrometer utilised in the following studies.

Figure 3.1 shows an image of the Magritek Spinsolve NMR spectrometer.

Figure 3.1 – Magritek Spinsolve spectrometer

The specifications supplied by Magritek for the spectrometer model used here are as follows:

- Larmor Frequency: 42.5 MHz ^{1}H
- Resolution: HHLW < 0.7 Hz (16 ppb)
- Lineshape: 0.55% linewidth < 20 Hz
- Dimensions: 58 x 43 x 40 cm
- Weight: 55 kg
- Magnet: Halbach array of permanent magnets at 1 T isothermal at 306 K
- Stray field: < 2 G all around system
- Field lock: Internal; monitoring of an unspecified ^{19}F compound
- Gradients: Not on this model
- Tuning/matching: Compromise between ^{1}H and ^{19}F, not accessible to user
- Shim coils: 15 coil geometries - Z, Z^2, Z^3, X, X^3, Y, Y^3, XY, X^2-Y^2, ZX, Z^2X, ZY, Z^2Y, ZXY, $Z(X^2-Y^2)$
- Shimming: adjusted via automated protocols using 10% H_2O in D_2O
- Channels: 1, either ^{1}H or ^{19}F
- Sample introduction: 5 mm NMR tubes or flowcell options

The spectrometer can be safely accommodated on a standard lab bench or placed on a mobile bench to allow co-location with chemistry reactor vessels for real time reaction monitoring. After a prolonged period of downtime the system needs to be left overnight to thermally equilibrate, but this is reduced to < 1 hour when the system has been turned off for a period of a few hours or less.

The cost of the Spinsolve spectrometer used in this work was *ca*. £60,000 at the time of purchase. Since purchase, Magritek Spinsolve spectrometers at 60 and 80 MHz [8] and with gradient coils have become available, as well as spectrometers with ^{13}C and ^{31}P channels.

Shimming

Magritek recommend the basic QuickShim routine to be run daily and the more advanced PowerShim when required. However, as the shimming is likely to be affected by the local laboratory environment, sample temperature, sample introduction method and the build of the individual magnet itself, there is no universal advice on likely behaviours over time. In a highly detailed study and protocol on the application of a new version of this spectrometer to diagnostics of metabolic conditions, Wilson *et al.* [9] state that they routinely perform shimming after every sample to maintain HHLW below 0.55 Hz using a mixture of 10% H_2O in D_2O. Advice in a review by Blümich *et al.* [10] makes the general suggestion that daily shimming is required for benchtop NMR spectrometers and quotes a HHLW specification of 0.7 Hz for the Spinsolve. In later work, Blümich and Singh [11] claim that a HHLW of 0.45 Hz was achieved post shimming. Work by Maiwald *et al.* [12] on reaction monitoring with the Spinsolve notes that after shimming, a HHLW of 0.6 Hz was achieved for neat ethyl acetate when using plastic tubing, and that this did not rise above 1.0 Hz over 48 hours. Whilst these reference indicates the general expectations, it is clear that these data will need to be determined for this particular spectrometer with the experiment setups used in this thesis.

Whereas Basic 1D and 2D NMR experiments are pre-installed in the Spinsolve software for T_1 inversion recovery, T_2 CPMG, COSY and JRES. It is possible to acquire ^{1}H - $\{^{19}F\}$ NMR data using a windowed decoupling experiment supplied for the Spinsolve Expert software.

Software

The spectrometer is controlled using Magritek's own software. Data can be processed using Magritek software, and also can be opened directly and processed with MestreNova (from Mestrelab Research).

Flowing samples and physical parameters

To run in flow mode, two options are available. Firstly, a custom-made glass Dewar can be inserted down the bore of the magnet and PTFE capillary, 2 mm OD (*ca.* 1.5 mm ID), then passes through the Dewar (see Figure 3.2).

Figure 3.2 – Top of spectrometer showing glass Dewar and capillary

This configuration has the disadvantage that the field homogeneity will be adversely affected if the capillary is not of uniform diameter or if it is moved, leading to broadened line shapes. The volume of liquid within the active volume is also not optimum, resulting in a lowering of SNR compared to using an NMR tube.

The second option is a custom glass flowcell that can be inserted into the magnet bore, with a capillary at each end and a widened section around the active volume (see Figure 3.3).

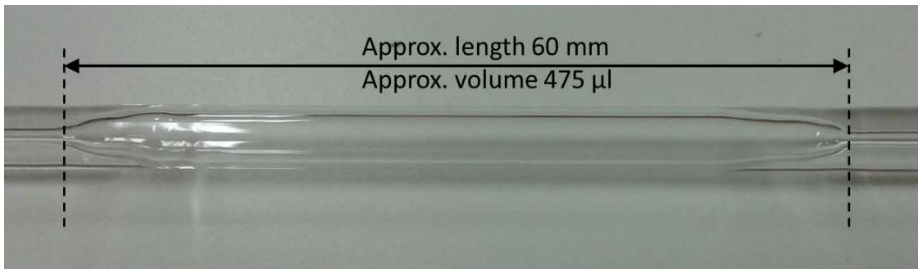


Figure 3.3 – Widened section of flowcell in active volume

It should be noted that the exact dimensions and construction details were not supplied by the vendor for either the glass flowcell, the magnet or the probe. Partial information pertaining to physical parameters of the spectrometer and glass flow cell was found to be distributed across a number of literature references. The details listed here are therefore collated from a combination of notes in literature work and first-hand observations. For example, the volumes in Figure 3.3 were measured here gravimetrically by filling sections of the flowcell with water.

The full dimensions and volumes for the flowcell, as measured here experimentally by partially filling sections of the tube with water (which is then removed and weighed), are:

- Total length of glass rod = 600 mm
- Flowcell length in centre of rod = 60 mm
- Capillary length between end of rod and flowcell (× 2) = 270 mm
- Total volume of glass rod = 750 µL
- Flowcell volume = 475 µL
- Flowcell maximum inner diameter ≈ 3 mm
- Capillary volume between end of rod and flowcell (× 2) = 138 µL

It is likely that the flowcell used here is not precision-made and differences in volumes and lengths between those determined here and those quoted in literature may be due in some part to variability in the fabrication process.

Danieli *et al.* [13] give the dimensions of a glass flow cell for the Spinsolve, with an inner diameter the increases from 1 to 4 mm over a transition region of 10 mm and an active sample volume of about 60 μL. Although no image of the flowcell is given, it is likely that this is of similar, or the same, design as the one used here.

Kappe *et al.* [14] provide an image in their supplementary information in their work that suggests they also use the same flow probe as used here, and quote an internal volume of 800 μL and a total length of 550 mm.

No consolidated details on magnet construction could be found in the literature. Magritek informally quoted the homogenous region within the magnet as a sphere *ca.* 7 mm in diameter. Blümich and Singh [11] calculate a prepolarisation length of 170 mm based on experimental works. Maiwald *et al*. [12] quote a magnet design of 140 mm length, a coil length of 7.5 mm and polarisation length of approximately 100 mm (the last two values based on experimental and calculated data).

Effects of flowing samples on signal intensity were discussed to differing extents in all the above references, and also in references [15-18]. However, to draw direct comparison between literature data and the data acquired here, the spectrometer and flowcell design and setup need to be same, or very similar. This limits the number of references that can be used as a basis of comparison in the following discussion section.

3.1.2.2. Non-linear instrument response

The role of the amplifier in an NMR spectrometer is to increase the amplitude of the synthesised RF pulses. This stage takes the output of the signal generating components, typically around 1 V, and increases the amplitude to hundreds of volts, as supplied to the probe components.

The exact configuration and design of the signal path in the Spinsolve spectrometer was not supplied, but there are a number of non-linear components typical in the RF circuits of spectrometers.

For example, crossed diodes are a common motif in NMR spectrometers [19] as they present low impedance to high voltages, such as the excitation pulse and high impedance to low voltages, such as the returning FID signal from the sample (see Figure 3.4).

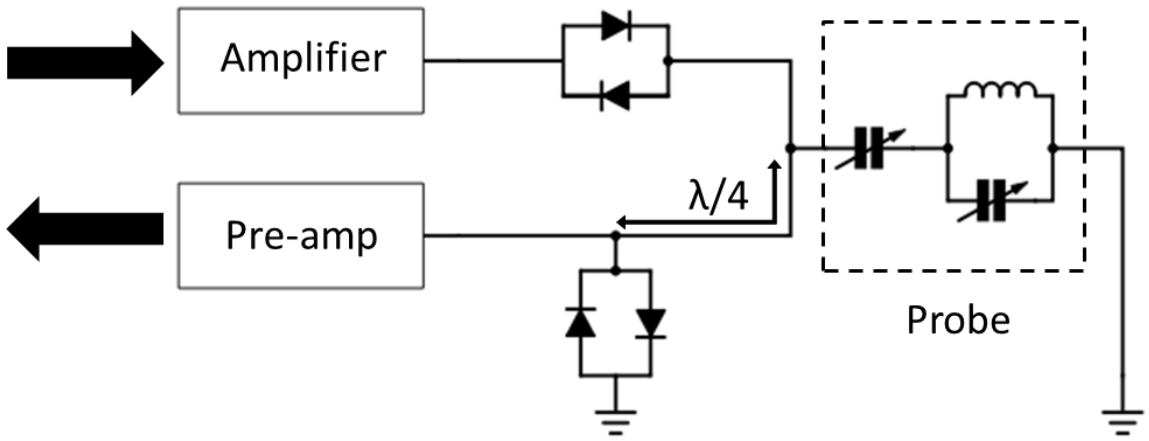


Figure 3.4 – Example NMR circuit diagram

During excitation, the high voltage RF pulse passes from the amplifier through a first set of crossed diodes (now behaving as an effective conductor with low impedance) and on to the probe. A second set of crossed diodes connecting the signal path to ground will also have low impedance, thus protecting the pre-amp from residual RF signal from the excitation pulse. The output impedance, Z_{output}, of the λ/4 path at its junction to the probe components can be characterised by the following equation (3.1) [20]:

$$Z_{output} = \frac{Z_0^2}{Z_{input}} \quad \textbf{(3.1)}$$

where Z_{input} is the input impedance at the λ/4 path junction with the crossed diodes and the preamp and Z_0 is the impedance of the transmission line, which for NMR spectrometers is usually 50 Ω. Under the conditions of excitation as discussed above, $Z_{input} \approx 0$ Ω, hence $Z_{output} \approx \infty$ Ω and the λ/4 path appears to the probe and amplifier as an open circuit, maintaining the voltage at this point.

After the sample has been excited, both sets of cross diodes now behave as high impedance insulators to the weak, low voltage FID signal. The first set of crossed diodes simultaneously block noise from passing from the amplifier into the pre-amp and stop the FID signal being passed back into the amplifier electronics. The FID signal now bypasses the second set of crossed diodes and passes through the λ/4 path to the pre-amp and into the receiver.

This arrangement works well for the square, high voltage, hard pulses typically used for non-selective excitation. For shaped, lower voltage, soft pulses of the type used for band-selective excitation, power may be lost from these components, resulting in non-linear excitation of the sample compared to the expected power.

For high field spectrometers, the non-linearity in amplifier/signal path behavior is measured and stored in correction tables, often as part of the initial installation and setup process. Any subsequent pulse powers are then corrected automatically by the software so that linearity of power response is ensured.

For the new range of low field spectrometers this automated correction is not currently available and so needs to be performed by the user. Furthermore, there appears to be no details in current benchtop NMR literature on the either the degree of amplifier non-linearity, or how to correct it. For example, work by Giraudeau *et al.* [21], extensively utilises shaped pulses in solvent suppression sequences. The work quotes the powers and widths of Gaussian pulses used, but note only that the values were experimentally adjusted in order to obtain the minimal residual solvent signal. It is possible that amplifier non-linearity and the options for mitigation are well

understood amongst advanced users of benchtop spectrometers, but lack of explanation of these aspects in literature is a significant barrier to using shaped pulses to new users and represents a gap in available knowledge.

Once the sample has been excited by the strong RF pulse from the amplifier, the weaker, decaying RF signal from the sample will then be picked up by the coils and enter the receiver. The receiver will pass the signal into an ADC that converts the analogue FID signal to a digital output, in the case of this spectrometer as a series of 16 bit numbers. Depending upon the nature of the sample and the nucleus being analysed, signals could have amplitudes over many orders of magnitude. For this reason, the gain on the receiver needs to be optimised such that the largest voltage from the FID can be represented by the maximum value of 16 bit in the ADC. If the gain is much, less then signal is lost to thermal noise, or if much more, the signal is clipped (see Chapter 2), leading to distorted and non-quantitative spectra.

As with amplifier power, the receiver gain setting on high field NMR spectrometers is corrected such that the relationship between the gain and the size of the digitised signal is linear. The response for the Spinsolve spectrometer is not known, and so will be determined here. Note that on the Spinsolve Expert software, gain is entered as a value between -20 and 70 dB.

3.2. Experimental

3.2.1. Sample preparation and acquisition details

All data were acquired either using the vendor's Spinsolve or Spinsolve Expert software. Using the Spinsolve software, ^{1}H QuickScan is a 'pulse-and-acquire' sequence with a single scan, a fixed pulse width of 8.2 μs at 0 dB, acquisition time of 6.4 s acquiring 32 k points in the FID and a repetition time of 8 s. The dwell time is 200 μs, resulting in spectral width of 5 kHz (~116 ppm). The pulse and acquisition occur at the start of the experiment, with any remaining delay up to the repetition time coming *after* the acquisition. Hence, a manual wait of 5 × T_1 is required before the starting acquisition. The 'Proton+' option allows the number of scans, acquisition time, repetition time and pulse angle to be set to one of a series of predetermined values. A pulse angle of 90° corresponds to a fixed pulse width of 8.2 μs.

T_1 experiments acquired on the Spinsolve software use an inversion recovery pulse sequence and allow the parameters of the number of scans, acquisition time, repetition time, maximum inversion time and number of steps to be set from a series of predetermined values. T_2 experiments use a CPMG sequence and allow the parameters of the number of scans, acquisition time, repetition time, number of steps, CPMG echo time and final echo time to be set from a series of predetermined values.

The Spinsolve Expert software allows full control over data acquisition parameters for ^{1}H pulse and acquire (referred to as 1Pulse-H) and T_1 experiments.

The sample temperature is not directly controlled, but in practice will equilibrate to the magnet temperature of 306 K.

For investigations in flow, samples were pumped through the flowcell with an Ismatec model ISM596D peristaltic pump as supplied with the spectrometer.

3.2.1.1. Low field spectra and spectrometer performance

A reference sample of acetonitrile/ethanol/chloroform in the ratio 80/10/10% v/v was prepared and dispensed into a 5 mm NMR tube to assess the general spectral quality of the spectrometer using the QuickScan experiment. Data were processed with MestreNova without apodization, with manual phase correction and using Bernstein 3rd order polynomial baseline correction.

A sample of 10% H_2O in D_2O in a sealed, 5 mm NMR tube was supplied by the vendor and was used to acquire linewidth and SNR data for long term stability and lineshape checks. The ‘CheckShim’ experiment automatically measures the current HHLW and SNR with a single scan pulse-and-acquire sequence. The ‘QuickShim’ and ‘PowerShim’ experiments perform adjustment of the shims using a simplex algorithm.

3.2.1.2. Effects of flowing samples

The location of the probe coil and its position within the magnet was estimated by dispensing a small quantity of deionised water into the bottom of a long 5 mm NMR tube such that the distance from the bottom of the tube to the meniscus was *ca*. 4 mm. ^{1}H NMR spectra were repeatedly acquired with the sample located at varying positions in the tube holder. The points of maximum signal intensity were noted. An arbitrary “zero position” was established as described in the results and discussion section.

Experiments to investigate the effect of flowcell location were performed by pumping deionised water from a reservoir into the flowcell and back again in a loop. Data were acquired using a reaction monitoring script running a series of Proton+ single scan experiments with 90° pulse angle, acquisition time of 6.4 s, a repetition time of 7 s and a recycle time of 15 s. Five experiments were acquired for each of the positions/conditions in the following investigations:

- Under static flow conditions, the flowcell was set at positions of -16, -8, 0, 8, 16, 24, 32, 40, 49 mm relative to the zero position.
- Under flowing conditions, the flowcell was set at positions of 0, 8, 16, 24 and 32 mm relative to the zero position at each flow rate of 0.5, 1, 2, 3 and 4 mL/min.
- The flowcell was fixed at the zero position and flow rates were set at 0, 0.5, 1, 2, 3, 4, 5, 6 mL/min.

T_1 data were acquired with 1 scan, acquisition time of 6.4 s, repetition time of 15 s, maximum inversion time of 10 s and 11 steps. Duplicate analyses were performed for each of the positions/conditions in the following investigations:

- Under static conditions, the flowcell was set at positions of 0, 16 and 32 mm relative to the zero position for each flow rate of 0, 2, 4 mL/min.

3.2.1.3. Non-linear instrument response

A sample of 10% v/v acetone in acetone-d_6 was dispensed into a 5 mm NMR tube.

All data were acquired with the Spinsolve Expert software using the 1Pulse-H experiment.

Experiments to establish amplifier response were acquired with an acquisition time of 6.4 s acquiring 32 k points in the FID with a dwell time of 200 μs and a repetition time of 25 s. The pulse was set at 7.575 μs over the range 0 to -24 dB, 121.2 μs over the range -30 to -45 dB, 242.4 μs over the range -48 to -51 dB, 484.8 μs at -54 dB and 969.6 μs over the range -54 to -63 dB. The full range of 0 to -63 dB was explored over 22 experiments in steps of -3 dB. The 90° pulse of 7.575 μs at 0 dB was calibrated with the acetone sample by manually varying the pulse width and observing the 360° null point.

Experiments to test the selective pulse were based on a modified 1Pulse-H with details described in results and discussions with the acquisition parameters set as for the amplifier response experiments and using the acetonitrile/ethanol/chloroform sample. The carrier frequency was set to the chemical shift of the chloroform resonance for each experiment. The Gaussian pulse shape was calculated over 50 points using Bruker Topspin Shapetool and implemented with a pulse width of 11880 μs at -45.91 dB, equal to a bandwidth of 150 Hz.

The receiver gain response was investigated using the acetone sample prepared above and the acquisition parameters set as for the amplifier response experiments. The receiver gain was varied over a range of 70 to -20 dB in 3 dB steps resulting in 31 experiments.

3.2.1.4. Lineshape

All experiments were acquired on the acetonitrile/ethanol/chloroform sample prepared above.

High field ^{1}H NMR data were acquired on a Bruker 400 MHz Avance-III NMR spectrometer, as described in Chapter 2. The ^{1}H NMR spectrum was acquired with 8 scans, a pulse width of 14.5 μs at 10 W, acquisition time of 8 s acquiring 128 k points in the FID and with a repetition time of 10 s. Data were processed using Topspin 3.0 with exponential apodisation of 0.3 Hz, manual phasing and automated baseline correction. The T_2 data were acquired using a CPMG sequence with 4 scans, refocusing delays of 0.1, 1.5, 3, 4.5, 6.1, 7.6, 9.1, 10.7, 12.2, 13.7, 15.2 s, and a relaxation delay of 15 s.

The low field ^{1}H NMR spectrum was acquired using the Proton+ experiment with 256 scans, 90° pulse angle, acquisition time of 6.4 s, and a repetition time of 10 s. T_2 data were acquired with 4 scans, acquisition time of 6.4 s, repetition time of 15 s, CPMG echo time of 10 ms and final echo time of 5 s over 10 steps. Data were

processed using MestreNova with no apodisation, zero filled 4 times, manually phase corrected and baseline corrected with a Bernstein 3rd order polynomial.

High and low field NMR data were deconvoluted using DMFit 2011 software.

3.3. Results and discussion

3.3.1. Low field NMR spectra and spectrometer performance

An example 42.5 MHz NMR spectrum of the acetonitrile/ethanol/chloroform sample is shown in Figure 3.5. The 9.4 fold decrease in spectrometer frequency compared to a typical 400 MHz NMR spectrometer results in a corresponding increase in the apparent coupling of peaks in the FT spectrum. For example, the ^{1}J (^{1}H-^{13}C) coupling for the methyl protons in acetonitrile is *ca*. 136 Hz, which equates to the ^{13}C satellites being separated by 3.24 ppm at 42.5 MHz, compared to a separation of only 0.34 ppm at 400 MHz. The lower ppm ^{13}C satellite for acetonitrile can be seen in Figure 3.5 at around 0 ppm, with the second satellite overlapping with the ethanol CH_2 multiplet at *ca*. 3.24 ppm.

The insets at the top of Figure 3.5 show the full spectrum with all peaks on scale and with vertical zoom to demonstrate the baseline and noise over this full range.

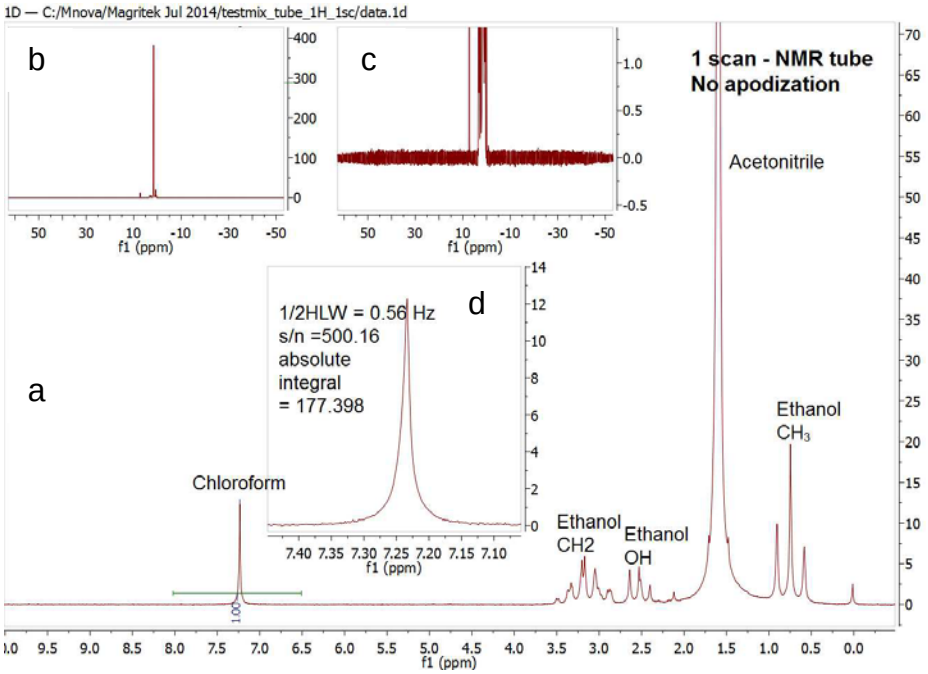


Figure 3.5 – a) 42.5 MHz 1D ^{1}H NMR spectrum of ethanol/chloroform/acetonitrile 10/10/80% v/v in a 5 mm NMR tube, b) Full spectrum with all peaks on scale c) Full spectrum with vertical zoom d) Close-up of chloroform peak

Shimming was performed regularly using the 10% H_2O in D_2O sample. From inspection of data acquired over 2 years it was found that the linewidth and SNR can drift over a period of 24 hours and the system requires the QuickShim routine to be run daily before use. The spectrometer is capable of routinely achieving a HHLW of 0.4 – 0.5 Hz and a SNR for the shimming sample of 9000 – 10000 : 1. This is consistent with ranges achieved by Wilson *et al.* [9] and Blümich and Singh [11]. Figure 3.6 shows an example of HHLW before and after the daily QuickShim routine over a period of several months.

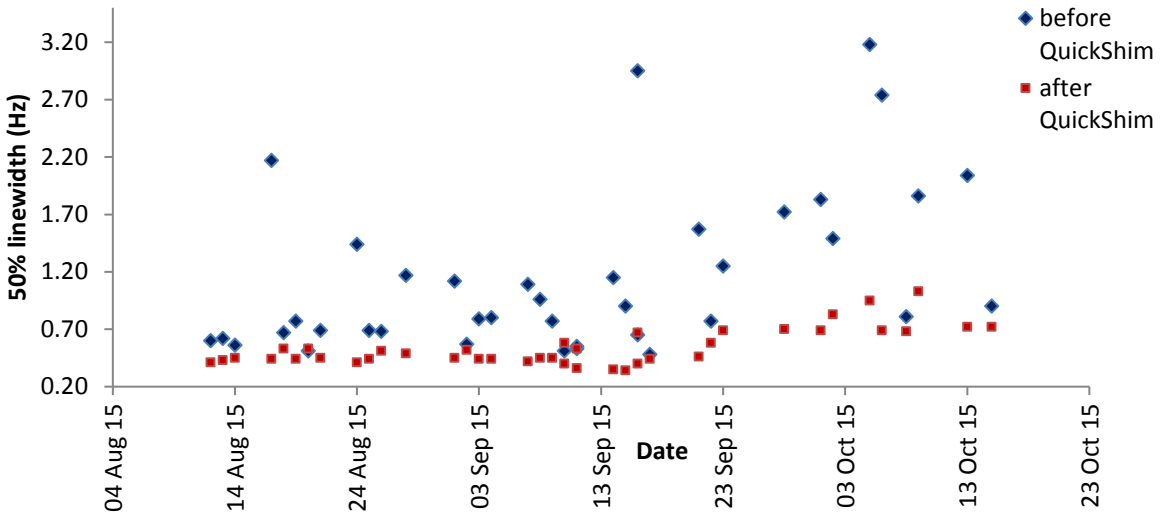


Figure 3.6 – Half height linewidth of 10% H_2O in D_2O before and after QuickShim routine

As can be seen, the HHLW after QuickShim increased over a period of weeks and therefore a PowerShim routine was subsequently run approximately every month. The regime of running daily QuickShim and monthly PowerShim routines maintains a HHLW of <= 0.5 Hz and SNR of > 9000 : 1. The knowledge is additional to that provided in the literature references and allows a comprehensive protocol for instrument shimming to be implemented on the Spinsolve spectrometer.

3.3.2. Effects of flowing samples

The magnetic field within the bore of the spectrometer magnet is not uniform and a gradient will exist across the volume of the NMR tube or flowcell. Spins in different spatial locations within this volume will experience different magnetic fields and therefore achieve equilibrium magnetisation and precess at different rates. Although spins will migrate between locations over time due to Brownian motion and convection currents, this will also reach an equilibrium and can effectively be considered as a constant for a given temperature, solvent, spectrometer, probe and sample tube.

The implications become more complex if the sample is moving through a flowcell, as the spins will experience a sequence of different magnetic fields before and during excitation from the RF pulse, and during acquisition. At minimum, literature sources on the subject note that flow NMR offers advantages in time efficiency [22, 23]. After an RF pulse the excited spins are replaced by new spins already at equilibrium magnetisation, effectively equating flow to an additional relaxation pathway, meaning shorter observed T_1 and hence faster repetition times are possible. This is, in part, offset by a similar effect upon T_2, leading to broader lineshapes. This effect is described by the equations (3.2) and (3.3) from [22, 23].

$$\frac{1}{T_{1flow}} = \frac{1}{T_{1static}} + \frac{1}{\tau} \quad \textbf{(3.2)}$$

$$\frac{1}{T^*_{2flow}} = \frac{1}{T_{2static}} + \frac{1}{T_{2\Delta B_0}} + \frac{1}{\tau} \quad \textbf{(3.3)}$$

where τ is the residence time, in this instance in the active volume, as defined in (3.4):

$$\tau = \frac{Volume}{Flow\ rate} \quad \textbf{(3.4)}$$

The assumption not stated here is that the spins entering the active volume have achieved full equilibrium magnetisation, which in turn assumes the flowing solution has spent sufficient time in a premagnetisation volume before entering the active volume. Most articles [24, 15, 13, 21, 16, 11] include this further level of complexity into the model and discuss the impact of up to four discrete volumes [24] that the solution may flow through. Of these, only two are directly relevant to the work here: the premagnetisation volume and the active volume/detection cell volume (described as “zone 2” and “zone 4” in this last reference, respectively).

The active volume with the flowcell installed was initially estimated through a combination of experimental data and assumptions based on these data. For spectra

acquired under static conditions, no intensity change would be expected whilst the flowcell is installed such that the coil sits within the widest section. Moving a sample of water in a NMR tube manually through the magnet in small increments gave a maximum signal over a distance of 6 mm, and the signal intensity dropped off to < 1% of the maximum over another 6 mm above and below this region. This is consistent with information from Magritek and [12] that the coil height is 7 – 8 mm in length and that the region with some degree of excitation is several millimeters outside this region. From the observed data and the supplied information, an active height of *ca*. 14 mm is estimated. Assuming from observation flowcell has an inner diameter of *ca.* 3 mm, when this accessory is in place the active volume is estimated to be *ca.* 100 µL. This is greater than that quoted in [13], which was authored in part by representatives from Magritek, although the precise volume may depend upon the exact spectrometer and iteration of magnet design.

In summary, as long as the flowcell was placed such that the coil sat over the two marked bands on the far ends of the flowcell, as show on the left in Figure 3.7, then no significant drop in intensity from the maximum was observed. If the flowcell was moved such that the coils were outside of these marked bands in regions where the flowcell diameter decreases, then the intensity was shown to drop, as expected.

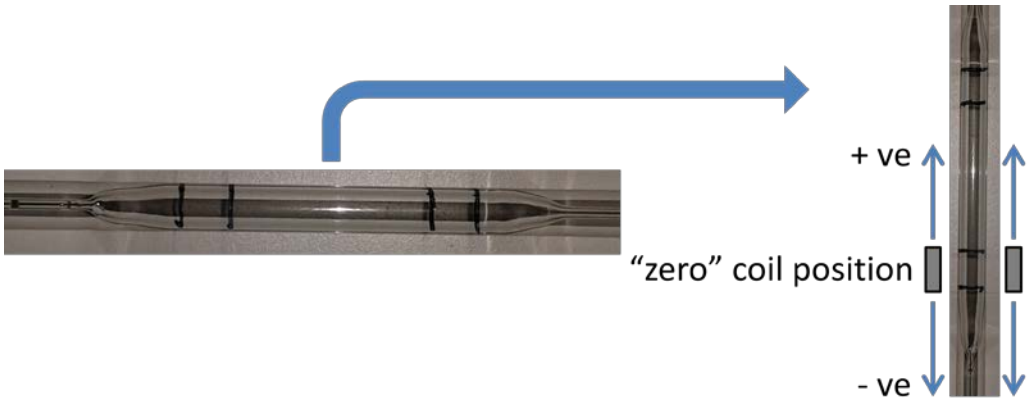


Figure 3.7 – Image of flowcell showing furthest extents of the region of maximum intensity (left) and the reference "zero" position for the coil (right)

The right-hand image in Figure 3.7 shows the "zero" position of the coil around the flowcell, used as a reference point for subsequent experiments.

The data for water peak intensity as the flowcell was moved through the magnet under static conditions is shown in Figure 3.8.

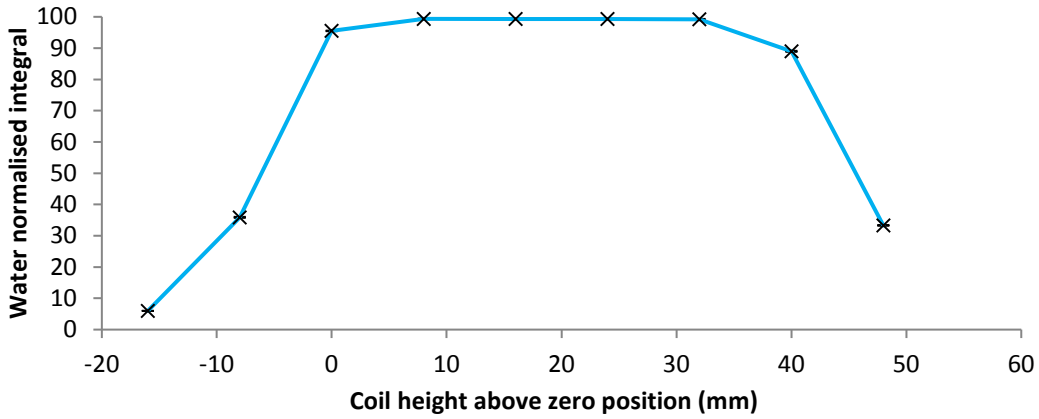


Figure 3.8 – Water peak integral as a function of location of the flowcell in the magnet (horizontal error bars denote ± 2 standard deviations, not visible on this scale)

Each point represents an average of normalised integrals for the five experiments with associated error bars, but in some cases the variation is too small to be

observed on this scale. As expected, a plateau was observed when the coils were located over the widest region of the flowcell.

The data for average water peak intensity as a function of flowcell position and flow rate is shown in Figure 3.9. Note that the flow through the magnet is from bottom to top, with reference to Figure 3.7.

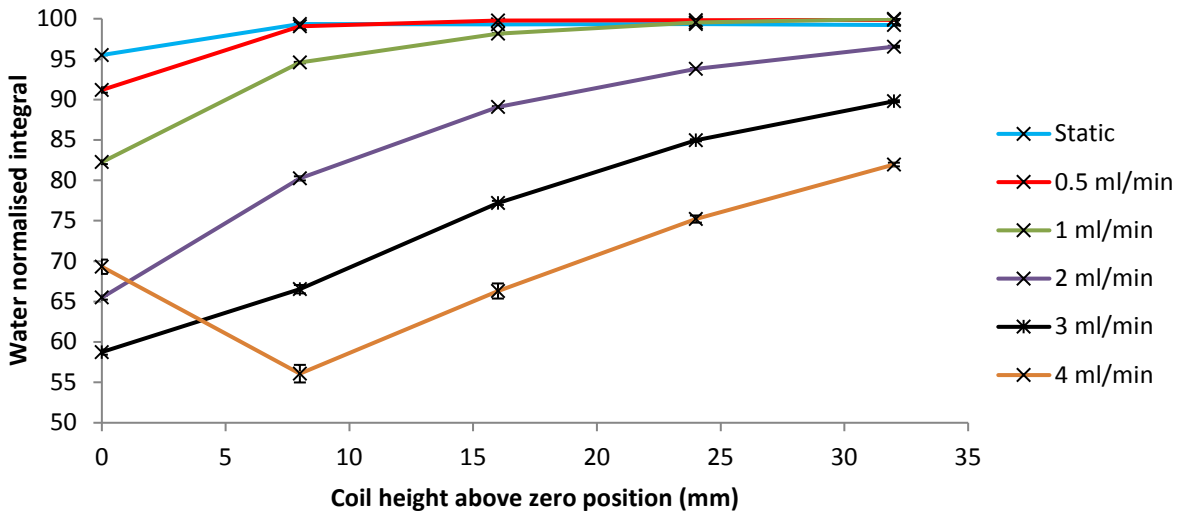


Figure 3.9 – Water peak integral with variable flowcell position and flow rate (error bars denote ± 2 standard deviations)

The slight changes observed in signal intensity under static conditions may be due to imperfections in the coils leading to a non-uniform diameter. For flowing samples there were several more notable trends. Firstly, a decrease in signal intensity was observed with increasing flow rate, at all sample positions with the exception of the zero position at 4 mL/min. In reference to the previous discussion, this suggested that the premagnetisation volume as described in [24] was such that the incoming spins do not achieve full equilibrium magnetisation, even at lower flow rates. Secondly, the effect that the flow had upon signal intensity was dependent upon the coil position over the flowcell. The optimum position of the flowcell was such that the coil was placed higher up the flowcell. In this way more of the wider section of the tube was upstream of the coil. This increased the effective premagnetisation volume and residence time in this volume and resulted in the spins having greater time to magnetise before entering the active volume for excitation.

Figure 2 in [21] shows a plot of normalised peak of a 40% ethanol in water at different flow rates and with varying recycle delays, Tr (give as ratio of Tr to T_1). The flowcell also appears to the same, or similar to the one used here. The peak area in the published work does not decrease significantly up to a flow rate of 1 mL/min and then drops increasingly with increased flow rate, as observed here. It appears from the published plot that the normalised area does not drop below the equivalent of 95% until 3 mL/min, whereas the area is already at 90% in the data generated here. These discrepancies are most probably due to differences in the relative position of the flow cell between the literature and experiments acquired here, leading to differences in pre-polarisation volume. This aspect is not explicitly discussed either in this reference, or other literature references given in this chapter on the subject. In contrast, the experiments here demonstrate this effect clearly and provide useful information for anyone replicating a similar flow setup for the first time.

For comparison with other flow options, Figure 3 in [13] shows a plot of the integral of peaks from isopropanol flowing through a PTFE tube of 0.93 mm ID *vs.* flow rate. In this work the decrease in peak area is similar to that observed in the experiments performed here, again demonstrating that these effects need to be measure for each individual instrument and flowcell to optimise performance.

The discrepancy observed here at zero height and 4 mL/min was not readily explained. The full set of experiments were repeated and the same results for all conditions were noted, including the observations at 4 mL/min. It may be that the field gradient below the active volume was more complex than a simple decrease over distance, and/or there is a change in the flow regime at this location and flow rate. For example, a lower volume may be subjected to a higher field. If this were the case, then a higher flow rate may introduce spins with a greater degree of polarisation into the active volume from this lower volume. To investigate this further, the data for average water peak intensity versus flow rate with the flowcell at the zero position was examined, as shown in Figure 3.10.

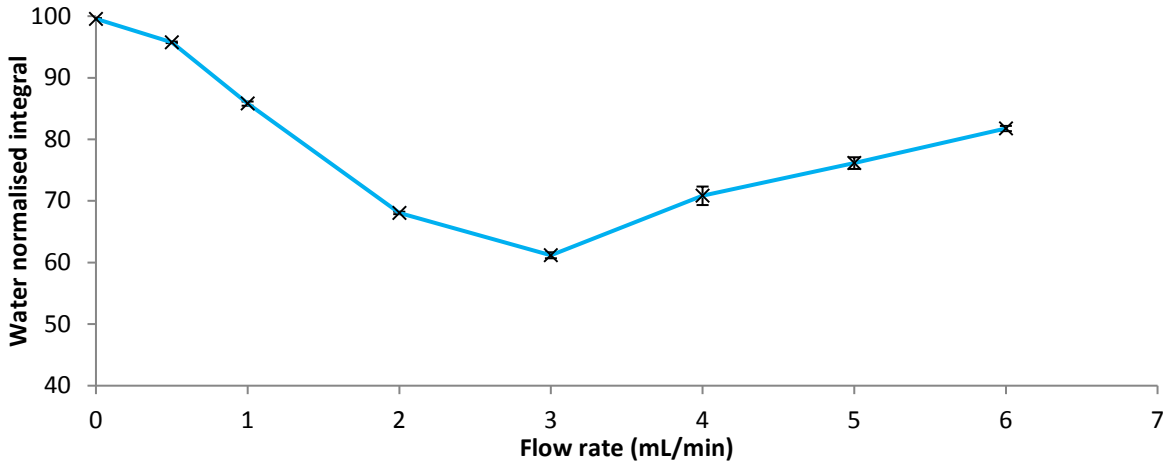


Figure 3.10 – Water peak integral with increasing flow rate and fixed zero coil position (error bars denote ± 2 standard deviations)

A progressive increase in intensity was observed between flow rates of 3 and 6 mL/min, which is consistent with, but not a confirmation of, the hypothesis above. Testing this hypothesis would require either magnetic field measuring instrumentation or vendor information not available at this time. Alternatively, if the experiments were repeated with the flow in the opposite direction (not conducted at this time), this might give more information on the symmetry of the field gradient within the magnet. It is also interesting to note that this increase is not observed in the data presented in figure 2 in [21] as discussed above, suggesting differences in some aspect, or aspects, of the setup between the literature work and the experiments shown here. As these higher flow rate were not likely to be required in the work planned here, this was not investigated further at this time.

Average ^{1}H T_1 as a function of flow rate and flowcell position is shown in Figure 3.11.

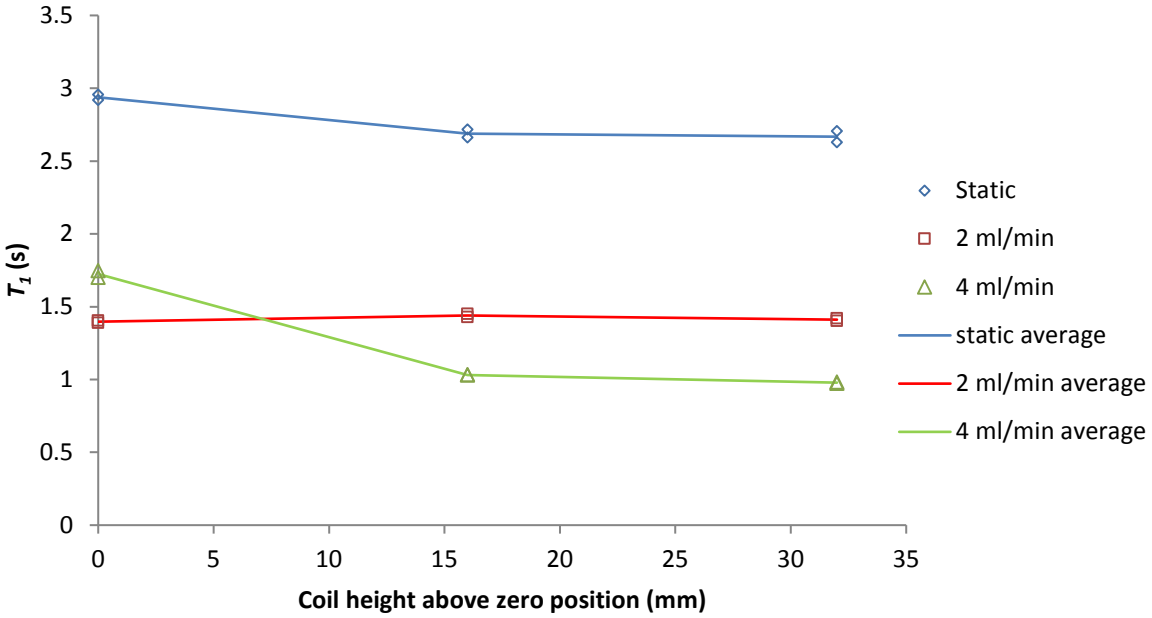


Figure 3.11 – ^{1}H T_1 of water with varying flow rates and coil position. Lines show trend of data averaged for each experiment.

Under static and 2 mL/min flow conditions, relatively minor changes in the T_1 values were observed as the coil position was changed. A different trend was observed when the coil was located at the zero position and water was flowing at 4 mL/min. This may further supported the hypothesis that at higher flow rates, spins from a region of higher magnetic field were being swept into the flowcell. The relationships in (3.2) and (3.4), as applied to the active volume, can be combined to give (3.5).

$$\frac{1}{T_{1flow}} = \frac{1}{T_{1static}} + flow\ rate \cdot \frac{1}{active\ volume} \qquad \textbf{(3.5)}$$

Thus if $1/T_{1flow}$ is plotted against flow rate, then a linear relationship with a slope of 1/active volume would be expected. T_1 data for static and 2 mL/min flow conditions from Figure 3.11 were averaged over the different cell positions. For a flow rate of 4 mL/min, only the mid and high point data were averaged as the zero position data represents a different regimen that is not fully understood. These data are shown in Figure 3.12.

