NMR CRYSTALLOGRAPHY OF DISORDERED COCRYSTALS

By

Hannah Elin Kerr

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

Department of Chemistry

2016

Abstract

Crystallographic disorder is common in the solid state but it is rarely investigated explicitly despite having a fundamental impact on the solid-state structure of a material. In this work, nuclear magnetic resonance (NMR) crystallography methods are utilised to achieve a detailed understanding of the structure and dynamics of solid organic systems containing disorder. Several new cocrystal systems are studied, each containing a topical drug molecule (caffeine, naproxen or furosemide) and each serving to demonstrate how NMR crystallography can be applied to a variety of structural questions. Hydrogen bonding motifs are identified using single crystal X-ray diffraction experiments, where possible, and are subsequently verified by solid-state NMR. Alternative hydrogen bonding models are ruled out by comparison of experimental solid-state NMR data with density functional theory calculated shieldings, and proton transfer can be investigated by monitoring the energy of the system with respect to proton position. In a particularly challenging case, 2D solid-state NMR experiments go some way to identify the hydrogen bonds in a system that cannot be crystallised. Dynamic disorder of fragments and solvent molecules are characterised by variable temperature solid-state NMR, including analysis of relaxation times to establish energy barriers and rates of motion. A mechanism of motion is also proposed for dynamic acetone molecules in a new cocrystal solvate, which is supported by good agreement between experimental and simulated ^{2}H static NMR line shapes. Finally, the current limit of NMR crystallography is identified with respect to the reproducibility of calculated NMR parameters following geometry optimisation. It is shown that the geometry optimisation protocol does not affect standard NMR crystallography investigations pertaining to atom assignment, but it is significant for cases where very subtle structural features are probed, such as NMR linewidths. Overall, NMR crystallography investigations allow a deeper understanding of solid materials to be achieved than would be possible with any single technique and this work highlights the applicability of such methods to complex materials containing disorder.

Acknowledgements

First and foremost I would like to thank my supervisor, Paul, for finding the perfect balance between “hands-on” and “hands-off” supervision and for always answering my silly questions with patience. I have learnt so much and will always be grateful for the belief placed in me. My sincere thanks to Ivana for providing a fresh perspective, teaching me about diffraction, and for providing me with most of the systems studied in this work. A special thank you must go to David Apperley, without whom I would not have been able to run a single NMR experiment. Thank you for taking so many hours to help me set up experiments, fix problems (both trivial and serious) and for our long chats during cryogen fills or while watching FIDs slowly acquire.

I am grateful to everyone who has been a member of the SSNMR group during my time at Durham; Robin (for sharing his years of experience over coffee), Martin (for teaching me how to run CASTEP and to use a command line for the first time), Eric, Fraser, Ilya, Marcin, Martins and Agris. A special thank you to Cory for keeping me company in the desk room and for all the fun debates we have had over coffee. I hope I can acquire an eye for detail to rival yours one day! Also to Will and Tavleen, for all our procrastination chats and the mountains of cake!

Thanks to my family for instilling a love of science in me from the very beginning. Finally, thank you to Kevin for all the love and support you have given me, as well as the late night chemistry debates and all your wonderful cooking.

Abbreviations

ABMS	Anisotropy of the bulk magnetic susceptibility
ADPs	Anisotropic displacement parameters
AL	Alanine
APD	4-aminopyridine
API	Active pharmaceutical ingredient
BABA	Back-to-back
BPY	Bipyridine
CAF-CA	1:1 caffeine citric acid
CAF-2CA	Caffeine-citric acid 1:2 cocrystal hydrate
2CAF-HBA	bis(caffeine) 4-hydroxybenzoic acid
CAF-HNA	caffeine 6-hydroxy-2-naphthoic acid
2CAF-PCA	bis(caffeine) *p*-coumeric acid hydrate
CASTEP	Cambridge Serial Total Energy Package
CBZ	Carbamazepine
CBZ-TFA	Carbamazepine trifluoroacetic acid
CP	Cross polarization
CRAMPS	Combined rotation and multiple pulse spectroscopy
CSA	Chemical shift anisotropy
CSD	Crystal Structure Database
CSP	Crystal structure prediction
CW	Continuous wave
CYT	Cytosine
DFT	Density functional theory
DMSO	Dimethyl sulfoxide
DMU-OA	Dimethylurea-oxalic acid
DNP	Dynamic nuclear polarisation
DQ/SQ	Double quantum/single quantum
DSC	Differential scanning calorimetry
DUMBO	Decoupling using mind-boggling optimisation
DVS	Dynamic vapour sorption
EAFUS	Everything added to food in the United States

EFG	Electric field gradient
EXPRESS	Exchange program for relaxing spin systems
FDA	Food and Drug Administration
FID	Free induction decay
FS	Furosemide
2FS-INA	Furosemide-isonicotinamide 2:1 cocrystal
FSLG	Frequency-switched Lee-Goldberg
FS-PA	Furosemide-picolinamide 1:1 cocrystal
FT	Fourier Transform
FWHM	Full width at half maximum
GIPAW	Gauge-included projector augmented wave
GRAS	Generally regarded as safe
HETCOR	Heteronuclear correlation
HMQC	Heteronuclear multiple quantum correlation
INADEQUATE	Incredible natural abundance double quantum transfer experiment
INA-N-O	isonicotinamide-*N*-oxide
IND-NA	Indomethacin-nicotinamide
IR	Infra-red spectroscopy
LAG	Liquid-assisted grinding
LG-CP	Lee-Goldberg cross polarization
lph	Litres per hour
MAS	Magic angle spinning
MD	Molecular dynamics
MQ	Multiple quantum
MU-TCA	N-methylurea trichloroacetic acid
NMR	Nuclear magnetic resonance spectroscopy
NPX	Naproxen
NPX-PA	Naproxen-picolinamide 1:1 cocrystal
NQR	Nuclear quadrupole resonance
NQS	Non-quaternary suppression
OGL	N-octylglucamine
OTFG	On-the-fly-generated
PABA	p-aminobenzoic acid
PBE	Perdew, Burke and Ernzerhof

ppm	Parts-per-million
PPZ	Piperazine
PR	Proline
PXRD	Powder X-ray diffraction
RF	Radiofrequency
rINADEQUATE	Refocused INADEQUATE
RMSDs	Root-mean-square differences
SCXRD	Single crystal X-ray diffraction
SOLA	Solids line shape analysis
SSNMR	Solid-state nuclear magnetic resonance spectroscopy
TBPE	Trans-1,2-bis(4-pyridyl)ethylene
TEM	Transmission electron microscopy
TMS	Tetramethylsilane
TOSS	Total sideband suppression
TP	Tryptophan
TY	Tyrosine
USP	Ultra-soft pseudopotential
XRD	X-ray diffraction

Table of Contents

Chapter 1: Introduction

1.1 General remarks

The majority of pharmaceutical materials are available on the market in a solid form, 80–90 %, and it is well known that the solid-state structure can affect important properties such as solubility, stability and manufacturability.[1] It is vital for pharmaceutical companies to characterise the solid structure of drugs in order to understand their properties and how these relate to biological function. The infamous case of polymorphism in ritonavir[2] serves as an example of how a lack of comprehensive understanding of the solid form can lead to costly and challenging reformulation problems.[3] Patent applications also require specific structural characterisation, and small structural changes can fulfil the requirements for a new patent on an old blockbuster system.[4] However, it is not straightforward to gain structural information for pharmaceutical materials. For example, the bulk active pharmaceutical ingredient (API) may be far from an ordered crystalline form. Indeed, it has been found that drugs containing crystallographic disorder offer potentially significant improvements to solubility and bioavailability, as highlighted in a review by Hancock.[5] Crystallographic disorder is common and comes in many forms including long-range disorder in amorphous systems and short-range disorder in crystalline organics, where atoms or molecules may be statically or dynamically disordered.

Solid-state nuclear magnetic resonance spectroscopy (SSNMR) plays an important role in the characterisation of pharmaceutical solids, particularly in the latter stages of drug discovery. Regulatory authorities such as the Food and Drug Administration (FDA) make specific mention of SSNMR as a recommended technique for monitoring polymorph formation in new drug substances[6,7] due to its sensitivity to local structure.[8] Even as a "fingerprint" SSNMR is a powerful technique for discriminating between different solid forms and it is now relatively common for researchers to extract atomic-level structural information, such as conformation and nuclear proximities. SSNMR provides a rich variety of structural handles, which has led to the emergence of the field of NMR crystallography.[9] This term describes studies of the atomic-level structure of materials using primarily NMR

spectroscopy, often in combination with complementary techniques like X-ray diffraction (XRD).

This work focuses specifically on the use of SSNMR, in the context of NMR crystallography, to study the structure and dynamics of organic compounds of pharmaceutical relevance that contain crystallographic disorder. Such systems are relevant to the ongoing search for more soluble forms of well-known drugs, and detailed structural studies of disordered molecular organics are still relatively under-reported in the literature. Therefore, there is scope for NMR crystallography to provide a more detailed understanding of the structure and dynamics of these materials with a view to understanding why crystallographic disorder is present in some systems but not in others.

The remainder of this introductory chapter introduces the main techniques utilised in this work. First, the theoretical basis of nuclear magnetic resonance spectroscopy (NMR) is introduced in the context of the solid state. Complementary techniques are then introduced, specifically diffraction methods and density functional theory calculations. This is followed by a discussion of solid form diversity, highlighting the information that can be gained from NMR crystallography studies using relevant examples from the literature.

1.2 Nuclear magnetic resonance spectroscopy

There are many text books that outline the theoretical background to NMR, so the following discussion will discuss the aspects that are relevant to this work. There are several approaches that describe the basics of the NMR experiment, the simplest of which is the vector model used in the discussion below. The reader is referred to a number of texts, in particular "Spin Dynamics" by M. Levitt[10] for a comprehensive review of NMR theory and also "Solid-State NMR: Basic Principles and Practise" by D. Apperley, R. K. Harris and P. Hodgkinson for a more practical approach.[11]

1.2.1 Vector model

Nuclei with a non-zero spin angular momentum, $\mathbf{I}$, have a nuclear magnetic moment, $\boldsymbol{\mu}$.* These two properties are related by $\mathbf{I} = \gamma\boldsymbol{\mu}$, where γ is the gyromagnetic ratio. The magnitude of the angular momentum is determined by the spin quantum number, *I,* which varies from zero upwards in half-integer steps. This is a fundamental property of every nucleus that determines the number of spin states, $2I + 1$, which are distinguished by a magnetic quantum number, m_I.

* In this discussion symbols in bold typeface are vectors.

For an ensemble of spins, in the absence of an external magnetic field, the magnetic moments are randomly distributed so there is no net nuclear magnetisation, **M**, and the different spin states are degenerate. If a strong magnetic field, $\mathbf{B}_0$, is applied the magnetic moments of the spins align such that the *z* components lie in the direction of $\mathbf{B}_0$, resulting in a non-zero **M**. The magnetic moments also precess about the axis of the applied field at ν_{NMR}, the Larmor frequency, Figure 1. By convention $\mathbf{B}_0$ is applied along the +*z* axis.

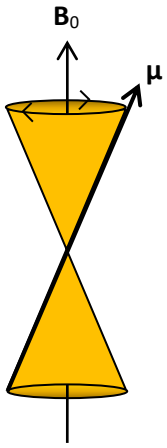

Figure 1 The nuclear magnetic moment, **μ**, of a single spin precessing about the axis of the external field, $\mathbf{B}_0$.

The Larmor frequency, ν_{NMR}, is related to the magnitude of the applied field, B_0, and the gyromagnetic ratio by Equation (1.0). The energies of the spin states are affected by the applied field and they lose their degeneracy by the Zeeman Effect, Equation (1.1).[12] NMR spectroscopy is based upon transitions between these non-degenerate Zeeman states.

$$\nu_{NMR} = -\frac{\gamma B_0}{2\pi} \qquad \textbf{(1.0)}$$

$$E = -m_I \hbar \gamma B_0 \qquad \textbf{(1.1)}$$

The population of the spin states, *N*, is determined by the Boltzmann distribution, Equation (1.2), where *N*(*upper*) and *N*(*lower*) are the populations and E(*upper*) and E(*lower*) are the energies of the upper and lower spin states respectively. At room temperature the difference in energy between the states is very small so the populations are almost equal. An illustrative example[13] is that of protons in a field of 11.74 T. The energy difference between the Zeeman states is roughly 3×10^{-25} J, which is four orders of magnitude smaller

than the available thermal energy at room temperature, $k_B T \approx 4 \times 10^{-21}$ J. As a result, NMR is an intrinsically insensitive technique. This is a particular problem for spins with a low natural abundance, such as ^{15}N at 0.36 %. Traditionally, higher magnetic fields are employed to enhance the sensitivity. Significant advances are currently being made in the field of dynamic nuclear polarisation (DNP), where large thermal polarisation is induced in the electron spins and then transferred to nuclear spins by microwave irradiation.[14,15]

$$\frac{N(upper)}{N(lower)} = e^{\frac{E_{upper} - E_{lower}}{k_B T}} \sim \frac{\hbar \gamma B_0}{k_B T} \qquad \textbf{(1.2)}$$

1.2.2 Rotating frame

During NMR experiments the populations of the spin states are perturbed by the application of radiofrequency (RF). These electromagnetic waves have an associated magnetic field, $\mathbf{B}_1$, which oscillates at a particular frequency, ν_{RF}. When applied to a spin system such that $\nu_{RF} = \nu_{NMR}$ the RF is said to be on resonance. At this point it is conventional to describe the NMR experiment using a rotating frame of reference that rotates about $\mathbf{B}_0$ at ν_{NMR}, Figure 2(i). In the rotating frame the applied field, $\mathbf{B}_1$, appears static along an axis in the *xy* plane by convention, Figure 2(ii). **M** interacts with $\mathbf{B}_1$ and precesses at a nutation frequency, ν_1, and $\mathbf{B}_0$ is effectively removed. Note that RF does not have to be exactly on resonance, in which case some component of $\mathbf{B}_0$ remains and the net magnetisation precesses about $\mathbf{B}_{eff}$.

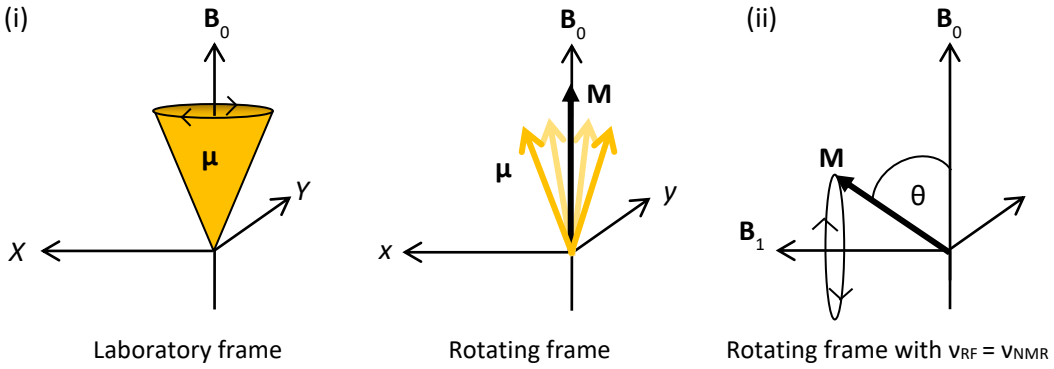

Figure 2 (i) Magnetic moments of multiple spins precess about $\mathbf{B}_0$ in the laboratory frame but each appear stationary in the rotating frame. **(ii)** On application of $\mathbf{B}_1$ (along the *y* axis in this schematic) **M** is tipped away from $\mathbf{B}_0$ by θ, the flip angle.

RF is usually applied in short pulses. The phase of the pulse is defined by the angle $\mathbf{B}_1$ makes with the *x* axis in the rotating frame. It is common for pulse sequences to be repeated with the phases of the pulses varied, known as phase cycling. The purpose of this is to select NMR signals with particular properties while removing unwanted signals.[16] The flip angle of the pulse, θ, is the angle **M** is perturbed by during the pulse, Figure 2(ii), and is related to the pulse duration, τ, by Equation (1.3).[17]

$$\theta = \gamma B_1 \tau \qquad \textbf{(1.3)}$$

NMR samples usually possess many different components with slightly different NMR frequencies, so the RF must be applied with a certain bandwidth. This bandwidth is inversely proportional to τ. Short pulses are on the μs timescale and they excite broad bandwidths (>10 kHz). Long pulses allow for selective excitation (<100 Hz bandwidth) and tend to be of ms duration. The power of an RF pulse is described using the nutation frequency ν_1, which is related to τ_{90°; the pulse duration that rotates **M** by 90˚ about the *x* axis in the rotating frame, Equation (1.4).[18] Note that ν_1, the nutation frequency resulting from the RF pulse, is not the same as ν_{RF}, the frequency at which the RF is applied.

$$\nu_1 = \frac{1}{4\tau_{90^\circ}} \qquad \textbf{(1.4)}$$

There is a maximum RF power for all NMR probes after which too high a voltage is generated, which can cause the surrounding air to be ionised, as well as affecting different components of the probe itself. This generates a discharge, or arc, that can be large enough to swamp the NMR signal, Figure 3. This not only renders the $\mathbf{B}_1$ field erratic but can also seriously damage the probe.

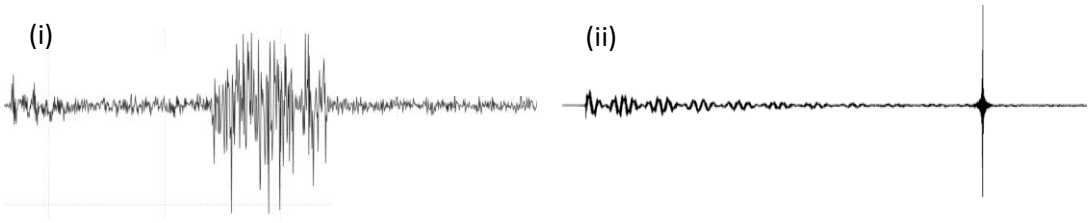

Figure 3 Screen captures of arcing in the FID during acquisition. **(i)** Arcing caused by high RF power ionising air near the probe coil. **(ii)** An arc thought to be caused by the sample rotor touching the coil following unstable spinning.

The signal that is detected in an NMR experiment is caused by the precession of **M** about $\mathbf{B}_0$ when it has been rotated into the *xy* plane by an RF pulse. The precessing magnetisation induces an oscillating electric current in a wire coil placed near the sample. This oscillating current is recorded along two of the orthogonal axes in the rotating frame, $+x$ and $-y$ for example. This is called quadrature detection and the combination of the signal from these two axes gives a complex time-dependent Free Induction Decay (FID). Fourier transform (FT) converts FIDs into the familiar spectra.

1.2.3 Relaxation

Once an RF pulse is switched off, **M** decays back to equilibrium through various relaxation mechanisms that each occur at different timescales. Spin-spin, or transverse, relaxation describes the dephasing of **M** in the *xy* plane perpendicular to $\mathbf{B}_0$, Figure 4. The mechanism of spin-spin relaxation involves the interaction of spins with one another without energy being transferred to the surrounding lattice. The time constant associated with this process is T_2, which is usually on the ms timescale for solids.

Spin-lattice, or longitudinal, relaxation describes the re-equilibration of the population of the spin states, resulting in **M** returning to the $\mathbf{B}_0$ axis, Figure 4. In this case, interactions between spins result in energy being transferred to the surrounding lattice. This relaxation is affected by slower local dynamics in the system. Fast rotation of methyl groups enhances the relaxation, as do other molecular motions such as alkyl chain dynamics. The time constant for this process is T_1, and it tends to be of the order of several seconds in solids. Note that T_1 is always greater than T_2 in the solid state, but in solution $T_1 = T_2$ as shown in Figure 4. The T_1 relaxation time of a system determines how quickly successive RF pulses can be applied; if the magnetisation has not fully relaxed before the next pulse then only a small portion of the spins will be re-excited. For full relaxation to occur the time between pulses should be at least $5T_1$, at which point the magnetisation has relaxed by 99 %.[19]

The final type of relaxation is $T_{1\rho}$, which follows the same principle as T_1 except the magnetisation decays to the axis of a spin-lock RF pulse in the rotating frame, rather than to $\mathbf{B}_0$. $T_{1\rho}$ is sensitive to molecular motions on kHz timescales, as opposed to T_1, which is sensitive to motions of the order of the NMR frequency.[19]

It is useful to mention here that the strong coupling between ^{1}H spins in solids results in spin diffusion.[20] This is the spatial redistribution of magnetisation from a local excited site to distant sites, without any physical diffusion of molecules.[21] It is mediated by dipolar

coupling, described in section 1.2.4. The effect of this is that all the ^{1}H spins in a homogeneous sample tend to behave similarly, with a uniform T_1 and $T_{1\rho}$, unless the MAS frequency is fast enough to remove a significant contribution of the dipolar coupling.

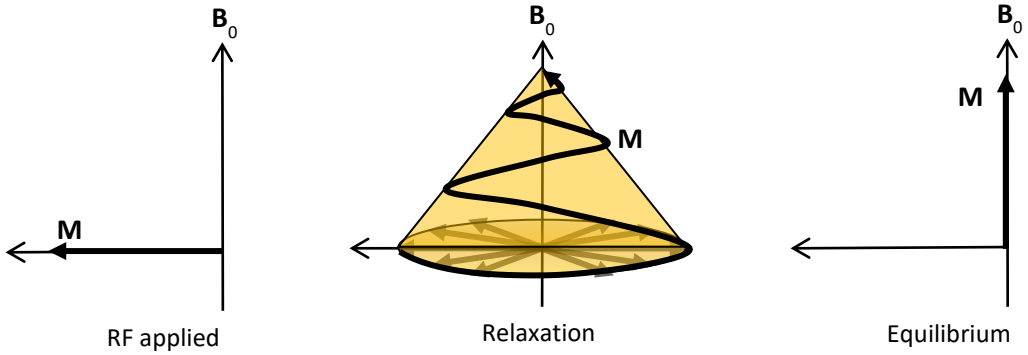

Figure 4 A schematic of the relaxation of **M** at the $T_1 = T_2$ condition. T_2 relaxation causes **M** to dephase in the *xy* plane, *i.e*. around the base of the cone. A combination of both T_2 and T_1 relaxation causes **M** to dephase around the surface of the cone until it returns to equilibrium.

1.2.4 NMR interactions

Chemical shift and indirect coupling

There are several interactions that are present in an NMR experiment and one of the most important is chemical shielding. Nuclei are surrounded by a non-symmetrical cloud of electron density and, in the presence of $\mathbf{B}_0$, this electron density induces a small magnetic field that opposes the applied field as the electrons circulate. This reduces the field felt by the nearby nuclei, $\mathbf{B}_{nuc}$, shown in Equation (1.5) where σ is a shielding constant.

$$\mathbf{B}_{nuc} = (1 - \sigma)B_0 \qquad \textbf{(1.5)}$$

The energies of the Zeeman states are also affected by this magnetic shielding effect, usually in an anisotropic manner due to the unsymmetrical nature of the local electron density. It is difficult to measure magnetic shielding directly[22] so chemical shifts, δ, are measured instead with units of parts-per-million (ppm). These are linearly field dependent to a good approximation,[23] and are usually reported relative to a reference compound such as tetramethylsilane (TMS), Equation (1.6). ν and ν_{ref} are the NMR frequencies of the compound of interest and the reference sample respectively, measured at the same

applied field. The orientation-dependent contribution to chemical shift is the chemical shift anisotropy (CSA) and is discussed in more detail in section 1.2.5.

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} \times 10^6 \quad \textbf{(1.6)}$$

The electron density surrounding nuclei allows neighbouring spins to interact with one another. This interaction is termed indirect spin-spin coupling or *J*-coupling. This is usually an intramolecular interaction and provides a direct link between NMR and chemistry. *J*-coupling is anisotropic due to the unsymmetrical nature of the electron density, but the magnitude is small compared to other interactions present in the solid state so it does not usually impact the appearance of spectra, though its relationship to connectivity can be exploited,[21] see section 2.1.4.

The dipolar interaction

Just as two classical magnets will interact with one another through space, so too do nuclear spins through a dipolar interaction. At high field the energies of the Zeeman states are affected as shown in Equation (1.7) where D_{IJ} is the dipolar coupling constant between two different nuclei, I and J, and θ is the angle of the internuclear vector with respect to $\mathbf{B}_0$. The coupling constant is related to $1/(d_{I\cdots J})^3$ where $d_{I\cdots J}$ is the through-space distance between the two spins. Generally, only nuclei that are closer than 3 Å show significant dipolar coupling.[24] The dipolar interaction does not have an impact on solution-state NMR spectra, other than on the relaxation of the spin system, because E_{IJ} averages to zero over the isotropic tumbling motion that molecules undergo in solution. However, it does play a significant role in SSNMR spectra as discussed further in section 1.2.5.

$$E_{IJ} = \frac{1}{2}(3\cos^2\theta - 1)D_{IJ} \quad \textbf{(1.7)}$$

The quadrupolar interaction

Quadrupolar nuclei make up more than 74 % of the periodic table[25] and possess $I > ½$. This means the charge distribution in the nucleus is not uniform, so it can interact with the local electric field gradient (EFG) that results from the nearby electron density. The quadrupolar interaction is, therefore, an electric interaction rather than a magnetic one and it is usually several orders of magnitude smaller than the Zeeman interaction. The magnitude is defined by the quadrupole coupling constant, C_Q, and the asymmetry, η, which can vary between zero and one. C_Q values have a large range,[26] from kHz to hundreds of MHz, but the

quadrupolar nucleus considered in this report, ^{2}H, generally has a moderate C_Q between 50–180 kHz.[27] The spin quantum number of ^{2}H is $I = 1$, so there are three spin states and two allowed transitions, Figure 5(i), and the spectrum is a doublet in the presence of a non-zero C_Q, Figure 5(ii).

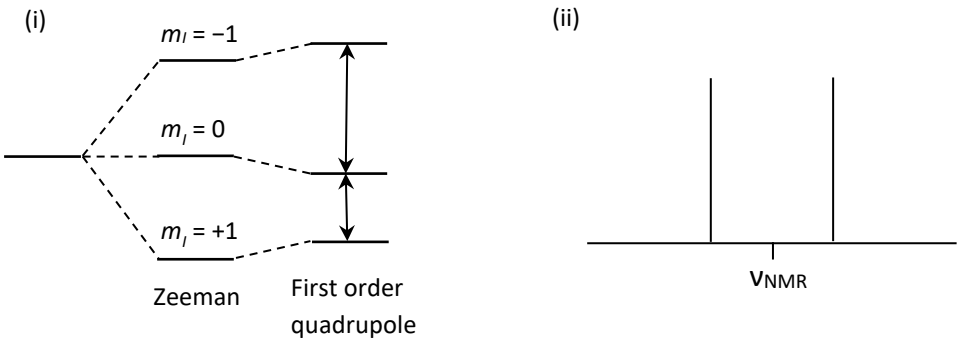

Figure 5 (i) Perturbation of the spin states by the quadrupolar interaction when $I = 1$. **(ii)** Schematic of an NMR spectrum when $I = 1$.

1.2.5 NMR in the solid state

The interactions described above are all potentially orientation-dependent with respect to $\mathbf{B}_0$. Molecules in solution tumble isotropically on a fast timescale relative to the NMR experiment so the anisotropic components average to zero. However, molecules in the solid state are generally locked in a specific orientation, meaning the interactions remain anisotropic. This adds an extra level of information that can be obtained but simultaneously complicates the appearance of spectra by broadening the spectral lines. The line broadening is so extensive that often a single broad lump is observed in a static SSNMR spectrum.

In the 1950s it was found that mechanically spinning a solid sample in the NMR probe at an angle of 54.74˚ with respect to the $\mathbf{B}_0$ field results in remarkable narrowing of the spectral lines.[28] This is because all the NMR interactions contain the orientation-dependent term $(3\cos^2\theta - 1)$. When $\theta = 54.74^\circ$, dubbed the magic angle, this term becomes zero thus simulating the isotropic tumbling that occurs in solution. Magic angle spinning (MAS) is now ubiquitous in SSNMR and ever faster speeds are being achieved with the development of state-of-the-art ultrafast MAS probes. At the time of writing, the current record for commercially available MAS probes is 111 kHz.[29] Despite the advances in MAS probe

design, further mechanical development is necessary. It was recently estimated that speeds of up to 250 kHz will be required to obtain solution-like resolution in SSNMR of proteins.[30] This is because the magnitudes of some of the NMR interactions are too large for current MAS speeds to fully average.

The dipolar interaction in organic solids is typically moderate at 2 kHz for ^{13}C–^{13}C coupling,[21] so low MAS speeds are sufficient (2–10 kHz). However, dipolar coupling involving protons in the solid state is a special case where the magnitude is particularly large; 25 kHz for ^{1}H–^{13}C coupling[12] and significantly more for ^{1}H–^{1}H coupling. This means that ^{1}H spectra require particularly fast MAS speeds to narrow the spectral lines.[31] Mechanically this is a challenge, so the dipolar coupling is perturbed by sequences of RF pulses, known as decoupling. This is discussed in more detail in Chapter 2.

Additionally, CSA contributions can be as large as 200 ppm for ^{13}C, requiring MAS speeds of 25 kHz to fully average. If lower MAS speeds are used the static powder pattern is broken up into a manifold of spinning sidebands, which appear on either side of the isotropic shift separated by the MAS frequency. Spinning sidebands are a result of rotational echoes: sample orientation changes with respect to $\mathbf{B}_0$ during MAS, so ν_{NMR} and the chemical shielding also change, only returning to their original values when the rotor returns to its original position with respect to $\mathbf{B}_0$. The FID, therefore, consists of a series of rotational echoes associated with this evolution.[32] In the presence of a large CSA at slow MAS speeds the isotropic peak is not necessarily the most intense,[18] Figure 6, and the sidebands can still be very broad if the dipolar contribution is also large. CSA scales with the NMR frequency, so even faster spinning is required at high magnetic fields.[33] Similarly the quadrupolar interaction is too large to average completely, resulting in extensive sideband manifolds under MAS.

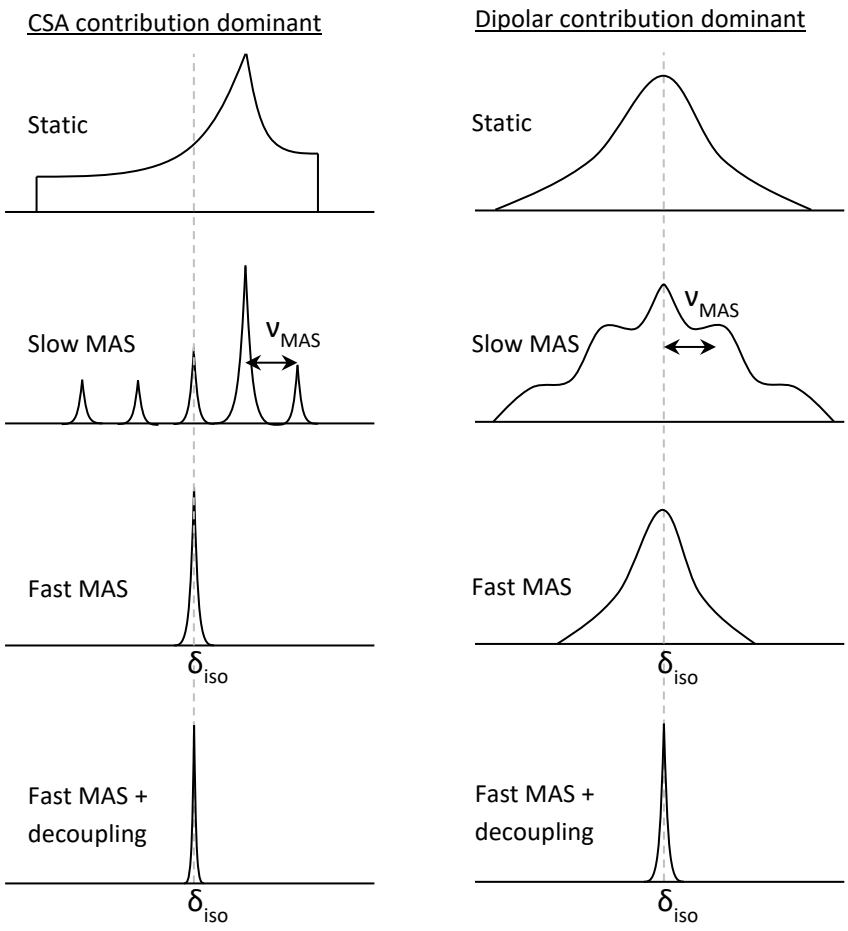

Figure 6 The effect of MAS and decoupling on the schematic spectra of two systems, one dominated by CSA and the other dominated by dipolar coupling.

1.2.6 Linewidths in SSNMR

There are many factors contributing to the linewidths of SSNMR spectra. T_2 determines the natural linewidth, or homogeneous linewidth, *i.e*. the narrowest linewidth theoretically possible for a particular system. In solution T_2 is relatively long so the natural linewidth is small. However, in the solid state there is an additional contribution from residual dipolar coupling that is not fully averaged by MAS or decoupling techniques. The homogeneous linewidth is, therefore, related to T_2' rather than T_2, see Equation (1.8) where FWHM stands for full-width at half-maximum.[34,35]

$$\text{FWHM} = \frac{1}{\pi T'_2} \qquad \textbf{(1.8)}$$

CSA and quadrupolar broadening effects are classified as inhomogeneous contributions, and they can each be manipulated by specific NMR experiments. Note that quadrupolar broadening can be very large if the nuclei of interest are close to quadrupolar nuclei with a large C_Q.[36-38] Glycine provides a good example: a ^{15}N-enriched sample shows a $^{13}CH_2$ signal with 7 Hz linewidth whereas a natural abundance sample has a linewidth of 90 Hz under the same experimental conditions.[39] Thus the presence of a single ^{14}N results in an increase in $^{13}CH_2$ linewidth of 1300 %.

Another inhomogeneous contribution is the anisotropy of the bulk magnetic susceptibility (ABMS), $\Delta\chi$. This is a bulk effect resulting from the different magnetic susceptibilities possessed by individual particles in a sample depending on their orientation with respect to $\mathbf{B}_0$. The net result is a uniform shift in the NMR frequencies and hence a uniform line broadening contribution across all the peaks in the spectrum. The specific effects of ABMS depend nontrivially on crystallite-air boundaries, particle size and sample packing.[40] Fast MAS and co-grinding with a substance like NaCl that has $\Delta\chi = 0$, *i.e*. the bulk magnetic susceptibility is isotropic, can reduce ABMS contributions.[41] A qualitative measure of the magnitude of the ABMS contribution can be given by the ABMS line broadening factor $|4\pi\Delta\chi|$, where $\Delta\chi$ is predicted computationally.[42] Systems that possess sharp lines tend to yield ABMS line broadening factors below 0.5 ppm, while those that have broader spectra yield ABMS factors above 1 ppm.

Finally, broadening can be caused by the local structure of a material such as peak overlap from chemically distinct sites that are only subtly different in environment. Structures containing crystallographic disorder can exhibit this sort of spectral broadening if the disorder is static in nature or slow on the NMR timescale, discussed further in section 1.4.3.

1.3 NMR crystallography

NMR crystallography is a term that describes studies where SSNMR is used to provide structural, and sometimes dynamic, information on a solid system. The NMR crystallography community is well established and the International Union for Crystallography recently introduced a commission called “NMR Crystallography and Related Methods”.[43] In such studies other complementary techniques are often employed, for example thermal methods, microscopy and diffraction experiments.[44] Additionally, it has become increasingly common to use computational methods to directly relate structural changes to observable NMR parameters.[45] This allows potential structures to be modelled and the predicted NMR parameters can be compared with experimental data,

which is particularly useful when the structure is partially or wholly unknown. For example, 3D structural characterisations of a dipeptide,[46] the small molecule thymol,[47] an indomethacin-nicotinamide (IND-NA) cocrystal,[48] and a diazole-based cocrystal[49] have been carried out by comparing experimental SSNMR parameters with those back-calculated from structures refined from diffraction data. In most cases ^{1}H chemical shifts are calculated to establish root-mean-square differences (RMSDs) between the model structure/s and experimental NMR data, though more complex experiments involving ^{1}H–^{1}H spin diffusion are also employed.[46,49] While total 3D structure characterisation *via* NMR crystallography is still the sole province of specialised research groups, the premise of combining SSNMR information with computational methods is carried through in this report. The following discussion will introduce the complementary methods utilised in this work, followed by an introduction to the phenomenon with which this work is chiefly concerned: disorder in cocrystal systems.

1.3.1 Density functional theory

In order to accurately calculate energies, geometries and NMR parameters a quantum chemical approach must be taken. The approach in this study utilises density functional theory (DFT) implemented in the program CASTEP, which stands for Cambridge Serial Total Energy Package.[50] It should be noted that the reproducibility of DFT calculations is a topical question; a recent publication discusses the pair-wise reproducibility of many commercially available DFT codes, including CASTEP.[51] It was concluded that most codes give essentially identical results, although the study focussed on energy *vs* volume prediction rather than NMR parameters. A comprehensive review of the use of CASTEP to calculate NMR parameters was published in 2012,[52] so only a brief discussion is given here to highlight the approximations and limitations that need to be considered.

DFT relies on the fact that the energy and electronic properties of an atom can be uniquely determined using the electron density, ρ, as the fundamental variable to solve the Kohn-Sham equations,[53] Equation (1.9). This equation describes a non-interacting set of particles where $\Psi_i(\boldsymbol{r})$ are mathematical single-particle orbitals, E_i is the orbital energy, and ∇^2 is the operator corresponding to the kinetic energy. The Kohn-Sham potential, $v(\boldsymbol{r})$, contains terms describing the interaction of ionic charges of the nuclei (V_{nuc}), the Coulomb interaction (V_C), and an exchange correlation function (V_{XC}) that describes the many-body interactions between electrons, Equation (1.10). The exact form of V_{XC} is unknown so an approximation must be used that satisfies known physical constraints. There have been

many approximations for V_{XC},[54] but this work makes use of the well-known PBE (Perdew, Burke and Ernzerhof)[55] functional. It is noted that V_{XC} functionals, like PBE, have been found to consistently underestimate the barriers to proton transfer in hydrogen bonds.[56-58] Never-the-less, PBE-optimised hydrogen positions show better agreement with neutron diffraction structures compared to those refined from XRD,[59] for reasons discussed in section 1.3.2. PBE also offers the option of including long-range van der Waals interactions *via* semi-empirical dispersion correction,[60] which has been shown to improve the accuracy of geometry optimisations in molecular organic systems, particularly when the unit cell parameters are allowed to relax.[61-64] The reader is referred to Tkatchenko and Scheffler[60] for a discussion of the dispersion correction scheme used in this work.

$$\left[-\frac{1}{2}\nabla_i^2 + v(\boldsymbol{r})\right]\Psi_i(\boldsymbol{r}) = E_i\Psi_i(\boldsymbol{r}) \qquad \textbf{(1.9)}$$

$$v(\boldsymbol{r}) = V_{nuc} + V_C[\rho] + V_{xc}[\rho] \qquad \textbf{(1.10)}$$

The advantage of DFT over methods that calculate the full electronic wavefunction is that ρ relies on only three coordinates (*x*,*y*,*z*). Nevertheless, for solid-state calculations it is necessary to find a way to model the enormous number of electrons in an essentially infinite crystal lattice, *i.e*. the orbitals $\Psi_i(\boldsymbol{r})$ must be expanded to a computationally feasible basis set. Some approaches used cluster models that reflect the local lattice symmetry but contain a manageable number of molecules/electrons.[65,66] Such calculations are computationally inexpensive and can yield accurate results in, for example, a set of 72 insulating molecular solids containing the nuclei ^{13}C, ^{15}N, ^{19}F and ^{31}P.[66] This approach is useful for non-periodic systems such as proteins,[67] but there are challenges associated with determining the size and termination of the cluster.[68] Instead, CASTEP employs the planewave pseudopotential method, where $\Psi_i(\boldsymbol{r})$ are described using a planewave basis set. The number of planewaves used, *i.e*. the size of the basis set, is controlled by cut-off energy E_{cut}. This utilizes Bloch's theorem so only the unit cell need be considered and periodic boundary conditions are applied. Physical properties are calculated as an average over a set number of *k*-points in the first Brillion Zone in reciprocal space.[52] The *k*-point spacing determines how fine the sampling of the Brillion Zone is; the finer the sampling the more computational resources are required but the predicted values are more reliable.

Pseudopotentials, otherwise known as effective core potentials,[69,70] are used to describe the core electrons. This utilises the frozen core approximation: an assumption that the core electrons are tightly bound and contribute little to the properties of a system.[71] Of course

NMR properties are highly sensitive to the electronic structure, including the core electrons, so this information must be recovered to calculate properties like chemical shift. In this work, the core electron properties are accounted for using the gauge-included projector augmented wave (GIPAW) method developed by Pickard *et al*.[72] The theoretical details are beyond the scope of this work and the reader is referred to a review by Charpentier.[73] The GIPAW method has become the method of choice for calculating NMR parameters in the NMR crystallography community, but there are some limitations. In terms of implementation in CASTEP, spin-orbit coupling has only recently been introduced and relativistic effects are not taken into account, though this is not usually a problem for small organic molecules. Spin polarised systems also cannot be handled, so paramagnetic NMR spectra cannot be predicted.

Several physical approximations are necessary in DFT methods, namely the choice of V_{XC} and the use of the frozen core approximation, and these introduce a systematic error in calculated NMR parameters compared to experimental data. In terms of the choice of V_{XC}, the errors in GIPAW-calculated shieldings using PBE are comparable to those of fragment-based methods that allow the use of modern hybrid V_{XC} functionals.[74,75] Historically, the errors in calculated chemical shieldings were considered to be roughly 2–3 % of the usual chemical shift range for a particular element,[52] for example 2 ppm for ^{13}C. This was confirmed by a recent benchmark study of five common SSNMR reference compounds, such as adamantane.[76]

An additional approximation that has not yet been discussed is that all DFT calculations are effectively carried out at 0 K. Neglecting thermal effects has been shown recently to be significant in the calculation of accurate chemical shifts.[76,77] By examining the effects of atomic vibrations on the chemical shielding tensor temperature corrections of up to 1 ppm for ^{1}H and 2–6 ppm for ^{13}C have been suggested for the test case of L-alanine. Whilst it is still not trivial to include temperature corrections, it is possible to carry out DFT molecular dynamics simulations (DFT-MD) in CASTEP at a finite temperature.[78] This allows small thermal fluctuations to be taken into account to bridge the temperature gap between structure determinations by routine single crystal X-ray diffraction (SCXRD), generally carried out between 100–150 K, and NMR experiments, which are usually limited to roughly 190 K and above. DFT-MD has been found to increase the accuracy of calculated NMR parameters,[76] but it cannot always capture dynamics that occur on the NMR timescale, such as methyl group rotation, since the computational cost is such that the

ensemble cannot be fully sampled. Other MD methods are available, such as classical MD, which can access timescales of ns but requires specific force fields,[79] and metadynamics.[80] Nuclear delocalisation effects are also not taken into account by DFT and these have also been shown to affect carbon chemical shifts, particularly when the carbon atom is directly bonded to one or more protons such as methyl groups.[81]

Despite its limitations, the agreement of DFT-calculated NMR parameters with experimental data is generally good and it has become increasingly common for structural and dynamic studies to include some degree of computation.[82] A prerequisite for these calculations is to have a sensible structural model from which to calculate physical properties. It is becoming possible to predict the 3D structure of small systems, as established in the recent crystal structure prediction (CSP) blind tests,[83] but the usual source of such a model is a structure refined from SCXRD data.

1.3.2 Conventional crystallography

There are many texts that outline the theoretical background of XRD experiments so this section will remain brief and the reader is referred to the comprehensive book by Giacovazzo *et al*.[84]

The spectral range of X-rays encompasses wavelengths of the order of 1 Å so they interact with atoms in a number of ways. The first is simply absorption of the X-ray photons by the electrons in an atom, which results in radiation damage to the sample if the photon energies are high enough. Alternatively, the photons may be scattered. Inelastic (or Compton) scattering results in energy transfer from the photons to the electrons of the sample, so a lower energy photon is emitted. This type of scattering is hard to control and is not very useful for structural studies. XRD relies upon the elastic (or Thomson) scattering of X-rays by the electrons in a solid sample, which provides information on the distribution of atomic electron density. The long-range periodicity of the electron density in crystalline solids results in coherent scattering of the X-rays from different planes of the crystal, Figure 7, which gives rise to a set of characteristic reflections whose intensities are recorded by a detector. The positions of these reflections provide information on the unit cell. Systematic absences of reflections give information on possible space groups, which define the symmetry of the structure. The intensities of the reflections reveal the distribution of electron density, *i.e*. the atomic positions.

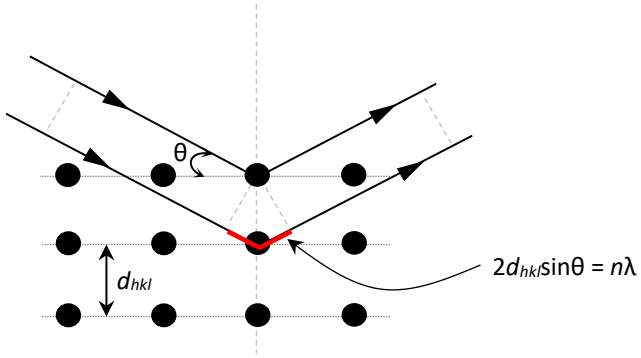

Figure 7 A schematic representation of Bragg's Law where d_{hkl} is the distance between two planes in the crystal characterised by Miller indices *h*,*k*,*l*. θ is the angle of incidence, *n* is an integer and λ is the wavelength of the incident beam.

The intensities measured in the XRD experiment are proportional to a structure factor, $\mathbf{F}_{hkl}$, Equation (1.11). This is a vector and so can be represented by Equation (1.12), where ϕ is the phase of the scattered wave from a particular lattice plane. In order to calculate the electron density from the observed reflections of a particular lattice plane $\mathbf{F}_{hkl}$ is expressed using Equation (1.13).[85] The parameter f_j denotes the individual atomic scattering factors, quantifying the scattering power of each element, and x_j,y_j,z_j are the fractional coordinates of the atoms. Each structure factor corresponds to one reflection and the electron density is related to these structure factors by Fourier transformation.

$$I_{hkl} \propto |F_{hkl}|^2 \quad \textbf{(1.11)}$$

$$\mathbf{F}_{hkl} = |F_{hkl}|e^{i\phi} \quad \textbf{(1.12)}$$

$$\mathbf{F}_{hkl} = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad \textbf{(1.13)}$$

It is not possible to directly measure the phase of a scattered wave in an XRD experiment, leading to the phase problem. There are well-established solutions to this,[84] during which a model structure is produced. The calculated structure factors associated with this model are compared to the experimental values, and the degree of agreement is expressed using an *R* factor. The model is then computationally adjusted to give the lowest possible *R* factor. Note that the *R* factor is a measure of precision, not accuracy, and incorrect structures may refine with a low *R* factor.[86] The refinement process has become automated over recent years; sophisticated software solves and checks structures with relatively small human input required in simple cases.[87]

SCXRD has been called the "gold standard" for solid-state structure determination,[87] but there are limitations to the technique. Most importantly, it is not always possible to grow suitable single crystals of the sample of interest. In such cases it is possible to carry out powder X-ray diffraction (PXRD). This relies upon the same principles as SCXRD but there are additional complications to structure solution because the reflections from a 3D structure are projected onto a single dimension in the resulting powder pattern. The reflections overlap extensively so individual intensities are challenging to determine.[9] It is not trivial for non-experts to carry out the analysis of PXRD patterns to yield full structural information, but it is common to use them as a fingerprint to compare between different samples. For example, it is easy to spot evidence of solid form diversity in PXRD data because changes in the 3D packing of the molecules results in very different peak shapes, intensities, and positions.[88]

A well-known limitation of XRD is the placement of hydrogen atoms. This is because there is only a single electron associated with a hydrogen atom, resulting in a small atomic scattering factor, and this electron tends to reside in covalent bonds rather than directly next to the atom nucleus. As a result, hydrogen atoms tend to be placed in geometrically sensible positions if they cannot be directly refined. Developments in XRD methods and equipment have reduced this difficulty somewhat, though the asymmetric nature of the electron density around hydrogen atoms still results in larger errors being placed upon H–X bond lengths.[89] Optimisation of the hydrogen atom positions is, therefore, necessary before NMR parameters can be calculated from SCXRD structures.[45] In many cases even the heavy atom positions must be optimised, though the root mean squared (RMS) displacements between the starting and optimised structures are generally small (< 0.05 Å).[90] Neutron diffraction can be carried out to determine the hydrogen positions more reliably, but this requires the use of a beam line at a nuclear reactor or a spallation source.

1.3.3 Complementary information

SSNMR is ideally suited to providing information that is complementary to XRD. In terms of investigations into dynamics, the two techniques probe very different timescales due to the difference in frequency of X-rays (10^{16}–10^{20} Hz) and RF (1–10^{11} Hz). SSNMR is therefore more sensitive to small molecular motions than XRD, for example ring flipping or methyl group rotation. However, XRD is sensitive to molecule translation, which is often slower than the NMR timescale. In XRD experiments the average distribution of the electron

density is modelled using anisotropic displacement parameters (ADPs),[91] and these can capture some evidence of dynamic sites. For example, unusually elongated ADPs might be an indication of the presence of atomic motion. Conversely, small amplitude motion may be disguised within ADPs. In contrast, local atomic motion is observable in variable temperature SSNMR experiments by changes in chemical shift, peak shape or relaxation times.

SCXRD experiments rely upon the existence of long-range order to probe the periodicity of a system, providing information on symmetry and the unit cell, whereas SSNMR probes only local structure in the order of up to 10 Å.[9] This combination of information provides a more complete picture of the structure compared to the information provided by a single method. DFT provides the bridge between XRD and SSNMR by allowing the computation of NMR parameters from structural models and subsequent comparison with experimental data. This allows solid-state structures to be validated and additional local structure/dynamic information can be identified. A recent example focuses on validating duplicate structures deposited in the Cambridge Structural database (CSD).[92] Two depositions for the API furosemide (FS) were investigated, which differed only in the position of a single hydrogen atom. It was found that only one of the two structures could be verified by considering the agreement between calculated and experimental ^{1}H and ^{13}C chemical shifts. The other was considered to be an incorrect structure, rather than a different form, on the basis that the energy was calculated to be 25 kJ mol^{-1} higher than the validated structure. The usual energy difference between different solid forms is < 5 kJ mol^{-1}.[93]

The atom positions in crystal structures solved from XRD data are often not quite in a DFT-calculated energy minimum, due to the different temperature windows accessible in DFT and XRD and also the challenges associated with hydrogen placement[94] from electron density information. This can result in poor agreement between experimental and calculated SSNMR parameters so it is conventional to optimise the atom positions of the input structure before NMR calculations are carried out. Hydrogen atom positions in particular always require optimisation prior to NMR calculations.[45] Once optimised, the agreement between calculated and experimental NMR parameters provides a tangible link between the average structure calculated from XRD data and the local structure probed by SSNMR.

1.4 Cocrystals and solid form diversity

Solid-state chemistry is intrinsically linked to the pharmaceutical industry because structure profoundly influences physicochemical properties.[95] Solubility is particularly important, being related to the bioavailability of an API.[96] However, many common drugs are poorly soluble so much research is directed at the discovery of new forms with increased solubility. A variety of disordered forms are emerging in this area, including solid solutions,[97] but it is more usual for crystalline forms to be chosen to limit issues with stability.

An approach that has become increasingly popular is the formation of cocrystals *via* co-crystallisation of an API with a small soluble coformer. The reader is referred to a number of reviews on cocrystals published in the last five years.[98-105] A generally accepted definition of the term "cocrystal" has been given by Shan and Zarowotko: "*… a multicomponent crystal in which all components are solid under ambient conditions when in their pure form. These components exist as a stoichiometric ratio of a target molecule or ion and a neutral molecular cocrystal former(s)*".[106] This distinguishes cocrystals from solvates, where one of the components is liquid under ambient conditions, and hydrates, in which one of the components as water. See Chapter 6 for a more detailed introduction to solvates. Salts are also distinct from cocrystals because one or all of the components in a salt are not neutral.[107] There has been discussion in the literature about the exact terminology pertaining to cocrystals,[108] with molecular complex suggested to be more accurate.[109] However, the term cocrystal will be used in this report to denote systems of the form described by Shan and Zarowotko.

The components of cocrystals interact through strong intermolecular interactions such as hydrogen bonds. The molecular fragments containing these interactions are known as synthons. A small selection of common synthons[110] are shown in Figure 8, highlighting that APIs and coformers with many hydrogen bonding donors/acceptors are likely to be good candidates for cocrystal formation. Carboxylic acids are known to form dimers, since they contain both a hydrogen bond donor and an acceptor, as do amides and other such functional groups. Additionally, ribbon-like chains are commonly observed in cocrystals, such as synthons **B**, **E**, and **F** in Figure 8.

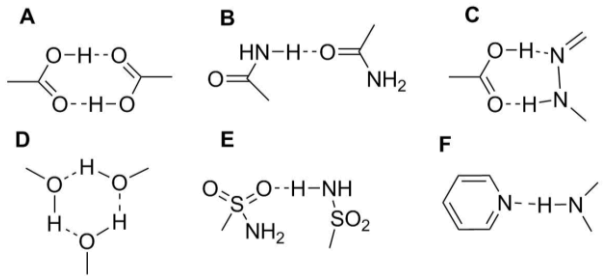

Figure 8 Some examples of supramolecular synthons that can be observed as building blocks in cocrystals.

It was observed as early as 1946 that the cocrystallisation of an API with a soluble coformer, theophylline and glycine respectively in this case, forms a cocrystal with enhanced solubility compared to the pure API.[111] This clinical study of the theophylline-glycine cocrystal also found that it was tolerated in "*unusually high amounts*" in humans while eliciting the usual API response. Thus, cocrystallisation is frequently employed as a technique to improve the physicochemical properties of known APIs. The solubility of cocrystals is frequently studied and has been found to be dependent on a number of factors, including drug-coformer interactions.[112] Co-crystallisation also has the advantage of increasing the diversity of solid forms available for a particular drug molecule: there are many compounds that can be suitable coformers, the only requirements being high solubility and low toxicity. Many compounds with GRAS (generally regarded as safe) status are suitable as listed in the EAFUS inventory (everything added to food in the United States),[113] and it is estimated that the number of potential coformers is of the order of several hundred.[114] Additionally, many APIs are non-ionisable, precluding salt formation but not cocrystal formation.[103]

Crystal engineering aims to design suitable API/coformer combinations.[115,116] Hydrogen bonding propensities are important, as well as halogen bonding, π–π stacking and other intermolecular interactions. New developments in CSP are beginning to accelerate cocrystal screening in cases like the cocrystals of succinic acid, amino-benzoic acid, and caffeine. Comparisons of the lattice energies of the pure components with those of the predicted cocrystals go some way to predicting likely cocrystal formation without the need for experimental screening. It has been found that cocrystals generally have lower lattice energies than the pure components but the energy differences were small, < 5 kJ mol^{-1}, and place heavy demands on the computational methods.[117,118] It is not yet routine to use

CSP methods to predict new cocrystal structures, though commercial CSP codes are in development.[119,*]

It is more common for cocrystal screening to be performed *via* slow solvent evaporation or mechanochemical routes.[95,120-122] Solvent evaporation can be limited by the solubility of the API, whereas physically grinding the components together provides a cheap and green pathway to cocrystal formation. Mechanochemistry is also more scalable for industrial applications than solvent-based methods.[122] It has been suggested that the mechanism of cocrystal formation through grinding involves amorphous intermediates,[123] but this is likely to be system-specific and the details of formation are not generally known.[95] Liquid-assisted grinding (LAG) can accelerate solid-state reactions,[122] or provide a route to solvate/hydrate formation. Generally cocrystals are stable under ambient conditions, though it is commonly observed that high humidity can induce degradation due to the differences in solubility between APIs and coformers.[124] The presence of additives that are structurally similar to the coformer can induce coformer replacement during processing,[125] so stability tests are necessary for cocrystals to be formulated for pharmaceutical use.

There are now several commercially available drugs that are formulated as cocrystals,[103] approved in accordance with the FDA cocrystal guidance paper on drug substances in 2013.[126] Entresto™ is a new drug marketed by Novartis for the treatment of chronic heart failure[127] and is composed of monosodium sacubitril, disodium valsartan and water. Other examples include the antidepressant escitalopram oxalate,[128] marketed as Lexapro™, and ertugiflozen 5-oxoproline cocrystal,[129] which is nearing the end of Phase III clinical trials for the treatment of Type II diabetes.[130]

1.4.1 Polymorphism and disorder in cocrystals

Cocrystals exhibit the same structural variability as all crystalline solids. Polymorphism is defined as the ability of a system to crystallise with different 3D packing or in different conformations, and cocrystals are susceptible to this phenomenon. Polymorphs tend to have similar lattice energies,[82,131] making them computationally demanding to predict by CSP methods,[82] and a kinetically favoured polymorph can spontaneously convert to the thermodynamically favoured form over time. The pharmaceutical industry, therefore,

[*] The most recent CSP blind test, in which academic groups compete to predict structures of molecules set by the Cambridge Crystallographic Data Centre, revealed that several CSP codes are capable of correctly predicting the structure and polymorphs of a cocrystal, a salt hydrate and a large flexible molecule.

carries out polymorph screening to ensure the stability of the chosen formulation.[132] Most polymorphic cocrystals are discovered serendipitously,[101] though Lemmerer *et al*. suggest that cocrystals formed from components that are polymorphic in their pure forms are also likely to be polymorphic.[133] It has been found that LAG with specifically chosen solvents[134] and polymer assisted grinding[135] can provide some control to polymorph selection in cocrystals. Generally though, there is little work devoted to the specific search for novel cocrystal polymorphs.

Disorder is another common cause of solid form diversity in molecular crystals. In fact, it has recently been hypothesised that most crystals are inherently disordered on a small scale, based on the distribution of geometrical parameters calculated from high-quality CSD structures of a representative set of six common inorganic molecules. A blind area of uncertainty was observed in these distributions, corresponding to random scattering between 0.02–0.04 Å for bond lengths and 2–3˚ for bond angles.[136] However, further investigation is required to distinguish between intrinsic disorder and experimental factors.

The introduction of disorder can affect pharmaceutically relevant properties like dissolution and stability, particularly when a system is made amorphous.[137,138] There are several non-crystalline drugs listed in the current European Pharmacopoeia (as reported by Hancock *et al.*)[5] including small organic molecules, such as lactose, and APIs like nelfinavir mesylate. Importantly, moving away from a crystalline form introduces new challenges, such as stability.[139] For example, grinding anhydrous carbamazepine (CBZ) to induce the formation of disordered regions has been found to enhance the rate of phase transformation to the dihydrate,[140] while other systems show a tendency to crystallise back to an ordered form over time.[141]

In the context of this report, “disorder” is not meant to imply a complete lack of order, as observed in amorphous systems. This work focuses on local disorder in crystalline organics, pertaining to atom displacements of the order of a few angstroms. Such disorder is not considered a problem for API development and is of more interest to fundamental studies of structure, though its impact on solubility has not been investigated. Local disorder is the result of a loss of positional order, orientational order, or conformational order. It results from the fact that conformationally flexible moieties possess many degrees of freedom so stable 3D packing may be achieved through a number of different conformations. Similarly, molecules such as caffeine[142] may have no thermodynamic preference for one orientation over another. Disorder can be static, where the disordered fragments remain in distinct

orientations, or dynamic, where the disordered component either jumps between discrete positions or undergoes continuous motion. Loosely bound fragments, such as solvent molecules in channels, are frequently dynamic since there are no strong intermolecular interactions holding them in place.[143] Some systems have been found to possess varying degrees of disorder depending on the crystallisation conditions, thought they tend to be challenging to synthesise and purify.[87,144-146]

The position of hydrogen atoms in the hydrogen bonds of a cocrystal is of interest, both for fundamental studies and for patent applications, because the hydrogen atom position determines whether the compound is best described as a cocrystal or a salt. It is possible to computationally predict whether a system will be more thermodynamically stable as a salt or a cocrystal,[147] and usually protons can be located using NMR crystallography[148] (see Chapters 4, 5, and 6) or with techniques like neutron diffraction.[149] However, there are some systems that exhibit hydrogen disorder or occupy the continuum between a pure salt and a pure cocrystal.[150] Such systems are of interest for fundamental studies and potentially useful in functional materials, but this specific form of disorder was not observed in this report.

1.4.2 Modelling disorder

Disorder implies that there is reduced long range order, which clearly presents challenges for structure solution from XRD data. For example, in the extreme case of an amorphous solid virtually no Bragg scattering can be detected. NMR crystallography methods can be applied to these compounds, for example to distinguish between two different amorphous forms of an API, valsartan, by comparing the SSNMR spectra.[151] However, XRD data can provide a useful average picture of the 3D structure when only local disorder is present.

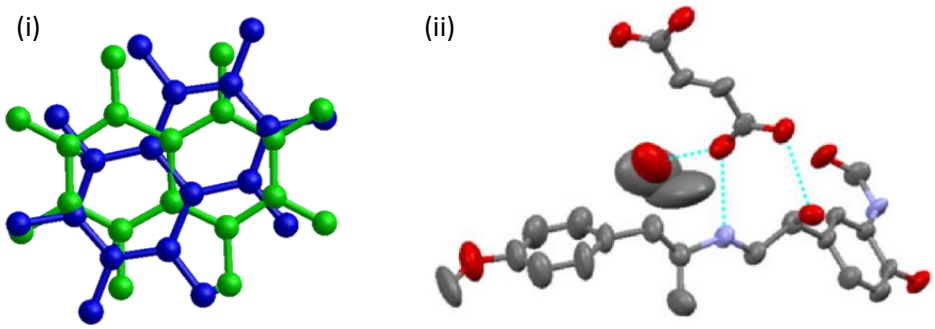

Figure 9 (i) Two-site disorder of octofluoronaphthelene with split-atom sites coloured for clarity. **(ii)** Solvent disorder in a formoterol fumarate solvate modelled with elongated ADPs, reproduced with permission from Apperley *et al*., Magn. Reson. Chem., 2012, 50, 10, 680–690. Copyright © 2012 John Wiley & Sons, Ltd.

Structures containing local disorder can be refined from SCXRD data, though it is not generally possible to distinguish between static and dynamic disorder without variable temperature experiments and specialised data approaches. Disordered components can be modelled using split sites if there are discrete regions of electron density, with each set of atomic coordinates associated with one conformation of the disorder, for example the two-orientation disorder of octofluoronaphthelene[152] in Figure 9(i). The fractional occupancies of these split sites are explicitly refined and they provide an insight into the population of the disordered components. Alternatively, disorder may be obscured (or revealed) by the ADPs. Unusually shaped ADPs can be an indication of the presence of disorder, for example the dynamically disordered ethanol solvate of formoterol fumarate,[153] Figure 9(ii). Solvate systems containing disorder can represent a challenge for structure refinement if the electron density of the solvent molecule is continuous, for example through channels. In such cases an atomistic model using ADPs or split sites is not suitable. There are crystallographic tools to deal with this situation,[154] where the structure is solved without the solvent molecules modelled at all, but in these circumstances no information is retained on the position or nature of the disordered molecules.

Analysis of diffuse scattering in XRD can also provide information on atomic motion and disorder. Diffuse scattering is observed as weak intensity between the main reflections in a SCXRD pattern,[155] Figure 10(i), and is caused by small correlated fluctuations away from the average long range order. Extracting information from this scattering is challenging, but highly disordered organics can be characterised from high quality diffraction data such as the proton disorder in phloroglucinol dihydrate, Figure 10(ii).[156] The hydrogen position on one of the OH groups is disordered, leading to a “domino effect” in the hydrogen bonding network whereby a hydrogen atom on the water molecule is also disordered. The site occupancies of the two disordered hydrogen atoms are 50:50 on average, but cooperative hydrogen bonding must occur over long distances in the unit cell. Geometry optimisation of two model structures, each containing one site of the disordered atoms, revealed small heavy-atom displacements of 0.04 Å depending on which sites the disordered hydrogen atoms occupied. This heavy atom perturbation was proposed to be the origin of the strong diffuse scattering. Few other techniques can distinguish random disorder from correlated disorder.

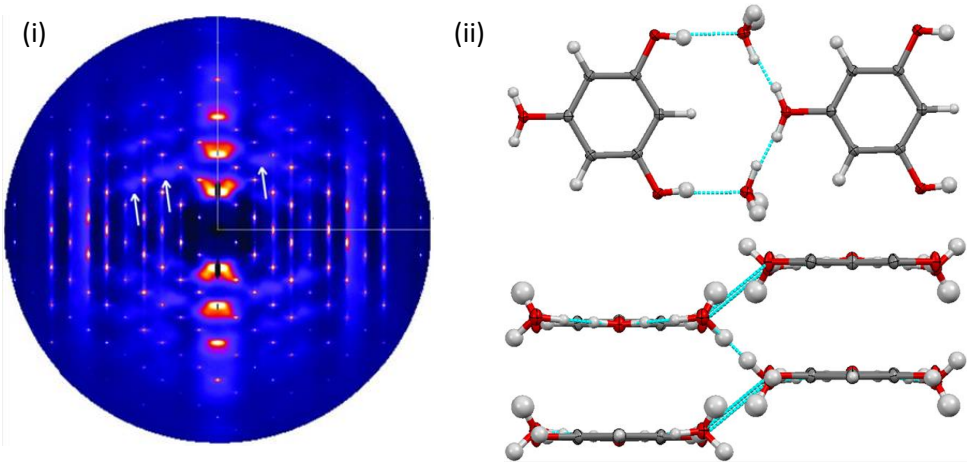

Figure 10 (i) Diffuse scattering observed from a pentachloronitrobenzene single crystal, from Thomas *et al.,* Acta Crystallogr. B, 2007, 63, 4, 663–673, reproduced with permission from the International Union of Crystallography. **(ii)** Hydrogen bonding chains in phloroglucinol containing correlated proton disorder.

Computational modelling bridges the gap between the average structure of a disordered solid from XRD data and the spectral variations observed in SSNMR.[94] Disordered systems are challenging to model computationally, since the exact structure is incomplete without some way of describing the nature of the disorder. It has already been stated that DFT methods do not account for thermal effects, so dynamic disorder cannot be explicitly modelled. Additionally, truly random static disorder should really be modelled with an infinitely large supercell. In order to deal with this, several approximations have been established. Configurational ensemble methods have been developed with the aim of modelling disorder to a high level of accuracy. This models the system as an ensemble over every possible disordered configuration, so physical properties may be computed for the whole ensemble. The symmetry-adapted ensemble method[157,158] is one such technique that models site-occupancy disorder by producing supercells with all possible symmetry-unique configurations of the disorder. The energies of different supercells can be compared and used to calculate probabilities for the presence of particular configurations in the true disordered structure. The structures of anhydrous β caffeine and isocaffeine have been investigated using this method,[159] discussed in detail in Chapter 3, providing insight as to why β caffeine is disordered while isocaffeine is not. However, this approach requires significant expertise and specifically written computer programs so it is not very accessible.

The simplest method to computationally model disorder is to split the disordered structure into its ordered component parts. Physical properties can then be calculated for the two components and then the results may be compared to the experimental data. Several

successful investigations have been carried out using this approach.[160-165] For example, a dimethyl sulfoxide (DMSO) solvate of CBZ was found to contain two-site disorder of the DMSO molecule in low temperature SCXRD data.[160] Several computational models of the system were made with the DMSO molecules placed in different orientations to model static disorder. It was found that, if the two lowest energy models were considered to be in equilibrium, then the relative populations of the two DMSO orientations was reproduced reasonably well at low temperature. NMR parameters can also be calculated with this method, as used by Ashbrook *et al*.[161] to gain information on site-occupancy disorder of an inorganic solid solution, though the technique is by no means limited to inorganic systems.

