Methods to Aid Robust Structural Characterisation Using Nuclear Magnetic Resonance Spectroscopy

CAITLIN LEIGH EVANS

Methods to Aid Robust Structural Characterisation Using Nuclear Magnetic Resonance Spectroscopy

Caitlin L. Evans

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Department of Chemistry

Supervisor: Prof. Paul Hodgkinson

May 2023

Abstract

The structural assignment of organic molecules in spectroscopy underpins our understanding of their related physical and chemical properties. This is particularly prominent for pharmaceuticals, where understanding the behaviour of a material in both solution and solid can dictate its applicability for future developments. Nuclear Magnetic Resonance (NMR) spectroscopy can play a critical role given its ability to elucidate local environment information whilst providing insight to conformation, polymorphism, dynamics, and exchange.

In this thesis, organic materials have been probed using both solid and solution state NMR methods with the aim of providing comprehensive and robust structural characterisation. This has been demonstrated through the determination of the correct structural model of indapamide using an NMR crystallographic approach alongside the evaluation of possible multiple-quantum-based pulse sequences to probe carbon connectivity both indirectly and directly. A Bayesian probability-based workflow has also been presented as a method to evaluate a proposed experimental assignment of a material where its proof of concept is discussed.

In the final chapter, the solution state behaviour of the pharmaceutical, ritonavir, has been revisited where the assignment, exchange behaviour and conformation has been evaluated. This has led to the extraction of kinetic information directly from ^{1}H-^{1}H exchange spectroscopy (EXSY) spectra to comprehensively understand the interconversion of conformers in the solution state.

Acknowledgements

First and foremost, I must thank my supervisor Paul for his knowledge, supervision and unwavering patience over these past three and half years to mould me into the researcher I am today. I would also like to thank Juan for allowing me to explore my interests in the solution-state side of NMR and for the guidance in the best ways to understand the experimental data. A big thank you must also be extended to both David and Sam for their teachings in the practicalities of NMR and allowing me to ask as many silly questions without a hint of judgement. I am forever grateful to you all!

I also need to thank all the members (past and present) of the SSNMR group for cryogen fills, discussions and socials. A special thank you must be extended to Helen who has been here since Day 1 where we battled stuck rotors, COVID and Brian misbehaving.

My time at Durham would not have been as memorable without the people. Billy and Emma, 13 Wynyard Grove will forever remain with me as our battle against the pandemic to keep ourselves entertained and I am eternally grateful that you both helped me find the best in a difficult situation. Louise, I am so glad we bonded in our first week in Durham and have remained as an organised chaos duo that I feel will remain for many years to come and Eilish, the synthesis to my materials chemistry, our coffee breaks and loops around the science site have helped in more ways than one and I am so thankful. She may not have been in Durham but Samara, you are my other half in more ways than one and I thank you so very much for being the icon you are.

Joe, I cannot thank you enough for your love and support as we reach the finish line in our respective times at Durham and I am so excited to see where we go in the future.

And a final big thank you needs to be said to my family for their encouragement, support and love throughout as you kept me motivated and focussed as this chapter closes in my life. Regardless of the time of day or topic, your unwavering support has kept me going (alongside the daily dog photos!) and I am forever thankful for everything you have done for me.

Abbreviations

Abbreviation	Definition	Page First Defined
1D/2D/3D	One Dimensional / Two Dimensional / Three Dimensional	2
2HCA	2-hydroxycinnamic acid	54
3HCA	3-hydroxycinnamic acid	54
4HCA	4- hydroxycinnamic acid	28
6-APA	6-Amino Penicillanic Acid	49
ADP	Atomic Displacement Parameters	41
AF	Agreement Factor	53
Ala	Alanine	90
CASTEP	Cambridge Serial Total Energy Package	26
CP	Cross Polarisation	6
CSA	Chemical Shift Anisotropy	9
CSD	Cambridge Structural Database	32
DFT	Density Functional Theory	26
DMA	Direct Matrix Analysis	22
DQ/SQ	Double Quantum / Single Quantum	3
EXSY	Exchange Spectroscopy	i
FID	Free Induction Decay	7
FSLG	Frequency-Switched Lee Goldberg	13
FWHM	Full-Width Half Maximum	7
GIPAW	Gauge-Induced Projector Augmented Wave	27
HBA	Hydrogen Bond Acceptor	115
HBD	Hydrogen Bond Donor	115
HETCOR	Heteronuclear Correlation	17
IF	Iterative Fit	22
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment	3
IND	Indapamide Hemihydrate	3
INEPT	Insensitive Nuclei Enhanced by Polarisation Transfer	76
INH	Isoniazid	28
IntraHB	Intramolecular Hydrogen Bond	128

LAG	Liquid-Assisted Grinding	54
MAS	Magic Angle Spinning	1
MC	Monte Carlo Analysis	135
MD	Molecular Dynamics	128
MQ	Multiple Quantum	
NMR	Nuclear Magnetic Resonance Spectroscopy	i
NOE	Nuclear Overhauser Effect	21
NPX	Naproxen	75
NQS	Non-Quaternary Suppression	15
PAS	Principal Axis System	9
PBE	Perdew, Burke and Ernzerhof	27
ppm	Parts-per-million	10
POST-C7	Permutationally Offset Stabilised C7	19
PXRD	Powder X-Ray Diffraction	23
RF	Radiofrequency	5
RMSD	Root-Mean-Square Deviation	31
RRF	rotor-RF resonance	107
RVR	Ritonavir	3
SCXRD	Single Crystal X-Ray Diffraction	23
SNR	Signal-To-Noise Ratio	15
SPC5	Supercycled POST-C5	19
SPINAL	Small Phase Incremental Alternation	13
SSB	Spinning Sideband	13
SSNMR	Solid state Nuclear Magnetic Resonance Spectroscopy	1
TMS	Tetramethylsilane	10
TOSS	Total Sideband Suppression	15
TPPM	Two-Pulse-Phase-Modulation	13
Tyr	L-Tyrosine Methyl Ester Hydrochloride	50
USP	Ultrasoft Pseudopotential	27
VCA	Virtual Crystal Approximation	38
VT	Variable Temperature	35
XRD	X-Ray Diffraction	28
ZQ	Zero-Quantum	18

Table of Contents

Chapter 1: General Remarks

1.1 Brief Introduction to NMR

NMR is a powerful and versatile spectroscopic technique to probe the local atomic environment of materials. The initial discovery of NMR was first published in 1946 for both solution[1] and solids.[2] However, there has been a clear discrepancy in the method development and application of NMR with the use of solution state methods vastly outweighing its solid state counterpart.

This is, in part, due to the increased complexity associated with solid state acquisition and interpretation. Static lineshapes tend to be broad and featureless, resulting in overlapped signals where it can be difficult to extract the required chemical information. Such lineshapes occur as there is an intrinsic lack of motion in solids which introduces an orientational dependence. This is where all possible orientations are observed on acquisition. In contrast, molecular tumbling in solution allows the orientational dependence to be averaged, resulting in highly resolved, narrow lineshapes that provide chemical insight into a studied material. This has led to solution-state NMR methods being commonplace in many branches of chemistry to elucidate structure, concentration, conformation, dynamics, exchange and impurities.

However, there is an abundance of information that can be extracted from solid state NMR (SSNMR). Through developments in both methodology and spectrometer hardware, such as magic angle spinning (MAS), the broad lineshapes can be reduced to resolved, individual signals that can be assigned to individual environments.[3,4] When MAS is used in conjunction with specialised pulse sequences, significant information can be extracted from the acquired data. Alongside this, data acquisition is also not limited to a single dimension. In 1971, Jeener conceptualised a two-dimensional (2D) NMR experiment[5] which was achieved experimentally by Ernst and co-workers 5 years later.[6] This was a significant advancement in the structural assignment of molecules as clearer distinction between signals was aided by the additional, orthogonal dimension.[7] This led to an increase of both solution state and solid state 2D NMR pulse sequences to gain further insight into molecules through both homonuclear and heteronuclear correlation. This has been taken a step further, particularly in biological NMR where advances in 3D and 4D pulse sequences have been described.[8,9]

1.2 Structural Assignment in NMR

1.2.1 In Solids

SSNMR, particularly 1D ^{13}C SSNMR spectra, can be used as a fingerprinting tool to determine the form present in the sample studied alongside deriving crystallographic parameters.[10] However, to achieve a robust structural assignment, signals must be assigned to particular environments. This plays a key role in the understanding of the physicochemical properties of the studied material. Through the complementary nature of crystallography with NMR, the field of NMR Crystallography has become a reliable structural characterisation tool.[11] Coupled with Density Functional Theory (DFT) first principles calculations, the approach can aid in comprehensive assignment where SSNMR chemical shifts are assigned, helped by the calculated chemical shifts that are generated from the crystallographic structural model. The breadth of applicability of the approach is vast where recent reviews have demonstrated its use in the assignment of pharmaceuticals, zeolites and proteins to name a few.[10,12–16]

With focus on pharmaceuticals, a large number of drugs are formulated as tablets so the understanding of the behaviour and related physicochemical properties is critical.[15] The NMR crystallographic approach can also investigate the solid state landscape to identify possible polymorphs. This would avoid repetition of the infamous cases of ranitidine hydrochloride, ritonavir and paroxetine hydrochloride.[17–20]

1.2.2 In Solution

Structural assignment in solution state NMR can be considered more straightforward due to information available in the acquired data. As there is greater resolution relative to SSNMR, the *J*-couplings associated with signals can provide insight into the local atomic environment where the ^{1}H chemical shifts are primarily used to assign the materials through use of a wide variety of different 1D and 2D NMR experiments.[21] This is discussed further in Chapter 6 where experimental data can be used to probe conformation and exchange behaviour associated with the studied material. As with SSNMR, assignment in solution can be aided with comparison to calculated NMR parameters. There are popular strategies that combine computational predictions with structural elucidation[22–26] alongside methods that utilise statistical parameters to indicate a preferential structure.[27–29]

1.3 Thesis Outline

This thesis describes the application of NMR to aid in robust structural understanding of organic materials in both the solid and solution state. As an accompaniment to the results presented, a Data Archive has been created and, as seen for Chapter 3, will be provided as the work is published or can be requested from Prof. Hodgkinson.

Chapter 2 details the relevant NMR theory applied within this work alongside general background information relating to conventional crystallography and DFT-based calculations used as part of the NMR crystallographic approach in Chapters 3 to 5. Alongside this, the chapter also outlines the methods used in Chapter 6 to probe assignment, exchange, and conformation using solution-state NMR. Chapter 3 describes the determination of the correct structural model of the pharmaceutical, indapamide (IND). Use of DFT-based calculations and lineshape analysis were critical in the identification of the correct structural model, whilst the crystallographic investigation uncovered the problems associated with the other possible structural models.

A Bayesian probability-based workflow to quantitatively assess the compatibility of a proposed solid state assignment has been presented in Chapter 4. The developed workflow aims to quantitatively consider additional experimental data, alongside chemical shift agreement, in order to evaluate the compatibility of a proposed assignment. A variety of materials have been evaluated to investigate its ability to aid assignment. In Chapter 5, the applicability of dipolar-mediated, multiple-quantum pulse sequences to elucidate the carbon connectivity of molecular organic solids has been reported and its scope explored. Two pulse sequences have been investigated; the previously published dipolar-based ^{13}C-^{13}C Incredible Natural Abundance Double Quantum Transfer Experiment (INADEQUATE) and a newly developed dipolar-based ^{1}H-^{13}C Double Quantum/Single Quantum (DQ/SQ) pulse sequence.

The final results chapter, Chapter 6, focuses on solution state NMR methodology with the investigation into the behaviour of the pharmaceutical, ritonavir (RVR) in solution. RVR was chosen as its behaviour in solution is not as well-documented as its infamous conformational polymorphism in the solid state. The assignment and initial observations in relation to its solution conformation in different solvents has been discussed in conjunction to the development of a quantitative workflow to calculate kinetic parameters from ^{1}H-^{1}H exchange spectroscopy (EXSY).

Chapter 2: Theory and Methodology

The theory described in this chapter has been adapted from *'Spin Dynamics'* by M. H. Levitt,[30] *'Solid-State NMR: Basic Principles and Practice'* by D. C. Apperley, R. K. Harris and P. Hodgkinson,[31] '*NMR: The Toolkit*' by P. G. Hore, J. Jones and S. Wimperis,[32] '*Understanding NMR Spectroscopy*' by J. Keeler,[33] *The Nuclear Overhauser Effect in Structural and Conformational Analysis'* by M. P. Williamson and D. Neumann,[34] and *'High-Resolution NMR Techniques in Organic Chemistry'*, by T. D. W. Claridge.[21]

2.1 General NMR Theory

2.1.1 Vector Model and the Rotating Frame

Spin-active nuclei have a non-zero intrinsic spin angular momentum ($\mathbf{I}$) whose magnitude is defined by the spin quantum number, I. I can be either an integer or half-integer, thus $\mathbf{I}$ has $2I + 1$ allowed spin states. When $\mathbf{I}$ is non-zero, the spins possess a nuclear magnetic moment ($\boldsymbol{\mu}$). This is defined by the gyromagnetic ratio (γ), which is nucleus specific, and the value of I, as seen in equation 2.1. The allowed spin states to vary between values of $-I$ and $+I$ for the nucleus under study and are defined by the magnetic quantum number, m_I, as according to equation 2.2. The selection rule to allow transitions is $\Delta m_I = \pm 1$. The transition describes the movement of spins from one energy level to another where, for an allowed, directly observable transition, the change in m_I does not exceed ± 1.

$$\mu = \gamma \mathbf{I} \qquad \text{Eq. 2.1}$$

$$\mathrm{I}_z = m_I \hbar \qquad \text{Eq. 2.2}$$

$$E = -\mu_Z B_0 = -\gamma m_I \hbar B_0 \qquad \text{Eq. 2.3}$$

$$\Delta E = E_{up} - E_{down} = -\gamma \hbar B_0 \ = \hbar \omega_0 \qquad \text{Eq. 2.4}$$

$$v_{NMR} = -\frac{\Delta E}{h} = -\frac{\gamma B_0}{2\pi} \qquad \text{Eq. 2.5}$$

In the absence of an external magnetic field ($\boldsymbol{B_0}$), the spin states are degenerate. However, upon application of $\boldsymbol{B_0}$, this degeneracy is broken. This results in a net excess of $\boldsymbol{\mu}$ being aligned in the same direction as $\boldsymbol{B_0}$, where this break in degeneracy is controlled by the Zeeman interaction in equation 2.3. The frequency of the transition can be described as a difference between the energy levels but also as the Larmor frequency (ω_0), as shown in equation 2.4 and frequency units (υ_{NMR}) in equation 2.5. When placed in an external magnetic field, $\boldsymbol{\mu}$ precesses about $\boldsymbol{B_0}$ at ω_0 (or υ_{NMR} in frequency units), where it is assumed that $\boldsymbol{B_0}$ is along the z-axis.

Assuming that the sample is at thermal equilibrium and the sign of γ is positive, there is a slight preference of the spin states to align with $\boldsymbol{B_0}$ rather than be opposed. If the sign of γ is negative, the spins oppose $\boldsymbol{B_0}$. This preference is described by the Boltzmann distribution in equation 2.6 and means that the lowest energy level of all the $2I$ + 1 states of a given nucleus is slightly more populated than the others. The mathematical expression for the Boltzmann distribution defines N_{up} as the population of the higher energy level and N_{down} as the population of the lowest energy level with other parameters defined in the previous equations. As the thermal energy at room temperature (k_BT) is 4-5 orders of magnitude larger than the related difference in the energies of the upper and lower energy states, resulting in a preference for spins to sit in the lower energy level of approximately 1 in 100,000. Therefore, NMR can be described as an inherently insensitive technique where the net bulk magnetisation ($\boldsymbol{M}$) is slightly preferentially aligned with $\boldsymbol{B_0}$ for the z-component at equilibrium, as shown in figure 2.1. In contrast, the xy component at equilibrium is zero.

$$\frac{N_{\mathrm{up}}}{N_{down}} = e^{\frac{E_{\mathrm{up}}-E_{down}}{k_BT}} = e^{\frac{\hbar\omega_0}{k_BT}} \qquad \text{Eq. 2.6}$$

To acquire experimental data, a radio-frequency (RF) pulse must be applied to the sample whilst it is in the presence of $\boldsymbol{B_0}$. This perturbs the spin state populations established at thermal equilibrium, rotating $\boldsymbol{M}$ towards the xy plane. If the pulse frequency (ω_{RF}) matches that of ω_0, it is considered on-resonance and can be straightforwardly visualised using the rotating frame of reference in figure 2.1. In the normal laboratory frame, 2.1 a), $\boldsymbol{M}$ precesses around $\boldsymbol{B_0}$ at ω_0 but when the RF pulse is applied, the rotating frame at ω_{RF} is more appropriate. As shown in 2.1 b) and c), $\boldsymbol{M}$ now precesses at frequency Ω, equation 2.7,

around $\boldsymbol{B_1}$. $\boldsymbol{B_1}$ appears static along the x-axis, assuming an on-resonance pulse, whilst $\boldsymbol{B_0}$ is effectively removed from consideration.

$$\Omega = \omega_0 - \omega_{RF} \qquad \text{Eq. 2.7}$$

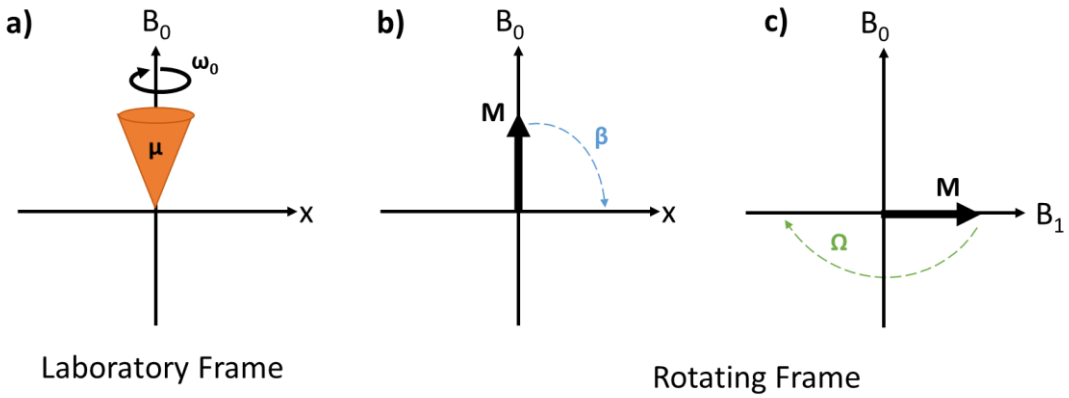


Figure 2.1: a) $\boldsymbol{\mu}$ of the spin states precess about $\boldsymbol{B_0}$ in the laboratory frame but appears stationary in the rotating frame b). When $\boldsymbol{B_1}$ is applied at a flip angle (β), of 90°, $\boldsymbol{M}$ is rotated away from $\boldsymbol{B_0}$ and the magnetisation precesses at $\boldsymbol{\Omega}$ in c).

The RF pulse can be defined by its flip angle (β), as shown in equation 2.8. This is the angle that $\boldsymbol{M}$ precesses relative to $\boldsymbol{B_1}$, where τ_p is the pulse duration. The most common pulses used are 90° (π/2) and 180° (π). A key pulse duration discussed in this work is τ_{CP}, which refers to the cross-polarisation (CP) contact time. This is a parameter which can be optimised in two-dimensional (2D) pulse sequences, as discussed further in Section 2.2.4.

$$\beta = \gamma B_1 \tau_p \qquad \text{Eq. 2.8}$$

Equation 2.8 can also be written in terms of nutation frequencies, as shown in equation 2.9 for the nutation frequency of ^{1}H spins. The relationship between the 90° pulse length (τ_{90}) at a given MAS rate and the nutation frequency plays a critical role in cross polarisation (CP) measurements, as discussed further in Section 2.2.2.

$$v_1^H = \frac{\gamma B_1}{2\pi} = \frac{1}{4\tau_{90}} \quad \text{Eq. 2.9}$$

2.1.2 Detection

As the system returns to thermal equilibrium after the RF pulse, the NMR signal is generated from the induction of current in a wire coil which is situated near the sample. The x and y components of the decaying magnetisation are detected by the receiver and in older spectrometers, the current is captured along two orthogonal axes in the rotating frame. This distinguishes between negative and positive frequencies and is called quadrature detection. Two free induction decays (FIDs) are the result of this, one is labelled 'real' whilst the other 'imaginary', where after phasing, the imaginary component can be discarded. In modern spectrometers, the signal tends to be mixed and the subsequent real and imaginary components are generated with software. This is Direct Quadrature Detection.

2.1.3 Relaxation

After the RF pulse, $\boldsymbol{M}$ decays back to equilibrium via relaxation mechanisms. These pathways all occur at different timescales. T_1, known as longitudinal (spin-lattice) relaxation, refers to the return to equilibrium of the z-component of $\boldsymbol{M}$ after perturbation. Assuming $\boldsymbol{M}$ is in the xy plane (i.e. a 90° pulse), the relaxation can be defined using equation 2.10.

$$M_z(t) = M_0\left(1 - \exp\left(\frac{-t}{T_1}\right)\right) \quad \text{Eq. 2.10}$$

The value of T_1 can range from milliseconds to seconds and has a direct impact on acquisition as all the spins need to return to equilibrium before repeating the experiment. The term which defines the time between successive experiments is the recycle delay. Therefore, the recycle delay for a material can then be determined from T_1 relaxation measurements, where for quantitative measurements, 5 x T_1 is recommended.

T_2 refers to transverse (spin-spin) relaxation and relates to the magnetisation decay of the xy-component of $\boldsymbol{M}$ after an RF pulse. This is an irreversible relaxation mechanism which has a direct impact on the resulting lineshape of the acquired data. Smaller values of T_2 result in broader lineshapes which can reduce resolution in the resulting peaks, as captured in equation 2.11, where $FWHM$ refers to full-width half maximum of a signal in Hz.

$$FWHM = \frac{1}{\pi T_2} \qquad \text{Eq. 2.11}$$

The last relaxation mechanism discussed is longitudinal relaxation in the rotating frame, $T_{1\rho}$. This is the return of the xy-component of $\boldsymbol{M}$ in the presence of spin-locking from the RF magnetic field, $\boldsymbol{B_1}$. A spin-lock describes the use of a second RF pulse to the initial RF pulse that rotates $\boldsymbol{M}$ around $\boldsymbol{B_1}$, in order to hold the magnetisation along $\boldsymbol{B_1}$ in one direction to prevent free evolution of transverse magnetisation in the rotating frame. $T_{1\rho}$ tends to be affected by molecular motion on the kHz timescale so can provide insight into dynamic processes on this timescale.

2.1.4 NMR Interactions

For diamagnetic materials, there are six NMR interactions of interest and these can be split into external and internal NMR interactions. The external NMR interactions are the Zeeman interaction ($\widehat{H}_{\mathrm{Z}}$) and the RF pulse ($\widehat{H}_{\mathrm{RF}}$) whilst the internal NMR interactions refer to the chemical shielding ($\widehat{H}_{\mathrm{CS}}$), indirect coupling ($\widehat{H}_{\mathrm{J}}$), dipolar coupling ($\widehat{H}_{\mathrm{D}}$) and quadrupolar coupling ($\widehat{H}_{\mathrm{Q}}$). As only spin-½ nuclei are studied in this thesis, the quadrupolar interaction will not be discussed further. The Zeeman interaction is by far the greatest in magnitude, but the other contributions still play a pivotal role in how a given nucleus behaves during the experiment.

To describe the interactions, the Hamiltonian ($\widehat{H}$) notation can be used, as in equation 2.12. I are the individual spin angular momentum operators, $\boldsymbol{R}$ are the tensors in a Cartesian representation and $\boldsymbol{N}$ refers to either the second spin angular momentum operator or the external magnetic field. This value is dependent on the interaction discussed. Tensors, such as $\boldsymbol{R}$, are often symmetric so in such circumstances, it is possible to form the principal axis system ($\boldsymbol{R_{PAS}}$) as shown in equation 2.13.

$$\widehat{H} = \begin{pmatrix} \hat{I}_x & \hat{I}_y & \hat{I}_z \end{pmatrix} \cdot \begin{pmatrix} R_{xx} & R_{xy} & R_{xz} \\ R_{yx} & R_{yy} & R_{yz} \\ R_{zx} & R_{zy} & R_{zz} \end{pmatrix} \cdot \begin{pmatrix} N_x \\ N_y \\ N_z \end{pmatrix} \qquad \text{Eq. 2.12}$$

$$\boldsymbol{R_{PAS}} = \begin{pmatrix} R_{XX} & 0 & 0 \\ 0 & R_{YY} & 0 \\ 0 & 0 & R_{ZZ} \end{pmatrix} \qquad \text{Eq. 2.13}$$

One simplification that is used in the following sections is the secular approximation. As discussed previously, the Zeeman interaction is generally the dominant interaction whilst the internal interactions, in contrast, are much smaller in magnitude. Therefore, the secular approximation can be applied where, for the internal interactions discussed in this thesis, are treated as first order perturbations of $\widehat{H}_Z$. This allows for simplification of $\widehat{H}$ but is only a valid approximation at large magnetic fields and also when the Zeeman interaction dominates.

Chemical Shift and Chemical Shift Anisotropy (CSA)

The chemical shift is a key NMR interaction studied for structural assignment. The local environment of each nucleus is influenced by the circulating electrons. The electrons induce a local magnetic field known as $\boldsymbol{B}_{\mathbf{induced}}$. $\boldsymbol{B}_{\mathbf{induced}}$ opposes $\boldsymbol{B}_{\mathbf{0}}$, therefore, the nucleus experiences a local field ($\boldsymbol{B}_{\mathbf{local}}$) that is defined by the difference between $\boldsymbol{B}_{\mathbf{0}}$ and $\boldsymbol{B}_{\mathbf{induced}}$. This can be described by the shielding tensor (σ), in equation 2.14, where $\boldsymbol{B}_{\mathbf{0}}$ is assumed to be along the z-axis. The full Hamiltonian is displayed in equation 2.15 but this can be simplified further through use of the secular approximation, resulting in equation 2.16.

$$\widehat{H}_{cs} = -\hat{\boldsymbol{I}} \cdot \boldsymbol{\sigma} \cdot \boldsymbol{\gamma} \boldsymbol{B}_{\mathbf{0}} \qquad \text{Eq. 2.14}$$

$$\widehat{H}_{cs} = -\begin{pmatrix} \hat{I}_x & \hat{I}_y & \hat{I}_z \end{pmatrix} \cdot \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ B_0 \end{pmatrix} \qquad \text{Eq. 2.15}$$

$$\widehat{H}_{cs} = -\hat{I}_z \sigma_{zz} \gamma B_0 \qquad \text{Eq. 2.16}$$

The distribution of electrons around each nucleus is not spherical. Therefore, it introduces anisotropy to the shielding tensor property. σ_{zz} then exhibits an orientational dependence, as seen in equation 2.17, with the $(3cos^2\theta - 1)$ term. The value of θ and ϕ refer to angles defining the orientation of the principal axis system (PAS) in the magnetic field with θ corresponding to the angle between the PAS relative to $\mathbf{B_0}$. The notations, R_{iso}, R_{aniso} and η refer to isotropic chemical shift, anisotropy and asymmetry, respectively. The parameters have been defined using equations 2.18 to 2.20 with the magnitude of the PAS values (σ_{XX}, σ_{YY} and σ_{ZZ}) listed in equation 2.21.

$$\sigma_{zz}(\theta) = R_{iso} + \frac{1}{2}R_{aniso}\{(3cos^2\theta - 1) + \eta(cos2\phi sin^2\theta)\} \quad \text{Eq. 2.17}$$

$$R_{iso} = \frac{1}{3}(\sigma_{XX} + \sigma_{YY} + \sigma_{ZZ}) \quad \text{Eq. 2.18}$$

$$R_{aniso} = \sigma_{ZZ} - R_{iso} \quad \text{Eq. 2.19}$$

$$\eta = \frac{(\sigma_{YY} - \sigma_{XX})}{R_{aniso}} \quad \text{Eq. 2.20}$$

$$|\sigma_{ZZ} - R_{iso}| \geq |\sigma_{XX} - R_{iso}| \geq |\sigma_{YY} - R_{iso}| \quad \text{Eq. 2.21}$$

R_{iso} is the average of the diagonal PAS values, whilst R_{aniso} is defined as the largest deviation in the chemical shift from the isotropic value. η is the difference between the other two principal values which are not involved in the R_{aniso} calculation and is always between 0 and 1. As each orientation is represented in the acquired experimental data in solids, the resulting lineshape can be broad and is called a powder pattern lineshape.

It is common to work with chemical shifts rather than chemical shieldings so they are converted using equation 2.22. $\nu_{\text{reference}}$ refers to the reference material used whilst ν_{sample} is the material under study. The standard reference for ^{13}C, ^{1}H and ^{29}Si. used is tetramethylsilane (TMS) whose chemical shift is at 0 parts-per-million (ppm) for the given nuclei.

$$\delta = \frac{\nu_{\text{sample}} - \nu_{\text{reference}}}{\nu_{\text{reference}}} \times 10^6 \quad \text{Eq. 2.22}$$

Dipolar coupling

Nuclei can also interact when they are close in space via through-space dipolar coupling. The energy of the interaction is defined by E_D, equation 2.23, where a distance dependence (r) is seen. As r is the distance between the two dipoles, dipolar coupling can provide direct insight into the internuclear distance of the studied pair. Using the secular approximation in equation 2.24, the calculation can be simplified further. D_{IS} is the dipolar coupling constant in Hz, defined by equation 2.25, and θ is the angle between the vector of the spins and $\boldsymbol{B_0}$.

$$E_D = \left\{\frac{\boldsymbol{\mu_I} \cdot \boldsymbol{\mu_S}}{r^3} - \frac{3(\boldsymbol{\mu_I} \cdot \mathrm{r})(\boldsymbol{\mu_S} \cdot \mathrm{r})}{r^5}\right\}\frac{\boldsymbol{\mu_0}}{4\pi} \qquad \text{Eq. 2.23}$$

$$E_{IS} = -\frac{1}{2}(3cos^2\theta - 1)D_{IS} \qquad \text{Eq. 2.24}$$

$$D_{IS} = \left(\frac{\gamma_I \gamma_S \hbar}{r_{IS}^{-3}}\right)\left(\frac{\mu_0}{4\pi}\right) \qquad \text{Eq. 2.25}$$

There is also a dependence on the spin-pair where the Hamiltonian varies based on whether the pair is homonuclear, as in equation 2.26, or heteronuclear, as in equation 2.27.

$$\widehat{H}_D(\text{homonuclear}) = E_{IS} \cdot (3I_{Iz}I_{Sz} - \hat{\boldsymbol{I}}_{\boldsymbol{I}}\hat{\boldsymbol{I}}_{\boldsymbol{S}}) \qquad \text{Eq. 2.26}$$

$$\widehat{H}_D(\text{heteronuclear}) = E_{IS} \cdot (2I_{Iz}I_{Sz}) \qquad \text{Eq. 2.27}$$

In an isolated spin-pair in solids, all possible values of θ are observed. This results in the powder lineshape pattern known as the Pake doublet.[35]

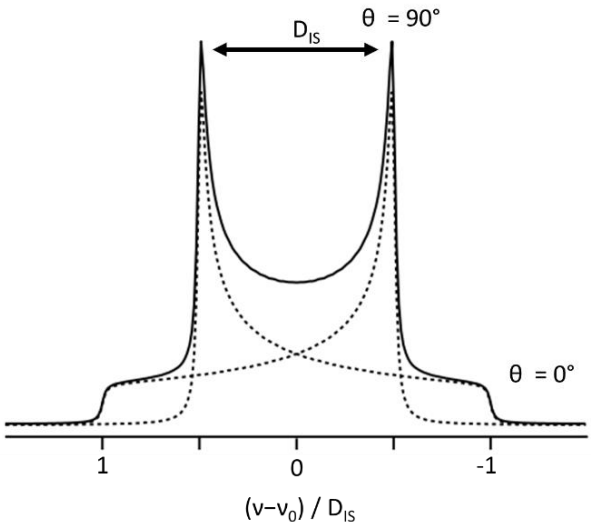


Figure 2.2: The heteronuclear Pake doublet for spin A in an AX spin system. Adapted from reference 31, the dashed lines highlight the two spectra for each value of m_I for spin X whilst the solid line shows the characteristic Pake doublet lineshape of their sum.

The outer limits in figure 2.2 are when θ = 0° whilst the horns correspond to θ = 90°. The distance between the two horns is the value of D_{IS} so the internuclear distance between I and S can be derived. However, in reality, acquired data that involves dipolar interactions is much more complex as broad lineshapes are observed. This is due to the studied spin also coupling with other spins within the system.

Indirect (*J*) coupling

The circulating electrons of neighbouring nuclei can interact with one another, resulting in the through-bond NMR interaction of indirect coupling. Known also as scalar or *J* coupling, its application to solid state NMR is mainly in multidimensional experiments to probe connectivity.[36] Due to its smaller magnitude relative to other interactions, indirect coupling has minimal impact to the signal appearance in the solid state so is not observed in the systems discussed in Chapters 3 to 5. In contrast, indirect coupling provides significant information in the solution-state through the lineshapes observed for particular signals, as shown in Chapter 6. The full Hamiltonian between spins I and S on the same material is given in equation 2.28. To note, there is no orientational dependence.

$$\widehat{H}_J = 2\pi(I_x \quad I_y \quad I_z) \cdot \begin{pmatrix} J_{xx}^{IS} & J_{xy}^{IS} & J_{xz}^{IS} \\ J_{yx}^{IS} & J_{yy}^{IS} & J_{yz}^{IS} \\ J_{zx}^{IS} & J_{zy}^{IS} & J_{zz}^{IS} \end{pmatrix} \cdot \begin{pmatrix} S_x \\ S_y \\ S_z \end{pmatrix} \qquad \text{Eq. 2.28}$$

2.2 Solid State NMR Methodology

2.2.1 Magic Angle Spinning (MAS) and Decoupling

Due to the lack of intrinsic motion in solids, resolution of signals in SSNMR data is poor. As discussed previously, the equations that define CSA, quadrupolar coupling and dipolar coupling contain the orientationally dependent term of $(3cos^2\theta - 1)$. As solids exhibit little to no motion, all values of θ are observed which results in a broad, featureless spectrum with limited chemical information. However, two techniques can be used to improve the information content of the acquired data.

MAS

The first technique is magic angle spinning (MAS). This requires the rotation of the sample at an angle of 54.74° relative to $\boldsymbol{B_0}$, the contributions from orientationally dependent interactions are minimised. This is because the $(3cos^2\theta - 1)$ term equates to 0.

The MAS rate does need to be equal or greater than the magnitude of the interaction being removed otherwise spinning sidebands (SSBs) can appear in the experimental data and complicate assignment.[31] As the difference in frequency between the sidebands is equal to the MAS rate used, a variation in MAS rate can be used to differentiate between the true signals and SSBs. The presence of SSBs can be minimised with faster spinning speeds but this is at the expense of rotor size. Smaller rotors have a faster maximum spin rate where at the time of writing, rates of >170 kHz have been reported.[37] This does, however, come at the expense of the amount of sample used, therefore use of smaller rotors can be somewhat counterproductive when studying dilute nuclei like ^{13}C.

Decoupling

The second technique employed to aid resolution of signals is decoupling. Decoupling pulse elements are particularly important to help in the suppression of the dipolar coupling contributions which cannot be fully minimised by MAS alone. The largest contributor tends to be couplings that involve ^{1}H spins and the application of pulses aims to disrupt the interactions to reduce their contribution to the acquired data. The combination of decoupling sequences alongside MAS is commonplace in structural assignment of molecular organics, as discussed within the literature.[10,12,38] Decoupling sequences can either be as homonuclear or heteronuclear.

Heteronuclear decoupling pulse sequences are arguably easier to implement with continuous wave and two-pulse phase-modulation (TPPM)[39] being popular choices. Method development of such pulse sequences remains an active area of research[40,41] and has resulted in sequences, such as the common pulse element of Small Phase Incremental Alternation (SPINAL-64).[42] This has been used in this work, where the key parameter to optimise is the ^{1}H decoupling pulse length.

In contrast, homonuclear decoupling is more complex and requires a selection of specific pulses in order to disrupt the homonuclear dipolar coupling of ^{1}H nuclei. This is also an area of active research, with the 2016 review by Mote, Agarwal and Madhu summarising five decades of method development.[43] In this work, frequency-switched Lee-Goldberg (FSLG)

has been used.[44–46] FSLG involves off-resonance pulses on the ^{1}H channel where rapid switching of both frequency and phase occurs throughout the pulse duration.[44] It is a robust sequence, where choice of offset frequency and interval duration can be maintained and adjusted if required.

2.2.2 Cross Polarisation and Modifications

Cross-polarisation (CP) has been used to increase the sensitivity of low abundant nuclei, namely ^{13}C, in this work. The majority of nuclei can be acquired directly via direct excitation experiments but for organic molecules, where ^{13}C is a critical nucleus to probe, CP is more useful and preferred.

For CP, ^{1}H spins are initially excited and then during the CP period, the magnetisation from the ^{1}H spins is transferred to the dilute ^{13}C spins in order to increase the resulting observed signal for the ^{13}C nuclei. To ensure the maximum efficiency during the magnetisation transfer stage, the Hartmann-Hahn matching condition, equation 2.29, must be met.[47] For maximum efficiency, the match condition should be optimised at the MAS rate desired as the match can be written in terms of nutation frequencies, as shown in equation 2.30 below where nv_r reflects the explicit MAS rate consideration. The relationship between pulse length and MAS rate has been discussed in equation 2.9 previously.

$$\gamma^H B_1^H = y^X B_1^X \qquad \text{Eq. 2.29}$$

$$v_1^H = v_1^X \pm nv_r \qquad \text{Eq. 2.30}$$

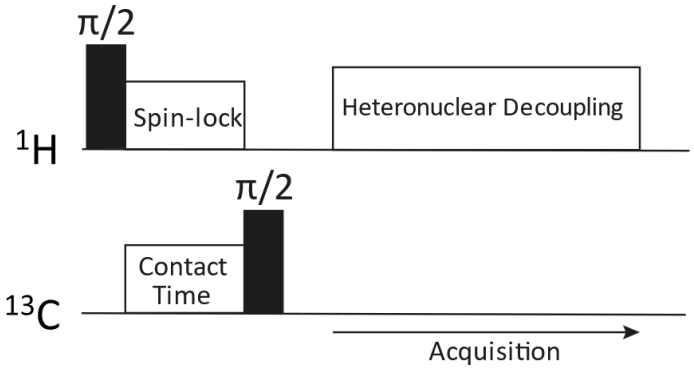


Figure 2.3: General pulse sequence for CP. Magnetisation is transferred from ^{1}H to ^{13}C where a spin-lock pulse is also applied on the ^{1}H channel during the contact time period.

The pulse sequence for a generalised CP experiment is shown in figure 2.3, where the ^{1}H spins are first excited with a 90° pulse whilst the second pulse, the spin-lock, is phase shifted relative to the initial pulse. The spin-lock holds the ^{1}H magnetisation along $\boldsymbol{B_1}$. Alongside the spin-lock period, there is also a pulse applied to the ^{13}C spins. The length of when the two pulses are applied is known as the contact time (τ_{CP}) and should be optimised prior to data acquisition. If the contact time is too short, inefficient magnetisation transfer will occur whilst signal decay via the heteronuclear dipolar coupling will be prevalent when the contact time is too long due to $T_{1\rho}$. However, this can be exploited using dipolar dephasing experiments, as discussed later in this section. A ramped spin-lock pulse tends to be used on the ^{1}H channel to sweep through the match conditions to allow for a robust Hartman-Hahn match.[48] During t_2, detection occurs on the ^{13}C channel whilst heteronuclear decoupling is applied to the ^{1}H channel.

A key benefit of CP experiments is that the acquisition parameters are based on the abundant spin rather than the dilute spin. This means that the recycle delay is defined by T_1 of the ^{1}H spins, which tends to be faster than the ^{13}C spins. This then results in an increase in the signal-to-noise ratio (SNR) of the spectrum which allows for quicker acquisition relative to direct excitation if identical acquisition parameters are used. There is also the signal enhancement benefit which is defined by the ratio between the gyromagnetic ratios of the studied nuclei. For γ_H/γ_C , the value is around 4-fold where, as shown in equation 2.4, the magnitude of the gyromagnetic ratio directly influences the frequency of the transition.

Unfortunately, the CP experiment is not completely quantitative relative to direct excitation. This is because the signal intensities may not reflect the full number of protons contributing to the CP signal at a given site. As the local dipolar coupling strength (equation 2.25) associated with the sites is not uniform, due to dependence on the protonation of a given environment, this may hinder quantitation. However, this variation in strength can be exploited through spectral editing experiments. Alongside this, standard 1D ^{13}C CP spectra can be complicated by the appearance of SSBs so total sideband suppression (TOSS[49]) has been used to remove the SSBs from the acquired data. A series of delays and π pulses are applied on the ^{13}C channel which average out the SSBs at low to moderate spinning rates. This can be taken one step further with non-quaternary suppression (NQS), where an additional, dipolar dephasing delay, is used.[50] Initially described by Opella and Frey, the heteronuclear decoupling within the dipolar dephasing time is switched off. The ^{13}C spins then precess in their local ^{1}H dipolar fields so the magnetisation of the nuclei decay, as mentioned previously. Protonated sites, such as CH_2 and CH groups, decay significantly whilst

quaternary carbons do not as there are no directly bonded protons. CH_3 groups would be expected to decay but as the group exhibits fast rotation, it disrupts the local dipolar couplings which in turn, reduces the decay observed so suppression is not achieved.[50]

2.2.3 Relaxation Measurements

Inversion-recovery relaxation measurements can be used to determine values of T_1 for the materials studied in this work. The pulse sequence is shown in figure 2.4.[31] The experiment involves a series of acquisitions over a range of delay times (τ) which are then fitted to equation 2.31, where S_0 reflects the signal intensity when τ = 0. It is important to ensure that the recycle delay used is much larger (4-5 times larger) than the suspected value of T_1 to allow the spins to return to equilibrium before successive experiments.

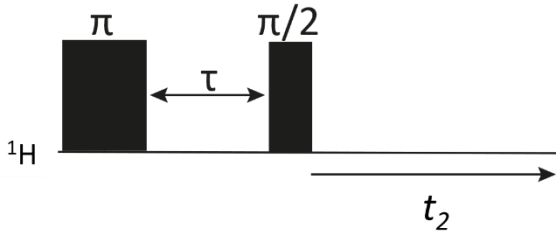


Figure 2.4: The inversion-recovery pulse sequence. Magnetisation is inverted prior to the delay period. The subsequent 90° pulse prepares the magnetisation for detection during t_2.

$$S(\tau) = S_0\left(1 - 2e^{-\frac{\tau}{T_1}}\right) \qquad \text{Eq. 2.31}$$

2.2.4 Homonuclear and Heteronuclear Correlation

Homonuclear

Homonuclear correlation experiments are either mediated by dipolar or *J*-couplings. In this work, the ^{13}C-^{13}C connectivity has been probed using the dipolar-based ^{13}C-^{13}C INADEQUATE and is discussed further in Section 5.2 of Chapter 5.

Heteronuclear

The standard experiment used to probe heteronuclear correlations is the ^{1}H-^{13}C HETCOR experiment which correlates the ^{1}H and ^{13}C spins via dipolar coupling.[44,51,52] As dipolar

coupling is distance dependent, it provides insight into the through-space distances of the environments.

The pulse sequence is shown in figure 2.5 where there is an initial 90° pulse on the ^{1}H spins which is then followed by the evolution period, t_1, on the ^{13}C channel with homonuclear decoupling. Homonuclear decoupling is essential during the evolution period to ensure decent resolution in the indirect ^{1}H dimension. Next is the CP step, where the contact time can be varied to limit the C-H distances detected during t_2. Shorter contact times allow only direct C-H correlations (where the proton is directly bonded to the associated carbon) to be observed whilst longer contact times allow for magnetisation transfer to occur over longer distances. During detection in t_2, heteronuclear decoupling is applied.

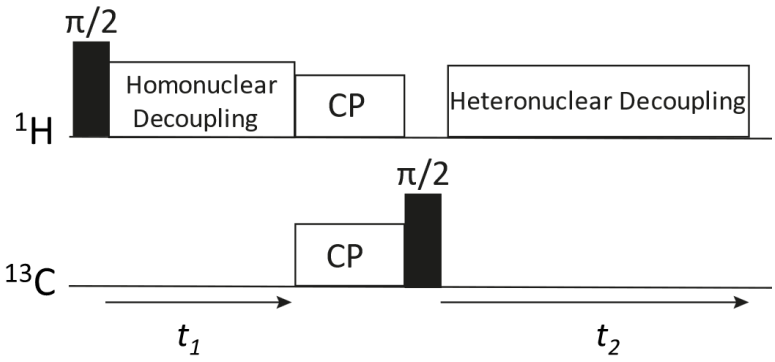


Figure 2.5: ^{1}H-^{13}C HETCOR pulse sequence. After the initial ^{1}H 90° pulse, the spins are left to evolve during t_1 prior to magnetisation transfer during CP and subsequent acquisition on the ^{13}C channel in t_2.

An alternative experiment is the ^{1}H-^{13}C double-quantum/single-quantum (DQ/SQ), where ^{1}H DQ coherences are correlated with ^{13}C SQ coherences to indirectly probe the carbon connectivity of a material.[53–56] The pulse sequence development and optimisation is described further in section 5.3 of Chapter 5.

Within these experiments, spin diffusion must be considered as its presence can complicate assignment. At longer contact times, apparent cross peaks can appear that are the result of strong homonuclear dipolar couplings between ^{1}H spins; this is spin diffusion. Spin diffusion can be exploited to provide insight into internuclear distances[57] however, its presence as relay peaks in ^{1}H-^{13}C HETCOR can be misleading. It is something to consider when evaluating weak, longer-range cross peaks to ensure the interpretation is correct.

2.2.5 Multiple-quantum (MQ) Coherences

In NMR, the only detectable coherences are SQ coherences as these obey the $\Delta m_I = \pm 1$ rule. SQ coherences have a coherence order (ρ) of 1 whilst other coherences, zero-quantum (ZQ) and double-quantum (DQ) have values of $\rho = 0$ and $\rho = 2$ respectively and are considered multiple quantum (MQ) coherences. These have been visualised in figure 2.6.

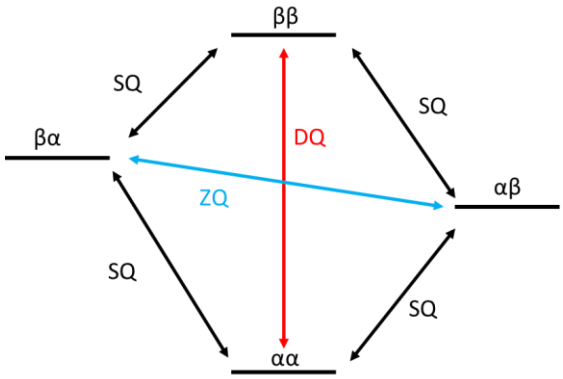


Figure 2.6: Coherence orders associated with a two-spin system with ZQ (blue), SQ (black) and DQ (red) coherences, respectively.

Using a generalised DQ/SQ pulse sequence in figure 2.7, the initial excitation recoupling element excites the DQ coherences which then evolve over the evolution period, t_1. The reconversion recoupling element then converts the DQ coherences into z-magnetisation which is necessary as MQ coherences cannot be detected directly. After reconversion, a 90° pulse generates the SQ coherences that are subsequently detected in t_2. To ensure only particular coherences are selected and observed, phase cycling is used.[58] Through a particular choice of phase and steps, particular coherences can be isolated and detected.

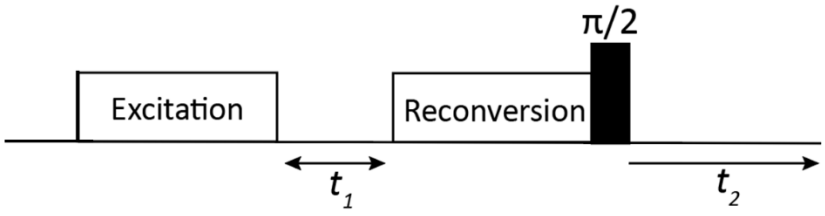


Figure 2.7: General representation of a DQ/SQ pulse sequence. The DQ coherences are initially excited and left evolve during t_1 prior to reconversion and subsequent acquisition during t_2.

There are a vast array of different recoupling pulse sequences but in this work, symmetry-based pulse sequences have been used, namely Supercycled POST-C5 (SPC5) and $R20_2^9$.[59,60] These are examples of γ-encoded pulse sequences which fall into two broad categories:

namely the CN_n^v and RN_n^v classes. The key variation between the two classes is how the phase alternates.[61]

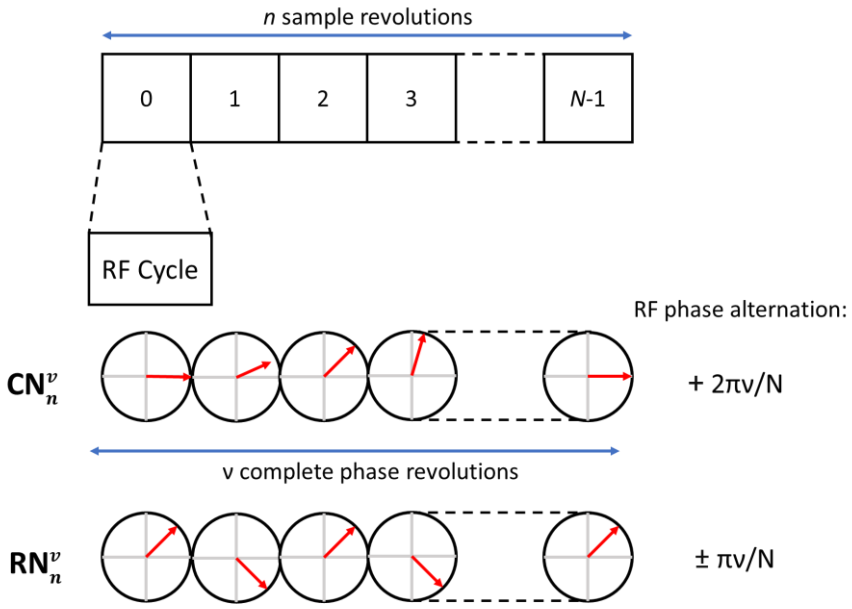


Figure 2.8: Adapted from reference 61, the construction of the CN_n^v and RN_n^v pulse sequences.[61]

CN_n^v

Examples of these sequences include Permutationally Offset Stabilised C7 (POST-C7)[62], a popular recoupling block within the literature[54,55,63] and SPC5[59], the recoupling block used in the proposed ^{1}H-^{13}C DQ/SQ pulse sequence in Chapter 5. The construction of the CN_n^v sequences is depicted in figure 2.8 where *n* rotational periods are subdivided into *N* equal intervals. Each interval contains an RF pulse cycle where the spins (and therefore the phase) are rotated incrementally from 0° to 360° by the angle 2πv/*N*. The choice of *n* and v can be altered to suit the needs of the recoupling sequence. For example, POST-C7 ($C7_2^1$) is known as a CSA compensated DQ recoupling sequence where there are seven equal intervals (*N*=7) across two rotor periods (*n*=2) with the phase of the spins incrementing by 2π/7.

RN_n^v

In contrast, the RN_n^v sequence consists of *N*/2 pairs of elements, R_ϕ R'_ϕ, where the overall phase shift is equal to πv/N radians. The construction of the sequence is similar to the CN_n^v sequences but instead of incrementally rotating the pulse in each element, each element is instead rotated 180° each time, as shown in figure 2.8. Initially, a sequence of RF pulses which rotate the spins through 180° about the x-axis is chosen, this is R. The signs of the RF pulses are then inverted to create the phase-inverted element, R'. An RF amplitude is selected to

ensure that *N* elements of R occupy the same time interval as *n* rotational periods of the sample where *N* must be an even number. $R20_2^9$ has been used in this work for the dipolar-based ^{13}C-^{13}C INADEQUATE and this corresponds to 10 pairs of elements over two rotor periods where the overall phase shift is equal to 9π/20 radians or 81°.[60]

2.2.6 Other General Experimental Details

Phase Cycling

As discussed in the previous section, the generation of coherence orders is only achieved by RF pulses. However, the RF pulses will not only generate the coherences of interest but also unwanted coherence pathways. In order to select the desired coherence pathway, phase cycling must be used. Through repetition of the pulse sequence and an incremental phase shift of certain pulses and the receiver, the overall sum of the FIDs will remove the unwanted coherence pathways whilst retaining the desired coherence pathways.

There are two key rules to phase cycling that must be followed, in order for the phase cycle to be successful. The first rule refers to the phase change of a pulse by an increment, defined as $\Delta\varphi$, which will result in a phase shift to any coherence (Δp) that is changing coherence order within the pulse sequence. This has been summarised with equation 2.32.

$$\Delta\varphi\ \Delta p \qquad \text{Eq. 2.32}$$

$$360°/N \qquad \text{Eq. 2.33}$$

The second rule, defined in equation 2.33, refers to the phase cycle steps where N is the number of repetitions of the pulse sequence itself. Alongside the desired pathways (Δp) that are selected, pathways of $\Delta p \pm N$ will also be selected but all other pathways blocked. In order for the phase cycle to be successful, it must reach completion, therefore, the total number of FIDs must be an integer multiple of N. Rule 2 does also select higher quantum order coherences as well as the desired SQ and/or DQ coherences, however, they tend to be ignored as the excitation efficiency does decrease with the increasing quantum order.

Spectrometer Calibration and Referencing.

Prior to pulse calibration, the magic angle was checked periodically using the Br nucleus (I = $^3/_2$) in KBr with the standard reference samples used for the acquired data being adamantane and glycine.

2.3 Solution State NMR Methodology

Chapter 6 discusses the investigation into the solution-state behaviour of ritonavir (RVR). A variety of 1D and 2D experiments were acquired but the key datasets used were the ^{1}H-^{1}H NOESY/ROESY and ^{1}H-^{1}H EXSY spectra.

2.3.1 Nuclear Overhauser Effect (NOE)

The NOE can be briefly defined as a change in intensity of one resonance, I, when the spin transitions of another resonance, S, are perturbed from its equilibrium population. An NOE can only arise if the two nuclei share a mutual dipolar coupling to undergo cross-relaxation. As dipolar couplings are intrinsically distance dependent, the presence of NOEs can provide insight into internuclear distances as the magnitude of the NOE falls rapidly with increasing distance.

Internuclear distances can be derived from ^{1}H-^{1}H NOESY (and ^{1}H-^{1}H ROESY) spectra. Pioneered by the Butts group[64,65], when it is assumed that the initial rate approximation is obeyed[66] and the length (r_{IS}) and cross peak intensity (η_{IS}) of one H-H distance is known, the length of a second H-H distance can be derived. This is defined in equation 2.34.

$$\frac{\eta_{IS}(\mathrm{ref})}{\eta_{IS}(\mathrm{x})} = \frac{r_{IS}^{-6}(\mathrm{ref})}{r_{IS}^{-6}(\mathrm{x})} \qquad \text{Eq. 2.34}$$

The reference H-H distance ideally should be a distance that is well-known within the literature or guided by crystallographic structures.[65] Caution is needed however for weaker NOE signals, where interproton distances are over 4 Å. These signals will be comparable in intensity to those associated with noise and artefacts so misidentification is much more likely.[65]

2.3.2 Calculation of Exchange Rates

2D ^{1}H-^{1}H EXSY spectra were used to understand the chemical exchange of RVR in Chapter 6. The ^{1}H-^{1}H EXSY was first published by Jeener, Meier, Bachmann and Ernst and it has become

a popular and reliable method in the understanding of chemical exchange.[67–70] Exchange rates can be extracted from the resulting exchange cross peaks as their presence indicate that slow exchange is occurring between conformers. Therefore, it can be quantified.[69,71] A brief overview of quantification is provided below but a comprehensive discussion can be found within the references provided.[67–74]

When exchange is classed as slow, the incoherent transfer of the magnetisation can be described using equation 2.35. $\boldsymbol{I}(\tau_m)$ represents the matrix of the diagonal (I_{ii}) and cross-peak (I_{ij}) intensities at a mixing time (τ_m) that is non-zero whilst $\boldsymbol{I}(0)$ is the intensity ratio when τ_m is zero. $\boldsymbol{L}$ is the dynamic matrix of exchange constants ($\boldsymbol{L_{ij}}$) and M_M and M_m are the intensities of the diagonal peaks when $\tau_m = 0$. As defined in equation 2.37, k_{Mm} is the rate of exchange between the major conformation ($\boldsymbol{M}$) and minor conformation ($\boldsymbol{m}$) whilst k_{mM} is the reverse. R_M and R_m are the relaxation rate constants and T_{1i} refers to the longitudinal time constants for M and m respectively.

$$\boldsymbol{I}(\tau_m) = \boldsymbol{I}(0)e^{\boldsymbol{L}\tau_m} \qquad \text{Eq. 2.35}$$

$$\boldsymbol{I}(\tau_m) = \begin{bmatrix} I_M & I_{Mm} \\ I_{mM} & I_m \end{bmatrix} \qquad \boldsymbol{I}(0) = \begin{bmatrix} M_M & 0 \\ 0 & M_m \end{bmatrix} \qquad \text{Eq. 2.36}$$

$$\boldsymbol{L} = \begin{bmatrix} -R_M - k_{Mm} & k_{mM} \\ k_{Mm} & -R_m - k_{mM} \end{bmatrix} \qquad \text{Eq. 2.37}$$

$$R_M = \frac{1}{T_{1M}} \quad R_m = \frac{1}{T_{1m}} \qquad \text{Eq. 2.38}$$

To determine the kinetic parameters from the experimental data, there are two approaches that have been used. These are the iterative fit (IF) and the direct matrix analysis (DMA).

For IF, the $\boldsymbol{I}(\tau_m)$ matrix, defined by the intensities and mixing time, can be iteratively adjusted to optimise the fit of the matrix.[75–78] The larger the number of mixing times considered, the more reliable the resulting fit will be.[77] The DMA approach is slightly different as it requires a maximum of one non-zero mixing time, thus the required spectrometer acquisition is less. Using two spectra, where $\tau_m = 0$ and $\tau_m = t$, the kinetic parameters can be derived with equations 2.39 and 2.40, which are derived from the re-arrangement of equation 2.35.

$$\boldsymbol{I}(t_m)\boldsymbol{I}(0)^{-1} = e^{\boldsymbol{L}t_m} = \boldsymbol{A} = \begin{bmatrix} \frac{I_M}{M_M} & \frac{I_{Mm}}{M_m} \\ \frac{I_{mM}}{M_M} & \frac{I_m}{M_m} \end{bmatrix} \quad \text{Eq. 2.39}$$

$$\boldsymbol{L} = \frac{1}{t_m}\boldsymbol{X}(ln\boldsymbol{\Lambda})\boldsymbol{X}^{-1} \quad \text{Eq. 2.40}$$

The equation can be defined by matrix $\boldsymbol{A}$ which is constructed of the intensity values derived from the experimental data. I_x refers to the intensities extracted from the $\tau_m = t$ t ^{1}H-^{1}H EXSY spectrum whilst the values of M_x are extracted from the $\tau_m = 0$ ^{1}H-^{1}H EXSY spectrum. **A** can be diagonalised to give a square matrix of its eigenvalues ($\boldsymbol{\Lambda}$) and the matrix of eigenvectors ($\boldsymbol{X}$). This allows the logarithm of $\boldsymbol{\Lambda}$ to be taken to give the solution for $\boldsymbol{L}$ which can then ultimately output the resulting exchange rate values.