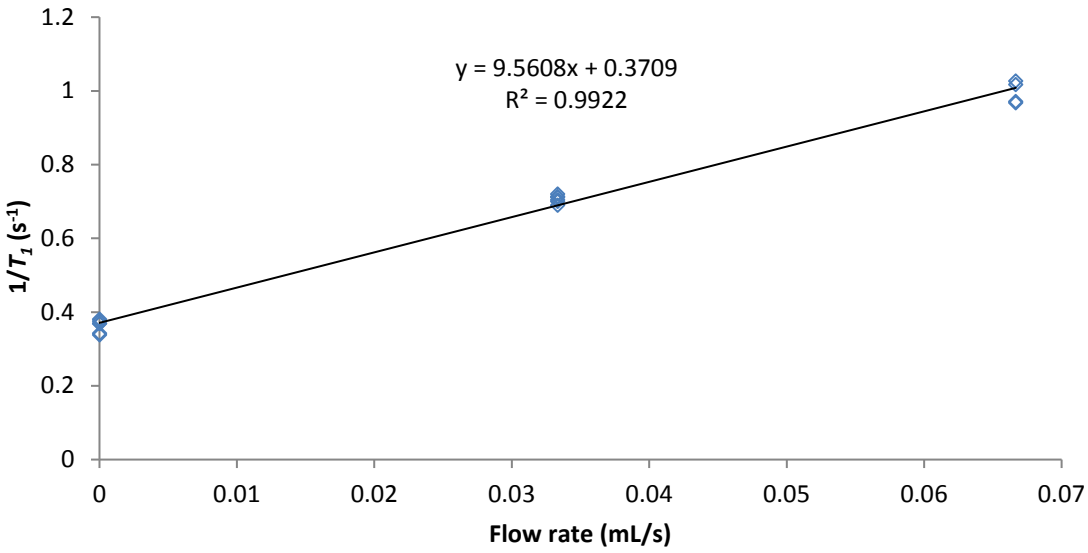


Figure 3.12 – Plot of 1/T_1 versus flow rate

The slope of 9.5608 equates to an active volume of 104 µL (from Equation (3.5)), which is consistent with the value of *ca*. 100 µL estimated earlier on in this section.

This result was calculated from a limited set of data with only 3 points for regression, but it does suggest that this approach could be used as a method for estimating the active volume experimentally. Data for T_1 at varying flow rate on the spectrometer and flowcell used here could not be found in literature, although figure 4 in [25] shows a plot of 1/T_1 *vs.* flow rate for a different spectrometer and flowcell. The plot in the published work also shows a increasing trend similar to that shown in Figure 3.12, but with differing parameters. This literature reference did not take the additional step of estimating the active volume as was done here, nor could an example of this be found in the literature presented in this chapter.

With the optimum position of the flowcell established (where the coil is situated at the upper end of the widened section), the premagnetisation volume in this configuration could be estimated. Equation (3.6) taken from [26] assumes laminar flow. The setup in this reference uses what appears to be the same, or similar flow cell to that used in the works here. Diagrams in the published work also suggest that the prepolarisation volume extends over the widened and capillary sections of the flow

cell. It is not apparent or explained how the linear velocity in distance per unit time is calculated for the variable diameter tube. For this reason, the equation as shown here is modified from the literature reference to take the flow rate in mL/min and hence yields a prepolarisation volume V_{pol}, rather than a length, as given in [26].

$$\frac{S}{S_0} = 1 - exp\left(-\frac{V_{pol}}{flow\ rate \times T_{1static}}\right) \quad \textbf{(3.6)}$$

where S is the observed signal amplitude.

The data for signal amplitude was taken from Figure 3.9, with the coil height 32 mm above the zero position and $T_{1static}$ taken from averaged values from Figure 3.11. The data were plotted and then fitted in Statistica by non-linear estimation (see Figure 3.13).

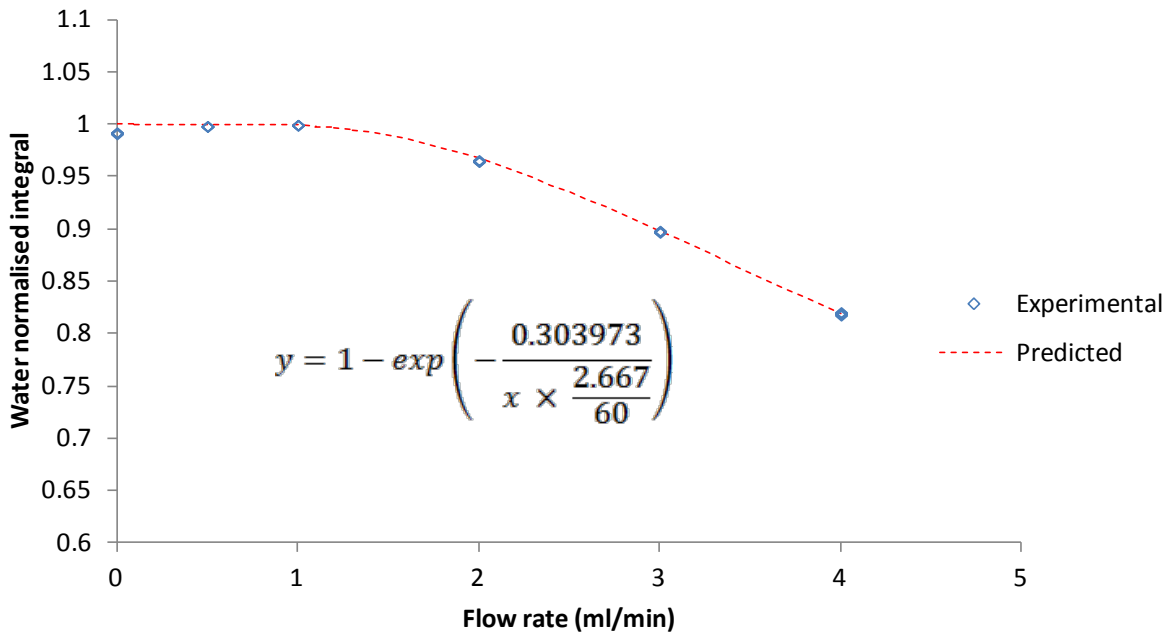


Figure 3.13 – Water peak integral as a function of flow rate, modelled to estimate magnetisation volume

The equation fits the data well and yields a prepolarisation volume of 304 μL. As noted this cannot be directly related to the linear polarisation length of 170 mm as described in the literature. Future work would need to repeat these experiments using tubing of constant diameter to flow the sample through the tube.

3.3.3. Amplifier response

A given amplifier will have a maximum possible power output, P_{max}, measured in Watt (W). For the Spinsolve spectrometer, P_{max} is 1 mW. Typically, all powers that the amplifier will need to deliver will be at or below this value, which is achieved by attenuating the power output. For this reason, it is common for spectrometer software to require the user to enter the required power as a relative attenuation in units of decibels (dB), rather than as an absolute power in W. The former expresses power in terms of how much P_{max} needs to be decreased, or attenuated, and uses a base 10 logarithmic scale (3.7) [27].

$$dB = 10 \times log_{10}\left(\frac{P_{final}}{P_{initial}}\right) \quad \textbf{(3.7)}$$

For the discussion above, $P_{initial} = P_{max}$, and P_{final} is the required power, in W.

To produce a correction table for the Spinsolve spectrometer, the relationship of P_{in} and P_{out} needs to be determined. These terms are defined here as the amplifier power as set by the user in the software and the power the sample actually experiences, respectively. A simple process to determine this relationship would be to measure the ^{1}H 90° pulse (PW_{90}) at P_{max}. The power would then be progressively decreased and PW_{90} measured.

The angle that the bulk magnetisation vector makes to the z-axis after an RF pulse is referred to as the flip angle, θ. In units of radians, this can be calculated as shown in (3.8) [28].

$$\theta = \gamma \times B_1 \times PW \quad \textbf{(3.8)}$$

where B_1 is the RF power of the pulse in Tesla (T), which can be defined as shown in (3.9) in terms of V, the potential difference in Volts.

$$B_1 = \frac{V.s}{m^2} \quad \textbf{(3.9)}$$

and from simple electrical theory (3.10):

$$P = \frac{V^2}{R} \quad \textbf{(3.10)}$$

$$V = \sqrt{PR} \quad \textbf{(3.11)}$$

substituting (3.11) into (3.9):

$$B_1 = \frac{\sqrt{PR}.s}{m^2} \quad \textbf{(3.12)}$$

and substituting (3.12) into (3.8):

$$\theta = \gamma \times \frac{\sqrt{PR}.s}{m^2} \times PW \quad \textbf{(3.13)}$$

Hence we can see from (3.13) that PW is inversely proportional to the square root of the power in Watts (3.14).

$$PW \propto \frac{1}{\sqrt{P}} \quad \textbf{(3.14)}$$

To express (3.7) in terms of a ratio of pulse widths it is necessary to invert the relationship and square the PW values as shown (3.15).

$$dB = 10 \times log_{10}\left(\frac{PW_{90\,initial}}{PW_{90\,final}}\right)^2 \quad \textbf{(3.15)}$$

or

$$dB = 20 \times log_{10}\left(\frac{PW_{90\ initial}}{PW_{90\ final}}\right) \quad \textbf{(3.16)}$$

where the $PW_{90\ initial}$ is the initial PW_{90} value measured at P_{max} (0 dB), and $PW_{90\ final}$ is the PW_{90} measured at an attenuated pulse power. (3.16) is also referenced in [27], but has been derived here for clarity and to lay a basic foundation for following derivations in this section.

PW_{90}, in this instance, is the theoretical pulse width that would be expected if the pulse power that the sample experienced was the same as the value set, *i.e.* perfect linearity of response from the amplifier.

These experiments were performed on the Spinsolve spectrometer using the Expert software on a sample of 10% v/v acetone in acetone-d_6. Data for P_{in} versus P_{out} as determined from 90° pulse calibration is shown in Figure 3.14.

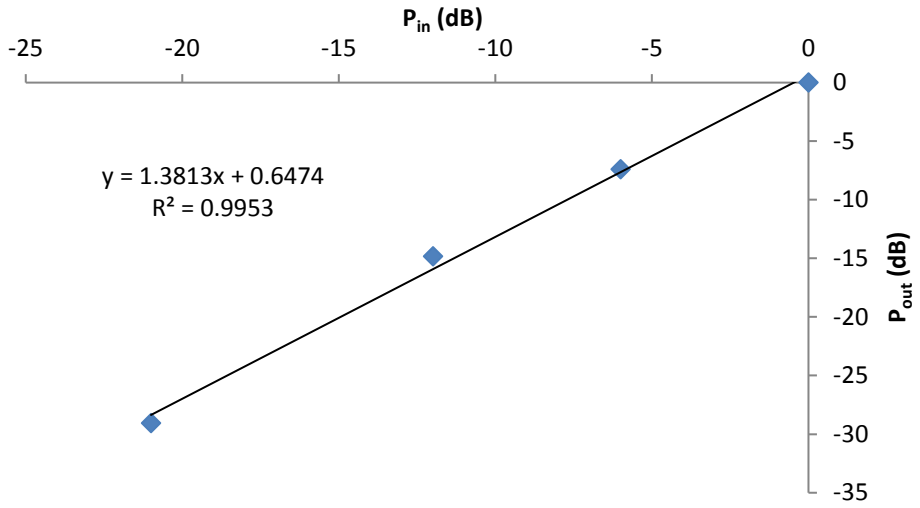


Figure 3.14 – P_{out} versus P_{in} as determined by change in PW_{90}

The data appears to show an approximately linear relationship over short range of powers set here. At -21 dB, the 90° pulse width is around 860 µs. The spectrometer

software has a limit of 1000 µs for pulses to protect the hardware, and so longer pulses are not readily achievable. Synthesising shaped pulses requires the amplifier to deliver very low pulse powers, so an alternative approach is required to understand amplifier response below -21 dB.

In the following approach it is shown that it is possible to measure the change in peak area between a spectrum acquired at a known initial state and one acquired at a state with different conditions and use this to calculate the difference in pulse power that the sample has experienced between these two states. This approach will take relationships from literature and from them derive a new formula that will enable amplifier linearity to be easily assessed over a wide range of powers.

It can be seen from (3.8) that θ is proportional to a change made to the pulse width, *e.g.* from the $PW_{90\ initial}$ at 0 dB to another value (3.17).

$$\theta \propto \frac{PW_{final}}{PW_{90\ initial}} \tag{3.17}$$

Where PW_{final} is now any pulse width set by the user for a particular experiment. It has already been shown that the relationship of power attenuation in dB to PW is expressed as (3.16). This can also be expressed in terms of the power that the sample experiences after a pulse P_{out} (3.18).

$$P_{out} = 20 \times log_{10}\left(\frac{PW_{final}}{PW_{90\ initial}}\right) \tag{3.18}$$

rearrange to:

$$\frac{P_{out}}{20} = log_{10}\left(\frac{PW_{final}}{PW_{90\ initial}}\right)$$

$$10^{\frac{P_{out}}{20}} = \frac{PW_{final}}{PW_{90\ initial}} \tag{3.19}$$

From (3.19) a relationship between θ and P_{out} can be described (3.20).

$$\theta \propto 10^{\frac{P_{out}}{20}} \tag{3.20}$$

Combining (3.17) and (3.20), the experimental flip angle that the sample experiences, θ_{exp}, can be determined for a given change in PW and power attenuation (3.21).

$$\theta_{exp} = \theta_{90} \times 10^{\frac{P_{out}}{20}} \times \frac{PW_{final}}{PW_{90\ initial}} \tag{3.21}$$

where

$$\theta_{90} = \frac{\pi}{2}\ (radians)$$

It can also be shown that the peak area in a given experiment, A_{exp}, is proportional to the sine of θ_{exp} (3.22).

$$A_{exp} = A_{90} \times \sin\theta_{exp} \tag{3.22}$$

Where A_{90} is the peak area with a 90° flip angle pulse, *i.e.* the maximum signal possible for a single pulse experiment. This can be rearranged to:

$$\theta_{exp} = sin^{-1}\frac{A_{exp}}{A_{90}} \tag{3.23}$$

and substituting (3.23) into (3.21) to give (3.24).

$$sin^{-1}\frac{A_{exp}}{A_{90}} = \theta_{90} \times 10^{\frac{P_{out}}{20}} \times \frac{PW_{final}}{PW_{90\ initial}} \tag{3.24}$$

Rearranging:

$$\frac{sin^{-1}\frac{A_{exp}}{A_{90}}}{\theta_{90}} = 10^{\frac{P_{out}}{20}} \times \frac{PW_{final}}{PW_{90\ initial}}$$

$$\frac{sin^{-1}\frac{A_{exp}}{A_{90}}}{\theta_{90}} \times \frac{PW_{90\ initial}}{PW_{final}} = 10^{\frac{P_{out}}{20}}$$

$$log_{10}\left(\frac{sin^{-1}\frac{A_{exp}}{A_{90}}}{\theta_{90}} \times \frac{PW_{90\ initial}}{PW_{final}}\right) = \frac{P_{out}}{20}$$

$$P_{out} = 20 \times log_{10}\left(\frac{sin^{-1}\frac{A_{exp}}{A_{90}}}{\theta_{90}} \times \frac{PW_{90\ initial}}{PW_{final}}\right) \quad \textbf{(3.25)}$$

This formula as derived here has not been found in the literature in this form for this purpose and it's utility will now be demonstrated.

In practice, $PW_{90\ initial}$ is first calibrated at 0 dB, a spectrum is acquired and the peak area measured. A series of spectra are then acquired with $PW_{final} = PW_{90\ initial}$ and increasing attenuation, P_{in}, and the peak area is then measured for each. At the point at which the peak area cannot be accurately measured due to low SNR, the pulse width (PW_{final}) is increased and the spectrum acquired. P_{out} can then be calculated from (3.25). P_{in} is then plotted versus P_{out} and the amplifier response can be evaluated as shown in Figure 3.15.

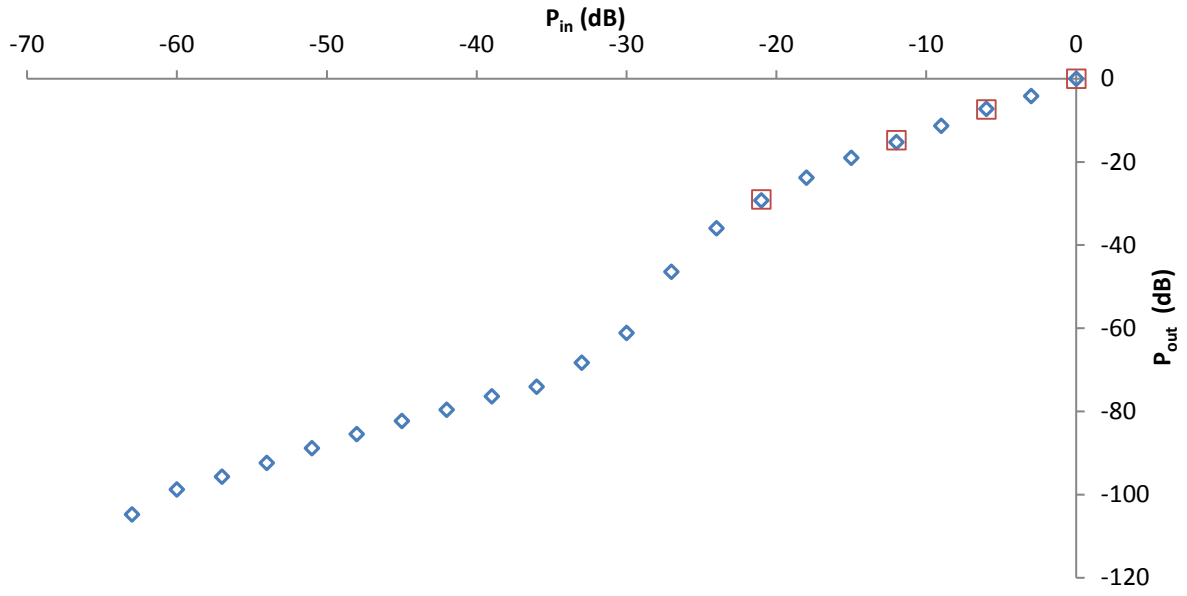


Figure 3.15 – P_{out} versus P_{in} as determined by changing pulse power attenuation. Red boxes denote data from Figure 3.14 for comparison

The validity of this alternative approach was supported by the good agreement with data points acquired by calibrating the PW_{90} at varying attenuation levels.

The relationship was not linear and was likely to represent a superimposition of the response from different components within the signal path and amplification circuits of the spectrometer. As a simple linear relationship could not be established, and as in practice the value of P_{in} required for a desired P_{out} needs to be determined, it was more practical to invert the axis of the graph and fit a polynomial to the curve. This representation of the data is shown in Figure 3.16 with the polynomial fitted in Excel.

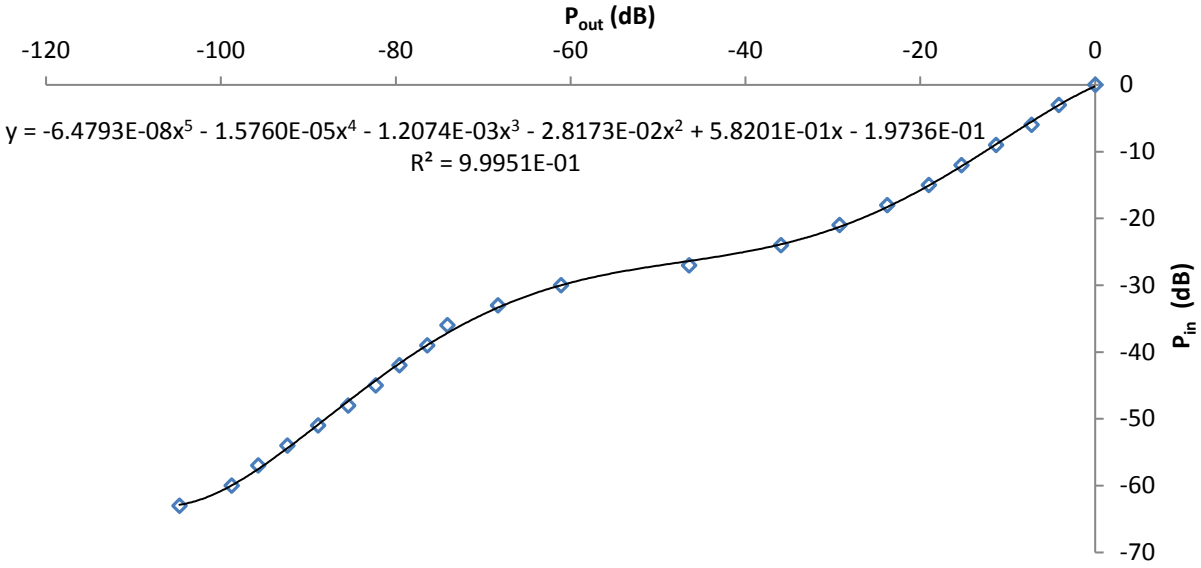


Figure 3.16 – P_{in} versus P_{out} with polynomial fitting

At lower power it can be seen that the amplifier deviates progressively away from linearity, which would be a significant issue for applications requiring accurate pulse powers to be set.

From a plot of predicted – P_{in} versus P_{out} values in Figure 3.17, there is a maximum error of approximately ±1.2 dB when using this equation.

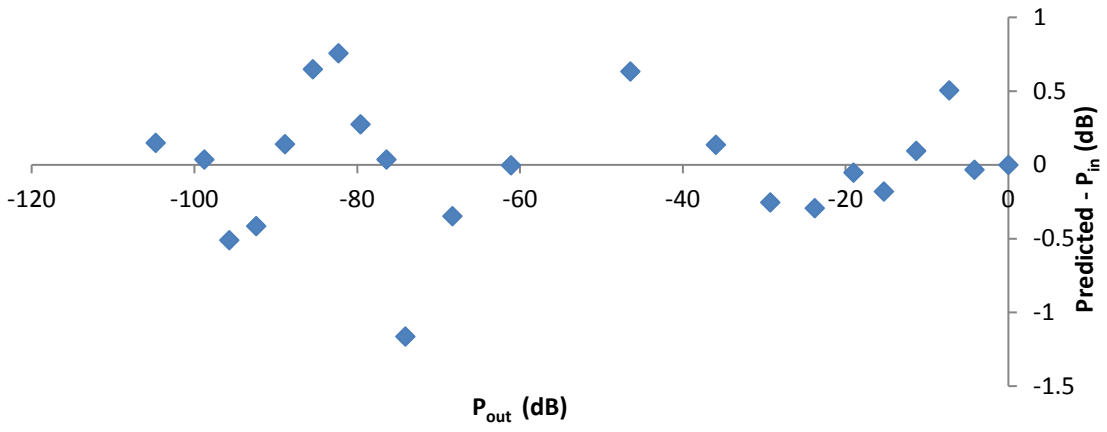


Figure 3.17 – Residual plot from polynomial fitting in figure 3.16

No further interpretation of this function is made here (and indeed may not be realistically possible given the unknown details of the signal path) and the purpose is simply to provide a practical correction factor to allow accurate pulse powers to be set over a wide range of attenuation.

No evidence was found in the literature of amplifier response for the Spinsolve (or any other benchtop NMR spectrometer) with the level of details described here, nor has any detail of an approach to mitigate the issue as shown here.

3.3.4. Receiver gain response

Absolute peak intensity as a function of receiver gain is shown in Figure 3.18.

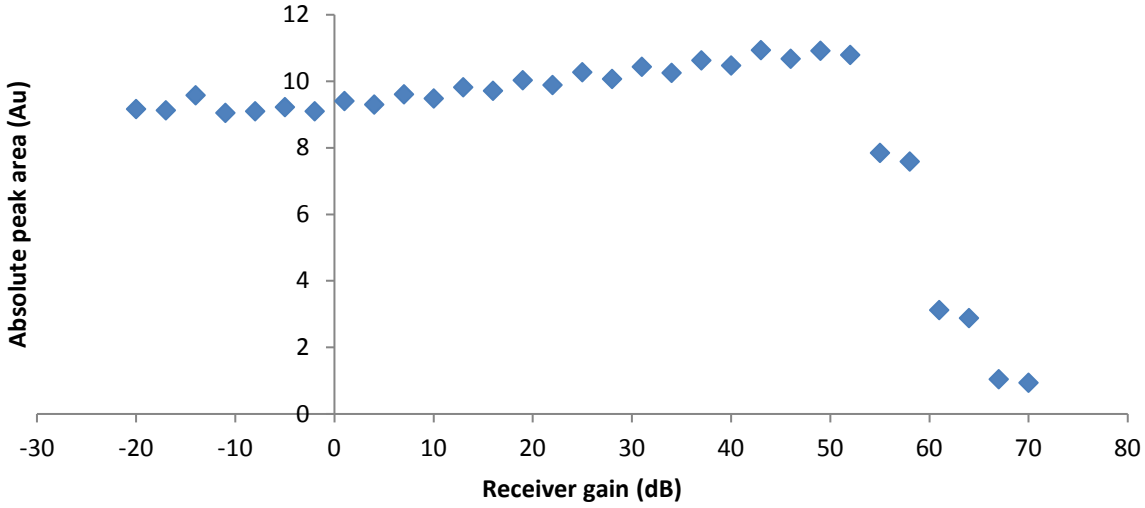


Figure 3.18 – Plot of absolute peak area versus receiver gain

The data suggests an approximately linear trend over the range -20 to 50 dB and a sharp drop in intensity thereafter. This is expected as inspection of data over 50 dB showed the main peaks to have truncation artifacts arising from signal clipping of the FID. In contrast to the high field data shown in Figure 2.8 however, the data here shows only a small relative change in peak area over the range of gain values.

The SNR values were plotted versus gain, with a more marked change noted as shown in Figure 3.19.

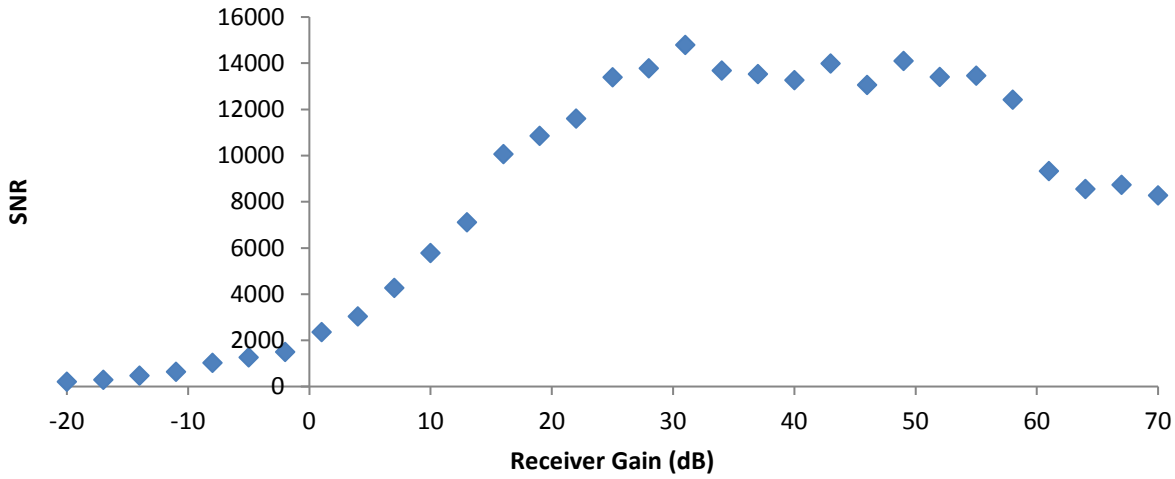


Figure 3.19 – Plot of SNR versus receiver gain

This response appeared to conflict with the data in Figure 3.18, but can be rationalised by looking at the signal and noise component of these values in Figure 3.20, which shows a relatively small increase in signal, but a significant decrease in noise with increasing gain.

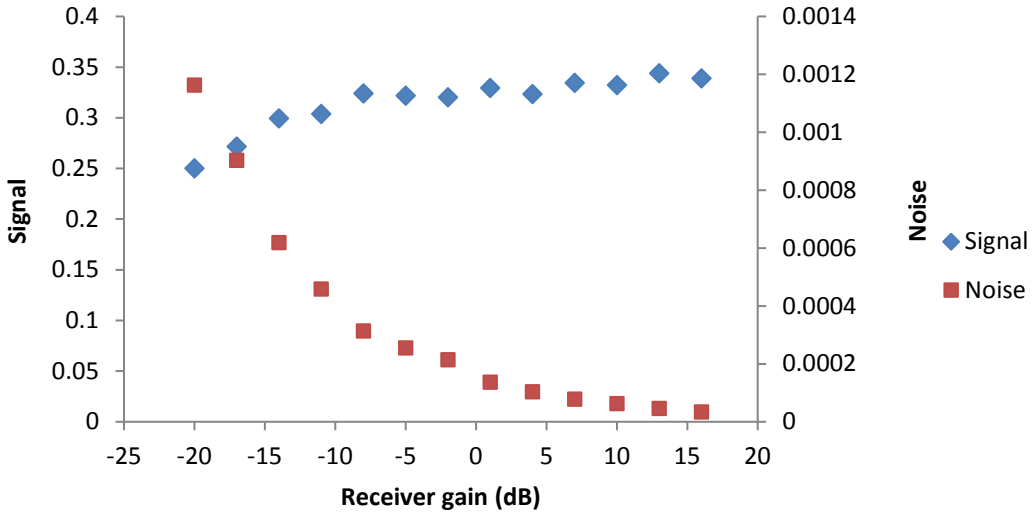


Figure 3.20 – Plot of signal and noise values versus receiver gain

Closer inspection of the data as plotted in Figure 3.18 showed an apparent oscillation of the absolute peak area as the gain increased in integer values of dB. The odd values of RG gave greater values of Au than the even values. The source of this oscillation is not known and may be hardware, firmware or software related.

The trends shown in Figure 3.20 are rationalised by information supplied by the manufacturer stating that for data acquisition using Spinsolve Expert software the data are scaled by the receiver gain when saved. This does not occur with the standard Spinsolve software. As an additional note, all the data shown for the receiver linearity experiments are from single scan spectra – for multiple scans, both software packages average the data from the number of transients rather than adding them together. To reconstruct the total absolute area from multiple scan experiments, the averaged area must therefore be multiplied by the number of scans.

It is again noted that there is no evidence in the literature of this level of detail on receiver gain response as shown here.

3.3.5. Selective pulses

One application which requires the accurate control of low amplifier powers is the synthesis of shaped pulses. In accordance with Heisenberg's uncertainty principle, for a pulse width of Δt:

$$\frac{1}{\Delta t} = \Delta \nu \qquad \textbf{(3.26)}$$

where Δν is the uncertainty in the applied frequency, approximately equivalent here to the bandwidth of excitation.

For a non-selective rectangular pulse experiment, the pulse is applied for a period on the order of ~10 μs, which is expected to result in a bandwidth of ~100 kHz (*ca*. 23

ppm on a 43 MHz NMR spectrometer). Due to the short period of the pulse, a relatively high power can be applied to achieve the desired flip angle (usually 30 - 90°). Conversely, if only a discrete portion of the spectrum was to be excited, for example 20 Hz, then a pulse width of around 50 ms would need to be applied. This would then require a much lower power to achieve the same flip angle.
To achieve an accurate shaped pulse, the correction model for the amplifier, as discussed in 3.3.3, was incorporated into a new pulse programme script (see eSuppl 2).

The scripting language for the pulse programme only appears to allow 50 values to be defined for a shaped pulse. The Gaussian pulse shape as defined using Topspin's Shapetool was stored in an array of normalised values called "s" within the script. The normalised "s" values were then converted into an array of dB values, "r," using the following equation:

$$r = \left(20 \times log_{10}(s)\right) + a90Amp \quad \textbf{(3.27)}$$

where a90Amp is the maximum power for the pulse as set in the user interface. These values are then corrected for amplifier non-linearity using the 5th order polynomial in Figure 3.16 and stored in an array "q". Although the user interface uses log scale dB values, when the pulse program is run, the spectrometer takes a series of linear values to set the amplifier power. Hence the dB array, "q" is finally converted into a linear scale in an array called "t1" (by convention of the software) with the following relationship:

$$t1 = 2^{14} \times 10^{\frac{q}{20}} - 1 \quad \textbf{(3.28)}$$

The "shapedrf" command then uses the "t1" values to modulate the pulse amplitude during the experiment. Figure 3.21 shows a comparison of a single scan pulse-and-acquire experiment with a hard rectangular pulse, and a selective pulse with and without correction. The selective Gaussian pulse has a 150 Hz bandwidth.

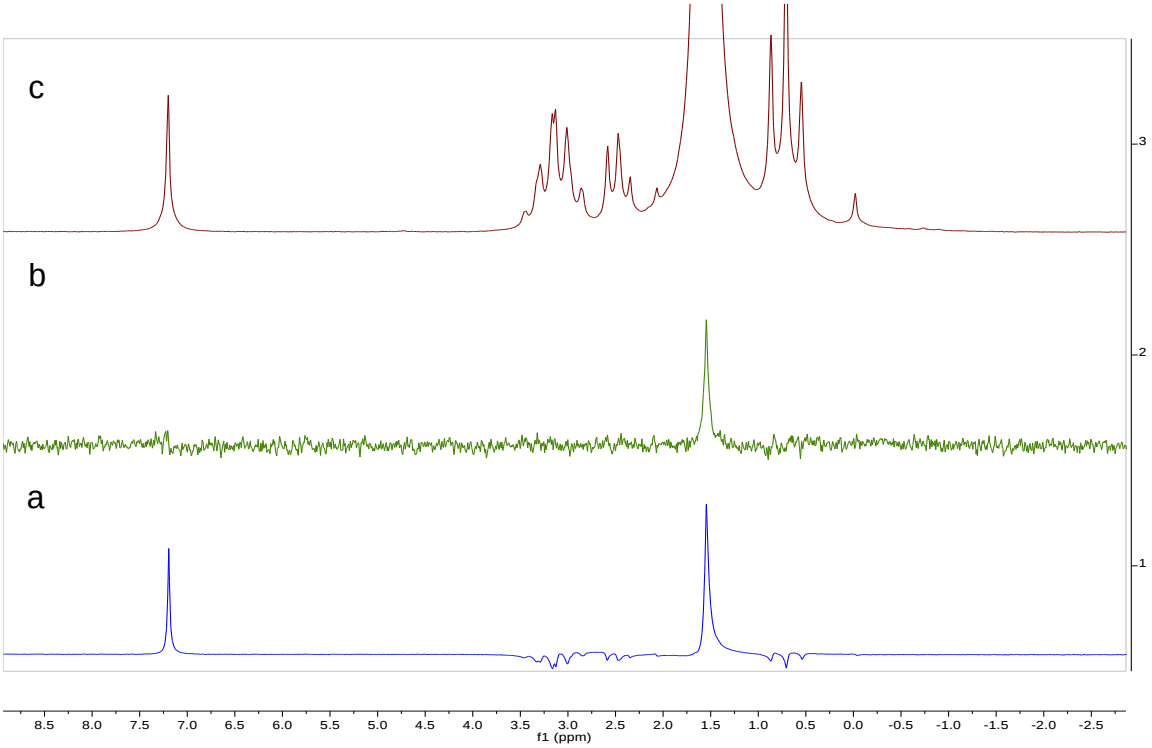


Figure 3.21 – Comparison of spectra of acetonitrile/ethanol/chloroform acquired with a) Gaussian shaped pulse (corrected) b) Gaussian shaped pulse (uncorrected), c) Rectangular hard pulse. The shaped pulse was centred on the chloroform resonance at ca. 7.2 ppm.

Practically no peak intensity is observed for chloroform with the uncorrected shaped pulse, such is the loss in signal. The signal to noise is significantly improved with the addition of amplifier power correction and a difference in peak height of *ca.* × 50 between chloroform and acetonitrile has been reduced to an approximate ratio of 1:1.

The initial application for accurate shaped pulse synthesis was to remove the solvent signal from solutions of weak analyte concentration. Doing so would allow the RG to be increased without clipping the intense solvent signals, and the weaker analyte signals would then fill more of the dynamic range of the ADC and potentially lead to increases in SNR. This is a common practice in high field NMR data, where FID signal to noise is relatively low and may be below the intensity of the smallest ADC

bit in the presence of high intensity solvent signals. However, at lower field, NMR signals are much less intense and in practice noise and high intensity peaks are usually both above the value of the lowest bit and so no additional advantage can be gained.

Other applications exist for shaped pulses including pulse sequences where single resonances are to be excited and the magnetisation subsequently manipulated to assist in assignment of complex spectra, or for the purposes of solvent suppression in field gradient experiments. For example, in the review and comparison of solvent-suppression methods in [21]; for characterisation of nano-emulsions using diffusion-based methods by D'Agostina *et al*. [29]; or in Ultrafast 2D experiments used in reaction monitoring [30]. Note that these, and other, literature references mention the ability of the Spinsolve spectrometer and software to produce shaped pulses and incorporate them into routine pulse sequences, but do not give further details on how this is achieved or how the issue of non-linearity is addressed. It may be that more recent iterations of the hardware and software of this, or other, benchtop spectrometers have automated the process, but this has not been explicitly recorded in the available literature.

3.3.6. Lineshape

As discussed, the main compromise encountered at low field, alongside SNR, is the decreased signal dispersion. At 43 MHz, a spectral window of 1 ppm represents 43 Hz, rather than the 400 Hz it would represent on a high field 400 MHz NMR spectrometer. One implication is that, given equivalent linewidth and coupling, signal overlap is more likely compared to high field data. If discrete peaks are unavailable for simple integration, then line fitting algorithms must be used to obtain relative peak areas by a process of deconvolution.

These fitting algorithms will assume the peak lineshapes conform to certain mathematical functions, with the complexity of the available functions being

dependent upon the processing software used. The FID of a single pulse ^{1}H NMR experiment is approximately exponential, which results in a Lorentzian lineshape after Fourier transformation. The FID may deviate from an ideal exponential due to factors such as field inhomogeneity or the application of a weighting function. In the case of high field NMR data, this would typically yield a peak with part Lorentzian and part Gaussian components, as exemplified in Figure 3.22.

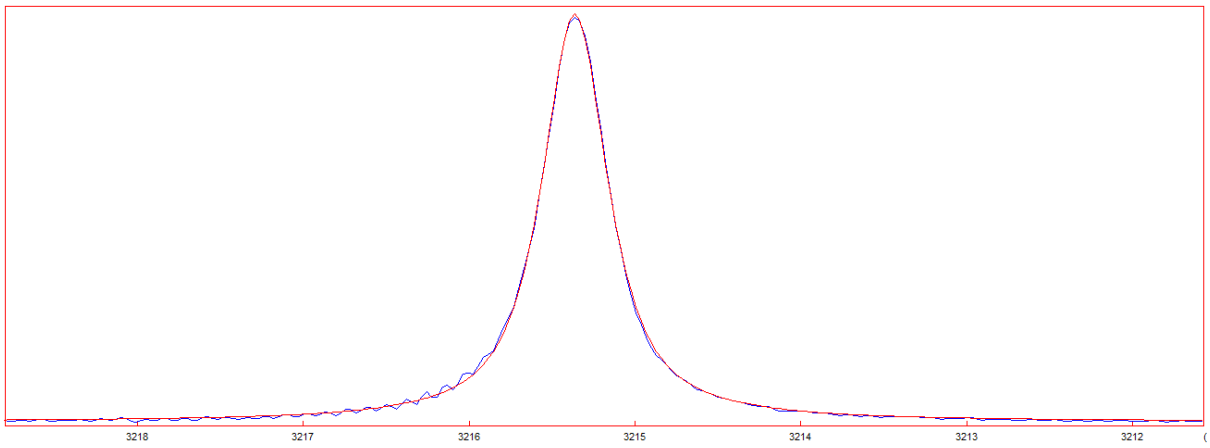


Figure 3.22 – Deconvolution of a ^{1}H singlet at 400 MHz

This peak can be fitted with a single Gauss/Lorentz function with a HHLW of 0.48Hz and a G/L value of 0.12, where:

$$f(x) = \frac{G}{L} \times \frac{x}{1-x} \tag{3.29}$$

Hence the lineshape is predominantly Lorentzian in character. Compare this with a typical singlet from a well-shimmed magnet at 43 MHz as shown in Figure 3.23.

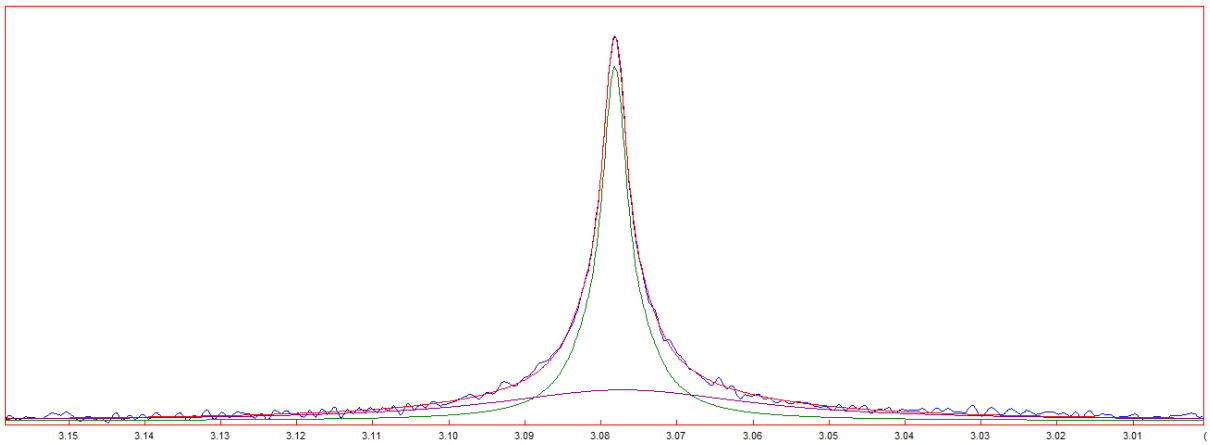


Figure 3.23 – Deconvolution of a ^{1}H singlet at 43 MHz

Now the peak requires two lines to be fitted, both with G/L value of 0, *i.e.* purely Gaussian functions. These have HHLW of 0.47 and 4.10 Hz and represent 57 and 43% of the total peak area, respectively. As discussed above, although the width of the narrower peak is comparable with the high field data, it will still appear broader in the spectrum at lower field. Also, there is a significant component of the peak that has a very broad lineshape, which from discussion with the manufacturer is due to inhomogeneity at the edges of the NMR tube within the active volume. This second broad component poses issues due to the large tails that are present in spectra from intense solvent peaks.

Lineshape can be expressed relative to the carrier frequency of the spectrometer. Determining the T_2 of the peaks via a CPMG experiment (see 2.4.2) can then determine the contribution from field inhomogeneity. For the spectra acquired here, at 400 MHz, a HHLW of 1.2 ppb was achieved, of which 1.1 ppb was due to field inhomogeneity (where the ratio of the HHLW to the spectrometer ^{1}H carrier frequency, in Hz, is expressed as parts per billion). In contrast, at 43 MHz a HHLW of 12.0 ppb was achieved, of which 8.5 ppb was due to field inhomogeneity. This demonstrates the lower relative resolution expected with low field bench top spectrometers.

3.4. Conclusions

In this chapter a set of detailed information pertaining to the capabilities and operation of the Magritek Spinsolve spectrometer have been collated and summarised from manuals, experimentation and conversations with the vendor company. New knowledge has been presented and practical, novel approaches not currently described in the literature have been described.

It has been shown that a basic regime of shimming maintains acceptable data quality in terms of lineshape and SNR, and that these properties are limited by the spectrometer design. The interdependency of flow rate and flow cell position and their effect on SNR and T_1 have been investigated. The results of these experiments have been interpreted to estimate flowcell volumes and determine optimum flowcell position for subsequent use. This information is consistent with current literature and also adds additional information pertinent to the current setup of the NMR spectrometer and it's intended use here.

It has also been shown that the pulse amplification and signal receiver stages are non-linear. The non-linearity has been measured and can now be corrected on-the-fly using a custom pulse sequence, the utility of which has been demonstrated by the application to allow accurate synthesis of selective shaped pulses. This correction process has not been currently published anywhere as far as can be reasonably ascertained and thus can be considered a novel approach.