1.4.3 Characterising disorder in cocrystals by SSNMR

Cocrystals containing disordered components are particularly amenable to NMR crystallography studies because the details of the structure must be investigated on a local level. Initially, cocrystal formation can be proven by analysis of the change in chemical shift of the product compared to the pure components and nuclear proximities may be probed by 2D experiments.[49] Subsequently, the nature of the disorder can be determined: static and dynamic disorders are distinguishable since they affect SSNMR spectra in different ways. Static disorder causes inhomogeneous line broadening or the appearance of multiple peaks for a particular site since the different chemical environments of disordered fragments results in a small chemical shift distribution.[146] In contrast, rapid dynamics that are faster than the timescale of the NMR experiment, will result in a time-averaged spectrum being observed,[94] often with sharp lines. Slower dynamics will result in incomplete averaging of the spectrum, causing line broadening, and variable temperature experiments can probe these dynamics by changing the rate of motion. Subsequent monitoring of linewidths[143] and relaxation times[166] can yield activation energies for the motion and also correlation times.

The structural investigation of carvedilol phosphate hemihydrate highlights how many of the techniques used in this work can be applied to solve important structural questions.[167] Carvedilol is a chiral API used in the treatment of high blood pressure and heart failure, Figure 11(i). The SCXRD structure of the racemic phosphate hemihydrate contains a complex combination of twinning and disorder, requiring both SSNMR and SCXRD to solve. SCXRD analysis revealed the presence of twinning, and the structure was solved with disorder of the fused ring system of carvedilol over three sites with roughly equal occupancies, Figure 11(ii). The ^{13}C SSNMR spectrum, Figure 11(iii) shows broadening of

peaks associated with sites close to the chiral centre, C16, of the carvedilol molecule. Specifically, the C11 peak was estimated to contain three components. These broadened peaks correlate well with the crystallographically disordered sites, though the three distinct disorder groups cannot be distinguished to validate the site occupancies by SSNMR. Few changes are observed in the ^{13}C spectrum over a 100 K temperature range, indicating that the disorder is static in nature. ^{1}H–^{13}C HETCOR correlations confirm the presence of edge-on aromatic π stacking for additional validation of the SCXRD structure. The mechanism of water incorporation was also probed. Successive hydration/dehydration cycles highlighted the reversibility of water uptake, and no evidence of water jump motions were observed by variable temperature ^{2}H SSNMR. This was ascribed to hydrogen bonding between the water and carvedilol and the narrow diameter of the channels.

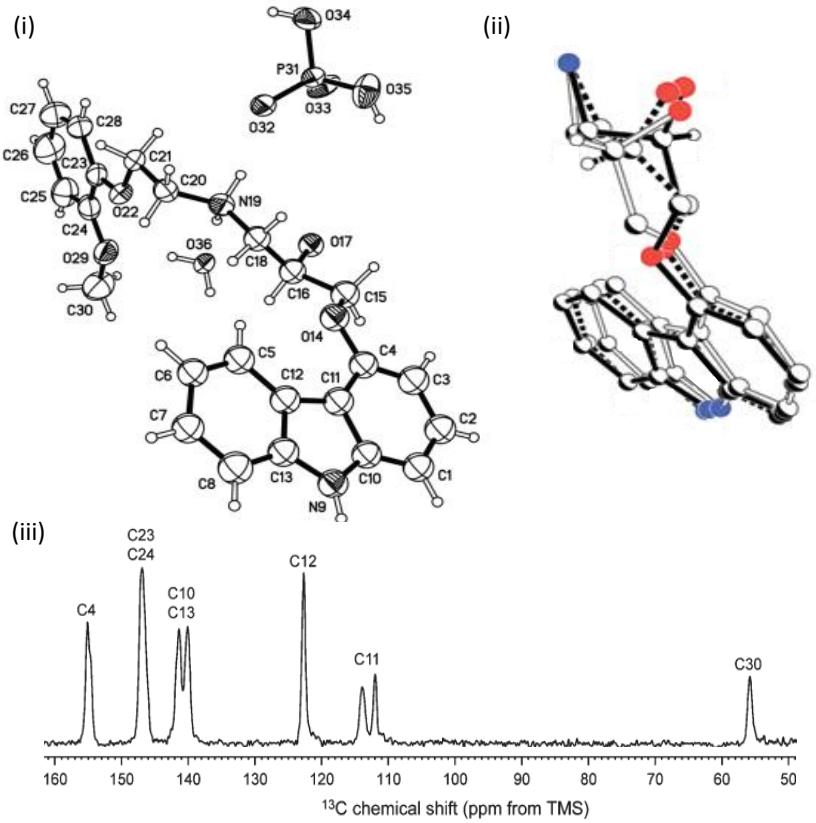

Figure 11 (i) Asymmetric unit of racemic carvedilol phosphate hemihydrate with thermal ellipsoids shown at the 50 % probability level. The whole water molecule is shown for clarity, though only half is present in the asymmetric unit. Only one component of the disorder is shown. **(ii)** Three-component disorder of part of the carvedilol molecule. **(iii)** ^{13}C CP/TOSS spectrum of racemic carvedilol phosphate hemihydrate acquired with a dipolar dephasing delay. Broad peaks are indicative of disorder, *e.g*. C4 and C11. Adapted with permission from Vogt *et al.,* Cryst. Growth Des., 2010, 10, 6, 2713–2733. Copyright © 2010 American Chemical Society.

1.5 Overview

The aim of this work is to use NMR crystallography to gain structural and dynamic information for disordered cocrystal systems. Chapter 2 describes the specifics of the methods employed, including NMR pulse sequences and details of the computational methods. Results are presented in the subsequent four chapters, each focussing on a specific system with different structural questions. Firstly, Chapter 3 examines an aspect of NMR crystallography methodology by investigating the limit of DFT geometry optimisations and what impact these have on predicted NMR parameters. This is with the view to modelling how static disorder contributes to the linewidths of the ^{13}C SSNMR spectrum of a new caffeine-citric acid cocrystal hydrate, CAF-2CA. This investigates the limit of NMR crystallography by finding the effective "error bar" on geometry optimisations, which is important in the study of subtle NMR features. Chapter 4 focusses on a new cocrystal of the API naproxen (NPX) with the coformer picolinamide, NPX-PA, where the positions of two hydrogen atoms are ambiguous in XRD data leading the question of whether the system is a cocrystal or a salt. SSNMR and DFT energy calculations are carried out in order to validate and refine the hydrogen atom positions. Chapters 5 discusses the nature of the disorder in a cocrystal of the API furosemide (FS) with isonicotinamide, 2FS-INA. Two distinct fragments of the structure are dynamically disordered, and the motion is characterised by NMR crystallography methods to estimate the rate of motion and energy barriers. Finally, Chapter 6 discusses the structure of a new cocrystal of FS with picolinamide, FS-PA, for which no SCXRD data is available along with its two isostructural solvates. Structural features are identified in FS-PA and the solvate dynamics are characterised, with an estimate for the mechanism of motion proposed.

Over the course of this work it is shown that a combination of SSNMR, DFT and XRD allows for detailed characterisation of the structure and dynamics of disordered cocrystal systems, including solvates. Both static and dynamic disorder are identified and characterised in this work and the origins of the disorder can be inferred from consideration of calculated energies of structural models that aim to simulate the disordered systems. The methods utilised in this work are applicable to future pharmaceutical systems for which detailed structural characterisation is required. Publications arising from aspects of this work are noted at the beginning of the relevant chapters.

1.6 References

(1) F. G. Vogt, *Future Med. Chem.*, **2010**, 2, 915.
(2) J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter and J. Morris, *Pharm. Res.*, **2001**, 18, 859.
(3) S. Datta and D. J. Grant, *Nat. Rev. Drug Discov.*, **2004**, 3, 42.
(4) *Newsletter September 2012*, http://www.crystallics.com/newsletter-september-2012, accessed 2016.
(5) B. C. Hancock, *J. Pharm. Pharmacol.*, **2002**, 54, 737.
(6) S. Byrn, R. Pfeiffer, M. Ganey, C. Hoiberg and G. Poochikian, *Pharm. Res.*, **1995**, 12, 945.
(7) U. S. Government, Q6A Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances., **1997**, *62*, 62890.
(8) P. A. Tishmack, D. E. Bugay and S. R. Byrn, *J. Pharm. Sci.*, **2003**, 92, 441.
(9) *"NMR Crystallography*", R. K. Harris, R. E. Wasylishen and M. J. Duer (Eds.), John Wiley & Sons Ltd: West Sussex, UK, 2009.
(10) *"Spin Dynamics: Basics of Nuclear Magnetic Resonance*", 1st ed., M. H. Levitt, John Wiley & Sons, Ltd: West Sussex, England, 2001.
(11) *"Solid State NMR Basic Principles and Practise*", D. C. Apperley, R. K. Harris and P. Hodgkinson (Eds.), Mentum Press: LLC, New York, US, 2012.
(12) "Introduction", D. C. Apperley, R. K. Harris and P. Hodgkinson in "*Solid State NMR Basic Principles and Practise"*, Mentum Press: LLC, New York, US, 2012, p 1.
(13) "Uncoupled Spins-1/2", M. H. Levitt in "*Spin Dynamics: Basics of Nuclear Magnetic Resonance",* John Wiley & Sons, Ltd: West Sussex, England, 2001, p 239.
(14) A. Abragam and M. Goldman, *Rep. Prog. Phys.*, **1978**, 41, 395.
(15) T. Maly, G. T. Debelouchina, V. S. Bajaj, K. N. Hu, C. G. Joo, M. L. Mak-Jurkauskas, J. R. Sirigiri, P. C. van der Wel, J. Herzfeld, R. J. Temkin *et al.*, *J. Chem. Phys.*, **2008**, 128, 052211.
(16) "The NMR Experiment", M. H. Levitt in "*Spin Dynamics: Basics of Nuclear Magnetic Resonance",* John Wiley & Sons, Ltd: West Sussex, England, 2001, p 69.
(17) "The Basics of Solid-State NMR", M. J. Duer in "*Solid-State NMR Spectroscopy: principles and applications"*, Blackwell Science Ltd: Oxford, UK, 2002, p 3.
(18) D. L. Bryce, G. M. Bernard, M. Gee, M. D. Lumsden, K. Eichele and R. E. Wasylishen, *Can. J. Anal. Sci. Spectros.*, **2001**, 46, 46.
(19) "Spin-1/2 Nuclei: A Practical Guide", D. C. Apperley, R. K. Harris and P. Hodgkinson in "*Solid State NMR Basic Principles and Practise"*, Mentum Press: LLC, New York, US, 2012, p 39.
(20) D. Suter and R. R. Ernst, *Phys. Rev. B*, **1985**, 32, 5608.
(21) "Basic NMR Concepts for Solids", D. C. Apperley, R. K. Harris and P. Hodgkinson in "*Solid State NMR Basic Principles and Practise",* Mentum Press: LLC, New York, US, 2012, p 17.
(22) K. Jackowski, M. Jaszunski and M. Wilczek, *J. Phys. Chem. A*, **2010**, 114, 2471.
(23) "Nuclear Spin Interactions", M. H. Levitt in "*Spin Dynamics: Basics of Nuclear Magnetic Resonance",* John Wiley & Sons, Ltd: West Sussex, England, 2001, p 125.
(24) P. Hodgkinson and L. Emsley, *J. Magn. Reson.*, **1999**, 139, 46.
(25) S. E. Ashbrook and S. Sneddon, *J. Am. Chem. Soc.*, **2014**, 136, 15440.
(26) "Quadrupolar Nuclei", D. C. Apperley, R. K. Harris and P. Hodgkinson in "*Solid State NMR Basic Principles and Practise"*, Mentum Press: LLC, New York, US, 2012, p 141.
(27) M. Shen, R. Roopchand, E. S. Mananga, J. P. Amoureux, Q. Chen, G. S. Boutis and B. Hu, *Solid State Nucl. Magn. Reson.*, **2015**, 66-67, 45.
(28) E. R. Andrew, A. Bradbury and R. G. Eades, *Nature*, **1959**, 183, 1802.
(29) P. Corkum, *JEOL Resonance Introduces Worlds Fastest and Smallest Solid State NMR Probe*, http://www.jeolusa.com/NEWS-EVENTS/Press-Releases/ID/270/JEOL-Resonance-Introduces-Worlds-Fastest-and-Smallest-Solid-State-NMR-Probe, accessed 2016.

(30) A. Bockmann, M. Ernst and B. H. Meier, *J. Magn. Reson.*, **2015**, 253, 71.
(31) V. E. Zorin, S. P. Brown and P. Hodgkinson, *J. Chem. Phys.*, **2006**, 125, 144508.
(32) "Essential Techniques of Spin-1/2 Nuclei", M. J. Duer in "*Solid-State NMR Spectroscopy: principles and applications",* Blackwell Science Ltd: Oxford, UK, 2002, p 73.
(33) "Going Further With Spin-1/2 Solid-State NMR", D. C. Apperley, R. K. Harris and P. Hodgkinson in "*Solid State NMR Basic Principles and Practise*", Mentum Press: LLC, New York, US, 2012, p 109.
(34) A. Lesage, M. Bardet and L. Emsley, *J. Am. Chem. Soc.*, **1999**, 121, 10987.
(35) S. Cadars, A. Lesage and L. Emsley, *J. Am. Chem. Soc.*, **2005**, 127, 4466.
(36) R. K. Harris and A. C. Olivieri, *Prog. Nucl. Magn. Reson. Spectrosc.*, **1992**, 24, 435.
(37) S. H. Alarcón, A. C. Olivieri and R. K. Harris, *Solid State Nucl. Magn. Reson.*, **1993**, 2, 325.
(38) R. K. Harris, M. M. Sünnetçioglu, K. S. Cameron and F. G. Riddell, *Magn. Reson. Chem.*, **1993**, 31, 963.
(39) J. G. Hexem, M. H. Frey and S. J. Opella, *J. Am. Chem. Soc.*, **1981**, 103, 224.
(40) M. Alla and E. Lippmaa, *Chem. Phys. Lett.*, **1982**, 87, 30.
(41) D. L. Vanderhart, W. L. Earl and A. N. Garroway, *J. Magn. Reson.*, **1981**, 44, 361.
(42) A. J. Robbins, W. T. Ng, D. Jochym, T. W. Keal, S. J. Clark, D. J. Tozer and P. Hodgkinson, *Phys. Chem. Chem. Phys.*, **2007**, 9, 2389.
(43) *NMR Crystallography Commission* http://www.iucr.org/iucr/commissions/nmr-crystallography, accessed February 2016.
(44) D. E. Bugay, *Adv. Drug Deliv. Rev.*, **2001**, 48, 43.
(45) R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates and V. Zorin, *Magn. Reson. Chem.*, **2007**, 45, 174.
(46) B. Elena, G. Pintacuda, N. Mifsud and L. Emsley, *J. Am. Chem. Soc.*, **2006**, 128, 9555.
(47) E. Salager, R. S. Stein, C. J. Pickard, B. Elena and L. Emsley, *Phys. Chem. Chem. Phys.*, **2009**, 11, 2610.
(48) D. V. Dudenko, P. A. Williams, C. E. Hughes, O. N. Antzutkin, S. P. Velaga, S. P. Brown and K. D. Harris, *J. Phys. Chem. C*, **2013**, 117, 12258.
(49) D. Luedeker, R. Gossmann, K. Langer and G. Brunklaus, *Cryst. Growth Des.*, **2016**, 16, 3087.
(50) S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson and M. C. Payne, *Z. Krystallogr.*, **2005**, 220, 567.
(51) K. Lejaeghere, G. Bihlmayer, T. Bjorkman, P. Blaha, S. Blugel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso *et al.*, *Science*, **2016**, 351, aad3000.
(52) C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azais, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri *et al.*, *Chem. Rev.*, **2012**, 112, 5733.
(53) W. Kohn and L. J. Sham, *Phys. Rev. A*, **1965**, 140, 1133.
(54) F. A. Hamprecht, A. J. Cohen, D. J. Tozer and N. C. Handy, *J. Chem. Phys.*, **1998**, 109, 6264.
(55) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, **1996**, 77, 3865.
(56) V. Barone and C. Adamo, *J. Chem. Phys.*, **1996**, 105, 11007.
(57) J. Ireta, J. Neugebauer and M. Scheffler, *J. Phys. Chem. A*, **2004**, 108, 5692.
(58) A. D. Boese, *ChemPhysChem.*, **2015**, 16, 978.
(59) V. L. Deringer, V. Hoepfner and R. Dronskowski, *Cryst. Growth Des.*, **2012**, 12, 1014.
(60) A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, **2009**, 102, 073005.
(61) P. G. Karamertzanis, G. M. Day, G. W. Welch, J. Kendrick, F. J. Leusen, M. A. Neumann and S. L. Price, *J. Chem. Phys.*, **2008**, 128, 244708.
(62) J. van de Streek and M. A. Neumann, *Acta Crystallogr. B*, **2010**, 66, 544.
(63) K. S. Thanthiriwatte, E. G. Hohenstein, L. A. Burns and C. D. Sherrill, *J. Chem. Theory Comput.*, **2011**, 7, 88.

(64) J. Klimes and A. Michaelides, *J. Chem. Phys.*, **2012**, 137, 120901.
(65) S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Phys.*, **2014**, 141, 164121.
(66) S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Theory Comput.*, **2015**, 11, 5229.
(67) J. D. Hartman, T. J. Neubauer, B. G. Caulkins, L. J. Mueller and G. J. Beran, *J. Biomol. NMR*, **2015**, 62, 327.
(68) J. Weber and J. Schmedt Auf der Gunne, *Phys. Chem. Chem. Phys.*, **2010**, 12, 583.
(69) D. Vanderbilt, *Phys. Rev. B*, **1990**, 41, 7892.
(70) J. R. Yates, C. J. Pickard and F. Mauri, *Phys. Rev. B*, **2007**, 76, 024401.
(71) E. C. Snow and J. H. Wood, *Chem. Phys. Lett.*, **1974**, 25, 111.
(72) C. J. Pickard and F. Mauri, *Phys. Rev. B*, **2001**, 63, 245101.
(73) T. Charpentier, *Solid State Nucl. Magn. Reson.*, **2011**, 40, 1.
(74) J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller and G. J. Beran, *Phys. Chem. Chem. Phys.*, **2016**, 18, 21686.
(75) J. D. Hartman, G. M. Day and G. J. Beran, *Cryst. Growth Des.*, **2016**, 16, 6479.
(76) X. Li, L. Tapmeyer, M. Bolte and J. van de Streek, *ChemPhysChem.*, **2016**, 17, 2496.
(77) B. Monserrat, R. J. Needs and C. J. Pickard, *J. Chem. Phys.*, **2014**, 141, 134113.
(78) "Ab Initio Molecular Dynamics: Theory and Implementation", D. Marx and J. Hutter in "*Modern Methods and Algorithms of Quantum Chemistry",* J. Grotendorst (Eds.), John von Neumann Institute for Computing, NIC Series: Julich, Germany, 2000, p 301.
(79) "Introduction to Molecular Dynamics Simulation", M. P. Allen in "*Computational Soft Matter: From Synthetic Polymers to Proteins John von Neumann Institute for Computing series",* N. Attig, K. Binder, H. Grubmuller, K. Kremer (Eds.), John von Neumann Institute for Computing, NIC Series: Julich, Germany, 2004, p 1.
(80) A. J. Ilott, S. Palucha, P. Hodgkinson and M. R. Wilson, *J. Phys. Chem. B*, **2013**, 117, 12286.
(81) M. Dracinsky and P. Hodgkinson, *Chem. Eur. J.*, **2014**, 20, 2201.
(82) G. J. Beran, *Chem. Rev.*, **2016**, 116, 5567.
(83) E. Gibney, *Nature*, **2015**, 527, 20.
(84) *"Fundamentals of Crystallography*", C. Giacovazzo, H. Monaco, D. Artioli, D. Viterbo, D. Ferraris, G. Gilli, G. Zanotti and M. Catti (Eds.), Oxford University Press, 2002.
(85) IUCr*, Online dictionary of crystallography: Structure Factor*, http://reference.iucr.org/dictionary/Structure_factor, accessed June 2016.
(86) IUCr*, Online dictionary of crystallography: R Factor*, http://reference.iucr.org/dictionary/R_factor, accessed August 2016.
(87) A. D. Bond, *CrystEngComm.*, **2012**, 14, 2363.
(88) B. Rodriguez-Spong, C. P. Price, A. Jayasankar, A. J. Matzger and N. Rodriguez-Hornedo, *Adv. Drug Del. Rev.*, **2004**, 56, 241.
(89) *"Structural Inorganic Chemistry*", 5th ed., A. F. Wells (Eds.), Oxford University Press: Oxford, UK, 1984.
(90) S. Cadars, A. Lesage, C. J. Pickard, P. Sautet and L. Emsley, *J. Phys. Chem. A*, **2009**, 113, 902.
(91) R. W. Grosse-Kunstleve and P. D. Adams, *J. Appl. Crystallogr.*, **2002**, 35, 477.
(92) C. M. Widdifield, H. Robson and P. Hodgkinson, *Chem. Commun.*, **2016**, 52, 6685.
(93) S. L. Price, *Chem. Soc. Rev.*, **2014**, 43, 2098.
(94) S. E. Ashbrook and D. McKay, *Chem. Commun.*, **2016**, 52, 7186.
(95) D. Tan, L. Loots and T. Friščić, *Chem. Commun.*, **2016**, 52, 7760.
(96) M. Banik, S. P. Gopi, S. Ganguly and G. R. Desiraju, *Cryst. Growth Des.*, **2016**, 16, 5418.
(97) S. Cherukuvada and A. Nangia, *Chem. Commun.*, **2014**, 50, 906.
(98) O. Almarsson, M. L. Peterson and M. Zaworotko, *Pharm Pat Anal*, **2012**, 1, 313.

(99) H. G. Brittain, *Cryst. Growth Des.*, **2012**, 12, 5823.
(100) R. Thakuria, A. Delori, W. Jones, M. P. Lipert, L. Roy and N. Rodriguez-Hornedo, *Int. J. Pharm.*, **2013**, 453, 101.
(101) S. Aitipamula, P. S. Chow and R. B. H. Tan, *CrystEngComm.*, **2014**, 16, 3451.
(102) P. A. Wood, N. Feeder, M. Furlow, P. T. A. Galek, C. R. Groom and E. Pidcock, *CrystEngComm.*, **2014**, 16, 5839.
(103) G. Bolla and A. Nangia, *Chem. Commun.*, **2016**, 52, 8342.
(104) S. Cherukuvada, R. Kaur and T. N. Guru Row, *CrystEngComm.*, **2016**, 18, 8528.
(105) N. K. Duggirala, M. L. Perry, O. Almarsson and M. J. Zaworotko, *Chem. Commun.*, **2016**, 52, 640.
(106) N. Shan and M. J. Zaworotko, *Drug Discov. Today*, **2008**, 13, 440.
(107) E. Grothe, H. Meekes, E. Vlieg, J. H. ter Horst and R. de Gelder, *Cryst. Growth Des.*, **2016**, 16, 3237.
(108) A. D. Bond, *CrystEngComm.*, **2007**, 9, 833.
(109) G. R. Desiraju, *CrystEngComm.*, **2003**, 5, 466.
(110) G. R. Desiraju, *Angew. Chem.*, **1995**, 34, 2311.
(111) J. C. Krantz, J. M. Holbert, H. K. Iwamoto and C. J. Carr, *J. Am. Pharm. Assoc.*, **1947**, 36, 248.
(112) D. J. Good and N. Rodriguez-Hornedo, *Cryst. Growth Des.*, **2009**, 9, 2252.
(113) *Everything Added to Food in the United States (EAFUS)*, http://www.accessdata.fda.gov/scripts/fcn/fcnnavigation.cfm?rpt=eafuslisting, accessed April 2016.
(114) G. P. Stahly, *Cryst. Growth Des.*, **2007**, 7, 1007.
(115) G. R. Desiraju, *J. Am. Chem. Soc.*, **2013**, 135, 9952.
(116) C. B. Aakeröy, N. R. Champness and C. Janiak, *CrystEngComm.*, **2010**, 12, 22.
(117) N. Issa, P. G. Karamertzanis, G. W. A. Welch and S. L. Price, *Cryst. Growth Des.*, **2009**, 9, 442.
(118) P. G. Karamertzanis, A. V. Kazantsev, N. Issa, G. W. Welch, C. S. Adjiman, C. C. Pantelides and S. L. Price, *J. Chem. Theory Comput.*, **2009**, 1432.
(119) CCDC*, CCDC Blind Test Showcases Major Advance in Crystal Structure Prediction Methods*, https://www.ccdc.cam.ac.uk/News/List/post-43/, accessed September 2016.
(120) D. R. Weyna, T. Shattock, P. Vishweshwar and M. J. Zaworotko, *Cryst. Growth Des.*, **2009**, 9, 1106.
(121) S. L. James, C. J. Adams, C. Bolm, D. Braga, P. Collier, T. Friščić, F. Grepioni, K. D. Harris, G. Hyett, W. Jones *et al.*, *Chem. Soc. Rev.*, **2012**, 41, 413.
(122) D. Braga, L. Maini and F. Grepioni, *Chem. Soc. Rev.*, **2013**, 42, 7638.
(123) A. Jayasankar, A. Somwangthanaroj, Z. J. Shao and N. Rodriguez-Hornedo, *Pharm. Res.*, **2006**, 23, 2381.
(124) M. D. Eddleston, N. Madusanka and W. Jones, *J. Pharm. Sci.*, **2014**, 103, 2865.
(125) M. B. Alsirawan, V. R. Vangala, J. Kendrick, F. J. J. Leusen and A. Paradkar, *Cryst. Growth Des.*, **2016**, 16, 3072.
(126) US-FDA*, Guidence for Industry: Regulatory Classification of Pharmaceutical Co-Crystals*, http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryi%20nformation/guidances/ucm281764.pdf, accessed 2016.
(127) Novartis*, Novartis' new heart failure medicine LCZ696, now called Entresto™, approved by FDA to reduce risk of cardiovascular death and heart failure hospitalisation*, https://www.novartis.com/news/media-releases/novartis-new-heart-failure-medicine-lcz696-now-called-entrestotm-approved-fda, accessed 2016.
(128) W. T. Harrison, H. S. Yathirajan, S. Bindya, H. G. Anilkumar and Devaraju, *Acta Crystallogr. C*, **2007**, 63, 129.

(129) D. Bernhardson, T. A. Brandt, C. A. Hulford, R. S. Lehner, B. R. Preston, K. Price, J. F. Sagal, M. J. St Pierre, P. H. Thompson and B. Thuma, *Org. Process Res. Dev.*, **2014**, 18, 57.
(130) Merck and Pfizer, *Efficacy and Safety of Ertugliflozin (MK-8835/PF-04971729) With Sitagliptin in the Treatment of Participants With Type 2 Diabetes Mellitus (T2DM) With Inadequate Glycemic Control on Diet and Exercise (MK-8835-017)*, https://clinicaltrials.gov/ct2/show/study/NCT02226003, accessed 2016.
(131) J. Nyman and G. M. Day, *CrystEngComm.*, **2015**, 17, 5154.
(132) J. Aaltonen, M. Alleso, S. Mirza, V. Koradia, K. C. Gordon and J. Rantanen, *Eur. J. Pharm. Biopharm.*, **2009**, 71, 23.
(133) A. Lemmerer, D. A. Adsmond, C. Esterhuysen and J. Bernstein, *Cryst. Growth Des.*, **2013**, 13, 3935.
(134) F. Fischer, A. Heidrich, S. Greiser, S. Benemann, K. Rademann and F. Emmerling, *Cryst. Growth Des.*, **2016**, 16, 1701.
(135) D. Hasa, E. Carlino and W. Jones, *Cryst. Growth Des.*, **2016**, 16, 1772.
(136) Y. L. Slovokhotov, *Cryst. Growth Des.*, **2014**, 14, 6205.
(137) M. Mosharraf, T. Sebhatu and C. Nystrom, *Int. J. Pharm.*, **1999**, 177, 29.
(138) B. C. Hancock and M. Parks, *Pharm. Res.*, **2000**, 17, 397.
(139) A. Newman and G. Zografi, *J. Pharm. Sci.*, **2014**, 103, 2595.
(140) D. Murphy, F. Rodriguez-Cintron, B. Langevin, R. C. Kelly and N. Rodriguez-Hornedo, *Int. J. Pharm.*, **2002**, 246, 121.
(141) D. Zhou, G. G. Zhang, D. Law, D. J. Grant and E. A. Schmitt, *J. Pharm. Sci.*, **2002**, 91, 1863.
(142) G. D. Enright, V. V. Terskikh, D. H. Brouwer and J. A. Ripmeester, *Cryst. Growth Des.*, **2007**, 7, 1406.
(143) A. Bērziņš and P. Hodgkinson, *Solid State Nucl. Magn. Reson.*, **2015**, 65, 12.
(144) R. C. B. Copley, S. A. Barnett, P. G. Karamertzanis, K. D. M. Harris, B. M. Kariuki, M. C. Xu, E. A. Nickels, R. W. Lancaster and S. L. Price, *Cryst. Growth Des.*, **2008**, 8, 3474.
(145) E. Skořepová, J. Čejka, M. Hušák, V. Eigner, J. Rohlíček, A. Šturc and B. Kratochvíl, *Cryst. Growth Des.*, **2013**, 13, 5193.
(146) M. Urbanová, A. Šturcová, J. Brus, H. Beneš, E. Skořepová, B. Kratochvíl, J. Čejka, I. Šeděnková, L. Kobera, O. Policianová *et al.*, *J. Pharm. Sci.*, **2013**, 102, 1235.
(147) S. Mohamed, D. A. Tocher and S. L. Price, *Int. J. Pharm.*, **2011**, 418, 187.
(148) X. Kong, A. Brinkmann, V. Terskikh, R. E. Wasylishen, G. M. Bernard, Z. Duan, Q. Wu and G. Wu, *J. Phys. Chem. B*, **2016**, 10.1021/acs.jpcb.6b08091.
(149) O. D. Putra, T. Yoshida, D. Umeda, K. Higashi, H. Uekusa and E. Yonemochi, *Cryst. Growth Des.*, **2016**, 16, 5223.
(150) S. L. Childs, G. P. Stahly and A. Park, *Mol. Pharm.*, **2007**, 4, 323.
(151) M. Skotnicki, D. C. Apperley, J. A. Aguilar, B. Milanowski, M. Pyda and P. Hodgkinson, *Mol. Pharm.*, **2016**, 13, 211.
(152) A. J. Ilott, S. Palucha, A. S. Batsanov, M. R. Wilson and P. Hodgkinson, *J. Am. Chem. Soc.*, **2010**, 132, 5179.
(153) D. C. Apperley, A. F. Markwell, R. K. Harris and P. Hodgkinson, *Magn. Reson. Chem.*, **2012**, 50, 680.
(154) A. L. Spek, *Acta Crystallogr. C*, **2015**, 71, 9.
(155) L. H. Thomas, T. R. Welberry, D. J. Goossens, A. P. Heerdegen, M. J. Gutmann, S. J. Teat, P. L. Lee, C. C. Wilson and J. M. Cole, *Acta Crystallogr. B*, **2007**, 63, 663.
(156) L. H. Thomas, G. A. Craig, C. A. Morrison, A. M. Reilly and C. C. Wilson, *Cryst. Growth Des.*, **2011**, 11, 2045.
(157) R. Grau-Crespo, S. Hamad, C. R. A. Catlow and N. H. de Leeuw, *J. Phys. Condens. Matter.*, **2007**, 19, 256201.
(158) M. Habgood, R. Grau-Crespo and S. L. Price, *Phys. Chem. Chem. Phys.*, **2011**, 13, 9590.

(159) M. Habgood, *Cryst. Growth Des.*, **2011**, 11, 3600.
(160) A. J. Cruz-Cabeza, G. M. Day and W. Jones, *Phys. Chem. Chem. Phys.*, **2011**, 13, 12808.
(161) S. E. Ashbrook and D. M. Dawson, *Acc. Chem. Res.*, **2013**, 46, 1964.
(162) A. D. Bond, C. Cornett, F. H. Larsen, H. Y. Qu, D. Raijada and J. Rantanen, *Cryst. Growth Des.*, **2013**, 13, 3665.
(163) X. Filip, I. G. Grosu, M. Miclăuş and C. Filip, *CrystEngComm.*, **2013**, 15, 4131.
(164) N. Wahlberg, P. Ciochon, V. Petricek and A. O. Madsen, *Cryst. Growth Des.*, **2014**, 14, 381.
(165) D. E. Braun, L. H. Koztecki, J. A. McMahon, S. L. Price and S. M. Reutzel-Edens, *Mol. Pharm.*, **2015**, 12, 3069.
(166) D. E. Braun, K. P. Nartowski, Y. Z. Khimyak, K. R. Morris, S. R. Byrn and U. J. Griesser, *Mol. Pharm.*, **2016**, 13, 1012.
(167) F. G. Vogt, R. C. B. Copley, R. L. Mueller, G. P. Spoors, T. N. Cacchio, R. A. Carlton, L. M. Katrincic, J. M. Kennady, S. Parsons and O. V. Chetina, *Cryst. Growth Des.*, **2010**, 10, 2713.

Chapter 2: Methods

2.1 Solid state nuclear magnetic resonance spectroscopy

All SSNMR experiments were carried out on a Varian InfinityPlus spectrometer or a Bruker Avance III HD spectrometer operating at 499.7 MHz for ^{1}H, 125.7 MHz for ^{13}C, 50.6 MHz for ^{15}N, and 76.7 MHz for ^{2}H. Three different probes were used to run NMR experiments. The main work-horse probe for most ^{13}C and all ^{2}H experiments was a 4 mm Bruker HX, spinning up to 14 kHz. ^{1}H experiments were generally performed on the 1.3 mm Bruker probe, which spins up to 62 kHz. Finally, a 5 mm Varian probe was used to acquire early ^{13}C and ^{15}N experiments due to the increased sensitivity afforded by the larger rotor size.

Chemical shifts were calibrated using standard replacement reference samples,[1] Table 1. The chemical shifts of the second axis in 2D experiments were referenced by equating the shifts of characteristic peaks to those observed in corresponding 1D spectra. All data was processed in Topspin, Bruker.

Table 1 Reference samples used in this work for chemical shift referencing.

Replacement reference sample	Chemical shift / ppm
Adamantane	^{1}H = 1.9
D_2O	^{2}H = 0.0
α-glycine	^{13}C (carboxyl) = 176.5
	^{15}N = −346.8
Saturated solution of NH_4NO_3	^{14}N (ammonium) = −3.8

2.1.1 Setting the magic angle

The magic angle was set periodically on all three probes during the course of this work. Generally the angle is not expected to drift significantly, but if the probes are handled it is possible to knock the magic angle settings. A sample of KBr was generally used because ^{79}Br has an NMR frequency very close to that of ^{13}C (125.2 MHz) and has a 50 % natural abundance. The ^{79}Br FID shows a large number of rotational echoes at slow MAS rates (*ca.*

4 kHz) that are very sensitive to changes in the magic angle. The FID is acquired continuously as the angle is adjusted until the maximum number of echoes are observed when the magic angle is set correctly, extending out to roughly 8 ms as in Figure 12(i). This method is not possible of the 1.3 mm probe, since the spin rate is unstable at <45 kHz. Instead, the ^{2}H spectrum of alanine-d_2 is used since the line shape of the centre band is sensitive to the magic angle settings, shown in Figure 12(ii). When the angle is set correctly the linewidth of the ^{2}H peak is roughly 55 Hz.

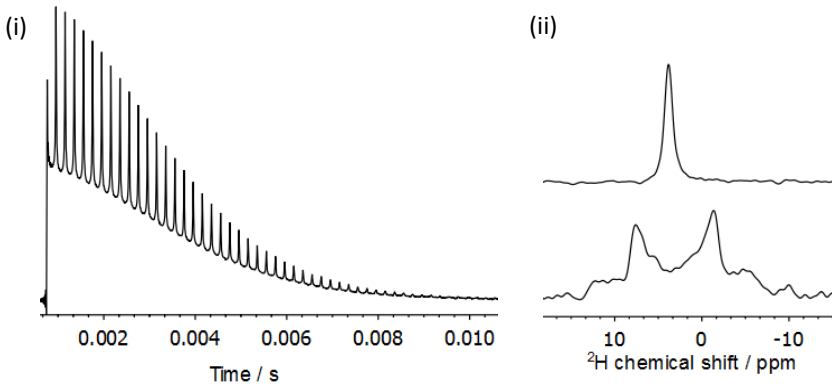

Figure 12 (i) ^{79}Br FID of KBr at 4 kHz MAS in the 4 mm probe on the magic angle. The rotary echoes extend below 8 ms. **(ii)** ^{2}H spectra of alanine-d_2 at 60 kHz MAS on the 1.3 mm probe on (top) and off (bottom) the magic angle.

2.1.2 Calibrating pulses

A sample of adamantane was used to calibrate ^{1}H pulses by systematically varying the pulse duration, τ, for a particular power. The resulting nutation curve is shown in Figure 13. The true 90˚ is at the maximum of the first peak in the nutation curve but this can be difficult to see precisely. It is easier to observe the first null point, where the magnetisation has been rotated by 180˚, and divide this value by two to get the 90˚ pulse length.* For accurate pulse calibration the RF must be applied on resonance and the adamantane rotor should be freshly packed, because sublimation of the sample will alter the shape of the nutation curve. An analogous procedure is carried out for X pulses with a labelled sample of α-glycine for ^{13}C and ^{15}N and using cross polarization (CP) to excite the dilute X spins. In this case the 90˚ is at the first null point. CP is discussed in detail below in section 2.1.4.

* Note that often the duration for a 90˚ pulse is slightly less than half that of the 180˚ pulse due to RF inhomogeneity.

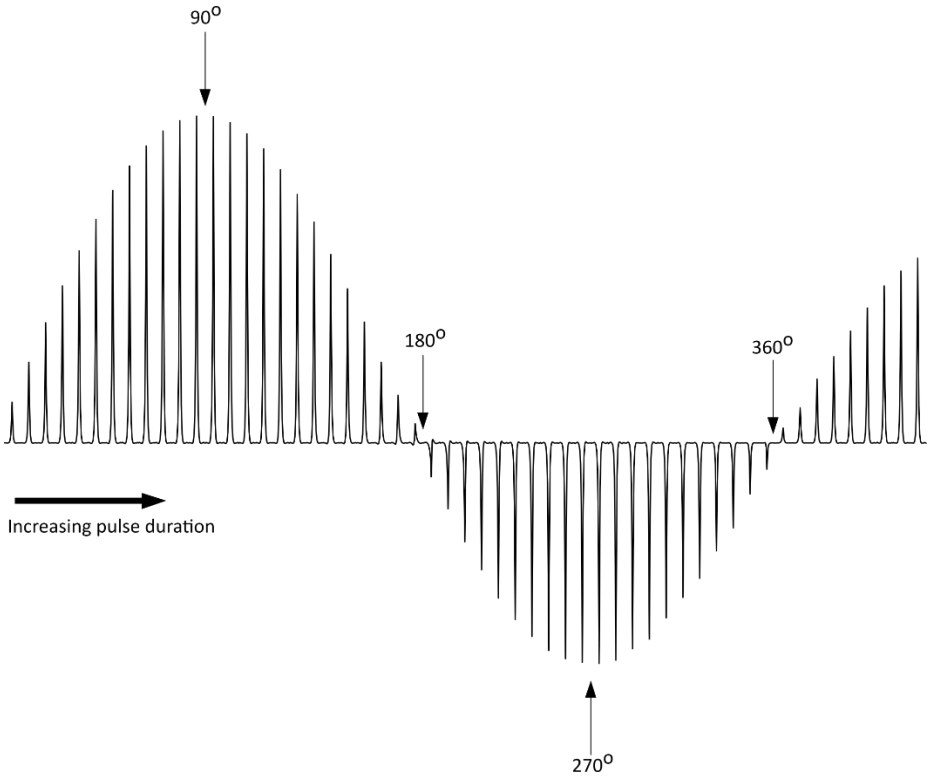

Figure 13 ^{1}H nutation curve of adamantane with τ ranging from 0.2 µs to 11 µs. The overall loss of signal intensity is due to RF inhomogeneity.

2.1.3 Decoupling

It has previously been stated that dipolar couplings involving ^{1}H spins lead to extensive line broadening in solid-state spectra. A key method of reducing this effect is by decoupling,[2-4] which aims to disrupt the dipolar coupling between the observed nuclei and the protons. Historically continuous wave (CW) decoupling was used to disrupt heteronuclear dipolar couplings;[3] a strong RF field is applied continuously on the proton channel during acquisition. At moderate MAS speeds this is generally effective if a high enough decoupling field can be applied. However, it is inefficient at fast MAS speeds because the ^{1}H spectrum breaks up into sidebands.[5] More complex heteronuclear decoupling methods have been developed over the years[2] and a common scheme is SPINAL-64,[6] though this is still an area of active research. SPINAL-64 is straight-forward to implement since the only parameter that usually needs optimising is the decoupling pulse length: ^{1}H τ_{180}. Note that adamantane is a poor setup sample for optimising decoupling because it is highly mobile so the ^{1}H peak is narrower than is usual for solid samples. A more representative sample is α-glycine and the CH_2 peak is particularly sensitive to the decoupling conditions, as demonstrated in Figure 14.

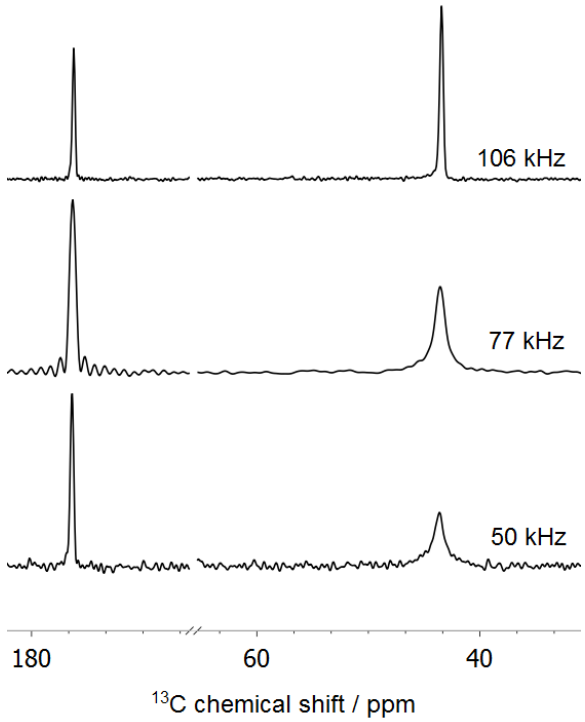

Figure 14 ^{13}C CP-MAS spectra of α-glycine acquired using SPINAL-64 heteronuclear decoupling with different ^{1}H decoupling fields. The experimental conditions are dissimilar, but the different decoupling strengths are the main contribution to the changes in CH_2 linewidth.

Homonuclear decoupling methods[7] involve multiple pulses applied to ^{1}H nuclei. In order to observe chemical shifts of homonuclear decoupled spectra the technique used is known as combined rotation and multiple-pulse spectroscopy (CRAMPS).[8] MAS is used to suppress the ^{1}H CSA contribution to the linewidth and multiple-pulse spectroscopy aims to suppress the homonuclear dipolar coupling. In this work the homonuclear decoupling sequence used is frequency-switched Lee-Goldberg (FSLG).[9] Briefly, FSLG decoupling involves the application of off-resonance pulses on the ^{1}H channel, with alternating phases, in order to simulate MAS in the rotating frame. The reader is referred to the original references and a recent review[7] for detailed descriptions of the theory of homonuclear decoupling sequences. Careful setup is required for these techniques to minimise artefacts and optimise performance,[10,11] but well-resolved ^{1}H spectra can be achieved. It is now possible to resolve two peaks separated by 0.13 ppm in the ^{1}H CRAMPS spectrum of adamantane,[12] which were previously considered to have a chemical shift difference that was too small to be visible by SSNMR.[13]

The ^{1}H nutation rates for the heteronuclear and homonuclear decoupling applied to each system in this work are detailed in Table 2. Experiments on CAF-2CA carried out on the 5 mm probe necessarily used lower powers.

Table 2 ^{1}H nutation rates used in this work for heteronuclear decoupling, SPINAL-64, and homonuclear decoupling, FSLG.

	SPINAL-64		FSLG
	1D / kHz	2D / kHz	2D / kHz
CAF-2CA	50[a]	50[c]	-
NPX-PA	78[a]	100[d],109[d]	100
2FS-INA	56[b] ,78[a]	83[e,f],109[d]	117
FS-PA	78[a]	83[e],100[d]	100

[a] `^{13}C observed, [b] ^{15}N observed, [c] ^{13}C T_2' relaxation time measurements, [d] ^{1}H–^{13}C heteronuclear correlation, [e] ^{13}C T_1 relaxation time measurements, [f] ^{13}C–^{13}C homonuclear correlation.

2.1.4 Pulse sequences

Direct excitation and cross polarization

Nuclei with large natural abundance and a high gyromagnetic ratio, such as ^{1}H, can be acquired with a direct excitation sequence, where a 90˚ pulse is applied and the resulting oscillating current is detected. Only a few FIDs need to be acquired since the signal is so large for ^{1}H. Other nuclei, such as ^{13}C and ^{15}N, have much lower natural abundances so the spectra cannot usually be acquired in a practical amount of time using direct excitation, unless the T_1 times are short.

The standard method of acquiring spectra of dilute nuclei is CP. This involves exciting an abundant spin, usually ^{1}H, and then transferring the ^{1}H magnetisation onto the nucleus of interest. In order for the magnetisation to be efficiently transferred the nutation rates of the two nuclei must be the same, known as the Hartmann-Hahn matching condition,[14] Equation (2.0).

$$\gamma^{H}B_1^{H} = \gamma^{X}B_1^{X} \text{ or } \nu_1^{H} = \nu_1^{X} \qquad \textbf{(2.0)}$$

At low spinning speeds (*ca*. 2 kHz) this is straightforward but at higher spinning speeds the Hartmann-Hahn match condition breaks down into a series of sidebands. It is, therefore, vital to optimise CP experiments at the target MAS speed. Figure 15 shows the CP pulse sequence. The ^{1}H spins are first excited with a 90˚ pulse. The second 90˚ pulse is phase

shifted compared to the initial pulse, and is called the spin-lock. This holds the ^{1}H magnetisation along $\mathbf{B}_1$ and the relaxation is slow, depending on ^{1}H $T_{1\rho}$ rather than T_2. It is usual to use a ramped spin-lock pulse on the ^{1}H channel in order to sweep through the sideband matching conditions for a robust Hartman-Hahn match.[15,16] During the spin-lock period a pulse is simultaneously applied to the X spins. The time over which these two pulses are applied is called the contact time, and is typically of the order of a few ms for ^{1}H–^{13}C CP, though the exact value is sample-dependent and should be optimised. If the contact time is too long then ^{1}H $T_{1\rho}$ relaxation will cause the spin-locked ^{1}H magnetisation to decay, but if the contact time is too short then the maximum magnetisation transfer will not have occurred. Signal arising from direct excitation of the X spins is averaged out by phase cycling.

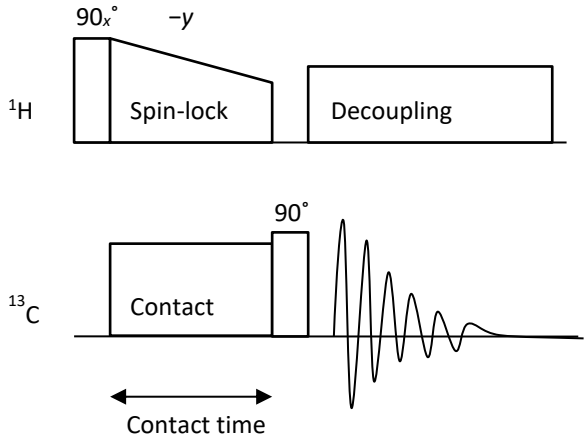

Figure 15 Schematic of the CP pulse sequence. In this example magnetisation is transferred from ^{1}H to ^{13}C and a ramped spin-lock pulse is applied on the ^{1}H channel. Not to scale.

The benefits of CP are two-fold. Firstly, the sensitivity of the experiment is determined by the sensitivity of the abundant spin, rather than that of the dilute one, so a much higher signal/noise ratio can be achieved. Secondly, the time between successive FID acquisitions is determined by the T_1 of the abundant spin. In solids ^{1}H T_1 times are usually significantly shorter than those of other spins so more FIDs can be acquired in a practical time than in a direct excitation experiment.

A drawback of CP is that the experiment is not quantitative, *i.e*. the signal intensities are not necessarily proportional to the number of spins contributing to the signal. Spin

diffusion means that the ^{1}H $T_{1\rho}$ times are essentially uniform if the sample is homogeneous, but the time constant for CP is site-dependent. This is because the local strengths of the H–X dipolar interactions are not uniform. For example, CP in a CH_2 group will be quicker than that of a quaternary carbon site since there are more protons in close proximity to the CH_2 carbon nucleus.[17] Quantitative information can be regained by acquiring several spectra over a range of contact times and fitting the resulting curves[15,18] but this is not always practical. There are occasions where CP is not effective at all. The magnetisation transfer relies upon dipolar coupling so anything that disrupts this will reduce the CP efficiency. For example, molecular motion can be sufficient to average out the dipolar coupling, rendering some sites invisible to CP experiments.

Modifications of cross polarisation

Spinning sidebands can complicate spectra if sufficiently fast MAS speeds are not available, especially if they overlap with peaks of interest. It is straightforward to acquire spectra at two different spinning speeds to identify the sidebands since these peaks will move as the MAS speed is changed. Alternatively, total sideband suppression (TOSS)[19,20] is a method that can be used in tandem with CP to average out the sideband manifold, leaving only the isotropic peak behind. Sideband manipulation using TOSS has been reviewed extensively by Antzutkin,[21] so the reader is referred to this paper and the references therein for a discussion of the theory. Practically, the CP/TOSS experiment is the same as the usual CP experiment with the addition of a series of π pulses on the ^{13}C channel. It is effective at removing small sideband manifolds at slow to moderate spinning speeds (up to *ca*. 8 kHz).

In addition to CP/TOSS, a dipolar dephasing time can be introduced before acquisition. During this time period the heteronuclear decoupling is turned off. This causes the magnetisation of X nuclei to decay *via* heteronuclear dipolar coupling. Little magnetisation is left on CH_2 and CH carbon sites, while quaternary sites retain some magnetisation as they are spatially distant from proton sites. This is known as CP/TOSS/NQS, where NQS stands for non-quaternary suppression. It should be noted that methyl carbon sites are not suppressed efficiently with NQS because the fast rotation of the group disrupts the heteronuclear dipolar coupling during the dephasing time.

Echo experiments

A certain amount of dead time is necessary before signal detection in order for ringing in the probe caused by the RF pulses to stop. However, spectra with very broad lines have rapidly decaying FIDs so a significant part of the signal is lost before acquisition can begin.

Spectral distortions caused by this require significant amounts of first order phase correction to fix, which can result in baseline roll.[22] Such a problem can be overcome using a spin-echo experiment,[23] Figure 16. The transverse magnetisation dephases during the first echo delay period and then refocuses during the second. The echo delay, τ, is set to be long enough to include the probe dead time and should be an integral number of rotor periods to make use of the rotational echoes from MAS. It is useful to optimise τ_2 such that acquisition starts at the maximum of the echo; τ_2 is usually slightly less than τ_1.

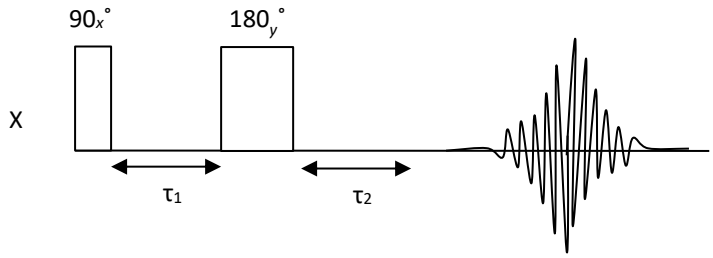

Figure 16 Schematic representation of a spin-echo experiment. τ_1 and τ_2 are the echo delays**.** Not to scale.

A related experiment is the quadrupolar echo*,* or solid echo, sequence. This is similar to the sequence in Figure 16, but the second pulse is a 90_y° pulse rather than 180_y°. A detailed explanation of the difference between spin echoes and quadrupolar echoes is beyond the scope of this work,[24,25] and relates to the fact that the quadrupolar interaction is affected by RF in a different way to the Zeeman interaction.

Relaxation measurements

As mentioned previously, measurement of T_2 relaxation times in the solid state is complicated by the additional contribution of residual dipolar coupling. In solution-state NMR it is straightforward to measure T_2 using Carr-Purcell-Meiboom-Gill spin-echo methods.[26] This is similar to the solid-echo scheme shown in Figure 16, but a series of 180° pulses are applied producing a series of echoes. Signal continues to be observed with sequential 180° pulses, but the amplitudes of these echoes decay exponentially. In the solid state this signal decay is not usually exponential and it is more usual to simply acquire a series of spin-echo experiments with an array of τ values, and fit the resulting signal intensities using Equation (2.1). This fit yields a time constant, T_2', which is the effective T_2

time constant under a spin echo. In this report T_2' values were measured for ^{13}C sites, so the echo sequence was combined with CP and decoupling, Figure 17.

$$S(\tau) = S_0 e^{-\frac{\tau}{T_2'}} \quad \textbf{(2.1)}$$

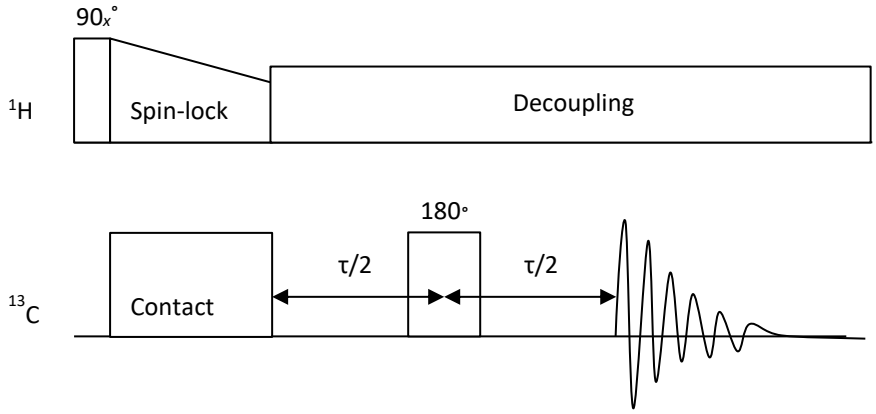

Figure 17 Schematic representation of the CP sequence with an additional 180˚ pulse to generate a spin-echo. Successive repetitions of this sequence with incremented delay values produce a pseudo-2D echo decay, which can be fit to extract T_2' relaxation times. Not to scale.

There are several different methods for measuring T_1 relaxation times.[27] Inversion-recovery, Figure 18(i), involves inverting the magnetisation along $-z$ and then waiting for τ, the recovery delay, before applying a 90˚ pulse to detect the z component of the magnetisation. This is carried out over an array of τ times, and the resulting signal intensity recovery is fit to Equation (2.2), where S_0 is the signal intensity when $\tau = 0$. Inversion recovery has the advantage of having a large range of possible intensities, from $-S_0$ to $+S_0$, but the recycle delay must be 4–5 times T_1 in order for the system to have returned to equilibrium before successive experiments.

$$S(\tau) = S_0(1 - 2e^{-\frac{\tau}{T_1}}) \quad \textbf{(2.2)}$$

An alternative experiment is saturation-recovery, Figure 18(ii). The first stage of this sequence is a train of 90˚ pulses that aims to fully saturate the spins. The delay between these pulses must be greater than T_2 for saturation to occur, meaning the experiment can

only be used for systems where T_2 is much less than T_1.* Since the magnetisation is saturated there is no need to wait between successive experiments, so the recycle delay can be very short.† At short τ times the signal is essentially zero and the signal increases as τ is increased. The recovering signal is fit using Equation (2.3).

$$S(\tau) = S_0(1 - e^{-\frac{\tau}{T_1}}) \quad \textbf{(2.3)}$$

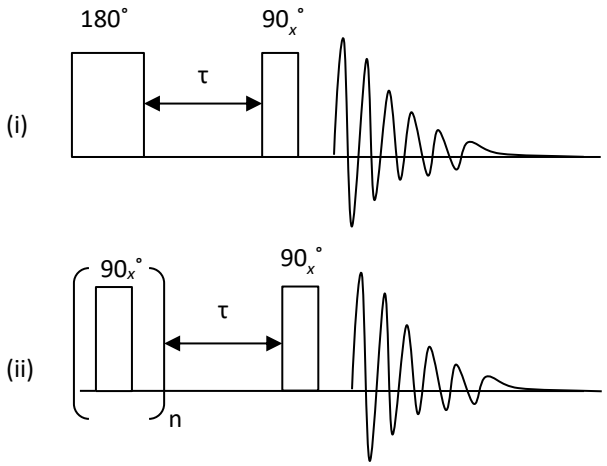

Figure 18 Schematic representations of the **(i)** inversion-recovery and **(ii)** saturation-recovery pulse sequences used to measure T_1 relaxation times. Not to scale.

Heteronuclear correlation

Neighbouring atom proximities can be probed using 2D correlation techniques. An example is the heteronuclear correlation (HETCOR) experiment,[28] which records correlations between ^{1}H spins and neighbouring X spins and is mediated by dipolar coupling, providing information on through-space distances. Figure 19 shows a schematic of the pulse sequence for the HETCOR experiment used in this work, which is based on the standard CP sequence but with an evolution period, t_1, which is incremented for 2D acquisition. Homonuclear decoupling is necessary during t_1 in order to obtain useful resolution in the ^{1}H dimension. In this work FSLG homonuclear decoupling was used.[28] The length of the contact time in the mixing period determines the distance between ^{1}H and X spins over which magnetisation can be transferred. Usually, a short contact time of 50 μs to 200 μs is

* This is usually the case for rigid solids.

† If decoupling is being used then the recycle delay must be long enough to prevent the probe overheating.

used to allow only directly bonded ^{1}H and X correlations to be visible. Longer contact times allow magnetisation transfer to occur over longer distances of up to 3 Å.[10]

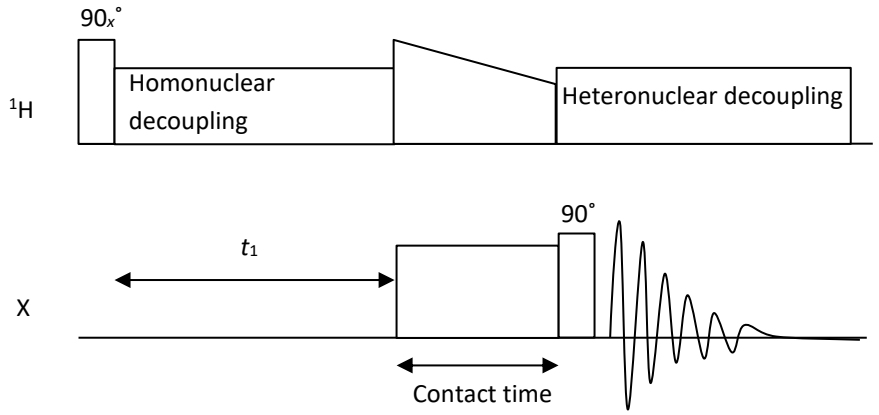

Figure 19 Schematic of a ^{1}H–X HETCOR experiment. FSLG homonuclear decoupling is applied as the ^{1}H magnetisation evolves before the magnetisation is transferred *via* CP to the X spins and subsequently detected. Not to scale.

Homonuclear correlation and coherences

Homonuclear correlation experiments rely upon couplings (dipolar or *J*-coupling) to mediate magnetisation transfer between neighbouring nuclei of the same type. Double quantum/single quantum (DQ/SQ) correlation experiments were used in this work to investigate ^{1}H–^{1}H connectivity. This type of experiment relies on the existence of coherences between spin states of different quantum number m_I in a coupled spin system. The origin of coherences is best described using a quantum mechanical approach rather than the vector model that was used in Chapter 1. In a two spin-1/2 system each spin can be in either the 'α' or 'β' state. There are four classical combinations of these two spins, Figure 20(i), |αα> with total $m_I = +1$, |αβ> and |βα> with total $m_I = 0$, and |ββ> with total $m_I = -1$. The quantum state of the ensemble of spins, on the other hand, is best described using a density operator, $\hat{\rho}$, Figure 20(ii). This takes the form of a matrix where the diagonal elements correspond to the populations of the spin states, *N*, and the off-diagonal elements are coherences between the states, *X*.

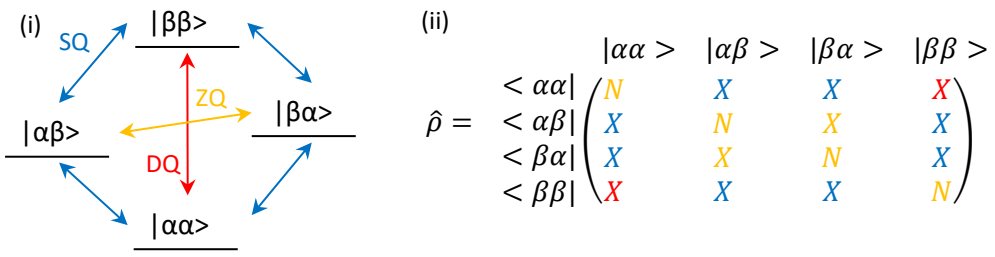

Figure 20 (i) Schematic representation of the origin of single quantum (blue), zero quantum (orange) and double quantum (red) coherences in a theoretical system containing two coupled spin-1/2 nuclei. **(ii)** Simple representation of the density matrix for the two spin-1/2 system, with the elements corresponding to different coherence orders coloured as in **(i)**. Diagonal elements, *N*, represent populations and off-diagonal elements, *X*, represent coherences.

The only detectable coherences are single quantum (SQ) coherences, shown in blue in Figure 20. These occur between states with $\Delta m_I = \pm 1$, described as having a coherence order of ±1, and are equivalent to transverse magnetisation.[29] The population elements and zero quantum (ZQ) coherences have a coherence order of 0, orange in Figure 20. Double quantum (DQ) coherences exist between states with a coherence order of ±2, red in Figure 20. Higher order coherences are termed multiple quantum (MQ) coherences. Coherence orders are transformed by RF pulses,[30] so DQ coherences can be converted into SQ coherences for detection. RF pulses excite a large number of different coherence pathways simultaneously so phase cycling is used to select only the magnetisation that has passed through the desired pathway.[29,31] For a more detailed description of the density operator approach and coherences the reader is referred to M. Levitt's textbook.[32]

A simplified schematic of a DQ/SQ pulse sequence to investigate ^{1}H-^{1}H correlations is shown in Figure 21. The initial block excites DQ coherence with rotor-synchronised 90˚ pulses. An evolution period allows the magnetisation to evolve under the influence of the homonuclear dipolar coupling network. This period is incremented for 2D detection. The reconversion period selectively converts the DQ coherence back to *z* magnetisation, and a final 90˚ pulse generates SQ coherence for detection. The time required for excitation and reconversion is dependent on the dipolar coupling strength: weaker couplings require a longer period. The particular pulse sequence used here was BABA (back-to-back), which utilises a train of back-to-back 90˚ pulses to excite DQ coherence over an number of rotor cycles, two in this case.[33] ^{1}H–^{1}H homonuclear correlation is complicated by the very broad

line shapes of ^{1}H spectra, so fast MAS is necessary for useful resolution; 50-60 kHz was used in this work.

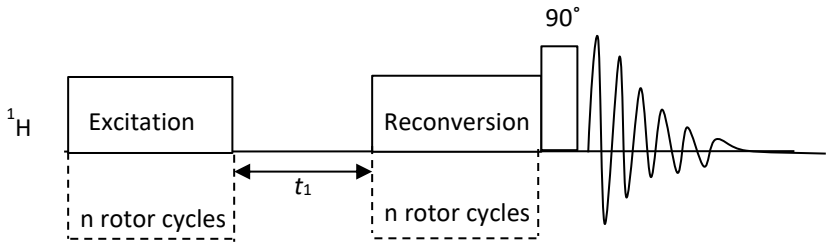

Figure 21 Schematic representations of a general DQ/SQ experiment. The evolution period, t_1, is incremented for 2D acquisition and the number of rotor cycles during excitation, n, is an integer. Not to scale.

Sites that are coupled will produce a pair of DQ/SQ correlation peaks at the same DQ frequency, Figure 22. The intensity of the correlation is related to the strength of the dipolar coupling between the sites, which is related to the internuclear distance.[34] However, the strong dipolar coupling between protons in a rigid CH_2 group can cause a large perturbation on the observed DQ correlations involving a CH_2 proton at a non-CH_2 SQ frequency, so the intensities cannot necessarily be quantitatively analysed to extract distance information.[35] This is known as dipolar truncation and it constrains the use of DQ/SQ experiments to measure long-range distances precisely.[36]

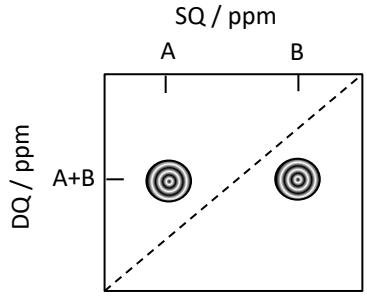

Figure 22 Schematic representation of the DQ/SQ spectrum of a coupled two-spin system. The coupled sites, A and B, result in a pair of correlation peaks at a DQ frequency equal to the sum of the two SQ frequencies.

An alternative experiment to study homonuclear correlations relies upon *J*-coupling to gain through-bond proximity information. An example is the INADEQUATE sequence[37]

(incredible natural abundance double quantum transfer experiment), Figure 23, which was used once in this work to examine ^{13}C–^{13}C connectivity. After an initial CP period the ^{13}C magnetisation evolves under the homonuclear *J*-coupling network for a period, 2τ. Dipolar couplings are removed by fast MAS and heteronuclear decoupling. DQ coherence is then excited, evolves during t_1, and is reconverted to SQ coherence for detection.[37] The use of DQ coherence means that strong signals from isolated ^{13}C atoms can be suppressed, leaving behind only the weak signals from directly bonded ^{13}C atoms that interact *via J*-coupling to produce DQ signals at the same frequency.[38,39] A refocused INADEQUATE (rINADEQUATE) sequence was used in this work, which contains additional refocussing 180˚ pulses in the middle of the 2τ periods, which removes inhomogeneous line broadening effects to increase the efficiency of the experiment. However, some signal is lost due to relaxation so the efficiency of the experiment depends on T_2'.[39]

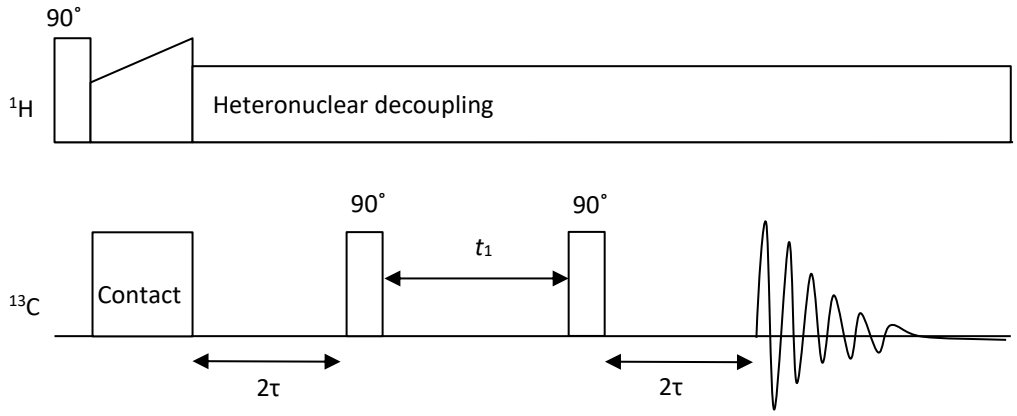

Figure 23 Simplified schematic of a ^{13}C–^{13}C INADEQUATE experiment. Magnetisation is transferred from ^{1}H to ^{13}C *via* CP and then evolves under ^{13}C–^{13}C *J*-coupling for a period of 2τ. DQ coherence is then excited by a 90˚ pulse, before reconversion to SQ coherence. A refocussing pulse is applied in the middle of each 2τ period. Decoupling is applied on the ^{1}H channel throughout to remove heteronuclear dipolar coupling effects. τ delays are rotor synchronised. Not to scale.