If thermodynamic parameters are of interest, ^{1}H-^{1}H EXSY spectra can be taken at different temperatures and resulting rates can be evaluated using the Arrhenius equation defined by the pre-exponential factor (A), activation energy (E_a), universal gas constant (R = 8.314 J K mol^{-1}) and temperature (T).

$$k = Ae^{\frac{-E_a}{RT}} \quad \text{Eq. 2.41}$$

2.4 Crystallography

Single-crystal X-Ray diffraction (SCXRD) data was acquired by the Durham University Crystallography Service by Dr. Dmitry Yufit whilst the power X-ray diffraction (PXRD) data was acquired by Mr. Gary Oswald. Chapter 3 in the textbook *'Fundamentals of Crystallography'* by C. Giacovazzo, H. L. Monaco, G. Artioli, D. Viterbo, M. Milanesio, G. Gilli, P. Gilli, G. Zanotti, G. Ferraris and M. Catti has been used to provide an overview for this section.[79]

A crystal structure can be defined as a regular arrangement of atoms or molecules in a repeating 3D pattern. Due to the long-range periodicity associated with crystal structures, the structure and symmetry can be captured by the unit cell. The unit cell is the smallest repeating unit that reflects the studied material where the symbol *Z* refers to the number of

formula units in the unit cell whilst *Z*' is used to define the number of molecules within the asymmetric unit. The asymmetric unit captures the number of independent formula units in the unit cell which are not related by symmetry.[31] SSNMR can be used to determine *Z*'.[80]

The unit cell is defined by *Z*' and the space group. There are 230 unique space groups and they have associated symmetry operations that relate the multiple copies of the asymmetric unit to one another. Possible symmetry operations include rotations, reflections, translations and inversions. They can behave independently or can be combined to form more complex elements such as glide planes (reflection and translation) and inversion axes (rotation and inversion).

There are two types of scattering of electromagnetic radiation, elastic and inelastic. X-ray diffraction relies upon the elastic scattering of the X-ray photons by the electrons within the sample to provide insight into the electron density of each atom. Bragg's Law,[81] equation 2.42, must be satisfied for coherent scattering to occur. This has been visualised in figure 2.9. Assignment of the correct Miller indices (hkl) from the reflections results in the determination of unit cell parameters via the Bragg equation. The intensities of reflections identify the atomic positions of the atoms whilst systematic absences provide insight into the symmetry.

$$2d_{hkl}\sin\theta = n\lambda \qquad \text{Eq. 2.42}$$

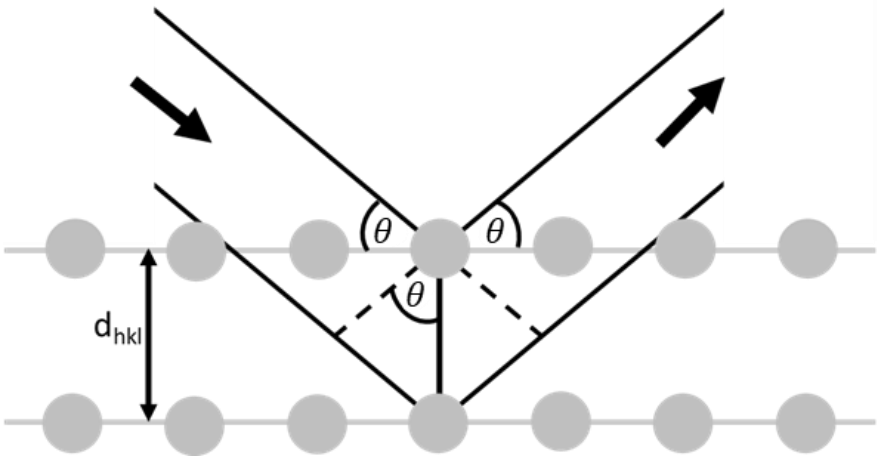


Figure 2.9: Depiction of Bragg's Law where d_{hkl} refers to the spacing between the two planes in the crystal structure, defined by the Miller indices (h,k,l) and ϑ is the incident angle used.

A conventional X-ray diffraction experiment measures intensities, where the magnitude is proportional to the structure factor $|\mathrm{F}_{hkl}|^2$**.** This is represented in equation 2.43. In equation 2.44, $\mathbf{F}_{\boldsymbol{hkl}}$ has an associated dependence on the phase of the scattered wave (ϕ) where, to calculate the electron density from the resulting reflections, equation 2.45 is needed. $\mathbf{F}_{\boldsymbol{hkl}}$ reflects both phase and magnitude of the diffracted wave from the crystal and this is characterised with the Miller indices. The value of f_j refers to the individual atomic scattering factors with x_j, y_j, z_j being the fractional coordinates for each atom.

$$\mathrm{I}_{hkl} \propto |\mathrm{F}_{hkl}|^2 \qquad \text{Eq. 2.43}$$

$$\mathbf{F}_{\boldsymbol{hkl}} = |\mathrm{F}_{hkl}|e^{i\phi} \qquad \text{Eq. 2.44}$$

$$\mathbf{F}_{\boldsymbol{hkl}} = \sum_j f_j e^{2\pi i(hx_j+ky_j+lz_j)} \qquad \text{Eq. 2.45}$$

Although the amplitude is obtained from the diffraction experiment, the phase cannot be measured directly; this is known as the phase problem, however, there are many methods in the literature that can be used to mitigate against the problem.[79] To obtain the structural model, the parameters used to describe the structure are varied and compared to the experimental structure factors until a good match is achieved. The agreement between the calculated structure factors and the experimental data can be evaluated to determine the agreement between the two. This quantified as the R-factor, equation 2.46, where F_{obs} and F_{calc} reflect amplitudes of the experimental and calculated structure factors, respectively. The R-factor can monitor the quality of the proposed structural model, where a lower R-factor usually indicates a better agreement. However, as discussed in Chapter 3, this can be misleading in determining the correct structural model.

$$R(F) = \frac{\sum |\mathrm{F}_{obs} - \mathrm{F}_{calc}|}{\sum |\mathrm{F}_{obs}|} \qquad \text{Eq. 2.46}$$

There are two common types of diffraction that are used for structure determination, SCXRD and PXRD. Both methods can obtain 3D structures however, SCXRD is preferred as the

structural determination is much easier using the resulting 3D diffraction data. However, there is a reliance on suitable crystals of a particular size and quality, which limits the types of materials studied.[82] When unavailable, PXRD can be used. Unfortunately, the data collected is compressed into a single dimension as a PXRD pattern so determination of a structural model solely from the PXRD pattern is challenging without complementary techniques, such as SSNMR, as reviewed by Harris in 2022.[80,82–84]

2.5 Computational Calculations

The calculations conducted in this work are DFT-based through the program of CASTEP (Cambridge Serial Total Energy Package).[85] A brief overview is provided here but several reviews have assessed the implementation of CASTEP in NMR crystallography of solids.[10,86,87] Specific methodology and parameter choice has been discussed within the relevant chapters.

2.5.1 General Theory

Density Functional Theory (DFT) has been instrumental in aiding robust structural assignment of molecular organics using NMR data in both the solid state and solution state. With focus on the solid state, DFT-based codes can compute NMR parameters for a variety of different nuclei directly from structural models. This has been incorporated into the NMR crystallographic approach and formed the basis theme for the reviews by Hodgkinson and Bonhomme et al.[10,86]

DFT is considered the best compromise between computational feasibility and calculation accuracy. The Kohn-Sham equations, equation 2.47, demonstrated that the electronic properties of an atom can be determined from the electron density (ρ).[88–90]

$$-\frac{1}{2}\nabla^2\boldsymbol{\Psi}^n(\boldsymbol{r}) + v_{eff}(\boldsymbol{r})\boldsymbol{\Psi}^n(r) = \varepsilon_n\boldsymbol{\Psi}^n(r) \qquad \text{Eq. 2.47}$$

$$v_{eff}(\boldsymbol{r}) = V_{nuc} + V_{hartree}[\rho] + V_{XC}[\rho] \qquad \text{Eq. 2.48}$$

Using the notation of Bonhomme et al.[86], ρ is determined from the sum of the occupied states at position, r. The ∇^2 operator refers to the kinetic energy of the non-interacting electrons whilst V_{nuc} is the electrostatic interaction with the nuclei and $V_{hartree}[\rho]$ is the mean field electrostatic interaction with the other particles in the system. These three terms

can be precisely calculated. However, the key issue for DFT calculations is that the last term, the exchange-correlation contribution ($V_{XC}[\rho]$) is not known so must be approximated. This term is critical as it calculates the exchange energy from the electron density so reliable approximations are needed which are achieved through choice of functional. A common functional used is the Perdew, Burke and Ernzerhof (PBE) functional[91] and this has been used in this work.

DFT allows the total energy of a system to be derived solely from the electron density whilst other methods, such as Hartree-Fock[92,93], use a many-body wave function. Ultimately, Hartree-Fock calculations are more accurate but they are extremely computationally expensive and become unfeasible for routine use.[10] However, DFT implementation must be adapted for the solid state as solid materials contain a significant number of electrons. Therefore, methods are needed that find a balance between computational expense and conservation of key information. CASTEP uses a planewave pseudopotential approach, where the wavefunctions are described using a planewave basis set.[85,94] The basis set itself is defined by the maximum kinetic energy of the waves it contains and its size can be controlled from a single parameter, E_{cut}.[10,86] This utilises Bloch's theorem where for single-particle wave functions, only the unit cell needs to be considered with periodic boundary conditions as the associated wavefunctions are quasi-periodic.[86,95] E_{cut} can be optimised through convergence, as described in the following section. An additional parameter that can also be optimised is the *k*-point sampling of the first Brillouin zone in reciprocal space. There is a balance however between choice of *k*-sampling and computational expense.

One problem with the planewave basis set is that to represent the core electrons, a significant number of planewave coefficients would be required. CASTEP mitigates for this by using pseudopotentials where the frozen core approximation is applied.[96] This assumes that the core electrons are tightly bound and therefore do not take part in chemical bonding and that they remain unperturbed in different chemical environments.[97] However, significant amounts of chemical and physical properties are influenced by the core electrons, including NMR parameters, so it is crucial that the core electron information is returned prior to the calculation of NMR properties. The method used in this thesis is the gauge-induced projector augmented wave (GIPAW).[94,98] This uses ultra-soft pseudopotentials (USP) which allow convergence at lower values of E_{cut}.[91,99] Overall, the CASTEP and GIPAW approaches allow for comprehensive calculation of NMR parameters for a variety of nuclei.

There are some caveats to the DFT-based method of calculating NMR parameters. These have been discussed in detail by Hodgkinson previously.[10] As the form of V_{XC} is approximated, there will be associated errors (around 2-3 % of the usual chemical shift range for a nucleus[86]) with the resulting NMR isotropic chemical shielding values. DFT calculations are also calculated at 0 K so thermal effects are omitted. This can be counteracted with DFT molecular dynamics simulations (DFT-MD) in CASTEP at different temperatures but this is a time consuming task so have not been used in this work.[10] It is also important to have a realistic structural model prior to calculations, otherwise, the resulting NMR parameters can be widely incorrect. The usual source of input structures is the Cambridge Structural Database[100] but it is important to ensure that the structural models are geometry optimised prior to the calculation of NMR parameters. In X-ray diffraction (XRD), the standard practise in structural refinement is to place hydrogen atoms at chemically plausible, geometric positions if they cannot be unambiguously refined. This is due to their small electron density. As NMR probes the local environment, chemical shifts are sensitive to atomic position so the structural model must be geometry optimised to ensure proper placement of the atoms.[80] The variation in hydrogen positions can be significant and can be quantified to help identify unrealistic structural models, as discussed in Chapter 3.[101,102]

2.5.2 Convergence

To determine the optimal calculation parameters, convergence testing was used where the size of the basis set, E_{cut}, and the *k*-point sampling could be varied, alongside the version of CASTEP used.[1] For a given parameter, convergence is observed when a plateau is present in plotted data points.

The co-crystal, isoniazid-4-hydroxycinnamic acid (INH-4HCA) has been used in the example below. The co-crystal features in Chapter 4 where reliable calculated chemical shifts are crucial to help aid assignment. As shown in figure 2.10, there is evaluation of both CASTEP version (17.2 against 22.1) and choice of E_{cut} for the NMR parameter calculation. The resulting plots depict the difference in calculated shieldings relative to the E_{cut} = 1200 eV calculation.

[1] *The version of CASTEP was evaluated due to an upgrade of the high-performance computer, which resulted in a variation in the standard version of CASTEP available.*

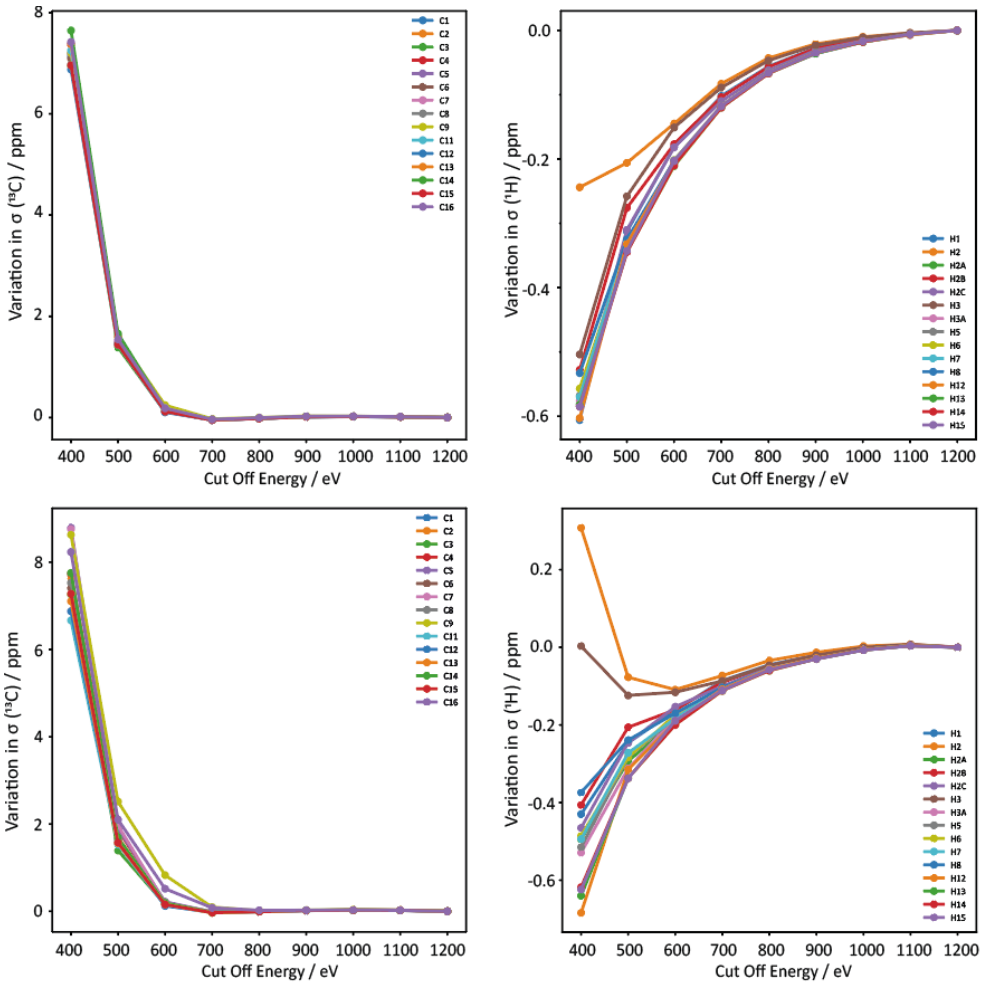


Figure 2.10: Example of convergence testing of the E_{cut} values for the NMR parameter calculation using CASTEP version 17.2 (top) and version 22.1 (bottom) for the ^{13}C chemical shifts (left) and ^{1}H chemical shifts (right) of INH-4HCA.

It was found that the version of CASTEP did not significantly vary the resulting isotropic chemical shieldings that were calculated. Alongside this, there was a plateau around 600 eV for ^{13}C and 900 eV for ^{1}H, where variation in shielding values were below 0.1 ppm. Due to the higher value of E_{cut} needed for ^{1}H, 900 eV has been used as the standard in the NMR parameter calculation unless specified otherwise.

Convergence was also attempted to determine the value of E_{cut} used in geometry optimisation where the standard value used is 600 eV. Taking both ^{13}C and ^{1}H environments for INH-4HCA, as shown in figure 2.11, there is a slight variation in calculated shielding values for each environment. However, this is due to the change of E_{cut} for the NMR parameter calculation rather than the value of E_{cut} for the geometry optimisation. Therefore, it was decided that the standard value of 600 eV would be used in geometry optimisation.

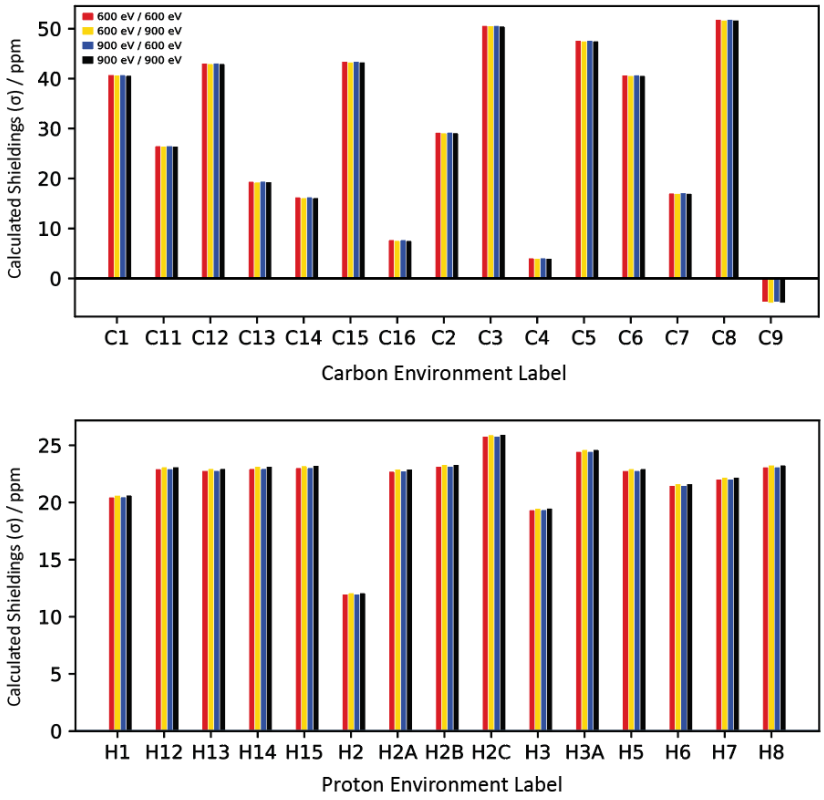


Figure 2.11: Comparison between the resulting chemical shieldings calculated for INH-4HCA on variation of the values of E_{cut} for geometry optimisation (600 eV in red and yellow, 900 eV in green and blue) and NMR parameter calculations (600 eV in red and green, 900 eV in yellow and blue).

The second parameter that can be optimised is the *k*-point spacing. There is an intrinsic link between the number of *k*-points sampled and the total computational time, as shown in table 2.1 for INH-4HCA. Therefore, unless specified otherwise, the spacing used was no greater than 0.1 $Å^{-1}$ with the *k*-point offset of (¼,¼,¼). This avoids the Γ point (0,0,0) as sampling here is particularly unrepresentative of the complete Brillouin zone.[10]

Table 2.1: Example of k-point spacing variation for INH-4HCA for the NMR chemical shift calculation using CASTEP version 17.2.

Material k-point spacing / $Å^{-1}$	**0.1**	**0.05**
Number of *k* points sampled	4	6
Total Calculation Time / min	93.1	144.1

2.5.3 Referencing

Once the NMR isotropic chemical shieldings are calculated, they need to be referenced for comparison to the experimental chemical shifts. This has been discussed in detail in Chapter 4, as a variety of methods can be used. One of these is on-the-fly linear scaling, where a plot of calculated isotropic shieldings (σ_{ref}) against the experimental chemical shifts ($\delta_{exp,i}$) results in a value for the gradient, m, and the intercept, σ_{ref}, as seen in equation 2.49. The values of m and σ_{ref} can then be fitted to convert the calculated shieldings into scaled and referenced GIPAW calculated chemical shifts, $\delta_{calc,i}$. Alternatively, m can be set to −1 and equation 2.50 can be used instead where $\langle\sigma_{\mathrm{calc}}\rangle$ and $\langle\sigma_{\mathrm{exp}}\rangle$ refer to the average of the calculated shieldings and experimental chemical shifts respectively. This method is advantageous as an assignment is not required. Both methods have been used in this thesis.

$$\delta_{exp,i} = \sigma_{ref} - m\sigma_{calc,i} \qquad \text{Eq. 2.49}$$

$$\sigma_{ref} = \langle\sigma_{calc}\rangle + \langle\delta_{iexp}\rangle \qquad \text{Eq. 2.50}$$

The agreement between the experimental and calculated chemical shifts can be evaluated using root-mean-square deviation (RMSD). As a simple metric, its magnitude can indicate whether there is an issue in a proposed assignment where in equation 2.51, N is the number of data points.[10] Salager et al. as well as Hartman et al. provided a benchmark to assess resulting RMSD values. [103,104] These were 1.9 ± 0.4 ppm for ^{13}C and 0.33 ± 0.16 ppm for ^{1}H. The two publications take slightly different approaches as on-the-fly linear scaling was used by Salager et al. on 15 organic compounds[103] whilst Hartman et al. derived ^{13}C and ^{1}H ranges from a fragment-based approach from predetermined regression parameters on a set of benchmark systems.[104]

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i}^{N}(\delta_{exp,i} - \delta_{calc,i})^2} \qquad \text{Eq. 2.51}$$

2.6 Software Used

Acquired experimental NMR data was processed in Topspin[105] and visualised using Topspin, MestreNova[106] and JASON.[107] The resulting spectral plots in this work were created either in ssNAKE[108], Python[109] or Adobe Illustrator.[110] Deconvolution of experimental peaks was completed using the open-source ssNAKE software and to view calculated NMR parameters from CASTEP, the resulting magres files were imported directly into MestreNova or JASON. To simulate NMR spectra, MagresView and MagresPython[111] were used to produce synthetic 1D and 2D spectra that were subsequently used to aid assignment. The supercomputer, Hamilton7 and Hamilton8, was used to run calculations using CASTEP versions 17.2, 19.1 and 22.1. Olex2[112] and Mercury[113], a programme within the CSD Software Package,[100] was used to visualise crystal structure files in .cif formats and simulate PXRD patterns. For RVR in Chapter 6, Avogadro[114] and Scigress[115] were used to draw and visualise the conformations of RVR.

This work directly benefitted from networking / training activities run by the Collaborative Computing Project for NMR Crystallography, funded by EPSRC grant EP/T026642/1.

2.7 Summary

This chapter has aimed to describe the general NMR theory and related methodology used within this thesis. This is in conjunction with the general background of both crystallography and computational methodologies used to aid structural characterisation across Chapters 3 to 5. As Chapter 6 probes the solution-state, the theory and methods that have been used have also been briefly discussed. Specific experimental acquisition, calculation parameters and synthetic information are included in subsequent chapters where relevant.

Chapter 3: Validation of Structural Models: The Case of Indapamide Hemihydrate

3.1 Introduction

The Cambridge Structural Database (CSD)[100] is an invaluable resource of crystal structures. In NMR crystallography, the CSD is a source of structural models for a system under investigation. Therefore, the choice of which model to use is important.[10] A survey by Widdifield and co-workers in 2020 identified that within the CSD, there are over 3000 'repeat structures' which reflect multiple structural determinations for what appears to be the same solid form.[102] As the CSD is a repository of crystal structures, for a given material, there can multiple structural models corresponding to different forms but also for the same form. One such example is furosemide where there are two structural deposits (**FURSEM01** and **FURSEM17**) for the same material. The only difference between them was the placement of a single hydrogen atom.[116] Without additional verification from alternate methods, such as solid state NMR, the choice of which structural model to use can be difficult.

During an investigation to identify the configurational preference of amide torsions in molecular organic solid pharmaceuticals, Indapamide hemihydrate (herein IND), was evaluated. IND is of interest due to its importance in the treatment of hypertension.[117,118] Recent literature has focused on improvements in bioavailability through amorphization studies[119–121] and co-crystallisations with other pharmaceuticals.[84,122]

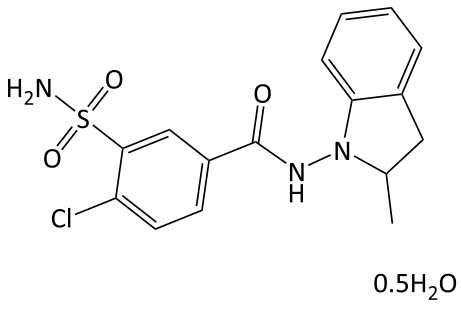
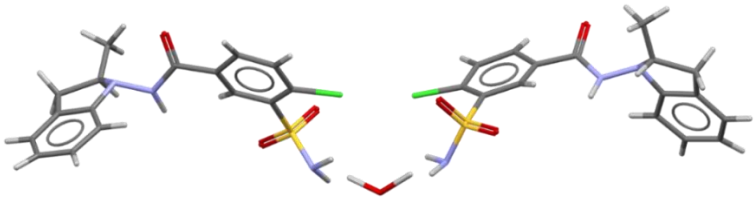


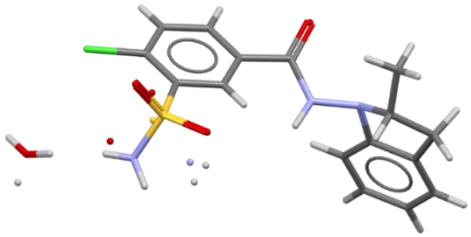
Figure 3.1: Formula unit of indapamide hemihydrate.

Additional forms have also been studied to characterise the solid state landscape of IND so it was not surprising that there were various depositions in the CSD.[123,124] The form of interest was the hemihydrate as this is the form used in tablet formulation.[125] It was expected that the related structural deposits would have incremental reference codes, as seen for furosemide.[116] Instead, there were three unique reference codes (**FOCCAD**, **VAGKUM** and **WOCPEM**) which all referred to the same solid form but, as shown in figure 3.2, the structural models varied considerably.[83,84,125] A key variation was the disorder in **VAGKUM** in figure 3.2(b).

a) FOCCAD



b) VAGKUM



c) WOCPEM

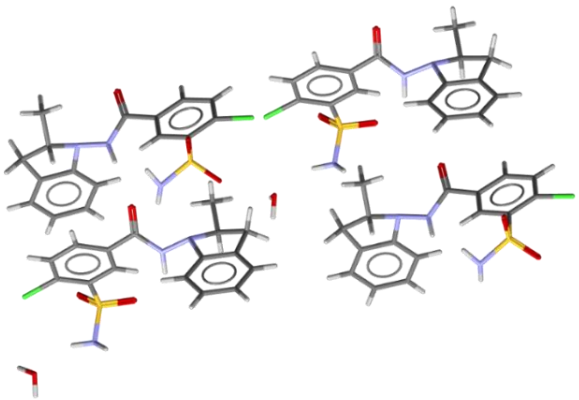


*Figure 3.2: Visualisation of structural deposits of a) **FOCCAD**, b) **VAGKUM** and c) **WOCPEM.** The crystal structures were obtained directly from the CSD and visualised in Mercury.*[100,113] *For **VAGKUM**, one of the sulfonamide positions is visualised whilst the other position is represented with spheres.*

Disorder is not uncommon and it can be difficult to characterise disordered structures due to the reduction in long-range order which can pose a challenge for crystallographic studies.[82,126,127] Complementary techniques, such as SSNMR, can provide insight and reinforcement to verify and validate disorder through its ability to probe short-range, local order.[10,82]

There are two types of disorder: static or dynamic.[10,126] Static disorder is when the atoms are located in different positions in adjacent unit cells whilst dynamic disorder is when the atom positions vary with time and the rate of exchange can be temperature dependent. Static disorder tends to cause inhomogeneous line broadening whereas dynamic disorder can result in sharp lines of a time-averaged spectrum when dynamics are faster than the NMR timescale.[126,128] Dynamic disorder can also be characterised further through variable-temperature (VT) experiments to probe the lineshape as the rate of motion varies.[126,128–130]

In relation to IND, the disorder on the sulfonamide group is chemically plausible as there is literature precedent that sulfonamide groups are prone to dynamic disorder.[131–133] The water molecule in **VAGKUM** is also disordered where one of the protons sits on a glide plane.[125] Both are examples of disorder by symmetry. This is where a section of the molecule is disordered across two (or more) atomic positions. It does not reflect the space group symmetry individually but through superposition of all atomic positions, the space group symmetry is then respected.[134–136]

3.1.1 Aims of This Work

The aim of this work was to determine firstly which is the correct structural model and then uncover why the structural models were so different. Using an NMR crystallographic approach, particularly ^{13}C solid state NMR and first-principles calculations, the correct model could be determined. Alongside additional crystallisation studies, the correct structural model was verified and validated whilst the issues with the other structural models were identified. The work in this chapter has been adapted from the publication, "Resolving alternative structure determinations of indapamide using ^{13}C solid state NMR".[80]

3.2 Materials and Methods

IND was purchased from Merck Life Sciences Ltd (CAS Number: 26807-65-8) and used without further purification. The following methodology has been adapted from reference 80.

3.2.1 Powder X-Ray Diffraction

The PXRD data was acquired by Mr. Gary Oswald. Data acquisition was performed on a Bruker D8 Advance diffractometer where the PXRD pattern was collected at room temperature for 60 minutes. The data was visualised and fitted using Topas Academic software.[137] Simulated PXRD patterns were calculated from the structural deposits in Mercury (Version 5.3.0).[113]

3.2.2 Single-Crystal X-Ray Diffraction

The SCXRD data was acquired and solved by Dr. Dmitry Yufit of the Durham Crystallography Service. Previous literature used the recrystallisation method of slow evaporation to obtain suitable crystals for SCXRD.[84,125] **VAGKUM** was produced from a hexane-acetonitrile solution[125] whilst **WOCPEM** was serendipitously resolved after a failed co-crystallisation of IND with gliclazide in MeOH.[84] Both conditions were attempted, however, only the **WOCPEM** conditions were successful.

10 mL of MeOH was used to dissolve IND and the resulting solution was left to slowly evaporate at room temperature for approximately two weeks to produce suitable crystals. The X-ray single crystal data was collected using MoKα radiation (λ = 0.71073 Å) on a Bruker D8 Venture 3-circle diffractometer equipped with a Cryostream (Oxford Cryosystems) open-flow nitrogen cryostat at the temperature 120.0(2) K. The structure was solved by direct methods and refined by full-matrix least squares on F^2 for all data using Olex2[112] and SHELXTL[138] software. Crystallographic data for the structure have been deposited with the Cambridge Crystallographic Data Centre as supplementary publication CCDC-2115849 and reference code **WOCPEM01**.

3.2.3 Solid State NMR

The ^{13}C CP/TOSS[49] spectrum was acquired at a MAS frequency of 9 kHz on a Oxford 11.7 T (^{13}C of 125.65 MHz) superconducting magnet, and a Bruker AVANCE III HD console, operating at frequencies of 499.69 (^{1}H) and 125.65 (^{13}C) MHz. The data was collected using a 4 mm HX magic-angle spinning probe at ambient temperature with a spectral width of 50 kHz taken across 512 co-added transients using a recycle delay of 7 s and a contact time of 1 ms. SPINAL-64 heteronuclear decoupling[42] was used during acquisition with a ^{1}H nutation frequency of approximately 73.5 kHz and a ^{1}H 90° pulse length of 3.4 µs. The ^{13}C chemical shifts were referenced using the high-frequency signal of adamantane ($\delta_{iso}(^{13}C)$ = 38.5 ppm). The resulting spectrum agreed well with previously published data.[124]

3.2.4 Computational Methods

First principles calculations were carried out using the GIPAW method implemented into CASTEP version 17.2/19.1[85] using the Perdew, Burke and Ernzerhof (PBE) functional and on-the-fly generated ultrasoft pseudopotentials.[91,99] Integrals were taken over the Brillouin zone using a Monkhorst-Pack grid.[139] Unless specified otherwise, the maximum k-point spacing was 0.1 $Å^{-1}$ and a k-point offset of (¼, ¼, ¼) was used. The value *of* E_{cut} was 600 eV. All atomic positions were geometry optimised with a fixed centre of mass and the unit cell parameters fixed at their diffraction-determined values. Input files were generated using CIF2cell,[140] with the original CIF labelling incorporated into the output .magres files.[111] NMR parameters were calculated[94,98] using the same parameters, and the resulting ^{13}C shielding values were converted to chemical shifts using a subset of ^{13}C resonances that could be unambiguously assigned in the experimental data.

Due to the large size of the unit cell of **WOCPEM** and **WOCPEM01** (166 atoms), the calculations were modified slightly to ensure that Hamilton could calculate NMR parameters. From the unit cell, the k-point grid was 2 × 1 × 1 therefore the k-point offset, (¼,¼,¼) could be removed. This meant the calculations could be performed on a single k-point at (½, 0, 0). Crucially, this was not at the gamma (**Γ**) point of (0,0,0).[10] Alongside this, "node dilution" was used to disperse the memory required for the calculations where the number of tasks allocated to each node was dropped from 24 processes to 18. This allowed the 54 GB of memory per compute node to be used effectively.

For **VAGKUM**, to model each possible orientation of the sulfonamide group, the **VAGKUM** CIF file was separated into two separate CIFs, as shown in figure 3.3. This approach has been adopted in previous publications.[131,141]

Unfortunately, this could not be done for the symmetry of the water molecule so the virtual crystal approximation (VCA) was used to model the 50% occupancy of the two proton positions of the water molecule.[142] Due to the approximations, the calculations themselves did lack physical significance and were treated with caution.

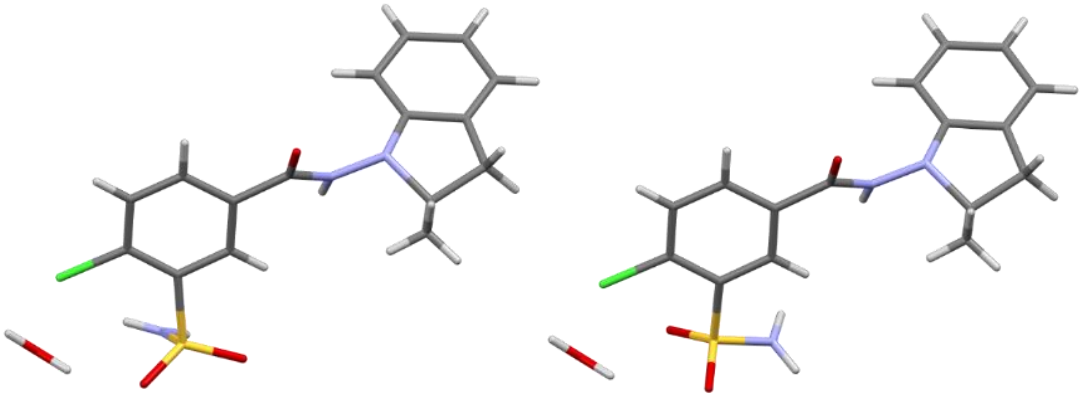


Figure 3.3: ***VAGKUM_A*** *(left) and* ***VAGKUM_B*** *(right) visualised to highlight the sulfonamide group disorder associated with* ***VAGKUM****.*

3.2.5 Lineshape Analysis

Lineshape analysis was conducted using ssNAKE (Version 1.3) where the standard parameters of the software were used.[108] These were Powell[143] minimisation across 500 evaluations coupled with a Lorentzian / Gaussian fitting method. For the fitting, the common lineshape and integral was kept constant, as detailed in table 3.1, to simplify the fitting procedure. It was assumed that the each of the four environments would exhibit similar lineshapes.

Table 3.1: Lineshape analysis fitting parameters for ***C16*** *and* ***C9*** *in ssNAKE.*[108]

Environment	**Me**				**CO**			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Chemical Shift / ppm	16.0	16.8	16.8	17.6	167.7	168.6	168.9	169.5
Integral	1.78e+10				8.88e+09			
Lorentz / Hz	83.3				114.0			
Gauss / Hz	25.3				19.8			

3.3 Initial Calculations and Variation in Structural Models

3.3.1 History of the Models

FOCCAD was the initial structural deposit by Smrkolj and Meden in 2006.[83] They were unable to grow suitable crystals for SCXRD so the structure was solved from PXRD data. A decade later, Bojarska and co-workers were successful in obtaining the SCXRD structure, **VAGKUM**. **VAGKUM** was described as 'pretty similar' to **FOCCAD** with the key variation being the

disorder on the sulfonamide group. Rather accidentally, Aljohani and co-workers then isolated the SCXRD structure of **WOCPEM**.[84]

*Table 3.2: Crystallographic parameters for **FOCCAD, VAGKUM, WOCPEM** and **WOCPEM01**.*

	FOCCAD	**VAGKUM**	**WOCPEM**	**WOCPEM01**
Formula Unit	$C_{16}H_{16}ClN_3O_3S{\cdot}0.35H_2O$	$C_{16}H_{16}ClN_3O_3S{\cdot}½H_2O$		
M_r / g mol^{-1}	372	374.83		
Temperature / K	298	100	150	120
Wavelength / Å	1.54	1.54	0.71	
Crystal System	Monoclinic	Monoclinic	Monoclinic	Monoclinic
Space Group	$P2/a$[2]	$I2/a$	$P2_1/c$	$P2_1/c$
a / Å	23.811(3)	15.059(9)	30.060(11)	30.140(10)
b / Å	9.6940(9)	9.6218(6)	9.6685(3)	9.6025(4)
c / Å	15.114(2)	23.508(14)	23.573(10)	23.461(8)
β / °	91.66(3)	92.60(16)	92.33(4)	92.59(10)
V / Å^3	3487.2(3)	3402.6(4)	6845.4(5)	6783.2(4)
Z, Z'	4,1	8,1	16,4	16,4
Data	-	3036	12492	18001
Restraints	-	3	41	61
Parameters	9	261	948	979
R_1 [$1 \geq 2\sigma(I)$][3]	0.0560	0.0369	0.0868	0.0842

The resulting PXRD patterns from both publications matched that of Smrkolj and Meden despite clear differences in the crystallographic parameters as shown in table 3.2. The key variation was the values of Z and Z' where **WOCPEM** has approximately double the unit cell volume relative to **VAGKUM** and **FOCCAD**. Interestingly, the simulated PXRD patterns for **VAGKUM** and **WOCPEM**, when overlaid, were virtually identical, as shown in figure 3.4.

[2] *To note, the space group of **FOCCAD** appears incorrect in the CSD .cif deposit. The original reference states a space group of I2/a.*[83]

[3] *Values can be found under _refine_ls_R_factor_gt within the related .cif files within the Data Archive.*

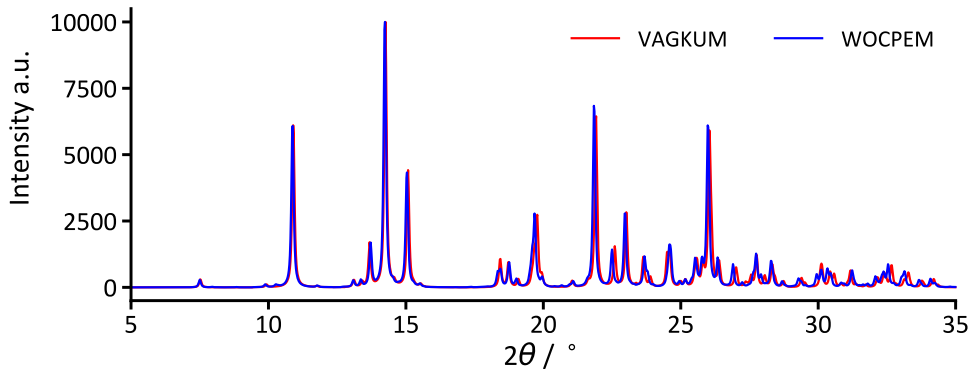


*Figure 3.4: Overlay of the simulated PXRD patterns for **VAGKUM** (red) and **WOCPEM** (blue).*

Alongside the variation in disorder, there was a difference in the *R*-factor between **VAGKUM** and **WOCPEM**. The *R*-factor is a metric that represents the quality of agreement and on first glance, **VAGKUM** appears the better model with its lower *R*-factor.[82,144,145] However, a lower *R*-factor does not necessarily mean a correct structural model.[146] SSNMR can be used to provide complementary information to allow distinction between the structural models.[100]

3.3.2 Initial Calculations

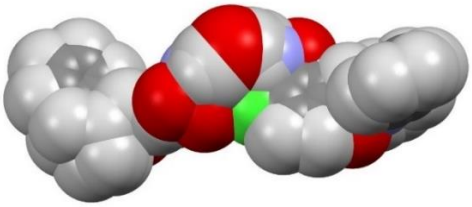
The CIF files for **FOCCAD**, **VAGKUM** and **WOCPEM** were used as downloaded from the CSD and the atomic displacements were evaluated after DFT geometry optimisation, based on the works by Widdifield et al. and van de Streek and Neumann.[101,102] If displacement was greater than 0.25 Å, this indicated a problematic structure where the choice of 0.25 Å as a threshold was to mitigate against the variation in proton positioning during geometry optimisation. From table 3.3, it was clear that **FOCCAD** was significantly problematic whilst **VAGKUM** saw larger RMSD values than desired but this is not surprising, as discussed in the previous section. In contrast, **WOCPEM** observes slight displacement.

*Table 3.3: Derived parameters from geometry optimisation and RMSD between the original atomic positions and geometry optimised positions for **FOCCAD**, **VAGKUM_A**, **VAGKUM_B** and **WOCPEM**.*

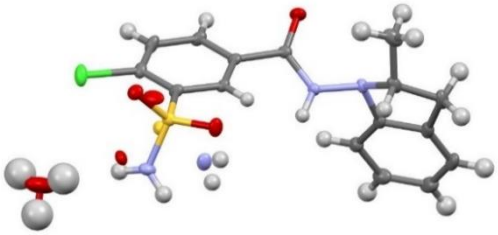
Structure	RMSD (non-H) / Å	Maximum Displacement / Å	Displaced Atom	Region of Molecule
FOCCAD	0.347	2.039	H15	Sulfonamide NH
VAGKUM_A	0.179	1.003	H2N3	Sulfonamide NH
VAGKUM_B	0.141	0.778	H1W	Water
WOCPEM	0.041	0.825	H13A	Water

Atomic displacement (ADPs) ellipsoids, as described by Müller, were also used to visualise the models to identify problematic regions and disorder.[147–150] The ADPs were visualised in figure 3.5 where it further reinforced the flaws in **FOCCAD**. In contrast, the ellipsoids for the water molecule in **VAGKUM** were large but this was expected due to the proposed disorder whilst for **WOCPEM**, the ellipsoids remained relatively small with larger ADPs for exchangeable protons. For both **VAGKUM** and **WOCPEM**, the ADPs do not indicate a severe issue with the structures whilst for **FOCCAD**, there is a clear problem. In combination with the RMSD and maximum displacement values discussed above, **FOCCAD** was not considered further in this work.

a) FOCCAD



b) VAGKUM



c) WOCPEM

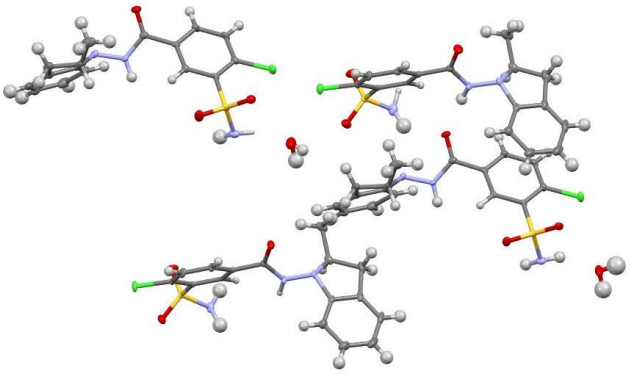


*Figure 3.5: From the standard ellipsoid visualisation in Mercury, the ADP visualisation of a) **FOCCAD**, b) **VAGKUM** and c) **WOCPEM**. The probability level for the ellipsoids was set at 30%.*

3.4 ^{13}C Solid State NMR Spectra

With **FOCCAD** omitted from further study, this left **VAGKUM** and **WOCPEM**. As shown in table 3.2, there are clear variations in the crystallographic parameters, particularly, the value of Z' (Z'= 1 for **VAGKUM** and Z' = 4 for **WOCPEM**). SSNMR was used to identify Z' as the multiplicity of signals present is indicative of the value of Z'.[31]

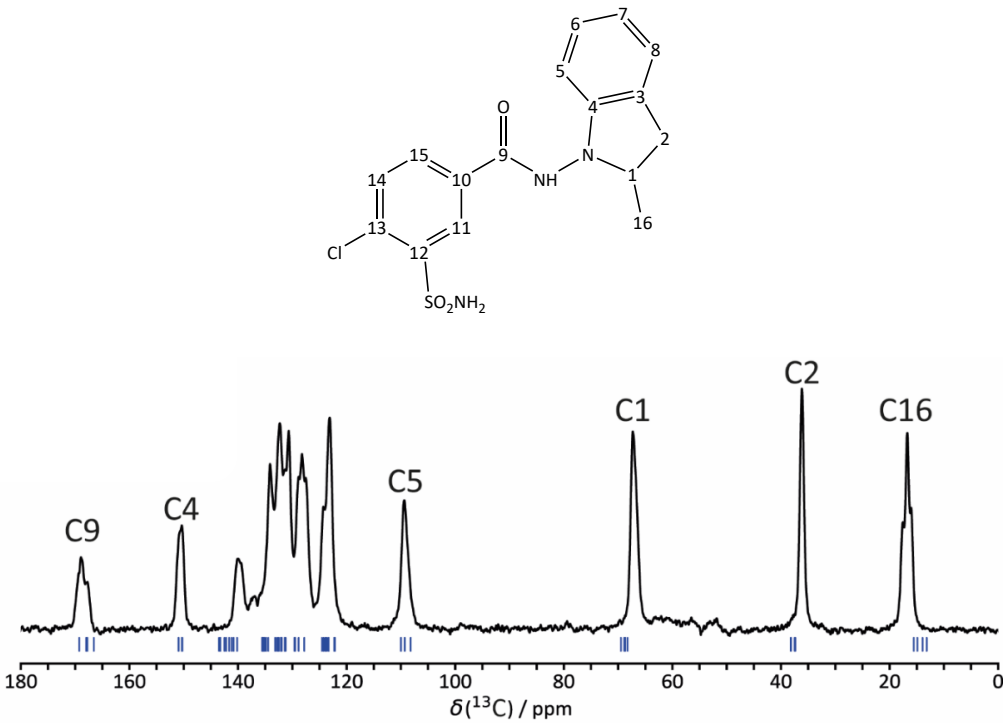


Figure 3.6: A ^{13}C CP/TOSS spectrum of IND acquired at a MAS frequency of 9 kHz with a contact time of 1 ms and recycle delay of 7 s. Further experimental parameters have been described in Section 3.2. The labelled, molecular structure is included alongside the assignment of the subset of environments that were used in the scaling procedure. The GIPAW calculated chemical shifts (blue, dashed lines) of ***WOCPEM*** *are also included.*

For IND, there are 16 carbon atoms therefore, if Z' = 1, there is the expectation of 16 peaks in the ^{13}C CP/TOSS spectrum however, if there are more, a Z' greater than 1 is likely.[151] In figure 3.6, there are more than 16 peaks present which indicates a Z' > 1 structure. The lineshapes of the highest frequency signal (168 ppm) and lowest frequency signal (16.8 ppm) are characteristic of multiple sets of signals. These were evaluated using lineshape analysis.

3.4.1 Lineshape Analysis

To further reinforce that Z' = 4, lineshape analysis of the lowest (**C16**) and highest (**C9**) frequency peaks in the ^{13}C CP/TOSS spectrum was attempted where, as shown in figure 3.7, Z' = 4 fits well with the observed experimental lineshapes.

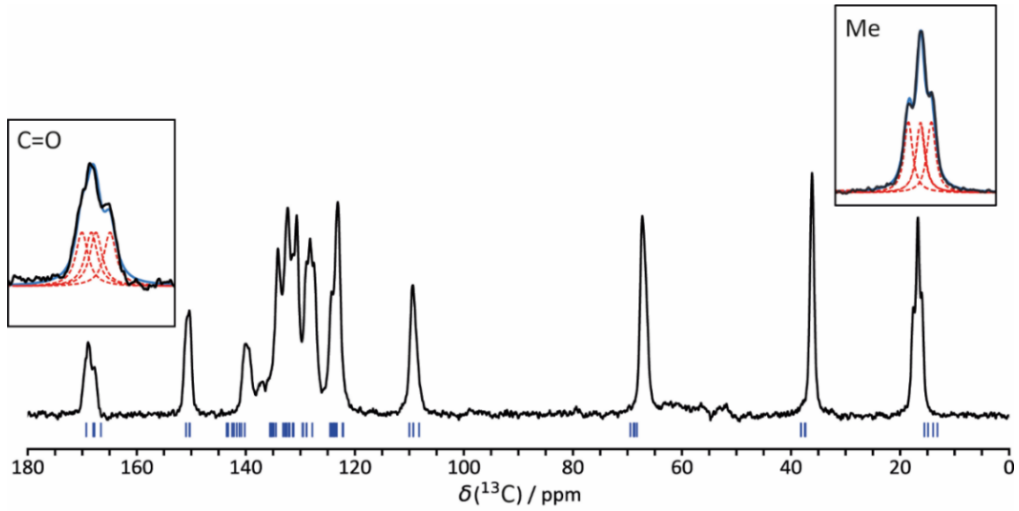


*Figure 3.7: A ^{13}C CP/TOSS spectrum of IND. Adapted from figure 3.6 where inset boxes contain the lineshape deconvolution of **C9** (CO) and **C16** (Me) into four individual components.*

3.4.2 Comparison to Calculations

Using the subset of ^{13}C resonances labelled in figure 3.6, the calculated ^{13}C shielding values were scaled and referenced relative to each structural model for comparison where RMSD was used to assess the agreement. As shown in table 3.4, **WOCPEM** saw the best agreement. However, this must be observed pragmatically as the proposed assignment assumed that the order of calculated chemical shifts correspond to the experimental chemical shifts. Therefore, it is likely that the RMSD values are not representative.[10] Nevertheless, it was apparent that **WOCPEM** is in much better agreement with the experimental data than **VAGKUM**.

*Table 3.4: RMSD values calculated using a subset of carbon environments as labelled in figure 3.6 for **VAGKUM A**, **VAGKUM B** and **WOCPEM.***

Structural Model	Scaling Parameters	RMSD / ppm
VAGKUM_A	$-1.0318x + 173.2$	12.5
VAGKUM_B	$-1.0052x + 173.2$	11.9
WOCPEM	$-0.9458x + 163.5$	1.4

This was reinforced further in figure 3.8 where the qualitative agreement between experimental and calculated chemical shifts was significantly better for **WOCPEM** (blue) relative to **VAGKUM_A** (purple) and **VAGKUM_B** (green). There does appear to be systematic deviations in the alkyl region but this could be the result of the sensitivity of the methyl carbons to nuclear quantum effects, as described by Dračínský and Hodgkinson,

where they observed deviations of up to 20 ppm.[152] Alternate scaling could mitigate against this,[53] however, this was not feasible for IND as there are not enough peaks below 100 ppm for the regression to be meaningful.[144]

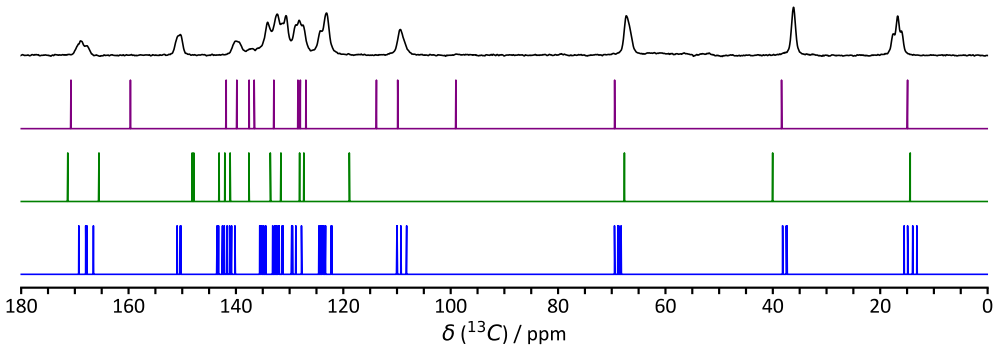


*Figure 3.8: Comparison of referenced GIPAW calculated chemical shifts to experimental ^{13}C SSNMR chemical shifts of **VAGKUM_A** (purple), **VAGKUM_B** (green) and **WOCPEM** (blue).*

3.4.3 Overall Insight from the NMR Crystallographic Approach

Without the requirement for repeat crystallography experiments, the NMR crystallographic approach has confidently demonstrated that **WOCPEM** is the more realistic structural model for IND. From the 1D ^{13}C CP/TOSS spectrum alone, it was apparent that $Z' > 1$ and this therefore ruled out **VAGKUM**. Alongside this, there are no lineshapes that are indicative for disorder. Therefore, it does not appear that IND exhibits any disorder that is modelled in **VAGKUM**. However, to understand the ambiguity surrounding the other structural deposits, additional crystallographic exploration was required.

3.5 WOCPEM01

A recrystallisation following the published method by Aljohani et al. was attempted.[84] This resulted in the crystallographic deposit of **WOCPEM01** which has been visualised in figure 3.9. As shown in table 3.2, **WOCPEM01** mirrored **WOCPEM** in its crystallographic parameters as well as the non-H RMSD and maximum displacement, as tabulated in table 3.6. The GIPAW calculated chemical shifts were also compared in figure 3.9 where the resulting RMSD value was again similar to **WOCPEM** ($-0.9489x + 164.4$, RMSD = 1.32 ppm).

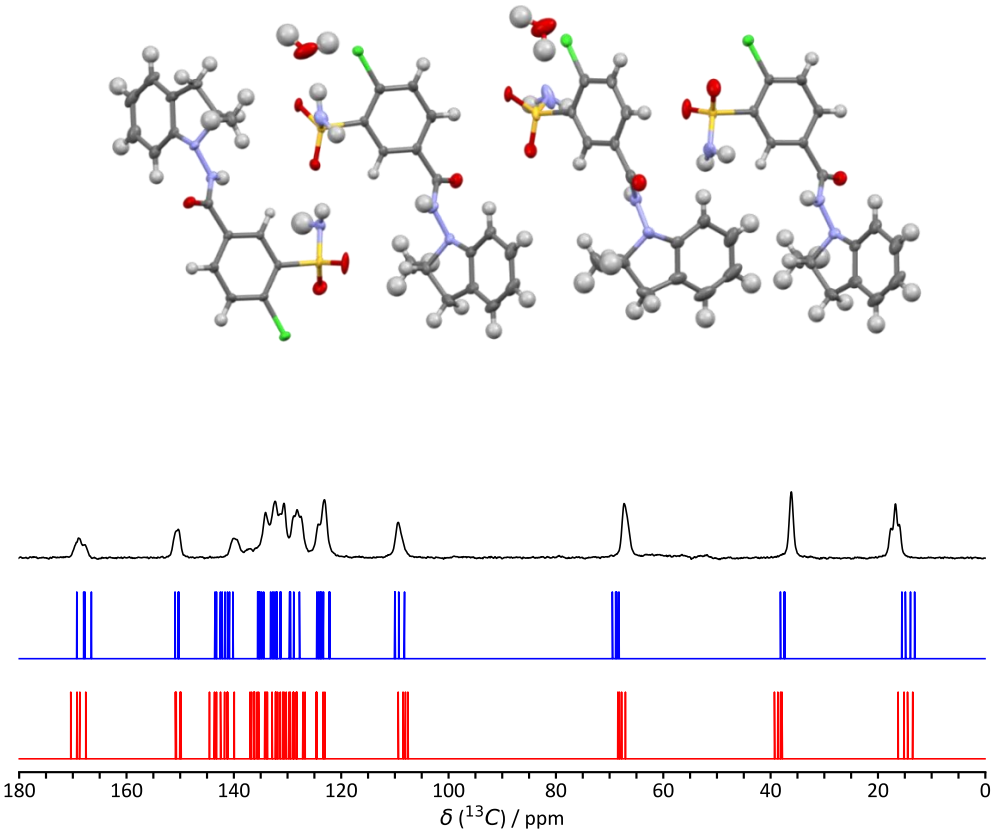


*Figure 3.9: Top is the structural deposit of **WOCPEM01** with standard ADP ellipsoids visualised in Mercury whilst on the bottom is the comparison of referenced GIPAW calculated ^{13}C chemical shifts to ^{13}C CP/TOSS data for **WOCPEM** (blue) and **WOCPEM01** (red).*

*Table 3.6: Derived parameters from geometry optimisation and RMSD between the original atomic positions and geometry optimised positions for **WOCPEM** and **WOCPEM01**.*

Structure	RMSD (non-H) / Å	Maximum Displacement / Å	Displaced Atom	Region of Molecule
WOCPEM	0.041	0.825	H13A	Water
WOCPEM01	0.093	0.826	H3AB	Sulfonamide NH

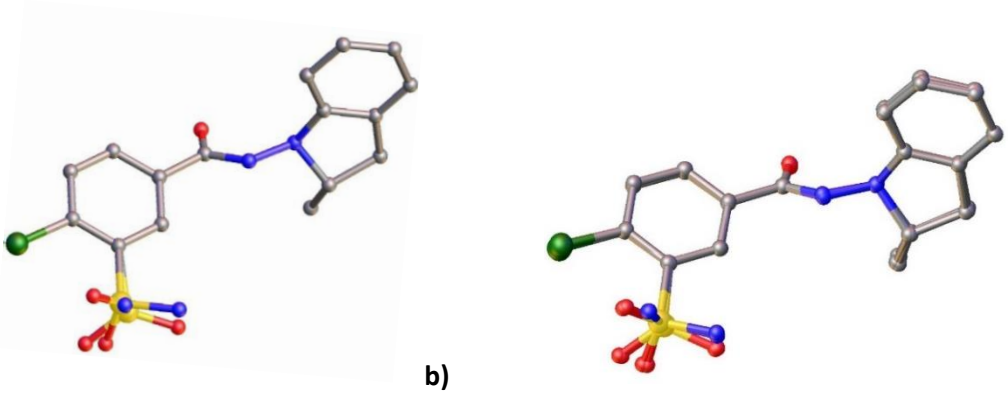
3.6 The Presence of Pseudo-symmetry

IND is an example of a crystal structure exhibiting pseudo-symmetry.[79,153–158] This is when crystallographically independent molecules deviate from regularity, causing an increase of independent molecules in the asymmetric unit.[153,156] Steed and Steed in their 2015 review of

high-*Z´* structures highlighted that 27% structures that have *Z´*> 1 have approximate symmetry elements.[156]

3.6.1 SCXRD

The four IND molecules in the asymmetric unit are crystallographically independent and related by symmetry except for the orientation of the sulfonamide group. The sulfonamide group sits slightly differently in each of the four IND molecules which disrupts the periodicity, breaks symmetry and results in an increase of *Z´* to 4.



a) **b)**

Figure 3.10: Comparison between a) ***VAGKUM*** *and b)* ***WOCPEM01*** *asymmetric units. There are similarities in the positioning of the sulfonamide group which may have resulted in the erroneous disorder conclusion for* ***VAGKUM****. The visualisation was achieved using Olex2.*[112]

An overlay of the four independent molecules is shown in figure 3.10 b). The presence of pseudo-symmetry can complicate the choice of space group.[153,156] If pseudo-symmetry is mistaken for true crystallographic symmetry, this will result in an erroneous structural solution.[156] Within the experimental XRD data, a pseudo-symmetric structure produces strong reflections arranged as expected, but the deviation from regularity produces additional weak reflections.[157] This was observed for IND. A direct comparison of the overlaid IND molecules in **WOCPEM01** to **VAGKUM** in figure 3.10 shows the similarity in how the sulfonamide group is modelled. It is theorised that **VAGKUM** and **WOCPEM** are modelling the same material, but the key difference is that **VAGKUM** has mistaken the pseudo-symmetry for disorder.