These data and conclusions are fundamental in the everyday use of the spectrometer and will inform future operation of the system. Specifically, the information on linewidths at low field and the understanding of RG and scaling using the Spinsolve software is utilised in the development of quantitative methods as exemplified in the next chapter.

Examples where more recent versions of hardware and software have bridged some of the gaps since the experiments were carried out here have been noted, such as the more wider incorporation of shaped pulses into experiments. However, the details as how these have been achieved are still absent from available literature.

References

[1] The Eft-60 & Eft-90 NMR Spectrometers, https://aiinmr.com/high-quality-durable-nmr-spectrometers-pmnmr-instruments/, Retrieved 12 Dec 2018.

[2] Nanalysis, NMReady 60, http://www.nanalysis.com/nmready60p.html, Retrieved 22 Jan 2019.

[3] Bruker, Minispec, http://www.bruker.com/products/mr/td-nmr/minispec-mq-series.html, Retrieved 22 Jan 2019.

[4] Thermo Scientific, Picospin 80, http://picospin.com/products/picospin-80/, Retrieved 22 Jan 2019.

[5] Thermo Scientific, PicoSpin 45, http://picospin.com/products/picospin-45/, Retrieved 22 Jan 2019.

[6] Resonance Systems, Spin Track-60, http://www.nmr-design.com/ft_nmr_spectrometer.html, Retrieved 22 Jan 2019.

[7] Oxford Instruments, Pulsar (60MHz), https://nmr.oxinst.com/products/nmr-spectrometers/pulsar, Retrieved 08 Feb 2019.

[8] Magritek, Spinsolve (43, 60 , 80 MHz), http://www.magritek.com/products-spinsolve, Retrieved 22 Jan 2019.

[9] K. Meyer, S. Kern, N. Zientek, G. Guthausen and M. Maiwald, Process control with compact NMR. TrAC Trends in Analytical Chemistry, 83 (2016) 39-52.

[10] K. Singh and B. Blümich, NMR spectroscopy with compact instruments. TrAC Trends in Analytical Chemistry, 83 (2016) 12-26.

[11] K. Singh and B. Blümich, Online monitoring of the kinetic isotope effect in chemical reactions with 1H and 19F low-field NMR spectroscopy. Analyst, 143 (2018) 4408-4421.

[12] N. Zientek, C. Laurain, K. Meyer, M. Kraume, G. Guthausen and M. Maiwald, Simultaneous 19-1H medium resolution NMR spectroscopy for online reaction monitoring. Journal of Magnetic Resonance, 249 (2014) 53-62.

[13] E. Danieli, J. Perlo, A. L. L. Duchateau, G. K. M. Verzijl, V. M. Litvinov, B. Blümich and F. Casanova, On-Line Monitoring of Chemical Reactions by using Bench-Top Nuclear Magnetic Resonance Spectroscopy. ChemPhysChem, 15 (2014) 3060-3066.

[14] P. Sagmeister, J. D. Williams, C. A. Hone and C. O. Kappe, Laboratory of the future: a modular flow platform with multiple integrated PAT tools for multistep reactions. Reaction Chemistry & Engineering, 4 (2019) 1571-1578.

[15] F. Dalitz, M. Cudaj, M. Maiwald and G. Guthausen, Process and reaction monitoring by low-field NMR spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy, 60 (2012) 52-70.

[16] D. Cortés-Borda, E. Wimmer, B. Gouilleux, E. Barré, N. Oger, L. Goulamaly, L. Peault, B. Charrier, C. Truchet, P. Giraudeau, M. Rodriguez-Zubiri, E. Le Grognec and F.X. Felpin, An Autonomous Self-Optimizing Flow Reactor for the Synthesis of Natural Product Carpanone. The Journal of Organic Chemistry, 83 (2018) 14286-14299.

[17] S. Kern, K. Meyer, S. Guhl, P. Gräßer, A. Paul, R. King and M. Maiwald, Online low-field NMR spectroscopy for process control of an industrial

lithiation reaction—automated data analysis. Analytical and Bioanalytical Chemistry, 410 (2018) 3349-3360.
[18] S. Kern, L. Wander, K. Meyer, S. Guhl, A. R. G. Mukkula, M. Holtkamp, M. Salge, C. Fleischer, N. Weber, R. King, S. Engell, A. Paul, M. P. Remelhe and M. Maiwald, Flexible automation with compact NMR spectroscopy for continuous production of pharmaceuticals. Analytical and Bioanalytical Chemistry, 411 (2019) 3037-3046.
[19] I. J. Lowe and C. E. Tarr, A fast recovery probe and receiver for pulsed nuclear magnetic resonance spectroscopy. Journal of Physics E: Scientific Instruments, 1 (1968) 320-322.
[20] E. Fukushima and B. W. Roeder (1981), NMR Hardware - Quarter wave lines and a quarter wave network. Experimental Pulse NMR - A nuts and bolts approach, Addison Wesley, Massachusetts, 400-407.
[21] B. Gouilleux, B. Charrier, S. Akoka and P. Giraudeau, Gradient-based solvent suppression methods on a benchtop spectrometer. Magnetic Resonance in Chemistry, 55 (2017) 91-98.
[22] J. F. Haw, T. E. Glass and H. C. Dorn, Conditions for quantitative flow FT-^{1}H NMR measurements under repetitive pulse conditios. Journal of Magnetic Resonance, 49 (1982) 22-31.
[23] K. Albert (2002), LC-NMR: Theory and Experiment. On-line LC-NMR and Related Techniques, John Wiley and Sons, Chichester, 1-22.
[24] J. L. Sudmeier, U. L. Günther, K. Albert and W. W. Bachovchin, Sensitivity Optimization in Continuous-Flow FTNMR. Journal of Magnetic Resonance, Series A, 118 (1996) 145-156.
[25] F. Dalitz, L. Kreckel, M. Maiwald and G. Guthausen, Quantitative Medium-Resolution NMR Spectroscopy Under Non-Equilibrium Conditions, Studied on the Example of an Esterification Reaction. Applied Magnetic Resonance, 45 (2014) 411-425.
[26] G. J. Krüger, A. Birke and R. Weiss, Nuclear magnetic resonance (NMR) two-phase mass flow measurements. Flow Measurement and Instrumentation, 7 (1996) 25-37.
[27] T. D. W. Claridge (1999), Practical aspects of high-resolution NMR. High-Resolution NMR Techniques in Organic Chemistry, Pergamon, Oxford, 45-110.
[28] P. J. Hore (2002), The Vector Model. Nuclear Magnetic Resonance, Oxford University Press, 74-77.
[29] C. D'Agostino, V. Preziosi, A. Khan, M. Mantle, E. Fridjonsson and S. Guido, Microstructure evolution during nano-emulsification by NMR and microscopy. Journal of Colloid and Interface Science, 551 (2019) 138-146.
[30] B. Gouilleux, B. Charrier, S. Akoka, F.-X. Felpin, M. Rodriguez-Zubiri and P. Giraudeau, Ultrafast 2D NMR on a benchtop spectrometer: Applications and perspectives. TrAC Trends in Analytical Chemistry, 83 (2016) 65-75.

Chapter 4

Automated Quantitation by Bench Top NMR

4.1. Introduction

4.1.1. Project aim

In this Chapter, the application of bench top NMR to rapid, atline analysis for process support by synthetic chemists will be explored and different approaches to quantitation investigated. In approximate order of complexity, these are: sum integration, deconvolution, external standardisation and multivariate analysis. The automation of the processing steps *via* scripting within existing processing software will be discussed alongside each approach. Finally, the design and coding of a single customised user interface developed to control the current acquisition and processing software will be considered.

Each approach to quantitation will be presented and explored in the context of a real-world example with the eventual aim of producing a comprehensive at line tool designed for ease of the non-expert user.

4.1.2. Background

4.1.2.1. Online versus off-line

Despite recent advances in PAT and evidence in the literature for an increasing number of techniques being adapted for online analysis, many of the analyses that take place to monitor and control processes in the pharmaceutical industry are still conducted offline. Processes are diverted, stopped, or otherwise held-up whilst samples are offloaded from the vessels, transferred to analytical laboratories, prepared, loaded onto instrumentation, analysed and then the data processed and recorded to render the final information required.

The ideal solution would be to implement online technology capable of continually measuring a given parameter throughout the process runtime, but this is not always feasible. On line instruments generally have more limited capability compared to their offline counterparts, not least in the requirements to analyse a smaller sample size, often in a flowcell and without any form of sample preparation. It is possible to automate sampling of a process and subsequent sample preparation, for example with on line HPLC, but this then necessitates additional valves, flow lines and other equipment to be installed, adding to the complexity of the process.

An approach that offers a compromise between online and current offline analysis is to find ways of reducing the time between sampling and receiving the desired information using offline techniques. This can be through a combination of choosing the right techniques, developing better, more efficient interfaces through method validation and software customisation and making instruments portable so that they can be co-located with ongoing processes.

The Magritek Spinsolve spectrometer is small enough that the spectrometer and PC can fit on a mobile bench, which can then be placed temporarily next to a chemical

process in a laboratory. The internal field lock and robust shimming also offer opportunities to develop rapid, bespoke analytical tools for the process chemist.

4.1.2.2. Quantitation

A number of key aspects in quantitation by NMR have already been introduced in preceding chapters. These include the importance of acquiring data under quantitative conditions, a summary of the acquisition parameters that need to be considered and validation protocols to determine the reliability of the results produced under given circumstances.

One particular aspect not yet explored is the process by which numerical data are extracted from NMR spectra which, in the examples shown here, are in the form of peak areas. The numerical data are then fed into a series of calculations to take into account parameters such as number of equivalent protons, molecular weight and density to yield a result in the format required for the user *e.g.* % w/w, % v/v, mg/mL and so forth.

There are many approaches for the initial step of yielding quantitative data. Four of the most common approaches are sum integration, deconvolution, external standardisation and multivariate analysis (MVA). Another common approach not discussed here is the use of internal standards where a well characterised standard of known concentration and purity is spiked into a known volume of the analyte solution. Since this requires accurate weights and volumes of materials to be dispensed at the point of analysis, and the introduction of a potentially interfering species into the prepared sample of analyte, it is not considered an ideal solution for developing rapid at line testing.

4.1.2.2.1. Sum integration

In an ideal case, the resolution of the ^{1}H NMR spectrum is such that discrete peaks can be integrated for all components of interest. This can be most simply achieved by sum integration, equivalent to adding the intensity of the data points in the real spectrum over a given range.
If peaks are not baseline resolved, the intensity of any one point on a peak may be affected if the peak sits on the tail or shoulder of another more intense peak leading to inaccuracies in measuring peak area. If the overlap is not excessive, it can be taken into account by a baseline correction conducted prior to integration, or by baseline correction during the integration process as set using the common “slope” and “bias” parameters interactively.

If the number of data points acquired in the FID are such that there are insufficient data points defining a peak in the real FT spectrum, then it will also not be possible to accurately integrate the peak. Additional zero values can be added to the FID before Fourier transformation (zero filling). These do not affect the data other than to increase the number of points in the resulting real spectrum as shown in Figure 4.1, allowing more accurate integration.

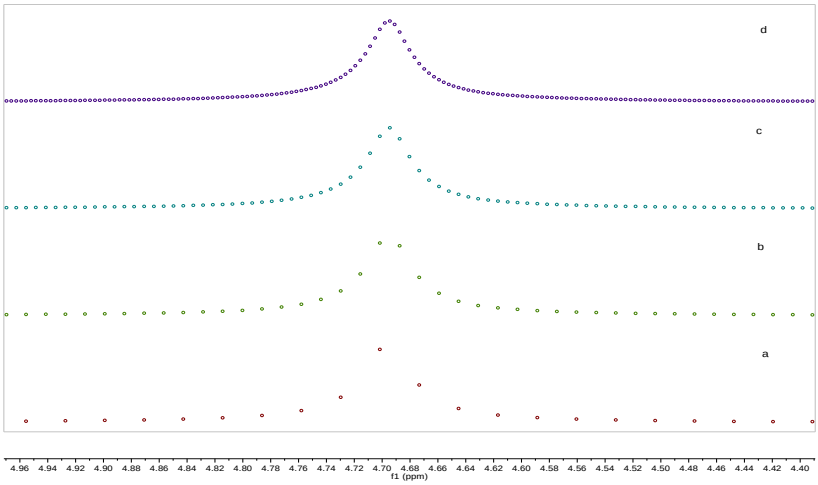


Figure 4.1 – Example peak with a) no zero filling (16 k points total), and zero filling to give a total of b) 32 k points, c) 64 k points, d) 128 k points.

Subsequent experimental sections will demonstrate this approach as applied to three different examples.

In the first example the final drug substance producing step from a multi-stage process resulted in a liquor containing the drug molecule of interest in a tertiarybutylmethylether (TBME). To produce the correct polymorphic solid form, the drug substance needed to be recrystallised from acetic acid, which was achieved by adding acid to the TBME liquours and distilling off the TBME. Rapid analysis was required to determine the point at which the residual TBME reached an allowable level and the distillation could be stopped. Benchtop NMR was investigated as a possible analytical solution. The analysis also needed to be able to cope with the presence of water as a contaminant from the acetic acid.

In the second two examples, new software was being investigated by an engineering team for the purposes of modelling distillation in pharmaceutical manufacturing process. To evaluate the software, distillations of a number of binary mixtures of solvents were planned. These required repeated analysis to determine the solvent ratios at time points during the distillation. Atline benchtop NMR methods were developed and used for two of these binary mixtures – although the data from the subsequent software modelling could not be included here.

4.1.2.2.2. Deconvolution

If baseline resolution of the peaks of interest cannot be achieved, then it may be possible to model the peaks with a series of idealised mathematical functions, as introduced in Chapter 3. Applied over an entire spectrum, this process can be used to calculate the total integral of regions of partially overlapped peaks to allow quantitation of species represented in that spectrum.

Depending upon the software used, initial parameters of the peaks (chemical shift, intensity, linewidth, Gauss/Lorentzian function) are either set by the user or the

software. An automated iterative process then takes place where the total model of all peaks is compared to the spectrum obtained and parameters are varied until a defined end point is reached. An example of this is the Global Sample Deconvolution (GSD) function available within MestreNova software.

Care must be taken as the iterative algorithms are designed only to reach the end point for the whole spectrum, usually a minimum residual. This can result in a series of peaks being fitted to the spectrum with parameters that do not represent the true areas of individual peaks. Examples of this are often found in the case of heavily overlapped regions, or overlap with very intense peaks, where additional large, broad resonances are fitted in an attempt to achieve the best residual at the expense of smaller peaks. In some cases parameters can be constrained by the user to force peaks to always be included in the calculation with specific locations and widths.

The enhanced capability provided by deconvoluting more complex spectra will be demonstrated by quantitation of analyte reaction concentration in a reaction liquor. The example liquors were taken from the initial stage of multi-step process and contain a reaction product in the form of an unstable organic molecule, dissolved in THF in the presence of process impurities and hexane. The reaction product concentration was required to determine the quantity of reagents to be charged in the next chemical step. Dilution of the liquor to allow HPLC analysis was not possible due to the instability of the reaction product in most HPLC solvents and so methods that involved direct analysis of the neat liquors were investigated, including benchtop NMR as discussed here.

4.1.2.2.3. External standardisation

As ^{1}H NMR spectra acquired at low field become more complicated, the situation may arise whereby only a single resonance for the species of interest is resolved well enough to integrate accurately and all other resonances are overlapped. At this level of complexity, it may not be possible to know exactly the identity of the

resonances that are in an extensively overlapped spectral region as, for instance, there may be an unknown contribution from, for example, water, exchangeable protons or impurities. In these circumstances it would still be possible to quantify the content of the species of interest by using an external standard.

This approach is often used for HPLC assays, where solutions of known concentration are prepared from a batch of well-characterised material and are analysed immediately before and after prepared solutions of the sample. The average response factor for the standards in terms of unit area per quantity of standard (taking into account standard purity) is then calculated and applied to determine the amount of material present in the sample solutions. By running the standards before and after the samples, any variation in instrument response is also accounted for.

Standards may be the same compound as the analyte, but a different batch. This has the advantage of avoiding issues of differing response factors. Alternatively, standards may also be of a different compound to the analyte, in which case conditions that effect both standard and analyte response factors must also be explored.

The compound that will be used as a basis for investigating external standardisation in benchtop NMR is the alkylating reagent hexyllithium as supplied by the manufacturer in hexane solution. Chemists using this reagent are supplied with a certificate of analysis that includes a solution concentration, but an orthogonal test was required to confirm the concentration at the point of use. One of the reasons behind the requirement for additional testing was the common practice of using older batches that had been previously part-used, for development work. Hexyllithium concentration was likely to change over time and with exposure to moisture, which results in the conversion of hexyllithium to hexane and lithium hydroxide. As with other examples in this chapter, the inherent instability of the solution prevented the use of many common analytical methods and hence the possibility of determining

concentration using benchtop NMR as a non-invasive approach was to be investigated.

4.1.2.2.4. Multivariate analysis

In more extreme cases of ^{1}H NMR spectral complexity, it may be that there are no fully resolved peaks in the spectrum and/or ambiguity in the identity of peaks within integratable regions meaning that the approaches discussed so far cannot reliably yield quantitative data.

A solution to this scenario is to determine a correlation, or correlations, between an observed change, or changes, in the spectrum and a change in the sample. It would not be necessary to fully characterise the change in the spectrum as arising from a specific resonance. Although there are obvious inherent risks to a completely “black box” approach, it can potentially allow a degree of quantitation to be made from an otherwise inaccessible spectrum.

Multivariate analysis (MVA) covers a number of different processes by which these correlations can be determined. These will be explored further and demonstrated in the work shown in Chapter 5.

4.1.2.3. User interface and automation

The ultimate aim of this project was to build a single user interface that would allow a non-expert user to obtain the desired analytical results with the minimum amount of interaction. To achieve this, the different options for data acquisition and processing on the Magritek Spinsolve spectrometer and software were considered.

Data can be acquired either through the Spinsolve software interface, which was designed to allow common experiments to be readily accessible to the non-expert user, or through the Expert Software interface. This was originally designed for

manufacturer instrument development, but has been made available to more advanced users. Both of these interfaces are based on the Prospa platform developed by Magritek for their imaging instruments and eventually their NMR spectrometers.

Some basic processing can be performed within the Spinsolve software, and more advanced processing is accessible in the Expert interface. MestreNova is considered one of the more popular NMR data processing tools available to NMR users, alongside Bruker's Topspin and ACD Spectrus. The latter two packages can each open all common file types and perform most calculations and processing procedures required.

In order to expedite development of the application and utilise as much capability from the Spinsolve and MestreNova software platforms, a third platform was developed in the Python language [1] to control the other two processes and provide a single user interface.

The Spinsolve software can be controlled remotely through a web socket interface using a Transmission Control Protocol – Internet Protocol (TCP-IP) protocol to transfer messages in xml. This allows experiments and associated parameters to be chosen and run and the location of data files and their names to be freely specified. As a result of this latter aspect, controlling the Spinsolve software remotely is more flexible than running the spectrometer directly through the software interface itself. Any actions that can be performed within the MestreNova GUI can be automated *via* scripts (written in the JavaScript programming language), as discussed in preceding sections. MestreNova contains its own script interpreter, where data are handled as objects and the processing actions as object methods. MestreNova can, in turn, be controlled remotely *via* the command line, which allows files to be opened, processed with specific scripts and the data saved under different file formats.

4.2. Experimental

All data were acquired on the Magritek Spinsolve NMR spectrometer, using Spinsolve version 1.11.3 with built-in experiments as described in Chapter 3. All data were processed using MestreNova version 10.0.2 and methods were automated using a custom programme as described in the results and discussions sections. All samples and solutions were dispensed into 5 mm diameter NMR tubes for analysis. All solvents, including water, are non-deuterated.

The structures and associated reaction schemes for the organic compounds, referred to as **1** and **2** in the following sections have not been included in this current report.

The specific details of individual sample preparations are detailed at the start of each sub-section of the result and discussion section.

4.3. Results and discussion

4.3.1. Quantitation by sum integration

The final stage of a chemical process as described in the introduction was investigated. The reaction solution contained an organic molecule (**1**), along with acetic acid, water and tertiarybutylmethylether (TBME). The solution underwent a subsequent distillation to remove TBME, where a target TBME : acetic acid molar ratio of < 0.119 : 1 was to be achieved. Prior work had shown this to be the upper limit of TBME that would result in final drug substance of an acceptable quality from the subsequent crystallisation step. The complete structure of **(1)** is not included here, but it will be stated that it contains a isopropyl moiety as well as aromatic and olefinic moieties.

4.3.1.1. Sample preparation

The following reference samples were prepared and dispensed into NMR tubes:

- Approximately 20 mg of **1** was dissolved in approximately 1 mL of acetic acid-d_4
- A few drops of TBME were mixed with approximately 3 mL of acetic acid
- Neat TBME, acetic acid and water

^{1}H NMR reference spectra were acquired on each sample using the QuickScan experiment (see Chapter 3). T_1 inversion recovery experiments were performed on each sample with 2 scans, acquisition time of 6.4 s, repetition time of 30 s, maximum inversion recovery delay of 15 s over 11 steps.

To assess linearity, a series of solutions were prepared containing TBME at molar ratios of TBME : acetic acid of 0, 0.015, 0.038, 0.061, 0.119, 0.179, 0.265 : 1. These correspond to TBME concentrations of 0, 3.0, 7.1, 10.9, 19.1, 26.0, 34.0% v/v in acetic acid and each solution was prepared such that they also contained water at a constant concentration of 4.0% v/v. Two aliquots from each solution were dispensed into 5 mm NMR tubes. Each solution was run with a single scan, acquisition time of 6.4 s, pulse angle of 90° and a repetition delay of 15 s (determined as adequate for full relaxation from the above T_1 experiments).

To assess repeatability, ^{1}H NMR data were acquired 15 times sequentially using the above parameters on the linearity solution at the TBME : acetic acid ratio of 0.119 : 1 as prepared previously.

Accuracy was assessed by preparing a solution containing TBME : acetic acid at a ratio of 0.119 : 1 as above (including water at 4.0% v/v). 1 mL of this solution was added to approximately 250 mg of **1**. ^{1}H NMR data were acquired using the same

parameters as for the linearity experiments. T_1 data were acquired using the same parameters as for the individual components above.

Data were processed using a custom script (see eSuppl 3) with the integrals as detailed in the results and discussion section. The FID was zero filled once, apodised with an exponential function of 0.3 Hz, and the FT spectrum manually processed with fixed zero and first order phase corrections and automatic baseline corrected using a Bernstein 3rd order polynomial.

4.3.1.2. Discussion

^{1}H NMR spectra of samples containing acetic acid, water, TBME and **(1)** are shown in Figure 4.2.

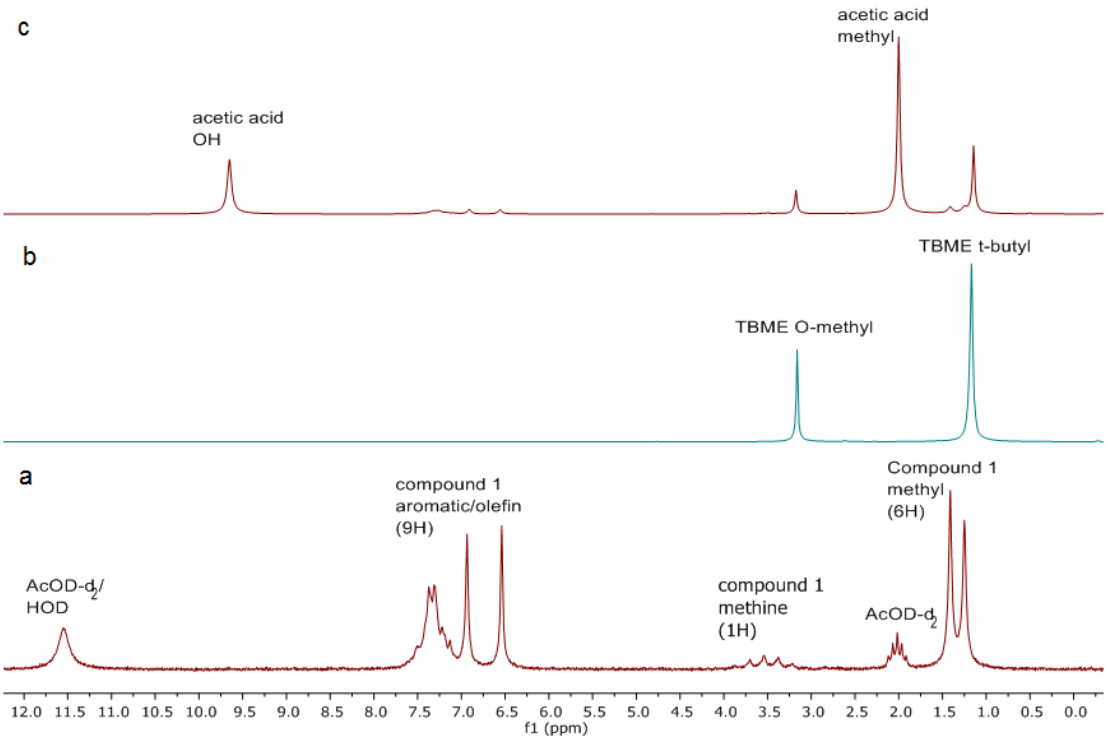


Figure 4.2 – 43 MHz ^{1}H NMR spectra of a) 1 in acetic acid-d$_4$, b) Pure TBME, c) A solution of 1 and TBME in acetic acid

Clearly resolved peaks were observed for acetic acid and **1** with overlap between the methyl peaks of **1** and TBME. The spectrum was integrated in the regions *ca.* 8 – 6

(**1** aromatic/olefin), 2.5 – 1.5 (acetic acid methyl) and 1.5 – 0.5 ppm (**1** methyl and TBME t-butyl). The TBME integral was determined by subtracting the expected integral for the methyl resonance of **1** from the region 1.5 – 0.5 ppm. This was equivalent to subtracting 6/9 of the aromatic for the aromatic region of the ^{1}H NMR spectrum of **1**. The relative integrals were then corrected for the number of equivalent protons they represent and a molar ratio of TBME : acetic acid was then calculated.

T_1 data for pure solutions of water and acetic acid yielded values of 2.3 and 1.8 s, respectively. A pure solution of TBME yielded T_1 values of 3.0 and 2.4 s for the methyl and t-butyl peaks, respectively. T_1 was measured at 1.0 s for **1** in solution in acetic acid with TBME and water present.

Linear regression of experimental *vs.* theoretical molar ratios of TBME : acetic acid yielded a slope of 0.9558, an intercept of 0.0089 and R^2 = 0.9995. A residual plot showed no obvious trends.

15 sequential analyses of the solution containing TBME : acetic acid at a ratio of 0.119 : 1 yielded a %RSD of 5.5%, demonstrating acceptable repeatability. Accuracy was assessed *via* recovery, as calculated from the data acquired on the linearity samples. Recovery of theoretical *vs*. experimental TBME : acetic acid ratio was 80 – 120% for all solutions except the data for the two solutions with TBME : acetic acid ratio of 0.015 : 1, which both gave recoveries of 150%. However, this is considerably below the target ratio of 0.119 required for the process and hence should not significantly affect the resulting TBME : acetic acid ratio. In summary, these data show the method will be acceptable for the required purpose.

Data processing was automated using a script in MestreNova. The acetic acid peak was referenced to 2.0 ppm, and the 3 regions of 8.00 – 6.00 ppm, 2.55 – 1.55 ppm and 1.55 – 0.55 ppm were integrated. The peaks were not observed to move significantly over the range of concentrations investigated here. The peak areas

were read into variables and then manipulated as discussed above to yield the molar ratio which was then displayed on the spectrum. The resulting spectrum with text result was automatically saved under the native MestreNova format (.mnova) and exported as a .pdf rendition and .png image (see Appendix 4).

4.3.2. Quantitation by sum integration – binary mixtures

Methods were developed for the relative quantitation of two binary mixtures of solvents for the purpose of modelling distillations.

4.3.2.1. Sample preparation

Methods were developed for 2 binary mixtures; isopropanol/methanol (IPA/MeOH) and 3-fluoro-2-methylphenyl isocyanate/toluene (FMPI/toluene).

For each method, ^{1}H NMR reference spectra were acquired on each pure component using the QuickScan experiment (see Chapter 3). T_1 inversion recovery experiments were performed on each pure component with 2 scans, acquisition time of 6.4 s, repetition time of 30 s, maximum inversion recovery delay of 15 s over 11 steps.

To assess linearity, the following binary mixtures were prepared:
IPA/MeOH in the ratios of 100/0, 90/10, 75/25, 50/50, 10/90, 0/100% w/w.
FMPI/toluene in the ratios of 100/0, 90/10, 85/15, 75/25, 50/50, 25/75, 10/90, 0/100.
^{1}H NMR spectra were acquired with a single scan, acquisition time of 6.4 s, pulse angle of 90° and a repetition delay of 15 s for 2-MeTHF/MeOH solutions and 30 s for FMPI/toluene solutions. Repetition delays were determined as adequate for full relaxation from the above T_1 experiments. All solutions were analysed in duplicate.

Data were processed using a custom script with the integrals as detailed in the results and discussion section. The FID was zero filled once, apodised with an

exponential function of 0.3 Hz, automated phase corrected and baseline corrected using a Bernstein 3rd order polynomial.

4.3.2.2. Discussion

In cases where the relative contents of only two species are to be calculated it is only necessary to have two discrete integral regions.

Furthermore, these regions do not need to have single well-resolved peaks. They can each encompass a number of overlapping peaks from both species, as long as 1) all the peaks under each region are known and accounted for and 2) the total number of protons represented by each species are different in at least one of the two regions. As long as these conditions are met, then simultaneous equations can be constructed and solved (see Appendix 5).

For the binary mixture of IPA and MeOH, linear regression of experimental % w/w values yielded slopes of 0.9737, and intercepts of 1.033 and 1.5937 for MeOH and IPA, respectively. R^2 = 1.000 and the residual plots showed no obvious trends. All recoveries for MeOH and IPA were in the range 90 – 110% with the exception of one IPA/MeOH 90:10 sample at 112%.

For the binary mixture of FMPI and toluene, linear regression of experimental % w/w values yielded slopes of 0.9773, and intercepts of 1.5766 and 0.6943 for FMPI and toluene, respectively. R^2 = 0.999 and the residual plots showed no obvious trends. All recoveries for FMPI and toluene were in the range 90 – 110%, with the exception of one of the FMPI/toluene 10:90 samples at 111%.

4.3.3. Quantitation by spectral deconvolution

The initial stage of another chemical process was investigated. The reaction liquor of the initial stage of the process contained an organic molecule **2**, synthetic impurities

and hexane in THF. It was necessary to determine the solution concentration of **2** to inform the charging of reagents in the subsequent chemical step: **2** is unstable and hence is not isolated.

4.3.3.1. Sample preparation

A sample of the final reaction liquor for production of **2**, with a concentration of 4.46% w/w (determined by external standard HPLC assay) was supplied. Initial ^{1}H NMR data for a pure sample of the reaction liquor were acquired using the QuickScan experiment. A T_1 inversion recovery experiment was performed with 16 scans, acquisition time of 6.4 s, repetition time of 60 s, maximum inversion recovery delay of 30 s over 11 steps.

Linearity, repeatability, accuracy, DL and QL were assessed by diluting aliquots of the reaction liquor with THF to give solutions of **2** at concentrations of 0.5, 1.2, 2.3 and 3.5 % w/w. Samples were prepared volumetrically in triplicate and weighed by difference with a Mettler 5 place balance along with 3 separate aliquots of the undiluted reaction liquor, resulting in a total of 15 samples.

^{1}H NMR spectra were acquired for each solution with 4 scans, acquisition time of 6.4 s, pulse angle of 90° and a repetition delay of 60 s (determined as adequate for full relaxation from the above T_1 experiments).

The resulting data were processed using a custom script in MestreNova (see Appendix 6) with the GSD parameters as discussed in the result and discussion section. The FID was zero filled once, apodised with an exponential function of 0.3 Hz, automated phase corrected using the Baseline Optimization algorithm, and automated baseline corrected using a Bernstein 3rd order polynomial.

4.3.3.2. Discussion

A ^{1}H NMR spectrum of the reaction liquor is shown in Figure 4.3. T_1 measurements of the aromatic region of **2**, the THF resonance region from 3.5 – 2.5 ppm and the THF/hexane region from 2.0 – 0.0 ppm yielded values of 3.6, 10.4 and 6.2 s, respectively. Good resolution was observed in the aromatic resonance region between signals for **2**, its related impurity and all other peaks in the NMR spectrum. A t-butyl peak from **2** overlapped completely with the lower ppm THF peak, which in turn overlapped with the partially resolved methyl and methylene peaks for hexane. One of the ^{13}C satellites from the higher ppm THF peak was clearly resolved at ca. 4.5 ppm and the second also overlapped with the lower ppm THF peak.

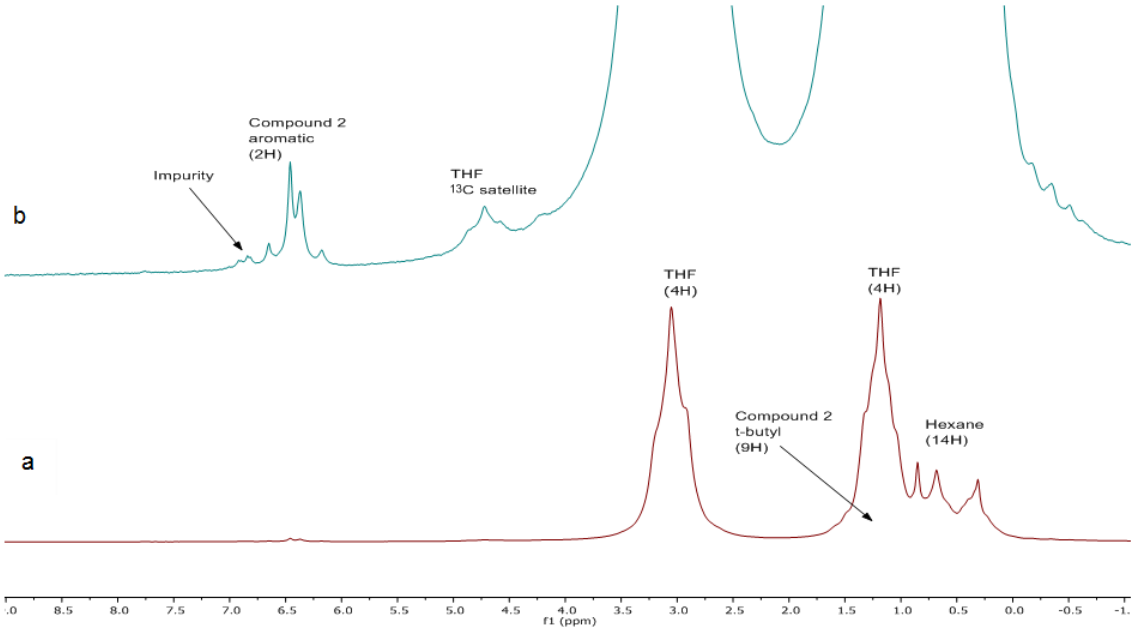


Figure 4.3 – 43 MHz ^{1}H NMR spectra of the reaction liquor associated with 2 at a) Full vertical scale, b) ×16 vertical zoom (upper).

As can be seen in the vertical expansion, none of the peaks were truly baseline resolved due to the tails of the intense THF peaks. This presented an issue with accuracy when determining the peak areas by sum integration alone, as signals from the peak tails artificially raised the baseline and led to an over-estimation of peak areas for **2** and its related impurity. It was possible to correct the baseline by defining

specific points between the peaks and setting these to zero, but this had the effect of distorting the integration values of the large THF peaks. Instead, the entire spectrum was deconvoluted using MestreNova's Global Spectral Deconvolution (GSD) as shown in Figure 4.4 and individual peaks summed for each region of interest.

The data were processed with an automated script. The lowest field THF peak was referenced to 3.58 ppm. GSD was then run and the script iterated through the peak table, summing peaks in the following regions: 7.27 to 6.60 ppm (**2** aromatics), 5.50 to 4.97 ppm (THF ^{13}C satellite), 4.97 to 2.64 ppm (THF), 2.64 to -1.00 ppm (THF, hexane and THF satellite).

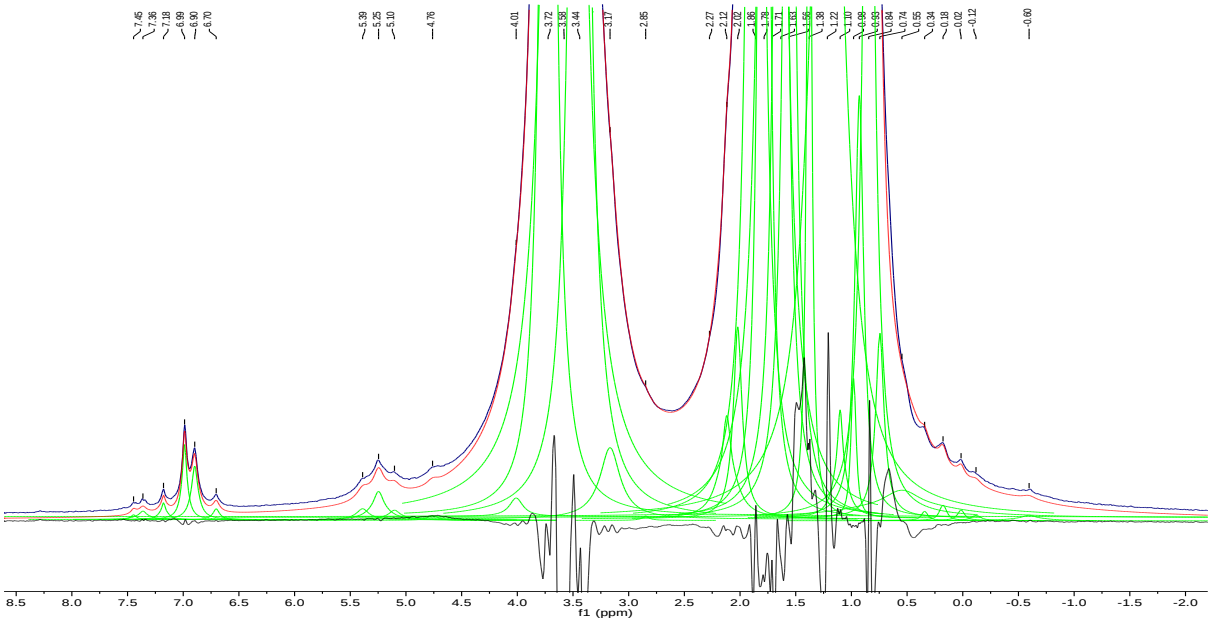


Figure 4.4 – MestreNova GSD of the ^{1}H NMR spectra of the reaction liquor associated with 2; blue = original spectrum, green = deconvoluted peaks, red = total model of the deconvolution, black = residual between model and original spectrum

The hexane integral was determined by subtracting the signals for the THF ^{13}C satellite, THF and **2** from the sum of the overlapped region 2.64 to -1.00 ppm. The peak used in the data subtraction process for compound **2** was calculated as a ratio of the equivalent number of protons for the t-butyl and aromatic groups (*i.e.* a factor of 9/2) multiplied by the total peak area of the summed aromatic peaks. The relative

integrals were then corrected for the number of equivalent protons they represent, converted to a weight ratio using the molecular weights and a %w/w of **2** with respect to THF and Hexane was then calculated. The contribution from the related impurity was not included as this is at a relatively insignificant level compared to the other species.
The resulting spectrum with the text summarising the numerical results was automatically saved and exported as per the sum integration method.

Since **2** was unstable, no isolated material existed and so linearity was assessed by diluting aliquots of reaction liquor with THF.

Linear regression of experimental *vs.* theoretical % w/w THF content yielded a slope of 1.0073, an intercept of -0.1348 and R^2 = 0.996. A residual plot showed no obvious trends.

Accuracy was assessed *via* recovery, as calculated from the data acquired on the 15 samples. Recovery was between 80 and 102%, with the exception of two of the lower concentration solutions. This is likely to be due to low SNR and should not be an issue as real solutions for analysis would not be in this range. If analysis of more dilute solutions was required, then data could be acquired with a greater number of scans.

Precision was assessed as repeatability of the 3 replicate analyses at each level. All replicate analyses were < 5% RSD, with the exception of the lowest concentration, which yielded an RSD of 20%. As with accuracy, this is not considered an issue for the sample concentrations likely to be encountered and in this case, if required, could be mitigated by acquiring additional scans.

DL and QL of **2** were estimated from the SNR of the highest concentration as 0.2% w/w and 0.7% w/w respectively, which are significantly below the concentrations likely to be encountered.

4.3.4. Quantitation by external standardisation

The intent of this investigation was to quantify the solution concentration of hexyllithium in hexane.

4.3.4.1. Sample preparation

A sample of hexyllithium in hexane was supplied by Sigma Aldrich with a CoA claiming a concentration of 2.5 mol/L (Molar) as determined by sec-butanol titration [2].

Initial ^{1}H NMR data of the hexyllithium solution were acquired at low field using the Proton+ experiment with 8 scans, acquisition time of 6.4 s, pulse angle of 90° and a repetition delay of 60 s. T_1 data were acquired with 2 scans, acquisition time of 6.4 s, repetition time of 30 s, maximum inversion recovery delay of 10 s over 11 steps. ^{1}H NMR data were also acquired at high field on a Bruker AV500-II NMR spectrometer equipped with a BBFO+ probe at 300 K with 4 scans, relaxation delay of 30 s, 64 k data points in the FID. The data were processed in Topspin 3.0 with 0.3 Hz exponential apodisation, manual phase correction and automated baseline correction.

Samples of paracetamol, butylatedhydroxytoluene (BHT) and maleic acid were supplied by Sigma Aldrich with CoAs claiming purities of 99.96, 99.9 and 99.9% w/w. respectively, as determined by HPLC assay. Solutions of each compound were prepared at a nominal concentration of 40 mg/mL in DMSO. Initial ^{1}H NMR data to assess overlap and response factor were acquired for three samples using the Proton+ experiment with 8 scans, acquisition time of 6.4 s, pulse angle of 90° and a repetition delay of 60 s. T_1 NMR data were acquired with 2 scans, acquisition time of 6.4 s, repetition time of 30 s, maximum inversion recovery delay of 15 s over 11 steps.

4.3.4.2. Discussion

A ^{1}H NMR spectrum of the hexyllithium solution in hexane is shown in Figure 4.5.