2.1.5 Temperature calibration

The Bruker Avance III HD spectrometer has variable temperature capabilities provided by a devoted gas flow and two temperature sensors, one measured near the gas flow entry (sensor 1) and the second near the gas flow exit after the sample (sensor 2). These sensors do not take into account frictional heating from sample rotation. Additionally, the two gas

flows associated with sample spinning, the bearing and drive, are not actively cooled during low temperature work and it was initially unclear how much of an effect this had on the true temperature of the sample. As a result, temperature calibration was carried out for the 4 mm probe, using neat methanol and neat ethylene glycol as reference samples. The chemical shift difference between the two ^{1}H signals, Δδ, of both methanol and ethylene glycol are highly temperature dependent, following Equations (2.4) and (2.5) respectively.[40] Neither compound can be used over the full temperature range: methanol can only be used between 178 K and 330 K, while ethylene glycol is useful between 273 K and 416 K.

$$T(\textit{methanol}) = 409.0\ \text{K} - 36.54\ \text{K} \times (\Delta\delta\ /\ \text{ppm}) - 21.85\ \text{K} \times (\Delta\delta\ /\ \text{ppm})^2 \quad \textbf{(2.4)}$$

$$T(\textit{ethylene glycol}) = 466.5\ \text{K} - 102.00\ \text{K} \times (\Delta\delta\ /\ \text{ppm}) \quad \textbf{(2.5)}$$

At room temperature the significance of frictional heating at various spinning speeds is evident in Figure 24. With no temperature control the true temperature of the sample can be described by Equation (2.6),* while temperature control at 298 K with a gas flow rate of 1000 liters per hour (lph) meant the true temperature followed Equation (2.7).

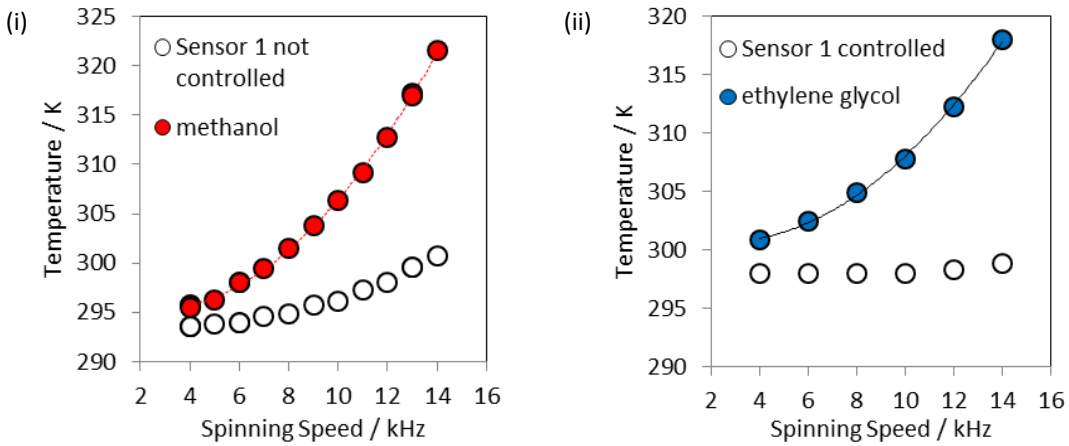

Figure 24 Temperature calibration curves at various MAS speeds on the 4mm probe. **(i)** Neat methanol (filled red points) spun with no temperature control and the corresponding reading for sensor 1 (open points). **(ii)** Neat ethylene glycol (filled blue points) spun with active temperature control set at 298 K with a gas flow rate of 1000 lph and the corresponding sensor 1 readings (open points).

* The ambient temperature in the room was roughly 20 ˚C.

$$T(\textit{no control}) = 296.34\ K - 0.93\ K \times (\nu_{MAS} / kHz) + 0.19\ K \times (\nu_{MAS} / kHz)^2 \quad \textbf{(2.6)}$$

$$T(\textit{control at 298 K}) = 301.4\ K - 0.61\ K \times (\nu_{MAS} / kHz) + 0.13\ K \times (\nu_{MAS} / kHz)^2 \quad \textbf{(2.7)}$$

The calibration data at 8 kHz MAS, Figure 25, shows that the temperature at the sample is close to the target temperature set on sensor 1 during variable temperature experiments with a variety of gas flow rates. The gradient of the line of best fit is 1.31(1) with an offset of −84(2) K. It is noted that the exact value of the gradient of Figure 25 will be different at other MAS speeds. Additionally, deviation away from linear behaviour is expected for cryogenic cooling because the set-up is different; bypassing the chiller unit and connecting the gas flow directly to a dewar of liquid nitrogen.

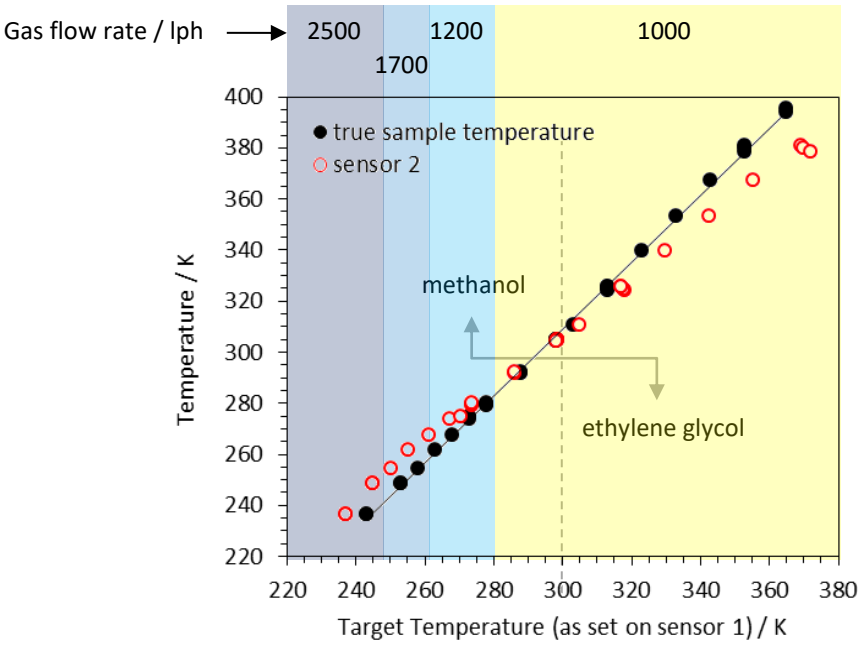

Figure 25 Temperature calibration data measured for the 4 mm probe at 8 kHz MAS using methanol and ethylene glycol as NMR thermometers. The temperature was set using sensor 1 and the gas flow rates are specified.

Temperature calibration has not yet been carried out on the 1.3 mm probe. However, Purusottam *et al*. have carried out calibration on their 1.3 mm Bruker probe using the temperature dependent shifts of ^{79}Br in a sample of adamantane physically mixed with KBr. The results indicate that, in the absence of any temperature control, a correction of 30 K is necessary at the maximum MAS speed of 62 kHz and a correction of 20 K is necessary at 50 kHz.[41]

2.2 Software

2.2.1 Density functional theory

DFT calculations of energy, geometry, and NMR parameters were carried out using CASTEP.[42] Several versions were available over the course of this work; 6.0, 8.0, and 16.1. The main differences between the software versions were the pseudopotentials available, though these differences are not expected to significantly impact the conclusions made in this work.

Aside from the physical approximations discussed in Chapter 1, there are also computational approximations that need to be considered when carrying out CASTEP calculations. In particular the *k*-point sampling and the size of the basis set, *i.e*. E_{cut}, must be optimised for each system as they are dependent on the identity of the elements present. Convergence testing was carried out by calculating the energy of the system of interest at several different E_{cut} values, generally using a *k*-point sample spacing of 0.1 $Å^{-1}$ as a starting point. Figure 26 shows convergence tests carried out for two different systems with respect to their final energy, using CASTEP version 6.0. Figure 26(i) was calculated using 2FS-INA, discussed in Chapter 5, and it is clear that there is a very large difference in the calculated final energy calculated at E_{cut} values of 100 eV to 400 eV. After this the difference is much smaller, highlighted in Figure 26(ii) for CAF-2CA, see Chapter 3. The final energy continues to decrease, even at E_{cut} values of 2000 eV, but the computational time increases roughly linearly so a suitable value must be chosen that balances these two factors. The forces acting on each atom were also monitored to ensure that subsequent geometry optimisation calculations would be suitably converged. These generally followed the same convergence behaviour as the final energy, though the forces on nitrogen and oxygen atoms were slower to converge (by roughly 100 eV).

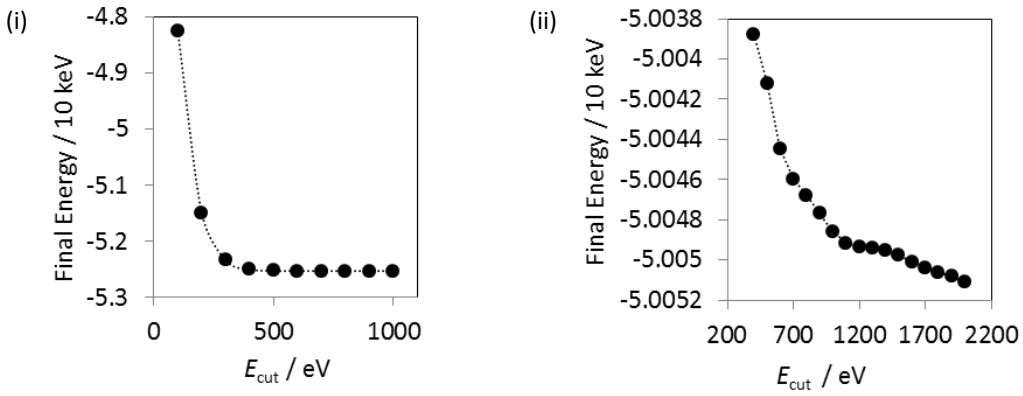

Figure 26 Examples of convergence testing of the final energy with respect to the size of the basis set, E_{cut}. **(i)** Final energy of 2FS-INA over a range of E_{cut} values using on-the-fly-generated pseudopotentials, see Table 3. **(ii)** Final energy of CAF-2CA over a range of E_{cut} values using ultrasoft pseudopotentials, see Table 3. Both used a *k*-point spacing of 0.1 Å^{-1}.

Once a suitable E_{cut} value was chosen the final energies were tested with respect to *k*-point spacing, see Figure 27 for CAF-2CA in CASTEP version 6.0. The final energy changes little with respect to *k*-point spacing in all cases, as do the forces on the atoms (not shown). A spacing of 0.5 Å^{-1} and 0.3 Å^{-1} is generally too coarse, as highlighted by the discrete jump in Figure 27(i), but there is essentially no difference when values below 0.1 Å^{-1} are used except in the computational time, Figure 27(ii).

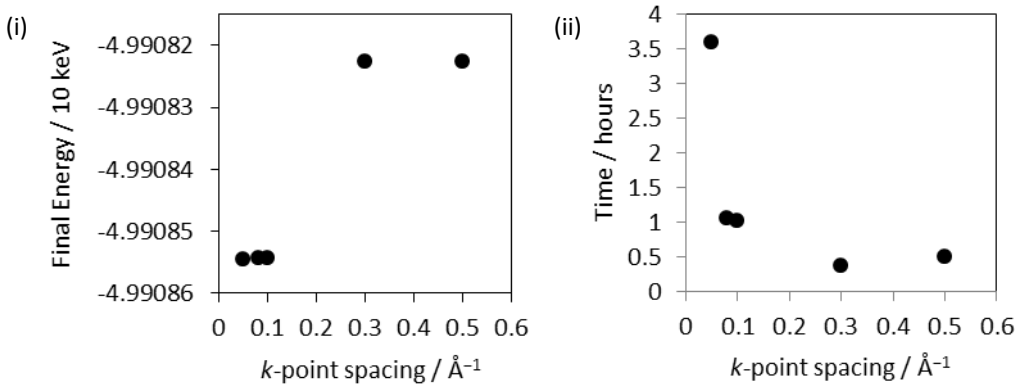

Figure 27 Examples of convergence testing of the *k*-point spacing for CAF-2CA. **(i)** Final energy at various *k*-point spacing values. **(ii)** Calculation times for single point calculations at various *k*-point spacing values. Note that these *k*-point spacing values correspond to a single *k*-point for this system, except for 0.05 Å^{-1} which corresponds to three *k*-points.

The convergence depends on the pseudopotentials used. On-the-fly generated (OTFG) pseudopotentials in CASTEP version 6.0 converged more slowly than Vanderbilt[43] ultrasoft (USP) pseudopotentials so the latter were chosen for geometry optimisations in Chapter 3. These were also the pseudopotentials used for the E_{cut} and *k*-point spacing convergence tests for CAF-2CA in Figure 26(ii) and Figure 27. However, CASTEP versions 8.0 and 16.1 contain improved OTFG pseudopotentials so these were used in the latter stages of this work, including the E_{cut} convergence testing of 2FS-INA in Figure 26(i). Table 3 shows the E_{cut} values, *k*-point spacing, and pseudopotentials chosen for geometry optimisation calculations for each system studied.

Table 3 E_{cut} values, *k*-point spacing and pseudopotentials chosen for geometry optimisations such that the final energy and atom forces were converged for the systems studied in this work.

System	E_{cut} / eV	*k*-point spacing / Å^{-1}	Pseudopotentials
CAF-2CA	400	0.1	USP
NPX-PA	700[a]	0.1	OTFG version 8.0
2FS-INA	500	0.1	OTFG version 6.0
FS-PA	700[a]	0.1	OTFG version 16.1

[a] A high E_{cut} value was chosen because the hydrogen bonding was of interest.

NMR calculations in CASTEP are more demanding than geometry optimisations and require a larger basis set and finer *k*-point spacing. As a result E_{cut} values of 600–700 eV were used with a *k*-point spacing of 0.05 Å^{-1}. An offset of (0.25, 0.25, 0.25) was also applied to avoid sampling the gamma point in reciprocal space. Additionally, only the OTFG pseudopotentials are currently compatible with the GIPAW method. The calculated shielding parameters were referenced in order to convert them into chemical shifts for comparison with experimental data. This was carried out by one of two methods, stated explicitly in the relevant chapters. The first was using Equation (2.8) where δ_{iso} is the calculated chemical shift, σ_{iso} is the calculated shielding and σ_{ref} is a constant that equates the average shielding value with the average experimental chemical shift.[44] σ_{ref} is dependent on the V_{XC} used.[45] A recent computational study of five of the common SSNMR reference compounds has suggested that σ_{ref} could be taken to be 168.9 ppm for ^{13}C, following a combination of DFT-MD and SSNMR calculations in CASTEP.[46] However, the accuracy of this result depends on the accuracy of the MD force fields used and the V_{XC} (PBE-D2), which was not the same as the V_{XC} used here (PBE). Therefore, σ_{ref} values were

calculated for each system individually in this report. The second method used the line of best fit obtained by linear regression of a plot of the calculated shielding *vs* the experimental shift, Equation (2.9). This is a more accurate referencing method since it takes into account the non-unity gradient.[46] However, accurate assignment of each atom site is necessary.

$$\delta_{iso} = \sigma_{ref} - \sigma_{iso} \quad \textbf{(2.8)}$$

$$\delta_{iso} = m\sigma_{iso} + c \quad \textbf{(2.9)}$$

2.2.2 Other software

The molecular viewer Mercury[47] was used to view crystal structure files, generally in cif format, as part of the Crystal Structure Database Software Package.[48] Editing of crystal structures was carried out using the free software Avogadro.[49] It is noted that Avogadro does not perceive the space group $P2_1/n$, but rather converts structures to the related space group $P2_1/c$. Therefore care must be taken that the correct space group is carried forward for CASTEP calculations. A locally modified molecular viewer, GDis,[50] was used to convert crystal structure files from cif format into the input file format for CASTEP (.cell).

Experimental SSNMR spectra were processed in Topspin and relaxation data was analysed using the T_1/T_2 relaxation centre. The resulting signal intensity *vs* delay curves were also fitted using QtiPlot[51] to extract error values. The errors on the signal intensities were estimated by establishing the noise in the spectrum with the longest delay and subsequently a weighting scheme was applied to the fit in QtiPlot using Equation (2.10) to give realistic errors on the T_1 values, where the value of the noise is c_i,. Experimental spectral images were produced using MestReNova[52] or the free vector graphics editor used to produce all other images: Inkscape™.[53]

$$w_i = \frac{1}{c_i^2} \quad \textbf{(2.10)}$$

Deconvolution of experimental spectra was carried out in Topspin using SOLA (solids line shape analysis) and also in the free software GSim,[54] to ensure robust fits were obtained. In both cases a mixture of Lorentzian and Gaussian line shapes were used. GSim was used as an interface to the simulation tool pNMRsim,[55] which was used to simulate and fit quadrupolar spectra in order to extract quadrupolar parameters, C_Q and η, along with associated errors. Static quadrupolar spectra for systems undergoing rotational dynamics

were simulated using EXPRESS,[56] exchange program for relaxing spin systems, as discussed in Chapter 6. In-house software (calcav2) was used to calculate the average effective NMR tensors corresponding to a two-site jump between orientations differing by 2θ under the limit of fast exchange using Equation (2.11). **A** represents the original tensor in the laboratory frame, **A**(θ) represents the tensor rotated through θ away from the *z* axis, and *f* denotes the fraction of time spent at each site. Diagonalising ⟨**A**⟩ gives the anisotropy and asymmetry parameters of the averaged tensor. When two uncorrelated dynamic process occur, a second jump motion of 2θ' can be applied by repeating the process. The situation is simpler if the dynamics involve an *n*-site jump between sites on a cone with half-angle θ, such as methyl rotation, where the anisotropy simply scales with ½(3cos^{2}θ − 1) and the asymmetry averages to zero.

$$\langle \mathbf{A} \rangle = f\mathbf{A}(-\theta) + (1 - f)\mathbf{A}(-\theta) \qquad \textbf{(2.11)}$$

Calculated NMR spectra produced by CASTEP version 6.0 could be read into GSim and viewed as simulated spectra. Calculated spectra in the newer format from versions 8.0 and 16.1 can be read by MagresView,[57] which was used to simulate 2D spectra.

2.3 Other experimental details

All the materials used in this work were purchased from Sigma Aldrich and used without further purification. The specific syntheses of the cocrystals, hydrates and solvates are given in the relevant chapters.

SCXRD data was collected by the Durham University crystallography service. SCXRD data for CAF-2CA was collected on an Xcalibur, Sapphire3, Gemini ultra diffractometer using graphite monochromated MoKα radiation. The raw data were collected using the CrysAlisPro (Agilent Technologies) software. Structure solution and refinement was carried out using SHELXS-97 and ShelXL2014.[58] SCXRD data for NPX-PA, discussed in Chapter 4, and the FS-PA solvates, discussed in Chapter 6, was collected on a Bruker Venture D8 instrument with an exposure time of 50 s / frame. Data reduction was carried out using the APEXII software.[59] Structure solution was carried out using SIR92 and structure refinement on F^2 was carried out using CRYSTALS.[60] PXRD data was acquired by Mr G. Oswald on a Bruker D8 Advance theta-theta diffractometer with $Cu_{K\alpha}$ radiation over a 2θ range of 5–50˚.

The solid-state FT-IR spectrum of FS-PA acetone solvate was acquired by Dr K. B. Vincent at the University of Huddersfield on a Thermo Scientific Nicolet 380 FT-IR spectrometer equipped with Diamond ATR.

2.4 References

(1) S. Hayashi and K. Hayamizu, *Bull. Chem. Soc. Jpn.*, **1991**, 64, 685.
(2) A. E. Bennett, C. M. Rienstra, M. Auger, K. V. Lakshmi and R. G. Griffin, *J. Chem. Phys.*, **1995**, 103, 6951.
(3) M. Ernst, *J. Magn. Reson.*, **2003**, 162, 1.
(4) P. Hodgkinson, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2005**, 46, 197.
(5) "Going Further With Spin-1/2 Solid-State NMR", D. C. Apperley, R. K. Harris and P. Hodgkinson in "*Solid State NMR Basic Principles and Practise*", Mentum Press: LLC, New York, US, 2012, p 109.
(6) T. Bräuniger, P. Wormald and P. Hodgkinson, *Monatsh. Chem.*, **2002**, 133, 1549.
(7) K. R. Mote, V. Agarwal and P. K. Madhu, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2016**, 97, 1.
(8) B. C. Gerstein, R. G. Pembleton, R. C. Wilson and L. M. Ryan, *J. Chem. Phys.*, **1977**, 66, 361.
(9) M. Lee and W. I. Goldburg, *Phys. Rev. A*, **1965**, 140, 1261.
(10) S. P. Brown, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2007**, 50, 199.
(11) A. J. Vega, *J. Magn. Reson.*, **2004**, 170, 22.
(12) D. H. Brouwer and M. Horvath, *Solid State Nucl. Magn. Reson.*, **2015**, 71, 30.
(13) I. Schnell and H. W. Spiess, *J. Magn. Reson.*, **2001**, 151, 153.
(14) S. R. Hartmann and E. L. Hahn, *Phys. Rev.*, **1962**, 128, 2042.
(15) G. Metz, M. Ziliox and S. O. Smith, *Solid State Nucl. Magn. Reson.*, **1996**, 7, 155.
(16) G. Metz, X. L. Wu and S. O. Smith, *J. Magn. Reson., Ser A*, **1994**, 110, 219.
(17) R. K. Harris, *Analyst*, **1985**, 110, 649.
(18) G. A. Stephenson, R. A. Forbes and S. M. Reutzel-Edens, *Adv Drug Deliv Rev*, **2001**, 48, 67.
(19) W. T. Dixon, J. Schaefer, M. D. Sefcik, E. O. Stejskal and R. A. Mckay, *J. Magn. Reson.*, **1982**, 49, 341.
(20) W. T. Dixon, *J. Chem. Phys.*, **1982**, 77, 1800.
(21) O. N. Antzutkin, *Prog. Nucl. Magn. Reson. Spectrosc.*, **1999**, 35, 203.
(22) "Spin-1/2 Nuclei: A Practical Guide", D. C. Apperley, R. K. Harris and P. Hodgkinson in "*Solid State NMR Basic Principles and Practise",* Mentum Press: LLC, New York, US, 2012, p 39.
(23) E. L. Hahn, *Phys. Rev.*, **1950**, 80, 580.
(24) I. Solomon, *Phys. Rev.*, **1958**, 110, 61.
(25) J. H. Davis, K. R. Jeffrey, M. Bloom, M. I. Valic and T. P. Higgs, *Chem. Phys. Lett.*, **1976**, 42, 390.
(26) H. Y. Carr and E. M. Purcell, *Phys. Rev.*, **1954**, 94, 630.
(27) "Relaxation, Exchange and Quantitation", D. C. Apperley, R. K. Harris and P. Hodgkinson in "*Solid State NMR Basic Principles and Practise",* Mentum Press: LLC, New York, US, 2012, p 177.
(28) B. J. van Rossum, H. Förster and H. J. M. de Groot, *J. Magn. Reson.*, **1997**, 124, 516.
(29) G. Bodenhausen, H. Kogler and R. R. Ernst, *J. Magn. Reson.*, **1984**, 58, 370.
(30) Y. S. Yen and A. Pines, *J. Chem. Phys.*, **1983**, 78, 3579.
(31) A. D. Bain, *J. Magn. Reson.*, **1984**, 56, 418.
(32) *"Spin Dynamics: Basics of Nuclear Magnetic Resonance*", 1st ed., M. H. Levitt (Eds.), John Wiley & Sons, Ltd: West Sussex, England, 2001.
(33) M. Feike, D. E. Demco, R. Graf, J. Gottwald, S. Hafner and H. W. Spiess, *J. Magn. Reson., Ser A*, **1996**, 122, 214.
(34) J. P. Bradley, C. Tripon, C. Filip and S. P. Brown, *Phys. Chem. Chem. Phys.*, **2009**, 11, 6941.

(35) 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.
(36) P. Hodgkinson and L. Emsley, *J. Magn. Reson.*, **1999**, 139, 46.
(37) A. Lesage, C. Auger, S. Caldarelli and L. Emsley, *J. Am. Chem. Soc.*, **1997**, 119, 7867.
(38) A. Bax, R. Freeman and S. P. Kempsell, *J. Am. Chem. Soc.*, **1980**, 102, 4849.
(39) A. Lesage, M. Bardet and L. Emsley, *J. Am. Chem. Soc.*, **1999**, 121, 10987.
(40) C. Ammann, P. Meier and A. Merbach, *J. Magn. Reson.*, **1982**, 46, 319.
(41) R. N. Purusottam, G. Bodenhausen and P. Tekely, *J. Magn. Reson.*, **2014**, 246, 69.
(42) S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson and M. C. Payne, *Z. Krystallogr.*, **2005**, 220, 567.
(43) D. Vanderbilt, *Physical review. B, Condensed matter*, **1990**, 41, 7892.
(44) R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates and V. Zorin, *Magn. Reson. Chem.*, **2007**, 45, 174.
(45) J. C. Johnston, R. J. Iuliucci, J. C. Facelli, G. Fitzgerald and K. T. Mueller, *J. Chem. Phys.*, **2009**, 131, 144503.
(46) X. Li, L. Tapmeyer, M. Bolte and J. van de Streek, *ChemPhysChem.*, **2016**, 17, 2496.
(47) C. F. Macrae, P. R. Edgington, P. McCabe, E. Pidcock, G. P. Shields, R. Taylor, M. Towler and J. van De Streek, *J. Appl. Crystallogr.*, **2006**, 39, 453.
(48) C. R. Groom and F. H. Allen, *Angew. Chem. Int. Ed.*, **2014**, 53, 662.
(49) M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminform.*, **2012**, 4, 17.
(50) S. Fleming and A. Rohl, *Z. Krystallogr.*, **2005**, 220, 580.
(51) I. Vasilef*, QtiPlot, Data Analysis and Scientific Visualisation*, www.qtiplot.com*,* accessed 2016.
(52) *MestReNova* v10.0.2-15465; Mestrelab Research S. L., 2015, www.mestrelab.com/software/mnova/nmr/.
(53) M. Owens*, Inkscape*, www.inkscape.org/en/, accessed 2016.
(54) V. Zorin*, GSim - a visualisation and processing program for solid-state NMR*, www.sourceforge.net/projects/gsim/, accessed 2016.
(55) P. Hodgkinson*, pNMRsim: a general simulation program for large problems in solid-state NMR*, www.dur.ac.uk/paul.hodgkinson/pNMRsim, accessed 2016.
(56) R. L. Vold and G. L. Hoatson, *J. Magn. Reson.*, **2009**, 198, 57.
(57) S. Sturniolo, T. F. 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.
(58) G. M. Sheldrick, *Acta Crystallogr. C*, **2015**, 71, 3.
(59) *APEX II* Bruker AXS Inc., 2012,
(60) P. W. Betteridge, J. R. Carruthers, R. I. Cooper, K. Prout and D. J. Watkin, *J. Appl. Crystallogr.*, **2003**, 36, 1487.

Chapter 3: Testing the limits of NMR crystallography

Based on the publication "*Testing the limits of NMR crystallography: the case of caffeine-citric acid hydrate*", H. E. Kerr, H. E. Mason, H. Sparkes and P. Hodgkinson, CrystEngComm., 2016, 18, 6700–6707, published by the Royal Society of Chemistry. Sections 3.3, 3.4 and 3.5 are lightly adapted from the publication.

3.1 Predicting static disorder effects

The average periodic structure of ordered crystalline systems is generally well characterised by Bragg diffraction data. However, disordered systems are not fully described by an average picture so local level information is required from complementary techniques like SSNMR. Dynamic disorder in organic systems can be probed by a variety of means including variable temperature relaxation times and multinuclear studies like ^{2}H SSNMR. Static disorder is more challenging to investigate. It introduces subtle differences in local environment, which leads to a distribution of chemical shifts for chemically-related sites, *i.e*. line broadening. Typically, this is of the order of a few ppm and can be too subtle to observe clearly.

Line broadening can provide hints of the local structure of statically disordered systems but it is challenging to interpret quantitatively. For example, trospium chloride[1] exhibits variable degrees of ^{13}C line broadening depending on its crystallisation conditions,[2] highlighted by blue boxes in Figure 28, but it is unclear whether this is due to varying degrees of disorder or the presence of impurities by SSNMR data alone. 2D SSNMR experiments have been used to increase the resolution of peaks broadening by disorder, especially when strong correlation is observed between the shifts of related sites.[3-6]

Analysis of these peaks can provide an indication as to the key contributing factor to the line broadening, by distinguishing between ABMS and local disorder effects. However, it is hard to extract quantitative structural information directly from the NMR data due to the indirect relationship between chemical shift and 3D structure.

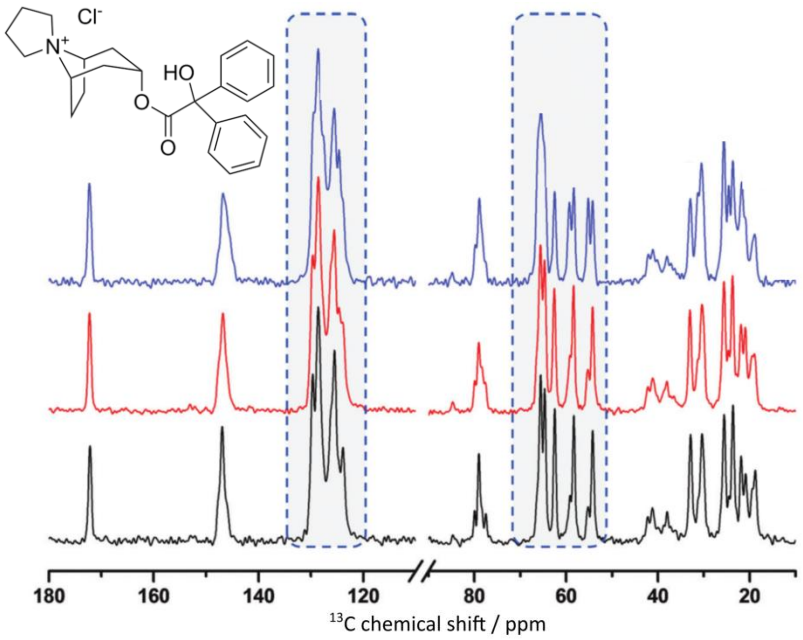

Figure 28 ^{13}C CP spectra of different samples of trospium chloride exhibiting different degrees of line broadening. Samples prepared from; a saturated solution of equimolar ethanol/water at 0 ˚C (blue), excess solvent (red), a saturated solution of ethanol (black). Boxed areas highlight regions with spectral differences and the molecular structure is shown top left. Reproduced from Urbanova *et al*., J. Pharm. Sci., 2013, 102, 4, 1235–1248. Copyright © 2013 Wiley Periodicals, Inc. and the American Pharmacists Association.

The prediction of NMR parameters based on a computational model allows spectral changes to be directly related to structure. As discussed in section 1.3.1, the systematic error on calculated shieldings is generally 2–3 % of the chemical shift range for each element. This error determines how well the predicted NMR parameters agree with experimental data. It has also been noted that different correlations are observed between experimental shifts and calculated shieldings for different carbon types.[7-9] Therefore, the systematic uncertainty for atoms of the same type must be lower than 2 %, for example in the case of chemically-related or symmetry-related sites, and it is likely to be affected by subtle considerations like convergence and the quality of the input structure. The input structures for CASTEP calculations are commonly structures refined from XRD data, so

geometry optimisations are ubiquitous to relax the atomic positions into an energetic minimum before NMR calculations are performed.[10,11]

It is currently unclear how geometry optimisations impact calculated NMR parameters. This chapter investigates the degree of uncertainty imparted to the calculated shieldings by geometry optimisation, and whether this uncertainty can be reduced enough to allow the prediction of SSNMR linewidths in an organic cocrystal containing static disorder. If such predictions are within the limit of NMR crystallography then the spectra of statically disordered systems could be predicted in order to directly relate line broadening to the degree of structural disorder. The test case for this investigation is a previously unreported cocrystal of 1:2 caffeine-citric acid hydrate (CAF-2CA), which contains statically disordered caffeine molecules within an ordered citric acid and water framework.

3.2 Properties and structure of caffeine

Caffeine, Figure 29, belongs to a class of three methylxanthines, along with theobromine and theophylline, which differ in the nitrogen substituents. Both caffeine and theophylline possess GRAS status[12] and the use of theobromine as an anticancer agent has been patented.[13] Caffeine is well known to stimulate the central nervous system as well as possessing mild diuretic properties, so it is used in the treatment of fatigue, orthostatic hypotension,[14] and headaches when combined with analgesics.[15] The mode of action is thought to be *via* antagonism of adenosine receptors in the plasma membrane of most cells, promoting neurotransmitter release.[16]

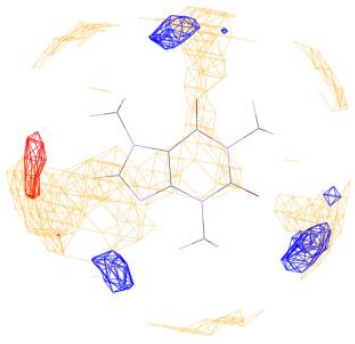

Figure 29 Molecular structure of caffeine showing the full interaction map with hydrogen bond acceptors in red, hydrogen bond donors in blue and hydrophobic interactions in orange.

Caffeine is an alkaloid with a pK_a of 14 at 20 ˚C and is readily absorbed after oral or parenteral administration, though it has limited aqueous solubility. The structure is characterised by a planar hydrophobic surface surrounded by hydrophilic (carboxyl) and hydrophobic (N-methyl) functionalities. Figure 29 shows the full interaction map[17] for caffeine calculated in Mercury,[18] with uncharged NH nitrogen atoms (blue), carbonyl oxygen atoms (red) and methyl carbon atoms (orange) as proxies for hydrogen bond acceptors, hydrogen bond donors and hydrophobic interactions respectively. In an aqueous environment aggregates tend to be formed by stacking of the hydrophobic rings,[19] similar to stacking observed for nucleotide bases.[20]

3.2.1 Caffeine in the solid state

The solid-state structure of caffeine is complex. Anhydrous caffeine exists as two polymorphs; a high temperature form (α) and the thermodynamically stable form (β).[21] However, it is challenging to produce phase-pure crystalline samples of either polymorph because crystals suitable for SCXRD are hard to grow and the composition is history-dependent.[22,23] Additionally, a hydrate is formed under humid conditions[24] that dehydrates under ambient conditions to the β form.[22]

The crystal structure of caffeine hydrate[25] was first solved in 1958, though the space group was corrected[22] in 1997 to $P2_1/c$ instead of $P2_1/a$. There is one caffeine molecule in the asymmetric unit and a water molecule with a site occupancy of 0.8. The water molecules form an extended chain through the crystal structure with hydrogen bonding to the unsubstituted nitrogen atom of the caffeine molecule, Figure 30(i). The dehydration mechanism is dependent on particle size and atmospheric vapour pressure, indicating that the system is a nonstoichiometric hydrate.[24]

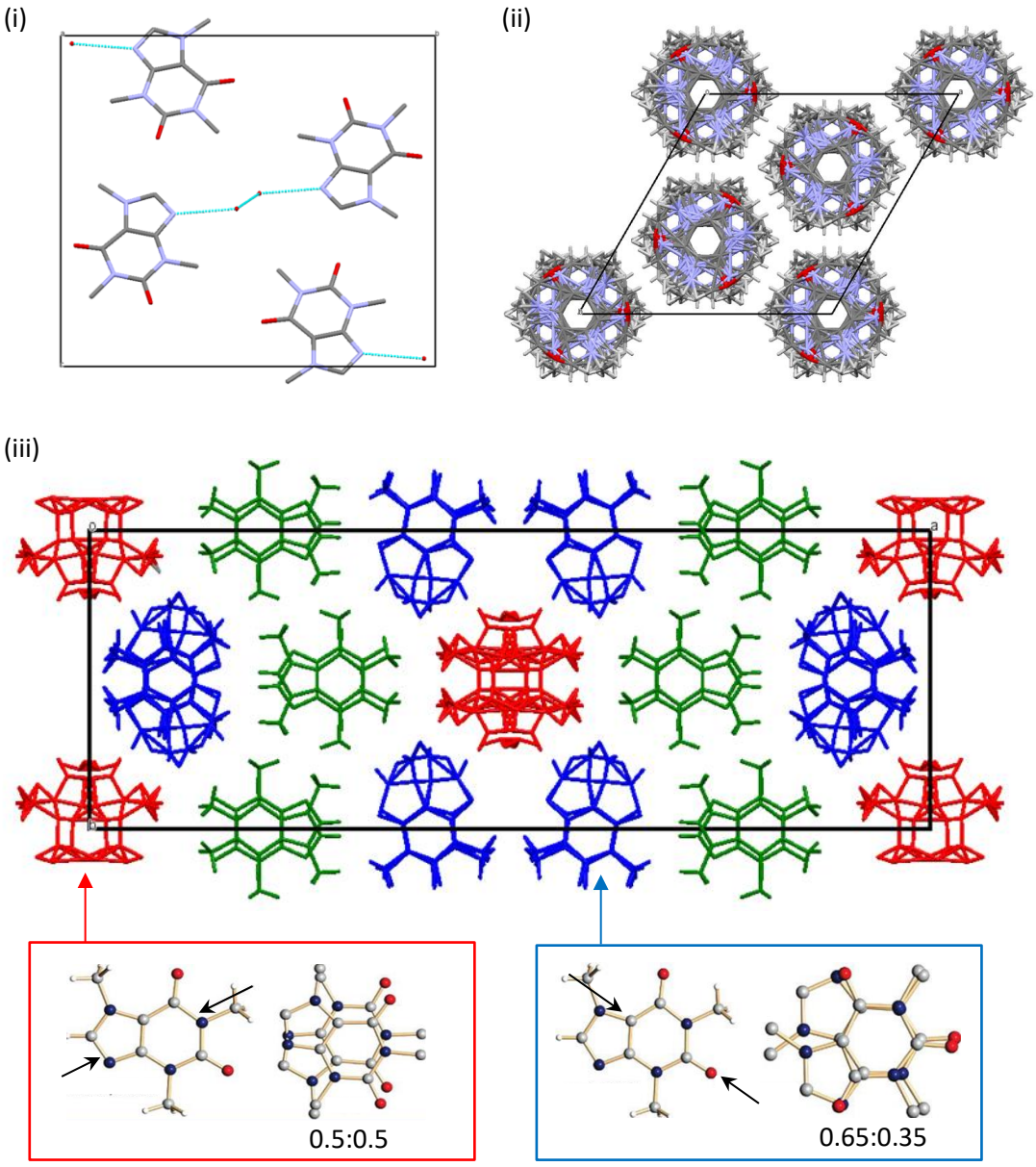

Figure 30 (i) Unit cell and hydrogen bonding in caffeine hydrate, viewed along the crystallographic *a* axis. **(ii)** Unit cell of α-caffeine viewed along the crystallographic *c* axis. **(iii)** Unit cell of β-caffeine viewed along the crystallographic *c* axis with the three unique caffeine sites coloured for clarity. The rotational axes relating the two positions of the disordered sites are shown along with site occupancies. Adapted with permission from Enright *et al*., Cryst. Growth Des., 2007, 7, 8, 1406–1410. Copyright © 2007 American Chemical Society.

At 150 ˚C α-caffeine is produced,[21] which has a highly disordered and symmetric structure, Figure 30(ii). The structure was initially solved from PXRD data in the R3c space group[26] but was later corrected to R$\bar{3}$c following analysis of SCXRD data.[23] There is one sixth of a caffeine molecule in the asymmetric unit and six molecules overlap in the *ab* plane at each

crystallographically unique site. The disorder has been characterised as being dynamic, particularly above the glass transition temperature, T_g = 250 K, in the so-called "rotator phase".[27-29] Monte Carlo simulations show that the caffeine molecules undergo free rotation in the *ab* plane with some additional out-of-plane dynamics at high temperature.[30] The dynamics persist as α-caffeine is cooled to a metastable glassy state at room temperature, before slow conversion to the thermodynamically stable β form.

Anhydrous β-caffeine also contains complex disorder, solved by combining high field SSNMR data with SCXRD analysis[23] to improve a structure previously solved from PXRD data.[31] The space group is identified as *C*2/*c* and the caffeine molecules exhibit π stacking in columns along the *c* axis, Figure 30(iii). There are 2.5 caffeine molecules in the asymmetric unit over three crystallographically distinct sites; one fully ordered and two exhibiting orientational disorder over two positions. Figure 30(iii) shows how the two positions are related in each disordered site. The nature of the disorder in β-caffeine has been investigated with dielectric relaxation spectroscopy[27,28] and Raman spectroscopy,[29] which suggest the presence of a dynamic component at room temperature. No evidence of dynamic disorder was found at low temperatures.[29] ^{1}H–^{14}N nuclear quadrupole resonance (NQR) studies have also been carried out. Briefly, NQR is analogous to NMR except that it is based upon the interaction of the nuclear quadrupole moment with the EFG. Just as NMR chemical shifts are sensitive to the local environment, so too are NQR frequencies. In β-caffeine discrepancies were observed between experimental and predicted NQR frequencies, C_Q and η values, which were ascribed to the presence of dynamic averaging.[12] It is possible that the dynamic component observed in these studies is due to small quantities of α-caffeine impurity, since α–β conversion is very slow.

The ^{13}C SSNMR linewidths of β-caffeine reflect the presence of disorder,[23] Figure 31(i), while caffeine hydrate has sharp ^{13}C lines in keeping with the ordered crystalline structure, Figure 31(ii). Additional resonances are observed for β-caffeine in the methyl region, 20–40 ppm, supporting the presence of Z'>1. The nature of the disorder is unclear from a single temperature ^{13}C spectrum so ^{2}H investigations have been carried out on methyl-deuterated β-caffeine,[32] revealing a thermally activated CD_3 reorientation through ^{2}H T_1 relaxation times. No molecular motion was observed *via* stimulated-echo spectroscopy below 370 K on the ms–s timescale, in contrast to the previous results from dielectric and Raman spectroscopy. ^{2}H T_1 relaxation times are shorter for α-caffeine, indicative of increased molecular mobility, and samples of β-caffeine that had been heated and subsequently

cooled were found to have shorter relaxation times compared to the original β sample. This supports a slow α–β phase change.

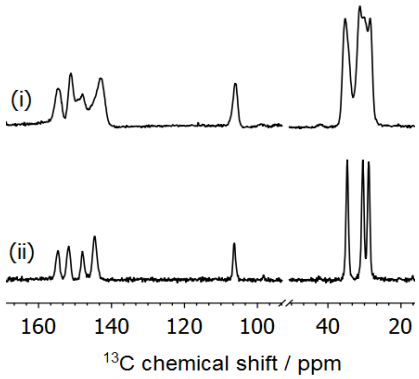

Figure 31 ^{13}C CP/MAS spectra of **(i)** anhydrous β-caffeine and **(ii)** caffeine hydrate, both acquired at 8 kHz MAS on the 5 mm probe with a contact time of 1 ms. Spectrum **(i)** was acquired with a 30 s recycle delay over 1000 transients, and spectrum **(ii)** was acquired with a 5 s recycle delay over 144 transients.

Several computational studies have explored why anhydrous caffeine does not form an ordered structure. The dominant packing mode is stacking of parallel rings, but the lateral interactions between neighbouring molecules are all energetically similar.[33] Kinetic factors in crystallisation also contribute, resulting in non-equal occupancies of some of the disordered sites instead of the theoretically predicted equal-occupancy structure.[34] A symmetry-adapted ensemble model[35] of β-caffeine reveals many unit cell variants of almost equal energy and the ordered crystal energy landscape shows several low energy structures. This is in contrast to isocaffeine, whose crystal energy landscape shows a single low energy structure in good agreement with experimental data. The difference is ascribed to the less symmetric structure imparted to isocaffeine by the different position of the C12 methyl group (labelling from Figure 29).

3.2.2 Caffeine cocrystals

Caffeine has a high propensity to form cocrystals[36-41] and so it is often used as a model compound to demonstrate the potential for improved mechanical properties following cocrystallisation,[42,43] as well as cocrystal reactivity.[44-46] It forms hydrogen bonds through the unsubstituted nitrogen atom, two carboxyl groups, and the NH. There are more than 75 caffeine-containing cocrystals in the CSD, of which 19 are hydrates. It is noted that disorder

of the caffeine molecule is found in only three of these entries, suggesting that the synthons formed on cocrystallisation tend to be extensive. The three cocrystals with disordered caffeine molecules are bis(caffeine) *p*-coumeric acid hydrate (2CAF-PCA),[47] caffeine 6-hydroxy-2-naphthoic acid[36] (CAF-HNA), and bis(caffeine) 4-hydroxybenzoic acid (2CAF-HBA).[38] In CAF-HNA an unusual carboxylic acid dimer is present; it is more usual for carboxylic acids to form hetero-synthons in cocrystals. The other two systems contain three-component assemblies where one of the two crystallographically unique caffeine molecules is disordered over two positions in non-equal ratios. In all three cases the two positions are related by the same pseudo-two-fold rotation axis and the caffeine imine group is not involved in hydrogen bonding.

Caffeine citrate is a salt marketed as a solution under the trade name Cafcit® for the treatment of migraines.[48] It is much more soluble in water than pure caffeine, due to the high solubility of citric acid. In the solid state, caffeine co-crystallises with citric acid to give an ordered 1:1 cocrystal (CAF-CA) crystallising in space group $P\bar{1}$.[37] OH···N hydrogen bonds form between the citric acid and the caffeine imine nitrogen atom with $d_{O\cdots N}$ = 2.6842(9) Å, as well as OH···O bonds to the caffeine ketone groups with $d_{O\cdots O}$ = 2.628(1) Å and $d_{O\cdots O}$ = 2.874(1) Å. Interestingly, this cocrystal can only be formed by neat grinding of caffeine hydrate with citric acid or by LAG with water. Neat grinding of anhydrous caffeine with citric acid yields only a mixture of starting components. This suggests that the water is crucial to facilitate synthon formation between caffeine and citric acid molecules, despite not being present in the final crystal structure. Attempts at forming a hydrated 1:1 cocrystal were previously unsuccessful. However, in this work the hydrate CAF-2CA was successfully synthesised by changing the stoichiometries of the starting components.[49]

3.2.3 Crystal structure of caffeine-citric acid hydrate

The crystal structure of CAF-2CA was solved from SCXRD data by H. Mason and H. Sparkes of Durham University in the monoclinic space group *C*2/*c*. The asymmetric unit contains a water molecule, a citric acid molecule, and half a caffeine molecule, which is refined with disorder over two positions of equal occupancy over a two-fold rotation axis, Figure 32(i). It is noted that some crystals refined better in the *Cc* space group, *i.e*. without a centre of inversion, because the site occupancies of the caffeine atoms refined to 0.35:0.75 rather than 0.5:0.5. These crystals came from different batches but the relationship between crystallisation conditions and caffeine site occupancies is not clear without more detailed crystallisation experiments, which is beyond the scope of this work. There are four caffeine

molecules in the unit cell and they form symmetric hydrogen bonds with neighbouring citric acid molecules, Figure 32(ii), with $d_{O10\cdots O7}$ = 2.550(8) Å and $d_{N2\cdots O7}$ = 2.735(10) Å. The citric acid and water framework is bound together in a six-membered ring conformation, Figure 32(iii). No changes were observed between the structures refined from SCXRD data acquired at 120 K and at 295 K, other than the expected non-directional increase in the size of the ADPs, suggesting that the disorder is static.

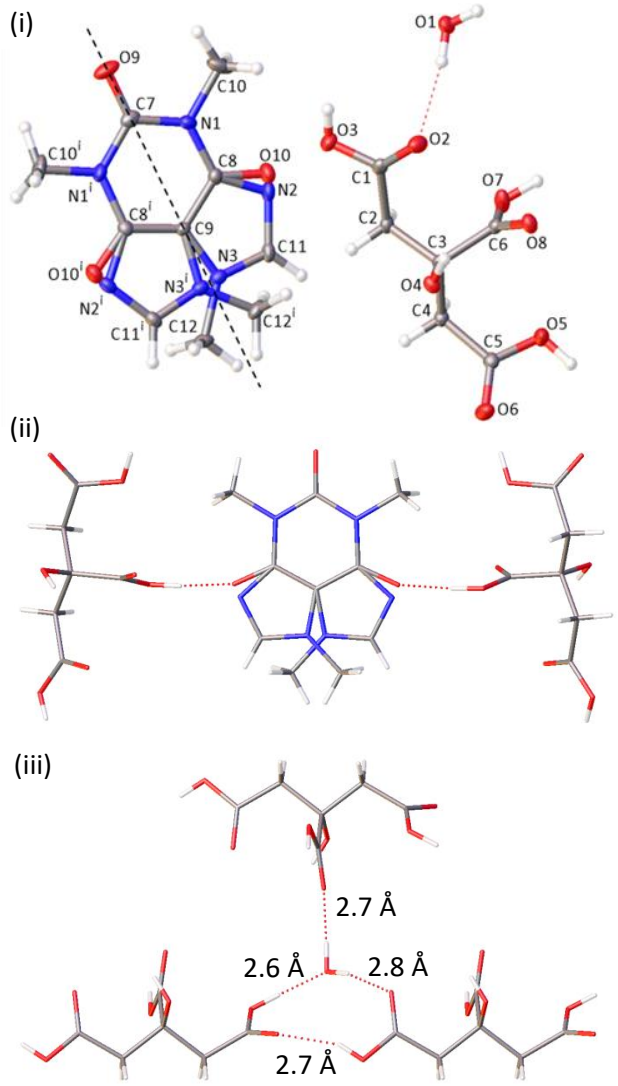

Figure 32 (i) Schematic of CAF-2CA solved in the *C*2/*c* space group with ellipsoids viewed at the 50 % probability level. The dashed line denotes the two-fold rotation axis relating the disordered caffeine orientations. **(ii)** Symmetric hydrogen bonding between caffeine and citric acid molecules. **(iii)** Hydrogen bonding between citric acid and water molecules viewed along the crystallographic *a* axis with donor-acceptor heavy atom distances labelled.

3.3 Preliminary structure investigations

The structure refined from SCXRD data provides only an average picture of the caffeine disorder in CAF-2CA. An NMR crystallography investigation is described in the following sections to gain a deeper insight into the disorder. This investigation leads to the question of whether it is possible to quantify how the disorder contributes to ^{13}C linewidths and whether the current NMR crystallography protocols are capable of predicting such subtle effects.

3.3.1 Experimental details

CAF-2CA was produced by neat grinding a 1:2 molar ratio of anhydrous β caffeine (CAF) and citric acid hydrate (CIT) in a mortar and pestle. This grinding product was used for all SSNMR experiments. LAG with a few drops of water added prior to grinding also yields CAF-2CA, but grinding at an ambient temperature of *ca*. 30 ˚C yields a mixture of the ordered anhydrous cocrystal,[37] CAF-CA, and citric acid. Single crystals suitable for SCXRD analysis were grown by slow cooling of a solution of the neat grinding product dissolved in hot nitromethane. This solution was cooled to room temperature over 2 hours, the resulting precipitate was filtered off (found to be CAF-CA) and the filtrate was left in the fridge for 3 weeks after which clear block crystals of CAF-2CA were obtained. Attempts to crystallise bulk quantities failed due to the propensity of CAF-CA to precipitate out.

Five other caffeine containing cocrystals were synthesised by recrystallisation; 2CAF-HBA, CAF-2HBA,[38] 2CAF-PCA, CAF-PCA,[47] and CAF-HNA.[36] 2CAF-HBA was synthesised by heating a 2:1 molar ratio of CAF and HBA in acetonitrile until the solids dissolved, transferring the hot solution to a sealed vial and leaving to cool/evaporate for one month, during which time colourless needles were produced (0.257 g). SCXRD carried out by D. Yufit, Durham University Crystallography Service, confirmed the product matched the previously reported structure, CSD label MOZCUA.[38] CAF-2HBA was synthesised by the same method, but using a 1:2 molar ratio of CAF and HBA, producing 0.293 g of product. 2CAF-PCA and CAF-PCA were also synthesised by the same method, with 2:1 and 1:1 molar ratios of CAF and PCA respectively and pale yellow plates were produced, 0.200 g and 0.230 g respectively. CAF-HNA was synthesised by dissolving a 1:1 molar ratio CAF and HNA in hot acetonitrile, cooling slowly to room temperature, filtering off the solid and leaving the filtrate to slowly evaporate in a sealed vial. Fine colourless needles were produced. All samples were ground in a mortar and pestle prior to SSNMR experiments.

All NMR experiments were acquired with a 5 mm Varian HX MAS probe. The CAF-2CA grinding product is stable on storage in air at ambient temperature over 9 months, but degradation is observed following several hours under MAS conditions *via* the appearance of new peaks. No noticeable line broadening is observed of the CAF-2CA peaks, indicating that the linewidths are unaffected by low levels of degradation. No degradation peaks are visible in any of the ^{13}C spectra reported here, which were acquired over short periods (<2 hours).

A transition state search was carried out in CASTEP version 6.0 with the "reactant" and "product" structures formed by splitting the disordered CAF-2CA unit cell into two ordered structures. A suggested "intermediate" structure was also imported to direct the calculation, created by manually rotating the caffeine molecules of the "product" structure out of plane about the molecules centre of mass in the Avogadro molecular editor.[50]

3.3.2 Characterising disorder

PXRD of the CAF-2CA grinding product (obtained by G. Oswald at Durham University) shows clear differences compared to a physical mixture of the starting components, Figure 33. In particular, the reflections at 26˚ and 27˚ are not visible in the cocrystal spectrum, indicating that structural changes have occurred on grinding. The CAF-2CA peaks are sharp so the product is crystalline and there is no evidence of an amorphous component in the baseline.

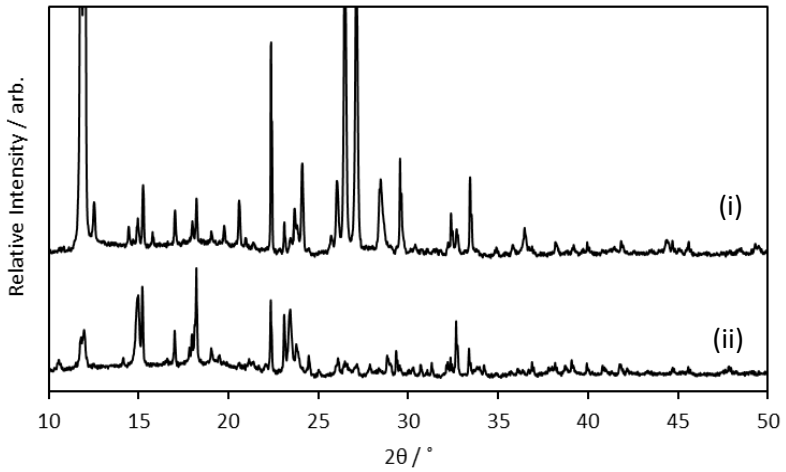

Figure 33 PXRD patterns of **(i)** a physical 1:2 mixture of caffeine and citric acid hydrate, **(ii)** CAF-2CA grinding product.

Evidence for the static nature of the caffeine molecule disorder is provided by the failure to computationally locate a theoretical transition state between the two disordered caffeine orientations. The void in which the caffeine molecules reside is too small to allow 90˚ out-of-plane rotation about the rotation axis shown previously without the surrounding citric acid/water framework being perturbed, as shown in Figure 34. There is a large energy penalty (1795 kJ mol^{-1}) associated with this perturbation and so dynamic disorder is implausible.

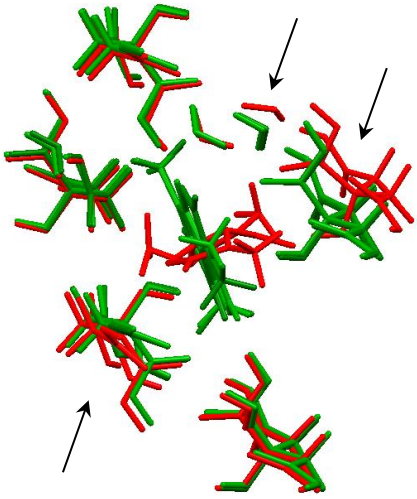

Figure 34 Overlay of the two disordered orientations of CAF-2CA (green) and a predicted transition state between the two (red), where the caffeine molecule is rotated out of plane by 90˚. Arrows highlight perturbations of the surrounding citric acid/water framework.

Static disorder can be random or there may be local domains in which the caffeine orientations are correlated. Such a phenomenon has been previously observed in pharmaceutical systems such as aspirin[51] and felodipine,[52] termed intergrowth polymorphism. Correlated disorder would be expected to give rise to diffuse scattering, which is not observed in the SCXRD data of CAF-2CA, though higher quality synchrotron data would be more conclusive. Transmission electron microscopy (TEM) images also show no evidence of domain boundaries on the µm scale, but the CAF-2CA crystals degrade when nm scale imaging is attempted, even under cryogenic cooling.[*] Overall, the disorder is

[*] Electron diffraction patterns obtained from TEM experiments were not investigated further due to the instability of the sample under the electron beam.

likely to be random in CAF-2CA but the presence of small non-diffracting domains cannot be ruled out.

The ^{13}C CP/MAS SSNMR spectrum of the CAF-2CA grinding product is distinct from that of the starting component β-caffeine and also the previously reported 1:1 cocrystal CAF-CA, Figure 35. A small quantity of excess citric acid is present, the peak marked 'x' on Figure 35(i), but no excess caffeine is visible and the peaks do not overlap with those of the cocrystal so no further purification was carried out. The chemical shifts of the excess citric acid peaks correspond to anhydrous citric acid[53] rather than the citric acid hydrate starting material. This suggests some water is lost during synthesis, possibly due to frictional heating during grinding. The disorder of the caffeine site renders C8/C8i and C10/C10i crystallographically indistinguishable but they are chemically distinct and so are observed in the ^{13}C spectrum as separate peaks. No changes are observed in the sideband manifold on heating the sample to 70 ˚C at a low spinning rate of 3 kHz, so the CSA is not temperature-dependent suggesting that there are no dynamics on the order of the sideband width (10^{-3} s). Additionally, no peaks are observed in a spectrum acquired with a short 1–5 s recycle delay, highlighting the lack of any fast dynamics in the system. ^{13}C T_1 relaxation times were not studied in detail due to the long experiment times required and the potential for sample degradation under MAS conditions.

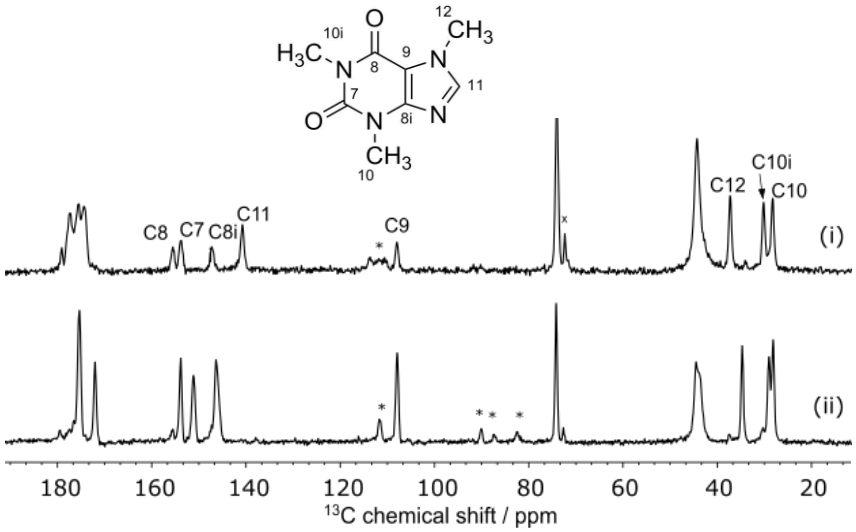

Figure 35 ^{13}C CP/MAS spectra of **(i)** CAF-2CA and **(ii)** CAF-CA acquired at 8 kHz MAS with a recycle delay of 15 s and contact times of 1 ms and 5 ms over 280 and 624 transients respectively. Spinning sidebands are marked with asterisks and excess citric acid is marked with an 'x'.

The static disorder manifests as line broadening in CAF-2CA, evidenced by a slightly broader average caffeine linewidth compared to that of ordered CAF-CA (98 Hz vs 86 Hz in Figure 35). The broadening is not as extensive as β-caffeine, Figure 31(i). As discussed previously, there are many contributions to SSNMR linewidths that might differ between CAF-CA and CAF-2CA including the presence of disorder and the ABMS. The ABMS line broadening factor,[5] $|4\pi\Delta\chi|$ introduced in section 1.2.6 and determined from the magnetic susceptibility calculated in CASTEP is 1.1 ppm for CAF-2CA compared to 2.0 ppm for CAF-CA. If the ABMS is the dominant factor in broadening the ^{13}C linewidths of CAF-2CA then it would be expected that the ABMS line broadening factor would be larger for CAF-2CA than CAF-CA, which is not the case. The ABMS line broadening factors of both cocrystals are smaller than those of pure caffeine, which are calculated to be 2.8 ppm for caffeine hydrate, 3.5 ppm for α-caffeine and 3.1 ppm for β-caffeine. Therefore, it is likely that the disorder contribution is significant in CAF-2CA though it is likely to be small, so the computational prediction of the ^{13}C linewidths is a challenging test of NMR crystallography protocols.

3.4 Exploring the limits of geometry optimisations

Geometry optimisation calculations are always carried out before NMR parameters are predicted. These geometry optimisations involve a specific set of convergence criteria, beyond the usual E_{cut} and *k*-point spacing, which must introduce a degree of uncertainty that is carried through to NMR calculations. This uncertainty has not previously been investigated. CAF-2CA is an ideal test case because the disordered structure can be split into 16 simulated-disorder models (theoretical ordered unit cells with all possible combinations of caffeine orientation), which form 8 symmetry-related pairs. These should be identical in energy, atom position and NMR parameters at the limit of perfect geometry optimisation. Figure 36 shows an example of one of these simulated disorder models, labelled as structure “1111” to denote the caffeine orientations.

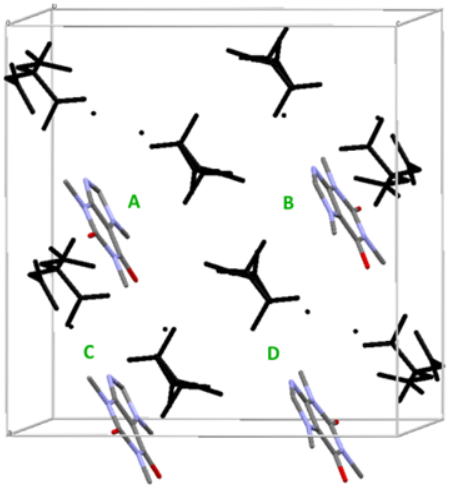

Figure 36 Schematic of the theoretical ordered unit cell of CAF-2CA (1111) with the citric acid and water network in black and hydrogen atoms omitted for clarity. Caffeine molecules are labelled such that the ABCD order reflects the order of the 1111 labels, *e.g.* flipping molecule C results in the 1121 structure.

3.4.1 Computational approach

The disordered structure of CAF-2CA, solved from SCXRD data, was split into two ordered unit cells; 1111 and 2222. The other 14 simulated-disorder models were produced by manually flipping the caffeine molecules 180˚ about their centre of mass, starting from the 1111 model, using the Avogadro molecular editor.[50] Configurational ensemble methods using supercells could be used to more accurately model the disordered nature of the material, as done for β-caffeine,[35] but this would be demanding for DFT calculations. Due to the manual creation of the simulated-disorder models a crude geometry optimisation was carried out in CASTEP version 6.0 using a low E_{cut}, 400 eV, to provide a sensible starting point for subsequent testing of tighter geometry optimisations.

DFT geometry optimisations were carried out to relax all 288 atomic positions with the centre of mass fixed. In total eight different optimisation methods were applied, see Table 4, where $dE_{electronic}$ is the energy per atom threshold for convergence of the electron self-consistent field minimisation and $dE_{geometry}$ is the corresponding threshold for convergence of the energy between successive geometry optimisation steps. The latter parameter cannot be meaningfully tightened beyond the former. $|F|_{max}$ is the convergence parameter for the atomic forces between successive geometry optimisation steps and the convergence window sets the number of successive iterations within which all tolerances must be satisfied for the calculation to be considered converged. The Tkatchenko-Scheffler

(TS) dispersion correction scheme[54] was applied in all "[–D]" optimisations and the unit cell parameters were kept fixed to the parameters refined from SCXRD data in all cases except [3-DC].

Table 4 Conditions used in different geometry optimisations of the simulated-disorder models of CAF-2CA. The final column contains the mean energy difference between the symmetry-related pairs.

Name[a]	E_{cut} / eV	d$E_{electronic}$ / 10^{-5} eV	d$E_{geometry}$ / 10^{-5} eV	$\|F\|_{max}$ / 10^{-2} eV	Mean ΔE between pairs / eV
[1]	400	1	2	5	0.036
[2]	700	0.1	0.2	1	0.009
[2-D]	700	0.1	0.2	1	0.012
[3]	700	0.001	0.002	0.01[b]	0.005
[3-D]	700	0.001	0.002	0.01[b]	0.004
[3-DO]	700	0.001	0.002	0.01[b]	0.002
[3-DW]	700	0.001	0.002	0.01[b]	0.004
[3-DC]	700	0.001	0.002	0.01[b]	0.005

[a] Codes used: D – dispersion correction applied, O – OTFG pseudopotentials used, W – convergence window increased from 2 to 10, C – unit cell parameters allowed to vary. [b] $|F|_{max}$ converged to within 0.1 x 10^{-2} eV per atom but did not reach the 0.01 x 10^{-2} eV per atom tolerance.

3.4.2 Effect on energy and atomic position

The final energies of the 16 simulated-disorder models do depend on the geometry optimisation, Figure 37. The energies are random when the default tolerances are used, [1], but tightening the energy tolerances by one order of magnitude, [2] and [2-D], results in the 1111/2222 pair (yellow in Figure 37) converging to approximately 0.03 eV higher energy than the other models. An additional matching pair 1122/2211 (green in Figure 37) emerge as the conditions are tightened further, [3], [3-D] and [3-DO]. The tighter conditions result in all symmetry-related pairs converging to more similar values, final column of Table 4.

