New methods for structure determination and speciation by NMR crystallography

Martins BALODIS

Acknowledgements

First, I would like to thank the jury members **Ulrich Lorenz**, **Paul Hodgkinson**, **Elodie Salager** and **Jiri Vanicek** for reading and evaluating my thesis.

Second, I would like to thank my thesis supervisor **Lyndon Emsley** for guiding me through these years. I am thankful for the opportunity that you gave me to work in your group and I'll be forever grateful that I had the pleasure to work under your guidance.

Next, I want to say thank you to all of my lab colleagues, both the previous and the current. **Manuel Cordova, Albert Hofstetter, Pierrick Berruyer, Andrea Bertarello, Federico De Biasi, Michael Allan Hope, Aditya Mishra, Anna Melgares, Pinelopi Moutzouri, Yu Rao, Bruno Simões de Almeida, Daria Torodii, Fanny Racine, Amrit Venkatesh, Federico Paruzzo, Gabrielle Stevanato, Dominik Kubicki, Baptiste Busi, Brennan Walder, Snaedis Bjoergvinsdottir, Jayasubba Reddy Yarava, Claudia Avalos and Jasmine Viger Gravel.** This would not have been possible without the discussions with you in group meetings, coffee area, private messages and while collaborating.

I want to say thank you for the help, support and discussions with the people listed next, both collaborators and visitors to our lab that participated or showed interest in our projects. **Cory Widdifield, Michele Cheriotti, Brad Chmelka** and **Graeme Day**.

Special thanks to those who we sometimes forget but who play an extremely crucial part. They helped to solve technical issues, build new instrumentation and guided us when we were in need. **Harald Holze, Benjamin Charles Le Geyt, Aurélien Bornet, Emilie Baudat** and others, it was pleasure to work with you.

I want to thank my family and friends for all the support you gave me, especially **Liga Ezera, Koks, Marika Kacare** and **Irina Agapova**.

And finally, I want to thank the unmentioned ones as this list is becoming too long, but I do remember you all.

Abstract

NMR crystallography has been around for half a century, but with the advent of NMR crystal structure determination protocols in the last decade it has shown perspectives that were not seen before. Amalgamation of NMR and crystal structure determination has been successful in predicting crystal structures *de novo*. Still, there are many challenges to make these methods universal and applicable to any molecular crystal at any stage of a structural investigation. Larger molecules still take too much time to be solved by the current methods up to the point of being impossible. Amorphous structures of molecular solids had not been solved in general for any method. Reaction mechanisms and structures involved in the formation of solids are still challenging to investigate due to their fast nature and the low concentration of the reaction intermediates in solution.

The current NMR crystal structure determination protocols involve NMR at the final step of the candidate crystal structure selection, but one of the biggest bottlenecks is actually at the first steps of structure determination that involve selecting gas phase conformers that will later be used in the trial crystal structures.

Here, we develop a series of new NMR methods to address these bottlenecks. We show how unambiguous constraints extracted from solid state NMR experiments can help to significantly reduce the initial conformer space and help to select conformers that correspond more closely to what is found in the crystal structure. We then show how machine learned chemical shifts can be included in the crystal structure determination process from the start, and we determine the crystal structure of two compounds, one of whom is polymorphic. Then, we show how machine learned shifts in combination with molecular dynamics can be used to solve the atomic level structure of an amorphous compound yielding insights into the hydrogen bonding and stabilization of the amorphous form. To further aid crystal structure determination, we present a new method to assign spectra based on machine learned chemical shifts and propose a new database for small organic molecular crystals that would help in the future crystallographic investigations. Finally, we investigate carbonate speciation by using dissolution DNP.

Keywords

Résumé

La cristallographie par RMN existe depuis un demi-siècle, mais avec l'arrivée des protocoles de détermination de structures cristallines par RMN au cours de la dernière décennie, elle a ouvert des perspectives qui n'avaient pas encore été explorées auparavant. L'association de la RMN et de la détermination de structures cristallines a permis de prédire des structures cristallines *de novo*. Il reste cependant de nombreux défis à relever pour rendre ces méthodes universelles et applicables à n'importe quel cristal moléculaire et à n'importe quel stade d'une étude structurelle. Les plus grosses molécules prennent encore trop de temps pour être résolues par les méthodes actuelles, au point d'être impossibles. Les structures amorphes des solides moléculaires n'ont été résolues en général par aucune méthode. Les mécanismes de réaction et les structures impliqués dans la formation des solides sont encore difficiles à étudier en raison de leur nature rapide et de la faible concentration des intermédiaires de réaction en solution.

Les protocoles actuels de détermination de la structure cristalline par RMN impliquent la RMN à l'étape finale de la sélection de structures cristallines candidates, mais l'un des principaux goulets d'étranglement se situe en fait aux premières étapes de la détermination de la structure, qui impliquent la sélection de conformères en phase gazeuse qui seront ensuite utilisés dans les structures cristallines d'essai.

Nous développons ici une série de nouvelles méthodes RMN pour résoudre ces goulets d'étranglement. Nous montrons comment des contraintes non ambiguës extraites d'expériences de RMN de solides peuvent aider à réduire de manière significative l'espace initial des conformères et à sélectionner des conformères qui correspondent plus étroitement à ce que l'on trouve dans la structure cristalline. Nous montrons ensuite comment les déplacements chimiques appris par machine peuvent être inclus dans le processus de détermination de la structure cristalline dès le début, et nous déterminons la structure cristalline de deux composés, dont l'un est polymorphique. Nous montrons ensuite comment les déplacements chimiques appris par machine, combinés avec de la dynamique moléculaire, peuvent être utilisés pour résoudre la structure au niveau atomique d'un composé amorphe, ce qui permet de comprendre les liaisons hydrogène et la stabilisation de la forme amorphe. Afin de faciliter la détermination de la structure cristalline, nous présentons une nouvelle méthode d'attribution des spectres basée sur les déplacements chimiques appris par la machine et proposons une nouvelle base de données pour des cristaux moléculaires organiques, qui pourrait aider les futures recherches cristallographiques. Enfin, nous étudions la spéciation du carbonate en utilisant la polarisation nucléaire dynamique par dissolution.

Mots clés

Cristallographie par RMN (NMRX), prédiction de la structure cristalline (CSP), ShiftML, RMN à l'état solide, base de données RMN, affectation spectrale, dissolution DNP, matériaux amorphes, matériaux cristallins, carbonates.

Table of contents

List of publications

The present thesis is based on the following publications:

1. Balodis, M.; Cordova, M.; Hofstetter, A.; Ceriotti, M.; Day, G. M.; Emsley, L., *De novo* crystal structure determination via machine learned chemical shifts. Submitted.

2. Cordova, M.; Balodis, M.; Simoes de Almeida, B., Ceriotti, M., Emsley, L., Bayesian Probabilistic Assignment of Chemical Shifts in Organic Solids. Sci. Adv., in press.

3. Cordova, M.; Balodis, M.; Hofstetter, A.; Paruzzo, F.; Lill, S. O. N.; Eriksson, E. S. E.; Berruyer, P.; de Almeida, B. S.; Quayle, M. J.; Norberg, S. T.; Ankarberg, A. S.; Schantz, S.; Emsley, L., Structure determination of an amorphous drug through large-scale NMR predictions. Nat Commun **2021**, 12 (1).

4. Hofstetter, A.; Balodis, M.; Paruzzo, F. M.; Widdifield, C. M.; Stevanato, G.; Pinon, A. C.; Bygrave, P. J.; Day, G. M.; Emsley, L., Rapid Structure Determination of Molecular Solids Using Chemical Shifts Directed by Unambiguous Prior Constraints. J Am Chem Soc **2019**, 141 (42), 16624-16634.

Chapter 1.   Introduction

First, a few words about what makes NMR so special to be chosen as the principal axis of my thesis. As a technique NMR finds its way back to the previous century when Isidor Isaac Rabi in 1938 described an experiment to measure a nuclear magnetic moment.[1] He introduced the idea of flipping the magnetic rotation of a nucleus while it is immersed in a strong static magnetic field with the application of a small oscillating magnetic field at the right angle. It took almost a decade for this to catch on, when in 1946 both Felix Block and Edward Mills Purcell published the first papers demonstrating NMR on water and paraffin.[2-4] At that time, it was considered only as a precise method for measuring nuclear magnetic moments. It took 4 years for others to notice that the resonance frequency of a nucleus depends on its environment, with the discovery of the chemical shift.[5-6] That finally enabled NMR to be what it is known for these days - a powerful tool to study chemical systems. Many years have since passed, and NMR now is an established method used in every major chemistry and physics branch. Today, every major science branch needs to understand the underlying atomic-level structures of the materials that are being investigated and used. And that is where NMR crystallography comes into play as one of the strongest structural investigation tools known to the mankind.

1.1   Principles of NMR

In a nutshell NMR is similar to other spectroscopic techniques, it measures absorption of electromagnetic radiation as a function of frequency. What makes it unique is that the frequency range where NMR operates perturbs the nuclear magnetic moments (1.1) which in return lets us directly probe the local magnetic environments. $\boldsymbol{\mu_i}$ is a magnetic moment of an i-th nucleus, γ_i the gyromagnetic ratio of the nucleus and $\boldsymbol{I_i}$ the nuclear spin of the nucleus.

$$\boldsymbol{\mu_i} = \gamma_i \boldsymbol{I_i} \qquad\qquad 1.1$$

The sum of magnetic moments is net magnetization (1.2) and this magnetization is what we measure in an NMR experiment. Let's take a look where comes from and how we can use it.

$$\boldsymbol{M} = \sum_i \boldsymbol{\mu_i} \qquad\qquad 1.2$$

The magnetization in NMR is generated by nuclear spins. The spin systems are described by an NMR Hamiltonian.[7] The components of the Hamiltonian relevant to this thesis are the following:

$$\mathcal{H}_{NMR} = \mathcal{H}_Z + \mathcal{H}_{CS} + \mathcal{H}_D + \mathcal{H}_J \qquad\qquad 1.3$$

$\mathcal{H}_Z$ is the Zeeman Hamiltonian, $\mathcal{H}_{CS}$ describes chemical shielding, $\mathcal{H}_J$ describes J couplings and $\mathcal{H}_D$ dipolar interactions.

The Zeeman Hamiltonian describes the interaction of a nucleus with the external magnetic field. In case of a static external field, it can be written as following:

$$\mathcal{H}_Z = -\gamma B_0 I_z \qquad\qquad 1.4$$

Zeeman interaction is important, but in itself it is insufficient to describe the phenomena that gives rise to NMR spectra, and if that would be the case the spectra would be very boring. The main reason is that a nucleus does not experience only the static external magnetic field, but also the induced local magnetic fields arising from surrounding electrons and atoms. The arising change in chemical shift can be looked as perturbations that subtracts or adds to the field of nucleus. The Hamiltonian describing chemical shift can be written as shown in equation 1.5:

$$\boldsymbol{\mathcal{H}_{CS}} = \gamma \boldsymbol{I\sigma} B_0 \qquad\qquad 1.5$$

where I is the spin vector and σ is chemical shielding tensor coupling the spin and the external magnetic field. Chemical shielding is a 3x3 matrix and after careful choosing of the axis frame (principal axis system, PAS) it can be diagonalized as in 1.6:

$$\begin{pmatrix} \sigma_{xx} & & \\ & \sigma_{yy} & \\ & & \sigma_{zz} \end{pmatrix} \qquad\qquad 1.6$$

The isotropic chemical shielding is the average of the diagonal elements as shown in 1.7:

$$\sigma_{iso} = \frac{\sigma_{xx} + \sigma_{yy} + \sigma_{zz}}{3} \qquad\qquad 1.7$$

The isotropic chemical shielding is converted to the chemical shift δ via a reference chemical shielding value, and the chemical shift is the value that is usually reported. The chemical shift is sensitive to the local atomic environment which itself is influenced both by short- and long-range interactions, with the latter being especially important in solids. It is unique for every structure and possibly the most important observable in NMR crystallography. The main application of the chemical shift will be mentioned in the **chapters 2 and 3**.
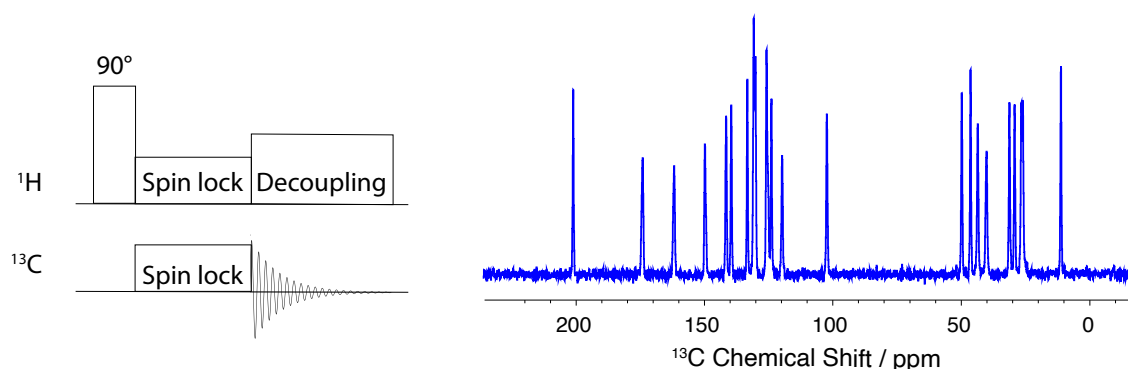
Figure 1-1 The workhorse experiment to measure the chemical shift in solid state NMR – CP (cross polarization) experiment. On the left a pulse sequence is shown and, on the right, a ^{1}H-^{13}C CP/MAS spectrum of an organic pharmaceutical AZD5718.

Another phenomenon which is very important in solid state NMR and crystallography is dipolar interaction described by dipolar coupling Hamiltonian $\mathcal{H}_D$ between two magnetic moments:

$$\mathcal{H}_D = \frac{\mu_0 \hbar^2 \gamma_1 \gamma_2}{4\pi |r|_{12}^3} \left(\mathbf{I_1} \cdot \mathbf{I_2} - \frac{3(\mathbf{I_1} \cdot \mathbf{r_{12}})(\mathbf{I_2} \cdot \mathbf{r_{12}})}{|r|_{12}^2} \right) \qquad 1.8$$

This equation can be furthermore transformed for the high field case where $\mathcal{H}_D \ll \mathcal{H}_Z$ and we arrive at:

$$\mathcal{H}_D = \frac{1}{2} \frac{\mu_0 \hbar^2 \gamma_1 \gamma_2}{4\pi |r|_{12}^3} (1 - 3cos^2\theta)(3I_{1z}I_{2z} - \mathbf{I_1} \cdot \mathbf{I_2}) \quad 1.9$$

There are two notable things in this equation. First, the dipolar interaction between two magnetic moments depends on distance, r^3. Second is that this interaction is angle dependant and if θ is 54.74 degrees, the term $(1 - 3cos^2\theta)$ goes to 0. Both of these things are very important in NMR crystallography. The r^3 dependence makes it possible to distinguish atoms that are close in space from atoms that are far in space. The angular dependence of the dipolar coupling on the other hand is one of central themes in NMR and led to the development of countless methods to reduce the linewidth of NMR signals in solids. From mechanical methods the most notable is magic angle spinning first described by E. R. Andrew and I. J. Lowe.[8-9] While it is possible to reduce dipolar couplings by rotating the sample in space, it is also possible to achieve a similar outcome by applying this principle in the spin space. There is a plethora of methods applied both for homonuclear dipolar decoupling[10-24] and heteronuclear dipolar decoupling.[25-32] They are always combined with magic angle spinning as the latter also removes chemical shift anisotropy and in case of homonuclear decoupling they are called combined rotation and multi-pulse spectroscopy (CRAMPS) methods.[33]

Last but not least are J couplings (also called indirect coupling) described by $\mathcal{H}_J$. It is a coupling that happens via bonds and therefore provides information on connectivity. It is widely used in solution state NMR, but not as much in solid state NMR because it is relatively small compared to other NMR interactions. Nevertheless, it still finds its uses and with recent NMR coherence lifetimes becoming longer and longer one can predict that the usage of the indirect coupling will become even more prevalent in the future. Even now, to understand the crystal structure the NMR peaks need to be assigned and one of the cornerstone experiments for that is refocused INADEQUATE[34-35] that relies on J-based interactions. J couplings also contain information on bond lengths and angles therefore giving complementary information to dipolar and chemical shift interactions.[36-41]

1.2   NMR crystallography

NMR crystallography, as the name suggests, is to obtain crystallographic data using NMR. The first example of NMR crystallography was recorded in 1948 when G. E. Pake measured the distance between hydrogen atoms in gypsum.[42] Since then NMR crystallography has evolved and there are many good reviews and articles[43-48] for those who want to delve deeper. Here I will try to give a more condensed overview about the field to transition into my own

contributions. The main areas that NMR crystallography covers can be broadly classified as 1) *de novo* structure determination using NMR data, 2) structure refinement against NMR data and 3) cross validation of structure models using NMR data.[44] Materials that can be examined with NMR crystallography include a wide array of chemical species: proteins, glasses, polymers and compound materials like cement, but from now on I will concentrate on NMR crystallography of small organic molecules (molecular solids).

The characterization of materials is important, especially in the pharmaceutical industry where appearance of new polymorphic forms can be disastrous, such as in the case of ritonavir[49] and rotigotine[50] where appearance of new polymorphs resulted in a temporary withdrawal of the drugs from the market. The field of science that characterizes the three-dimensional arrangement of atoms in solids is crystallography. When scientists speak about crystallography, they still mainly think about single crystal X-ray diffraction. And there is a reason for that because single X-ray diffraction is one of the most practical methods to determine the structure of a crystalline compound and has been since Max von Laue announced the discovery.[51-52] Every method has its drawbacks and so does X-ray diffraction. First, a single crystal of a proper size is needed for XRD, and while in some cases it is hard to make a proper size single crystal, in many cases it is impossible to make one without changing the structure itself. Amorphous materials and heterogenous systems such as pharmaceutical formulations are a few important examples where it is impossible to use single crystal X-ray diffraction. Powder X-ray diffraction is usually much less informative. Many new materials are created via mechano-chemistry by grinding and the crystallites made by grinding are usually too small for single crystal X-ray diffraction. Powder X-ray diffraction is often used to work with microcrystalline samples, but due to the much less information than single crystal X-ray diffraction, structure solving is challenging and often NMR is employed to help in structure determination and to validate the determined structures.[53-56]

NMR naturally complements diffraction methods. While diffraction methods get the information from long range periodicity, NMR probes the local environment. Both single crystal and powder X-ray diffraction are sensitive to both static and dynamic disorder happening in the system as they disturb the quality of the solution.[57] For this reason, single crystal X-ray diffraction is usually done at low temperatures to quench the motions and remove the dynamic disorder. That can induce phase transitions or conformational changes different from the room temperature. NMR in opposite is embracing dynamics and is often used to understand the motions and disorder in the system.[58-59] XRD struggles to find H atoms in structure due to weak scattering of only one electron that H atom possesses. It also struggles to differentiate isoelectronic species such as OH and F and CH_3. Neutron diffraction can help both with locating H and differentiating isoelectronic species, but it also requires much larger samples and is less readily accessible than X-ray diffraction or NMR.[43] In last years a significant contribution from electron diffraction have been made in the structure determination[60-64] but this technique is considered challenging and not always applicable. All of the mentioned techniques have their weak sides which make them either inapplicable to solve the structure or they need to be applied in tandem with NMR. A good example to illustrate combined efforts of diffraction methods and NMR was a study of di- and tri-*p*-benzamides.[65] As these chemicals crystallize poorly, which excluded the use of single crystal XRD, they were solved by the combined effort of NMR, electron diffraction and powder XRD.

Initially NMR in crystallography was used as a complementary tool for XRD, and still is an excellent tool to provide answers to questions that x-ray diffraction methods can't answer or struggle with. It is also a great validation tool for structures solved by powder X-ray diffraction. A relatively straight forward example being the number of inequivalent molecules in asymmetric unit. A simple 1D spectrum in most cases is enough to see how many molecules the asymmetric unit contains[66-68] while sometimes a more complicated investigation can reveal a mistake in the asymmetric unit in already determined structures.[69] NMR can validate structures obtained by PXRD and single crystal XRD especially in cases where the solution is ambigious.[55, 70-73] Hydrogen bonding patterns can be determined by NMR crystallography[65] and notably hydrogen positions can be correctly located such as in the case for furosemide[53] or tautomeric cyameluric acid.[74] NMR crystallography is applied to explain the structure formation, for example, co-crystal formation by investigating intra- and intermolecular bonds[75] or using dipolar information to confirm predictions.[76] NMR and diffraction protocols can be combined to drive the structure determination and to give better answers than the diffraction methods alone could give.[77-83]

One of the reasons that made NMR so powerful for small organic molecules was the introduction of the gauge-induced projector augmented wave (GIPAW) approach.[84-86] It enabled calculations of chemical shielding in periodic systems that were not so easily available before. In comparison to previous models that used molecular orbitals[87] GIPAW accounted for periodic conditions using plane wave-based pseudopotential formalism of the density functional theory (DFT). Suddenly, it was possible to calculate the shielding from the first principles with high accuracy and at a reasonable time. The calculations were implemented and made readily available to public via software packages like CASTEP[88] and Quantum Expresso.[89] The implementation of GIPAW in different codes vary slightly, mainly because of the difference pseudo-potentials used. A comprehensive study done by Lejaeghere and coworkers[90] showed that the most recent codes and methods converge to a single calculated value with errors comparable to those of experiment. For that calculated values of 71 crystals were compared using 15 most widely used DFT codes and employing 40 different potentials. The results gave confidence that the current codes are close to the intrinsic limits of PBE-based DFT. The ability to reliably optimise crystalline structures and calculate NMR chemical shifts led a rise of structure determination using only NMR data. DFT-GIPAW approach now is the standard to calculate NMR shifts for periodic crystals, but occasionally DFT-GIAO based approach can be used when working with small clusters of molecules.[91]
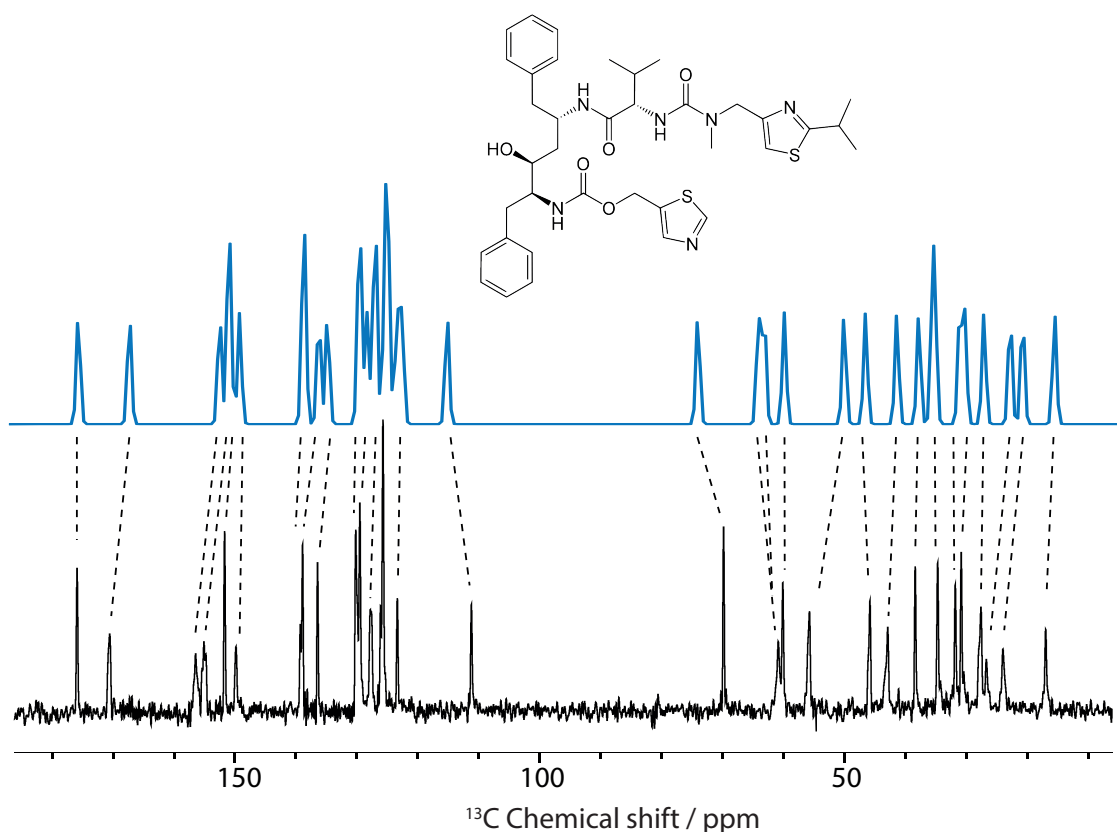
Figure 1-2 An example of a ^{13}C chemical shift DFT calculation done on a drug molecule ritonavir. The top spectrum is the computed spectrum and the bottom spectrum is the experimental ^{13}C CP/MAS spectrum shown for comparison. The dotted lines connect the corresponding resonances.

One obvious way to do crystal structure determination would be to use distances calculated from the NMR data. As mentioned in the previous chapter, dipolar interactions are distance dependant and therefore in theory it is possible to extract interatomic distances from them as Pake showed.[42] Indeed, a lot of progress has been done to measure the internuclear distances[92-102] and also bond [103-106] and dihedral angles.[107-109] While possible, it is still quite challenging. First, one needs to isolate individual couplings which is usually done by selective labelling. Second, magic angle spinning that is usually applied in the experiments to resolve peaks averages dipolar couplings so a recoupling[110] is needed to re-introduce the anisotropic interactions.[110-111] Still, information arising from dipolar couplings have influenced much work in crystal structure determination mentioned in the next paragraphs.

Proton spin diffusion is a process where the magnetization is exchanged between protons according to a process driven by the internuclear distance dependent dipolar coupling.[112] Elena introduced[113] a model of spin diffusion using a phenomenological kinetic rate matrix approach:

$$k_{ij} = A \left(\frac{\mu_0 \gamma_H^2 \hbar}{4\pi} \right)^2 \Sigma_m \frac{1}{\left(r_{ij}^n \right)_\lambda} \qquad\qquad 1.10$$

Where k_{ij} is the rate of exchange between two spins, r_{ij} s the internuclear distance between atoms i and j, λ indicates the sum over exchange between sites i and j in different molecules in the crystalline lattice, A is a phenomenological scaling constant and n, although being a variable was set to 6.

The model allowed to model proton spin diffusion curves and it was shown that the fit between calculated and experimental diffusion curves were sensitive to the local environment up to 6 Å which is enough for crystal structure refinement. The first successful attempt to solve the structure only by NMR and molecular modelling was done by Salager *et al.* in 2006 where she and co-workers did a full structure determination of β-l-aspartyl-l-alanine using proton spin diffusion curves and comparing the calculated ones with the experimental ones.[114] After this it was shown that the quality of the calculated structures can be further improved by DFT geometry optimisation.[115] That led to Salager *et al.* combining the use of proton spin diffusion curves as optimisation restraints with the refinement of the structures with DFT. In the same publication they introduced the usage of DFT calculated chemical shifts as a validation tool with comparing the calculated structures against the experimental ones.[13] One must note that the recording of proton spin diffusion curves is experimentally challenging and requires either well separated proton signals or selective labelling, so shortly after a more general approach was introduced that required only chemical shifts from 1D spectra of ^{1}H and ^{13}C.[116] It combined the recent progress in crystal structure determination from first principles[117] with NMR. At that time the crystal structure prediction computational community had reached a point where they were able to predict a set of structures that usually contained the actual structure. It was shown on thymol that NMR can reliably select the correct structure from this set of candidates by comparing the predicted and experimental ^{1}H shifts. For the first time uncertainties of calculated values were alos estimated, resulting in an average rms error of 0.33 ppm (±0.16 ppm) for ^{1}H chemical shifts and 1.9 ppm (±0.4 ppm) for ^{13}C calculated on 15 organic compounds. I was also shown that the right structure is not necessarily the lowest energy one found computationally. This work established the current NMR structure determination scheme shown in **Figure 1-3**.
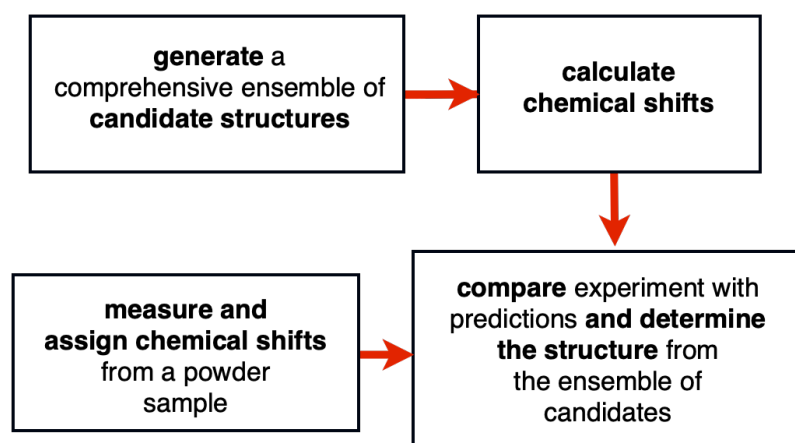
Figure 1-3 The general scheme of the current NMR crystal structure determination approach using chemical shifts.