_R_-factor Variation

WOCPEM does have a considerably larger *R*-factor (8.7%) relative to **VAGKUM** (3.7%). Several factors do contribute to the final *R* factor which may result in a higher value for a better structure.[146,159] One of which is the wavelength used as this can increase the number

of reflections considered.[79] This correlates with the number of reflections considered by **WOCPEM** relative to **VAGKUM** as the former reflects the inclusion of additional weak reflections present in the **WOCPEM** and **WOCPEM01** data.[160]

3.6.2 PXRD

Alongside the SCXRD study, the PXRD pattern was analysed further with the help of Prof. Ivana Radosavljevic Evans to determine whether the weak reflections are observed. The (*hkl*) file from SCXRD data acquisition can provide insight about the intensities of the reflections in the CIF file.[161]

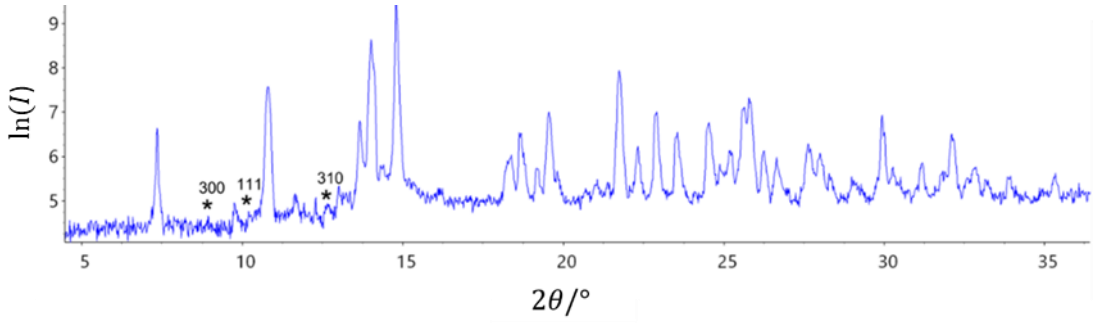


*Figure 3.11: The truncated experimental PXRD pattern of **WOCPEM01** with the positions and very weak intensities for the (300), (310) and (111) reflections of unit cell marked with an asterisk (*). The data has been visualised in Topas Academic Software by Prof. I. R. Evans.*[137]

The presence of superstructure reflections is indicative of a pseudo-symmetric structure. The simulated PXRD patterns of **VAGKUM** and **WOCPEM** in figure 3.4 look identical on first glance but further inspection of the simulated patterns saw additional superstructure reflections for **WOCPEM**. As NMR indicated that **WOCPEM** (and **WOCPEM01**) was the correct structural model, there was the possibility that the very weak reflections would be seen in the acquired PXRD pattern.

As shown in figure 3.11, relative to the other reflection intensities, the highlighted superstructure reflections for **WOCPEM01** were around 3-4 orders of magnitude less than other reflections but critically, they were present. To exaggerate the weak intensities, a log scale was used to compress the vertical axis.

Table 3.7 also reflects this with comparison of the average intensities between expected reflections and identified superstructure reflections from the *hkl* file. Overall, the evidence of superstructure reflections in the PXRD data further supported that **WOCPEM** (and **WOCPEM01**) is the correct structural model of IND.

*Table 3.7: Tabulation of (hkl) intensities for **WOCPEM01** which are superstructure reflections.*[4]

(*hkl*) Reflection	Average Intensity (I)	σ(I)	2θ / °
(3,0,0)	9.9	1.2	8.6
(3,1,0)	13.9	1.8	10.2
(1,1,1)	25.4	2.4	12.6
(4,0,0)[5]	*119.0*	*9.3*	
(16,0,0)[4]	*3053.1*	*228.9*	

3.7 Conclusions

This chapter aimed to demonstrate the power of NMR crystallography in the validation of structural models through the example of IND. From the ^{13}C NMR alone, it was straightforward to identify that *Z´* is greater than 1, hence determining that **WOCPEM** is the most likely structural model. This was reinforced with lineshape analysis as well as first principles calculations. To determine why there are discrepancies between the models, a crystallographic investigation was conducted. It identified that IND is a molecular organic solid that exhibits pseudo-symmetry and that **VAGKUM** mis-identified this as disorder.

[4] *To note, for each reflection, the absolute average of all related intensities and related averaged errors (σ) were calculated from the individual values located in Table A1 in the Appendix.*

[5] *The hkl reflections of (4,0,0) and (16,0,0) have been included solely as a comparison of the intensity magnitude to the identified superstructure reflections in figure 3.11. The 2θ values have not been included.*

Chapter 4: Development of a Bayesian Probability-Based Workflow to Aid Solid State NMR Assignment

4.1 Introduction

A key step in any structural study is the assignment of chemical shifts to molecular sites. This is particularly important when the acquired experimental data is being used for comprehensive structural studies rather than as a conventional fingerprint tool. The NMR crystallographic approach can facilitate this where agreement between experimental chemical shifts and GIPAW calculated chemical shifts, reinforced by additional experimental data, can indicate the compatibility of a proposed assignment.[10,12] The benefit of additional experimental data to aid assignment has been highlighted in several reviews.[10,12,38,82]

The use of calculated chemical shifts has been instrumental in the identification of mis-assignments within the literature.[10] Such examples include naproxen[162–164] and 6-aminopencillanic acid (6-APA)[165–167] whose initial assignment were revised when the inclusion of additional experimental data and corresponding calculated chemical shifts resulted in a modified assignment. Unfortunately, mis-assignments are somewhat common, as highlighted in the 2020 review by Hodgkinson.[10] Therefore, methods by which to avoid such scenarios are desirable.

The 2020 review also highlighted how spectral overlap can complicate assignment. This was seen for naproxen where the disputed assignment region involved partial overlap of the ^{13}C signals.[162–164] Methods, such as MAS and robust decoupling sequences, can be used to improve signal resolution, but particularly for ^{1}H signals, overlap can remain. As a result, assignments tend to be partial where well-resolved ^{13}C or ^{1}H signals are confidently assigned whilst less resolved signals require additional data acquisition. This means replication of a published assignment can be tedious and difficult to achieve. The problem is only exacerbated when the number of signals increase for more complex materials, such as $Z' > 1$ materials.

To tackle overlap, experimental methods, such as spectral-editing and multidimensional spectra, can be used to provide additional insight. This is demonstrated for L-tyrosine methyl ester hydrochloride (Tyr) in figure 4.1. As seen in the 50-55 ppm region, there is partial overlap of two signals so the environments (marked with asterisks) cannot be assigned confidently. However, with the ^{13}C NQS spectrum in red, a confident assignment can be made where the Me group is assigned to the higher frequency signal of 53.7 ppm and the CH signal to 53.1 ppm. The ^{13}C NQS can be used to distinguish between quaternary and protonated carbons, as discussed in section 2.2.2, however, the fast rotation of CH_3 groups reduces the suppression observed.[50] This then allows for the distinction between the CH and the CH_3 peaks in Tyr, as seen in figure 4.1 below. Interestingly, if only the GIPAW calculated chemical shifts were considered, the assignment would be incorrect.

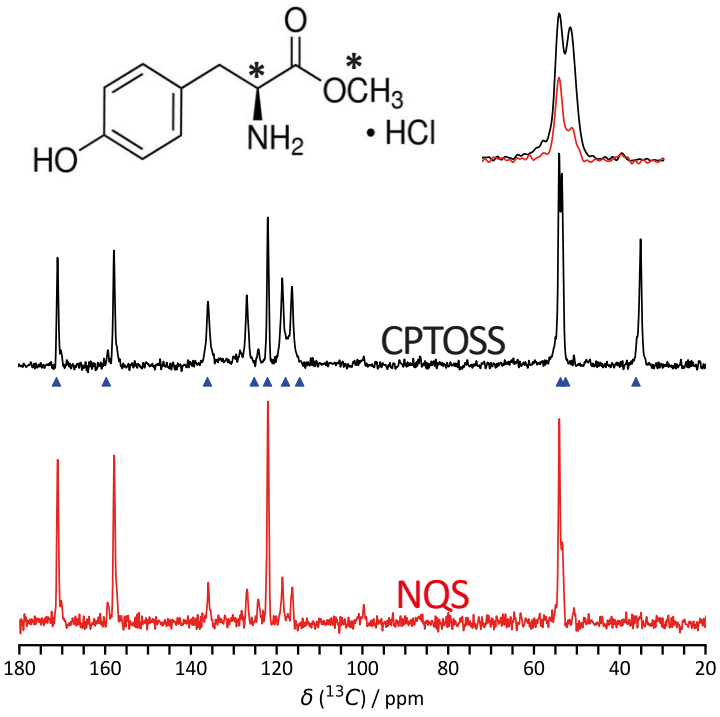


Figure 4.1: A ^{13}C CP/TOSS (top) and ^{13}C CP/TOSS/NQS (NQS) of L-tyrosine methyl ester hydrochloride with the GIPAW calculated chemical shifts visualised with blue triangles. Acquisition parameters can be found in Section 4.2.

4.1.1 Quantitative Methods to Assess Solid state NMR Assignment

The quantification of assignment compatibility tends to be limited to the agreement between the experimental and calculated chemical shifts. There are several methods to evaluate agreement including Mean Absolute Error (MAE) in equation 4.1, and root-mean-square deviation (RMSD) in equation 4.2. The outputted values reflect how much the proposed experimental assignment deviates from the calculated chemical shifts.[10]

$$\mathrm{MAE} = \frac{\sum_i |\delta_{\mathrm{exp},i} - \delta_{\mathrm{calc},i}|}{N} \qquad \text{Eq. 4.1}$$

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i}^{N}(\delta_{\mathrm{exp},i} - \delta_{\mathrm{calc},i})^2} \qquad \text{Eq. 4.2}$$

These metrics tend to pragmatically evaluate and potentially validate global assignments but they tend to falter when faced with overlap as the agreement between possible assignments do not indicate a preference. Typical values for a respectable agreement for ^{13}C and ^{1}H chemical shifts is around 1.9 ± 0.4 ppm and 0.33 ± 0.16 ppm, respectively.[103,168] So, if possible assignments fall within these bounds, it is difficult to argue which assignment is most preferable. As an assignment queries small differences between signals in a similar region of the spectrum, the metrics do not aid in the process unless additional weighting parameters are introduced. However, this can complicate interpretation more, as decisions must be made as to how the weighting is done which can introduce ambiguity.[10] For context, the two possible assignments of Tyr discussed in figure 4.1, the resulting RMSD values (which are equally weighted) were 1.13 and 1.06 ppm. Both values fall within the proposed ^{13}C bounds and only vary by 0.07 ppm so no distinction is achieved. Additional quantitative methods have also been proposed to consider other pieces of experimental data such as the approach by Brus and Czernek to evaluate ^{1}H-^{13}C HETCOR quantitatively. However, this method suffers poor convergence and is not straightforward to implement so will not be discussed further.[164,169,170]

4.1.2 Bayesian Probability in NMR

In Solution

Another way to evaluate the compatibility of an assignment is through Bayesian probability-based methods. The DP4 parameter, first proposed by Smith and Goodman in 2010, has been embraced within the solution-state NMR community.[27] Its implementation has resulted in over 650 citations at time of writing, with several iterations published since.[28,29,179,171–178]

DP4 was designed to calculate the relative probability between alternative structures, mainly diastereoisomers, through comparisons between the experimental and calculated ^{13}C and ^{1}H chemical shifts.[27] To note, this has been taken a step further with the DiCE workflow where ^{15}N chemical shifts can also be considered.[176] Critically, DP4 was demonstrated to outperform

alternate metrics, such as MAE, in identifying the correct diastereoisomer from the possibilities.[27]

There are two key assumptions made when using DP4. First, the workflow assumes that the error (*e*), defined as the absolute difference between the experimental and calculated chemical shift values, is random and uncorrelated. The second assumption is that the resulting error values can be fitted to either a Student's *t* distribution, defined by the mean (μ), standard deviation (σ) and degrees of freedom (ν)[27] or a normal distribution, defined by μ and σ.[172] Most recent iterations of the DP4 script have adopted the latter distribution function.[172–174,177] The mathematical expressions for distribution functions used in DP4 and iterations are shown in equations 4.3 and 4.4. The μ is assumed to be 0, due to the associated linear scaling of calculated chemical shifts, and the σ values (and ν values where appropriate) are nucleus specific. $P_{\text{shift,A}}$ refers to the individual probability of assignment A, T represents the Student's t distribution and N reflects the normal distribution. These are cumulative distribution functions. σ (and ν in equation 4.3) are chosen based on the nucleus under study, as discussed further in section 4.3.2.

$$P_{shift,A} = \prod_{k=1}^{N} 1 - T^{\nu}\left(\frac{|\delta_{calc,k} - \delta_{expt,k}|}{\sigma}\right) \qquad \text{Eq. 4.3}$$

$$P_{shift,A} = \prod_{k=1}^{N} 1 - N\left(\frac{|\delta_{calc,k} - \delta_{expt,k}|}{\sigma}\right) \qquad \text{Eq. 4.4}$$

$$AF(A\ not\ B) = \frac{P_A}{P_A + P_B} \qquad \text{Eq. 4.5}$$

Any chemical shift for any nucleus can be included in this way as long as the distribution parameters have been derived.[176] To calculate the percentage, henceforth known as the agreement factor (AF), Bayes theorem can be used to discriminate between alternative possibilities. This is shown in equation 4.5 but explained further with the following example. Take two assignments, A and B, with the individual probabilities ($P_{shift,A}$ and $P_{shift,B}$) which are calculated using equation 4.4. The AF for each assignment, in percentage units, can then be calculated using equation 4.5 to determine the probability that either A or B is correct. As it is a relative probability, any systematic limitations are cancelled out and the sum of the two AF values must be equal to 100%.

In solids

Within the solid state, the application of Bayesian probability-based methods is not novel as the Ceriotti and Emsley groups were the first to propose a Bayesian workflow.[180,181] Engel et al. initially proposed a Bayesian approach to help determine confidence in the identification of experimental crystal structures through the agreement between experimental and GIPAW calculated chemical shifts.[180] This has been built on in the subsequent 2021 publication by Cordova et al. where the Bayesian approach has been used to determine the compatibility of a proposed experimental assignment to a 2D structure.[181] This was achieved with the construction of a chemical shift database that used both the CSD and a machine learning model to generate chemical shift predictions. From the assembled database, the probabilistic assignment of the considered materials was then assessed.[181] An alternative approach is OPTICS, which is more reflective of the original DP4 paper, where the error distribution is evaluated using the Student's *t* distribution.[182] The approach was developed using the fundamentals of the solution-state DiCE workflow,[176] where the aim of OPTICS is to evaluate ^{13}C chemical shifts in order to identify the correct polymorph present in the acquired experimental data.[182] However, neither approach quantitatively considers additional experimental data, such as spectral-editing and 2D experiments.

4.1.3 Aims of this Work

The aim of this work was to establish a proof of principle for a Bayesian probability-based workflow that allows the incorporation of additional experimental data to quantitatively determine the percentage compatibility of an assignment. The workflow implementation will be discussed through a worked example of Tyr alongside a selection of model systems. It is hoped that, with the establishment of precedent, future work will focus on the development of a semi-automated workflow. This will allow not only collation of acquired data in one place, but also a straightforward evaluation of compatibility for a proposed assignment within a processing software.

4.2 Materials and Methods

4.2.1 Synthesis

L-tyrosine methyl-ester hydrochloride (Tyr) (CAS Number: 2417-91-2), isoniazid (INH) (CAS Number: 54-85-3) and co-formers 2-hydroxycinnamic acid (2HCA) (CAS Number: 614-60-8), 3-hydroxycinnamic acid (3HCA) (CAS Number: 588-30-7) and 4-hydroxycinnamic acid (4HCA) (CAS Number: 501-98-4) were purchased from Merck Life Sciences Ltd and used without further purification.

The co-crystals, INH-2HCA, INH-3HCA and INH-4HCA were successfully synthesised via liquid assisted grinding (LAG) of a 0.5 g 1:1 stoichiometric ratio of INH and the related co-former with minimal quantities of MeOH (< 1 mL) for 35 minutes.[183] To ensure the co-crystals were generated successfully, the 1D ^{13}C CP/TOSS spectrum was compared to reference 183.[183] Crystallographic data was acquired previously[183] (Supplementary publications CCDC 1993881-1993887) for INH-2HCA (**PUMFOV**), INH-3HCA (**PUMGOW**) and INH-4HCA (**PEHFUF01**). For tyrosine methyl-ester hydrochloride, the CSD deposit, **YEFTUZ**,[184] was used. For published data discussed in Section 4.5, the relevant publications have been cited alongside the specified crystallographic files used to calculate the isotropic shieldings.

4.2.2 Solid state NMR

For Tyr and the INH co-crystals, the experimental data was acquired using an Oxford superconducting magnet operating at 11.7 T (^{13}C of 125.65 MHz), and a Bruker AVANCE III HD console which corresponds to frequencies of 499.69 (^{1}H) and 125.65 (^{13}C) MHz. The data was collected using a 4 mm HX magic-angle spinning probe at ambient temperature. The ^{13}C chemical shifts were referenced using the high frequency signal (δ_{iso}(^{13}C) = 38.5 ppm) of adamantane. The chemical shifts for the second axis in the 2D ^{1}H-^{13}C HETCOR spectra were referenced by using the relevant 1D ^{1}H MAS spectra or by using calculated characteristic ^{1}H chemical shifts. For the INH co-crystals, ^{1}H NMR spectra were obtained under fast MAS with a recycle delay of 180 s over 16 transients using the 9.4 T Bruker AVANCE III instrument. The experimental data was acquired by Dr. David Apperley.

^{13}C CP/MAS measurements were recorded with[49] and without sideband suppression between 9 and 10 kHz. The recycle delay and contact time were optimised for each sample but typical values for the contact time were 0.1 ms where acquisition was over 64-256 transients with SPINAL-64[42,185] used as the heteronuclear decoupling element at a ^{1}H nutation frequency of approximately 73 kHz and a ^{1}H 90° pulse length of 3.4 µs (INH-2HCA, Tyr) or 3.6 µs (INH-2HCA, INH-4HCA). Non-quaternary suppression (dipolar dephasing / NQS) spectra[50] were acquired with a typical dephasing delay of 50 µs whilst ^{1}H-^{13}C HETCOR spectra were recorded at 10 kHz MAS with a Hartmann-Hahn[47] CP under sample-specific optimised conditions.[44,51,52] In general, the contact times used were 0.1 ms and 0.4 ms with a sample optimised recycle delay. An FSLG pulse element[44] was used for homonuclear decoupling (FSLG 2π pulse of 19.8 µs) whilst SPINAL-64[42,185] used as the heteronuclear decoupling element at a ^{1}H nutation frequency of approximately 73-75 kHz (sample dependent) and a ^{1}H 90° pulse length of 3.2 µs (INH-3HCA, INH-4HCA), 3.4 µs (Tyr) or 3.6 µs (INH-2HCA). 32-80 transients were collected for each of the 64 (INH-2HCA, INH-3HCA, INH-4HCA) or 24 (Tyr) t_1

increments with a spectral width of 50 kHz and approximately 19-20 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 8 step phase cycle.

4.2.3 Calculations

In this work, the GIPAW method implemented in CASTEP, version 22.1 was used.[85] Calculations were performed as described in previous chapters. The cut-off energy for the geometry optimisation was set to 600 eV with a maximum k-point spacing of 0.1 $Å^{-1}$ and a k-point offset of (¼,¼,¼) in fractional reciprocal space. The NMR parameter calculation used the same parameters, except the value E_{cut}, was increased to 900 eV after convergence testing on a select group of molecules, as seen in section 2.5.2.[94,98] Only the H-atom positions were geometry optimised.

4.3 Workflow Development

4.3.1 Methodology Overview

An essential feature of the Bayesian approach is that information from multiple experiments can be incorporated simply by multiplying the resulting values together to determine the overall AF. Chemical shifts are evaluated using equation 4.4 but in order to incorporate additional experimental data, an alternative term would be required. The individual error values were instead assigned based on the type of experiment[6] ($e_{\lambda,i}$), as shown in equation 4.6 where λ reflects the type of experiment under consideration.

$$P_{\lambda,A} = \prod_{i=1}^{N} e_{\lambda,i} \qquad \text{Eq. 4.6}$$

$$P(A) = P_{shift}(A)P_{NQS}(A)P_{SCHET}(A) \ldots \qquad \text{Eq. 4.7}$$

The individual values of $e_{\lambda,i}$ have been summarised in table 4.1 which are specific to the type of experiment and the information that is obtained. The additional "evidence" from these experiments can then be easily incorporated into $P(A)$ of equation 4.5 through multiplication of the resulting probabilities, as shown in equation 4.7.

[6] *The values of $e_{\lambda,i}$ stated in table 4.1 have been determined based on the level of confidence one can place on a given experiment in relation to the information it provides. For example, in this chapter, a higher $e_{\lambda,i}$ value is used for the scHETCOR spectrum relative to the ^{13}C NQS spectrum as the ^{13}C NQS experiment can suffer from inefficient suppression, as discussed further in Section 4.4.1 for Tyr.*

Table 4.1: The type of experiment, information derived and probability determination.

Experiment	Information	Determination of Probabilities ($P_{\lambda,A}$)	
^{13}C & ^{1}H Chemical Shifts	Agreement between experimental and calculated chemical shifts	Equation 4.4	
^{13}C NQS	Suppression of CH and CH_2 signals	Equation 4.6 where e_{NQS} is 0.8	
		Peak is suppressed for CH/CH_2 signal	e_{NQS}
		Peak is present for C/Me signal	e_{NQS}
		Peak is suppressed for C/Me signal	$1 - e_{NQS}$
		Peak is present for CH/CH_2 signal	$1 - e_{NQS}$
^{1}H-^{13}C scHETCOR (Short Contact Time)	Whether an environment is protonated	Equation 4.6 where e_{SCHET} is 0.9	
		Cross peak present for $CH/CH_2/CH_3$ signal	e_{SCHET}
		Cross peak absent for C signal	e_{SCHET}
		Cross peak absent for $CH/CH_2/CH_3$ signal	$1 - e_{SCHET}$
		Cross peak present for C signal	$1 - e_{SCHET}$
^{1}H-^{13}C lcHETCOR (Long Contact Time)	Long-range dipolar correlations dependent on C-H distance (r)	Equation 4.6 where e_{LCHET} is 0.75 AND is C-H distance dependent (r)	
		Cross peak present where $r < 1.5$ Å	$e_{LCHET} + 0.15$
		Cross peak present where 1.5 Å $< r <$ 2.5 Å	e_{LCHET}
		Cross peak present where 2.5 Å $< r <$ 5.0 Å	$e_{LCHET} - 0.15$
		Cross peak absent where 2.5 Å $< r <$ 5.0 Å	$1.15 - e_{LCHET}$
		Cross peak absent where 1.5 Å $< r <$ 2.5 Å	$1 - e_{LCHET}$
		Cross peak absent where $r < 1.5$ Å	$0.85 - e_{LCHET}$

4.3.2 Workflow Implementation

Objectives

The complete, or as near as complete, assignment of the ^{13}C chemical shifts is the primary objective of the workflow. The secondary objective is to aim for a complete, or as near as complete, assignment of the ^{1}H chemical shifts. This is because a near-complete ^{1}H assignment is sometimes critical to understand features of the material studied, such as intermolecular contacts.[186,187]

Stages

The workflow is an iterative approach that is split into stages. This allows for the introduction of new data until the desired objective is reached. This does vary from OPTICS, which focuses on ^{13}C chemical shifts whilst Cordova et al. focuses on characterisation based on a 2D structure.[180,181] The latter approach is ambitious and tackles the problem of 'de novo' chemical shift-based NMR crystallography.[10] This varies from the proposed workflow, where it is assumed that the 3D crystallographic structure is known but the assignment is not.

Choice of Thresholds

For this workflow, a set of AF thresholds were used to reduce the number of possible assignments under consideration and to identify a confident assignment. If an AF exceeds 90%, it is considered a confident assignment so no further experimental validation is required. In contrast, if an AF value falls below 1%, the assignment is discarded, thus reducing the possible assignments under consideration. This has been exemplified for the worked example Tyr in section 4.4. The magnitude of the raw probabilities can also be used to limit the number of possible assignments, as seen for naproxen in section 4.5.

Choice of Distribution and Associated Parameters

So, as to mirror the current iteration of DP4, the normal distribution has been used to evaluate error.[177,180] The values of σ used in DP4 were 2.27 ppm and 0.19 ppm respectively for ^{13}C and ^{1}H but in order to reflect the variation in calculation methods (B3LYP functional[188] against PBE-based GIPAW calculations[94,98]), more conservative values of σ were adopted.

The values used were 2.3 ppm and 0.5 ppm for ^{13}C and ^{1}H respectively which were chosen based on the upper uncertainty limits proposed by Salager et al.[103] In contrast, OPTICS, used values of 0.07 ppm, 1.71 ppm and 6.25 for μ, σ and ν which were determined from a fit of

1111 chemical shifts to a Student's *t* distribution.[182] The choice of Student's *t* distribution was appropriate as the value of v was small.

Referencing

First principles calculations of chemical shieldings require conversion to scaled and referenced GIPAW calculated chemical shifts prior to comparison to experimental data. There are a variety of different approaches to referencing within the literature, which is reflected in the publications that discuss solid state Bayesian probability approaches.[180–182]

The method used in this work was system-specific, 'on-the-fly' linear regression of the calculated shieldings against the experimental chemical shifts where the gradient (*m)* was set to –1. Engel et al.[180] did describe how allowing the gradient to deviate from –1 can be helpful to remove system-dependent systematic errors but, in order to be effective, it requires a spread of values across a wide frequency range and it is more appropriate when calculating RMSD.[10] When environments are clustered, as seen in section 4.5, the reference scaling term (σ_{ref}) can vary considerably when *m* is allowed to deviate therefore, regression can be poor. It was therefore judged to use scaling that set *m* to –1.

An alternative referencing approach is benchmark studies where a large collection of systems are used to establish referencing parameters.[104] This is a preferred approach when a solid material has only a few environments, such as ^{15}N, so as to avoid scaling to a straight line.[176] There is no correct approach but as reference parameters can be system-dependent and vary noticeably for particular environments, system-specific scaling is used for this workflow.[104,180] The wide variation for environments was noted by the authors of the OPTICS approach, where despite their preference of benchmark scaling, deviations of up to 10.6 ppm were seen between calculated and experimental chemical shifts.[182]

Grouping

Rather than considering all of the possible global assignments, the workflow is greatly simplified by the grouping of ^{13}C (and ^{1}H where appropriate) resonances according to their ^{13}C (or associated ^{13}C) chemical shift. This then can identify which groupings require evaluation but also environments that can be omitted as their assignment is certain. This is a similar approach to Cordova et al., where the resonances were grouped based on their protonation state.[181] This type of grouping was suitable as their input was a 2D molecular structure which carries a greater risk of incorrect chemical shift prediction.

For this chapter, the groupings were capped at 4 environments to keep the possible assignments limited to 24 unique possibilities (as *N* distinct environments results in *N*! possible assignments). If there is an increase of 1, the possible assignments increases to 120. This will ultimately be unfeasible to calculate by hand but will be a key area that will be developed and automated once precedent for the workflow has been determined.

Consideration of Overlap

As the workflow focuses on a known crystal structure, the calculated chemical shifts, in theory, should match the number of the experimental chemical shifts present in the data. Unfortunately, overlap remains a common limitation.

For DP4, the recommendation for overlapped resonances was to find the midpoint and assign all relevant environments to the same chemical shift as there was limited effect on the overall metric.[27] However, this is not practical in the solid state as overlap is much more common and averaging such environments would result in the loss of key information. Instead, overlap has been incorporated into this workflow, where more than one environment can be assigned to a particular chemical shift. The resulting probabilities and subsequent AFs also need to take the presence of overlap into account. This is best demonstrated for the INH co-crystals in section 4.5, where similarities in the co-crystal structure result in a narrow range of both ^{1}H and ^{13}C frequencies which exhibit overlap.

4.4.4 Worked Example

To demonstrate how the workflow calculates compatibility of a proposed assignment, Tyr has been used as an example. There are 10 carbon environments, labelled from **C1** to **C10**, which were grouped based on their ^{13}C chemical shift, as shown in table 4.2.

Table 4.2: Proposed groupings of the experimental chemical shifts of Tyr alongside the GIPAW calculated ^{13}C and ^{1}H chemical shifts for each environment.[7]

Grouping	Experimental ^{13}C (^{1}H) Chemical Shifts / ppm	Possible Assignments	Calculated ^{13}C(^{1}H) Chemical Shifts / ppm
1	171.0	C1	172.6
2	157.8	C7	160.7
3	135.8	C5	136.6
4	126.8 (5.4) 121.9 118.5 (5.5) 116.2 (6.2)	C4, C6, C8, C9	C9 – 125.6 (5.1) C4 – 122.4 C8 – 118.1 (5.5) C6 – 114.8 (6.3)
5	53.7 (3.1) 53.1 (3.9)	C2, C10	C2 – 52.6 (4.2) C10 – 51.4 (3.1)
6	34.7 (2.4)	C3	34.7 (1.9, 2.6)

The overall workflow is shown in figure 4.2 with the considered experimental data in figure 4.3.. The two groupings of interest were Groups 4 and 5 and these have been evaluated in figure 4.2. The other groupings did not need to be considered as their assignment can be confidently identified based on qualitative comparison between experimental and calculated chemical shift in table 4.2.

[7] *The CSD deposit, **YEFTUZ** was used as the input .cif file for the calculations. The σ_{ref} values used for ^{13}C and ^{1}H were 167.9 and 29.3 ppm.*

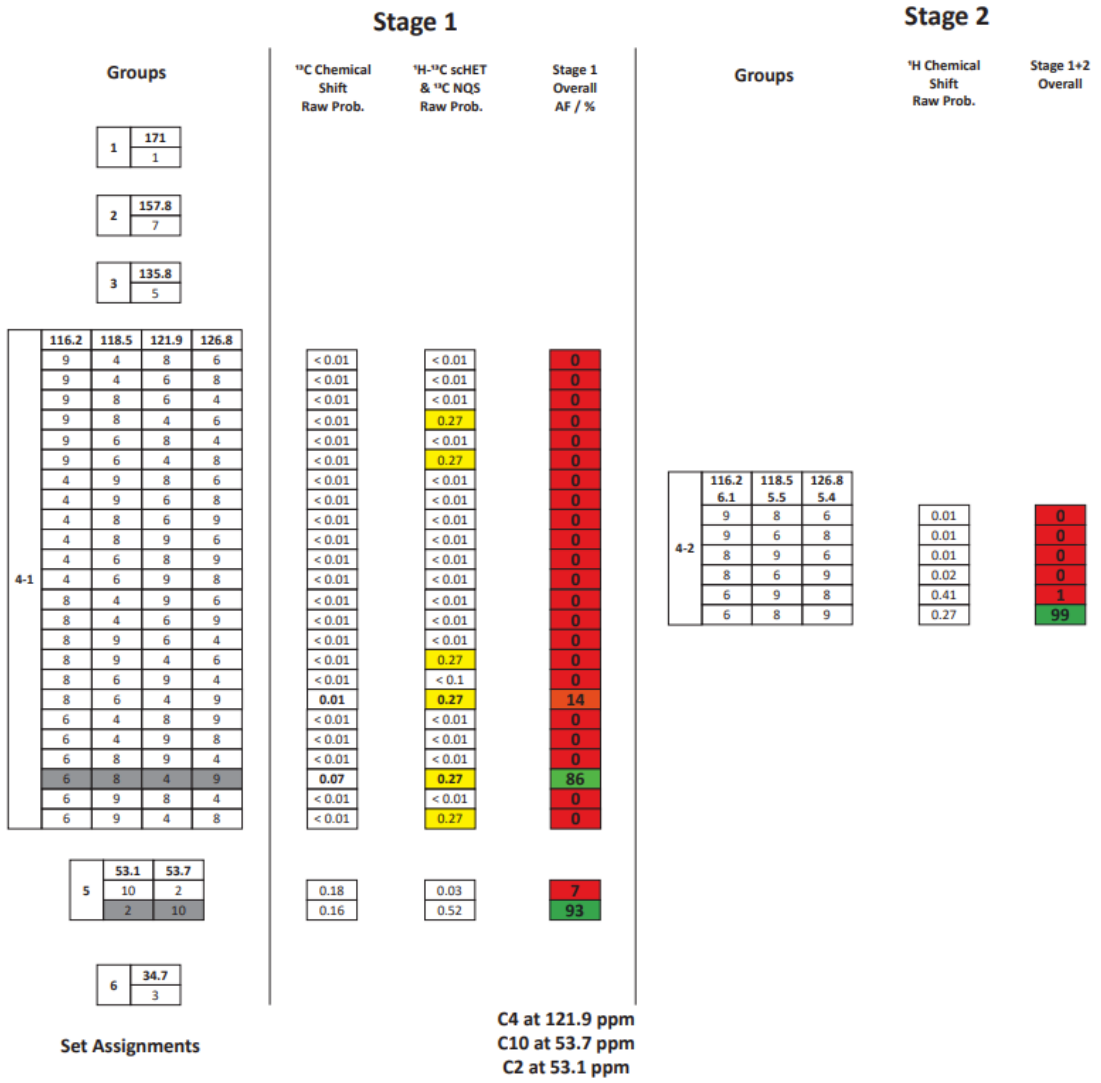


Figure 4.2: Graphical representation of the assignment workflow for Tyr. For each group of resonances with similar shifts, the possible assignments of carbon sites to ^{13}C chemical shifts (in ppm) are listed. At each stage, the probability of a given assignment is determined and from this, an overall colour-coded AF value is calculated. If an assignment exceeds the 90% threshold, a confident assignment can be made. Where this threshold is not met, additional data is added in subsequent stages, and the overall AF re-calculated. Note, in the fourth group, the number of possible assignments has been reduced from 24 to 6 due to the application of AF < 1% filter. The relevant raw probability values have been highlighted in yellow whilst the most probable assignment in each group is highlighted in grey.

4.4.1 Workflow Implementation – Stage 1

Stage 1 of the workflow takes into account the ^{13}C chemical shifts alongside the combination of short-contact ^{1}H-^{13}C HETCOR and ^{13}C NQS experiments. The latter two experiments are considered together as they are expected to provide complementary information in relation to the presence of protonated and quaternary carbons. The short-contact ^{1}H-^{13}C HETCOR experiment also provides the experimental ^{1}H chemical shift assignment, as demonstrated in figure 4.3.

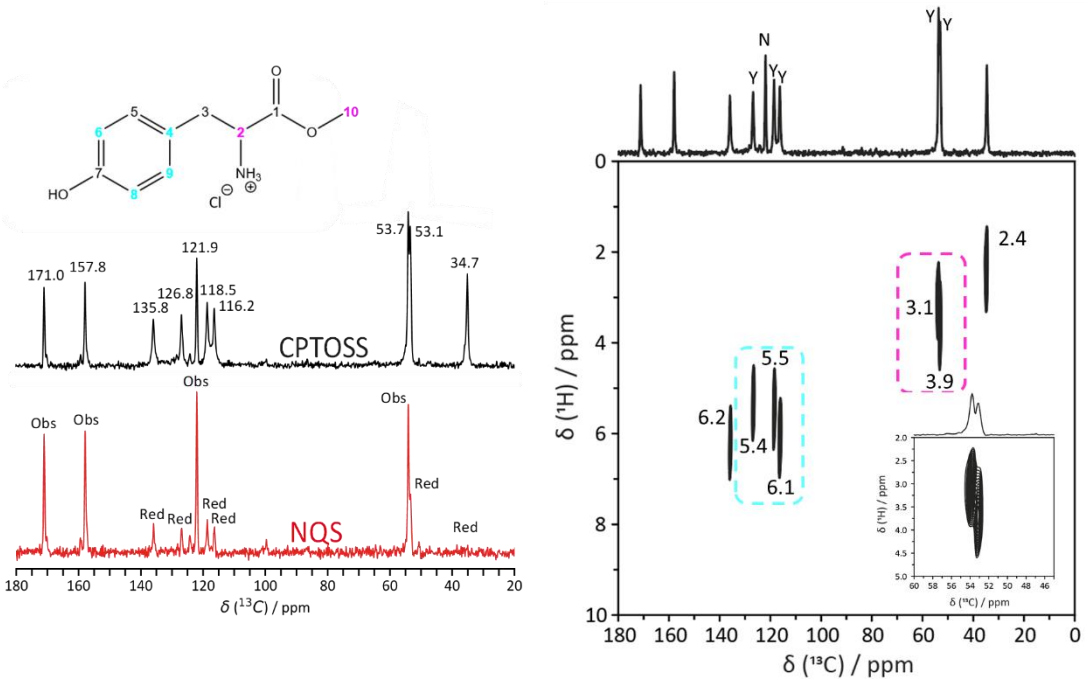


Figure 4.3: Experimental data for Tyr used in Stages 1 and 2 alongside the molecular structure of Tyr with Group 4 in blue and Group 5 in pink. Annotations of the ^{13}C NQS data reflect whether the signal is observed (Obs) or suppressed (Red) whilst Y/N refer to the presence of a cross peak in the short-contact ^{1}H-^{13}C HETCOR spectrum where Groups 4 (blue dashed box) and 5 (pink dashed box) are identified. Inset is a zoomed in region focussing on the 50–54 ppm region of the partially overlapped signals, demonstrating the resolved ^{1}H chemical shifts that can be used in Stage 2.

To evaluate the ^{13}C NQS spectrum, the annotations in figure 4.3 refer to whether the environment has experienced a reduction in intensity relative to the ^{13}C CP/TOSS spectrum or not. To note, the ^{13}C NQS experiment did not fully suppress the aromatic CH signals relative to the known CH_2 signal at 34.7 ppm, as the suppression is disrupted by the mobility of the ring and has been documented previously.[60] This is why the 'red' annotation has been used rather than 'suppressed'. From figure 4.3, values of e_{NQS} can be assigned for each environment in each possible assignment and combined using equation 4.6.

The short-contact ^{1}H-^{13}C HETCOR experiment has been used for two purposes in this example. Firstly, in stage 1, it is used to determine whether a cross peak is present at a particular carbon signal and this is independent of the associated ^{1}H chemical shift. Once it is determined whether a cross peak is present or not, the appropriate e_{SCHET} value can be assigned for each environment in each possible assignment.

Using the information derived from the ^{13}C NQS and short-contact ^{1}H-^{13}C HETCOR spectra, equation 4.7 can be used and this resulted in the confident assignment of group 5 as the 90% threshold was achieved. The quantitative inclusion of the ^{13}C NQS data was pivotal as suppression was expected for the CH signal but not for the Me signal. Group 5 also demonstrates how the ^{13}C NQS spectrum and short-contact ^{1}H-^{13}C HETCOR spectrum can provide different information. In this example, the short-contact ^{1}H-^{13}C HETCOR data provided no additional insight in Stage 1 as both environments are protonated. There are other techniques that can be used to distinguish between different protonation states, such as 1D variants of ^{13}C MAS-J-HSQC experiment used by Cordova et al.[181] but also ^{13}C CPPI spectra, as discussed for piroxicam in section 4.5.4.[182]

The 90% AF threshold was not reached for Group 4 but filtering of the possible assignments could be used. As shown in figure 4.2, the majority of the possible assignments in Group 4 do not exceed the AF < 1% threshold. Therefore, they can be disregarded. The remaining possible assignments for consideration in Stage 2 consistently assigned **C4** to 121.9 ppm, therefore **C4** could confidently be assigned to 121.9 ppm and not included in Stage 2.

4.4.2 Workflow Implementation – Stage 2

The second purpose of the short-contact ^{1}H-^{13}C HETCOR experiment in this example was the determination of the ^{1}H chemical shifts. These were used in Stage 2. Their incorporation into the workflow reflects the ^{13}C chemical shifts, where the only variation is the choice of σ in equation 4.4. For the 6 possible assignments in Group 4, the AF threshold of 90% is achieved, therefore no additional experimental data was needed. At this point, the most probable assignment for Tyr has now been proposed.

4.5 Model Systems

To demonstrate different scenarios where the workflow can be used, a selection of model systems have been studied.

4.5.1 INH Co-crystals

Isoniazid (INH) is an antitubercular drug with poor stability in the solid state. This has led to interest in co-crystallisation of INH with other materials in order to improve its stability.[183,189,190] One set of coformers studied for co-crystallisation were cinnamic acid derivatives (2HCA, 3HCA and 4HCA), as detailed by Mashhadi et al.[183] A variety of different co-crystals were prepared and characterised with a selection of techniques, including SSNMR, however, due to significant overlap, there was no structural assignment, so this provided a great opportunity to apply the workflow.

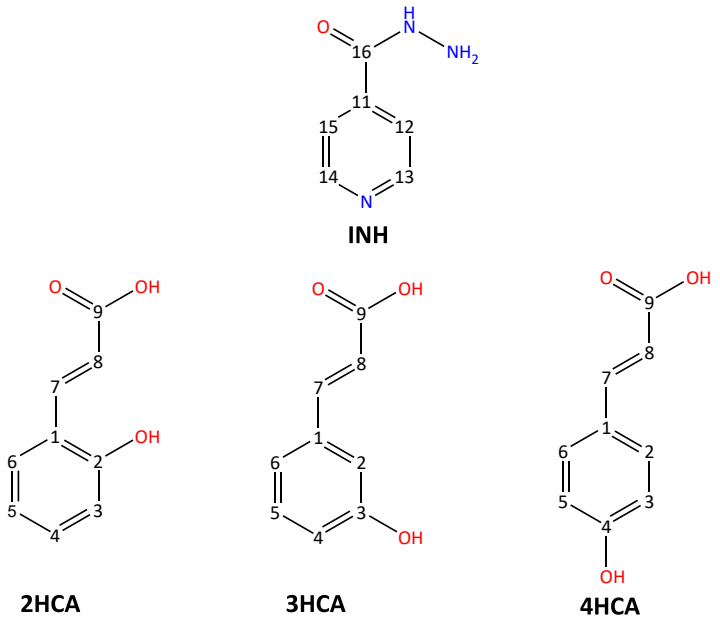


Figure 4.4: Labelled molecular structures of INH and the associated coformers 2HCA, 3HCA and 4HCA.

The experimental data considered were the ^{13}C and ^{1}H chemical shifts, ^{13}C NQS and both short-contact and long-contact ^{1}H-^{13}C HETCOR spectra. The choice of referencing, where *m* is set to −1 was appropriate for these systems due to clustering of both ^{1}H and ^{13}C chemical shifts.[191] For example, in INH-4HCA, the σ_{ref} used was 30.1 ppm but if *m* was allowed to deviate, the σ_{ref} decreased to 24.7 ppm, which can have a significant impact on the resulting calculated chemical shifts. Therefore, *m* of −1 was used.

Inclusion of Long Contact ^{1}H-^{13}C HETCOR Spectra

In order to include information from the long-contact ^{1}H-^{13}C HETCOR spectrum, the predicted C-H distances for the calculated cross peaks need to be considered. This is because of the $1/r^6$ relationship between the C-H distance (r) and cross peak intensity, where shorter C-H distances result in stronger cross peaks.[186,187] This is the distinction between the short-

contact and long-contact ^{1}H-^{13}C HETCOR experiments within the workflow. The short-contact variant informs on whether a particular environment is protonated or not (independent of ^{1}H chemical shift) as well as providing the ^{1}H chemical shifts where appropriate. In contrast, the long-contact variant takes into account both the ^{1}H chemical shift in the indirect dimension alongside the associated C-H distance in order to identify firstly, whether a cross peak is present, and second, to decide the appropriate e_{LCHET} value to use.

4.5.1.1 INH-2HCA

Beginning with INH-2HCA, there were 7 groupings, 4 of which have been considered in the workflow, as shown in table 4.3 and figure 4.5 below. The considered experimental data can be seen in figure 4.6.

Table 4.3: Proposed groupings of the experimental chemical shifts of INH-2HCA alongside the GIPAW calculated ^{13}C and ^{1}H chemical shifts for each environment.[8]

Grouping	Experimental ^{13}C (^{1}H) Chemical Shifts / ppm	Possible Assignments	Calculated ^{13}C(^{1}H) Chemical Shifts / ppm
1	172.6	C9	169.8
2	165.6	C16	162.8
3	156.3	C2	158.3
4	150.1(8.8) 146.5(8.7)	C13-H13, C14-H14	C13 – 152.0(8.8) C14 – 146.2(8.6)
5	139.8(8.2) 133.2(7.0)	C4-H4, C7-H7, C11	C7 – 142.4(8.5) C11 – 139.0 C4 – 135.2(6.8)
6	124.3(7.9) 120.7(8.4)	C1, C6-H6, C12-H12, C15-H15	C6 – 124.1(7.6) C12 – 123.7(8.1) C15 – 122.2(8.6) C1 – 120.7
7	119.0(6.3) 118.2(7.5) 117.2(7.5)	C3-H3A, C5-H5, C8-H8	C3 – 118.2(7.5) C5 – 117.9(5.8) C8 – 115.7(8.0)

[8] *The parameters used in the GIPAW calculations can be found in section 4.2.3. The σ_{ref} values used for ^{13}C and ^{1}H were 166.3 and 30.7 ppm, respectively.*

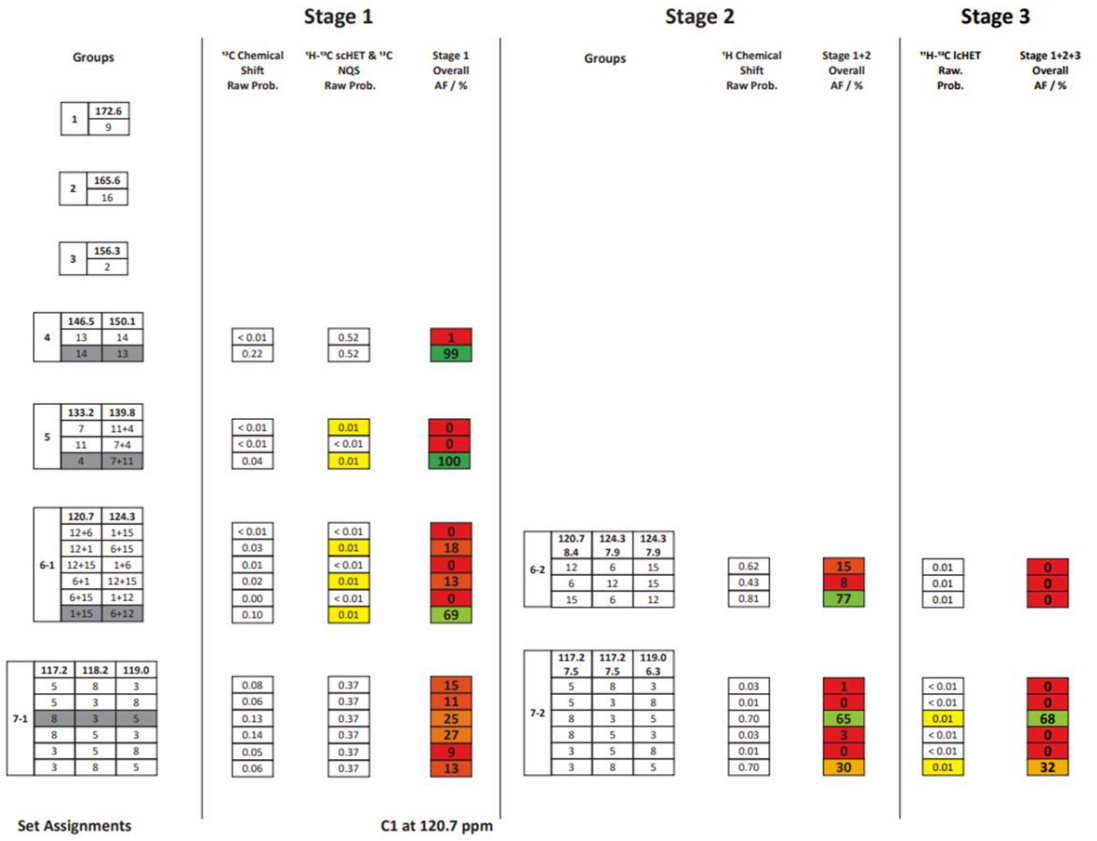


Figure 4.5: Graphical representation of the assignment workflow for INH-2HCA. For each group of resonances with similar ^{13}C chemical shifts, the possible assignments of carbon sites to ^{13}C chemical shifts (in ppm) are listed. At each stage, the probability of a given assignment is determined and from this, an overall colour-coded AF value is calculated. The relevant raw probability values used in the AF < 1% filter have been highlighted in yellow whilst the most probable assignment in each group is highlighted in grey.

At the end of Stage 1, Groups 4 and 5 saw AF values of over 90% so a confident assignment could be made. On the other hand, Groups 6 and 7 needed additional experimental verification. Group 6, in particular, was affected by overlap, as shown in the experimental data in figure 4.6. This resulted in raw probability values which appear to contradict one another for the 120.7 ppm signal. At 120.7 ppm, there is a signal in the ^{13}C NQS spectrum, which indicates a quaternary carbon, whilst the short-contact ^{1}H-^{13}C HETCOR spectrum displays a cross peak, indicative of a protonated carbon. This contradiction does provide key information that at 120.7 ppm, it is likely that there are at least two environments assigned to this signal and at least one of these environments will be quaternary. This was reinforced

further when the AF < 1% filter was applied as the number of possible assignments halved and resulted in the confident assignment of **C1** to 120.7 ppm, as shown in figure 4.5.

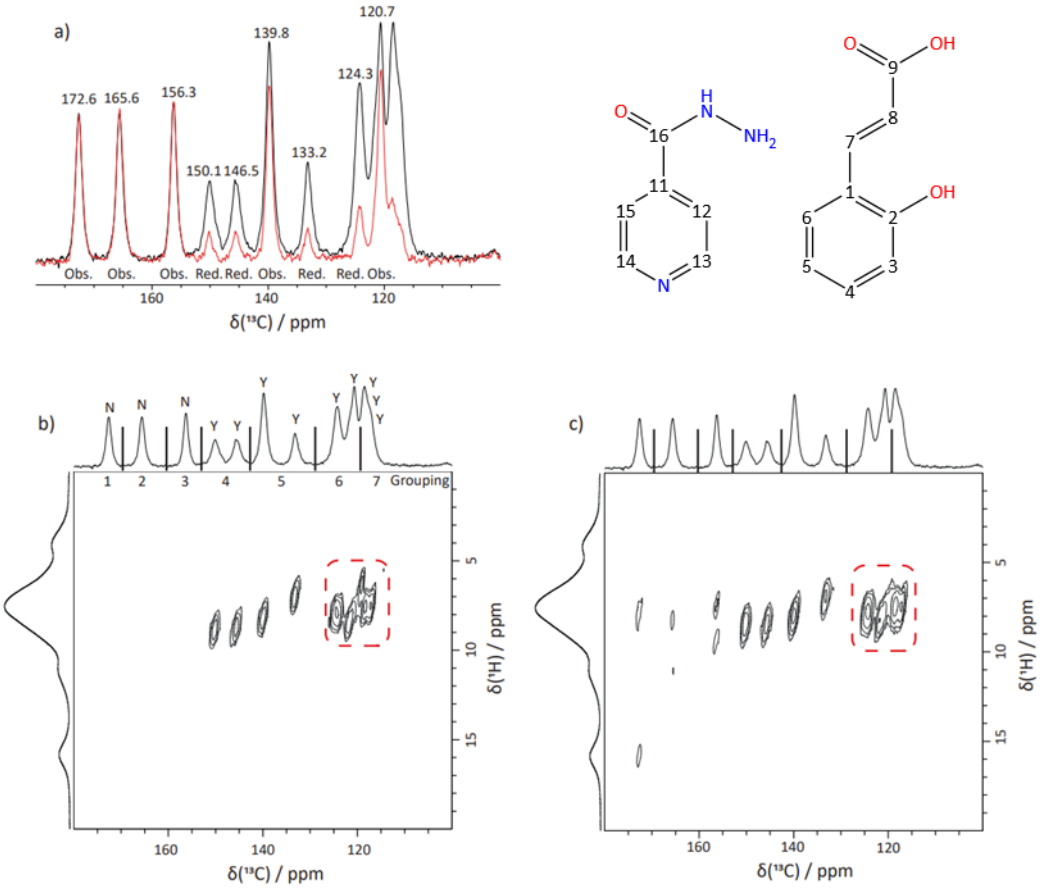


Figure 4.6: Experimental data for INH-2HCA with a) ^{13}C CP/TOSS and ^{13}C NQS, b) short-contact ^{1}H-^{13}C HETCOR and c) long-contact ^{1}H-^{13}C HETCOR spectra. Annotations of the ^{13}C NQS spectrum reflect whether the signal is observed (Obs) or suppressed (Red) whilst Y/N refer to the presence or not of a cross peak in the short-contact ^{1}H-^{13}C HETCOR spectrum. ^{1}H-^{13}C HETCOR spectra have dashed red boxes indicating the key cross peaks that influenced the resulting AFs after Stage 2 and 3. Groupings have been marked on the ^{1}H-^{13}C HETCOR spectra.

The reduction in the possible assignments for Group 6 was carried forward into Stage 2 alongside Group 7. However, neither grouping achieved the 90% threshold so Stage 3 was conducted with the additional consideration of a long-contact ^{1}H-^{13}C HETCOR spectrum. A long-contact ^{1}H-^{13}C HETCOR spectrum is shown in figure 4.6 where the dashed box focuses on the cross peaks associated with Groups 6 and 7. For each environment within each group, the predicted cross peaks were calculated. These were then used to determine whether a cross peak was present or not for a particular assignment. In order to determine whether a

predicted cross peak could be present, the related experimental ^{1}H chemical shift is used. Once it has been estimated whether a cross peak is present or not, the calculated C-H distance can be used to assign the e_{LCHET} value according to table 4.1. Unfortunately, the 90% threshold was not reached for both groups, as seen in figure 4.5, so a confident assignment cannot be specified but an assignment preference can be seen.

4.5.1.2 INH-3HCA

The workflow for INH-3HCA shown in figure 4.7 follows the same pattern as for INH-2HCA. It is a key example to demonstrate how important the quantitative consideration of additional data is. The critical grouping was Group 2, **C3** and **C16** in table 4.4, where at the end of Stage 1, there is clear preference towards one particular assignment but by the end of Stage 3, once the 90% threshold has been achieved so the most probable assignment flips.

Table 4.4: Proposed groupings of the experimental chemical shifts of INH-3HCA alongside the calculated ^{13}C and ^{1}H chemical shifts for each environment.[9]

Grouping	Experimental ^{13}C (^{1}H) Chemical Shifts / ppm	Possible Assignments	Calculated ^{13}C(^{1}H) Chemical Shifts / ppm
1	172.1	C9	170.5
2	162.7 159.3	C3, C16	C3 – 161.7 C16 – 160.0
3	149.4(8.3,8.0) 146.0(7.6)	C7-H7, C13-H13, C14-H14	C13 – 151.1(8.2) C7 – 150.9(8.1) C14 – 149.8(7.4)
4	137.9 136.1 132.1(7.6)	C1, C5-H5, C11	C1 – 136.4 C11 – 135.1 C5 – 132.3(7.7)
5	124.2(7.4) 122.4(7.1) 121.3(6.6)	C2-H2A, C12-H12, C15-H15	C12 – 123.5(7.1) C15 – 122.9(6.9) C2 – 121.4(6.5)
6	118.2(6.6) 113.5(7.8)	C4-H4, C6-H6, C8-H8	C8 – 118.9(7.2) C4 – 117.8(6.4) C6 – 110.7(7.9)

[9] *The parameters used in the GIPAW calculations can be found in section 4.2.3. The σ_{ref} values used for ^{13}C and ^{1}H were 166.8 and 30.3 ppm, respectively.*

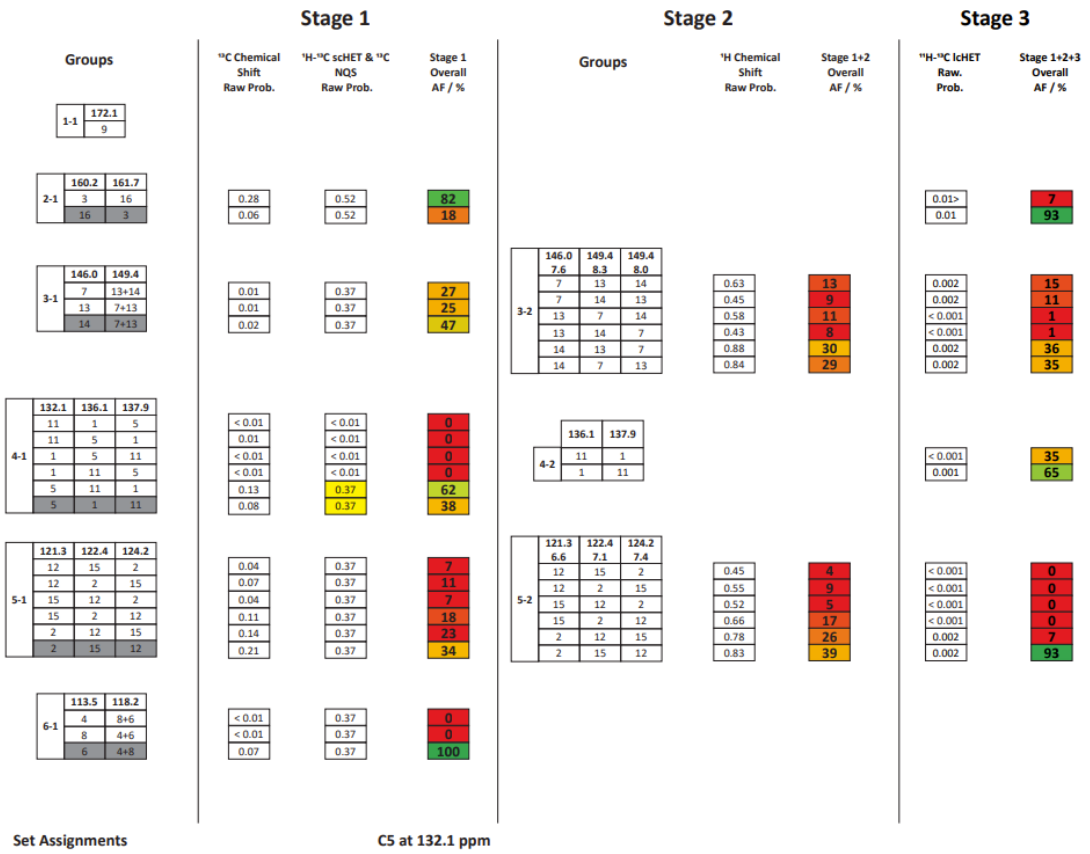


Figure 4.7: Graphical representation of the assignment workflow for INH-3HCA. For each group of resonances with similar ^{13}C chemical shifts, the possible assignments of carbon sites to ^{13}C chemical shifts (in ppm) are listed. At each stage, the probability of a given assignment is determined and from this, an overall colour-coded AF value is calculated. The relevant raw probability values used in the AF < 1% filter have been highlighted in yellow whilst the most probable assignment in each group is highlighted in grey.