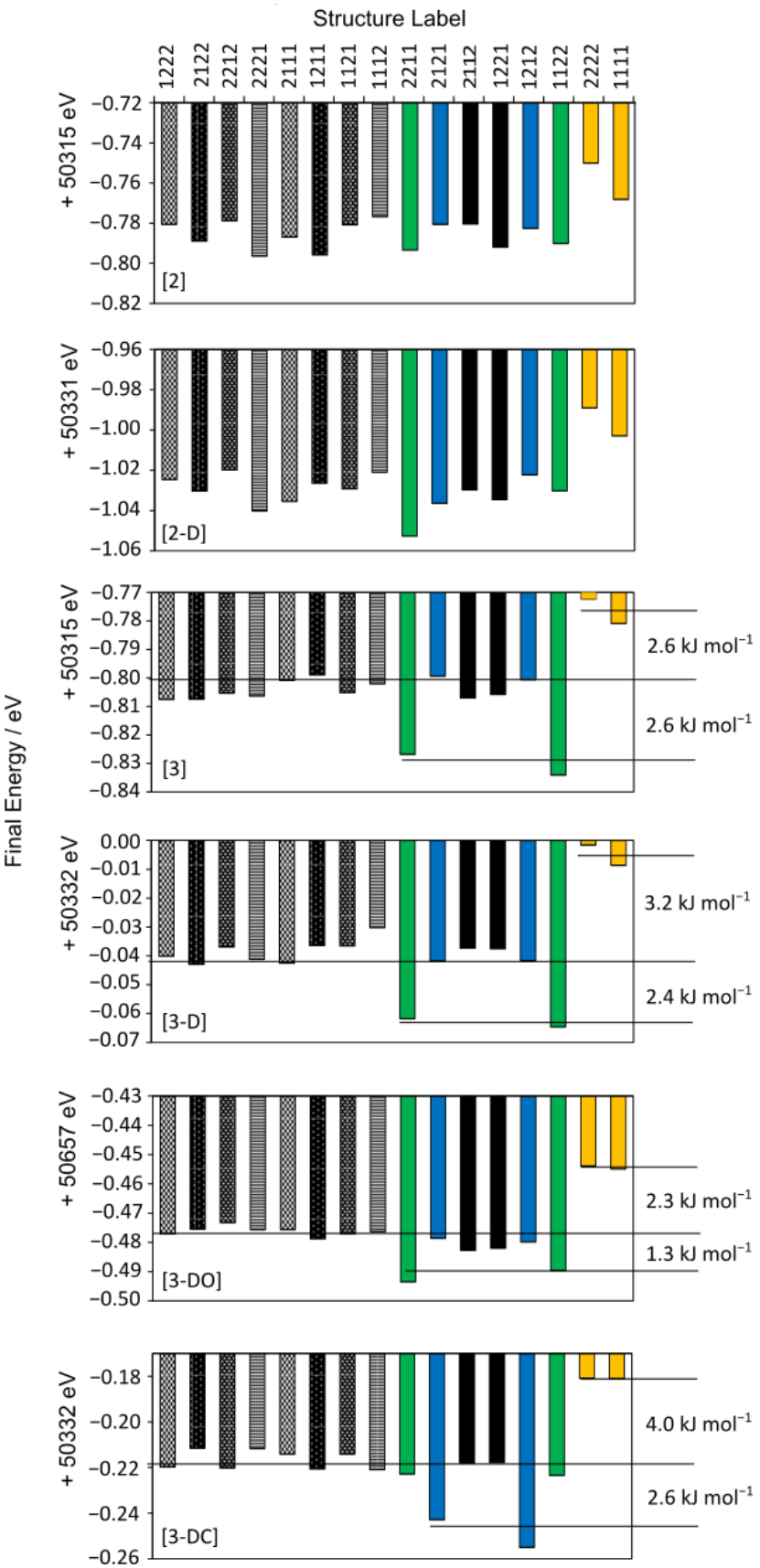

Figure 37 Final energies per unit cell of each simulated-disorder model following different geometry optimisation conditions. Symmetry related pairs discussed in the main text are coloured for clarity. Energy differences between sets of structures are given on the right.

Including dispersion correction systematically lowers the energies but does not change the relative energetic ordering of symmetry-related pairs. A combination of dispersion correction and OTFG pseudopotentials results in the relative energy differences between different sets of symmetry-related pairs being smaller, Table 4, reducing to 2 meV on average. Increasing the convergence window so the convergence criteria must be satisfied over 10 optimisation steps instead of 2 did not reduce the energy differences between symmetry-related pairs. Instead none of the calculations converged with respect to $dE_{geometry}$, indicative of over-tight convergence criteria. The results of this run are not discussed further.

A slightly different pattern is observed when the unit cell is allowed to vary, [3-DC], with the 1212/2121 pair converging to a lower energy. This correlates with a larger percentage change in the α and γ unit cell angles away from 90˚ for the 1212/2121 pair compared to the other pairs, blue in Figure 38. The *c* axis reduced by a moderate 2 % compared to the values fitted from SCXRD data in all cases. However, this result is less physically significant than the optimisations with the unit cell parameters constrained because the true unit cell of the CAF-2CA system is an average over all disorder combinations, measured best in the SCXRD experiment.

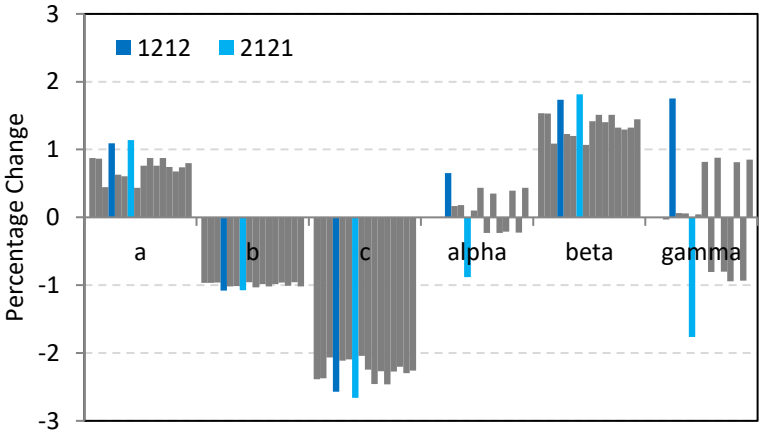

Figure 38 Percentage change of the unit cell parameters of the 16 simulated-disorder models following geometry optimisation with variable unit cell parameters (relative to unoptimised value). The 1212 symmetry-related pair is highlighted in blue.

The energy differences between symmetry-related pairs are similar to those between polymorphs at 1–4 kJ mol^{-1} per unit cell.[55] Although the absolute energy depends on the DFT protocol and individual parameters used, the difference between symmetry-related pairs converges to 2 meV per unit cell on average giving confidence to the comparison between lattice energies of chemically similar disordered structures. It might be expected that CAF-2CA should adopt the lowest energy structures (1122/2211) but in the tightest optimisations, [3-D] and [3-DO], these are only 2.4 kJ mol^{-1} and 1.3 kJ mol^{-1} per unit cell lower in energy than the other simulated-disorder structures and k_BT is 2.5 kJ mol^{-1} at room temperature. There is an entropic bias towards a disordered structure that contains all 16 structures in equal occupancy; crudely, the Boltzmann ratio for disordered:1122/2211 is about three at room temperature. Similar energetic arguments have been made to rationalise the presence of disorder in some, but not all, solvates of droperidol[56] and the presence of disorder in β-caffeine but not isocaffeine.[34]

The degree of convergence between symmetry-related pairs can also be probed by examining the RMSD between heavy atom positions.* As expected these show the same trends as Figure 37, with the smallest average RMSD between symmetry-related pairs coming from the geometry optimisations with the tightest convergence criteria, Figure 39(i). Note the pseudo-rotation axis is not precisely retained following geometry optimisation, Figure 39(ii). This is because the symmetry of the system is reduced to P1 in the simulated-disorder models, except the 1111/2222 pair whose symmetry is *Cc*. The atomic deviations are within the boundaries of the thermal ellipsoids shown in Figure 32(i), so the optimised structures are indistinguishable from the original SCXRD structure within the experimental uncertainties.

* All but one hydrogen atom reside in methyl groups so RMSDs between the hydrogen atom positions are not meaningful.

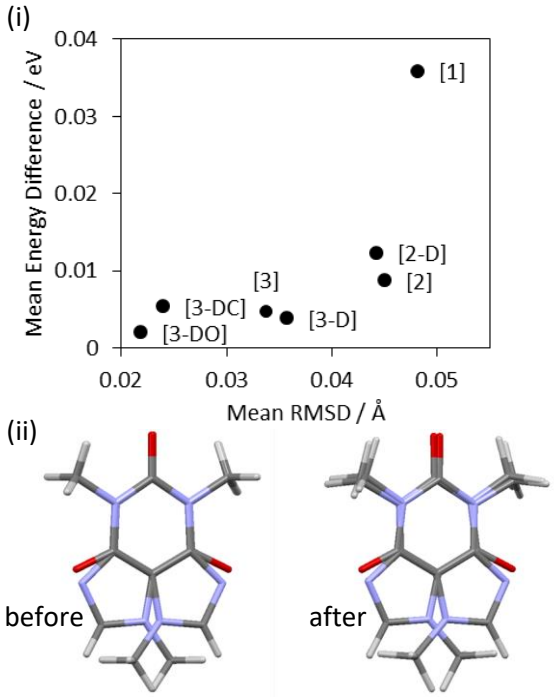

Figure 39 (i) Correlation between energy differences and heavy-atom RMSD between symmetry-related pairs following different geometry optimisations. **(ii)** Overlays of the two orientations of the disordered caffeine molecules before and after geometry optimisation. Atoms on the six-membered ring no longer perfectly overlay after optimisation.

3.4.3 Effect on NMR parameters

The NMR parameters of the symmetry-related pairs should be identical at the limit of perfect geometry optimisation. Figure 40 shows that the average difference in ^{13}C shielding between related caffeine carbon sites in the symmetry-related pairs is non-negligible, but reduces as the geometry optimisation conditions are tightened. Specifically, optimisation using the default conditions, [1], results in a large spread of ^{13}C shielding differences, 0.05–0.35 ppm, while tighter optimisation, [3-DO], results in smaller more consistent differences of 0.02–0.1 ppm. No other pattern is observed in Figure 40(i); no symmetry-related pair shows consistently larger or smaller average ^{13}C shielding differences than the others. Therefore, Figure 40(ii) simply plots the average over all related sites in all pairs. Even the default optimisation, [1], results in average ^{13}C shielding differences that are an order of magnitude smaller than the typical uncertainty on DFT calculated shieldings relative to experimental shifts. This is reduced further to roughly 0.1 ppm with the tightest geometry optimisation, [3-DO]. In general the difference is always smaller when only the caffeine carbon sites are considered. This likely reflects the rigid nature of the caffeine molecule compared to the citric acid, thus reducing the complexity of the optimisation surface.

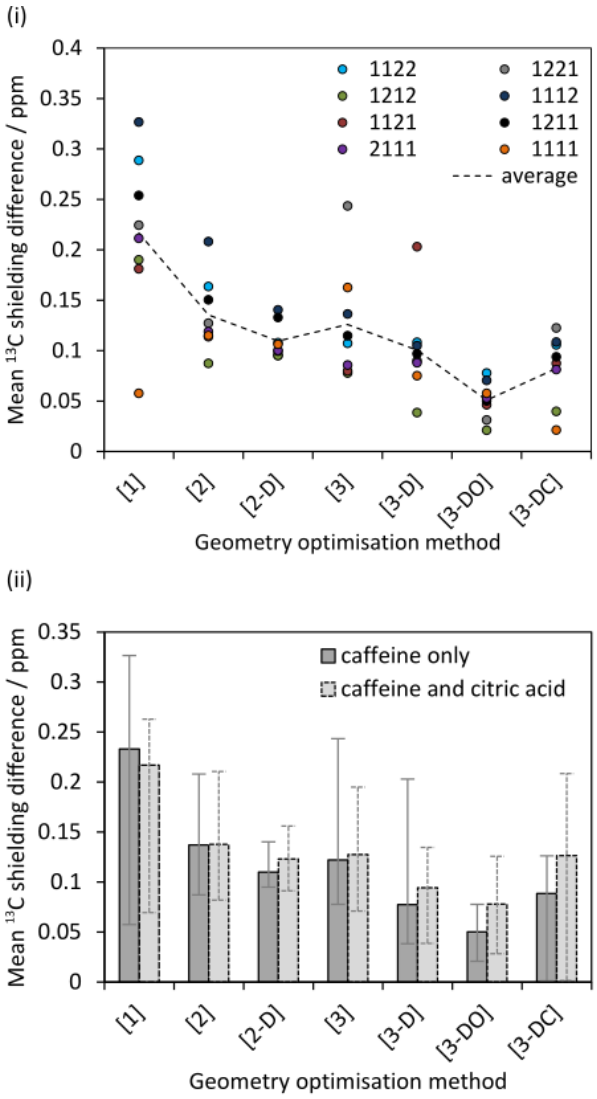

Figure 40 (i) Mean difference in calculated isotropic ^{13}C shielding between related caffeine carbon sites of symmetry-related pairs following different geometry optimisations. **(ii)** Average of these differences over all symmetry-related pairs. The "caffeine only" dataset corresponds to the dotted line in **(i)** and the error bars indicate the maximum and minimum value in **(i)**.

Optimisation with dispersion correction consistently lowers the mean ^{13}C shielding difference, as observed for the energy difference and RMSDs. The pattern of Figure 40(ii) generally reflects that of Figure 39(i) with the best convergence of symmetry-related pairs observed for optimisations [3-D] and [3-DO]. Fewer iterations are necessary for the calculations with dispersion correction, Table 5, which may reflect a smoother energy landscape. This is supported by the increase in "complex landscape" error warnings when dispersion correction is included. It is clear that very subtle differences in chemical shift can be predicted between chemically-related sites. This is in keeping with a recent investigation

into the role of water in sildenafil citrate, where differences of <2 ppm were found to be computationally predictable.[57]

Table 5 Additional computational details for the geometry optimisation calculations, averaged over all 16 calculations for each method.

Geometry Optimisation Method	Average CPU time / hours	Average number of iterations	Complex landscape error	Bisection search failed	Repeat reset of inverse Hessian
[1]	8	49	3	0	0
[2]	48	113	7	0	1
[2-D]	24	116	11	4	5
[3]	106	373	0	0	0
[3-D]	44	126	10	4	2
[3-DO]	61	142	13	7	5
[3-DC]	473	694	0	0	0

The conditions of the geometry optimisation do not impact the agreement of the calculated ^{13}C shielding parameters with experimental ^{13}C chemical shifts, Figure 41. The calculated ^{13}C shieldings were averaged for each carbon site and referenced using $\delta_{iso} = m\sigma_{iso} + c$, with m and c calculated from linear regression. Due to the overlap of citric acid resonances with trace amounts of excess citric acid, only the caffeine sites are considered. Since the effects of imperfect geometry optimisation are of the order of 0.1–0.3 ppm depending on the convergence criteria, it is unsurprising that no correlation is observed between the protocol and the quality of agreement with experimental data. The inclusion of dispersion correction shows no improvement, in keeping with previous reports that show only a slight improvement in ^{15}N shift with dispersion correction if the unit cell parameters are fixed.[58] The use of OTFG pseudopotentials in [3-DO] shows the largest ^{13}C shift RMSD compared to the optimisations using USP pseudopotentials. This is consistent with the systematic limitations of DFT being the major source of disagreement with experiment,[8,59-61] rather than subtle convergence considerations in geometry optimisation. Therefore, the limit of NMR crystallography protocols involving atom assignment is unlikely to be affected by the geometry optimisation conditions.

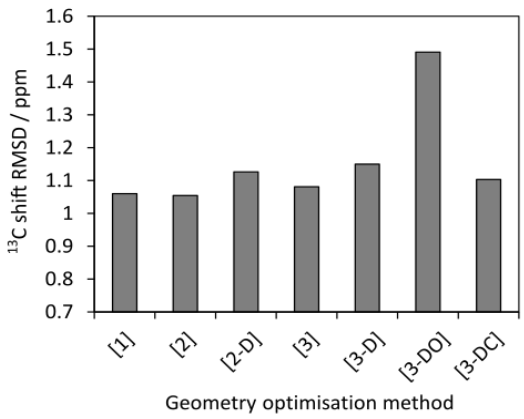

Figure 41 RMSDs between experimental ^{13}C shifts and average calculated ^{13}C shieldings referenced as described in the text.

3.5 Simulating linewidths

Since the systematic uncertainty on the predicted shielding of chemically-related sites is found to be small, it is feasible to simulate the ^{13}C linewidths of a disordered system in a quantitative manner. However, the systematic uncertainty is of the order expected for the disorder contribution. The following section discusses whether the quantitative prediction of disorder effects on NMR linewidths is beyond the current limit of NMR crystallography.

3.5.1 Methodology

The linewidth simulation for disordered systems uses pNMRsim[62] to calculate synthetic ^{13}C spectra for each site using the isotropic chemical shieldings read in from the .magres files of simulated-disorder models, as output by CASTEP, using the MagresPython library.[63] This is not a full calculation of the density matrix, but rather a simple addition of the Gaussian and Lorentzian contributions to the linewidth, discussed below. The sum of these synthetic spectra gives a single spectrum per carbon site, Figure 42, the linewidths of which can be deconvoluted in Gsim.[64]

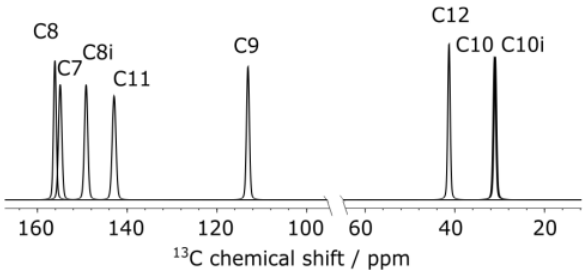

Figure 42 Simulated ^{13}C spectra for each caffeine carbon site in CAF-2CA summed over the spectra of all 16 simulated-disorder models optimised using the [3-DO] method. Spectra referenced using $\delta_{iso} = \sigma_{ref} - \sigma_{iso}$ with $\sigma_{ref} = 174$ ppm.

The input linewidth for the synthetic spectra contains a homogeneous contribution with a Lorentzian shape, which was calculated from T_2' experiments,[4,65] Table 6. These experiments used a ^{13}C τ_{180} of 10 μs and the echo delay τ was varied from 1–16 ms over 7 increments of 1.25 ms (equal to 10 τ_r where τ_r is the rotor period). 48 transients were acquired for each increment to minimise sample degradation under MAS conditions.

Table 6 Breakdown of the linewidth contributions to CAF-2CA with uncertainties on the final digit given in parentheses.

Caffeine Label	No. neighbouring N atoms	CAF-2CA exp. linewidth/ Hz	CAF-2CA non-refocussable linewidth / Hz
C10	1	87(1)	20(5)
C10i	1	85(1)	17(2)
C12	1	87(1)	18(3)
C9	1	91(1)	16(3)
C11	2	104(1)	17[a]
C8	2	107(1)	18(4)
C7	2	112(1)	16(2)
C8i	2	107(1)	16(5)

[a] Signal/noise ratio too poor to accurately determine so average of all other T_2' values used.

The inhomogeneous contribution has a Gaussian shape and contains ABMS, quadrupolar broadening, and disorder effects. However, the quadrupolar and ABMS contributions are challenging to measure directly so an estimate is required from a closely related ordered system. For CAF-2CA the inhomogeneous contribution is taken from CAF-CA, Table 7. Note that residual dipolar interactions caused by imperfect decoupling will also contribute to the linewidths, so the same decoupling conditions and MAS frequencies were used in the T_2' and CP/MAS experiments.

Table 7 Breakdown of the linewidth contributions to the ordered cocrystal CAF-CA with errors on the final digit given in parentheses.

Caffeine Label	CAF-CA exp. linewidth / Hz	CAF-CA non-refocussable linewidth / Hz	CAF-CA inhomogeneous linewidth / Hz
C10	76(1)	9(1)	67(1)
C10i	80(1)	11(1)	73(1)
C12	69(1)	9(1)	60(1)
C9	85(1)	5.5(5)	80(1)
C11	104(1)	9(1)[a]	95(1)
C8	100(1)	9(1)[a]	91(1)
C7	100(1)	4.5(2)	97(1)
C8i	75(1)	4.8(6)	71(1)

[a] T_2' values for C8/C11 cannot be distinguished individually due to peak overlap.

3.5.2 Disordered linewidths of caffeine-citric acid hydrate

There is a tentative correlation between the simulated and experimental caffeine ^{13}C linewidths, $R^2 = 0.54$, Figure 43. The error bars denote the 0.1 ppm uncertainty imparted on the calculated ^{13}C shieldings by imperfect geometry optimisation and the magnitude is such that they limit the quality of the observed correlation. Note that the contribution of ABMS is also crudely estimated from the ordered CAF-CA cocrystal, but this is a uniform broadening applied to all sites and so will not affect the quality of the correlation in Figure 43.

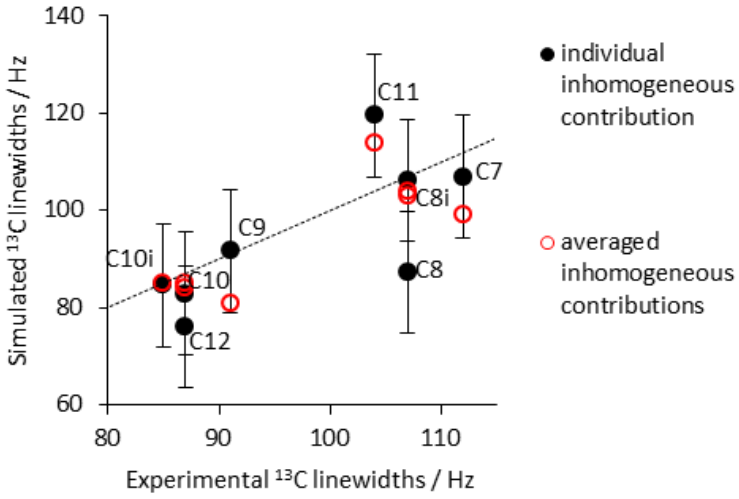

Figure 43 Experimental *vs* simulated ^{13}C linewidths of CAF-2CA derived from the simulated spectra of individual caffeine sites in Figure 42. Error bars denote the 0.1 ppm uncertainty associated with imperfect geometry optimisation and the dotted line corresponds to $y = x$.

The observation of a correlation is promising, given the crude estimation of the inhomogeneous contribution, and the data points in Figure 43 follow the target $y = x$ trend line. However, the disorder contribution to the caffeine ^{13}C linewidths of CAF-2CA is small enough to be on the limit of what is possible with current NMR crystallography methods. Other systems might be more amenable to computational prediction and that the approach could be used in combination with 2D SSNMR methods to probe the correlation of shifts in disordered systems.[3,4,6]

Using the inhomogeneous contribution of CAF-CA as an estimate for the inhomogeneous contribution of CAF-2CA is not ideal, as shown by applying the same methodology to the linewidth simulation of two ordered caffeine cocrystals, CAF-2HBA and CAF-PCA, Figure 44. In the limit of perfect linewidth simulation the data should follow the respective $y = x - c$ trend lines. However, the average standard deviation of the residuals is 6 ppm, indicating that imperfect estimation of the inhomogeneous contribution may be the limiting factor in predicting the disorder contribution to the ^{13}C linewidths. The linewidths in Figure 44 cluster depending on the number of neighbouring nitrogen atoms. This clustering is also observed in Figure 43 for CAF-2CA, suggesting that the quadrupolar broadening may be the main contribution to the linewidths. The red dataset on Figure 43 was calculated using average values for the inhomogeneous contribution depending on the number of neighbouring nitrogen atoms; 69 Hz for one and 88 Hz for two. A similar level of correlation is observed to that of the black dataset where individual CAF-CA inhomogeneous contributions are used, though the scatter is reduced with an $R^2 = 0.69$. This is mainly due to C8, which has a particularly low inhomogeneous contribution because the overlap of C8 and C11 in the CAF-CA spectrum makes it challenging to deconvolute individual linewidths.

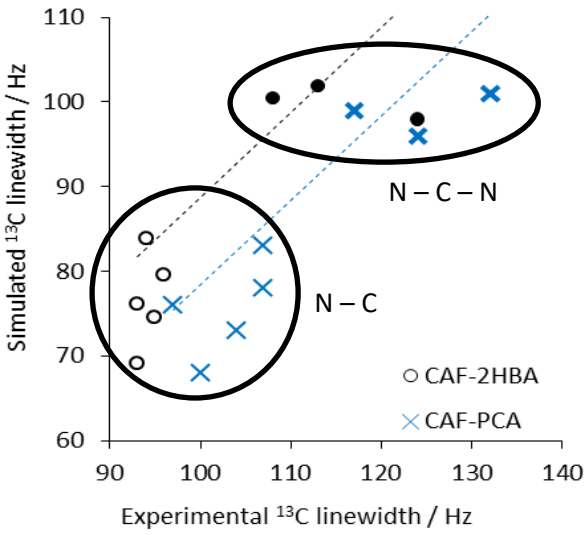

Figure 44 Experimental *vs* simulated ^{13}C linewidths for two ordered caffeine cocrystals derived from the simulated spectra of individual caffeine sites with the inhomogeneous contribution estimated from CAF-CA.

3.5.3 In search of more suitable systems

There are three cocrystals in the CSD that contain disorder of the caffeine molecule.[36,38,47] In all three cases, the relationship between the two disordered orientations of the caffeine is similar to that of CAF-2CA. Additionally, two of these cocrystals are each related to a cocrystal containing no disorder that differs in the stoichiometry of the components, akin to CAF-2CA and CAF-CA. Figure 45 shows the ^{13}C CP/MAS spectra of these five caffeine cocrystals, and differences can be seen between the related systems. CAF-2HBA is ordered and has a relatively sharp average ^{13}C linewidth of 116 Hz compared to the average linewidth of the disordered 2CAF-HBA, 147 Hz. A large difference in linewidth of 132 Hz is observed between the ordered CAF-PCA and the disordered 2CAF-PCA. The disorder contribution may be larger in these systems than in CAF-2CA, so the 0.1 ppm uncertainty imparted by geometry optimisation may be less significant in the simulation of the ^{13}C linewidths.

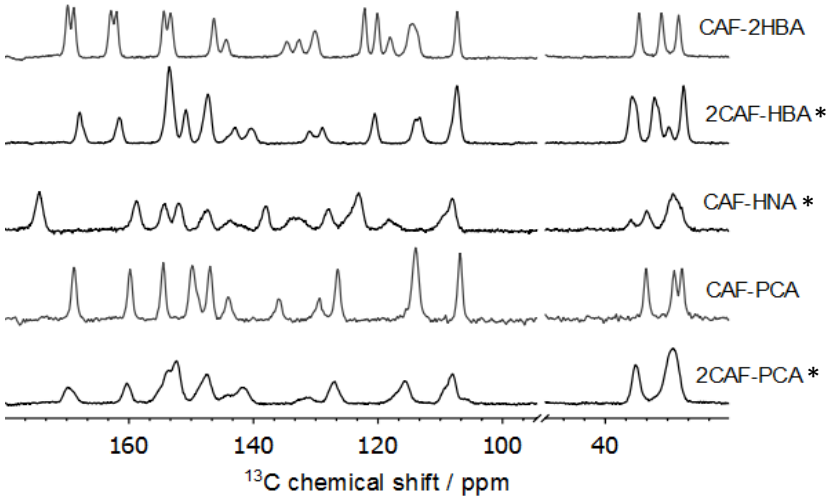

Figure 45 ^{13}C CP/TOSS spectra of five caffeine cocrystals. A star denotes systems that contain disorder of the caffeine molecule. Spectra acquired at 7–8 kHz MAS with a recycle delay of 20 s and contact times of 4–6 ms over 800–1000 transients.

However, CAF-HNA and 2CAF-PCA have very broad ^{13}C spectra that preclude the precise measurement of experimental linewidths, possibly due to a large ABMS contribution, Figure 45. The broad linewidths of 2CAF-PCA are similar to those of previously patented PCA-NA and 2PCA-NA cocrystals.[66] The third cocrystal, 2CAF-HBA, has two caffeine molecules in the asymmetric unit of which only one is disordered, so the SSNMR ^{13}C caffeine peaks overlap preventing accurate determination of experimental linewidths or T_2' relaxation times.

Several factors must be considered in the search for a system that is more amenable to linewidth prediction. Firstly, the disorder contribution should be larger than 0.1 ppm in order for the experimental and simulated linewidths to cover a range large enough to observe a correlation. A system with extensive disorder would be desirable, such as β-caffeine. However, the unit cell of β-caffeine is prohibitively large so the number of simulated-disorder models required would be 2^{12}. Automation tools would need to be produced to make such calculations feasible, which is beyond the scope of this work. Secondly, the linewidths of the caffeine carbon atoms are complicated by quadrupolar broadening due to the presence of four ^{14}N atoms. This might be reduced by using a higher magnetic field, but a system without quadrupolar nuclei would be more desirable.

3.6 Conclusions

This chapter presents the structure of a previously unpublished caffeine-citric acid hydrate. The caffeine molecules are statically disordered and the ^{13}C linewidths reflect the distribution of chemical environments. The system is a good test case to investigate the uncertainties in calculated NMR parameters associated with geometry optimisations, which are carried out in most NMR crystallography studies. It is established that the uncertainty is non-negligible at 0.3 ppm, but can be reduced to 0.1 ppm by careful tightening of the convergence criteria, particularly the energy tolerances. This is at the expense of significantly longer calculation times (2 to 3 times longer). The uncertainties are an order of magnitude smaller than the RMSDs between calculated and experimental shifts, confirming the limiting factor for atom assignment of chemically distinct sites to be the systematic errors in DFT, not the geometry optimisations. Hence, subtle effects can be investigated such as differences in shift between chemically-related sites that are less than 2 ppm. Some insight is also gained into why CAF-2CA forms a disordered structure rather than an ordered one. The energies of the 16 simulated-disorder models are all close in energy, with energy differences similar to those observed between polymorphs. Therefore there is likely to be an entropic bias towards a disordered structure.

The uncertainty from geometry optimisation is significant in relation to the computational prediction of ^{13}C linewidths of the disordered CAF-2CA system. A tentative correlation is observed between experimental and simulated linewidths, though the quality of the correlation is hampered by the 0.1 ppm uncertainty imparted by imperfect geometry optimisation. The quadrupolar contribution from ^{14}N nuclei is an important factor in the simulation of the ^{13}C linewidths, so systems with a larger disorder contribution and no quadrupolar nuclei might be more amenable to prediction. This investigation serves as proof of principle.

3.7 References

(1) E. Skořepová, J. Čejka, M. Hušák, V. Eigner, J. Rohlíček, A. Šturc and B. Kratochvíl, *Cryst. Growth Des.*, **2013**, 13, 5193.
(2) M. Urbanová, A. Šturcová, J. Brus, H. Beneš, E. Skořepová, B. Kratochvíl, J. Čejka, I. Šeděnková, L. Kobera, O. Policianová *et al.*, *J. Pharm. Sci.*, **2013**, 102, 1235.
(3) D. Sakellariou, S. P. Brown, A. Lesage, S. Hediger, M. Bardet, C. A. Meriles, A. Pines and L. Emsley, *J. Am. Chem. Soc.*, **2003**, 125, 4376.
(4) S. Cadars, A. Lesage and L. Emsley, *J. Am. Chem. Soc.*, **2005**, 127, 4466.
(5) A. J. Robbins, W. T. Ng, D. Jochym, T. W. Keal, S. J. Clark, D. J. Tozer and P. Hodgkinson, *Phys. Chem. Chem. Phys.*, **2007**, 9, 2389.
(6) S. Cadars, A. Lesage, C. J. Pickard, P. Sautet and L. Emsley, *J. Phys. Chem. A*, **2009**, 113, 902.

(7) S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Phys.*, **2014**, 141, 164121.
(8) S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Theory Comput.*, **2015**, 11, 5229.
(9) J. D. Hartman, S. Monaco, B. Schatschneider and G. J. Beran, *J. Chem. Phys.*, **2015**, 143, 102809.
(10) R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates and V. Zorin, *Magn. Reson. Chem.*, **2007**, 45, 174.
(11) S. E. Ashbrook and D. McKay, *Chem. Commun.*, **2016**, 52, 7186.
(12) J. N. Latosińska, M. Latosińska, G. A. Olejniczak, J. Seliger and V. Zagar, *J. Chem. Inf. Model.*, **2014**, 54, 2570.
(13) H. J. Lee, K. W. Lee, K. S. Kang, D. Y. Kim, P. H. H., M. J. Lee, H. S. Kim and I. B. Kwon, *Theobromine with an anti-carcinogenic activity*, 2004, US6693104 B2.
(14) *DrugBank*, http://www.drugbank.ca/drugs/DB00201, accessed July 2016.
(15) J. Armellino, *Compositions containing the nonprescription combination of acetaminophen, aspirin and caffeine to alleviate the pain and symptoms of migraine*, 1999, United States Patent and Trademark Office: US5972916 A.
(16) A. Nehlig, J. L. Daval and G. Debry, *Brain Res. Brain Res. Rev.*, **1992**, 17, 139.
(17) P. A. Wood, T. S. G. Olsson, J. C. Cole, S. J. Cottrell, N. Feeder, P. T. A. Galek, C. R. Groom and E. Pidcock, *CrystEngComm.*, **2013**, 15, 65.
(18) C. F. Macrae, P. R. Edgington, P. McCabe, E. Pidcock, G. P. Shields, R. Taylor, M. Towler and J. van De Streek, *J. Appl. Crystallogr.*, **2006**, 39, 453.
(19) M. Falk, W. Chew, J. A. Waiter, W. Kwiatkowski, K. D. Barclay and G. A. Klassen, *Can. J. Chem.*, **1998**, 76, 48.
(20) L. Tavagnacco, S. Di Fonzo, F. D'Amico, C. Masciovecchio, J. W. Brady and A. Cesaro, *Phys. Chem. Chem. Phys.*, **2016**, 18, 13478.
(21) A. Cesàro and G. Starec, *J. Phys. Chem.*, **1980**, 84, 1345.
(22) H. G. M. Edwards, E. Lawson, M. deMatas, L. Shields and P. York, *J. Chem. Soc., Perkin Trans. 2*, **1997**, 10, 1985.
(23) G. D. Enright, V. V. Terskikh, D. H. Brouwer and J. A. Ripmeester, *Cryst. Growth Des.*, **2007**, 7, 1406.
(24) U. J. Griesser and A. Burger, *Int. J. Pharm.*, **1995**, 120, 83.
(25) D. J. Sutor, *Acta Crystallogr.*, **1958**, 11, 453.
(26) P. Derollez, N. T. Correia, F. Danede, F. Capet, F. Affouard, J. Lefebvre and M. Descamps, *Acta Crystallogr. B*, **2005**, 61, 329.
(27) M. Descamps, N. T. Correia, P. Derollez, F. Danede and F. Capet, *J. Phys. Chem. B*, **2005**, 109, 16092.
(28) J. J. Moura Ramos, N. T. Correia, H. P. Diogo and M. Descamps, *J. Phys. Chem. B*, **2006**, 110, 8268.
(29) A. Hedoux, A. A. Decroix, Y. Guinet, L. Paccou, P. Derollez and M. Descamps, *J. Phys. Chem. B*, **2011**, 115, 5746.
(30) N. A. Murugan and A. Sayeed, *J. Chem. Phys.*, **2009**, 130, 204514.
(31) C. W. Lehmann and F. Stowasser, *Chemistry*, **2007**, 13, 2908.
(32) A. Nowaczyk, R. C. Nath, H. Zimmermann and R. Bohmer, *Z. Phys. Chem.*, **2012**, 226, 1115.
(33) L. Carlucci and A. Gavezzotti, *Chemistry*, **2004**, 11, 271.
(34) M. Habgood, *Cryst. Growth Des.*, **2011**, 11, 3600.
(35) M. Habgood, R. Grau-Crespo and S. L. Price, *Phys. Chem. Chem. Phys.*, **2011**, 13, 9590.
(36) D. K. Bučar, R. F. Henry, X. Lou, R. W. Duerst, T. B. Borchardt, L. R. MacGillivray and G. G. Zhang, *Mol. Pharm.*, **2007**, 4, 339.
(37) S. Karki, T. Friščić, W. Jones and W. D. Motherwell, *Mol. Pharm.*, **2007**, 4, 347.

(38) D.-K. i. Bučar, R. F. Henry, X. Lou, R. W. Duerst, L. R. MacGillivray and G. G. Z. Zhang, *Cryst. Growth Des.*, **2009**, 9, 1932.
(39) M. Habgood and S. L. Price, *Cryst. Growth Des.*, **2010**, 10, 3263.
(40) D. K. Bučar, G. M. Day, I. Halasz, G. G. Z. Zhang, J. R. G. Sander, D. G. Reid, L. R. MacGillivray, M. J. Duer and W. Jones, *Chemical Science*, **2013**, 4, 4417.
(41) T. Leyssens, N. Tumanova, K. Robeyns, N. Candoni and S. Veesler, *CrystEngComm.*, **2014**, 16, 9603.
(42) C. C. Sun and H. Hou, *Cryst. Growth Des.*, **2008**, 8, 1575.
(43) S. Ghosh and C. M. Reddy, *Angew. Chem. Int. Ed. Engl.*, **2012**, 51, 10319.
(44) T. Friščić, L. Fabian, J. C. Burley, W. Jones and W. D. Motherwell, *Chem. Commun.*, **2006**, 48, 5009.
(45) G. He, P. S. Chow and R. B. H. Tan, *Cryst. Growth Des.*, **2010**, 10, 3763.
(46) T. Leyssens, G. Springuel, R. Montis, N. Candoni and S. Veesler, *Cryst. Growth Des.*, **2012**, 12, 1520.
(47) N. Schultheiss, M. Roe and S. X. M. Boerrigter, *CrystEngComm.*, **2011**, 13, 611.
(48) *National Center for Biotechnology Information. PubChem Compound Database CID: 6241*, https://pubchem.ncbi.nlm.nih.gov/compound/6241, accessed August 2016.
(49) H. E. Kerr, H. E. Mason, H. A. Sparkes and P. Hodgkinson, *CrystEngComm.*, **2016**, 18, 6700.
(50) M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminform.*, **2012**, 4, 17.
(51) A. D. Bond, R. Boese and G. R. Desiraju, *Angew. Chem. Int. Ed. Engl.*, **2007**, 46, 618.
(52) M. K. Mishra, G. R. Desiraju, U. Ramamurty and A. D. Bond, *Angew. Chem. Int. Ed. Engl.*, **2014**, 53, 13102.
(53) J. W. Fischer, L. H. Merwin and R. A. Nissan, *Appl. Spectrosc.*, **1995**, 49, 120.
(54) A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, **2009**, 102, 073005.
(55) J. Nyman and G. M. Day, *CrystEngComm.*, **2015**, 17, 5154.
(56) A. Bērziņš and P. Hodgkinson, *Solid State Nucl. Magn. Reson.*, **2015**, 65, 12.
(57) 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.
(58) N. Mercadal, S. P. Day, A. Jarmyn, M. B. Pitak, S. J. Coles, C. Wilson, G. J. Rees, J. V. Hanna and J. D. Wallis, *CrystEngComm.*, **2014**, 16, 8363.
(59) A. J. Cohen, P. Mori-Sanchez and W. Yang, *Chem. Rev.*, **2012**, 112, 289.
(60) R. Laskowski, P. Blaha and F. Tran, *Phys. Rev. B*, **2013**, 87, 195130.
(61) X. Li, L. Tapmeyer, M. Bolte and J. van de Streek, *ChemPhysChem.*, **2016**, 17, 2496.
(62) P. Hodgkinson, *pNMRsim: a general simulation program for large problems in solid-state NMR*, www.dur.ac.uk/paul.hodgkinson/pNMRsim, accessed 2016.
(63) S. Sturniolo, T. F. 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.
(64) V. Zorin, *GSim - a visualisation and processing program for solid-state NMR*, www.sourceforge.net/projects/gsim/, accessed 2016.
(65) G. De Paepe, A. Lesage, S. Steuernagel and L. Emsley, *ChemPhysChem.*, **2004**, 5, 869.
(66) M. J. Bevill and N. Schultheiss, *Cocrystals of p-coumaric acid*, 2014, United States Patent and Trademark Office: US 20140073674A1.

Chapter 4: Investigating proton positions in hydrogen bonds

Based on work published in "*Structure and physicochemical characterisation of a naproxen-picolinamide cocrystal*", H. E. Kerr, L. K. Softley, K. Suresh, P. Hodgkinson, I. R. Evans, Acta Cryst. C, 2016, accepted manuscript. Sections 4.2 and 4.3 are lightly adapted from the publication.

4.1 Hydrogen bond definitions and challenges

Hydrogen bonding is one of the key intermolecular interactions exploited in crystal engineering and the location of hydrogen atoms is essential in fundamental studies of structure. In particular, the location of hydrogen atoms in the synthons of cocrystals must be known because this determines where a material is a salt or a cocrystal; an important discrimination for the pharmaceutical industry. It is common for acid-base interactions to be manipulated to form cocrystals and it is generally accepted that a neutral cocrystal will tend to be formed, rather than a salt, if the ΔpK_a between the acid and base is small.[1,*] However, there are several examples of systems that cannot be simply classified as either a salt or a cocrystal but exist in the continuum between the two classes.[2] These tend to occur when $0 \leq \Delta pK_a \leq 3$[1,3] but other factors are also important, for example lattice packing effects.[2,3] Systems classified as intermediary between a salt and a cocrystal are undesirable for pharmaceutical use because the FDA cocrystal guidance paper does not consider such a protonation state.[4] Therefore, the location of hydrogen atoms is of paramount importance.

The hydrogen bond was recently defined in a technical report[5] by IUPAC as "…*an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in*

[*] pK_a values are measured in solution and it is generally assumed that these values are transferable to the solid state. However, the relationship between solution and solid pK_a values is likely to be less straightforward, so the ΔpK_a shift should be viewed as a rough indicator, not a strict rule.

which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation." The specified evidence for bond formation includes (1) the XH···Y geometry, which tends to linearity, (2) the type of interaction, specifically there should be an element of electrostatics and the interactions must not be entirely dispersive in nature, and (3) spectroscopic data such as a red shift in the XH vibrational frequency or ^{1}H deshielding in NMR spectra. The potential energy curve of a hydrogen bond plotted as a function of hydrogen position also provides a fundamental criterion for classifying hydrogen bonds,[6] Figure 46. Some systems containing strong hydrogen bonds with low/no energy barrier have been observed experimentally to exhibit proton migration, such as 3,5-pyridine dicarboxylic acid[7,8] and lauric acid,[9] which are potentially useful as functional materials.

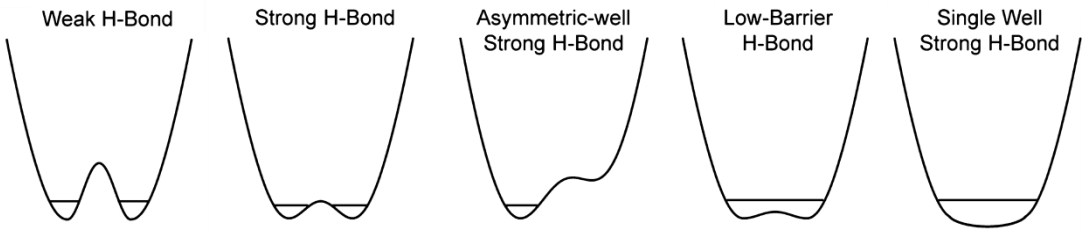

Figure 46 Potential energy as a function of proton position for various different types of hydrogen bond. Reproduced with permission from White *et al.*, J. Phys. Chem. B, 2015, 119, 35, 11581–11589. Copyright © 2015 American Chemical Society.

Hydrogen atoms are challenging to locate by XRD methods due to their very low scattering power to X-rays. When they can be located by XRD the refined position is different to the position determined by neutron diffraction, especially in the context of hydrogen bonds. Commonly, hydrogen atoms are refined to be closer to the heavy atom to which they are bonded compared to neutron diffraction data. For example, a difference of 0.15 Å is common for an O–H proton position.[10] The reason for this discrepancy is the asymmetric electron distribution around the covalent X–H bond, with the refinement of XRD data placing the atom at the centre of mass of the electron density. It is also challenging to computationally predict whether a system will crystallise as a salt or cocrystal because the energy differences can be very small, placing heavy demands on the computational methods, and temperature effects can be significant.[11]

The question of proton location is intrinsically a local structure problem, and has therefore been probed by SSNMR methods in various cocrystals.[12-20] A combination of ^{1}H and ^{15}N SSNMR is commonly used due to the sensitivity of chemical shifts to protonation state. ^{1}H chemical shifts of hydrogen-bonded protons span a large range,[21] from 8–20 ppm, and the magnitude of the shift is related to the length of the hydrogen bond, which is a proxy for bond strength.[22] It has been shown that the highest ^{1}H chemical shifts for OH···O, OH···N and NH···O hydrogen bonds occur for short donor–donor distances and a symmetric central position of the proton.[23] OH···N and NH···O hydrogen bonds are a common and robust intermolecular interaction in cocrystal synthons,[24] so ^{15}N nuclei can be probed to establish proton position. Dipolar dephasing ^{15}N experiments can distinguish between protonated and unprotonated tertiary amines[25] and, more recently, spectral editing techniques have been developed based on H–N indirect coupling.[26] In these experiments ^{15}N magnetisation evolves during a spin echo period under the influence of the one bond H–^{15}N indirect coupling. Nitrogen atoms bonded to two protons experience a positive modulation, while those bonded to one or three protons are inverted so ionised and unionised states can be differentiated. Other SSNMR methods also probe H–N distances *via* 2D HETCOR experiments or solution-like heteronuclear multiple quantum correlation (HMQC) experimnts.[16,26,27]

The standard SSNMR techniques used to investigate hydrogen bonding in cocrystals are highlighted by a study on a cocrystal of 3,5-dimethylpyrazole with 4,5-dimethylimidazole.[28] Both compounds are important templates for biological and pharmaceutical applications, as well as displaying promiscuous hydrogen bonding behaviour. Strong hydrogen bonds were identified in the cocrystal by the presence of ^{1}H resonances at high chemical shifts, 13 ppm and 15 ppm, attributed to two different NH···N interactions. Assignment of these sites was assisted by DFT calculations of the ^{1}H chemical shieldings and ^{1}H–^{13}C HETCOR experiments to observe long range H···C distances. Proton proximities were also investigated using a DQ/SQ experiment, Figure 47, which shows that equivalent NH protons are more than 3.5 Å apart, while non-equivalent NH protons are close enough to produce DQ/SQ correlations. This information was useful for discriminating between different models of the structure refined from PXRD data.

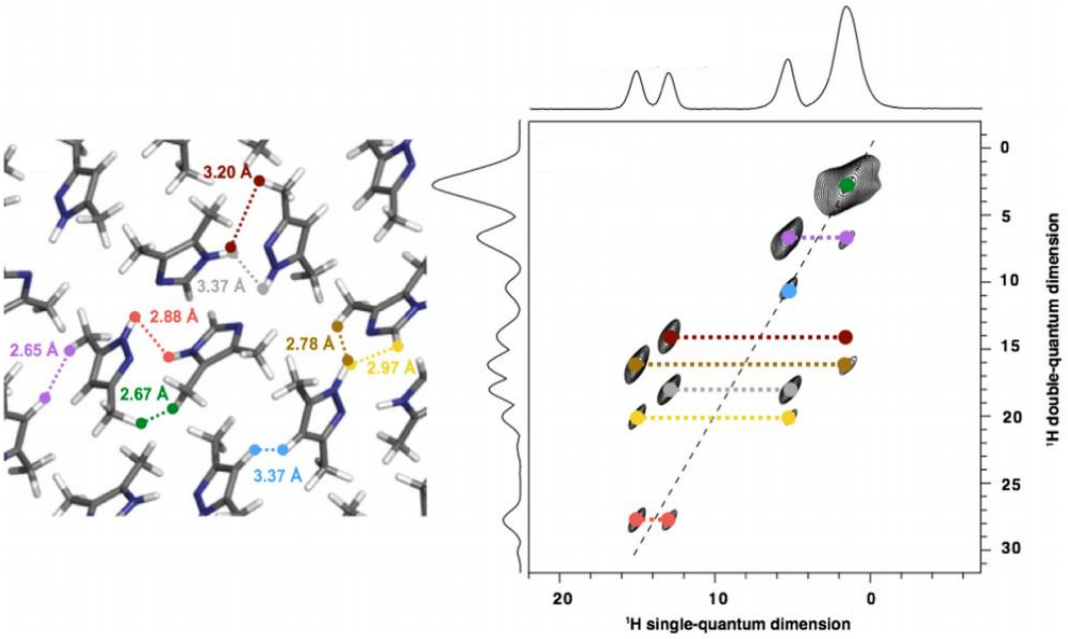

Figure 47 ^{1}H–^{1}H DQ/SQ spectrum of a 1:1 cocrystal of 3,5-dimethylpyrazole with 4,5-dimethylimidazole, with correlations assigned by colour matching those indicated in the crystal structure. Atom distances corresponding to each correlation are also given. Reproduced with permission from Sardo *et al*., Solid State Nucl. Magn. Reson., 2015, 65, 49–63. Copyright © 2014 Elsevier Inc.

This chapter focusses on the use of NMR crystallography methods, such as those mentioned above, to investigate the position of hydrogen-bonded protons in a new cocrystal of NPX. Both XRD and SSNMR methods are necessary to unambiguously place the hydrogen-bonded protons and the potential energy curve is compared to systems containing similar hydrogen bonding motifs to examine the possibility of proton migration.

4.2 Naproxen in the pharmaceutical industry

Naproxen, (*S*)-2-(6-methoxynaphthalen-2-yl) propanoic acid, Figure 48(i), is a non-steroidal anti-inflammatory drug used to relieve pain and fever symptoms.[29] It is regularly used in the treatment of arthritis, dysmenorrhea, and acute gout.[30] Side effects of NPX are generally minor at low doses, including constipation and nausea, but prolonged high doses can lead to gastro-intestinal bleeding.[31] NPX is a weak acid[29] with pK_a = 4.2, leaving it unionised in gastro-intestinal fluids. The hydrophobic aromatic rings in the molecule disfavour interactions with polar molecules so it is insoluble in aqueous media. It is formally classified as a low-solubility high-permeability drug,[32] but the solubility is dependent on the pH of the surrounding environment. In environments with high pH, such as cell membranes, NPX becomes ionised. The charged COO^{-} forms more favourable interactions

with water, thus enhancing aqueous dissolution. Therefore, NPX is currently marketed as a sodium salt to overcome the solubility problems of the neutral molecule. However, alternative solutions to improving the solubility of NPX across all pH values are desirable. With this in mind, inclusion complexes have been synthesised with cyclodextrins with molecules of NPX incorporated in the cavities.[33] The polar surface of the complex makes it soluble in aqueous media, and subsequent targeted delivery of NPX has been shown to increase the bioavailability. Alternatively, the dissolution rate of NPX can be increased four-fold by formation of an amorphous binary system with cimetidine, which is often co-administered with NPX to reduce gastro-intestinal problems.[34] However, the COOH group means NPX is also amenable to cocrystal formation in order to boost solubility and this has been an active area of research over recent years, as discussed below.

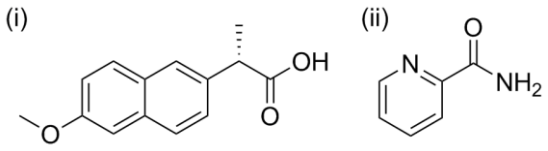

Figure 48 Molecular structures of **(i)** (*S*)-naproxen and **(ii**) picolinamide.

4.2.1 Naproxen and its known cocrystals

The crystal structure of NPX[35] has been known since 1985. There are two NPX molecules in the unit cell and one in the asymmetric unit. Each molecule forms two symmetry-equivalent hydrogen bonds to neighbouring molecules through the COOH group, synthon **A** in Figure 49. The aromatic rings give NPX a rigid structure and this, coupled with the presence of the hydrogen bonding carboxyl group on one side, means that NPX is not prone to local disorder over the whole molecule in the solid state. In fact, it has not been possible to produce an amorphous sample of NPX by grinding the crystalline form.[34]

NPX is suitable for cocrystal formation with hydrogen bonding possibilities *via* the COOH group, which can form a range of synthons.[36] Additionally, π-stacking interactions are made possible by the aromatic region. At the time of writing 16 cocrystals of NPX have been synthesised and characterised using a variety of techniques including XRD, infra-red spectroscopy (IR), and differential scanning calorimetry (DSC). Cocrystals with the following coformers have been reported: nicotinamide (NA), isonicotinamide (INA), picolinamide (PA),[37,38] trans-1,2-bis(4-pyridyl)ethylene (TBPE),[39] duloxetine,[40] tramadol,[41] bipyridine (BPY)

and piperazine (PPZ),[42] and several chiral amino acids including alanine (AL), zwitterionic prolinium (PR), tyrosine (TY), tryptophan (TP), and N-octylglucamine (OGL).[43-45] The structures of many of these cocrystals have been determined using SCXRD and are reported in the literature. The relevant synthons are shown in Figure 49 and summarised for all 16 reported cocrystals in Table 8.

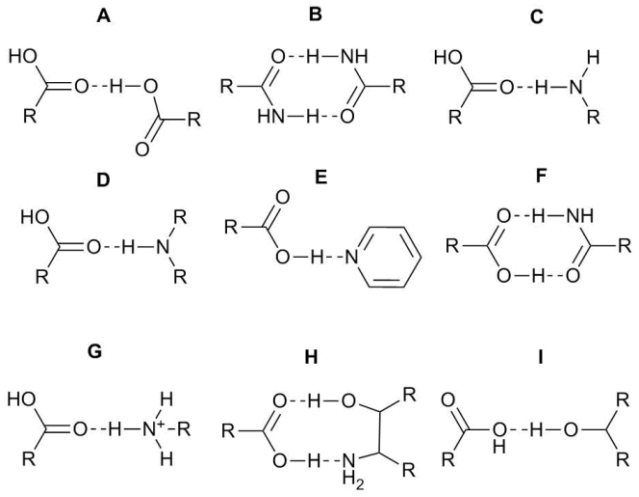

Figure 49 Synthons present in NPX and its cocrystals.

Table 8 Synthons present in NPX and the cocrystals previously reported in the literature. Relevant references are given in the main text.

	A	B	C	D	E	F	G	H	I
S-NPX	*								
S-2NPX.NA			*		*	*			
S-NPX.INA		*	*		*				
S-NPX.TBPE					*				
RS-NPX.BPY					*				
S-NPX.BPY					*				
RS-NPX.PPZ				*					
S-NPX.PPZ				*					
S-NPX.*L*-AL	*						*		
S-NPX.*D*-AL	*						*		
S-NPX.*D*-TY	*						*		*
S-NPX.*D*-TP	*			*			*		
S-NPX.*L*-PR	*		*						
S-NPX.*D*-PR	*		*						
RS-NPX.*L*-PR	*								
RS-NPX.*DL*-PR	*		*						
S-NPX.*D*-OGL				*				*	*

The coformers of five of the reported cocrystals contain a pyridine ring, which readily interacts with the carboxylic acid group on NPX *via* synthon **E**. The importance of synthon **E** is highlighted by observing the synthons present in cocrystals composed of isomeric pyridine carboxamides NA and INA coformers. Formation of synthon **E** in 2NPX-NA prevents the formation of synthon **B**, which is seen in NPX-INA.[46] Consequently, synthon **F** is formed instead. Synthon **A** is present in most of NPX-amino acid cocrystals, highlighting the preference for the formation of COOH···COOH dimers. In all cases the packing is governed by π-π stacking interactions between aromatic rings in NPX and/or the coformer, resulting in a herringbone arrangement.

NMR crystallography studies have been employed extensively to characterise hydrogen bonding in a variety of cocrystals,[12,13,19,27,47-49] including cocrystals of NPX. For example, 2D ^{1}H–^{13}C SSNMR HETCOR experiments of 2NPX-NA confirmed that the hydrogen bonding network contained a relatively rare example of a single carboxyl group, giving rise to two different intermolecular synthons in the same cocrystal, probably due to the 2:1 stoichiometry.[46] Pure NPX and its sodium salt have also been studied by SSNMR methods, with full chemical shift assignment and ring current effects investigated[50] in 2013, while ^{23}Na SSNMR experiments were also used to study NPX hydrates and solvates.[51] The ^{13}C and ^{1}H spectral assignments of NPX have been critically analysed recently using calculated NMR parameters with DFT, with three correlations from the paper by Ando *et al.* reassigned in the crowded region of the ^{1}H–^{13}C HETCOR spectra.[52] The tetrahydrate sodium salt of NPX has a complex disordered sodium and water network, which can be interpreted in terms of either stacking faults or multiple twinning following an NMR crystallography study utilising SSNMR, XRD and computational methods.[53]

Several of the NPX cocrystals have been found to have enhanced solubility compared to pure NPX. In particular, the 2NPX-NA cocrystal was found to be three times more soluble than the commercially available NPX salt and also absorbed water to a lesser degree, making it a desirable cocrystal for pharmaceutical formulation.[46] NPX-duloxetine[54] and NPX-tramadol[55] are patented so solubility information is not available, but it can be assumed that the physicochemical properties are favourable.

4.2.2 Structure of naproxen-picolinamide

The preparation of cocrystals between NPX and the third pyridine carboxamide isomer PA, Figure 48(ii), has been suggested by Castro *et al*.; however, single crystals could not be formed in their study and so no crystal structure was determined.[37] The data acquired that

supported cocrystal formation included DSC data, which indicated the melting point (91 ˚C) was significantly lower than that of the pure components or a physical mixture. IR and PXRD patterns were also found to be distinctly different from those of pure NPX and PA, but no structural data was obtained.

In order to solve the structure of this cocrystal a sample was synthesised through grinding by L. Softely of the I.R.E group at Durham University. Comparison of the ^{13}C spectra of the pure components with that of the grinding product confirms the formation of a cocrystal, Figure 50, which was acquired by the Durham SSNMR service. The physical mixture and pure NPX show identical spectra, consistent with a lack of interaction between components,* but the spectrum of the NPX-PA cocrystal is distinctly different. Firstly, some of the resonances arising from NPX have shifted relative to the pure NPX spectrum. Secondly, signals are now observed from PA carbon sites as a result of the intimate association of PA and NPX, which shortens the ^{1}H T_1 relaxation time of the PA resonances. The doubling of most signals is consistent with the presence of two crystallographically non-equivalent molecules of NPX and PA in the asymmetric unit.

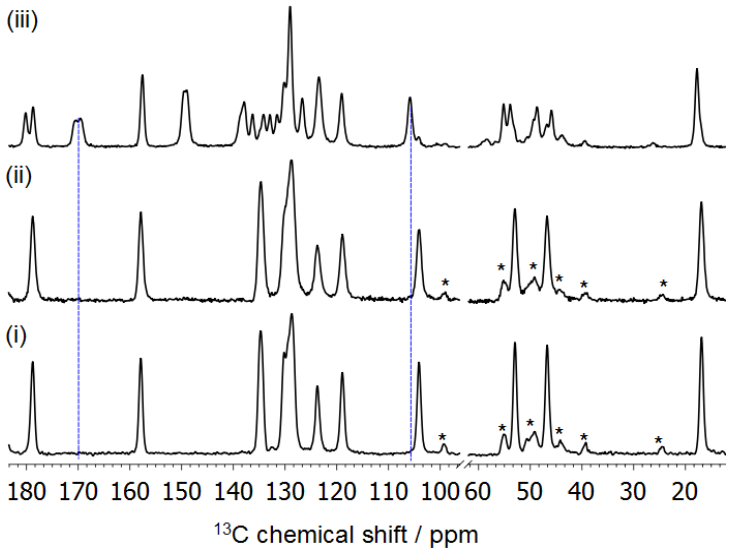

Figure 50 ^{13}C CP spectra of **(i)** pure NPX, **(ii)** a physical mixture of NPX and PA and **(iii)** the NPX-PA cocrystal. The three spectra were acquired by the Durham University solid-state NMR service at a ^{13}C frequency of 100.6 MHz with 8 kHz MAS, a recycle delay of 5 s and a contact time of 5 ms. Decoupling was applied with a ^{1}H nutation frequency of 74 kHz over 200 transients. The blue lines are guides for the eye indicating peaks that shift or appear on cocrystal formation. Asterisks denote spinning sidebands.

* Peaks from PA are not observed because it has a long ^{1}H T_1 relaxation time compared to the recycle delay of 5 s used in the acquisition of these spectra.

The crystal structure for NPX-PA was solved by the I.R.E group at Durham University and the cocrystal was found to crystallise in the space group $P2_1$, with four molecules in the asymmetric unit, Figure 51(i). The molecules form heterodimers *via* synthon **F**. The presence of four molecules in the asymmetric unit yields two inequivalent **E** heterodimers, arranged in a zigzag pattern along the crystallographic *c* axis, Figure 51(ii). These dimers interact with each other *via* synthon **C** with $d_{O\cdots N}$ = 3.022(4) Å. The details of the hydrogen bonding in NPX-PA are summarised in Table 9. A CSD search* yielded 237 structures containing the synthon **F** fragment and the NPX-PA hydrogen bond distances fall within the range observed over these CSD structures, Figure 52.

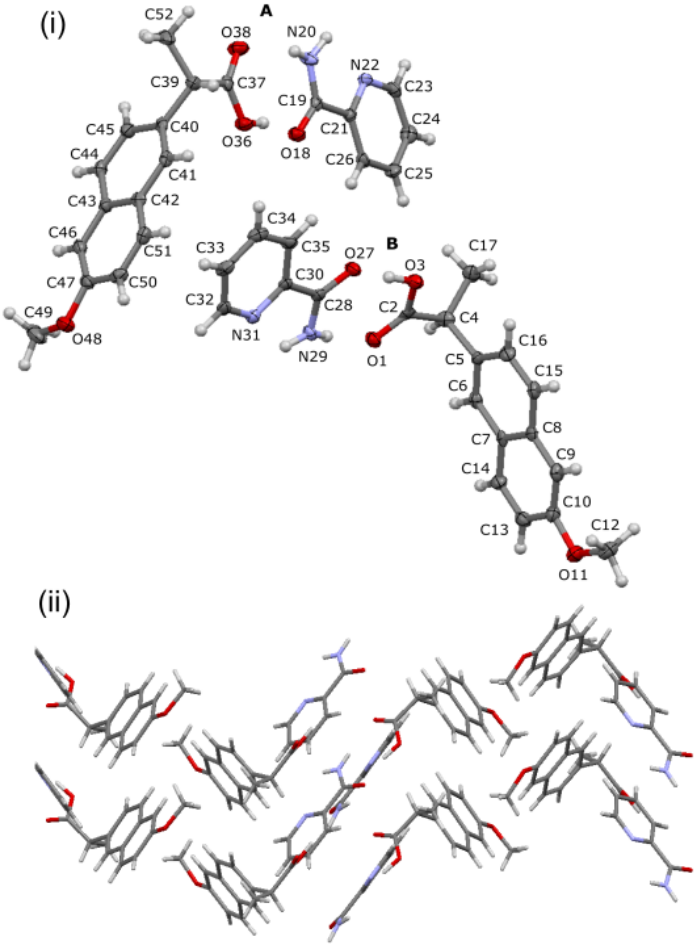

Figure 51 (i) Asymmetric unit of NPX-PA with atom labelling and ADPs shown at the 50 % probability level. Hydrogen atoms are shown with a fixed radius of 0.15 Å. **(ii)** Schematic of the herringbone packing in NPX-PA viewed along the *c* crystallographic axis.

* Subject to the following restrictions: 'only organics', '3D coordinates determined', 'no powder structures' and 'not disordered'.

Table 9 Hydrogen bond geometry in the NPX-PA cocrystal.

Dimer	Hydrogen bond	d_{X-H} / Å	$d_{H\cdots A}$ / Å	$d_{X\cdots A}$ / Å	Bond angle / °
A	O36—H361···O18	0.85(4)	1.74(4)	2.572(5)	166(4)
	N20—H202···O38	0.85(4)	2.15(4)	2.965(5)	158(4)
B	O3—H31···O27	0.97(6)	1.62(6)	2.579(5)	169(6)
	N29—H291···O1	0.81(4)	2.08(4)	2.890(5)	175(4)

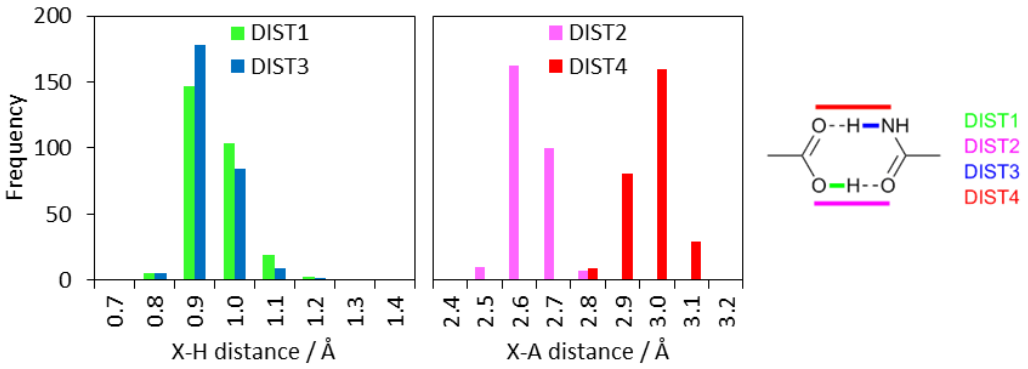

Figure 52 Histograms showing the ranges for the bond distances of synthon **F**, right, for the 237 structures obtained from a CSD search of the synthon **F** fragment.

The structure of NPX-PA is consistent with the trends seen in previously determined NPX cocrystal structures. In particular, the COOH···CONH heterodimer formation is reminiscent of the synthon formation in 2NPX-NA and NPX-INA. However, synthon **E** is not observed in NPX-PA and alternatively synthon **F** is formed, in contrast to the structures of 2NPX-NA and NPX-INA. This demonstrates the role of basicity in NPX cocrystal formation. Among the three pyridine carboxamide coformers, NA and INA are more basic (pK_a of 3.63 and 3.45) so the acid-pyridine synthon **E** is observed. However, PA exhibits the *ortho* effect and is less basic (pK_a of 1.17) so the PA carboxamide group interacts with NPX *via* synthon **F** rather than synthon **E**.

The DSC trace of NPX-PA, also acquired by L. Softley, shows a single endothermic peak at 93 °C, which represents the melting of a single solid phase in keeping with the endothermic peak observed at 91 ˚C by Castro *et al.*[37] The intrinsic dissolution rate of the NPX-PA cocrystal is essentially the same as commercial NPX, as established by A. Nangia and coworkers at the University of Hyderabad, India.

4.3 Hydrogen bond discrimination, refinement and validation

A question that arises from the SCXRD data concerns the OH···O hydrogen bonds in the two synthon **F** dimers. While the donor-acceptor distances are essentially the same, Table 9, the donor-hydrogen bond lengths are less precise given the larger standard uncertainties of these parameters and systematic issues with locating proton positions using XRD. In particular, the X–H distances in Table 9 might suggest that the two crystallographically unique COOH···CONH dimers are significantly different, but the large standard errors of the hydrogen positions preclude any definitive conclusions. Additionally, $|\Delta pK_a| = 3$ between NPX and PA indicating that proton transfer might be possible.[1,3] These uncertainties arise because the quality of the XRD data was poorer than usual: the ratio between numbers of observed reflections to number of refined parameters is low. As a result, an NMR crystallography investigation is discussed below aiming to characterise the hydrogen bonding with more confidence and to conclusively reject the possiblity of proton disorder.