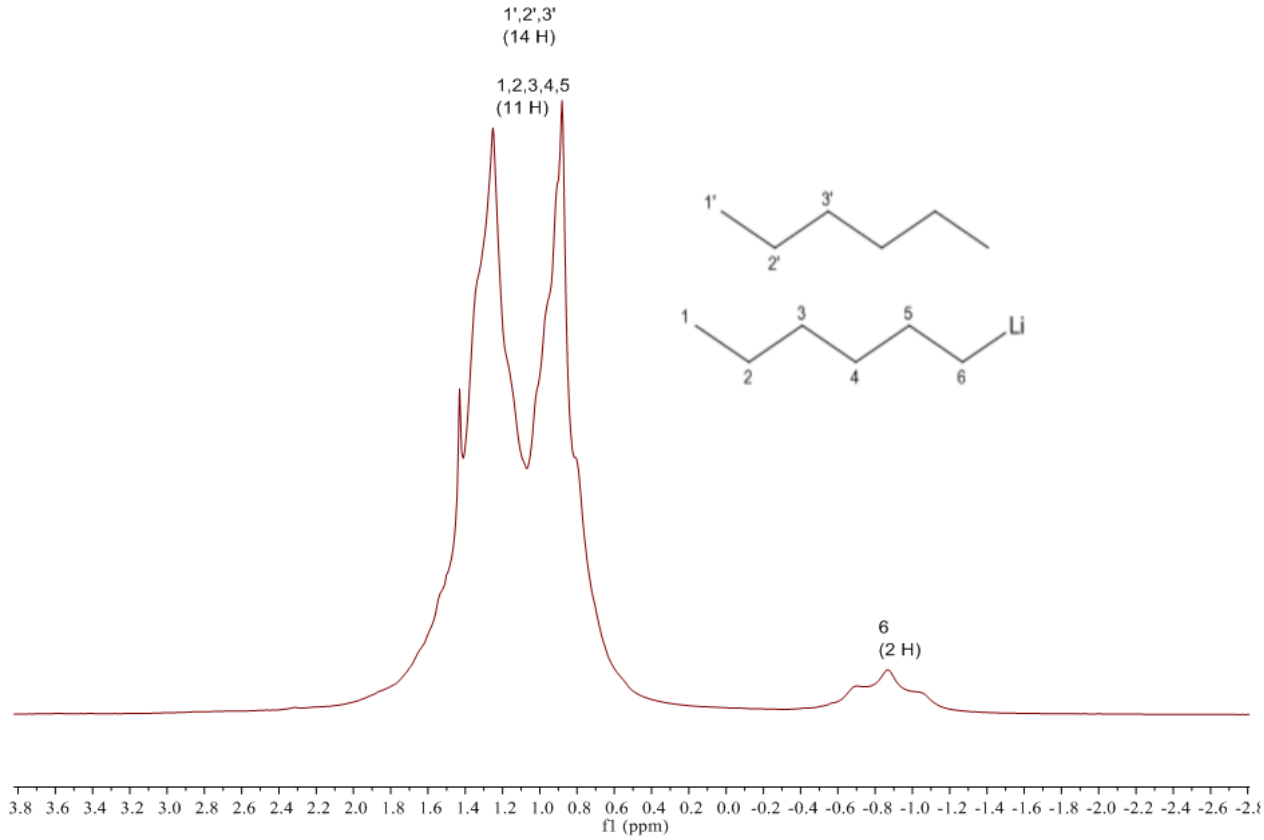


Figure 4.5 – 43 MHz ^{1}H NMR spectrum of hexyllithium in hexane

All resonances overlap in the region 2.0 – 0.4 ppm with the exception of the resonances centred at -0.9 ppm arising from the methylene group adjacent to the lithium. As shown in the figure it should be possible to calculate the concentration by subtracting the integral of the hexyllithium contribution from that of the overlapped region. However, this yields a hexyllithium content of 4.4 mol/L, which is inconsistent with the CoA value.

High field NMR data were acquired on the same neat sample of hexyllithium in hexane at 500 MHz (see Figure 4.6). At this field strength the spectrum showed a large number of resonances in the aliphatic region. This showed that hexyllithium

was supplied not in pure n-hexane, but in *hexanes*, which are a mixture of C_6H_{14} isomers.

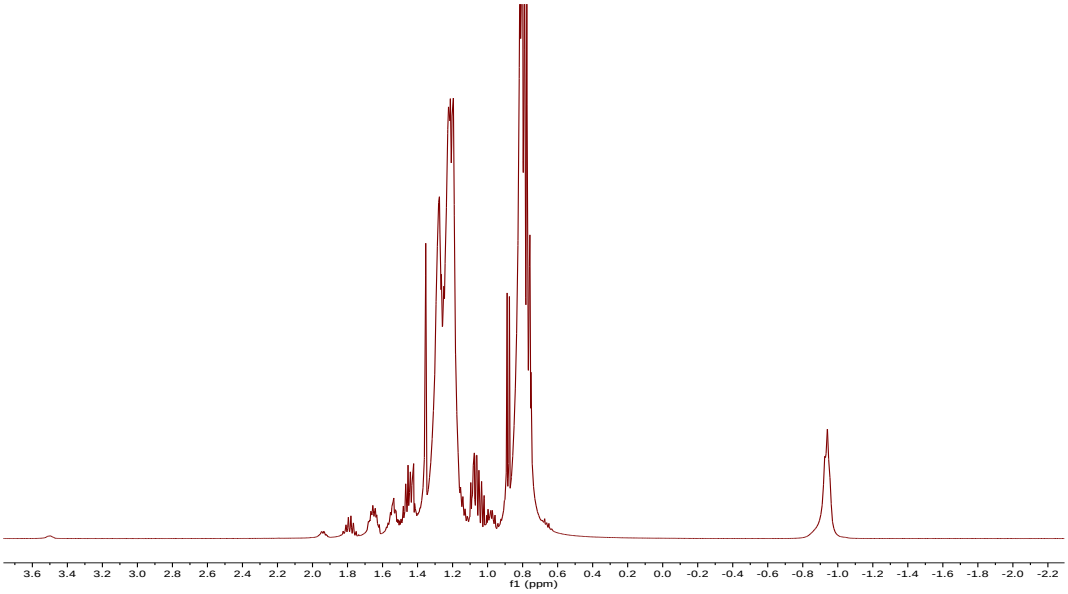


Figure 4.6 – 500 MHz ^{1}H NMR spectra of hexyllithium in hexane

Unless this mixture was well characterised, the composition of each contributing structure known as well as an assurance that the sample contained only C_6H_{14} alkyl chains, then it would not be possible to use this integral region to allow relative quantitation.

The presence of a single well-resolved peak for the species of interest means that it would still be possible to quantify this versus a suitable external standard. A summary of the development of an external standard method is given in the introduction to this chapter and first requires a suitable standard compound of known purity to be identified.

^{1}H NMR spectra for solutions of known concentration of the standard compound and the sample compound are then acquired contemporaneously. A response factor in terms of concentration per unit area can be calculated for the standard solution and this is applied to the sample solution to calculate its concentration. The response of the spectrometer in terms of area units per mole of protons in the active volume should be constant, given that certain parameters are either constant, or have a

negligible response (*e.g*. temperature, quality factor, tuning/matching, receiver gain, *etc.*). Following this reasoning, the standard used can be a solution of any proton containing material of known purity. In practice, the solution would ideally yield a spectrum with a minimum number of well resolved peaks. Where possible these should be singlets, to maximise SNR and peak width, aspects that are particularly important at lower magnetic field strengths.

Three potential external standards were assessed; BHT, paracetamol and maleic acid. 43 MHz ^{1}H NMR spectra for solutions of the compounds in DMSO are shown in Figure 4.7.

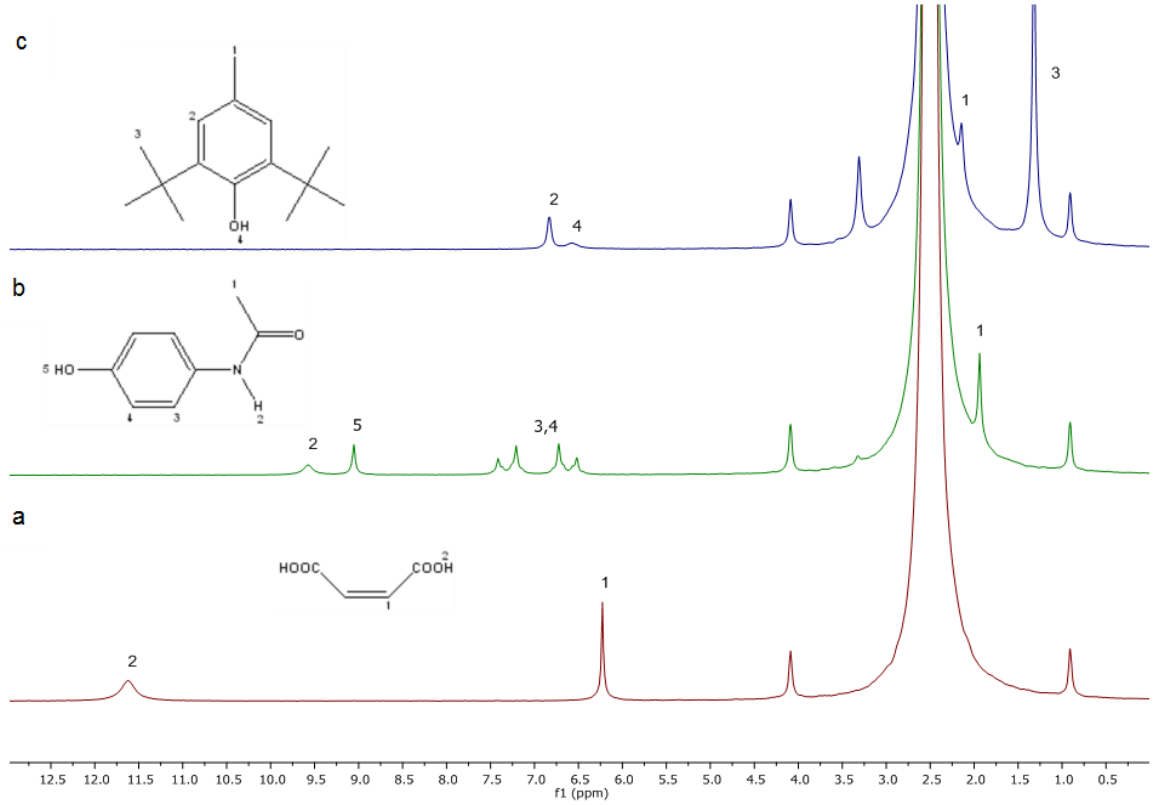


Figure 4.7 – 43 MHz ^{1}H NMR spectra of a) Maleic acid, b) Paracetamol, c) BHT in DMSO. Labels assign resonances arising from protons in their respective structures.

The T_1 value measured for maleic acid was 1.4 s. For paracetamol, T_1 values of the aromatic and acetyl methyl peaks were measured as 1.2 and 1.6 s, respectively. For

BHT, T_1 values of the aromatic and t-butyl peaks were measured at 1.0 and 0.6 s, respectively.

Initial investigation found the integration of the paracetamol peaks to be variable and highly sensitive to slight changes in baseline and phase correction. This is likely to be due to the lower SNR of the coupled aromatic peaks and overlap of the acetyl methyl peak. Paracetamol was discounted and was not investigated further.

Response factors were calculated from equation (4.1) for both spectra using the integrals of the aromatic and t-butyl peaks for BHT and the olefin peak for maleic acid.

$$Response\ Factor = \frac{W \times N_{eq}}{V \times I \times Mw} \times \frac{Purity}{100} \quad \textbf{(4.1)}$$

where W is sample weight (g), N_{eq} is the number of equivalent protons for a given peak, V = volume (mL), I = peak integral (AU), Mw = molecular weight ($gmol^{-1}$) and Purity is the % purity by HPLC mass balance as stated in the supplied CoA. Response factor as calculated from replicate analysis is shown in Figure 4.8.

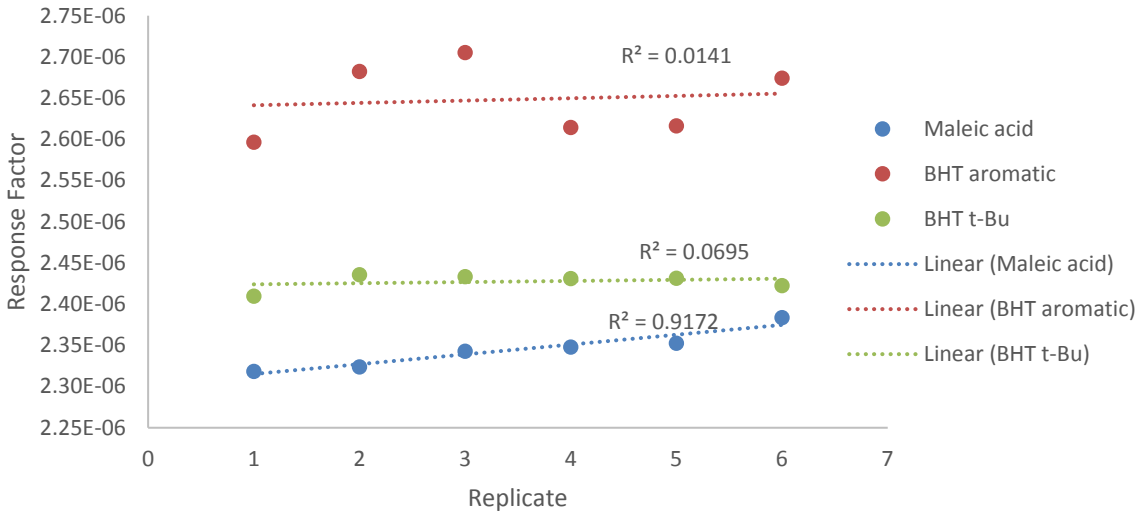


Figure 4.8 – Response Factors of standard solutions

The data shows differences in response factor between species and between the peaks used for integration. This suggested that small, but consistent, factors that bias the peak integrals such as peak shape and overlap were still having a significant effect on the response factor.

To optimise the SNR of the peaks for the standards, the use of pure solvents was considered. These have the advantages of having more intense peaks, less overlap from multiple species, and require no preparation. Eight solvents were chosen as potential standards, each with well-resolved peaks.

As discussed in Chapter 2, when absolute intensity is to be used for quantitative measurements, several factors need to be considered, including pulse width calibration, tuning and matching and temperature.

Pulse width can only be adjusted in set fractions of a nominal 90° value of 8.3 μs (see Chapter 3) on the automated Spinsolve software. The 90° pulse widths were calibrated for all six solvents in Spinsolve expert. The maximum deviation from the nominal pulse width of 8.3 μs was found with DMSO and would result in a loss in signal intensity of 0.03% compared with data acquired with the true 90° pulse. Hence pulse width deviation is not considered a significant issue for this application.

Temperature is a particularly complex factor as it effects the parameters of T_1, T_2, Boltzmann distribution and density, all of which, in turn, will affect signal intensity. As noted in Chapter 3, although the magnet array in the Spinsolve Spectrometer is held isothermal, there is no direct control over the sample temperature. Sample temperature was calculated from ^{1}H spectra of neat methanol acquired over time in the magnet [3]. The temperature was found to equilibrate to *ca.* 28°C over a period of 10 minutes. However, repeat measurement of the peak area of a neat acetone sample over time showed that 99.6% of peak area at thermal equilibrium was achieved after 3 minutes.

As it is not possible to hold the sample at specific temperatures whilst T_1 and T_2, are measured, this relationship could not be assessed directly on this spectrometer. As the acetone intensity over time could be modelled well by considering only the temperature effect on the solvent density and Boltzmann distribution, it will be assumed that T_1 and T_2 changes are not having a significant effect on the data. The response factor values for each of the 8 solvents were acquired after 3 minutes in the magnet to assess variability with solvent (see Figure 4.9). For pure solvents, equation (4.1) is not appropriate as the solvents are assumed to be 100% pure and density now replaces concentration. Response factor values were therefore calculated using equation (4.2).

$$Response\ Factor = \frac{D\ \times N_{eq}}{I\ \times Mw} \tag{4.2}$$

where:

$$D = \frac{D_{20}}{1 + \left(VTEC\ \left(T_{sample} - 20\right)\right)} \tag{4.3}$$

Densities are typically quoted in data tables at 20°C, D_{2O}, hence the requirement to calculate density, D, at the sample temperature, T_{sample}, using VTEC (Volumetric Thermal Expansion Coefficient) in units of $°C^{-1}$. These equations assume equilibrium magnetisation and temperature have been achieved for the sample before data acquisition.

% response factor values were then calculated with respect to an average of the 18 analyses of acetone.

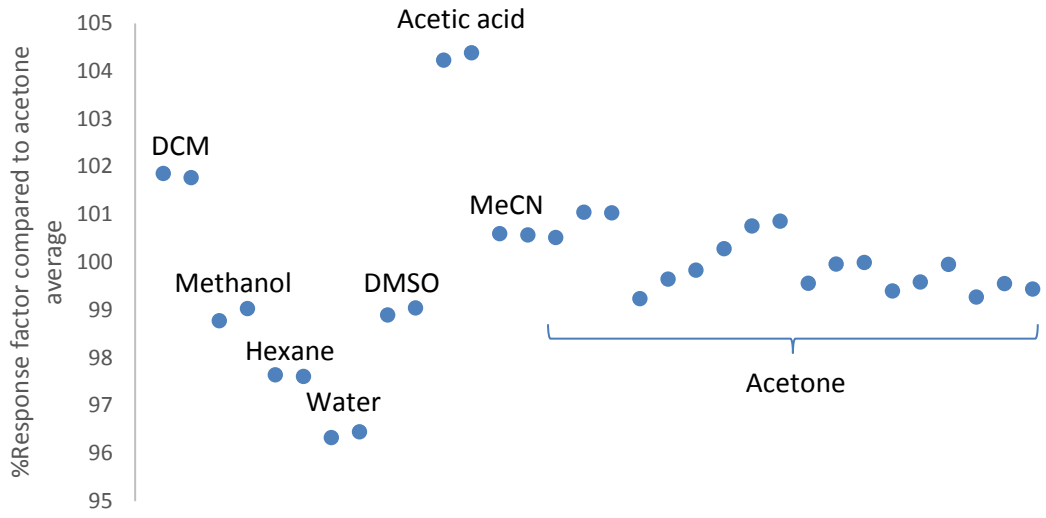


Figure 4.9 – % relative response factors for 8 solvents

Although variation is observed between acetone analyses, there is a much greater variation between different solvents, with a range of 8% between water and acetic acid. There is no obvious trend between the response factor values of the solvents and other parameters such as conductivity, magnetic susceptibility or whether a solvent is protic/aprotic.

As it is not possible to tune or match the spectrometer, the possibility that variation in the tuned state of the probe was affecting the response cannot be discounted. The Wobb curve can be monitored in Spinsolve Expert and some variation was observed, but changes in appearance were difficult to quantify and no values could be recorded or trended against the calculated response factor values.

The experimental work was unable to establish a fixed relationship between the molar concentration of protons in the sample (or standard) under analysis and the observed peak area. As a consequence, there would be a significant variation in the experimentally determined hexyllithium content depending upon the standard compound used, with no obvious explanation. A method could still potentially be developed by establishing the relationship between the response factors of the

hexyllithium solution and a specific standard sample, but significant additional work would be required to confirm the validity of this approach and the priority of this application was such that further experimental was discontinued at this time.

4.3.5. User interface and automation

The overall structure of the Python interface GUI (developed as part of the work in this thesis) is shown in Figure 4.10, where each box represents a different module, each containing one or more functions or objects (see eSuppl 5 for full source code). The code was written in Python 3.5.2 from Anaconda using PyCharm Community edition 2016.3.2 as the developer environment. The GUI utilises PyQt 5.6.0 libraries and was constructed with Qt designer 5.6.0.

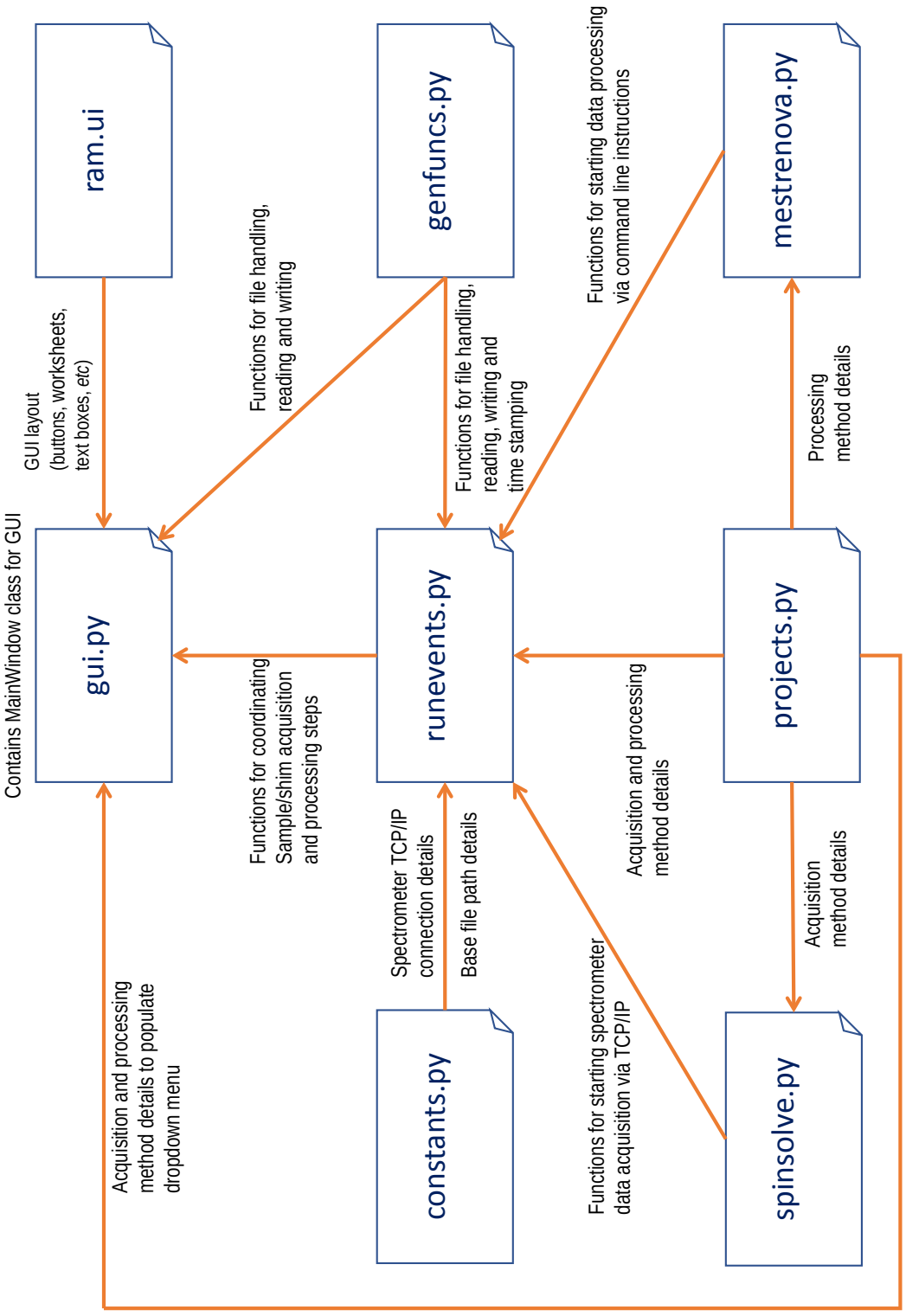


Figure 4.10 – Diagram of module interaction for the Python user interface

Arrows denote where one module imports functions or object attributes from another. The **ram.ui** file contains details of the widgets that appear in the GUI and is in xml format as interpreted by PyQt. The GUI is then initiated and controlled by the module **gui.py**. All other modules are contained in a library, **distmonfuncs**, and are subsequently called as modules with reference to this library. The module **gui.py** takes information from the user in the form of the method to run and sample name and responds to the “Go” button. Once the “Go” button has been pressed, subsequent processes are coordinated by the **runevents.py** module that in turn accesses acquisition and processing packages *via* **Spinsolve.py** and **mestrenova.py**, respectively. The attributes of method name to appear on the GUI, Spinsolve protocol, number of scans, acquisition time, repetition name, pulse angle and MestreNova processing script for each method are stored in an object within **projects.py** along with methods that allow these to be readily accessed by other modules. General functions not specific to any one module are found in **genfuncs.py** and the host and server names for the spectrometer are stored in **constants.py**.

The general layout of the GUI window is shown in Figure 4.11.

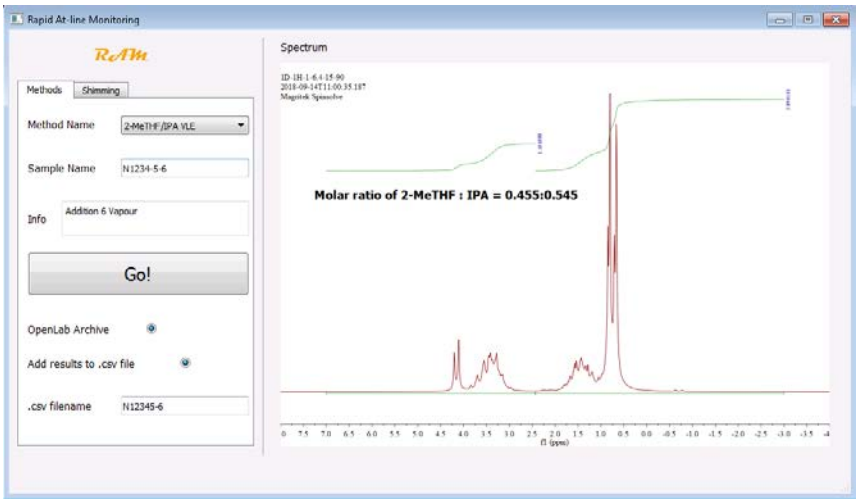


Figure 4.11 – User interface for the Rapid At line Monitoring application, general layout

In terms of a work flow and how the user interacts with the interface, there are two tabs, one for running shimming routines at the start of the day, and one for running a sample as shown in Figure 4.12, left and right, respectively.

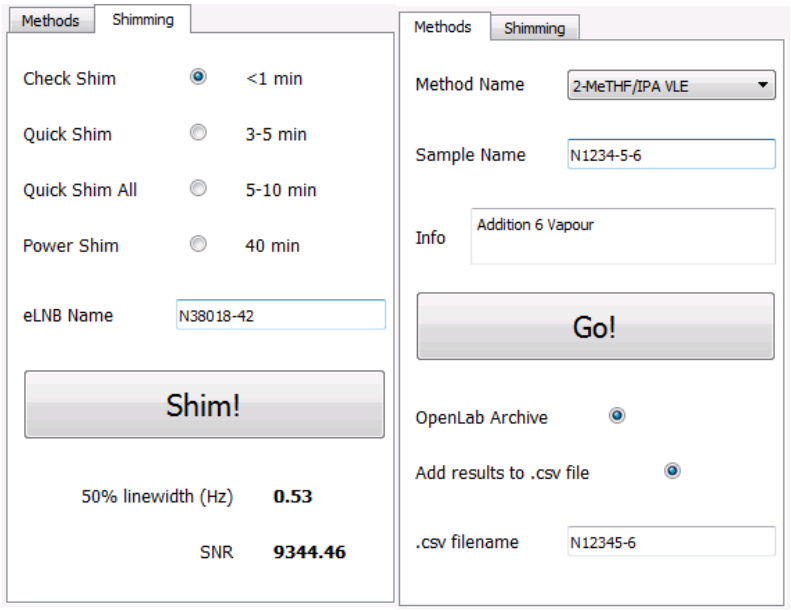


Figure 4.12 – Details of Shimming (left) and Methods (right) tabs

Figure 4.13 shows the work flow for shimming routines.

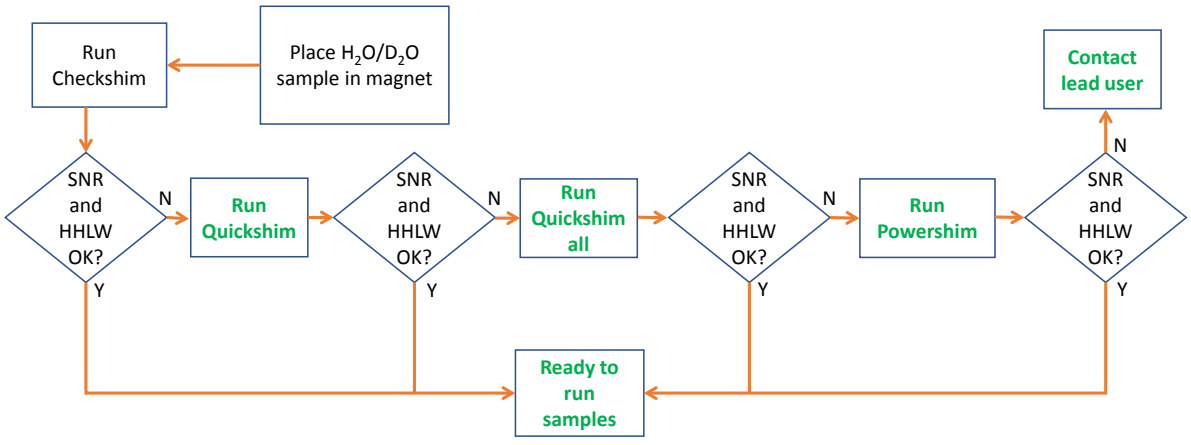


Figure 4.13 – Flow diagram of user interaction with shimming routines (green text denotes messages to user).

First the user would run a Checkshim routine that measures current HHLW and SNR. As with all the built-in pulse sequences accessed through Spinsolve, the

shimming routines do not have any pre-scan delay to allow the sample to achieve equilibrium magnetisation when placed in the magnet. For this reason a 30 s delay is automatically inserted when the “Shim” button is pressed.

The SNR and HHLW are extracted from the shim.par file and displayed at the bottom of the window. If the HHLW is <0.6 Hz, or the SNR < 8000, then a message is displayed on the screen asking the user to run a Quickshim routine. The SNR and HHLW are, again, assessed against the same criteria and progressively more comprehensive shimming routines are recommended until the criteria are met. If this is not achieved after the final Powershim routine, the message asks the user to contact a lead user as there may be other issues with the spectrometer that would require attention.

After the spectrometer has been shimmed satisfactorily, the user is then ready to acquire data on a sample. Figure 4.14 shows a flow diagram for sample data acquisition.

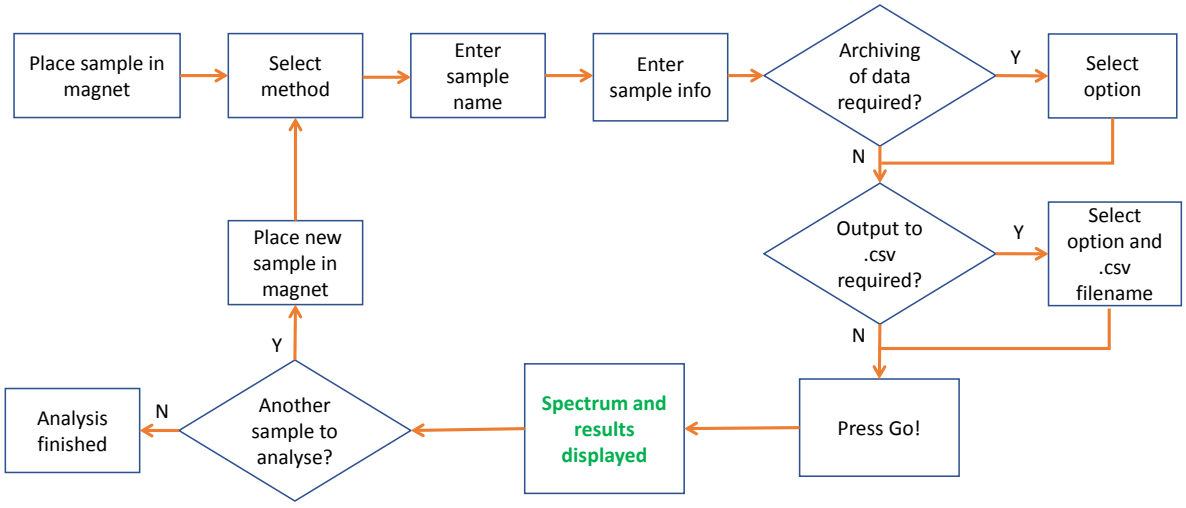


Figure 4.14 – Flow diagram of user interaction when running samples (green text denotes messages to user).

To run a sample, the user dispenses the sample directly into a 5 mm NMR tube and this is placed in the magnet, the correct method is chosen and the sample name

typed in. Additional sample information may be typed in for use later (*e.g.* sample time point, temperature, *etc*.).

If archiving is required then the “OpenLab” radio button is selected. The data will then be saved in a folder that is automatically swept by a remote system each evening for long-term secure archiving. If a number of samples are to be run and the results compiled, this can be facilitated by selecting the option to output to a .csv file. Whilst this button is selected, results from successive sample runs are added to a .csv file named as entered in the user text box. Any text entered as additional information will also be recorded in the .csv file against the relevant sample. This feature allows results from a series of samples (*e.g.* taken as timepoints in a distillation) to be collated in a single file, from where it can be easily copied into an experimental write-up.

Once “Go” is pressed, all acquisition and processing proceeds automatically. After data acquisition is complete the raw data files are saved under the supplied sample name, and the image on the GUI is updated with the new sample spectrum including the result in the desired format, in this case a solvent ratio. A copy of this image is also automatically saved as a PDF file that can be attached to any experimental documentation for archiving later.

As with shimming, an additional delay is added before the first experiment to allow equilibrium magnetisation to be achieved. In this case, this is achieved through running a single dummy scan. From pressing “Go” to the spectrum and results being displayed typically takes between 1 and 4 minutes (depending upon the required number of scans and relaxation delay) and requires no further user intervention.

4.3.5.1. Programming languages and approaches

There are many hundreds of programming languages available, each created at different stages in the development of the computer and each with it’s own intended

use. Although many of the popular medium/high-level languages such as C++, Java, Visual Basic and Matlab have evolved to include modules and libraries such that they can, in principle, perform the same tasks, there are still reasons to choose particular languages for a particular task. The use of Python for this application was deliberate.

Python, created by Guido van Rossum in 1991, is a high-level, interpreted language with support for object orientated, functional and procedural coding structures. The syntax has an emphasis on code readability and is succinct by design, meaning that functions and applications can be put together quickly with relatively few lines of code, compared to languages such as C, or it's variants. It is open source, so unlike Matlab, another popular choice of language for this type of application, it does not require payment for licences to use.

There is a large and comprehensive standard library, including access to GUI tools, and new libraries are created to allow access to more advanced maths functions and new technology standards and interfaces. When creating a new application, modules for specific tasks (for example, reading and writing to .csv files) can be downloaded from online repositories and incorporated into the code for free, again, allowing applications to be developed quickly – as was the case here. As Python interfaces easily with different technologies and other programming languages, it is sometimes referred to as a "glue language", meaning that Python code can be used to tie together other programmes, functions and applications.

As with any application, there is always scope for further improvement. The final application as shown here had several changes made, following suggestions from the end users – for example the inclusion of the .csv and "additional information" functions. Following the development of multivariate quantitative methods (such as those discussed in the following chapter), these could be incorporated into the interface in future by either directly calling Matlab functions from within Python, or by coding the processing in Python itself.

Other areas for improvement could include allowing lists of experiment names and information to be imported as a queue, although this may have limited impact for the current spectrometer as it has no autosampler. In terms of improving the “user experience”, even simple additions such as a status bar to show how long an experiment has left to run can help with user/software interaction.

4.4. Conclusions

The investigations in this chapter have identified four approaches to extract quantitative data from an NMR spectrum, depending upon the complexity of the spectrum and the degree of peak overlap.

Methods were developed for two of the cases, namely sum integration and deconvolution, to provide acceptable quantitative data for specific problems.

The developed methods were then incorporated into a single user interface that allows the required result to be generated and reported with minimum interaction from the non-expert user. The interface was extended to include other routine actions such as shimming and archiving to further integrate these functions into a single package.

Although each quantitation approach was selected as a solution for a specific problem, between them, and in conjunction with the new user interface, they represent a toolkit that will aid in the development and implementation of rapid at line analytical solutions for synthetic chemists. The aims of the investigations in this chapter have therefore been achieved and the system as a whole has been deployed to the local synthetic chemistry and engineering groups where it is now used as walk-up system to generate solvent quantitation by non-expert users. Whilst it is likely that other industrial users of the Spinsolve have produced similar consolidated GUI solution for end users, non have been reported in the literature as far as can be practically ascertained. This system can therefore be considered a

novel approach with respect to current literature and could be of considerable utility to other Spinsolve users, even if only as a basis for further development of interfaces customised to the specific needs of a user.

Initial investigations into quantitation by external standardisation were conducted, but differences in response were observed between different compounds and these could not be correlated to measurable parameters. The differences may have been due to factors that cannot be user controlled in this spectrometer, such as tuning and matching. If future investigation found the differences in response factors are due to parameters which cannot be measured, controlled or otherwise corrected for, it may not be possible to conduct accurate quantitation by external standardisation on the current model of this NMR spectrometer.

MVA was identified as a solution for situations where there is no clear resolution of any peaks of interest. This is a complex subject and the application of MVA to NMR to produce quantitative data will be investigated in depth in Chapter 5.

References

[1] Python Software Foundation, https://www.python.org, Retrieved 30 August.
[2] S. C. Watson and J. F. Eastham, Colored indicators for simple direct titration of magnesium and lithium reagents. Journal of Organometallic Chemistry, 9 (1967) 165-168.
[3] C. Ammann, P. Meier and A. Merbach, A simple multinuclear NMR thermometer. Journal of Magnetic Resonance, 46 (1982) 319-321.

Chapter 5

Multivariate Analysis of Bench Top NMR data

5.1. Introduction

5.1.1. Project aim

The previous chapter explored quantitative approaches of increasing complexity as applied to NMR data, all of which required the analyst to first identify individual peaks in a sample and be able to distinguish them in integratable regions.

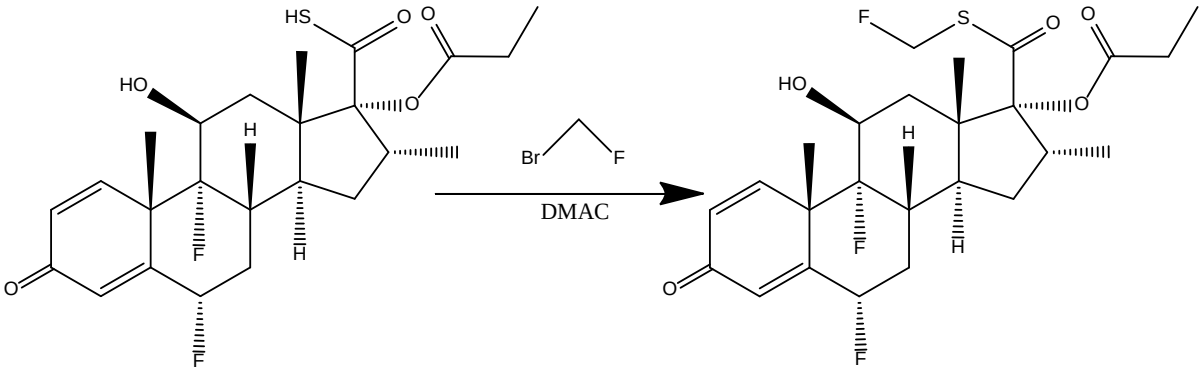
This chapter will now explore Multivariate Analysis as a mathematical tool to enable quantification of spectra where peaks from different species overlap in a single region with no clear baseline resolution. Using the example of a fluorinated steroid synthesis, the approach of Partial Least Squares will be applied to low field ^{1}H NMR data to attempt to quantify starting material and final product in simulated reaction mixtures. The samples will also contain other species known to be present in genuine reaction mixtures and methods to minimise the influence of these species on the quantitation of product and starting material will be investigated.

Finally, the ^{1}H NMR results determined using Multivariate Analysis will be compared to those acquired using simple integration of peaks from ^{19}F spectra of the same samples.

5.1.2. Background

5.1.2.1. Fluticasone propionate synthesis

In the final stage of an industrial process, fluticasone propionate (FP, **4**) [1] is synthesised by alkylation of a steroid thioacid propionate (TAP, **3**) with bromofluoromethane (BFM).



(**3**) (**4**)

The reaction takes place in dimethylacetamide (DMAC) in the presence of several other components carried through from a previous stage. The predominant components are pyridine and imidazole.

Monitoring the current process requires offline HPLC analysis, which in turn requires sample preparation. It has already been shown in Chapter 4 that analysis and processing steps can potentially be automated by bench top NMR if an appropriate quantitation method can be developed, and that this can take place using neat reaction solutions.

In this instance, both starting material, reagent and product have distinguishable ^{19}F spectra (see later), but there are many other related steroid API molecules for which this is not the case. A method will be developed using ^{1}H NMR to quantify the relative contents of TAP and FP in neat reaction liquours. The BFM reagent is a low

boiling point liquid and difficult to handle and prepare in analytical samples. As it is also very reactive, it will be assumed that it is unlikely to be present at appreciable concentrations in real samples, and so for these reasons will not be included in the following experimentally prepared samples.

5.1.2.2. Multivariate Analysis

All analytical techniques fundamentally produce a response, and it is the role of the analyst to find a way to correlate this response to an attribute of interest for the sample under analysis. This is achieved through method development, where trends and mathematical relationships between the instrument response and sample attributes are determined from analysis of a series of prepared calibration samples. In the simplest case, a single variable, obtained from an instrument response, is examined – for example, a % water content by Karl Fisher, an HPLC peak area *via* UV detection, or a relative peak area of a multiplet by ^{1}H NMR. These are examples of *univariate data* and can be readily visualised and interpreted by plotting the response variable versus the attribute variable of the prepared samples (typically an analyte concentration).

In more complex cases there may be multiple variables associated with each sample. For example, the quantified levels of several trace metals determined simultaneously by ICP-OES, multiple ions in an MS spectrum or a series of overlapped and unresolved peaks in an NIR spectrum. In the latter case, each individual data point in the spectrum could be classed as a variable, leading to many tens of thousands of variables per sample. These are examples of *multivariate data* [2], and require more complex approaches to visualise, interpret and quantify.

One of the issues encountered when interrogating multivariate data is that it is not possible to visualise more than three dimensions simultaneously. To explore all the variables of a multivariate data set, many plots must be created and cross referenced. An alternative approach is to apply the process of dimensionality

reduction by using one or more methods of MultiVariate Analysis (MVA). The application of MVA to chemical data is referred to as *chemometrics* – a term coined by a Svante Wold in 1974 [3]. This application of the field emerged in the 1970s and is now a broad subject with a wide range of applications outside the field of pure chemistry. There are a limited number of main MVA approaches, including Principle Components Analysis, Linear Discrimination Analysis, Partial Least Squares Analysis, Support Vector Machines and Soft Independent Modelling of Class Analogy [4]. Through combining elements of these approaches with other statistical and mathematical transformations, there are many other named MVA approaches. The following paragraphs include a subset of these methods that will aim to demonstrate the overarching principles and focus on approaches most applicable to low field NMR data.

5.1.2.2.1. Principal Components Analysis

Principal Components Analysis (PCA) is an exploratory data tool for obtaining a visual understanding of underlying data structures [5]. In essence, correlations between analytical responses for a number of real variables are used to construct a series of abstract variables. These abstract variables are known as principal components (PC), or latent variables (LV) from the Latin *latent* meaning “being hidden”. Latent variables are calculated with the intention that the variation in the data due to the changing attributes of interest in the samples can be modelled with a lower number of LVs than the original real variables. In this way sample attributes can be described by a smaller number of LVs than real variables and the dimensionality is reduced.

By definition, multivariate analysis is performed on arrays of data, which are handled as matrices. In the case of spectroscopic data, an X matrix is defined as containing rows of measured variables – here in the form of spectral data points, *n*, with each row representing the data points from one specific sample, *s*. In the case of PCA,

this response matrix can be decomposed into two smaller matrices, T and P as shown in Figure 5.1.

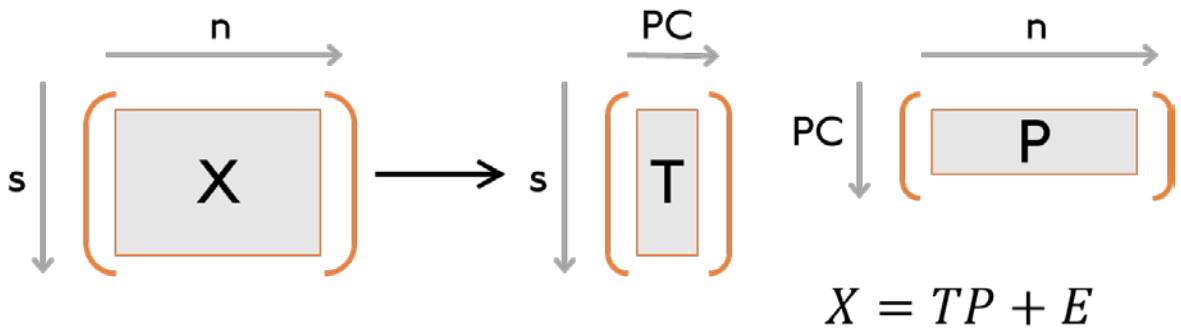


Figure 5.1 – Matrix representation of Principal Components Analysis

where PC are Principal Components. The T matrix then contains information on the relationships between different samples and the P matrix describes any relationships between individual spectral data points. These are also known as the *scores matrix* and *loadings matrix*, respectively. E is the error in the observed data points, including instrumental noise, variation in sampling, error in sample preparation, and so forth.