Baias *et al.* used the same approach when determining the structure of cocaine, flutamide, flufenamic acid and theophylline.[118] Again, crystal structure prediction was used to determine the possible candidate structures without any prior experimental knowledge, except chemical

connectivity. On the other hand, the method failed to sufficiently discriminate amongst theophylline candidate structures, which didn't come as a surprise knowing that theophylline contains only 4 distinct proton chemical shift sites. At the same time Baias *et al*. also did the first *de novo* NMR crystal structure determination on a pharmaceutical AZD8329.[119] With the molecular weight of 422 g/mol AZD8329 was the largest molecule so far tackled by NMR crystal structure determination, **Figure 1-4**. The determination was complicated by the partial assignment of chemical shifts therefore all possible assignments were kept in mind when comparing with the experimental shifts. Also, it was not known if the conformer found in the crystal is in *cis* or *trans* conformation, but comparison with proton chemical shifts effectively solved this problem by only giving good agreements with the *cis* conformer.
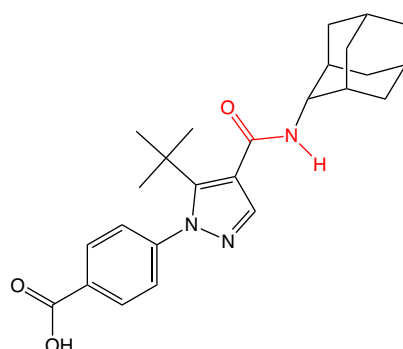
Figure 1-4 The structural formula of AZD8329. The red fragment is the fragment under questioning in the work done by Baias *et al*.[119] where cis and trans conformations were compared.

As an alternative to chemical shifts, ^{13}C-^{13}C dipolar couplings were proposed by Mollica *et al*. as a tool to discriminate between CSP generated crystal structures. They showed the possibility to use ^{13}C-^{13}C dipolar couplings to narrow down the possible space groups and even to determine the unit cell parameters.[120] They also showed that by employing ^{13}C-^{13}C dipolar couplings they are able to discriminate between different polymorphs of theophylline.[118, 121] In later work they were able to reduce the CSP generated crystal structures to just two as the possible structural candidates.[122] An another computer-generated structure discrimination tool was proposed by Harper *et al*. and involved usage of chemical shift tensors to complement isotropic chemical shifts[123-124] which were crucial in finding the most likely structures of ^{129}Xe clathrates.[125]

Brus *et al*. validated the NMR crystallography approach developed used by Salager and Baias.[126] After comparing ^1H and ^{13}C calculated and experimental values of decitabine they were left with two possible candidate structures. To find the correct structure they further restricted the structural space by analysing ^1H–^{13}C HETCOR and ^1H–^1H double-quantum/single-quantum NMR correlation spectra and determined the covariance. Following that, they employed the same method to solve the structure of decitabine and sebacic acid within polymeric microbeads.[127] They showed that NMR crystal structure determination can solve the structure of multicomponent polymeric solids with complex nanodomain architecture. As these systems exist at the borderline between crystalline and amorphous solids, for diffraction techniques solving them is extremely challenging.

Enzyme active sites are other difficult and elusive systems to investigate. Mueller and Dunn showed in their work that the synergistic combination of NMR crystallography, diffraction methods and computational methods can probe the active site of a protein with high accuracy. By combining these methods and comparing the experimental and calculated chemical shift they were able to give a new acid-form hypothesis for the indoline quinonoid intermediate in tryptophan synthase.[128]

The above shows the importance of chemical shifts in NMR crystallography and in particular, in NMR crystal structure determination. With NMR crystallography providing answers with comparable accuracy to more traditional methods like X-ray diffraction, the need for fast and accurate calculations of chemical shifts is becoming more and more prevalent. While traditional DFT based methods are accurate their cubic scaling of computational cost with system size makes them anything but fast and renders their use applicable only to a relatively small set of up to medium size crystals. Usage of machine learned properties are becoming more and more prevalent and almost every month we see a new breaking discovery, like the recent introduction of AlphaFold that is used for highly accurate protein structure predictions.[129] Paruzzo et al. recognized the potential of machine learning and invented a fast and accurate model to calculated chemical shifts dubbed ShiftML.[130] 2000 diverse structures selected from the Cambridge Structural Database (CSD) were described by the smooth overlap of atomic positions (SOAP) kernel[131-132] and the Gaussian process regression (GPR) framework was used to predict the chemical shift **Figure 1-5**.[133] With the later introduced uncertainty estimation[134] the resulting model was able to accurately predict chemical shifts in seconds that would otherwise take hours or even days. In **Chapter 2.2 and 2.3** we use ShiftML to rapidly calculate chemical shifts for our large-scale calculations.
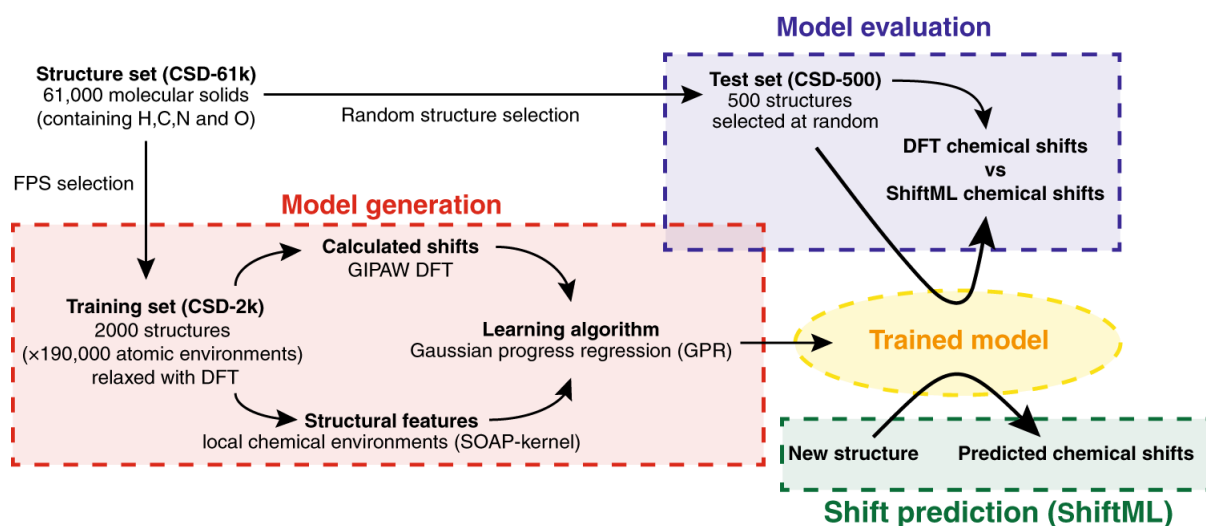
Figure 1-5 The scheme of the machine learning model used for chemical shift predictions.[130]

Since 1999 the Cambridge Crystallographic Data Centre (CCDC) has held 6 crystal structure prediction blind tests.[117, 135-139] Each of these tests has brought together computational scientists from different fields to use state of the art methods to predict previously unknown crystal structures and compare the methods against each other in doing so. In the latest, the 6th blind test[139] 5 different chemical systems were given for the participants to solve. All systems were flexible except the first and contained a semi-rigid molecule, a multi-

polymorphic system, a salt, co-crystal and a large molecule with 5 rotatable bonds. All but one of the molecules were predicted by at least one of the blind test participants. The molecule that was not predicted was a polymorph that contained 2 molecules in the asymmetric unit. With all this success possible structure generation is limited by the computational cost that scales up drastically with the introduction of more variables that usually correlates with the size of the system. Studies that involve larger molecules than the ones in the 6th blind test usually involve some experimental knowledge. For example, a study on one of the forms of AZD7624 used conformational information derived from sXRD as the molecule itself contained 11 non-trivial rotatable bonds and 2 molecules in the asymmetric unit therefore being too big for the standard CSP methods to solve in a reasonable time.[69] When trying to solve for the structure of the form I of metergoline the best matching structure generated by CSP gave ^{1}H rmsd of 0.85 ppm while the XRD determined structure gave 0.39 ppm thus being in the limits estimated by Salager.[116, 140] Authors also investigated the form II of metergoline, but didn't even try to do crystal structure prediction on it as the form II contains two molecules in the asymmetric unit thus being a considerable challenge for CSP algorithms. This example illustrates the biggest bottleneck in the NMR crystallography approaches presented here - fast and reliable generation of possible candidate structures and high computation cost associated with that. The first way to deal with the currently high computational cost is development of faster, better algorithms and methods helped by the ever-increasing computational power. Recently there has been work done on machine learning algorithms to predict and analyse the energy landscape and properties of organic molecules in solids.[141-143] The second way is the usage of the experimental parameters in the structure prediction algorithms and we introduce new methods to do that in **Chapters 2.1** and **2.2** of this thesis.

1.3 Dissolution dynamic nuclear polarisation

One of the biggest drawbacks of NMR is its intrinsically low sensitivity caused by the small thermal equilibrium population difference between spin states. This is relevant to a broad audience because many chemical species are available only in small quantities, such as proteins or surface catalysts. The detected signal is proportional to the polarization, which in turn is governed by the Boltzmann factor:

$$P = tanh\left(\frac{\gamma h B_0}{4\pi k_B T}\right)$$

<div align="right">1.11</div>

Here, P is polarization, γ is gyromagnetic ratio, h is Planck constant, k_B is Boltzmann constant, B_0 is the external magnetic field and T is temperature.

From equation 1.11 it follows that it is possible to decrease the temperature and increase the external magnetic field to make polarization gains, but that is a limited playfield. While NMR magnets increase their strength gradually over years, the increase is slow and linear, and for solutions there is only so far one can go down in temperature before the phase changes. Another possibility is to use polarization transfer from large γ nuclei to low γ nuclei and this method is commonly applied, but the overall gains are still small, e.g., maximum 4 in case of transfer from ^{1}H to ^{13}C. Small in comparison to polarization enhancements gained from hyperpolarization techniques that can generate much larger, non-thermal population

differences. Amongst these are para-hydrogen induced polarization (PHIP)[144], chemically induced dynamic nuclear polarization (CIDNP)[145], signal amplification by reversible exchange (SABRE)[146], optical pumping[147], spin-exchange optical pumping (SEOP)[148] and dynamic nuclear polarization (DNP)[149]. The latter is the most versatile method and currently has the broadest applicability.

In 1953 Overhauser proposed that the much larger polarization from electron spins can be transferred to nuclear spins by microwave irradiation[149] with a theoretical maximum enhancement factor of around 657 for ^{1}H at room temperature. Shortly after it was verified by T. R. Carver and C. P. Slichter in metals.[150] As the acquisition time scales inversely with the square of the enhancement factor, tremendous time savings can be achieved and systems unreachable before may now be explored. This knowledge led to the pioneering work of R. G. Griffin who introduced advanced instrumentation consisting of a high power/frequency gyrotron to drive the DNP effect in a high-field DNP NMR spectrometer.[151-154] A decade later commercial companies started to build the instrumentation for DNP and that led to DNP being accessible for the rest of the solid-state NMR community.[155-156]

Today DNP is usually done with glassy solids around 100K under magic angle spinning, but in 2003 Ardenkjaer-Larsen introduced a novel method, dissolution DNP.[157] Dissolution DNP is based on the notion that the sample can be hyperpolarized at low temperatures, usually 1-4 K, then dissolved in a fraction of a second, transferred to an NMR spectrometer or an MRI scanner and a at room temperature a spectrum can be acquired on the still hyperpolarized liquid state sample, **Figure 1-6**.
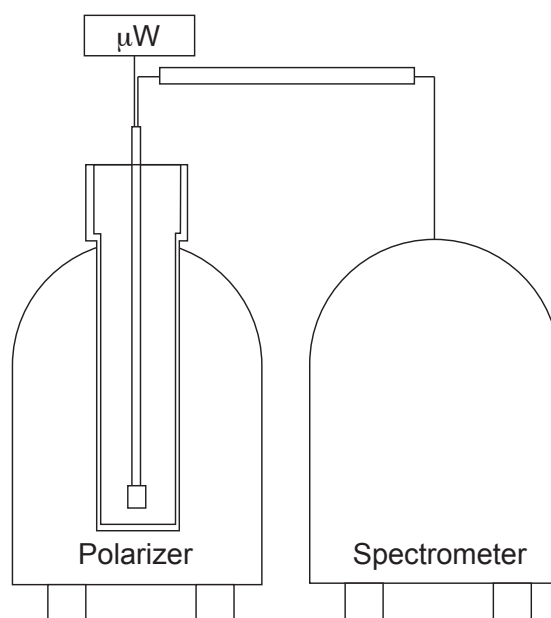
Figure 1-6 A simplified dissolution DNP scheme. The polarizer is where the sample is irradiated with microwaves to hyperpolarize it. The sample is held at low temperature, typically 1-4K. After the polarisation the sample is flash dissolved and transferred to a spectrometer where the spectral acquisition is done, typically at room temperature.

First, dissolution DNP was proposed for applications of metabolic imaging[158], but the possible applications range all across the spectrum of chemistry and biochemistry due to the high sensitivity boost of the method.[159-169] Currently the main drawbacks of dissolution DNP are that signal averaging is not possible and the dissolution and injection process limits the maximum sample concentration and minimum spin-lattice relaxation time of the analyte. Therefore, compounds with good solubility and long relaxation times are preferred. The lack of the possibility to do many scans renders the ability to use conventional 2D experiments, but a variety of ultrafast 2D experiments have recently been developed.[170-172]

In the light of NMR crystallography, dissolution DNP might be a method to provide access to the nucleation and growth mechanisms of solid structures which we investigate in **chapter 5**.

1.4 Outline of the present thesis

In the introductory part of this thesis, I have given an outlook for solid state NMR and shown the importance of NMR crystallography. Also, the concepts of dissolution DNP have been introduced.

In **Chapter 2** I will show the advances in NMR crystallography that have been made in the course of this thesis. First, I will show how NMR can be used to accelerate the crystal structure determination by usage of unambiguous experimentally determined constraints. Second, I'll show how machine learned isotropic chemical shifts can be used in the structure determination procedure, something that has not been possible until now. And lastly, I will move from crystalline to amorphous substances where machine learned chemical shifts will be used to determine the structure of an amorphous compound.

In **Chapter 3** a new assignment method for organic molecules in solid state is introduced. It is based of probabilistic assignment of machine learned chemical shifts on structures taken from Cambridge Structural database. Next, progress on the first solid state NMR database for organic small molecules is shown.

In **Chapter 4** I will speak about investigation of carbonate speciation, specifically, prenucleation clusters of calcium carbonate via dissolution DNP.

In **Chapter 5** I will summarize the results.

Chapter 2.   Chemical shift driven crystal structure determination

The 40,000-60,000 new crystal structures published every year[173-175] perfectly illustrate the importance of the knowledge of atomic level structures of solids, which is key to understanding and predicting their properties. For example, in pharmaceutical compounds crystal structures guides the understanding of physicochemical and pharmacokinetic properties such as bioavailability and/or solubility.[176] However, many active pharmaceutical ingredients (APIs) are only available as powders, and therefore are not amenable to resolution with typical X-ray diffraction methods. Structure elucidation can be complicated further if, for example, the crystallites are sub-micron in size, or the structure contains elements of disorder. Therefore, NMR crystallography together with CSP can be used as described in **Chapter 1.2**.

A common feature of most CSP-NMRX methods developed to date is that they exploit constraints from solid-state NMR only in the final step, in order to select the correct crystal structure from an ensemble of predicted structures. **Introducing experimental constraints earlier** in the CSP process would be an obvious way to guide and accelerate structure determination.

The bottleneck for CSP of flexible molecules usually relates to the size of the molecular conformational space, so guidance to constrain the size of the search space would be most valuable if it relates to single molecule conformations. However, it is not immediately clear how experimental measurements on the crystalline samples would be relevant to restrict the single molecule conformational space.

In **Chapter 2.1** we introduce a CSP-NMRX method to determine crystal structures in which we use unambiguous constraints from solid-state NMR on microcrystalline samples to restrict the CSP search space to relevant regions of single molecule conformational space. The approach directs the determination procedure from the first steps towards the correct crystal structure, without the need for assumptions. We parametrize the approach on the crystal structures of cocaine, flutamide, and flufenamic acid and demonstrate a significant acceleration in computational times for these compounds. Most significantly, using chemical shifts calculated with both DFT and machine learning,[177] we correctly determine the crystal structure of powdered ampicillin, for which the usual approach to CSP-NMRX would have failed.

Established approaches to *de novo* structure determination, for example by single-crystal X-ray diffraction of large molecules or by solution NMR, usually involve an iterative process where a (random) starting structure is optimized under the combined effect of an (empirical) energetic potential and a penalty term that compares the computed observables with the measured values at every step of the optimization.[178] This is a very powerful approach to finding the correct structure, and is enabled by the fact that the calculation of observables from any trial structure is very rapid. So far, this has not been possible, with a few notable exceptions where chemical shifts were incorporated and derived from parametrized force-fields.[179-180] To make this approach general the calculation of chemical shifts so far would have required highly accurate but very time consuming DFT calculations.[84-86, 181-182] As described in

the Chapter 1, this results in structure determination requiring first the generation of a large ensemble of credible candidate structures, usually done with some form of computational crystal structure prediction protocol[183-187], followed by DFT chemical shift calculations for the set of candidates, and only at the end of this process is there a comparison with the experimental shifts to determine which is the correct structure. While powerful, this is a time consuming and laborious approach which is obviously not ideal.

In **Chapter 2.2** we successfully determine the crystal structures of two polymorphs of the drug molecule AZD8329 and cocaine using on the fly generated machine-learnt isotropic chemical shifts to directly guide a Monte-Carlo based structure determination process starting from random gas phase conformers.

Chapters 2.1 and **2.2** concentrate on crystalline solids, but solid-state NMR is also amongst the most popular methods to study the structure of amorphous materials. Still, while two-dimensional correlation experiments are able to identify intermolecular contacts between atom pairs,[188-190] obtaining complete atomic-level structures is a challenge due to the disordered molecular environments present in amorphous solids.

Candidate structures of amorphous compounds can be generated using MD simulations,[191-193] however the large cell sizes required to model amorphous structures, together with the large number of structures needed to describe amorphous ensembles, have so far prevented the computation of chemical shifts in these systems. The recent introduction of ShiftML[130] described in Chapter 1 can now be used to predict chemical shifts for any molecular solid (currently limited to solids comprised of C, H, N, O, S atoms) in a matter of seconds,[130, 194] which opens up the possibility for large-scale shift computations on large structures.

In **Chapter 2.3** we investigate the structure of anhydrous crystalline AZD5718 form A, by combining measured ^{1}H, ^{13}C and ^{15}N chemical shifts obtained using DNP-enhanced NMR experiments from a powder sample, CSP, and DFT chemical shift computations. The structure is validated with that obtained from single-crystal XRD. We then model the hydrated amorphous drug with different water contents using MD simulations and obtain predicted NMR spectra for large structural ensembles using machine learned chemical shifts. We then analyse the ensembles to identify the different hydrogen bonding motifs present in the amorphous structures, by comparison of the experimental and predicted chemical shift distributions associated with these structural motifs. From the amorphous structures we also compute the interaction energy between molecules of AZD5718 and their environment, and we relate the energies to the local hydrogen bonding motifs.

2.1 Rapid structure determination of molecular solids using chemical shifts directed by unambiguous prior constraints

This chapter has been adapted from Hofstetter, A.; Balodis, M.; Paruzzo, F. M.; Widdifield, C. M.; Stevanato, G.; Pinon, A. C.; Bygrave, P. J.; Day, G. M.; Emsley, L., Rapid Structure Determination of Molecular Solids Using Chemical Shifts Directed by Unambiguous Prior Constraints. J Am Chem Soc 2019, 141 (42), 16624-16634.

My contributions to the publication. Performing NMR experiments, participating in the development of the methodology and investigation and participating in the writing and editing of the paper.

2.1.1 Introduction

NMR based crystallography approaches involving the combination of crystal structure prediction methods, ab-initio calculated chemical shifts and solid-state NMR experiments are a powerful approach for crystal structure determination of microcrystalline powders. However, currently structural information obtained from solid state NMR is usually included only after a set of candidate crystal structures has already been independently generated, starting from a set of single molecule conformations. Here, we show with the case of ampicillin that this can lead to failure of structure determination. We propose a crystal structure determination method that includes experimental constraints during conformer selection. In order to overcome the problem that experimental measurements on the crystalline samples are not obviously translatable to restrict the single molecule conformational space, we propose constraints based on the analysis of absent cross-peaks in solid-state NMR correlation experiments. We show that these absences provide unambiguous structural constraints on both the crystal structure and the gas phase conformations, and therefore can be used for unambiguous selection. The approach is parameterized on the crystal structure determination of flutamide, flufenamic acid, and cocaine, where we reduce the computational cost by around 50%. Most importantly, the method is then shown to correctly determine the crystal structure of ampicillin, which would have failed using current methods because it adopts a high energy conformer in its crystal structure.

2.1.2 Methods

Figure 2-1a schematically illustrates the workflow in a successful case for the current CSP-NMRX approaches.[195-197] In the first step, the torsional degrees of freedom are explored to generate a comprehensive ensemble of energetically stable single molecule conformers. The ensemble is then sorted according to the calculated conformational energies and the lowest energy conformers are selected to proceed to the next step, based on an empirical cut-off energy. Although flexible molecules often do not assume their lowest energy molecular conformation in their observed crystal structures,[198] the assumption here is that low energy crystal structures, including the correct (observed) polymorph, will generally result from low

energy molecular conformers. However, this is not always the case, as will be demonstrated below.

The selected conformations are then each subjected to a crystal structure search, during which trial structures are generated by varying the unit cell dimensions, center of mass in the cell, packing symmetry, and the number of molecules per asymmetric unit. This process can lead to hundreds or thousands of possible crystal structures from each single molecular conformer. The energy of each structure is then minimized, typically using atom-atom force fields and DFT.[199]

Next, this ensemble is ranked by calculated lattice energy and again only the structures below a given cut-off energy are retained. In the final step, these structures are further optimized, typically using periodic boundary DFT calculations, and then the chemical shifts (or other experimental data such as dipolar couplings or chemical shift anisotropies)[123, 200-204] for this sub-ensemble of crystal structures are calculated and compared to experimental chemical shifts measured on a powder sample. The differences between the calculated and the experimental chemical shift data are then used to determine the crystal structure that is in best agreement with the experimental NMR data acquired using the powder sample. For this structure, positional errors are then calculated using a molecular dynamics approach.[205] Note, that the computational cost rises sharply when moving from the energy calculations of a single molecule to lattice energy calculations to GIPAW DFT chemical shift calculations, thus requiring the use of successive selection steps to reduce the number of candidate structures at each stage.

From the description of the CSP-NMRX procedure above, it is evident that a gas phase conformer similar to the one present in the correct crystal structure must be among those initially selected.

Figure 2-1b illustrates a case where the current CSP-NMRX method fails. Analogously to the previous case, a large ensemble of single molecule conformers is generated and sorted by conformational energy. However, here *the molecular conformer present in the crystal structure is of very high relative energy in the gas phase, and thus does not pass the selection criteria by energy*. An illustrative example of this case could be when intra-molecular hydrogen bonds stabilize the most stable conformations in the gas phase, while the crystal structure conformation is stabilized through inter-molecular hydrogen bonds or other interactions only present in the solid phase. Thus, following the normal selection steps based on the conformational energy, the correct conformer is not included in the crystal packing and lattice energy calculation steps, and as a consequence is not present in the trial crystal structures that are compared to the experimental data.

Taking this into account, one could extend the crystal structure determination procedure, and we consider two ways below. One option is to loosen the initial selection criteria, thus allowing more conformers to proceed to the following steps. This approach will increase the computational cost, often prohibitively. Even moderately flexible molecules can have hundreds of conformations, each requiring significant computing; for example, for ampicillin, one of the molecules studied here, the CSP procedure required, on average, just over 3 days of computing on 200 dedicated CPUs per conformer (yielding a total of 54 days for all the

conformers). Thus, this approach either involves very long timescales or requires access to very large-scale computing. The second option is to use a different initial selection criterion including information from experiment.

Figure 2-1c illustrates this second approach, which we introduce here. Contrary to the standard CSP methods, no assumptions based on calculated energy are made in the initial conformer selection process. Instead, a sub-ensemble of conformers is selected using experimental constraints from solid-state NMR experiments on the powdered microcrystalline sample. This approach guides the conformational sub-ensemble selection towards the correct crystal conformer, and thus reduces the chances that the structure determination is limited by possibly erroneous assumptions.

(a)

single molecule conformations

conformational energy (gas phase DFT)

energy cutoff

molecular crystals

lattice energy (force-field and solid state DFT)

energy cutoff

comparison to experiment

chemical shift error (ab-initio vs. experiment)

shift uncertainty

crystal structure determined!

+

positional error analysis

$^1H\ \delta_{GIPAW\text{-}DFT}$ / ppm

$^1H\ \delta_{Experimental}$ / ppm

(b)

single molecule conformations

conformational energy (gas phase DFT)

energy cutoff

molecular crystals

lattice energy (force-field and solid state DFT)

energy cutoff

comparison to experiment

chemical shift error (ab-initio vs. experiment)

shift uncertainty

structure determination fails!

$^1H\ \delta_{GIPAW\text{-}DFT}$ / ppm

$^1H\ \delta_{Experimental}$ / ppm

(c)

single molecule conformations

experimental constraints from the crystal structure

+

^1H, ppm

^{13}C, ppm

molecular crystals

energy cutoff

lattice energy (force-field and solid state DFT)

comparison to experiment

chemical shift error (calculated vs. experiment)

shift uncertainty

crystal structure determined!

+

positional error analysis

$^1H\ \delta_{GIPAW\text{-}DFT}$ / ppm

$^1H\ \delta_{Experimental}$ / ppm

Figure 2-1 Schematic of current and proposed CSP-NMRX methods. (a) an example of a successful structure prediction using the current CSP-NMRX method. (b) an example of a failed structure prediction using the current CSP-NMRX method. (c) an example of the proposed experimentally constrained CSP-NMRX method, which successfully overcomes the failure of the current CSP-NMRX method shown in panel (b). In each panel the structures in the first line depict single molecule gas phase conformations sorted by their conformational energy. After applying a given selection criterion a reduced conformer set is used to generate an ensemble of possible crystal structures (represented by the 2nd line in each panel). The coloured boxes are intended as a guide to the eye, as to which conformer results in which crystal structures. The 3rd line in each panel represents crystal structures picked from the 2nd line after a further selection criterion is applied. This final set of structures is then compared to the experimental chemical shifts, to determine the correct crystal structure. In each panel the scatterplot shows the experimental 1H chemical shift plotted against the DFT-calculated 1H chemical shift for the trial structure with the lowest error between DFT and experimental chemical shifts.

However, experimentally we only have access to the full crystal structures and cannot probe the underlying "virtual" gas phase conformations independently. Thus, we need to measure experimentally accessible constraints that would be unambiguously fulfilled both in the crystal structure as well as in the gas phase conformations. Note that commonly used solid-state NMR constraints, such as the presence of (dipolar-coupling mediated) cross peaks in NMR correlation experiments[114, 206-216]due to internuclear proximity, do not contain unambiguous information about the gas phase conformations. This is because a cross peak could arise either from intra or inter molecular proximity.
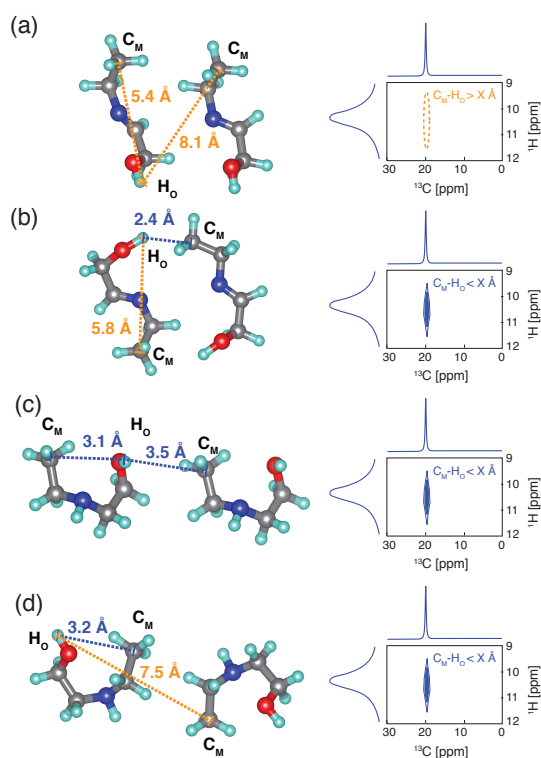
Figure 2-2 Schematic illustrations of ^{1}H-^{13}C HETCOR spectra (right) for four different structural fragments (left) and the derived constraints. Structures (a) and (b) contain an "open"

conformer. Structures (c) and (d) contain a "closed" conformer. Blue dotted lines are sufficiently short C-H distances between C_M and H_O to generate peaks in the spectra. Orange dotted lines are too long to generate peaks. After applying the constraints with a threshold distance of X=3.5 Å, we see that the absence of a peak in fragment (a) is the only unambiguous constraint.