4.3.1 Experimental details

Samples of NPX-PA for SSNMR experiments were synthesised by LAG grinding by L. Softley of the I.R.E. group, Durham University. The structural model obtained from SCXRD data was used to fit PXRD data obtained on the bulk sample. A Rietveld fit, not shown, had no peaks unaccounted for so the polycrystalline material prepared by mechanochemistry is a single-phase product and identical to the single crystal form of NPX-PA.

DFT calculations were carried out on three different hydrogen bonding models, Figure 53. The first was the structure refined from SCXRD data. NMR calculations were carried out after optimisation of all 184 atom positions. The two alternative models, Model 1 and Model 2, were produced by manually moving the hydrogen-bonded protons along the OH···O and NH···O bond vectors by editing the Cartesian coordinates of the optimised SCXRD structure. The hydrogen atoms in both COOH···CONH dimers were moved simultaneously. The energy was calculated as the protons were stepped along the hydrogen bond vectors in 16 stages. NMR calculations for Model 2 were performed after optimisation of the hydrogen atom positions, but no optimisation could be carried out for Model 1 because the hydrogen atoms reverted back to the XRD structure positions. Geometry optimisation including dispersion correction did not significantly change the hydrogen atom position or predicted shielding values.

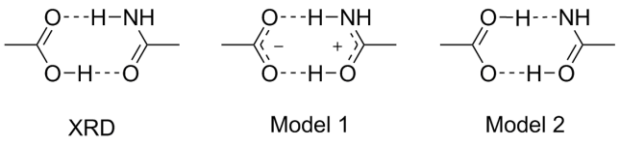

Figure 53 Hydrogen bonding motifs in the XRD structure of NPX and the two alternative hydrogen bonding models, Model 1 and Model 2, used in DFT investigations.

4.3.2 Spectral assignment

The full assignment is given in Figure 54(i) and Table 10, aided by a ^{13}C CP/TOSS/NQS spectrum, Figure 54(ii), 2D ^{1}H–^{13}C HETCOR experiments acquired with a contact time of 100 µs, Figure 55(i), and a contact time of 1 ms, Figure 55(ii), as well as CASTEP-calculated shieldings, discussed below. The labelling scheme follows that of Figure 51(i). A direct excitation ^{13}C experiment with a short recycle delay of 0.5 s, Figure 54(iii), shows that the only obviously dynamic sites are the methyl groups.

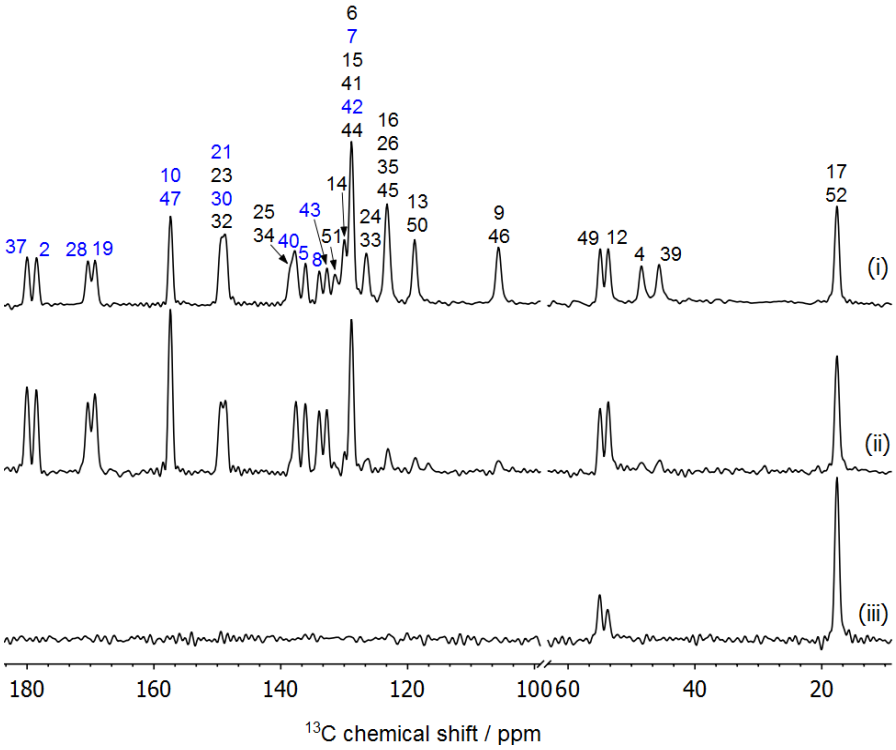

Figure 54 ^{13}C spectra of NPX-PA all at 8 kHz MAS. **(i)** CP/TOSS and **(ii)** CP/TOSS/NQS each acquired over 200 transients with a 5 s recycle delay and 2.5 ms contact time. A dipolar dephasing delay of 80 µs was used to suppress non-quaternary sites. **(iii)** Direct excitation experiment acquired over 100 transients with a recycle delay of 0.5 s. All peaks are assigned, though the assignment at 130 ppm is ambiguous due to peak overlap. Quaternary carbon peaks are labelled in blue.

Table 10 Carbon chemical shifts and assignments, along with predicted shieldings and the associated .magres labelling.* The shifts are taken from the CP/TOSS spectrum acquired at a ^{13}C frequency of 125.67 MHz.

Atom label	Chemical shift / ppm	Magres	Notes on assignment[a,b]	Atom Label	Chemical shift / ppm	Magres	Notes on assignment[a,b]
C17	17.5	14	C, S	C14	130	11	C?, S?, L?
C52		40	C, S	C51	131.4	39	C?, S?, L?
C39	45.7	28	C, S	C43	132.8	32	C, Q
C4	48.6	2	C, S	C8	134	6	C, Q
C12	53.7	9	C, S	C5	136.1	3	C, Q
C49	55	37	C, S	C40	137.6	29	C, Q
C9	105.7	7	C, S?	C25	138.3	18	C
C46		35	C, S?	C34		24	C
C13	118.9	10	C, S?	C21	148.9	16	C, Q
C50		38	C, S?	C30		22	C, Q
C16		13	C, S?	C23	149.3	20	C, S?
C26		17	C, S?	C32		26	C, S?
C35	123.3	23	C, S?	C10	157.4	8	C, L H91: 2.16 Å, H131: 2.16 Å
C45		34	C, S?	C47		36	C, L H461: 2.16 Å, H501: 2.16 Å
C24	126.5	19	C, S?	C19	169.4	15	C, Q
C33		25	C, S?	C28	170.4	21	C, Q
C6		4	C?, S?, L? H41: 2.56 Å	C2	178.5	1	C, Q
C7		5	C?, Q	C37	180	27	C, Q
C15	128.9	12	C?, S?, L? H52: 2.37 Å				
C41		30	C?, S?, L? H391: 2.30 Å				
C42		31	C?, Q				
C44		33	C?, S?				

[a] Symbols used to indicate the basis of assignment: C = CASTEP-calculated ^{13}C shielding, Q = ^{13}C peak in non-quaternary suppression spectrum, S = cross peaks in HETCOR experiment with a short contact time, L = cross peaks in HETCOR experiment with a long contact time, ? = evidence is suggestive rather than definitive.

[b] Distances to selected non-bonded hydrogen atoms are given up to 2.70 Å.

* Magres labelling denotes the atom labelling that is output from CASTEP in both geometry and NMR calculations. The ordering is the generally the same as the crystallographic labelling.

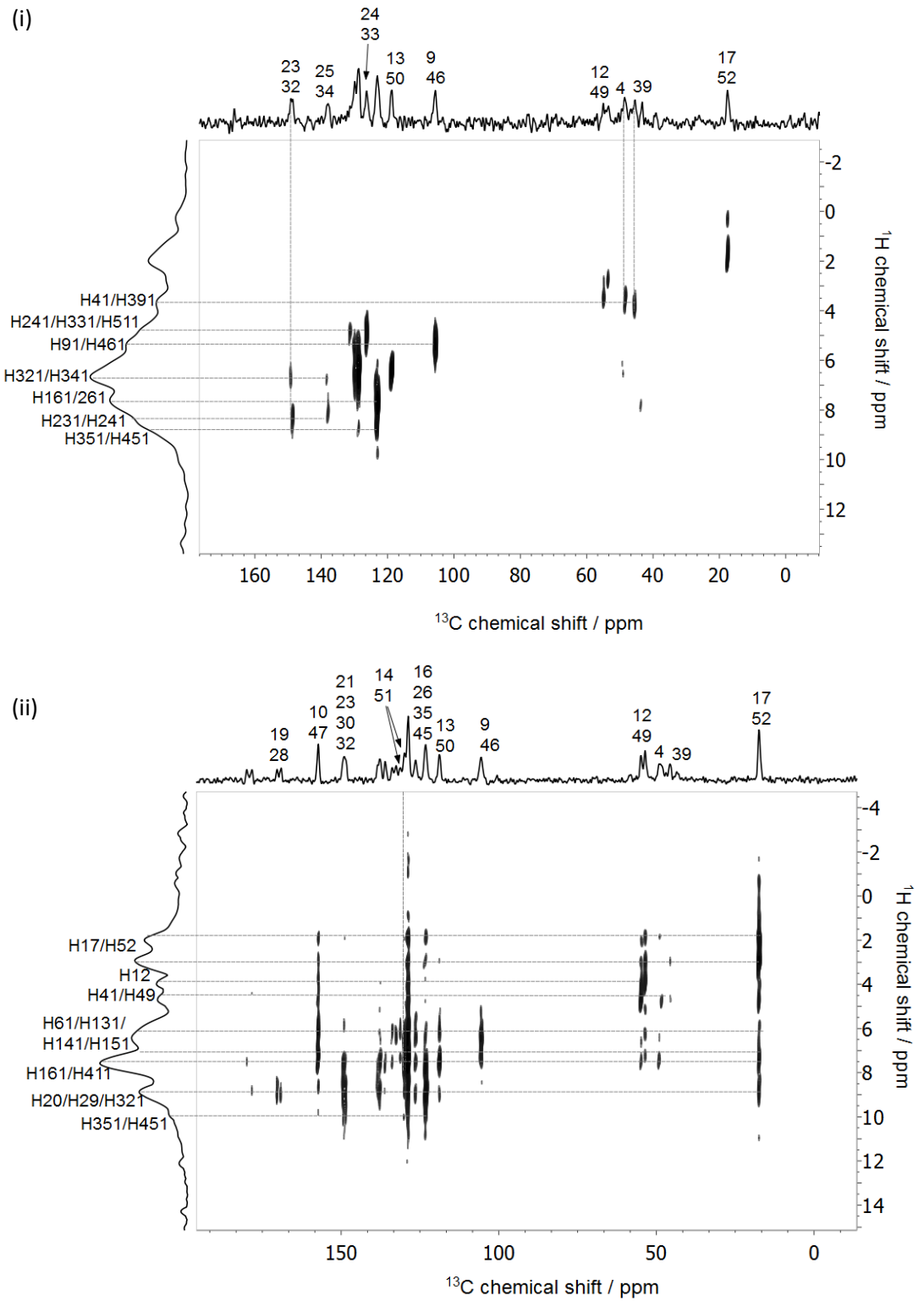

Figure 55 ^{1}H–^{13}C HETCOR experiments acquired at 10 kHz MAS with a recycle delay of 2 s with 64 t_1 increments each acquired with 100 transients. The contact times were **(i)** 100 µs and **(ii)** 1 ms. The traces are projections summed over the rows and columns. Hydrogen atom labelling is shown in Figure 56(iii).

The assignment at 130 ppm is ambiguous due to the overlap of 8 signals, but C14 and C51 are assigned to 130 ppm and 131.4 ppm respectively by the lack of a correlation between these peaks and any low shift protons (less than 4 ppm) in the ^{1}H–^{13}C HETCOR spectrum acquired with a long contact time of 1 ms, Figure 55(ii). These carbon sites are the only sites in the 130 ppm region that are further than 3 Å away from low shift protons. Other assignments were straight forward. Figure 55(i) shows the ^{1}H–^{13}C HETCOR spectrum acquired with a short contact time of 100 µs so only correlations between directly bonded atoms are visible. Differences can be observed in ^{1}H chemical shift between the two crystallographically unique dimers A and B. Note that the hydrogen atom labelling is given in Figure 56(iii). The correlation between C4/C39 and H41/H391 reveals a chemical shift difference between H41 and H391 of less than 0.5 ppm. However, the correlation between C23/C32 and H231/H321 reveals a chemical shift difference of *ca*. 1.5 ppm between H231 and H321. This is consistent with the packing arrangement of the NPX and PA molecules: H41 and H391 are in near identical environments in the two dimers, but H231 is packed closer to an oxygen atom (O1) than H321, which is near the C52 methyl group.

4.3.3 Hydrogen bonds

The ^{1}H spectrum of NPX-PA acquired with fast MAS of 60 kHz, Figure 56(i), shows comparable resolution with that of the 2NPX-NA spectrum presented by Ando *et al*.[46] It is likely that the chemical shifts of some of the protons are affected by intermolecular ring current effects due to the edge-to-face herringbone structure of NPX-PA, which is similar to that of pure NPX studied by Carignani *et al*.[50] The broad linewidths of ^{1}H SS-NMR can hinder the ability of SSNMR to distinguish between protons in hydrogen bonds of the same type, but there are many cocrystals/salts for which it is possible.[12,13,28,56] The NPX-PA cocrystal is one such case. Not all of the ^{1}H peaks could be assigned, but the peaks of the two protons involved in the synthon **F** dimers, H31 and H361, can be distinguished in Figure 56(i). The T_1 relaxation times of the hydrogen-bonded protons (roughly 10 s) are longer than those of other hydrogens (less than 1 s).

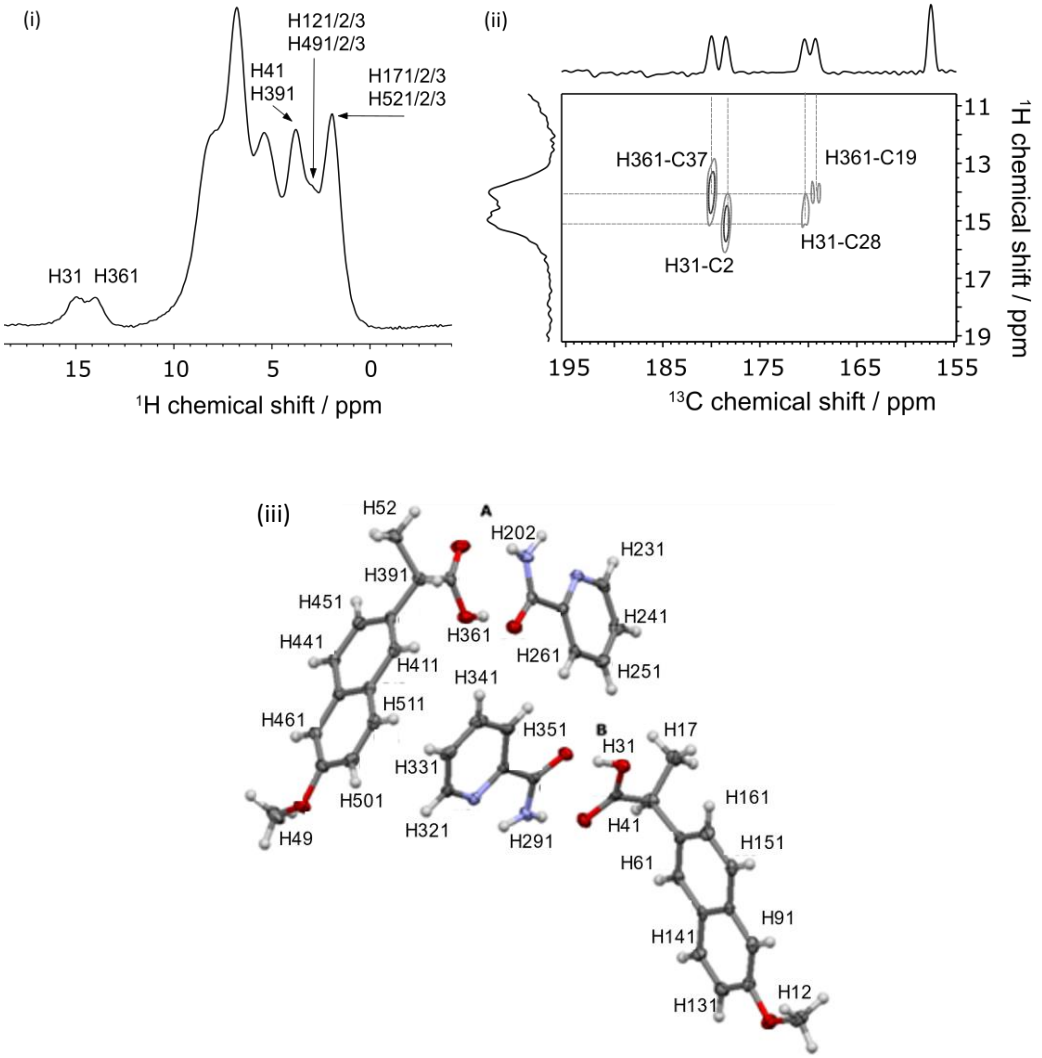

Figure 56 (i) ^{1}H spectrum of NPX-PA acquired at 60 kHz MAS with 96 transients. **(ii)** ^{1}H–^{13}C HETCOR spectrum acquired at 10 kHz MAS with a contact time of 1 ms. 64 t_1 increments were each acquired with 48 transients. A 20 s recycle delay was used for both spectra. **(iii)** Hydrogen atom labelling is given such that the hydrogen label is HX1, where X is the label of the directly bonded heavy atom.

The positioning of H31 and H361 within the hydrogen bonds is further investigated by comparing the experimental NMR data to predicted shift values. Optimisation of all atomic positions in the structure obtained from XRD results in a small heavy-atom RMSD between the optimised and original structures of 0.09 Å. However, the positions of the hydrogen atoms in the dimers are significantly displaced upon optimisation, Table 11. The average hydrogen displacement of 0.2 Å is of the order of the usual displacements observed by neutron diffraction compared to XRD data.[10] The hydrogen bonds become more symmetric and the difference between the two OH···O hydrogen bond lengths largely disappears.

However, the small difference of 1 ppm in the ^{1}H shifts for H31 and H361 observed in Figure 56(i) shows that the two dimers are chemically very similar but not identical. The difference in environment may be due to ring current effects.[50]

Table 11 Geometry of the hydrogen bonds in NPX-PA following geometry optimisation of all 184 atom positions.

Dimer	Hydrogen Bond	$d_{X—H}$ / Å	$d_{H\cdots A}$ / Å	$d_{X\cdots A}$ / Å	Bond angle / °
A	O36—H361···O18	1.05	1.49	2.534	169
	N20—H202···O38	1.03	1.89	2.897	166
B	O3—H31···O27	1.05	1.51	2.546	170
	N29—H291···O1	1.03	1.83	2.844	167

Two hydrogen bonding models can be proposed as alternatives to the XRD refinement, introduced in Figure 53 and shown in Figure 57(i). Model 1 corresponds to a salt where H31 and H361 are transferred across the hydrogen bonds onto the PA amide oxygen atoms, O27 and O18 respectively. Model 2 corresponds to an alternative cocrystal where H31 and H361 are transferred across their hydrogen bonds, along with H201 and H291 being transferred from the PA amide onto the NPX carboxylic acid, O38 and O1 respectively. The potential energy of the system calculated as a function of hydrogen position for both of these models is shown in Figure 57(ii). There is a single energy minimum in both cases corresponding to the structure determined from XRD following all atom optimisation. The curves both show a single steep minimum so the alternative positions for the hydrogen atoms are implausible, and temperature-dependent proton transfer over the hydrogen bonds is highly unlikely. The system could be classified as possessing strong asymmetric hydrogen bonds if the potential energy curve is compared with those shown previously in Figure 46. This contrasts with a previously reported system, 3,5-pyridinedicarboxylic acid, where an NMR crystallography study revealed the presence of quantum tunnelling of a proton across a hydrogen bond with a shallow double minimum in the potential energy surface.[8]

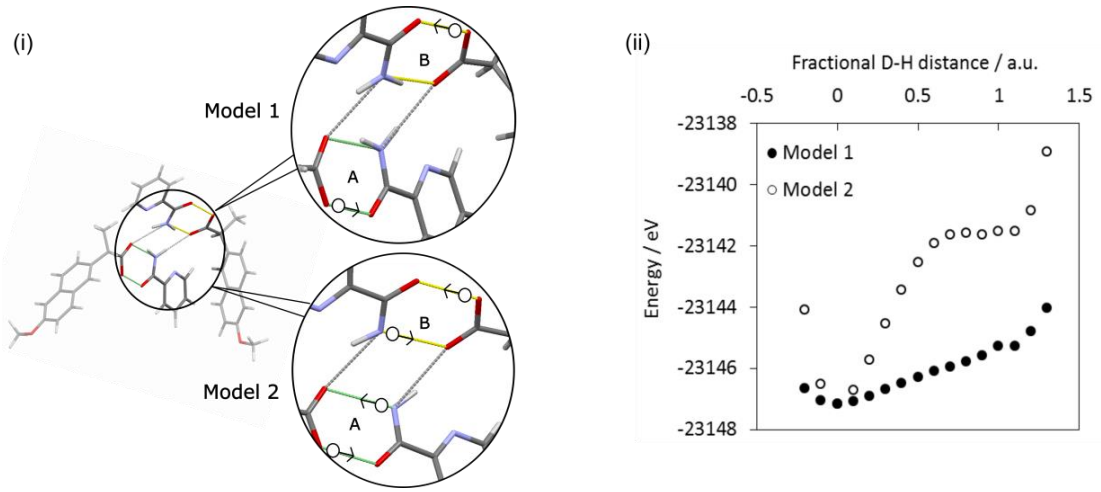

Figure 57 (i) Schematic showing hydrogen atoms that are moved in tandem to form Model 1 and Model 2. White circles indicate the initial position of the hydrogen atoms from the XRD structure and arrows indicate the direction in which the atoms were incrementally moved. **(ii)** Energy of NPX-PA as a function of hydrogen atom position. Fractional distance 0 corresponds to the DFT-optimised XRD position, while 1 corresponds to the hydrogen atom being at the equivalent position on the other side of the hydrogen bond.

In addition to the energy considerations, only the NMR parameters calculated from the NPX-PA geometry-optimised XRD structure are compatible with the experimental data, Figure 58. In particular, the RMSDs between experimental and calculated ^{13}C shifts, Table 12, are inconsistent with a valid structure for Model 1 or Model 2, being greater than 2 ppm.[57] The predicted ^{1}H spectra for Model 1 and Model 2 also poorly match the experimental ^{1}H spectrum as the amide protons, H202/H291, are predicted to give rise to peaks that are not present experimentally; 11 ppm in Model 1 and 14 ppm in Model 2. The calculated spectra were referenced using $\delta_{iso} = \sigma_{ref} - \sigma_{iso}$ where σ_{iso} is the CASTEP-calculated shielding value and σ_{ref} was calculated to equate the average calculated average experimental shifts, Table 12.

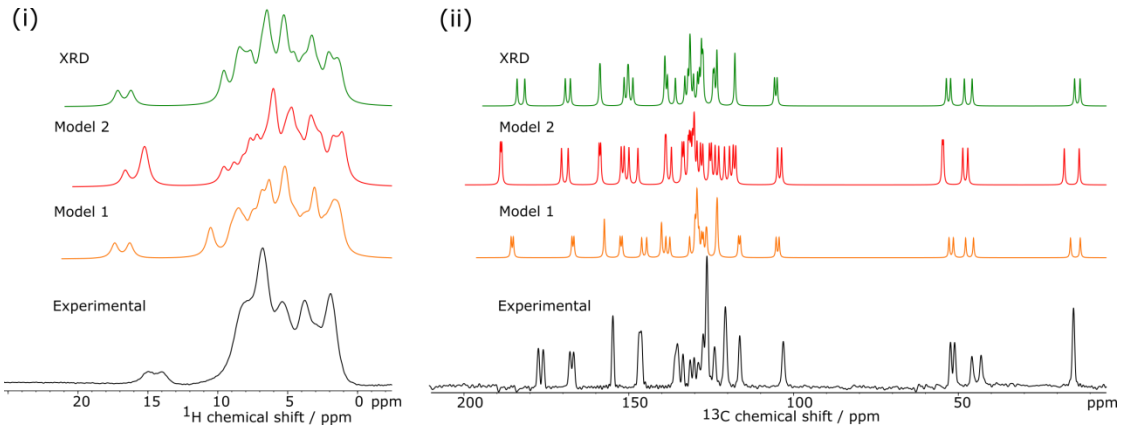

Figure 58 Comparison of experimental and calculated **(i)** ^{1}H and **(ii)** ^{13}C spectra for the three hydrogen bonding models. Calculated spectra were referenced as described previously and modelled with 300 Hz and 50 Hz Lorentzian line broadening for ^{1}H and ^{13}C respectively.

Table 12 σ_{ref} values used to reference the CASTEP-calculated ^{1}H and ^{13}C spectra, and the total RMSD in ^{13}C shifts for the three hydrogen bonding models.

Structure	^{1}H σ_{ref} / ppm	^{13}C σ_{ref} / ppm	^{13}C RMSD / ppm
XRD	31	170	1.6
Model 1	30	169	2.4
Model 2	30	171	2.7

^{1}H–^{1}H DQ/SQ correlation peaks, Figure 59, were assigned after measuring the ^{1}H–^{1}H proximities in the geometry optimised structure, Table 13. Four correlations are visible to the hydrogen-bonded protons, H31 and H361. The CASTEP-calculated DQ/SQ correlations to H31 and H361 for the XRD-refined structure are shown in blue and these are compatible with the experimental spectrum. It is noted that two strong correlations between H31–H41 and H31–H251 are not visible experimentally despite the distances being relatively small at 2.56–2.57 Å, as indicated by the black arrow in Figure 59. This may be due to dipolar truncation.[58,59] Neither Model 1 nor Model 2 result in a simulated DQ/SQ spectrum that is compatible with the experimental spectrum. In particular, a strong correlation to H361 is predicted in Model 1 at a SQ frequency of *ca*. 10 ppm that is not observed experimentally. There is also no correlation predicted for H361 at a DQ frequency less than 2 ppm, but the H361–H391 correlation is clearly visible experimentally. Model 2 results in a predicted correlation at a DQ shift of *ca*. 8 ppm, which is also not visible experimentally. Therefore, the comparison of the simulated 1D and 2D SSNMR spectra with experimental data allows the protonation state of NPX-PA to be validated; the data converge to the only viable structure being that determined from SCXRD data.

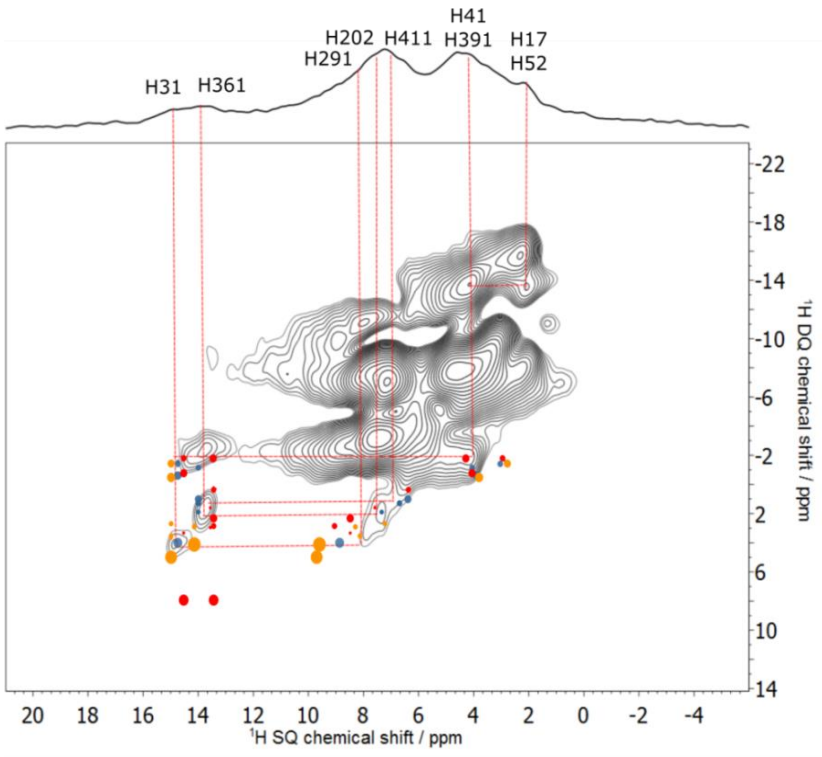

Figure 59 ^{1}H–^{1}H DQ/SQ spectrum of NPX-PA acquired at 60 kHz MAS using 8 rotor cycles in the recoupling period of the BABA sequence and an evolution time of 1 µs. 128 t_1 increments were acquired with 16 transients per increment and a 20 s recycle delay was used. 200 Hz Gaussian line broadening was applied prior to FT. Simulated DQ/SQ spectra were produced using MagresView[60] and the point sizes are determined by the magnitude of the dipolar coupling, with a cut-off of 3.5 Å. The three simulated spectra were overlaid and then scaled as one to overlay with the experimental spectrum by eye. The structure determined by SCXRD is in blue, Model 1 is in orange and Model 2 is in red. Hydrogen atom labelling given in Figure 56(iii).

Table 13 Selected ^{1}H–^{1}H distances to the hydrogen-bonded protons, H31 and H361, up to 3.0 Å, measured in the optimised structures for the XRD structure and Model 2.

Correlation	XRD[a] / Å	Model 1 / Å	Model 2[b] / Å
H31 – H291	2.39	2.29	2.21
H361 – H202	2.42	2.23	2.18
H31 – H41	2.56	2.64	2.53
H31 – H251	2.57	2.75	2.81
H361 – H411	2.59	2.72	2.68
H31 – H173	2.80	2.82	2.79
H361 – H391	2.84	2.69	2.68
H361 – H341	2.85	> 3.00	> 3.00
H31 – H351	2.89	> 3.00	> 3.00
H31 – H261	> 3.00	2.62	2.72
H361 – H351	> 3.00	2.96	> 3.00
H17/H52 – H41/391[c]	2.68	2.68	2.68

[a] All atom positions optimised, [b] only hydrogen atom positions optimised, [c] averaged over H171/2/3 and H41/2/3.

4.4 Database systems

A CSD search* for the Model 1 and Model 2 fragments yields only three structures[1,61,62] for Model 1 and none for Model 2. The comparative lack of experimental structures containing these alternative hydrogen bonding models implies that they are not energetically favourable, in keeping with the steep potential energy curves of NPX-PA in Figure 57(ii). All three Model 1 structures are reported to exhibit proton disorder or migration, so the potential energy curves were calculated to compare with the potential energy of NPX-PA, Figure 60.

* The search was subject to the following restrictions: 'only organics' and '3D coordinates determined'.

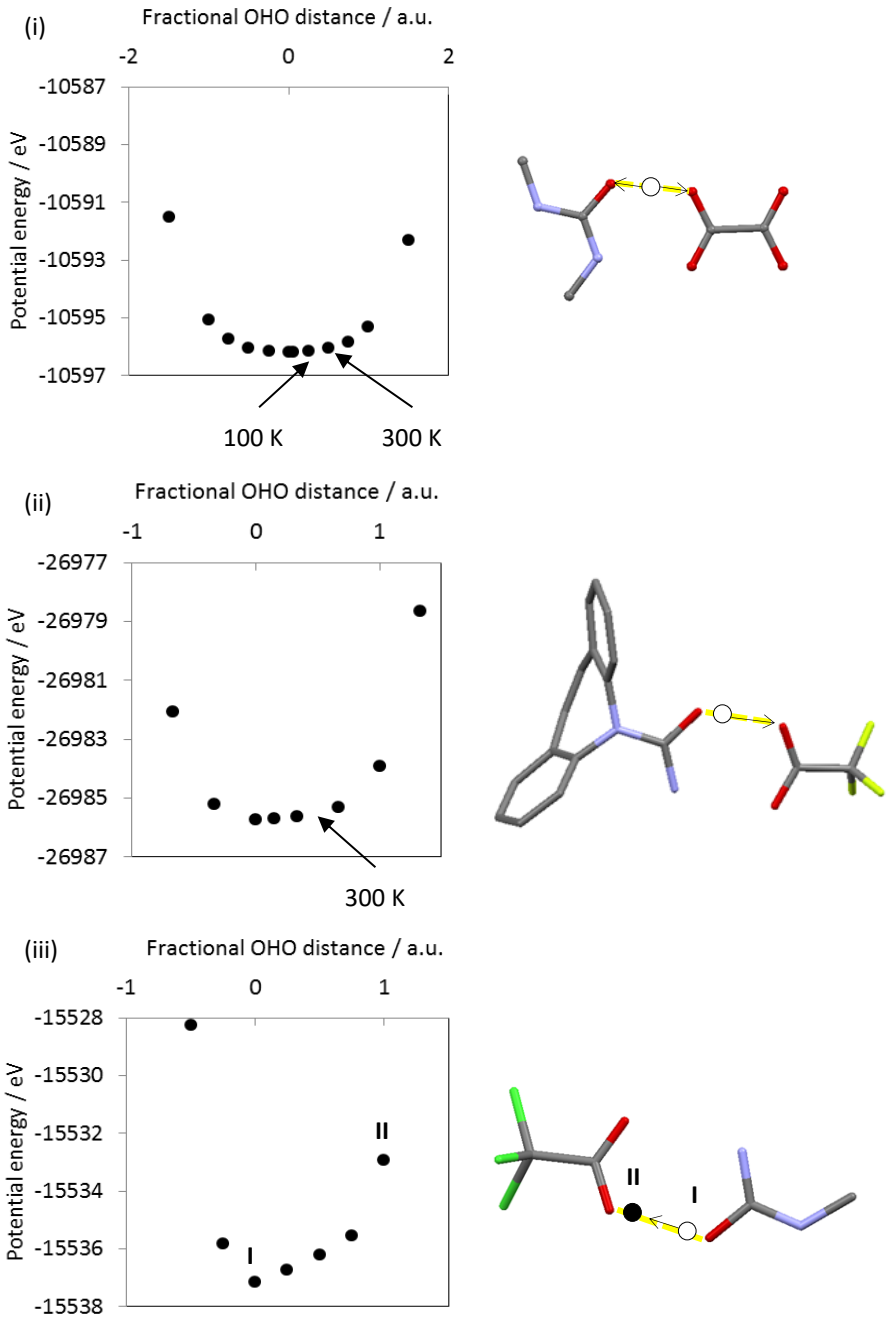

Figure 60 Potential energy curves of **(i)** DMU-OA, **(ii)** CBZ-TFA and **(iii)** MU-TCA. Schematics of the structures are given on the right, with the hydrogen atom of interest drawn as a white circle and the arrows denoting the direction of movement over the hydrogen bond. Fractional coordinate 0 corresponds to the optimised SCXRD structure at 100 K. The proton position of CBZ-TFA at 300 K is indicated in **(ii)**. The two positions of the disordered proton in MU-TCA are indicated in **(iii)**.

The first, a dimethylurea-oxalic acid complex DMU-OA, was found to exhibit temperature-dependent proton migration over a short OH···O hydrogen atom in the carboxylic acid–amide dimer similar to synthon **F** ($d_{O\cdots O}$ = 2.457 Å) by extensive XRD and neutron diffraction studies.[61] All atom geometry optimisation of DMU-OA results in the hydrogen atom being placed centrally in the OH···O hydrogen bond, rather than being associated with the OA as is observed experimentally at 300 K, Figure 60(i). This corresponds to a change in the OA d_{O-H} of 0.10 Å compared to the SCXRD structure refined at 100 K, and 0.17 Å compared to the structure refined at 300 K. The potential energy curve is broad and flat in Figure 60(i), indicating that there is no significant energy penalty for the hydrogen atom to be in any particular position along the short hydrogen bond. This is in keeping with the experimental evidence of proton migration from variable temperature neutron diffraction experiments.[61] It must be noted that the precise nature of the broad flat potential energy profile may not be described reliably by DFT in this case, since temperature is likely to be important.

The second system, carbamazepine trifluoroacetate CBZ-TFA, was found to be in an intermediary state between a neutral cocrystal and a salt.[1] The trifluoro group was also refined as being disordered in a similar manner to methyl group disorder. Optimisation of both the high and low temperature SCXRD structures of CBZ-TFA resulted in the hydrogen bonds both converging to the low temperature hydrogen position, Figure 60(ii). The potential energy well is shallow and slightly asymmetric; the hydrogen positions refined at 100 K and 300 K have energies that are less than 1 eV apart, so it is unsurprising that some movement is possible over this short hydrogen bond. Proton migration has not been extensively investigated in this system; only two temperatures have been used to collect XRD data. Therefore, it is possible that SSNMR data might provide further insight into the possibility of proton migration in this system. Very low temperatures are generally required to slow proton dynamics enough to observe an effect in the ^{1}H T_1 relaxation times. For example, proton tunnelling in short hydrogen bonds in 3,5-pyridine dicarboxylic acid is observed *via* ^{1}H T_1 relaxation measurements at temperatures below 71 K, and classical migration is observed at temperatures of *ca*. 100 K.[8] Such temperatures are not accessible with the available SSNMR setup.

The final system, N-methylurea trichloroacetic acid MU-TCA, was refined from SCXRD to contain proton disorder over two sites in a 0.75:0.25 ratio in favour of a salt form ($d_{O\cdots O}$ = 2.524 Å).[62] Optimisation of MU-TCA required the disorder to be separated into two ordered structures; **I** corresponding to the salt and **II** corresponding to the neutral

cocrystal. Both **I** and **II** converged to **I** upon optimisation. The potential energy well is steep and asymmetric, Figure 60(iii), and site **II** is *ca*. 4 eV higher in energy than **I**. Therefore, it is not initially clear why the experimental structure should contain proton disorder; 4 eV is roughly two orders of magnitude higher than k_BT at room temperature. However, upon closer inspection the analysis of the crystallographic data for MU-TCA raised some questions. Firstly, the proton positions of **I** and **II** appear to have been constrained rather than freely refined: no standard uncertainties are given on the $d_{O–H}$ values. This suggests the experimental data was made to fit the model, rather than the other way around. Additionally, analysis of the experimentally determined $d_{C–O}$ lengths, Figure 61(i), indicates that hydrogen position **I** is more reasonable. The MU $d_{C–O}$ length lies within the expected range for a single bond, Figure 61(ii), and the two TCA $d_{C–O}$ lengths lie within the range for double bonds. If the proton was truly disordered then one might expect all of the $d_{C–O}$ lengths to lie centrally between the usual ranges for single and double bonds, red star in Figure 61(ii).

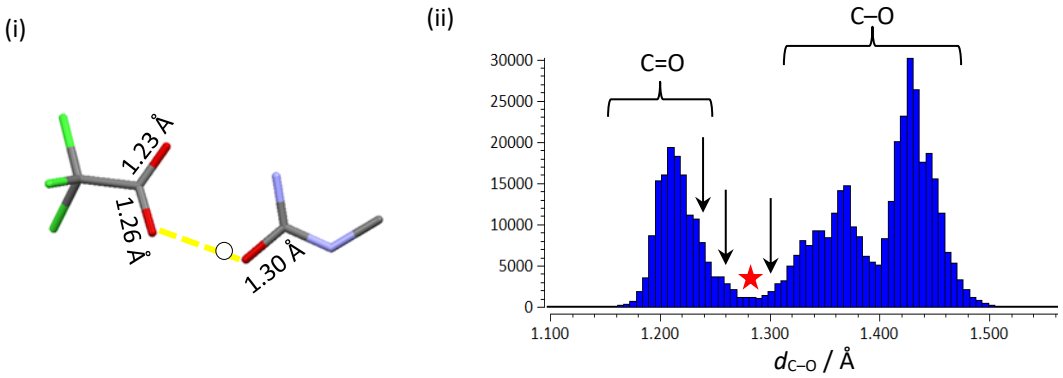

Figure 61 (i) Experimental MU-TCA $d_{C–O}$ lengths with the hydrogen-bonded proton in position **I**. **(ii)** Histogram of $d_{C–O}$ lengths found in the CSD* with the three MU-TCA lengths indicated by black arrows. The red star indicates the expected region for the bond lengths if the proton were disordered between **I** and **II**.

* The search was subject to the following restrictions: 'only organics', '3D coordinates determined', 'no errors', 'not disordered', 'not polymeric' and 'no powder structures' with the R factor ≤ 0.05.

4.5 Conclusions

This chapter highlights how an NMR crystallography approach is useful to both distinguish between chemically identical hydrogen bonds and to refine the positions of hydrogen atoms in hydrogen bonds, using the crystal structure of a new NPX-PA cocrystal as an example. NPX-PA contains two crystallographically unique NPX and PA molecules that form two COOH···CONH dimers. These two dimers are distinguishable by fast MAS ^{1}H SSNMR, indicating that they are in unique environments, though the difference in chemical shift is small at 1 ppm. This demonstrates the sensitivity of NMR spectra to hydrogen positioning. DFT geometry-optimisation refines the positions of the hydrogen atoms in the dimers of NPX-PA, resulting in an excellent agreement with experimental ^{13}C and ^{1}H SSNMR. The calculated potential energy of NPX-PA as a function of hydrogen positioning validates the overall positioning determined by XRD, confirming that NPX-PA is a cocrystal and not a salt. This is also supported by 2D SSNMR experiments. The steepness of the potential energy well is likely to preclude temperature-dependent proton transfer, so any possiblity of proton disorder in NPX-PA is conclusively rejected. This contrasts to two systems found in the CSD that possess similar synthons to NPX-PA, DMU-OA and CBZ-TFA, whose potential energy curves are broad and shallow, in keeping with experimentally observed proton migration. The NPX-PA potential energy curve is more similar to that of a final system found in the CSD with similar hydrogen bonding, MU-TCA, which is refined with proton disorder. However, questions are raised about the quality of this structure refinement and it is likely that MU-TCA, like NPX-PA, has a single favourable hydrogen position.

4.6 References

(1) A. R. Eberlin, M. D. Eddleston and C. S. Frampton, *Acta Crystallogr. C*, **2013**, 69, 1260.
(2) S. L. Childs, G. P. Stahly and A. Park, *Mol. Pharm.*, **2007**, 4, 323.
(3) S. Mohamed, D. A. Tocher, M. Vickers, P. G. Karamertzanis and S. L. Price, *Cryst. Growth Des.*, **2009**, 9, 2881.
(4) US-FDA*, Guidence for Industry: Regulatory Classification of Pharmaceutical Co-Crystals*, http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryi%20nformation/guidances/ucm281764.pdf, accessed 2016.
(5) E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza *et al.*, *Pure Appl. Chem.*, **2011**, 83, 1619.
(6) P. B. White and M. Hong, *J. Phys. Chem. B*, **2015**, 119, 11581.
(7) S. J. Ford, O. J. Delamore, J. S. Evans, G. J. McIntyre, M. R. Johnson and I. Radosavljevic Evans, *Chem. Eur. J.*, **2011**, 17, 14942.
(8) I. Frantsuzov, S. J. Ford, I. Radosavljevic Evans, A. J. Horsewill, H. P. Trommsdorff and M. R. Johnson, *Phys. Rev. Lett.*, **2014**, 113, 018301.
(9) J. Powell, K. Kalakewich, F. J. Uribe-Romo and J. K. Harper, *Phys. Chem. Chem. Phys.*, **2016**, 18, 12541.
(10) *"Structural Inorganic Chemistry*", 5th ed., A. F. Wells (Eds.), Oxford University Press: Oxford, UK, 1984.

(11) S. Mohamed, D. A. Tocher and S. L. Price, *Int. J. Pharm.*, **2011**, 418, 187.
(12) R. Gobetto, C. Nervi, M. R. Chierotti, D. Braga, L. Maini, F. Grepioni, R. K. Harris and P. Hodgkinson, *Chemistry*, **2005**, 11, 7461.
(13) F. G. Vogt, J. S. Clawson, M. Strohmeier, A. J. Edwards, T. N. Pham and S. A. Watson, *Cryst. Growth Des.*, **2009**, 9, 921.
(14) K. Maruyoshi, D. Iuga, O. N. Antzutkin, A. Alhalaweh, S. P. Velaga and S. P. Brown, *Chem. Commun.*, **2012**, 48, 10844.
(15) D. V. Dudenko, J. R. Yates, K. D. M. Harris and S. P. Brown, *CrystEngComm.*, **2013**, 15, 8797.
(16) A. S. Tatton, T. N. Pham, F. G. Vogt, D. Iuga, A. J. Edwards and S. P. Brown, *Mol. Pharm.*, **2013**, 10, 999.
(17) R. Koike, K. Higashi, N. Liu, W. Limwikrant, K. Yamamoto and K. Moribe, *Cryst. Growth Des.*, **2014**, 14, 4510.
(18) D. Ludeker and G. Brunklaus, *Solid State Nucl. Magn. Reson.*, **2015**, 65, 29.
(19) D. Luedeker, R. Gossmann, K. Langer and G. Brunklaus, *Cryst. Growth Des.*, **2016**, 16, 3087.
(20) O. D. Putra, T. Yoshida, D. Umeda, K. Higashi, H. Uekusa and E. Yonemochi, *Cryst. Growth Des.*, **2016**, 16, 5223.
(21) M. R. Chierotti and R. Gobetto, *Chem. Commun.*, **2008**, 14, 1621.
(22) R. K. Harris, P. Jackson, L. H. Merwin, B. J. Say and G. Hagele, *J. Chem. Soc. Farad. Trans.*, **1988**, 84, 3649.
(23) S. Sharif, G. S. Denisov, M. D. Toney and H. H. Limbach, *J. Am. Chem. Soc.*, **2007**, 129, 6313.
(24) M. Khan, V. Enkelmann and G. Brunklaus, *J. Am. Chem. Soc.*, **2010**, 132, 5254.
(25) M. Bauer, R. K. Harris, R. C. Rao, D. C. Apperley and C. A. Rodger, *J. Chem. Soc., Perkin Trans. 2*, **1998**, 3, 475.
(26) A. S. Tatton, T. N. Pham, F. G. Vogt, D. Iuga, A. J. Edwards and S. P. Brown, *CrystEngComm.*, **2012**, 14, 2654.
(27) S. P. Brown, *Solid State Nucl. Magn. Reson.*, **2012**, 41, 1.
(28) M. Sardo, S. M. Santos, A. A. Babaryk, C. Lopez, I. Alkorta, J. Elguero, R. M. Claramunt and L. Mafra, *Solid State Nucl. Magn. Reson.*, **2015**, 65, 49.
(29) *DrugBank*, www.drugbank.ca/drugs/DB00788, accessed July 2016.
(30) H. F. Hill, A. G. Hill, A. G. Mowat, B. M. Ansell, J. A. Mathews, M. H. Seifert, J. M. Gumpel and G. A. Christie, *Ann. Rheum. Dis.*, **1974**, 33, 12.
(31) *"AHFS Drug Information 2009*", G. K. McEvoy (Eds.), American Society of Health System Pharmacists: Bethesda, MD, 2009.
(32) T. Takagi, C. Ramachandran, M. Bermejo, S. Yamashita, L. X. Yu and G. L. Amidon, *Mol. Pharm.*, **2006**, 3, 631.
(33) M. E. Davis and M. E. Brewster, *Nat. Rev. Drug Discov.*, **2004**, 3, 1023.
(34) M. Alleso, N. Chieng, S. Rehder, J. Rantanen, T. Rades and J. Aaltonen, *J. Control. Release*, **2009**, 136, 45.
(35) K. Ravikumar, S. S. Rajan and V. Pattabhi, *Acta Crystallogr. C*, **1985**, 41, 280.
(36) G. R. Desiraju, *Angew. Chem.*, **1995**, 34, 2311.
(37) R. A. E. Castro, J. D. B. Ribeiro, T. M. R. Maria, M. R. Silva, C. Yuste-Vivas, J. Canotilho and M. E. S. Eusebio, *Cryst. Growth Des.*, **2011**, 11, 5396.
(38) C. Neurohr, M. Marchivie, S. Lecomte, Y. Cartigny, N. Couvrat, M. Sanselme and P. Subra-Paternault, *Cryst. Growth Des.*, **2015**, 15, 4616.
(39) D. R. Weyna, T. Shattock, P. Vishweshwar and M. J. Zaworotko, *Cryst. Growth Des.*, **2009**, 9, 1106.

(40) H. H. D. Buschmann, C. L. Solà, B. J. Benet and B. J. C. Cerón, *Co-crystals of duloxetine and co-crystal formers for the treatment of pain*, 2009, European Patent Office: EP20080384009.
(41) C. L. Solà, B. J. C. Cerón, B. J. Benet and H. H. Buschmann, *Cocrystals of tramadol and NSAIDs*, 2010, European Patent Office: 09736862.5.
(42) K. Manoj, R. Tamura, H. Takahashi and H. Tsue, *CrystEngComm.*, **2014**, 16, 5811.
(43) X. J. Yuan, J. G. Li, Y. Q. Tian, G. H. Lee, X. M. Peng, R. G. Zhu and X. Z. You, *Tetrahedron: Asymmetry*, **2001**, 12, 3015.
(44) A. Tilborg, G. Springuel, B. Norberg, J. Wouters and T. Leyssens, *CrystEngComm.*, **2013**, 15, 3341.
(45) N. Tumanova, N. Tumanov, K. Robeyns, Y. Filinchuk, J. Wouters and T. Leyssens, *CrystEngComm.*, **2014**, 16, 8185.
(46) S. Ando, J. Kikuchi, Y. Fujimura, Y. Ida, K. Higashi, K. Moribe and K. Yamamoto, *J. Pharm. Sci.*, **2012**, 101, 3214.
(47) *"NMR Crystallography*", R. K. Harris, R. E. Wasylishen and M. J. Duer (Eds.), John Wiley & Sons Ltd: West Sussex, UK, 2009.
(48) M. R. Chierotti and R. Gobetto, *CrystEngComm.*, **2013**, 15, 8599.
(49) J. S. Stevens, S. J. Byard, C. C. Seaton, G. Sadiq, R. J. Davey and S. L. Schroeder, *Phys. Chem. Chem. Phys.*, **2014**, 16, 1150.
(50) E. Carignani, S. Borsacchi, J. P. Bradley, S. P. Brown and M. Geppi, *J. Phys. Chem. C*, **2013**, 117, 17731.
(51) K. M. Burgess, F. A. Perras, A. Lebrun, E. Messner-Henning, I. Korobkov and D. L. Bryce, *J. Pharm. Sci.*, **2012**, 101, 2930.
(52) J. Czernek, *Chem. Phys. Lett.*, **2015**, 619, 230.
(53) A. D. Bond, C. Cornett, F. H. Larsen, H. Y. Qu, D. Raijada and J. Rantanen, *Cryst. Growth Des.*, **2013**, 13, 3665.
(54) H. H. Buschmann, C. L. Solà, B. J. Benet, B. J. C. Cerón and A. J. Ramirez, *Co-crystals of duloxetin and naproxen*, 2013, European Patent Office: EP20090749634.
(55) H. H. Buschmann, C. L. Solà, B. J. Benet and B. J. C. Cerón, *Co-crystals of tramadol and NSAIDs*, 2010, European Patent Office: EP20090736862.
(56) R. K. Harris, P. Hodgkinson, V. Zorin, J. N. Dumez, B. Elena-Herrmann, L. Emsley, E. Salager and R. S. Stein, *Magn. Reson. Chem.*, **2010**, 48 Suppl 1, 103.
(57) C. M. Widdifield, H. Robson and P. Hodgkinson, *Chem. Commun.*, **2016**, 52, 6685.
(58) P. Hodgkinson and L. Emsley, *J. Magn. Reson.*, **1999**, 139, 46.
(59) 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.
(60) S. Sturniolo*, MagresView: visualisation and processing of NMR data calculated in CASTEP*, www.ccpnc.ac.uk/magresview/magresview/magres_view.html, accessed 2016.
(61) A. O. Jones, M. H. Lemee-Cailleau, D. M. Martins, G. J. McIntyre, I. D. Oswald, C. R. Pulham, C. K. Spanswick, L. H. Thomas and C. C. Wilson, *Phys. Chem. Chem. Phys.*, **2012**, 14, 13273.
(62) A. J. Rybarczyk-Pirek, *Struct. Chem.*, **2012**, 23, 1739.

Chapter 5: NMR crystallography to elucidate the nature of molecular disorder

Based on work published in "*A furosemide-isonicotinamide cocrystal: an investigation of properties and extensive structural disorder*", H. E. Kerr, L. K. Softley, K. Suresh, A. Nangia, P. Hodgkinson, I. R. Evans, CrystEngComm., 2015, 35, 6707–6715, published by the Royal Society of Chemistry. Sections 5.2 and 5.4 are lightly adapted from the publication.

5.1 Furosemide in the pharmaceutical industry

Furosemide (FS), Figure 62, is an API with potent and fast-onset diuretic properties, used in the treatment of hypertension and edema. The mechanism of action involves the inhibition of water reabsorption in the nephron of the kidney by competitive inhibition at the chloride binding site in a sodium-potassium-chloride cotransporter.[1] It is commercially available in liquid form for intravenous or oral administration, and also as a tablet. However, the solid form has a low aqueous solubility (6 mg dm^{-3}) and low permeability so the bioavailability is limited. The pharmokinetic profile is variable between patients[2] so prescribed doses must be high:[3] up to 80 mg. This increases the likelihood of side effects such as dehydration, electrolyte depletion, rapid weight loss and even permanent hearing damage.[1] This poor solubility is probably linked to the strong hydrogen bonding network in the crystalline form, which resists facile solvation. As a result, there have been attempts at improving the solubility by preparing non-crystalline forms of FS. One method utilises hydrophilic mesoporous materials, where the large surface area introduced by the presence of large pores allows FS to absorb onto the surface.[4] The FS-surface interactions are weak so fast drug delivery is possible at the target sites; the stomach and small intestine. Recently, novel hydrated sodium and potassium salts of FS have been found to have aqueous solubility up to four orders of magnitude higher than the commercial form of

FS. However, these salts are only stable with respect to the starting components for 1–2 weeks in humid conditions[5] so they are not suitable for pharmaceutical applications.

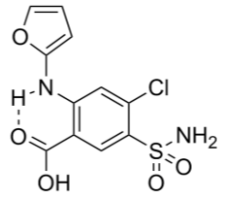

Figure 62 Molecular structure of furosemide.

Numerous cocrystals and solvates of FS have been synthesised with a wide range of coformers, some displaying polymorphism.[3,6-11] Of the reported cocrystals, four display enhanced solubility compared to commercial FS and are stable in slurry conditions for over 48 hours.[3,12] The cocrystallisation method shows promise to improve the pharmokinetic properties of FS, whilst simultaneously maintaining suitable stability and synthetic reproducibility.

5.2 Furosemide and its known cocrystals

The first crystal structure of FS with 3D coordinates[13] was published in 1978. An intramolecular NH···O bond gives rigidity to the central portion of the molecule, Figure 62, but there is conformational flexibility about the furan ring and sulphonamide torsion angles. FS is also capable of forming a variety of hydrogen bonds through both the carboxylic acid and the sulphonamide group, and there is potential for halogen bonding to the chlorine atom. Therefore, it is unsurprising that FS exists in three different polymorphic forms (**I**, **II** and **III**), which have all been characterised by SCXRD.[14]

Similar hydrogen bonding arrangements can be found in all three polymorphs. In particular synthon **A** and synthon **J** in Figure 63 are ubiquitous. Note that the synthon labelling here is the same as Figure 49 in Chapter 4 for synthons **A**, **B**, **E**, and **F**, but the identities of the other synthons differ. Polymorphs **I** and **III** are both triclinic, crystallising in space group $P\bar{1}$, but form **I** has $Z' = 2$ compared to form **II** which has $Z' = 1$. Form **II** is monoclinic with $Z' = 1$. Synthon **A** is present in all three forms but, instead of following the layered structure of form **I** and **III**, form **II** has a helical synthon **J** motif. The sulphonamide and furan ring torsion angles differ in all three polymorphs, highlighting the conformational flexibility. Form **II** and **III** possess higher aqueous solubility than the commercial form **I**, but these polymorphs are

metastable and so are not well suited to formulation where stability under storage is necessary.[14]

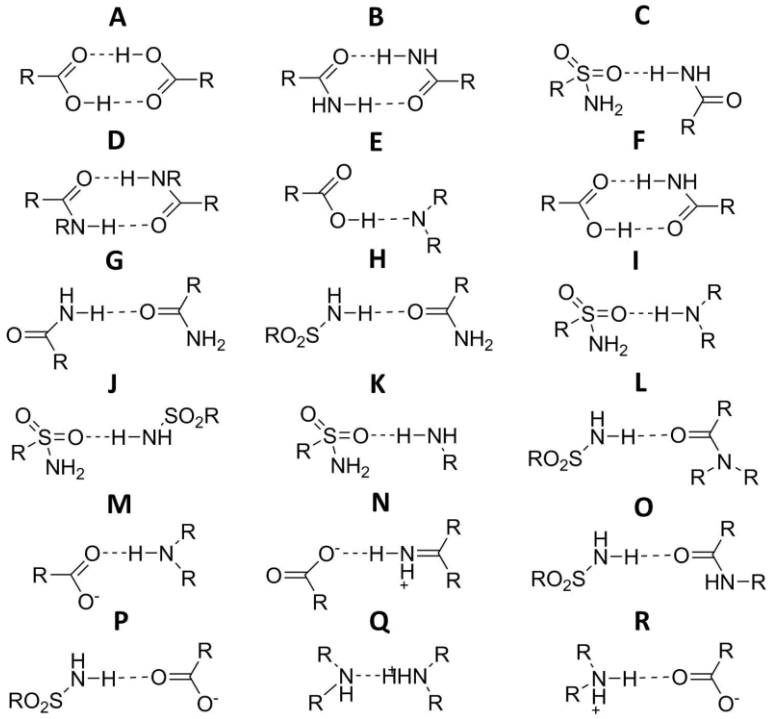

Figure 63 Synthons present in pure FS and the cocrystals of FS.

Many coformers have been used to synthesise cocrystals of FS, Figure 64, and 12 structures been reported in the literature, not including polymorphs,[7,10,14] solvates[8,9,11,12], hydrates[5,15] or the drug-drug cocrystal with the painkiller pentoxifylline.[9] The synthons relevant to these cocrystals are labelled in Table 14, referring to Figure 63 for synthon labels. FS-cytosine (CYT), FS-4-aminopyridine (APD) and FS-PPZ are distinct due to proton transfer in synthons **N**, **M**, **Q** and **R** resulting in salts rather than cocrystals. Both FS-CYT and FS-CAF crystallise in the space group $P\bar{1}$, while the cocrystals of FS with pyridine carboxamide coformers, NA and INA, all crystallise in $P2_1/n$. Changing the stoichiometry from FS-INA to 2FS-INA alters the number of hydrogen bonding groups and so the synthons differ. Most of the aforementioned cocrystals contain synthon **E**, except FS-CYT where the proton is transferred resulting in synthon **M**. The only cocrystal to exhibit the carboxylic acid dimer, synthon **A**, which is prevalent in pure FS, is FS-*p*-aminobenzoic acid (PABA). These dimers are held together by a weak interaction, synthon **K**.

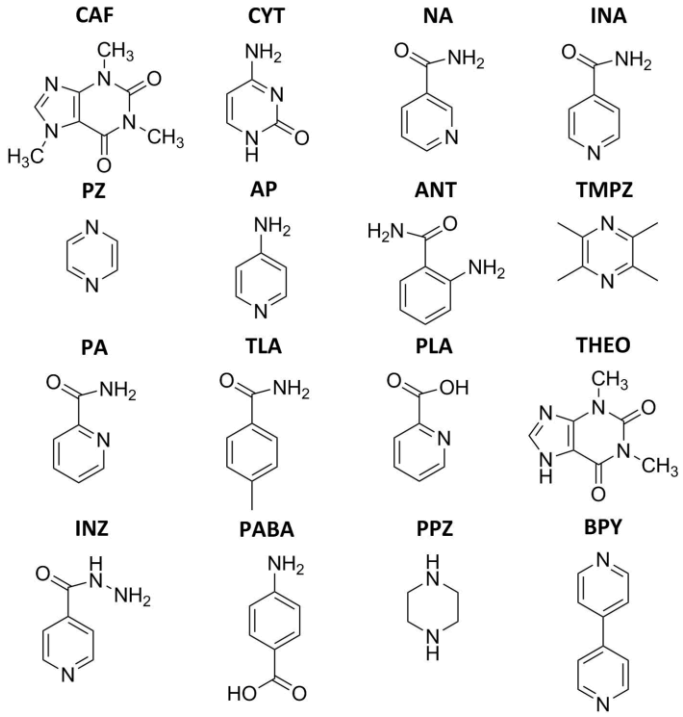

Figure 64 Small molecule coformers of the cocrystals of FS present in the literature, not including solvates or hydrates.

Table 14 Synthons present in the cocrystals of FS whose crystal structures are reported in the literature. Polymorphs, solvates and hydrates are not included.

Cocrystal	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
FS-CAF[a]					*				*			*						
FS-CYT[a]				*									*	*	*			
FS-NA[a]					*		*			*								
FS-INA[a]		*	*		*			*										
2FS-INA[b]			*		*	*				*								
FS-INA-N-oxide[c]						*						*						
FS-AP[d]												*	*					
FS-ANT[e]						*					*							
2FS-TMPZ[e]					*					*								
FS-PABA[a]	*									*	*							
FS-PPZ[e]										*						*	*	*
2FS-PPZ[e]																*		*
FS-BPY[d]					*					*								

[a] From the structures reported by Goud[3] *et al.*, [b] from the structure reported by Harriss[6] *et al.*, [c] from the structure reported by Goud[15] *et al.*,[d] from the structures reported by Sangtani[7] *et al.*, [e] from the structures reported by Banik[12] *et al*.

5.2.1 Disorder in solid furosemide forms

The conformational flexibility of FS can be attributed to its rich supramolecular chemistry, but it also results in the presence of disorder. The 293 K SCXRD structure of pure FS form **II** contains disorder of the furan ring over two positions at 0.25:0.75 occupancies, differing by an angle of 172˚ between the ring planes. Form **I** has also been reported to possess similar disorder,[16] but more recent analysis has shown that the structure is better described as being a doubled unit cell containing two unique FS molecules with different furan ring torsion angles.[14,17] This furan ring disorder is observed in several FS cocrystals including FS-CAF,[3] FS-NA hydrate,[10] FS-INA-N-oxide,[15] and 2FS-INA.[6] The nature of the furan disorder in FS-CAF has been described as dynamic because the low temperature SCXRD structure (100 K) could be solved with disorder over two orientations with occupancies of 0.35:0.65, while the structure at 298 K was too disordered for the furan ring to be successfully refined with split sites.[3] The nature of the disorder is not discussed in the other cocrystals.

The disorder and dynamics in forms **I** and **II** of FS have been investigated by SSNMR and NQR to provide information complementary to the SCXRD data.[18,19] Form **I** is clearly distinguishable from form **II** in ^{35}Cl NQR by the presence of two NQR lines in the spectrum corresponding to the two crystallographically unique chlorine atoms.[18] The ^{35}Cl NQR lineshape of FS form **II** is broad (104 Hz FWHM) over a large temperature range, reported to be due to the presence of static disorder.[18] The SCXRD structure of form **II** at 298 K supports the presence of disorder, but repeated SCXRD experiments at 100 K show only a single, ordered, orientation of the ring.[14] This is more in keeping with previous reports of FS-CAF, where furan ring dynamics are partially frozen out at low temperature.

The conformational flexibility about the C–S bond of the sulphonamide group of FS has been noted in previous reports.[11,14] However, disorder of the whole sulphonamide group is not commonly observed in FS cocrystals. This is probably due to the propensity of this group to hydrogen bond, as evidenced by a systematic study of sulphonamide – pyridine-N-oxide synthons.[15] However, there has been found to be dynamics of the NH_2 protons in the sulphonamide group. SSNMR ^{1}H T_1 relaxation times have been analysed to propose a relaxation mechanism for form **I**, consisting of temperature-dependent proton transfer over the short OH···O hydrogen bond in synthon **A**, as well as proton jumps in the sulphonamide NH_2 group.[19] NQR ^{35}Cl T_{1Q} relaxation times show evidence of a reorientation mechanism in FS form **II**, ascribed to NH_2 jumps as reported in form **I**. The fitting of SSNMR ^{1}H T_1 relaxation times yields a similar E_a value to that of the T_{1Q} relaxation times.[18]

5.2.2 Furosemide-isonicotinamide

The structure of the cocrystal 2FS-INA has recently been reported.[6] It exhibits complex disorder at 120 K, affecting both of the crystallographically unique FS molecules (**FS1** and **FS2** in Figure 65). The furan ring on **FS2** is disordered, as seen previously in other solid forms of FS, and it is modelled with split sites over two orientations with 0.67(2):0.33(2) occupancies. A model using single atomic sites is also satisfactory, albeit with elongated anisotropic displacement parameters perpendicular to the plane of the ring, discussed later. The sulphonamide group on **FS1** is disordered over two positions with 0.49(1):0.51(1) occupancies, differing by an angle of 32˚ about the C–S bond. This is the first time disorder of the whole sulphonamide group of FS has been observed. A final interesting feature is a short hydrogen bond between the carboxylic acid of **FS1** and the ring nitrogen atom on INA, $d_{O\cdots N}$ = 2.546 Å. This hydrogen bond is similar in length to a previously reported system, 3,5-pyridinedicarboxylic acid, which showed proton migration across a short strong hydrogen bond (SSHB) of length $d_{O\cdots N}$ = 2.50 Å.[20,21]

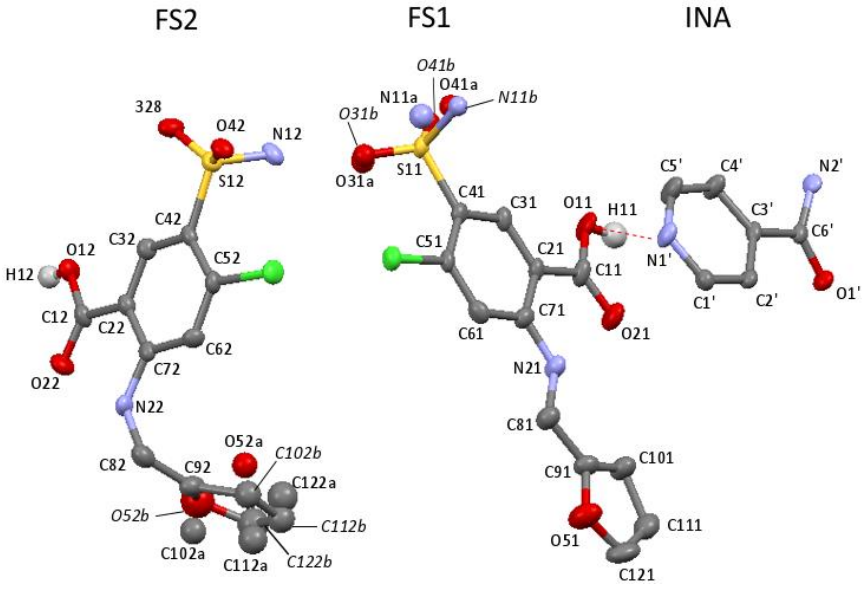

Figure 65 Asymmetric unit of 2FS-INA with thermal ellipsoids shown at the 50 % probability level. Major occupied sites of the sulphonamide and furan ring are shown as floating atoms and the short hydrogen bond is indicated by the red dashed line. FS atom labelling follows that of Babu *et al*. with a 1 or 2 denoting FS1 or FS2. A trailing apostrophe denotes INA sites.

It has been established that 2FS-INA does not possess advantageous physical properties for pharmaceutical use.[22] The intrinsic dissolution rate (IDR) is a kinetic measurement of the rate at which a substance dissolves, and the IDR of 2FS-INA is 44×10^{-3} mg cm^{-2} min^{-1}; indistinguishable from that of commercial form **I** FS over the experiment time period. The apparent solubility is 5.6 times higher than that of pure FS, but the cocrystal is unstable in slurry conditions and decomposes into its starting components over a 24-hour period.