Once viewed in this form, several key points of PCA become apparent. Firstly, that the data can be interrogated with a smaller number of LVs than the original X matrix. Secondly, that the decomposed matrices allow different aspects of the data structure to be visualised against these LVs and models can be built from here. Finally, there is a notable absence of variables corresponding to analyte concentration included in the PCA model. The implication of this last point is that a model *may* be built with no prior knowledge of identity or content of the analytes in the samples under analysis and hence apparent correlations may be observed that do not relate to any underlying structure of the data. Indeed, there may be no underlying structure. This is not to say that the process of PCA cannot be scientifically rigorous, but that prior knowledge of the system (ideally using characterised samples) must be applied in developing and assessing the model (see later discussion).

Literature examples of PCA applications to NMR data often cover identification and classification of food and drink such as edible oils [6], wines [7] and fruit juices [8].

5.1.2.2.2. Multiple Linear Regression

Once a trend between spectral data points, or a grouping of certain samples has been visualised, it may be required to establish a quantitative relationship between the observed spectra and the sample preparation variables. One method of achieving this is through Multiple Linear Regression (MLR) using Classical Least Squares (CLS) or Inverse Least Squares (ILS) [9]. CLS-MLR assumes analyte concentrations from preparation are of high accuracy and the error is in the measured spectral data points. The opposite is true for ILS-MLR, where the assumption is that modern spectrometers have superior stability and SNR and hence the source of error to consider is the sample preparation.

For CLS-ML, as before, a sample matrix of spectral data, X, is decomposed into smaller matrices (see Figure 5.2).

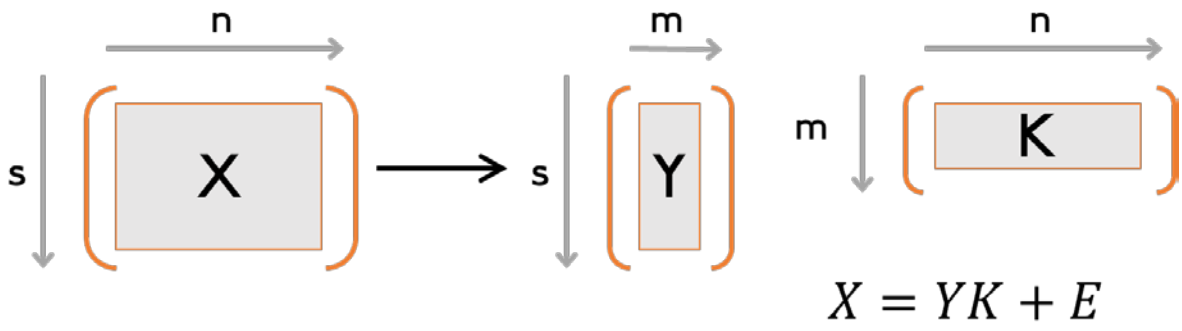


Figure 5.2 – Matrix representation of Classical Least Squares – Multiple Linear Regression

Y is a matrix of predictors, and contains rows of analyte concentrations, with one row per sample, one column per analyte, m. K then contains rows of regression

coefficients. CLS-MLR essentially decomposes a spectrum of overlapping species into a linear combination of sub-spectra, where the regression coefficients in K are the sub-spectra of the component analytes. If pure spectra are available for the analytes, then these can be directly entered into the matrix K. The equation shown in Figure 5.2 is then rearranged to determine the composition of a given sample and no calibration is required.

If pure samples are not available, spectra must be acquired on a set of samples containing all species with known concentrations. The number of samples must at least equal the number of analytes. From here, the K matrix can be calculated to complete the model.

In ILS-MLR, the concentrations of the analytes (in the predictor matrix, Y), are considered to be a function of the spectral response (from the spectral variable matrix, X) as shown in Figure 5.3.

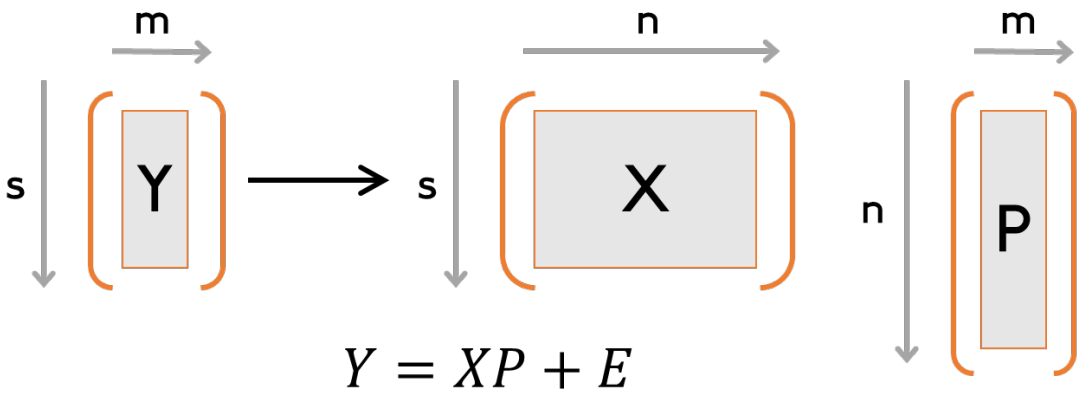


Figure 5.3 – Matrix representation of Inverse Least Squares – Multiple Linear Regression

The regression coefficients matrix, P, now contains *n* rows of values in *m* columns as calculated to fit the sample concentrations included in Y to the resultant spectral variables in X. Since the matrix P is no longer a direct list of sub-spectra, calibration must first take place to determine its values. Calibration models are built for each

species of interest, but other species must be included in the calibration solution to be modelled implicitly.

Comparing the two approaches, CLS-MLR is more intuitive to interpret as the regression coefficients are simply sub-spectra of the sample spectra and, if these are available, it is not necessary to prepare mixed samples for calibration. However, this necessitates the identification of all species in the sample spectra, including minor components. ILS-MLR, conversely, always requires initial calibration, but calibration solutions could be prepared by spiking known concentrations of species of interest into existing samples without knowing the identity of all other species present. ILS-MLR is also more tolerant of minor differences between calibration solutions and real samples. In terms of restrictions, both approaches require at least as many calibration spectra as species of interest (*i.e.* $s \geq m$). There is an additional restriction for ILS in that in order to allow the calculation of the P matrix, the number of observed variables can't be greater than the number of calibration samples (*i.e.* $n \leq s$). In practice, selected spectral points or binned regions may be chosen from the spectra to accommodate this.

5.1.2.2.3. Partial Least Squares

Partial Least Squares (also known as Projection to Latent Structures) combines the approaches of PCA and ILS-MLR in a single stage [10]. Both the X and Y matrices, as defined above, are decomposed by PCA and the PCs are rotated to maximise correlation between the scores (T and U) matrices as shown in Figure 5.4.

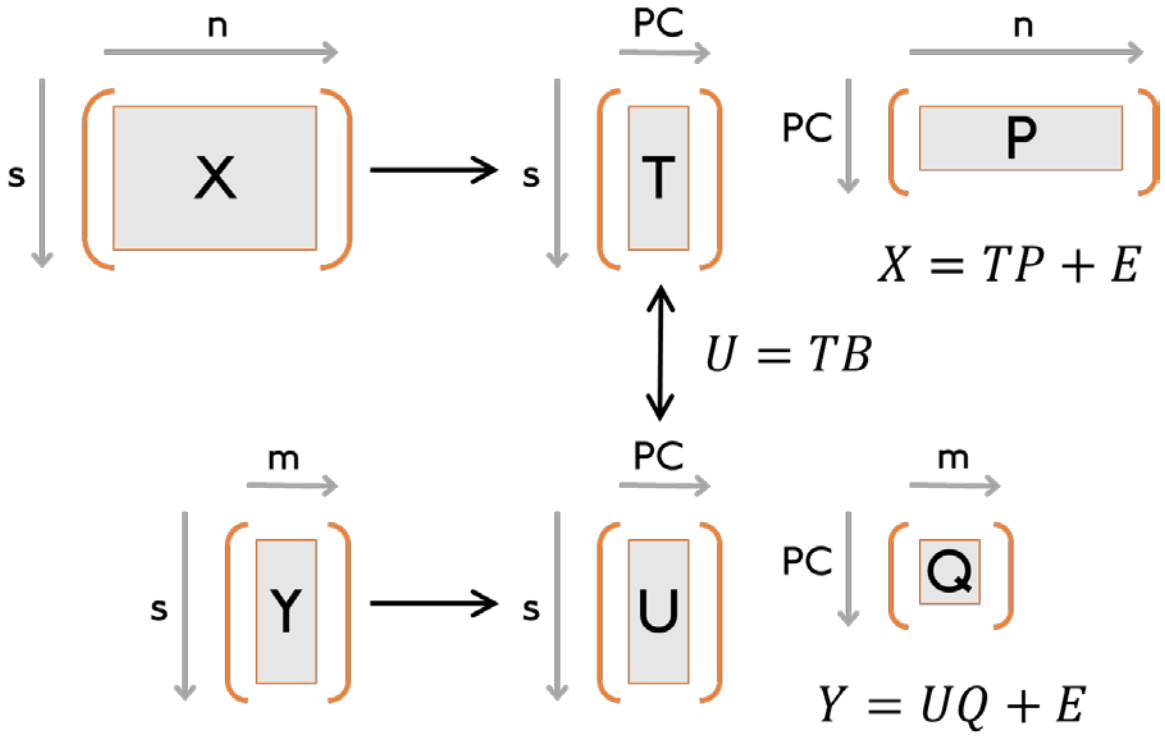


Figure 5.4 – Matrix representation of Partial Least Squares

This results in a new regression coefficient matrix B.

As with ILS-MLR, calibration solutions are required to build the regression coefficients in matrix B and calibration is only required for the species of interest. As before, the calibration samples should include all analytes for implicit modelling, but there is now no restriction on the upper number of variables and the full spectral width can be used. Further advantages include the option to build models optimised for each species of interest (PLS1) rather than inclusion of all analytes in one model (PLS2), and there is further tolerance of differences between calibration and real samples.

As expected, literature examples of the application of PLS to NMR data are usually for the purposes of analysing inherently complex mixtures. Work by Duarte *et al.* [11] covers the modelling of overlapped ^{1}H NMR data of 74 crude oil samples from a

specific location to predict distillation temperatures. Similarly, work by Danieli *et al.* [12] uses the Spinsolve to model ^{1}H spectra of 57 sample of commercial diesel using PLS with aim of predicting parameters such as specific gravity and distillation temperature.

Other natural material have been modelled with PLS, such as the work by Singh and Blümich [13], who modelled the ^{1}H spectra of 108 samples of raw rubber to quantify different monomer units. This example also selected specific regions of the spectrum to model, rather than the whole range of observed signals. In another example by Fanizzi *et al.* [14], 89 olive oil cultivars are classified from high field ^{1}H NMR spectra using a process of PLS-DA. Partial Least Squares-Discriminants Analysis extend the PLS process by rotating the PCs such that maximum separation between classes is achieved.

Another variant of PLS is Orthogonal Partial Least Squares-Discriminants Analysis (OPLS-DA), in which a regression model is constructed between the multivariate data and a variable that only contains class information, for example in work by Wilson *et al.* [15]. In this example, OPLS-DA is applied to ^{1}H benchtop NMR spectra of patient urine to detect markers of diabetic control with the eventual aim of developing a point-of-care analytical tool.

The application of PLS to small molecule mixtures relevant to the pharmaceutical industry appears less common in the literature. This is possibly because quantitation is more commonly achieved by spectral deconvolution, as demonstrated previously in chapter 4. One reaction monitoring example by Nordon *et al.* [16] on a sub – 1 T NMR spectrometer demonstrates quantitation of species in a sulfonation reaction by PLS of ^{1}H spectra. The work also compares the errors and accuracies of NIR PLS data and univariate analysis of ^{19}F NMR spectra acquired from the same reaction mixtures. In another example by Alcantara *et al*. [17], triplicate analysis of 20 pediatric syrups by ^{1}H NMR at low field were processed used PLS to quantify the

sucrose content of the samples. Other multivariate techniques including PCA were used to investigate other sweetener and flavourings in the same samples.

Notably common across all of these above references is this use of a relatively large number of samples with known parameters in order build reliable models.

5.1.2.2.4. General considerations in model building

If a sufficiently large data set is examined enough then some correlation will eventually become manifest. This is independent of whether any real structure actually exists in the data.

To construct an accurate and robust model capable of yielding quantitative results from an analytical instrument it is critical to ensure that any observed trends in a data set are truly causal. This is achieved through combining experience of the analytical technique, understanding of mathematical models used to process the data and consideration of the sample data it will be required to handle in the future. This could be summarised in three general principles:

1) A model is only as good as the data used to build it
2) A model must be built with the minimum number of assumptions
3) A model must be assessed on its ability to predict new samples

To address the first principle, it is not enough for the analyst to consider the data as a string of abstract numbers. In-depth knowledge of the data is required, and the dependency of its features on the analytes, sample attributes and acquisition conditions must be understood. From here, the data can be pre-processed appropriately to maximise the variance of interest and minimise other unwanted variance.

The second principle requires careful selection of the parameters used to build the model. The “goodness-of-fit” of a model may increase significantly by including a higher number of LVs, but this may then result in over-fitting the calibration data. An over-fitted model is less likely to be able to cope with small variations from real-world samples. This consideration can also be applied to the selection of pre-processing algorithms as discussed in the first principle.

To satisfy the last principle, the model must be tested with data typical of real-world samples that encompass the variations that might be expected *e.g.* concentration, pH, presence of additional species, different batches of analyte and so forth. It is not sufficient to select a model based solely on is ability to predict the samples used to build it.

There is obvious overlap and interaction between these principles and they should not be considered in isolation, but they serve as useful guidance during the process of model building and assessment.

5.2. Experimental

All the samples and solutions described below were dispensed into 5 mm NMR tubes for analysis.

5.2.1. PLS model calibration sample preparation

The PLS model was built with a series of calibration samples in full factorial design with two factors at three levels, resulting in $3^2 = 9$ individual experiments (see Table 5.1). Samples were prepared with TAP/FP composition of 0/0, 0/50, 0/100, 50/0, 50/50, 50/100, 100/0, 100/50, 100/100% mol/mol in dimethylacetamide (DMAC), where 100% represents a concentration of 4.0×10^{-4} mol/mL. The stock DMAC diluent also contained imidazole and pyridine at a concentration of 4.0×10^{-5} mol/mL.

Table 5.1 – Experimental design

Experimental design*		% mol/mol nominal	
TAP	FP	TAP	FP
-1	-1	0	0
-1	0	0	50
-1	+1	0	100
0	-1	50	0
0	0	50	50
0	+1	50	100
+1	-1	100	0
+1	0	100	50
+1	+1	100	100

* -1, 0 and +1 denotes low, intermediate and high levels, respectively.

5.2.2. Test sample preparation

To test the PLS model, a series of solutions were prepared with TAP/FP at 10/90, 25/75, 50/50, 75/25, 90/10% mol/mol relative in DMAC, where the total molar concentration of TAP and FP in solutions remained constant at 4.0×10^{-4} mol/mL. The stock DMAC diluent also contained imidazole and pyridine at a concentration of 4.0×10^{-5} mol/mL.

5.2.3. Effects of pyridine and imidazole

To test the effect of the presence of pyridine and imidazole, three solutions were prepared with TAP/FP at 50/50% mol/mol relative in DMAC, with a total molar concentration of TAP and FP of 4.0×10^{-4} mol/mL. The solutions also contained imidazole and pyridine at 0, 4.0×10^{-5} and 8.0×10^{-5} mol/mL, respectively.

5.2.4. Experimental conditions and processing

All data were acquired on the Magritek Spinsolve spectrometer, using the Spinsolve software and built-in experiments as described in Chapter 3.
^{1}H NMR data were acquired in duplicate for all above solutions, using the Proton+ experiment with 8 scans, acquisition time of 6.4 s, pulse angle of 90° and a repetition delay of 15 s.

The resulting ^{1}H data were processed in MestreNova. The FID was zero filled once, apodised with an exponential function of 0.3 Hz, and the FT spectrum manually phase corrected, and automated baseline correction using a Bernstein 3rd order polynomial. The acetyl peak of the DMAC was referenced to 2.00 ppm. The data were exported as a series of .csv files and data shifted such that the highest point in each spectrum corresponds to point 33978 in the list (the highest point of the first spectrum acquired) and compiled into a single data matrix text file using a custom Python application (see eSuppl 6). Data were then imported into Matlab and the PLS

model was built and tested using Eigenvector's Matlab PLS toolbox, as detailed in the result and discussion section.

^{19}F data were acquired in duplicate for all the above samples, using the Fluorine+ experiment with 32 scans, acquisition time of 1.64 s, pulse angle of 90° and a repetition delay of 4 s.

The resulting ^{19}F data were processed in MestreNova. The FID was zero filled once, apodised with an exponential function of 2 Hz and the FT spectrum manually phase corrected, and automatically baseline corrected using a Bernstein 3rd order polynomial.

5.3. Results and discussion

5.3.1. ^{1}H reference data

An overlay of the ^{1}H NMR spectra for FP and TAP in DMAC at 43 MHz shows considerable complexity (see Figure 5.5). This is due to a combination of chemical similarity of the structures, relatively large solvent resonances, broadness of all peaks and the appearance of strongly coupled peaks at low field strength.

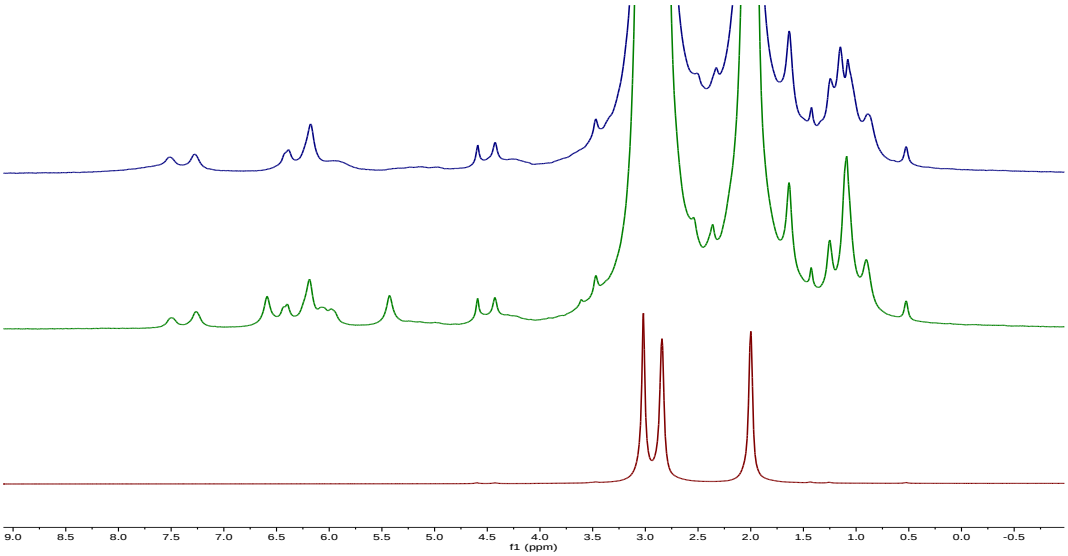


Figure 5.5 - 43 MHz ^{1}H NMR spectra of DMAC (lower), fluticasone propionate (middle) and thioacid propionate (upper).

When pyridine and imidazole are added (see Figure 5.6), the complexity increases further and there are no baseline resolved peaks for the main two species available for simple integration. In addition, although peaks have not been explicitly assigned in these spectra, there are exchangeable peaks present for most species and there is a high likelihood that even non-exchangeable peaks are going to be subject to some degree of concentration and/or pH effects in real solutions.

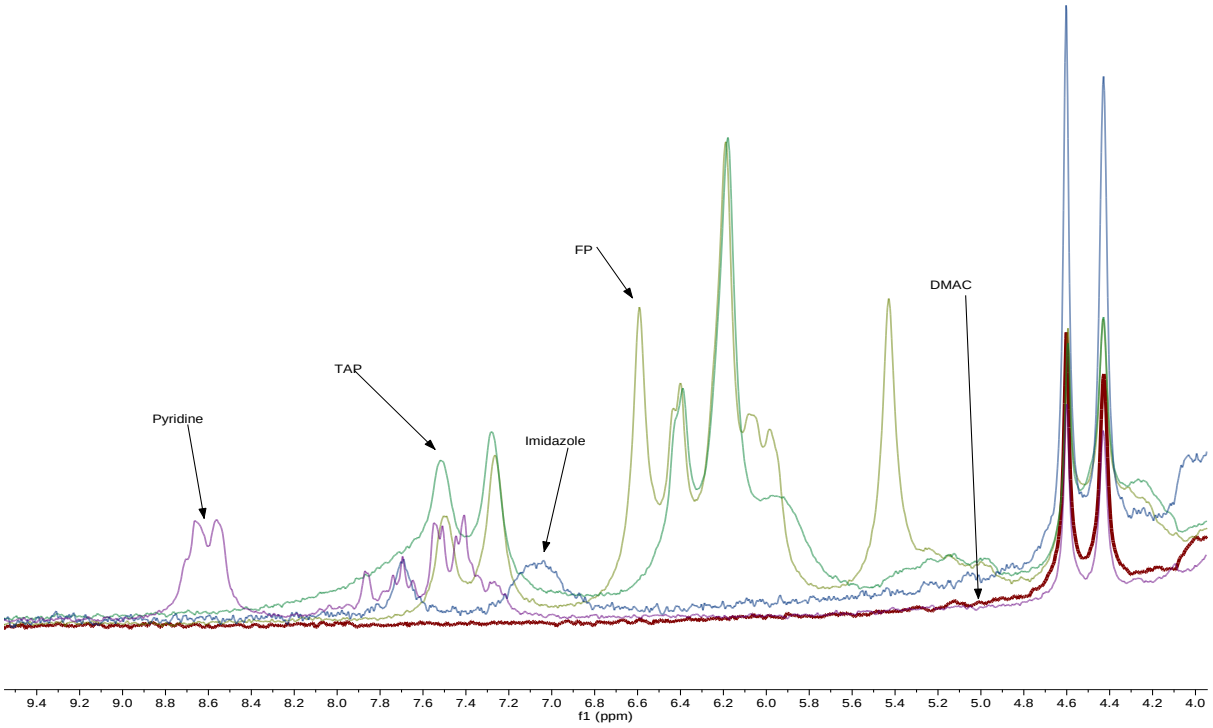


Figure 5.6 - 43 MHz ^{1}H NMR spectra, horizontal and vertical expansion overlay of neat DMAC and solutions of fluticasone propionate, thioacid propionate, pyridine and imidazole in DMAC.

^{1}H T_1 data for solutions in DMAC yielded values of up to 1.8 s for the FP peaks, up to 1.9 s for the TAP peaks and up to 3.0 s for the DMAC peaks. Overall, a 15 s relaxation delay was selected to result in a quantitative response from all peaks from the species of interest.

For these reasons, more advanced modelling of the data to take into account these aspects will be required to yield quantitative data.

5.3.2. Building a PLS model

MestreNova can export ^{1}H spectra as a .csv file in the format {ppm}{tab}{real}, where {ppm} and {real} are the chemical shift and intensity values, respectively, for each point in the real FT spectrum. When a peak in a spectrum is referenced to a specific

chemical shift in MestreNova, then the list of {ppm} values changes, but the {real} values do not. This becomes an issue when the {real} values are plotted against their index and compared for multiple samples, as the spectra are no longer aligned. For this reason, a short Python application (see eSuppl 6) was coded to align the spectra and save them together in a single .csv file to be imported into Matlab.

Initially, the full data range available (*i.e.* 64 k real data points) from all samples was used to build a PLS2 model for TAP and FP. Mean centering was applied as pre-processing. Cross validation was performed using the Venetian blinds method with 9 data splits of 1 sample per blind. The Root Mean Squared Error of Cross validation (RMSECV) plot for TAP and FP (see Figure 5.7) gives a measure of the error observed when using a given model to predict the results of samples used to build the model itself. The y-axis shows the number of LVs used to build each model. In this example, the plot shows that a larger number of LVs are required to model the data than would be expected for just two varying species of interest.

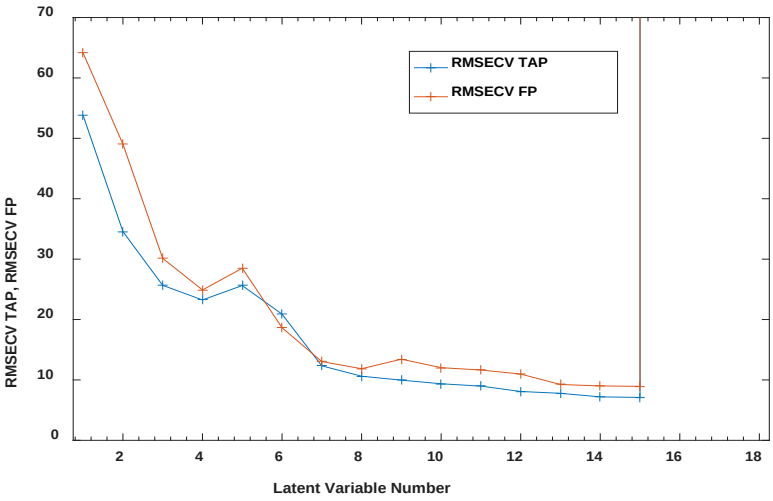


Figure 5.7 – TAP and FP PLS2 calibration RMSECV plot for full data range

This indicates that a significant proportion of the variance in the calibration data set arises from sources other than the TAP and FP contents. A PLS2 model was constructed from 4 LVs, which accounted for 95.4% of the variance. Figure 5.8 shows a plot of the regression coefficients for each point in the spectra. The coefficients are greatest in magnitude for the region 0 – 3 ppm, which predominantly

contains DMAC solvent peaks. As the model has identified that a large proportion of the variance in the data arises from this region, even small changes in the appearance of these peaks (*e.g.* intensity, peak width, chemical shift or asymmetry) would have a significant effect on the predicted results.

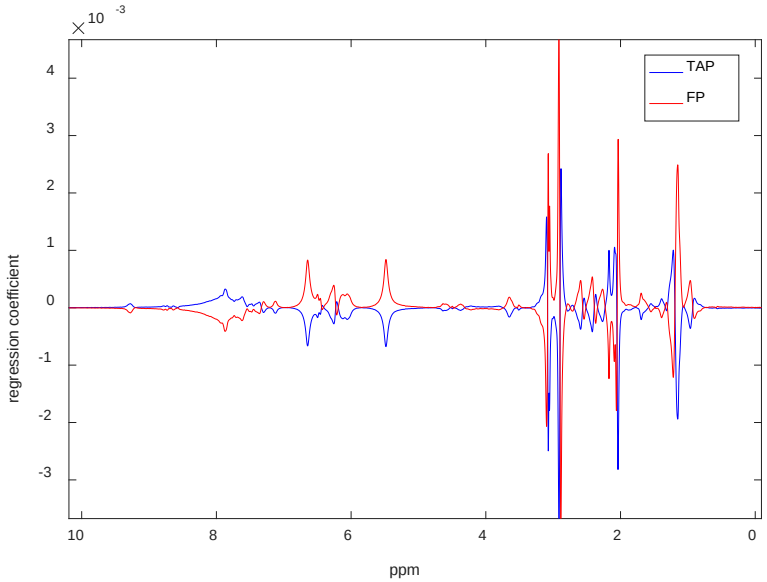


Figure 5.8 – Plot of TAP and FP regression coefficients for a PLS2 model using 4 LVs on full data range

Inspection of the measured *vs.* predicted plots for TAP and FP based on this model shows grouping of data at high, middle and low points, but there is overlap between the levels, most notably with FP (Figure 5.9). Although the R^2 values for the overall calibration are > 0.95, the values for the cross validation are lower as would be expected for the scatter observed.

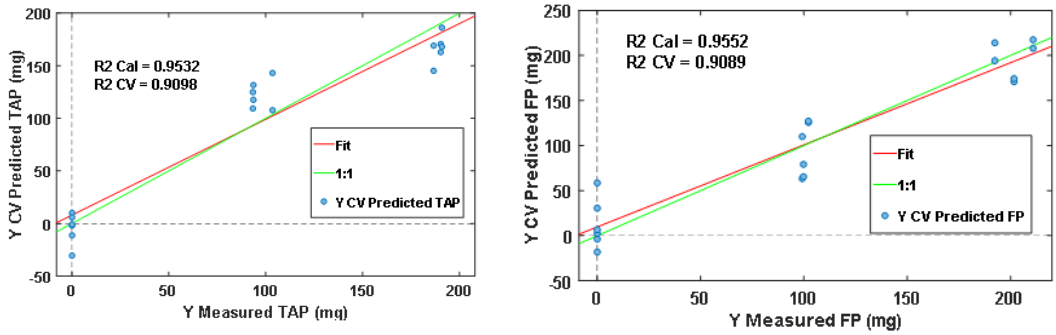


Figure 5.9 - Measured *vs*. predicted plot for TAP (left) and FP (right) for a 4LV PLS-2 model on the full data range

Inspection of a full scale ^{1}H NMR spectrum of one of these samples (see Figure 5.10) shows that most of the data used for this model represent either noise (due to the wide spectral width) or solvent peaks (DMAC).

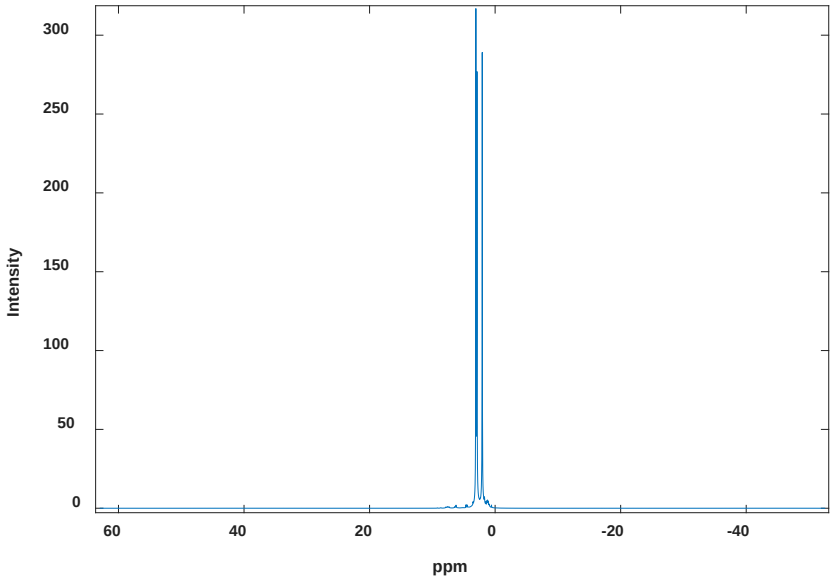


Figure 5.10 – Example 43 MHz ^{1}H NMR spectra of a calibration sample, full scale

Excluding the regions of the spectrum that are predominantly noise (> 10 ppm and < 0 ppm) or solvent (2 – 4 ppm) would therefore allow the model to be built using more relevant data. An expansion of the region of the spectra containing peaks (see Figure 5.11, left) shows that the methyl resonances are overlapping with the low-ppm tail of the solvent peaks and their satellites around 0 – 2 ppm. There is another

large satellite in the region 4 – 5 ppm on the tail of the solvent peaks plus several small peaks that appear to move significantly between the calibration samples at > 8.5 ppm, which are likely to be exchangeables. These regions will also be excluded and a new PLS2 model built from the region between *ca*. 4.9 and 8.4 ppm, consisting of 2001 individual data points (see Figure 5.11, right).

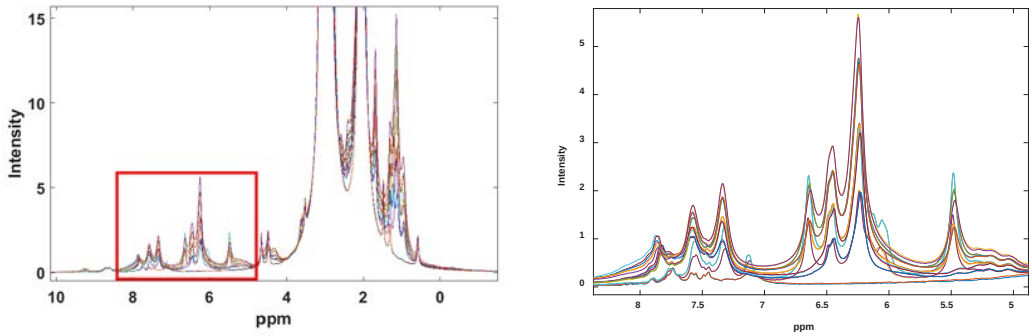


Figure 5.11 – Overlay of ^{1}H spectra of all calibration samples showing main peaks (left) and expansion of boxed region (right) excluding solvents and noise

The RMSECV plot (see Figure 5.12, left) now shows relatively low prediction error using only 2 LVs, with 98.1% of the variance being accounted for. The plot of regression coefficients (see Figure 5.12, right) shows the model is now dependent predominantly upon TAP and FP peaks.

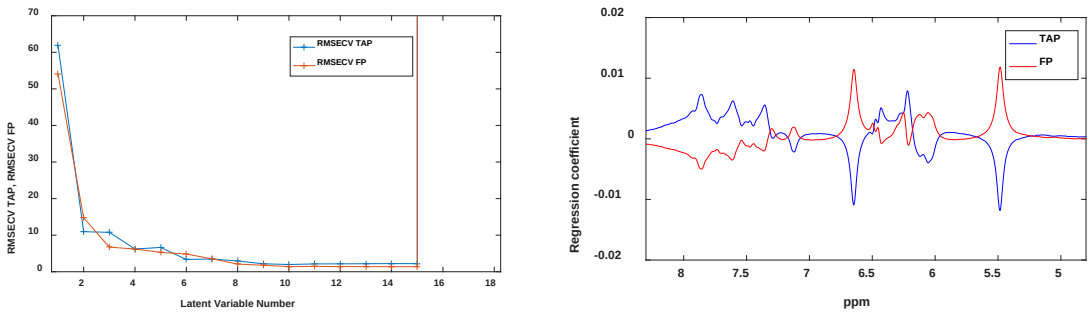


Figure 5.12 - TAP and FP PLS2 model RMSECV plot (left) and regression coefficients for a PLS2 model using 2 LVs (right) for reduced data range

The measured *vs*. predicted plots for TAP and FP in Figure 5.13 now show less scatter and there is a subsequent increase in the R^2 values for calibration and cross validation.

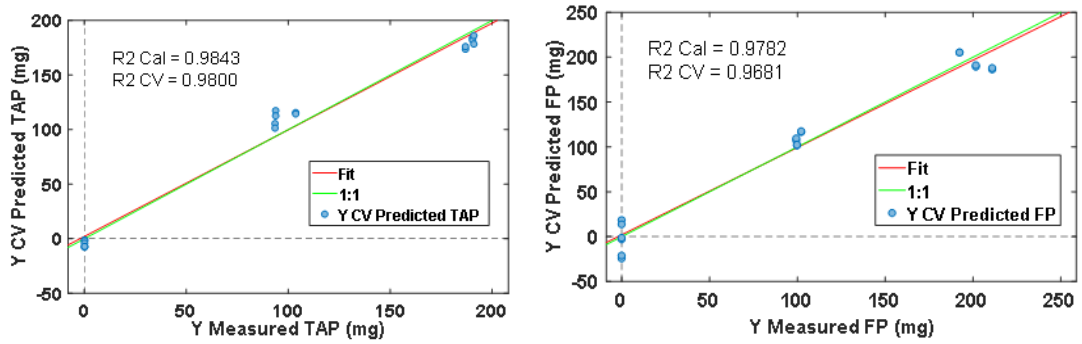


Figure 5.13 – Measured *vs*. predicted plot for TAP (left) and FP (right) for a 2LV PLS-2 model for data range excluding noise and solvents

There is, however, scatter still apparent in the data. Further inspection of the ^1H overlay in Figure 5.11 (right), shows the peaks are on raised (or non-zero) baseline from the tail of the neighbouring, high intensity, DMAC solvent peaks. There is also no complete baseline resolution in this region as a result of low signal resolution due to the broadness of the base of peaks on this low field instrument (see Chapter 3, 3.3.6).

These two factors result in an effective vertical offset between the different data sets as observed in Figure 5.11 (right). Taking the first derivative of the spectral data can minimise the influence on this apparent offset. PLS toolbox includes additional smoothing filters before differentiation process, for which the Savitzky-Golay algorithm was applied here with second order polynomial and a filter width of 15. An example of the effect of derivatisation on the spectra is shown in Figure 5.14.

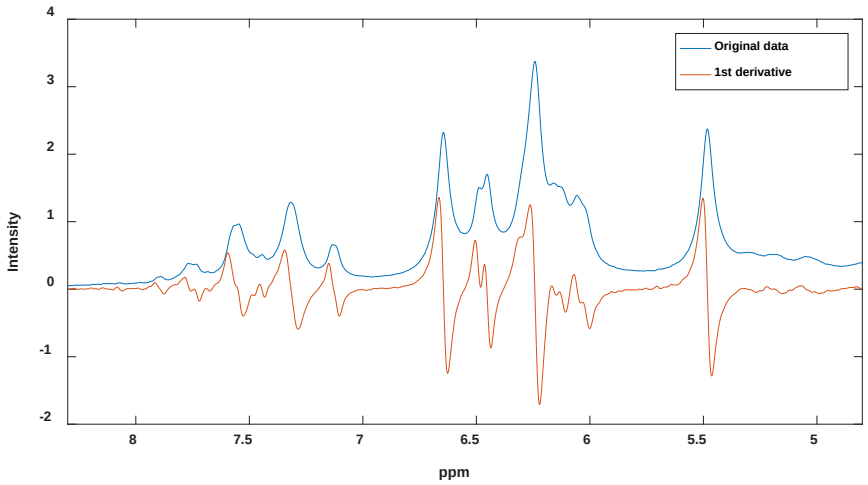


Figure 5.14 – Example of 1st derivative processing on ^{1}H NMR data

Compared to the original data, the peaks now appear “anti-phase” and the baseline is centred on zero. This is because the plot now shows the *rate of change* across the data. Moving left to right (as shown here) from the baseline to the apex of a peak, the rate of change increases. As the apex is approached, the rate of change drops rapidly to zero, then into a negative rate of change on moving from the apex back down the other side of the peak to the baseline. Any section of data having a vertical offset, but otherwise little gradient, has negligible rate of change and so will appear at, or close to, zero.

The RMSECV plot in Figure 5.15 (left) now shows that 4 LVs would be required to minimise the error of prediction, the regression coefficient plot is shown Figure 5.15 (right).

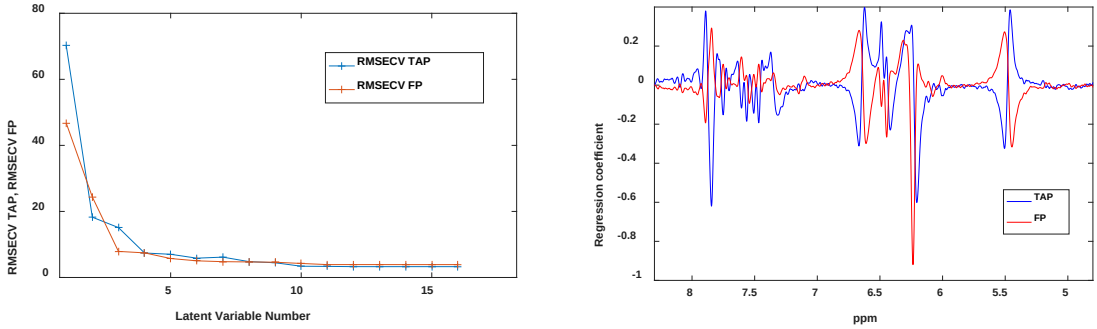


Figure 5.15 - TAP and FP PLS2 model RMSECV plot (left) and regression coefficients for a PLS2 model using 4 LVs (right) for reduced data range with 1st derivative pre-processing

The resulting measured *vs.* predicted plots are shown in Figure 5.16.

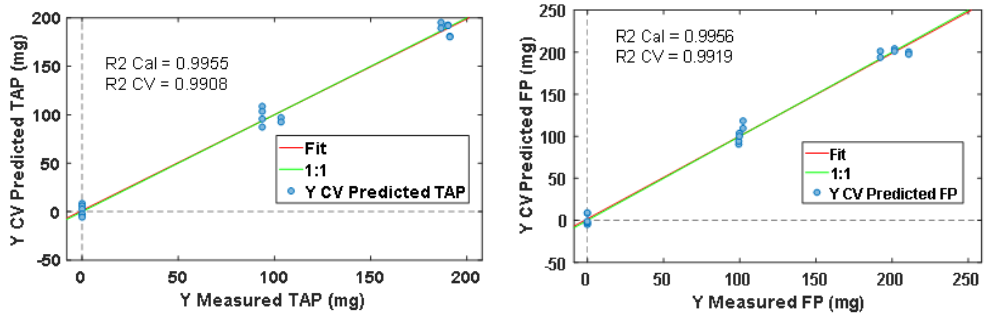


Figure 5.16 – Measured *vs*. predicted plot for TAP (left) and FP (right) for a 4LV PLS-2 model with 1st derivative pre-processing

There is still scatter in the data, but the regression fit goes through more of the points and all R^2 values are ≥ 0.99, suggesting a better linear response from the processed data.

The number of LVs used to model the data has been increased from 2 to 4, suggesting further influence from other species. Figure 5.15 (right) shows relatively high regression coefficients for the region > 7 ppm. This region contains not only peaks from TAP and FP, but also peaks from imidazole and pyridine (see Figure 5.6). Although these latter species are present at a constant level in all the

calibration samples, the peaks from these species will be variable in appearance and dependent upon pH and concentration. It is therefore likely that even in the calibration samples there will be variability in this region from pyridine and imidazole. This will be greater when analysing different samples as the concentrations of the species themselves start to vary. These concentration dependencies may be difficult to model, and so it would be advantageous not to include this region in the model.

For that reason, the data region was reduced further to *ca*. 4.8 – 6.9 ppm (1201 data points), to exclude pyridine and imidazole peaks, as shown in Figure 5.17.