Here we introduce a novel approach that extracts unambiguous conformational constraints on the single molecule conformations present in crystalline samples. The approach is schematically illustrated in **Figure 2-2**, where we differentiate between two conformers ("open" and "closed") by analysing a ^{1}H-^{13}C HETCOR spectrum.

The ^{1}H-^{13}C HETCOR spectrum contains two different types of information. First, cross-peaks that are present indicate atoms that are close in space. Second, absent cross-peaks contain information about atoms that are more than a certain distance "X" apart, where "X" possibly depends on the CP contact time, experimental setup and the investigated system. **Figure 2-2** shows that only the information from the absent cross-peaks in the solid-state spectra can be directly transferred to constraints on the single molecule conformations. This is demonstrated with a thought experiment. If the heteroatoms C_M and H_O are close in space, the cross-peak at C_M-H_O will be present in the HETCOR spectra. However, the cross-peak can result either from a short intra-molecular C_M-H_O distance (i.e. the "closed" conformer) (**Figure 2-2c**-**d**) or from a short inter-molecular interatomic distance (which can be from the "closed" or the "open" conformer) (**Figure 2-2b**-**c**). Thus, the presence of a cross peak does not contain unambiguous information about the single molecule conformer, as the fragments in **Figure 2-2b**-**d** contain both possible conformations.

An absent cross-peak for C_M-H_O however indicates that C_M and H_O are at least "X" angstroms apart, for both intra- and inter-molecular C_M-H_O distances (**Figure 2-2a**). This can only happen for the "open" conformer. Thus, information from the absent cross-peaks is unambiguous regarding the single molecule conformation and can be used as a constraint on trial structure generation.

Note that the fragment in **Figure 2-2b** also contains the "open" conformation, but would be expected to contain a cross-peak for C_M-H_O and thus will not result in a constraint on the distance between C_M and H_O. However, such cases only result in fewer constraints on the single molecule conformer but do not induce any incorrect constraints.

Note also that it is not *a priori* clear what the threshold distance "X" is. In general, we expect to reliably see all ^{1}H-^{13}C HETCOR cross-peaks at least up to 3.0 Å.[217] To establish a reliable value for the threshold distance "X", accessible in the ^{1}H-^{13}C HETCOR experiments used here, we investigate the correlation between interatomic ^{1}H-^{13}C distances and signal intensities of the cross-peaks in the HETCOR experiments recorded for cocaine, flutamide and flufenamic acid.

For these three compounds the experiments were performed at different contact-times, magic-angle spinning-rates and on different spectrometers. **Figure 2-3a** shows that for cocaine we have signal-to-noise ratios (SNR) of up to 80, while flufenamic acid has a maximum SNR of around 10. Additionally, for a ^{1}H-^{13}C HETCOR experiment, where the signal is

transferred from the ^{1}H to the ^{13}C, the SNR also depends on the number of protons involved in the transfer, as well as the number of protons overlapping at a given frequency.
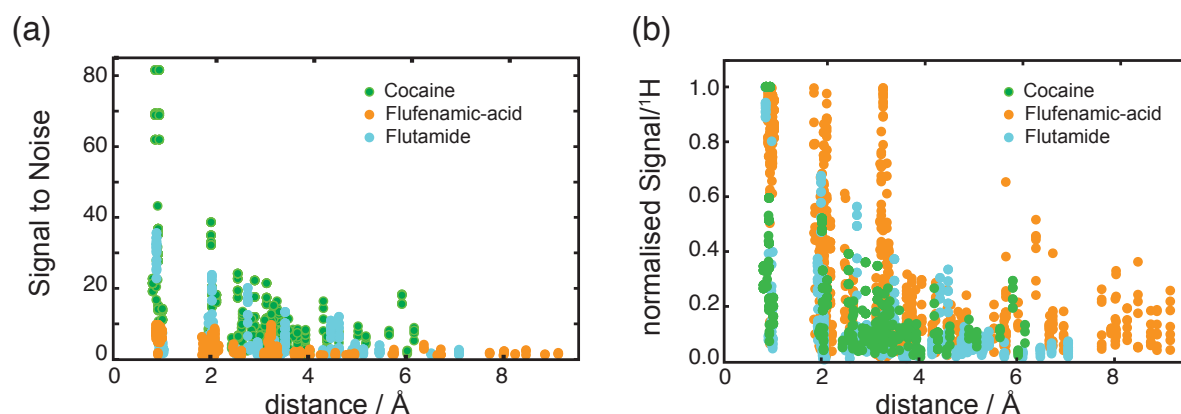
Figure 2-3 Signal intensity of ^{1}H-^{13}C HETCOR cross-peaks plotted against the corresponding interatomic distance for cocaine (green), flufenamic acid (orange) and flutamide (cyan). **(a)** The SNR is extracted directly for all ^{1}H-^{13}C HETCOR at different contact-times and different experimental setups. **(b)** The normalised SNR per ^{1}H allows a direct comparison across different experimental setups and for cross-peaks corresponding to a different amount of protons.

To make different spectra comparable, we first estimate that the number of active protons for a given cross-peak in a spectrum is proportional to the maximum signal intensity at a given frequency in ω_1. The signal intensity of each cross-peak is then re-normalised by this number of a protons. Then, we consider the difference in overall SNR between spectra by re-normalising each cross-peak with respect to the maximum proton-normalised SNR per spectra. This leads to a normalised SNR per ^{1}H, which is comparable across all experiments and which is shown in **Figure 2-3b.**

Once we have selected a reliable threshold distance X Å for a given SNR cut-off (this process is described below), the selected threshold distance in combination with each absent HETCOR cross peak is transformed into a constraint on the conformer space as, "*if the HETCOR cross peak between C_x and H_y is below the SNR cut-off it is classified as absent and so the distance between the atoms C_x and H_y must exceed X Å.*"

For each single molecule conformer all of the generated constraints were checked and the conformers were sorted according to the number of constraints violated. This procedure allows us to select conformers for the subsequent CSP procedure. If we are confident in the extracted constraints, it is sufficient to only select the sub-ensemble with the lowest number of constraint violations. However, if this sub-ensemble is very small or if additional computational resources are available, the selected sub-ensemble can easily be extended to include structures with a progressively higher number of constraint violations. Accepting conformations with a small number of constraint violations can allow for moderate changes in molecular geometry between the gas phase and crystal structure.

2.1.3 Results and discussion

We first establish the range of reliable threshold distances "X" for a given SNR cut off S_{norm}. For this we investigate the correlation between S_{norm} and the corresponding inter-atomic distances for the three trial compounds cocaine, flufenamic acid and flutamide. Then, we investigate the application of the parametrised constraints to the CSP-NMRX structure determination of these three compounds. Finally, we perform the full CSP-NMRX crystal structure determination including the unambiguous constraints on microcrystalline ampicillin where the parametrisation of the threshold distance ("X") and the normalised signal-to-noise cut-off value (S_{norm}) was done without using any prior knowledge regarding the crystal structure.

Parametrization using known structures. For cocaine, flufenamic acid and flutamide, ^{1}H-^{13}C HETCOR experiments were performed with ^{1}H-^{13}C contact times of 0.5, 0.75, 1.0 and 1.5 ms; 0.1, 0.5, 1.5, 2.0, 3.0 and 3.5 ms; and 0.1, 0.3, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75 and 2.0 ms respectively. We re-normalised the spectra as described above, (see **SI** for details). The resulting normalised SNR per ^{1}H is then comparable between compounds, see **Figure 2-3b**.

However, **Figure 2-3b** shows that although there is a correlation between the normalised SNR and the corresponding inter-atomic distance, there are significant fluctuations. This is expected since the HETCOR experiment is quite simple but is subject to spin diffusion relayed transfer, among others. We find that the effect of these fluctuations can be minimised by only considering correlations/distances from protons that are situated towards the extremities of the molecules. (We note here, that this currently results in a reduced number of extracted constraints. If the constraints could be extracted in a more quantitative manner, e.g., by accounting for changes in peak intensities due to ^{1}H-^{1}H spin diffusion, then the selection criteria could be made stronger). However, these distances are the most information-rich in terms of the overall molecular conformations. We thus only consider cross-peaks resulting from the "*terminal*" protons shown in **Table 2-5**, and marked with a green dotted-line in **Figure 2-4a**. This results in a clearer correlation between normalised SNR and the corresponding inter-atomic distances, as shown in **Figure 2-4b**.
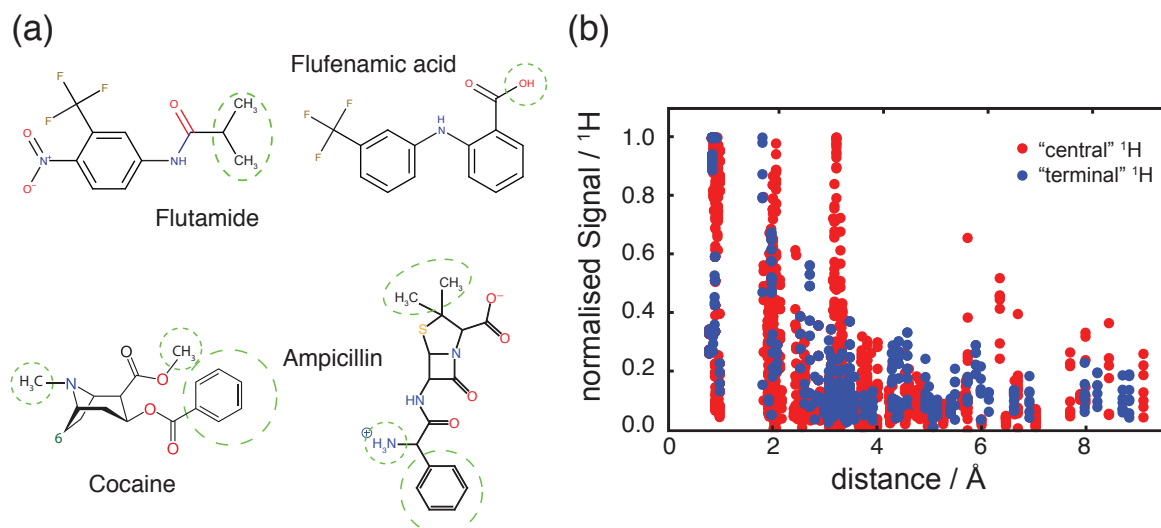
Figure 2-4 (a) Illustration of 'terminal' protons, which contribute to eventual conformational constraints. **(b)** normalised SNR of ^{1}H-^{13}C HETCOR cross-peaks plotted against the corresponding interatomic distance for 'centre' protons (red) and 'terminal' protons (blue), which are used to generate conformational constraints.

From **Figure 2-4a** it is clear that only a very limited number of inter-atomic distances below 3 Å result in a normalised *SNR* above 0.2. We then test a range of S_{norm} values from 0.08 to 0.22 with threshold distances "*X*" ranging from 2.0 to 5.0 Å. For this we used the single molecule conformer ensembles previously generated for the successful CSP-NMRX structure determination protocol described by Baias *et al.*[196] Our goal was to verify that the proposed parameterisation can select the gas-phase conformer that leads to the correct crystal structure while at the same time significantly reducing the total amount of conformers that have to be considered.

Figure 2-5a shows the set of parameters for which the selection procedure was successful for all three molecules simultaneously. **Figure 2-6** shows the set of successful parameters for each molecule individually. The dashed orange line in **Figure 2-5a** shows the limit at which the selection process starts to fail. To obtain maximal selection power, the parameters should be chosen as close as possible to this limit. For cocaine, flufenamic acid and flutamide the highest selectivity within the investigated conformer ensembles explored here was obtained using S_{norm} =0.14 and "*X*" = 3.5 Å.
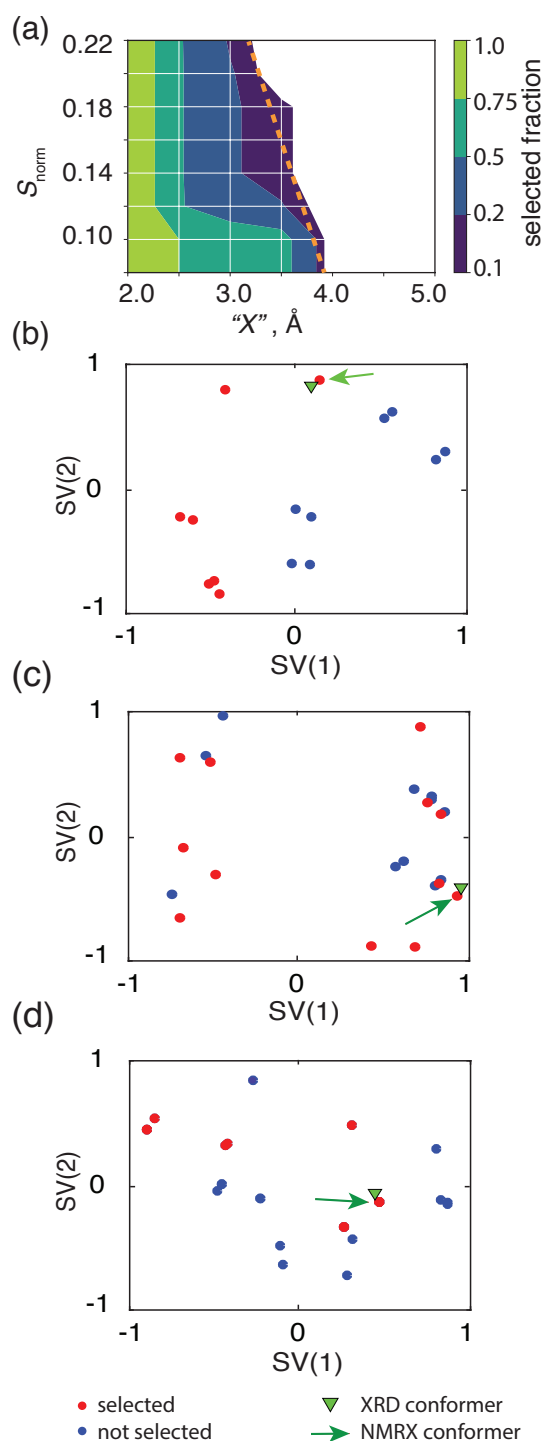
Figure 2-5 (a) Grid search results of the threshold distance "X" and S_{norm} cut-off values for flutamide, cocaine and flufenamic acid. The colour-map shows the percentage of selected structures from within the conformer ensemble. The white area indicates the region where the correct conformer is not selected. Optimal selection parameters should select the smallest conformer ensemble that still contains the correct structure. This corresponds to the dark blue regions within the different panels. The dashed orange line denotes the boundary at which the selection process starts to fail. (b-d) Conformer selection for flutamide (b), flufenamic acid (c) and cocaine (d). The panels show the sketch-map projections of the gas-phase ensembles. Red dots represent the structures that are selected where a threshold

distance of 3.5 Å and a S_{norm} cut-off value of 0.14 were used. The green triangle shows the conformer found in the XRD-generated crystal structure. The green arrow points to the gas-phase conformer which results in the correct crystal structure after the CSP procedure.
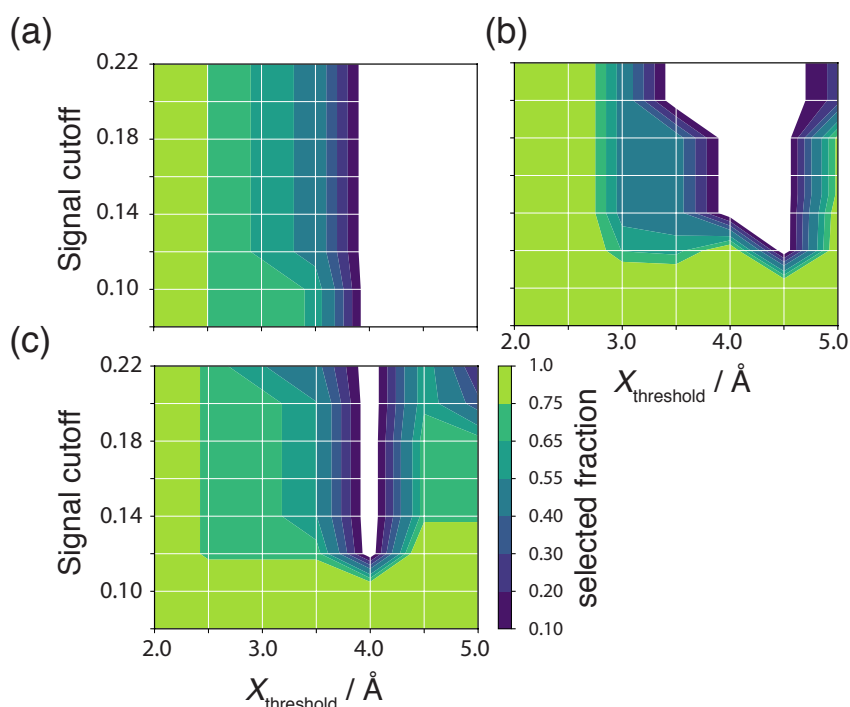
Figure 2-6 Grid search results of the threshold distance "X" and S_{norm} cut-off values for **(a)** flutamide, **(b)** cocaine and **(c)** flufenamic acid. The colour-map shows the percentage of selected structures from within the conformer ensemble. The white area indicates the region where the correct conformer was not selected. Optimal selection parameters should select the smallest conformer sub-ensemble, while still containing the correct structure. This corresponds to the dark blue regions within the different panels.

To aid our interpretation of the selection procedure we apply a sketch-map[218-221] analysis to the gas-phase conformer ensembles. The details of the sketch-map analysis including an interpretation of the underlying conformational changes for cocaine, flutamide and flufenamic acid are given in **Figure 2-7**, **Figure 2-8** and **Figure 2-9** .
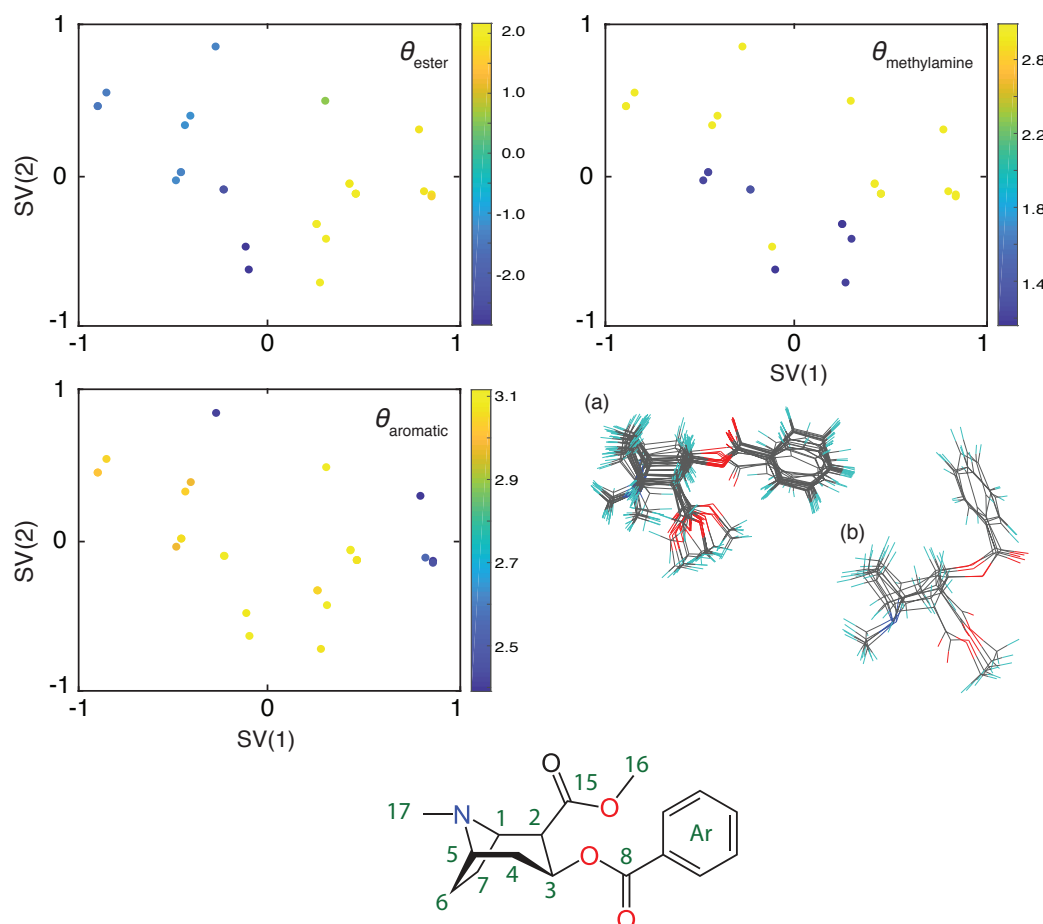
Figure 2-7 Top) Sketch-map representation of the locally stable cocaine conformations. To show the extent of the sub-clustering, the panels are coloured according to different torsion angles reporting on different torsion angle values in the molecule. θ_{ester} is defined as the C1-C2-C15-O4 torsion angle and reports on rotations of the ester group. $\theta_{methylamine}$ is defined as the C2-C1-N-C17 torsion angle and reports on rotations of the methyl group attached to the nitrogen. $\theta_{aromatic}$ is defined as the C(ortho)-C(ipso)-C8-O2 torsion angle and specifies the orientation of the phenyl ring plane relative to the nearby carboxylate group plane. The lower right panel shows the overlapped conformation with the aromatic ring roughly in the plane defined by the carboxylate group **(a)** and roughly perpendicular to the carboxylate group plane **(b)**. **Bottom)** 2D structure of cocaine with the labelling scheme used.
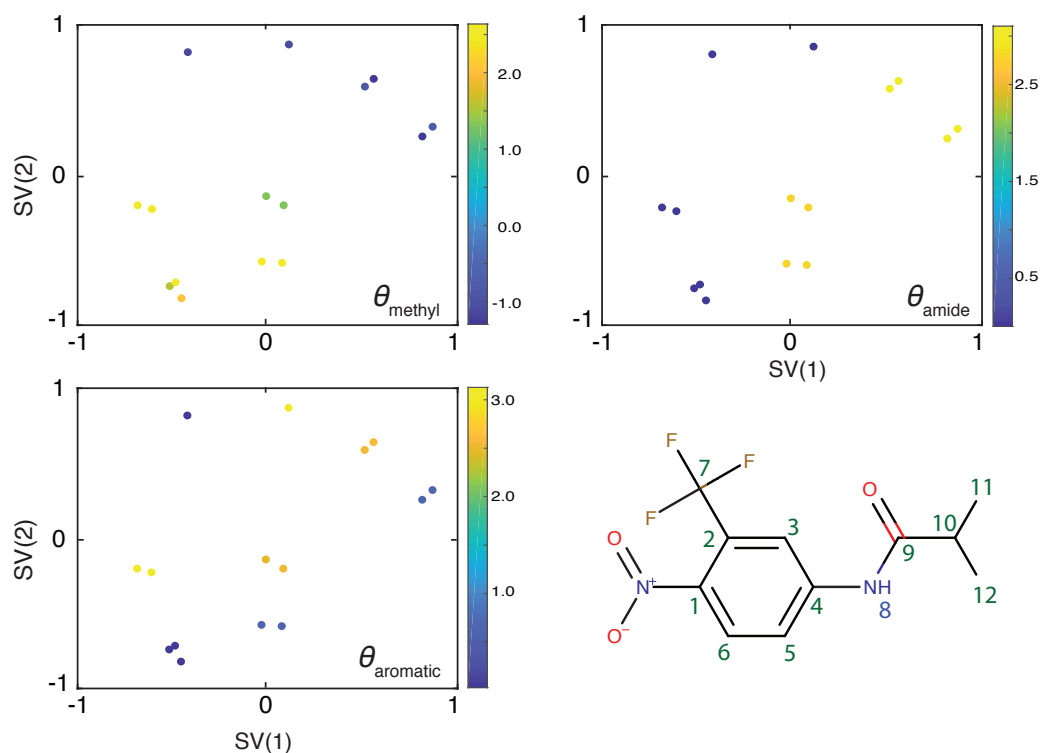
Figure 2-8 Sketch-map representation of the gas-phase flutamide conformer ensemble. To show the extent of the sub-clustering, the panels are coloured according to different torsion angles reporting on different torsion angle values in the molecule. θ_{methyl} is defined as the C11-C10-C9-N(H) torsion angle and reports on rotations of the methyl carbons. θ_{amide} is defined as the C4-N(H)-C9-O1 torsion angle and reports on the amide conformation. $\theta_{aromatic}$ is defined as the C3-C4-N(H)-C9 torsion angle and reports on rotations of the aromatic group. The lower right panel shows the 2D structure of flutamide with the labelling scheme used.
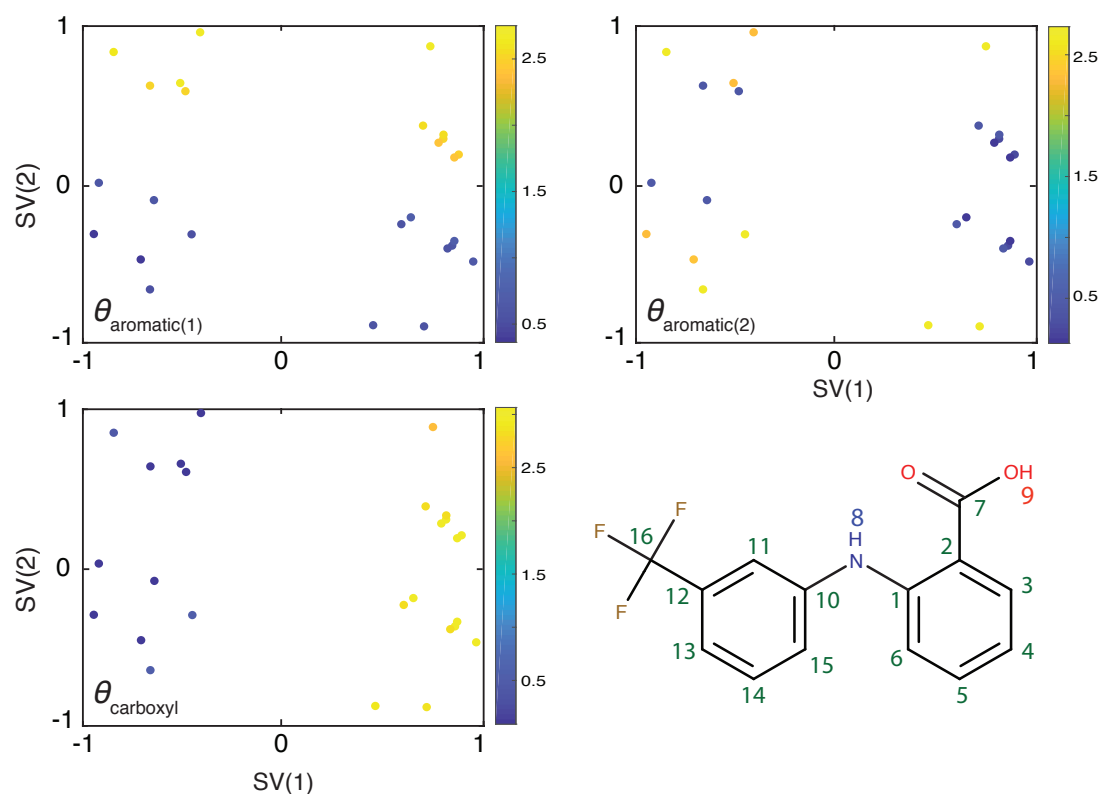
Figure 2-9 Sketch-map representation of the gas-phase flufenamic acid conformers. To show the extent of the sub-clustering, the panels are coloured according to different torsion angles reporting on different torsion angle values in the molecule. $\theta_{aromatic(1)}$ is defined as the C15-C10-N(H)-H(N) torsion angle and reports on rotations of aromatic ring with the attached trifluormethyl. $\theta_{aromatic(2)}$ is defined as the C7-C2-N(H)-H(N) torsion angle and reports on rotations of aromatic ring with the attached carboxyl. $\theta_{carboxyl}$ is defined as the C1-C2-C7-O(H) torsion angle and reports on rotations of the carboxyl group. The lower right panel shows the 2D structure of flufenamic acid with the labelling scheme used.