The ^{1}H chemical shifts were determined from the short-contact ^{1}H-^{13}C HETCOR spectrum, where, as shown in figure 4.8 b), the 149.4 ppm signal in Group 3 has two distinct ^{1}H chemical shifts. This is reflected in the workflow where there are 3 possible assignments in Stage 1 but 6 possible assignments in Stage 2. Unfortunately, Stage 3 for Group 3 did not exceed the 90% AF threshold but this was largely due to the ^{1}H chemical shift assignment. If just the ^{13}C chemical shifts are considered, there is a clear preference for **C14** at 146 ppm and **C13** and **C7** at 149.4 ppm. The AF < 1% filter could be used on Group 4 to reduce the number of possible assignments from 6 to 2 where **C5** could be confidently assigned to 132.1 ppm. Stage 2 was not required, as both **C1** and **C11** are quaternary, but at the end of Stage 3, a preference is seen. Groups 5 and 6 were run as usual where the 90% threshold was met in Stage 3 and 1, respectively.

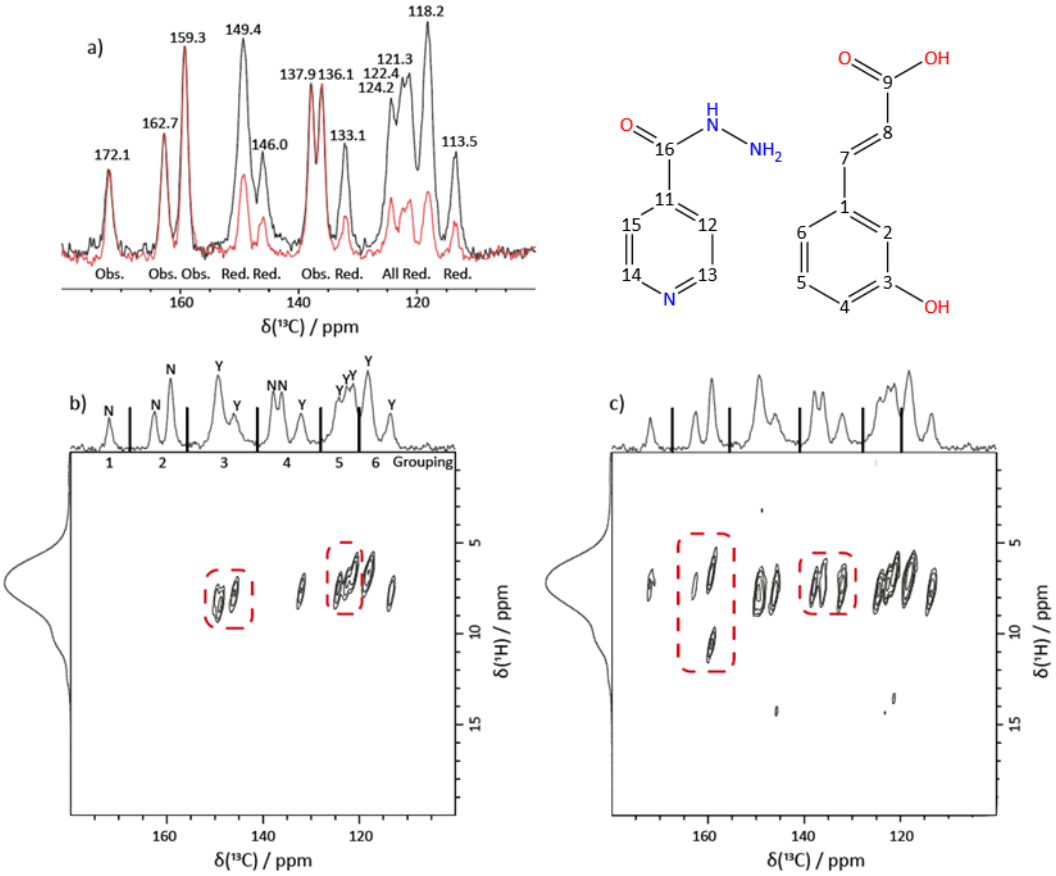


Figure 4.8: Experimental data for INH-3HCA with a) ^{13}C CP/TOSS and ^{13}C NQS, b) short-contact ^{1}H-^{13}C HETCOR and c) long-contact ^{1}H-^{13}C HETCOR spectra. Annotations of the ^{13}C NQS spectra reflect whether the signal is observed (Obs) or suppressed (Red) whilst Y/N refer to the presence or not of a cross peak in the short-contact ^{1}H-^{13}C HETCOR spectrum. ^{1}H-^{13}C HETCOR spectra have dashed red boxes indicating the key cross peaks that influenced the resulting AFs after Stage 2 and 3. Groupings have been marked on the ^{1}H-^{13}C HETCOR spectra.

4.5.1.3 INH-4HCA

5 groups were also evaluated for INH-4HCA, as shown in the workflow in figure 4.9, using the experimental data in figure 4.10. In similar fashion to INH-3HCA, Group 2 of C4 and C16 in table 4.5, saw a variation in assignment preference between Stage 1 and Stage 3, again highlighting the importance of including additional data.

Table 4.5: Proposed groupings of the experimental chemical shifts of INH-4HCA alongside the calculated ^{13}C and ^{1}H chemical shifts for each environment.[10]

Grouping	Experimental ^{13}C (^{1}H) Chemical Shifts / ppm	Possible Assignments	Calculated ^{13}C(^{1}H) Chemical Shifts / ppm
1	172.5	C9	170.8
2	161.7 160.2	C4, C16	C4 – 162.2 C16 – 158.6
3	149.8(7.5) 146.1(7.6)	C7-H7, C13-H13, C14-H14	C14 – 150.1(7.0) C7 – 149.2(7.9) C13 – 146.9(7.2)
4	140.6 135.6(7.3)	C2-H2A, C11	C11 – 139.8 C2 – 137.1(7.2)
5	125.5(8.1) 123.6(7.0,7.2)	C1, C6-H6, C12-H12, C15-H15	C6 – 125.6(8.5) C1 – 125.5 C12 – 12.2(7.0) C15 – 122.9(6.9)
6	119(7.0) 116.1(5.7) 114.9(6.4)	C3-H3A, C5-H5, C8-H8	C5 – 118.7(7.2) C3 – 115.7(5.5) C8 – 114.5(6.8)

[10] *The parameters used in the CASTEP-GIPAW calculations can be found in section 4.2.3. The σ_{ref} values used for ^{13}C and ^{1}H were 166.1, 30.0 ppm, respectively.*

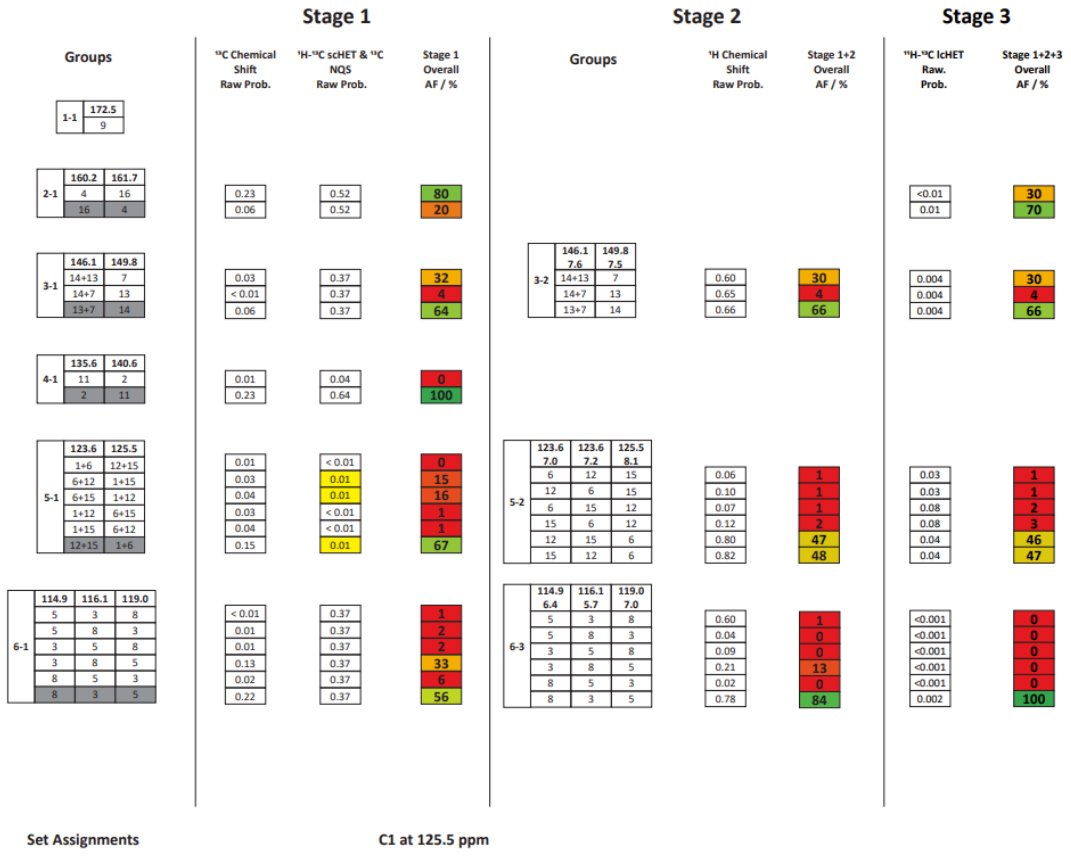


Figure 4.10: Graphical representation of assignment workflow for INH-4HCA. For each group of resonances with similar ^{13}C chemical shifts, the possible assignments of carbon sites to ^{13}C chemical shifts (in ppm) are listed. At each stage, the probability of a given assignment is determined and from this, an overall colour-coded AF value is calculated. The relevant raw probability values used in the AF < 1% filter have been highlighted in yellow whilst the most probable assignment in each group is highlighted in grey.

As with Group 2, Group 3 saw a preference towards an assignment at the end of Stage 3 whilst Group 4 achieved the 90% threshold at the end of Stage 1. The AF < 1% filter was used in Group 5 where **C1** could be set to 125.5 ppm and this reduced the number of carbon assignments from 6 to 3. However, as there were two distinct ^{1}H chemical shifts for the signal at 123.6 ppm, as marked with red crosses in figure 4.10, the number of possible ^{1}H assignments remained at 6. As with Group 3 in INH-3HCA, only a preference towards an assignment was seen for Group 5 at the end of Stage 3 but this again was controlled by the ^{1}H chemical shift assignment. If just the ^{13}C chemical shifts are considered, there is a clear preference for **C12** and **C15** at 123.6 ppm and **C6** at 125.5 ppm. The most probable assignment for Group 6 was achieved with certainty by the end of Stage 3.

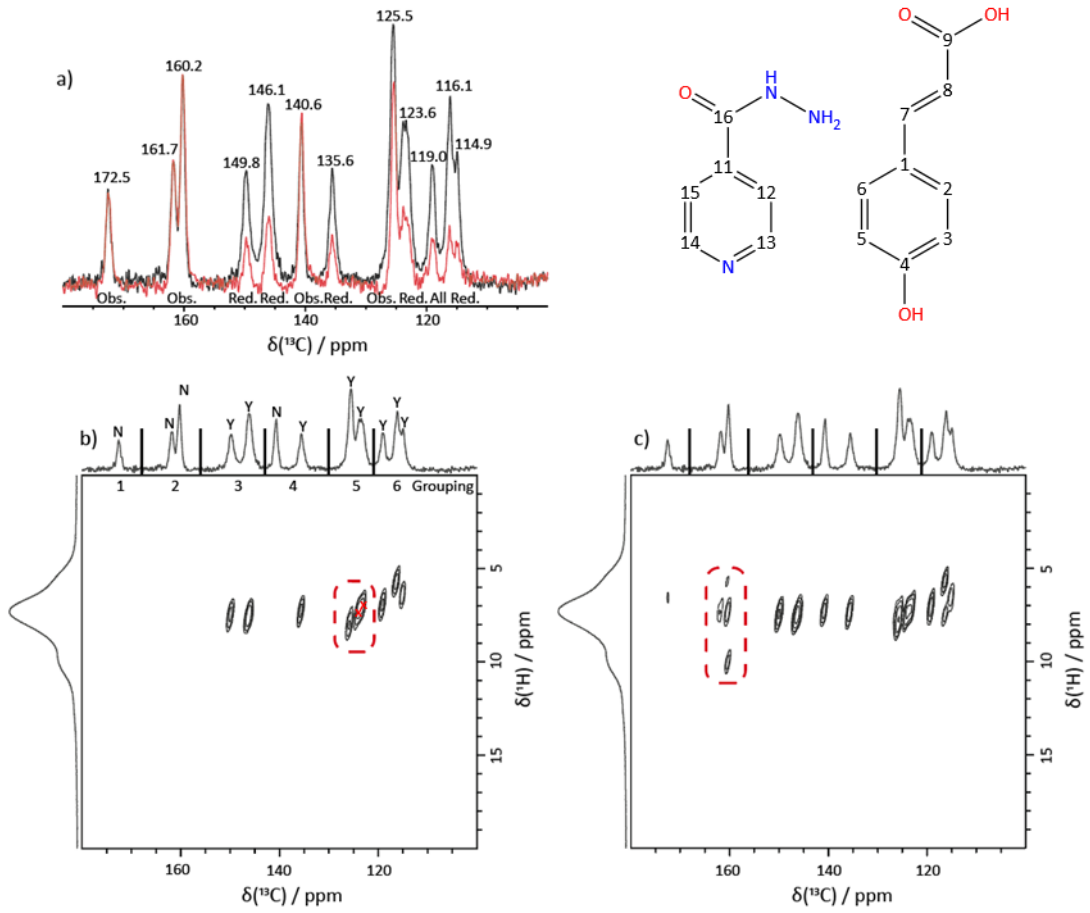


Figure 4.9: Experimental data for INH-4HCA with a) ^{13}C CP/TOSS and ^{13}C NQS, b) short-contact ^{1}H-^{13}C HETCOR and c) long-contact ^{1}H-^{13}C HETCOR spectra. Annotations of the ^{13}C NQS spectrum reflect whether the signal is observed (Obs) or suppressed (Red) whilst Y/N refer to the presence or not of a cross peak in the short-contact ^{1}H-^{13}C HETCOR spectrum. ^{1}H-^{13}C HETCOR spectra have dashed red boxes indicating the key cross peaks that influenced the resulting AFs after Stage 2 and 3. Groupings have been marked on the ^{1}H-^{13}C HETCOR spectra.

4.5.1.4 INH Co-crystal Summary

A summary of the proposed assignments for the co-crystals, as determined from the workflows, has been collated in table 4.6. For groupings where the 90% AF threshold has been achieved or exceeded, these have been put in bold.

Table 4.6: Proposed experimental assignment of the ^{13}C and ^{1}H chemical shifts for INH-2HCA, INH-3HCA and INH-4HCA alongside the calculated ^{13}C and ^{1}H RMSD values for the proposed assignments. Assignments which have exceeded the 90% threshold are in bold.

Grouping	INH-2HCA ^{13}C (^{1}H) / ppm	INH-3HCA ^{13}C (^{1}H) / ppm	INH-4HCA ^{13}C (^{1}H) / ppm
1	**C9 - 172.6**	**C9 - 172.1**	**C9 - 172.5**
2	**C16 – 165.6**	**C16 – 162.7** **C3 - 159.3**	C16 - 161.7 C4 - 160.2
3	**C2 – 156.3**	C13 - 149.4 (8.0) C7 – 149.4 (8.3) C14 - 146.0 (7.6)	C14 - 149.8 (7.5) C7- 146.1 (7.6) C13 - 146.1 (7.6)
4	**C13 – 150.1 (8.8)** **C14 – 146.5 (8.7)**	C11 – 137.9 C1 – 136.1	**C11 – 140.6** **C2 – 135.6 (7.3)**
5	**C7 – 139.8 (8.2)** **C11 – 139.8** **C4 – 133.2 (7.0)**	**C12 – 124.4 (7.4)** **C15 – 122.4 (7.1)** **C2 – 121.3 (6.6)**	**C1 – 125.5** C6 – 125.5 (8.1) C12– 123.6 (7.2) C15 – 123.6 (7.0)
6	C6 – 124.3 (7.9) C12 – 124.3 (7.9) **C1 – 120.7** C15 – 120.7 (8.4)	**C4 – 118.2 (6.6)** **C8 – 118.2 (6.6)** **C6 – 113.5 (7.8)**	**C5 – 119.0 (7.0)** **C3 – 116.1 (5.7)** **C8 – 114.9 (6.4)**
7	C5 – 119 (6.3) C3 – 118.2 (7.5) C8 – 117.2 (7.5)		
RMSD – ^{13}C	**1.66**	**1.89**	**1.43**
RMSD – ^{1}H	**0.28**	**0.28**	**0.32**

Across the three co-crystals, at least half of the considered groupings have exceeded the 90% threshold, i.e., where the workflow can be used to fully consider and incorporate overlap quantitatively. This was key to achieving the primary objective of determining the most probable ^{13}C assignment. Alongside this, the use of the filtering threshold (AF < 1%) reduced the number of possible assignments in all three workflows, thus allowing for particular assignments to be set, thus simplifying the subsequent stages.

Inclusion of additional data has helped avoid mis-assignment when just the ^{13}C chemical shifts are considered, as seen for Group 2 in both INH-3HCA and INH-4HCA. The workflow also highlights how each piece of experimental data can be used, but also when additional experimental data may be needed. For example, for Group 2 of INH-4HCA, a ^{13}C CPPI spectrum would provide no additional insight but a ^{13}C-^{13}C INADEQUATE spectrum could.

Overall, the workflow was able to confidently assign 10 of the 15 carbons in INH-2HCA, 9 of the 15 carbons in INH-3HCA and 10 of the 15 carbons in INH-4HCA. This number does reduce slightly for the ^{1}H assignments for INH-4HCA however, around ⅔ of the environments can be confidently determined where, within this number, overlapped resonances are included. In order to achieve the 90% AF threshold for all groupings, additional experiments would be needed.

4.5.2 Naproxen (NPX)

In the 2020 review by Hodgkinson, naproxen (NPX)[162–164] was highlighted as an example of structural mis-assignment within the literature.[10] The initial assignment for NPX was first published by Ando et al. but an alternative was proposed by Carignani and co-workers a year later.[162,163] Czernek in 2015 aimed to determine the correct assignment of NPX, concluding that the latter assignment was correct using a selection of qualitative and quantitative approaches.[164] The two assignments have been summarised in table 4.8.

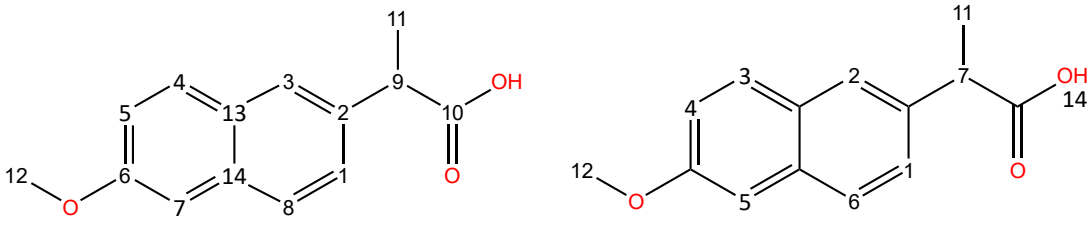


Figure 4.11: Molecular structure of NPX with carbon labels (left) and proton labels (right) that follow the numbering of reference 164.

It was of interest as to whether the proposed workflow could identify the more probable assignment as the RMSD for ^{13}C could not. There are four environments (**C13**, **C8**, **C4** and **C3**) which vary between the assignments and as they are similar in chemical shift, it is not surprising that the RMSD values cannot distinguish between the two proposed assignments.

However, this does allow the environments to be grouped together in one group and allows the workflow to be tailored to just focus on groupings where ambiguity remains.

Table 4.8: Proposed experimental assignments of NPX from the publications of Ando et al.[162] *and Carignani et al.*[163] *alongside the GIPAW calculated chemical shifts are in brackets.*[11] *The environments of interest have been coloured red.*

Environment	Ando et al. Experimental (Calculated) Chemical Shifts / ppm		Carignani et al. Experimental (Calculated) Chemical Shifts / ppm	
	^{13}C	^{1}H	^{13}C	^{1}H
C10	179.2 (177.8)		179.0 (178.5)	
C6	157.8 (161.5)		158.1 (162.2)	
C14	134.5 (134.4)		134.9 (135.1)	
C2	134.5 (138.4)		134.9 (139.1)	
C3-H2	130.1 (130.0)	4.2 (6.8)	129.1 (130.7)	6.1 (6.2)
C4-H3	129.3 (131.6)	6.2 (3.8)	130.6 (132.3)	3.8 (3.1)
C8-H6	128.5 (132.0)	6.7 (6.1)	129.1 (132.6)	5.9 (5.4)
C13	128.5 (129.0)		134.9 (129.7)	
C1-H1	123.7 (124.3)	7.6 (8.2)	124.0 (125.0)	7.0 (7.6)
C5-H4	118.9 (120.7)	5.3 (5.1)	119.2 (121.4)	4.5 (4.5)
C7-H5	104.1 (102.6)	4.8 (4.7)	104.3 (103.2)	4.1 (4.1)
C12-H12	53.1 (49.1)	3.0 (2.9)	53.2 (49.8)	2.3 (2.3)
C9-H7	46.8 (45.7)	4.2 (4.3)	47.0 (46.4)	3.2 (3.7)
C11-H11	17.1 (9.1)	2.4 (2.4)	17.5 (9.8)	1.8 (1.7)
RMSD	3.1	1.2	3.4	0.4

Experimental Data

From the paper by Ando et al., alongside the ^{13}C and ^{1}H chemical shifts, conventional short-contact ^{1}H-^{13}C HETCOR experiment and dipolar 1D ^{13}C Insensitive Nuclei Enhanced by

[11] *The CSD deposit,* ***COYRUD12*** *was used as the input .cif file for the calculations. The* σ_{ref} *values used for* ^{13}C *and* ^{1}H *were 166.8, 30.6 ppm and 167.5, 30.0 ppm for Ando et al. and Carignani et al. respectively.*

Polarisation Transfer (INEPT) measurements were used.[162] In contrast, Carignani and co-workers used short-contact and long-contact ^{1}H-^{13}C HETCOR spectra.[163]

In order to consider the 1D ^{13}C dipolar INEPT data, a new set *of* $e_{\lambda,i}$ values were needed based on the information the experiment can provide. From the Ando et al. paper, the ^{13}C dipolar INEPT data was used to determine protonation states of environments,[162] as seen for the ^{13}C NQS experiment. Therefore, it was deemed appropriate that the value of e_{NQS} (0.8) was used. The quoted percentages for the ratio between the signals at two different delay (τ) times were 69%, 6%, 136%, and 217%-303% for CH, CH_2, CH_3, and C groups, respectively.[162] In order to account for variation from these ratios, appropriate ranges for the e_{INEPT} were proposed in table 4.8.

Table 4.8: Summary of e_{INEPT} *values used for the dipolar INEPT, where the rules are loosely based on the Wickramasinghe and Ishii ratios.*[192]

Wickramasinghe and Ishii Ratios	e_{INEPT} (=0.8)
Ratio below 30% where the environment is CH_2	e_{INEPT}
Ratio between 30-85% where the environment is CH	
Ratio between 85-175% where the environment is CH_3	
Ratio over 175% where the environment is C	
Ratio below 30% where the environment is not CH_2	$1 - e_{INEPT}$
Ratio between 30-85% where the environment is not CH	
Ratio between 85-175% where the environment is not CH_3	
Ratio over 175% where the environment is not C	

Workflow

The workflows for the two groupings are shown in figure 4.12 where the proposed assignment from each paper is highlighted in grey.[162,163] The filtering threshold of AF < 1% was not possible therefore the magnitude of the raw probabilities in the ^{13}C NQS + ^{1}H-^{13}C scHET column in figure 4.12 could be used instead to reduce the number of assignments considered in Stage 2. In the Carignani et al. assignment, the ^{13}C NQS + ^{1}H-^{13}C scHET column either had a raw probability of 0.01 or 0.66 so those with a value of 0.01 could be discarded. This then allowed **C13** to be assigned to 134.9 ppm. A similar method was used for the Ando

et al. assignment where assignments which fell below the raw probability value of 0.001 were discarded. This left 6 possible assignments where **C13** could also be assigned to 128.5 ppm.

The assignment of **C13** does vary considerably between the two experimental datasets. However, identification of **C13** from the dipolar ^{13}C INEPT ratios is ambiguous as a ratio of 104% was seen for 128.5 ppm. The expected ratio should be between 217-303%.[162,192] It could be argued that overlap affected the ratio but this has not been discussed by the authors so this assignment can be considered sceptical.

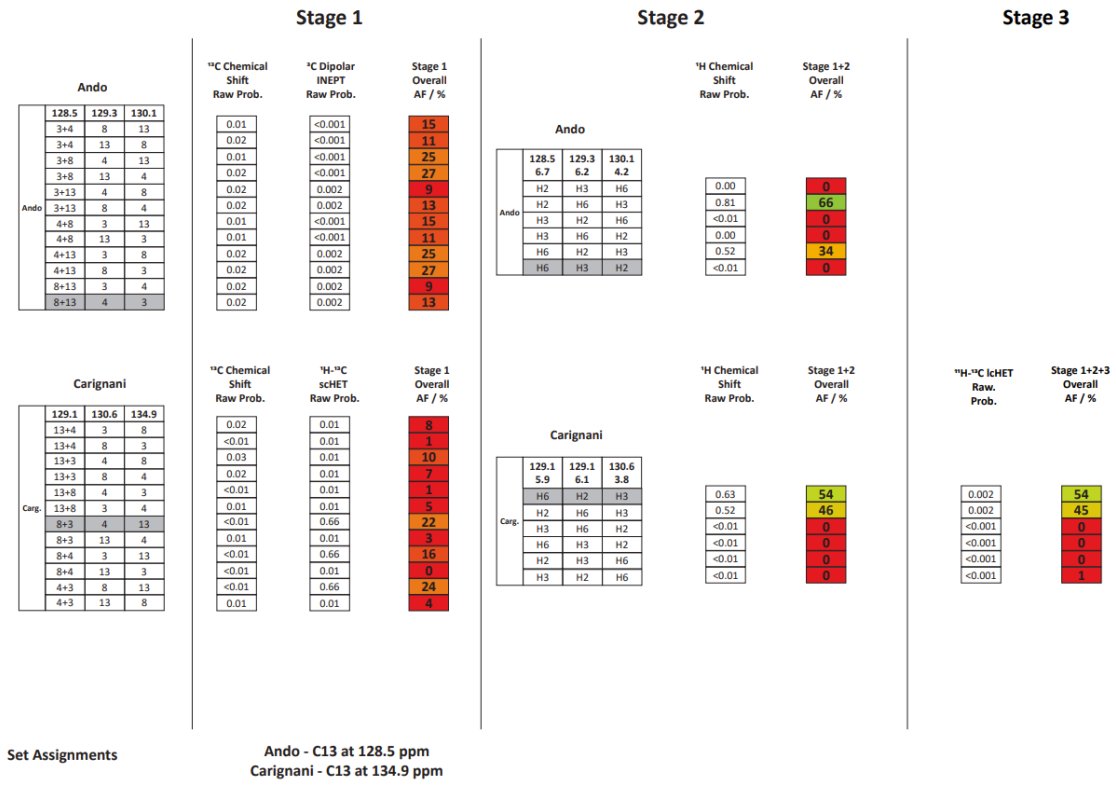


Figure 4.12: Graphical representation of assignment workflow for NPX where the assignment in grey is the proposed assignment from the publications.[162,163] *For the group, the possible assignments of carbon sites to ^{13}C chemical shifts (in ppm) are listed. At each stage, the probability of a given assignment is determined and from this, an overall colour-coded AF value is calculated. After Stage 1, a different filter could be applied based on the raw probabilities in the ^{13}C NQS + short-contact ^{1}H-^{13}C HETCOR column, as discussed in-text.*

In Stage 2 of the workflow, it becomes apparent that the proposed Ando et al. assignment is not probable. If the filtering threshold of AF < 1% is applied at the end of Stage 2, the published assignment would actually be discarded. In contrast, at the end of Stage 3 for the Carignani et al. assignment, the proposed experimental assignment of both carbons and

protons is the most probable. Interestingly, the most probable carbon assignment seen in figure 4.12 from the Ando et al. workflow was more reflective of the Carignani et al. assignment, particularly for **C4-H3** at 130 ppm.

The workflow clearly demonstrated that the published experimental assignment proposed by Ando et al. is incorrect. Alongside this, the proposed assignment by Carignani et al. was calculated as the most probable. Even though the 90% threshold was not reached, this was largely due to the ^{1}H chemical shift assignment with the overlapped signal of 129.1 ppm.

4.5.3 6-APA

Another mis-assignment within the literature was for 6-APA, where only the ^{13}C chemical shifts have been published.[165–167] The initial assignment by Clayden et al. was guided by the corresponding solution-state chemical shift assignment but was contradicted by Aguiar et al. where, from CP build-up rates, the assignment of two sets of signals were switched. These were **C9/C10** and **C3/C5**.[165,166] Over 20 years later, D. de Aguiar et al. used the NMR crystallographic approach to reinforce the initial assignment by Clayden et al.[167] The experimental assignment proposed by each publication is tabulated in table 4.9 alongside the exemplar workflow, based on the most recent assignment that is shown in figure 4.13.[167]

Table 4.9: Proposed experimental assignments of 6-APA from the related publications[165–167] *alongside the calculated ^{13}C chemical shifts in brackets.*[12] ***C2** has been highlighted in red alongside the associated RMSD values.*

Environment	Clayden et al.	M. Aguiar et al.	D. de Aguiar et al.
	Experimental (Calculated) Chemical Shifts / ppm		
C11	174.3 (175.3)	174.7 (175.5)	174.1 (175.0)
C7	167.1 (167.7)	167.3 (167.9)	166.7 (167.4)
C3	73.7 (75.5)	66.3 (75.7)	73.3 (75.2)
C2	67 (75.7)	67.3 (75.9)	66.8 (75.4)
C5	66.2 (70.2)	73.8 (70.4)	65.7 (69.8)
C6	59 (56.9)	59.2 (57.1)	58.7 (56.5)
C9	36.9 (35.5)	28.3 (35.8)	36.6 (35.2)
C10	27.9 (24.0)	37.2 (24.3)	27.6 (23.7)
RMSD	3.7	7.0	3.7
RMSD – C2 omitted	2.4	6.9	2.5

[12] *The CSD deposit, **AMPEN01** was used as the input .cif file for the calculations.*[165–167] *The σ_{ref} values used were 169.0 ppm (Clayden), 169.2 ppm (M. Aguiar) and 168.6 ppm (D. de Aguiar) respectively.*

6-APA is a great example to demonstrate the strength of the workflow in recognising when there is insufficient data available and also how sole reliance on chemical shift can skew interpretation. From qualitative comparison between the experimental and calculated ^{13}C chemical shifts, **C9** and **C10** could be assigned. Therefore, the workflow focussed on the mis-assignment of **C3/C5** where there is a severe discrepancy between experimental and calculated chemical shifts for **C2**. This then had a significant effect on the resulting AF values. One explanation for the 4-5 ppm deviation for **C2** is the effects of the neighbouring sulfur atom on the calculation but to verify this, specialist calculations would be needed.[193–195]

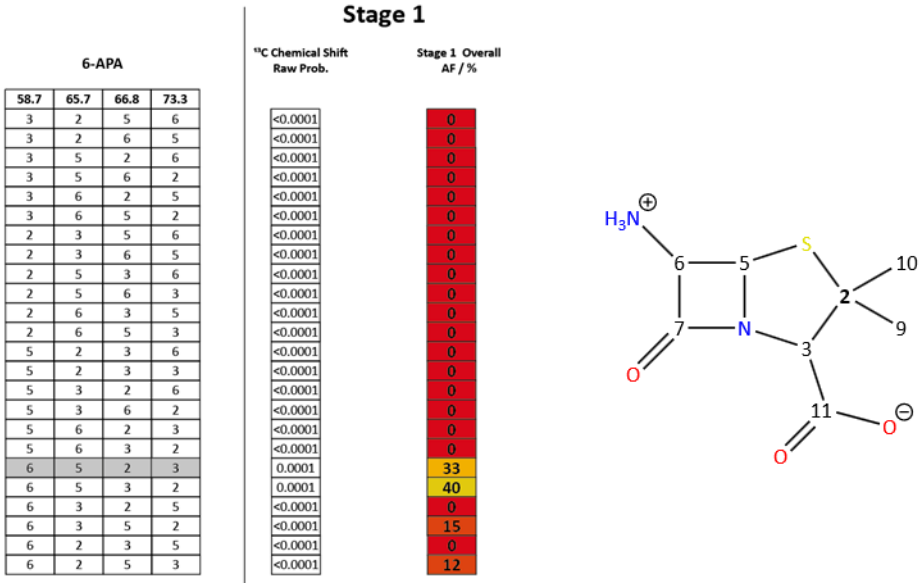


Figure 4.13: Graphical representation of assignment workflow for 6-APA where the most probable assignment for Group 2 is in grey. The labelled molecular structure of 6-APA is also included. For each group of resonances with similar ^{13}C chemical shifts, the possible assignments of carbon sites to ^{13}C chemical shifts (in ppm) are listed. As only the ^{13}C chemical shifts were available, these are the sole contributor to the AF values.

As the ^{13}C chemical shifts are the only piece of experimental data considered, the dramatic skew by **C2** is observed in both initial RMSD values in table 4.6 and in the workflow. Once the AF < 1% filter threshold has been applied, this leaves 4 possible assignments, where the only confident assignment is **C6** at 58.7 ppm. Because of this, 6-APA demonstrates one of the strengths of the workflow whereby it can identify when additional experimental information is required and, in some respects, indicate which type of experiments should be acquired. However, if acquisition of additional data is not possible, the workflow can indicate a preference towards a particular assignment. In this instance, there is a clear preference to assign **C6** to 58.7 ppm. Alongside this, the assignment of **C5** to 65.7 ppm is also indicated as

the top two assignments (in relation to AF percentages) assign **C5** at 65.7 ppm. However, as the 90% threshold has not been achieved, it only indicates a preferential assignment. However, relative to the mis-assignment within the literature, both contested pairs have been confidently assigned.

4.5.4 Piroxicam

The final example discussed is piroxicam as it is a material that was also considered by the OPTICS approach.[182] In the OPTICS paper, it is stated that the correct form of a material could be identified solely from the ^{13}C chemical shifts.[182] Therefore, it was of interest as to whether the proposed workflow could achieve a similar feat where, from one set of experimental data, the correct form of piroxicam could be identified.

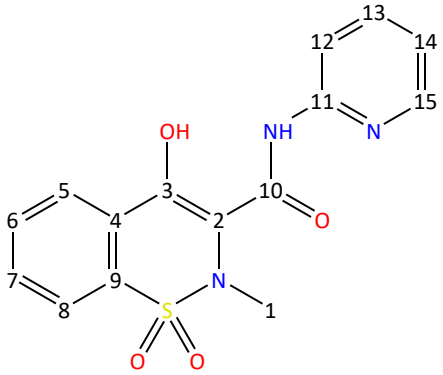


Figure 4.14: Labelled molecular structure for piroxicam

Standard Workflow with AF_{global}

In the OPTICS paper, four forms of piroxicam (Form I, II, III and IV) were compared to one set of assigned, experimental data. The proposed experimental assignment from reference 182 is presented in table 4.10. From reference 182, the data considered was the ^{13}C chemical shifts, ^{13}C CPPI and short-contact ^{1}H-^{13}C HETCOR spectra for piroxicam Form I. As the ^{13}C CPPI was used to determine the protonation state of the signals, as seen previously for the ^{13}C NQS and ^{13}C dipolar INEPT, the value of e_{CPPI} used was 0.8.

Table 4.10: Proposed experimental assignments of piroxicam from the OPTICS paper alongside the GIPAW calculated ^{13}C chemical shifts in brackets for each of four forms under consideration.[13]

Experimental ^{13}C Chemical Shift / ppm	Environment	Calculated ^{13}C Chemical Shift / ppm			
		Form I	Form II	Form III	Form IV
167.1	C10	164.0	163.6	164.1	162.6
158.7	C3	160.1	159.0	164.1	160.7
150.6	C11	149.8	148.2	149.6	154.7
147.4	C15	150.2	151.7	152.1	153.0
135.9	C13	137.2	137.2	139.0	138.8
135.9	C6	137.1	132.4	132.6	134.4
133.9	C9	133.5	136.2	138.2	132.9
132.3	C7	134.0	131.6	134.4	130.7
127.9	C4	127.8	130.0	127.5	125.1
126.8	C5	127.7	125.3	128.2	128.3
124.1	C8	125.3	125.9	125.2	123.4
119.6	C14	120.0	121.4	119.2	120.7
113.8	C12	114.2	113.3	110.9	110.8
112.4	C2	111.8	113.9	110.7	113.5
40.1	C1	33.7	36.8	30.7	36.9

In contrast to OPTICS, the proposed workflow does not 'know' the correct experimental assignment and this is reflected in table 4.11. When the assignment is allowed to deviate,

[13] *The CSD deposits,* ***BIYSEH13*** *(form I),* ***BIYSEH05*** *(form II),* ***BIYSEH07*** *(form III) and* ***BIYSEH08*** *(form IV) were used as the input .cif files for the calculations The ^{13}C and ^{1}H σ_{ref} values used were 165.0, 30.2 ppm (BIYSEH13), 166.2, 30.4 ppm (BIYSEH05), 165.2, 30.2 ppm (BIYSEH07) and 167.8, 30.4 ppm (BIYSEH08), respectively.*

there is variation in the signal assignment in Groups 4 and 5. However, when compared to the published experimental assignment, only Forms I and IV saw the ordering of the chemical shifts match.[182]

Table 4.11: The most probable assignments calculated for each of the solid forms with their associated AF (%). Those which exceeded the 90% AF threshold are in bold.

Group	Exp. Chemical Shift / ppm	Form I		Form II		Form III		Form IV	
		Env.	AF / %	Env.	AF / %	Env.	AF / %	Env.	AF / %
1	167.1	**C10**	**100**	**C10**	**100**	**C10**	**100**	**C10**	**100**
2	158.7	**C3**	**100**	**C3**	**100**	**C3**	**100**	**C3**	**100**
3	150.6 147.4	**C11** **C15**	**100**	**C11** **C15**	**99**	**C11** **C15**	**100**	**C11** **C15**	**100**
4	135.9 135.9 133.9 132.3	C13 C6 C9 C7	75	C13 C6 C9 C7	39	C13 C7 C9 C6	58	C13 C6 C9 C7	86
5	127.9 126.8 124.1	C4 C5 C8	76	C4 C8 C5	61	C4 C5 C8	78	C4 C5 C8	77
6	119.6	**C14**	**100**	**C14**	**100**	**C14**	**100**	**C14**	**100**
7	113.8 112.4	**C12** **C2**	**100**	**C12** **C2**	**100**	**C12** **C2**	**100**	**C12** **C2**	**100**
8	40.1	**C1**	**100**	**C1**	**100**	**C1**	**100**	**C1**	**100**
AF_{global} / %		87		13		0		0	

However, the conventional workflow focuses on local regions rather than the global assignment. Therefore, in order to distinguish which form was the most probable, the relative AF percentage for the global assignment (AF_{global}) needed to be defined. To calculate AF_{global}, the raw probability for the most probable assignment in each of the 8 groupings was selected and subsequently combined to create a global cumulative raw probability value for each form. The AF_{global} value for each form was then calculated, as seen previously in equation 4.5, where form I, at 87%, is the most probable. This is, however, smaller than the OPTICS percentage of 99.56%.

Consideration of ^{13}C Chemical Shifts Only

OPTICS focuses solely on the ^{13}C chemical shifts and therefore, this was attempted in table 4.12. The resulting AF and $AF_{\text{global I}}$ percentages saw a severe deviation from table 4.11, where Form II is identified as the most probable form despite the experimental data being from Form I. This swing in preference was suspected to be a consequence of allowing the assignment to vary. In OPTICS, it is assumed that the assignment is known whilst in the proposed workflow here, the assignment is assumed to not be known. Therefore, an additional workflow was attempted.

Table 4.12: The most probable assignment for each solid form with their associated AF (%) when only the ^{13}C chemical shifts are considered. Environments in bold are those which deviate from table 4.7.

Group	Exp. Chemical Shift / ppm	Form I		Form II		Form III		Form IV	
		Env.	AF / %	Env.	AF / %	Env.	AF / %	Env.	AF / %
1	167.1	C10	100	C10	100	C10	100	C10	100
2	158.7	C3	100	C3	100	C3	100	C3	100
3	150.6 147.4	**C15** **C11**	59	**C15** **C11**	93	**C15** **C11**	82	C11 C15	70
4	135.9 135.9 133.9 132.3	C13 C6 **C7** **C9**	37	C13 **C9** **C6** C7	41	C13 **C9** **C7** C6	51	C13 C6 C9 C7	38
5	127.9 126.8 124.1	C4 C5 C8	39	C4 C8 C5	45	**C5** **C4** C8	44	**C5** **C4** C8	52
6	119.6	C14	100	C14	100	C14	100	C14	100
7	113.8 112.4	C12 C2	74	**C2** **C12**	59	C12 C2	51	**C2** **C12**	73
8	40.1	C1	100	C1	100	C1	100	C1	100
AF_{global} / %		10		90		0		0	

Experimental Assignment is Set

Assuming that the experimental assignment proposed is certain,[182] the workflow was attempted twice again, as shown in table 4.13. The AF values in table 4.13 correspond to when just the ^{13}C chemical shifts are considered whilst the values in brackets are for when additional data is considered. When the experimental assignment is set, the resulting AF_{global} values are more reflective of those published previously,[182] where the 90% threshold has been exceeded with a preference for Form I in both adaptations. Coincidentally, both AF_{global} values were the same for both iterations.

Table 4.14: The most probable assignments calculated for each of the solid forms with their associated AF (%) when only the ^{13}C chemical shifts (or all additional data) are considered. The AF_{global} values have also been included where the values calculated when all additional data is considered are in brackets.

Group	Experimental Chemical Shift / ppm	Environment	Form I AF / %	Form II AF / %	Form III AF / %	Form IV AF / %
1	167.1	C10	100 (100)			
2	158.7	C3	100 (100)			
3	150.6 147.4	C11 C15	41 (100)	7 (99)	18 (100)	70 (100)
4	135.9 135.9 133.9 132.3	C13 C6 C9 C7	29 (75)	6 (39)	2 (12)	38 (86)
5	127.9 126.8 124.1	C4 C5 C8	39 (76)	28 (37)	33 (78)	21 (77)
6	119.6	C14	100 (100)			
7	113.8 112.4	C12 C2	74 (100)	41 (100)	51 (100)	27 (100)
8	40.1	C1	100 (100)			
AF_{global} / %			**92**	**8**	**0**	**0**

However, this demonstrates the influence that an assumed experimental assignment can have on the resulting AF values. This can possibly be a limitation of OPTICS because to achieve such high values (99.56% for piroxicam form I), there is an intrinsic bias that the proposed experimental assignment is correct. The workflow in this chapter aims to avoid this and answer a different question, where the assignment is not known but the form of the material is assumed correct.

4.5.5 Section Summary

Overall, in this section, 6 model systems have been used to demonstrate the applicability of the workflow to a variety of different assignment questions. It is believed that the precedent of the workflow has been established whereby key insight can be provided to aid assignment.

4.6 Conclusions and Future Work

The aim of this chapter was to evaluate the scope and precedent of a Bayesian probability based workflow to aid robust structural characterisation. Tyr has demonstrated how overlap can be incorporated quantitatively into the probability calculation and this was reiterated further in the INH co-crystals. For NPX and 6-APA, as the workflow focuses primarily on local regions of the spectrum, it allows for tailored evaluation of ambiguous regions within an assignment. Particularly for 6-APA, an additional strength of the workflow has been recognised whereby it can identify when (and the type of) additional data that is needed. For piroxicam, the workflow can be adapted in order to consider the global assignment if desired but this case also highlighted how the objective of the workflow is different from the previously published approaches. Rather than to assume that the experimental assignment is known, this workflow assumes that the material (and form) is pure and is known, therefore the calculated chemical shifts are reflective of the material.

Critically, the precedent of the workflow has been demonstrated in relation to aiding assignment, therefore, future work will focus on automation of the workflow in order to incorporate into a software package. This will allow for expansion of the current workflow, particularly the size of groupings, which will allow for larger, more complex materials to be studied.

Chapter 5: Exploring The Use of Dipolar-Based Pulse Sequences With ^{13}C-^{13}C INADEQUATE & ^{1}H-^{13}C DQ/SQ

5.1 Introduction

As discussed in the previous chapter, multidimensional solid state NMR plays a key role in the robust characterisation of molecular organic solids.[10,12,38] The additional dimension in both homonuclear and heteronuclear experiments can distinguish between environments that may be overlapped or difficult to assign from calculated chemical shifts alone. One example is the ^{1}H-^{13}C HETCOR experiment which provides insight into C-H correlations through variation of the contact time.[44,51,52] Shorter contact times can restrict the cross peaks observed to those associated with direct C-H pairs (protons which are directly bonded to the associated carbon) whilst longer contact times allow for longer-range C-H correlations to be seen. The longer-range correlations tend to result in cross peaks being observed for a proton that is not always directly bonded to the carbon associated with the ^{13}C chemical shift.[186] This then provides additional structural insight. *J*-mediated alternatives, such as the ^{1}H-^{13}C INEPT[36], are also used to investigate C-H correlations but tend to be limited to direct C-H pairs.[10,12]

An alternate probe of connectivity is through multi-quantum coherences with double quantum (DQ) / single quantum (SQ) experiments. The homonuclear, *J*-based ^{13}C-^{13}C INADEQUATE experiment is one of the best to investigate connectivity but the low natural abundance of ^{13}C (1.1%) does result in data acquisition spanning several days.[56,60,196,197] There is also a reliance on optimal acquisition parameters, where materials with small values of T_1 are the ideal candidates, so the incorporation of such experiments into a standard structural workflow is unfeasible. Despite these limitations, the wealth of information provided by the *J*-based ^{13}C-^{13}C INADEQUATE is unrivalled. This is especially important for more complex materials, such as $Z' > 1$ materials, where signal overlap and similarities in calculated chemical shifts make robust assignment difficult.[198,199]

A recent alternative to the *J*-based ^{13}C-^{13}C INADEQUATE is a dipolar-based variant of the ^{13}C-^{13}C INADEQUATE, published by Dekhil et al. in 2016.[60] The authors describe one of the benefits as a reduction in acquisition time which increases the range of materials that can be studied.[60] The build-up of magnetisation transfer is faster in dipolar-mediated experiments (and corresponding faster dephasing during recoupling) which, when optimised, results in greater signal intensities within the same time frame, as demonstrated in figure 5.1 for L-tyrosine ethyl ester.[60,200] Despite this clear improvement, the use of the dipolar-based ^{13}C-^{13}C INADEQUATE has been limited.[201–203]

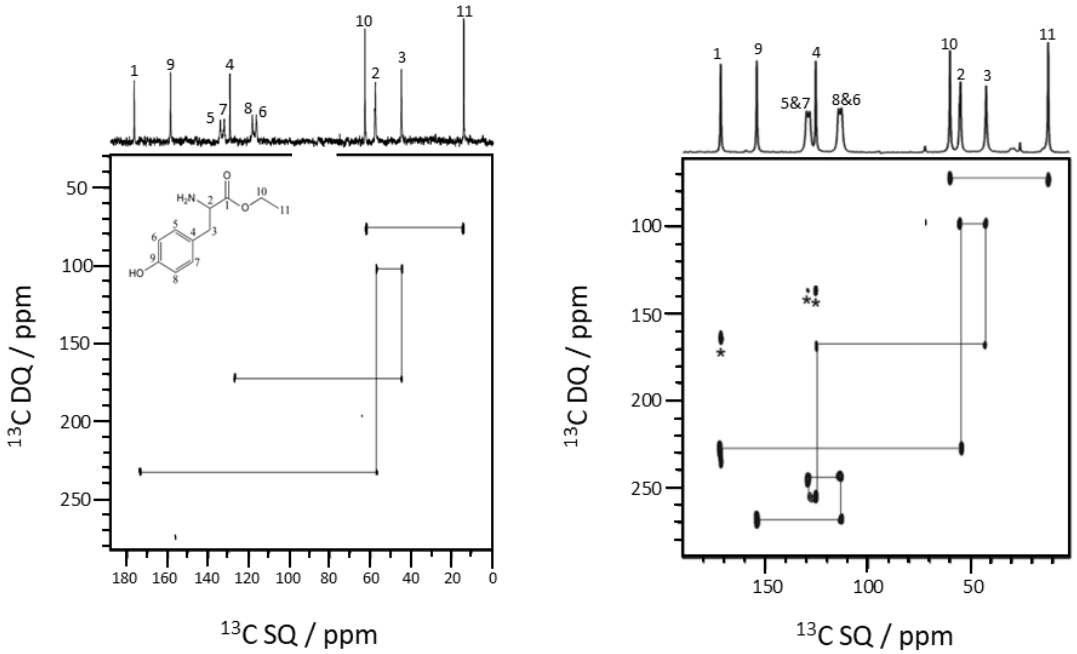


Figure 5.1: Comparison between the J-based ^{13}C-^{13}C INADEQUATE (left) acquired by De Paëpe et al.[200] and dipolar-based ^{13}C-^{13}C INADEQUATE (right) acquired by Dekhil et al.[60] of L-tyrosine ethyl ester at natural abundance and ambient temperature. Both spectra have been adapted from the relevant publications and were acquired in approximately 3 days.[60,200]

An alternate set of homonuclear experiments probe ^{1}H-^{1}H correlations where ^{1}H-^{1}H DQ/SQ has been used to aid structural assignment and also identify hydrogen bonding motifs in materials such as co-crystals.[10,131,204,205] However, due to the low resolution associated with ^{1}H signals, very fast MAS is required to tackle the broad lineshapes but with its narrow chemical shift range, overlap can persist.[186,206,207]

As ^{1}H-^{1}H DQ/SQ suffers from broad lineshapes and ^{13}C-^{13}C DQ/SQ is affected by low natural abundance, the ^{1}H-^{13}C DQ/SQ is a possible alternative to balance resolution and sensitivity as the desired information is captured from both nuclei in a single spectrum. In the acquired

data, the direct dimension of ^{13}C SQ coherences are correlated with the indirect dimension of ^{1}H DQ coherences, where the presence of a ^{1}H DQ cross peak indicates that two protons are close to one another in space. The ^{1}H DQ correlations are then linked to two different carbon chemical shifts, which indirectly provides insight into the carbon skeleton of the material.[54,55,63]

Lesage et al. were among the first to publish a dipolar-based ^{1}H-^{13}C DQ/SQ experiment, the DQ-edited ^{1}H-^{13}C HETCOR, in 2006[55] which was then followed by a *J*-based alternative, the ^{1}H DQ-DUMBO – ^{13}C SQ refocussed INEPT by Webber et al. in 2010.[63] Using the latter sequence, the full ^{1}H assignment of β-maltose was proposed but unfortunately, the CH_2 signals were lost during data acquisition.[63] This loss of signal was a consequence of the duration used for the spin-echo periods within the INEPT element.[36,63] The duration (1.12 ms) was to target the ^{13}C signals that would correlate with a ^{1}H DQ cross peak that involved a direct C-H pair. Shorter duration (and ^{1}H DQ recoupling times) would improve CH_2 intensity at the expense of the CH groups.[63] Nevertheless, the full assignment of β-maltose was achieved. Another *J*-based sequence published was the MAS-*J*-^{1}H DQ – ^{13}C-HMQC SQ pulse sequence by Reddy and co-workers where the acquired data was used to identify ^{1}H DQ correlations, as shown in figure 5.2.[54,208,209]

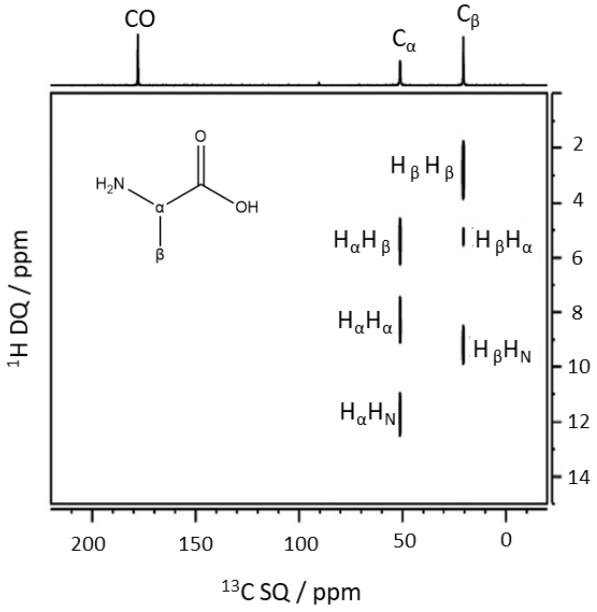


Figure 5.2: A MAS-J-^{1}H DQ – ^{13}C-HMQC of L-alanine, adapted from reference 54. 256 transients were collected for each of the 170 t_1 increments with a recycle delay of 4 s and a contact time of 2 ms at a MAS rate of 11 kHz. For each of the correlations, the associated H-H pair is labelled relative to the molecular structure.

As the 2D ^{1}H-^{13}C DQ/SQ spectrum correlates ^{1}H DQ coherences with ^{13}C SQ coherences, the correlation peaks, in principle, appear as pairs at a common ^{1}H DQ frequency. This then indicates that the two carbon sites, which are associated with the pair of ^{1}H DQ cross peaks, are connected via the ^{1}H dipolar network. Therefore, the peak pairs, in theory, can help trace the carbon skeleton. This is seen in figure 5.2 where the ^{1}H DQ correlation, H_{β}-H_{α}, appears as a pair, indicating that the two low frequency carbons are close together in space. Alongside this, it is also not unusual to see isolated ^{1}H DQ correlations that correlate to only one carbon signal, as seen for H_{β}-H_{β}. This is because it is not a requirement for two carbons to be present at each correlation. Auto-correlation cross peaks are generally observed mainly for carbons with multiple bonded protons, such as methyl resonances.

5.1.1 Aims of this Work

The aims of this work was to explore both heteronuclear and homonuclear DQ/SQ pulse sequences to understand whether there were reasons for their lack of prevalence within the literature. The dipolar-based ^{13}C-^{13}C INADEQUATE experiment was evaluated in relation to its applicability to a standard structural assignment workflow, where the pulse sequence used was from the publication *"Determining carbon–carbon connectivities in natural abundance organic powders using dipolar couplings"*.[60] On the other hand, the dipolar-based ^{1}H-^{13}C DQ/SQ experiment has been developed in-house where the optimisation and application has been discussed to determine the scope of the experiment. The data presented in figure 5.18 was supplied by Dr. Nghia Tuan Duong at the University of Marseille.

5.1.2 Model Systems Used

The model systems used in this work were L-alanine (Ala) and L-tyrosine methyl ester hydrochloride (Tyr). These were chosen due to their prevalence in the literature[54,60,200] with specific experimental parameters included within the captions. Overall, the experimental data was acquired using an Oxford superconducting magnet operating at 11.7 T (^{13}C of 125.65 MHz) which corresponds to frequencies of 499.69 (^{1}H) and 125.65 (^{13}C) MHz. The data was collected using a 4 mm HX magic-angle spinning probe at ambient temperature. The ^{13}C chemical shifts were referenced to the high-frequency signal of adamantane (38.5 ppm) whilst for the indirect dimension, the ^{13}C DQ chemical shifts or ^{1}H DQ chemical shifts were referenced by adding two known SQ signals together. In the ^{1}H-^{13}C DQ/SQ experiment, this was 2.8 ppm for the Me-Me ^{1}H DQ signal in Ala whilst for Tyr, the 6.2 ppm for the Me-Me ^{1}H DQ signal was used. To reference the ^{13}C DQ chemical shifts, two signals in the ^{13}C CP/MAS spectrum were added together (example correlations were 224 ppm for Tyr and 228 ppm for Ala). As described in the previous chapter, SPINAL-64[42,185] used as the heteronuclear

decoupling element at ^{1}H nutation frequencies of approximately 71-73 kHz and a ^{1}H 90° pulse length of around 3.4-3.6 µs whilst an FSLG pulse element[44] was used for homonuclear decoupling (FSLG 2π pulse of 19.8 µs). All calculations conducted follow the standard CASTEP protocol described in previous chapters.[85,94,98]

5.2 Dipolar-Based ^{13}C-^{13}C INADEQUATE Experiments

5.2.1 Overview

The dipolar-based ^{13}C-^{13}C INADEQUATE experiment was of interest due to its simplicity in parameter optimisation and speed of acquisition.[60] It was hypothesized that the dipolar variant would be more suited to incorporation into a workflow and allows for a wider range of molecules to be studied. The original pulse sequence published by Dekhil et al. has been used with no variation to the pulse elements.[60] Ideally, a 2.5 µs ^{13}C 90° pulse is used which corresponds to a MAS rate of 10 kHz.[61,210,211] For the excitation and reconversion blocks, $R20_2^9$ was used by the authors due to its robustness against ^{1}H-^{13}C dipolar coupling and ^{13}C CSA.[60,211,212]

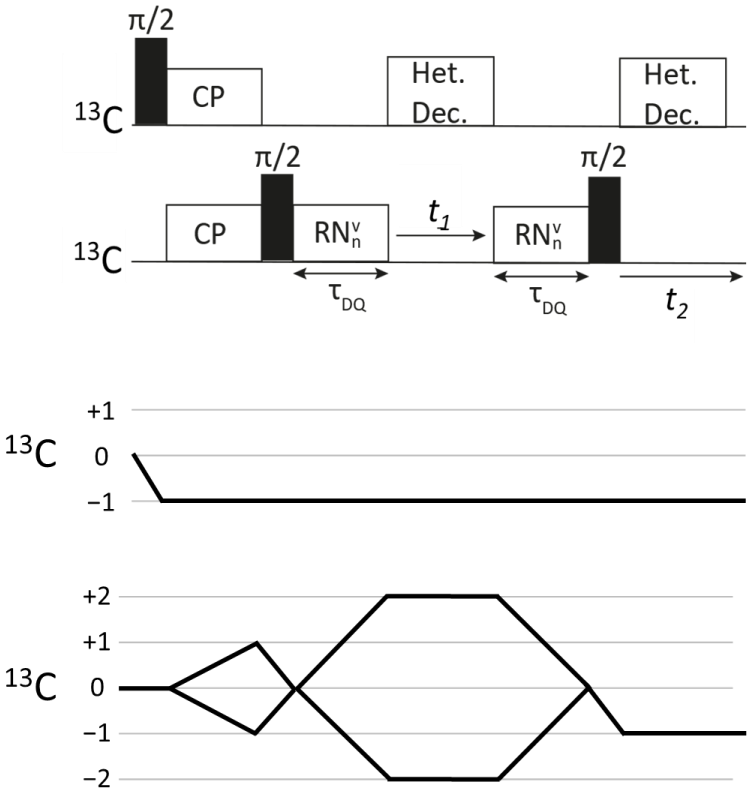


Figure 5.3: A dipolar-based ^{13}C-^{13}C INADEQUATE pulse sequence adapted from reference 60 with corresponding coherence transfer pathway. A RN_n^v recoupling block is used to excite and reconvert the DQ coherences prior to acquisition under heteronuclear decoupling. $R20_2^9$ was used which is 10 elements across 2 rotor periods.