5.3 Characterising molecular disorder

Only the average structure of 2FS-INA is known from the previous SCXRD study so the nature of the disorder and the characteristics of the short hydrogen bond are yet to be determined. This is therefore an ideal system to demonstrate how an NMR crystallography approach can be used to characterise the nature of complex molecular disorder.

5.3.1 Experimental details

A polycrystalline sample of 2FS-INA was prepared by L. Softely of the I.R.E. group at Durham University. This method of synthesis has previously been shown to reliably yield a single phase of 2FS-INA.[6] The cocrystal was partially deuterated at the exchangeable proton sites by storage in a D_2O atmosphere for two weeks prior to SSNMR experiments. Initial attempts to deuterate the sample by a "D_2O shake" failed due to degradation of the cocrystal into the starting components. Figure 66 shows the ^{2}H and ^{13}C spectra of three deuterated batches of 2FS-INA, with batches **(i)** and **(ii)** prepared by the "D_2O shake" method. The red lines on the ^{13}C spectra are guides for the eye showing the position of additional peaks that were present upon deuteration in batches **(i)** and **(ii)**, corresponding to the degradation products. It is clear from the ^{2}H spectra, showing only the centre peaks, that the D_2O shake method was not reproducible. This is not surprising since 2FS-INA has been found to be unstable in slurry conditions in H_2O during dissolution studies.[22]

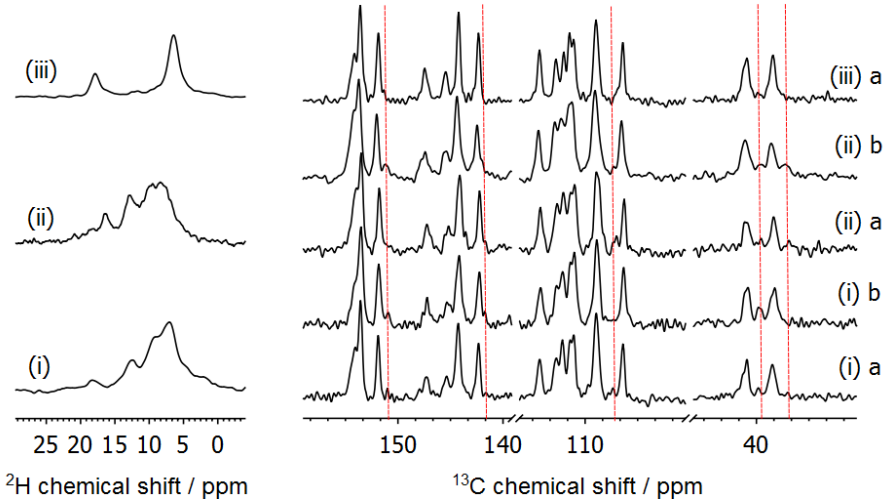

Figure 66 ^{2}H centre peak (left) and ^{13}C CP/TOSS spectra (right) of 2FS-INA after three deuteration attempts at 8 kHz MAS. **(i)** Sample left in 4 mL of D_2O overnight, **(ii)** Sample left in 4 mL D_2O for 5 days, **(iii)** Sample left in a D_2O atmosphere for 2 weeks. ^{13}C experiments labelled "a" are fresh samples and those labelled "b" were acquired after heating/cooling. ^{2}H spectra were acquired with a 1 s recycle delay and ^{13}C spectra were acquired with a contact time of 4 ms and recycle delays of **(i)** 35 s **(ii)** 15 s and **(iii)** 10 s.

To carry out DFT calculations on 2FS-INA the original disordered structure was split into four ordered structures, each containing a different combination of the disordered group orientations (r1/r2 for the disordered ring and s1/s2 for the disordered sulphonamide), resulting in the four structures r1s1, r1s2, r2s1 and r2s2. The behaviour of the short hydrogen bond was problematic in geometry optimisations, with the proton being transferred across the bond in some of the calculations. Similar issues have been found with other related functionals.[23,24] To avoid this, the proton was also fixed in position in the two problematic optimisations: r1s2-f and r2s2-f.

5.3.2 Preliminary structure

The ^{13}C SSNMR spectrum of 2FS-INA is shown in Figure 67(i). Most of the FS peaks are doubled, in keeping with the presence of two FS molecules in the asymmetric unit. Many of the carbon atoms are in aromatic environments so there is some peak overlap, but all the peaks can be assigned. The assignment is aided by the ^{13}C dipolar dephasing spectrum to identify quaternary carbon sites, Figure 67(ii), ^{1}H–^{13}C HETCOR experiments with varied contact times, Figure 68, as well as predicted shieldings from CASTEP. The full assignment with accompanying explanatory notes is given in Table 15. There is no evidence of small peaks that can be assigned to the minor occupied furan ring orientation on **FS2**, though they could be obscured by peak overlap in the region 110–120 ppm. The broadening of the

C51 and C52 peak at 142 ppm is likely to be due to the close proximity of these carbon atoms to the quadrupolar chlorine atom.[25-27] A ^{13}C–^{13}C rINADEQUATE experiment was attempted, Figure 68(iii), but after 10 days only a few signals are visible and these are barely distinguishable above the noise. Of the few correlations that can be picked out, the most useful of are the correlations C21–C71 and C22–C12, which help to resolve the assignment order.

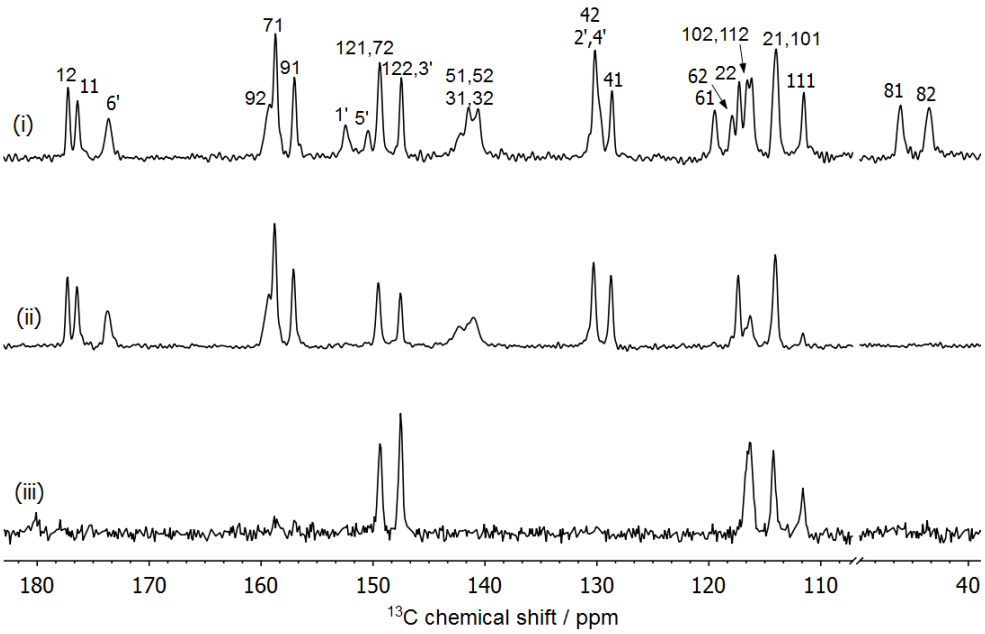

Figure 67 ^{13}C spectra of FS-INA at 8 kHz MAS. **(i)** CP/TOSS spectrum acquired over 48 transients with a recycle delay of 35 s and a contact time of 4 ms. **(ii)** CP/TOSS spectrum with a dipolar dephasing time of 80 µs acquired over 180 transients. **(iii)** Direct excitation spectrum acquired over 400 transients with a 1 s recycle delay.

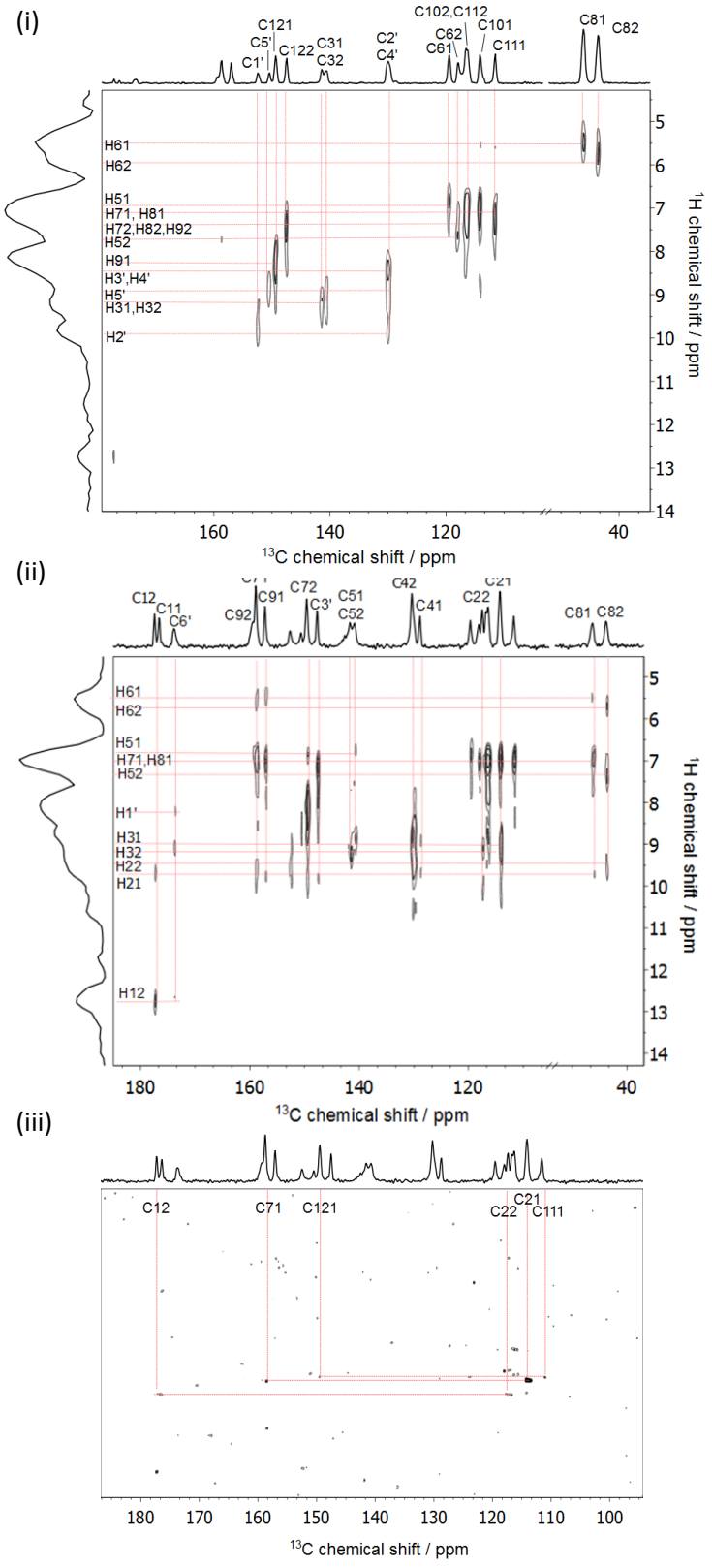

Figure 68 ^{1}H–^{13}C HETCOR spectra of 2FS-INA acquired with contact times of **(i)** 0.2 ms and **(ii)** 1 ms. Spectra were acquired at 9.5 kHz MAS with a recycle delay of 10 s with 64 t_1 increments each with 128 transients. **(iii)** ^{13}C–^{13}C rINADEQUATE spectrum of 2FS-INA acquired at 14 kHz MAS with a recycle delay of 7 s with 128 t_1 increments each with 800 transients.

Table 15 List of all carbon atoms in 2FS-INA with assignments and accompanying notes.

Atom Label	Chemical shift / ppm	Notes[a,b]	Atom Label	Chemical shift / ppm	Notes[a,b]
C82	43.6	C, S, L H5: 2.16 Å, H3: 2.64 Å.	C32		C?,S.
C81	46.3	C, S, L H39: 2.16 Å, H38: 2.66 Å.	C52	140.9	C?, Q, broad due to proximity to Cl.
C111	111.6	C, S.	C51	141.6	C?, Q, broad due to proximity to Cl.
C21	114.2	Q?, L? H32: 2.13 Å, H39: 2.46 Å.	C72	147.6	C?, Q,? L? H5: 2.01 Å
C101		C, S.	C122		C?, S.
C102		C?, S.	C121		C, S.
C112	116.4	C?, S.	C71	149.5	C, Q?, L? H21: 1.99 Å, H51: 2.17 Å.
C22	117.4	Q?, L H16: 2.14 Å, H5: 2.54 Å.	C5’	150.6	C, S.
C62	118.0	C, S.	C1’	152.6	C, S.
C61	119.6	C, S.	C91	157.1	C, Q, L H40a,H40b: 2.14 Å, H45: 2.20 Å, H39: 2.50 Å.
C41	128.8	C, Q, L? H32: 2.14 Å.	C3’	158.8	C?, Q?, L? H4’: 2.16 Å, H3’: 2.18 Å, H1’: 2.67 Å.
C42		C,Q.	C92	159.4	C?, Q?, L H6a,H6b: 2.13 Å, H103: 2.19 Å.
C2’	130.3	C, S.	C6’	173.7	C, Q, L H48a,H48b: 2.06 Å, H14: 2.52 Å, H54?: 2.70 Å.
C4’		C, S.	C11	176.5	C, Q.
C31	140.9	C?, S.	C12	177.3	C, Q, S, L H14: 1.93 Å, H5: 2.43 Å.

[a] Symbols used to indicate the basis of assignment: C = calculated ^{13}C shielding from CASTEP, Q = ^{13}C peak in non-quaternary suppression spectrum, S = cross peak visible in short contact time HETCOR experiment, L = cross peak visible in long contact time HETCOR experiment, ? = evidence is suggestive rather than definitive.
[b] Distances to non-bonded hydrogen atoms are given up to 2.70 Å.

Subtle changes are visible in the ^{13}C spectrum between −50 ˚C and 70 ˚C, Figure 69. This is indicative of the presence of a dynamic process affecting some of the carbon sites, but it is challenging to link the changes to a specific process. The C41 peak at 129 ppm drifts by 1.3 ppm over the temperature range shown, indicating that the local environment changes. C21 and C101 also become inequivalent in terms of chemical shift at low temperature. Dynamic ring flipping of the INA is not observed since C5’ and C1’ remain two distinct peaks with no evidence of coalescence.

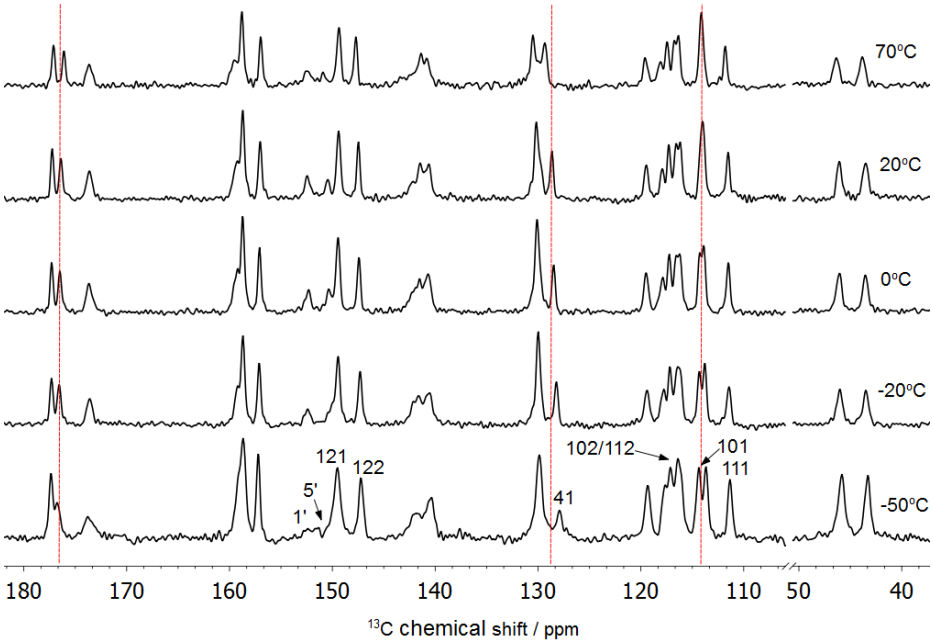

Figure 69 Variable temperature ^{13}C CP/TOSS spectra of 2FS-INA. All spectra were acquired at 8 kHz MAS with a recycle delay of 35 s except for the −50 ˚C spectrum, which was acquired at 5 kHz MAS with 15 s. Dashed lines are guides for the eye indicating reversible spectral changes, centred on peaks in the 20 ˚C spectrum.

5.3.3 Hydrogen position

The hydrogen-bonded protons, H11 and H12, are observed in the ^{1}H spectrum in Figure 70 at significantly higher chemical shifts than the bulk of the protons. H11 in particular has a chemical shift of 19.3 ppm. This is likely due to the fact it resides in the short hydrogen bond, since ^{1}H chemical shifts are intimately linked to the length of the hydrogen bond in which they reside.[28] H11 is also somewhat broader than H12, which might be indicative of some dynamics. However, the evidence is not conclusive and variable temperature experiments are restricted to a small range on the 1.3 mm probe.

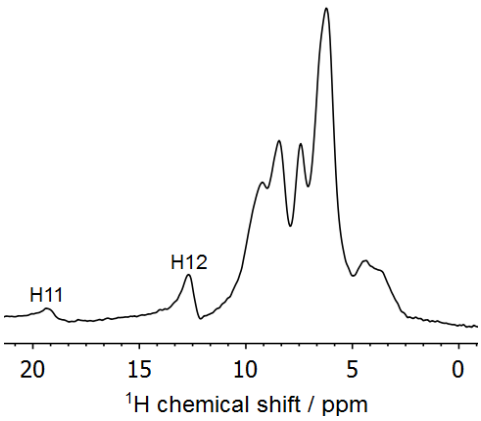

Figure 70 ^{1}H spectrum of 2FS-INA at 60 kHz MAS acquired with a recycle delay of 10 s over 40 transients. Hydrogen-bonded protons are labelled.

The nature of the short hydrogen bond is further investigated by comparing the experimental ^{15}N spectrum with calculated ^{15}N chemical shifts, Figure 71, since ^{15}N shifts are known to be sensitive to protonation state.[29-34] Calculated shieldings were converted to chemical shifts using $\delta_{iso} = \sigma_{ref} - \sigma_{iso}$, where σ_{iso} is the CASTEP-calculated shielding value and σ_{ref} is calculated to equate the average calculated shift and average experimental shift.[35] σ_{ref} is −161.7 ppm when H11 was transferred to INA and −169.0 ppm when H11 was fixed. The experimental shift of N1' (−116 ppm in Figure 71) matches the predicted shifts when H11 remains on FS1 and not when proton transfer has occurred across the short hydrogen bond. This supports the placement of H11 in the SCXRD data and also indicates that the proton position is not averaged over fast transfer. The calculated difference in N1' shielding upon protonation would be expected to be observed experimentally at 36.3 ppm, which it is not, and this is within the usual experimental range reported in the literature for salt/cocrystal transformations.[36] 2FS-INA can therefore be confidently described as a cocrystal and not a salt.

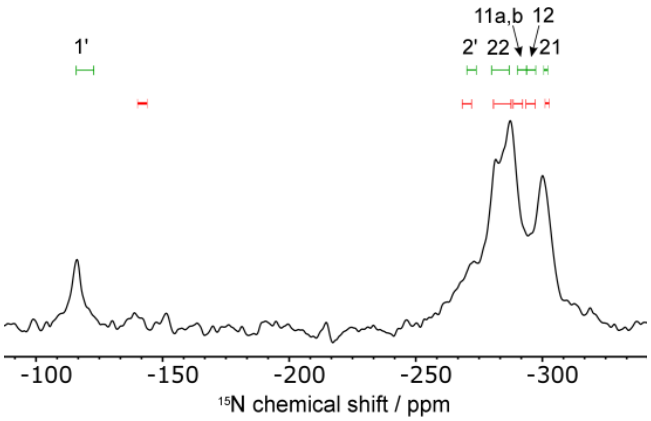

Figure 71 ^{15}N CP spectrum of 2FS-INA at 8 kHz MAS acquired with a recycle delay of 30 s and a contact time of 4 ms over 8640 transients. Calculated ^{15}N shieldings are overlaid with H11 associated with FS1 (green) and with transfer of H11 onto INA (red). The length of the lines represents the spread of shieldings over the four simulated structures.

5.3.4 Sulphonamide disorder

The ^{15}N spectrum is too heavily overlapped in the low shift region to draw conclusions about the sulphonamide disorder, and there are no nitrogen atoms near the furan ring. Variable temperature ^{1}H CRAMPS spectra acquired over the temperature range −24–60 ˚C showed no significant change, though the information available from these experiments was limited by the overlap of the broad ^{1}H lines. On the other hand, the sulphonamide group contains exchangeable protons on the NH_2 so the sulphonamide disorder can be investigated explicitly by ^{2}H SSNMR. Figure 72 shows that the shape of the ^{2}H spectra changes significantly with temperature: peaks broaden as the temperature is lowered to −60 ˚C. These changes are reversible so the signal intensity loss is due to the peak broadening rather than sample degradation/transformation.

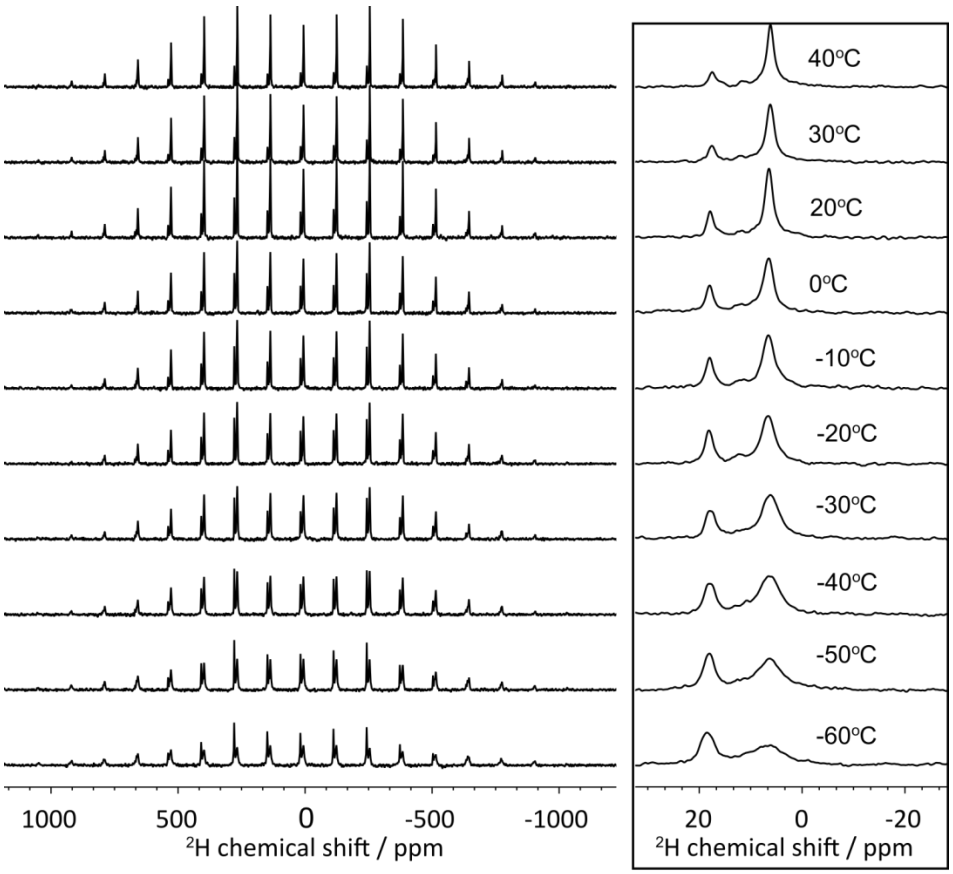

Figure 72 Variable temperature direct excitation ^{2}H spectra at 10 kHz MAS. The centre band of each spectrum is shown on the right. Spectra were acquired with a recycle delay of 1 s over 100 transients. Baselines were corrected for this figure using Whittaker smoother in MestReNova.

The broadening of the NH_2 peak of the sulphonamide group is particularly notable and is indicative of the freezing out of a dynamic process at low temperature. Figure 73 shows the sulphonamide peak broadening follows an Arrhenius-type curve with an activation barrier of *ca*. 9.5 kJ mol^{-1} (temperatures include +15 K to account for frictional heating during MAS). This is only a rough estimate of the barrier to the dynamic process since it is assumed that any other contributions to the sulphonamide peak are negligible, or at least are independent of temperature.

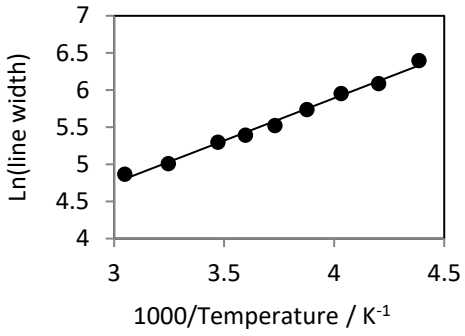

Figure 73 Variation in sulphonamide ^{2}H NMR peak width with temperature.

The quadrupolar parameters in Table 16 were obtainied by fitting the ^{2}H sideband patterns using pNMRsim.[37] Note that the −60 ˚C spectrum cannot be fit satisfactorily due to the extensive peak broadening. The hydrogen-bonded protons, H11 and H12, are fit to very different C_Q values, the H11 value being particularly low at 79 kHz at 20 ˚C. It has been reported that reducing the donor-acceptor distance in a hydrogen bond ($d_{D\cdots A}$) lowers the C_Q.[38-41] The fitted C_Q values are in keeping with the fact that the H11 hydrogen bond is very short ($d_{O11\cdots N1'}$ = 2.546 Å), while the H12 hydrogen bond is longer ($d_{O12\cdots O1'}$ = 2.619 Å) as measured by SCXRD. However, it is challenging to determine quantitative correlations between the C_Q and structural parameters because most studies into hydrogen bonds in molecular solids focus on many types of hydrogen bond with different strengths.[40] It is possible in individual cases, for example an isolated molecule containing a CO···HO hydrogen bond has been used to show that η is sensitive to bond angles as opposed to $d_{D\cdots A}$.[42] However, the H11 hydrogen bond in 2FS-INA is an N···HO bond so cannot be directly compared.

Table 16 Fitted ^{2}H quadrupolar parameters at three temperatures and the average, calculated parameters. One standard deviation of each fitted value is given in parentheses, though these errors are likely to be underestimated.

Site	−50 ˚C		20 ˚C		40 ˚C		CASTEP-average [a]	
	C_Q / kHz	η	C_Q / kHz	η	C_Q/ kHz	η	C_Q / kHz	η
Sulphonamide	88.4(6)	0.51(2)	87.1(2)	0.417(6)	87.8(2)	0.421(7)	222	0.16
H12	170(2)	0.22(5)	161(2)	0.34(4)	168(2)	0.26(5)	138	0.16
H11	68.8(5)	0.39(3)	78.9(5)	0.34(3)	81.0(9)	0.32(5)	48	0.52

[a] Structures showing transfer of H11 across the hydrogen bond are not included in the average.

The calculated quadrupolar parameters for the hydrogen-bonded protons H11 and H12 are somewhat comparable to the experimental fits. However, the calculated C_Q for the disordered sulphonamide site is *ca*. 2.5 times larger than the values obtained from fitting the spectrum at −50 ˚C. This can only be explained if there is extensive dynamics at the sulphonamide site, in keeping with the significant reversible line broadening discussed previously. NH_2 dynamics have been studied previously, see section 5.2.1, and were described as consisting of hindered NH_2 jumps that are rapid at ambient temperature.[19] The calculated quadrupolar parameters, C_Q = 222 kHz and η = 0.16, can be averaged over a jump of 112˚, corresponding to the exchange of the two NH_2 hydrogen atoms, giving dynamically-averaged values of C_Q = 112 kHz and η = 0.66. These are still somewhat higher than the experimental results. More consistent values can be obtained by averaging over two jump motions. For example, jumps of both 112˚ and 40˚ give C_Q = 89 kHz and η = 0.57 while jumps of 112˚ and 30˚ give C_Q = 100 kHz and η = 0.61. This second jump angle is consistent with the magnitude of the angle measured crystallographically between the two disorder sulphonamide positions. High frequency librational motions will also contribute to the dynamics[43] of the sulphonamide group and the actual atomic motion at ambient temperature will be more complex than instantaneous jumps between discrete sites. However, it is still possible to conclude that the observed ^{2}H quadrupolar parameters are consistent with rapid rotations between the two orientations of the sulphonamide group that were observed by SCXRD, as well as rapid exchange of the NH_2 protons. It is also important to note that the sulphonamide groups are relatively close in space, $d_{N\cdots N} \approx 2.5$ Å, so the motion is likely to be further complicated by correlations between the orientations. This would be very challenging to model computationally.

Complementary information is available by fitting the ^{2}H T_1 relaxation times over the temperature range −40–60 ˚C, Figure 74. The relaxation data is fitted to direct calculations of the ^{2}H relaxation rate under MAS, using the calculated C_Q of 222 kHz and a two-site jump model as described by Apperley *et al*.,[44] along with Arrhenius-type temperature dependence, Equations (5.0) and (5.1). This model uses the jump rate, k, between two equivalent orientations subtended by a half angle Θ. The functions A_1 and B_{1-5} are given in full by Torchia[45] and depend on Θ and the polar angles defining the crystallite orientations, θ and ϕ. τ_c is the correlation time for the jump motion, equal to $1/2k$, ω is the Larmor angular frequency of the nuclei involved, τ_0 is a frequency factor, E_a is the activation energy of the motion, and R is the gas constant.

$$\text{Rate} = \frac{(2\pi C_Q)^2}{8} A_1 \left\{ [B_4 - (0.75B_1 - B_2)\cos 2\phi] \frac{\tau_c}{(1+\omega^2\tau_c^2)} + [4B_5 - 4B_2\cos 2\phi] \frac{\tau_c}{(1+\omega^2\tau_c^2)} \right\} \quad \textbf{(5.0)}$$

$$\tau_c = \tau_0 e^{\left(\frac{E_a}{RT}\right)} \quad \textbf{(5.1)}$$

The fit in Figure 74 corresponds to E_a = 16.7(7) kJ mol^{-1} with a jump angle of 26.5(5)˚. The error bars are one standard deviation of the value, and are calculated using the average uncertainty of the fitting T_1 values, 0.7 ms. When a correction for sample heating under MAS is applied, the E_a is lowered slightly (14.9 kJ mol^{-1}) and there are only marginal effects on the other fitting parameters. This activation energy is lower than that of the hindered NH_2 jumps in pure FS[19] (31 kJ mol^{-1}), but the order of magnitude is consistent with the same type of process being involved. At 298 K, τ_c = 86 x 10^{-11} s corresponding to an effective jump rate 1/τ_c of 1.2 GHz, which is consistent with the fact that the C_Q is observed to be averaged in the experimental ^{2}H spectrum. These values should be treated with some caution as the true molecular motion is more complex than a simple two-site jump. Additionally the non-zero ^{2}H η value is not accounted for in the expressions by Torchia *et al.*,[45] though this is not expected to have a significant impact on the fitted parameters.

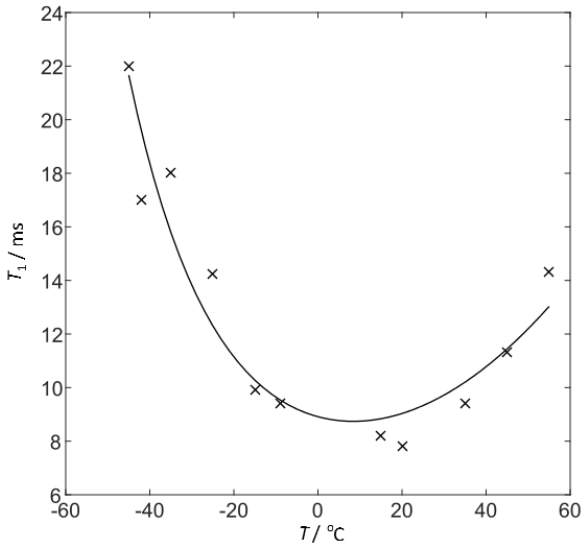

Figure 74 Sulphonamide ^{2}H T_1 relaxation times as a function of temperature fit to a two-site jump model with an Arrhenius-type temperature dependence of the jump rate. Temperatures include a correction of +15 °C for sample heating due to MAS. ^{2}H T_1 times were measured by applying a saturation train of 20 pulses separated by a delay of 0.2 ms at 10 kHz MAS and the τ delay was varied between 0.0001–2 s in 30 steps. Errors on fitted T_1 values are of the order of hundreds of μs, of the order of the size of the data points.

5.3.5 Furan disorder

The ^{13}C spectra in Figure 67 and Figure 69 are too overlapped to draw clear conclusions about the furan ring disorder. The ^{1}H–^{13}C HETCOR experiments might be expected to be more useful due to the extra information provided by the ^{1}H dimension. As such, the HETCOR spectrum acquired with a short contact time is compared with simulated ^{1}H–^{13}C spectra that were set up to model both static and dynamic furan ring disorder, Figure 75. The simulated spectra for static disorder of the furan ring, Figure 75(i), were calculated by averaging the NMR parameters from r1s1 and r1s2-f (blue), and the parameters from r2s1 and r2s2-f (red). The chemical shifts for all four were averaged to simulate dynamic disorder of the furan rings in Figure 75(ii). Note that averaging the chemical shifts in this way also assumes the sulphonamide group is dynamic, in keeping with the ^{2}H NMR experiments discussed above. The correlation between C122–H92 is more accurately predicted in the simulated dynamic disorder spectrum, but this is not conclusive evidence of the presence of dynamics because the differences are subtle. Other systems with less overlapped ^{13}C spectra might benefit from this spectral modelling method. It is noted that the three simulated spectra were referenced manually to get the best visual agreement.

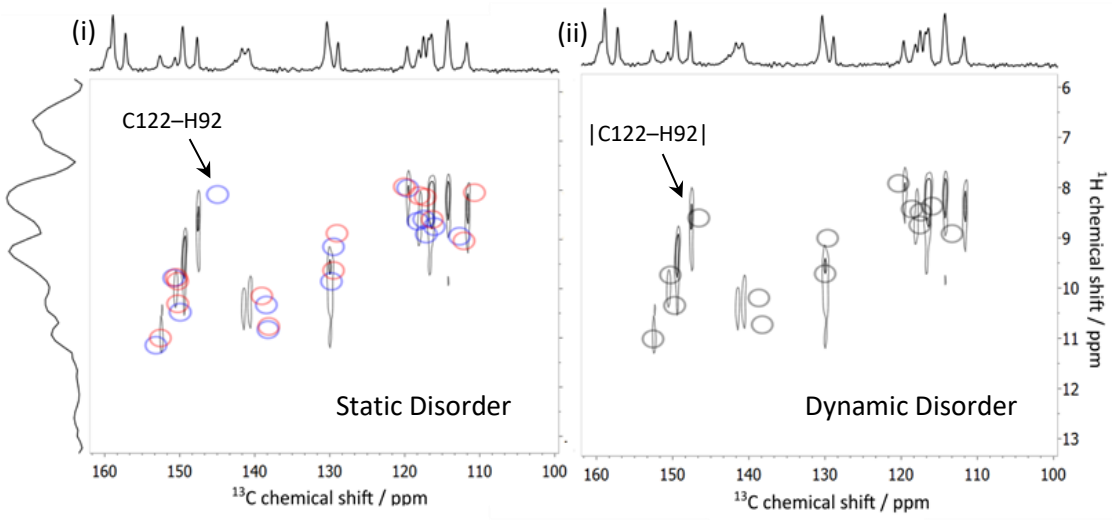

Figure 75 Simulated ^{1}H–^{13}C HETCOR spectra for static disorder **(i)**, where the red and blue circles denote the correlation peaks for the two furan ring orientations, and dynamic disorder **(ii)**. Only correlations between directly bonded ^{1}H and ^{13}C atoms are shown.

The ^{1}H–^{1}H proximities would be expected to be affected by the furan ring disorder. Figure 76 shows the ^{1}H–^{1}H DQ/SQ spectrum of 2FS-INA and Four correlations are visible to the

carboxylic acid protons of **FS1** and **FS2**; H11 and H12 respectively. The weak correlation **D** is assigned as H11–H2' because H2' is the closest neighbouring hydrogen atom to H11 with $d_{H11–H2'}$ = 2.53 Å, see inset of Figure 76 for the hydrogen atom labelling. Correlation **A** is assigned as H12–H1' with $d_{H12–H1'}$ = 2.53 Å. The other two correlations to H12, **B** and **C**, correspond to either H4' on the INA molecule ($d_{H12–H4'}$ = 2.69 Å) or H72 on the disordered furan ring ($d_{H12–H72}$ = 2.66 Å in one orientation, > 4 Å in the other orientation). **B** is assigned as the H12–H72 correlation because the heteronuclear C102/C112–H72/H82 correlation is observed in the ^{1}H–^{13}C HETCOR spectrum in Figure 68(i) at a lower ^{1}H chemical shift than the C2'/C4'–H3'/H4' correlation. **C** is therefore assigned as the H12–H4' correlation.

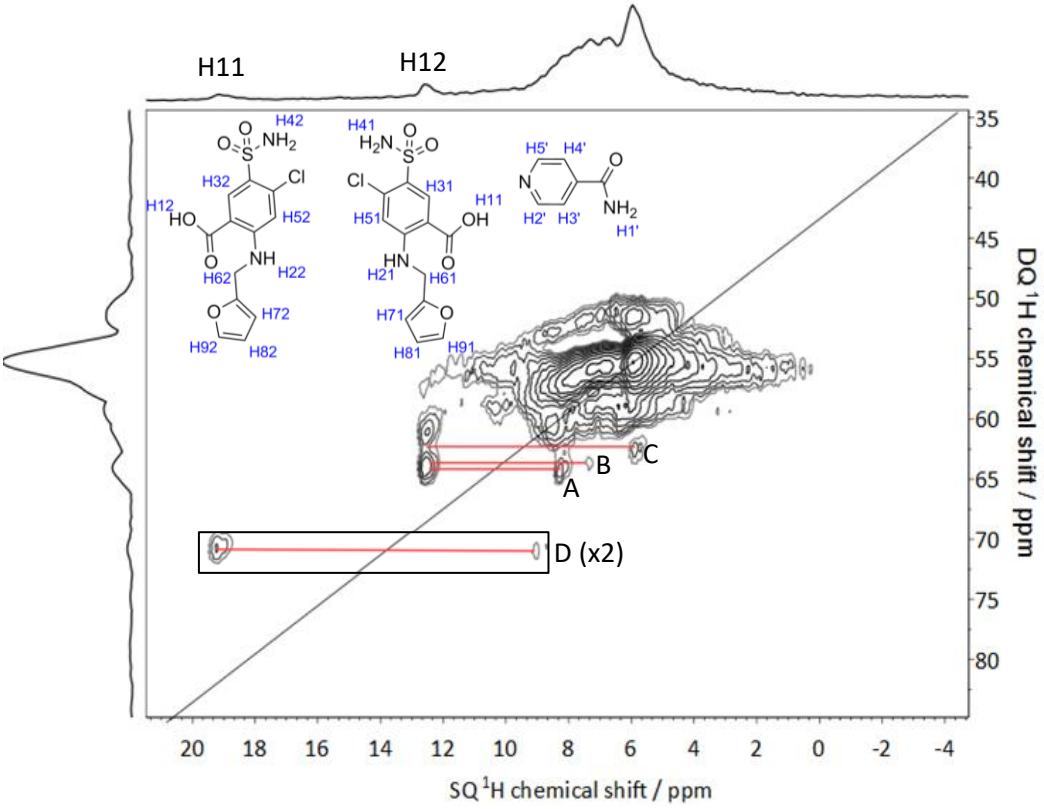

Figure 76 ^{1}H–^{1}H DQ/SQ spectrum of FS-INA at 50 kHz MAS acquired with a 10 s recycle delay. All 32 t_1 increments were each acquired over 1024 transients. No significant improvement in resolution was found when the MAS speed was increased to 62 kHz. The correlation D is overlaid as an inset magnified by two. Hydrogen atom labelling is also shown.

Again, it is challenging to draw conclusions about the furan ring disorder from the spectrum in Figure 76. Rapid dynamic disorder of 180˚ rotation of the **FS2** furan ring, as modelled by

the split atom sites in the SCXRD structure, would reduce the dipolar coupling between H12 and H72 by close to a factor of two. Specifically, averaging the dipolar tensors corresponding to $d_{H–H}$ distances of 2.66 Å and 5.43 Å using "calcav2" gives a mean dipolar coupling of 3.26 kHz, taking into account the difference of 50˚ in the tensor orientations. This compares to a coupling of 6.38 kHz for the 2.66 Å distance in the closest disordered conformation. This reduction is relatively small and it is difficult to measure ^{1}H–^{1}H dipolar couplings accurately due to dipolar truncation effects,[46] so it is hard to be sure whether correlation **B** is affected by dynamic disorder of the furan ring or not. This would be even more difficult if **FS2** is undergoing small angle motion. More definitive results might be obtained by this modelling method for other systems where the major occupied position results in a weak dipolar coupling while the minor position resulted in a strong coupling, *i.e*. the opposite case to 2FS-INA. Better still would be a system that contained an isolated spin pair, for example if H12 was substituted for a fluorine, in which case the ^{19}F–^{13}C dipolar coupling strength could be accurately measured.

A simple, but diagnostic, direct-excitation ^{13}C experiment with a short recycle delay selects only rapidly relaxing carbon signals, *i.e*. the sites that are likely to be dynamic. Figure 67(iii) shows that the furan ring carbon sites are visible for both **FS1** and **FS2** with a recycle delay of only 1 s, indicating that the ^{13}C T_1 relaxation times are significantly shorter than those of the other carbon sites. Since T_1 relaxation is driven by dynamics on the order of ν_{NMR} of the relevant nuclei, the furan ring carbon sites must exhibit dynamics of the order of 125 MHz at 298 K. The temperature dependence and magnitude of the ^{13}C T_1 times is similar for both furan rings, Figure 77, implying that similar dynamic processes are involved.

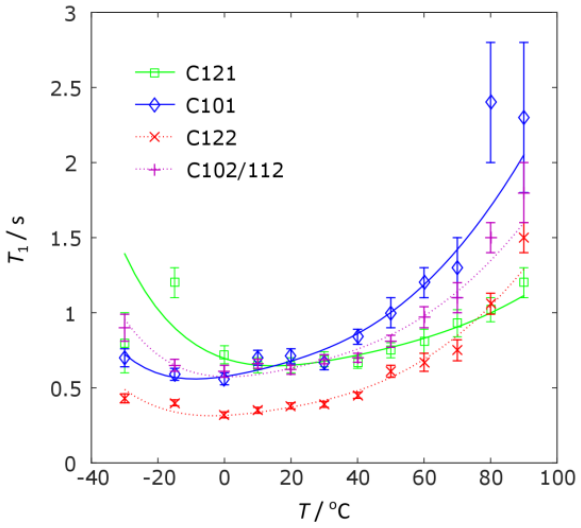

Figure 77 Variable temperature ^{13}C T_1 relaxation times for the furan ring carbon atoms on FS1 and FS2. Temperatures include a correction of +10 °C for sample heating due to MAS. The ^{13}C T_1 times were measured by applying a saturation train of 20 pulses separated by a delay of 10 ms at 8 kHz MAS and the τ delay was varied between 0.05–3 s in 25 steps.

The ^{13}C T_1 relaxation times can be fit using expressions for dipolar relaxation *via* rotational diffusion on a cone[45] to yield the parameters in Table 17. This model is clearly a simplification of the true motion, but the agreement between motional parameters fitted for carbon atoms on the same furan ring is satisfactory. It is hard to say if the differences between the furan rings of **FS1** and **FS2** are statistically significant. It might be said that the cone angle for **FS2** is larger than that of **FS1**, corresponding to a larger amplitude motion, particularly for C122. However, it is clear that magnitude of the motion is similar for both rings, *i.e.* the fitted parameters are inconsistent with one ring undergoing very large amplitude motion of up to 180˚ while the other undergoes small amplitude motion.

Table 17 Parameters describing the dynamics at the carbon atoms on the two furan rings from ^{13}C T_1 measurements. [a]

		E_a / kJ mol^{-1}	Log τ_0 / a.u.	Cone angle / degrees
FS1	C121	20(2)	−12.7(3)	19.8(5)
	C101	22(1)	−13.3(2)	21.6(4)
FS2	C122	25(1)	−13.7(2)	31.9(9)
	C102/C112	22(1)	−13.2(2)	21.3(4)

[a] ^{13}C T_1 data for C111 could not be satisfactorily fit due to the low signal intensity

The SCXRD experiment resolved additional electron around the **FS2** furan ring, prompting the use of a split site model with isotropic displacement ellipsoids while the **FS1** furan ring was modelled with single atom sites and ADPs. The ADPs on **FS1** are slightly elongated along the axis perpendicular to the plane of the ring, which is suggestive of small-amplitude motion. Since the ^{13}C T_1 relaxation time data suggests the motion of the two rings is similar, it seems contradictory to model the furan ring on **FS1** with small amplitude motion and the other on **FS2** with a rotation of 180˚. However, it is important to note than an equally good fit to the experimental SCXRD data can be obtained by fitting the **FS2** ring with single atom sites and ADPs, as shown in Figure 78, since it is challenging to discriminate between oxygen and carbon atoms in disordered fragments. Using this fit, the ring disorder looks to be similar on both **FS1** and **FS2** with some small difference in amplitude. Therefore, the SCXRD experiment is unable to distinguish between small and large-amplitude motion, but the SSNMR data suggests that the two rings are undergoing very similar small-amplitude motion.

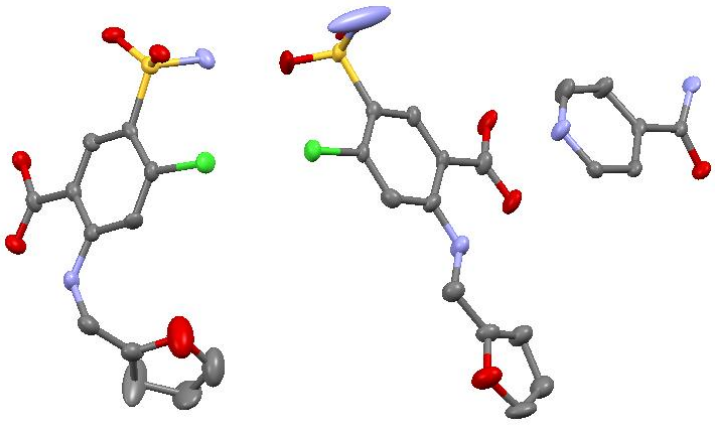

Figure 78 Asymmetric unit of 2FS-INA with the furan ring fragments modelled with single atom sites and ADPs.

5.4 Conclusions

Three points of interest were investigated regarding the disordered solid-state structure and dynamics of 2FS-INA that could not be addressed with SCXRD alone. Firstly, the nature of a short OH···N hydrogen bond was investigated in order to establish the precise position of the hydrogen atom in the bond. ^{15}N SSNMR showed conclusively that the hydrogen atom is associated only with the FS carboxylic acid group following comparison with

computationally-predicted ^{15}N shifts for both the SCXRD structure and the alternative salt structure. No evidence of dynamic proton transfer was found. Additionally, the disorder of the sulphonamide group was readily probed with variable temperature ^{2}H experiments, and the data was found to be consistent with a motional model involving fast exchange of the NH_2 protons as well as rotation of the whole group sulphonamide about the C–S bond. The activation energy of *ca*. 17 kJ mol^{-1} was estimated, though the motion is certainly more complex than a simple two-site jump. Finally, the carbon atoms of the furan rings of both of the unique FS molecules were found to possess very short relaxation times, consistent with the presence of dynamics. Fitting of the variable temperature relaxation times provided estimates for the energy barriers to rotation and both rings were found to behave similarly within the experimental uncertainties.

This chapter highlights how an NMR crystallography approach can be implemented to provide information complimentary to SCXRD on the disorder of fragments of different molecules in the asymmetric unit of a cocrystal, as well as confirming the location of hydrogen atoms in short hydrogen bonds. The dynamic nature of the disordered moieties at ambient temperature simplifies the NMR spectra in this case by sharpening line widths and shortening relaxation times allowing for detailed information to be obtained.

5.5 References

(1) http://www.drugbank.ca/drugs/DB00695, accessed January 2016.
(2) G. E. Granero, M. R. Longhi, M. J. Mora, H. E. Junginger, K. K. Midha, V. P. Shah, S. Stavchansky, J. B. Dressman and D. M. Barends, *J. Pharm. Sci.*, **2010**, 99, 2544.
(3) N. R. Goud, S. Gangavaram, K. Suresh, S. Pal, S. G. Manjunatha, S. Nambiar and A. Nangia, *J. Pharm. Sci.*, **2012**, 101, 664.
(4) V. Ambrogi, L. Perioli, C. Pagano, F. Marmottini, M. Ricci, A. Sagnella and C. Rossi, *Eur. J. Pharm. Sci.*, **2012**, 46, 43.
(5) U. B. R. Khandavilli, S. Gangavaram, N. R. Goud, S. Cherukuvada, S. Raghavender, A. Nangia, S. G. Manjunatha, S. Nambiar and S. Pal, *CrystEngComm.*, **2014**, 16, 4842.
(6) B. I. Harriss, L. Vella-Zarb, C. Wilson and I. R. Evans, *Cryst. Growth Des.*, **2014**, 14, 783.
(7) E. Sangtani, S. K. Sahu, S. H. Thorat, R. L. Gawade, K. K. Jha, P. Munshi and R. G. Gonnade, *Cryst. Growth Des.*, **2015**, 15, 5858.
(8) V. K. Srirambhatla, A. Kraft, S. Watt and A. V. Powell, *CrystEngComm.*, **2014**, 16, 9979.
(9) D. Stepanovs and A. Mishnev, *Acta Crystallogr. C*, **2012**, 68, 488.
(10) T. Ueto, N. Takata, N. Muroyama, A. Nedu, A. Sasaki, S. Tanida and K. Terada, *Cryst. Growth Des.*, **2012**, 12, 485.
(11) V. S. Minkov, A. A. Beloborodova, V. A. Drebushchak and E. V. Boldyreva, *Cryst. Growth Des.*, **2014**, 14, 513.
(12) M. Banik, S. P. Gopi, S. Ganguly and G. R. Desiraju, *Cryst. Growth Des.*, **2016**, 16, 5418.
(13) J. Lamotte, H. Campsteyn, L. Dupont and M. Vermeire, *Acta Crystallogr. B*, **1978**, 34, 1657.
(14) N. J. Babu, S. Cherukuvada, R. Thakuria and A. Nangia, *Cryst. Growth Des.*, **2010**, 10, 1979.

(15) N. R. Goud, N. J. Babu and A. Nangia, *Cryst. Growth Des.*, **2011**, 11, 1930.
(16) W. Shin and G. S. Jeon, *Proc. Coll. Nat. Sci. SNU*, **1983**, 8, 45.
(17) S. Karami, Y. Li, D. S. Hughes, M. B. Hursthouse, A. E. Russell, T. L. Threlfall, M. Claybourn and R. Roberts, *Acta Crystallogr. B*, **2006**, 62, 689.
(18) A. Wolfenson, S. C. Perez, M. J. Zuriaga, C. Garnero, A. Miranda, M. Longhi and S. N. Faudone, *Chem. Phys. Lett.*, **2015**, 641, 163.
(19) J. N. Latosińska, M. Latosińska, W. Medycki and J. Osuchowicz, *Chem. Phys. Lett.*, **2006**, 430, 127.
(20) I. Frantsuzov, S. J. Ford, I. Radosavljevic Evans, A. J. Horsewill, H. P. Trommsdorff and M. R. Johnson, *Phys. Rev. Lett.*, **2014**, 113, 018301.
(21) S. J. Ford, O. J. Delamore, J. S. Evans, G. J. McIntyre, M. R. Johnson and I. Radosavljevic Evans, *Chem. Eur. J.*, **2011**, 17, 14942.
(22) H. E. Kerr, L. K. Softley, K. Suresh, A. Nangia, P. Hodgkinson and I. R. Evans, *CrystEngComm.*, **2015**, 17, 6707.
(23) J. Ireta, J. Neugebauer and M. Scheffler, *J. Phys. Chem. A*, **2004**, 108, 5692.
(24) A. D. Boese, *ChemPhysChem.*, **2015**, 16, 978.
(25) R. K. Harris and A. C. Olivieri, *Prog. Nucl. Magn. Reson. Spectrosc.*, **1992**, 24, 435.
(26) S. H. Alarcón, A. C. Olivieri and R. K. Harris, *Solid State Nucl. Magn. Reson.*, **1993**, 2, 325.
(27) R. K. Harris, M. M. Sünnetçioglu, K. S. Cameron and F. G. Riddell, *Magn. Reson. Chem.*, **1993**, 31, 963.
(28) R. K. Harris, P. Jackson, L. H. Merwin, B. J. Say and G. Hagele, *J. Chem. Soc. Farad. T. I*, **1988**, 84, 3649.
(29) J. Leclaire, G. Poisson, F. Ziarelli, G. Pepe, F. Fotiadu, F. M. Paruzzo, A. J. Rossini, J. N. Dumez, B. Elena-Herrmann and L. Emsley, *Chemical Science*, **2016**, 7, 4379.
(30) R. M. Claramunt, *ARKIVOC*, **2005**, 2006, 5.
(31) S. Sharif, G. S. Denisov, M. D. Toney and H. H. Limbach, *J. Am. Chem. Soc.*, **2007**, 129, 6313.
(32) A. S. Tatton, T. N. Pham, F. G. Vogt, D. Iuga, A. J. Edwards and S. P. Brown, *CrystEngComm.*, **2012**, 14, 2654.
(33) P. B. White and M. Hong, *J. Phys. Chem. B*, **2015**, 119, 11581.
(34) S. L. Veinberg, K. E. Johnston, M. J. Jaroszewicz, B. M. Kispal, C. R. Mireault, T. Kobayashi, M. Pruski and R. W. Schurko, *Phys. Chem. Chem. Phys.*, **2016**, 18, 17713.
(35) R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates and V. Zorin, *Magn. Reson. Chem.*, **2007**, 45, 174.
(36) J. S. Stevens, S. J. Byard, C. C. Seaton, G. Sadiq, R. J. Davey and S. L. Schroeder, *Phys. Chem. Chem. Phys.*, **2014**, 16, 1150.
(37) P. Hodgkinson*, pNMRsim: a general simulation program for large problems in solid-state NMR*, www.dur.ac.uk/paul.hodgkinson/pNMRsim, accessed 2016.
(38) T. Chiba, *J. Chem. Phys.*, **1963**, 39, 947.
(39) T. Chiba, *J. Chem. Phys*, **1964**, 41, 1352.
(40) R. Webber and G. H. Penner, *Solid State Nucl. Magn. Reson.*, **2012**, 47-48, 10.
(41) R. Blinc and D. Hadži, *Nature*, **1966**, 212, 1307.
(42) G. W. Turner, R. L. Johnston and K. D. M. Harris, *Chem. Phys.*, **2000**, 256, 159.
(43) M. Dracinsky and P. Hodgkinson, *CrystEngComm.*, **2013**, 15, 8705.
(44) D. C. Apperley, A. F. Markwell, I. Frantsuzov, A. J. Ilott, R. K. Harris and P. Hodgkinson, *Phys. Chem. Chem. Phys.*, **2013**, 15, 6422.
(45) D. A. Torchia and A. Szabo, *J. Magn. Reson.*, **1982**, 49, 107.
(46) P. Hodgkinson and L. Emsley, *J. Magn. Reson.*, **1999**, 139, 46.

Chapter 6: Characterising a new cocrystal and its isostructural solvates

6.1 Solvates in the pharmaceutical industry

Organic solvates are common;[1,2] it has been estimated that up to a third of all APIs are capable of forming a hydrate.[3] APIs are also frequently exposed to organic solvates during pharmaceutical production, processing, and storage, so it is important to understand the solid-state structure and behaviour. Such information is pertinent to fundamental studies of structure and dynamics, but it may also act as a guide for more directed crystal engineering methods.

As defined briefly in Chapter 1, a solvate is a multicomponent crystal where one of the components is liquid at room temperature. A search* of the CSD yields 52000 entries containing the word "solvate" or "hydrate", some of which are multiple entries for compounds that have a particular propensity for solvate formation. For example, the antibiotic sulfathiazole has been observed to form over 100 solvates with great structural diversity.[4] Many API solvates have been patented for drug delivery,† though the scope is limited to biologically compatible solvents like propylene glycol[5] and water.[6] Hydrates are particularly relevant for pharmaceutical matters[7] because they offer advantages such as higher stability, solubility, and crystallinity compared to non-solvated forms of APIs.[1,8-10] However, solvate stability varies widely, depending on temperature and the partial pressure of the solvent. Therefore, they are usually avoided in the pharmaceutical industry and systematic screenings are carried out to identify potentially problematic solvate forms.[11,12] More often solvates are precursors to solvent-free forms, since polymorphism of solvates appears to be rare.[13] Additionally, desolvation can be controlled in some cases to yield systems with a narrow range of particle sizes, so industrial uses tend to focus on the

* Search subject to the following restrictions: 'only organics', '3D coordinates determined' and 'not polymeric'.

† A search of Espacenet for patents whose title contained the word '"solvate" yielded 580 results worldwide.

use of solvates for purification and solid form selection as a stepping stone towards another desired product.[11]

There are two main driving forces for the formation of a solvate.[2] Firstly, a system may crystallise such that there is unexploited potential for intermolecular interactions. In this case the incorporation of solvent molecules can compensate, particularly if there are hydrogen bonding possibilities available. Secondly, the host crystal lattice may pack in a relatively inefficient manner, resulting in channels or voids in the lattice that can be reduced by the inclusion of guest solvent molecules. Both driving forces have the net effect of reducing in the free energy of the crystal compared to the non-solvated form.[2] Water is particularly proficient of fulfilling unexploited intermolecular interactions and filling voids, due to its small size and multidirectional hydrogen bonding capabilities, though many organic solvates also satisfy these properties.

6.1.1 Classification and properties

Two classes are generally defined: stoichiometric solvates and nonstoichiometric solvates. The former have a fixed ratio of solvent/compound, which need not be an integral number but is well-defined between different samples, *e.g*. the water molecules in caffeine hydrate have a site occupancy of 0.8.[14] The solvent is an integral part of the crystal network, so desolvation leads to a breakdown of the structure, yielding a different packing morphology or a disordered state. In contrast, nonstoichiometric solvates are a type of inclusion compound where the solvent molecule is a guest within channels or cavities in the host crystal network. An important feature of this class is that the host network is robust enough to survive variable degrees of desolvation. In some cases complete desolvation can yield an isostructural material termed a desolvated (or isomorphous) solvate, which retains all of the packing features of the original solvate but without the guest solvent molecules.[15,16] It is common for such systems to be metastable and they are prone to taking up other solvent molecules from the environment.[11,15]

Different solvates of a particular system can also be isostructural,[17-20] and they generally form when the solvent molecules undergo a specific interaction with the host lattice[21] or are of a certain molecular size/shape to fit in voids/channels in the lattice, Figure 79.[22] Examples of APIs that form isostructural solvates are droperidol,[18,23,24] benperidol,[25] and cocrystals of FS and 4-toluamide.[26]

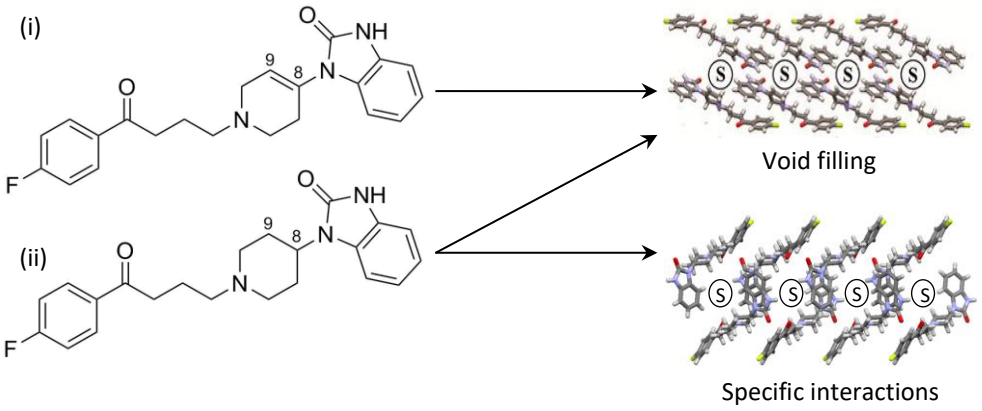

Figure 79 Molecular structures of **(i)** droperidol and **(ii)** benperidol. The type of isostructural solvate formed for each molecule is indicated. The 'void filling' solvate image is taken with permission from Bērziņš *et al.,* Cryst. Growth Des., 2014, 14, 5, 2654–2664. Copyright © 2014 American Chemical Society.

The stability of solvates is dependent on several properties, including the temperature and partial pressure of solvent[11] but, generally, stoichiometric solvates are more stable than nonstoichiometric solvates becuase the solvent molecules are intimately bound to the lattice. Desolvation is a complex kinetic phenomenon, depending on uncontrollable parameters such as particle size and crystal defects. Complete desolvation may result in an amorphous system being formed, which may itself be metastable and prone to crystallisation or hydration.

6.1.2 Solvate characterisation

Accurate characterisation of solvates requires detection of the solvent. This is not always trivial because solvent molecules are often volatile and desolvation can occur under comparatively mild conditions.[11] Thermal methods provide insight into pharmaceutically-relevant properties. For example, DSC is used to establish the heat of desolvation, and dynamic vapour sorption (DVS) probes solvate formation and behaviour over a range of temperatures and solvent concentration. TGA measures the weight loss of a system as a function of applied temperature and so can be used to quantify the amount of solvent present.

Many other techniques are used in combination with thermal methods. By observing changes in crystal habit under heating, microscopy can distinguish between stoichiometric and nonstoichiometric solvates: the former become opaque on desolvation due to the collapse of the crystalline structure.[11] CSP has also been applied to solvate systems, though

this is subject to the same challenges as polymorph prediction, *i.e*. the energy differences between different structures are small. With that being said, the structure of a chloride salt hydrate has been successfully predicted by at least one CSP method during the most recent (6th) CSP blind test.[27] Some indication can also be given for systems that may be prolific solvate formers by considering CSP calculations of the pure compound, which might indicate that a system does not pack efficiently on its own.[28]

Solvates often exhibit structural disorder, especially nonstoichiometric solvates where there are few or no intermolecular interactions between the solvent and the host lattice. The disorder may be modelled as discussed previously using split sites or ADPs, but in some cases the atom positions can be challenging to define in XRD data. An alternative solution is to use the PLATON SQUEEZE approach,[29] which back-calculates the solvent contribution to the structure factors from the electron density. This takes the contribution of the solvent into account in determining the structure of the host lattice, lowering the *R* index, but does not provide any information on the behaviour of the solvent itself. The reader is referred to the original reference for a description of the theory.

Identifying solvent and distinguishing between different solvates is straightforward with routine ^{13}C SSNMR in the context of an NMR crystallography approach. Figure 80 shows an example of five different solvates of phenolbarbitol in which the solvent peaks are clearly identified.[30] Solvent–lattice or solvent–solvent interactions can be probed by 2D experiments, which may help to locate solvent molecules in disordered systems like diterbutaline sulphate diacetic acid.[31] Changes in chemical shift that are correlated to the solvent content can also provide information on the location of solvent molecules.[32,33] For example, subtle changes in ^{13}C shift are observed in topotecan hydrochloride as the water content is varied, highlighting the regions of the topotecan molecule that are close to the water channels.[31] Desolvation is frequently monitored using multinuclear SSNMR,[9,31,34-36] especially for hydrates.[37,38] Details of the mechanism of desolvation can be ascertained, for example subtle changes in ^{15}N spectra have revealed the preferential desolvation of H_2O from only one of two possible sites in a nonstoichiometric hydrate.[39]

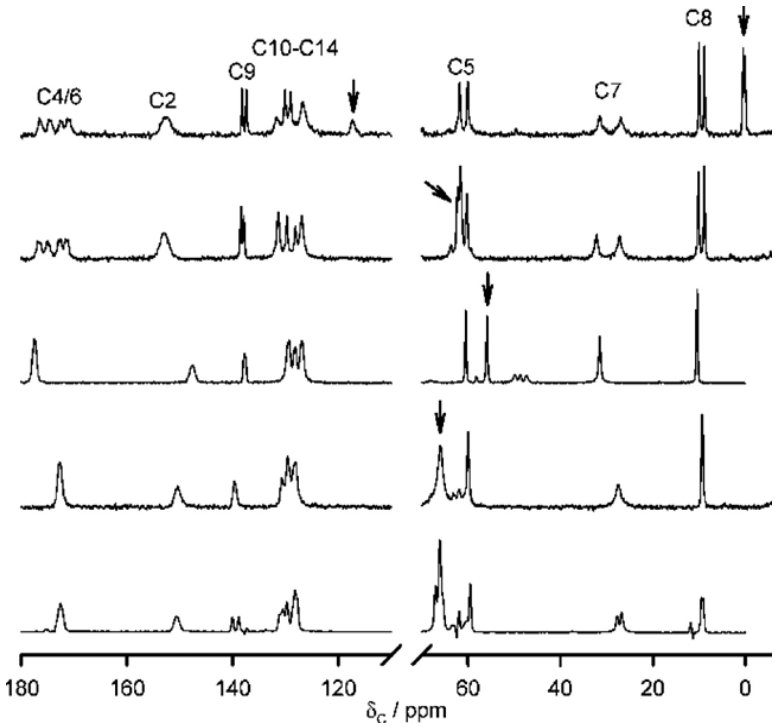

Figure 80 ^{13}C CP/MAS spectra of five different solvates of phenolbarbitol with the solvent peaks indicated by arrows. Spectra reproduced with permission from Zencirci *et al*., J. Phys. Chem. B, 2014, 118, 12, 3267–3280. Copyright © 2014 American Chemical Society.

A nucleus of particular utility in the context of solvate characterisation by SSNMR is ^{2}H. Lattice-incorporated solvent molecules can be easily identified by the observation of a ^{2}H spectrum with an extensive sideband manifold or a broad static line shape following deuterium-labelled vapour exchange of the solvent molecules.[40] The ^{2}H NMR parameters are also affected by dynamic processes on a broad range of timescales.[41] In the slow motion regime, where the exchange rate between two discrete sites is less than 10 kHz, distinct sideband manifolds are associated with the different sites. The sideband linewidth is directly proportional to the exchange rate in this regime.[42] In the fast motion regime, faster than 10 MHz, the ^{2}H spectrum is averaged leading to narrow sideband manifolds with low C_Q values, *e.g*. fast moving D_2O molecules in a channel hydrate give rise to a motionally narrowed line shape with 12 kHz FWHM.[32] In this regime the sideband linewidth is inversely proportional to the exchange rate. The intermediate regime, 10^4–10^7 Hz, is the accessible window for static ^{2}H experiments. Quadrupole echo experiments can be used to identify motion in this regime because diagnostic line shape distortions are observed as a function of the echo pulse spacing.[43] These distortions can be correlated with the exchange rate.[44] It

is also possible to model the effect of specific dynamics on ^{2}H NMR parameters in order to estimate the timescale and relative orientation of motion.[41]

The bulk of this chapter discusses the use of NMR crystallography methods, including ^{2}H NMR, to investigate the structure and dynamics of two isostructural solvates of a previously unreported cocrystal, FS-PA. However, before discussing the solvates it is pertinent to examine the parent cocrystal. Both FS-PA and its solvates are particularly challenging to characterise, since FS-PA cannot be crystallised and SCXRD data of the solvates does not provide clear atom positions for the solvent molecules. As a result, NMR crystallography methods are applied to these systems to provide additional information on the structure and dynamics. In the case of the parent cocrystal FS-PA SSNMR data is intended to go some way to help solve the structure, with the view of combining NMR data with future PXRD experiments. This is a challenging problem and it is beyond the scope of this work to provide a full structure solution, but some constraints are established as described in section 6.2. In the latter part of this chapter, section 6.3, the location and behaviour of the solvent molecules in both solvates are identified.