Flutamide. The initial gas-phase ensemble of flutamide conformers generated in the first step of CSP contains 15 conformers,[196] of which 7 are in the *trans* and 8 are in the *cis* conformation with respect to the amide group (**Figure 2-7**). All the absent cross-peaks in a series of ¹H-¹³C HETCOR spectra (**Figure 2-11a**) were used to generate the conformational constraints shown in **Figure 2-11a**. **Figure 2-5b** shows the selected sub-ensemble of gas-phase conformers in the sketch map that fulfil the most constraints. The sub-ensembles with the lowest number of violations (2 of 10 total constraints) are selected for the subsequent CSP procedure.

Note, that these two constraints are violated for all conformers and do not correspond to significant changes in the conformation, as the involved atoms are not separated by more than 2 bonds. The reduced ensemble contains the gas-phase conformer that led to the correct crystal structure during the subsequent CSP procedure,[196] while being able to reduce the gas-phase conformer ensemble from 15 to 7 conformations. This significantly reduces the computational cost of the following CSP steps by approximately 54% (assuming that all conformers lead to similar numbers of putative crystal structures), while still including the

correct gas-phase conformer that leads to the observed crystal structure. Additionally, the constraints from the absent cross-peaks uniformly selected the trans form in all 7 conformers (see **Figure 2-10**).
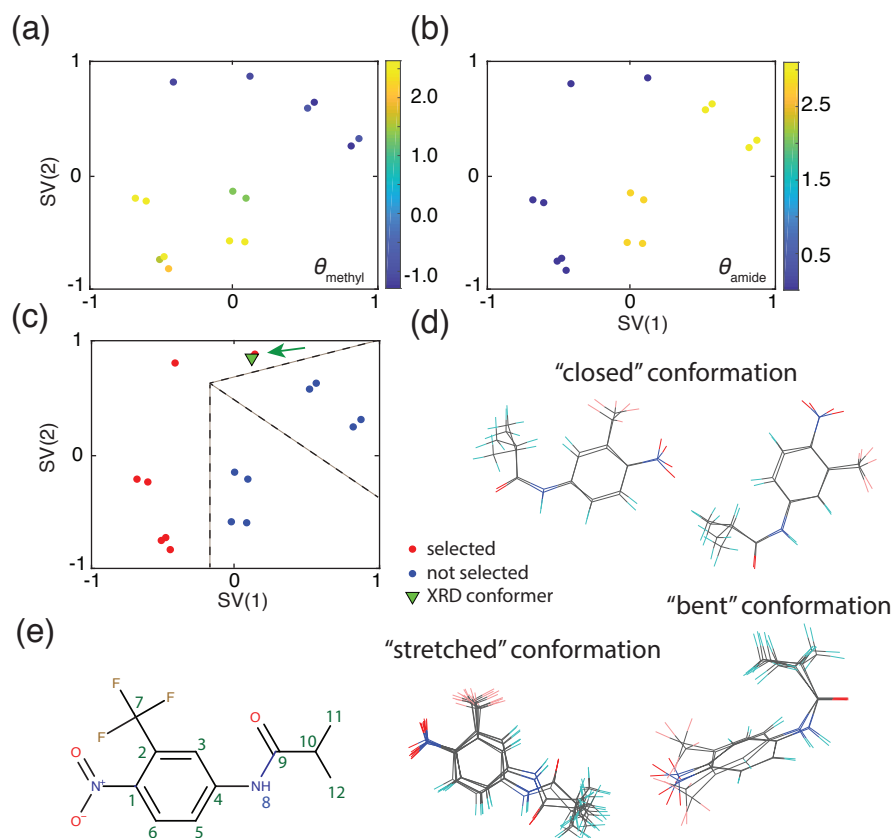

Figure 2-10 **(a-b)** Sketch-map representation of the gas-phase flutamide conformer ensemble. To show the extent of the sub-clustering, the panels are coloured according to different torsion angles reporting on different torsion angle values in the molecule. θ_{methyl} is defined as the C11-C10-C9-N(H) torsion angle and reports on rotations of the methyl groups. θ_{amide} is defined as the C4-N(H)-C9-O1 torsion angle and reports on the amide conformation. **(c)** Sketch-map projection of the gas-phase flutamide ensemble. Red dots represent the conformers with the lowest number of constraint violations, and are thus selected. The green triangle shows the conformer present in the XRD-determined crystal structure. The green arrow points to the gas-phase conformer, which resulted in the correct crystal structure after the CSP procedure. The black dashed lines indicate the regions where the different conformer sub-ensembles, shown in (d) are located. **(d)** Overlay of the structures within the different sketch-map clusters. The "stretched" conformations correspond to the trans conformers and are all selected. The "bent" and "closed" conformations correspond to the cis conformers and are not selected. **(e)** 2D structure of flutamide with the labelling scheme used.

Flufenamic acid. The gas-phase conformer ensemble for flufenamic acid contains 26 molecular conformations.[196] **Figure 2-5c** shows the selection of the sub-ensembles with the lowest number of violations (0 of 2 total constraints) using ^{1}H-^{13}C HETCOR. The extracted constraints are shown in **Figure 2-11c**. Note that, for flufenamic acid, there are only two non-

aromatic protons and that the cross-peaks from the aromatic protons are not distinguishable due to overlap in the ^{1}H dimension. However, the distance constraints extracted solely from the carboxyl proton (see **Figure 2-11c** and **Figure 2-9**) were sufficient to reduce the number of relevant conformers by 46% (from 26 to 14 conformers), while still selecting the correct conformer, leading to the observed crystal structure.
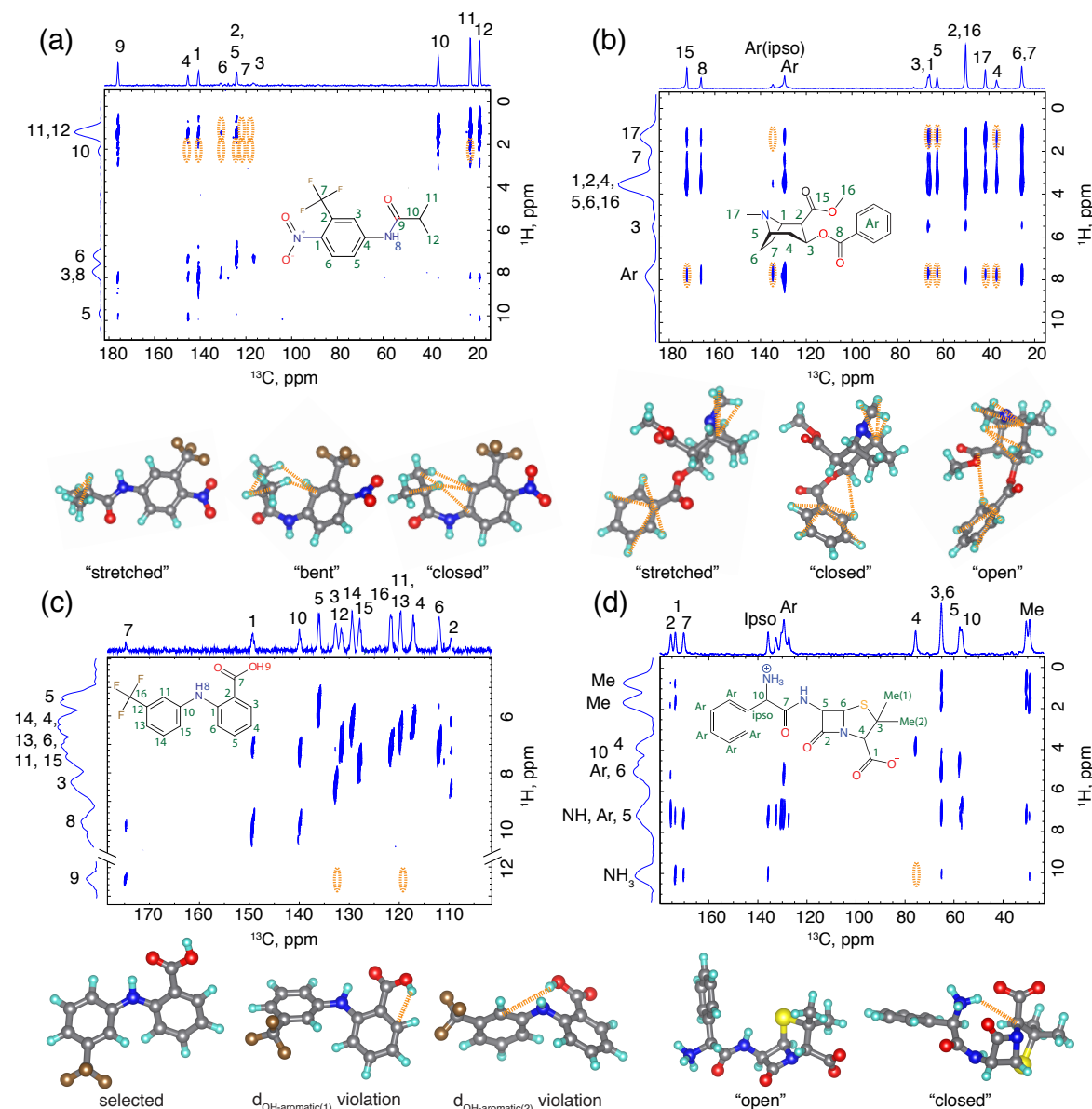
Figure 2-11 The top part in each panel shows the ^{1}H-^{13}C HETCOR spectrum of: flutamide with a 1.25 ms ^{1}H-^{13}C cross-polarization contact time **(a)**, cocaine with a 1.0 ms contact time **(b)**, flufenamic acid with a 1.5 ms contact time **(c)** and ampicillin with a 1.5 ms contact time **(d)** (further details and raw data in SI). ^{13}C peaks are assigned based on the literature[222] and ^{1}H peaks are assigned from HETCOR spectra and DFT chemical shift calculations (see SI). The cross-peaks from the terminal protons (**Figure 2-4a**) below a S_{norm} of 0.14 were used as constraints on the conformer ensembles, and are indicated as orange ellipsoids. The lower part of each panel shows the violated constraints extracted from all of the ^{1}H-^{13}C HETCOR cross-peaks for different example conformers within the ensembles.

Cocaine. The initial CSP conformer ensemble for cocaine contains 27 conformers.[196] **Figure 2-5d** shows the selection of the sub-ensembles with the lowest number of constraint violations (2 out of 10 total constraints) extracted from the ^{1}H-^{13}C HETCOR NMR spectra (**Figure 2-11b**). As with flutamide, these two constraints were violated for all conformers and do not correspond to significant changes in the conformation, as the involved atoms are separated by only 3 bonds. **Figure 2-7** shows that the ^{1}H-^{13}C HETCOR constraints were able to distinguish between the folding (closed and bent form) and stretching of the cocaine molecule with respect to the aromatic group as well as a flip in the methylamine group. Here, the relevant conformer ensemble is reduced by around 55% (from 27 to 12 conformations), while retaining the conformer that leads to the correct crystal structure.

Crystal structure determination of ampicillin. In contrast to the three cases above, the crystal structure determination of ampicillin would have failed using the usual CSP-NMRX protocol. In the first step, an ensemble of 16 locally stable gas-phase conformers is generated (for details, see SI) and the ensemble is then sorted according to the isolated molecule conformational energy. **Figure 2-12b** and **Figure 2-16** show that all the conformers within 25 kJ · mol^{-1} of the lowest energy structure are stabilized through an intra-molecular hydrogen bond between the amino nitrogen and oxygen atoms of the carboxyl group, whose strength is enhanced by the zwitterionic nature of the molecule. However, in the known single-crystal XRD structure, these intra-molecular hydrogen bonds between charged ends of the molecule are sacrificed to allow the formation of strong, charge-assisted inter-molecular hydrogen bonds, with the molecule adopting a more extended, open conformation.
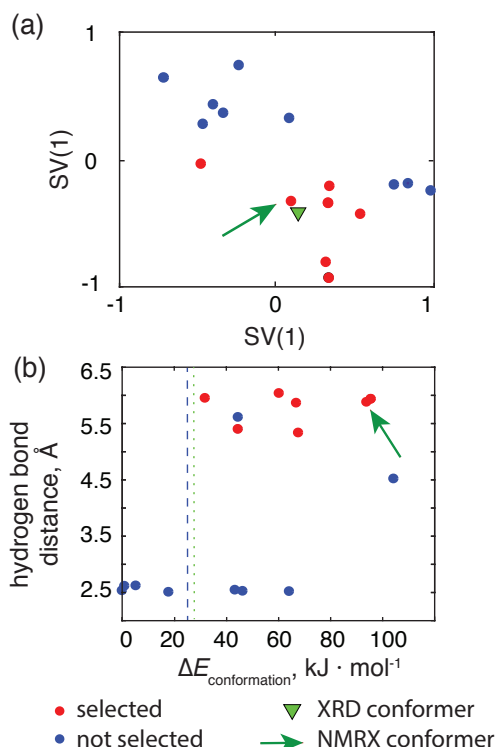
Figure 2-12 Conformer selection for ampicillin. (a) The panel shows the sketch-map projections of the gas-phase conformer ensemble. Red dots represent the structures which are selected using a threshold distance "X" of 3.5 Å and a S$_{norm}$ of 0.14. The green triangle denotes the conformer found in the XRD-determined crystal structure. The green arrow

points to the gas-phase conformer which results in the correct crystal structure after the CSP procedure. (b) Scatterplot showing the relative difference in the energy (ΔE) for the single molecule conformers of ampicillin against the shortest intra-molecular hydrogen-bond distance (N-O distance). The blue dashed line is the typical cut off energy (25 kJ/mol) used for conformer selection in CSP. The green dotted line is a guide to the eye to show at which ΔE the conformers with inter-molecular hydrogen bonds become accessible. The green arrow shows the conformer which results in the correct crystal structure.

For each conformer remaining within this reduced gas-phase ensemble, we generated a crystal structure ensemble using a quasi-random sampling[223] of lattice parameters, molecular positions and orientations within the commonly observed space groups. All 154,000 generated crystal structures were first optimized using an atomic-multipole based force field,[224] followed by dispersion corrected DFT-D re-optimization of the lowest energy crystal structures, producing a final set of 75 candidate crystal structures. The full procedure is detailed in the SI.

^{1}H chemical shift values were then calculated with GIPAW DFT and a machine learned method (ShiftML)[177] for each candidate structure and compared to the experimental chemical shifts (details are given in the SI). **Figure 2-13** shows the RMSE between DFT calculated and measured ^{1}H chemical shifts together with the calculated relative lattice energies for the candidate set. Based on currently-accepted metrics we expect a valid structure to have a ^{1}H RMSE of 0.33 ppm ($\pm$0.16 ppm) or lower.[225] This is indicated as the grey zone in **Figure 2-13**. Predicted structures with ^{1}H chemical shift difference within this zone are thus considered to be indistinguishable from experiment with a confidence of 1σ.
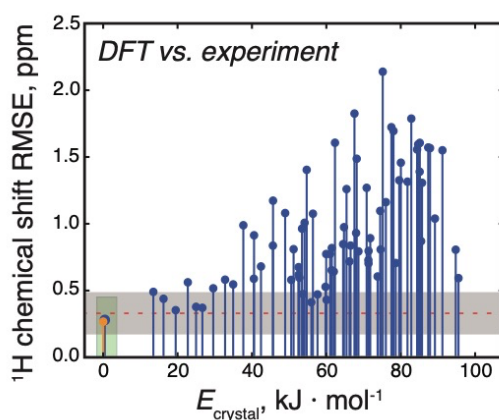
Figure 2-13 Comparison of crystal structure candidates. The structures are sorted according to their relative lattice energy, as specified on the horizontal axis. The vertical axis shows 1H chemical shift RMSE between DFT calculated and experimental chemical shifts. The orange marker shows the 1H chemical shift RMSE for the single-crystal XRD structure. The red line shows the mean of the current difference between experimental and DFT calculated 1H chemical shifts with the distinguishability limits (at the 1σ level) indicated as grey shaded zone, as described in the main text.

Figure 2-14 shows the RMSE between ShiftML calculated and measured ^{1}H chemical shifts together with the DFT calculated relative lattice energies for the candidate set. Using a benchmark set of 11 molecular crystal structures with around 150 experimental ^{1}H chemical

shifts we expect a correct structure to have a ^{1}H RMSE of 0.346 ppm ($\pm$0.195 ppm) or lower. Note that the RMSE between experiment and the predicted chemical shifts follows broadly similar relative to the DFT-calculated shifts (**Figure 2-13**).

Based on the agreement between experimental and calculated ^{1}H chemical shifts, both for ShiftML and DFT, we find that the crystal structure lowest in lattice energy, with a large gap in energy to the next predicted structure, also best produces the experimental NMR chemical shifts from the powdered microcrystalline sample used in the present study (**Figure 2-13, Figure 2-14**). Thus, we identify this structure as the correct candidate structure. Using chemical shifts calculated either directly from DFT or using ShiftML, several higher energy putative crystal structures produce ^{1}H chemical shifts within the acceptable error bounds. However, none of these alternative structures falls within the usual energy range of observed polymorphism (typically up to 7-8 kJ/mol)[226] above the best candidate structure. Thus, our final structure selection relies on both the chemical shifts and calculated lattice energies.
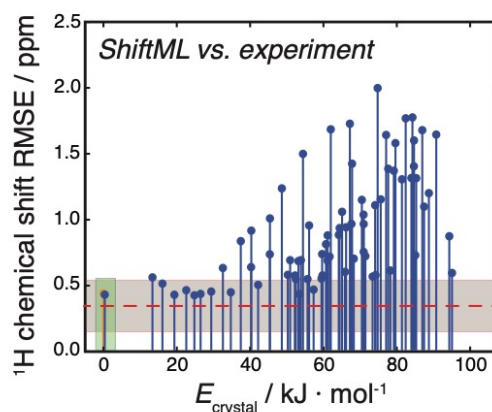
Figure 2-14 Comparison of crystal structure candidates. The structures are sorted according to their relative lattice energy, horizontal axis. The vertical axis shows ^{1}H chemical shift RMSE between ShiftML calculated and experimental chemical shifts. The orange marker shows the ^{1}H chemical shift RMSE for the single-crystal XRD structure. The red line shows the mean of the current error (0.346 ppm) between experimental and ShiftML calculated ^{1}H chemical shifts with the limits at one standard deviation (0.195 ppm) indicated as grey shaded zone, as described below.

The structure determined here agrees very well with the known reference structure determined by single-crystal XRD,[227] as illustrated in **Figure 2-15a.** The deviation in atomic positions in the NMR structure from the powder is 0.278 Å, measured as the RMSD of all heavy atoms (excluding protons) in a 20-molecule cluster taken from the two structures. The single-molecule heavy atom RMSD is 0.068 Å. The largest deviation in the lattice parameters is a contraction of 6.8% in the b lattice parameter, and a unit cell volume of the CSP-NMRX structure 7.4% smaller than the single crystal structure. This difference in volume is not unexpected as the NMRX structure is a temperature-free structure resulting from lattice energy minimization, while the single crystal structure was determined at room temperature. The slightly shorter lattice parameters in the NMRX structure are in line with the expected thermal expansion of an organic molecular crystal.
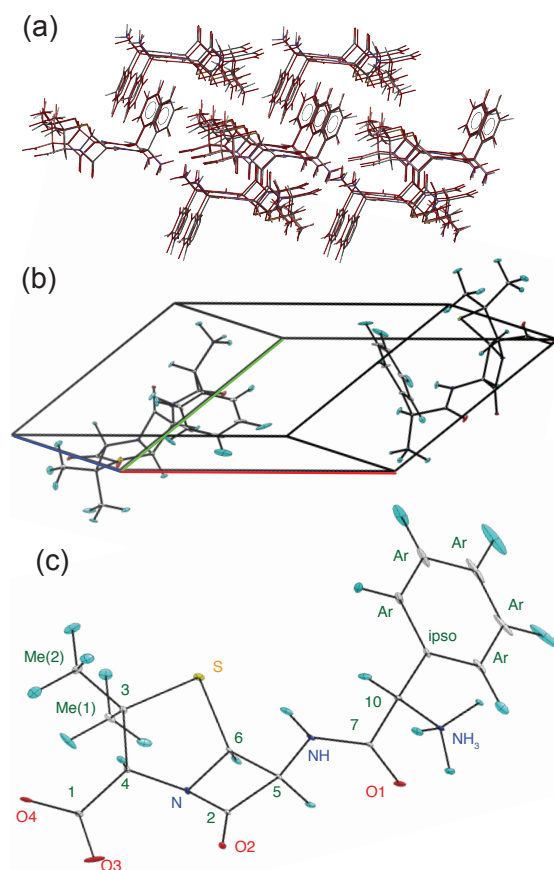
Figure 2-15 (a) Comparison between the structure of ampicillin as determined by the constrained powder ^{1}H CSP-NMRX and the single crystal XRD determined structure.[227] (b-c) ORTEP plot of the ampicillin crystal (b) and single molecule (c) structure drawn at the 90% probability level. The anisotropic ellipsoids correspond to a ^{1}H chemical shift RMSE of 0.49 ppm and to an average positional RMSE of $\langle r_{av} \rangle = 0.144$ Å.

Finally, we proceed with a positional error analysis that leads to the fully determined structure shown in **Figure 2-15c-d**. The positional error analysis is performed using the DFT-calculated ^{1}H chemical shifts following the procedure outlined by Hofstetter and Emsley[205] and is detailed in the SI (using DFT-MD here). The average positional RMSE on the NMR powder structure is $\langle r_{av} \rangle = 0.176$ Å, which corresponds to an average equivalent displacement parameter $U_{eq} = 0.0103$ Å^2. This compares with $\langle r_{av} \rangle = 0.149$ Å and $U_{eq} = 0.0074$ Å^2 for the single-crystal XRD structure.[227] Note that the positional RMSE on the single-crystal XRD structure only considers the heavy atoms, while the positional RMSE on the NMR powder structure also includes the ^{1}H atoms.

2.1.4 Conclusions

The most severe limitations of CSP based NMR crystallography are encountered when a molecule has many possible conformers and the molecular conformation adopted in the crystal could be significantly higher in energy than the most stable gas-phase conformation. In such cases, the usual energetic thresholds applied to the conformational ensemble used to generate candidate crystal structures create a risk of missing the true conformer as well as the crystal packing. Here we have demonstrated how the usual CSP-NMRX approach would

have failed for a powdered sample of ampicillin due to excluding the required conformer in the first step of CSP.

However, removing any conformer selection and including all possible conformers during crystal structure generation can lead to prohibitively high computational costs. To overcome this, we have proposed a modified CSP-NMRX method which includes unambiguous prior NMR constraints, in this case ^{1}H-^{13}C correlations, at the conformer search stage within CSP. The key development is a novel approach that extracts unambiguous conformational constraints on the single molecule conformations present in crystalline samples. We parametrised the proposed method on the crystal structure determinations of three flexible molecules that were previously studied using CSP-NMRX: cocaine, flutamide and flufenamic acid. For all of these compounds we found that the method reproduces CSP-NMRX results and determines the correct crystal structure, while reducing the computational cost by between 46 and 55%. Note that these three molecules are relatively small and the savings in computational expense will be greater for larger molecules with more conformational degrees of freedom.

We also demonstrated the capability of the novel constrained CSP-NMRX method by successfully determining the crystal structure of powdered ampicillin, which would have been very challenging for previous methods and either requiring that no energetic limit was applied to conformational energy, or likely missing the correct crystal structure. Here, a rough estimation shows that to run the CSP-NMRX calculations, including CSP search, DFT optimization and chemical shift calculations, for all 16 conformers would take approximately 54 days on 200 dedicated CPUs. By constraining the structural search space, we were able to more than halve this for the full crystal structure determination, while ensuring that the correct conformer is not excluded. We also emphasize that the large reduction in computational resources, demonstrated here, paves the way for the CSP-NMRX based determination of larger and more flexible molecules, which would previously have been out of the scope of the CSP-NMRX approach.

The compounds studied here were not subjected to any modification prior to the experiments, and they were investigated using powder samples at natural isotopic abundance. The resulting structures have a positional accuracy that is comparable to structures from, for example, single crystal XRD, while including the positions of the light atoms.

We note that the experimentally guided CSP method demonstrated here is not limited to pure NMRX applications but that the derived constraints can be used in any crystal structure determination methodology, which needs to limit the number of investigated conformations in order to reduce its computational cost.

We believe that the method is robust and we have chosen the experimental constraints, based on ^{1}H-^{13}C NMR correlation experiments, for their relative simplicity and ease of access. However, we note that ^{1}H-^{13}C correlation-based experiments are not the only ones that can give conformational constraints. Future work could incorporate other types of experiments such as ^{13}C-^{13}C correlations, or more accurate ^{1}H-^{13}C correlation experiments, which could be simpler to parameterize. Here the extraction of the constraints was performed in a fairly basic

and straightforward manner. We believe that if the constraints could be extracted in a more quantitative manner, e.g. by accounting for changes in peak intensities due to ^{1}H-^{1}H spin diffusion, the selection criteria can be made stronger, further reducing the conformational space and improving the computational efficiency and reliability of the methodology.

2.1.5 Supplementary

Samples

The powdered samples of anhydrous ampicillin ((2S,5R,6R)-6-([(2R)-2-amino-2-phenylacetyl]amino)-3,3-dimethyl-7-oxo-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid, purity > 98.0%) and free base cocaine (Methyl (1R,2R,3S,5S)-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylate, purity > 98.0%) were purchased from Sigma-Aldrich and Toronto Research Chemicals respectively, while the powdered samples of flutamide (2-Methyl-N-[4-nitro-3-(trifluoromethyl)phenyl]propenamide, purity > 98.0%) and flufenamic acid (2-((3-(Trifluoromethyl)phenyl)amino)benzoic acid, purity > 98.0%) were purchased from Tokyo Chemical Industry. All samples were used without further purification. For all compounds, the reference crystal structures were previously determined by single-crystal XRD. [227-230]

The reference crystal structure of ampicillin (CSD entry: AMCILL) is monoclinic, space group $P2_1$, with unit cell parameters a = 12.40 Å, b = 6.20 Å, c = 12 Å, and 2 molecules in the unit cell.

The reference structure of flutamide, (CSD entry: WEZCOT) contains 4 molecules in the unit cell, and it is orthorhombic, space group $Pna2_1$, with unit cell parameters a = 11.856(2) Å, b = 20.477(3) Å, c = 4.9590(9) Å.

The crystal structure of cocaine, (CSD entry: COCAIN10) contains 2 molecules in the unit cell, it is monoclinic, space group $P2_1$, with unit cell parameters a = 10.130(1) Å, b = 9.866(2) Å, c = 8.445(1) Å.

The flufenamic acid structure (CSD entry: FPAMCA11) is monoclinic, space group P21/c, with unit cell parameters a = 12.523(4) Å, b = 7.868(6) Å, c = 12.874(3) Å and 4 molecules in the unit cell.

Solid-state NMR experimental setup

In general, NMR experiments were performed at room temperature on a Bruker 500 wide-bore Avance III and a Bruker 900 US^2 wide-bore Avance Neo NMR spectrometers operating at Larmor frequencies of 500.43 and 900.13 MHz, equipped with H/X/Y 3.2 mm and H/C/N/D 1.3 mm probes.

The 2D ^{1}H-^{13}C dipolar heteronuclear correlation (HETCOR) experiments were performed using a 12.5 kHz MAS frequency for ampicillin, flutamide and cocaine and at 24.0 kHz MAS

frequency for flufenamic acid. In all experiments, SPINAL-64 was used for heteronuclear decoupling during t2 and eDUMBO-1$_{22}$ for homonuclear decoupling in the indirect dimension. 16 and 128 transients with 256 t1 increments were acquired for ampicillin, 64 transients and 256 t1 increments for flutamide, 4 transients with 64 t1 increments for flufenamic acid and 16 transients with 256 t1 increments for cocaine.

For the assignment of ampicillin, additional NMR experiments were required. These experiments were performed on either a standard-bore Bruker 700 Avance III or a wide-bore Bruker 500 Avance III operating at Larmor frequencies of 700.04 MHz and 500.16 MHz, respectively. Experiments on the 700 used a 3.2 mm HCN probe, while those on the 500 used a 4.0 mm HX probe. Recycle delays were between 1.0 and 1.3 s for all experiments outlined below.

The 11.7 T 2D ^{13}C-^{13}C refocused Incredible Natural Abundance Double Quantum Transfer Experiment (INADEQUATE) was performed using a 13.0 kHz MAS frequency at a temperature of 295 K. Prior to the indirect evolution period, cross-polarization (CP) from the ^{1}H nuclei was carried out (contact time of 2.5 ms). SPINAL-64 heteronuclear decoupling (100 kHz nutation frequency) was used during both evolution dimensions. 1760 transients with 128 t_1 increments were used. Each τ delay during the indirect dimension evolution was set to 3.84 ms, the length of the z-filter was 1.0 ms, and the recycle delay was 1.0 s.