5.2.2 L-Alanine (Ala)

To ensure the pulse sequence was correctly set up, Ala was used where in figure 5.4, the acquired data was achieved with a 3.0 µs ^{13}C 90° pulse at a MAS rate of 8.33 kHz.[213,214] The expected correlations at 71 ppm (Me and CH) and 228 ppm (CH and CO) were observed within a total acquisition time of 12 hrs.[213,214]

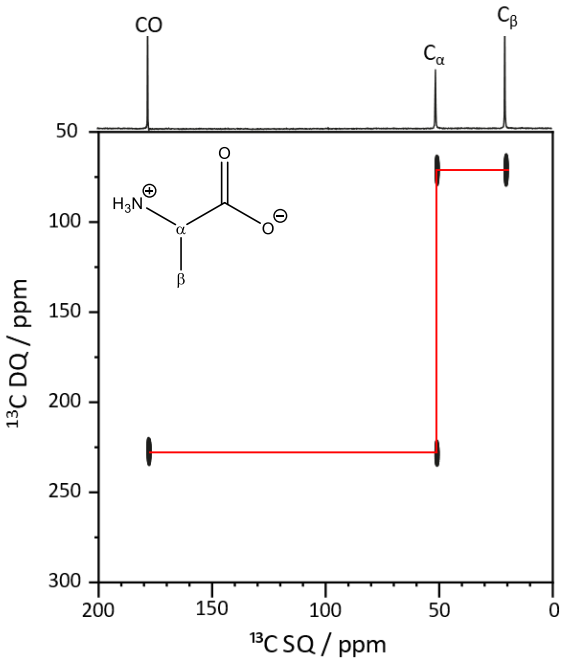


Figure 5.4: A dipolar-based ^{13}C-^{13}C INADEQUATE spectrum of Ala recorded on a 500 MHz Bruker Avance III HD spectrometer. A recycle delay of 1 s and contact of 1 ms was used at a MAS rate of 8.3 kHz with a 4 mm probe. Magnetisation transfer duration (τ_{DQ}) of 0.48 ms was used after optimisation. 1600 transients were collected for each of the 28 t_1 increments with a spectral width of 50 kHz and 40 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle[60] which resulted in a total experimental time of approximately 12 hrs. The carbon connectivity has been traced in red.

5.2.3 L-tyrosine methyl ester hydrochloride (Tyr)

As a more complex molecule, Tyr was investigated where a 2.5 µs ^{13}C 90° pulse was used. This did require significantly higher powers relative to Ala so care was taken as to avoid damage to the probe. Alongside this, an increase in both the number of transients and number of t_1 increments improved resolution where the latter improved spectral appearance the most. As shown in figure 5.5 a), all expected ^{13}C DQ correlations were observed, and the carbon skeleton was traced out. The majority of the correlations were also observed in figure 5.5 b) where the total acquisition time was reduced from 96 hrs to 15 hrs.

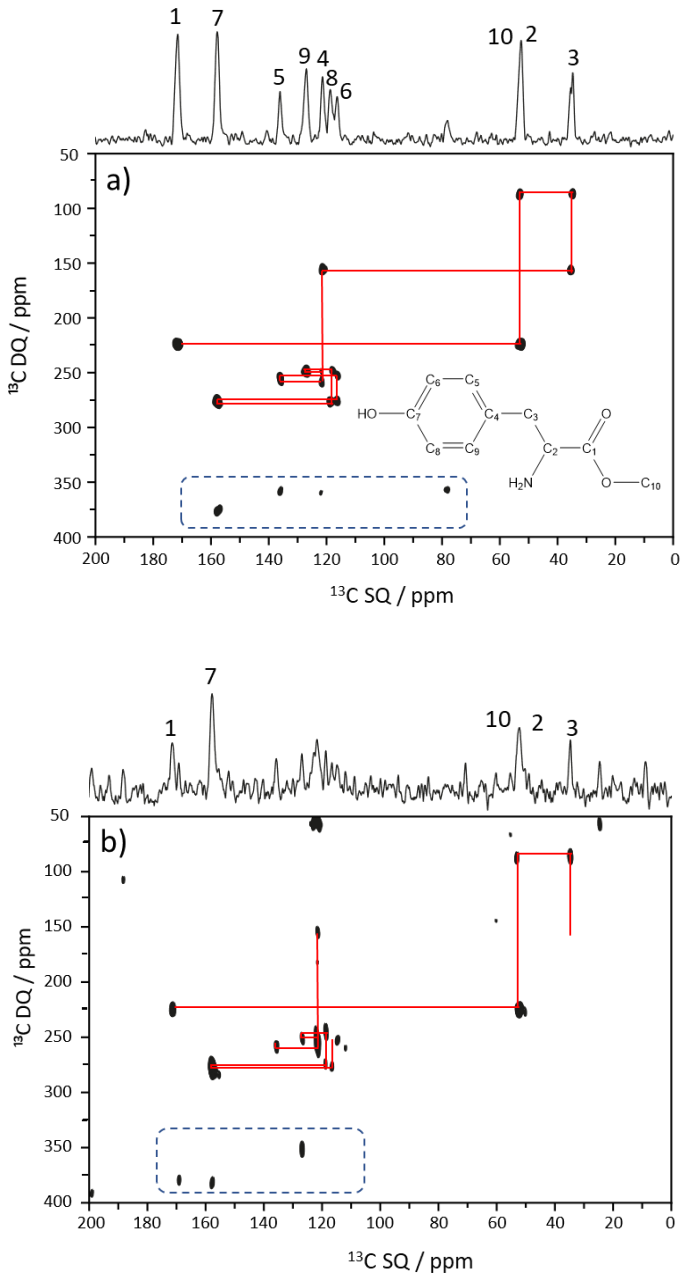


Figure 5.5: A dipolar-based ^{13}C-^{13}C INADEQUATE spectra for Tyr on a 400 MHz Bruker Avance III HD spectrometer. A recycle delay of 2.5 s and contact time of 4 ms was used at a MAS rate of 10 kHz with a 4 mm probe. Magnetisation transfer duration (τ_{DQ}) was set to 4 ms with 90 kHz heteronuclear decoupling after optimisation. a) 1600 or b) 400 transients were collected for each of the a) 86 or b) 69 t_1 increments with a spectral width of 50 kHz in both F_2 and F_1. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle[60] which resulted in a total acquisition time of approximately a) 96 hrs and b) 16 hrs. Additional signals are highlighted with the dashed box and the carbon connectivity is traced out in red whilst the missing cross peaks in (b) are clear from the lack of connectivity line. The horizontal trace is the skyline projection of the ^{13}C dimension.

However, as observed in the dashed boxes in both spectra, there is the presence of additional cross peaks which are suspected to be spinning sidebands (SSBs). The approximate distances between the expected ^{13}C DQ cross peaks and additional cross peaks in figure 5.6 were around 10 kHz, which reinforced the suspicion that the additional cross peaks were suspected SSBs. SSBs can cause a reduction in the centre band intensity as the magnetisation is shared. This may have contributed to the loss of signal intensity in 5.5 b).[215] Alongside their presence, the SSBs exhibit different asymmetric peak shapes relative to their related centre band cross peak. This could be due to the variation in phase in t_1 but to explore further, comprehensive simulations would be required.[216–218]

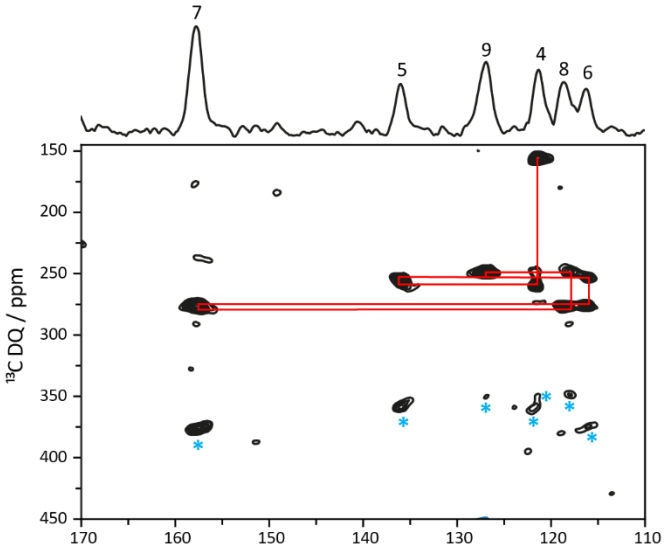


Figure 5.6: Aromatic region of the dipolar-based ^{13}C-^{13}C INADEQUATE for Tyr acquired in approximately 96 hrs. Carbon connectivity is traced in red and the suspected SSBs are marked with blue asterisks. The horizontal trace is the skyline projection of the ^{13}C dimension.

5.2.4 Section Summary

In this section, the applicability of the dipolar-based ^{13}C-^{13}C INADEQUATE experiment was investigated using two exemplar molecules. It was found that the experiment is a worthy alternative to the standard *J*-based pulse sequence. Noticeably, the total acquisition time is shorter and there is scope to adapt the pulse sequence to spectrometer requirements without sacrificing information, as observed for Ala. Parameter optimisation is straightforward but care must be taken if the desired 2.5 µs ^{13}C 90° is used as the power requirement may be significant for particular materials. An additional consideration is the presence of suspected SSBs in the acquired data. These have the possibility of reducing the intensity of the resulting cross peaks, limiting its applicability at higher magnetic fields. However, the suspected SSBs could also be exploited to derive information in relation to

$^{1}H/^{13}C$ CSA and how it is correlated between sites. This would require extensive spin simulations that fall outside of the scope of this work.[219–221] Nevertheless, the dipolar-based ^{13}C-^{13}C INADEQUATE demonstrated its capability of providing carbon connectivity information in a fraction of the time which increases its applicability in a structural assignment workflow.

5.3 Dipolar-Based ^{1}H-^{13}C DQ/SQ Experiments

5.3.1 Overview of Pulse Sequence

Rather than using a previously published pulse sequence, a new, dipolar-based ^{1}H-^{13}C DQ/SQ pulse sequence was developed as shown in figure 5.7. As with previous sequences, it is a 2D correlation experiment that is characterised by two dipolar mixing times. The ^{1}H-^{13}C correlation is achieved with CP, as seen in a conventional ^{1}H-^{13}C HETCOR, whilst the ^{1}H-^{1}H correlation is generated via the recoupling time (τ). A 16-step phase cycle has been used to select the DQ coherences where choice of contact time and total recoupling time, defined as the sum of excitation and reconversion durations ($\tau_{total} = \tau_{exc} + \tau_{rec}$) control the observed correlations.

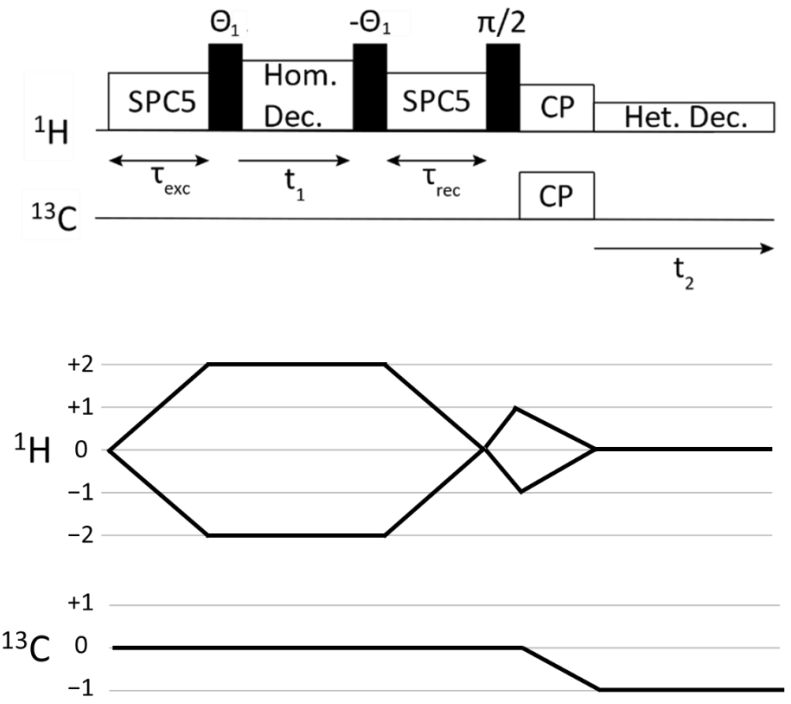


Figure 5.7: Schematic of the developed ^{1}H-^{13}C DQ/SQ pulse sequence with corresponding coherence transfer pathway. The first block of SPC5 generates the DQ coherences, which evolve under homonuclear decoupling before being reconverted to SQ coherences in the second SPC5 block. The pre-pulses (Θ) were used to rotate proton magnetisation from the

tilted transverse plane under homonuclear decoupling to the (xy) plane of the laboratory frame and back. After CP, the signal is detected in t_2 under heteronuclear decoupling.

As in a ^{13}C-^{13}C INADEQUATE spectrum, the correlation peaks in principle will appear in pairs at a common ^{1}H DQ frequency. This would indicate that the two carbon sites are close to one another, connected via the ^{1}H dipolar network. In theory, such peak pairs can thus help trace out the carbon skeleton in a similar way as a ^{13}C-^{13}C INADEQUATE experiment, but it is important to remember that not all cross peaks are observed as a pair as these are auto-corelation ^{1}H DQ signals. The types of correlations expected are visualised in figure 5.8. The sensitivity of the experiment should be comparable to a conventional ^{1}H-^{13}C HETCOR experiment, so should be much better than the classic ^{13}C-^{13}C INADEQUATE experiment, albeit with some sensitivity loss associated with generation of ^{1}H DQ coherence.

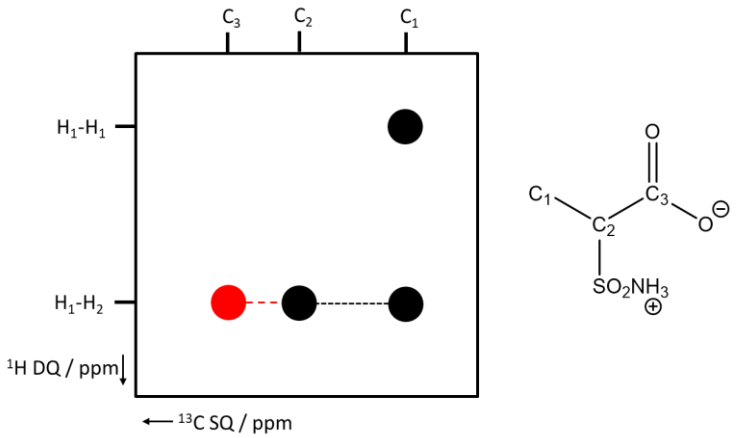


Figure 5.8: Schematic depicting the two expected types of ^{1}H DQ correlations using the exemplar molecule, $C_3H_7NO_4S$. From the molecular structure, at shorter contact times, C_1 is expected to observe the auto-correlation ^{1}H DQ signal of H_1-H_1 and the ^{1}H DQ frequency pair of H_1-H_2 with C_2. These correlations are visualised in black. At longer contact times, it is possible that the H_1-H_2 ^{1}H DQ correlation can extend to C_3 and this has been visualised in red.

5.3.2 Optimisation

Using Ala as the exemplar molecule, a number of optimisations were completed to evaluate different parameters of the ^{1}H-^{13}C DQ/SQ pulse sequence.

Recoupling – Choice of Element

A key difference to previous pulse sequences was the choice of recoupling block used. In the literature, the standard recoupling element used is POST-C7. It is described as a well-applied, symmetry-based, rotor-synchronised pulse element used in a vast array of publications.[53–55,62,187,210] However, the size of the ratio required between the nutation frequency and spin

frequency, at 7:1, does pose limits on the pulse powers and MAS rates used.[61,62] Therefore, in this work, SPC5 has been used instead. SPC5 has lower power requirements where, the symmetry requirement drops from sevenfold to fivefold.[59] This reduces the demand on the decoupling so higher spin rates can be achieved if desired. For homonuclear decoupling, FLSG was used as this is the standard decoupling element used in the conventional ^{1}H-^{13}C HETCOR experiments in this thesis.[44]

Recoupling – Choice of Duration

The number of SPC5 recoupling elements used for the excitation (τ_{exc}) and reconversion (τ_{rec}) blocks in figure 5.7 are defined by L1 and L2 respectively. SPC5 involves 5 complete elements for both excitation and reconversion, however, when compared to previous pulse sequences, the total recoupling time was considerably longer, as detailed in table 5.1.[54,59,61,63]

Table 5.1: Summary of the total recoupling times observed in the literature.

Experiment	τ_{total} / µs	Recoupling Method
^{1}H DQ-DUMBO – ^{13}C SQ refocussed INEPT[63]	68.6	POST-C7
MAS-J-^{1}H DQ – ^{13}C-HMQC[208]	52	POST-C7
^{1}H-^{13}C DQ/SQ (this work)	400 (5 elements)	SPC5
	80 (1 element)	

Reduction in signal intensity has also been noted previously by Webber and co-authors, where dipolar dephasing at longer recoupling durations resulted in signal loss, particularly for CH_2 groups.[10,63] This reduction in signal intensity was observed when the total recoupling time increased from figure 5.9 a) to 5.9 d) and was reinforced further with 1D cross section slices of the ^{1}H DQ correlation at 2.8 ppm. As a dipolar-mediated experiment, the magnetisation build up is quicker but this is at the expense of the faster decay during recoupling.[222] Therefore, it was decided that 1 recoupling element for both excitation and reconversion would be used.

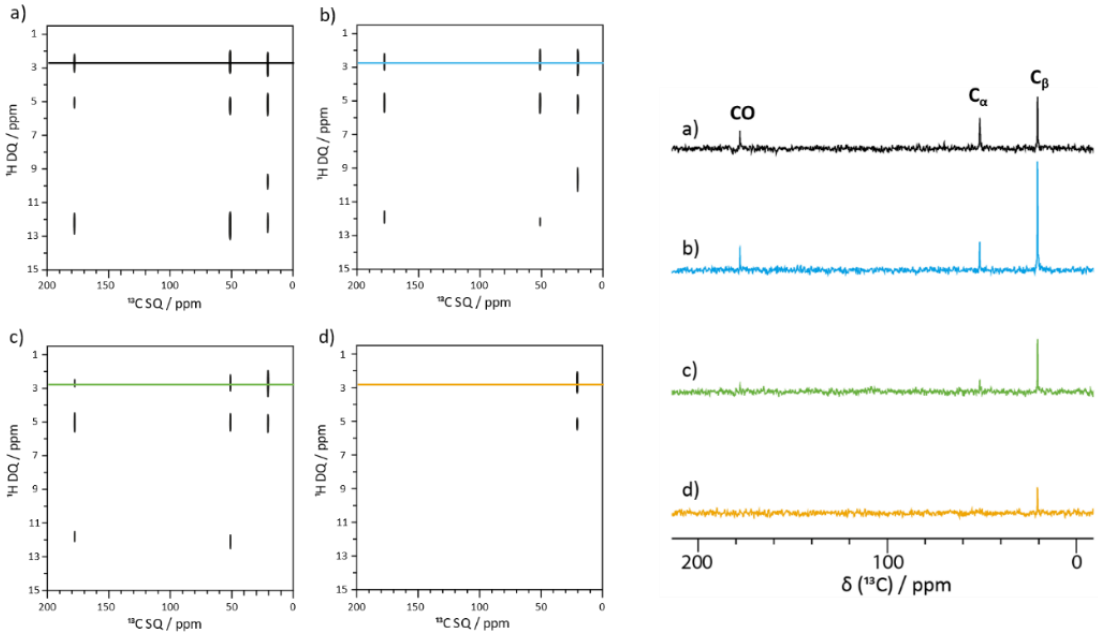


Figure 5.9: Left are the ^{1}H- ^{13}C DQ/SQ spectra at varying values of τ_{total} for Ala, recorded on a 500 MHz Bruker Avance III HD spectrometer using the pulse sequence in figure 5.7. A recycle delay of 1 s and a contact time of 0.4 ms was used at a MAS rate of 10 kHz with a 4 mm probe. 1152 transients were collected for each of the 48 t_1 increments with a spectral width of 50 kHz and approximately 19.3 kHz for F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle to achieve DQ selection. The τ_{total} (and choice of L1 and L2) were as follows: a) τ_{total} of 160 µs (L1/L2 = 2), b) τ_{total} of 240 µs (L1/L2 = 3), c) τ_{total} of 320 µs (L1/L2 = 4) and d) τ_{total} of 400 µs (L1/L2 = 5).

However, this is at the expense of the appearance of additional, unwanted cross peaks outside the expected spectral range, as highlighted in figure 5.10. It was initially suspected that the correlations were unwanted coherences, so a z-filter was inserted into the pulse sequence.[213,223–226] This was an attempt to suppress their presence, however, no change was observed. The unwanted cross peaks are suspected to be homonuclear decoupling artefacts. In order to avoid observing them in the acquired datasets, they can be omitted through choice of vertical plot range. This was used in figure 5.9 but also in the subsequent figures in this chapter.

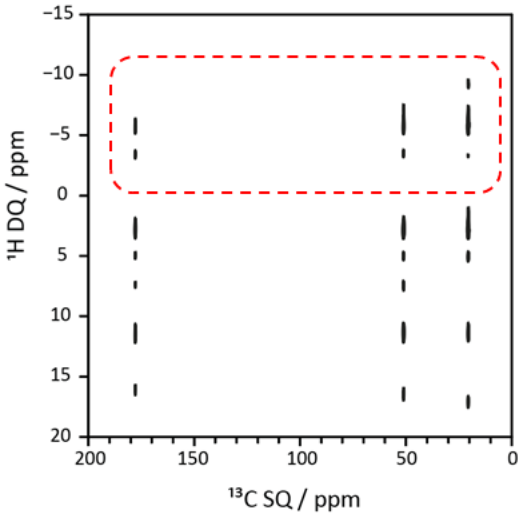


Figure 5.10: A ^{1}H-^{13}C DQ/SQ spectrum of Ala recorded on the 500 MHz Bruker Avance III HD spectrometer. A recycle delay of 1 s, contact time of 0.4 ms and τ_{total} of 80 µs was used at a MAS rate of 10 kHz using the 4 mm probe. 800 transients were collected for each of the 48 t_1 increments with a spectral width of 50 kHz and approximately 19.3 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle to achieve DQ selection. The suspected homonuclear decoupling artefacts are within the red dashed box.

Choice of Contact Time

As demonstrated in figure 5.11, the choice of contact time will influence the cross peaks that are observed. Shorter contact times, such as 0.1 ms in figure 5.11 a), limits the cross peaks to C_α and C_β, the protonated carbon environments of Ala. On the other hand, when the contact time is increased, more cross peaks are observed. The ^{1}H DQ frequencies have extended to include the quaternary carbon (CO) as well as the protonated environments. This was seen for the 2.8 ppm ^{1}H DQ frequency in 5.11 b), where the auto-correlation ^{1}H DQ correlation of H_β-H_β extends to both CO and C_α.

However, too long a duration of both contact and recoupling time can increase the risk of apparent cross peaks as a direct result of spin diffusion.[186,187] Spin diffusion can occur during the magnetisation transfer steps so choice of duration must be suitable otherwise the apparent cross peaks will complicate assignment.[186,187]

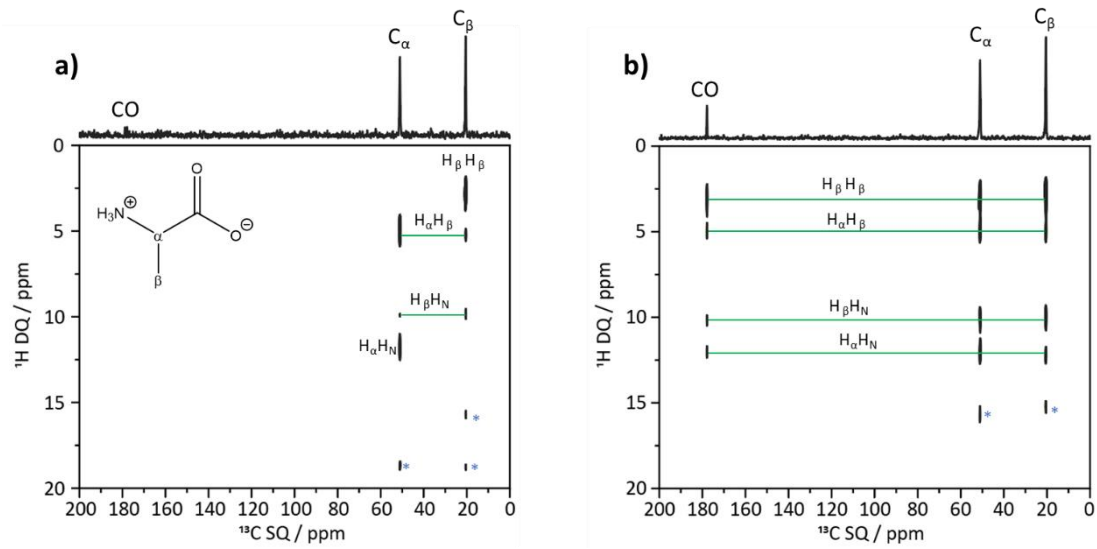


Figure 5.11: The ^{1}H-^{13}C DQ/SQ spectra of Ala where the contact time is varied where a) is 0.1 ms and b) is 0.4 ms. A contact time of 0.4 ms and τ_{total} of 80 µs was used at a MAS rate of 9434 Hz using the 4 mm probe. 512 transients were collected for each of the 48 t_1 increments with a spectral width of 50 kHz and approximately 19.3 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle to achieve DQ selection. The horizontal trace is the skyline projection of the ^{13}C dimension. Assignment labels reflects reference 54 in green whilst suspected SSB-like artefacts are marked with an blue asterisk (*).

5.3.3 Ala – Comparison to Previous Literature

Ala was used in reference 54 as an example to demonstrate the applicability of the developed pulse sequence, MAS-J-^{1}H DQ – ^{13}C-HMQC.[54] This presented the opportunity to directly compare the published data to the acquired dipolar-based ^{1}H-^{13}C DQ/SQ in figure 5.12. There are clear similarities between the two datasets where the published data is more reminiscent of the short-contact ^{1}H-^{13}C DQ/SQ in figure 5.11 a). This was somewhat expected as the pulse sequence is *J*-mediated so the magnetisation transfer is expected to be slower despite the longer contact time (2 ms).[54] Alongside this, in both datasets, there is the presence of an intermolecular correlation, labelled H_α-H_α, as marked in the blue dashed box of 5.12 b).

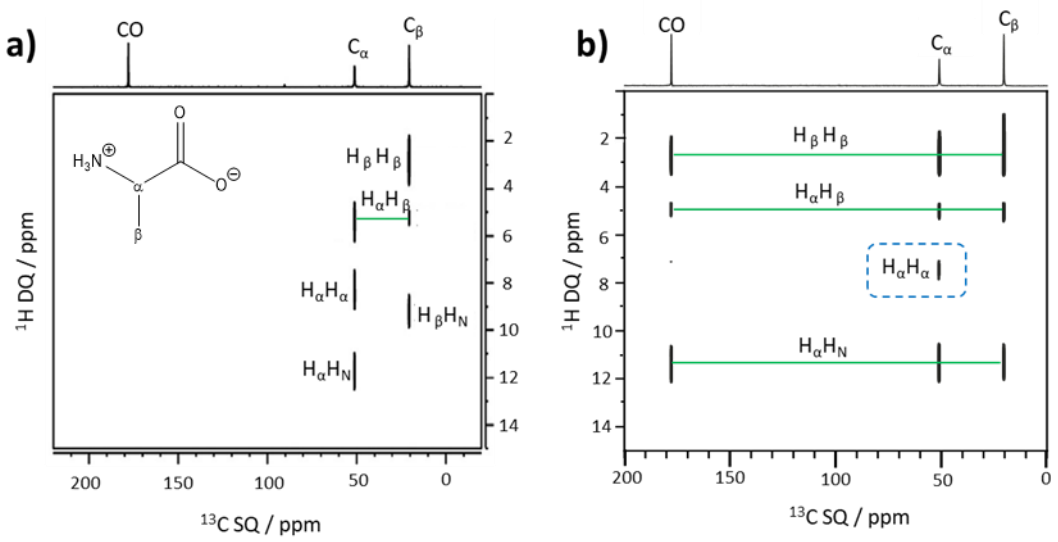


Figure 5.12: a) is the MAS-J-^{1}H DQ – ^{13}C-HMQC of Ala adapted from reference 54 and acquisition parameters are in the caption of figure 5.2. b) is the ^{1}H-^{13}C DQ/SQ of Ala recorded on the 500 MHz Avance III HD spectrometer. A recycle delay of 1 s, contact time of 0.4 ms and τ_{total} of 80 µs was used at a MAS rate of 10 kHz using the 4 mm probe. 800 transients were collected for each of the 48 t_1 increments with a spectral width of 50 kHz and approximately 19.3 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle to achieve DQ selection. Assignment labels used reflect reference 54 in green whilst suspected SSB-like artefacts are marked with an blue asterisk (*) and dashed box includes the intermolecular correlation.

This intermolecular correlation has been documented previously in the 2006 paper by Brown[186] but to ensure its plausibility, the correlation was verified through visualisation of the molecular packing in figure 5.13. The measured H_α-H_α distance is around 3.5 Å. Conventionally, ^{1}H-^{1}H DQ/SQ spectra are expected to be observed for H-H distances of up to 3.5 Å so the correlation is feasible and unlikely to be a product of spin diffusion.[186,187] The observed ^{1}H DQ intermolecular correlation does provide unexpected insight into crystal packing of Ala but also highlights that the interpretation of the ^{1}H DQ signals will not always be straightforward, particularly for larger molecules, as seen for Tyr.

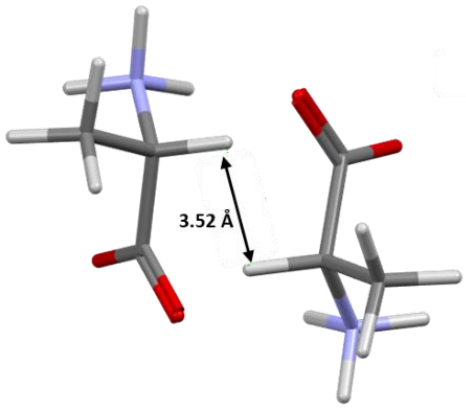


Figure 5.13: Molecular packing of Ala where the H_α-H_α intermolecular distance between the two H_α protons has been visualised in Mercury.[113] *Crystal structure used was* ***LALNIN23*****.**

5.3.4 Tyr – Evaluation and Assignment

For Tyr, it was hoped that the ^{1}H-^{13}C DQ/SQ experiment would provide similar levels of insight as the dipolar-based ^{13}C-^{13}C INADEQUATE experiment in Section 5.2. Three sets of experimental data were acquired which were a short-contact ^{1}H-^{13}C DQ/SQ at 9434 Hz and two long-contact ^{1}H-^{13}C DQ/SQ at 9434 Hz and 10 kHz. The variation in spin rate reflects the variation in power used during the SPC5 recoupling element, where a lower power is associated with a slower MAS rate.

To help decipher which ^{1}H DQ correlations are present, a combination of ^{13}C and ^{1}H calculated chemical shifts were used alongside simulated ^{1}H-^{13}C DQ/SQ spectra. To generate the simulated ^{1}H-^{13}C DQ/SQ spectra, the related magres file was required alongside knowledge of the C-H and H-H distances, specifically the shortest distances, which were also extracted from the magres file. The resulting simulated spectra are shown in figure 5.14. The choice of C-H and H-H distance can be used to control the correlations observed where the average ^{1}H chemical shift for the CH_3 and NH_3 protons have been used.

It is important to note however, that the synthetic ^{1}H-^{13}C DQ/SQ spectra are relative to the choice of C-H distance and H-H distance only. Conversely, in the acquired experimental data, the dipolar couplings will be affected by dynamics, particularly those of the Me group, so the observed ^{1}H DQ frequencies will vary.[10,31,227,228]

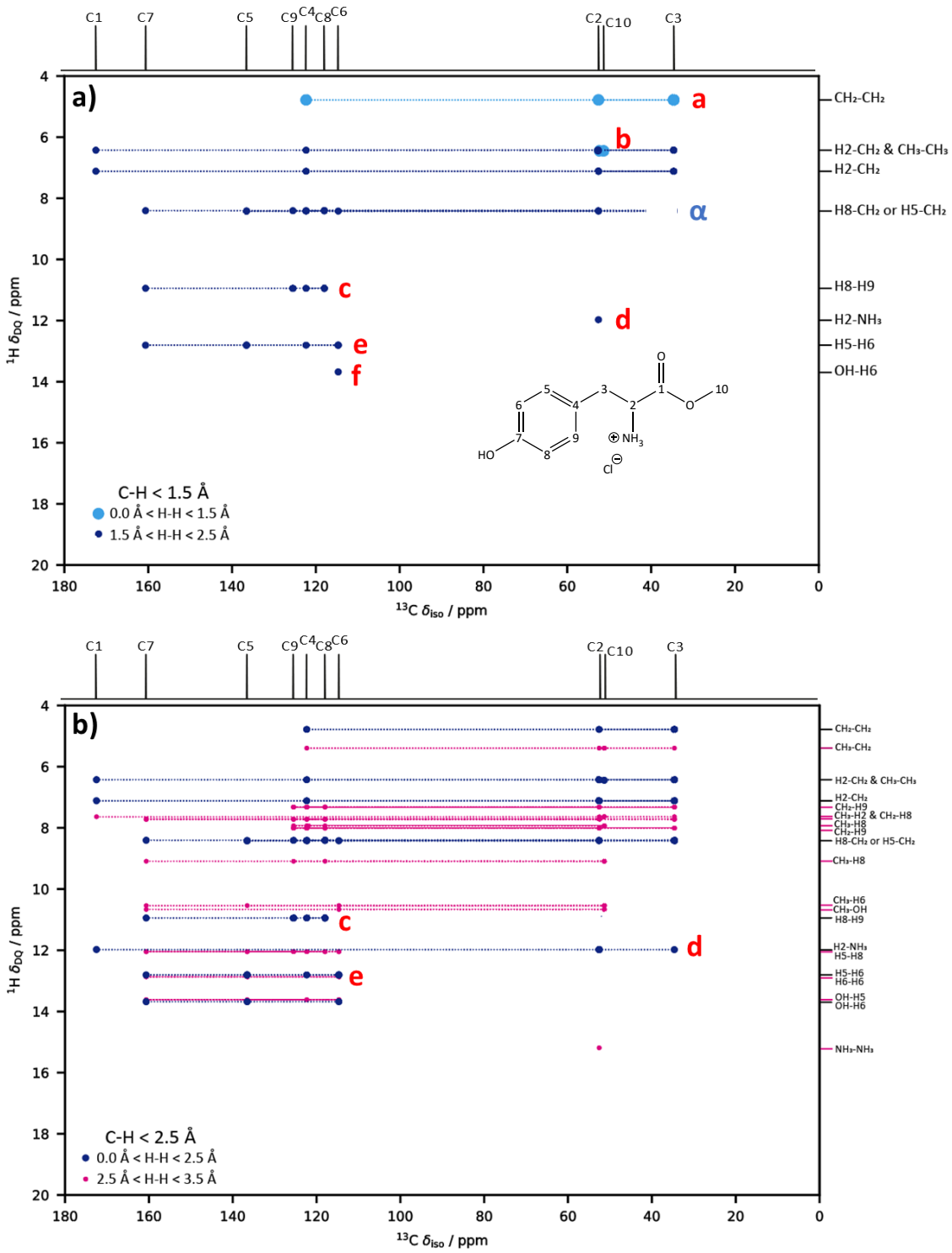


Figure 5.14: Simulated short-contact a) and b) long-contact ^{1}H-^{13}C DQ/SQ spectra where the C-H distance has been restricted to 1.5 Å and 2.5 Å, respectively. ^{1}H DQ correlations are reported on the virtual axis where the H-H distances thresholds are listed within the legend. Labels a to f in red refer to the correlations observed in the short-contact ^{1}H-^{13}C DQ/SQ spectrum in figure 5.15.

Short-Contact ^{1}H-^{13}C DQ/SQ spectrum

Beginning with the short-contact ^{1}H-^{13}C DQ/SQ spectrum, there are 6 identified ^{1}H DQ correlations labelled **a** to **f** in figure 5.15 whose assignments are tabulated in table 5.2.

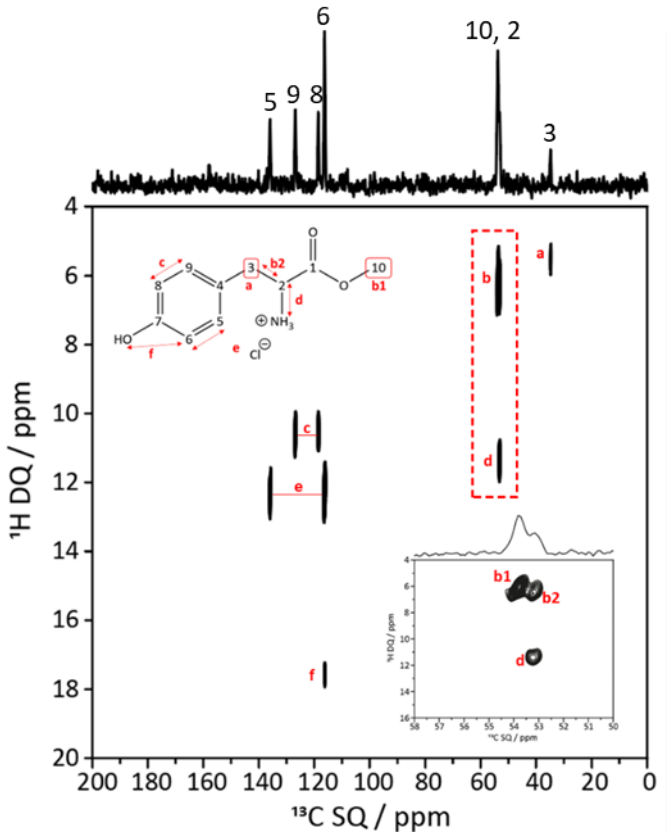


*Figure 5.15: A short-contact ^{1}H-^{13}C DQ/SQ spectrum of Tyr recorded on the 500 MHz Bruker Avance III HD spectrometer. A recycle delay of 7 s, contact time of 0.1 ms and τ_{total} of 80 µs was used at a MAS rate of 9434 Hz using the 4 mm probe. 512 transients were collected for each of the 48 t_1 increments with a spectral width of 50 kHz and approximately 19.3 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle to achieve DQ selection. The correlation peaks are labelled **a** to **f**, with the corresponding hydrogens annotated on the molecular structure. The inset corresponds to the red dashed box, and highlights that the **b** correlation contains two distinct peaks, **b1** and **b2**. The horizontal traces are skyline projections of the ^{13}C dimension.*

Using the simulated ^{1}H-^{13}C DQ/SQ spectra, alongside the calculated ^{13}C chemical shifts, assignments **a**, **c**, **e** and **f** were straightforwardly made. At shorter contact times, the ^{1}H DQ correlations were limited to protonated environments, where the direct C-H pair is involved in the ^{1}H DQ correlation. For example, **e**, refers to the **H5-H6** correlation and involves **C5** and **C6**. From the calculated ^{13}C chemical shifts, **C5** is distinct in its ^{13}C SQ frequency so can be

used to distinguish the **H5-H6** correlation from the **H8-H9** correlation. This is used in conjunction with the predicted ^{1}H DQ frequency to reinforce the assignment, where there is good agreement between ^{1}H-^{13}C HETCOR derived and ^{1}H-^{13}C DQ/SQ derived chemical shifts.

Table 5.2: Summary of the identified correlations in the short-contact ^{1}H-^{13}C DQ/SQ.

Label	^{1}H DQ correlation	Associated carbons	^{1}H DQ / ppm
a	CH_2-CH_2	C3	5.4
b1	CH_3-CH_3	C10	6.2
b2	H2-CH_2	C2	6.6
c	H8-H9	C8, C9	10.7
d	H2-NH_3	C2	11.5
e	H5-H6	C5, C6	12.3
f	OH-H6	C6	17.6

The assignment of **b1**, **b2** and **d** in figure 5.15 required the inset where focus was limited to the partially overlapped region of 50-55 ppm. Using the assignment of **d**, **b1** and **b2** could be identified. **C10** refers to the OMe group and, as shown in figure 5.14 a), only one ^{1}H DQ correlation is expected whereas **C2** is predicted to observe multiple correlations, including **H2-NH_3**, which has been assigned as **d**. Therefore, **C2** can be assigned to the lower ^{13}C SQ frequency signal and **C10** must correspond to the higher frequency ^{13}C SQ signal. This distinction can also be made using a conventional ^{13}C NQS and short-contact ^{1}H-^{13}C HETCOR, as discussed in Chapter 4. The overall assignment of experimental ^{1}H chemical shifts has been summarised in table 5.3.

*Table 5.3: Proposed experimental assignment for Tyr alongside the related GIPAW calculated chemical shifts[14] which have been calculated in-house from the CSD deposit **YEFTUZ**.*[162,163]

Environment	^{13}C Chemical Shift / ppm		^{1}H Chemical Shift / ppm		
	Experimental Chemical Shift / ppm	Calculated Chemical Shift / ppm	Experimental (^{1}H-^{13}C HETCOR)[15]	Experimental (^{1}H-^{13}C DQ/SQ)	Calculated
C1	171.0	172.6	-	-	-
C2 (CH)	53.1	52.6	3.7	3.9	4.2
C3 (CH_2)	34.7	34.7	2.4	2.7	2.3
C4	121.9	122.4	-	-	-
C5 (CH arom.)	135.8	136.6	6.2	6.2	6.2
C6 (CH arom.)	116.2	114.8	6.1	6.1	6.3
C7	157.8	160.7	-	-	-
C8 (CH arom.)	118.5	118.1	5.5	5.4	5.5
C9 (CH arom.)	126.8	125.6	5.4	5.3	5.1
C10 (CH_3)	53.7	51.4	3.1	3.1	3.1
NH_3^+				7.6	7.5
OH				11.5	7.1

Long-contact ^{1}H-^{13}C DQ/SQ spectra

Two long-contact ^{1}H-^{13}C DQ/SQ spectra were considered in figure 5.16. The correlations of **b2, d** and **e** are extended to the adjacent quaternary carbons in the molecular structure, as predicted in the simulated spectra. It is particularly powerful as the extension of a given correlation in a reliable interpretation is a rare occurrence in conventional ^{1}H-^{13}C HETCOR

[14] *The σ_{ref} values used for ^{13}C and ^{1}H were 167.9 and 29.35 ppm for ^{13}C and ^{1}H, respectively. The gradient, m, was set to −1.*

[15] *The experimental ^{1}H chemical shifts in this column are from the short-contact ^{1}H-^{13}C HETCOR, as described in Chapter 4.*

spectra. Particularly, as at longer contact times in ^{1}H-^{13}C HETCOR spectra, it can be difficult to determine which cross peaks are robust as it can sometimes appear that all environments are correlating with one another. This is discussed further in the context of figure 5.17.

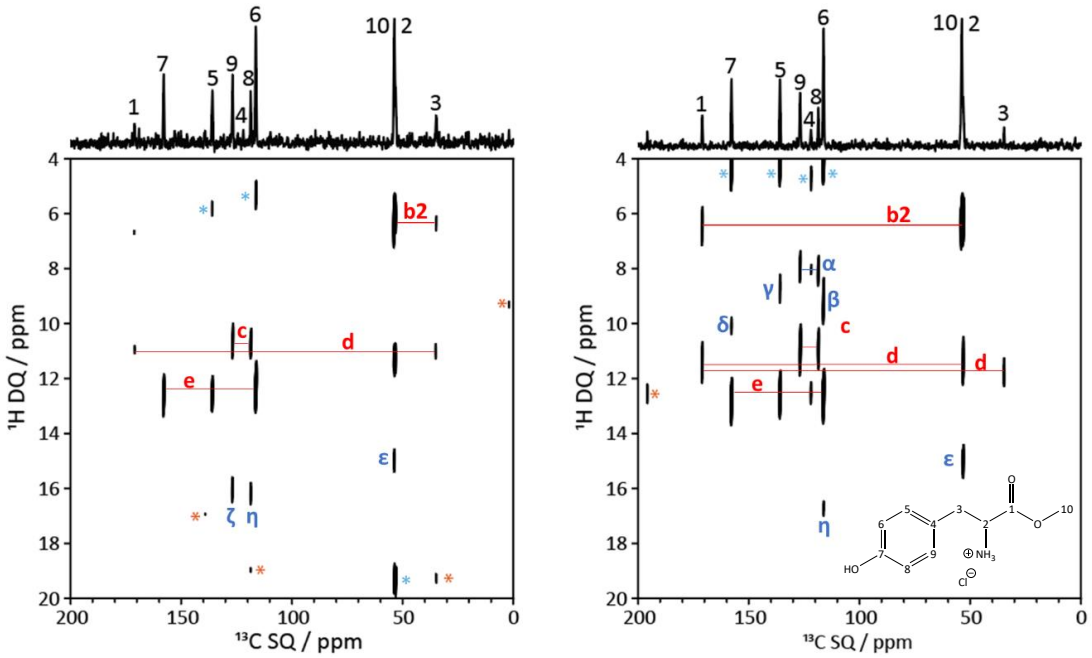


Figure 5.16: A long-contact ^{1}H-^{13}C DQ/SQ spectrum recorded of Tyr at 9434 Hz (left) and 10 kHz (right) where the contact time was 0.4 ms with a τ_{total} of 80 µs in both datasets. The experimental parameters for the 9434 Hz spectrum are within the figure 5.15 caption. For the 10 kHz spectrum, a recycle delay of 4 s was used on the 4 mm probe where 800 transients were collected for each of the 48 t_1 increments with a spectral width of 50 kHz and approximately 19.3 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle to achieve DQ selection. Previously identified DQ correlations are in red, whilst additional ^{1}H DQ correlations are in blue. The horizontal trace is the skyline projection of the ^{13}C dimension.[16]

As labelled in figure 5.16 b), there appears to be additional ^{1}H DQ correlations (**α** to **ζ**). These could be assigned using the simulated ^{1}H-^{13}C DQ/SQ spectra in figure 5.14 however, as only the contact time has been extended (from 0.1 ms to 0.4 ms) rather than the total recoupling time, it is unlikely that additional correlations are observed relative to the ones already seen. It is more likely that the additional cross peaks are artefacts from the pulse sequence itself.

[16] *Both spectra with the F_1 skyline projection are presented in Section A3 of the Appendix which demonstrates both the SSB artefacts that have been omitted from the plots through choice of F_1 spectral width alongside the suspected RRF resonances.*

With further analysis of the data in figure 5.16, there are indications that the additional cross peaks are rotor-rf (RRF) resonances.[229] The correlations labelled **ε** and **η**, which are present in both datasets in figure 5.16, have variable ^{1}H DQ chemical shift values alongside the lack of consistency of the cross peaks labelled in blue between the experiments. The additional correlations observed for alanine in figure 5.11 could also possibly be explained by these RRF resonances. Also, the proposed ^{1}H SQ chemical shift for the OH signal is over 4 ppm higher than the predicted value so the predicted ^{1}H DQ correlations in the simulated ^{1}H-^{13}C DQ/SQ spectrum that include the OH proton are not realistic. This is important, especially as the majority of the predicted ^{1}H DQ correlations fall in the 8-12 ppm region.

The presence of such artefact cross peaks has been documented previously by Leskes, Madhu and Vega where they described the phenomenon as RRF resonances.[229] As a consequence of second averaging of the dipolar interaction, they describe how RRF peaks are difficult to predict prior to data acquisition so when they appear experimentally, it can be difficult to determine what is a real cross peak and what is an RRF-resonance.[229] One method suggested to mitigate against their presence is by shifting the RF offset and adjusting the spectral window that the RRF lines fall outside which by coincidence, this work has done post-acquisition for both the ^{13}C-^{13}C INADEQUATE and ^{1}H-^{13}C DQ/SQ datasets.

Comparison to long-contact ^{1}H-^{13}C HETCOR spectrum

The relationships between the ^{1}H DQ correlations in figures 5.15 and 5.16 can be related to the ^{1}H-^{13}C HETCOR spectrum, as annotated in figure 5.17. Similarities can be seen between the protonated environments but in order to tie the information together to the observed ^{1}H DQ pairs, a ^{1}H-^{1}H DQ/SQ spectrum would be required. This is why the annotations are marked with an asterisk. As noted previously, at longer contact times, it can be a struggle to identify which long-range cross peak correlations are real or apparent. An example of this is seen in figure 5.17, where the lower ^{1}H SQ cross peaks appear to correlate for all environments of Tyr, where it is clear that cannot be the case.

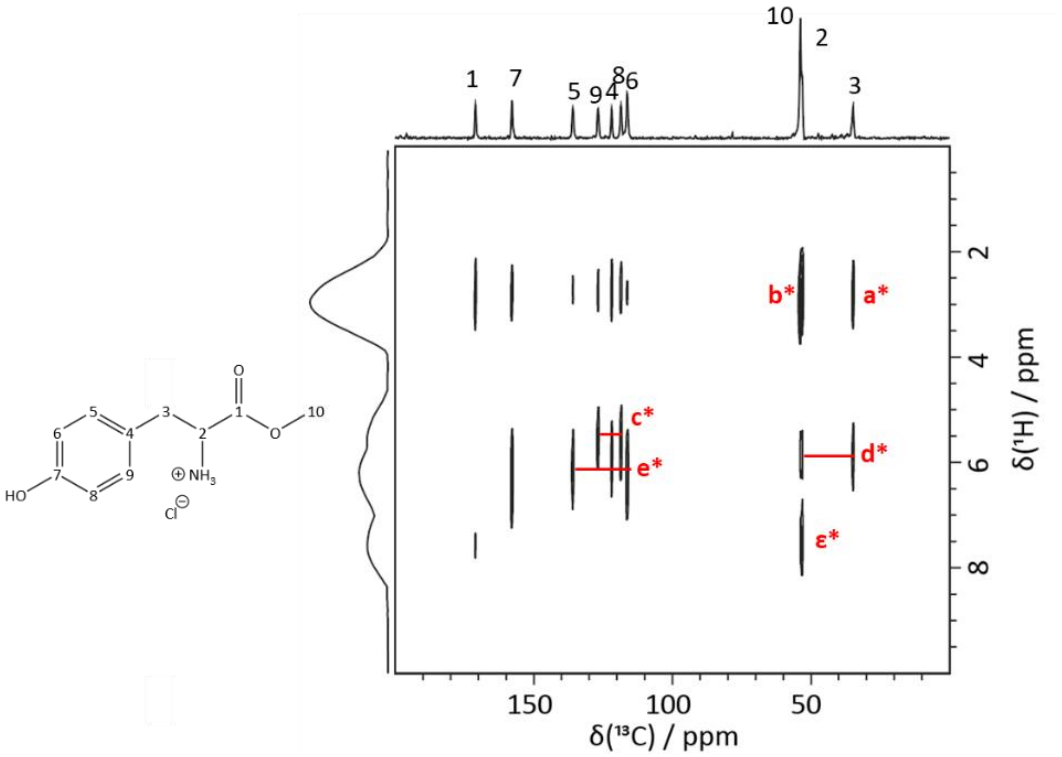


Figure 5.17: An annotated long-contact ^{1}H-^{13}C HETCOR spectrum of Tyr recorded on a 500 MHz Avance III HD spectrometer. A recycle delay of 2 s and contact time of 1 ms was used at a MAS rate of 10 kHz on the 4 mm probe. 80 transients were collected for each of the 24 t_1 increments with a spectral width of 50 kHz and approximately 19.3 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination. The horizontal and vertical traces are the skyline projections of the ^{13}C and ^{1}H dimensions respectively. Annotations in red refer to 'quasi' ^{1}H DQ correlation assignments observed in the ^{1}H-^{13}C DQ/SQ datasets in figure 5.16 .

Comparison to Marseille data

To test the reproducibility of the acquired data, a ^{1}H-^{13}C DQ/SQ spectrum was acquired by the researchers at the University of Marseille on their 400 MHz Avance III spectrometer. The acquired experimental data in figure 5.18 reflects the previous observations in the datasets discussed where, despite the longer contact time (0.5 ms), the resulting ^{1}H DQ correlations reflect the short-contact ^{1}H-^{13}C DQ/SQ in figure 5.15. The signal intensity variation can be attributed to differences in SNR in the acquired datasets. Nevertheless, it was clear that the assigned cross peaks in figure 5.15 were reproducible whilst the suspected artefact peaks in figure 5.16 were not. This reinforced further that that additional correlations in figure 5.16 were not reproducible (this likely to be RRF resonances) and that the pulse sequence set-up and acquisition plays a direct role in whether they are present experimentally or not.

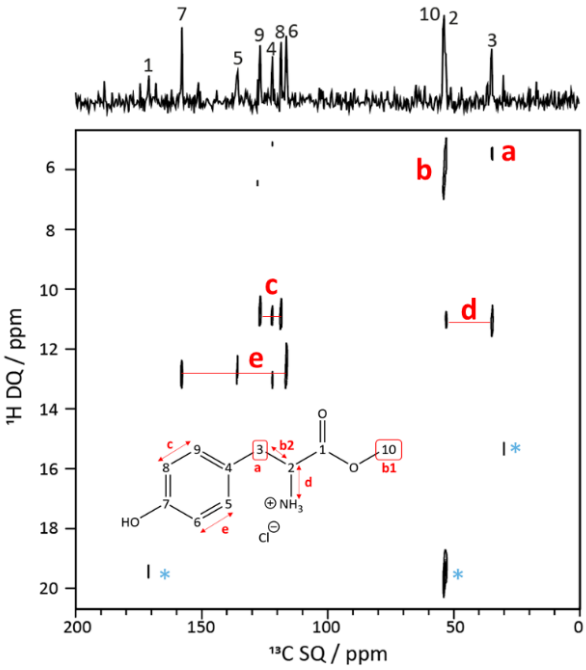


Figure 5.18: A ^{1}H-^{13}C DQ/SQ spectrum of Tyr recorded on a 400 MHz Avance III spectrometer. A recycle delay of 2 s, contact time of 0.5 ms and τ_{total} of 80 µs at a MAS rate of 10 kHz using a 4 mm probe. 768 transients were collected for each of the 48 t_1 increments with a spectral width of approximately 35.7 kHz and 13.2 kHz in F_2 and F_1 respectively. The States-TPPI method was used to achieve sign discrimination with a 16-step phase cycle to achieve DQ selection.

5.3.5 Section Summary

The initial aim of the ^{1}H-^{13}C DQ/SQ experiment was to provide an alternate method to probe carbon connectivity. As in the conventional ^{1}H-^{13}C HETCOR, the length of contact time can be changed but it is suspected that any new correlations, relative to the short-contact variant, are likely artefacts rather than new correlations.[229] Using short contact times, a number of ^{1}H DQ frequencies can be identified and tend to be restricted to protonated environments. A similar level of insight can be provided by the combination of ^{1}H-^{13}C HETCOR and ^{1}H-^{1}H DQ/SQ but the benefit of the ^{1}H-^{13}C DQ/SQ is that it is one experiment which is acquired at a given MAS rate. A summary of the assigned ^{1}H DQ correlation peaks for Tyr has been summarised in figure 5.19, especially as the presence of RRF resonances can complicate assignment.

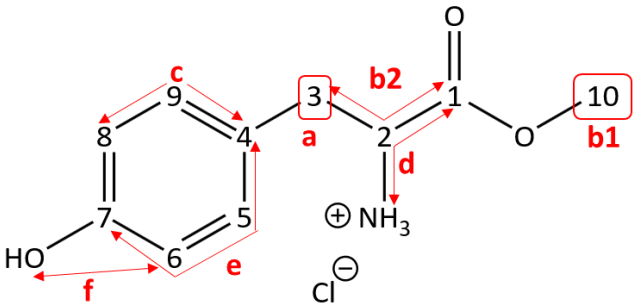


Figure 5.19: Visualisation of the DQ correlations identified where the red and blue annotations refers to assigned short-contact ^{1}H-^{13}C DQ/SQ and long-contact ^{1}H-^{13}C DQ/SQ spectra in figures 5.15 and 5.16 respectively.

For some ^{1}H DQ correlations, direct comparison of the ^{1}H DQ frequencies to the simulated ^{1}H-^{13}C DQ/SQ spectrum can identify the H-H correlation. This provides insight into the carbons that are close in space to one another and can be reinforced with longer contact times. Additional cross peaks for the same ^{1}H DQ frequency can be extended to other protonated environments but more interestingly, to quaternary carbons. A great example of this was for Tyr where, for the **H5-H6** correlation, two additional ^{1}H DQ cross peaks were observed in the long-contact ^{1}H-^{13}C DQ/SQ spectrum in figures 5.16 and 5.18, linking **C4** and **C7**. This gives connectivity information to assign the carbons of the aromatic ring and is a result unlikely to be achieved from the conventional ^{1}H-^{13}C HETCOR alone. At longer contact times, as seen in figure 5.17, it can become difficult to interpret the observed cross peaks confidently, particularly those that indicate a long-range correlation; Additional data acquisition is likely to be required to be convincing. In contrast, the ^{1}H-^{13}C DQ/SQ confidently achieves this in a single experiment. The simulated spectra helps reiterate a particular assignment but care must be taken as the predicted correlations are solely based on C-H and H-H distances with no consideration of other factors.

A key benefit of the ^{1}H-^{13}C DQ/SQ experiment is its ability to provide insight into the intermolecular behaviour of materials in relation to molecular packing and possible hydrogen bonding. The ^{1}H-^{13}C HETCOR can also probe hydrogen bonding with the appearance of high frequency ^{1}H SQ cross peaks, but tends to be limited to exchangeable protons.[162,204] Here, the ^{1}H-^{13}C DQ/SQ experiment is observing intermolecular correlations between aromatic and aliphatic protons as well as exchangeable protons.

Further optimisation is required to reduce the observed artefacts. Numerical simulations would be helpful to identify problems, develop strategies to mitigate against the loss of signal

intensity but also determine whether the artefact features may provide additional information. Variation of the RF offset, as discussed by Leskes, Madhu and Vega[229] could also be attempted experimentally to mitigate against the RRF artefact peaks. Simulations will also help understand the relationship between the two durations (contact time and recoupling time) to fully comprehend how one affects the other to allow the experiment to be run to its full potential.