6.2 NMR crystallography without single crystals

NMR crystallography studies can go beyond the identification of local structural details, such as disorder and dynamics, as has been done in the main bulk of this work. One goal is 3D structure solution; a challenging problem that requires large amounts of SSNMR data. This is most useful for complex systems that do not crystallise readily, which is becoming an increasing problem in the pharmaceutical industry as drug molecules become larger and more complex.[45] There are many reports in the literature where SSNMR has been used to provide constraints to a structural model obtained from fits of PXRD data.[46-52] An example is the structure solution of a cocrystal of IND with NA.[48] PXRD data of the powdered cocrystal was fit to a structural model, which was then used as an input to CASTEP for structure optimisation and NMR calculations. Geometry optimisation of the PXRD structure led to a small average heavy atom displacement of 0.08 Å, indicating that the model was near an energy minimum. The calculated NMR parameters were found to agree well with both 1D and 2D experimental ^{1}H and ^{13}C data, therefore validating the PXRD model.

It is also possible to solve the structure of powdered solids by combining SSNMR and DFT without any XRD data. Proton spin diffusion methods have been utilised to measure distances,[53,54] even in complex systems like proteins,[55] but these are experimentally demanding. CSP has also been combined with SSNMR to solve the crystal structure of a

potential diabetes drug[45] by comparing the calculated ^{1}H chemical shifts of predicted models with experimental NMR data. The ^{1}H chemical shift RMSD between calculated and predicted shifts of 35 trail models was plotted and only one model was found to have an RMSD within one standard deviation of the accepted error on calculated ^{1}H shifts (±0.3 ppm). This approach requires access to custom CSP programs and the associated expertise, so it is not yet routine to carry out structure solution in this manner.

All of the published structure solutions by NMR crystallography require a computational model with which to compare predicted and experimental NMR parameters. Without such models it is challenging to achieve a full 3D structure because NMR is not sensitive to key crystallographic parameters such as unit cell or space group. Instead, the asymmetric unit can be identified and interatomic distances and orientations might be characterised, but the average structure is not accessible.

6.2.1 Furosemide-picolinamide

The API FS is a prevalent cocrystal former and 12 cocrystals have been reported, as discussed in Chapter 5.[26,56-58] Of these cocrystals, three consist of FS cocrystallised with pyridine carboxamides, NA and INA, and an additional hydrate with the third pyridine carboxamide, PA, was published recently.[26] However, the anhydrous cocrystal of FS and PA has not been reported. FS-PA was synthesised by neat grinding by L. Softley of the I.R.E group at Durham University, with the intention of establishing the structure and pharmaceutical properties. However, no single crystals could be grown of FS-PA in a 1:1 ratio; the system preferentially crystallises with solvent molecules trapped in the lattice, see section 6.3. PXRD data of FS-PA was acquired by the I.R.E group, but a Reitveld fit of the data was only suitable to yield unit cell parameters, see Table A1 in the Appendix. An NMR crystallography approach is described in the following section, aiming to identify the asymmetric unit of FS-PA and key structural features that might aid structure solution with future, high quality, PXRD data.

6.2.2 Experimental details

A fresh sample of FS-PA was synthesised for SSNMR investigations by grinding 0.127 g PA and 0.333 g FS by hand in a mortar and pestle for 1.5 hours. A PXRD of the fresh sample, Figure 81, was acquired by G. Oswald of the Durham University XRD service and the peak positions and intensities showed no obvious differences when compared to the original sample synthesised by the I.R.E group. The sample was partially deuterated by storage in a D_2O atmosphere for seven days prior to ^{2}H SSNMR experiments. No evidence of sample

degradation was observed by ^{13}C SSNMR over the course of nine months, but subtle changes were observed by PXRD over six months. This is similar to the degradation observed for 2FS-INA under a D_2O atmosphere and further highlights the instability of FS cocrystals in a humid atmosphere.

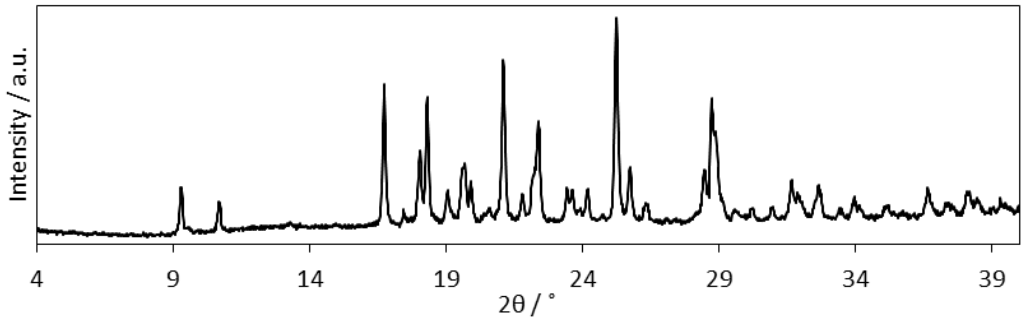

Figure 81 PXRD pattern of FS-PA after synthesis by grinding. No peaks are observed that correspond to the starting materials.

6.2.3 Investigating an unknown structure

The ^{13}C spectrum of FS-PA, Figure 82, has relatively narrow peaks (100 Hz FWHM on average) indicating that the sample is crystalline. The shifts of all of the peaks differ to those of the pure starting materials, indicating that a cocrystal is formed, with the components bound together *via* intermolecular interactions. The asymmetric unit can be identified as containing a single FS molecule and a single PA molecule as there is no splitting observed, most notably at the FS carboxylic acid and PA carboxamide peaks, C1 and C1'. There is a degree of peak overlap, but it is possible to fully assign the ^{13}C sites using the numbering scheme in Figure 82(i) and Table 18. This is aided by the ^{13}C dipolar dephasing spectrum to identify quaternary carbon sites, Figure 82(ii), as well as ^{1}H–^{13}C HETCOR experiments at both long and short contact times, Figure 83.

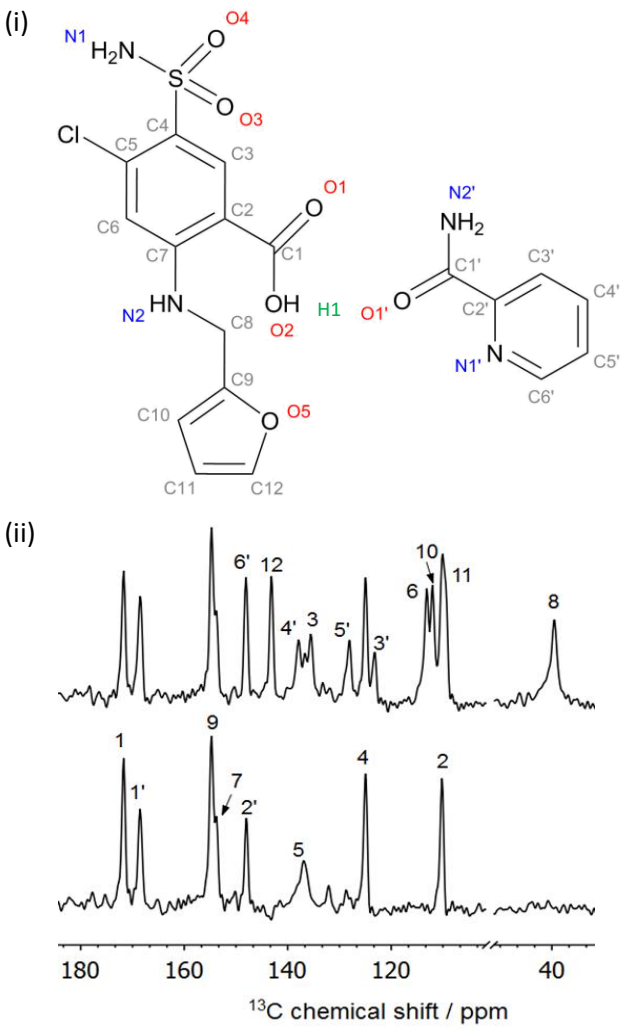

Figure 82 (i) Molecular structure of FS-PA with atom labels. **(ii)** ^{13}C CP/TOSS spectrum with non-quaternary sites labelled (top) and ^{13}C CP/TOSS spectrum acquired with an 80 µs dipolar dephasing time and quaternary sites labelled (bottom). Both spectra were acquired at 5 kHz MAS over 100 transients with a recycle delay of 5 s and a contact time of 2.5 ms.

Table 18 List of all carbon atoms in FS-PA with assignments and accompanying notes.

Atom Label	Chemical shift / ppm	Notes[a]	Atom Label	Chemical shift / ppm	Notes[a]
C8	39.5	S?, F	C5	136.7	Q
C11	109.4	S?, F	C4'	137.9	S?, P
C2	110.1	Q, F, L: H3	C12	143.2	S, F
C10	112.0	S?, F	C6'	148.1	Q, F
C6	113.2	S?, F	C2		S, F
C3'	123.2	S, F	C7	153.7	Q, L: H6
C4	125.0	Q, L: H3	C9	154.7	Q, L: H8
C5'	128.1	S, P	C1'	168.5	Q, P, L:PA NH_2
C3	135.5	S, F	C1	171.7	Q, F, L: PA NH_2

[a] Symbols used to indicate the basis of assignment: Q = ^{13}C peak in non-quaternary suppression spectrum, S = cross peak visible in short contact time HETCOR experiment, L = cross peak visible in long contact time HETCOR experiment, ? = evidence is suggestive rather than definitive, F = ordering follows that of 2FS-INA, P = ordering follows that of NPX-PA.

The carbon atom assignment is mostly straightforward, but peak overlap introduces ambiguity at 154 ppm. The order of C7 and C9 is determined by observing the long-range HETCOR correlations in Figure 83(ii); the peak assigned to C9 correlates with the neighbouring CH_2 protons and the C7 peak does not. Comparison can also made with the assignments[56,59] of pure FS form **I** and 2FS-INA with the assumption that the ordering of C10, C11 and C12 is the same in FS-PA. Additionally, C5 can be seen to be broader than the other quaternary peaks, which is likely due to quadrupolar broadening from the neighbouring chlorine atom as observed in 2FS-INA.[60-62] The PA peak assignment is aided by comparison with the peak ordering in the NPX-PA cocrystal, for example both carbon atoms adjacent to the pyridine nitrogen atom on PA fall at essentially the same chemical shift in NPX-PA, so it is reasonable to expect that the corresponding sites in FS-PA (C2' and C6') will also be of similar chemical shift.

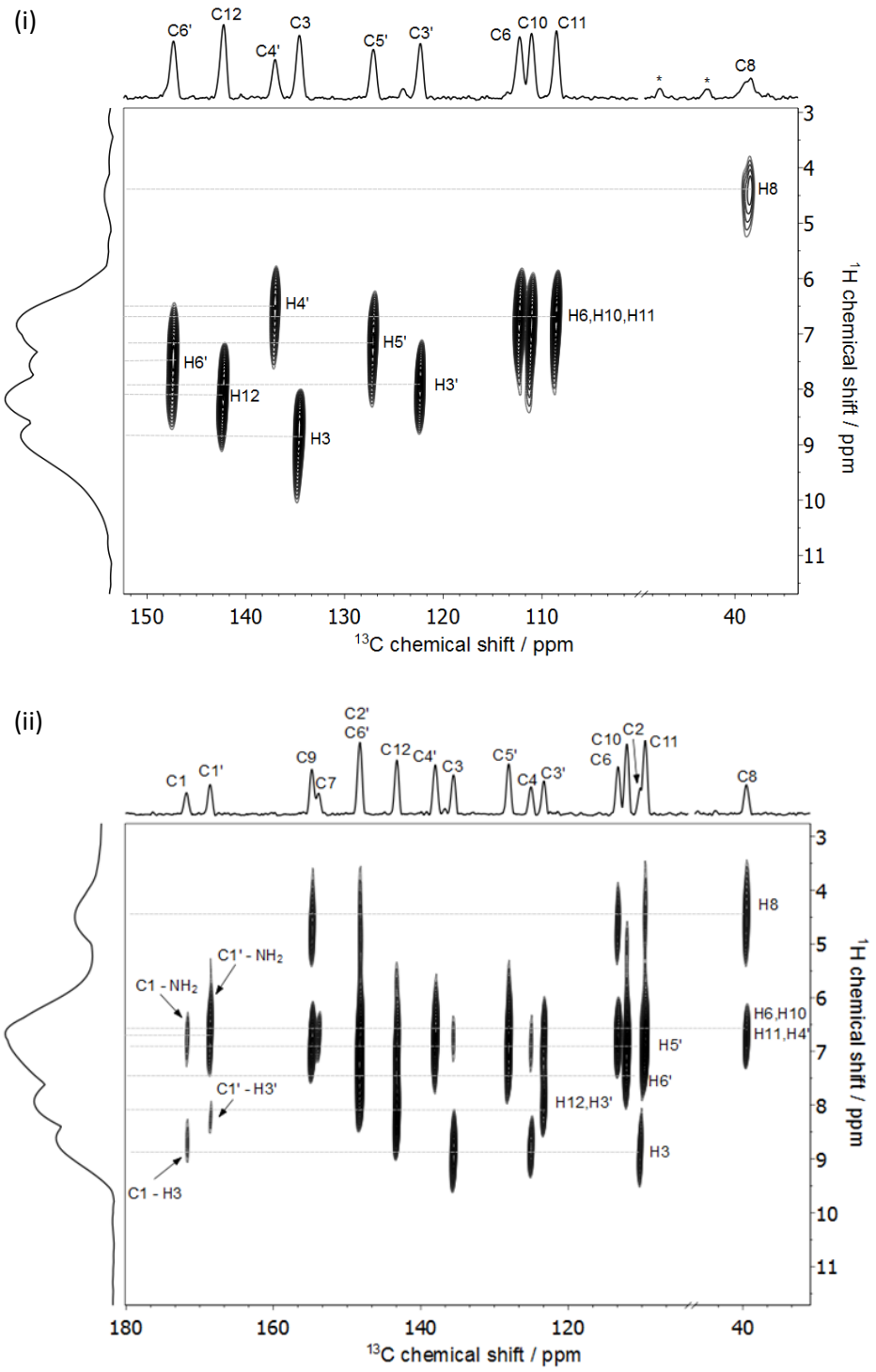

Figure 83 ^{1}H–^{13}C HETCOR spectra of FS-PA acquired with contact times of **(i)** 0.1 ms and **(ii)** 1 ms. Spectra were acquired at 10 kHz MAS with a recycle delay of 5 s with 64 t_1 increments each with 160 transients. Hydrogen labelling follows that of the carbon labels, *i.e*. the hydrogen atom bonded to C6' is labelled as H6'.

Hydrogen bonding

With the asymmetric unit of FS-PA identified, a further structural question is the identity of the synthons present and, subsequently, what the long-range hydrogen bonding motifs might be. Figure 84 shows synthons involving the FS carboxylic acid that have been observed in previous FS cocrystals, with the synthon labels following that of Figure 63 in Chapter 5 for consistency. An additional synthon, labelled as **S**, has not previously been observed in FS cocrystals, but has been observed in two PA cocrystals.[63,64] Chapter 4 of this work displayed the sensitivity SSNMR to hydrogen bonding in cocrystals, so it should be possible to distinguish between synthons **A**, **E**, **F** and **S** in FS-PA.

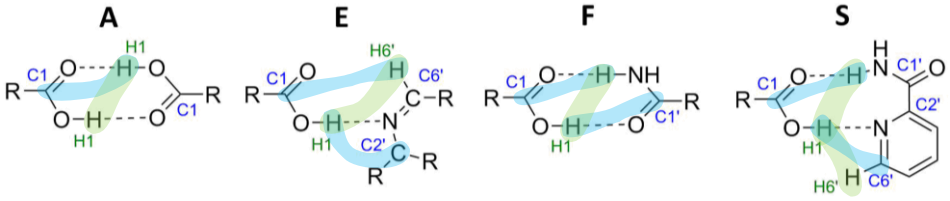

Figure 84 Possible synthons in FS-PA involving the carboxylic acid group of FS. Expected dipolar coupling pairs are indicated; ^{1}H–^{1}H green, ^{1}H–^{13}C blue.

The 1D ^{1}H spectrum of FS-PA is shown in Figure 85(i) and the carboxylic acid proton, H1, is clearly visible at 13.8 ppm. The high chemical shift is consistent with the proton being involved in a hydrogen bond.[65] There is little resolution in the rest of the spectrum due to overlap of the aromatic environments, though the FS CH_2 peaks may be assigned to the low shift peak at 4.4 ppm by comparison with the ^{1}H spectrum of 2FS-INA, see Figure 70 in Chapter 5.[59] It is noted that the sample used for ^{1}H experiments was stored in a D_2O atmosphere for 1 month prior to acquisition, but this is not expected to induce any changes in the structure and no mobile HDO is visible in the spectrum.

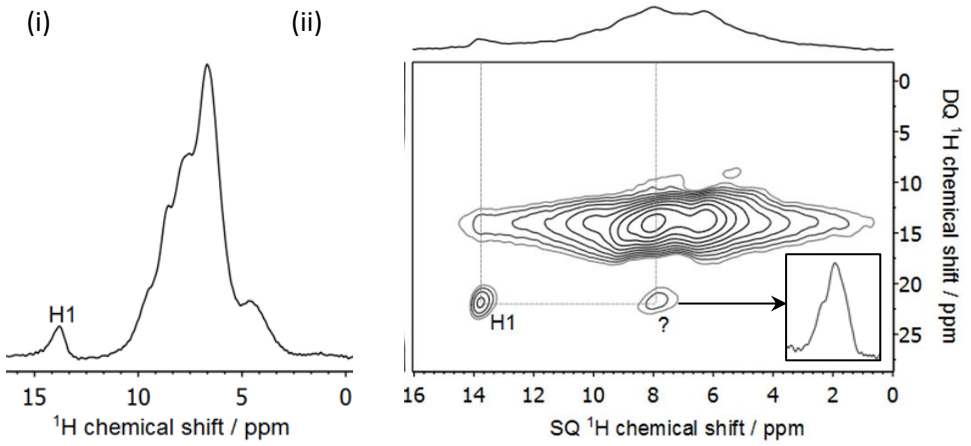

Figure 85 (i) ^{1}H spectrum of FS-PA acquired at 60 kHz MAS with 16 transients. **(ii)** ^{1}H–^{1}H DQ/SQ spectrum of FS-PA acquired at 60 kHz MAS using 8 rotor cycles in the recoupling period and an evolution time of 1 µs. 128 t_1 increments were acquired with 16 transients per increment and a 30 s recycle delay was used. Extracting a row through the correlation peak labelled "?" is shown as an inset.

Investigation of the homonuclear dipolar coupling to H1 is useful to help distinguish between the four possible synthons. A single correlation is observed in the DQ/SQ spectrum between H1 and a proton site at a SQ frequency of 8 ppm, Figure 85(ii). This is, again, indicative of H1 being involved in a hydrogen bond since there are no protons in close proximity to H1 on the FS molecule. Synthon **A** cannot be present because an interaction between two FS carboxyl groups would result in a correlation at a DQ frequency of 27.6 ppm, but the DQ frequency of the observed correlation is only 21.8 ppm. It is more challenging to discriminate between synthons **E**, **F** and **S** using the DQ/SQ spectrum. Synthon **S** is expected to give rise to two correlations, one to the NH_2 protons and one to H6', but only one correlation is observed to H1. However, a 1D slice taken through the correlation indicates that there may be multiple peaks overlapped, see Figure 85(ii) inset. The SQ frequency of 8 ppm is in a crowded region of the ^{1}H spectrum, making it hard to assign the signal. In the ^{1}H–^{13}C HETCOR spectra in Figure 83(ii) H6' is assigned to a shift of 7.5 ppm, while the PA NH_2 protons are assigned at 6.7 ppm. However, this difference in chemical shift is not sufficient to discriminate between the possible synthons using only the DQ/SQ spectrum.

More conclusive synthon discrimination is given by the long-range ^{1}H–^{13}C correlation observed between the carboxylic acid C1 and the PA NH_2 protons, Figure 83(ii). No correlation is observed between C1 and H6'. This is indicative of synthon **F**, though again

the proton spectrum is densely overlapped in this region. Additional evidence for synthon **F** is obtained by repeating the $^{1}H–^{13}C$ HETCOR with a longer recycle delay to observe long-range correlations to the carboxylic acid proton H1, Figure 86. The expected H1–C1 correlation is observed along with a second, weaker, correlation to C1'. No correlation is observed between H1 and C6'. This supports the presence of synthon **F** and not synthons **E** or **S**. Further attempts to directly observe the synthon *via* 2D SSNMR methods employing correlations beteen H1 and nearby nitrogen atoms were unsuccessful.

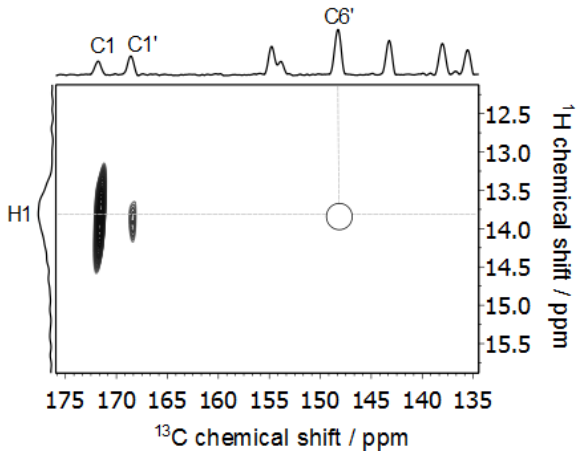

Figure 86 $^{1}H–^{13}C$ HETCOR spectrum of FS-PA acquired with contact times of 1 ms at 10 kHz MAS with a recycle delay of 30 s acquired over 64 t_1 increments each with 160 transients. The open circle denotes a missing H1–C6' correlation.

The pyridine nitrogen N1' is tentatively assigned to −80 ppm in Figure 87, which is consistent with the site being involved in a hydrogen bond.[59,66-69] The weak signal is possibly a result of poor CP transfer due to dynamic averaging, despite a long contact time of 5 ms. It is also likely that the CSA of this site is large since the solvates of FS-PA, discussed later in section 6.3, have a CASTEP-calculated N1' CSA of 500 ppm.

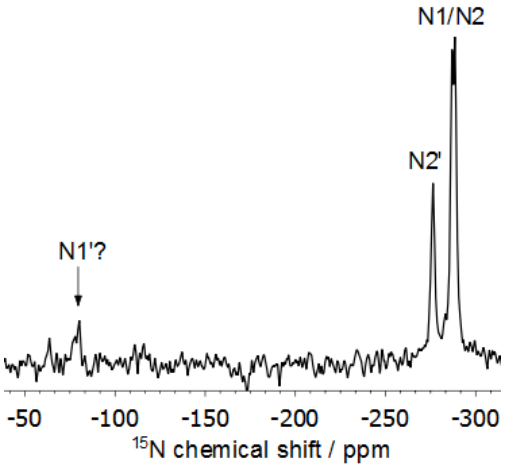

Figure 87 ^{15}N CP/MAS spectrum of FS-PA at 13.5 kHz MAS, acquired over 5552 transients with a contact time of 5 ms and a recycle delay of 30 s.

The chemical shifts of H1 and N1' are expected to be sensitive to the identity of the synthon in which they reside and so the chemical shift should be characteristic of the hydrogen bonding motif. A CSD search for the synthon **F** fragment, discussed previously in section 4.2.2, yields 200+ structures and a similar search for the synthon **E** fragment yields 600+ structures. NMR calculations on 62 of these database systems, following all-atom geometry optimisation, provide characteristic ranges of chemical shielding for each synthon, Figure 88. However, it is clear that the ranges of ^{1}H chemical shielding overlap extensively for the two synthons. These calculations are preliminary and cover only a small sample of the 800+ structures extracted from the CSD, so more conclusive results might be possible with a more complete structure set. The shielding ranges for N1' are more discrete, though again the sample size is very small and few structures possess synthon **F** in combination with a pyridine nitrogen atom that is not involved in a hydrogen bond. Crudely, the average shielding value of N1' in structures containing synthon **F** is −93 ppm, equivalent to a chemical shift of −76 ppm using $\delta_{iso} = \sigma_{ref} - \sigma_{iso}$ with $\sigma_{ref} = -169$ ppm.[*] The average N1' shielding value for structures containing synthon **E** is −58 ppm, equivalent to a chemical shift of −111 ppm. The experimental chemical shift on N1' in Figure 87 is −80 ppm, in keeping with synthon **F** rather than synthon **E**.

[*] Value taken from the σ_{ref} of 2FS-INA in Chapter 5.

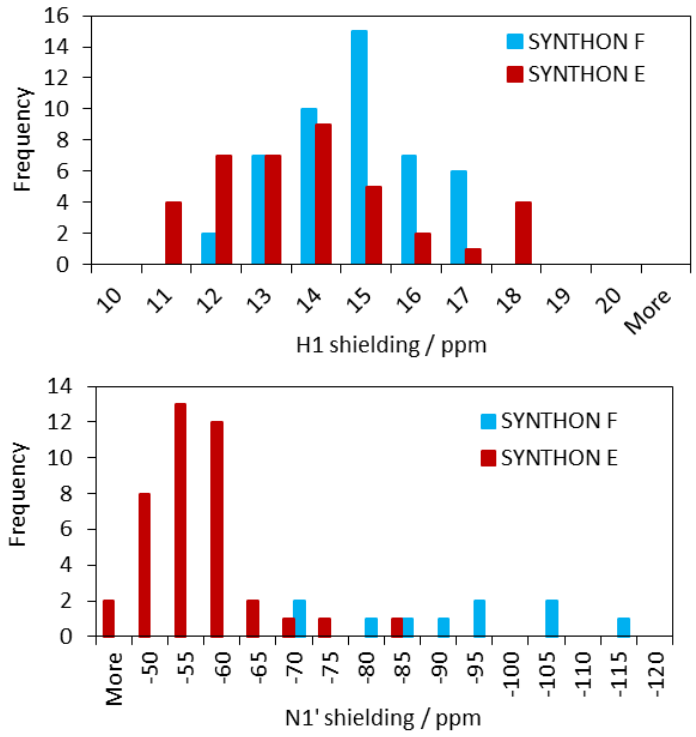

Figure 88 Histograms of calculated shielding for the carboxylic acid proton, H1, and the pyridine nitrogen atom, N1', over 62 structures extracted from the CSD containing the synthons **E** or **F**.

Disorder

There is a possibility that FS-PA contains disordered components, as observed previously for 2FS-INA at the sulphonamide group and furan ring.[59] Sulphonamide dynamics are challenging to observe by ^{1}H or ^{13}C SSNMR so the FS-PA sample was deuterated, as described in section 6.2.2, and the ^{2}H spectra are shown in Figure 89 at several temperatures. The high shift peak in the centre band, right of Figure 89, is the deuterated H1 and the low shift peak is likely to contain contributions from both the FS sulphonamide and the PA amide groups. Both peaks can be seen to broaden on cooling to −25 ˚C; the low shift peak broadens from 320 Hz to 460 Hz and the H1 peak broadens from 210 Hz to 320 Hz. This is indicative of the presence of some dynamics that are partially frozen out on cooling. The sideband manifold also changes subtly on cooling, particularly for the low shift peak as shown by the blue lines in Figure 89, which denote the maxima of the low shift peak in each sideband. However, only small changes are observed in the quadrupolar parameters following fitting in pNMRsim,[70] see Table 19. The C_Q of the sulphonamide peak is lower than that of the carboxyl peak, indicating that some dynamics may be present, but it is not as low as the C_Q of the 2FS-INA sulphonamide peak that was previously shown to

be dynamically disordered.[59] The dynamics of 2FS-INA are complex, comprising exchange of the NH_2 protons as well as rotation of the whole group about the C–S bond. The moderate averaging of the C_Q values of FS-PA suggest that the dynamics are less extensive, possibly involving only NH_2 exchange.

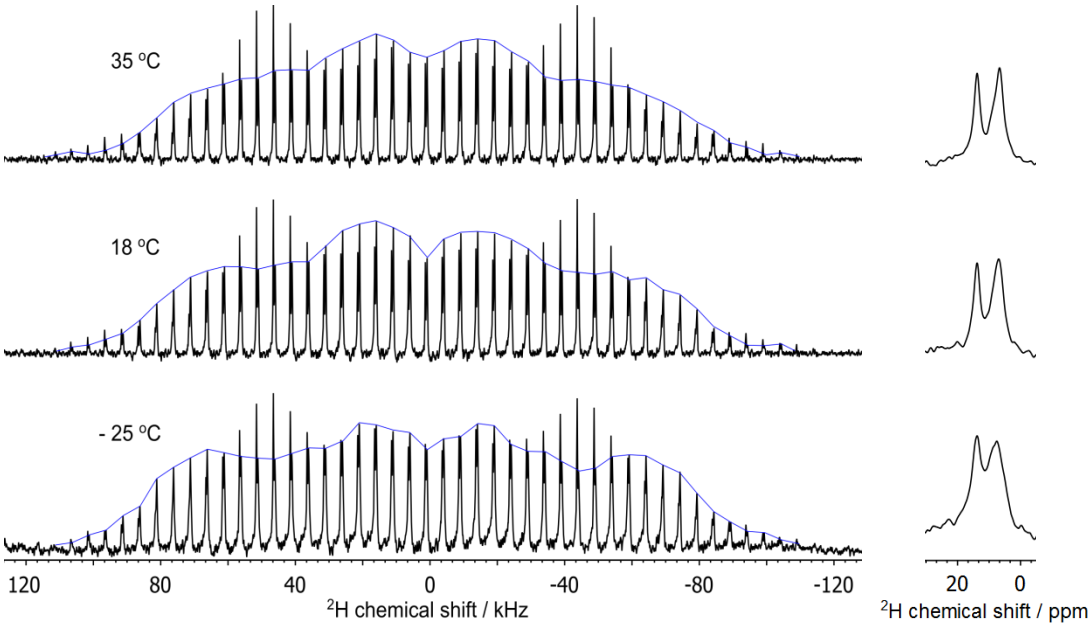

Figure 89 ^{2}H spectra of FS-PA acquired with a simple pulse-acquire sequence at 5 kHz MAS with a 5 s recycle delay. Centre peaks are shown on the right. The blue lines are guides for the eye to show the maxima of the low shift peak in the sideband manifold.

Table 19 Fitted ^{2}H quadrupolar coupling parameters at three temperatures plus the parameters of the 2FS-INA sulphonamide peak. One standard deviation of each fitted value is given in parentheses.

	−25 ˚C		18 ˚C		35 ˚C	
^{2}H site	C_Q / kHz	η	C_Q / kHz	η	C_Q / kHz	η
sulphonamide	121.9(2)	0.70(6)	121.6(2)	0.712(1)	120.2(2)	0.702(5)
carboxyl	153.3(2)	0.124(4)	152.1(2)	0.119(1)	152.0(2)	0.127(6)

Complementary information is available through the ^{2}H T_1 relaxation times, Figure 90. The relaxation times of the first pair of sidebands are shown, because the centre peak overlapped with a signal at 0 ppm, possibly arising from free D_2O that had absorbed onto the surface during storage for 9 months. This absorbed moisture is not expected to alter the bulk dynamics significantly, though this would need to be verified by comparing the

following relaxation analysis on a dry sample. A T_1 minimum is apparent for both sidebands at 280 K in Figure 90. Since the sidebands should have identical T_1 values, the average of the two datasets was fit to direct calculations of the ^{2}H relaxation rate under MAS, dotted line in Figure 90, using the procedure described by Apperley *et al*.[71] and also in section 5.3.4. The average difference between T_1 values for the two datasets is used as an estimate for the error on individual T_1 values (1.5 ms). The C_Q cannot be calculated in CASTEP without a crystal structure, so it is estimated to be roughly the same as 2FS-INA (222 kHz). The fit in Figure 90 corresponds to E_a = 33(3) kJ mol^{-1} and a jump angle of 102.9(5)˚. This angle is in keeping with the angle between the two protons in the sulphonamide NH_2 group. The correlation time for the motion at 20 ˚C is 108 x 10^{-11} s, corresponding to a jump rate of 900 MHz, which is consistent with the fact that the C_Q is observed to be averaged in the experimental ^{2}H spectrum. It should be noted that the lowest temperature data point for the −1 sideband was excluded to improve the quality of the fit; χ^2 lowered from 16 to 4. The calculated E_a for FS-PA is very similar to the E_a estimated for sulphonamide disorder in pure FS form **I** of 31 kJ mol^{-1},[72] supporting the presence of NH_2 proton exchange in FS-PA with a similar mechanism of motion to that of pure FS.

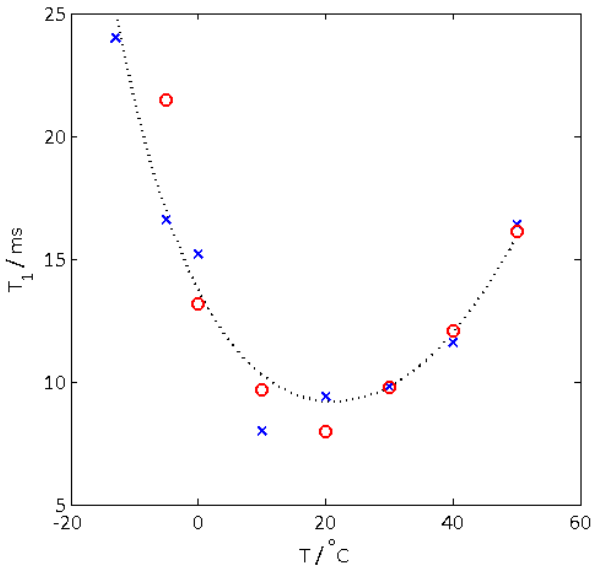

Figure 90 FS-PA sulphonamide ^{2}H T_1 relaxation times for the first pair of sidebands (blue crosses denote the +1 sideband and the red circles denote the –1 sideband) as a function of temperature fit to a two-site jump model with an Arrhenius-type temperature dependence of the jump rate. Temperatures include a correction of +10 K for sample heating from MAS at 10 kHz. ^{2}H T_1 times were measured by applying a saturation train of 20 pulses separated by a delay of 50 ms and the τ delay was varied between 1 µs–200 ms in 33 steps.

Another potentially disordered component is the FS furan ring. A direct excitation ^{13}C experiment with a 5 s recycle delay reveals three sharp peaks and a fourth is visible amongst the spinning sidebands, Figure 91(i). These are assigned as the carbon atoms on the furan ring of the FS molecule and indicate that these sites are dynamic, as observed in 2FS-INA.[59] The ^{13}C T_1 relaxation times are also fast, Figure 91(ii), suggesting the presence of dynamics of the order of 125 MHz. The T_1 relaxation times are fit to a simple Arrhenius-type equation in Figure 91(ii) to give $E_a = 22(2)$ kJ mol^{-1} on average. More sophisticated fitting is not possible since the precise T_1 minimum is not observed in the measured temperature range and higher temperatures were avoided to prevent sample degradation. Despite the simple model, the E_a values for all three furan sites of FS-PA are consistent with those calculated for 2FS-INA, indicating that a similar small amplitude motion is likely to be present.

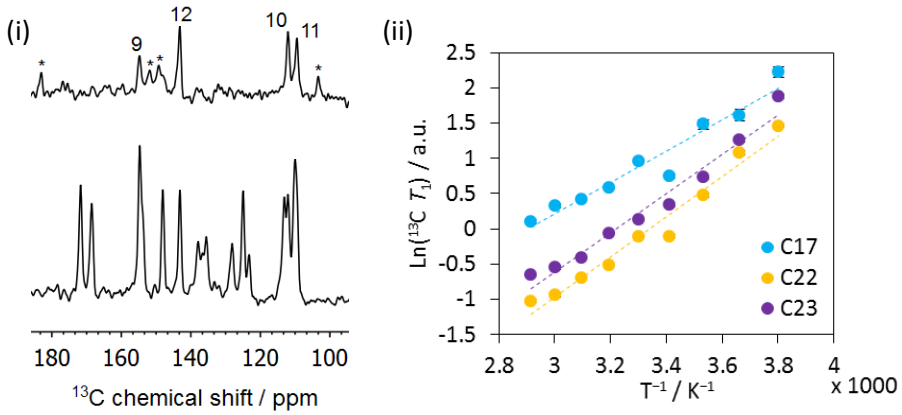

Figure 91 (i) ^{13}C direct excitation spectrum (top) and CP spectrum (bottom) of FS-PA acquired at 5 kHz MAS with a recycle delay of 5 seconds over 100 transients. **(ii)** An Arrhenius plot of ^{13}C T_1 relaxation times for the furan ring carbon. Temperatures do not include correction from sample spinning, since this is small at 5 kHz MAS. The ^{13}C T_1 times were measured by applying a saturation train of 20 pulses separated by a delay of 50 ms at 5 kHz MAS and the τ delay was varied between 0.05–5 s in 31 steps.

In summary, several important structural features of a new cocrystal FS-PA have been identified *via* SSNMR studies, despite the lack of diffraction data. The asymmetric unit is clearly observed in ^{13}C experiments as being Z' = 1 and a tentative indication of hydrogen bonding is observed between the carboxylic acid group of FS and the amide group of PA. Disordered components are also identified, namely the sulphonamide group and furan ring

of FS. This information should serve as a starting point for structural models fitted to future PXRD data.

6.3 Isostructural solvates

Two solvates of FS-PA have been identified following crystallisation experiments carried out by L. Softley of the I.R.E group. SCXRD data reveals that these acetone and ethanol solvates are isostructural with one another, Figure 92, and they are also isostructural with a recently published FS-PA hydrate.[26] All three structures crystallise in the space group P$\bar{1}$ with one FS and one PA molecule in the asymmetric unit. The unit cell parameters and other crystallographic details are shown in Table A1 in the Appendix. The sulphonamide group is refined with disorder over two positions with equal occupancies, similar to the disorder observed in 2FS-INA[57,59] but with a larger angle between the two orientations, 110˚ instead of 32˚. The shortest hydrogen bonding interaction is synthon **F** between the carboxylic acid of FS and the amide of PA (O1H1···O1'), with $d_{O\cdots O}$ = 2.55(1) Å in the acetone solvate and reported hydrate, and $d_{O\cdots O}$ = 2.592(5) Å in the ethanol solvate, Figure 92. Rotation about the FS C1–C2 bond is constrained by the presence of an intramolecular hydrogen bond (N2H···O2, $d_{N\cdots O}$ = 2.726(9) Å in the acetone solvate, $d_{N\cdots O}$ = 2.733(7) Å in the hydrate, and $d_{N\cdots O}$ = 2.767(5) Å in the ethanol solvate).

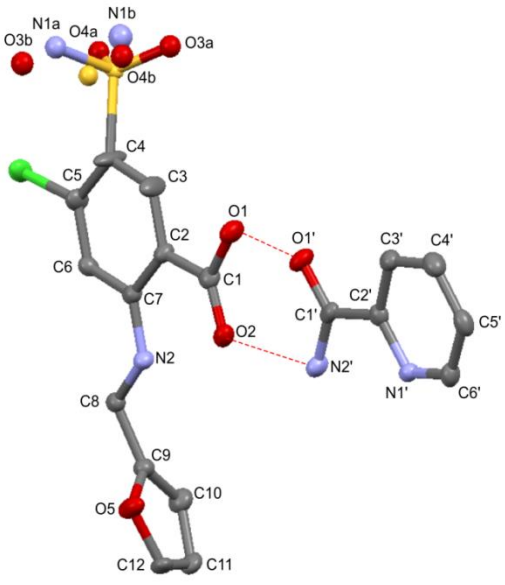

Figure 92 Asymmetric unit of the isostructural FS-PA solvates with thermal ellipsoids shown at the 50 % probability level. Disorder of the sulphonamide group is shown with split sites and hydrogen bonds are indicated by the red dashed line. Solvent molecules could not be modelled.

The system forms ribbons of alternating pairs of FS molecules connected to pairs of PA molecules. Neighbouring FS molecules form close contacts between sulphonamide groups ($d_{O3\cdots O3}$ = 2.16(2) Å), and neighbouring PA molecules interact in a dimer formation ($d_{N2'\cdots N1'}$ = 3.10(1) Å), Figure 93(i). The ribbons are loosely connected through the sulphonamide groups *via* synthon **J** ($d_{N1\cdots O3}$ = 2.68(3) Å), Figure 93(ii). The bond distances given here are for the acetone solvate, but the same hydrogen bonding is present in the ethanol solvate with heavy atom distances changing by less than 2 %. Synthon **J** is the exception with $d_{N1\cdots O3}$ increasing by 12 % in the ethanol solvate, reflecting a small elongation in the *a* axis, see Table A1.

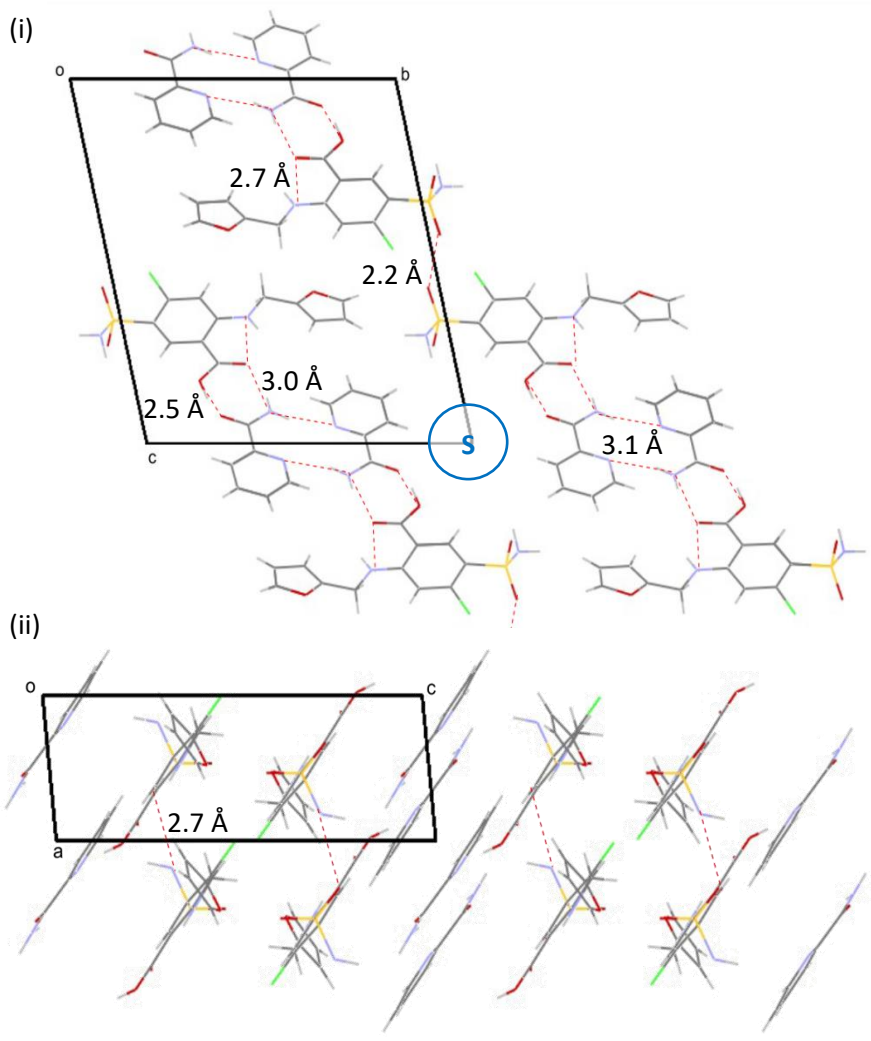

Figure 93 Diagrams of the packing in the FS-PA acetone solvate with hydrogen bonds indicated by dashed red lines and heavy atom distances given. Only one orientation of the disordered sulphonamide group is shown for clarity. **(i)** FS-PA ribbons viewed along *a* axis with the solvent molecule shown schematically as circles labelled "S". **(ii)** Perpendicular view along *b* axis showing synthon **J** linking the ribbons.

Fourier difference maps of both solvates showed evidence of electron density that is separate from the FS and PA framework. This is assumed to be the solvent, but the geometry could not be refined for either the acetone or ethanol molecules. As a result, the SCXRD structures discussed above were solved with the PLATON SQUEEZE approach.[29] The FS-PA framework packs in such a way that channels are present in the structure, Figure 93(i), in which the additional electron density resides. This is suggestive of a channel solvate. NMR crystallography investigations are necessary to identify whether solvent molecules are present. Additionally, it is likely that the solvent molecules are disordered so the nature of this disorder is investigated to characterise dynamics and to identify any interactions between the solvent molecules and the host lattice.

6.3.1 Experimental details

Fresh FS-PA acetone and ethanol solvate samples were prepared by LAG grinding of 1:1 molar quantities of the starting materials, 0.165 g FS (0.5 mmol) and 0.061 g PA (0.5 mmol). For the acetone solvate, FS and PA were ground in 1 mL acetone for 5 mins. A further 0.5 mL acetone was added every 10 mins until the total LAG time was 30 mins. Deuterated acetone solvate was produced by the same method using acetone-d_6. For the ethanol solvate, the FS and PA components were ground for 5 mins before 10 drops ethanol were added every 5 mins until the total grind time was one hour. Both syntheses yielded slightly damp cream powders, which were left to dry in air for one hour before packing in 4 mm rotors. Samples were kept stored in NMR rotors to limit desolvation.

To carry out CASTEP calculations the solvent molecules were modeled in the position where electron density was observed by XRD using the Avogadro molecular editor.[73] Only one orientation of each solvent molecule was modelled (a test calculation found there is < 0.2 eV between the energies with two different orientations of the ethanol molecule). Geometry optimisation was carried out for all atom positions, and NMR parameters were averaged for each site. Note that all-atom geometry optimisation cannot be carried out for the structure with the solvent molecules deleted because the PA heavy atoms are displaced by 1.8 Å on average in an attempt to fill the void left by the solvent molecule. Therefore, the “dry” structure was created by deleting the solvent molecules from the optimised acetone solvate structure and then optimising the hydrogen positions only.

The EXPRESS program[41] was used to model the effect of acetone dynamics on the ^{2}H spectrum of the FS-PA acetone solvate. This simulates the effects of Markovian jump dynamics between discrete positions. A motional model is described by defining the

number of sites, their relative populations, and their relative orientations with Euler angles (α,β,γ), as well as their connectivity if more than two sites are considered. The theory is detailed by Vold and Hoatsen,[41] and EXPRESS is implemented as a module in MATLAB. The ^{2}H spectra of the acetone solvate were simulated by modelling a two-site jump motion of 115˚ corresponding to a C_2 rotation about the C=O axis using the parameters in Table 20.

Table 20 Parameters for a simple two-site jump model of motion in FS-PA acetone solvate using EXPRESS to simulate the ^{2}H spectrum.

Experiment	^{2}H quadrupole echo		
τ_{90° / μs	3		
Echo delay / μs	60		
Dwell time / μs	2.5		
ν_{2H} / MHz	76.71		
Recycle delay / s	1		
Euler angles/ ˚	α	β	γ
Site 1	0	0	0
Site 2	0	115	0
Relative weighting	0.5		
C_Q [a]/ kHz	62		
η	0		
Tiling scheme	ZCW 500		
Left shift	24		

[a] The effective quadrupole tensor of all acetone ^{2}H sites averaged over C_3 methyl rotation.

6.3.2 Solvate structure

Proof of cocrystal formation is seen in Figure 94; the FS-PA acetone and ethanol solvates show distinct ^{13}C SSNMR spectra compared to the starting materials and the parent FS-PA cocrystal. Note that the peaks of PA are weak because the ^{1}H T_1 relaxation time is of the order of 100 s, since the rigid nature of the molecule results in few relaxation processes. The solvate spectra are broad compared to FS-PA and pure FS, 150–200 Hz, but no peak splitting is observed at the carbonyl region (C1, 172.2 ppm and C1', 169.9 ppm). This is in keeping with the asymmetric unit containing a single FS and a single PA molecule. The coloured arrows indicate additional peaks that cannot be assigned to either FS or PA and so are assumed to be the solvent molecules. The acetone carbonyl peak is not observed, but should appear at around 200 ppm. Previous literature of acetone solvates has also shown weak acetone carbonyl signals, ascribed to inefficient CP due to motional processes interfering with the magnetisation transfer during the contact time.[36] Little difference is

observed in the FS and PA chemical shifts of the two solvates, in keeping with the isostructural relationship observed from SCXRD data.

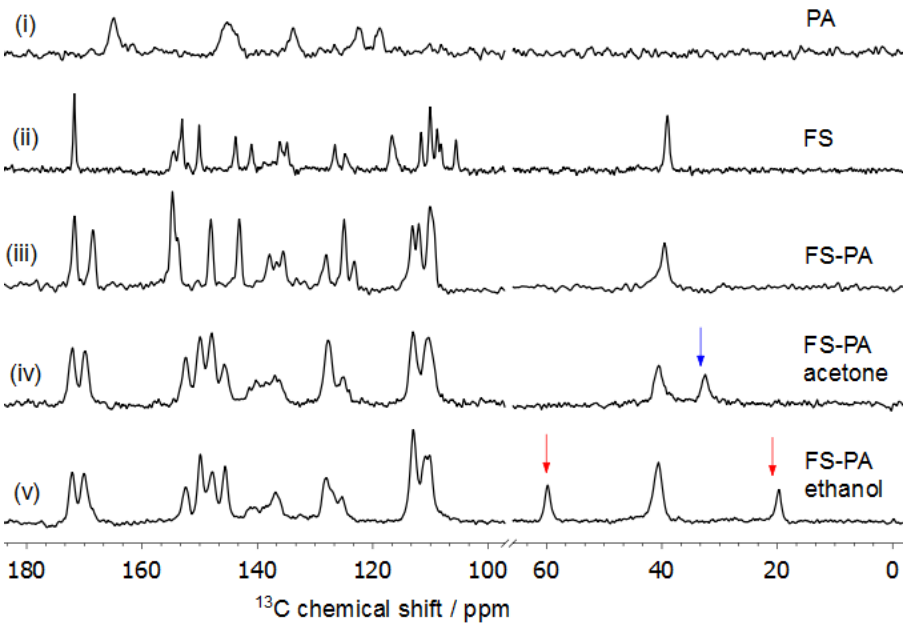

Figure 94 ^{13}C CP/MAS spectra of **(i)** PA, **(ii)** FS, **(iii)** FS-PA, **(iv)** FS-PA acetone solvate and **(v)** FS-PA ethanol solvate. Spectra **(i)**, **(ii)** and **(iv)** acquired at 8 kHz MAS and spectra **(iii)** and **(v)** acquired at 5 kHz MAS. The recycle delay was 5 s and contact time 2.5 ms over 100 transients for all except **(i)** recycle delay 30 s with contact time 4 ms, **(ii)** recycle delay 15 s with contact time 4 ms and **(v)** 480 transients. The blue arrow denotes the acetone CH_3 peak and the red arrows denote the ethanol peaks.

Full assignment of the ^{13}C spectra for the acetone solvate is given in Figure 95 and Table 21, and the same labelling applies to the ethanol solvate. The assignment was achieved with the aid of ^{13}C CP/TOSS/NQS spectra, Figure 95(ii), 2D ^{1}H–^{13}C HETCOR experiments with contact times of 0.2 ms, Figure 96, and 1 ms, Figure 97. The carbon assignment order is the same as that of the parent FS-PA cocrystal, though there is a higher degree of peak overlap. fFor example, C6 and C10 are indistinguishable in the solvates but can be individually assigned in FS-PA.

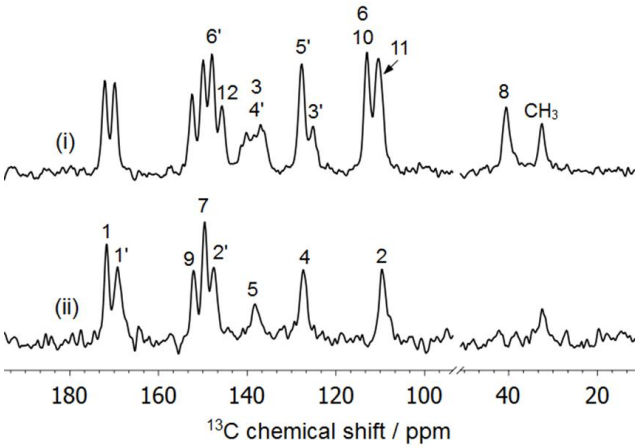

Figure 95 ^{13}C spectra of the FS-PA acetone solvate with labelling as in Figure 92. **(i)** CP/TOSS at 8 kHz MAS with non-quaternary sites labelled and **(ii)** CP/TOSS/NQS at 5 kHz MAS with quaternary sites labelled. Spectra were acquired over 100 transients with a 5 s recycle delay and 2.5 ms contact time. A dipolar dephasing delay of 80 µs was used to suppress non-quaternary sites.

Table 21 Carbon atom assignment for the FS-PA solvates and accompanying notes.

Atom Label	Chemical shift / ppm	Notes[a,b]	Atom Label	Chemical shift / ppm	Notes[a,b]
C8	40.5	F, L: NH 2.0 Å	C12	145.8	S?, F
C2	109.6	F, L: H3 2.0 Å, NH? 2.6 Å	C2’	147.5	Q, F, L? H3’ 2.0 Å, NH_2 2.6 Å
C11	110.8	Q, F, S?, L? sulph NH_2 1.9 Å, H12 2.4 Å	C6’	147.9	S?, F
C6	113.0	S?, F	C7	149.7	Q, L? NH 1.9 Å, H6 2.0 Å
C10			C9	152.2	Q
C3’	124.9	S, F, L? H4’ 2.0 Å	C1’	169.2	Q, F, L: NH_2 1.9 Å, H3’? 2.6 Å
C4	127.4	Q, F, L? H3 2.0 Å	C1	171.8	Q, F, L: NH 2.6 Å, H3 2.6 Å
C5’	127.7	S, F, L? H4’,H6’ 2.0 Å	Acetone CH_3	32.7	L: CH_3
C3	136 - 141	F?, L? H11 2.0 Å	Acetone C=O	Not observed	
C4’			Ethanol CH_3	19.7	
C5	138.2	Q	Ethanol CH_2	59.8	

[a] Symbols used to indicate the basis of assignment: Q = ^{13}C peak in non-quaternary suppression spectrum, S = cross peak visible in short contact time HETCOR experiment, L = cross peak visible in long contact time HETCOR experiment, ? = evidence is suggestive rather than definitive, F = ordering follows that of FS-PA.

[b] Distances corresponding to long distance correlations are given from the acetone solvate structure to one decimal place.

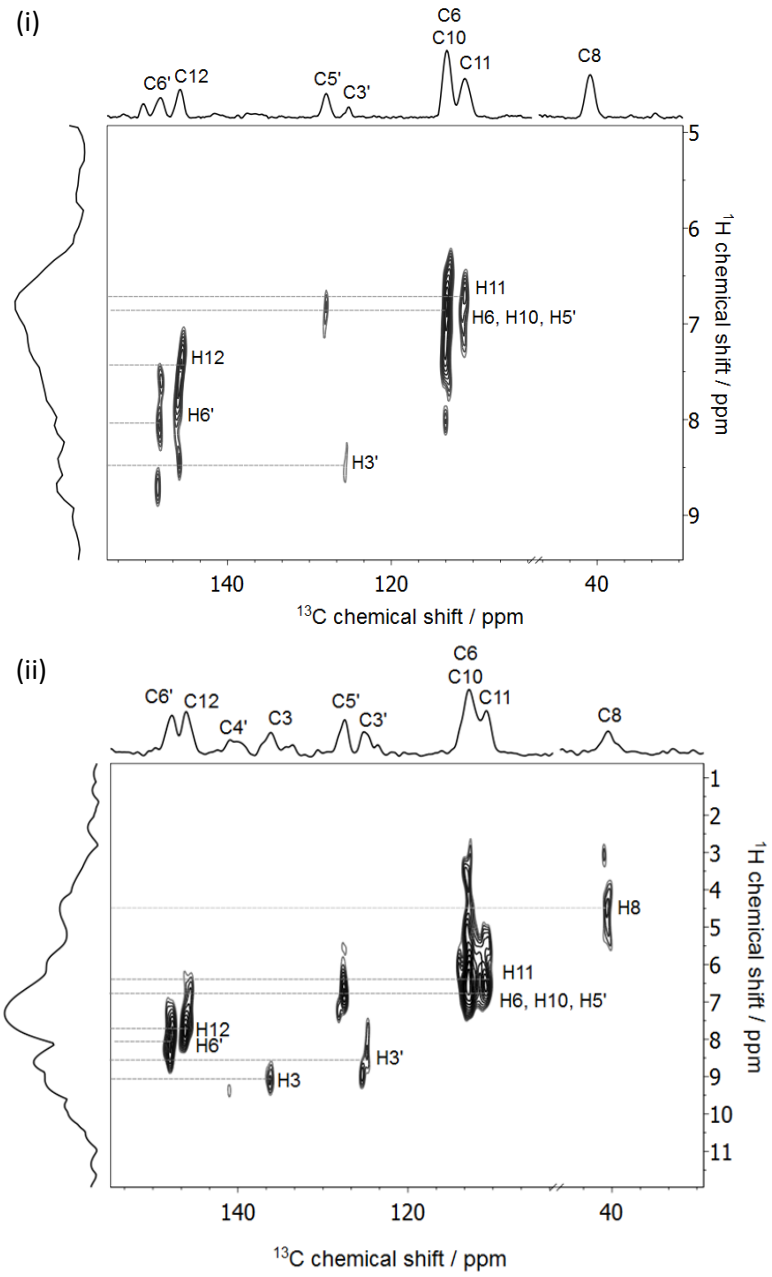

Figure 96 ^{1}H–^{13}C HETCOR spectra of FS-PA solvates. **(i)** Fresh acetone solvate acquired at 8 kHz MAS with contact times 0.2 ms over 128 t_1 increments each with 64 transients. **(ii)** 5 month old ethanol solvate acquired at 10 kHz MAS with contact time 0.1 ms over 64 t_1 increments each with 100 transients. Spectra were acquired with a recycle delay of 5 s. The hydrogen labelling follows that of the carbon labels.

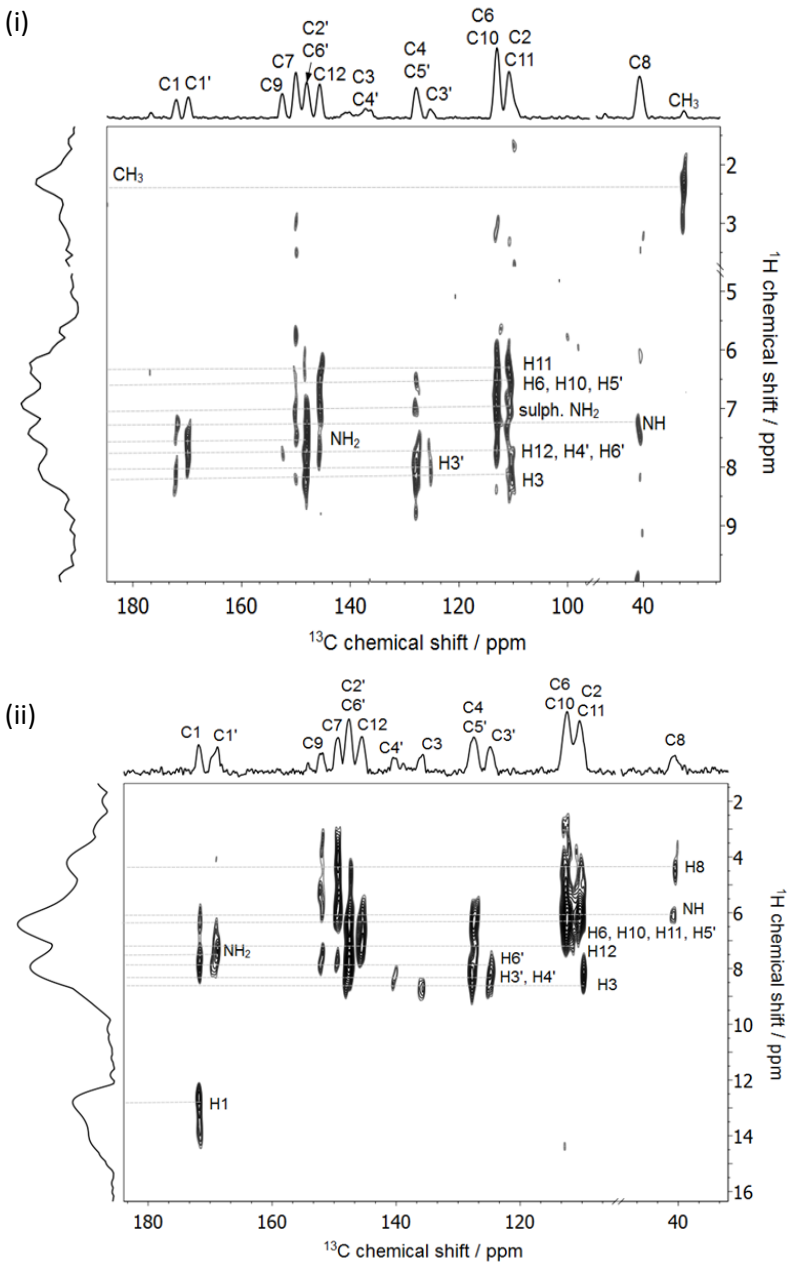

Figure 97 ^{1}H–^{13}C HETCOR spectra of FS-PA solvates acquired with a recycle delay of 5 s and a contact time of 1 ms. **(i)** Fresh acetone solvate acquired at 8 kHz MAS over 128 t_1 increments each with 64 transients. **(ii)** 5 month old ethanol solvate acquired at 10 kHz MAS over 64 t_1 increments each with 100 transients.

The ethanol sample used to acquire the HETCOR spectra was five months old, so the solvent peaks are too weak to be observed. However, the acetone spectra were acquired with a fresh sample so the solvent CH_3 group is visible in the long contact time spectrum, Figure 97(i), with a ^{1}H shift of *ca*. 2.5 ppm. No other correlations are visible to this peak, highlighting the lack of intermolecular interactions between the FS-PA framework and the

solvent molecules. The FS-PA solvates can be described as "void-filling" solvates, similar to the isostructural solvates of droperidol.[23] The FS carboxylic acid proton, H1, is visible in the long contact time HETCOR of the ethanol solvate, Figure 97(ii), with a chemical shift of *ca*. 13 ppm. This is characteristic of a hydrogen-bonded proton, in keeping with the presence of synthon **F** as observed by XRD. It is noted that the chemical shift is 0.8 ppm lower than that of H1 in the parent FS-PA cocrystal, which is hypothesised to also contain synthon **F**.

6.3.3 Desolvation behaviour

To characterise the FS-PA solvates as stoichiometric or nonstoichiometric, the amount of solvent in the lattice must be quantified and then the desolvation process monitored to identify degradation. SSNMR is not an ideal technique for quantification, but the ^{13}C T_1 relaxation times of the solvent molecules are relatively short; 7 s for the acetone CH_3, 0.4 s and 0.5 s for the ethanol CH_3 and CH_2 respectively. Additionally, the peak between 110–113 ppm has four carbon sites with a short relaxation time of *ca*. 2 s. Therefore, a direct excitation experiment with a recycle delay of 1 min is sufficient to allow quantitative integration of the solvent peaks and also peaks in the FS-PA framework for reference. The acetone, Figure 98(i), and ethanol, Figure 98(ii), solvates are both found to contain solvent:FS ratios of less than 1:1. The acetone CH_3 peak integrates to 0.82, meaning there are *ca*. 0.4 molecules of acetone for every one of FS or PA after five months. The same number of ethanol molecules is present after two weeks.

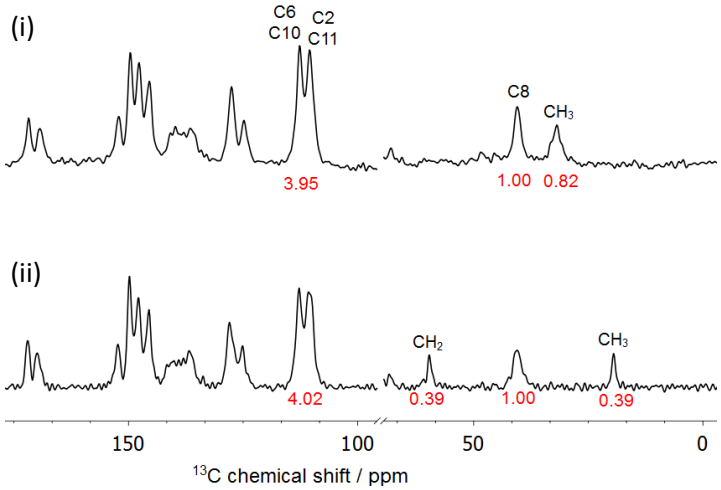

Figure 98 ^{13}C direct excitation experiments of **(i)** 5 month old acetone solvate and **(ii)** 2 week old ethanol solvate. Spectra acquired at 10 kHz MAS with a recycle delay of 1 min. Peak integrations are shown in red.

While individual CP experiments are not quantitative, it is reasonable to use the peak heights of successive experiments to compare relative solvent quantities, provided the experimental parameters are kept identical. The solvent peak intensities reduce over time, even with the sample stored in a sealed 4 mm rotor and no changes are observed to the FS or PA peaks. The ratio of peak heights is shown in red in Figure 99, rather than peak integrations, since the evidence is only qualitative. Nevertheless, it is indicative of the FS-PA solvates being nonstoichiometric, where the lattice can incorporate a variable number of solvent molecules in the channels without undergoing degradation. No evidence of degradation is observed in the ^{13}C spectra of either of the solvates after 9 months, indicating that only a trace amount of solvent is necessary to support the channel structure. Subsequent regrinding of a 9 month old acetone sample with 0.5 mL acetone increased the quantity of acetone incorporated in the structure, again highlighting the nonstoichiometric nature of the system.

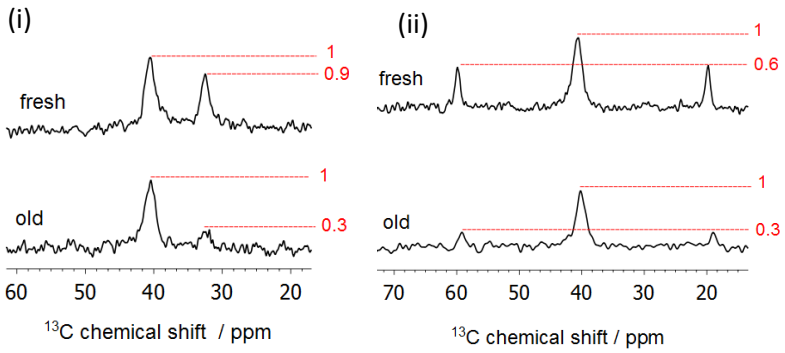

Figure 99 Comparison of relative peak heights of ^{13}C CP/TOSS spectra of the FS-PA solvates acquired at 5 kHz MAS with a 5 s recycle delay and a 2.5 ms contact time over 100 transients. **(i)** Acetone solvate where "old" denotes the sample after 3 month storage in a 4 mm rotor. **(ii)** Ethanol solvate where "old" denotes the sample after 4 month storage in a sealed vial. The FS CH_2 peak intensity is set to 1 as a reference.