The 16.4 T ^{1}H-^{15}N CP-HETCOR NMR experiment was carried out at T = 265 K using a 15 kHz MAS rotation frequency, while a ^{15}N magic-angle-turning (MAT) experiment was performed at T = 266 K and a 1.90 MAS rotation frequency. For the ^{1}H-^{15}N HETCOR experiment, SPINAL-64 heteronuclear decoupling was used during the t_2 dimension (83 kHz nutation frequency), and eDUMBO-1$_{22}$ was used for homonuclear decoupling in the indirect dimension (the scaling factor was set to 0.564). Prior to the indirect evolution period, CP from the ^{1}H nuclei was done (contact time = 300 μs). 1440 transients with 64 t_1 increments were used. For the ^{15}N MAT experiment, SPINAL-64 heteronuclear decoupling was used during both the t_1 and t_2 dimensions (100 kHz nutation frequency). Prior to the indirect evolution period, CP from the ^{1}H nuclei was done (contact time = 5.5 ms), with 1024 transients being acquired and averaged per t_1 increment, and with 125 t_1 increments being used.

The ^{1}H and ^{13}C chemical shifts were referenced indirectly to tetramethylsilane using the methyl signals of L-alanine at 1.3 ppm (^{1}H) and 20.5 ppm (^{13}C),[231] while ^{15}N chemical shifts were referenced using glycine at −347.54 ppm. ^{1}H chemical shifts were corrected for the scaling factor due to homonuclear decoupling, which was determined using ^{1}H 1D spectra acquired under fast spinning on a Bruker 900 spectrometer. Post-processing was done using Topspin 3.5 or 3.6.1.

Assignment of experimental NMR spectra

The assignment of ^{13}C and ^{1}H chemical shifts for flutamide, flufenamic acid and cocaine was taken from the paper by M. Baias *et al*[232] and verified by us.

The assignment of the ^{13}C spectra of ampicillin has been done by Clayden *et al.*[233] and then revised by Antzutkin *et al.*[222], but as the above authors mentioned, the assignment remains ambiguous, and so we revised it. To assign the ^{13}C NMR spectra at natural abundance a [13]C-^{13}C INADEQUATE experiment was done. To assign the ^{1}H directly attached to ^{13}C, the [1]H-^{13}C HETCOR spectra were used. To assign the ^{1}H directly attached to ^{15}N, a [1]H-^{15}N HETCOR experiment was done, which also helped for the assignment of ^{15}N resonances. To distinguish the ^{15}N chemical shifts belonging to NH and NH_3 resonances, a ^{15}N CP-MAT experiment was done, from which it was possible to tell that the NH_3 resonance corresponds to the peak with negligible chemical shift anisotropy due to the fast exchange of the three attached ^{1}H atoms. The assignment was cross-validated by comparing the experimental chemical shifts to shifts calculated with the GIPAW DFT method using the XRD crystal structure, albeit with optimized hydrogen positions.

Experimental chemical shifts

Label	[1]H, ppm	[13]C, ppm
1	3.5	66.0
2	3.5	50.2
3	5.5	66.7
4	3.3	36.7
5	3.4	62.6
6	3.4	25.6
7	2.4	25.6
8	-	165.9
Ar	7.8	129.4
Ar (ipso)	-	134.5
15	-	172.2
16	3.5	50.2
17	1.2	41.5

Table 2-1 Cocaine experimental chemical shifts.

Label	[1]H, ppm	[13]C, ppm
1	-	149.3
2	-	109.7
3	8.3	133.0
4	6.0	117.2
5	5.4	136.3
6	6.8	112.0
7	-	175.0
8	9.6	-
9	-6.6	-
10	-	139.9
11	6.9	121.7

12	-	131.7
13	6.2	119.8
14	5.9	129.5
15	7.3	128.1
16	-	124.1

Table 2-2 Flufenamic acid experimental chemical shifts.

Label	^{1}H, ppm	^{13}C, ppm
1	-	145.4
2	-	124.5
3	7.9	130.9
4	-	140.9
5	9.9	124.5
6	7.1	116.7
7	-	122.0
8	8.0	-
9	-	176.1
10	2.3	35.7
11	1.3	17.7
12	1.3	21.7

Table 2-3 Flutamide experimental chemical shifts.

Label	^{1}H, ppm	^{13}C, ppm	^{15}N, ppm
Me$_1$	0.6	30.1	-
Me$_2$	1.6	28.9	-
4	4.0	75.3	-
10	4.8	57.4	-
6	5.2	64.8	-
Ar	5.4	128.3	-
5	6.6	56.5	-
Ar	7.1	129.0	-
Ar	7.2	132.0	-
Ar	7.3	129.9	-
Ar	7.6	126.9	-
N	-	-	Around -210
NH	7.5	-	Around -270
NH$_3$	10	-	Around -340
3	-	64.8	
Ar(ipso)	-	135.4	
7	-	169.8	
1	-	173.2	
2	-	175.0	

Table 2-4 Ampicillin experimental chemical shifts.

Signal to Noise analysis

The signal to noise ratio (SNR) extraction and analysis were done using the Signals extracted directly from TopSpin 4.0.5 in text file format together with a home-written python script. The SNR was extracted as:

$$SNR = maxval(S)/(2*noise)$$

where maxval(S) is the maximum intensity at a given ^{1}H and ^{13}C chemical shifts coordinate ±0.2 ppm.- Note, that after a first extraction of maxval(S) the ^{1}H and ^{13}C coordinates were centered above maxval(S) and a refined maxval(S) was extracted.

The noise was extracted as the variance of the intensity for 100 areas $(0.4 \times 0.4\ ppm)$ within the spectra. The initial 10 noise-areas were chosen manually, as to not contain any cross-peaks. The subsequent 90 noise-areas were chosen at random and were included in the noise intensity if the maximum signal intensity within the random area was less-than or-equal to two times the maximum signal intensity in area previously selected. **Figure 2-3a** shows the extracted SNR of all ^{1}H-^{13}C HETCOR spectra for cocaine, flufenamic acid and flutamide against the corresponding inter-atomic distance.

First, we normalise each cross-peak by the number of protons. For this we estimate the number of protons for a given cross-peak in a spectrum by the maximum signal intensity at the given frequency, which is given from the maximum SNR at a given ^{1}H coordinate. In a next step, we take into account the difference in sensitivity between the spectra, due to the specific experimental setups, by normalising each cross-peak with respect to the maximal proton-normalised SNR per spectrum. This leads to a normalised SNR per ^{1}H, which is comparable across all experiments and is shown in **Figure 2-3b.**

Molecule	'terminal' ^{1}H
Cocaine	Ar
	16
	17
Flufenamic acid	9
Flutamide	10
	11
	12
Ampicillin	Ar
	NH$_3$
	Me(1)
	Me(2)

Table 2-5 Protons contributing to conformational constraints for cocaine, flufenamic acid, flutamide and ampicillin

Gas-phase conformer generation

For cocaine, flutamide and flufenamic acid, the CSP conformers and crystal structures were generated as described by M. Baias *et al*.[232]

For ampicillin, we generated as complete and unbiased a set of gas phase conformers as possible using a low-mode conformational search (LMCS) method,[234-235] as implemented in MacroModel.[236] Energies were calculated during the conformer search using the OPLS3 force field.[237] The only prior knowledge used was that the molecule was present in the zwitterionic configuration throughout the conformer search. Minimum and maximum move distances of 3 and 6 Å were applied and 12,000 search steps were performed (2,000 per flexible dihedral angle). Duplicate molecular geometries were identified and removed using an all-atom RMS deviation of atomic positions, with a 0.05 Å tolerance.

All conformers were re-optimized in Gaussian09 using dispersion-corrected density functional theory (DFT-D) at the B3LYP/6-311G** level of theory with the D3BJ dispersion correction.[238] The N-H bond lengths at the amino nitrogen atom were constrained to 1.035 Å to keep the molecule in its zwitterionic form. Without this constraint, a hydrogen atom transfers from the amino to the carboxyl group during DFT reoptimization of many of the conformers. Importantly, the resulting non-zwitterionic conformers are not relevant to the known polymorphs of ampicillin.

In analysing the conformers resulting from the search, we found that the configuration around chiral centres could be reversed during the LMCS search. Therefore, all possible diastereomers of ampicillin were found to be present in the results. All conformers of a different diastereomer to that of interest were removed from the conformational ensemble before CSP was undertaken.

Sketch-map analysis

Sketch-map is a dimensionality reduction algorithm that is well suited to explore the results of atomistic simulations due to the algorithm's non-linear nature. We can apply this algorithm to explore the similarity between the generated gas-phase conformers. For that we describe our systems by torsional degrees (high dimensionality representation) that are converted via sketch-map algorithm to collective variables (CVs) (low-dimensionality representation). Similar to principal components found by applying principal component analysis (PCA) CVs describe the similarity between the analysed data sets. While the underlying assumption of PCA is that the main differences between the data sets lie in a linear subspace of the full dimensionality space, sketch-map assumes that the subspace is non-linear. As CVs are derived from a combination of different structural features, they are not straightforward to interpret, but plotting them against each and against known structural features can give insight in structural similarity.

The cluster generation and analysis were performed with home-written Python and MATLAB codes and using the sketch-map package.[218-221] The sketch-map parameters are given in **Table 2-6**. They were chosen following the procedure described in Ceriotti *et al*.[219] and the tutorial

on sketchmap.org. The sketch-map analysis was not sensitive to small variations in the chosen parameters, as was already noted in the references.[219-221] As starting point for the sketch-map analysis we used all dihedral angles, not containing protons, over the full 2π range. This gives 55, 47, 31 and 35 dihedral angles for ampicillin, cocaine, flutamide and flufenamic acid, within a range of $-\pi$ to π.

Structure	$\Sigma = \sigma$	A	B	a	b
Cocaine	13	4	4	1	2
Ampicillin	6	2	2	1	1
Flutamide	6	3	3	1	1
Flufenamic Acid	6	2	2	1	1

Table 2-6 Sketch-map parameters for all compounds.

Ampicillin. The gas-phase CSP conformer ensemble of ampicillin contains 16 locally stable conformations (after DFT-D geometry optimization). The conformers are labelled according to increasing force-field energy. Conformer **14** is the most similar to the conformer in the crystal and resulted in the correct crystal structure after the remaining CSP procedure. **Figure 2-16** shows the sketch-map analysis of the ampicillin gas-phase ensemble.
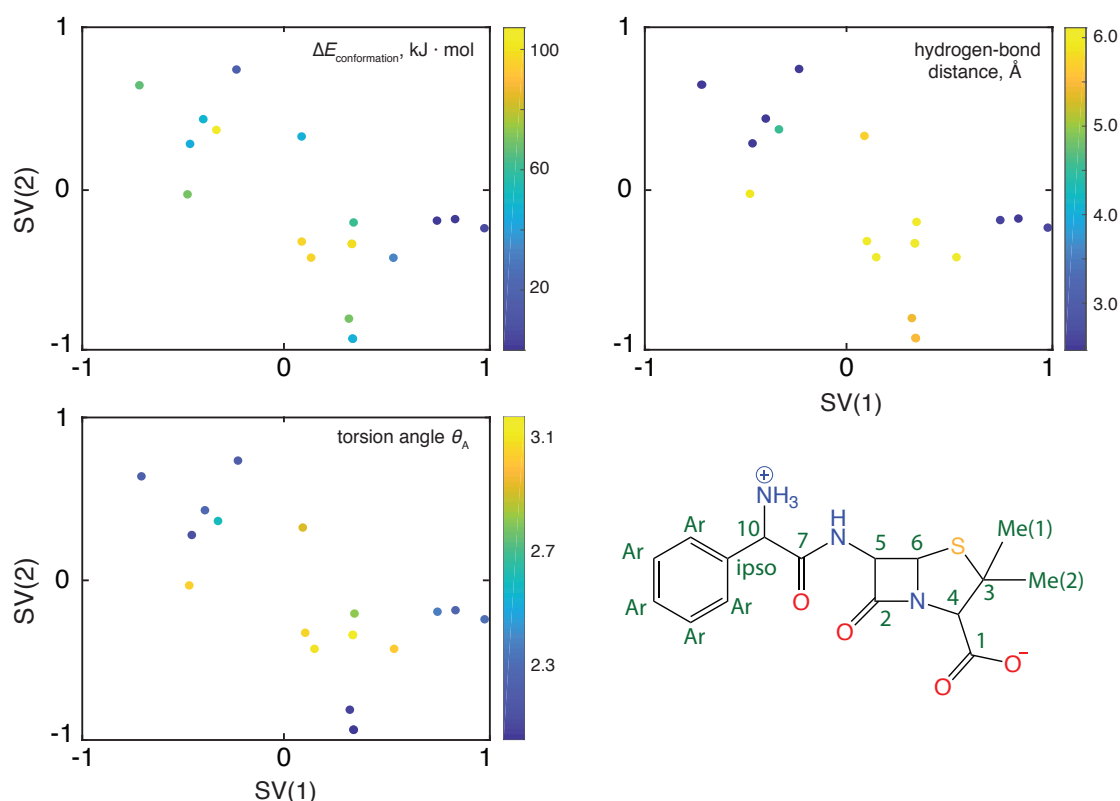
Figure 2-16 Sketch-map representation of the locally stable ampicillin conformers and the conformer from the single crystal XRD structure. To show the extent of the sub-clustering the panels are coloured according to different molecular properties. **Top left** shows the difference in conformational energy ($\Delta E_{conformation}$). **Top right** shows the shortest intra-molecular

hydrogen-bond distance between either NH_3 or NH and the carboxyl group. **Bottom left** shows the torsion angle θ_A. θ_A is defined as the torsion angle between C_{10}-C_7-$N(H)$-$(N)H$. In general, the clustering seems to correspond to conformational changes along the C_{ipso}-C_{10}-C_7-$N(H)$-C_5 chain and to relative changes between the methyl and carboxyl groups. **Bottom right,** shows the 2D structure of ampicillin with the labelling scheme used

Cocaine. The gas-phase CSP conformer ensemble of cocaine contains 27 locally stable conformations (after DFT-D geometry optimization). The conformers are labeled according to increasing force-field energy. Conformer **2** resulted in the correct crystal structure after the remaining CSP procedure.[232] **Figure 2-7** shows the sketch-map representation of the locally stable cocaine conformers. The main changes along the sketch-map principle components are rotations of the ester group (along SV(1)) and rotations within the methylamine group (along SV(2)).

Flutamide. The gas-phase CSP conformer ensemble of flutamide contains 15 locally stable conformations (after DFT-D geometry optimization). Of those, 7 are in the trans and 8 in the cis conformation with respect to the amide group. The conformers are labeled according to increasing force-field energy. Conformer **1** resulted in the correct crystal structure after the remaining CSP procedure.[232] **Figure 2-8** shows the sketch-map representation of the locally stable flutamide conformers. The sketch-map representation shows a relatively distinct clustering along the sketch-map axes, which correspond to the cis and trans conformations and rotations of the methyl groups. The SV(2) axis also partially corresponds to rotations of the aromatic ring.

Flufenamic acid. The initial CSP conformer ensemble of flufenamic acid contains 26 locally stable conformations (after DFT-D geometry optimization). The conformer **3** resulted in the correct crystal structure after the remaining CSP procedure.[232] **Figure 2-9** shows the sketch-map representation of the flutamide gas-phase conformer ensemble. The main changes along the sketch-map principle components correspond to rotations of the carboxyl group (along SV(1)) and rotations of the two aromatic groups (along SV(2)).

Parametrization of the constraints

It is not a priori clear as to what the threshold distance "X" should but in general we expect ^{1}H-^{13}C HETCOR cross-peaks in solid-state NMR spectra to appear for all interatomic distances of up to 3.5 Å. Here, we investigate the use of threshold distances ("X") from 2.0 to 5.0 Å in steps of 0.5 Å and for S_{norm} cut-off values from 0.08 to 0.22 in steps of 0.02 for the polymorphs of cocaine, flutamide, flufenamic acid. Figure 2-6 shows the set of successful parameters for each molecule individually.

Conformer selection

The ensemble selection was done with home-written Python codes. The HETCOR cross peaks below a S_{norm} value of 0.14 were interpreted as hydrogen-carbon distances greater than 3.5 Å (the "threshold" distance, "X"). For each conformation, the number of fulfilled constraints was counted and the conformers were sorted in decreasing order.

Flutamide. Conformer selection for flutamide was done based on constraints from multiple HETCOR NMR experiments, with variable contact times of 0.1, 0.3, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75 and 2.0 ms. The ^{1}H and ^{13}C cross-peaks from the two methyl groups could not be distinguished. Also, the ^{1}H cross peaks from H3 and H8 as well as the ^{13}C cross peaks from C5 and C2 are too close and thus indistinguishable. Therefore, if a cross-peak was observed it was attributed to all atoms in the given group.

Cocaine. The ^{1}H-^{13}C HETCOR NMR experiments were performed on cocaine at variable contact times of 0.5, 0.75, 1.0 and 1.5 ms. The ^{1}H and ^{13}C cross-peaks from the aromatic group could not be distinguished. Also, the ^{13}C cross-peaks from C6 and C7, the ^{13}C cross-peaks from C2 and C16, as well as the ^{1}H cross-peaks from H1, H2, H4, H5, and H6 were too close and hence indistinguishable. Therefore, if a cross-peak was observed it was attributed to all atoms in the given group.
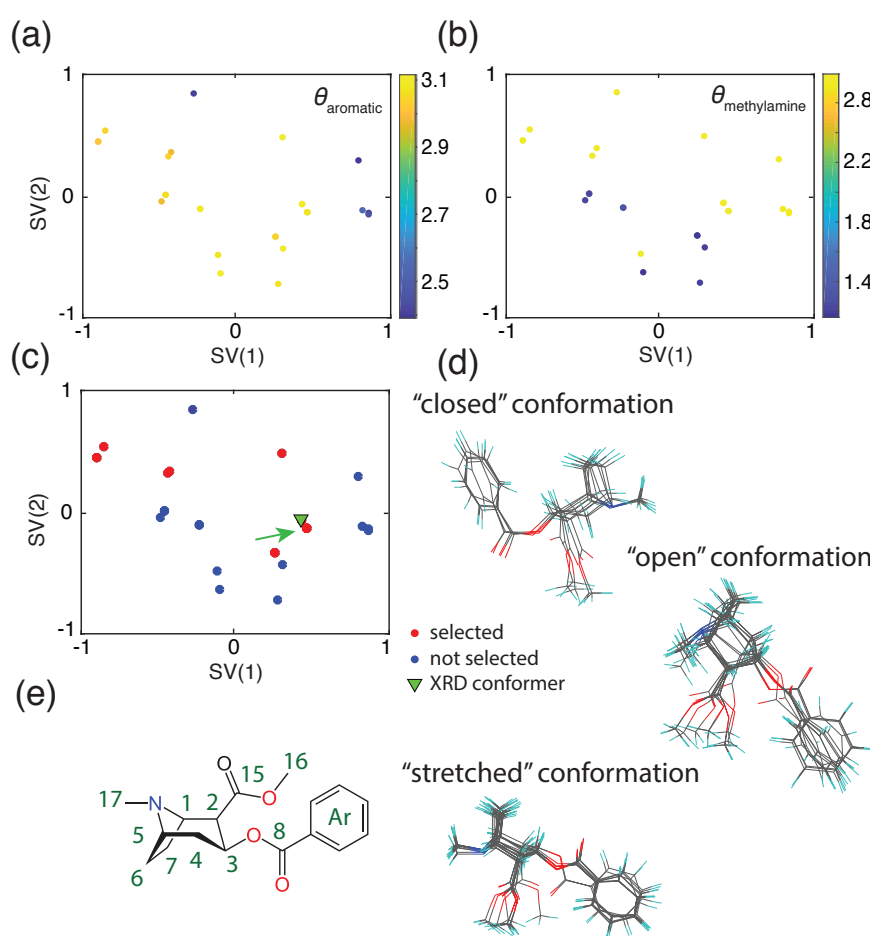
Figure 2-17 (a-b) Sketch-map representation of locally stable cocaine conformers. To show the extent of the sub-clustering, the panels are coloured according to different torsion angles reporting on different torsion angle values in the molecule. $\theta_{methylamine}$ is defined as the C2-C1-N-C17 torsion angle and reports on rotations of the methyl group attached to the nitrogen. $\theta_{aromatic}$ is defined as the C(ortho)-C(ipso)-C8-O2 torsion angle and reports on flips of the aromatic group. **(c)** Sketch-map projection of the gas-phase cocaine ensemble. Red dots represent the structures with the lowest number of constraint violations, and are thus selected. The green triangle shows the conformer present in the XRD-determined crystal

structure. The green arrow points to the gas-phase conformer, which resulted in the correct crystal structure after the CSP procedure. **(d)** Overlay of the structures within the different sketch-map clusters. The "stretched" conformations correspond to the selected conformers. The "closed" conformations contain different $\theta_{aromatic}$ torsional angle and are not selected. The "open" conformation contain a different $\theta_{methylamine}$ torsional angle and are not selected. **(e)** 2D structure of cocaine with the labelling scheme used.

Flufenamic Acid. Conformer selection for flufenamic acid was done based on constraints from multiple HETCOR NMR experiments, with variable contact times of 0.1, 0.5, 1.0, 1.5, 3.0 and 3.5 ms. The ^{1}H cross-peaks from H4, H13 and H14 as well as the ^{1}H cross-peaks from H6, H11 and H15 are too close and thus indistinguishable. Therefore, if a cross-peak was seen it was attributed to all atoms in the given group.
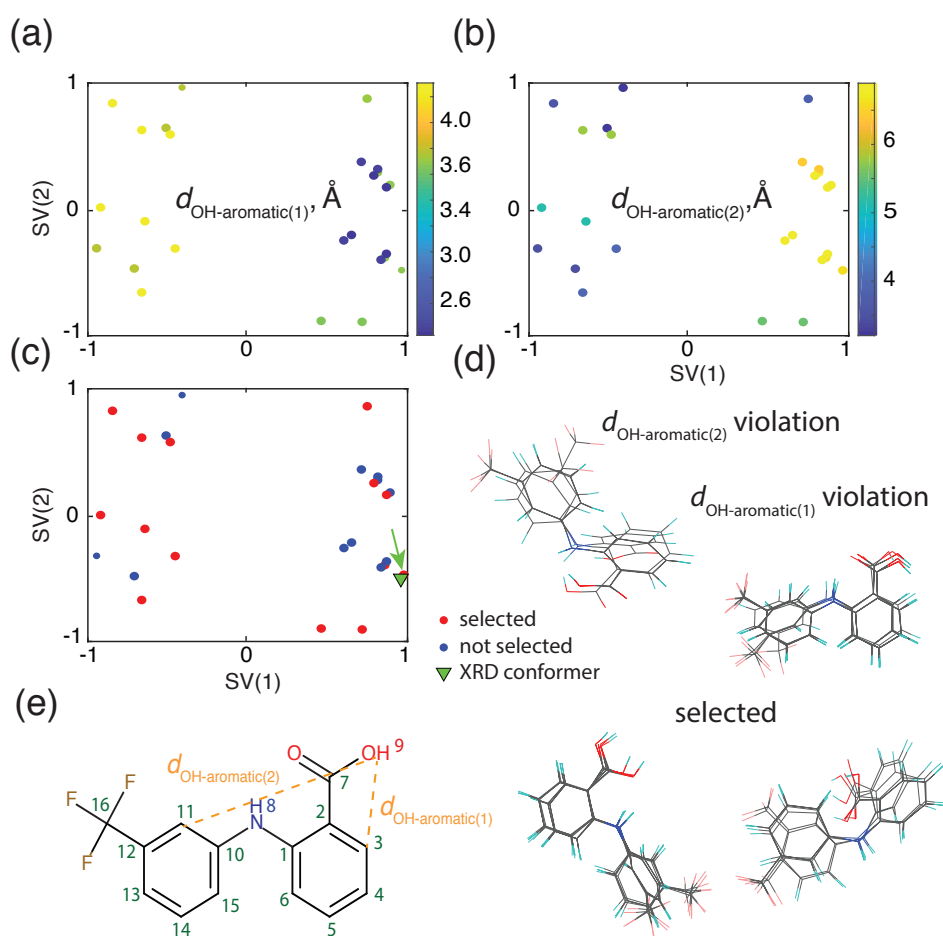
Figure 2-18 (a-b) Sketch-map representation of the gas-phase flufenamic acid conformer ensemble. To show the extent of the sub-clustering, the panels are coloured according to the distance [Å] between the OH group and the two aromatic rings. The distance is expressed as the distance between the carboxyl proton and C3/C11 (as shown in **e**). **(c)** Sketch-map projection of the gas-phase flufenamic acid conformer ensemble. Red dots represent the conformers with the lowest number of constraint violations, and are thus selected. The green triangle shows the conformer present in the XRD-determined crystal structure. The green arrow points to the gas-phase conformer, which resulted in the correct crystal structure after

the CSP procedure. **(d)** Overlay of the structures within the different sketch-map clusters. **(e)** 2D structure of flufenamic acid with the used labelling scheme.

Ampicillin. The conformer selection for ampicillin was done using constraints from ^1H-^{13}C HETCOR NMR experiments with variable contact times of 0.1, 0.3, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75 and 2.25 ms. The ^1H and ^{13}C cross-peaks from the two methyl groups could not be distinguished. Also, the ^1H cross-peaks from Ar2-6, H5 and NH, the ^1H cross-peaks from Ar1, H10 and H6, the ^{13}C cross-peaks from C3 and C6 as well as the ^{13}C cross-peaks from Ar1-5 were too close and hence indistinguishable. Therefore, if a cross-peak was observed it was attributed to all atoms in the given group.
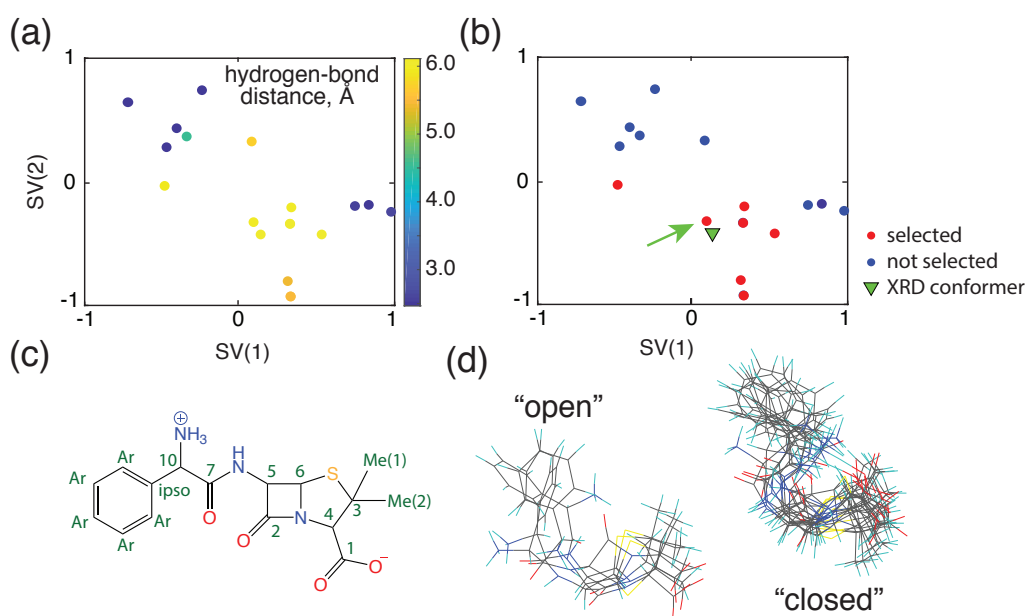
Figure 2-19 (a) Sketch-map representation of the locally stable ampicillin conformations. To show the extent of the sub-clustering, the panel is coloured according to the shortest intra-molecular hydrogen-bond distance [Å] between either NH_3 or NH and the carboxyl group. **(b)** Sketch-map projection of the gas-phase ampicillin ensemble. Red dots represent structures with the lowest number of constraint violations, and are therefore selected. The green triangle shows the conformer found in the XRD-determined crystal structure. The green arrow points to the gas-phase conformer that resulted in the correct crystal structure after the CSP procedure. **(c)** 2D structure of ampicillin with the labelling scheme used. **(d)** Overlay of the structures within the different sketch-map clusters. The "open" conformations correspond conformers without an intra-molecular hydrogen bond and were selected. The "closed" conformations mostly contain an intra-molecular hydrogen bond and were not selected.

2.2 De Novo crystal structure determination using NMR shifts

This chapter has been adapted from Martins Balodis, Manuel Cordova, Albert Hofstetter, Graeme M. Day, and Lyndon Emsley, De novo crystal structure determination from machine learned chemical shifts. *Submitted*.

My contributions to the publication. Performing simulations, writing the principal code, developing the methodology, investigating the data and being the main author of writing and editing the paper.