5.4 Conclusions

In this chapter, two different approaches to determine ^{13}C-^{13}C connectivity have been explored. A dipolar-based ^{13}C-^{13}C INADEQUATE experiment was found to be a suitable alternative to the *J*-based ^{13}C-^{13}C INADEQUATE where for both exemplar molecules, the carbon skeleton could be traced in a fraction of the acquisition time. In contrast, the ^{1}H-^{13}C DQ/SQ experiment has been conceptualised and explored where its initial aim as a direct alternative to the ^{13}C-^{13}C INADEQUATE experiment has morphed into a more complex study. Alongside indirect carbon connectivity, it has been found that additional intermolecular information can be derived. For future work, robust simulations would be required to fully optimise both pulse sequences, however, this chapter has demonstrated the precedent of both pulse sequences to aid in the robust structural characterisation of molecular organics solids.

Chapter 6: Exploring the Solution-State Behaviour of Ritonavir

6.1 Introduction

Solution state NMR is one of the methods used within industry and academia to understand the structure, purity, conformation and conformational exchange of a given material.[230–233] In this chapter, the focus is on the ability of solution-state NMR to probe the behaviour of the protease inhibitor, Ritonavir (RVR). RVR is primarily used in the treatment of acquired immune deficiency syndrome (AIDS)[234] but more recently in Paxlovid, a combination therapy for treatment of COVID-19.[235–237]

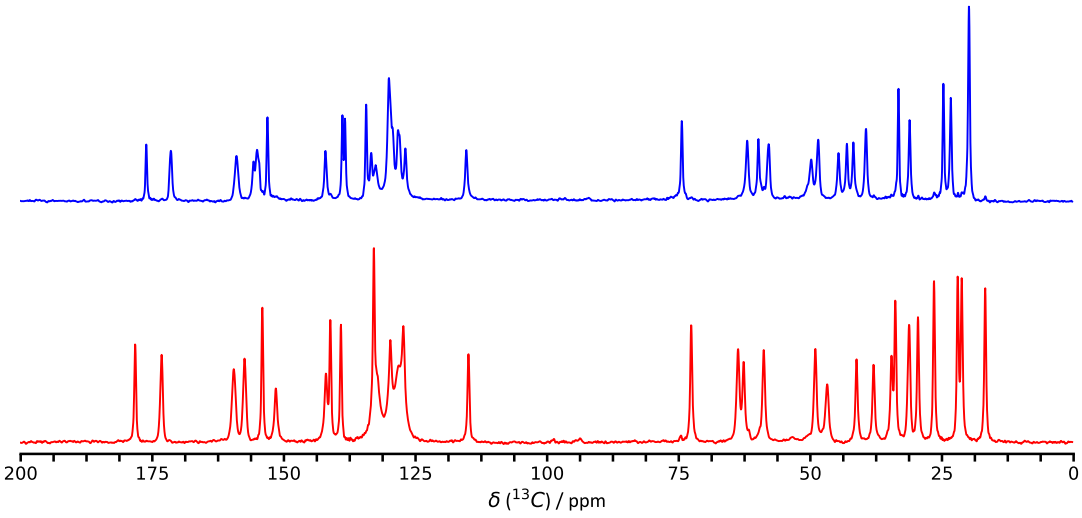


Figure 6.1: The ^{13}C CP/TOSS[49] spectra of RVR Form I (top) and RVR Form II (bottom). 512 transients were taken at a recycle delay of 4 s (Form I) or 6 s (Form II) at 10 kHz on the 4 mm HX MAS probe at ambient temperature. Data was acquired on the Oxford 11.7 T superconducting magnet, and Bruker AVANCE III HD console, operating at frequencies of 499.69 (^{1}H) and 125.65 (^{13}C) MHz with SPINAL-64 heteronuclear decoupling of approximately 73.5 kHz.[42] All spectra were processed in TopSpin and reproduced from data in reference 238.[238]

Within the solid state community, RVR represents an infamous example of a disappearing polymorph due to its conformational polymorphism.[17] The appearance of a second form, Form II, after the initial drug development pipeline required an entire re-design of the manufacturing process in order to isolate the desired Form I.[18,19] This was due to the inability to isolate Form I once Form II was present in the synthetic route.[17] The variation in the two forms can be identified using ^{13}C SSNMR, as demonstrated in figure 6.1.

Interest in the solid state behaviour of RVR remains, where there has been work both experimentally and computationally to understand the solid state landscape.[239–241] Most recently, a new form, Form III has been characterised by Zhang, Yao and Henry, where the configurational preferences for the key torsions have been highlighted in figure 6.2.[242] Interestingly, Form III may have been discovered previously in 2014 but had not been realised at the time.[243,244] Kawakami et al. identified the form as Form IV, in line with the 2003 work by Morissette et al., but there were discrepancies in the related PXRD patterns.[243,245] However, when the PXRD pattern was overlaid with Form III, the isolated form by Kawakami and co-workers appears to be Form III mixed with amorphous material.[242,243] Even after two decades of exploration, the solid state conformational landscape of RVR is not yet fully understood.

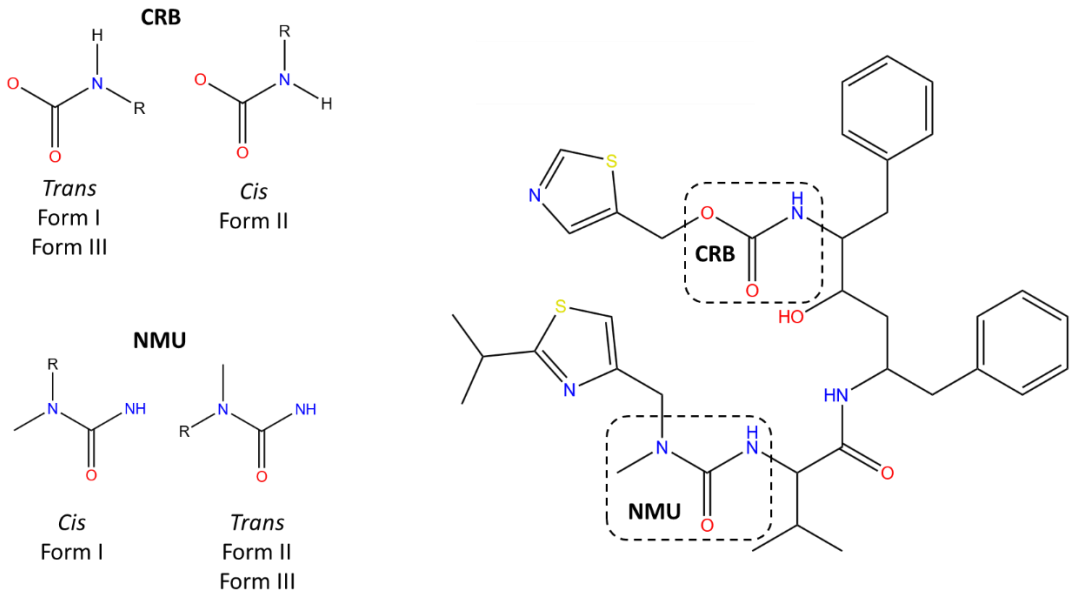


Figure 6.2: Variation in torsional preference of the carbamate (CRB) and N-Methyl Urea (NMU) moieties in RVR across the three identified solid forms. RVR has been drawn in the solid state form I conformation. Figure has been adapted from reference 239.[239]

The carbamate (CRB) and N-Methyl Urea (NMU) torsions, shown in figure 6.2, play a significant role in the physicochemical behaviour of RVR. As noted by the initial publication by Bauer et al., the more stable Form II has a more unusual conformation for the CRB torsion, where it sits cis.[19,239] They found that 2% of deposited structures within the CSD exhibited a cis carbamate torsion in 2001 and this number increased to 5% in 2021 where Wang et al. conducted a similar study.[19,239] Despite its unusual configuration, the cis CRB configuration in Form II allows all strong hydrogen bond donors (HBD) and acceptors (HBA) to be satisfied, which results in its increase in stability.[19] Form III on the other hand, was identified as the least stable form of the three.[242]

The solution-state behaviour of RVR is less well documented and little is known about the conformational variability relative to solvent. The seminal publication by Bauer et al. does describe an initial study of RVR in DMSO where they found evidence of the existence of at least two conformations in solution at a ratio of 99 to 1.[19] However, attempts to extract kinetic information from variable temperature experiments were unsuccessful.[19] There has also been little discussion regarding the suspected dominant conformation as it has been assumed that it would reflect the solid state Form I conformation.[19,239,241] Despite both CRB and NMU torsions varying between the forms, focus within the literature has been on the CRB torsion. The study by Chakraborty, Sengupta and Wales focused on purely computational estimations of kinetic and thermodynamic parameters for the interconversion of the CRB torsion.[241]

The spectral assignment of RVR has also been limited to a handful of publications and patents within a limited solvent range.[246–248] Despite the work by Bauer et al. highlighting the possible presence of at least two conformations in solution,[19] the published assignments tend to assume that there is only one conformation present. If there is the presence of additional signals, they have not been considered as a result of conformation but of the study itself. One example of this is the work by Jeong and co-workers, which acknowledges additional signals in the 1D ^{1}H spectrum, but they were suspected to reflect the different possible binding sites between RVR and the chosen catalyst for SABRE measurements.[247]

6.1.1 Aims of this Work

As literature tends to focus on the solid state behaviour of RVR, the aim of this chapter was to explore the solution-state behaviour through robust spectral assignment, quantification of the observed exchange and preliminary conformational analysis. Three different solvents were used ($CDCl_3$, DMSO and EtOD) to establish whether there is solvent dependence for

both exchange and conformation. The work in this chapter has been completed in collaboration with Ms. Sophia Hassan, Dr. Juan Aguilar, Professor Aurora Cruz Cabeza and Professor Mark Wilson.

6.2 Materials and Methods

6.2.1 Materials

RVR (CAS Number: 155213-67-5) was purchased from Merck Life Sciences Ltd and used without further purification. The provided solid was confirmed as Form I by PXRD, as shown in section A4 of the Appendix. All solvents used were of analytical grade and used as received. The solvents were $CDCl_3$ (CAS Number: 865-49-6) EtOD-d_6 (CAS Number: 1516-08-1) and DMSO-d_6 (CAS Number: 2206-27-1).

6.2.2 Methods

A significant number and variety of experiments were conducted in this chapter to probe the assignment, conformation, and exchange of RVR. The data was acquired by the solution-state NMR facility at Durham University, namely by Dr. Juan A. Aguilar and sample preparation was aided by Ms. Sophia V. S. Hassan. Around 15 mg of sample was dissolved in approximately 750 µL of each solvent and data was acquired mainly on Bruker Neo 700 MHz spectrometer with an operating frequency of 699.74 MHz for ^{1}H. All experimental data was processed in either Topspin or MestreNova.

For the exchange investigation, initial ^{1}H spectra were acquired over 32 transients, each consisting of 6536 complex data points with a spectral width of 19.8 kHz. The ^{1}H-^{1}H EXSY spectra were acquired at ambient temperature unless specified otherwise over 4 transients, each consisting of 4096 (t_2) and 1024 (t_1) data points with a spectral width of 7.8 kHz. The number of increments was 512 with a repetition time of 2.5 s and 0.5 s acquisition time. The duration of the experiments varied depending on the mixing time used but were approximately 3–3.5 hrs. Mixing times used were between 0.05 s and 0.5 s. Prior to integration, the baselines of the 2D EXSY spectra were flattened using an automated polynomial baseline correction whilst the peak intensities were obtained from either 2D volume integration or 1D cross section slices with no line broadening applied. The structural assignment was accomplished using 1D and 2D NMR experiments which included: 1D ^{1}H, 1D ^{13}C{^{1}H}, 1D and 2D ^{1}H-PSYCHE, ^{1}H-^{1}H COSY, ^{1}H-^{1}H NOESY, ^{1}H-^{1}H ROESY, ^{1}H-^{13}C edited-HSQC, ^{1}H-^{13}C HMBC, ^{1}H-^{15}N HSQC and ^{1}H-^{15}N HMBC.

6.3 Assignment

6.3.1 Overall Assignment

The structural assignment for RVR in each solvent has been tabulated in table 6.1 where values in brackets reflect the associated minor signal(s) with the atomic numbering in figure 6.3. In DMSO, the assignment in table 6.1 matches what has been previously reported.[246] To note, the notation 'M' refers to the major set of signals in each solvent whilst 'm' refers to the minor set of signals in each solvent.

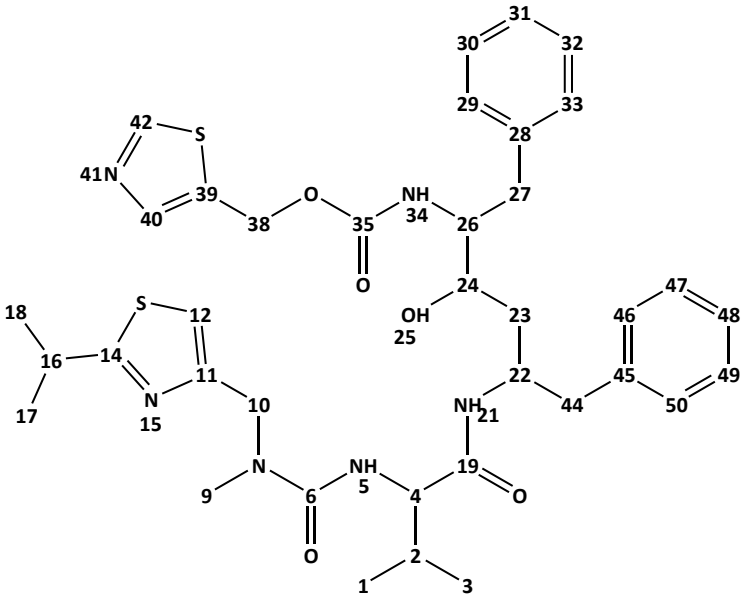


Figure 6.3: Molecular labelling for RVR where the conformation drawn reflects Form I (i.e the CRB is trans and NMU is cis).[18,19]

Table 6.1: Spectral assignment of RVR in $CDCl_3$, DMSO and EtOD with the minor assignment of environments in brackets. The full assignment can be found in the MestreNova files in the Data Archive.

Label	Type	$CDCl_3$ ^{13}C or ^{15}N / ppm ^{1}H / ppm	DMSO ^{13}C or ^{15}N / ppm ^{1}H / ppm	EtOD ^{13}C or ^{15}N / ppm ^{1}H / ppm
1,3	^{i}Pr	17.61,19.72 (17.63, 19.81) 0.80, 0.89 (0.84, 0.93)	18.14, 19.39 0.74	18.80, 20.01 (18.81, 20.06) 0.82, 0.84 (0.88)
2	CH	29.41 (29.44) 2.28 (2.31)	30.39 1.88	31.61 (31.63) 1.95 (2.00)
4	CH	60.69 (60.91) 3.99	59.49 3.93	61.24 (61.29) 3.95 (3.99)
5	NH	84.10 6.26	84.24 6.01 (6.05)	85.38 6.15 (6.21)
6	C=O	159.30	157.42	159.51
8	N	80.46	75.87	78.01
9	Me	35.03 (35.08) 2.95 (2.97)	34.48 2.87	35.20 2.95 (2.97)
10	CH_2	49.32 4.33, 4.40	48.23 4.41,4.47	49.54 4.44, 4.54 (4.44, 4.55)
11	C	151.63	152.79	153.43
12	CH	114.46 (114.55) 7.00 (7.02)	114.03 7.18	115.18 (115.24) 7.14 (7.16)
14	C	179.47	177.20	179.52
15	N	309.42	315.40	310.42
16	CH	33.21 3.25	32.42 3.23	34.04 3.27
17,18	Me	23.22, 23.05 1.36 (1.38)	22.77, 22.83 1.30	23.53 1.36
19	C=O	172.45 (172.27)	171.14	173.41 (173.05)
21	NH	126.11 (125.26) 6.61 (6.62)	126.23 7.68 (7.70)	129.00
22	CH	49.50 (49.42) 4.17 (4.21)	47.02 (47.10) 4.15 (4.03)	49.03 (48.82) 4.27 (4.16)
23	CH_2	40.45 (40.08) 1.59, 1.65 (1.57)	38.24 1.45	39.75 (40.31) 1.63 (1.68)

24	CH	69.61 (69.84) 3.64 (3.59)	68.98 (56.48) 3.58 (3.54)	70.17 (70.35) 3.69
25	OH	-	4.61 (4.65)	5.28
26	CH	57.25 (58.17) 3.74 (3.59)	55.46 (56.22) 3.82 (3.72)	56.86 (57.34) 3.97 (3.91)
27	CH_2	38.87 (39.76) 2.83 (2.73,2.78)	37.15 2.68	39.08 (40.31) 2.78, 2.81 (2.74)
28	C	138.17 (138.25)	139.43	139.70
29,33	CH	129.35 7.18	128.97 7.19	130.12 7.21
30,32	CH	128.36 7.24	127.93 7.21	128.89 7.17
31	CH	126.27 7.18	125.75 7.15	126.83 7.11
34	NH	84.58 (86.09) 5.21 (5.14)	86.14 (86.81) 6.88 (6.41)	84.50 6.44 (5.92)
35	C=O	155.76 (155.71)	155.55	157.39 (157.13)
38	CH_2	58.03 (58.29) 5.21 (5.01, 5.11)	57.16 (57.55) 5.15,5.12 (5.03)	58.58 (58.93) 5.18 (4.99, 5.03)
39	C	133.55	134.05	135.38 (134.83)
40	CH	143.26 (143.19) 7.82 (7.75)	143.0 (142.76) 7.85 (7.78)	143.44 (143.59) 7.80 (7.74)
41	N	320.20	323.93	314.45
42	CH	154.44 (154.41) 8.79 (8.77)	155.46 (155.34) 9.05 (9.02)	156.28 (156.39) 8.94 (8.93)
44	CH_2	41.84 (42.07) 2.72 (2.62, 2.73)	39.78 2.67,2.60	41.42 (41.78) 2.67,2.75
45	C	137.60 (137.64)	138.73	139.25
46,50	CH	129.14 7.07	129.25 (129.12) 7.11 (7.03)	130.32 7.10 (7.02)
47,49	CH	128.31 7.14	127.81 7.16	128.78 7.11
48	CH	126.41 7.11	125.69 7.13	126.73 7.06

It was clear in the experimental data in all solvents that there are at least two sets of signals present and an example of this is shown in figure 6.4 for the ^{1}H environments **22**, **24** and **26** in DMSO. The choice of experiment was also used to distinguish between exchange cross peak pairs and NOEs where a difference in phase is seen in the ^{1}H-^{1}H ROESY spectrum.[33] For example, the major (M) and corresponding minor (m) signal of environment **26** was identified in the ^{1}H-^{1}H ROESY spectrum as the corresponding cross peak is the same phase as the diagonal peak. If the cross peak was an NOE, it would be a different phase to the diagonal, as seen between **26M** and **22M** below.

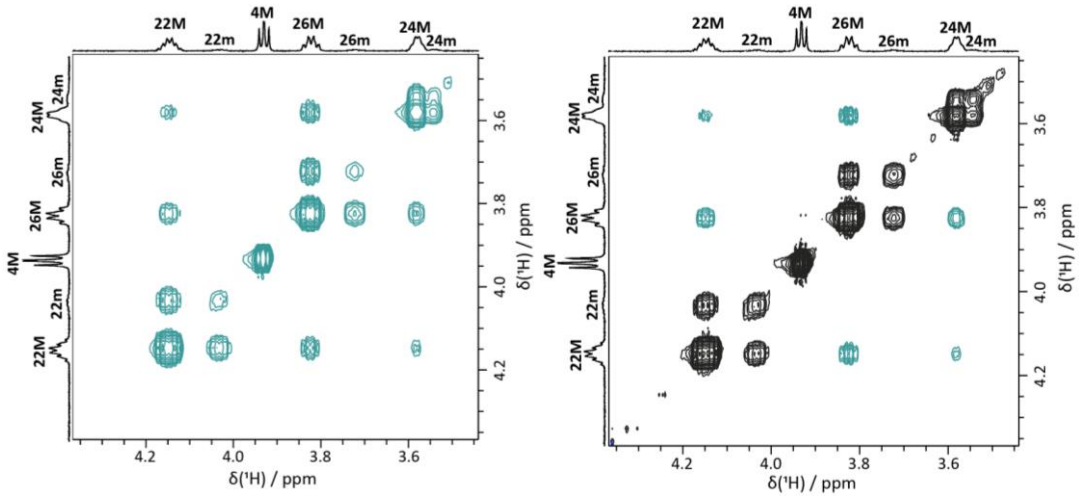


Figure 6.4: ^{1}H-^{1}H NOESY (left) and ^{1}H-^{1}H ROESY (right) spectra for RVR in DMSO. The benefit of ^{1}H-^{1}H ROESY is the distinction between exchange cross peaks (black) from NOEs (green). The environments have been labelled on both axes.

6.3.2 Conformer Ratio

Using a selection of resolved exchange pairs in the 1D ^{1}H spectra, conformer ratios were approximated. The overall average for each torsion has been tabulated in table 6.2 where a significant deviation from the previous 99 to 1 estimate is seen.[19] This is discussed further in Section 6.5.

Table 6.2: Summary of the average conformer ratios for the CRB and NMU torsions in each solvent (with associated standard error).[17]

Solvent	Type	CRB Torsion		NMU Torsion	
		Environments Considered	Approximate Ratio	Environments Considered	Approximate Ratio
$CDCl_3$	Non-polar	1	6.5 : 1	2	5.4 ± 0.3 : 1
DMSO	Polar, aprotic	6	10.9 ± 0.4 : 1	0	
EtOD	Polar, protic	4	8 ± 0.5 : 1	1	6.1 : 1

6.4 Preliminary Conformational Analysis of RVR

As noted in section 6.1, the CRB and NMU torsions play a pivotal role in the conformational behaviour of RVR in the solid state.[19,239,241] In this section, a preliminary conformational analysis has been conducted based on the acquired experimental data, quantitative methods and visualisation of proposed conformations in Avogadro.[114]

6.4.1 NMU and CRB Torsion

NMU Torsion

Focussing first on the NMU region of RVR, the ^{1}H-^{1}H NOESY spectra were used to determine whether an NOE was observed between the NMe group (**9**) and the NH environment (**5**). As shown in figure 6.5, an NOE was present between **9** and **5** in DMSO, as seen in figure 5.6 b), but not in $CDCl_3$ and EtOD, as shown in figure 5.6 a) and c). This indicates that there is a difference in the configuration in DMSO relative to $CDCl_3$ and EtOD, where the lack of an NOE corresponds to a cis configuration. The ^{13}C chemical shifts for the CO group (**6**) also appear linked to this variation in conformation where in EtOD and $CDCl_3$, the signal is 2 ppm higher than in DMSO (159.3 ppm and 159.5 ppm relative to 157.4 ppm, respectively).

[17] *To calculate the average, a subset of exchange pairs within the 1D ^{1}H spectrum that are associated with the CRB and NMU torsions respectively. As there are no distinct exchange pairs for the NMU region in DMSO, this approximate ratio was not calculated. Individual environment values can be found in section A5 of the Appendix.*

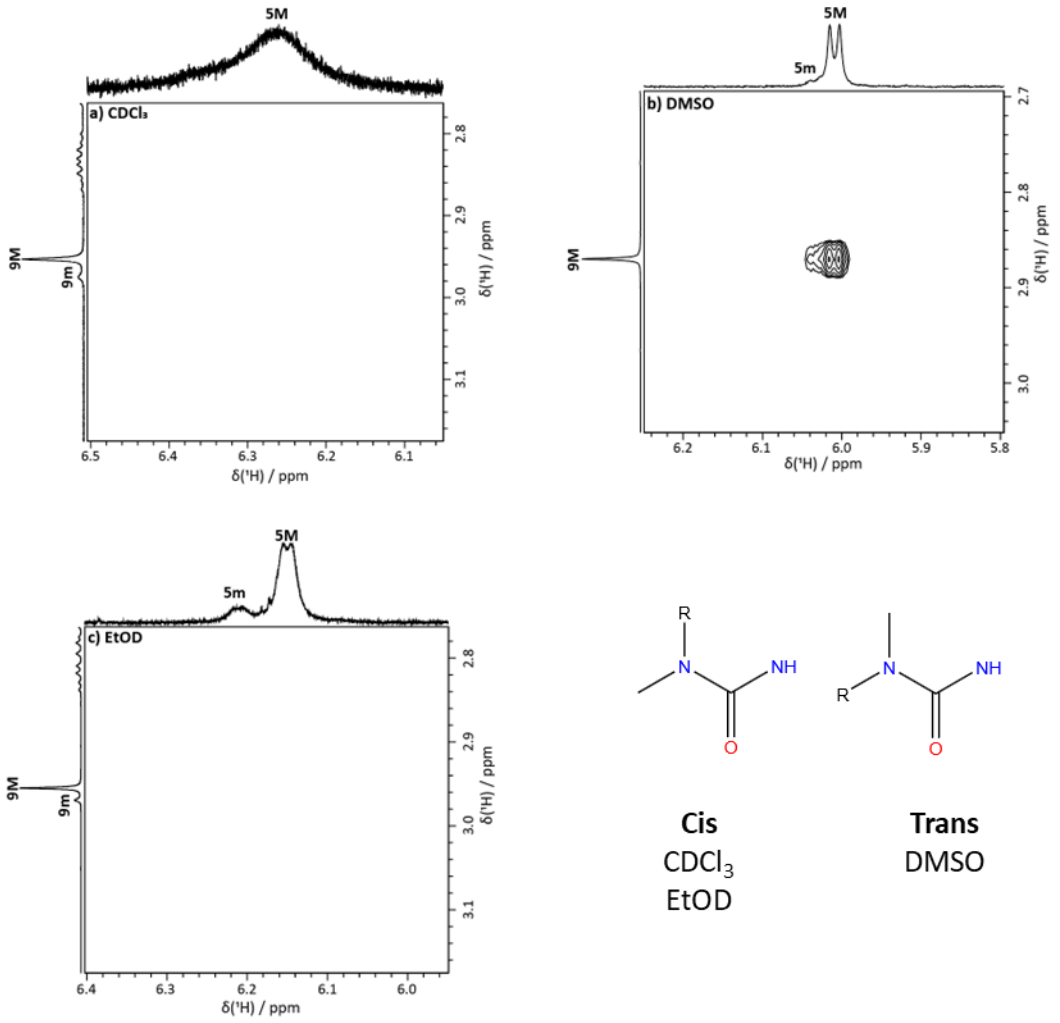


*Figure 6.5: Section of the ^{1}H-^{1}H NOESY spectrum of RVR in a) $CDCl_3$, b) DMSO and c) EtOD, focussing on the presence or absence of a cross peak between **5** and **9**. To note, environment **5** in $CDCl_3$ is broad, indicating exchange with solvent or water within the sample.*

Initial Evidence for CRB Torsion

Also in the ^{1}H-^{1}H NOESY spectra, environments **40** and **42** do not show any long-range NOEs in any of the solvents. This indicates that the CRB torsion does not bring the triazole ring in close proximity to the main body of the molecular structure nor is it close to either of the phenyl rings. This does deviate from the suggested global minimum in the gas phase, where the conformation is collapsed in on itself with stacking of the thiazole rings for both CRB configurations.[241]

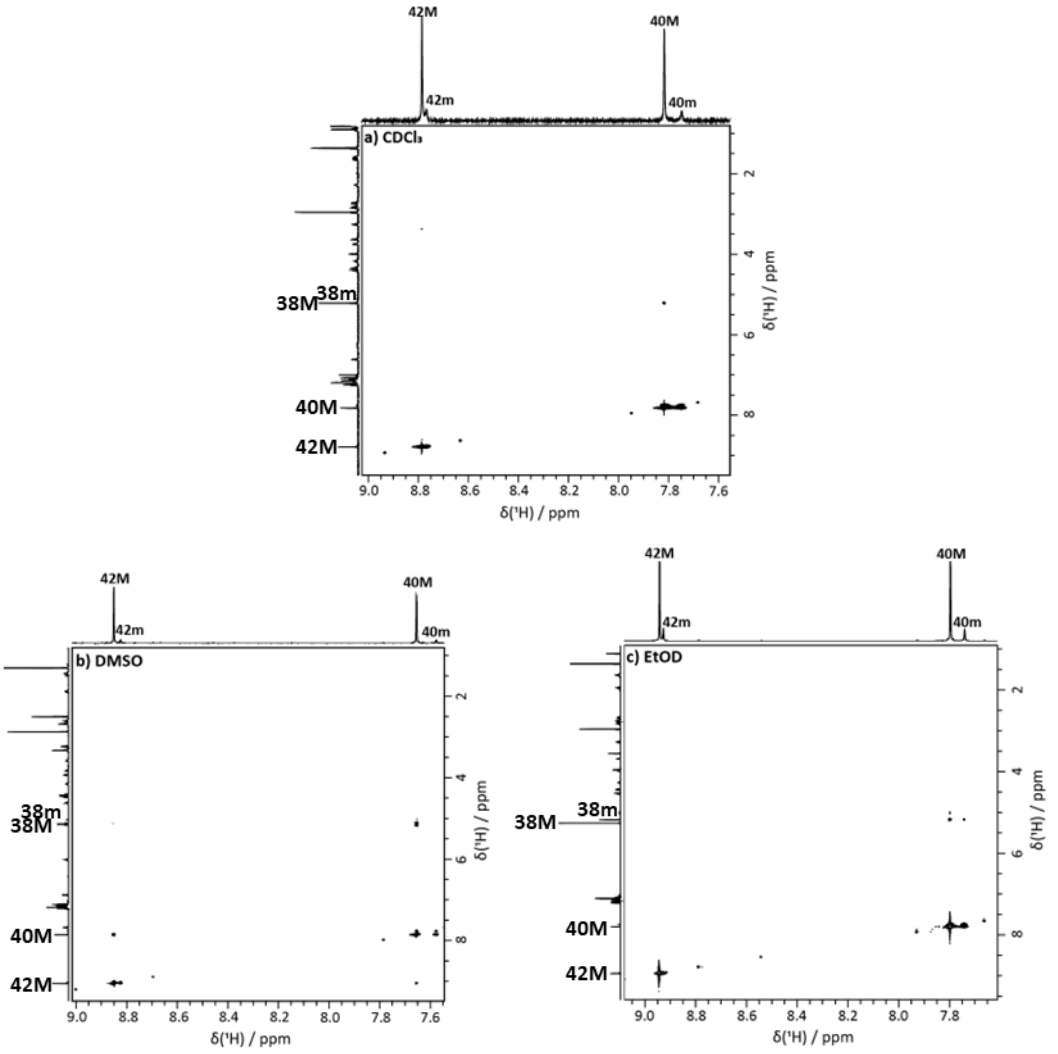


*Figure 6.6: Section of ^{1}H-^{1}H NOESY spectrum focussing on environments **42** and **40** of RVR in a) $CDCl_3$, b) DMSO and c) EtOD. There is an absence of NOEs (bar the expected **38-40** NOE), which rules out a dominant closed conformation. Environments have been labelled in the direct dimension.*

Additionally, there is significant change in ^{1}H chemical shift of the NH environment **34** of the CRB torsion. In figure 6.7, the ^{1}H chemical shift variation in DMSO ($\Delta\delta$ = 0.47 ppm) and EtOD ($\Delta\delta$ = 0.52 ppm) is significant whilst in $CDCl_3$, the variation is small ($\Delta\delta$ = 0.07 ppm). However, to further understand the behaviour of the CRB torsion, the surrounding environments required consideration. Environments of particular interest were **22**, **23**, **24** and **26** which are herein named the 'quasi-alkyl' chain and the focus was centred on the dominant conformation.

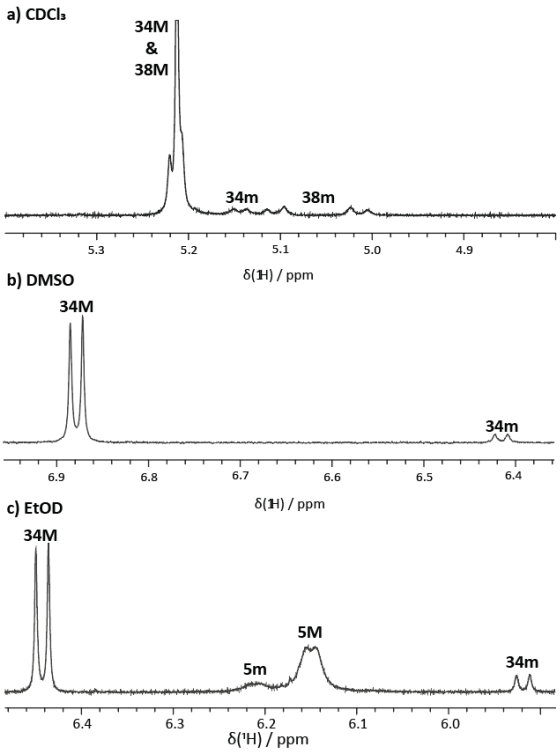


*Figure 6.7: Variation in ^{1}H chemical shift of **34** in $CDCl_3$ a), DMSO b) and EtOD c) with relevant environments labelled.*

Environments 22, 23, 24 and 26 in the Dominant Conformation – The 'quasi-alkyl' chain.

The environments **22**, **23**, **24** and **26** link together the CRB and NMU regions via an additional amide torsion. Wang et al. had previously identified their involvement in two of the four torsions that noticeably vary between the solid state forms I and II.[239]

In relation to the dominant conformation of the CRB torsion, it was found that the 'quasi-alkyl' chain plays a pivotal role in the preference of the CRB configuration. Based on the molecular structure in figure 6.3, it would be expected that the environments would show local NOEs with their neighbouring environments. But, as shown in figure 6.8, this was not the case. In DMSO and EtOD, a variation in NOE intensity was seen between the NOEs of **22** and **26** (**22-26**) and **22** and **24** (**22-24**).

It was expected that a stronger NOE would be seen for **22-24** compared to **22-26**. However, as shown in the red boxes of figures 6.8 b) and 6.8 c), the NOE for **22-26** is greater. This means that the 'quasi-alkyl' chain has a clear conformational preference that allows **22** and **26** to be close in space to one another. In contrast, in $CDCl_3$, the **22-24** NOE is stronger than the **22-26** NOE, which is expected. Therefore, the configurational preference is different in $CDCl_3$ relative to DMSO and EtOD.

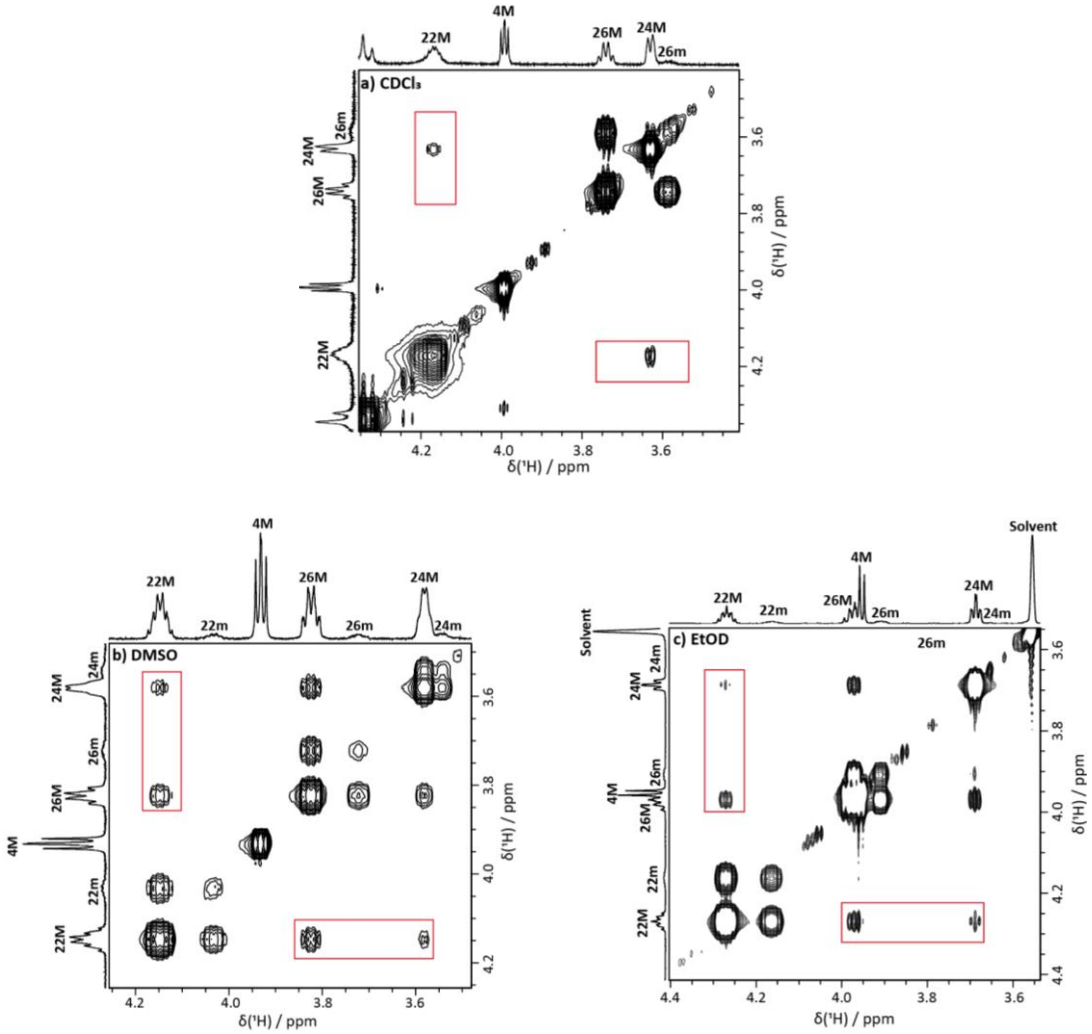


*Figure 6.8: ^{1}H-^{1}H NOESY spectra of RVR in a) $CDCl_3$, b) DMSO and c) EtOD with red boxes used to highlight the NOEs of interest between **22-26** and **22-24.** Environments have been labelled on both axes.*

This can be quantified further by extracting the internuclear distances from the NOE intensities, as described previously in Section 2.3.1. In order to estimate the distances for the **22-24** and **22-26** NOEs, a reference intensity and distance needed to be defined. In DMSO and EtOD, **44** was used (H-H = 1.78 Å) whilst **10** was used in $CDCl_3$ (H-H = 1.76 Å). The reference H-H distances were based upon the distance between the relevant CH_2 protons in the geometry optimised structure of RVR Form II (**YIPGIO03**). The relevant internuclear distances have been presented. Alongside this, the strong NOE of **5-9** in DMSO was included to demonstrate what internuclear distances are expected for a strong NOE.

Table 6.3 : Experimentally Derived Interproton Distances for the 22-23-24-26 chain.[18]

NOE Correlation	Experimentally Derived Distance / Å		
	DMSO	EtOD	$CDCl_3$
22-26	2.7	2.9	3.9
22-24	3.3	3.3	2.9
23-22	2.8	3.0	3.0
23-24	2.9	3.2	3.0
5-9	2.6	-	-

In table 6.3, the H-H distance for the **22-26** NOE is almost 1 Å shorter in DMSO and EtOD relative to $CDCl_3$. In $CDCl_3$, the environments are not close in space at all. The derived distances in table 6.3 can then be used to model the region to determine the preferred CRB configuration in the dominant conformation.

Regional Models of the 'quasi-alkyl' Chain and CRB Torsion

To understand how this difference in the quasi-alkyl chain configuration affects the overall conformation, models of the specific region were created in Avogadro. As shown in figure 6.9, the variation in the NOEs observed for the 'quasi-alkyl' chain report that the conformation has an impact on the surrounding moieties, which, in turn, affect the preference for the CRB configuration. This can play a role in the magnitude of the chemical shift difference observed for environment **34**.

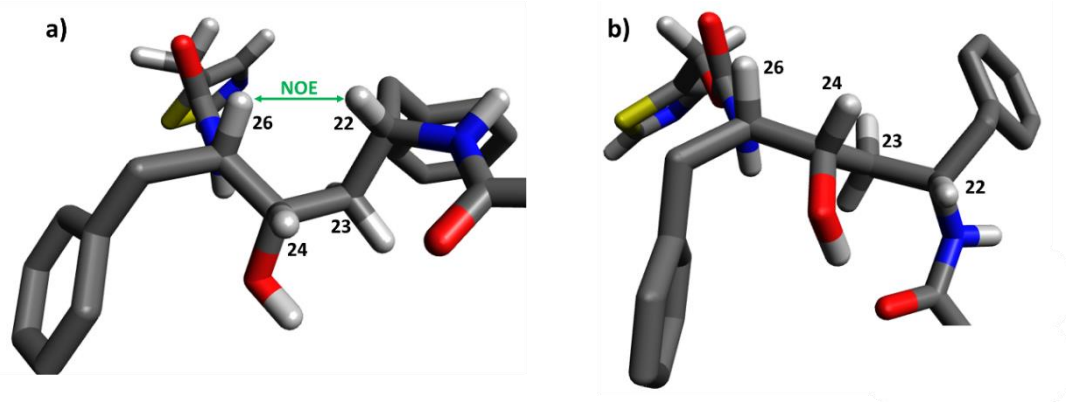


*Figure 6.9: Variation of the 'quasi-alkyl' chain when a) **22** and **26** are close in space (as in DMSO and EtOD) whilst b) is when **22** and **26** are not close in space (as in $CDCl_3$).*

[18] *Errors were determined based on references 64 and 65 where the % error quoted was approximately 3%.*[64,65] *In this instance, this would be around ± 0.1 Å.*

Focusing first on the 'quasi-alkyl' chain in DMSO and EtOD, figure 6.10 captures the two possible configurations for the CRB torsion. Arrows have been marked on bonds which are expected to exhibit some or complete rotational freedom, based on table 6.2, and the type of moiety. Within the CRB region, the thioazole ring (containing environments **40** and **42**) is expected to be under free rotation as are the phenyl rings which are connected to the main body of RVR by the CH_2 groups of **27** and **44**. **27** and **44** do have associated inequivalence for the CH_2 protons, as seen in table 6.2, but will have the ability to rotate whilst maintaining the inequivalence. The same is also seen for **38** where inequivalence is seen for the dominant conformation in DMSO whilst equivalence is seen in EtOD. These will all play a part in the understanding of the overall CRB configuration.

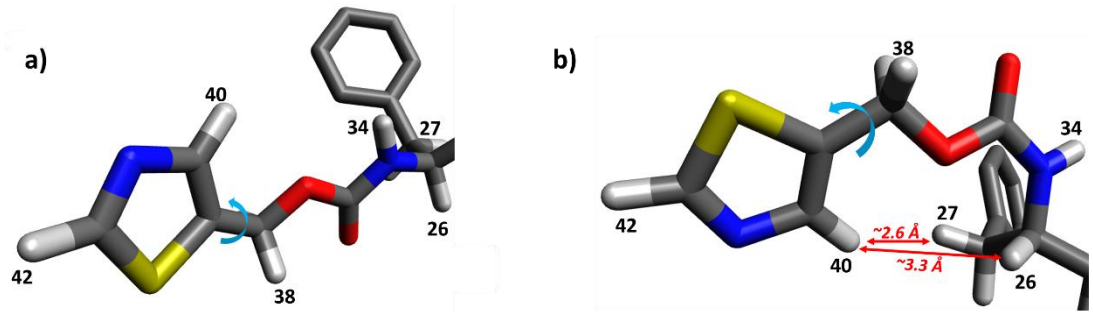


Figure 6.10: Visualisation of the CRB torsion when the 'quasi-alkyl' chain is arranged as shown in figure 6.9 a) in DMSO and EtOD. Here, the configuration of the CRB torsion is trans in a) and cis in b). The relevant environment labels are in black, possible rotatable bonds are in blue and approximate H-H distances for possible long-range NOEs are in red. The trans CRB configuration for the dominant solution conformation is in better agreement with the previously presented acquired data. The phenyl ring protons have been omitted for clarity.

When the CRB configuration is trans, as in figure 6.10 a), environments **40** and **42** are not expected to observe any long-range NOEs and this agrees with the acquired data in figure 6.6. In contrast, when the CRB configuration is cis, environment **40** would likely observe an NOE with **26** and **27**. The approximated distances fall within the range of those calculated in table 6.3 so would be expected to be observed in the acquired data. However, they were not, so it can be said that the dominant CRB configuration in DMSO and EtOD is trans.

When the 'quasi-alkyl' chain is modelled as expected in $CDCl_3$, similar observations can be made. In the trans CRB configuration, environments **40** and **42** do not observe any long-range NOEs whilst in the cis CRB configuration, there is the possibility of an NOE between **26** and

40 but this is not seen in the experimental data. The NOE between **40** and **27** is unlikely, as displayed in grey in figure 6.11 b), as its H-H distance exceeds the 3.5 Å cut-off described by the methodology.[64,65] Therefore, the CRB torsion is trans in the dominant conformation, despite the variation in the 'quasi-alkyl' chain in $CDCl_3$.

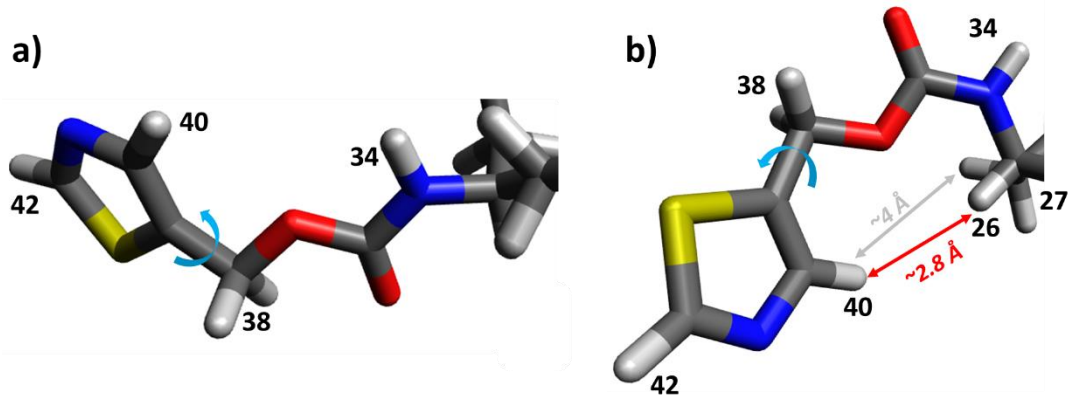


Figure 6.11: Visualisation of the CRB torsion when the 'quasi-alkyl' chain is arranged as shown in figure 6.9 b) in $CDCl_3$. Here, the configuration of the CRB torsion is trans in a) and cis in b). The relevant environment labels are in black, possible rotatable bonds are in blue and approximate H-H distances for possible NOEs are in red and grey.

Despite its lack of impact on the CRB torsion, the conformation of the 'quasi-alkyl' chain can be explained by the variation seen for environment **34**. As discussed previously, the chemical shift difference between **34M** and **34m** in $CDCl_3$ was significantly less relative to DMSO and EtOD. This variation appears to be independent of the CRB torsional preference so an initial calculation to determine its likelihood to engage in intramolecular hydrogen bonding was evaluated (Section A6 of Appendix), but this found no possibility.[249–251] Interestingly, the NH environment **5** indicated that there was the possibility of an intramolecular hydrogen bond (intraHB) but in order to investigate its plausibility, further work would be needed with molecular dynamics (MD) simulations to understand the significance of the result. This is also the case with the possibility of the conformational preference of the 'quasi' alkyl chain and resulting chemical shift variation of environments such as **34**. The experimental data indicate this possibility but in order to be confident, it would require MD simulations to determine the plausibility of such an effect.

Summary

The discussed conformational preferences for RVR in each solvent have been tabulated in table 6.4. In EtOD and $CDCl_3$, the dominant conformation reflects the solid state Form I conformation whilst, in DMSO, the dominant conformation is more similar to Form III. It is

clear that there is an associated solvent dependence in regard to the behaviour of RVR in solution and that the CRB torsion appears independent of this. The solvent dependence of RVR has been discussed further with lineshape analysis in the following section.

Table 6.4: Summary of observations for the dominant conformation of RVR in each solvents.

Solution-State Solvent	CRB Torsion	NMU Torsion	'Quasi-Alkyl' Chain (NOE Strength)	34 ^{1}H Chemical Shift Difference / Δppm
$CDCl_3$	Trans	Cis	**22-24** > **22-26**	0.07
DMSO	Trans	Trans	**22-26** > **22-24**	0.47
EtOD	Trans	Cis	**22-26** > **22-24**	0.52

6.4.2 Lineshapes

A subset of environments (**38**, **9** and **1&3**) have been chosen to demonstrate how particular lineshapes in DMSO differ from $CDCl_3$ to EtOD. This can be correlated to the preference of the NMU torsion, as seen in table 6.4. Figure 6.12 collates together the environments considered, where those in italics refer to environments that are either associated with the NMU region (**9**) or neighbour the torsion (**1&3**) whilst **38** is involved in the CRB torsion.

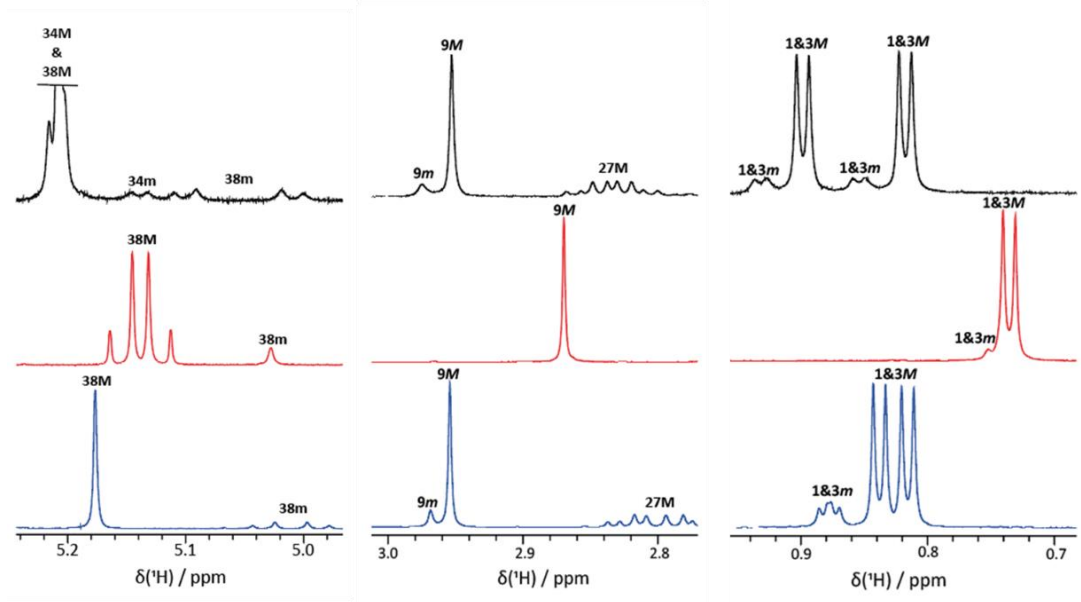


*Figure 6.12: Lineshape variation for environments **38** (left), **9** (middle) and **1&3** (right) in $CDCl_3$ (black), DMSO (red) and EtOD (blue). **27M** in DMSO is not observed in the middle plot as its ^{1}H chemical shift (2.68 ppm) falls out of the axis range whilst **9** and **1&3** are in italics due to their association with the NMU region of RVR.*

Across the subset of environments, the ^{1}H chemical shift in DMSO is consistently lower relative to $CDCl_3$ and EtOD. The distinction of environments, such as **9**, vary in DMSO relative to $CDCl_3$ and EtOD, where the resolution of the minor signal is seen in the latter two solvents. Alongside this, **38** and **1&3** exhibit different lineshapes, indicating a variation in mobility and molecular shape. The possible solvent dependence on molecular shape has been touched upon previously with the variation in the 'quasi-alkyl' chain and its subsequent influence on the ^{1}H chemical shift of **34**.

The ^{i}Pr group, **1&3**, was particularly intriguing as the lineshapes observed in $CDCl_3$ and EtOD were unexpected. It was expected that free rotation of the two Me groups would be seen, leading to equivalence with a lineshape of a simple doublet (due to coupling to the neighbouring CH). In RVR, this would correspond to free rotation about the **C2-C4** axis. This expected doublet lineshape is seen in DMSO but, in $CDCl_3$ and EtOD, a pair of doublets is observed instead, indicating that the individual Me groups are inequivalent. The CH_2 group, **38**, also exhibited inequivalence despite its direct involvement in the CRB torsion and, more interestingly, the inequivalence varied between conformations. In the dominant conformation in DMSO, **38** is equivalent but becomes inequivalent in the minor conformation. The opposite relationship is observed in $CDCl_3$ and EtOD where the dominant conformation saw equivalence whilst the minor conformation was inequivalent. As the CRB torsion, as summarised in table 6.4, is suspected to be trans in the dominant conformation, the variation could possibly be linked to the NMU torsion but to reinforce this possibility, extensive MD studies would be required.

6.4.3 Section Summary

In this section, the conformational behaviour of RVR in $CDCl_3$, DMSO and EtOD has been investigated. Solvent dependence plays a significant role in the conformation of RVR where variations are observed across the solvents. The configurational preference of the NMU torsion in the dominant conformation in DMSO differs from $CDCl_3$ and EtOD where this variation appears to be conserved in resulting lineshapes for a selection of environments.

Solvent also plays a critical role in the conformational preference of the 'quasi-alkyl' chain where NOEs clearly show that **22** and **26** are close in space to one another in DMSO and EtOD. As the 'quasi-alkyl' chain also contains an OH group (bonded to carbon **24**), the behaviour of the surrounding environments may be dictated by hydrogen bonding ability between the OH and the solvent. The OH group in the solid state plays a key role in the more stable Form II, where it behaves as both HBD and HBA, resulting in stronger overall hydrogen

bonding relative to Form I.[18,19] It is possible a similar effect is occurring in solution, resulting in a preferred orientation of the environments in DMSO and EtOD relative to $CDCl_3$.

To comprehensively determine the preferred conformation of RVR in each solvent, extensive MD simulations would be required. This would help fully capture the complexity and flexibility associated with RVR. The work presented in this section is only the beginning in the understanding of its behaviour in solution and how solvent plays a role in its conformational preference. Critically, this section has identified the torsional preferences of NMU and CRB torsions which can be probed further to determine kinetic information. As there are two distinct sets of signals in the acquired datasets, this indicates that the exchange is slow on the NMR timescale. Therefore, it can be probed quantitatively using ^{1}H-^{1}H EXSY spectra.

6.5 Exchange

6.5.1 Introduction to Exchange Spectroscopy (EXSY)

The ^{1}H-^{1}H EXSY experiment was first proposed by Jeener and co-workers[70] and allows for a visual representation of exchange processes that are slow on the NMR timescale.[67,68] For one set of exchange peaks, acquisition and analysis of the 1D ^{1}H EXSY spectrum is simpler, where a plot of intensity against delay time, can output the relevant exchange rate.[252] However, the benefit of the 2D ^{1}H-^{1}H EXSY experiment is that multiple exchange pairs can be captured and quantified in a single experiment.

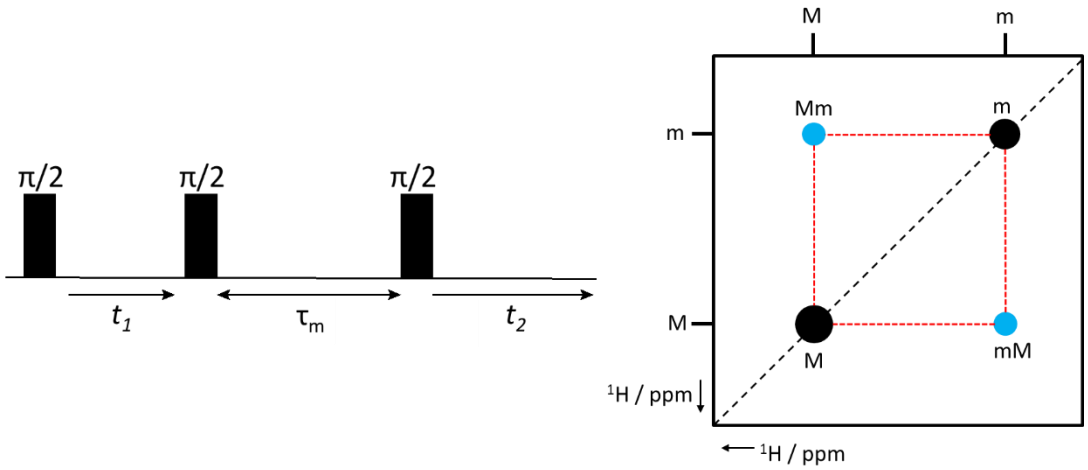


Figure 6.13: ^{1}H-^{1}H EXSY pulse sequence with a schematic of the resulting 2D ^{1}H-^{1}H EXSY spectrum. The black dots reflect the diagonal peaks of the major (M) and minor (m) components alongside the resulting off-diagonal cross peaks (mM and Mm) with their associated labels when the exchange cross peak pair are viewed along the direct dimension (horizontally). The size of the cross peaks reflect the approximate intensities.

For effective acquisition of data, the choice of mixing time (τ_m) is key. If τ_m is too short, inefficient magnetisation transfer occurs whilst too long a duration would result in signal decay due to T_1 relaxation dominating.[252,253] Perrin and Dwyer defined the optimal mixing time using equation 6.3, where T_1 is the longitudinal relaxation time constant and k is the sum of the forward (k_{Mm}) and reverse (k_{mM}) exchange rate constants. In the context of this work, the forward reaction (k_{Mm}) primarily refers to the interconversion of the CRB torsion from trans (M) to cis (m) and K refers to the equilibrium constant; This has been used to determine whether the outputted k_{Mm} values are realistic relative to the experimental conformer ratios.

$$\tau_m(\text{optimal}) = \frac{1}{\frac{1}{T_1} + k} \qquad \text{Eq. 6.3}$$

6.5.2 Exchange Calculations

Integration Choice

In order to calculate kinetic parameters, the intensities of the exchange pair cross peaks are required.[19] Two methods have been used. The first was direct integration of the 2D cross peaks whilst the second method was the extraction of 1D slices from the diagonal cross peaks of the exchange pairs. The latter method reflects what was used previously to extract internuclear distances from ^{1}H-^{1}H NOESY spectra.[64,65]

Slice integration was beneficial for signals where overlap was observed. Environments **46&50** and **24** in DMSO have been used as examples of this in figure 6.14 where one of the signals in the exchange pair is overlapped either with itself (**24**) or with another environment (**46&50**). The 1D cross sections can distinguish the signal but does not guarantee realistic K values. This is discussed further in Section 6.5.3.