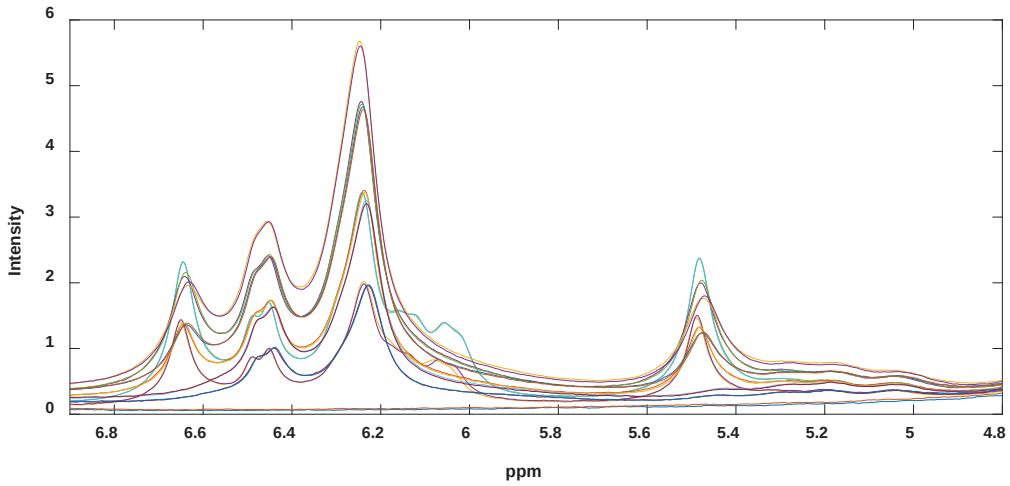


Figure 5.17 – Overlay of all ^{1}H spectra of model samples showing reduced region excluding pyridine and imidazole peaks

The model was then rebuilt using this region, with first derivative and mean centring as pre-processing. The resulting RMSECV and regression coefficient plots are shown in Figure 5.18 (left and right, respectively).

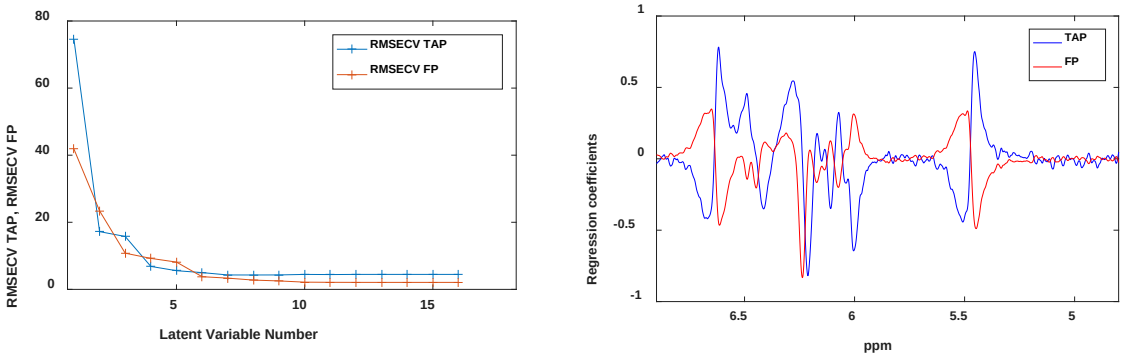


Figure 5.18 – TAP and FP PLS2 model RMSECV plot (left) and regression coefficients for a PLS2 model using 4 LVs (right) for data range excluding pyridine and imidazole peaks

Again, the RMSECV plot demonstrated the need to use 4 LVs to model the data with the current processing conditions, which now accounts for 99.57% of the variance in the data. The regression coefficients used in the model are now taken from data corresponding only to TAP and FP peaks.

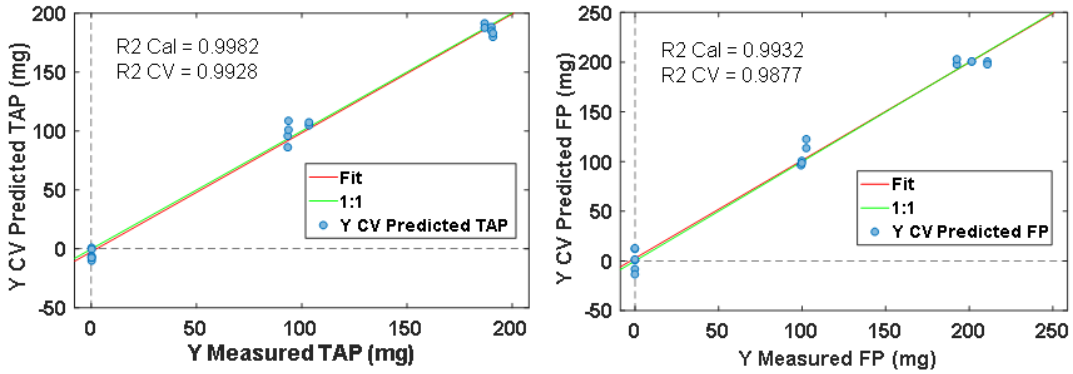


Figure 5.19 – Measured *vs*. predicted plot for TAP (left) and FP (right) for a 4LV PLS-2 model on the data region excluding pyridine and imidazole

The measured *vs*. predicted plots for TAP and FP using this data region are shown in Figure 5.19. There is some decrease in R^2 values for both species, but overall this model still appears to fit the data well, with the advantage that it should now be independent of pyridine and imidazole content.

As noted above, it would be expected that, for 2 varying species, only 2 LVs would be required to model the data. The requirement for higher numbers of LV suggest that even with a narrow data range and derivatisation, there are still additional sources of variance in the data. One source could be a variation in SNR between the calibration spectra. The samples were left in the magnet for sufficient time before analysis to allow the spins to reach equilibrium, but the 15 s delay is much less than 5 – 6 m required for the temperature of the sample to equilibrate with that of the magnet (see Chapter 3). Since the temperature of the sample is directly proportional to the signal intensity (see Chapter 1), then the calibration samples may have been acquired at different temperatures, leading to a varying in SNR between the samples.

5.3.3. Testing the PLS model

When developing a method using multivariate techniques, it is important that the final model should not be selected on its ability to fit the calibration data, but rather on its ability to predict new samples. To assess the method against this requirement, a set of test samples were prepared to include TAP and FP contents that met the following criteria:

1) Reflect real-world samples (*i.e*. TAP% mol/mol + FP% mol/mol = 100)
2) Encompassed by the design space from the model samples
3) Include “centre point” values (*i.e.* TAP/FP 50/50% mol/mol)
4) Include values that do not overlap those from the model

The PLS2 model built previously with 1st derivative pre-processing was applied to the resulting ^{1}H spectra and the measured *vs.* predicted data overlaid with the calibration data in Figure 5.20.

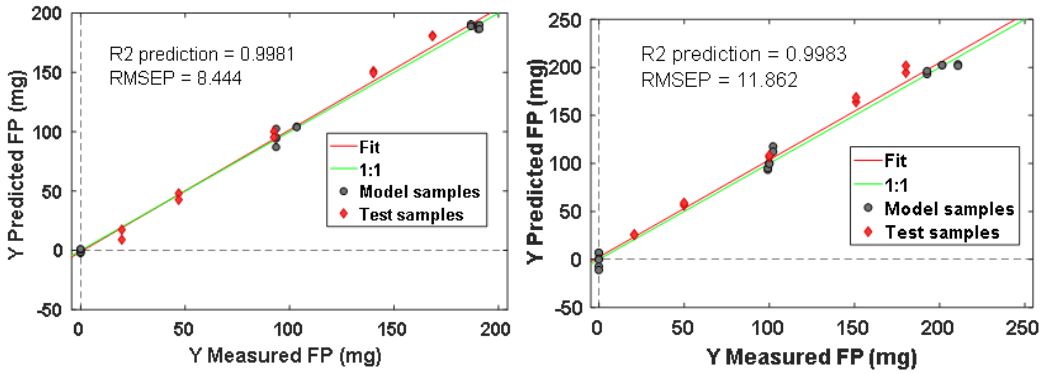


Figure 5.20 – Measured *vs*. predicted plot for TAP (left) and FP (right) for a 4LV PLS-2 model with 1st derivative pre-processing. Calibration samples in grey, test samples in red.

The test data follows the trend of the calibration data and qualitatively shows no greater variation than the calibration data.

5.3.4. Effect of pyridine and imidazole

The spectral region used in the modelling was chosen to exclude the influence of pyridine and imidazole. This was tested explicitly by preparing solutions containing 0, 100 and 200% of the nominal levels of these species compared to the calibration samples. The samples also contained “centre point” levels of TAP and FP (*i.e.* TAP/FP 50/50 %mol/mol). The PLS2 model was applied to the resulting ^{1}H spectra and the measured *vs.* predicted data overlaid with the calibration data in Figure 5.21.

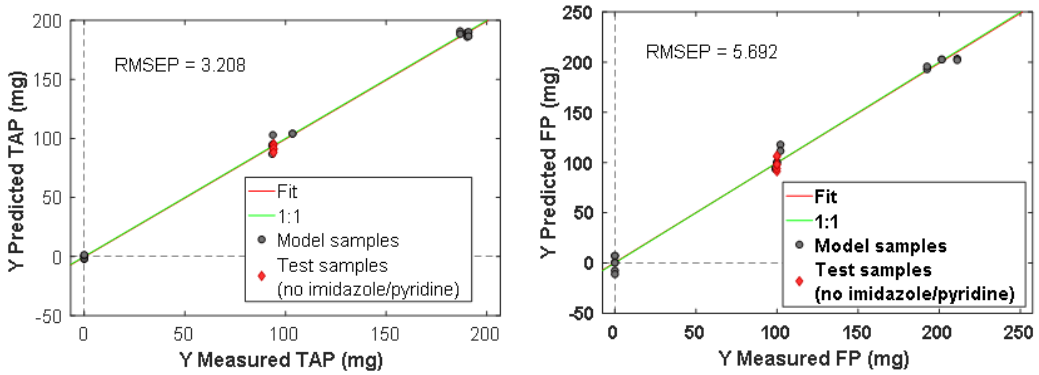


Figure 5.21 – Measured *vs*. predicted plot for TAP (left) and FP (right) for a 4LV PLS-2 model with 1st derivative pre-processing. Calibration samples in grey, test samples with varying levels of imidazole and pyridine in red.

Qualitatively, any trend in test samples with varying imidazole and pyridine is less than the variation observed between calibration samples. It can therefore be concluded that the processing and model yield predicted results for TAP and FP independent of imidazole and pyridine (within the range 0 to 200% nominal).

5.3.5. ^{19}F NMR data

An overlay of the ^{19}F NMR spectra for FP and TAP in DMAC at 43 MHz shows complete overlap between the two species of the peaks at *ca*. -163 and -186 ppm corresponding to the fluorine atoms on the steroid backbone (see Figure 5.22). However, the peak at *ca*. -191 ppm corresponding to the CH_2F fluorine is resolved and hence simple subtraction of integrated areas will yield relative quantitation of FP and TAP. Hence the FP integral is taken directly from area of the triplet at *ca*. -191 ppm, and by subtracting this integral from that of either of the two multiplets at *ca*. -163 or -186 ppm, the integral of TAP is calculated. In this instance the multiplet at *ca.* -163 ppm was used as it produced more accurate results.

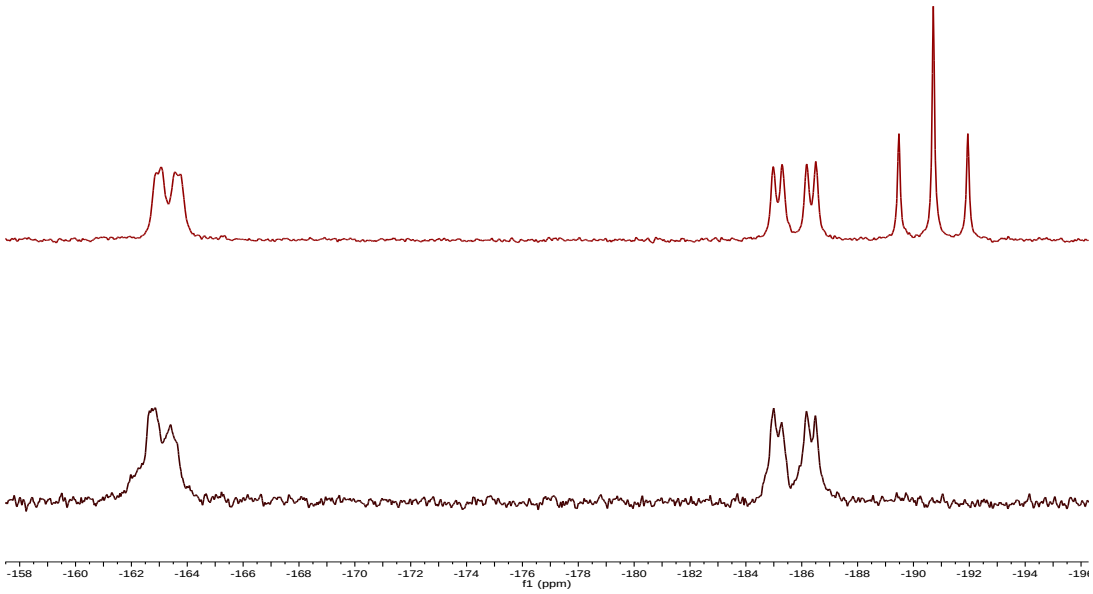


Figure 5.22 - 43 MHz ^{19}F NMR spectra of thioacid propionate (lower) and fluticasone propionate (upper).

^{19}F T_1 data for solutions in DMAC yielded values of up to 0.4 s for the FP peaks and up to 0.3 s for the TAP peaks. Overall, a 3 s relaxation delay should result in a quantitative response from all peaks of the species of interest.

The previous ^{1}H PLS model was generated using the absolute intensities of the ^{1}H spectra. To enable a like-for-like comparison of quantitation by ^{1}H PLS modelling and ^{19}F integrals, the absolute integral areas from the ^{19}F spectra were plotted against the sample weights in mg. Linear regression was applied for TAP and FP integrals as shown in Figure 5.23 and Figure 5.25, with residual plots shown in Figure 5.24 and Figure 5.26, respectively.

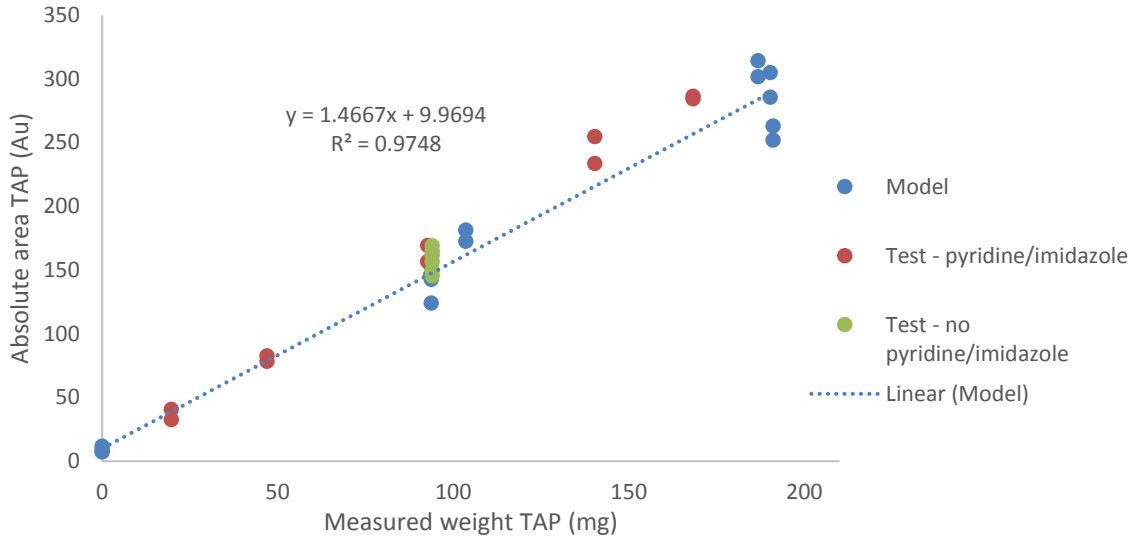


Figure 5.23 – Linearity plot of measured TAP weight vs. absolute ^{19}F peak area. Test data included in plot, but not regressed.

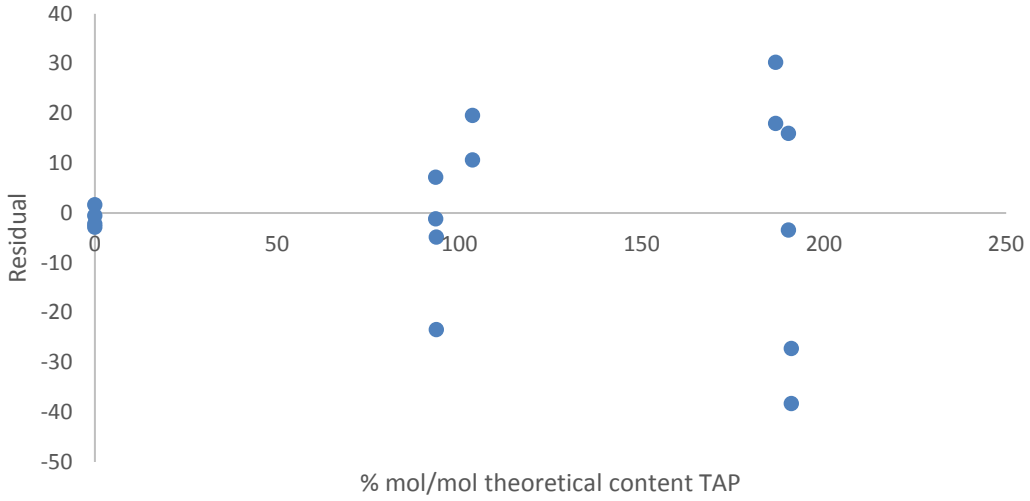


Figure 5.24 – Residual plot for TAP weight *vs*. absolute ^{19}F peak area linear regression

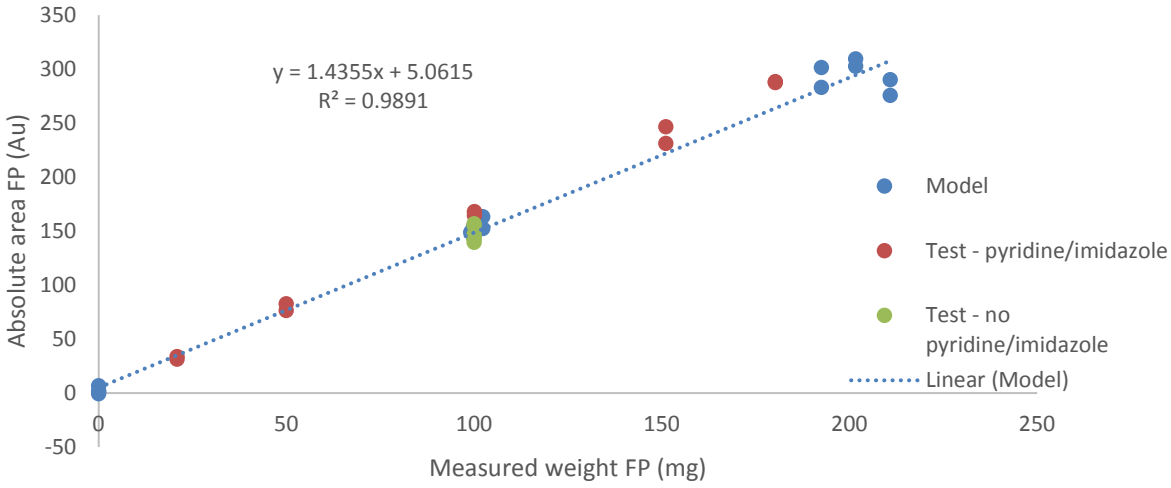


Figure 5.25 – Linearity plot of measured FP weight vs. absolute ^{19}F peak area. Test data included in plot, but not regressed.

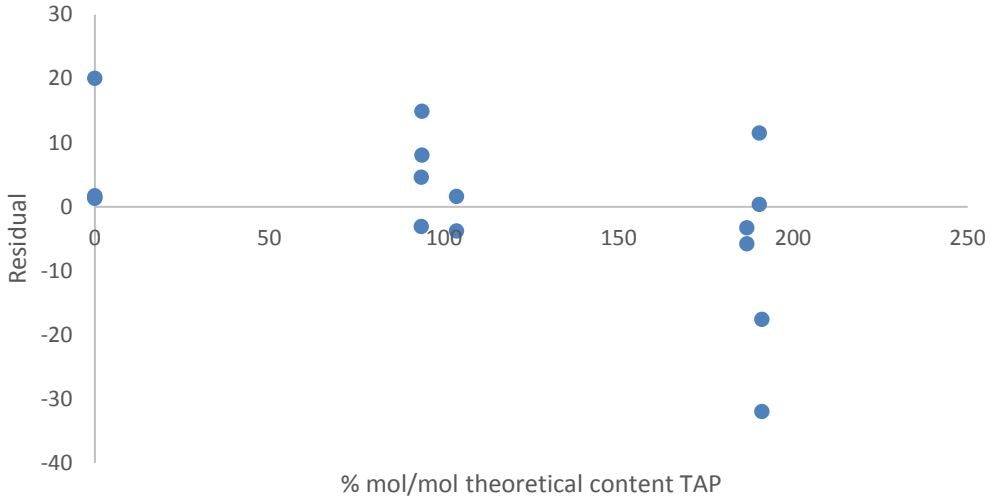


Figure 5.26 – Residual plot for FP weight *vs*. absolute ^{19}F peak area linear regression

Test data are also shown on the plots, but are not included in the linear regression.

Overall, a good linear trend is shown. The test samples containing varying levels of imidazole and pyridine show comparable variation with the calibration samples. There is some deviation in the trend of the test samples containing nominal levels of pyridine/imidazole compared to the overall linearity. Scatter is observed in the data

for higher sample weights in plots for both species, and this is reflected in the increased range of the residuals at the higher levels.

Plots of measured sample weights *vs.* predicted weights for TAP and FP are shown in figures Figure 5.27 and Figure 5.28.

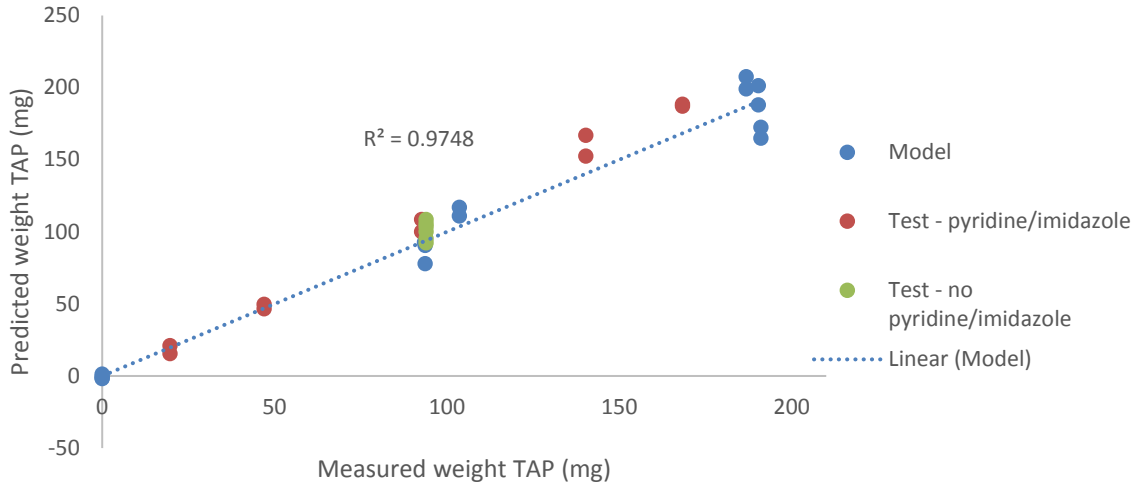


Figure 5.27 – Plot of measured *vs.* predicted sample weights of TAP.

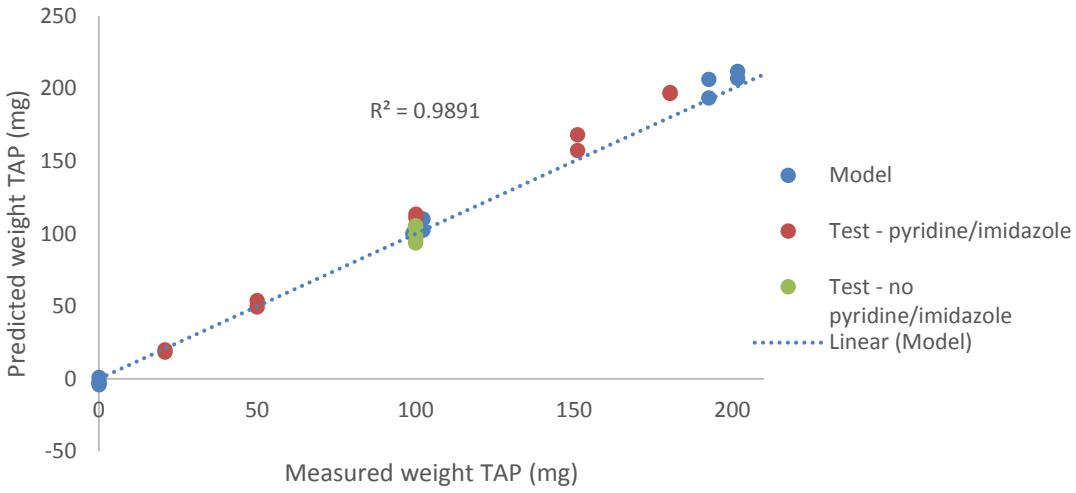


Figure 5.28 – Plot of measured *vs.* predicted sample weights of FP.

R^2 values are comparable to those from plots of measured *vs.* predicted for the ^1H PLS model. Note that because linear regression applies the sample first degree polynomial equation to each measured weight to achieve the predicted weight, the

equation of least squares in these plots is equivalent to y = x and hence is not shown.

Comparison of all predicted data (calibration and test samples) from ^{1}H PLS modelling and ^{19}F absolute integrals shows a linear trend with a slope close to unity and a relatively low intercept which demonstrates a good correlation between the approaches (see Figure 5.27 and Figure 5.28). There is, however, scatter in the data in both dimensions.

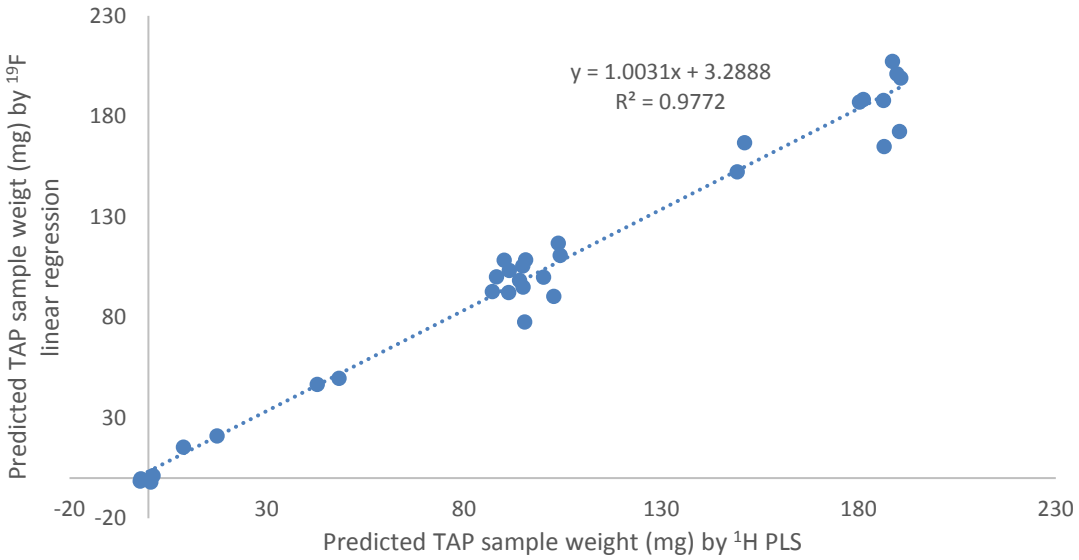


Figure 5.29 – Plot of all TAP predicted results by ^{1}H PLS and ^{19}F absolute intensity.

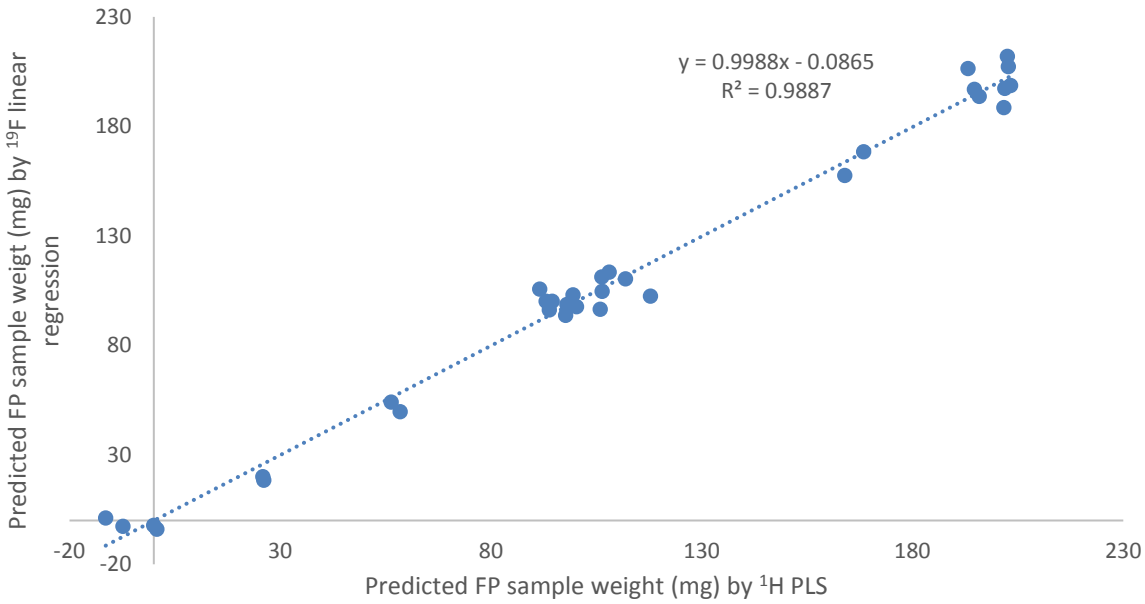


Figure 5.30 – Plot of all FP predicted results by ^{1}H PLS and ^{19}F absolute intensity.

In cases where individual peak areas can be determined directly from spectral integrals, the typical approach for quantitation would use relative integrals, resulting in relative % results. There are several reasons for using absolute integrals, most of which are a direct result of the number of factors that will affect the absolute integral of peaks in a spectrum, but do not affect the relative integrals. For example, as discussed previously, the calibration samples were not acquired under conditions of complete temperature equilibrium. It has also been noted in Chapter 4 that there are still gaps in the understanding of all the factors affecting the absolute signal intensity obtained on this spectrometer, such as the implications of fixed sample tuning.

The ^{19}F data were then processed using the more typical relative integral approach to yield % mol/mol values of TAP and FP. Linear regression was applied to relative % mol/mol values for TAP and FP in the calibration samples (see Figure 5.31 and Figure 5.33), and residual plots generated (see Figure 5.32 and Figure 5.34).

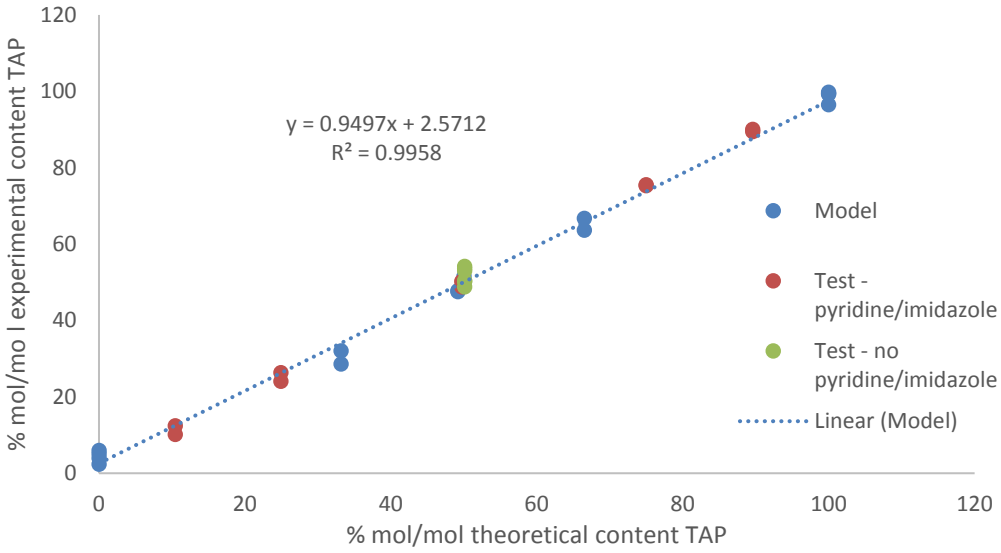


Figure 5.31 - Linear regression of TAP content in calibration samples. Test data plotted for comparison.

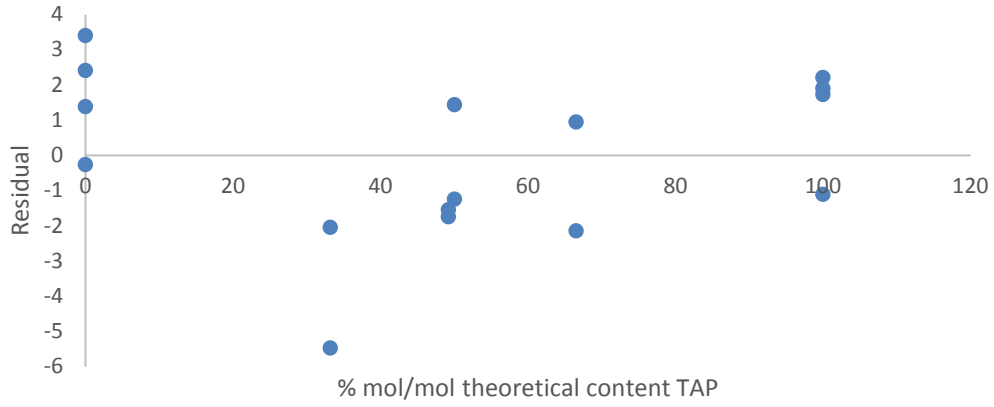


Figure 5.32 - Residual plot for TAP linear regression

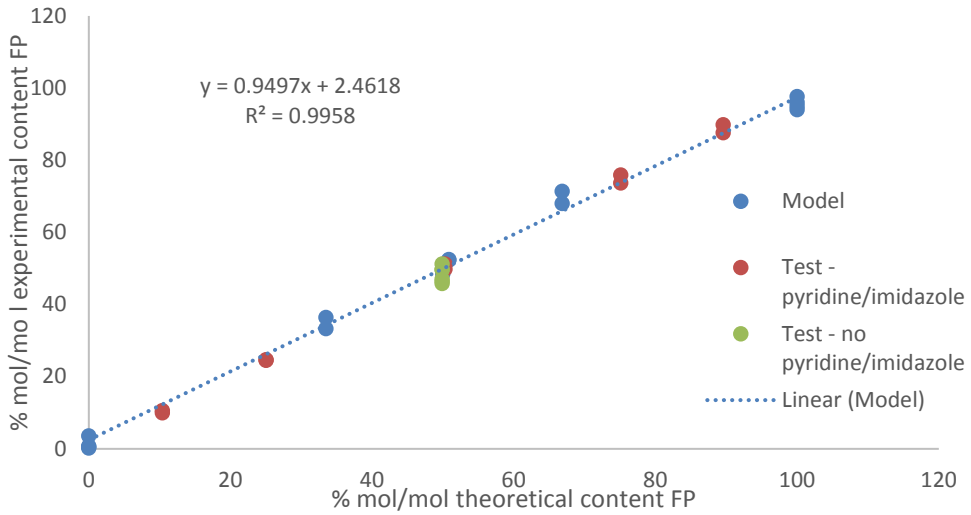


Figure 5.33 - Linear regression of FP content in calibration samples. Test data plotted for comparison.

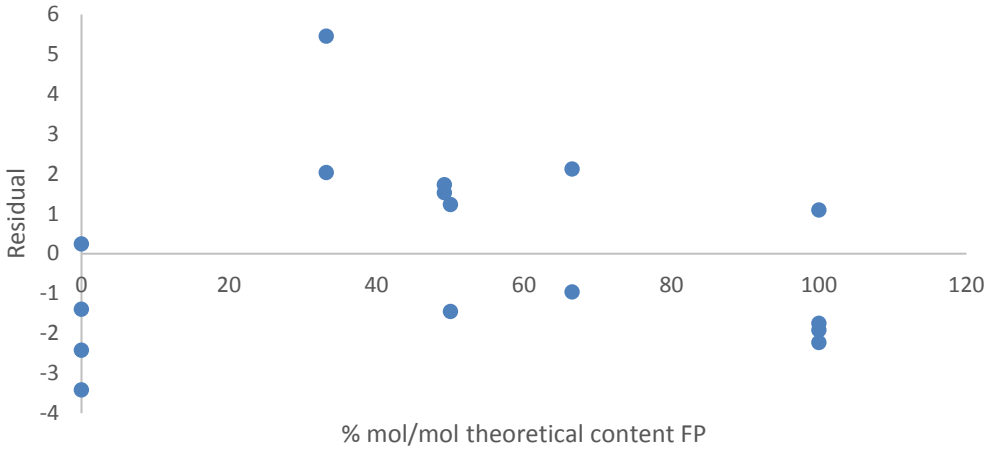


Figure 5.34 - Residual plot for FP linear regression

There is a clear improvement in the linear trends of all data in these plots compared to the ^{19}F absolute intensity data. R^2 is now >0.99, slopes are relatively close to unity and intercepts relatively low. No obvious trends are observed in the residual plots. By observation, the test samples lie closer to the regressed line from the calibration samples, and there is only minor variation in the samples containing varying levels of imidazole and pyridine.

To compare the relative ^{19}F integral approach to the ^{1}H PLS model, predicted absolute weights from the latter were converted to mole values and then % mol/mol results for TAP and FP calculated relative to each other. These were then plotted against % mol/mol results from the ^{19}F relative integrals, as shown in Figure 5.35 and Figure 5.36.

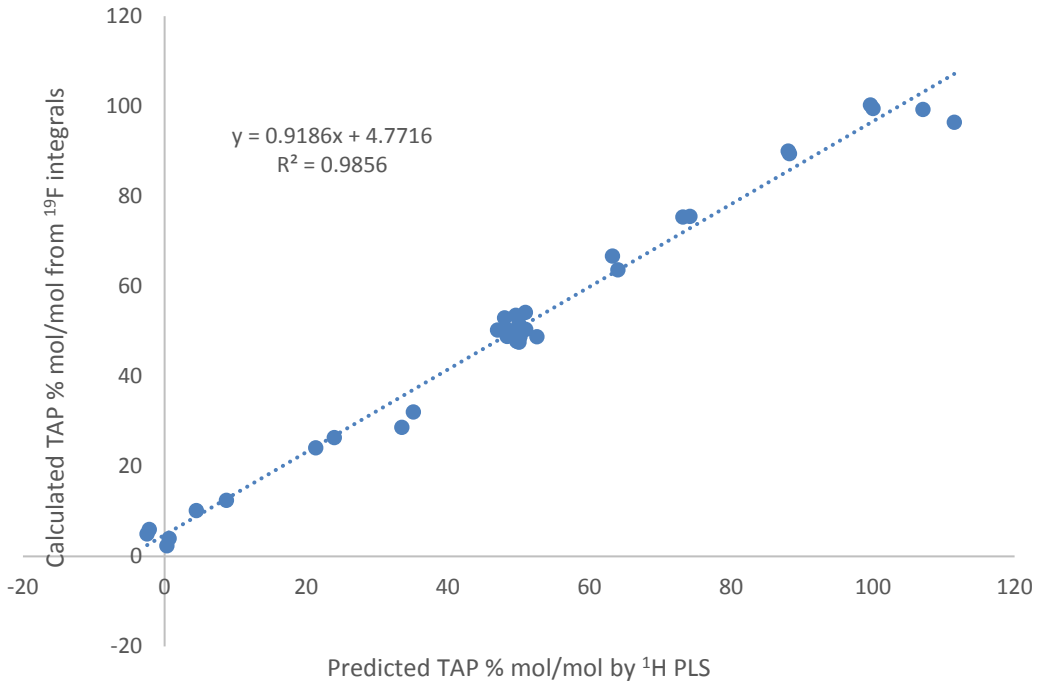


Figure 5.35 – Plot of all TAP results by ^{1}H PLS and ^{19}F relative integrals.

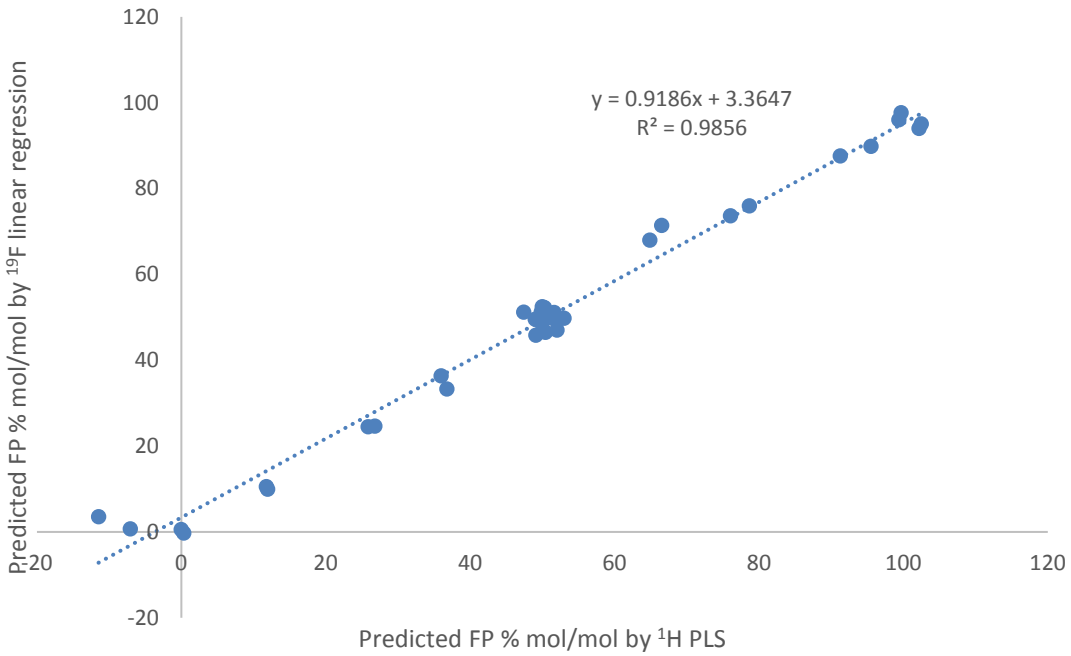


Figure 5.36 – Plot of all FP predicted results by ^{1}H PLS and ^{19}F relative integrals.

Overall the correlation is comparable to the previous plots of absolute intensity data for ^{1}H PLS and ^{19}F absolute intensity data. The slopes are further from unity for TAP and FP data compared to the absolute intensity data, possibly suggesting a slight bias between the two approaches.

From these observations and comparisons, it can be seen that the ^{19}F data supports the PLS model built from the ^{1}H spectra.

5.3.6. Accuracy comparison

The accuracy of the above approaches to quantitation can be measured by calculating the recovery of TAP and FP contents for each sample. As there was a large number of data sets to compare, these were summarised as the total number of recoveries that fell within successively greater ranges – see Table 5.2 and Table 5.3. All values in the table are out of a maximum of 28.

Table 5.2 – Cumulative recoveries for TAP values

	90 - 110%	80 - 120%	70 - 130%	60 - 140%
^{1}H predicted absolute (mg)	26	27	27	27
^{1}H predicted relative (% w/w)	24	27	27	27
^{19}F predicted absolute (mg)	17	27	28	28
^{19}F calculated relative (% w/w)	26	28	28	28

Table 5.3 – Cumulative recoveries for FP values

	90 - 110%	80 - 120%	70 - 130%	60 - 140%
^{1}H predicted absolute (mg)	21	26	28	28
^{1}H predicted relative (% w/w)	26	28	28	28
^{19}F predicted absolute (mg)	23	28	28	28
^{19}F calculated relative (% w/w)	28	28	28	28

FP quantitation was overall marginally more accurate than TAP with 14 recoveries outside the range 90 – 110% and only 2 outside the range 80-120%. TAP quantitation had 17 recoveries outside the range 90 – 110% and 3 outside the range 80-120%. The most accurate method overall was calculating % mol/mol from ^{19}F integrals, with all recoveries 90 – 100% for FP and 2 recoveries outside this range for TAP.