2.2.1 Introduction

In order to further simplify structure determination of molecular solids, here we show how by using a recently introduced machine learning model to predict chemical shifts, the structure of powdered organic solids can be determined in a manner fully analogous to the methods used in solution NMR or X-ray diffraction, and which completely removes the need for the prior CSP step.

We do this by integrating on-the-fly solid-state NMR shift calculations into a Monte Carlo simulated annealing optimization protocol. The approach is demonstrated to successfully determine the crystal structures of cocaine (**1**) and two different polymorphs of the drug molecule AZD8329 (**2**) (**Figure 2-20**).
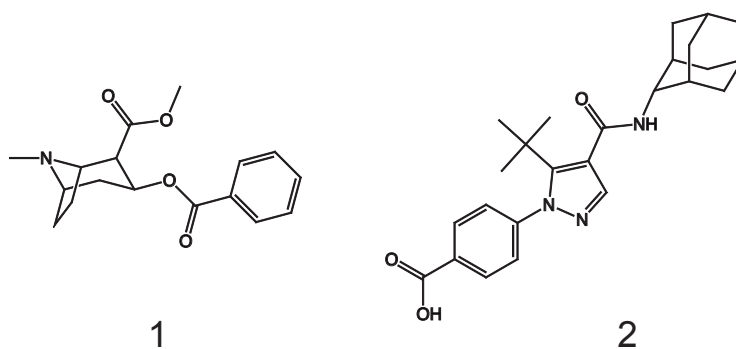
Figure 2-20 Molecular structures of cocaine (**1**) and AZD8329 (**2**).

2.2.2 Methods

Crystal structure determination.
The crystal generation and optimization were done by using a home-written Python script. The structure determination process followed the scheme provided in the **Figure 2-21**, an initial conformer was generated with random torsional angles. The generated conformer was placed in a unit cell in a randomly chosen position and orientation, with random cell lengths and angles. The rest of the details of the structure generation are given in the supplementary.

After the initial generation of a random crystal 4000 Monte Carlo loops were run where the temperature was linearly varied during from 2500 K to 50 K where in each of the steps. In each step one of the crystal structure defining parameters were changed by a random number from 0 to max step size. If the change was accepted the maximum step size was divided by a step factor α and if not, then multiplied by α, where α = 2.0. See SI for the step sizes. Every 500 loops hydrogen positions were optimized with DFTB method. In each step the system energy and chemical shifts were calculated.

The energy calculations were done using DFTB+ package that was implemented in the code.[239] The chemical shieldings were predicted using ShiftML version 1.2. (publicly available at https://shiftml.epfl.ch).[130] They were converted to chemical shifts via function:

$$\delta = a + \sigma b \qquad\qquad 2.1$$

where δ is the chemical shift, a and b are the experimentally determined calibration constants (see SI for details) and σ is the calculated chemical shielding. To account for ambiguity when comparing chemical shifts of protons for CH_2 groups, shifts were compared using the best matching criteria. Shifts which are hard or impossible to distinguish experimentally such as aromatic protons or CH_3 groups were averaged when making the comparison.

Crystal structure comparison.

The optimized structure comparison was done by the COMPACK algorithm[240] included in the commercial CSD package.[173] A cluster of 20 molecules was used for the comparison.

2.2.3 Results and Discussion

The optimization scheme introduced here is summarized in **Figure 2-21**.
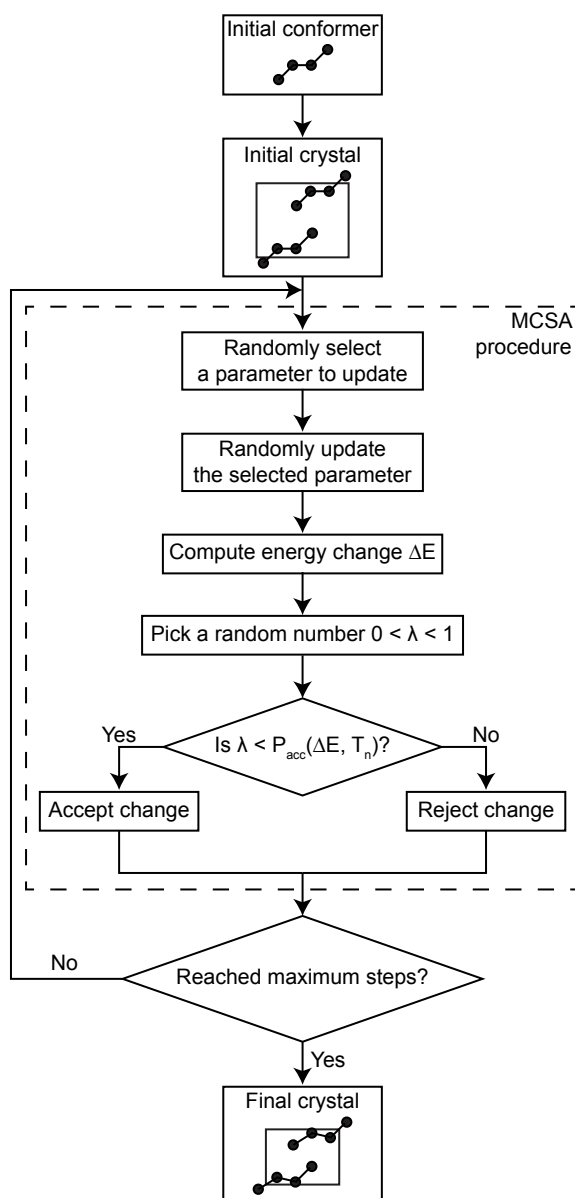
Figure 2-21 The scheme for crystal structure determination used in this study where $P_{acc} = e^{\frac{\Delta E}{RT}}$.

In the first step, a viable conformation of the single molecule is generated, and bond-angles and lengths are optimized using, here, DFTB3-D3H5 which provides a good compromise between accuracy and computational cost (on the same timescale as ShiftML chemical shift calculations) (see SI for details). Then, for each run, a random conformer is generated by randomizing the flexible torsion angles, and a starting crystal structure was generated by randomly selecting cell parameters in a given space group (cell lengths, cell angles, position and orientation of the molecule). This was done for 2500 trial structures for **1** and 10000 trial structures in the case for each of **2** polymorphs. Each structure was then optimized by a Monte Carlo simulated annealing process where in each step one of the parameters defining the crystal structure were randomly changed and chemical shifts and the DFTB system energy were calculated following the change.

Here, to enable the possibility to calculate shifts at each step, the ShiftML prediction algorithm was used.[130] ShiftML is a fast and accurate method to compute chemical shifts in a matter of seconds even for the largest of molecular crystals. It was recently (2018) developed in EPFL using a machine learning framework and was trained on DFT optimized structures derived from Cambridge Structural Database (CSD). The current version can predict ^{1}H, ^{13}C, ^{15}N, ^{17}O and ^{32}S chemical shifts.

The cost function used in the Monte Carlo process is:

$$E_{tot} = E_{DFTB} + E_{cs} \qquad\qquad 2.2$$

where

$$E_{cs} = c \left(\sqrt{\frac{\sum_{i=1}^{n}\left(\delta_{i,expt}-\delta_{i,shiftML}\right)^2}{n}} \right) \qquad\qquad 2.3$$

where $\delta_{i,expt}$ is the experimentally measured chemical shift of the ith nucleus in the molecule containing n nuclei and $\delta_{i,shiftML}$ is the corresponding shift computed using the ShiftML model. c is an empirically adjusted constant (in kJ/mol) that weights the relative contribution of the internal energy and the agreement with experiment in the cost function. (Note that the values of E_{cs} are independent of the size of the molecule, but will change from one type of nucleus to another, and that E_{DFTB} will depend on the size of the molecule. In the examples here, satisfactory results were found with vales of c such that $\Delta E_{DFTB} \sim \Delta E_{cs}$, where ΔE is the difference observed between two Monte Carlo steps at the end of the optimization process.) The other parameters in the simulated annealing process are given in the methods section and SI.

Figure 2-22 shows the results for AZD8329 Form I, AZD8329 Form IV and cocaine. For each case, the optimization was done including both DFTB energy and chemical shift differences in the penalty function, and then it was repeated using only DFTB energy.

For all the compounds we note that the majority of Monte-Carlo runs do not yield any results with either low DFTB energy or with a low chemical shift RMSD to experiment. Indeed, if we define a region of acceptable structures to have simultaneously a DFTB energy within 20 kJ/mol of the lowest energy structure in the Monte Carlo set and a chemical shift RMSD to experiment below 0.5 ppm, then the pure Monte Carlo approach using only DFTB energy as the driving force does not find any structures for either form of AZD8329.

Including chemical shifts in the penalty function yields three structures for Form IV within the acceptable ranges, and one structure for Form I.

Two molecules of the both Forms are shown in **Figure 2-23**, superimposed on the known structures for these two Forms, and we see that they are in excellent agreement with the correct structure.

Cocaine is an interesting example, since it is significantly less flexible than AZD8329. In this case, the Monte Carlo approach with only energy already produced four molecules in the acceptable region and adding chemical shift term produced the same number of molecules while bringing them closer in the energy-shift space. Two molecules of these structures are shown in **Figure 2-23** superimposed on the known structure for cocaine, and we again see that they are in excellent agreement with the correct structure

We explain these observations as cocaine having a relatively simple energy landscape, with only one known stable polymorph. AZD8329 on the other hand has a much richer energy landscape with at least 4 anhydrous polymorphs known.[119] Nevertheless, by using the chemical shifts of the two different Forms, we were able to successfully determine both structures.
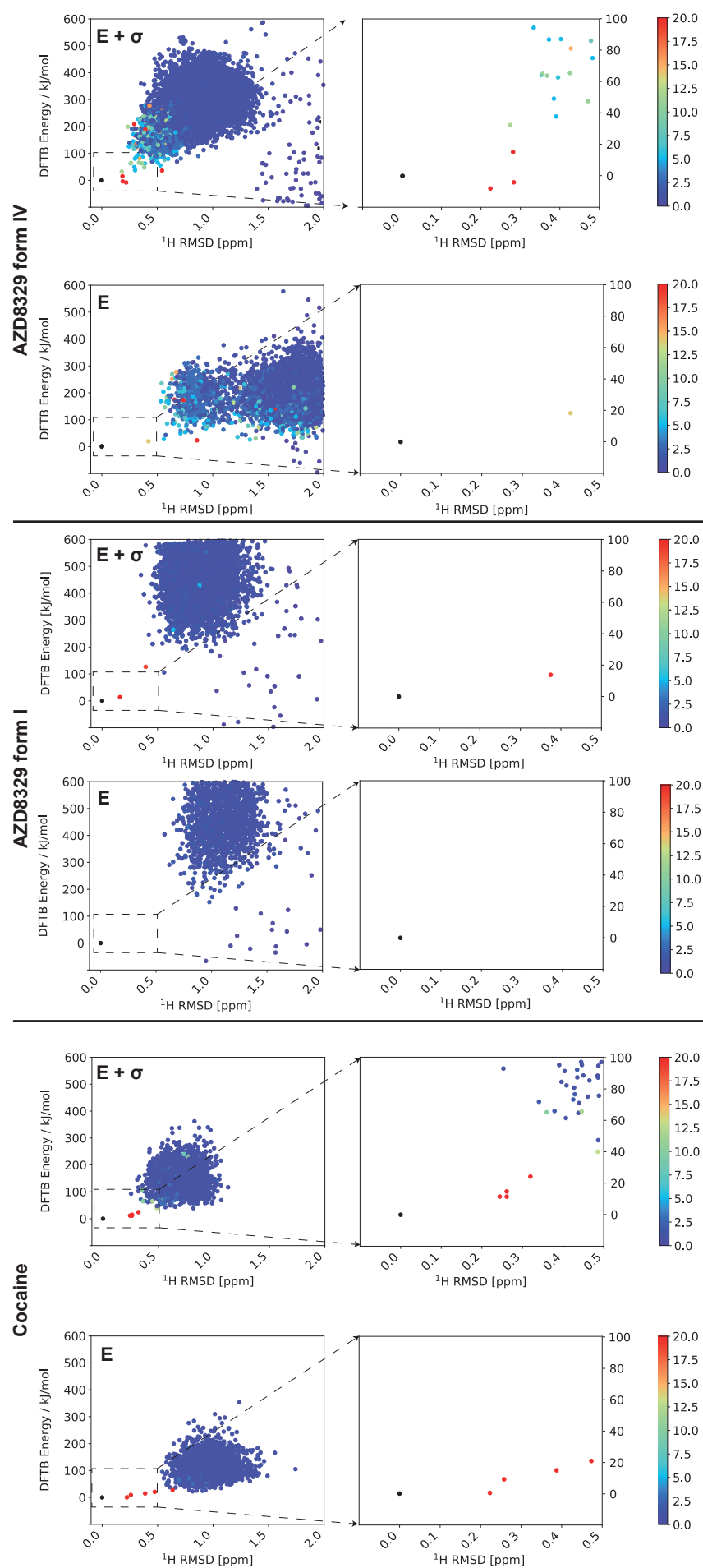
Figure 2-22 Plots of DFTB energy vs ^{1}H chemical shift rmsd for the results of 10000 simulated annealing runs on AZD8329 form IV, 10000 runs on AZD8329 form I and 2500 runs on cocaine. Each point represents a structure optimized as described in the methods section. The right column is a zoom on the low energy region. The vertical axis shows DFTB energies with respect to the known experimental structure which is set to 0. The color of each point reflects the similarity between each of the calculated structures and the reference structure, according to the scale on the right and as described in the methods section. Results for the optimization of AZD5718 form IV with (a) chemical shifts and DFTB, and (b) only DFTB used in the optimization process. Results for the optimization of AZD5718 form I with (c) chemical shifts and DFTB, and (d) only DFTB used in the optimization process. Results for the optimization of cocaine with (e) chemical shifts and DFTB, and (f) only DFTB used in the optimization process.

Figure 2-23 shows overlay of the reference structure with one of the found structures for each of compounds, AZD8329 form IV, AZD8329 form I and cocaine from left to right. By visual inspection it can be seen that the agreement is very good. $RMSD_{20}$ values were the following, 0.28, 0.14 and 0.17 Å for AZD8329 form IV, AZD8329 form I and cocaine respectively.
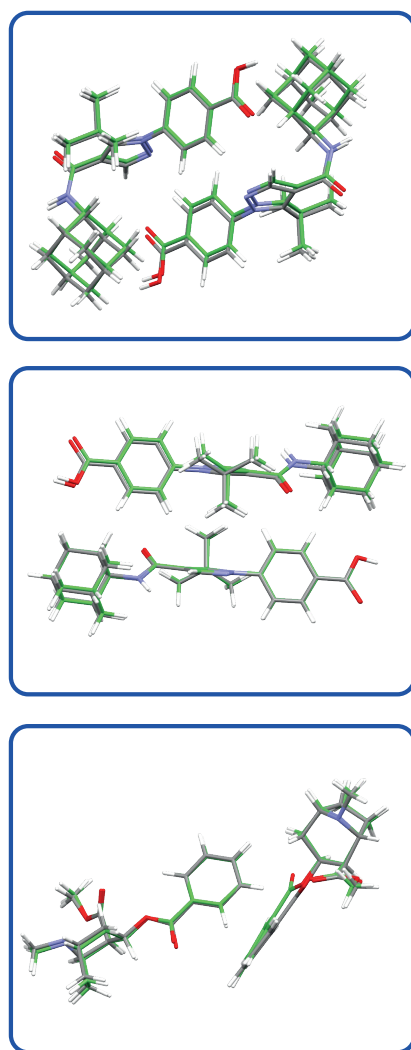
Figure 2-23 Overlay of two randomly taken molecules out of crystal lattice between one of the found structures for each of the system investigated in this study and the reference structure for each. From top to bottom: overlay of AZD8329 form IV, AZD8329 form I, cocaine.

2.2.4 Conclusions

We have shown the inclusion of machine learned chemical shifts directly into a crystal structure determination protocol. We have illustrated that for AZD8329 inclusion of machine chemicals shifts helped to solve the structure for two of its polymorphic forms. We have tested the method also for cocaine. In this study we choose to use Monte Carlo simulated annealing algorithm due to its relatively straightforward nature, but machine learned chemical shifts can be used with any method as they are easy to add as an additional pseudo energy term.

2.2.5 Supplementary

Trial crystal structure generation
First, a gas phase conformer was generated and the non-trivial torsional angles were randomized. In the case for AZD8329 the OCNH angle that corresponds to the amide bond was kept the same as in the experimentally found structure knowing that it can take only cis or trans position and if needed the other configuration can be explored too.

The randomized gas phase conformer was then introduced into a randomly generated crystal in the selected space group. Cell lengths, cell angles, position and the orientation of the asymmetric unit where changed. The maximum volume of the crystal was not larger than twice the sum of the van der Waals spheres of each atom in the molecule times the number of the molecules in the unit cell.

Both in the gas phase conformer and the crystal generation steps it was checked for interatomic clashes and the generation was run until no serious clashes were detected. The clash was defined as two atoms coming closer than 0.85 * sum of covalent radii.

Monte Carlo run
After the generation of the trial crystal structure, it was subjected to a Monte Carlo optimisation protocol. 4000 Monte Carlo runs were done with varying linearly temperature from 2500 K in the first run to 50 K in the last run. In each run one of the structure defining parameters were randomly selected (cell lengths, cell angles, position of the asymmetric unit, orientation of the asymmetric unit and torsional angles) and were randomly changed from 0 to the step size value. The initial step size was 2 Å for the cell lengths, 20 degrees for the cell angles, 0.05 fraction of the unit cell length in each direction for the translation, 30 degrees for the orientation and 40 degrees for the torsional angles.

After each random change a new, updated crystal was generated and the cost function was calculated (see main text). If the new cost function is lower in energy than the previous one the step is automatically accepted. If not, then the step is accepted if the following equation holds true:

$$\varepsilon < e^{\frac{E_{tot,old} - E_{tot,new}}{R \cdot T}}$$

where ε is a random number from 0 to 1, $E_{tot,old}$ is the old cost, $E_{tot,new}$ is the new cost, R is gas constant and T is temperature. If the step is accepted the step size is increased by a step factor and if it rejected, the step size is decreased by a step factor. We used step factor of 2 in our calculations. The max step size was set to 20 Å, 360 degrees, 1, 360 degrees and 360 degrees for cell lengths, cell angles, position, orientation and dihedral angles respectively. Also, min and max values for the cell lengths and angles were chosen to not produce very squished structures. These values were 1 to 50 Å for the cell lengths and 45 to 135 degrees for the cell angles.

Each 500 steps proton positions were optimised with DFTB.

Chemical shift referencing

To convert from chemical shieldings calculated by ShiftML to chemical shifts a set of 7 small organic molecules with known experimental chemical shifts were chosen.[241] Their shieldings were calculated and the calibration constants a and b for the equation 1.1 provided in the main text were found.

Data on the lowest energy / shift RMSD structures

Structure	$RMSD_{20}$, Å	Relative energy, kJ/mol	^1H shift RMSD, ppm	Matching molecules
AZ8329, form IV, E+σ	0.42	-4.3	0.28	20
AZ8329, form IV, E+σ	0.64	15.0	0.28	20
AZ8329, form IV, E+σ	0.28	-8.0	0.22	20
AZ8329, form IV, E	0.58	18.0	0.42	14
AZ8329, form I, E+σ	0.14	13.7	0.37	20
Cocaine, E+s	0.30	11.5	0.24	20
Cocaine, E+s	0.25	11.5	0.26	20
Cocaine, E+s	0.29	14.8	0.26	20
Cocaine, E	0.17	0.4	0.22	20
Cocaine, E	0.60	9.1	0.26	20
Cocaine, E	0.21	14.8	0.39	20

Table 2-7 $RMSD_{20}$, relative energy and 1H shift RMSD on the structures lower than 20 kJ/mol and 0.5 ppm ^1H RMSD after the Monte Carlo optimisation**.**

Scripts

The protocol to calculate the structures was home written from a scratch and the main part of the code is given in Appendix A.

2.3 Structure Determination of an Amorphous Drug through Large-Scale NMR Predictions

This chapter has been adapted from Cordova, M.; Balodis, M.; Hofstetter, A.; Paruzzo, F.; Lill, S. O. N.; Eriksson, E. S. E.; Berruyer, P.; de Almeida, B. S.; Quayle, M. J.; Norberg, S. T.; Ankarberg, A. S.; Schantz, S.; Emsley, L., Structure determination of an amorphous drug through large-scale NMR predictions. Nat Commun 2021, 12 (1).

My contributions to the publication. Performing NMR experiments, participating in the development of the methodology and investigation and participating in the writing and editing of the paper.

2.3.1 Introduction

Knowledge of the structure of amorphous solids can direct, for example, the optimization of pharmaceutical formulations, but atomic-level structure determination in amorphous molecular solids has so far not been possible. Solid-state nuclear magnetic resonance (NMR) is among the most popular methods to characterize amorphous materials, and Molecular Dynamics (MD) simulations can help describe the structure of disordered materials. However, directly relating MD to NMR experiments in molecular solids has been out of reach until now because of the large size of these simulations. Here, using a machine learning model of chemical shifts, we determine the atomic-level structure of the hydrated amorphous drug AZD5718 (**1**) by combining dynamic nuclear polarization-enhanced solid-state NMR experiments with predicted chemical shifts for MD simulations of large systems. From these amorphous structures we then identify H-bonding motifs and relate them to local intermolecular complex formation energies.

2.3.2 Methods

NMR experiments

Both crystalline and amorphous forms of **1** were provided by AstraZeneca. The samples were stored at equilibrium with the environment at approximately 22°C and 20% relative humidity prior to NMR analysis. The room temperature NMR experiments were performed on Bruker Ascend 500 wide-bore Avance III, Bruker 800 Ultrashield plus narrow-bore and 900 US^2 wide-bore Avance Neo NMR spectrometers. DNP solid-state NMR spectroscopy experiments were performed on a 400 MHz Avance III HD Bruker spectrometer. The spectrometer is equipped with a low temperature magic angle spinning (LTMAS) 3.2 mm probe and connected through a corrugated waveguide to a 263 GHz gyrotron capable of outputting ca. 5-10 W of continuous wave microwaves. All chemical shifts were referenced to alanine. For more details including experimental setup and the sample preparation see Supplementary Methods.

NMR crystallography

The candidate crystal structures were generated using a Monte-Carlo parallel tempering method[242] followed by lattice energy minimisation using an internally developed force-field. The 190 most stable candidates were selected for full DFT-D optimization at the PBE level of theory. Chemical shifts for the ten lowest energy candidates were computed at the PBE0 level of theory using the fragment- and cluster-based approach developed by Hartman *et al.*[243-245] The conversion from isotropic shielding to chemical shift was performed by linear regression between the obtained shieldings and experimental isotropic chemical shifts. The analysis of the positional uncertainty of the crystal structure was performed as described by Hofstetter *et al.*[246] by computing shifts of perturbed crystal structures obtained through low-temperature MD simulations of candidate #1 and relating chemical shift deviations to positional deviations. Chemical shift computations were also performed on an extended set of the 81 following lowest energy candidates, but did not lead to lower shift RMSEs than candidate #1.

Molecular dynamics simulation of amorphous structures

The amorphous structure of AZD5718 was modelled by carrying out MD simulations on periodic amorphous cells containing 128 molecules of **1** and a variable number of water molecules. Five cells of each water content; 0, 0.5, 1.0 and 2.0% (w/w, 0, 16, 32 and 65 water molecules in each cell, respectively), and two cells of 4% water (w/w, 132 water molecules in each cell) were generated. After equilibration for 1 ns using the canonical NVT ensemble at 298 K followed by 10 ns using the isothermal-isobaric ensemble (NPT) at 298 K and 1 bar, production simulations were carried out for 600 ns using the NPT ensemble at 298 K and 1 bar. Models of the amorphous structure were obtained by extracting 1,001 evenly spaced snapshots from the last 100 ns of each MD simulation, corresponding to 100 ps time steps between the extracted snapshots.

Chemical shift predictions and hydrogen bonding motifs

The predicted chemical shieldings of all snapshots extracted from the MD simulations (168,799,631 total shifts) were obtained using ShiftML version 1.2 (publicly available at https://shiftml.epfl.ch).[130, 194]

H-bonded N-H groups were identified in 11 snapshots from each MD simulation, spaced by 10 ns each. The corresponding bonding motifs were extracted by defining hydrogen bonds as N-H···X (X = O, N) patterns with an N-H-X angle above 130° and H-X bond length shorter than 2.5 Å, typically corresponding to moderate to strong hydrogen bonds in organic solids.[247] If the first H-bonded neighbor was found to be a water molecule, then secondary water-bound neighbors were searched for using the same criteria to define hydrogen bonds.

In addition, the N-H groups yielding predicted ^{1}H chemical shifts above 11 ppm were identified within each snapshot of the 4% water MD simulations (2002 total snapshots), and the corresponding hydrogen bonding patterns were extracted as described above.

Formation energies in the amorphous simulations

The formation energy of the intermolecular complex made of one molecule of AZD5718 and its local environment was computed for each molecule in the same 11 snapshots per

simulation as used for identification of all the hydrogen bonding motifs. The environment of a molecule was defined as all molecules having at least one atom within 5 Å from any atom in the probe molecule. The formation energy was computed as the difference between the energy of the total intermolecular complex and the energy of the isolated environment. The obtained formation energy thus contains both the ground-state energy of the isolated probe molecule, which includes its conformational energy, and the interaction energy between the probe molecule and its environment. The single-point energy computations were performed at the DFTB3-D3H5 level of theory using the 3ob-3-1 parameter set and the DFTB+ software version 20.1.[248-253]

2.3.3 Results and discussion

NMR crystallography

No polymorphism was observed for anhydrous crystalline **1**, nor were crystalline hydrates identified. The crystal structure of anhydrous crystalline **1** Form A was determined using a chemical shift-based NMR crystallographic approach. This involves the combination of the assigned experimental chemical shifts with CSP and computed chemical shifts. The ^{1}H, ^{13}C and ^{15}N resonances of AZD5718 (**Figure 2-24e**) were assigned using one-dimensional proton, carbon and nitrogen MAS NMR experiments (**Figure 2-24a-c**), as well as two-dimensional refocused ^{13}C-^{13}C INADEQUATE and ^{1}H-^{13}C HETCOR experiments (**Figure 2-24d-e**) as detailed in Supplementary Discussion. A set of DFT-D optimized candidate structures was generated using an internally developed rapid CSP approach, then the assigned experimental chemical shifts were compared to the shifts computed using the cluster- and fragment-based DFT approach introduced by Hartman et al.[243-245] for each structure in order to determine the experimental structure from the set of candidates. The structure determined using single-crystal X-ray diffraction was included and compared to the CSP set. The lowest energy CSP candidate (structure #1) was found to be structurally similar to the X-ray structure (as discussed in Supplementary Discussion).
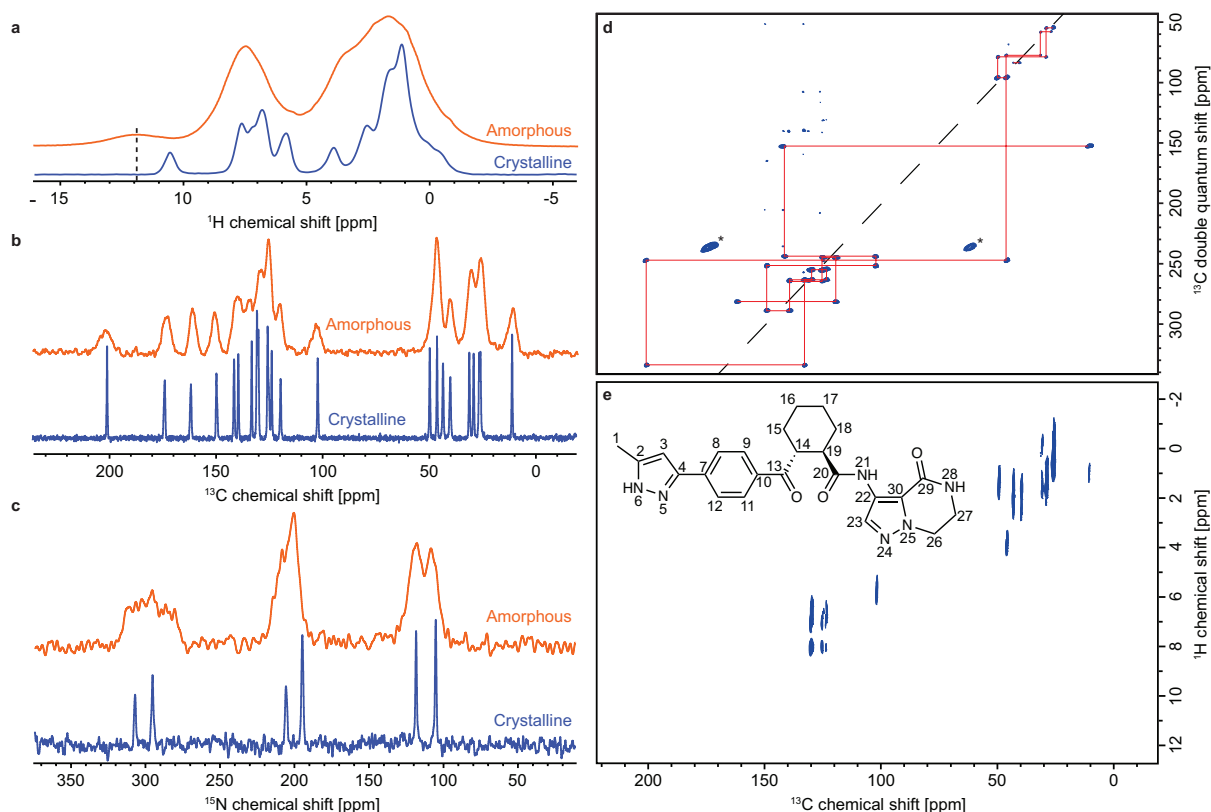
Figure 2-24 a) ^{1}H, b) ^{13}C and c) ^{15}N MAS NMR spectra of crystalline (blue) and amorphous (orange) **1** d) ^{13}C-^{13}C DNP enhanced solvent suppressed INADEQUATE and e), ^{1}H-^{13}C HETCOR spectra of crystalline **1**. The dashed black line in (a) indicates the chemical shift assigned to the proton bound to N6 in the amorphous sample. In (d), the 13C peaks denoted by a star at 60 and 170 ppm are attributed to impurities introduced during the NMR sample preparation. The chemical structure and labelling scheme of **1** is shown in (e).