[19] *The ^{1}H-^{1}H EXSY spectra in EtOD were acquired in two separate batches with two reference spectra. As the intensities were sufficiently similar, the reference intensities were averaged and evaluated as normal.*

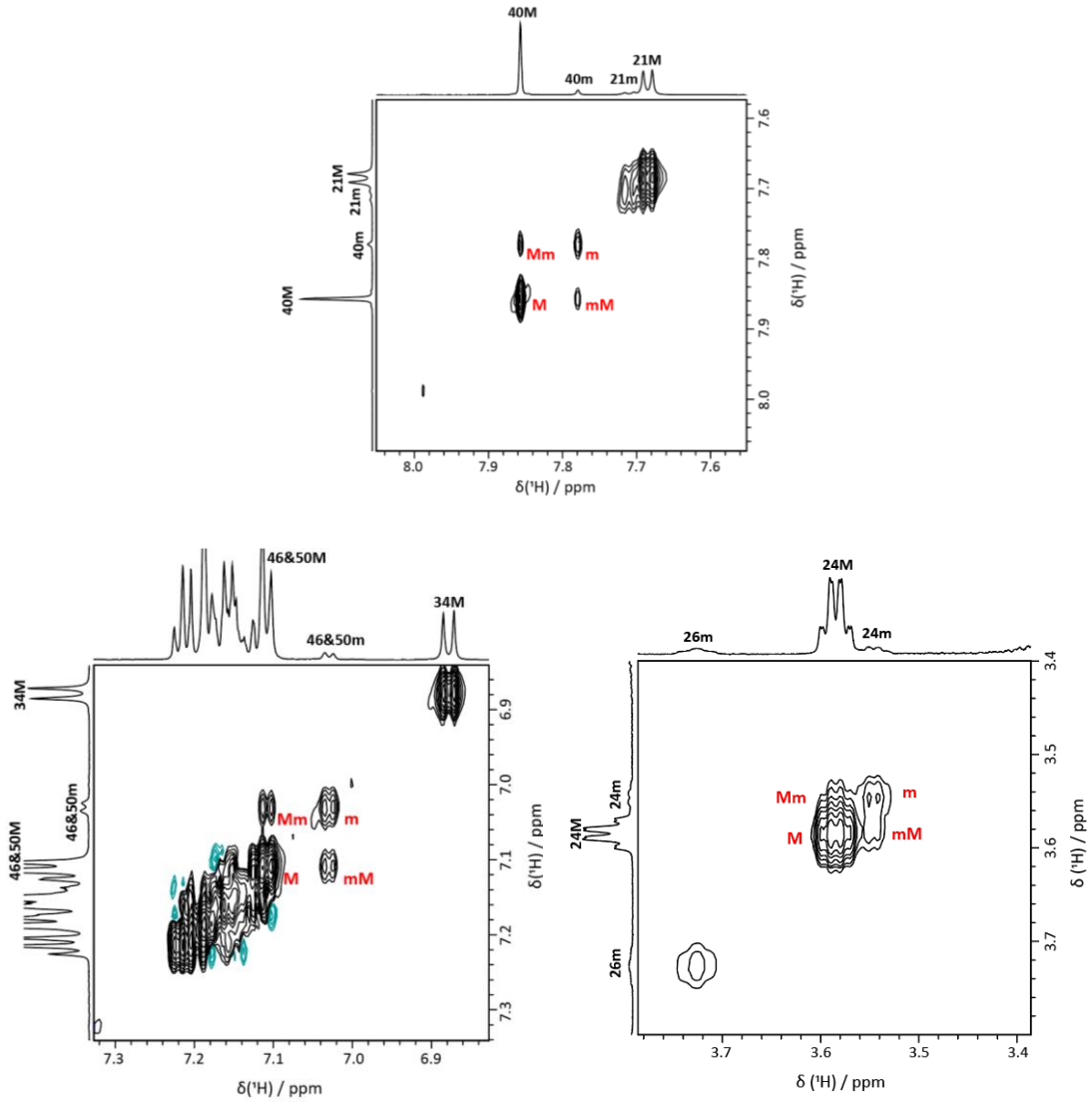


*Figure 6.14: ^{1}H-^{1}H EXSY spectra for RVR in DMSO at τ_m of 0.1 s of environment **40** (top), **46&50** (left) and **24** (right) where the exchange pair has been labelled in red with environment labels on both axes.*

Slice integration was the preferred method to extract intensities but when the shape of the peak is different within the exchange pair, the slice taken cannot be considered meaningful. This is shown in figure 6.15 for **38** in both DMSO and EtOD. The inequivalence of the CH_2 protons results in a pair of cross peaks for one environment, such as the **38M** and **38Mm** in DMSO and **38m** and **38 mM** in EtOD. Therefore, for environment **38**, volume integration was the better choice whilst, for the other environments, slice integration has been used.

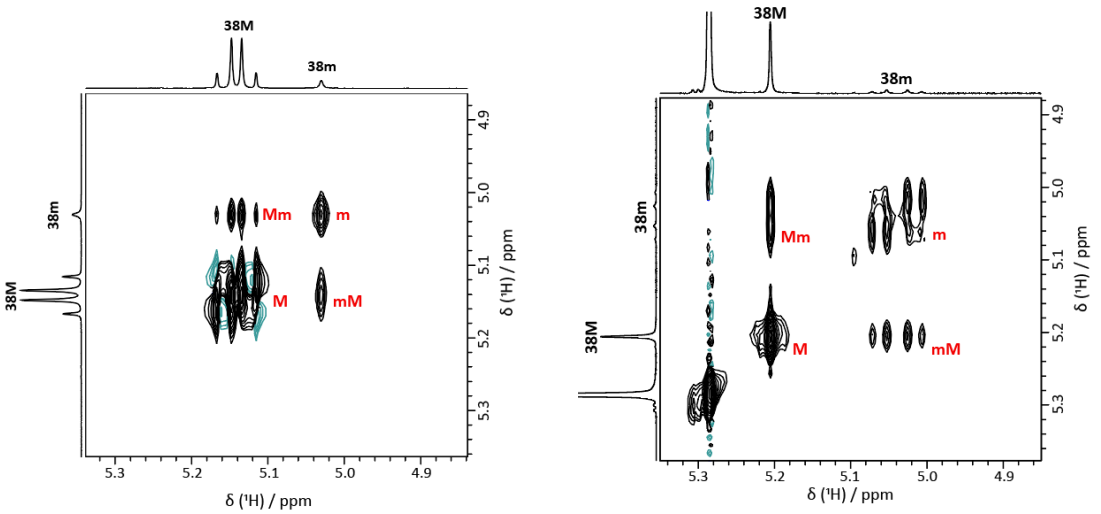


*Figure 6.15: ^{1}H-^{1}H EXSY spectra for RVR in DMSO (left) and EtOD (right) at τ_m of 0.1 s of environment **38** where the exchange pair has been labelled in red with environment labels on both axes. Cross peaks of interest are **38M/38Mm** in DMSO and **38m/38mM** in EtOD.*

Approaches Used

Within the literature, there are three broad categories of analysis that can be used to extract kinetic information: the initial rate approximation, iterative fit (IF) and direct matrix analysis (DMA).[67,252] The two most common approaches are IF and DMA so have been evaluated in this work where the theory has been discussed previously in section 2.3.

The free software, EXSYcalc uses the DMA approach,[254,255] as does an earlier programme, D2DNMR.[256] EXSYcalc has become a frequently used resource within the literature to estimate exchange rates in organic and inorganic materials.[257–262] The method requires the acquisition of two ^{1}H-^{1}H EXSY spectra, one at zero mixing time ($\tau_m = 0$) and one at a non-zero mixing time ($\tau_m = t$). It is essential that the mixing time chosen is appropriate otherwise the results will be widely over- or under-representative.[72] It can be difficult to identify whether a mixing time is appropriate as the EXSYcalc implementation of DMA does not consider experimental noise. Therefore, the uncertainty estimates are not provided with the quoted values. To determine whether the calculated kinetic information is reliable, an additional step is required.

In contrast, IF considers experimental uncertainty[20]. The uncertainty estimates are automatically calculated from the non-linear fitting of the inputted data to generate the kinetic parameters.[76] For IF, at least two non-zero mixing times datasets are preferred as the

[20] *The term 'noise' has also be used to describe the uncertainty in this work.*

greater number of data points considered, the more realistic the fit (assuming all choices of τ_m are appropriate).

Table 6.6: Choice of mixing time and number of environments considered.

Solvent	Number of Environments	Mixing Time / s
$CDCl_3$	2	0.05, 0.1, 0.2, 0.5
DMSO	8	0.1, 0.2, 0.3
EtOD (VT)	6	0.05, 0.1(+VT), 0.2

As mentioned previously, the choice of mixing time is important.[252,253] The mixing times used are tabulated in table 6.6. The IF approach considers all the supplied experimental data in one calculation so the resulting intensity against mixing time plots can be used to determine whether mixing times are suitable for DMA. From the intensity build-up curves in figure 6.16, the 0.5 s datapoint in $CDCl_3$ could be discarded. There is no information to be gained about the exchange kinetics as the intensities begin to plateau so can be omitted from the subsequent DMA calculation.

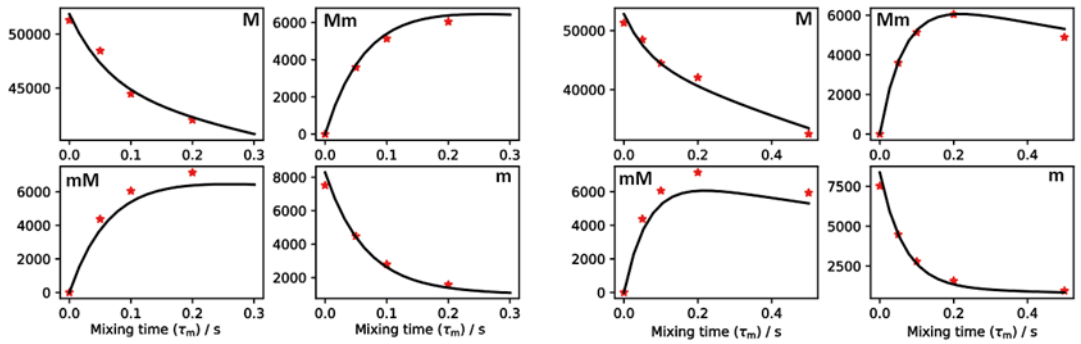


*Figure 6.16: IF intensity curves for environment **40** where the 0.5 s datapoint has been excluded (left) or included (right).*

Approach to Uncertainties

Both IF and DMA were used to calculate kinetic parameters for mainly the interconversion of the CRB torsion but also the NMU torsion with environment **5** in EtOD. However, in order to estimate the uncertainties for the DMA values, a Monte Carlo analysis (MC)[21] has been

[21] *Monte Carlo (MC) analysis is a computational method that uses random sampling to determine the likelihood of a result occurring.*[275,276]

used. MC was used to estimate the associated error of the calculated kinetic parameters by retrospectively determining the uncertainty by adding random noise to the calculated data. The calculation is then repeated around 1000 times to determine the standard deviation of the calculated parameters. However, in order to use MC, an input noise level was needed. It was important to define a noise level that was realistic to avoid under or over estimating the associated error as the error scales with the noise level.

Initially, the noise level was based on the random noise of the acquired experimental data where the random noise was estimated from the standard deviation of the intensities of a signal-free region of the 1D ^{1}H cross section for each exchange pair in the reference ^{1}H-^{1}H EXSY spectrum. As the SNR was high, this resulted in values that were underestimated as the magnitude of the signal to random experimental noise was around 1 to 10000. Alternatively, the difference in the non-diagonal intensities can be used, which should combine any random error with systematic error in determining the integrals. This resulted in a more realistic SNR magnitude of around 1 to 100, where the ratio reflects the sum of the diagonal intensities at $\tau_m = 0$ against the difference in the off-diagonal intensities at non-zero ($\tau_m = t$) mixing times. However, in this work, the residuals associated with the iterative fitting in IF were used instead to estimate the noise level.

Noise Levels – Specific or Average?

As more than two non-zero mixing times are considered for each solvent, the noise levels could be determined by the residuals associated with the fitting. However, for DMA, they can either be incorporated specific to each environment or averaged across the solvent. Figure 6.17 presents the MC distribution plots from the DMA calculation for two environments in DMSO, **40** and **25** at 0.1 s mixing time. The corresponding kinetic information has also been annotated within the relevant plot.

It was found that when the average noise level is significantly different to the environment specific noise: as seen for **40**, the resulting MC distribution plots become skewed, resulting in significant error estimates, as demonstrated in figure 6.17 a) where the average noise level resulted in an error estimate of K three times the magnitude relative to when the specific noise level is used. This is not representative as **40** is considered a well-resolved exchange pair, as seen in the direct dimension of figure 6.6, so the associated noise level will be small. However, when the average noise level is similar to the environment specific noise level, as seen for **25**, the resulting MC distribution plots reflect one another. In this work, to avoid unrealistic error estimates, environment specific noise levels have been used. The skew of

the MC distribution plot has also been used as a characteristic shape to indicate an ambiguous calculation of kinetic parameters.

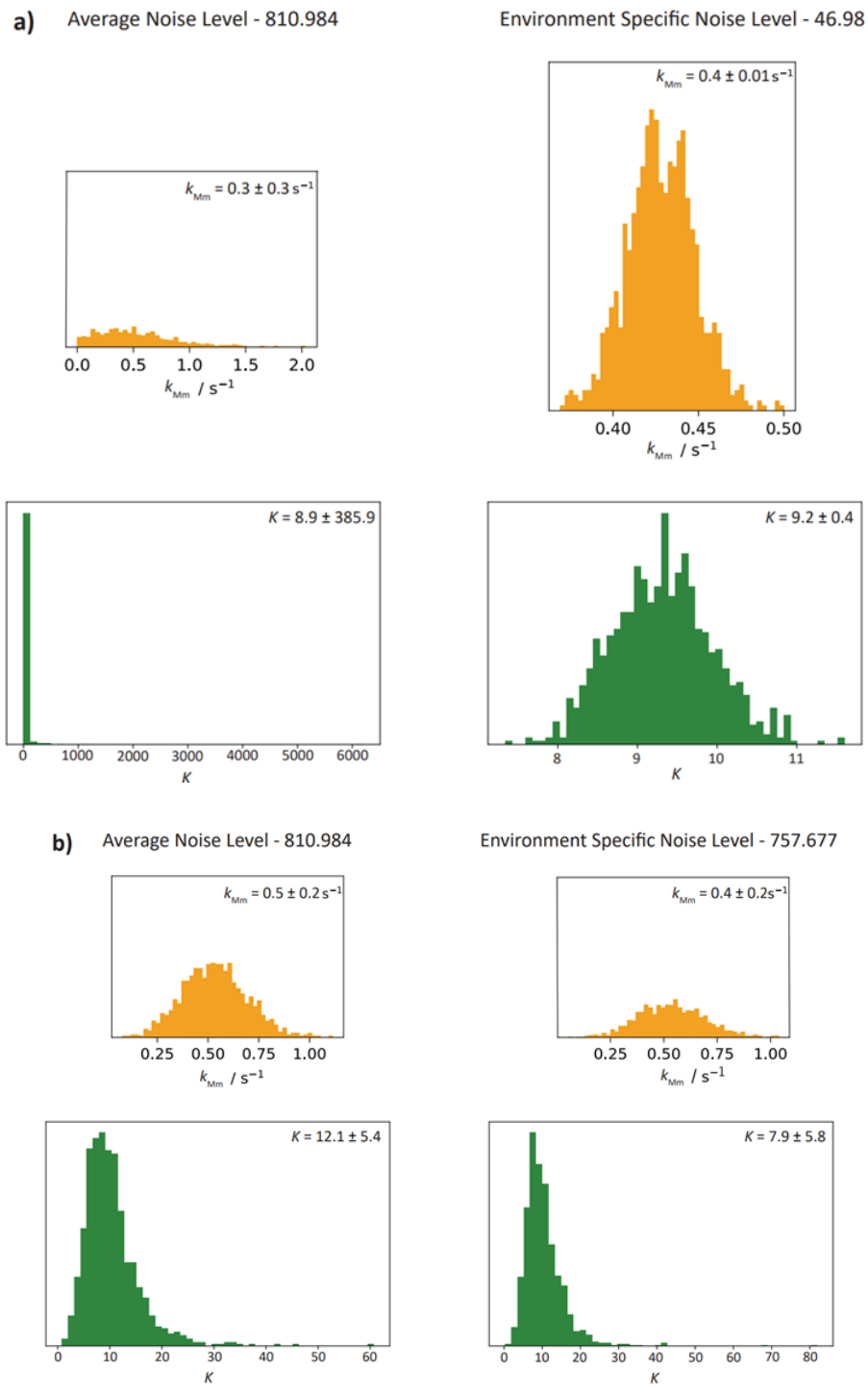


*Figure 6.17: Resulting MC distribution plots for a) **40** and b) **25** at τ_m = 0.1 s. The top distribution plots are when environment specific noise levels are used **40** (noise level = 46.98) and **25** (noise level = 757.677) whilst the bottom distribution plots are for the same calculation where instead an average noise level has been used (noise level = 810.984). The expected shape of the distribution is a bell-shaped curve.*

6.5.3 Solvent Dependence On Exchange

Results

The IF approach was used first and the resulting values for k_{Mm} and K have been tabulated in table 6.9. The inputted intensities have been determined using slice integration except for **38**, where volume integration has been used. Values of K can be compared to the approximate conformer ratios in table 6.2 to determine whether the outputted kinetic parameters are realistic.

Table 6.9: Summary of the calculated values of K and k_{Mm} for RVR in $CDCl_3$, DMSO and EtOD using the IF methodology.[22]

Solvent	Environments	K	k_{Mm} / s^{-1}
$CDCl_3$	40	6.3 ± 0.3	2.0 ± 0.3
	26	5.7 ± 0.5	1.5 ± 0.3
DMSO	46&50	12.7 ± 1.0	0.3 ± 0.05
	40	10.0 ± 0.1	0.4 ± 0.01
	38	11.2 ± 0.2	0.4 ± 0.02
	34	10.5 ± 1.2	0.4 ± 0.1
	26	10.2 ± 1.4	0.4 ± 0.1
	25	10.2 ± 0.7	0.4 ± 0.1
	24	8.8 ± 1.0	0.7 ± 0.2
	22	8.8 ± 0.7	0.4 ± 0.1
EtOD	46&50	14.9 ± 1.2	0.2 ± 0.05
	40	5.8 ± 0.1	0.6 ± 0.03
	38	6.1 ± 0.1	0.6 ± 0.03
	22	5.9 ± 0.5	0.6 ± 0.1
	5*	6.2 ± 0.9	1.1 ± 0.4

Using the noise levels calculated from the residuals, the DMA calculation could be completed. DMA outputs a distinct set of parameters for each mixing time considered therefore for each solvent, three sets of kinetic parameters were calculated. To note, the

[22] ** refers to an environment which is associated with the NMU torsion. The related uncertainties are estimated using the residuals associated with the iterative fitting of the data points.*

datapoint at 0.5 s for $CDCl_3$ was omitted due to the plateau observed in figure 6.16. The values of K have been presented in table 6.10.

Table 6.10: Summary of the calculated values of K for RVR in $CDCl_3$, DMSO and EtOD from DMA. The associated uncertainties were determined from the environment specific noise level derived from the iterative fitting.[23]

Environment	*K*		
	Mixing Time / s		
$CDCl_3$	**0.05**	**0.1**	**0.2**
40	3.8 ± 1.9	4.3 ± 1.5	8.2 ± 1.2
26	5.0 ± 3.9	4.0 ± 3.2	7.4 ± 2.6
DMSO	**0.1**	**0.2**	**0.3**
46 & 50	5.1 ± 28.8	9.6 ± 7.1	11.2 ± 4.9
40	9.4 ± 0.6	9.2 ± 0.4	10.0 ± 0.4
38	8.1 ± 1.7	12.9 ± 1.4	10.9 ± 1.2
34	23.7 ± 147.9	6.01 ± 12.1	23.8 ± 9.7
26	19.1 ± 62.4	26.2 ± 70.9	2.6 ± 47.3
25	7.9 ± 5.8	17.0 ± 3.8	11.2 ± 3.6
24	6.1 ± 17.9	18.3 ± 11.7	11.5 ± 161.7
22	5.9 ± 104.5	50.6 ± 7.5	11.9 ± 8.2
EtOD	**0.05**	**0.1**	**0.2**
46 & 50	7.9 ± 244.2	14.3 ± 391.8	27.6 ± 10.5
40	6.9 ± 1.9	4.6 ± 1.0	5.3 ± 0.6
38	11.3 ± 2.1	5.3 ± 1.0	6.2 ± 0.7
34	1.4 ± 53.7	6.7 ± 222.7	1.7 ± 50.7
22	3.1 ± 31.5	10.6 ± 57.2	4.7 ± 41.7

Despite the majority of the resulting values of k_{Mm} (Section A7) being in great agreement with the values in table 6.9, the values of K indicate serious deviations in the DMA approach. The approximate conformer ratios in each solvent (to the nearest integer) were 7:1 for $CDCl_3$, 11:1 for DMSO and 6:1 for EtOD and there is no consistency of the values in table 6.10. Alongside this, the magnitude of the associated uncertainties are considerably larger for the

[23] *Associated k_{Mm} values can be found in Section A7 of the Appendix.*

majority of the environments relative to the IF values. The appearance of such large uncertainties can be characterised by the one sided skew from the expected bell-shaped curve in the related MC distribution plots as seen previously for environment **40** in figure 6.17. Therefore, in this work, the IF approach has been used to evaluate exchange.

A visual summary of the kinetic parameters calculated for the CRB torsion is shown in figure 6.19 where two environments have been omitted from the plot (and subsequent average bars). These were **5** and **46&50**. Environment **5** is associated with the NMU torsion so was not appropriate to include whilst **46&50** suffered a particular type of overlap, as shown in figure 6.14, that resulted in skewed K values despite realistic k_{Mm} values. For the remaining environments considered, they were acceptable as they achieved desirable attributes: The exchange pair are well-resolved from one another (an example is **40**) and if there is overlap, the overlap only concerns the exchange pair evaluated (as seen for **24**).

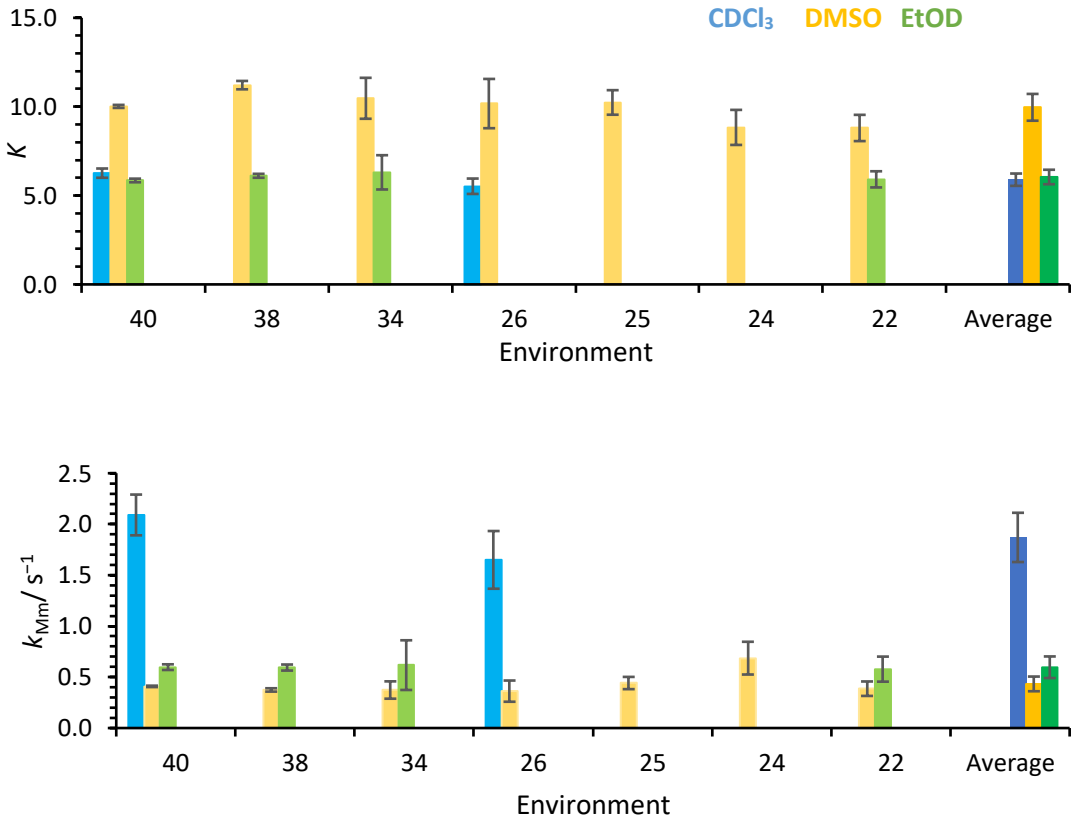


*Figure 6.19: Values of K (top) and k_{Mm} (bottom) obtained from iterative fitting and slice integration for each resolved environment with the exception of **38**, where the value of K was obtained from volume integration. The bars in blue refer to $CDCl_3$, those in yellow refer to DMSO and those in green are for EtOD in green. Darker bars on the right show the average value and associated standard error.*

The variation in K for **46&50** has been summarised in table 6.11 where it was hypothesised that variation of K was the result of the imbalance of intensities, as the diagonal cross peak, **46&50M**, has the additional contribution of the unrelated signal to the integral intensity. This then contributes to the deviation.

*Table 6.11: Values of K and k_{Mm} for environment **46&50** in DMSO and EtOD. Errors have been estimated from the residuals of the iterative fitting.*

Solvent	Conformer Ratio (nearest integer)	K	k_{Mm} / s^{-1}
DMSO	11:1	12.7 ± 1.0	0.3 ± 0.05
EtOD	6:1	14.9 ± 1.2	0.2 ± 0.05

To test this hypothesis, an artificial offset was applied to the well-resolved exchange pair of **40**. This was to determine whether overlap with another (unrelated) environment would skew the calculated K value without severely affecting the resulting k_{Mm} value. As shown in table 6.12, an increase of the **40M** diagonal cross peak intensity resulted in an increase in the K value, whilst the k_{Mm} values maintained consistency. Therefore, it was clear that an unwanted contribution was occurring so **46&50** was omitted from the final value.

*Table 6.12: Variation in calculated K and k_{Mm} values of environment **40M** diagonal cross peak when an offset was used.*

Solvent	Original		Offset		Difference	
	K	k_{Mm} / s^{-1}	K	k_{Mm} / s^{-1}	K	k_{Mm} / s^{-1}
$CDCl_3$	6.3 ± 0.3	2.0 ± 0.3	6.9 ± 0.3	1.9 ± 0.2	0.6	0.1
DMSO	10.0 ± 0.1	0.4 ± 0.01	11.7 ± 0.1	0.3 ± 0.01	1.7	0.1
EtOD	5.8 ± 0.1	0.6 ± 0.03	6.2 ± 0.1	0.6 ± 0.03	0.4	0.0

The K and k_{Mm} of the remaining environments were consistent so average values could be found and are tabulated in table 6.13, where the unweighted average standard error was used. It is clear that $CDCl_3$ exhibits faster exchange between the CRB conformations relative to DMSO and EtOD, whose values fall within error with one another. The exchange discussed refers to the transformation of the trans CRB configuration to the cis CRB configuration in each solvent. In contrast, **5** in EtOD is suspected to refer to the transformation of the cis NMU configuration to the trans NMU configuration.

*Table 6.13: Average fitted rate coefficient and equilibrium constant for the suspected CRB interconversion in each solvent where **46&50** is omitted from the calculation.*

Solvent	$CDCl_3$	DMSO	EtOD
Average *K*	6.0 ± 0.4	10.0 ± 0.8	6.0 ± 0.4
Average k_{Mm} / s^{-1}	1.8 ± 0.3	0.4 ± 0.1	0.6 ± 0.1

Discussion

The values of K calculated have been used as a check to ensure that the calculations are representative of the experimental data and this was the case. Also, despite involvement in the NMU torsion, environment **5** has a value of K that agrees with the other environments considered.

Conversely, the value of k_{Mm} for **5** in EtOD was 1.1 ± 0.4 s^{-1}. This was larger than the average value calculated for the other environments in EtOD at 0.6 ± 0.1 s^{-1}. The values do fall within one standard error of one another so to confidently suggest two different rates of interconversion, additional environments would be needed. An initial analysis of environment **9** in both EtOD and $CDCl_3$ was attempted where the k_{Mm} values in table 6.14 do vary from the average in table 6.13. However, the associated error is relatively large so in order to improve on the accuracy of these calculations, clearer distinction of the intensities would be required. This could be achieved with initial optimisation of the mixing time to the NMU torsion but also via low temperature experiments to allow further distinction between the exchange pair cross peaks and to reduce line broadening. Nevertheless, the values of k_{Mm} do indicate a faster interconversion and is further evidence to indicate that NMU and CRB torsions have different rates of exchange and are thus independent of one another.

Table 6.14: K and k_{Mm} values calculated for environment 9 in $CDCl_3$ and EtOD.

Environment 9	K	k_{Mm} / s^{-1}
$CDCl_3$	5.0 ± 0.7	6.0 ± 3.1
EtOD	3.8 ± 0.8	4.7 ± 2.9

RVR does exhibit solvent dependence in regard to both its conformation and exchange. This is not unheard of for Beyond Rule of 5 (Bro5) molecules, as discussed by Sebastiano et al. where they describe a phenomenon of chameleonicity.[263–265] RVR is classed as a Bro5 molecule as it violates at least one of Lipinski's criteria.[265] In fact, RVR violates three of the

four criteria as its molecular weight is greater than 500, there are more than 10 HBA and its octanol-water partition coefficient, log(*P*), exceeds 5.[239,265] This also results in its classification as a Class IV drug compound where it displays both low solubility and low permeability.[266] In contrast, most Ro5 molecules are smaller, more rigid and soluble. Chameleonicity is where a molecule hides its polar groups in non-polar environments as a means to improve its lipophilicity and reduce polarity.[263,264,267,268] Coincidentally, to monitor whether chameleonicity and the conformational adjustment of a molecule in polar and non-polar media, the authors suggested NMR as a local probe of conformational space.[269] As NMR probes the local order, particular environments can be used as molecular fingerprints to indicate a conformational change as well as monitor the rate of exchange and quantify the interconversion that is occurring.[67,252,270]

6.5.4 Temperature Dependence on Exchange

Results

Alongside the standard exchange calculations, variable temperature (VT) ^{1}H-^{1}H EXSY experiments were also acquired to derive thermodynamic parameters which could be qualitatively compared to the previously published computational study.[241] EtOD was used based on its chemical properties and its presence in the marketed formulation of RVR.[271] The ^{1}H-^{1}H EXSY spectra were acquired at three temperatures: 298 K, 308 K and 318 K.

Four environments were evaluated: **40**, **38**, **22** and **5**, with **46&50** omitted due to the intensity imbalance. To obtain integral intensities, slice integration was used except for **38**, where volume integration was used. Table 6.15 summarises the kinetic information calculated from the VT ^{1}H-^{1}H EXSY spectra. Environment **40** at 318 K was omitted from the average value as the uncertainty estimate was four times the magnitude of the corresponding k_{Mm} value. This can be attributed to the bad fit, as shown in figure 6.20. This indicates that a 0.1 s mixing time at 318 K was not appropriate for environment **40**. The residuals from the IF were used to define the associated noise levels.

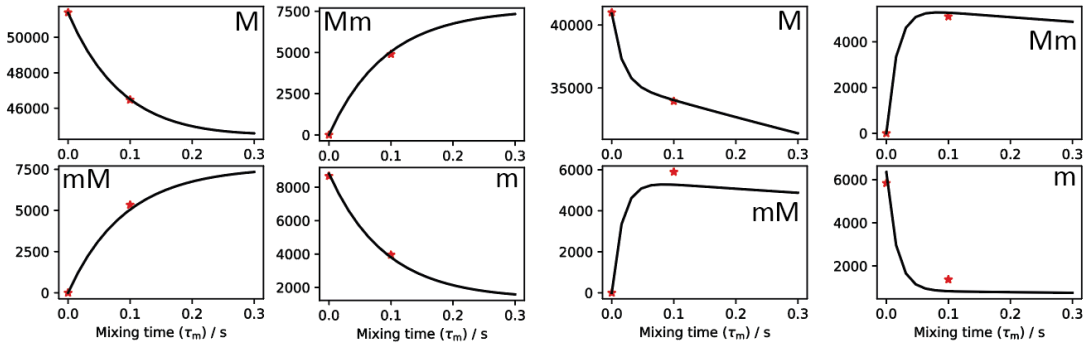


*Figure 6.20: IF intensity curves for environment **40** at 308 K (left) and 318 K (right).*

Table 6.15: Kinetic information calculated for each environment from VT ^{1}H-^{1}H EXSY spectra where those associated with the NMU torsion are in italics.

Temperature / K	Environment	K	k_{Mm} / s^{-1}	Average k_{Mm} / s^{-1}
298 K	**40**	6.7 ± 0.2	0.6 ± 0.05	0.6 ± 0.02
	38	6.1 ± 0.05	0.6 ± 0.003	
	22	5.9 ± 0.03	0.6 ± 0.01	
	5	*5.4 ± 0.3*	*1.4 ± 0.5*	
308 K	**40**	5.8 ± 0.1	1.6 ± 0.1	1.6 ± 0.1
	38	5.9 ± 0.1	1.5 ± 0.1	
	22	5.9 ± 0.1	1.6 ± 0.1	
	5	*4.9 ± 0.3*	*2.5 ± 0.5*	
318 K	**40**	6.4 ± 0.5	8.1 ± 45.3	4.6 ± 2.5
	22	6.0 ± 0.5	4.6 ± 2.5	
	5	*6.3 ± 0.5*	*6.8 ± 14.0*	

With the same logic, the 318 K value for **5** would also be omitted, however, to avoid fitting to a straight line, the value was included in the Arrhenius plots (Section A8 in the Appendix). The resulting values of the activation energy (E_a) and pre-exponential constant (A) for each torsion are in table 6.16.

Table 6.16: Calculated values of E_a and A from Arrhenius plots of $\ln(k_{Mm})$ against inverse T.

Torsion	E_a / kJ mol^{-1}	A / s^{-1}
CRB	80 ± 6	6.9 ± 0.3 x10^{13}
NMU	60 ± 4	4.9 ± 1.3 x10^{10}

Discussion

The derived parameters in table 6.16 can be compared to the computational study by Chakraborty et al.[241] This was a purely computational study of an isolated gas phase molecule so quantitative agreement was not expected. For RVR, Chakraborty et al. proposed the calculated rates of the interconversion as 0.048 s^{-1} for trans to cis transformation (k_{Mm}) whilst the reverse was considerably larger at 64 s^{-1} (k_{mM}).[241] Compared to the values in table 6.15, neither is similar to the experimental data, however, the magnitude of the cis to trans interconversion, k_{mM}, is much larger which indicates it is more likely that the exchange that is occurring is trans to cis for the CRB torsion.

The computational study also proposed a value of E_a for the cis to trans transformation, which was calculated at 58.6 kJ mol^{-1}.[241] This does differ considerably to the CRB torsion (80 ± 6 kJ mol^{-1}) but is more reflective of the NMU torsion (60 ± 4 kJ mol^{-1}). This correlates with the previously calculated values of k_{Mm} for **5** and **9** but also in relation to the configurational change that is expected to be occurring. In EtOD, the dominant conformation of RVR has the NMU torsion as cis so the exchange must be transformation from the cis NMU configuration to trans. This then agrees with the published E_a value for a cis to trans transformation, however, to fully reinforce this, further experimental exploration would be required.

6.6 Conclusions

This chapter has set out to provide additional insight into the solution-state behaviour of RVR through structural assignment, conformational analysis and quantification of exchange behaviour. It is clear that RVR is a complex molecule whose behaviour arises from its conformational flexibility with varying rotational barrier heights. Two critical torsions have been investigated and these were the carbamate (CRB) and N-Methyl Urea (NMU). It has been found that the CRB torsion primarily dictates the rate of interconversion, as observed in the approximate conformer ratios in the 1D ^1H spectra, whilst the NMU torsion also exhibits interconversion but its dominant configuration is dictated by solvent.

It is clear that solvent plays a pivotal role in both conformation and exchange where variations have been observed in lineshapes, NOEs and speed of exchange. Interestingly, the CRB configuration appears independent of solvent polarity but for other sections of RVR, polarity plays a key role. This then continues into the exchange behaviour where RVR exhibits faster interconversion in $CDCl_3$ as compared to DMSO and EtOD for the environments

associated with the CRB torsion. Initial thermodynamic parameters have also been calculated in EtOD where, again, differences were seen between the CRB and NMU torsions.

NMR can provide a comprehensive overview of the conformational behaviour of RVR, however, to further understand the behaviour, comprehensive MD simulations are needed. MD simulations can be used to help understand the molecular origin of the differences between the conformations in each solvent where the experimental observations in this chapter can be used to narrow down possible candidates.

Chapter 7: Concluding Remarks

This thesis aimed to explore how NMR and the NMR crystallographic approach can be used to aid robust structural characterisation of organic materials. Chapters 3 to 5 have focussed on methodology to aid structural characterisation in solids whilst Chapter 6 discusses the ability of solution-state NMR to give insight into the complex behaviour of ritonavir.

Chapter 3 discussed the importance of determining the correct structural model for the pharmaceutical, indapamide hemihydrate. Using a combined approach of ^{13}C SSNMR and SCXRD, the correct structural model was firstly identified and the reason as to why the other models were incorrect was elucidated. This chapter also demonstrated how the combination of SSNMR and PXRD can provide complementary insight into a solid material to robustly determine the correct structural model. Crystal structure determination using PXRD is an emerging area within the NMR crystallography community, as highlighted in the 2022 review by Harris.[82] In conjunction with the structural information derived from SSNMR, there is the possibility of determining structural models from PXRD patterns which, as efforts to create sustainable synthetic methods (such as mechanical grinding to create co-crystals[183,189,272]), allows for the structural model for a greater number of materials to be robustly determined. With reliable structural models, this allows for greater confidence in the accuracy of the resulting calculated chemical shifts, thus improving confidence in the structural assignment of SSNMR data.

In Chapter 4, a Bayesian-probability based workflow that assesses the compatibility of a proposed assignment has been described. It works through the quantitative incorporation of additional data alongside chemical shifts. The precedent of the workflow has been discussed using a selection of model systems, while future work will focus on automation of the workflow with the hope of incorporation into a spectral processing software. Recent literature has demonstrated the desire for a quantitative workflow to evaluate SSNMR spectra in relation to polymorph identification[182] and determination of the correct 3D structural model from either a 2D molecular structure and/or from CSP.[180,181] The workflow discussed in this thesis aims to quantify the likelihood of a proposed assignment through quantitative consideration of all the experimental data acquired as to reflect how robust structural characterisation using the NMR crystallographic approach is achieved within the literature.[10,12,82,144,163,273]

Robust structural assignment in the solid state benefits greatly from multidimensional experiments and pulse sequence development, so Chapter 5 explored two dipolar-based pulse sequences to determine their suitability for use in a structural assignment workflow. The first was the previously published ^{13}C-^{13}C INADEQUATE experiment[60], where a similar level of insight is gained as its *J*-based predecessor but in a shorter total acquisition time. With the shorter acquisition time, the dipolar-based ^{13}C-^{13}C INADEQUATE experiment becomes feasible in both academia and industry as the gold-standard experiment to establish carbon connectivity and determine *Z*' amongst other applications.[56,60,181] The structural information extracted from the experiment also has applications to aid in the derivation of structural models from PXRD[82] as well as adaptability of the pulse sequence to different nuclei, such as ^{29}Si[201] and incorporation into quantitative workflows such as the developed workflow in Chapter 4.

The second pulse sequence was the ^{1}H-^{13}C DQ/SQ experiment that was developed in-house and its optimisation and assignment has been discussed. Pulse sequence development is essential to improve in resolution of nuclei, particularly for ^{1}H. In conjunction with improvements in hardware, this has resulted in resolution enhancement where ^{1}H linewidths are reduced at faster MAS rates.[37,186,187] However, this requires specialist hardware which may not be readily available to all so alternatives must be found as a means to provide similar and/or orthogonal insight into materials. Despite the discussed limitations of the ^{1}H-^{13}C DQ/SQ experiment, the spectra demonstrate orthogonal insight through intermolecular correlations reflecting the molecular packing.[54,55,223] As seen for ^{1}H-^{1}H DQ/SQ spectra, the ^{1}H-^{13}C DQSQ spectra could then provide further insight into hydrogen bonding[133,206] but with the advantage of the additional ^{13}C dimension which can aid assignment.

In contrast, chapter 6 focussed on the characterisation of the solution-state landscape of ritonavir in a variety of solvents. The structural assignments have been presented, alongside the preliminary conformational analyses and associated quantitative kinetic and thermodynamic parameters. It was found that solvent plays a critical role in its conformation and exchange behaviour, where particular torsions were affected more by solvent than others, which can have an impact on the physicochemical properties associated with material. As noted by Sebastiano et al., conformational flexibility and solvent dependence can play a pivotal role in the drug properties in the Bro5 space and this has been observed in the context of RVR.[264,268] Robust understanding of exchange can also aid understanding of the conformational behaviour both experimentally and computationally. ^{1}H-^{1}H EXSY experiments are utilised within the literature to provide key insight into conformational

exchange processes[257–262] but with no consistent evaluation of error; this then introduces ambiguity into the repeatability of a given study. The methodology applied in this chapter aims to be robust and reproducible, thus increasing confidence and robustness in both method and the outputted values. Building on the work in this thesis, additional work would require MD simulations to probe the solution-state landscape further in conjunction with the observations from the experimental NMR data.

References

1. F. Bloch, W. W. Hansen and M. Packard, *Phys. Rev.*, 1946, **70**, 474–485.
2. E. M. Purcell, H. C. Torrey and R. V. Pound, *Phys. Rev*, 1946, **69**, 37–38.
3. E. R. Andrew, A. Bradbury and R. G. Eades, *Nature*, 1958, **182**, 1659.
4. I. J. Lowe, *Phys. Rev. Lett.*, 1959, **2**, 285–287.
5. J. Jeener and G. Alewaeters, *presented at the Ampere International Summer School,* Basko Polje, September, 1971.
6. W. P. Aue, E. Bartholdi and R. R. Ernst, *J. Chem. Phys.*, 1976, **64**, 2229–2246.
7. P. Giraudeau, *Magn. Reson. Chem.*, 2014, **52**, 259–272.
8. K. Grudziąż, A. Zawadzka-Kazimierczuk and W. Koźmiński, *Methods*, 2018, **148**, 81–87.
9. D. Marion, *Mol. Cell. Proteomics*, 2013, **12**, 3006–3025.
10. P. Hodgkinson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2020, **118**–**119**, 10–53.
11. *NMR Crystallography*, ed. R. K. Harris, R. E. Wasylishen and M. J. Duer, Eds., John Wiley & Sons Ltd., West Sussex, UK, 1st edn., 2009.
12. S. A. Southern and D. L. Bryce in *Recent Advances in NMR Crystallography and Polymorphism*, ed. G. A. Webb, Elsevier Ltd., Amsterdam, Netherlands, 2021, vol 102, ch. 1, pp. 1–80.
13. Y. Xu, P. M. J. Szell, V. Kumar and D. L. Bryce, *Coord. Chem. Rev.*, 2020, **411**, 213237.
14. C. Martineau, *Solid State Nucl. Magn. Reson.*, 2014, **63**–**64**, 1–12.
15. M. Li, W. Xu and Y. Su, *Trends Anal. Chem.*, 2021, **135**, 116152.
16. D. J. Kubicki, S. D. Stranks, C. P. Grey and L. Emsley, *Nat. Rev. Chem.*, 2021, **5**, 624–645.
17. D.-K. Bučar, R. W. Lancaster and J. Bernstein, *Angew. Chem. Int. Ed.*, 2015, **54**, 6972–6993.
18. S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley and K. McFarland, *Org. Process Res. Dev.*, 2000, **4**, 413–417.
19. J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter and J. Morris, *Pharm. Res.*, 2001, **18**, 859–866.
20. W. A. Rakoczy and D. M. Mazzochi, *J. Generic Med.*, 2006, **3**, 131–139.
21. T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier, Amsterdam, Netherlands, 2nd edn., 2016.

22. E. Jonas, S. Kuhn and N. Schlörer, *Magn. Reson. Chem.*, 2022, **60**, 1021–1031.
23. M. O. Marcarino, M. M. Zanardi, S. Cicetti and A. M. Sarotti, *Acc. Chem. Res.*, 2020, **53**, 1922–1932.
24. P. H. Willoughby, M. J. Jansma and T. R. Hoye, *Nat. Protoc.*, 2014, **9**, 643–660.
25. C. S. Kim, J. Oh and T. H. Lee, *Arch. Pharm. Res.*, 2020, **43**, 1114–1127.
26. A. Navarro-Vázquez, R. R. Gil and K. Blinov, *J. Nat. Prod.*, 2018, **81**, 203–210.
27. S. G. Smith and J. M. Goodman, *J. Am. Chem. Soc.*, 2010, **132**, 12946–12959.
28. N. Grimblat and A. M. Sarotti, *Chem. Eur. J.*, 2016, **22**, 12246–12261.
29. M. O. Marcarino, S. Cicetti, M. M. Zanardi and A. M. Sarotti, *Nat. Prod. Rep.*, 2022, **39**, 58–76.
30. M. H. Levitt, *Spin Dynamics*, John Wiley & Sons Ltd., West Sussex, UK, 2nd ed., 2007.
31. D. C. Apperley, R. K. Harris and P. Hodgkinson, *Solid-State NMR: Basic Principles & Practice*, Momentum Press LLC, New York, 1st edn., 2012.
32. P. G. Hore, J. Jones and S. Wimperis, *NMR: The Toolkit How Pulse Sequences Work*, Oxford University Press, Oxford, UK, 2015.
33. J. Keeler, *Understanding NMR Spectroscopy*, John Wiley & Sons Ltd., West Sussex, UK, 2nd edn.,2011.
34. M. P. Williamson and D. Neumann, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, New York, 2000.
35. G. E. Pake, *J. Chem. Phys.*, 1948, **16**, 327–336.
36. B. Elena, A. Lesage, S. Steuernagel, A. Böckmann and L. Emsley, *J. Am. Chem. Soc.*, 2005, **127**, 17296–17302.
37. A. Samoson, *J. Magn. Reson.*, 2019, **306**, 167–172.
38. D. L. Bryce, *IUCrJ*, 2017, **4**, 350–359.
39. A. E. Bennett, C. M. Rienstra, M. Auger, K. V. Lakshmi and R. G. Griffin, *J. Chem. Phys.*, 1995, **103**, 6951–6958.
40. P. K. Madhu, *Isr. J. Chem.*, 2014, **54**, 25–38.
41. P. Hodgkinson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2005, **46**, 197–222.
42. B. M. Fung, A. K. Khitrin and K. Ermolaev, *J. Magn. Reson.*, 2000, **142**, 97–101.
43. K. R. Mote, V. Agarwal and P. K. Madhu, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2016, **97**, 1–39.
44. B.-J. van Rossum, H. Förster and H. J. M. de Groot, *J. Magn. Reson.*, 1997, **124**, 516–519.
45. A. Bielecki, A. C. Kolbert and M. H. Levitt, *Chem. Phys. Lett.*, 1989, **155**, 341–346.

46. A. Bielecki, A. C. Kolbert, H. J. M. de Groot and M. H. Levitt, *Adv. Magn. Opt. Reson.*, 1990, **14**, 111–124.
47. S. R. Hartmann and E. L. Hahn, *Phys. Rev.*, 1962, **128**, 2042–2053.
48. G. Metz, M. Ziliox and S. O. Smith, *Solid State Nucl. Magn. Reson.*, 1996, **7**, 155–160
49. W. T. Dixon, J. Schaefer, M. D. Sefcik, E. O. Stejskal and R. A. McKay, *J. Magn. Reson.*, 1982, **49**, 341–345.
50. S. J. Opella and M. H. Frey, *J. Am. Chem. Soc.*, 1979, **101**, 5854–5856.
51. B. J. van Rossum, G. J. Boender, F. M. Mulder, J. Raap, T. S. Balaban, A. Holzwarth, K. Schaffner, S. Prytulla, H. Oschkinat and H. J. M. de Groot, *Spectrochim. Acta A.*, 1998, **54**, 1167–1176.
52. B. J. van Rossum, C. P. de Groot, V. Ladizhansky, S. Vega and H. J. M. de Groot, *J. Am. Chem. Soc.*, 2000, **122**, 3465–3472.
53. A. L. Webber, L. Emsley, R. M. Claramunt and S. P. Brown, *J. Phys. Chem. A*, 2010, **114**, 10435–10442.
54. Y. Jayasubba Reddy, V. Agarwal, A. Lesage, L. Emsley and K. V. Ramanathan, *J. Magn. Reson.*, 2014, **245**, 31–37.
55. A. Lesage, L. Emsley, F. Penin and A. Böckmann, *J. Am. Chem. Soc.*, 2006, **128**, 8246–8255.
56. A. Lesage, M. Bardet and L. Emsley, *J. Am. Chem. Soc.*, 1999, **121**, 10987–10993
57. A. A. Shcherbakov, J. Medeiros-Silva, N. Tran, M. D. Gelenter and M. Hong, *Chem. Rev.*, 2022, **122**, 9848–9879.
58. A. D. Bain, *J. Magn. Reson.*, 1984, **56**, 418–427.
59. M. Hohwy, C. M. Rienstra, C. P. Jaroniec and R. G. Griffin, *J. Chem. Phys.*, 1999, **110**, 7983–7992.
60. M. Dekhil, G. Mollica, T. T. Bonniot, F. Ziarelli, P. Thureau and S. Viel, *Chem. Commun.*, 2016, **52**, 8565–8568.
61. M. H. Levitt, *Symmetry-Based Pulse Sequences in Magic-Angle Spinning Solid-State NMR*, eMagRes, John Wiley & Sons Ltd., 2007.
62. M. Hohwy, H. J. Jakobsen, M. Edén, M. H. Levitt and N. C. Nielsen, *J. Chem. Phys.*, 1998, **108**, 2686–2694.
63. A. L. Webber, B. Elena, J. M. Griffin, J. R. Yates, T. N. Pham, F. Mauri, C. J. Pickard, A. M. Gil, R. Stein, A. Lesage, L. Emsley and S. P. Brown, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6970-6983.
64. C. R. Jones, C. P. Butts and J. N. Harvey, *Beilstein J. Org. Chem.*, 2011, **7**, 145–150.

65. C. P. Butts, C. R. Jones, E. C. Towers, J. L. Flynn, L. Appleby and N. J. Barron, *Org. Biomol. Chem.*, 2011, **9**, 177–184.
66. H. Hu and K. Krishnamurthy, *J. Magn. Reson.*, 2006, **182**, 173–177.
67. K. Nikitin and R. O'Gara, *Chem. Eur. J.*, 2019, **25**, 4551–4589.
68. C. L. Perrin and T. J. Dwyer, *Chem. Rev.*, 1990, **90**, 935–967.
69. A. D. Bain, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2003, **43**, 63–103.
70. J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546–4553.
71. M. Ababneh-Khasawneh, B. E. Fortier-McGill, M. E. Occhionorelli and A. D. Bain, *J. Phys. Chem. A*, 2011, **115**, 7531–7537.
72. Z. Zolnai, N. Juranić, D. Vikić-Topić and S. Macura, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 611–621.
73. J. Lu, D. Ma, J. Hu, W. Tang and D. Zhu, *J. Chem. Soc., Dalt. Trans.*, 1998, **1**, 2267–2273.
74. S. R. Hussaini, A. Kuta, A. Pal, Z. Wang, M. A. Eastman and R. Duran, *ACS Omega*, 2020, **5**, 24848–24853.
75. G. E. Hawkes, L. Y. Lian, E. W. Randall, K. D. Sales and S. Aime, *J. Magn. Reson.*, 1985, **65**, 173–177.
76. S. Zhang, Z. Kovacs, S. Burgess, S. Aime, E. Terreno and A. D. Sherry, *Chem. Eur. J.*, 2001, **7**, 288–296.
77. C. L. Perrin and R. E. Engler, *J. Magn. Reson.*, 1996, **123**, 188–195.
78. W. H. Bernskoetter, S. K. Hanson, S. K. Buzak, Z. Davis, P. S. White, R. Swartz, K. I. Goldberg and M. Brookhart, *J. Am. Chem. Soc.*, 2009, **131**, 8603–8613.
79. C. Giacovazzo, H. L. Monaco, G. Artioli, D. Viterbo, M. Milanesio, G. Ferraris, G. Gilli, P. Gilli, G. Zanotti and M. Catti, *Fundamentals of Crystallography*, Oxford University Press, Oxford, UK, 3rd edn., 2011.
80. C. L. Evans, I. R. Evans and P. Hodgkinson, *Chem. Commun.*, 2022, **58**, 4767–4770.
81. W. H. Bragg and W. L. Bragg, *Proc. R. Soc. London, Ser. A*, 1913, **88**, 428–438.
82. K. D. M. Harris, *Crystals*, 2022, **12**, 1277–1297.
83. M. Smrkolj and A. Meden, *Pharmazie*, 2006, **61**, 999–1004.
84. M. Aljohani, P. MacFhionnghaile, P. McArdle and A. Erxleben, *Int. J. Pharm.*, 2019, **561**, 35–42.
85. S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson and M. C. Payne, *Z. Kristallogr.*, 2005, **220**, 567–570.
86. C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azaïs, S. E.

Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri and C. J. Pickard, *Chem. Rev.*, 2012, **112**, 5733–5779.

87. P. J. Hasnip, K. Refson, M. I. J. Probert, J. R. Yates, S. J. Clark and C. J. Pickard, *Phil. Trans. R. Soc. A*, 2014, **372**, 20130270.
88. W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, 1133–1138.
89. W. Kohn, *Phys. Rev. Lett.*, 1996, **76**, 3168–3171.
90. P. Hohenberg and W. Kohn, *Phys. Rev. B*, 1964, **136**, 864–871.
91. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
92. J. C. Slater, *Phys. Rev.*, 1951, **81**, 385–390.
93. P. Lykos and G. W. Pratt, *Rev. Mod. Phys.*, 1963, **35**, 496–501.
94. C. J. Pickard and F. Mauri, *Phys. Rev. B.*, 2001, **63**, 245101.
95. C. G. Van de Walle and P. E. Blöchl, *Phys. Rev. B*, 1993, **47**, 4244–4255.
96. E. C. Snow and J. H. Wood, *Chem. Phys. Lett.*, 1974, **25**, 111–113.
97. U. von Barth and C. D. Gelatt, *Phys. Rev. B*, 1980, **21**, 2222–2228.
98. J. R. Yates, C. J. Pickard and F. Mauri, *Phys. Rev. B*, 2007, **76**, 024401.
99. D. Vanderbilt, *Phys. Rev. B*, 1990, **41**, 7892–7895.
100. C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta B.*, 2016, **72**, 171–179.
101. J. van de Streek and M. A. Neumann, *Acta Crystallogr. B,* 2010, **66**, 544–558.
102. C. M. Widdifield, J. D. Farrell, J. C. Cole, J. A. K. Howard and P. Hodgkinson, *Chem. Sci.*, 2020, **11**, 2987–2992.
103. E. Salager, G. M. Day, R. S. Stein, C. J. Pickard, B. Elena and L. Emsley, *J. Am. Chem. Soc.*, 2010, **132**, 2564–2566.
104. J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller and G. J. O. Beran, *Phys. Chem. Chem. Phys.*, 2016, **18**, 21686.
105. Bruker, Bruker TopSpin (version 3.6.4), USA, 2017
106. M. R. Willcott, *J. Am. Chem. Soc.*, 2009, **131**, 13180–13180.
107. JEOL, JASON (version 2.3), Japan, 2017.
108. S. G. J. van Meerten, W. M. J. Franssen and A. P. M. Kentgens, *J. Magn. Reson.*, 2019, **301**, 56–66.
109. Python, (version 2.7 and 3.9), Anaconda Inc., USA, 2023.
110. Adobe, Adobe Illustrator (version 26.3.1), Adobe Inc., USA, 2022.
111. S. Sturniolo, T. F. G. Green, R. M. Hanson, M. Zilka, K. Refson, P. Hodgkinson, S. P. Brown and J. R. Yates, *Solid State Nucl. Magn. Reson.*, 2016, **78**, 64–70.
112. O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann, *J. Appl.*

Cryst., 2009, **42**, 339–341.

113. C. F. MacRae, I. Sovago, S. J. Cottrell, P. T. A. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler and P. A. Wood, *J. Appl. Crystallogr.*, 2020, **53**, 226–235.
114. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchinson, *J. Cheminfomatics*, 2014, **4**, 1–17.
115. N. Marchand, P. Lienard, H. U. Siehl and H. Izato, *Fujitsu Sci. Tech. J.*, 2014, **50**, 46–51.
116. C. M. Widdifield, H. Robson and P. Hodgkinson, *Chem. Commun.*, 2016, **52**, 6685–6688.
117. G. C. Roush and D. A. Sica, *Am. J. Hypertens.*, 2016, **29**, 1130–1137.
118. G. C. Roush, M. E. Ernst, J. B. Kostis, S. Tandon and D. A. Sica, *Hypertension*, 2015, **65**, 1041–1046.
119. A. Drogoń, M. Skotnicki, A. Skotnicka and M. Pyda, *Pharmaceutics*, 2020, **12**, 800–815.
120. Z. Wojnarowska, K. Grzybowska, L. Hawelek, M. Dulski, R. Wrzalik, I. Gruszka, M. Paluch, K. Pienkowska, W. Sawicki, P. Bujak, K. J. Paluch, L. Tajber and J. Markowski, *Mol. Pharmaceutics.*, 2013, **10**, 3612–3627.
121. M. Skotnicki, A. Drogoń, J. J. Calvin, P. F. Rosen, B. F. Woodfield and M. Pyda, *Thermochim. Acta*, 2019, **674**, 36–43.
122. S. Allu, K. Suresh, G. Bolla, M. K. C. Mannava and A. Nangia, *CrystEngComm*, 2019, **21**, 2043–2048.
123. P. Ghugare, V. Dongre, P. Karmuse, R. Rana, D. Singh, A. Kumar and Z. Filmwala, *J. Pharm. Biomed. Anal.*, 2010, **51**, 532–540.
124. L. M. Rus, I. Kacso, G. Borodi, M. Aluas, I. Tomuta, C. Iuga, S. Simon, I. Bratu and M. Bojita, *AIP Conf. Proc.*, 2012, **1425**, 39–42.
125. J. Bojarska, A. Fruziński and W. Maniukiewicz, *J. Mol. Struct.*, 2016, **1116**, 22–29.
126. R. F. Moran, D. M. Dawson and S. E. Ashbrook, *Int. Rev. Phys. Chem.*, 2017, **36**, 39–115.
127. B. Dittrich, *IUCrJ*, 2021, **8**, 305–318.
128. S. E. Ashbrook and D. McKay, *Chem. Commun.*, 2016, **52**, 7186–7204.
129. B. Kojić-Prodić and Ž. Ružić-Toroš, *Acta Crystallogr. B.*, 1982, **38**, 2948–2951.
130. A. Běrziņš and P. Hodgkinson, *Solid State Nucl. Magn. Reson.*, 2015, **65**, 12–20.
131. H. E. Kerr, L. K. Softley, K. Suresh, A. Nangia, P. Hodgkinson and I. R. Evans, *CrystEngComm*, 2015, **17**, 6707–6715.
132. X. Filip, G. Borodi and C. Filip, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17978–17986.