The lack of significant changes in the ^{1}H or ^{13}C spectra over several months is perhaps unexpected, since the solvent molecules reside in channels that are bordered by the aromatic PA rings, Figure 93(i). However, CASTEP-calculated shielding differences demonstrate that the ^{1}H shieldings are predicted to change little going from the theoretical "dry" cocrystal to the two solvates, Figure 100(i), in keeping with the experimental data. The ^{13}C shieldings are also only subtly affected by the presence of solvent, Figure 100(ii),

and it has previously been observed in sildenafil citrate hydrate that differences such as these are overestimated compared to experimental changes in shift upon sample drying.[74] Therefore, the FS-PA framework is effectively independent of the nature/presence of solvent molecules in keeping with classification as a nonstoichiometric solvate system.

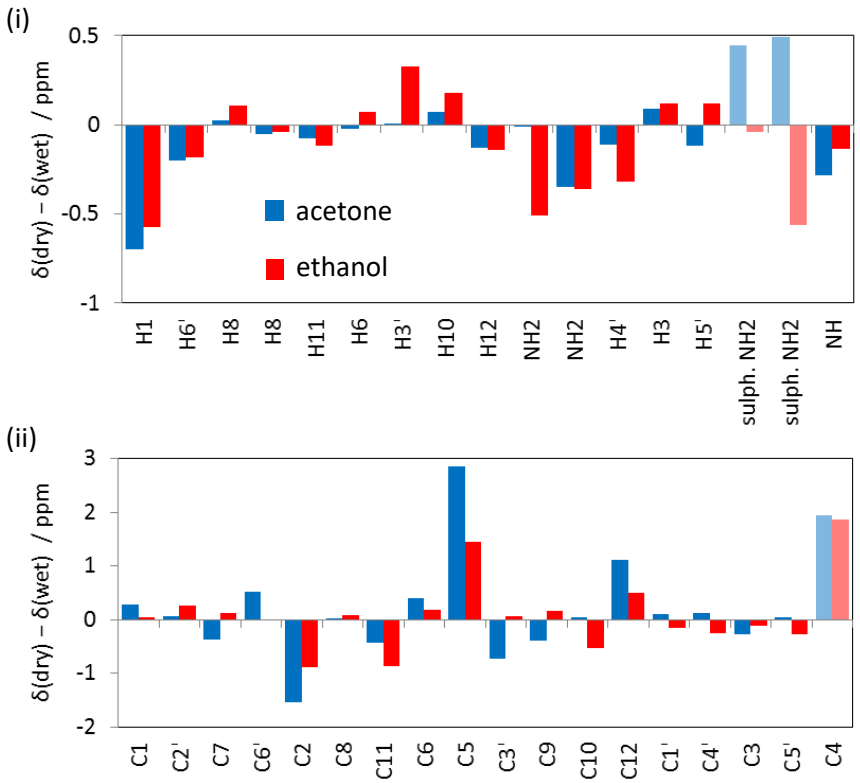

Figure 100 Differences between **(i)** ^{1}H and **(ii)** ^{13}C chemical shieldings of the "dry" structure and the two FS-PA solvates. The average is taken for each site over the unit cell and translucent data correspond to sites dominated by dynamics.

Degradation of the solvates is induced by heating to 80 ˚C, Figure 101(i). No changes are observed in the ^{13}C spectrum until 60 ˚C, at which point new peaks appear at 141 ppm and 151 ppm. At 65 ˚C a further peak appears at 114 ppm and the acetone CH_3 peak reduces in intensity. Additional changes are observed throughout the spectrum as the sample is heated further and maintaining a temperature of 80 ˚C for 21 hours results in complete degradation. The degradation products appear to be a combination of the pure components, FS and PA, and the parent FS-PA cocrystal, Figure 101(ii), though other techniques like DVS would be more conclusive. This suggests that the FS-PA solvates cannot form an isostructural desolvate and the channels in the FS-PA lattice collapse without the

solvent molecules. The degradation product is not amorphous, as can be the case for some desolvated solvates,[75] since the ^{13}C SSNMR linewidths are similar to those of the parent FS-PA cocrystal.

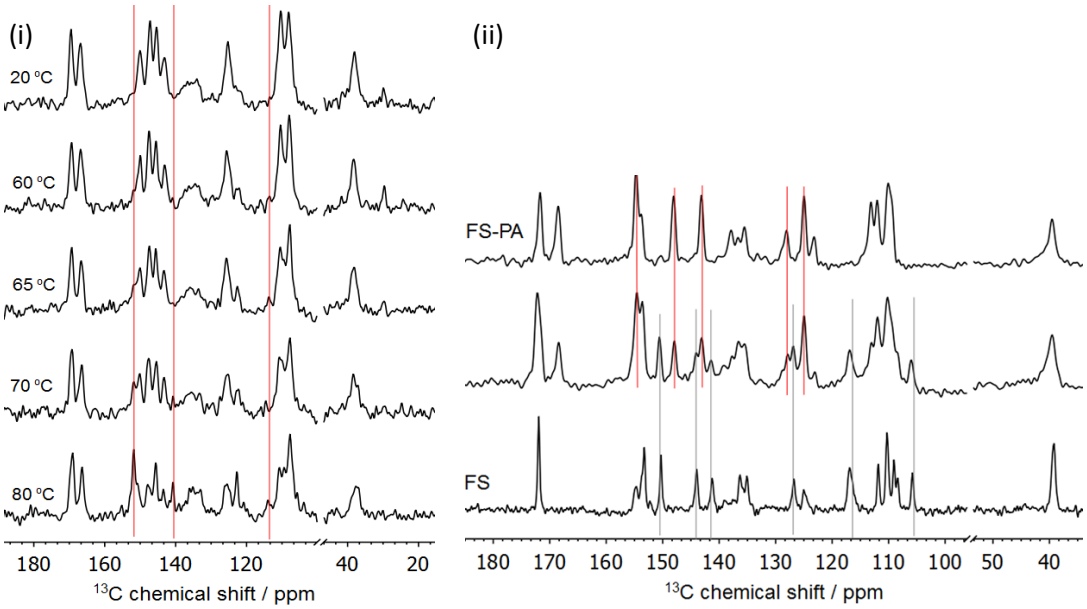

Figure 101 (i) Variable temperature ^{13}C CP/TOSS spectra of FS-PA acetone solvate at 5 kHz MAS with a recycle delay of 5 s and a contact time of 2.5 ms acquired over 200 transients. Red lines are guides for the eye indicating the positions of degradation product peaks. **(ii)** ^{13}C CP/TOSS spectra of FS-PA (top), acetone solvate following degradation by heating at 80 ˚C for 21 hours (middle) and pure FS (bottom). Lines are guides for the eye indicating peaks in the degradation product corresponding to FS-PA (red) and pure FS (grey).

The thermal behaviour is more accurately quantified using TGA and DSC. The TGA curve in Figure 102(i) shows a 30 % mass loss for the acetone solvate between 100–170 ˚C. If this is caused only by the loss of acetone from the sample then it would correspond to 3.3 acetone molecules for every one FS and PA molecule. This contradicts the NMR evidence above of *ca*. 0.4 molecules of solvent per asymmetric unit. However, further thermal analysis using DSC indicates that there are two thermal events over the 100–170 ˚C temperature range: a small endothermic event centred at 110 ˚C followed by a large endothermic event at 138 ˚C, Figure 102(ii). The small peak is in keeping with previously observed desolvation behaviour in other pharmaceutical solvates[39,40,76,77] and is therefore ascribed to solvent evaporation. The second corresponds to an endothermic melt or decomposition. No evidence of sample melting is observed in the variable temperature

SSNMR above, but decomposition is evident upon full desolvation. Additionally, a visual inspection of the sample on heating to 150 ˚C showed discoloration and degradation occurred between 135–150 ˚C. Therefore, the 30 % mass loss in the TGA curve is a combination of two closely spaced thermal events: first the onset of solvent evaporation at 100 ˚C, resulting in a 6.7 % mass loss, followed by decomposition at 140 ˚C associated with a further 23 % mass loss. The mass loss of 6.7 % corresponds to 0.5 acetone molecules per asymmetric unit, which is in keeping with the solvent quantification estimated by SSNMR.

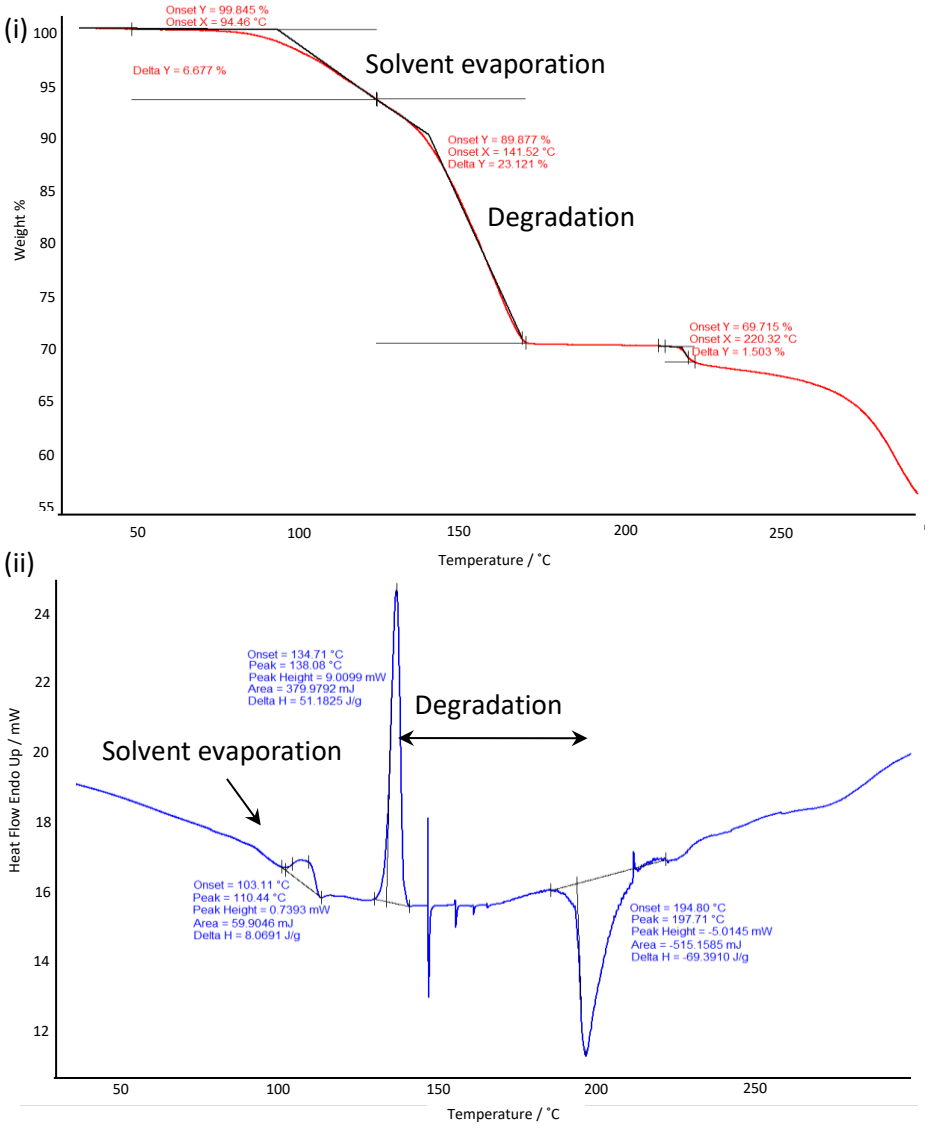

Figure 102 (i) TGA curve and **(ii)** DSC trace of FS-PA acetone solvate, acquired by D. Carswell, over the temperature range 25–300 ˚C with a sample heating rate of 5 ˚C min^{-1}. Major thermal events are annotated.

6.3.4 Solvent dynamics

It is common for nonstoichiometric solvates to be dynamic,[9,74,78] and the channel structure of the FS-PA solvates combined with the lack of strong intermolecular interactions between the framework and the solvent molecules are both likely to facilitate solvent motion. Few changes were observed in the ^{13}C spectra of the acetone solvate above on heating. However, some subtle changes are observed on cooling to −25 ˚C, Figure 103. The C2/C11 and C4/C5' peaks split and broaden, and the intensities of the solvent peaks reduce. These changes are reversible indicating that they are caused by the freezing out of some dynamics in the system, rather than desolvation.

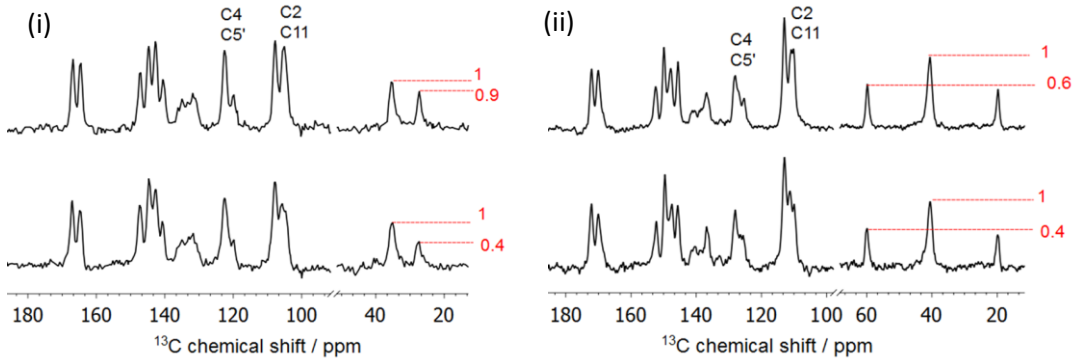

Figure 103 ^{13}C CP/TOSS spectra of FS-PA **(i)** acetone solvate and **(ii)** ethanol solvate, acquired at room temperature (top) and −25 ˚C (bottom) with 5 kHz MAS, a 5 s recycle delay and a 2.5 ms contact time over 100 transients. Relative peak intensities are shown in red with the FS CH_2 peak set to 1 as a reference.

It is challenging to link the subtle spectral changes to a motional model. The ^{13}C T_1 relaxation of the acetone CH_3 peak is also unlikely to be useful as it will be dominated by fast methyl rotation, masking any larger motion. However, it has previously been shown that the ^{13}C linewidth of the ethanol CH_2 peak can provide information on the dynamics of solvent molecules.[24] The linewidth of the ethanol CH_2 peak in the FS-PA solvate increases from 160 Hz to 190 Hz as the temperature is lowered to −30 ˚C, though the correlation is not good enough to be fit to estimate an E_a. A better correlation is observed in the ^{13}C T_1 relaxation times of the two ethanol peaks, Figure 104. The relaxation times are short for both of the ethanol carbon sites, compared to 2–3 s for the FS furan ring carbon atoms and more than 15 s for the rest. The decrease in ^{13}C T_1 relaxation times on cooling is consistent with approaching a T_1 minimum. The data were fit to Arrhenius curves, dashed lines of

Figure 104, yielding E_a values of 8(2) kJ mol^{-1} for the CH_2 and 23(3) kJ mol^{-1} for the CH_3. It is noted that more robust values would be obtained by fitting more extensive relaxation data encompassing the T_1 minima, and the errors stated are taken from linear regression and are therefore likely to be underestimated. The CH_3 E_a is consistent with rotational diffusion[71,79] but the CH_2 E_a is much lower. It is likely that there is a contribution to the CH_2 relaxation from cross-relaxation of the rapidly relaxing CH_3 protons,[24] but the observed line broadening means there must also be a dynamic component at −30 ˚C.

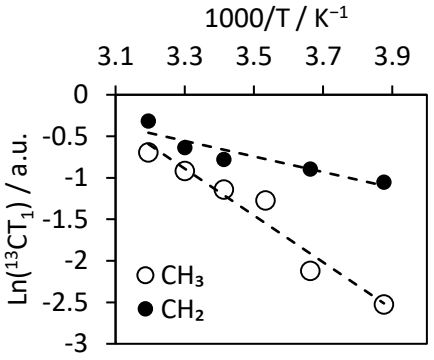

Figure 104 ^{13}C T_1 relaxation times of the ethanol solvent fit to an Arrhenius model. Temperatures include a +10 K correction from sample spinning at 10 kHz. ^{13}C T_1 times were measured by applying a saturation train of 20 pulses separated by a 50 ms delay at 10 kHz MAS, and the τ delay was varied between 0.05–5 s in 31 steps with 160 transients in each.

^{2}H SSNMR experiments are not expected to be informative on the dynamics of the ethanol solvate, since the use of ethanol-d_6 would cause exchange of all of the labile protons and ethanol-d_5 is prohibitively expensive. The acetone solvate is more readily studied with ^{2}H SSNMR by incorporating acetone-d_6. The ^{13}C CP/TOSS spectrum of the deuterated acetone solvate is identical to that of the non-deuterated sample, except that the acetone CH_3 peak is no longer visible because there are no protons on the acetone molecule from which magnetisation can be transferred *via* CP. The ^{2}H spectrum has a broad sideband manifold, highlighting the fact that the acetone is incorporated in the solvate structure, Figure 105. However, the sidebands do not extend as far out as would be expected for a rigid ^{2}H site[80] and the quadrupolar parameters reflect this; C_Q = 24.9(2) kHz and η = 0.80(4). The C_Q is low compared to the CASTEP-calculated C_Q of 186 kHz. Cooling to −45 ˚C induces subtle spectral changes and the fitted quadrupolar parameters become C_Q = 42.3(2) kHz and η = 0.33(2). This increase in C_Q is indicative of the freezing out of some dynamics.

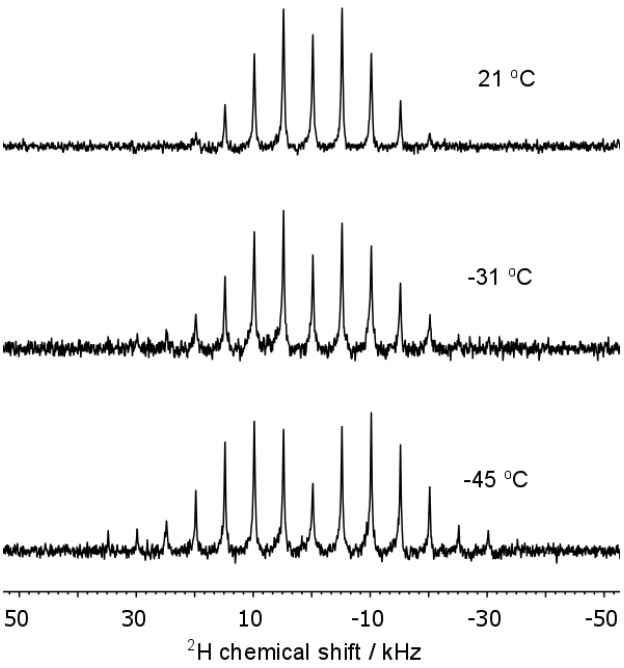

Figure 105 Variable temperature ^{2}H spectra of FS-PA acetone-d6 solvate acquired at 5 kHz MAS with a 1 s recycle delay over 80 transients.

Static ^{2}H spectra are more sensitive to subtle changes in quadrupole parameters because the whole line shape can be fitted. The turning points of the static ^{2}H spectrum of the FS-PA acetone solvate are broad, Figure 106, indicating that an inhomogeneous spread of motions could be present in the system.[81] The dynamically-averaged C_Q and η values can be estimated from these turning points, shown in red in Figure 106, using Equations (6.0) and (6.1). The values calculated from the static powder pattern are C_Q = 24 kHz and η = 0.56, which are similar to the parameters obtained by fitting the MAS spectrum above.

$$C_Q = \frac{2\Delta\nu_{zz}}{3} \qquad \textbf{(6.0)}$$

$$\eta = \frac{\Delta\nu_{yy} - \Delta\nu_{xx}}{\Delta\nu_{zz}} \qquad \textbf{(6.1)}$$

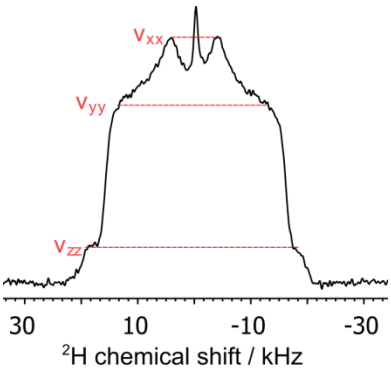

Figure 106 Static ^{2}H spectrum of FS-PA acetone-d6 solvate acquired with a 1 s recycle delay and echo delays τ_1 = 60 µs, τ_2 = 53.5 µs over 72000 transients. The turning points used to calculate the quadrupolar parameters are shown in red.

The acetone molecule is likely to undergo methyl rotation, so the CASTEP-calculated C_Q = 186 kHz and η = 0.025 can be averaged over a C_3 rotation using Equation (6.2), where θ = 70.53˚. This gives a C_3-averaged C_Q* = 62 kHz and η* = 0, which is still higher than the experimental C_Q. More consistent values can be achieved by including a second two-site jump motion of 105˚ using the in-house software "calcav2", giving C_Q = 31.0 kHz and η = 0.78. It is hypothesised that the whole acetone molecule might undergo a C_2 rotation about the C=O axis, as has been observed in DMSO[82] and DMSO solvates of APIs like CBZ.[83] The angle of the second jump motion, 105˚, is on the order of a C_2 flip in acetone, 120˚.

$$C_Q^* = \frac{C_Q}{2}(3\cos^2\theta - 1) \qquad \textbf{(6.2)}$$

Variable temperature ^{2}H experiments show significant changes occur on cooling to −65 ˚C, Figure 107(i), indicating that the motion of the acetone molecules must occur in the intermediate regime, 10^4–10^7 Hz. At low temperature the spectral width is twice that of the room temperature spectrum, suggesting that some aspect of the dynamics is frozen out. The spectra cannot be fit to a single site since the dynamics are in the intermediate regime.

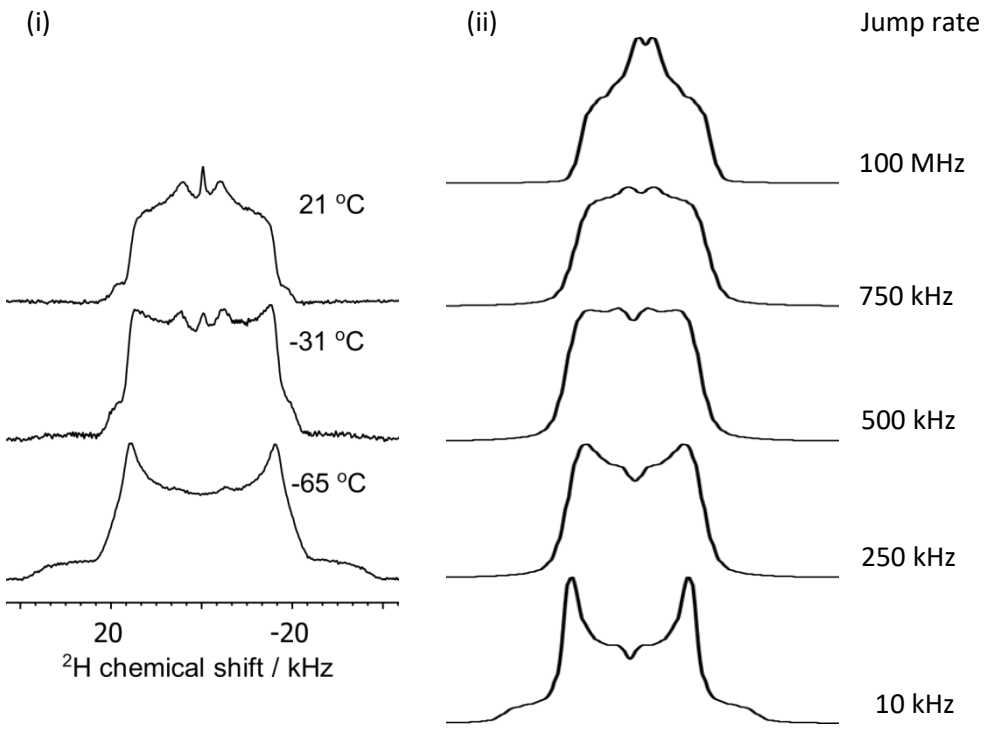

Figure 107 (i) Variable temperature static ^{2}H spectra of FS-PA acetone-d_6 solvate acquired with echo delays τ_1 = 60 µs, τ_2 = 53.5 µs over 72000 transients. **(ii)** EXPRESS-simulated static ^{2}H spectra over a two-site jump motion of 115˚ using the parameters in Table 20. The jump rates are given that provided the best match, by eye, to the experimental spectra in **(i)**.

The effect of the hypothesised C_2 motion on the ^{2}H spectrum can be simulated using EXPRESS.[41] It is possible to simulate a composite of many different jump motions, for example the C_2 jump and rotating methyl groups in DMSO must be modelled using two frames; in one the deuterons of the methyl group interchange and in the second the first frame undergoes two-site hops.[41] Acetone is expected to have similar motion to DMSO, though in the solvate it is likely to be complex. It is almost certain that the methyl groups undergo three-site exchange and that this motion is fast on the NMR timescale, even at −65 ˚C. Therefore, it is assumed that this contribution to the overall dynamics is fast enough to result in a single effective quadrupolar tensor (C_Q*),[84] taken to be the C_3-averaged value calculated above; C_Q* = 62 kHz and η* = 0. A simple two-site jump of 115˚, with each site equally weighted, emulates the flipping of the acetone molecule about the C=O axis, see Table 20 in section 6.3.1 for details. A jump rate of 750 kHz gives a simulated spectrum that matches the experimental room temperature spectrum well, Figure 107. Reducing the jump rate to 500 kHz and then 250 kHz, keeping all other parameters the same, also results in good agreement with the lower temperature spectra. These jump

rates are in the intermediate regime, as expected. Model spectra were also simulated at a fast jump rate (100 MHz) and a slow jump rate (10 kHz) to show the extremes of the fast and slow limits. At −65 ˚C the experimental spectrum appears to be nearing the slow limit regime. Note that the sharp peak in the centre of the experimental spectra is not modelled as this is likely to be trace amounts of acetone on the surface of the sample.

Further confirmation of the validity of the simple two-site jump model is given by considering the orientation-dependence of the T_2 relaxation.[44] In the intermediate regime the rate of motion is comparable to the quadrupole coupling constant, and T_2 relaxation becomes anisotropic due to fluctuations in the quadrupole interaction.[85] This causes a distortion in the echo decay compared to the FID. If $T_2 < \tau_1$ in some regions then this magnetisation may not be completely refocused, while the echo may refocus efficiently, causing line shape distortions. The echo-refocussing efficiency, and hence the line shape, depends on the exchange rate and geometry of motion.[81,86] The line shape of the FS-PA acetone-d_6 solvate shows T_2 distortion as the echo delay, τ_1, is altered in Figure 108(i). Using the simple two-site jump model of 115˚ and the previously established jump rate of 750 kHz, the EXPRESS-simulated ^{2}H spectra again match the experimental spectra using the same τ_1 delays, Figure 108(ii).

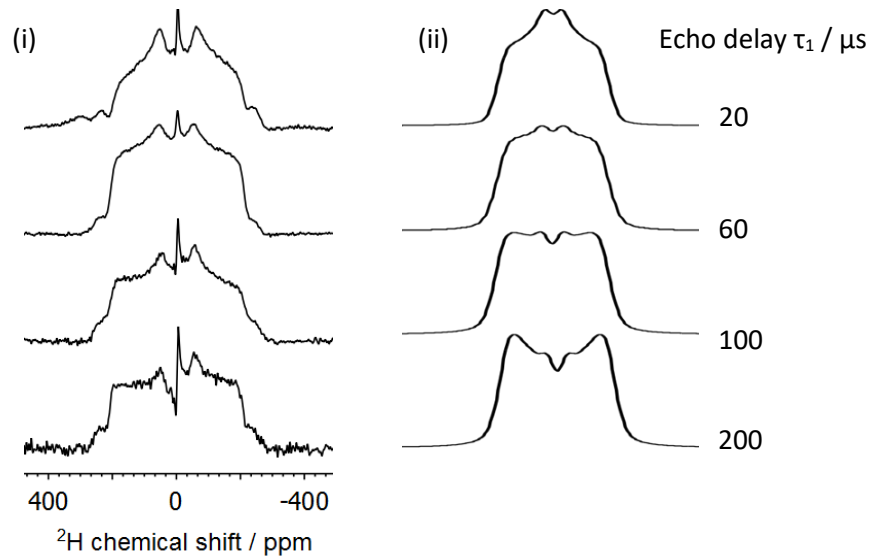

Figure 108 (i) Static ^{2}H spectra of FS-PA acetone-d_6 solvate acquired with variable echo delays over 18000 transients. **(ii)** EXPRESS-simulated static ^{2}H spectra of a two-site jump motion of 115˚ using the parameters in Table 6 and a jump rate of 750 kHz, with variable echo delays. τ_1 delays are given on the right and τ_2 delays were optimised for the FID to start at the top of the echo: from top to bottom, 13.5 μs, 53.5 μs, 93.5 μs, and 193.5 μs.

6.4 Conclusions

This chapter discusses the structure and dynamics of a new cocrystal FS-PA and its two isostructural solvates. In addition to the identification of structural features and disordered regions, NMR crystallography methods were used to gain insights into the timescale of motions and provide a simple motional model that showed surprisingly good agreement with experimental data.

No structure could be obtained for FS-PA *via* XRD methods, yet SSNMR was able to provide some insights. The asymmetric unit was identified as containing a single FS molecule and a single PA molecule. Evidence of hydrogen bonding was identified *via* long range ^{1}H–^{13}C HETCOR correlations to the FS carboxylic acid carbon and proton sites and also in the high shift proton peak observed at 13.8 ppm. Further evidence of hydrogen bonding is given by a ^{1}H–^{1}H DQ/SQ correlation and the chemic al shift of the pyridine nitrogen atom of PA. The synthon is tentatively identified as synthon **F**, *i.e*. a dimer is formed between the FS carboxylic acid group and the PA carboxamide group. The presence of dynamic disorder of the FS sulphonamide group was suggested by subtle changes in variable temperature ^{2}H spectra. ^{2}H T_1 times show a clear minimum, indicating that some dynamic motion is present, and direct fits of the ^{2}H relaxation data to a two-site jump model yield E_a = 33(3) kJ mol^{-1} averaged over the first pair of sidebands. The jump angle is 102.9(5)˚ and so the motion is found to be similar to that of pure FS form **I**, consisting of NH_2 proton exchange. Dynamic disorder was also identified at the FS furan ring, indicated by the short ^{13}C T_1 relaxation times that decay with temperature in keeping with the approach to a T_1 minimum. An estimated E_a of *ca*. 20 kJ mol^{-1} was obtained by fitting to an Arrhenius-type model, which is similar to the E_a associated with furan ring dynamics in 2FS-INA. These structural details might be useful to direct future structural investigations of FS-PA *via* synchrotron PXRD experiments.

The isostructural nature of the acetone and ethanol solvates of FS-PA was confirmed by ^{13}C SSNMR. While SCXRD data only observes a blur of electron density in channels within the FS-PA framework, ^{13}C SSNMR peaks were assigned to the acetone and ethanol molecules residing in the lattice. No long-range interactions were observed between the lattice and the solvent molecules *via* ^{1}H–^{13}C HETCOR experiments, so they were characterised as void-filling solvates. Quantitative ^{13}C experiments suggested that the FS-PA solvates contain *ca*. 0.4 solvent molecule for every one FS and PA molecule, supported by thermal analysis. Over time the solvent peak intensities decreased with no changes observed to the rest of

the spectrum. Therefore, the solvates are nonstoichiometric: the structure is capable of containing a variable number of solvent molecules without breaking down. No degradation was observed *via* ^{13}C SSNMR over a 9 month period, indicating that only trace amounts of solvent are required to stabilise the structure. However, heating to 80 ˚C for 21 hours resulted in degradation.

The solvent molecules were both characterised as being dynamic. Variable temperature ^{13}C T_1 times of the ethanol CH_2 peak show a decrease in keeping with the approach to a T_1 minimum. A very low E_a of 7 kJ mol^{-1} was estimated by fitting the T_1 times to an Arrhenius-type model, though a more complete dataset encompassing the T_1 minimum would provide much more robust figures. ^2H SSNMR was particularly informative for the acetone solvate, using acetone-d6. Line-shape changes were observed on cooling to −65 ˚C, indicating that the acetone molecule is dynamic and the timescale of these dynamics is in the intermediate regime. Averaging the CASTEP-calculated C_Q of 186 kHz over two motions, one being a jump of 72˚ corresponding to methyl rotation and the other being a jump of 105˚ corresponding to rotation about the C=O axis, results in a motionally averaged C_Q of 31.0 kHz, that is in keeping with the fitted experimental C_Q at room temperature. The effect of a simple two-site jump of 115˚ was modelled using EXPRESS, aiming to model a rotation about the C=O axis. Good agreement was observed with the room temperature experimental line shapes with a jump rate of 750 kHz. This reduced to 250 kHz at −65 ˚C, though the system may be nearing the slow motion regime. It was assumed in this model that methyl rotation would be fast on the NMR timescale at all temperatures. The simple two-site jump model was further ratified by observing good agreement between experimental and EXPRESS-simulated T_2-distorted line shapes using variable echo delays.

6.5 References

(1) R. K. Harris, *Analyst*, **2006**, 131, 351.
(2) C. P. Price, G. D. Glick and A. J. Matzger, *Angew. Chem. Int. Ed.*, **2006**, 45, 2062.
(3) S. R. Vippagunta, H. G. Brittain and D. J. Grant, *Adv Drug Deliv Rev*, **2001**, 48, 3.
(4) A. L. Bingham, D. S. Hughes, M. B. Hursthouse, R. W. Lancaster, S. Tavener and T. L. Threlfall, *Chem. Commun.*, **2001**, 7, 603.
(5) M. Tawa, O. Almarsson and J. F. Remenar, *Pharmaceutical propylene glycol solvate compositions*, 2010, US 7790905 B2.
(6) R. Kandarapu, V. D. Nasare, I. Bhushan and M. S. Mohan, *Paroxetine compositions* 2007, WO 2007035816 A2.
(7) "Structural Aspects of Hydrates and Solvates", K. R. Morris in "*Polymorphism in Pharmaceutical Solids",* H. G. Brittain (Eds.), Marcel Dekker Inc.: New York, US, 1999, p 125.
(8) M. B. Hickey, M. L. Peterson, E. S. Manas, J. Alvarez, F. Haeffner and O. Almarsson, *J. Pharm. Sci.*, **2007**, 96, 1090.

(9) S. Byard, A. Abraham, T. B. PJ, R. K. Harris and P. Hodgkinson, *J. Pharm. Sci.*, **2012**, 101, 176.
(10) E. Shefter and T. Higuchi, *J. Pharm. Sci.*, **1963**, 52, 781.
(11) "The Importance of Solvates", U. J. Griesser in "*Polymorphism in the Pharmaceutical Industry",* R. Hilfiker (Eds.), WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2006, p 211.
(12) J. Aaltonen, M. Alleso, S. Mirza, V. Koradia, K. C. Gordon and J. Rantanen, *Eur. J. Pharm. Biopharm.*, **2009**, 71, 23.
(13) V. S. Minkov, A. A. Beloborodova, V. A. Drebushchak and E. V. Boldyreva, *Cryst. Growth Des.*, **2014**, 14, 513.
(14) R. Gerdil and R. E. Marsh, *Acta Crystallogr.*, **1960**, 13, 166.
(15) G. A. Stephenson, E. G. Groleau, R. L. Kleemann, W. Xu and D. R. Rigsbee, *J. Pharm. Sci.*, **1998**, 87, 536.
(16) J. R. Authelin, *Int. J. Pharm.*, **2005**, 303, 37.
(17) N. Madusanka, M. D. Eddleston, M. Arhangelskis and W. Jones, *Acta Crystallogr. B*, **2014**, 70, 72.
(18) A. Bērziņš, T. Rekis and A. Actiņš, *Cryst. Growth Des.*, **2014**, 14, 3639.
(19) R. Prohens, R. Barbas, A. Portell, M. Font-Bardia, X. Alcobe and C. Puigjaner, *Cryst. Growth Des.*, **2015**, 15, 475.
(20) W. Xu, N. Gong, S. Yang, N. Zhang, L. He, G. Du and Y. Lu, *J. Pharm. Sci.*, **2015**, 104, 1256.
(21) N. Stieger, W. Liebenberg, J. C. Wessels, H. Samsodien and M. R. Caira, *Struct. Chem.*, **2010**, 21, 771.
(22) T. Hosokawa, S. Datta, A. R. Sheth, N. R. Brooks, V. G. Young and D. J. W. Grant, *Cryst. Growth Des.*, **2004**, 4, 1195.
(23) A. Bērziņš, E. Skarbulis, T. Rekis and A. Actiņš, *Cryst. Growth Des.*, **2014**, 14, 2654.
(24) A. Bērziņš and P. Hodgkinson, *Solid State Nucl. Magn. Reson.*, **2015**, 65, 12.
(25) A. Bērziņš, E. Skarbulis and A. Actiņš, *Cryst. Growth Des.*, **2015**, 15, 2337.
(26) M. Banik, S. P. Gopi, S. Ganguly and G. R. Desiraju, *Cryst. Growth Des.*, **2016**, 16, 5418.
(27) A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car *et al.*, *Acta Crystallogr. B*, **2016**, 72, 439.
(28) S. L. Price, D. E. Braun and S. M. Reutzel-Edens, *Chem. Commun.*, **2016**, 52, 7065.
(29) A. L. Spek, *Acta Crystallogr. C*, **2015**, 71, 9.
(30) N. Zencirci, U. J. Griesser, T. Gelbrich, V. Kahlenberg, R. K. Jetti, D. C. Apperley and R. K. Harris, *J. Phys. Chem. B*, **2014**, 118, 3267.
(31) D. C. Apperley, A. S. Batsanov, S. J. Clark, R. K. Harris, P. Hodgkinson and D. B. Jochym, *J. Mol. Struct.*, **2012**, 1015, 192.
(32) F. G. Vogt, P. C. Dell'Orco, A. M. Diederich, Q. Su, J. L. Wood, G. E. Zuber, L. M. Katrincic, R. L. Mueller, D. J. Busby and C. W. Debrosse, *J. Pharm. Biomed. Anal.*, **2006**, 40, 1080.
(33) D. C. Apperley, P. A. Basford, C. I. Dallman, R. K. Harris, M. Kinns, P. V. Marshall and A. G. Swanson, *J. Pharm. Sci.*, **2005**, 94, 516.
(34) D. E. Braun, K. P. Nartowski, Y. Z. Khimyak, K. R. Morris, S. R. Byrn and U. J. Griesser, *Mol. Pharm.*, **2016**, 13, 1012.
(35) D. C. Apperley, A. F. Markwell, R. K. Harris and P. Hodgkinson, *Magn. Reson. Chem.*, **2012**, 50, 680.
(36) M. O. Miclăuş, I. E. Kacso, F. A. Martin, L. David, M. M. Pop, C. Filip and X. Filip, *J. Pharm. Sci.*, **2015**, 104, 3782.
(37) D. Raijada, A. D. Bond, F. H. Larsen, C. Cornett, H. Qu and J. Rantanen, *Pharm. Res.*, **2013**, 30, 280.

(38) D. E. Braun, L. H. Koztecki, J. A. McMahon, S. L. Price and S. M. Reutzel-Edens, *Mol. Pharm.*, **2015**, 12, 3069.
(39) F. R. Kang, F. G. Vogt, J. Brum, R. Forcino, R. C. B. Copley, G. Williams and R. Carlton, *Cryst. Growth Des.*, **2012**, 12, 60.
(40) F. G. Vogt, J. Brum, L. M. Katrincic, A. Flach, J. M. Socha, R. M. Goodman and R. C. Haltiwanger, *Cryst. Growth Des.*, **2006**, 6, 2333.
(41) R. L. Vold and G. L. Hoatson, *J. Magn. Reson.*, **2009**, 198, 57.
(42) A. Schmidt and S. Vega, *J. Chem. Phys*, **1987**, 87, 6895.
(43) "Nuclear Magnetic Resonance Probes of Molecular Dynamics", R. Tycko in "*Understanding Chemical Reactivity",* R. Tycko (Eds.), Kluwer Academic Publishers: Dordrecht, Boston, US, 1994.
(44) M. Hologne and J. Hirschinger, *Solid State Nucl. Magn. Reson.*, **2004**, 26, 1.
(45) M. Baias, J. N. Dumez, P. H. Svensson, S. Schantz, G. M. Day and L. Emsley, *J. Am. Chem. Soc.*, **2013**, 135, 17501.
(46) D. H. Brouwer, R. J. Darton, R. E. Morris and M. H. Levitt, *J. Am. Chem. Soc.*, **2005**, 127, 10365.
(47) E. Davies, M. J. Duer, S. E. Ashbrook and J. M. Griffin, *J. Am. Chem. Soc.*, **2012**, 134, 12508.
(48) D. V. Dudenko, P. A. Williams, C. E. Hughes, O. N. Antzutkin, S. P. Velaga, S. P. Brown and K. D. Harris, *J. Phys. Chem. C*, **2013**, 117, 12258.
(49) P. Li, Y. Y. Chu, L. Wang, R. M. Wenslow, K. C. Yu, H. L. Zhang and Z. W. Deng, *CrystEngComm.*, **2014**, 16, 3141.
(50) M. Miclăuş, I. G. Grosu, X. Filip, C. Tripon and C. Filip, *CrystEngComm.*, **2014**, 16, 299.
(51) D. Ludeker and G. Brunklaus, *Solid State Nucl. Magn. Reson.*, **2015**, 65, 29.
(52) A. E. Watts, K. Maruyoshi, C. E. Hughes, S. P. Brown and K. D. M. Harris, *Cryst. Growth Des.*, **2016**, 16, 1798.
(53) B. Elena, G. Pintacuda, N. Mifsud and L. Emsley, *J. Am. Chem. Soc.*, **2006**, 128, 9555.
(54) E. Salager, R. S. Stein, C. J. Pickard, B. Elena and L. Emsley, *Phys. Chem. Chem. Phys.*, **2009**, 11, 2610.
(55) F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbein and H. Oschkinat, *Nature*, **2002**, 420, 98.
(56) N. R. Goud, S. Gangavaram, K. Suresh, S. Pal, S. G. Manjunatha, S. Nambiar and A. Nangia, *J. Pharm. Sci.*, **2012**, 101, 664.
(57) B. I. Harriss, L. Vella-Zarb, C. Wilson and I. R. Evans, *Cryst. Growth Des.*, **2014**, 14, 783.
(58) E. Sangtani, S. K. Sahu, S. H. Thorat, R. L. Gawade, K. K. Jha, P. Munshi and R. G. Gonnade, *Cryst. Growth Des.*, **2015**, 15, 5858.
(59) H. E. Kerr, L. K. Softley, K. Suresh, A. Nangia, P. Hodgkinson and I. R. Evans, *CrystEngComm.*, **2015**, 17, 6707.
(60) R. K. Harris and A. C. Olivieri, *Prog. Nucl. Magn. Reson. Spectrosc.*, **1992**, 24, 435.
(61) S. H. Alarcón, A. C. Olivieri and R. K. Harris, *Solid State Nucl. Magn. Reson.*, **1993**, 2, 325.
(62) R. K. Harris, M. M. Sünnetçioglu, K. S. Cameron and F. G. Riddell, *Magn. Reson. Chem.*, **1993**, 31, 963.
(63) H. C. S. Chan, G. R. Woollam, T. Wagner, M. U. Schmidt and R. A. Lewis, *CrystEngComm.*, **2014**, 16, 4365.
(64) Y. H. Luo and B. W. Sun, *Spectrochim. Acta. A Mol. Biomol. Spectrosc.*, **2014**, 120, 228.
(65) R. K. Harris, P. Jackson, L. H. Merwin, B. J. Say and G. Hagele, *J. Chem. Soc. Farad. T. I*, **1988**, 84, 3649.
(66) S. Ando, J. Kikuchi, Y. Fujimura, Y. Ida, K. Higashi, K. Moribe and K. Yamamoto, *J. Pharm. Sci.*, **2012**, 101, 3214.

(67) K. Maruyoshi, D. Iuga, O. N. Antzutkin, A. Alhalaweh, S. P. Velaga and S. P. Brown, *Chem. Commun.*, **2012**, 48, 10844.
(68) F. G. Vogt, J. S. Clawson, M. Strohmeier, A. J. Edwards, T. N. Pham and S. A. Watson, *Cryst. Growth Des.*, **2009**, 9, 921.
(69) H. E. Kerr, L. K. Softley, K. Suresh, P. Hodgkinson and I. R. Evans, *Acta Crystallogr. C*, **2016**, accepted manuscript.
(70) P. Hodgkinson*, pNMRsim: a general simulation program for large problems in solid-state NMR*, www.dur.ac.uk/paul.hodgkinson/pNMRsim, accessed 2016.
(71) D. C. Apperley, A. F. Markwell, I. Frantsuzov, A. J. Ilott, R. K. Harris and P. Hodgkinson, *Phys. Chem. Chem. Phys.*, **2013**, 15, 6422.
(72) J. N. Latosińska, M. Latosińska, W. Medycki and J. Osuchowicz, *Chem. Phys. Lett.*, **2006**, 430, 127.
(73) M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminform.*, **2012**, 4, 17.
(74) 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.
(75) H. Mimura, S. Kitamura, T. Kitagawa and S. Kohda, *Colloids Surf., B*, **2002**, 26, 397.
(76) U. B. R. Khandavilli, S. Gangavaram, N. R. Goud, S. Cherukuvada, S. Raghavender, A. Nangia, S. G. Manjunatha, S. Nambiar and S. Pal, *CrystEngComm.*, **2014**, 16, 4842.
(77) R. Chadha, A. Kuhad, P. Arora and S. Kishor, *Chem. Cent. J.*, **2012**, 6, 114.
(78) N. H. M. Hogg, P. J. T. Boulton, V. E. Zorin, R. K. Harris and P. Hodgkinson, *Chem. Phys. Lett.*, **2009**, 475, 58.
(79) A. J. Horsewill, *Prog. Nucl. Magn. Reson. Spectrosc.*, **1999**, 35, 359.
(80) M. Shen, R. Roopchand, E. S. Mananga, J. P. Amoureux, Q. Chen, G. S. Boutis and B. Hu, *Solid State Nucl. Magn. Reson.*, **2015**, 66-67, 45.
(81) R. J. Schadt, E. J. Cain and A. D. English, *J. Phys. Chem.*, **1993**, 97, 8387.
(82) M. S. Solum, K. W. Zilm, J. Michl and D. M. Grant, *J. Phys. Chem.*, **1983**, 87, 2940.
(83) A. J. Cruz-Cabeza, G. M. Day and W. Jones, *Phys. Chem. Chem. Phys.*, **2011**, 13, 12808.
(84) L. Vugmeyster, D. Ostrovsky, J. J. Ford, S. D. Burton, A. S. Lipton, G. L. Hoatson and R. L. Vold, *J. Am. Chem. Soc.*, **2009**, 131, 13651.
(85) H. W. Spiess and H. Sillescu, *J. Magn. Reson.*, **1981**, 42, 381.
(86) J. Hirschinger and A. D. English, *J. Magn. Reson.*, **1989**, 85, 542.

Chapter 7: Concluding remarks

7.1 Context

This thesis aims to show how NMR crystallography investigations can be used to understand the structure and dynamics of pharmaceutically-relevant cocrystals that contain crystallographic disorder in a level of detail that is not attainable using standard diffraction methods alone. Disorder in organic cocrystals is not often explicitly investigated but, in the context of fundamental structural studies, solid-state characterisation is incomplete without some indication as to the nature and behaviour of the disordered components. The systems studied in this report contain topical APIs (caffeine, furosemide, and naproxen) in cocrystal form so the structural investigations here are transferrable to the ever-increasing variety of cocrystal systems published annually. More general organic systems may also be studied using the methods displayed here to answer structural questions pertaining to the characterisation of disordered components, hydrogen bond identification, polymorphism *etc*.

7.2 Conclusions and future directions

Chapter 5 demonstrates the principles used throughout this work particularly clearly. SCXRD data of a furosemide cocrystal indicates the presence of complex disorder over several components in the unit cell, but gives no indication as to the nature of the disorder. Variable temperature SSNMR reveals both disordered components to be dynamic at ambient temperature, and a combination of experimental and calculated NMR parameters reveals details about the rate and mechanism of motion. Specifically, the temperature-dependent behaviour of ^{2}H quadrupolar parameters and relaxation times reveals that the disordered sulphonamide undergoes a combination of two motions: fast exchange of the NH_2 hydrogen atoms as well as a 30˚ rotation of the entire group about the C–S bond. ^{13}C T_1 relaxation analysis also reveals the rate and energy barrier of the rotation of a furan ring.

These methods can also be applied to challenging systems like nonstoichiometric solvates, as shown in Chapter 6. The presence (or lack) of intermolecular interactions between the

solvent molecules and the cocrystal lattice are probed by 2D SSNMR experiments, and the desolvation behaviour can be monitored over a wide temperature range. A computational model is also developed to go some way to describe the mechanism of motion of the solvent molecules, with validation possible by comparing experimental and computationally predicted ^{2}H spectra. The solvates investigated here would benefit from further validation of the motional model by molecular dynamics simulations. This would allow the motion to be visually inspected over a range of timescales (ps *via* DFT-MD and ns–ms *via* classical MD). The rate of motion might also be inferred to corroborate the rates estimated in this work by NMR.

As well as disorder characterisation, structural features pertinent to crystal engineering (*e.g*. hydrogen bonding) are shown to be accessible *via* SSNMR in Chapter 4 for a system with subpar SCXRD data. A combination of experimental and predicted SSNMR data allows the hydrogen bonding pattern that was observed in the diffraction data to be confirmed. Additionally, similar features are characterised in Chapter 6 for a system where diffraction data is not attainable. 2D SSNMR experiments allow the atom-atom proximities to be studied, leading to direct evidence of synthon formation in a cocrystal whose 3D structure is entirely unknown. This system would benefit from future work that is beyond the scope of this report, namely crystal structure determination. High quality PXRD data acquired at a synchrotron is required for structure determination *via* Rietveld refinement, or alternatively SSNMR approaches might be possible using spin diffusion methods or crystal structure prediction. Such future work would require the input from experts in the respective techniques, but structure solution will be aided by the identification of structural features (*i.e*. the hydrogen bonding and dynamic components) that was achieved in this work.

A final aspect of this work is discussed in Chapter 3; finding the limit to which chemical shielding can be predicted using a specific DFT package (CASTEP) and how the standard procedure of geometry optimisation affects this limit. Computational prediction of chemical shieldings is ubiquitous in NMR crystallography studies to allow the comparison of structural models with experimental NMR data. However, the importance of the geometry optimisation procedure had not previously been quantitatively investigated in this context. The findings of this investigation confirm the generally accepted “error” between experimental and predicted ^{13}C shieldings to be of the order of 2 ppm. The systematic error between the computed parameters of chemically identical sites caused by imperfect

geometry optimisation is significantly smaller at 0.3 ppm. This error is small, but non-negligible, and can be lowered to 0.1 ppm by tightening the geometry convergence criteria. Standard NMR crystallography investigations are not expected to be affected by this imperfect geometry optimisation error, since it is usually challenging to distinguish experimental differences below 0.5 ppm. However, studies into subtle NMR features like ^{13}C linewidths are shown to require consideration of this error. That being said, it is possible to gain insight into the disorder contribution to ^{13}C linewidths and the methodology introduced here might be extended to 2D techniques in the future.

Appendix

Crystallographic data

The following crystallographic data for the FS-PA cocrystal and related solvates, discussed in Chapter 6, were acquired and refined by the I.R.E group at Durham University. The crystallographic data for FS-PA hydrate is taken from Banik *et al*.[1] and is included for comparative purposes, due to its isostructural relationship with the two solvates.

Table A1 Crystal data and unit cell refinement for FS-PA and its solvates.

	FS-PA	FS-PA acetone	FS-PA ethanol	FS-PA sesquihydrate[a]
Empirical formula[b]	$C_{18}H_{17}N_4O_6SCl$	$C_{18}H_{17}N_4O_6SCl$	$C_{18}H_{17}N_4O_6SCl$	$C_{18}H_{17}N_4O_6SCl$[c]
Molar mass[b] / g mol^{-1}	452.9	452.9	452.9	477.73[c]
Temperature / K	298	120	120	150
Crystal system	Monoclinic	Triclinic	Triclinic	Triclinic
Space group	$P2_1$	$P\bar{1}$	$P\bar{1}$	$P\bar{1}$
a / Å	9.8313(7)	5.0474(17)	5.2086(17)	5.117(2)
b / Å	19.812 (2)	14.428(2)	14.641(5)	14.475(6)
c / Å	5.2574 (4)	14.778(3)	14.656(5)	14.859(6)
α / ˚	-	76.344(14)	76.378(5)	76.363(7)
β / ˚	104.226 (8)	83.49(2)	87.592(3)	82.502(8)
γ / ˚	-	86.854(19)	83.243(3)	87.226(8)
Volume / Å^3	-	1038.5(4)	1078.6(6)	1060.1(8)[c]
Z	-	2	2	2
Reflections collected	-	10937	3660	8489
Observed reflections	-	2585	3197	1614
Parameters	-	268	267	338
Goodness of fit	-	1.0321	1.0794	1.004
R indices (I/σ = 2)	-	R = 0.1298 R_w = 0.1485	R = 0.1011 R_w = 0.1052	R = 0.1077 R_w = 0.2694
R indices (I/σ = 0)	-	R = 0.3712 R_w = 0.2225	R = 0.1105 R_w = 0.1052	R = 0. 2468 R_w = 0.2063

[a] Data taken from Banik *et al*.[1], [b] Solvent molecules not included. [c] Value in deposited .cif file differs from value presented in Table 1 of Banik *et al*.[1]

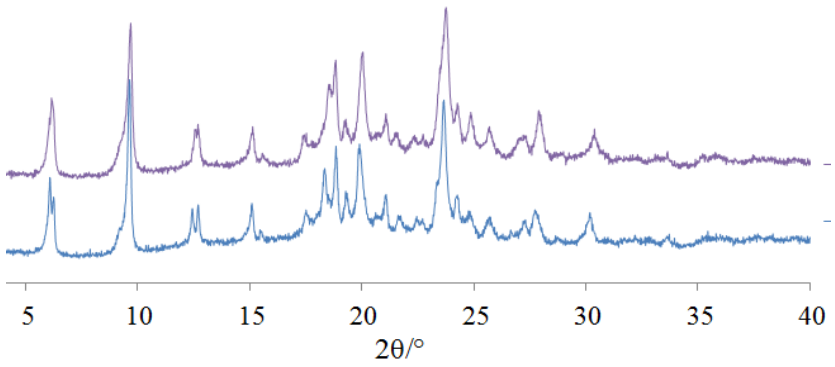

Figure A1 Experimental PXRD patterns for the acetone (purple) and ethanol (blue) solvates of FS-PA, taken from the Mchem report of L. Softley.

Hydrate *vs* acetone solvate

The isostructural relationship of the acetone and ethanol solvates is observed in the experimental PXRD patterns previously acquired by the I.R.E. group at Durham University, Figure A1. These patterns are similar, by eye, to the PXRD patterns calculated from the crystal structures determined by SCXRD, both with (Figure A2) and without (Figure A3) the disordered solvent molecules included. The FS-PA hydrate[1] also has very similar PXRD patterns calculated from the crystal structure deposited in the CSD, highlighting the isostructural relationship between the three systems.

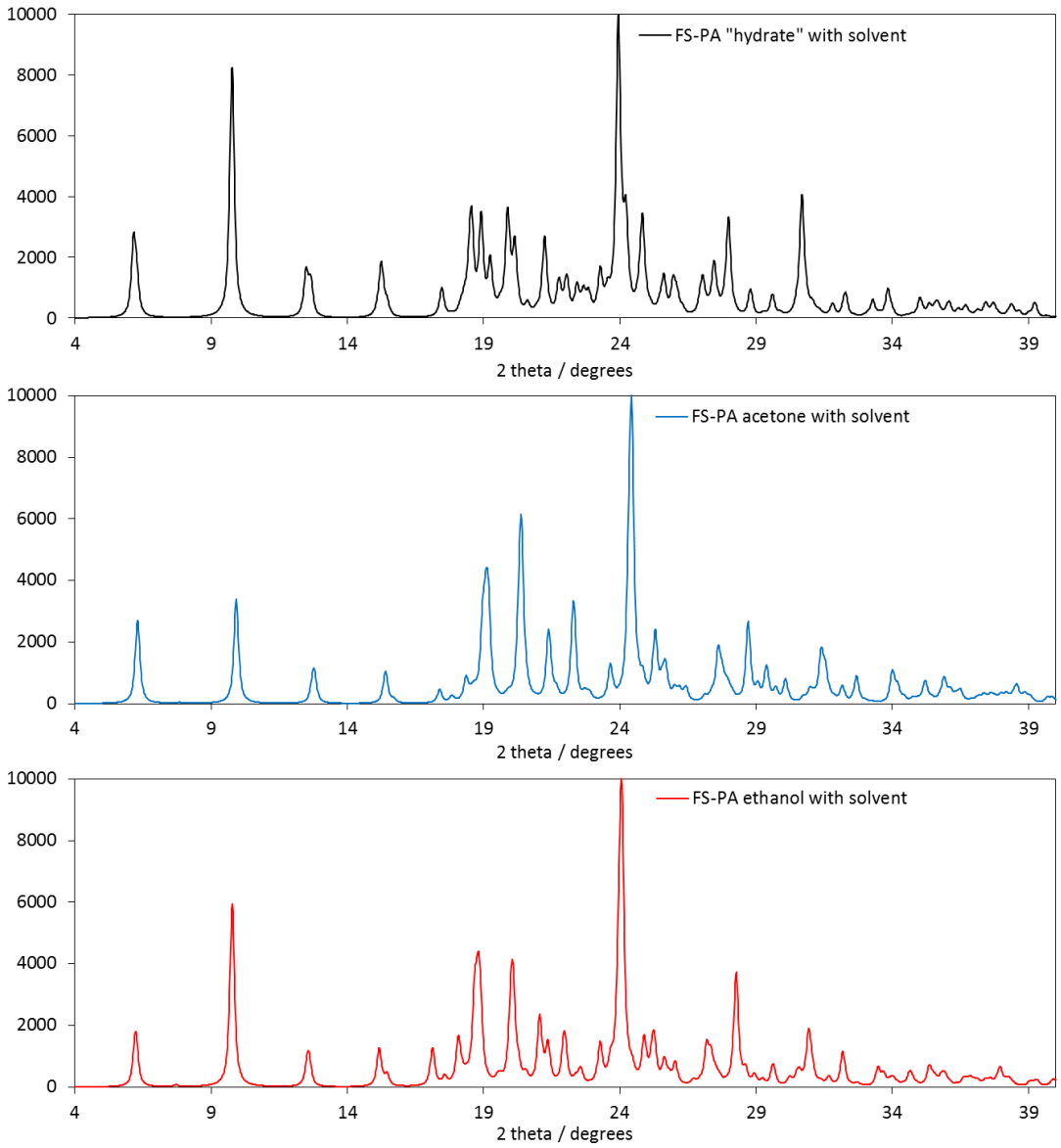

Figure A2 Calculated PXRD patterns for FS-PA solvates with electron density associated with the disordered solvent sites included.

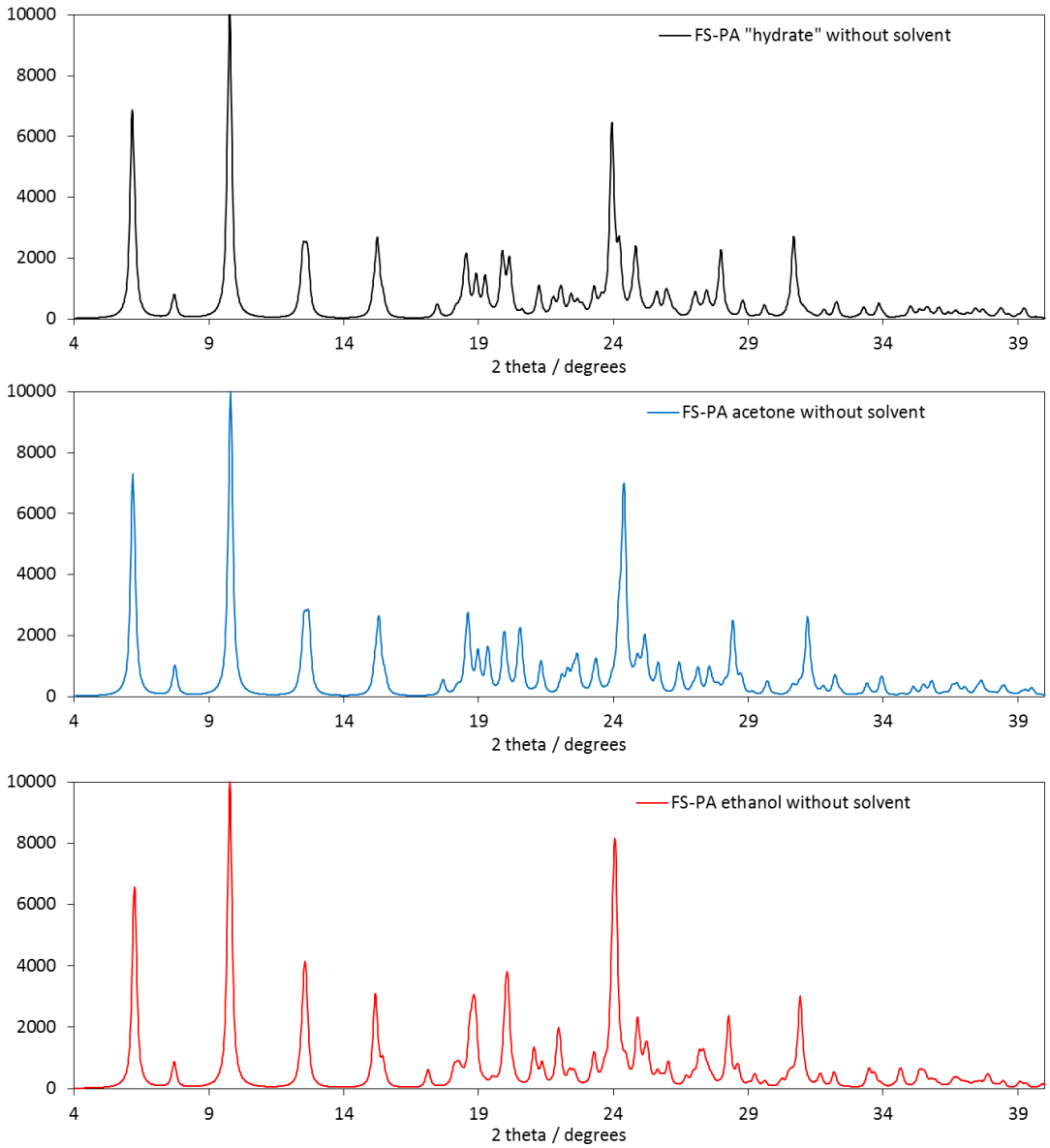

Figure A3 Calculated PXRD patterns for FS-PA solvates with electron density associated with the disordered solvent sites removed.

It is noted that the FS-PA hydrate[1] was reportedly synthesised by crystallisation from acetone, not water, leading to a question as to the identity of the disordered solvent molecules in the reported structure. It is not possible to conclusively determine whether the reported hydrate is the same as the acetone solvate or not without access to the original sample. However, the reported synthesis[1] is the same as that used here for the acetone solvate and it is unlikely that the system would preferentially form a hydrate in the hands of one group and a solvate in the hands of another. The calculated PXRD pattern is similar to those of the solvates synthesis in this work, though small differences in peak

position and intensity are expected due to the small differences in unit cell parameters between the three systems. Furthermore, the crystal structure determination for the hydrate exhibits some flaws. The number of observed reflections is low, resulting in R indices of the order of 20 %, and there are discrepancies in the cell volume and formula weight between Table 1 in the publication and the values reported in the cif deposited in the CSD. Other characterisation data provided in the supplementary information does little to resolve these queries. The experimental PXRD pattern of the hydrate is visually very similar to those of the acetone and ethanol solvates. An IR spectrum is presented in the supplementary informaton, but not discussed, and the peaks are also very similar to those of the FS-PA acetone solvate, Figure A4 and Table A2.

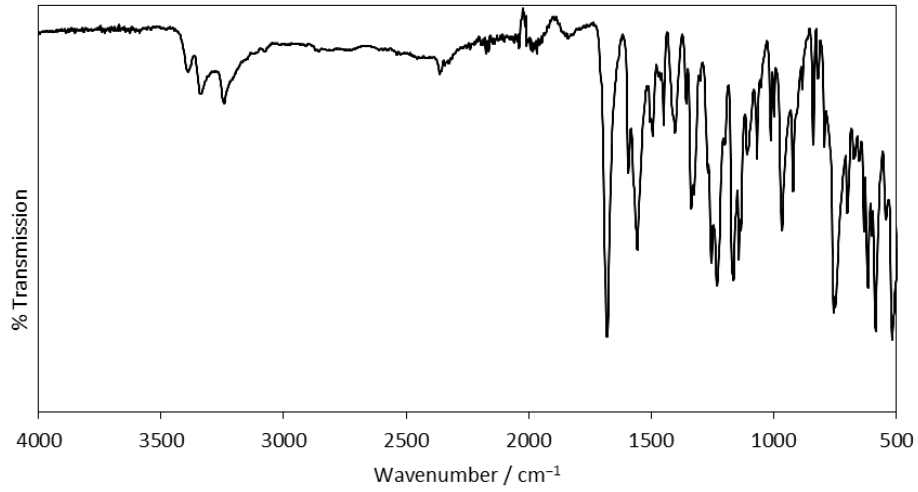

Figure A4 Solid-state FT-IR spectrum of the FS-PA acetone solvate.

Table A2 Peaks in the FT-IR spectra of the hydrate and acetone solvate of FS-PA. Hydrate peaks taken from Banik *et al*.[1]

FS-PA hydrate / cm^{-1}	FS-PA acetone solvate / cm^{-1}
3394	3390
3339	3342
3285	-
3242	3242
1682	1682
1591	1593
1562	1556
1492	1495
1445	1448
1402	1402
1324	1336
-	1254
1238	1230
-	1163
1143	1142
1070	1068
-	1012
969	966
922	922
839	837
758	756
618	617

The reported TGA curve of the hydrate is also similar to that of the FS-PA acetone solvate acquired in this work, Figure 102(i) in the main text, with a 30 % mass loss between 100–170 ˚C followed by a plateau until a 2 % mass loss at 220 ˚C and total mass loss at 275 ˚C. On the other hand, the DSC traces of the hydrate and acetone solvates show some significant differences, in particular at 100 ˚C. A small endothermic peak is observed for the acetone solvate, corresponding to solvent evaporation. In contrast, the hydrate is reported to display a very large endothermic peak that is attributed to water loss, followed by a small endothermic peak attributed to sample melting. This difference suggests that the two samples are not the same; though it is possible that slow desolvation under storage might alter the thermal behaviour of both the hydrate and acetone solvate. Without the original sample synthesised by Banik *et al*., or knowledge of the sample history associated with the reported DSC data, no firm conclusion can be made.

References

(1) M. Banik, S. P. Gopi, S. Ganguly and G. R. Desiraju, *Cryst. Growth Des.*, **2016**, 16, 5418.