The comparisons of the experimental and computed ^{1}H and ^{13}C chemical shifts are shown in **Figure 2-25a** and **b**, respectively. The root-mean-square errors (RMSEs) obtained for ^{1}H suggested that structure #1 best matches the experiment, while ^{13}C chemical shift results identified the X-ray structure as the best match. Additionally, the DFT-D energy per molecule of structure #1 was found to be the lowest among the CSP set (x-axis in **Figure 2-25a** and **b**). This also indicates that the force field used for the CSP procedure accurately describes the crystalline system, and supports the identification of candidate #1 as being the crystal structure. In order to elucidate the ambiguity between candidate #1 and the XRD structure, and to obtain a quantitative comparison of all candidates, a Bayesian probabilistic analysis was carried out using the approach introduced by Engel *et al.*[194] The two main advantages of using this method to determine the structure that best matches experiment are the quantitative determination of the confidence in the identification of the experimental structure on a continuous scale from 0 to 100%, and the combined use of NMR results for several elements, increasing the accuracy of the identification.
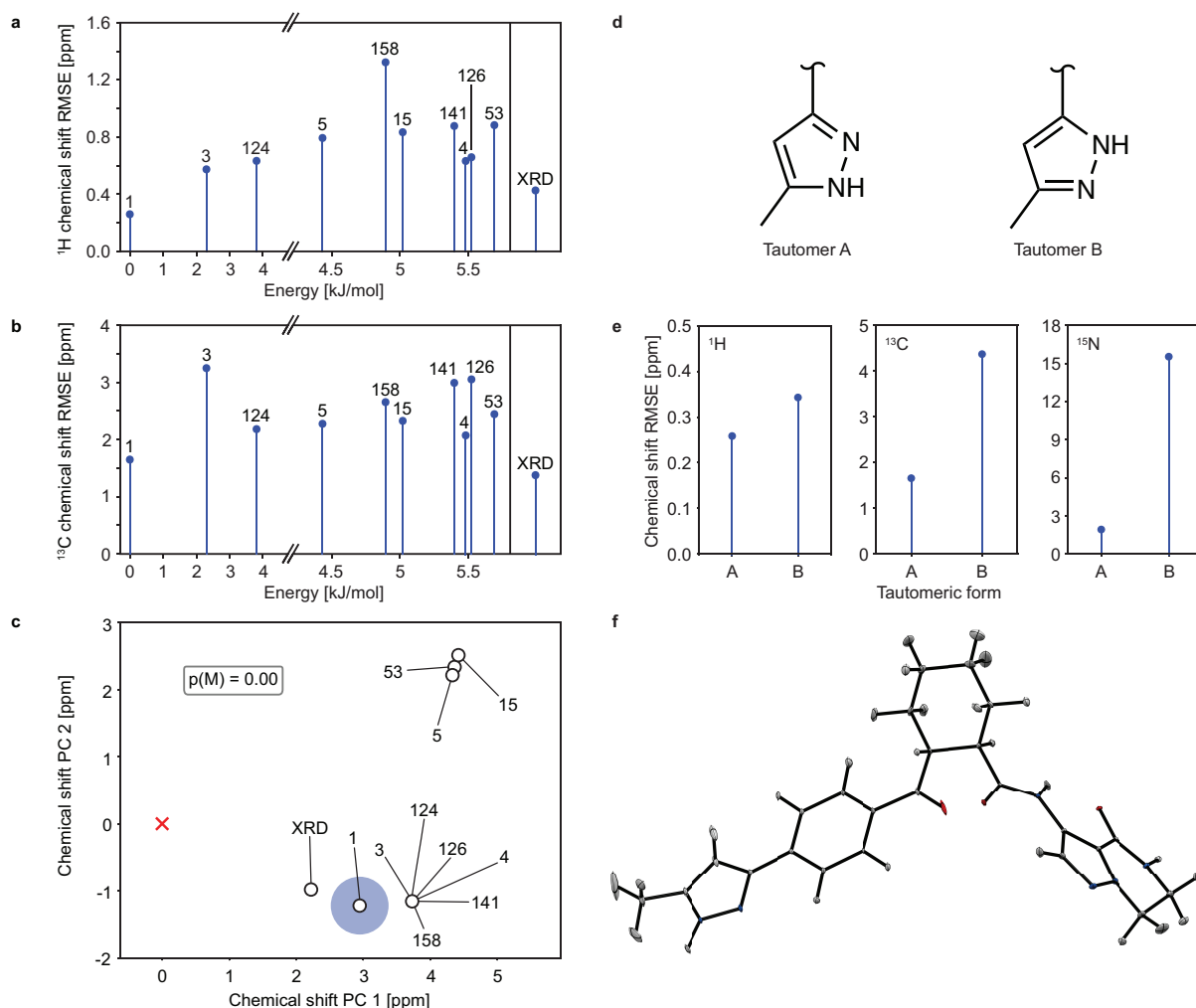
Figure 2-25 a) ¹H and b) ¹³C chemical shift RMSEs of the ten lowest DFT-D energy CSP candidates and of the crystal structure determined by X-ray diffraction (labelled XRD). The label rank in terms of increasing force-field energy per molecule of (**1**) of each candidate is indicated above each point. c) Two-dimensional projection of the similarity of the computed ¹H and ¹³C chemical shifts of the candidate structures to the experimental data (red cross). The probability of each candidate to match experiment is represented by the area of the blue disk. *p(M)* denotes the probability that a virtual candidate, which represents structures potentially missing from the CSP candidate pool, matches experiment. The virtual candidate is represented by the mean of the shifts of all CSP candidates, and a high probability *p(M)* suggests that the CSP candidate pool may not include the experimentally observed structure. d) Chemical structures of the two tautomers of **1** considered, labelled as A and B. e) Agreement between ¹H, ¹³C and ¹⁵N experimental and DFT computed chemical shifts for the two tautomers. F) ORTEP plot of the ADP tensors for the NMR structure of **1** drawn at the 90% probability level.

Figure 2-25c shows the results obtained with the Bayesian approach, represented as a principal component analysis (PCA) plot. This plot is a two-dimensional representation of the similarity of the different candidate structures according to their computed chemical shifts. The computed Bayesian probability of each structure to be the experimental crystal structure is represented by the area of the blue disk around each point. Using both ¹H and ¹³C chemical shifts, candidate #1 is found to be the most probable crystal structure, with 99.7% confidence.

Although the structure determined by X-ray diffraction (labelled XRD) appears closer to the experimental results (red cross) in the first two chemical shift principal components in **Figure 2-25c**, including the complete chemical shift space identifies candidate #1 as the structure that best matches experiment, as indicated by its associated confidence. **Figure 2-25c** highlights the similarity of the selected structures in terms of their chemical shifts, in the two dimensions that display the largest variance.

Comparison of the structures determined via XRD and NMR crystallography yielded a $RMSD_{15}$ (root mean square deviation of the atomic positions in 15 molecules, ignoring hydrogen positions) of 0.42 Å. The main difference between the two structures lies in the conformation of the bicyclo ring (see Supplementary **Figure 2-31**). Single-molecule heavy atom RMSD was found to be 0.22 Å, and was decreased to 0.15 Å after omitting the two carbons of the bicyclo ring (labelled 26 and 27 in **Figure 2-24e**).

Unlike X-ray diffraction, NMR is highly sensitive to hydrogen nuclei, making it the method of choice for validating the tautomeric form of **1**. Indeed, any of the two nitrogen atoms of the pyrazole ring (labelled 5 and 6 in **Figure 2-24e**) can be protonated in the crystalline sample. After computing ^1H, ^{13}C and ^{15}N shifts for the two possible tautomers displayed in Fig. 2d and comparing them with the experimental shifts (**Figure 2-24e**), the resulting chemical shift RMSE was found to be consistently lower for tautomer A, by a factor of 1.3 for ^1H, 2.6 for ^{13}C and 8.1 for ^{15}N. This unambiguously identifies tautomer A as the crystal structure. The position of the N-H proton on the pyrazole ring is crucial in setting up amorphous structures able to describe the properties of the amorphous phase of **1**.

The atomic displacement parameter (ADP) tensors of all atoms in the structure determined by NMR crystallography were obtained as described in Ref. [246]. Simulation details are given in Supplementary Methods. **Figure 2-24f** shows the ORTEP plot of the ADP tensors corresponding to a ^1H chemical shift RMSE of 0.34 ppm. This value corresponds to the estimated error of ^1H chemical shifts computed with the fragment- and cluster-based approach.[244] The average value of the ADPs is 0.00025 Å^2.

Hydrogen bonding motifs in the amorphous phase
Knowledge of the structure of AZD5718 in the amorphous phase is key to understanding its physicochemical properties. Investigation of the amorphous structure of **1** was performed using NMR experiments combined with ShiftML predicted chemical shifts for MD ensembles.

Comparison of the proton, carbon and nitrogen NMR spectra in crystalline and amorphous **1** shown in **Figure 2-24a-c** displays the overall broadening of the NMR signal typical of amorphous compounds. Apart from this observation, the chemical shifts do not display a significant change between the two phases of the compound. This suggests that **1** does not undergo large amplitude structural rearrangements upon transition from the crystalline to the amorphous state. The main difference between the NMR spectra of the two phases of **1** lies in the displacement of the ^1H resonance corresponding to the proton attached to the nitrogen labelled 6 (see **Figure 2-24e**) from 10.6 ppm in the crystalline sample to 11.8 ppm in the amorphous form. This suggests a change in the hydrogen bonding network in the structure.[254]

To better understand the structural differences between the amorphous and crystalline phases of **1**, we generated MD models of the amorphous structure at different hydration levels ranging from 0% to 4% (w/w) water content. This range of water content is representative of the experimental water content under real conditions, as confirmed by dynamic vapor sorption. Considering the large size of the simulation cells (128 molecules of **1** and up to 132 water molecules) and the large number of structures generated by MD, DFT computation of chemical shifts in these model systems would not be feasible. The machine learning model ShiftML was thus used to predict chemical shifts in these structures. The predicted spectra obtained for the crystalline structure and obtained by summing the spectra from 202 full cell snapshots of the 4% water MD simulations (i.e., 25,856 molecules of **1** and 26,664 water molecules) are displayed in **Figure 2-26**. The ^{1}H chemical shift RMSE obtained by comparing shifts predicted by ShiftML from the crystal structure with the experiment was found to be 0.61 ppm.
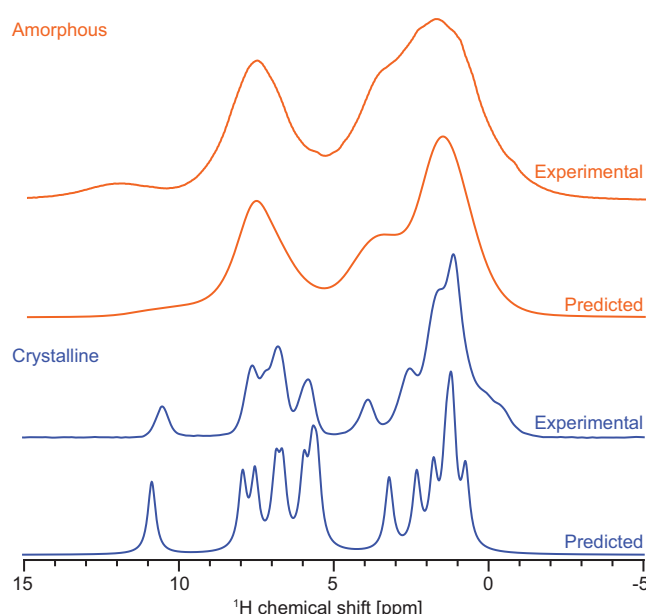
Figure 2-26 Predicted and experimental ^{1}H NMR spectra of crystalline (blue) and amorphous (orange) **1**. The predicted spectrum of amorphous **1** was obtained by considering only the 4% w/w water MD simulations.

Although no clear peak is observed at 11.8 ppm for the amorphous structure, the population of predicted shifts above 11 ppm was found to increase slightly with increasing water content (see Supplementary **Figure 2-32**). This behaviour suggests that interaction of **1** with water molecules does promote deshielding of the proton attached to the nitrogen labelled 6.

The predicted chemical shifts obtained were related to structural motifs in the model amorphous structures by identifying the different hydrogen bonding patterns (where the criteria are given in the methods section below) present in the structures and associated predicted ^{1}H shifts of the hydrogen bond donor groups. Figure 4a displays the chemical shift distributions of the most often occurring hydrogen bonding motifs. The atom most commonly bound to the N6-H group was found to be O20 (where O20 is the oxygen bound to C20), which corresponds to the hydrogen bond found in crystalline AZD5718. This is an indication that the structure of the amorphous compound is broadly similar to that of its crystalline counterpart. Over all analysed simulation snapshots (corresponding to an average water content of 1.16%

71

(w/w), or about 3.4 times more molecules of **1** than water), water was found to be the fourth most occurring hydrogen bonding partner to the N6-H group, and was found to lead to the most pronounced deshielding of the hydrogen bond donor proton (motif D in **Figure 2-27b**).

a

non-bonded (26.0%)

A (18.1%)

B (16.4%)

C (10.6%)

D (10.4%)

E (9.6%)

F (8.0%)

^{1}H chemical shift [ppm]

Occurrence

b

A (N6-H···O20)

B (N6-H···O29)

C (N6-H···O13)

D (N6-H···OH$_2$)

E (N6-H···N5)

F (N6-H···N24)

c

Number of occurrences

N6–H···O20
N6–H···O29
N6–H···N5
N6–H···OH$_2$
N6–H···OH$_2$/O20
N6–H···OH$_2$/O20
N6–H···O13
N6–H···N24
N6–H···OH$_2$, O13/O20
N6–H···OH$_2$/OH$_2$
N6–H···OH$_2$/O29
N6–H···OH$_2$/O20/O29
N6–H···OH$_2$, O13/O29*
N6–H···OH$_2$/O13
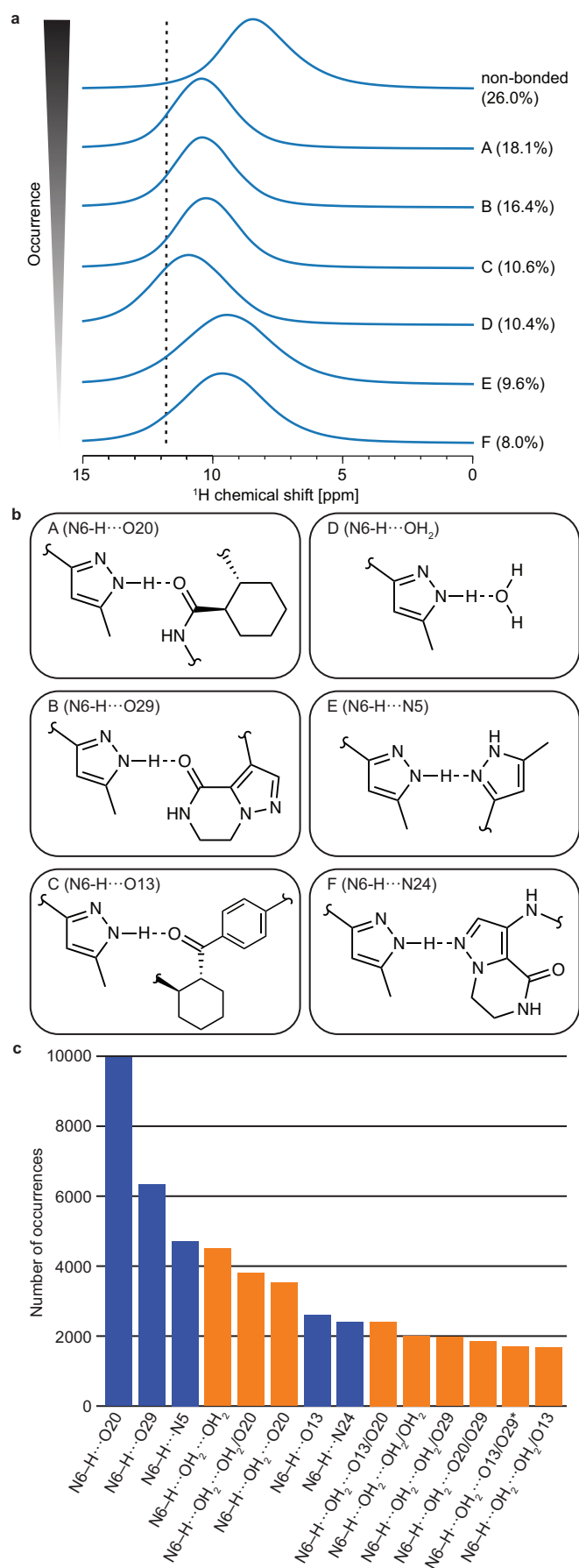
Figure 2-27 a) Predicted spectra obtained using the predicted ^{1}H chemical shifts of the most often occurring N-H bonding motifs involving N6 for 11 evenly spaced snapshots of all amorphous simulations of each water content. The percentages next to the spectra denote the fraction of bonding motifs their corresponding pattern represents, including the instances where no H-bonded neighbour was identified. The dashed vertical line indicates the experimental shift observed in amorphous **1** and assigned to the proton bound to N6. b) Hydrogen bonding motifs associated with the spectra in (a). c) Number of occurrences of extended H-bonding motifs yielding a predicted chemical shift above 11 ppm for every snapshot of the 4% water simulations. Only the patterns corresponding to the top 75% of all shifts above 11 ppm were selected. The orange bars represent the bonding motifs involving water, and the blue ones correspond to the motifs that do not involve water. Two secondary neighbours from the same molecule are indicated by an asterisk. In (b) and (c), On indicates the oxygen atom bonded to carbon n.

Because a single water molecule can form two hydrogen bonds involving its hydrogen atoms and two additional ones involving its oxygen atom, more extended hydrogen bonding motifs are likely to be observed for AZD5718 molecules bound to water. In order to investigate these extended patterns, we extracted N6-H···OH$_2$ motifs yielding a chemical shift above 11 ppm in all snapshots of the amorphous 4% water MD simulations, and obtained the secondary neighbours, bonded to the water protons. We restricted this analysis to the simulations with the highest water content, as bonding of water was found to lead to the largest deshielding of the proton attached to N6 (see **Figure 2-27a**), and as a larger number of water molecules in the simulation promotes extended hydrogen bonding motifs. **Figure 2-27c** shows the occurrences of extended hydrogen bonding patterns involving water, as well as the motifs made of pairs of H-bonded molecules of **1**, yielding a predicted shift above 11 ppm. The most often occurring pattern is the hydrogen bond present in the crystalline phase of the compound (N6-H···O20). When the H-bonded molecule is water, secondary neighbours are often found to be other water molecules, suggesting the formation of small clusters of water between AZD5718 molecules. Superpositions of ten instances of two hydrogen bonding motifs, N6-H···O20 and N6-H···OH$_2$···OH$_2$/O20, are shown in Fig. 5. It was observed that molecules of **1** being secondary H-bonded neighbours of N6-H generally lie away from the molecule bearing the hydrogen bond donor, indicating that steric clashes may constrain the possible geometries of hydrogen bonding in the amorphous form.
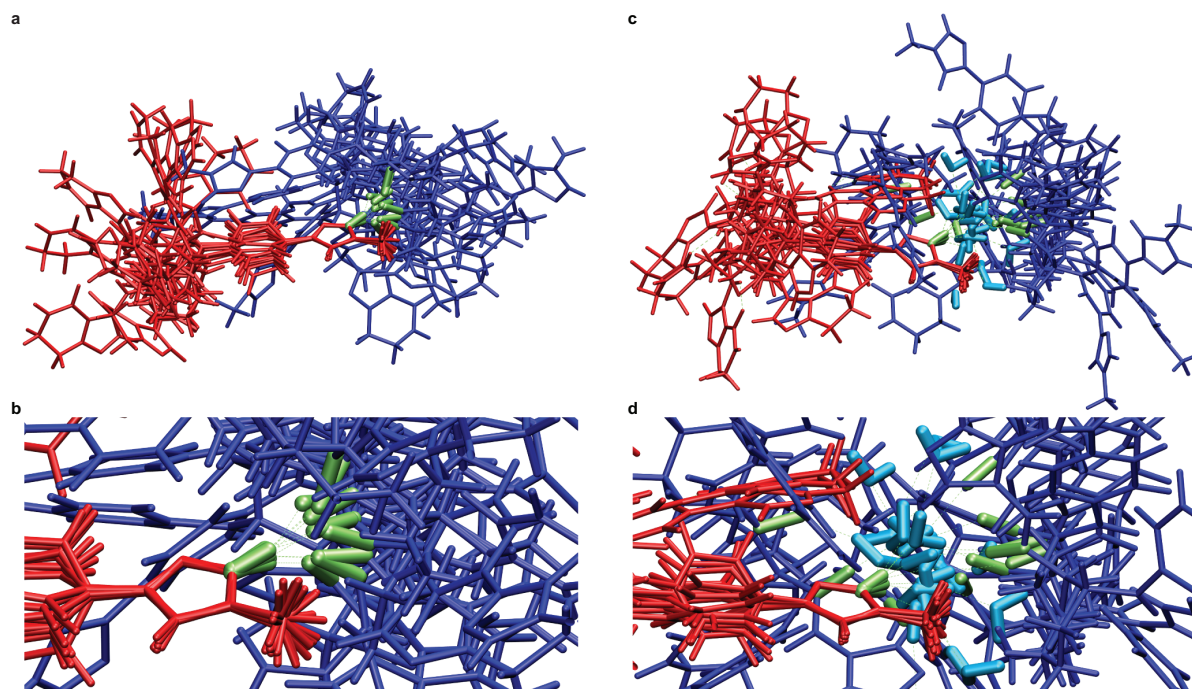
Figure 2-28 a) Superposition of 10 instances of the N6-H···O20 bonding motif. b) close-up view of the hydrogen bonding region in (a). c) Superposition of 10 instances of the N6-H···OH$_2$···OH$_2$/O20 bonding motif. d) close-up view of the hydrogen bonding region in (c). The red molecule represents **1** bearing the hydrogen bond donor (N6-H), the dark blue molecule represents **1** bearing the hydrogen bond acceptor, water molecules are coloured in cyan and the atoms of **1** involved in the hydrogen bonding motif are coloured in green.

Formation energies of intermolecular complexes

Obtaining the formation energies of the supramolecular complexes of molecules of **1** and their surroundings can help determine which hydrogen bonding pairs lead to overall more favourable intra- and intermolecular interactions. After computing the formation energies, including the conformational energy of the probe molecule, using the semiempirical DFTB3-D3H5 method, the results were gathered as a function of the hydrogen bonding motifs in which the probe molecule was involved as the hydrogen bond donor. The relative formation energy was defined as the difference between the formation energy of each instance and the mean formation energy of all instances where no hydrogen bond acceptor was found for the selected hydrogen bond donor.

Figure 6 shows the relative formation energy for each hydrogen bond donor and acceptor identified. Bonding of any N-H group to water was found to yield the most favourable interactions. Over all the simulation snapshots analysed for hydrogen bonding motifs, 6.1% of N21-H chemical groups were found to form an intramolecular hydrogen bond with the carbonyl labelled 29. This number may however be underestimated, as the same intramolecular hydrogen bond is found in the crystal structure, with a bond angle of 128.8°. This hydrogen bond would thus not be identified using the cut-off values selected here.
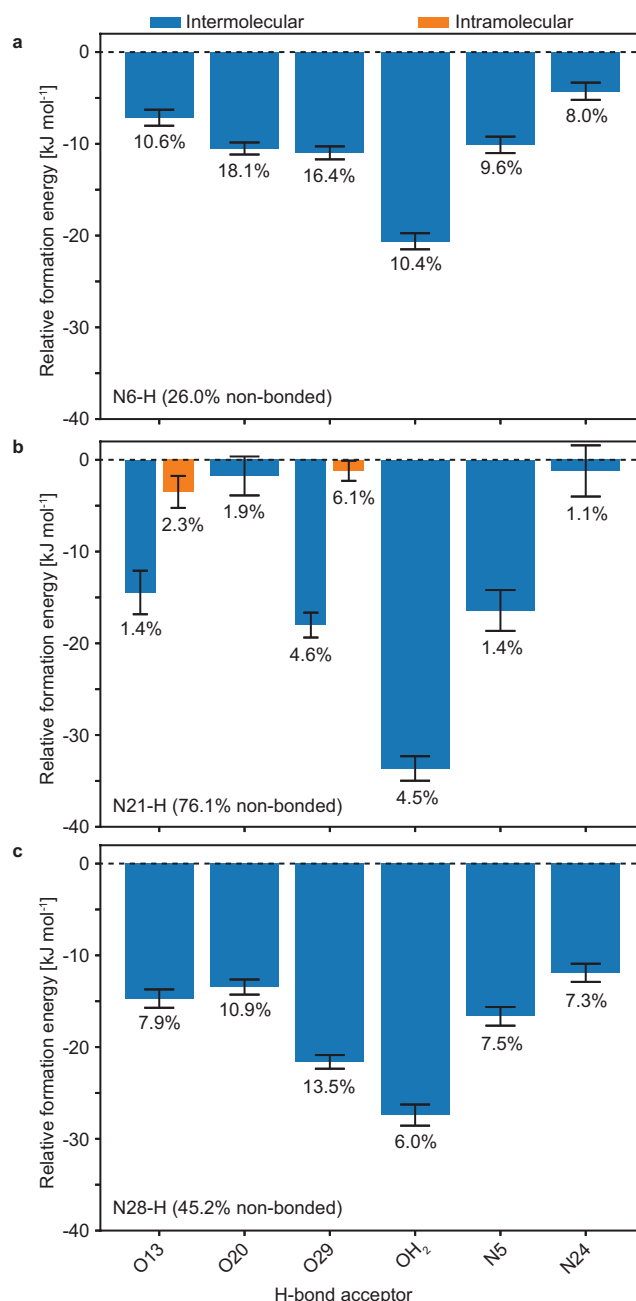
Figure 2-29 Mean computed formation energies of intermolecular complexes for the H-bond acceptor connected to **a** N6-H, **b** N21-H or **c** N28-H of the probe molecule. The percentage under each bar indicates the fraction of the N-H group bonded to the corresponding H-bond acceptor. Only the H-bond acceptors making up at least 1% of all instances analysed are displayed. The error bars indicate the standard error of the mean of the relative formation energies.

2.3.4 Conclusions

The atomic-level structure and hydrogen bonding patterns of the hydrated amorphous phase of AZD5718 were determined through solid-state NMR chemical shifts, MD simulations with various water contents, and machine learned chemical shifts. The chemical shifts associated with possible hydrogen bonding motifs generated from MD simulations were compared to

experimental NMR spectra in order to identify the most commonly occurring intermolecular interactions in the amorphous material. Bonding of N6-H to water was found to yield the largest deshielding of the proton involved in the hydrogen bond, and best described the experimental shift observed in the amorphous sample. This intermolecular bond to water was also associated with more favourable intermolecular complex formation energies as compared to direct H-bonding between two AZD5718 molecules. These favourable water-AZD5718 interactions highlight the potential ability of water to prevent physical aging of the amorphous drug.