133. B. I. Harriss, L. Vella-Zarb, C. Wilson and I. R. Evans, *Cryst. Growth Des.*, 2014, **14**, 783–791.
134. A. Kogan, I. Popov, V. Uvarov, S. Cohen, A. Aserin and N. Garti, *Langmuir*, 2008, **24**, 722–733.
135. S. Mahapatra, Y. Azim and G. R. Desiraju, *J. Mol. Struct.*, 2010, **976**, 200–204.
136. K. Chadwick, G. Sadiq, R. J. Davey, C. C. Seaton, R. G. Pritchard and A. Parkin, *Cryst. Growth Des.*, 2009, **9**, 1278–1279.
137. A. A. Coelho, J. Evans, I. R. Evans, A. Kern and S. Parsons, *Powder Diffr.*, 2011, **26**, S22–S25.
138. G. M. Sheldrick, *Acta Crystallogr. C.*, 2015, **71**, 3–8.
139. H. J. Monkhorst and J. D. Pack, *Phys. Rev. B.*, 1976, **13**, 5188–5192.
140. T. Björkman, *Comput. Phys. Commun.*, 2011, **182**, 1183–1186.
141. A. J. Al-Ani, P. M. J. Szell, Z. Rehman, H. Blade, H. P. Wheatcroft, L. P. Hughes, S. P. Brown and C. C. Wilson, *Cryst. Growth Des.*, 2022, **22**, 4696–4707.
142. L. Bellaiche and D. Vanderbilt, *Phys. Rev. B.*, 2000, **61**, 7877–7882.
143. M. J. D. Powell, *Comput. J.*, 1964, **7**, 155–162.
144. A. S. Tatton, H. Blade, S. P. Brown, P. Hodgkinson, L. P. Hughes, S. O. Nilsson Lill and J. R. Yates, *Cryst. Growth Des.*, 2018, **18**, 3339–3351.
145. C. J. H. Smalley, H. E. Hoskyns, C. E. Hughes, D. N. Johnstone, T. Willhammar, M. T. Young, C. J. Pickard, A. J. Logsdail, P. A. Midgley and K. D. M. Harris, *Chem. Sci.*, 2022, **13**, 5277–5288.
146. C. Buchsbaum and M. U. Schmidt, *Acta Crystallogr. B.*, 2007, **63**, 926–932.
147. A. L. Spek, *Acta Crystallogr. D.*, 2009, **65**, 148–155.
148. P. Müller, *Crystallogr. Rev.*, 2009, **15**, 57–83.
149. J. D. Dunitz, V. Schomaker and K. N. Trueblood, *J. Phys. Chem.*, 1988, **92**, 856–867.
150. K. N. Trueblood, H. B. Bürgi, H. Burzlaff, J. D. Dunitz, C. M. Gramaccioli, H. H. Schulz, U. Shmueli and S. C. Abrahams, *Acta Crystallogr. A.*, 1996, **52**, 770–781.
151. C. M. Widdifield, S. O. Nilsson Lill, A. Broo, M. Lindkvist, A. Pettersen, A. Svensk Ankarberg, P. Aldred, S. Schantz and L. Emsley, *Phys. Chem. Chem. Phys.*, 2017, **19**, 16650–16661.
152. M. Dračínský and P. Hodgkinson, *Chem. Eur. J.*, 2014, **20**, 2201–2207.
153. P. H. Zwart, R. W. Grosse-Kunstleve, A. A. Lebedev, G. N. Murshudov and P. D. Adams, *Acta Crystallogr. D.*, 2008, **64**, 99–107.
154. A. A. Lebedev and M. N. Isupov, *Acta Crystallogr. D.*, 2014, **70**, 2430–2443.

155. J. Li and J. Sun, *Acc. Chem. Res.*, 2017, **50**, 2737–2745.
156. K. M. Steed and J. W. Steed, *Chem. Rev.*, 2015, **115**, 2895–2933.
157. S. Radaev, J. Agniswamy and P. D. Sun, *Acta Crystallogr. D.*, 2009, **65**, 1334–1340.
158. Y. V. Nelyubina, I. L. Dalinger and K. A. Lyssenko, *Angew. Chem.*, 2011, **123**, 2944–2946.
159. C. Schlesinger, A. Fitterer, C. Buchsbaum, S. Habermehl, M. R. Chierotti, C. Nervi and M. U. Schmidt, *IUCrJ*, 2022, **9**, 406–424.
160. D. Watkin, *J. Appl. Crystallogr.*, 2008, **41**, 491–522.
161. A. Brink and J. R. Helliwell, *IUCrJ*, 2019, **6**, 788–793.
162. S. Ando, J. Kikuchi, Y. Fujimura, Y. Ida, K. Higashi, K. Moribe and K. Yamamoto, *J. Pharm. Sci.*, 2012, **101**, 3214–3221.
163. E. Carignani, S. Borsacchi, J. P. Bradley, S. P. Brown and M. Geppi, *J. Phys. Chem. C*, 2013, **117**, 17731–17740.
164. J. Czernek, *Chem. Phys. Lett.*, 2015, **619**, 230–235.
165. N. J. Clayden, C. M. Dobson, L. Y. Lian and J. M. Twyman, *J. Chem. Soc. Perkin Trans. 2*, 1986, **12**, 1933–1940.
166. M. R. M. P. Aguiar, S. M. C. Menezes, A. L. Gemal and R. A. S. San Gil, *Solid State Nucl. Magn. Reson.*, 1995, **4**, 179–185.
167. D. L. M. de Aguiar, R. A. da S. San Gil, R. B. de Alencastro, E. F. de Souza, L. B. Borré, V. da Silva Vaiss and A. A. Leitão, *Chem. Phys. Lett.*, 2016, **660**, 214–220.
168. J. D. Hartman, S. Monaco, B. Schatschneider and G. J. O. Beran, *J. Chem. Phys.*, 2015, **143**, 102809.
169. J. Czernek and J. Brus, *Chem. Phys. Lett.*, 2013, **586**, 56–60.
170. J. Czernek and J. Brus, *Chem. Phys. Lett.*, 2014, **608**, 334–339.
171. N. Grimblat, M. M. Zanardi and A. M. Sarotti, *J. Org. Chem.*, 2015, **80**, 12526–12534.
172. K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, *Org. Biomol. Chem.*, 2016, **14**, 3943–3949.
173. K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, *Org. Biomol. Chem.*, 2017, **15**, 8998–9007.
174. K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, *Org. Biomol. Chem.*, 2019, **17**, 5886–5890.
175. N. Grimblat, J. A. Gavín, A. Hernández Daranas and A. M. Sarotti, *Org. Lett.*, 2019, **21**, 4003–4007.
176. D. Xin, P. J. Jones and N. C. Gonnella, *J. Org. Chem.*, 2018, **83**, 5035–5043.

177. A. Howarth, K. Ermanis and J. M. Goodman, *Chem. Sci.*, 2020, **11**, 5351–4359.
178. A. Howarth and J. M. Goodman, *Chem. Sci.*, 2022, **13**, 3507–3518.
179. Y. Tsai, M. Amichetti, M. M. Zanardi, R. Grimson, A. H. Daranas and A. M. Sarotti, *Org. Lett.*, 2022, **24**, 7487–7491.
180. E. A. Engel, A. Anelli, A. Hofstetter, F. Paruzzo, L. Emsley and M. Ceriotti, *Phys. Chem. Chem. Phys.*, 2019, **21**, 23385–23400.
181. M. Cordova, M. Balodis, B. Simões de Almeida, M. Ceriotti and L. Emsley, *Sci. Adv.*, 2021, **7**, eabk2341.
182. G. A. Valdivia-Berroeta, K. Sarpal and N. C. Gonnella, *Cryst. Growth Des.*, 2021, **21**, 5904–5913.
183. S. M. A. Mashhadi, D. Yufit, H. Liu, P. Hodgkinson and U. Yunus, *J. Mol. Struct.*, 2020, **1219**, 128621.
184. I. Bryndal, M. Jaremko, Ł. Jaremko and T. Lis, *Acta Crystallogr. C.*, 2006, **62**, o111–o114.
185. G. Comellas, J. J. Lopez, A. J. Nieuwkoop, L. R. Lemkau and C. M. Rienstra, *J. Magn. Reson.*, 2011, **209**, 131–135.
186. S. P. Brown, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2007, **50**, 199–251.
187. S. P. Brown, *Solid State Nucl. Magn. Reson.*, 2012, **41**, 1–27.
188. W. J. Hehre, L. Radom, P. V. R. Schleyer and J. A. Pople, *Ab initio Molecular Orbital Theory*, Wiley, New York, 1986.
189. B. Swapna, D. Maddileti and A. Nangia, *Cryst. Growth Des.*, 2014, **14**, 5991–6005.
190. S. M. A. Mashhadi, A. S. Batsanov, S. A. Sajjad, Y. Nazir, M. H. Bhatti and U. Yunus, *J. Mol. Struct.*, 2021, **1226**, 129388.
191. G. N. M. Reddy, A. Marsh, J. T. Davis, S. Masiero and S. P. Brown, *Cryst. Growth Des.*, 2015, **15**, 5945–5954.
192. N. P. Wickramasinghe and Y. Ishii, *J. Magn. Reson.*, 2006, **181**, 233–243.
193. J. Vĺcha, J. Novotný, S. Komorovsky, M. Straka, M. Kaupp and R. Marek, *Chem. Rev.*, 2020, **120**, 7065–7103.
194. Y. Y. Rusakov and I. L. Rusakova, *Magn. Reson. Chem.*, 2018, **56**, 716–726.
195. P. Lantto, J. Vaara, A. M. Kantola, V.-V. Telkki, B. Schimmelpfennig, K. Ruud and J. Jokisaari, *J. Am. Chem. Soc.*, 2002, **124**, 2762–2771.
196. M. Baias, J. N. Dumez, P. H. Svensson, S. Schantz, G. M. Day and L. Emsley, *J. Am. Chem. Soc.*, 2013, **135**, 17501–17507.
197. M. Deschamps, S. Cadars, E. Gilbert, P. Azaïs, E. Raymundo-Pinero, F. Béguin and D.

Massiot, *Solid State Nucl. Magn. Reson.*, 2012, **42**, 81–86.

198. R. K. Harris, S. A. Joyce, C. J. Pickard, S. Cadars and L. Emsley, *Phys. Chem. Chem. Phys.*, 2006, **8**, 137–143.
199. R. K. Harris, S. Cadars, L. Emsley, J. R. Yates, C. J. Pickard, R. K. R. Jetti and U. J. Griesser, *Phys. Chem. Chem. Phys.*, 2007, **9**, 360–368.
200. G. De Paëpe, A. Lesage, S. Steuernagel and L. Emsley, *ChemPhysChem*, 2004, **5**, 869–875.
201. R. W. Dorn, E. A. Marro, M. P. Hanrahan, R. S. Klausen and A. J. Rossini, *Chem. Mater.*, 2019, **31**, 9168–9178.
202. P. Thureau, M. Juramy, F. Ziarelli, S. Viel and G. Mollica, *Solid State Nucl. Magn. Reson.*, 2019, **99**, 15–19.
203. K. Märker, S. Paul, C. Fernández-de-Alba, D. Lee, J. M. Mouesca, S. Hediger and G. De Paëpe, *Chem. Sci.*, 2017, **8**, 974–987.
204. H. E. Kerr, L. K. Softley, K. Suresh, P. Hodgkinson and I. R. Evans, *Acta Crystallogr. C.*, 2017, **73**, 168–175.
205. R. Zhang, K. H. Mroue and A. Ramamoorthy, *Acc. Chem. Res.*, 2017, **50**, 1105–1113.
206. Y. Nishiyama, *Solid State Nucl. Magn. Reson.*, 2016, **78**, 24–36.
207. N. T. Duong, V. Agarwal and Y. Nishiyama, *Solid State Nucl. Magn. Reson.*, 2022, **117**, 101774.
208. J. R. Yarava, R. Sonti, K. Kantharaju, S. Raghothama and K. V. Ramanathan, *Chem. Commun.*, 2017, **53**, 1317–1320.
209. J. R. Yarava, Y. Nishiyama, S. Raghothama and K. V. Ramanathan, *Chem. Biol. Drug Des.*, 2020, **95**, 394–407.
210. M. H. Levitt, *J. Chem. Phys.*, 2008, **128**, 052205.
211. M. Carravetta, M. Edén, O. G. Johannessen, H. Luthman, P. J. E. Verdegem, J. Lugtenburg, A. Sebald and M. H. Levitt, *J. Am. Chem. Soc.*, 2001, **123**, 10628–10638.
212. I. Marin-Montesinos, D. H. Brouwer, G. Antonioli, W. C. Lai, A. Brinkmann and M. H. Levitt, *J. Magn. Reson.*, 2005, **177**, 307–317.
213. S. Cadars, J. Sein, L. Duma, A. Lesage, T. N. Pham, J. H. Baltisberger, S. P. Brown and L. Emsley, *J. Magn. Reson.*, 2007, **188**, 24–34.
214. A. Lesage, C. Auger, S. Caldarelli and L. Emsley, *J. Am. Chem. Soc.*, 1997, **119**, 7867–7868.
215. Y. Nishiyama, G. Hou, V. Agarwal, Y. Su and A. Ramamoorthy, *Chem. Rev.*, 2023, **123**, 918–988.

216. M. G. Munowitz and R. G. Griffin, *J. Chem. Phys.*, 1982, **76**, 2848–2858.
217. P. Tekely, *Solid State Nucl. Magn. Reson.*, 1999, **14**, 33–41.
218. P. Tekely, *Solid State Nucl. Magn. Reson.*, 2015, **72**, 41–49.
219. I. Frantsuzov, M. Ernst, S. P. Brown and P. Hodgkinson, *Solid State Nucl. Magn. Reson.*, 2015, **70**, 28–37.
220. H. J. Hogben, M. Krzystyniak, G. T. P. Charnock, P. J. Hore and I. Kuprov, *J. Magn. Reson.*, 2011, **208**, 179–194.
221. C. Bengs and M. H. Levitt, *Magn. Reson. Chem.*, 2018, **56**, 374–414.
222. J. P. Bradley, C. Tripon, C. Filip and S. P. Brown, *Phys. Chem. Chem. Phys.*, 2009, **11**, 6941–6952
223. A. L. Webber, S. Masiero, S. Pieraccini, J. C. Burley, A. S. Tatton, D. Iuga, T. N. Pham, G. P. Spada and S. P. Brown, *J. Am. Chem. Soc.*, 2011, **133**, 19777–19795.
224. A. Chakraborty, L. D. Fernando, W. Fang, M. C. Dickwella Widanage, P. Wei, C. Jin, T. Fontaine, J. P. Latgé and T. Wang, *Nat. Commun.*, 2021, **12**, 6346–6358.
225. F. A. Perras, A. Venkatesh, M. P. Hanrahan, T. W. Goh, W. Huang, A. J. Rossini and M. Pruski, *J. Magn. Reson.*, 2017, **276**, 95–102.
226. S. Laage, A. Marchetti, J. Sein, R. Pierattelli, H. J. Sass, S. Grzesiek, A. Lesage, G. Pintacuda and L. Emsley, *J. Am. Chem. Soc.*, 2008, **130**, 17216–17217.
227. P. Hodgkinson and L. Emsley, *J. Magn. Reson.*, 1999, **139**, 46–59.
228. M. J. Bayro, M. Huber, R. Ramachandran, T. C. Davenport, B. H. Meier, M. Ernst and R. G. Griffin, *J. Chem. Phys.*, 2009, **130**, 114506.
229. M. Leskes, P. K. Madhu and S. Vega, *Chem. Phys. Lett.*, 2007, **447**, 370–374.
230. A. Emwas, K. Szczepski, B. G. Poulson, K. Chandra, R. T. Mckay, M. Dhahri, F. Alahmari, L. Jaremko, J. I. Lachowicz and M. Jaremko, *Molecules*, 2020, **25**, 4597.
231. B. Diethelm-Varela, *ChemMedChem*, 2021, **16**, 725–742.
232. Q. Li and C. B. Kang, *Molecules*, 2020, **25**, 2974.
233. L. Shi and N. Zhang, *Molecules*, 2021, **26**, 576.
234. D. J. Kempf, H. L. Sham, K. C. Marsh, C. A. Flentge, D. Betebenner, B. E. Green, E. Mcdonald, S. Vasavanonda, A. Saldivar, N. E. Wideburg, W. M. Kati, L. Ruiz, C. Zhao, L. Fino, J. Patterson, A. Molla, J. J. Plattner and D. W. Norbeck, *J. Med. Chem.*, 1998, **41**, 602–617.
235. S. Meini, A. Pagotto, B. Longo, I. Vendramin, D. Pecori and C. Tascini, *J. Clin. Med.*, 2020, **9**, 2050.
236. W. Wen, C. Chen, J. Tang, C. Wang, M. Zhou, Y. Cheng, X. Zhou, Q. Wu, X. Zhang, Z.

Feng, M. Wang and Q. Mao, *Ann. Med.*, 2022, **54**, 516–523.

237. Pfizer Shares In Vitro Efficacy of Novel COVID-19 Oral Treatment Against Omicron Variant, https://www.pfizer.com/news/press-release/press-release-detail/pfizer-shares-vitro-efficacy-novel-covid-19-oral-treatment, (accessed February 2023).
238. P. Sacchi, S. Wright, P. Neoptolemou, G. I. Lampronti, A. K. Rajagopalan, W. Kras, C. L. Evans, P. Hodgkinson and A. J. Cruz-Cabeza, *ChemRxiv*, 2023, preprint, DOI: 10.26434/chemrxiv-2023-gz7kb
239. C. Wang., I. Rosbottom, T. D. Turner, S. Laing, A. G. P. Maloney, A. Y. Sheikh, R. Docherty, Q. Yin and K. J. Roberts, *Pharm. Res.*, 2021, **38**, 971–990.
240. J. M. Miller, B. M. Collman, L. R. Greene, D. J. W. Grant and A. C. Blackburn, *Pharm. Dev. Technol.*, 2005, **10**, 291–297.
241. D. Chakraborty, N. Sengupta and D. J. Wales, *J. Phys. Chem. B*, 2016, **120**, 4331–4340.
242. X. Yao, R. F. Henry and G. G. Z. Zhang, *J. Pharm. Sci.*, 2023, **112**, 237–242.
243. K. Kawakami, T. Harada, K. Miura, Y. Yoshihashi, E. Yonemochi, K. Terada and H. Moriyama, *Mol. Pharmaceutics.*, 2014, **11**, 1835–1843.
244. S. D. Parent, P. A. Smith, D. K. Purcell, D. T. Smith, S. J. Bogdanowich-Knipp, A. S. Bhavsar, L. R. Chan, J. M. Croom, H. C. Bauser, A. McCalip, S. R. Byrn and A. Radocea, *Cryst. Growth Des.*, 2023, **23**, 320–325.
245. S. L. Morissette, S. Soukasene, D. Levinson, M. J. Cima and Ö. Almarsson, *Proc. Natl. Acad. Sci.,* 2003, **100**, 2180–2184.
246. *US Pat.*, US8598216B1, 2013.
247. H. J. Jeong, S. Min, H. Chae, S. Kim, G. Lee, S. K. Namgoong and K. Jeong, *Sci. Rep.*, 2020, **10**, 14290.
248. D. N. Nguyen and G. Van den Mooter, *Int. J. Pharm.*, 2014, **475**, 214–226.
249. M. H. Abraham, R. J. Abraham, J. Byrne and L. Griffiths, *J. Org. Chem.*, 2006, **71**, 3389–3394.
250. W. L. Whaley, E. M. Okoso-Amaa, C. L. Womack, A. Vladimirova, L. B. Rogers, M. J. Risher and M. H. Abraham, *Nat. Prod. Commun.*, 2013, **8**, 85–98.
251. M. H. Abraham, R. J. Abraham, W. E. Acree, A. E. Aliev, A. J. Leo and W. L. Whaley, *J. Org. Chem.*, 2014, **79**, 11075–11083.
252. Y. Ben-Tal, P. J. Boaler, H. J. A. Dale, R. E. Dooley, N. A. Fohn, Y. Gao, A. García-Domínguez, K. M. Grant, A. M. R. Hall, H. L. D. Hayes, M. M. Kucharski, R. Wei and G. C. Lloyd-Jones, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2022, **129**, 28–106.
253. C. L. Perrin, *J. Magn. Reson.*, 1989, **82**, 619–621.

254. C. L. Perrin and R. K. Gipe, *J. Am. Chem. Soc.*, 1984, **106**, 4036–4038.
255. S. Macura and R. R. Ernst, *Mol. Phys.*, 2002, **100**, 135–147.
256. E. W. Abel, T. P. J. Coston, K. G. Orrell, V. Šik and D. Stephenson, *J. Magn. Reson.*, 1986, **70**, 34–53.
257. F. Zaccaria, C. Zuccaccia and A. Macchioni, *Inorganica Chim. Acta*, 2021, **527**, 120565.
258. H. T. Ang, A. A. Ponich, M. Paladino, M. Miskolzie and D. G. Hall, *J. Am. Chem. Soc.*, 2022, **144**, 10570–10581.
259. M. Garberová, I. Potočňák, M. Tvrdoňová, M. Bago-Pilátová, S. Bekešová, Z. Kudličková, E. Samoľová, A. Kešeľáková, J. Elečko and M. Vilková, *J. Mol. Struct.*, 2023, **1271**, 134112.
260. M. Dangalov, P. Petrov and N. G. Vassilev, *J. Mol. Struct.*, 2021, **1230**, 129944.
261. K. Helttunen, *European J. Org. Chem.*, 2022, **42**, e202200647.
262. M. M. Shoshani, S. Xiong, J. J. Lawniczak, X. Zhang, T. F. Miller and T. Agapie, *Organometallics*, 2022, **41**, 2119–2131.
263. A. Whitty, M. Zhong, L. Viarengo, D. Beglov, D. R. Hall and S. Vajda, *Drug Discov. Today*, 2016, **21**, 712–717.
264. M. Rossi Sebastiano, B. C. Doak, M. Backlund, V. Poongavanam, B. Over, G. Ermondi, G. Caron, P. Matsson and J. Kihlberg, *J. Med. Chem.*, 2018, **61**, 4189–4202.
265. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 2001, **46**, 3–26.
266. A. Charalabidis, M. Sfouni, C. Bergström and P. Macheras, *Int. J. Pharm.*, 2019, **566**, 264–281.
267. P. Matsson, B. C. Doak, B. Over and J. Kihlberg, *Adv. Drug Deliv. Rev.*, 2016, **101**, 42–61.
268. M. Rossi Sebastiano, D. Garcia Jimenez, M. Vallaro, G. Caron and G. Ermondi, *J. Med. Chem.*, 2022, **65**, 12068–12083.
269. Y. Atilaw, V. Poongavanam, C. Svensson Nilsson, D. Nguyen, A. Giese, D. Meibom, M. Erdelyi and J. Kihlberg, *ACS Med. Chem. Lett.*, 2021, **12**, 107–114.
270. A. Y. S. Balazs, N. L. Davies, D. Longmire, M. J. Packer and E. Chiarparin, *Magn. Reson.*, 2021, **2**, 489–498.
271. J. B. Morris, D. A. Tisi, D. C. T. Tan and J. H. Worthington, *Int. J. Mol. Sci.*, 2019, **20**, 1718.
272. P. Ying, J. Yu and W. Su, *Adv. Synth. Catal.*, 2021, **363**, 1246–1271.
273. A. Abraham, D. C. Apperley, S. J. Byard, A. J. Ilott, A. J. Robbins, V. Zorin, R. K. Harris

and P. Hodgkinson, *CrystEngComm*, 2016, **18**, 1054–1063.

274. M. Shalaeva, G. Caron, Y. A. Abramov, T. N. O'Connell, M. S. Plummer, G. Yalamanchi, K. A. Farley, G. H. Goetz, L. Philippe and M. J. Shapiro, *J. Med. Chem.*, 2013, **56**, 4870–4879.

275. S. Raychaudhuri, *presented at the 2008 Winter Simulation. Conference (IEEE)*, Broomfield, 2008.

276. R. L. Harrison, R. L. Harrison, *AIP Conf. Proc.*, 2009, **1204**, 17–21.

Appendix

A1: *hkl* Tables for WOCPEM01

Alongside the visual plot of the superstructure reflections in Chapter 3, the *hkl* tables for the identified superstructure reflections are provided below.

Table A1: Summary of Superstructure reflections discussed in Chapter 3 in Table 3.7 where σ reflects the standard error.

***hkl* Reflections**	**Intensity (I)**				**Average (I)**	**σ(I)**				**Average (σ(I))**
300	9.4	10.4			9.9	1.3	1.6			1.45
−300	9.8	10			9.9	0.9	1.0			0.95
310	13.2	13.0	14.8	14.6	13.9	1.9	2.0	2.1	1.6	1.9
−310	11.6	15.5	14.5	14.6	14.05	2.0	1.7	1.7	1.7	1.775
3−10	12.7	16.9	12.0	12.5	13.525	1.6	1.7	2.0	1.7	1.75
−3−10	13.1	17.1	13.4	12.4	14.0	1.9	2.2	2.0	1.5	1.9
111	15.3	14.6	14.6	10.6	13.78	1.4	1.5	1.9	1.4	1.55
−111	34.8	35.5	38.7	-	36.33	3.1	3.4	3.1	-	3.2
1−11	12.3	12.0	14.7	15.7	13.68	1.5	1.7	1.3	1.5	1.5
11−1	37.3	38.3	38.0	34.3	37.0	3.1	2.9	3.4	3.3	3.18
−1−11	37.9	35.2	36.5	36.8	36.60	3.4	2.8	3.3	3.2	3.18
−11−1	14.6	14.0	14.6	14.5	14.43	1.3	1.8	1.6	1.4	1.53
1−1−1	38.1	33.6	39.4	-	37.03	3.1	3.2	3.4	--	3.23
−1−1−1	13.7	15.6	13.0	14.4	14.18	1.4	1.8	1.6	1.4	1.55

A2: Expanded Workflow for Group 4 of Tyr and Group 7 of INH-2HCA.

To complement that worked example in Section 4.4.4, additional insight is provided below in relation to how the resulting AF values are calculated for Group 4 of Tyr.

To begin, the calculated isotropic shieldings require scaling to calculated ^{13}C and/or ^{1}H chemical shifts. The method of choice in this chapter was use of equation 2.50 in Section 2.5.3, where σ_{ref} was determined from the average of the calculated isotropic shieldings (σ_{calc}) and experimental chemical shifts ($\delta_{i,exp}$). The benefit of this method is that the assignment does not need to be known. Once scaled and referenced, the environments are then grouped according to their experimental ^{13}C chemical shift. For Tyr, this resulted in 6 groupings but in this section, only group 4 will be discussed with the experimental and calculated ^{13}C and ^{1}H chemical shifts shown in table A2 below. There are 4 distinct experimental chemical shift values, therefore, there are 24 possible assignments.

Table A2: The experimental and calculated ^{13}C and ^{1}H chemical shifts for Group 4 of Tyr. Relevant calculation parameters are included in the main text in section 4.4.4.

Experimental ^{13}C (^{1}H) Chemical Shifts / ppm	**Possible Assignments**	**Calculated ^{13}C(^{1}H) Chemical Shifts / ppm**
126.8 (5.4)		C9 – 125.6 (5.1)
121.9	C4, C6-H6,	C4 – 122.4
118.5 (5.5)	C8-H8, C9-H9	C8 – 118.1 (5.5)
116.2 (6.2)		C6 – 114.8 (6.3)

A2.1 Stage 1

In Stage 1 of the proposed workflow, the agreement between the experimental and GIPAW calculated ^{13}C chemical shifts is evaluated first, followed by assessment of the ^{13}C NQS and short-contact ^{1}H-^{13}C HETCOR spectra. The data is displayed in figure 4.3 of in Section 4.4. As discussed previously, the ^{13}C NQS and short-contact ^{1}H-^{13}C HETCOR data should complement one another, therefore, are considered with each other in Stage 1.

^{13}C Chemical Shifts

Equation 4.4 is used to evaluate the agreement between a particular experimental and calculated ^{13}C chemical shift. The equation is reproduced in equation A1 below. The key

parameter to define is σ, which is nucleus specific. Here, as the ^{13}C environments are under consideration, the value of 2.3 is used.

$$P_{shift,A} = \prod_{k=1}^{N} 1 - N\left(\frac{|\delta_{calc,k} - \delta_{expt,k}|}{\sigma}\right) \quad \text{Eq. A1}$$

For each possible assignment of group 4, four individual probability values are determined and these are multiplied to determine P_{shift} for each possible assignment, as shown in figure 4.2. Table A3 demonstrates this where 24 P_{shift} values are calculated and subsequently evaluated using Bayes Theorem (Equation 4.5) to produce the AF values which only consider the ^{13}C chemical shift agreement.

^{13}C NQS and short-contact ^{1}H-^{13}C HETCOR Datasets

As described in Section 4.3, the individual error values ($e_{\lambda,i}$) are assigned based on the experiment type rather than calculated. This is demonstrated in tables A3 and A4 for both ^{13}C NQS and short-contact ^{1}H-^{13}C HETCOR datasets, where P_{NQS} and P_{SCHET} values are determined for each possible assignment and subsequent AF values that are experiment specific.

As shown in figure 4.2, there is an overall AF value for Stage 1. This is achieved using equation A2 and has been shown in table A5.

$$P_{Stage\,1}(A) = P_{shift,C}(A)P_{NQS}(A)P_{SCHET}(A) \quad \text{Eq. A2}$$

The methodology assumes that error is independent and random, therefore, this allows for easy multiplication of values to result in one AF value that incorporates all available information. Using an adaption of equation 4.7 in equation A2, the $P(A)$ values for each possible assignment of the group 4 environments can be determined from the multiplication of the relevant P_{shift}, P_{NQS} and P_{SCHET} values. Once the combined $P(A)$ values are determined for each of the 24 possible assignments, which are then evaluated using equation 4.5 to produce AF values which consider the data from the ^{13}C chemical shift agreement, the ^{13}C NQS spectrum and short-contact ^{1}H-^{13}C HETCOR spectrum.

Table A2: Calculation of $P_{shift,C}$ values for each of the possible assignments for Group 4 of Tyr alongside the subsequent AF values.[24] The possible assignments run from left to right.

Possible Assignments	$e_{shift,C}$ Experimental ^{13}C Chemical Shifts / ppm				$P_{shift,C}$	AF / %
	116.2	118.5	121.9	126.8		
C9 / C4 / C8 / C6	0.00	0.10	0.11	0.00	0.00	0
C9 / C4 / C6 / C8	0.00	0.10	0.00	0.00	0.00	0
C9 / C8 / C6 / C4	0.00	0.60	0.00	0.07	0.00	0
C9 / C8 / C4 / C6	0.00	0.60	0.58	0.00	0.00	0
C9 / C6 / C8 / C4	0.00	0.12	0.11	0.07	0.00	0
C9 / C6 / C4 / C8	0.00	0.12	0.58	0.00	0.00	0
C4 / C9 / C8 / C6	0.01	0.00	0.11	0.00	0.00	0
C4 / C9 / C6 / C8	0.01	0.00	0.00	0.00	0.00	0
C4 / C8 / C6 / C9	0.01	0.60	0.00	0.47	0.00	0
C4 / C8 / C9 / C6	0.01	0.60	0.12	0.00	0.00	0
C4 / C6 / C8 / C9	0.01	0.12	0.11	0.47	0.00	0
C4 / C6 / C9 / C8	0.01	0.12	0.12	0.00	0.00	0
C8 / C4 / C9 / C6	0.35	0.10	0.12	0.00	0.00	0
C8 / C4 / C6 / C9	0.35	0.10	0.00	0.47	0.00	0
C8 / C9 / C6 / C4	0.35	0.00	0.00	0.07	0.00	0
C8 / C9 / C4 / C6	0.35	0.00	0.58	0.00	0.00	0
C8 / C6 / C9 / C4	0.35	0.12	0.12	0.07	0.00	0
C8 / C6 / C4 / C9	0.35	0.12	0.58	0.47	0.01	13
C6 / C4 / C8 / C9	0.42	0.10	0.11	0.47	0.00	3
C6 / C4 / C9 / C8	0.42	0.10	0.12	0.00	0.00	0
C6/ C8 / C9 / C4	0.42	0.60	0.12	0.07	0.00	3
C6 / C8 / C4 / C9	0.42	0.60	0.58	0.47	0.07	81
C6 / C9 / C8 / C4	0.42	0.00	0.11	0.07	0.00	0
C6 / C9 / C4 / C8	0.42	0.00	0.58	0.00	0.00	0

[24] *For example, C9 / C4 / C8 / C6 means C9 is assigned to 116.2 ppm, C4 is assigned to 118.5 ppm, C8 is assigned to 121.9 ppm and C6 is assigned to 126.8 ppm.*

Table A3: Calculation of P_{NQS} values for each of the possible assignments for Group 4 of Tyr alongside the subsequent AF values.[25] The possible assignments run from left to right.

Possible Assignments	e_{NQS} Experimental ^{13}C Chemical Shifts / ppm				P_{NQS}	AF / %
	116.2	118.5	121.9	126.8		
C9 / C4 / C8 / C6	0.8	0.2	0.2	0.8	0.03	1
C9 / C4 / C6 / C8	0.8	0.2	0.2	0.8	0.03	1
C9 / C8 / C6 / C4	0.8	0.8	0.2	0.2	0.03	1
C9 / C8 / C4 / C6	0.8	0.8	0.8	0.8	0.41	14
C9 / C6 / C8 / C4	0.8	0.8	0.2	0.2	0.03	1
C9 / C6 / C4 / C8	0.8	0.8	0.8	0.8	0.41	14
C4 / C9 / C8 / C6	0.2	0.8	0.2	0.8	0.03	1
C4 / C9 / C6 / C8	0.2	0.8	0.2	0.8	0.03	1
C4 / C8 / C6 / C9	0.2	0.8	0.2	0.8	0.03	1
C4 / C8 / C9 / C6	0.2	0.8	0.2	0.8	0.03	1
C4 / C6 / C8 / C9	0.2	0.8	0.2	0.8	0.03	1
C4 / C6 / C9 / C8	0.2	0.8	0.2	0.8	0.03	1
C8 / C4 / C9 / C6	0.8	0.2	0.2	0.8	0.03	1
C8 / C4 / C6 / C9	0.8	0.2	0.2	0.8	0.03	1
C8 / C9 / C6 / C4	0.8	0.8	0.2	0.2	0.03	1
C8 / C9 / C4 / C6	0.8	0.8	0.8	0.8	0.41	14
C8 / C6 / C9 / C4	0.8	0.8	0.2	0.2	0.03	1
C8 / C6 / C4 / C9	0.8	0.8	0.8	0.8	0.41	14
C6 / C4 / C8 / C9	0.8	0.2	0.2	0.8	0.03	1
C6 / C4 / C9 / C8	0.8	0.2	0.2	0.8	0.03	1
C6/ C8 / C9 / C4	0.8	0.8	0.2	0.2	0.03	1
C6 / C8 / C4 / C9	0.8	0.8	0.8	0.8	0.41	14
C6 / C9 / C8 / C4	0.8	0.8	0.2	0.2	0.03	1
C6 / C9 / C4 / C8	0.8	0.8	0.8	0.8	0.41	14

[25] *P_{NQS} values are assigned based on whether the peak in the ^{13}C NQS spectrum has reduced in intensity relative to the ^{13}C CP/TOSS spectrum. This does require the acquisition parameters to be as identical as possible.*

Table A4: Calculation of P_{SCHET} values for each of the possible assignments for Group 4 of Tyr alongside the subsequent AF values.[26] The possible assignments run from left to right.

Possible Assignments	e_{SCHET} Experimental ^{13}C Chemical Shifts / ppm				P_{SCHET}	AF / %
	116.2	118.5	121.9	126.8		
C9 / C4 / C8 / C6	0.9	0.1	0.1	0.9	0.01	0
C9 / C4 / C6 / C8	0.9	0.1	0.1	0.9	0.01	0
C9 / C8 / C6 / C4	0.9	0.9	0.1	0.1	0.01	0
C9 / C8 / C4 / C6	0.9	0.9	0.9	0.9	0.66	16
C9 / C6 / C8 / C4	0.9	0.9	0.1	0.1	0.01	0
C9 / C6 / C4 / C8	0.9	0.9	0.9	0.9	0.66	16
C4 / C9 / C8 / C6	0.1	0.9	0.1	0.9	0.01	0
C4 / C9 / C6 / C8	0.1	0.9	0.1	0.9	0.01	0
C4 / C8 / C6 / C9	0.1	0.9	0.1	0.9	0.01	0
C4 / C8 / C9 / C6	0.1	0.9	0.1	0.9	0.01	0
C4 / C6 / C8 / C9	0.1	0.9	0.1	0.9	0.01	0
C4 / C6 / C9 / C8	0.1	0.9	0.1	0.9	0.01	0
C8 / C4 / C9 / C6	0.9	0.1	0.2	0.9	0.02	0
C8 / C4 / C6 / C9	0.9	0.1	0.1	0.9	0.01	0
C8 / C9 / C6 / C4	0.9	0.9	0.1	0.1	0.01	0
C8 / C9 / C4 / C6	0.9	0.9	0.9	0.9	0.66	16
C8 / C6 / C9 / C4	0.9	0.9	0.1	0.1	0.01	0
C8 / C6 / C4 / C9	0.9	0.9	0.9	0.9	0.66	16
C6 / C4 / C8 / C9	0.9	0.1	0.1	0.9	0.01	0
C6 / C4 / C9 / C8	0.9	0.1	0.1	0.9	0.01	0
C6/ C8 / C9 / C4	0.9	0.9	0.1	0.1	0.01	0
C6 / C8 / C4 / C9	0.9	0.9	0.9	0.9	0.66	16
C6 / C9 / C8 / C4	0.9	0.9	0.1	0.1	0.01	0
C6 / C9 / C4 / C8	0.9	0.9	0.9	0.9	0.66	16

[26] *P_{SCHET} values are determined based on whether a cross peak is observed in the short-contact ^{1}H-^{13}C HETCOR spectrum (independent of ^{1}H chemical shift).*

Table A5: Calculation of combined $P_{Stage\,1}(A)$ and subsequent AF values for each of the possible assignments for Group 4 of Tyr in Stage 1 of the proposed workflow. The possible assignments run from left to right. The cells highlighted in yellow refer to the values which remain after the AF > 0.01 filter is applied.

Possible Assignments (^{13}C) 116.2 / 118.5/ 121.9 / 126.8 / ppm	$P_{shift,C}$	$P_{NQS}P_{SCHET}$	$P_{Stage\,1}(A)$	AF / %
C9 / C4 / C8 / C6	0.00	0.00	0.00	0
C9 / C4 / C6 / C8	0.00	0.00	0.00	0
C9 / C8 / C6 / C4	0.00	0.00	0.00	0
C9 / C8 / C4 / C6	0.00	0.27	0.00	0
C9 / C6 / C8 / C4	0.00	0.00	0.00	0
C9 / C6 / C4 / C8	0.00	0.27	0.00	0
C4 / C9 / C8 / C6	0.00	0.00	0.00	0
C4 / C9 / C6 / C8	0.00	0.00	0.00	0
C4 / C8 / C6 / C9	0.00	0.00	0.00	0
C4 / C8 / C9 / C6	0.00	0.00	0.00	0
C4 / C6 / C8 / C9	0.00	0.00	0.00	0
C4 / C6 / C9 / C8	0.00	0.00	0.00	0
C8 / C4 / C9 / C6	0.00	0.00	0.00	0
C8 / C4 / C6 / C9	0.00	0.00	0.00	0
C8 / C9 / C6 / C4	0.00	0.00	0.00	0
C8 / C9 / C4 / C6	0.00	0.27	0.00	0
C8 / C6 / C9 / C4	0.00	0.00	0.00	0
C8 / C6 / C4 / C9	0.01	0.27	0.00	14
C6 / C4 / C8 / C9	0.00	0.00	0.00	0
C6 / C4 / C9 / C8	0.00	0.00	0.00	0
C6/ C8 / C9 / C4	0.00	0.00	0.00	0
C6 / C8 / C4 / C9	0.07	0.27	0.02	86
C6 / C9 / C8 / C4	0.00	0.00	0.00	0
C6 / C9 / C4 / C8	0.00	0.27	0.00	0

As mentioned in section 4.3.2, an AF value needs to exceed 90% to be considered a confident assignment where no further experimental validation is required. This has not been achieved by group 4 at the end of Stage 1. However, the AF > 0.01 (AF > 1%) threshold can be applied to the $P_{NQS}P_{SCHET}$ values to reduce the number of possible assignments considered in Stage 2. As highlighted in yellow, only 6 of the possible 24 assignments exceed this threshold. Within these 6 possible assignments, **C4** is consistently assigned to 121.9 ppm. Therefore, prior to Stage 2, **C4** can be confidently assigned to 121.9 ppm, thus removing the 121.9 ppm from consideration in Stage 2. This leaves 6 possible assignments for Stage 2.

A2.2 Stage 2

In Stage 2, the ^{1}H chemical shift agreement is evaluated. This uses the same methodology discussed previously for the ^{13}C chemical shift agreement with one key variation; the value of σ used. For ^{1}H, the value of 0.5 is used instead.

The 6 possible assignments were evaluated using equation A1. To note, for ease in the multiplication step, the order of the ^{1}H chemical shifts in table A6 reflects the ordering of the experimental ^{13}C chemical shifts in table A2 as the experimental ^{1}H chemical shifts were extracted from the short-contact ^{1}H-^{13}C HETCOR so each cross peak will have an experimental ^{13}C and ^{1}H chemical shift.

Table A6: Calculation of $P_{shift,H}$ values for each of the possible assignments for Group 4 of Tyr alongside the subsequent AF values.[27] The possible assignments run from left to right.

Possible Assignments	$e_{shift,H}$ **Experimental ^{1}H Chemical Shifts / ppm**			$P_{shift,H}$	AF / %
	6.1	5.5	5.4		
H9 / H8 / H6	0.5	0.98	0.61	0.30	9
H9 / H6 / H8	0.50	0.68	0.96	0.33	10
H8 / H9 / H6	0.79	0.88	0.61	0.42	13
H8 / H6 / H9	0.79	0.68	0.92	0.49	15
H6 / H9 / H8	0.95	0.88	0.96	0.81	25
H6 / H8 / H9	0.95	0.98	0.92	0.85	27

[27] *For example, H9 / H8 / H6 means H9 is assigned to 6.1 ppm, H4 is assigned to 5.5 ppm and H6 is assigned to 5.4 ppm.*

In order to incorporate the additional information from the ^{1}H chemical shift agreement, equation A2 can be adapted further, as shown in equation A3, where the Stage 1 $P(A)$ values for the relevant possible assignments can be simply multiplied by the $P_{shift,H}(A)$ to produce the $P_{Stage\ 1+2}(A)$ values. The subsequent AF values determined in table A7 now consider the ^{13}C and ^{1}H chemical shift agreement, the ^{13}C NQS data and short-contact ^{1}H-^{13}C HETCOR data. At the end of Stage 2, the 90% AF threshold has been reached so a confident assignment has been found for Group 4 of Tyr and nothing further is required.

$$P_{Stage\ 1+2}(A) = P_{Stage\ 1}(A)P_{shift,H}(A) \qquad \text{Eq. A3}$$

Table A7: Calculation of combined $P_{Stage\ 1+2}(A)$ and subsequent AF values for each of the possible assignments for Group 4 of Tyr at the end of Stage 2 of the proposed workflow. The possible assignments run from left to right.

Possible Assignments (^{13}C, ^{1}H) 116.2,6.1 / 118.5,5.5 / 126.8,5.4 / ppm	$P_{Stage\ 1}(A)$	$P_{shift,H}$	$P_{Stage\ 1+2}(A)$	**AF / %**
C9-H9 / C8-H8 / C6- H6	0.00	0.30	0.00	**0**
C9-H9 / C6-H6 / C8-H8	0.00	0.33	0.00	**0**
C8-H8 / C9-H9 / C6-H6	0.00	0.42	0.00	**0**
C8-H8 / C6-H6 / C9-H9	0.00	0.49	0.00	**9**
C6-H6 / C9-H9 / C8-H8	0.02	0.81	0.02	**91**
C6-H6 / C8-H8 / C9-H9	0.00	0.85	0.00	**0**

A2.3 Stage 3 using INH-2HCA as a worked example.

Stage 3 was not required for the worked example of Tyr, therefore, INH-2HCA has been used as an example to demonstrate the consideration of the long-contact ^{1}H-^{13}C HETCOR in Section 4.5.1. The incorporation of the long-contact ^{1}H-^{13}C HETCOR varies to the short-contact variant where the C-H distance needs to be known to evaluate against the observed cross peaks. Using group 7 of INH-2HCA as a worked example, the values of $e_{\lambda,i}$ are still assigned but the value used is relative to the predicted CH distance for a given correlation.

There are 6 possible assignments for environments **C3**, **C5** and **C8**. At the end of Stage 2, the 90% AF threshold has not been reached, therefore, the long-contact ^{1}H-^{13}C HETCOR is

required so the possible correlations, related C-H distances and predicted ^{1}H chemical shift for each environment have been calculated, as shown in table A8.

Table A8: Predicted ^{1}H-^{13}C correlations, related C-H distances, predicted ^{1}H chemical shifts and e_{lcHET} assignment if the cross peak is present or not.

Correlation	C-H Distance	GIPAW Calculated ^{1}H Chemical Shift / ppm	e_{lcHET} Assignment (Present / Not Present)
C3-H3	2.4	10.5	0.75 / 0.25
C3-H3A	1.1	7.5	0.9 / 0.1
C3-H4	2.1	6.8	0.75 / 0.25
C5-H4	2.2	6.8	0.75 / 0.25
C5-H5	1.1	5.8	0.9 / 0.1
C5-56	2.1	7.6	0.75 / 0.25
C8-H2A	2.8	4.0	0.6 / 0.4
C8-H6	2.7	7.6	0.6 / 0.4
C8-H7	2.1	8.5	0.75 / 0.25
C8-H8	1.1	8.0	0.9 / 0.1

Take the possible assignment, **C3** at 117.2 ppm, **C5** at 118.2 ppm and **C8** at 119.0 ppm. For **C3** at 117.2 ppm, the 3 predicted C-H correlations and their related ^{1}H chemical shifts are compared to the experimental data in figure 4.6. This is to determine whether the predicted cross peak could be present in the experimental data or not. Once evaluated, the appropriate e_{lcHET} is assigned and this is based on the predicted C-H distance. Once this is completed for each predicted CH correlation for each carbon environment in all possible assignments, the $P(A)$ and AF values can be calculated, as seen in Stage 1 and 2. In order to incorporate into overall AF value, equation A3 is adapted again to include the generated data, as shown below.

$$P_{Stage\ 1+2+3}(A) = P_{Stage\ 1+2}(A)P_{LCHET}(A) \qquad \text{Eq. A4}$$

A3: ^{1}H-^{13}C DQ/SQ Data for Tyr.

The F_1 projections of the long-contact ^{1}H-^{13}C DQ/SQ spectra in figure 6.15 are shown in figure A1 below. The −10 to 4 ppm region has also been shown to demonstrate how the cross peaks within this region are clearly SSB artefacts from the repetition in the F_1 projection alongside the RRF resonances discussed previously. The 10 kHz ^{1}H-^{13}C DQ/SQ spectrum in figure A1 b) is affected more by the RRF resonance cross peaks relative to the 9434 Hz ^{1}H-^{13}C DQ/SQ spectrum in figure A1 a) which in turn, complicates the spectrum.

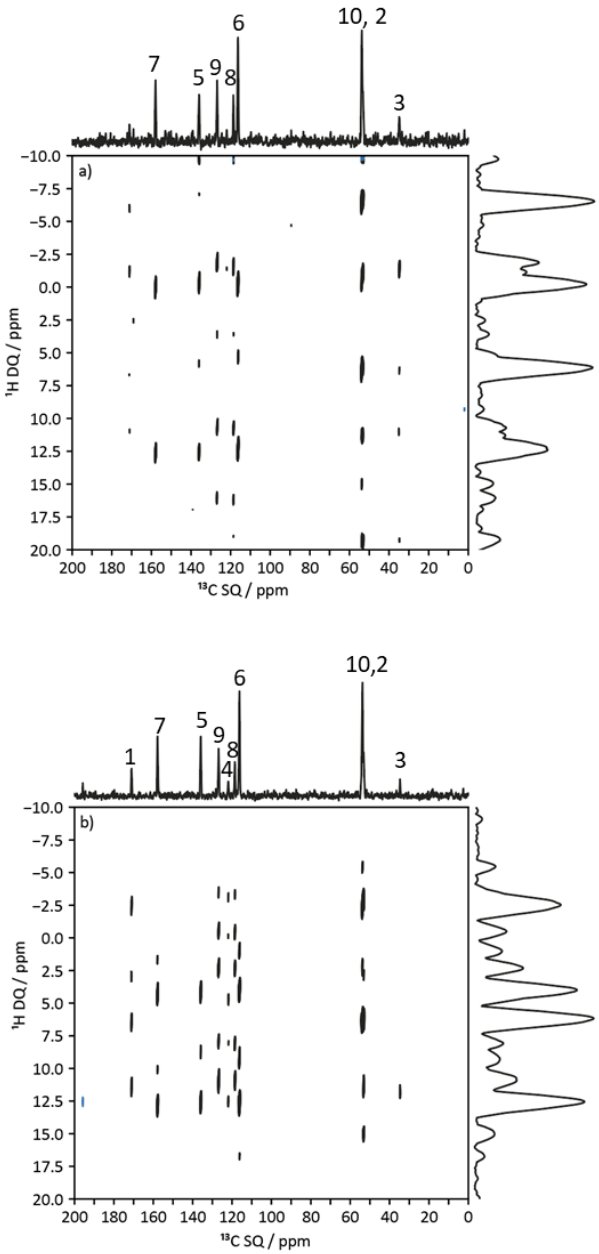


Figure A1: A long-contact ^{1}H-^{13}C DQ/SQ spectrum recorded of Tyr at a) 9434 Hz and b) 10 kHz (right). Acquisition parameters are captured in figure 5.16 in Chapter 5 where the projections for both F_1 and F_2 are displayed.

A4: Verification of RVR Starting Material

The PXRD pattern to verify the form of RVR is shown in figure A2. Relative to the literature, the purchased form is Form I.[19]

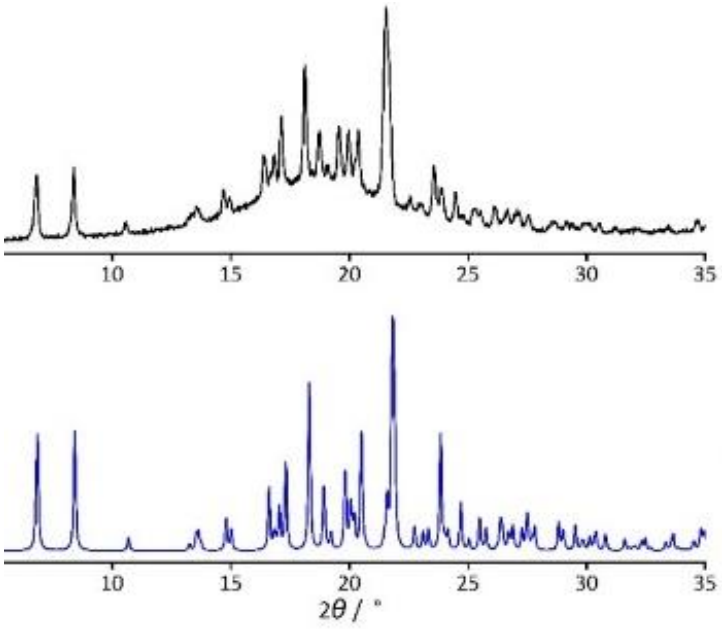


*Figure A2: Experimental PXRD pattern (top) compared to the simulated PXRD spectrum for Form I (**YIGPIO02**) of RVR. Simulated PXRD pattern was created in Mercury.*[113]

A5: Conformer Ratios

To determine the quoted approximate conformer ratios in section 6.3.2, a selection of environments were considered and have been tabulated in table A9.

Table A9: Environments considered in the conformer ratio calculation in each solvent.

Solvent	^{1}H Environment	Ratio (relative to environment 40)
$CDCl_3$	**40**	6.5 : 1
	1&3	5.7 : 1 / 5.2: 1
DMSO	**40**	12.2 : 1
	38	10.7 : 1
	22	9.9 : 1
	42	12.0 : 1
	34	10.7 : 1
	26	9.6 : 1
EtOD	**40**	6.8 : 1
	38	5.2 : 1
	22	5.3 : 1
	1&3	6.1 : 1

A6: Calculation of Intramolecular Hydrogen Bonding (intraHB)

In principle, the ^{1}H chemical shifts can be used to determine whether an intramolecular hydrogen bond (intraHB) is present. The method, published by Abraham and co-workers uses the difference in chemical shift of an exchangeable proton in DMSO and $CDCl_3$ to determine whether an intraHB could be possible.[249–251] The hydrogen bond acidity (A_{NMR}) is defined as the tendency of a molecule to act as a hydrogen bond donor (HBD) or hydrogen bond acceptor (HBA), where lower values indicate the presence of an intraHB.[251] The subscript notation of NMR is used to distinguish the method from alternate methodologies to determine *A*, such as from water-solvent partition coefficients (log *P*).[274]

Using equation A1, values of A_{NMR} could be determined for **34**, **21** and **5** in RVR. Unfortunately, the OH group (**25**) could not be evaluated as the ^{1}H chemical shift in $CDCl_3$ was not assigned. Using table A10, **34** and **21** had no intraHB (A_{NMR} = 0.23 and 0.14, respectively) whilst **5** (A_{NMR} = −0.25) indicated that there was a possibility of an intraHB. However, the plausibility of **5** engaging in an intraHB is ambiguous so further MD simulations would be required to verify whether the intraHB is a possibility.

$$A_{NMR} = 0.0065 + 0.133\left(\delta_H(DMSO) - \delta_H(CDCl_3)\right) \qquad \text{Eq. A6}$$

Table A10: Proposed values of A_{NMR} in the assessment of intramolecular hydrogen bonding (intraHB) for different exchangeable proton environments. Adapted from reference 251.

Proton Type	Possible intraHB	No intraHB
NH	$A_{NMR} < 0.05$	$A_{NMR} > 0.15$

A7: Calculation of k_{Mm} values Using the DMA Approach

In Chapter 6, from the values of *K* alone, it was concluded that the DMA approach was not being used to estimate the exchange rates of RVR in each solvent. The related values of k_{Mm} have been tabulated in table A11.

Table A11: Summary of the calculated values of k_{Mm} for RVR in $CDCl_3$, DMSO and EtOD from DMA. The associated uncertainties are determined from the environment specific noise level derived from the iterative fitting.

Environment	**k_{Mm} / s^{-1}**		
	Mixing Time / s		
$CDCl_3$	**0.05**	**0.1**	**0.2**
40	2.4 ± 0.5	2.5 ± 0.6	1.1 ± 0.7
26	2.3 ± 0.9	1.7 ± 0.8	1.0 ± 0.7
DMSO	**0.1**	**0.2**	**0.3**
46 & 50	0.4 ± 0.02	0.4 ± 0.01	0.4 ± 0.01
40	0.4 ± 0.02	0.4 ± 0.01	0.4 ± 0.01
38	0.4 ± 0.05	0.4 ± 0.03	0.5 ± 0.05
34	0.3 ± 0.2	0.4 ± 0.2	0.2 ± 0.2
26	−0.04 ± 0.3	0.2 ± 0.3	0.2 ± 0.3
25	0.4 ± 0.2	0.3 ± 0.1	0.4 ± 0.2
24	1.3 ± 0.3	0.5 ± 0.3	0.6 ± 0.3
22	0.7 ± 0.2	0.1 ± 0.2	0.4 ± 0.2
EtOD	**0.05**	**0.1**	**0.2**
46 & 50	0.3 ± 0.2	0.3 ± 0.1	0.2 ± 0.1
40	0.5 ± 0.1	0.6 ± 0.1	0.7 ± 0.04
38	0.5 ± 0.1	0.6 ± 0.1	0.6 ± 0.1
34	-1.5 ± 0.7	1.0 ± 0.7	0.8 ± 0.5
22	-0.1 ± 0.4	0.5 ± 0.3	0.6 ± 0.3

A8: Arrhenius Plots for CRB and NMU Environments

From the VT ^{1}H-^{1}H EXSY experiments conducted in EtOD, thermodynamic parameters could be derived from the resulting Arrhenius plots for the CRB torsion (figure A3) and NMU torsion (figure A4).

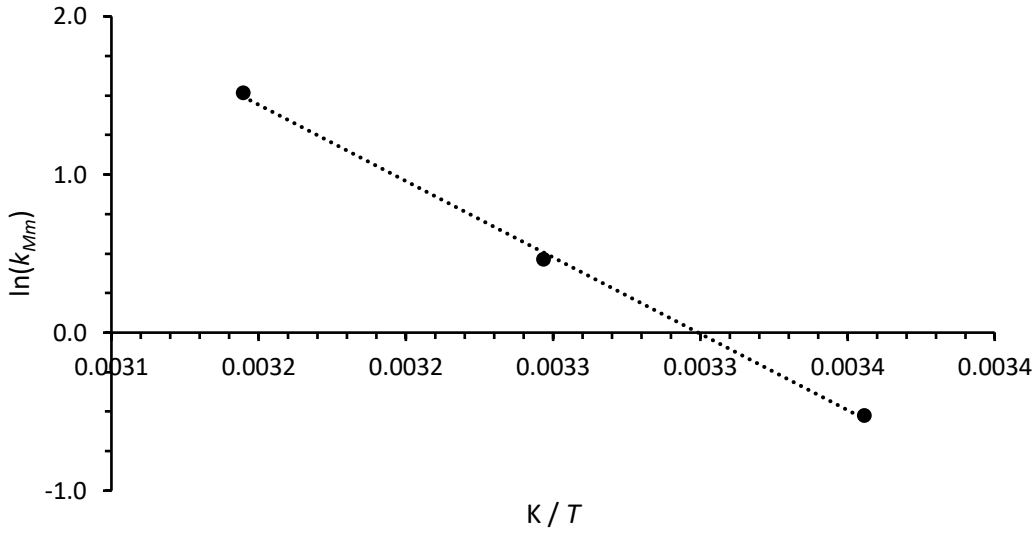


Figure A3: Arrhenius plot for the CRB interconversion (y = −9657.6x + 31.863).

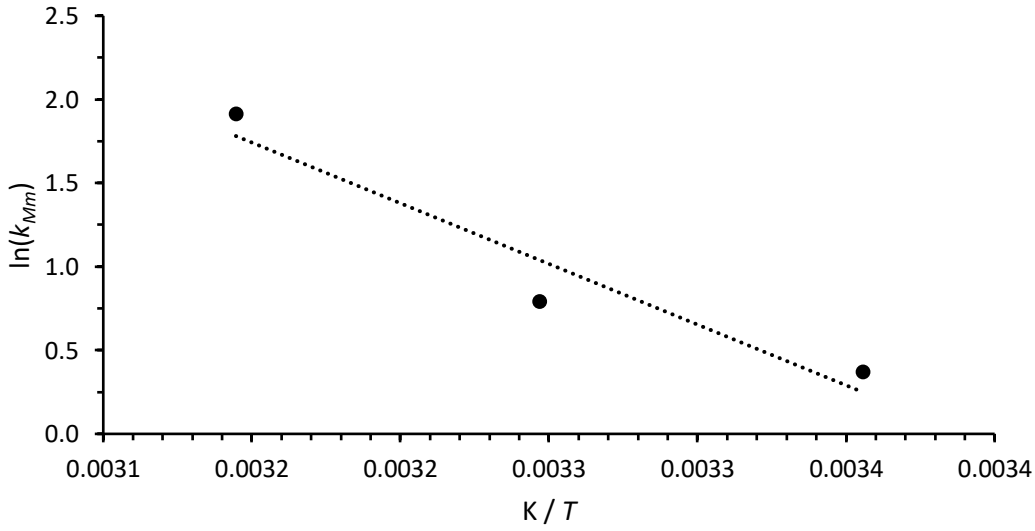


Figure A4: Arrhenius plot for the NMU interconversion (y = −7258.6x + 24.607).