The accuracy for TAP content as determined from absolute ^{19}F integrals was notably lower then the contents determined for TAP by either ^{19}F relative integrals or the ^{1}H PLS approaches. Although less overlapped, the ^{19}F spectra had much lower overall SNR than the ^{1}H spectra due to compromises on the fixed tuning on the probe, and the peaks for TAP were broad due to coupling. Added to this is the fact that TAP does not have any peaks in the ^{19}F spectra that do not overlap with FP, and can only be determined by subtraction. These combined influences are likely to render the ^{19}F absolute integrals more sensitive to slight changes in intensity or noise from one analysis to the next and introduce greater error. Corresponding results as calculated from ^{19}F relative integration have much better accuracy, further supporting the hypothesis that the greater error is due to changes in the overall spectrum intensity, which would have a much less significant effect on the relative areas.

The ^{1}H methods using PLS models were less accurate than ^{19}F methods, but were capable of producing quantitative results that may still be acceptable for purposes of reaction monitoring during process development. This would be of utility in circumstances where no fluorine is present in the analytes of interest.

5.4. Conclusions

The aim of the work presented in this chapter was to assess the applicability of multivariate approaches to low field ^{1}H NMR spectra. Partial Least Squares was identified as a suitable approach and was shown to be capable of extracting quantitative results from spectra that would have been inaccessible by univariate techniques.

A source of high variability was identified as arising from species additional to the analytes of interest. It was also apparent that the variability could not be readily modelled, and so the decision was taken to eliminate the spectral regions for these species before the data were modelled. In doing so a significant improvement of the model prediction was observed and it was demonstrated to be independent of the concentration of these additional species. Careful application of pre-processing, by taking the first derivative, was also found to improve variability in the data by minimising the apparent baseline offsets from non-baseline resolved peaks.

The resulting model was tested with data typical of “real-world” samples and then with samples exhibiting variation in the levels of additional species. In both cases it was shown that quantitation was possible and results generated by the ^{1}H PLS modelling were consistent with results acquired on the same samples using simple integration of ^{19}F peaks.

The work shown here is a demonstration of what can be achieved through the application of MVA to moderately complex ^{1}H spectra using off-the-shelf mathematical software and tools. From previous chapters, examples have been shown where manual manipulation of NMR data has been fully automated, and this could be repeated for the method developed here. For example, embedding the PLS processing from this chapter in the application developed in Chapter 4 would expand the breadth of samples that could be quantified and increase the utility of the spectrometer for the end user.

References

[1] Flixonase, https://public.gsk.co.uk/products/flixonase.html, Retrieved 14 Jun 2018.

[2] K. Varmuza, Filzmoser, P. (2009), Multivariate Data. Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, Florida, 31-33.

[3] S. Wold, Chemometrics; what do we mean with it, and what do we want from it? Chemometrics and Intelligent Laboratory Systems, 30 (1995) 109-115.

[4] R. G. Brereton (2009), Introduction. Chemometrics for Pattern Recognition, John Wiley & Sons, Ltd, Chichester, 1-13.

[5] S. Wold, K. Esbensen and P. Geladi, Principal component analysis. Chemometrics and Intelligent Laboratory Systems, 2 (1987) 37-52.

[6] S. L. Anderson, D. Rovnyak and T. G. Strein, Identification of Edible Oils by Principal Component Analysis of ^{1}H NMR Spectra. Journal of Chemical Education, 94 (2017) 1377-1382.

[7] S. Fan, Q. Zhong, C. Fauhl-Hassek, M. K. H. Pfister, B. Horn and Z. Huang, Classification of Chinese wine varieties using ^{1}H NMR spectroscopy combined with multivariate statistical analysis. Food Control, 88 (2018) 113-122.

[8] M. Koda, K. Furihata, F. Wei, T. Miyakawa and M. Tanokura, Metabolic Discrimination of Mango Juice from Various Cultivars by Band-Selective NMR Spectroscopy. Journal of Agricultural and Food Chemistry, 60 (2012) 1158-1166.

[9] R. G. Brereton (2003), Calibration. Chemometrics - Data Analysis for the Laboratory and Chemical Plant, John Wiley & Sons Ltd, Chichester, 272 - 338.

[10] R. G. Brereton (2009), Two Class Classifiers. Chemometrics for Pattern Recognition, John Wiley & Sons, Ltd, Chichester, 177-231.

[11] L. M. Duarte, P. R. Filgueiras, J. C. M. Dias, L. M. S. L. Oliveira, E. V. R. Castro and M. A. L. de Oliveira, Study of Distillation Temperature Curves from Brazilian Crude Oil by 1H Nuclear Magnetic Resonance Spectroscopy in Association with Partial Least Squares Regression. Energy & Fuels, 31 (2017) 3892-3897.

[12] M. H. M. Killner, E. Danieli, F. Casanova, J. J. R. Rohwedder and B. Blümich, Mobile compact ^{1}H NMR spectrometer promises fast quality control of diesel fuel. Fuel, 203 (2017) 171-178.

[13] K. Singh and B. Blümich, Compact low-field NMR spectroscopy and chemometrics: A tool box for quality control of raw rubber. Polymer, 141 (2018) 154-165.

[14] S. Piccinonna, R. Ragone, M. Stocchero, L. Del Coco, S. A. De Pascali, F. P. Schena and F. P. Fanizzi, Robustness of NMR-based metabolomics to generate comparable data sets for olive oil cultivar classification. An inter-laboratory study on Apulian olive oils. Food Chemistry, 199 (2016) 675-683.

[15] B. C. G. Percival, M.; Gibson, M.; Osman, Y.; Molinari, M.; Jafari, F.; Sahota, T.; Martin, M.; Casanova, F.; Mather, M.L.; Edgar, M.; Masania, J.; Wilson, P.B., Low-Field, Benchtop NMR Spectroscopy as a Potential Tool for Point-of-

Care Diagnostics of Metabolic Conditions: Validation, Protocols and Computational Models. High-Throughput, 8 (2019) 2.

[16] A. Nordon, C. Meunier, C. A. McGill and D. Littlejohn, Comparison of Calibration Methods for the Monitoring of a Fluorobenzene Batch Reaction Using Low-Field ^{19}F NMR, ^{1}H NMR, NIR, and Raman Spectrometries. Applied Spectroscopy, 56 (2002) 515-520.

[17] E. G. Alves Filho, L. M. A. Silva, N. V. P. Araújo, E. G. Alves, L. M. Lião and G. B. Alcantara, Qualitative and quantitative control of pediatric syrups using Nuclear Magnetic Resonance and chemometrics. Journal of Pharmaceutical and Biomedical Analysis, 153 (2018) 29-36.

Concluding Remarks

Summary of thesis

The work included in this thesis has shown how extension of a simple ^{35}Cl NMR experiment to include aspects of method validation, integrated user input, and automated acquisition and processing can result in the creation of a valuable tool for synthetic chemists. Only by fully understanding, measuring and controlling the parameters of an NMR experiment can fast, efficient and reliable analytical systems be developed for the benefit of the non-expert user. This work augments the only literature example of chloride analysis by NMR with additional knowledge and describes a practical implementation for synthetic chemists.

The same concept was applied to the development of a low field bench top NMR for rapid at line analysis using ^{1}H spectra. First, a comprehensive understanding of the spectrometer performance capability and performance was gained. Then, four primary quantitation approaches were identified to cover a range of spectral complexities and quantitative methods were developed for two of these which were presented to the end user through a simple graphical user interface that incorporates automated processing. The granularity of detail on the Spinsolve hardware and software capabilities do not appear in current literature, at least not in a consolidated form. The novel approaches developed here to synthesis accurate shaped pulses and allow the spectrometer to be used a simple walk-up analytical tool are also novel with respect to literature and are likely to be of value for current and future Spinsolve users.

Finally, multivariate analysis, the most complex of the quantitation approaches, was shown to be capable of extracting quantitative data from complex and overlapped ^{1}H NMR spectra. This last work exemplifies the more advanced mathematical approaches required to allow bench top NMR to reach its full potential in the arena of PAT. Very few examples of the application of multivariate analysis to small molecule reaction monitoring were found in the literature and the work presented here provides an additional practical example.

NMR and the future of quantitation and automation

The discussions running through this thesis have considered the development and interplay of existing NMR spectrometer hardware and software, and how these then communicate with software systems in the outside world. The accompanying work has then sought to find new ways of bringing these aspects together to create novel and valuable analytical tools for the industrial chemist.

As the work in any practical scientific thesis represents development and enhancement of existing technology, it then follows that once these have been demonstrated, the final remarks should discuss the new opportunities that now arise to continue this development into the future. These will now be discussed in the following section.

Multinuclear ion quantitation

Chloride, although a common ion in synthetic reaction mixtures, is just one of many and ion chromatography is still the industry standard for inorganic anion and cation analysis. Limitations will always exist for NMR spectroscopy of elements with lower abundancies of non-zero spin nuclei and high quadrupolar moments, such as sulphur and oxygen. However, elements such as iodine, bromine and nitrogen are becoming more accessible with higher field, cryogenically cooled magnets and so the method described in this thesis may be extended to other common ions. It may, quite rightly, be questioned if this is the best use of high-end spectrometers. The response would be that if these methods are developed similar to the one described in Chapter 2, such that no standards need to be run before and after samples, then they can be queued alongside routine ^{1}H and multinuclear structural experiments. Automation systems also have the opportunity to prioritise certain experiments, and so “lower priority” ion quantitation can be queued up during times when the instrument is at lower capacity.

Automated multivariate analysis

The integration of multivariate analysis, as developed in Chapter 5, into an automated GUI, such as that described in Chapter 4, was not achieved during the experiment work conducted in this thesis, but should be relatively trivial.

Scripts can be written in JavaScript within MestreNova that will save the processed FT spectrum as a comma separated value text document. The PLS processing from Matlab can then be embedded within further Matlab code to take a file name and path as an argument and then report the necessary quantification results as an output to a text file. The Matlab code can be compiled and called by the Python GUI seamlessly such that the end user simply choses the required method depending on the sample and is presented with the results regardless of the level of complexity of the data analysis required.

Data Integrity and the flow of information

Modern industry is increasingly turning its attention to the improvement of data flow, acquisition, interpretation and storage as potential areas to enhance and expedite business decisions. An additional benefit is the inherent improvement of data integrity that this can bring by storing more meta data and decreasing manual data entry and transcription. This has been facilitated by technological developments which have increased the availability and density of data collection, sought new application for complex data analysis (*e.g.* multivariate analysis or chemometrics) and decreased the cost of data storage media. In this context, the level of automation developed as part of the work in this thesis is just part of a longer chain of information and activity; this chain is in turn part of a wider “web” of activities that take place within the pharmaceutical development pipeline.

Further integration of the walk-up applications developed here could take batch and process information directly from a synthetic chemist’s notebook and record these details in the experimental records as meta data with the raw data and processed results. In this way a synthetic chemist could transfer the details directly from

laboratory or desk to the analytical system without manual transcription. At the other end, data could then be exported directly to the same notebook system for storage in a suitable data base, where all information pertaining to any sample generated in the company is readily accessible and searchable.

This level of automation poses several challenges around interfacing of different software packages and is also only feasible at a company level if there is harmonisation of electronic notebook and databasing systems. Further challenges are presented if the data needs to be acquired in accordance with Good Manufacturing/Laboratory Practices. Firstly, a greater level of method validation would be required to assure the confidence in the method and understand it's limitations. Secondly, the security of the information needs to be tightly controlled to prevent fraudulent tampering with the data and ensure long term accessibility of the data. Lastly, review and sign-off would be required at various stages to check the data is acceptable, interpretations are scientifically sound and the results meet the specifications required by the regulators. Again, the reports produced from all processing steps can be customised such that they present the data and calculation in a way that is human readable in a clear and understandable format for the reviewer.

Customisability of off-the-shelf hardware and software

The automated applications developed in this thesis were only possible because a certain level of customisation was possible within the vendors software. This is commonly in the form of an internal scripting language (*e.g.* C within Bruker Topspin and JavaScript with MestreNova) and the ability to control the inputs and outputs of the software externally (*e.g.* TCP/IP for Spinsolve and command line syntax for MesteNova). However, in all examples shown here, some degree of limitation was encountered. For example, although not documented here, additional phase correction routines were added to more recent version of MestreNova, but the accompanying user accessible objects were not added until much later. In other examples, values generated by processing software such as area integrals and

deconvolution fit details and residuals are displayed in the GUI, but are not accessible from within the scripting language.

Closer collaboration with vendors would help to improve the access users have to all features of the software and customise and augment these for specific applications were possible. Similarly, as highlighted in Chapter 4, assumptions were made on the hardware configuration of the Spinsolve spectrometer as these details were not available or shared by the vendor. Access to full schematics of instruments would allow the capabilities and limitations of a spectrometer to be determined from first principles, without having to rely on empirical observations as surrogates.

Industrialised solutions to quantitation

All the above suggestions for future developments have been on the assumption of using the built-in customisability of off-the-shelf hardware and software. The logical next step is to industrialise these systems using bespoke hardware and/or software. Technology similar to the automated tuning and matching accessory (ATMA) as developed by Bruker for high field NMR spectrometers (see Chapter 2) could be included in bench top systems. If the spectrometer can adjusted to a well-tuned state for each sample, then the variation in response between different solutions as observed in Chapter 4 may decrease, or at least can be ruled out as a source of variance. This may also allow an approach such a PULCON (pulse length concentration determination) to be used [1].

It was noted that Spinsolve contains a fluorinated internal lock substance. No further details of the compound used and the accompanying locking hardware have been shared by Magritek. If the identity of the compound was known and the spectrometer hardware was such that accurate integrals could be obtained, then this could potentially be used as an internal standard, either used to correct for the response of an external standard, or even to directly quantify the concentration of a sample. This could greatly enhance the built-in capability of the spectrometer as a quantitative instrument.

The automated methods developed in Chapters 2 and 4 utilise existing vendor software with Python as a “glue language” to control these through a GUI. Some, or all, of these software packages could be replaced by a single software language or platform. Although this may require significant work to integrate all current aspects into the one language, the rewards would be a more robust and seamless system for the end user, with complete control over capability and algorithms for the expert user. Spinsolve Expert and it’s underlying software platform, Prospa, do potentially allow for a user to programme direct control of the spectrometer and subsequent processing GUI [2], but the literature available is not expansive and would almost certainly require training from Magritek.

In terms of processing, rNMR [3] is an open-source package developed as an add-on to the statistical language R [4] and hence could be modified to include the processing and GUI aspects of automation. As a comprehensive and popular language for data analysis, it would be expected that approaches such as PLS would be readily accessible within the available libraries.

References

[1] G. Wider and L. Dreier, Measuring Protein Concentrations by NMR Spectroscopy. Journal of the American Chemical Society, 128 (2006) 2571-2576.
[2] B. Gouilleux, B. Charrier, S. Akoka and P. Giraudeau, Gradient-based solvent suppression methods on a benchtop spectrometer. Magnetic Resonance in Chemistry, 55 (2017) 91-98.
[3] I. A. Lewis, S. C. Schommer and J. L. Markley, rNMR: open source software for identifying and quantifying metabolites in NMR spectra. Magnetic resonance in chemistry : MRC, 47 Suppl 1 (2009) S123-S126.
[4] The R Project for Statistical Computing, https://www.r-project.org/, Retrieved 11 Feb 2019.

Appendix 1

Signal to Noise Ratio as Calculated by Bruker Topspin Software

Summary

Signal to noise is calculated using Bruker's Topspin software. Regions for the signal of interest and a typical section of baseline for noise are chosen and then the command "sino real" is used (which calculates the result based on real data, rather than "sino" which performs a magnitude calculation first).

The acceptable signal-to-noise value as calculated by the command "sino real" for quantitation limit (QL) is **12.5 : 1** and for detection limit (DL) is **3.75 : 1**.

Note – it is important that the baseline of the signal and noise regions are corrected to give a flat baseline before applying the "sino real" command.

Derivation of Sino Real Noise equation

The sino real calculation by Topspin (see Topspin processing manual) is:

Sino real = Signal / (2 x Sino Noise)

where Sino Noise is calculated by Topspin as:

$$Sino\ Noise = \sqrt{\frac{\sum_{i=-n}^{n} y(i)^2 - \frac{1}{N}\left((\sum_{i=-n}^{n} y(i))^2 + \frac{3.(\sum_{i=1}^{n} i(y(i) - y(-i)))^2}{N^2 - 1}\right)}{N - 1}}$$

Where N is the total number of points in the noise region, y(i) is the i^{th} point in the noise region and n =(N-1)/2.
Squaring both sides:

$$Sino\ Noise^2 = \frac{\sum_{i=-n}^{n} y(i)^2 - \frac{1}{N}\left((\sum_{i=-n}^{n} y(i))^2 + \frac{3.(\sum_{i=1}^{n} i(y(i) - y(-i)))^2}{N^2 - 1}\right)}{N - 1}$$

Then splitting up the fraction:

$$Sino\ Noise^2 = \frac{\sum_{i=-n}^{n} y(i)^2}{N-1} - \frac{(\sum_{i=-n}^{n} y(i))^2}{N(N-1)} - \frac{3.(\sum_{i=1}^{n} i(y(i)-y(-i)))^2}{N(N-1)(N^2-1)}$$

For large N, N-1 ≈ N and N^2-1 ≈ N^2, so:

$$Sino\ Noise^2 = \frac{\sum_{i=-n}^{n} y(i)^2}{N} - \frac{(\sum_{i=-n}^{n} y(i))^2}{N^2} - \frac{3.(\sum_{i=1}^{n} i(y(i)-y(-i)))^2}{N^4}$$

$$= \frac{\sum_{i=-n}^{n} y(i)^2}{N} - \left(\frac{\sum_{i=-n}^{n} y(i)}{N}\right)^2 - \frac{3}{N^4}\left(\sum_{i=1}^{n} i(y(i)-y(-i))\right)^2$$

RMS Noise is defined as the standard deviation, σ, of the noise region:

$$\sigma^2 = RMS\ Noise^2 = \frac{\sum_{i=-n}^{n} y(i)^2}{N} - \left(\frac{\sum_{i=-n}^{n} y(i)}{N}\right)^2$$

Then:

$$Sino\ Noise^2 = RMS\ Noise^2 - \frac{3}{N^4}\left(\sum_{i=1}^{n} i(y(i)-y(-i))\right)^2$$

Considering the second term:

$$\frac{3}{N^4}\left(\sum_{i=1}^{n} i(y(i)-y(-i))\right)^2$$

Since i = 1 to n then i ≤ n, so:

$$\frac{3}{N^4}\left(\sum_{i=1}^{n} i(y(i)-y(-i))\right)^2 \leq \frac{3}{N^4}\left(\sum_{i=1}^{n} n(y(i)-y(-i))\right)^2$$

$$= \frac{3}{N^4}\left(\frac{\sum_{i=1}^{n} n^2(y(i) - y(-i))}{n}\right)^2 = \frac{3n^4}{N^4}\left(\frac{\sum_{i=1}^{n} y(i) - y(-i)}{n}\right)^2$$

$$= \frac{3n^4}{N^4}\left(\bar{y}_{right} - \bar{y}_{left}\right)^2$$

where:

$$\bar{y}_{right} = \frac{\sum_{i=1}^{n} y(i)}{n}$$

$$\bar{y}_{left} = \frac{\sum_{i=1}^{n} y(-i)}{n}$$

This term then describes taking the noise region, and splitting it in half with the left half being the y(-i) values and the right half being the y(i) values. If the region of the spectrum used to calculate noise has been baseline corrected so that the baseline is flat across the noise region, then the average of the left half is equal to the average of the right half:

$$\bar{y}_{right} = \bar{y}_{left}$$

Therefore:

$$\frac{3}{N^4}\left(\sum_{i=1}^{n} i(y(i) - y(-i))\right)^2 = 0$$

and:

$$Sino\ Noise^2 = RMS\ Noise^2$$

or:

$$Sino\ Noise = RMS\ Noise$$

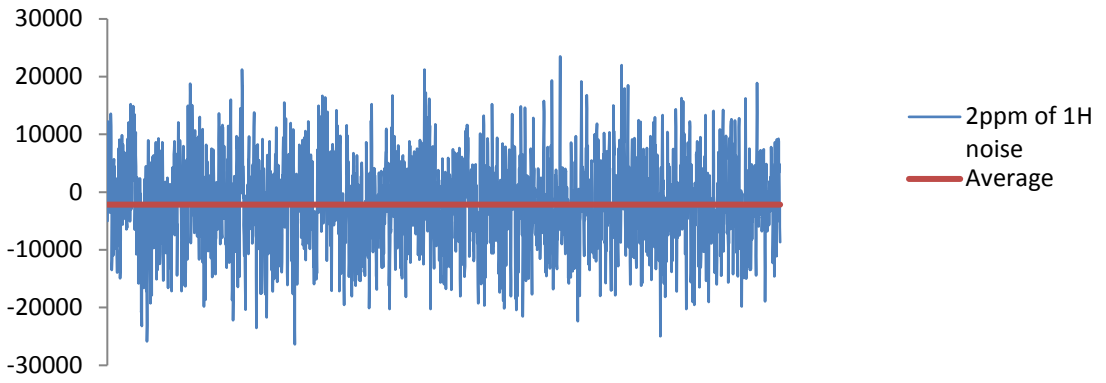
Finally,

$$Sino\ Noise = \frac{Signal}{2 \times RMS\ Noise}$$

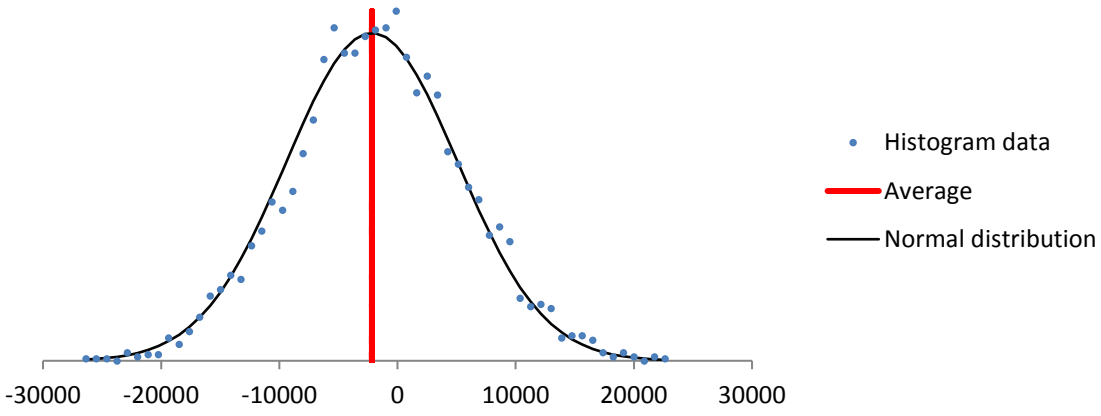
For illustration purposes, we have compared the results of signal-to-noise ratio calculated for an experimental ^{1}H NMR spectrum by using the Topspin “sino real” command and by using the equation above. The two results were identical; both the Topspin “sino real” command and the above equation yielded the signal-to-noise ratio of 121.2.

Statistical distribution of noise

A 2 ppm region of noise, 3250 data points, from a ^{1}H NMR spectrum was taken:

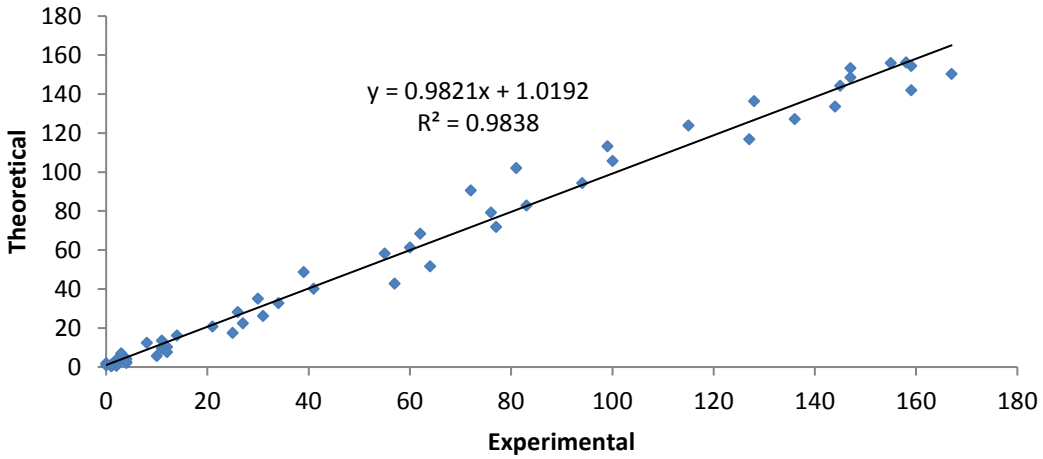


The data were then plotted in the form of a histogram:



It can be seen that the random noise in an NMR spectrum can be approximated to a normal distribution (known as Gaussian noise), where RMS Noise = σ. This can be

proved by plotting experimental values from the histogram with those calculated from a model normal distribution:



Comparison with GSK and ICH limits

ICH2 and internal GSK documents define acceptable signal-to-noise for QL and DL of an analytical method as 10 : 1 and 3 : 1 respectively. This is for situations where the noise is calculated as peak-to-peak.

GSK documentation derives acceptable signal-to-noise ratio (SNR) for QL when RMS noise is used:

$$SNR(GSK)^{QL} = \text{Signal / RMS Noise} = 25 \quad \text{i.e. SNR} = 25$$

Since the Topspin sino real calculation uses:

$$\text{Sino real}^{QL} = \text{Signal / (2 x Sino Noise)} = \text{Signal / (2 x RMS Noise)}$$

$$= ½ \times SNR(GSK)^{QL} = 25/2 = 12.5$$

Following the calculation from GSK documentation, acceptable s/n for DL with RMS noise would then be:

$$SNR(GSK)^{DL} = Signal / RMS\ Noise = 7.5 \quad i.e.\ SNR = 7.5$$

$$Sino\ real^{DL} = ½ \times SNR(GSK)^{DL} = 7.5/2 = 3.75$$

Therefore, the acceptable signal to noise ratios for QL and DL calculated by sino real of Topspin are 12.5 and 3.75, respectively.

It can be shown that if a data point which gives a sino real figure at the DL criteria, i.e. sino real = 3.75, it is highly unlikely that the data point is due to noise:

$$Sino\ real^{DL} = 3.75 = ½ \times Signal / RMS\ Noise = ½ \times Signal / \sigma$$

$$Signal = 2 \times 3.75\sigma = 7.5\sigma$$

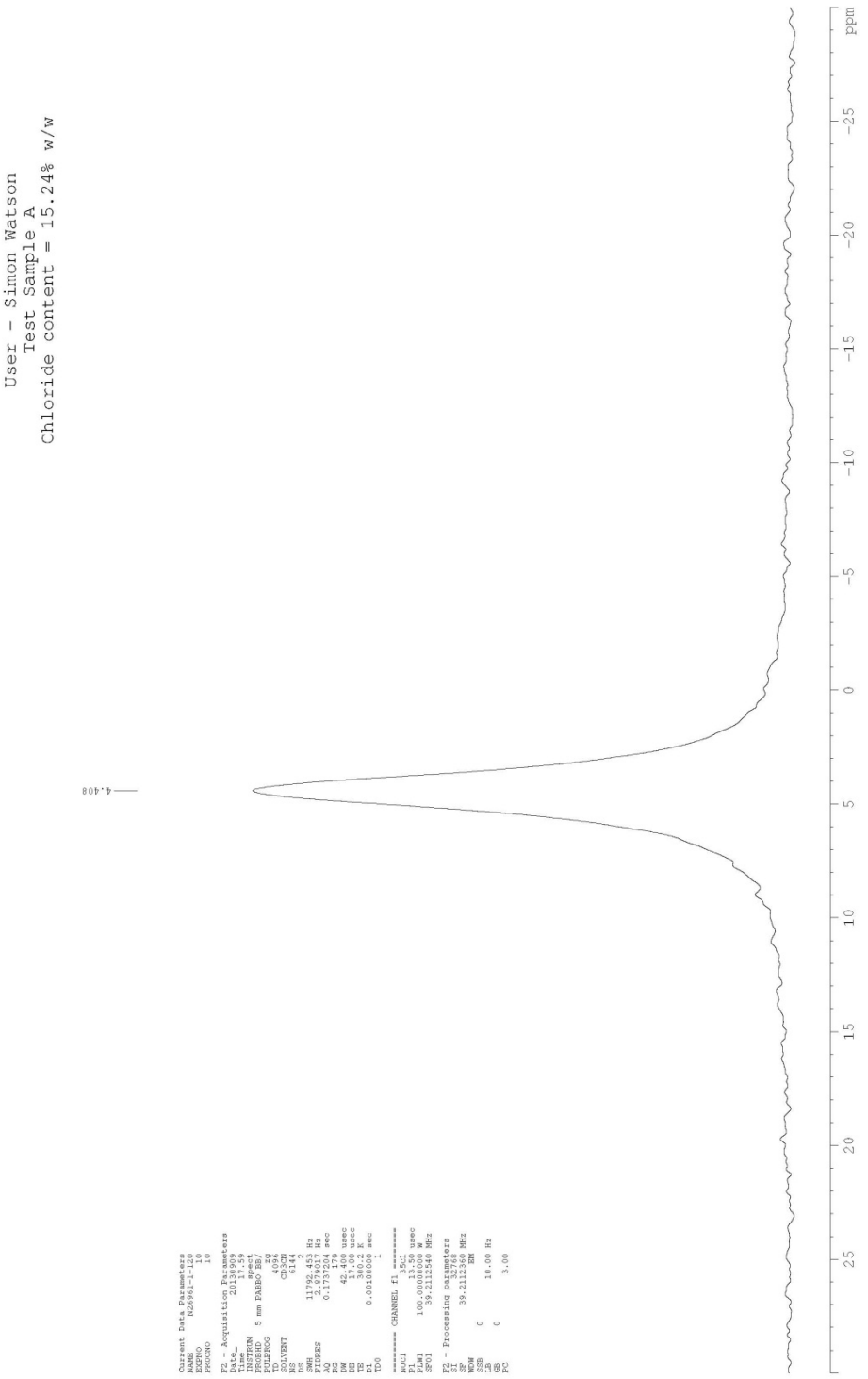
Calculating the probability from of the area outside the range $\{-7.5\sigma,7.5\sigma\}$ of a normal distribution:

$$P = 1 - Integral\ \{-7.5\sigma,7.5\sigma\} = 1 - 0.9999999999999303 \approx 0$$

Where Integral $\{-7.5\sigma,7.5\sigma\}$ is the integral of the normal distribution of the noise for the interval $\{-7.5\sigma,7.5\sigma\}$. Therefore, the probability of a point of noise outside the interval $\{-7.5\sigma,7.5\sigma\}$ is extremely small, practically zero.

Appendix 2

Example PDF Output from 35cl_quant AU Programme


User - Simon Watson
Test Sample A
Chloride content = 15.24% w/w
4.408
ppm
25
20
15
10
5
0
-5
-10
-15
-20
-25

Appendix 3

Standardless, Automated Determination of Chlorine-35 by ^{35}Cl Nuclear Magnetic Resonance

Analytical Letters

ISSN: 0003-2719 (Print) 1532-236X (Online) Journal homepage: http://www.tandfonline.com/loi/lanl20

Standardless, Automated Determination of Chlorine-35 by ^{35}Cl Nuclear Magnetic Resonance

Simon A. Watson, Andy J. Edwards & John A. Parkinson

To cite this article: Simon A. Watson, Andy J. Edwards & John A. Parkinson (2017) Standardless, Automated Determination of Chlorine-35 by ^{35}Cl Nuclear Magnetic Resonance, Analytical Letters, 50:1, 161-172, DOI: 10.1080/00032719.2016.1172636

To link to this article: http://dx.doi.org/10.1080/00032719.2016.1172636

Accepted author version posted online: 24 May 2016.
Published online: 24 May 2016.

ANALYTICAL LETTERS
2017, VOL. 50, NO. 1, 161–172
http://dx.doi.org/10.1080/00032719.2016.1172636

Taylor & Francis
Taylor & Francis Group

SPECTROSCOPY

Standardless, Automated Determination of Chlorine-35 by ^{35}Cl Nuclear Magnetic Resonance

Simon A. Watson[a], Andy J. Edwards[a], and John A. Parkinson[b]

[a]GlaxoSmithKline Medicines Research Center, Stevenage, Hertfordshire, United Kingdom; [b]WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, United Kingdom

ABSTRACT

A robust, fully automated, walk-up method is reported to quantify chloride in samples using ^{35}Cl nuclear magnetic resonance. Minimal user input is required, no standards are acquired at the time of analysis; and the submission, acquisition, processing, and production of results are seamlessly integrated within existing software. The method demonstrated good linearity with $R^2 = 0.999$ over three orders of magnitude of analyte concentration. The results were highly independent of analyte functionality, and the stability of instrument response was sufficient that analyses of additional standards were not required for a period of several months. At a nominal sample concentration of 10 mg/ml in D_2O at 400 MHz, detection and quantitation limits of 0.1 and 0.5% (w/w) were achieved in a 1-h analysis time. Robust methodology was achieved by applying a rigorous approach to method development and validation to determine and evaluate fully the time- and sample-dependent factors that affect quantitation in these measurements.

ARTICLE HISTORY
Received 17 November 2015
Accepted 27 March 2016

KEYWORDS
^{35}Cl; chloride; nuclear magnetic resonance; pharmaceutical analysis

Introduction

The maxim that "nuclear magnetic resonance (NMR) is inherently quantitative" has become a central dogma in the literature and presentations. Although correct, it may be more accurate to state that the *phenomenon* of NMR is inherently quantitative, as it is not necessarily true that any given NMR spectrum of a sample as acquired under a specific set of experimental conditions will yield quantitative results that are accurate and precise (Freeman 1987; Claridge 1999). To develop and validate a method capable of yielding acceptably quantitative data, two approaches can be considered.

In the first approach, only the basic parameters affecting quantitation are determined during method development (*e.g.*, T_1, T_2). Conditions are then either set to ensure a quantitative response directly from the signal integrals or are optimized for time efficiency and any expected decrease in response is calculated using the measured relaxation parameters and experiment delays. All other instrument-, sample-, and time-dependent factors are taken into account by performing additional validation and calibration at the time of analysis through spiking known quantities of internal standards into the sample, or acquiring spectra of external standards immediately pre and post sample analysis. Alternatively, an artificial reference peak can be induced in the spectrum using electronic reference to access

CONTACT Simon A. Watson simon.a.watson@gsk.com GlaxoSmithKline Medicines Research Center, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK.

in vivo concentrations—ERETIC (Akoka, Barantin, and Trierweiler 1999), or quantification by artificial signal—QUANTAS (Farrant et al. 2010). Quantitation is then achieved by comparing the relative integrals of the analyte and artificial reference peaks, in a process analogous to using an internal standard. Another way to determine instrument and sample variation at the time of analysis is to apply the principles of reciprocity (Hoult and Richards 1976) by measuring the pulse width for each sample and comparing the values to those acquired on a previous reference. This has been demonstrated by application of pulse length-based concentration determination—PULCON (Wider and Dreier 2006).

In the second approach, rather than performing continual validation and/or calibration at the time of analysis for each sample, the factors affecting signal response are considered and explored and then a region over which the method can operate without further validation is defined. In this way, a method can be developed and validated, such that it is robust, accurate, and precise but does not require additional experimentation at the time of analysis.

For either approach, the final time-saving component in developing a completely automated method is the creation of walk-up automation software that provides a simple interface requiring minimal user interaction. Although more complex NMR data can require many hours of manual processing, for routine processes and calculations, it can generally be assumed that if a task or series of tasks can be performed by an analytical expert using a computer, then it can be automated. This may require either the writing of entirely new pieces of software or, more frequently, the coding of custom programs to enable more flexible access to the capabilities within existing automation procedures.

For most analysts, NMR would not be considered, by first intent, for routine chloride analysis. Quantification of chloride ions would usually be performed by ion chromatography (IC) (Christian 2004). IC is a highly industrialized technique that can be hyphenated to advanced automation systems capable of weighing, dissolving, and injecting samples into the instrument, processing data, and reporting results. Run times can also be short (<10 min per chromatogram) for basic separation of anions.

While IC is a common technique, it may not, however, be available in all analytical laboratories and, if present, may not be set up as an automated walk-up system. If not automated, an IC system would require human intervention in the form of analytical experts to prepare samples, queue them on an autosampler, integrate the chromatographic data, and report the results. Regardless of automation, standards and diluent blanks need to be prepared and analyzed before and after the genuine samples, increasing the overall analysis time. As with all liquid chromatography techniques, IC also needs a constant flow of mobile phase to operate, requiring further time on the part of the analytical expert to prepare solutions, flush columns, and dispose of waste.

High-field NMR spectrometers are now a standard feature of many chemical laboratories and are ideally suited for automation and walk-up analysis. With the exception of scheduled routine maintenance and housekeeping (*e.g.*, regular magnet fills of liquid nitrogen and helium, emptying of the autosampler, and checking overall instrument performance), the systems require minimal or no daily intervention. The proceeding work gives an example of extensive method development for the quantification of chloride in solutions containing other organic or inorganic analytes by ^{35}Cl NMR. The work demonstrates how the seamless integration of simple programming code into established automation software can allow the handling of sample preparation information, data

processing, quantification, and reporting of data within a single piece of software. There are relatively few examples in the literature of using ^{35}Cl quantitatively in this way.

In one reported example (Lim and Lee 2006), quantitation was achieved by loading the sample into a 5-mm NMR tube into which was co-axially placed a sealed 3-mm NMR tube containing a solution of tetraethylammonium chloride as an internal standard. While this work demonstrates the general principle, they are no examples of extensive validation or automation for this particular application. It should, however, be noted that there are examples of extensive validation of ^{1}H quantitative methods conducted across multiple laboratories and considering many aspects of uncertainty (Malz and Jancke, 2005).

The factors considered likely to affect signal response and the final results were investigated, many of which are in line with general guidance for method validation (ICH Q2 (R1) 1996) and some that are specific to NMR spectroscopy, namely, linearity, detection, and quantitation limits, repeatability, coil loading, accuracy (including sample matrix effects), and processing.

Experimental

All data were acquired on a 400 MHz Bruker Avance-III NanoBay NMR spectrometer equipped with a BCU-05 temperature controller using a 5-mm probe with a broadband channel tunable to resonant frequencies in the range ^{31}P to ^{15}N (also including ^{19}F), automatic tuning and matching, z-gradients, and a BACS-60 autosampler. The spectrometer was controlled by Bruker Topspin version 3.0, running under Microsoft Windows 7, and using software module IconNMR for automated spectrometer control. Data were typically acquired over a frequency width of 11792 Hz into 4096 data points using a pulse and acquire sequence with a 90° pulse calibrated at 13.4 µs at 100 W, an acquisition time of 174 ms, a sample temperature of 300 K, and a dead time of 17 µs to minimize ring-down artifacts at the start of the free induction decay. The ^{35}Cl T_1 of a solution of NaCl in D_2O at a concentration of 10 mg/ml chloride was measured to be 32 ms (comparable to a calculated literature value of 40 ms in water at infinite dilution (Hertz 1973)). As the acquisition time is $>5 \times T_1$, an arbitrarily short relaxation delay of 1 ms was set. Data were acquired using 6144 transients per experiment, requiring a total of approximately 21 min of instrument time. Data were Fourier transformed with 30 Hz exponential line broadening function, and the free induction decay was zero filled with a factor of 4. All data processing was handled *through* a custom written program running under Topspin version 3.0. The program is available from the author on request.

Linearity

Solutions of sodium chloride in D_2O were prepared with chloride concentrations of 0.01, 0.02, 0.06, 0.1, 0.2, 0.4, 0.6, 1.0, 1.5, 2.0, 2.5, 5.0, and 10.0 mg/ml.

Accuracy

Solutions were prepared by dissolving 20 compounds, representing a range of common organic compound functionalities and salts, in a stock solution of sodium chloride in acetonitrile/D_2O 50/50 (v/v) with a nominal concentration of organic compound of 10 mg/ml

and a nominal chloride content of 2.0% (w/w). Further solutions of inorganic chlorides were prepared containing different counter ions. Solutions of magnesium chloride, potassium chloride, lithium chloride, ammonium chloride, calcium chloride, and cesium chloride were prepared in acetonitrile/D_2O 50/50 (v/v), so that a nominal concentration of 2.0% (w/w) of chloride was obtained for each.

Repeatability and instrument stability

A solution of sodium chloride in D_2O/acetonitrile 50/50 (v/v) with chloride concentration of 0.2 mg/ml (equivalent to 2.0% (w/w) with respect to a nominal sample concentration of 10 mg/ml) was analyzed periodically over a period of 24 h (24 spectra in total), then over a period of 7 days (54 spectra in total), and finally over a period of 10 weeks (139 spectra in total).

A subset of the linearity experiments was repeated approximately 3 years after the original analyses. Solutions of sodium chloride in D_2O were prepared with chloride concentrations of 0.1, 0.2, 0.5, 1.0, 2.5, 5.0, and 10.0 mg/ml.

Results and discussion

Linearity

Chloride signal (area units) measured as a function of chloride concentration (mg/ml) in the range 0.01–10 mg/ml showed a linear response fit with a correlation coefficient >0.999 and an equation of $y = 2.9307 \times 10^8 x + 3.7840 \times 10^6$. Linearity was demonstrated across a range representing 0.1–100% (w/w) compared with a nominal concentration of organic analyte of 10 mg/ml. These data were comparable to linearity achieved previously (Lim and Lee 2006) over the range 0.02 mg/ml to 5 mg/ml with a correlation coefficient >0.999 on a 700 MHz NMR spectrometer.

The response of the receiver across a range of gain values from 0.25 to 179 was also assessed and found to be acceptably linear with $R^2 = 0.999$. The regression curve was then used in the processing program to determine the chloride concentration from the absolute peak area, post-Fourier transformation, and phase correction (*vide infra*).

Signal-to-noise ratio

The minimum signal-to-noise ratios acceptable to determine limit of detection and limit of quantitation are commonly quoted as 3:1 and 10:1, respectively. However, these values are often taken from guidance intended for chromatographic data and by measuring peak-to-peak noise using a calculation of signal-to-noise$_{p\text{-}p} = 2H/h$ where H is the peak height from the zero point of the baseline and h is the height of the peak-to-peak noise. Topspin, as used to process these data, determines noise based on a root mean square calculation and determines signal-to-noise$_{\text{sino real(RMS)}} = H/2$ h. Applying a conversion of $h_{\text{peak-to-peak}} = 5 \times h_{RMS}$, then this equates to minimum detection and quantitation limits using the "sino real" command of 3.75:1 and 12.5:1, respectively.

Under the experimental conditions used, the 0.01 mg/ml chloride standard yielded an NMR response with signal-to-noise$_{\text{sino real}} = 3.92$:1. This is greater than the required value

of 3.75:1 and so is acceptable for the detection limit. The 0.06 mg/ml chloride standard yielded an NMR response with signal-to-noise$_{\text{sino real}} = 16.04{:}1$. This is above the limit of 12.5:1, and so is acceptable for the quantitation limit. It is possible to estimate the percentage chloride content that a peak at 12.5:1 would represent and thus estimate the theoretical quantitation limit:

$$\text{Estimated quantitation limit} = \frac{12.5}{16.04} \times 0.06\% = 0.0468\,\text{mg/ml chloride} \qquad (1)$$

Thus, if a sample of nominal concentration 10 mg/ml was analyzed using the conditions described for the current method, then detection and quantitation limits for chloride would be 0.1 and 0.5% (w/w), respectively. This will hold true for any solutions giving rise to a chloride response with the same half-height linewidth. As the actual samples are more likely to be dissolved in acetonitrile-d_3/D_2O 50/50 (v/v) (*vide infra*), the signal-to-noise ratio and detection and quantitation limits are calculated for each spectrum and used as limits against which to report data within the processing program. This step improves the robustness of the method by taking into account the quality of each spectrum at the time of analysis and will avoid reporting data against an inappropriate detection and quantitation limit if there is excessive noise in a sample spectrum due to unforeseen instrument problems.

The detection and quantitation limits shown above were obtained with run times of approximately 20 min; the rationale that longer run times for data acquired during working hours on a walk-up analytical instrument would be impractical for reasons of maintaining sample throughput. Longer run times submitted as "overnight experiments" would then achieve lower detection and quantitation limits and could be used on samples with limited availability or solubility. Sensitivity can also be increased in IC methods by increasing the injection volume. However, NMR offers the additional advantage that while there will be a maximum injection volume for a given IC method, there is theoretically no upper limit to the number of transients that can be acquired in an NMR experiment.

Repeatability

Replicates over the initial 24-h time period returned an average result of $1.57 \pm 0.06\%$ (w/w) chloride for a total of 25 results, where the error is quoted as ± 1 standard deviation. While this level of variability would be considered acceptable for a quantitative method, linear regression analysis of the data showed possible evidence of a decreasing trend in response over time with $R^2 = 0.3922$.

There was no evidence of precipitation in the sample, as the NMR tube was sealed with parafilm to prevent evaporation, and insolubility would not be an expected issue for an aqueous solution of sodium chloride at the concentration prepared here.

The presence of a continually decreasing response is a major issue where the goal of a method is to quantify based on an absolute response over a period of time. For this reason, a spectrum was acquired on the same sample once every day for further 7 days and then once every hour for a period of 24 h.

Error bars of ± 1 standard deviation for the two 24-h periods overlap with each other, and so although there is a drop in the average response over the 1-week period, the change is not significant compared to the degree of variability of response over 1 day.

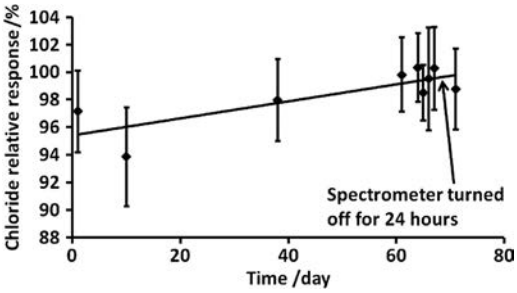


Figure 1. Change in chloride response over a period of 10 weeks, error bars show ±one standard deviation. The variation is given by $y = 6.18 \times 10^{-2}\ x + 9.5435 \times 10^{1}$, $R^2 = 0.6551$.

To definitively establish and quantify the trend, data were acquired over a period of approximately 10 weeks. The data were typically acquired in blocks of at least 12 experiments per 24 h-period during this time. Less trend was apparent over this 10-week duration compared to the data acquired every hour over a 24-h period. This fact is confirmed by taking a plot of the average values for each 24-h time period (Figure 1). There is a variation in the average values for a given day but no overall trend over time as there is overlap between error bars for all days. Furthermore, the standard deviation of the average values across the days is 2.03%, whereas the standard deviations for the values for each day range from 2.0 to 3.8%. This shows that there is more intraday variation than interday variation in the data. As shown in Figure 1, the spectrometer was turned off for a period of approximately 24 h, and no significant differences were observed in the data acquired before and after this time.

To further exemplify the effect of the level of variation upon actual measurement, the day with the greatest level of variation in results would yield a chloride concentration of $1.50 \pm 0.08\%$ (w/w), equating to 5.0% relative standard deviation. This is an acceptable level of variation for a typical analytical method and will be considered to be the uncertainty in measurement for this method. For comparison, the average of all data over this period yielded a chloride content of 1.58% (w/w).

As all data discussed here were acquired with the same NMR tube, and probe tuning and matching were not adjusted between consecutive data accumulations on the same day, it follows that the sources of the observed variations must lie in either the instrument response or the data processing. It should also be noted that day-to-day variability is usually assessed as part of intermediate precision. This should include instrument-to-instrument and analyst-to-analyst variability. In this instance, only one spectrometer was available on which to use the method, and the automated processing was such that the only analyst variation that would be encountered here would be due to sample preparation.

The method, and therefore by inference the response of the spectrometer, was shown to be robust over a period of at least 10 weeks without the need for further acquisition of standards. This included 24 h of complete spectrometer downtime during the 10-week period.

The repeat linearity experiments after 3 years yielded a regression equation that showed a change in instrument response. This change was significant enough to require the regression parameters to be updated in the experiment at this time. Although the goal of this work is to demonstrate that modern NMR spectrometers are capable of quantitation by absolute response without the need for standards, it is advisable to perform regular

performance checks for any analytical method in use over time. For example, a system suitability check, in the form of acquisition and recording of data on a standard sample on the day of analysis, would confirm the overall performance of the instrument and monitor any progressive changes, such as drift in amplifier response that may require regression data to be re-acquired.

Additional replicate analysis was also performed with a total of seven batches of NMR tubes across three manufacturers. It was found that there was no significant variation between batches of the tubes from the manufacturer typically used for this analysis, but a difference in response was observed with tubes from different manufacturers. This shows that for a method that relies on absolute response, care must be taken to use NMR tubes of the same manufacturer and specification for both validation and subsequent analyses.

Coil loading

Here we consider the application of the principle of reciprocity to NMR spectroscopy (Hoult and Richards 1976): "RF susceptibility and conductivity effects which influence the coupling of the processing nuclear magnetization to the coil will have the same effect on the radiofrequency field (produced by the same coil at the same frequency) that causes the excitation of the same parcel of nuclei." (Burton, Quilliam, and Walter 2005). In other words, anything affecting energy transfer between the nuclear spins within a sample and the coil within the probe will affect the transfer equally in both directions. The efficiency of this transfer is characterized by the quality factor that is proportional to signal intensity. For this reason, the spectrometer response may be affected not only by the presence of other species in the sample matrix but also by the concentration of the analyte.

If, for example, increased salt concentration leads to a lowering of the quality factor, then more energy will be required by the amplifier to achieve a true 90° pulse. From the other perspective, a change in quality factor with the same pulse power and width will lead to a decrease in the effective flip angle in the sample and a reduced response. For this reason, methods have previously been devised to account for these differences by measuring the 90° pulse width, p_{90}, for each sample and then using the difference between this value and the p_{90} of a reference sample to determine the increase or decrease in response with respect to that reference sample. In this way, a quantitative response may be calculated for each sample.

The extent of the effect was investigated with the selected analytes to determine if pulse calibration was required as part of the method. This was investigated indirectly, by observing the difference in the tuning/matching curve, and directly, by measuring the change in p_{90} for different samples.

The tuning and matching curve was recorded for all samples used in the linearity measurements and no significant differences were observed over this range. Although this does not take into account any differences from the presence of organic molecules, it is expected that the inorganic ions of Na^+ and Cl^- will have the greatest effect on the quality factor. We can conclude that the effect on tuning and matching, and hence the quality factor, is negligible for the concentration range used in this work.

The ^{35}Cl p_{360} was determined for a series of samples with varying chloride concentration, and the p_{90} subsequently calculated as $p_{360}/4$ (Figure 2). A nonlinear relationship is observed between pulse width and chloride concentration, with the lower concentrations

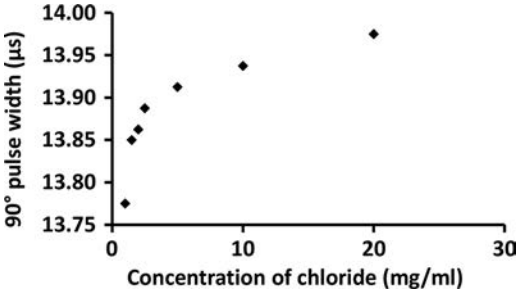


Figure 2. Effect of chloride concentration on ^{35}Cl pulse width.

having shorter pulse widths as the quality factor increases. The overall change in p_{90} over this range is no more than 0.2 µs. As the flip angle deviates from p_{90} by θ, so the intensity of the signal decreases from the maximum value by the factor cos θ (Burton, Quilliam, and Walter 2005). For a change of ±1 µs, this equates to a decrease in the signal of 2.51% relative to the maximum. For a sample containing 1.58% (w/w) chloride (*i.e.*, the average value from the repeatability experiments), a decrease in 0.04% (w/w) will result. This is less than the error of ±0.08% (w/w) obtained from replicate analysis of the same sample.

Due to low signal-to-noise ratio, it was not possible to accurately measure p_{90} for samples of lower concentration than 1 mg/ml chloride. However, no significant changes in the tuning and matching curve were observed at these concentrations and a linear response has already been established down to a chloride concentration of 0.1% (w/w).

For the work discussed here in an aqueous solution, it has been shown that the method is robust with respect to the effect of chloride ion concentration upon the quality factor for this probe. This further supports the proposition that subsequent analyses do not require standards of matching chloride concentration to be run at the time of data acquisition or an approach such as PULCON is required to correct for the effects of coil loading.

Accuracy

The recovery results for the samples prepared containing *ca*. 2% (w/w) chloride are presented in Table 1. The theoretical chloride content is calculated from the weight of the analyte, dilution volume, and quantity of fortified chloride. The measured chloride content is the result as determined by the automated processing method. Recovery is then calculated as the percentage by which the experimentally measured value differs from the theoretical value.

Only one value falls just outside 90–110% recovery, namely, butylhydroxy toluene, although it should be noted that only single preparations were made for each sample. Analysis of the solutions of different inorganic chlorides yielded recovery values 90–110% of the theoretical chloride concentration, based on the formula for each compound, for all samples.

For the purposes of the method, we have demonstrated that the method is robust with respect to sample matrix, for the presence of a variety of common chemical functionalities, and different chloride counterions. The effect of chlorides containing paramagnetic metals was not explored due to the hygroscopic character of the available metal chlorides, hence

Table 1. Recovery values for analyte solutions in acetonitrile-d_3/D_2O containing chloride.

Analyte	Theoretical chloride concentration % (w/w)	Experimentally measured chloride concentration % (w/w)	Recovery (%)
Acesulfame potassium	1.93	1.83	94.8
Quinine	1.93	1.87	97.0
Succinimide	1.80	1.78	99.0
Paracetamol	1.91	1.89	99.1
Benzoic acid	1.81	1.73	95.5
Succinic acid	1.83	1.80	98.4
Butylhydroxytoluene	1.84	1.65	89.7
Ascorbic acid	2.13	2.09	98.3
Vanillin	1.83	1.87	102.3
2-Hydroxy-5-methylbenzaldehyde	1.95	1.84	94.6
Benzophenone	1.88	1.82	96.8
3-Methyl-4-nitrobenzoic acid	1.87	1.89	101.1
Benzenesulphonic acid	1.94	1.92	98.8
Glucose	1.94	1.86	96.0
Benzeneboronic acid	1.88	1.71	90.9
Glycine	1.90	1.81	95.5
Sodium dodecyl sulfate	1.84	1.78	96.7
Potassium carbonate	1.92	1.79	93.4
Di-Sodium hydrogen aphosphate	1.90	1.83	96.5
Sodium acetate	1.90	1.85	97.3

this aspect should be considered and investigated before the method is applied to these samples.

Automation

The ultimate goal of this work was to develop a fully automated ^{35}Cl NMR quantitative chloride experiment for a walk-up NMR spectrometer. To achieve this, the method not only had to be developed with respect to experimental conditions and parameters but also had to be capable of data acquisition, processing, and reporting with minimal user interaction required.

For most measurements in a high throughput, automation setting, the user is usually only required to input a dataset name, the solvent, the type of experiment required, and title information such as the batch and/or sample preparation details. For the automated ^{35}Cl experiment, however, the user is also required to input sample weight and solution volume. The calculation to produce the result would then need to include other information from acquisition parameters, such as receiver gain and number of scans, and also the regression parameters determined from the linearity experiments. Finally, a result would be reported back to the user in the correct format. A processing program was therefore written to handle the different user inputs, process the data, and calculate and report the result.

The Topspin user acquisition parameters (USERA1 and USERA2) allow users to input the sample weight and volumes directly into the walk-up IconNMR software. Post acquisition, the program then Fourier transforms the data and performs phase correction and baseline correction. The main peak was integrated and the final chloride result was calculated, added to the title of the experiment and the spectrum, including title, is printed to a

PDF file. The only user intervention required is the sample preparation and entering the initial weight and volumes into IconNMR. Two specific aspects of the processing program are discussed in more detail below.

Determining the optimal integration region

During the development of the processing package, it was found that the peak width of the chloride resonance varied considerably depending on the other species present in a solution. For example, solutions of NaCl in D_2O gave half-height linewidths of *ca.* 20 Hz, whereas for some analytes dissolved in D_2O/acetonitrile 50/50 (v/v), this would rise to as much as 80 Hz. Setting an arbitrarily broad region for integration increases the variability of the results by enhancing the noise region included in the integral for more narrow resonances.

To resolve this issue, an additional subroutine was written into the program to determine the approximate linewidth of the chloride resonance. This function directly accesses the binary file that contains the real Fourier transform data produced by Topspin to determine values of individual data points.

In this function, the absolute intensity of the highest peak (apex) is first read in from the binary file. Then, the next two data points to the right of the apex (*i.e.*, at lower ppm values) are read. If both of these are greater than half of the apex intensity, then the process is repeated for each subsequent data point, moving stepwise down the right-hand side of the peak (Figure 3). When point 2 is less than the half-height intensity, the program determines whether peak 1 or 2 is closest to the half-height and then stores the index of this point. The program moves the pointer back to the apex and repeats the process moving stepwise down the left-hand side of the peak, stopping again when the point nearest to half-height intensity is found.

By subtracting the index value from the half-height point on each side of the peak, the approximate peak width may be determined in units of data points. Finally, multiplying this number by the spectral width in Hz and dividing by the number of points in the real spectrum yield the approximate peak width in Hz.

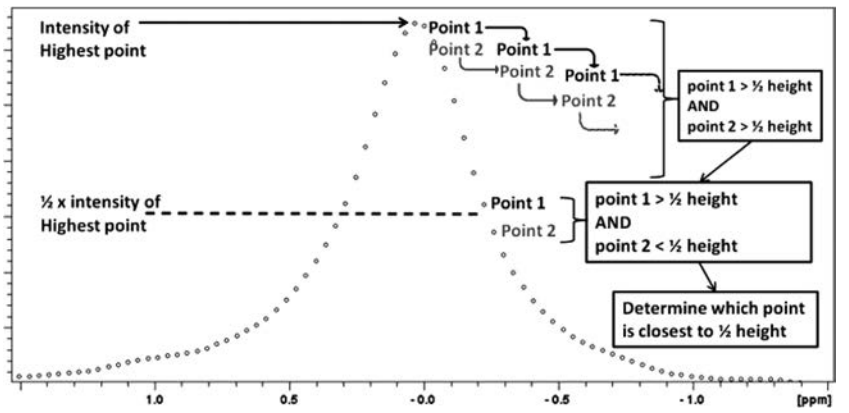


Figure 3. Schematic of the principles to measure linewidth at half-height.

The integral region is then set such that minimum noise is included in the integration calculation. Note that integration is achieved not by the built-in functions in Topspin but by a similar process of reading the intensity of the data points over the region of interest and adding these values to produce a total absolute area.

Automatic phase correction

In the ^{35}Cl NMR spectra discussed here, with the current experimental set up, the zero-order phase correction required to produce a resonance in the FT NMR spectrum of pure absorption phase is close to 180°. However, slight changes in the relative phase difference between the pulse and the receiver from one sample to the next require the zero-order phase correction to be adjusted for each spectrum.

This can be performed manually, which can be highly subjective, or by use of automated phase correction routines. The automated routines available in Bruker Topspin were not capable of achieving a consistent absorption phase peak in this case. Custom routines have been developed that achieve robust phase correction of ^{1}H spectra through a combination of baseline correction, area minimization, and negative area penalization (De Brouwer 2009). However, the ^{35}Cl spectra acquired here contain a single peak, and the baseline only needs to be considered over a narrow region around this peak. For this reason, a more simplified approach proved acceptable and a custom routine was written into the processing program.

A basic measure of ideal peak shape is the symmetry of the peak. If an imaginary line is drawn down the peak from the apex, and the distance from half way up the line to the edge of the peak on each side is taken, then the difference between these two distances will be at a minimum when the peak is symmetrical.

This approach has the advantage that it can be completely controlled and run *through* a processing program that directly manipulates the sample spectrum and then determines the required values by accessing the binary data. Figure 4 shows a plot of results as acquired by a subsequent modification of the processing program.

The asymmetry factor is calculated by the absolute difference between the apex line at half-height and each edge of the peak.

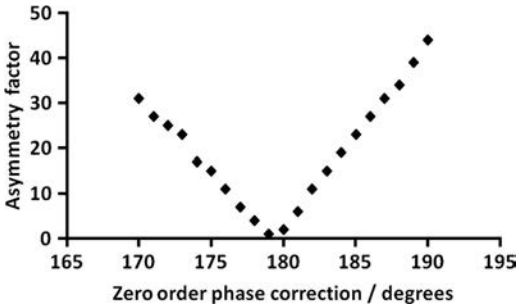


Figure 4. Plot of asymmetry factor as a function of zero-order phase correction.

Conclusion

We have shown here that specific, robust, and quantitative NMR methodology can be developed by rigorously investigating the factors affecting signal response. This has been conducted informally in this work, but it could take the form of a comprehensive risk assessment prior to method development. We have also shown that it is possible to quantify species without the need for standards at the time of analysis, as long as the receiver/transmitter coil response can be demonstrated to be sufficiently independent of sample matrix and analyte loading.

Finally, it was demonstrated that when existing processing macros are unable to produce the required data quality, custom programs can be embedded into automation software to ensure a seamless walk-up user experience from submission to results. In summary, we have shown that existing NMR spectrometers can be used flexibly to allow synthetic chemists to save considerable time and effort in acquiring accurate and sensitive data on a common analyte. The time expended in the development and validation of this and similar methods is returned from new and powerful walk-up tools to more rapidly analyze samples.

Acknowledgment

Andy Gibbs of Bruker UK is thanked for his advice on Bruker AU Programming.

References

Akoka, S., L. Barantin, and M. Trierweiler. 1999. Concentration measurement by proton NMR using the ERETIC method. *Analytical Chemistry* 71 (13):2554–57. doi:10.1021/ac981422i

Burton, I. W., M. A. Quilliam, and J. A. Walter. 2005. Quantitative 1H NMR with external standards: Use in preparation of calibration solutions for algal toxins and other natural products. *Analytical Chemistry* 77 (10):3123–31. doi:10.1021/ac048385h

Christian, G. D. 2004. Liquid chromatography. In *Analytical chemistry*, ed. D. Brennan, 625–27. Oxford, UK: Wiley.

Claridge, T. D. W. 1999. One-dimensional techniques. In *High-resolution NMR techniques in organic chemistry*, 114. Oxford, UK: Elsevier.

De Brouwer, H. 2009. Evaluation of algorithms for automated phase correction of NMR spectra. *Journal of Magnetic Resonance* 201 (2):230–38. doi:10.1016/j.jmr.2009.09.017

Farrant, R. D., J. C. Hollerton, S. M. Lynn, S. Provera, P. J. Sidebottom, and R. J. Upton. 2010. NMR quantification using an artificial signal. *Magnetic Resonance in Chemistry* 48 (10):753–62. doi:10.1002/mrc.2647

Freeman, R. 1987. Intensities. In *A handbook of nuclear magnetic resonance*, ed. J. E. Baldwin and R. M. Williams, 101–05. Harlow, UK: Longman Scientific and Technical.

Hertz, H. G. 1973. Magnetic relaxation by quadrupole interaction of ionic nuclei in electrolyte solutions Part II: Relaxation at finite ion concentrations. *Berichte der Bunsengesellschaft für physikalische Chemie* 77 (9):688–97.

Hoult, D. I., and R. E. Richards. 1976. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *Journal of Magnetic Resonance (1969)* 24 (1):71–85. doi:10.1016/0022-2364(76)90233-x

ICH Q2(R1). 1996. Validation of Analytical Procedures: Text and Methodology (CPMP/ICH281/95)

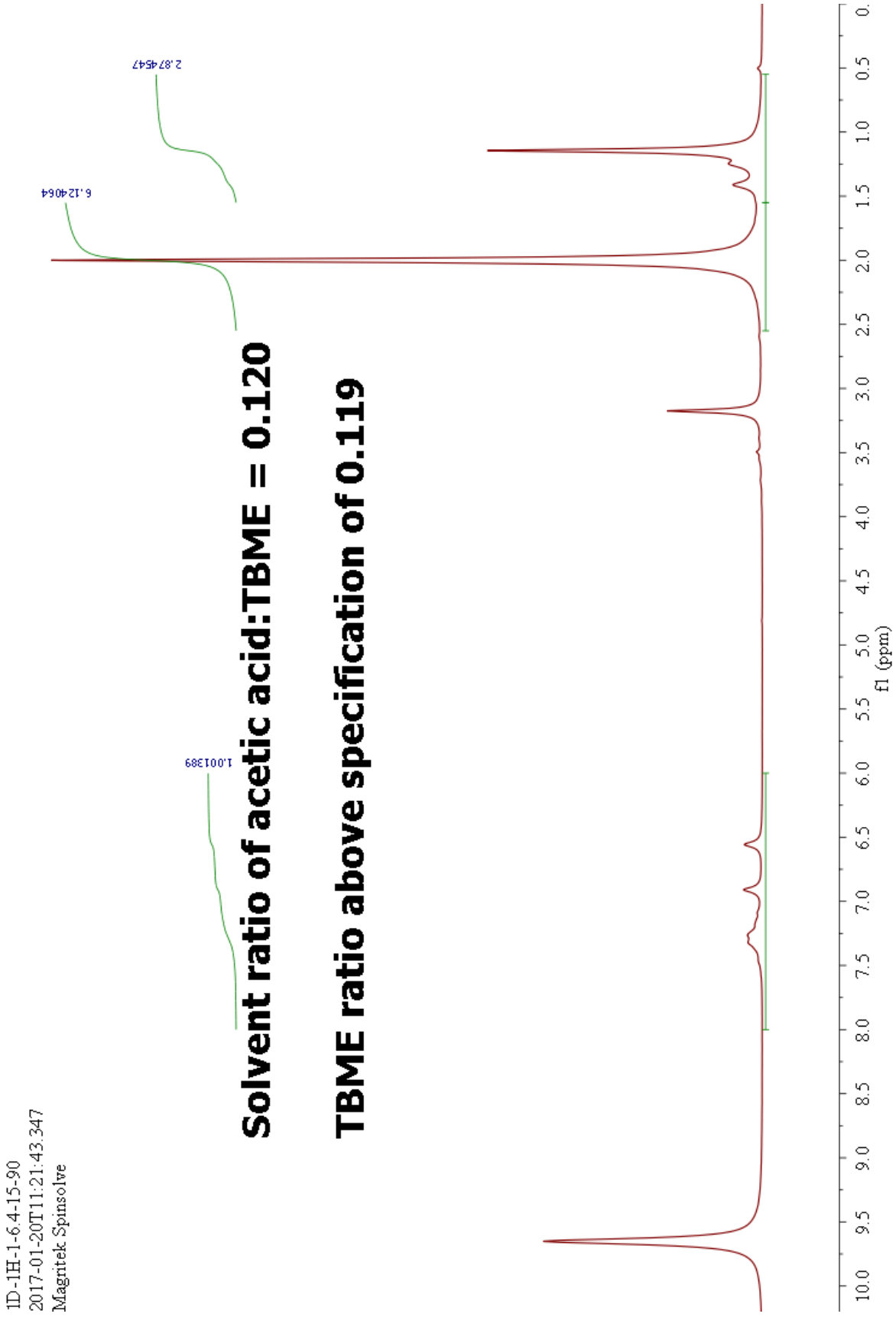
Lim, H., and S. Lee. 2006. Quantitative analysis of chloride by chlorine-35 NMR spectroscopy. *Bulletin of the Korean Chemical Society* 27 (7):972–73.

Malz, F., and H. Jancke. 2005. Validation of quantitative NMR. *Journal of Pharmaceutical and Biomedical Analysis* 38 (5):813–23. doi:10.1016/j.jpba.2005.01.043

Wider, G., and L. Dreier. 2006. Measuring protein concentrations by NMR spectroscopy. *Journal of the American Chemical Society* 128 (8):2571–76. doi:10.1021/ja055336t

Appendix 4

Example PDF output from automated method for determining acetic acid:TBME molar ratio


1D-1H-1-6.4-15-90
2017-01-20T11:21:43.347
Magritek Spinsolve
Solvent ratio of acetic acid:TBME = 0.120
TBME ratio above specification of 0.119
1.001389
6.124064
2.874547
f1 (ppm)

Appendix 5

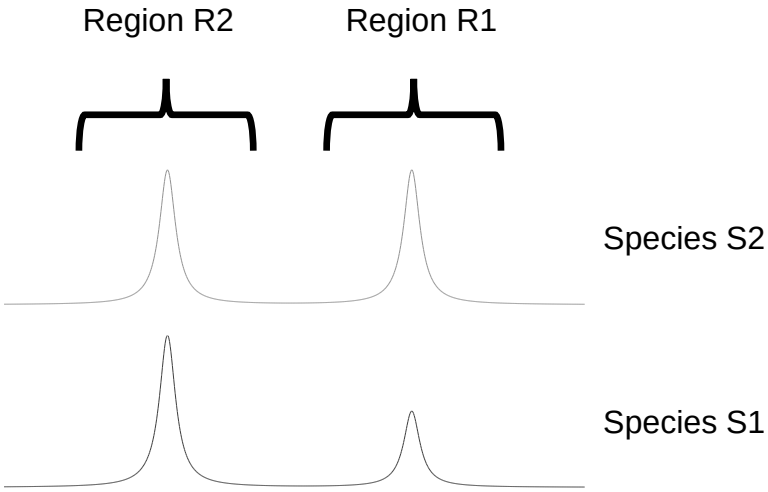
Calculating mole fractions from 43 MHz ^{1}H NMR spectra of binary mixtures

Calculating mole fractions from 43 MHz ^{1}H spectra of binary mixtures

First the general case of two overlapping species integrated in two regions is considered. Then, constants specific for the binary mixture are determined and the equations to calculate an integral for one peak from each species produced.

Calculating one integral for each species from the overlapped regions – general case

For the general case of two integrated regions with two overlapping species.



Either or both species could have multiple peaks in each region as long as the number of protons for each species is different for *at least 1 region.*

The area for one of the peaks for S1 and S2 can then be calculated for each region (referred to as A_{S1} and A_{S2}, respectively) as shown below.

First, the total area for each region, R1 and R2, are defined with respect to A_{S1} and A_{S2}. Each use of A_{S1} and A_{S2} is multiplied by a constant as determined specific for each binary mixture (see later) and are referred to here as a, b, c and d such that:

$$R1 = a \cdot A_{S1} + b \cdot A_{S2} \quad (1)$$

and

$$R2 = c \cdot A_{S1} + d \cdot A_{S2} \quad (2)$$

Rearrange (1) to make A_{S1} the subject:

$$R1 - b \cdot A_{S2} = a \cdot A_{S1}$$

$$A_{S1} = \frac{R1 - b \cdot A_{S2}}{a} \quad (3)$$

Substitute (3) into (2):

$$R2 = c\left(\frac{R1 - b \cdot A_{S2}}{a}\right) + d \cdot A_{S2}$$

Rearrange to make A_{S1} the subject:

$$R2 = \frac{c \cdot R1 - bc \cdot A_{S2}}{a} + d \cdot A_{S2}$$

$$R2 - d \cdot A_{S2} = \frac{c \cdot R1 - bc \cdot A_{S2}}{a}$$

$$a(R2 - d \cdot A_{S2}) = c \cdot R1 - bc \cdot A_{S2}$$

$$a \cdot R2 - ad \cdot A_{S2} = c \cdot R1 - bc \cdot A_{S2}$$

$$a \cdot R2 - c \cdot R1 = ad \cdot A_{S2} - bc \cdot A_{S2}$$

$$a \cdot R2 - c \cdot R1 = A_{S2}(ad - bc)$$

$$\boxed{A_{S2} = \frac{a \cdot R2 - c \cdot R1}{ad - bc}} \quad (4)$$

Now rearrange (1) to make A_{S2} the subject:

$$R1 - a \cdot A_{S1} = b \cdot A_{S2}$$

$$A_{S2} = \frac{R1 - a \cdot A_{S1}}{b} \quad (5)$$

Substitute (5) into (2):

$$R2 = c \cdot A_{S1} + d\left(\frac{R1 - a \cdot A_{S1}}{b}\right)$$

Rearrange to make A_{S2} the subject:

$$R2 = c \cdot A_{S1} + \frac{d \cdot R1 - ad \cdot A_{S1}}{b}$$

$$R2 - c \cdot A_{S1} = \frac{d \cdot R1 - ad \cdot A_{S1}}{b}$$

$$b(R2 - c \cdot A_{S1}) = d \cdot R1 - ad \cdot A_{S1}$$

$$b \cdot R2 - bc \cdot A_{S1} = d \cdot R1 - ad \cdot A_{S1}$$

$$b \cdot R2 - d \cdot R1 = bc \cdot A_{S1} - ad \cdot A_{S1}$$

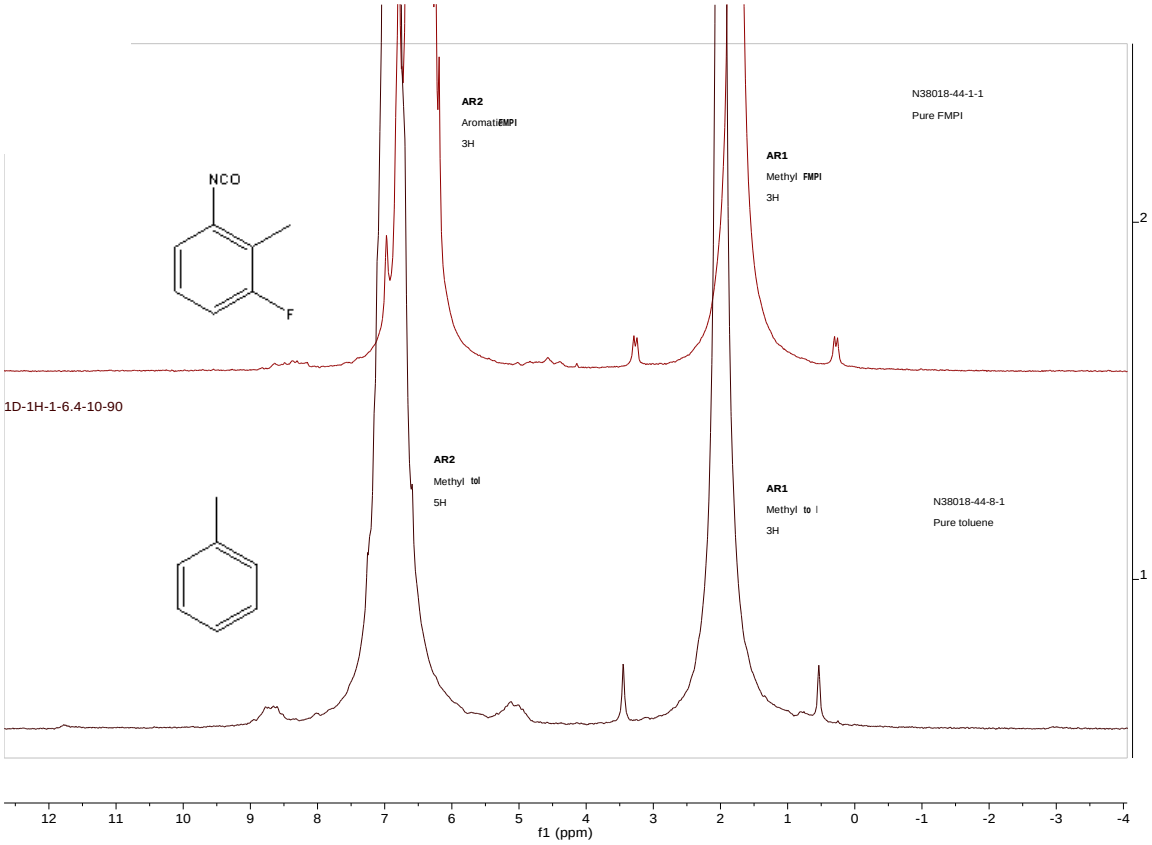
$$b \cdot R2 - d \cdot R1 = A_{S1}(bc - ad)$$

$$\boxed{A_{S1} = \frac{b \cdot R2 - d \cdot R1}{bc - ad}} \quad (6)$$

Hence the area for one peak from each species can now be calculated from (4) and (6).

Calculating the constants a, b, c and d for binary mixtures of FMPI and Toluene

Overlay of FMPI and Toluene at 43 MHz:



If region 1 and region 2 are defined as -1.0 to 4 ppm and 4.0 to 10.0 ppm, respectively, equations can be constructed that describe the overall integral of these regions in terms of the overlapping peaks:

$$R1 = AR1_{FMPI} + AR1_{tol} \quad (7)$$

$$R2 = AR2_{FMPI} + AR2_{tol} \quad (8)$$

There are terms for three peaks between these two equations. These need to be reduced down to one peak per species, in this case $AR1_{FMPI}$ and $AR2_{tole}$ for FMPI and toluene, respectively.

Since:

$$AR2_{FMPI} = AR1_{FMPI} \tag{9}$$

And:

$$AR2_{tol} = \frac{5}{3} AR1_{tol} \tag{10}$$

Substituting (9) and (10) into (8):

$$R2 = AR1_{FMPI} + \frac{5}{3} AR1_{tol} \tag{11}$$

From here, (7) and (11) can be written in the form:

$$R1 = a \cdot AR1_{FMPI} + b \cdot AR1_{tol}$$

$$R2 = c \cdot AR1_{FMPI} + d \cdot AR1_{tol}$$

Which, from the rearrangement of (1) and (4) to (2) and (6) demonstrated above, means they can be rearranged to:

$$AR1_{tol} = \frac{a \cdot R2 - c \cdot R1}{ad - bc}$$

$$AR1_{FMPI} = \frac{b \cdot R2 - d \cdot R1}{bc - ad}$$

Where:

$$a = 1$$

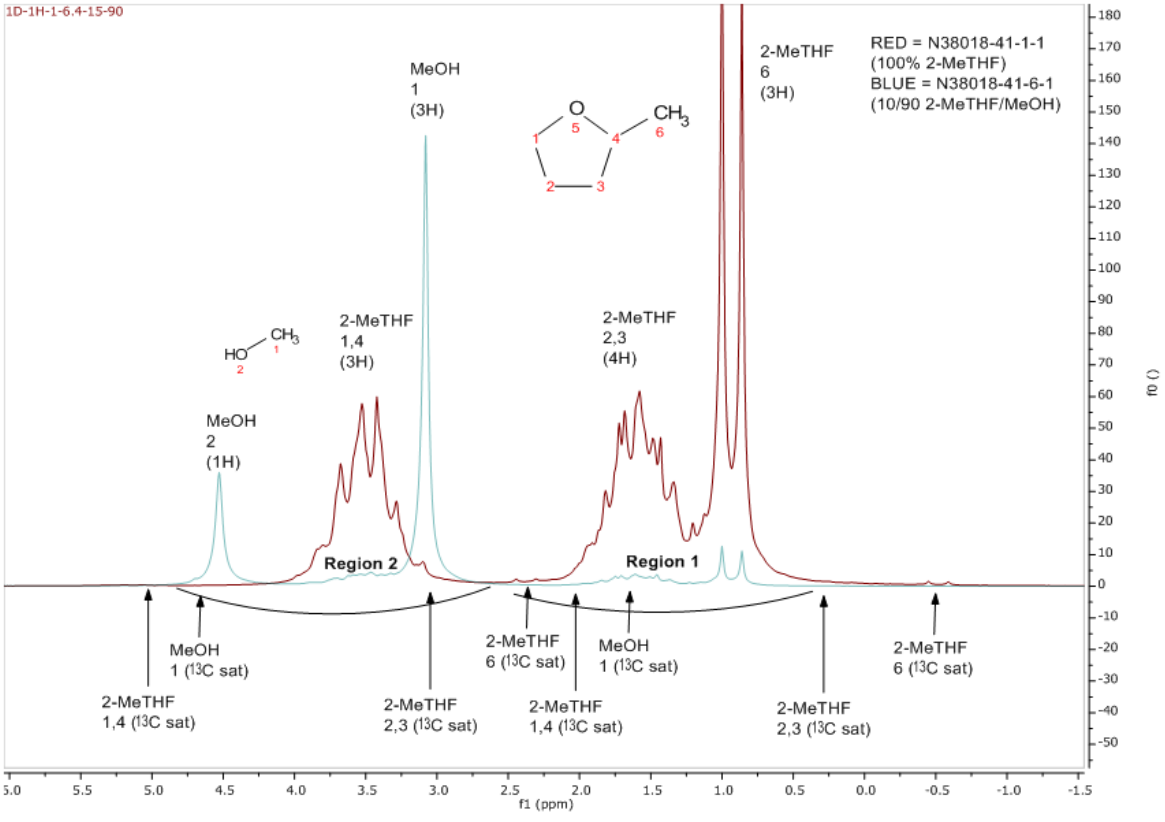
$$b = 1$$

$$c = 1$$

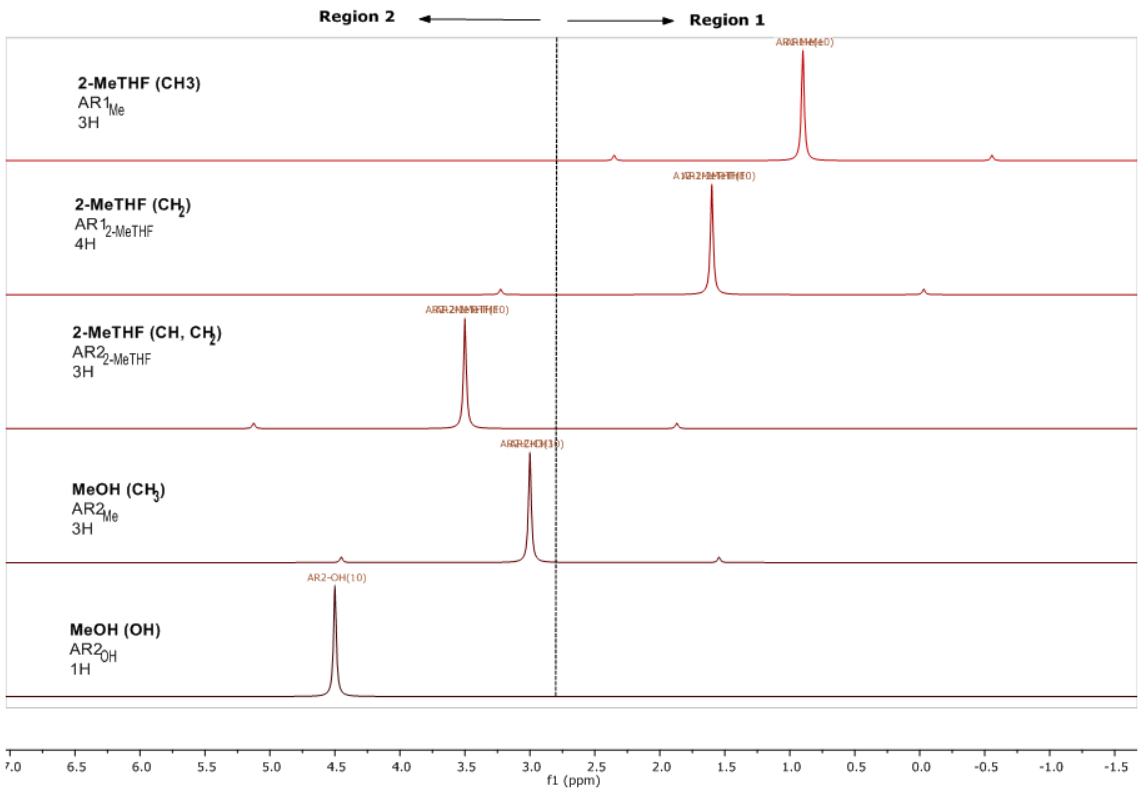
$$d = \frac{5}{3}$$

Calculating the constants a, b, c and d for binary mixtures of 2-MeTHF and MeOH

Overlay of 2-MeTHF and MeOH at 43 MHz:



This can be depicted as a simplified model showing only approximate peak shifts and ^{13}C satellite information (no coupling shown on main peaks and all intensities equal):



From here, equations can be constructed that describe the overall integral of regions 1 and 2 in terms of the overlapping peaks:

$$R1 = AR1_{Me} + AR1_{2-MeTHF} - Sf \cdot AR1_{2-MeTHF} + Sf \cdot AR2_{2-MeTHF} + Sf \cdot AR2_{Me} \quad (7)$$

$$R2 = Sf \cdot AR1_{2-MeTHF} + AR2_{2-MeTHF} - Sf \cdot AR2_{2-MeTHF} + AR2_{Me} - Sf \cdot AR2_{Me} + AR2_{OH} \quad (8)$$

There are terms for five peaks between these two equations. These need to be reduced down to one peak per species, in this case $AR1_{Me}$ and $AR2_{Me}$ for 2-MeTHF and MeOH, respectively.

For the MeOH peaks:

$$AR2_{OH} = \frac{1}{3} AR2_{Me} \tag{9}$$

For the 2-MeTHF peaks:

$$AR1_{2-MeTHF} = \frac{4}{3} AR1_{Me} \tag{10}$$

$$AR2_{2-MeTHF} = AR1_{Me} \tag{11}$$

Substituting (10) and (11) into (7):

$$R1 = AR1_{Me} + \frac{4}{3} AR1_{Me} - Sf \cdot \frac{4}{3} AR1_{Me} + Sf \cdot AR1_{Me} + Sf \cdot AR2_{Me}$$

Collecting together the terms:

$$R1 = (1 + \frac{4}{3} - Sf \cdot \frac{4}{3} + Sf) AR1_{Me} + (Sf) AR2_{Me} \tag{12}$$

Then substituting (9), (10), (11) into (8):

$$R2 = Sf \cdot \frac{4}{3} AR1_{Me} + AR1_{Me} - Sf \cdot AR1_{Me} + AR2_{Me} - Sf \cdot AR2_{Me} + \frac{1}{3} AR2_{Me}$$

Collecting together the terms:

$$R2 = (Sf \cdot \frac{4}{3} + 1 - Sf)AR1_{Me} + (1 - Sf + \frac{1}{3})AR2_{Me} \quad (13)$$

From here, (12) and (13) can be written in the form:

$$R1 = a \cdot AR1_{Me} + b \cdot AR2_{Me}$$

$$R2 = c \cdot AR1_{Me} + d \cdot AR2_{Me}$$

Which, from the rearrangement of (1) and (4) to (2) and (6) demonstrated above, means they can be rearranged to:

$$AR2_{Me} = \frac{a \cdot R2 - c \cdot R1}{ad - bc}$$

$$AR1_{Me} = \frac{b \cdot R2 - d \cdot R1}{bc - ad}$$

Where:

$$a = 1 + \frac{4}{3} - Sf \cdot \frac{4}{3} + Sf$$

$$b = Sf$$

$$c = Sf \cdot \frac{4}{3} + 1 - Sf$$

$$d = 1 - Sf + \frac{1}{3}$$