The combination of the three techniques presented here was crucial in elucidating the structure of this amorphous material through a large-scale direct comparison of experimental chemical shifts with predicted shifts from MD structures. While solid-state NMR has already been used in tandem with MD simulations of amorphous materials, previous work have generally used molecular dynamics either to relate relative NMR peak areas to statistical ratios of different types of interactions,[193, 255] or to generate conformational ensembles from which small supramolecular clusters are extracted for DFT shift computation.[256] Overall, the method presented here can be applied to a wide range of disordered organic systems to determine their complete atomic-level structures from their NMR spectra.

The structure of the crystalline form was also determined using NMR crystallography to within a positional error of 0.1 Å and was confirmed to be almost identical to the structure obtained with single crystal X-Ray diffraction.

2.3.5 Supplementary

Sample preparation for DNP NMR

In DNP MAS experiments, the high thermal polarization is transferred from unpaired electrons to nuclei (typically ^{1}H) which results in enhanced NMR signals. For organic powders, this is achieved by impregnating the powdered solid with an otherwise inert polarizing solution.[257-258] **1** dissolves both in water and in most organic solvents, so most typical polarizing solutions, such as 16 mM TEKPOL in 1,1,2,2-tetrachloroethane (TCE), were found to be incompatible. *Ortho*-terphenyl was found to be a suitable non-solvent for AZD5718, and 16 mM TEKPOL in *ortho*-terphenyl 99.5%-d$_{14}$ (OTP-d$_{14}$) was used as a polarization source. The sample was prepared according to the procedure described in references [259] and [260] by mixing a solid solution of 16 mM TEKPOL in OTP-d$_{14}$ with powdered **1**, then transferring it to a sapphire rotor sealed with a PTFE insert and capped with a zirconia drive cap. The rotor was then heated at ca. 65°C in a hot water bath in order to melt the OTP and allow the liquid to impregnate the API. It was then quickly inserted into the pre-cooled LT-MAS DNP probe to rapidly freeze the sample in order for the OTP to form a glass.[260] DNP enhancements of about 5 as measured on crystalline AZD5718 signals through (^{1}H)^{13}C DNP CPMAS were obtained, which was sufficient to allow the natural abundance INADEQUATE spectra to be recorded.

NMR spectroscopy

Experiments were performed on Bruker Ascend 400 and Ascend 500 wide-bore Avance III, and on Bruker 800 Ultrashield plus narrow-bore, and 900 US^2 wide-bore Avance Neo NMR spectrometers. The spectrometers operate at ^{1}H Larmor frequencies of 400.13, 500.43,

800.13, and 900.13 MHz respectively, and are equipped with H/X/Y 3.2 mm, H/C/N/D 1.3 mm and H/C/N 0.7 mm CPMAS probes. When the 3.2 mm probe was used, the samples were restricted to the central third of a rotor with an inner diameter of 2.2 mm, in order to maximize rf homogeneity.

DNP solid-state NMR spectroscopy experiments were performed on a 400 MHz Avance III HD Bruker spectrometer. The spectrometer is equipped with a low temperature magic angle spinning (LTMAS) 3.2 mm probe and connected through a corrugated waveguide to a 263 GHz gyrotron capable of outputting ca. 5-10 W of continuous wave microwaves.[155] The sweep coil of the main magnetic field was optimized so that the microwave irradiation gave the maximum positive proton DNP enhancement with binitroxide cross effect-based polarizing agents (e.g. AMUPOL, TEKPOL). DNP enhancements were determined based on the ratio of the area of the spectra acquired with and without microwave irradiation.

1D ^{1}H MAS NMR spectra were recorded at a temperature of 298 K using rotor spinning rates (v_r) up to 111 kHz. 1D ^{13}C cross-polarization[261] (CP) MAS NMR spectra were acquired at 298 K with v_r of 22 kHz. The CP contact time was 2 ms and during the signal acquisition SPINAL-64 decoupling[29] was applied with a ^{1}H rf field amplitude of 100 kHz. 1D ^{15}N CP NMR spectra were acquired at 100 K under DNP MAS conditions with v_r = 12.5 kHz for crystalline AZD5718, and similar measurements were made on amorphous AZD5718 using LT-MAS conditions (without DNP) in the same instrument. Variable amplitude cross-polarization[262] was used to transfer polarization from ^{1}H (60% to 100% ramp) to ^{15}N (constant amplitude). For the ^{15}N CPMAS spectra of crystalline AZD5718, 360 scans were acquired with DNP spaced by a recycling delay of 20 s leading to a total acquisition time of 2 h. For amorphous AZD5718, 14,720 scans were acquired without DNP, spaced by a recycling delay of 5 s leading to a total acquisition time of 21 h.

2D ^{1}H-^{13}C HETCOR experiments were carried out at 298 K using v_r = 22 kHz. 96 points were acquired in the indirect dimension with the States acquisition method,[263] and with indirect sampling intervals (Δt_1) of 96 μs. For the crystalline sample the recycle delay was 32 s (T_1 ~ 22 s) and 64 scans were collected for each t_1 point. For the amorphous sample the recycle delay was 4 s (T_1 ~ 3 s) and 769 scans were collected for each t_1 point. During t_1 100 kHz eDUMBO-1$_{22}$ was applied to decouple the ^{1}H-^{1}H dipolar coupling,[23] and during t_2 100 kHz SPINAL-64 decoupling was applied.

The 2D ^{13}C-^{13}C refocused INADEQUATE[18, 34] spectrum of crystalline **1** was acquired using DNP MAS NMR.[264] For the ^{13}C-^{13}C refocused INADEQUATE experiment, the probe was configured into ^{1}H/^{13}C double resonance mode. Variable amplitude cross-polarization[262] was used to transfer polarization from ^{1}H to ^{13}C. SPINAL-64[29] heteronuclear ^{1}H decoupling with RF fields of 100 kHz was applied in all cases.

The DNP enhancement allowed to record a ^{13}C-^{13}C refocused ^{13}C-^{13}C INADEQUATE spectrum at natural abundance for **1** in about 2 days of signal averaging. Moreover, using a ^{1}H spin-lock of 30 ms between the ^{1}H excitation pulse and the CP, the otherwise dominant OTP solvent signal was efficiently removed,[265] allowing to record a 2D spectrum ^{13}C-^{13}C refocused DNP INADEQUATE clean from the solvent signal. The spectrum was acquired in about 45 h with 128 points recorded in the indirect dimension with 256 scans each separated by recycling

time of 5 s. The increment in the indirect dimension was 40 µs, allowing a total indirect acquisition time of 5.12 ms using the States-TPPI method.[266] The tau period for J evolution was optimized and set to 4 ms. SPINAL-64 was used for heteronuclear decoupling. All chemical shifts were referenced via alanine.

Solid-state NMR experimental setup

	^{1}H	^{13}C	^{15}N
MAS rate	111 kHz	22 kHz	12 kHz
Recycle delay (d_1)	10 s	32 s	20 s
^{1}H to X CP			
Spin lock duration	-	2 ms	10 ms
Total acquisition time	5.5 ms	30 ms	25 ms
Dwell time	2.8 µs	13.2 µs	12.3
Number of points	1964	2268	2032
Number of scans	4	128	360
Acquisition mode	DQD	qsim	qsim

Table 2-8 Experimental parameters for 1D experiments on AZD5718 form A anhydrous

	^{1}H-^{13}C HETCOR	^{13}C-^{13}C INADEQUATE
MAS rate	22 kHz	12.5 kHz
Recycle delay (d_1)	32 s	5 s
^{1}H to X CP		
Spin lock duration	0.1 ms	3 ms
Acquisition in the indirect dimension (t_1)		
Total acquisition time	4.6 ms	2.6 ms
Dwell time	96 µs	20 µs
Number of points	96	256
Acquisition in the direct dimension (t_2)		
Total acquisition time	33 ms	15 ms
Dwell time	9.9 µs	5 µs
Number of points	3328	128
Number of scans per increment	64	128
Acquisition mode	States	States-TPPI
Delay t	-	5 ms

Table 2-9 Experimental parameters for 2D experiments on AZD5718 form A anhydrous

	^{1}H	^{13}C	^{15}N
MAS rate	62.5 kHz	22 kHz	8 kHz
Recycle delay (d$_1$)	6.5 s	4 s	5 s
^{1}H to X CP			
Spin lock duration	-	2 ms	10 ms
Total acquisition time	8.2 ms	30 ms	25 ms
Dwell time	1.0 µs	9.9 µs	12.3
Number of points	8192	3024	2032
Number of scans	4	128	30720
Acquisition mode	DQD	qsim	qsim

Table 2-10 Experimental parameters for 1D experiments on AZD5718 amorphous

	^{1}H-^{13}C HETCOR
MAS rate	22 kHz
Recycle delay (d$_1$)	4 s
^{1}H to X CP	
Spin lock duration	0.1 ms
Acquisition in the indirect dimension (t$_1$)	
Total acquisition time	4.6 ms
Dwell time	96 µs
Number of points	96
Acquisition in the direct dimension (t$_2$)	
Total acquisition time	33 ms

Table 2-11 Experimental parameters for 2D experiments on AZD5718 amorphous

Comparison of the structures determined via X-ray diffraction and NMR crystallography

The crystal structure determined using single-crystal X-ray diffraction (**Figure 2-30**) was compared to the structure obtained through NMR crystallography. The superposition of the two structures is shown in **Figure 2-31**. The two structures were found to be highly similar except for the conformation of the bicyclo ring (on the left of **Figure 2-31**).
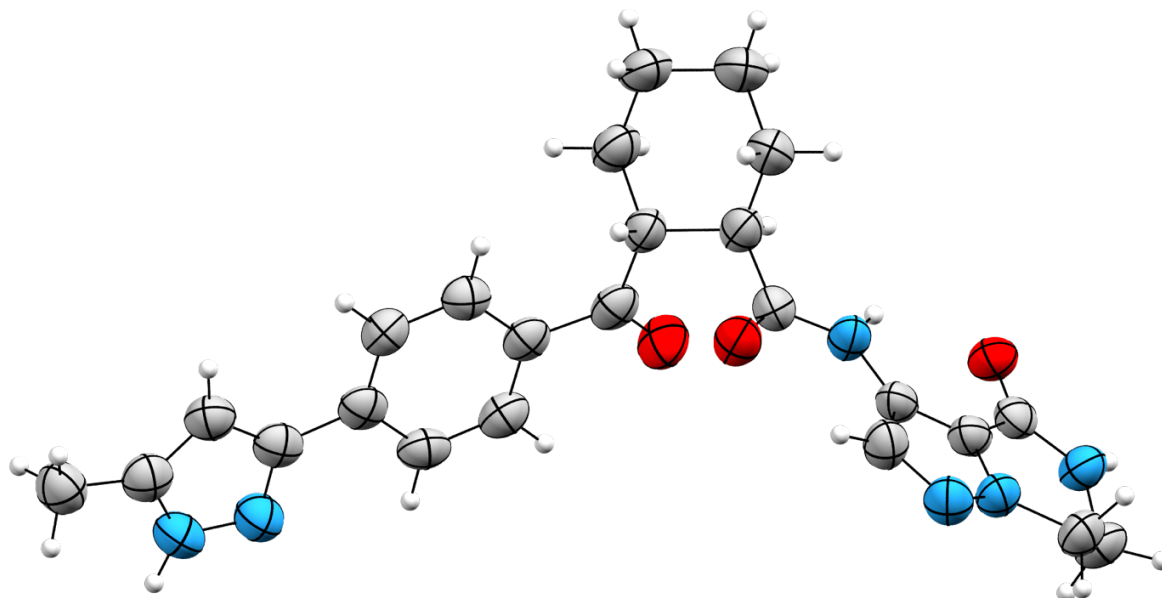
Figure 2-30 Positional uncertainty of the X-ray determined structure of AZD5718. ORTEP plot of the heavy atom ADP tensors for the crystal structure of 1 determined using single crystal X-ray diffraction, drawn at the 90% probability level.
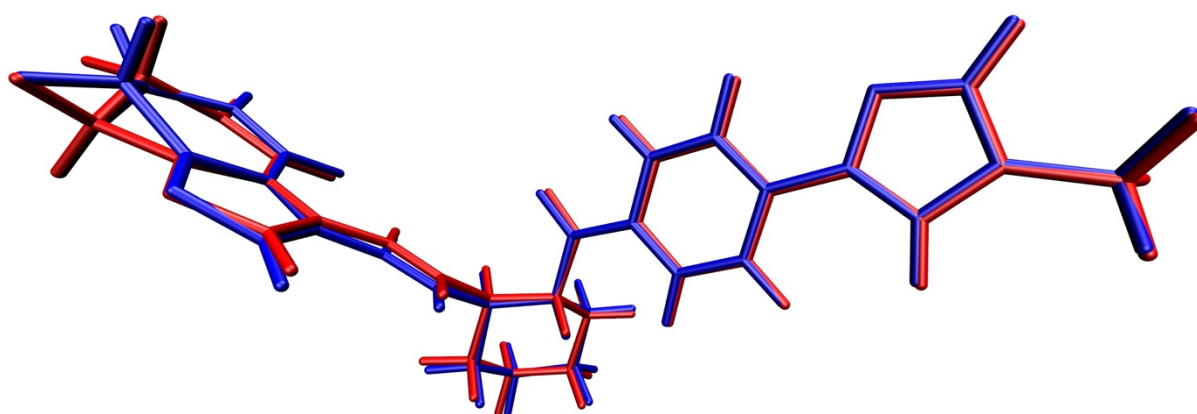
Figure 2-31 Similarity between the XRD and NMR crystallography structures. Comparison between the structure of **1** determined using X-ray diffraction (red) and NMR crystallography (blue).

Simulated spectra of AZD5718 amorphous MD simulations with different water contents

The simulated spectrum for each water content was computed by summing Lorentzian functions centred at the predicted shifts, and with a width of 0.3 ppm. The parameters for the conversion from shielding to shift were extracted by comparing the 4.0% water simulated

spectrum as described in section 1.8, and were applied to all simulations of different water contents. The spectra were normalised such that their maximum is one. Although the experimental peak observed at 11.8 ppm was not observed in the simulated spectra, a larger population of the shifts above 11 ppm was observed with increasing water content (**Figure 2-32**).
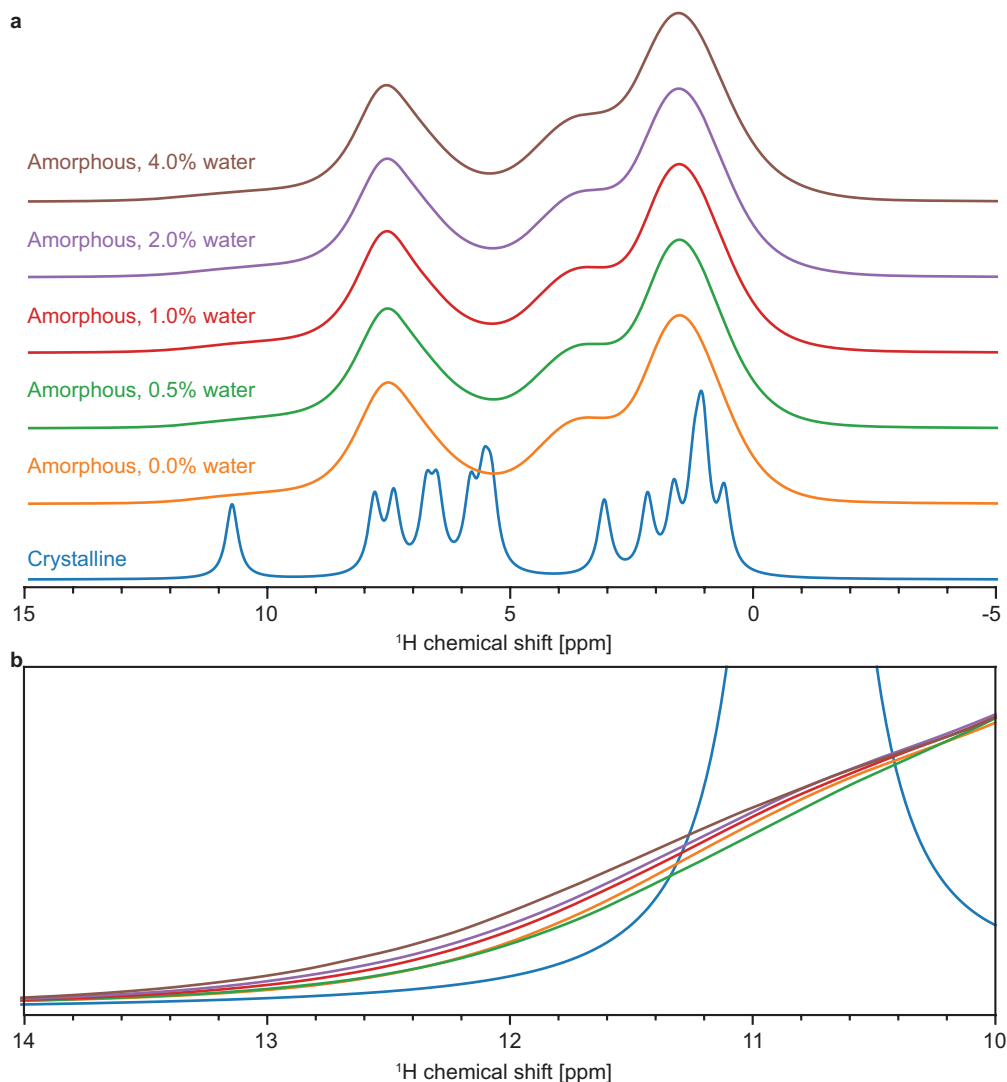
Figure 2-32 Effect of the water content on simulated ^{1}H NMR spectrum. a) Simulated ^{1}H NMR spectra of crystalline and amorphous AZD5718. b) Close-up view of the spectra in the region between 10 and 14 ppm.

Chapter 3.   Chemical shift assignment

Chemical shift assignment is the starting point of any detailed NMR study.[267] In organic solids at natural isotopic abundance, this is still a laborious and often challenging process. In particular, ^{13}C resonance assignment typically requires the use of the through-bond ^{13}C-^{13}C INADEQUATE experiment.[268-269] For materials for which the crystal structure is already known, the assignment can be determined at least partially by comparing the experimental chemical shifts with shifts computed using DFT in the gauge invariant projector augmented wave (GIPAW) method [270-271], or fragment-based methods.[272-273] However, in most applications the full structure is not known, and in particular as discussed above *de novo* chemical shift-based NMR crystallography, which is the main focus of this theis, relies on chemical shift assignment in order to either directly drive structure determination or to identify the crystal structure from among a set of candidates generated, for example, through crystal structure prediction (CSP). [118, 274-276]

Chemical shift assignment of biomolecules such as proteins and RNA can be obtained directly from their sequence through statistical analysis of chemical shifts[277-279]. In addition, simultaneous chemical shift assignment and structure determination can be obtained from matching atomic contacts to Nuclear Overhauser Effect (NOE) experiments[277]. These approaches rely on the existence of a large database of experimental chemical shifts and molecular structures, such as the Biological Magnetic Resonance Data Bank (BMRB)[280] and Protein Data Bank (PDB)[281], respectively. For example, the BMRB contains over 9.4 million instances of experimental chemical shifts for 279 types of proton, carbon and nitrogen sites in the 20 amino acids that make up proteins, with e.g. over 89 000 instances of the NH shift in alanine alone! Such large and diverse chemical shift databases however do not exist, to our knowledge, for organic crystals.

In **Chapter 3.1** we show how combining this model with a database of three-dimensional structures such as the Cambridge Structural Database (CSD)[173] enables the probabilistic assignment of organic crystals using chemical shift statistics without any knowledge of the 3D structure. We generate a large database of chemical shifts for organic crystals by predicting shifts using ShiftML on structures extracted from the CSD. By relating the shifts obtained to molecular fragment descriptors, we obtain probabilistic assignments of organic crystals directly from their molecular structure.

Databases are foundational for the scientific community when it comes to quickly find, evaluate, compare and analyse published data in one place. That is especially true now, in the era of deep-learning approaches that build upon existing experimental or computational data to predict new properties and materials.[129, 177] One can find a handful of established databases in the NMR community. The biological magnetic resonance data bank (BMRB) is a large database for the biological NMR data.[280] A solution state NMR database by the name of Spectral Database for Organic Compounds (SDBS) is commonly used by organic chemists and others.[282] NMRShiftDB is similar to SDBS and also incorporates spectral predictions[283]. These are just a few of the largest databases, but there are others, either being dedicated only to NMR and covering some aspects of chemistry[284] or NMR being a part of a much wider database that includes infrared spectra, mass spectra and others.[285] With that being said

there is still no centralised database to deposit raw solid-state NMR data for small molecules and materials.[267]

In **Chapter 3.2** we present the preliminary results towards creating a solid-state NMR database for materials and small molecules. We describe the structure of the database, initial design concepts and also compile some ^{13}C and ^{1}H isotropic shifts published so far in the scientific literature.

3.1 Bayesian Probabilistic Assignment of Chemical Shifts in Organic Solids

This chapter has been adapted from Cordova, M.; Balodis, M.; de Almeida, B. S., Ceriotti, M., Emsley, L., Bayesian Probabilistic Assignment of Chemical Shifts in Organic Solids. Sci. Adv., in press.

My contributions to the publication. Performing NMR experiments, participating in the development of the methodology and investigation and participating in the writing and editing of the paper.

3.1.1 Introduction.

A pre-requisite for NMR studies of organic materials is assigning each experimentally-determined chemical shift to a set of geometrically equivalent nuclei. Obtaining the assignment experimentally can be challenging and typically requires time-consuming multi-dimensional correlation experiments. An alternative solution for determining the assignment involves statistical analysis of experimental chemical shift databases, but no such database exists for molecular solids. Here, by combining the Cambridge structural database with a machine learning model of chemical shifts, we construct a statistical basis for probabilistic chemical shift assignment of organic crystals by calculating shifts for over 200 000 compounds. We then relate the shifts obtained to a topological descriptor of covalent environment to obtain the probabilistic assignment of the NMR spectra of organic crystals directly from their two-dimensional chemical structure. The approach is demonstrated with the ^{13}C and ^{1}H assignment of eleven molecular solids with experimental shifts, and benchmarked on 100 crystals using predicted shifts. The correct assignment was found among the two most probable assignments in over 80% of cases.

3.1.2 Results

The framework presented here was applied to a set of various organic molecules for which the carbon chemical shift assignment was already (at least partially) determined experimentally. The selected set is composed of theophylline [118], thymol [286], cocaine [118], strychnine, AZD5718 [276], lisinopril [287], ritonavir, the K salt of penicillin G [288], β-piroxicam [289], decitabine [126] and simvastatin [290]. The experimental spectra used for the assignment of strychnine and ritonavir are shown in **Figure 3-8** and **Figure 3-9**. Experimental shifts of lisinopril were obtained from a dihydrate form [287]. Experimental shifts of ritonavir were obtained from the polymorphic form II.
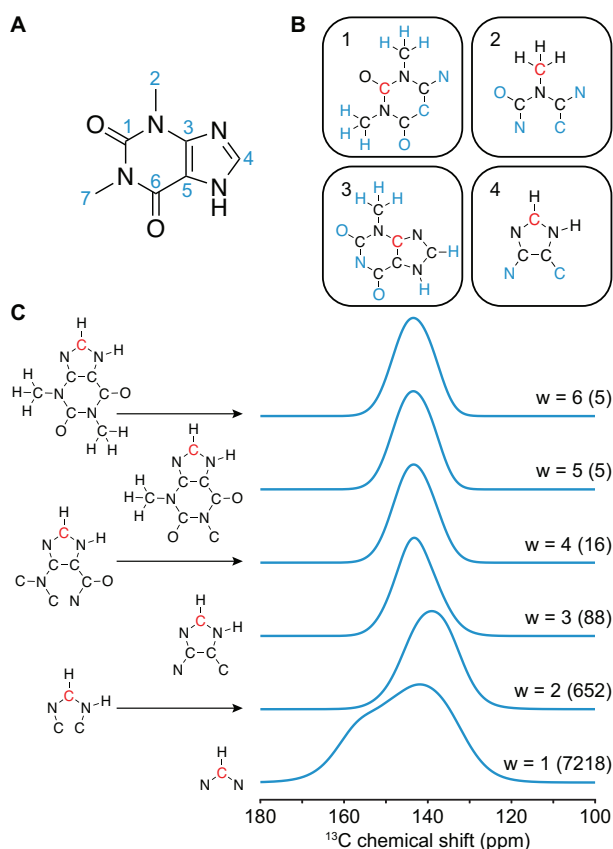
Figure 3-1 Graph and statistical distribution. (A) Two-dimensional structure and carbon labelling scheme of theophylline. **(B)** Graphs of carbons 1, 2, 3 and 4 of theophylline constructed at a depth $w = 3$. In each graph, the red vertex corresponds to the central atom (for which the chemical shift distribution is extracted), and blue vertices indicate the atoms at the maximum shortest path from the central vertex. **(C)** Chemical shift distributions obtained corresponding to the carbon labelled 4, with different graph depths w. The number of instances from the database used to construct each distribution is indicated in parentheses.

Graph generation is the starting point of statistical assignment and can be performed directly from the two-dimensional representation of the molecule. **Figure 3-1A-B** shows the graphs generated for illustrative carbon atoms in theophylline with a depth $w = 3$. The chemical shift distributions of the carbon labelled 4 in theophylline corresponding to different graph depths are shown in fig. 2C, together with the corresponding graphs. As expected, the distribution changes significantly as w is increased, until at $w = 3$ and above where they are found to be highly similar, with a width dominated by the uncertainty in the ShiftML prediction. We thus selected a minimum number of ten instances to construct each probability distribution, and used the maximum graph depth that fulfils this requirement for each nucleus.

The prior statistical distribution of chemical shifts for each atom in a molecule can be constructed from the shifts predicted for all atoms in the database that share the same graph. Evaluating the obtained statistical distributions at the observed shifts yields the probability of observing each shift originating from each nucleus in the molecule (**Figure 3-2A-B**). The possible combinations of individual assignments, based on a Bayesian construction, makes it possible to associate a probability to each global assignment of all shifts. After obtaining the probability for each global assignment in the set, marginalization yields individual assignment probabilities (**Figure 3-2C**). In this case, the most probable individual assignment for each

86

carbon, as well as the most probable global assignment, were found to correspond to the experimental assignment of theophylline (black dots in **Figure 3-2C**).
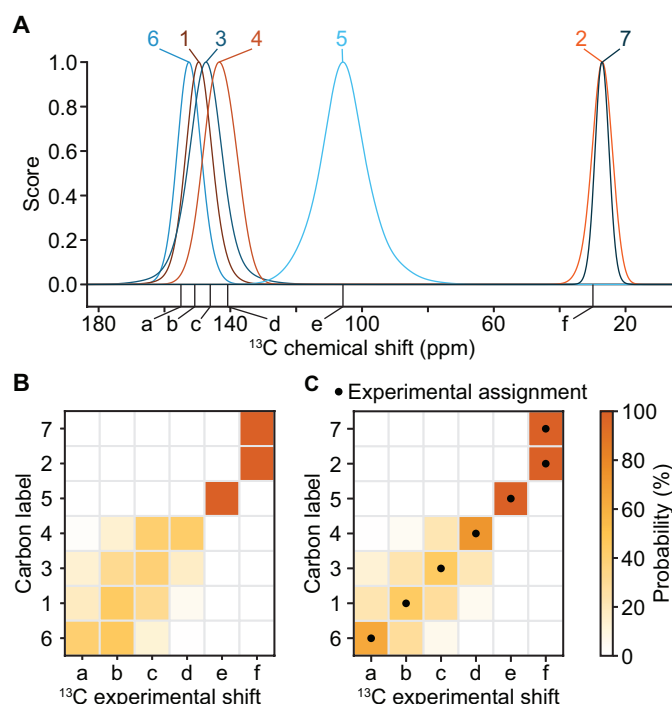
Figure 3-2 Probabilistic assignment of theophylline. (A) Statistical ^{13}C chemical shift distributions for theophylline (coloured lines). The carbon labels follow fig. 2A. Experimental shifts are indicated by black vertical lines below the distributions and are labelled a through f in order of decreasing chemical shift. **(B)** Probabilities of observing each chemical shift of theophylline for a given carbon nucleus. **(C)** Marginal individual assignment probabilities of the ^{13}C chemical shifts of theophylline after Bayesian inference of the possible global assignments. The dots indicate the experimentally determined correct assignment.

Overlap of the chemical shift distributions can lead to highly ambiguous assignments. A common method to separate overlapping NMR signals consists in spreading them along multiple dimensions. The HETCOR experiment yields high-sensitivity correlated ^1H and ^{13}C chemical shifts of dipolar coupled nuclei, and can be tuned to obtain a spectrum dominated by one-bond correlations [291-292]. The correlated statistical distributions of chemical shifts corresponding to a simulated HETCOR can be obtained by considering bonded CH pairs in the molecule. This additional dimension often helps separate overlapping one-dimensional statistical distributions and chemical shifts by incorporating the additional information given by the ^1H chemical shift. In addition, this can also be used to simultaneously assign ^{13}C and ^1H chemical shifts.
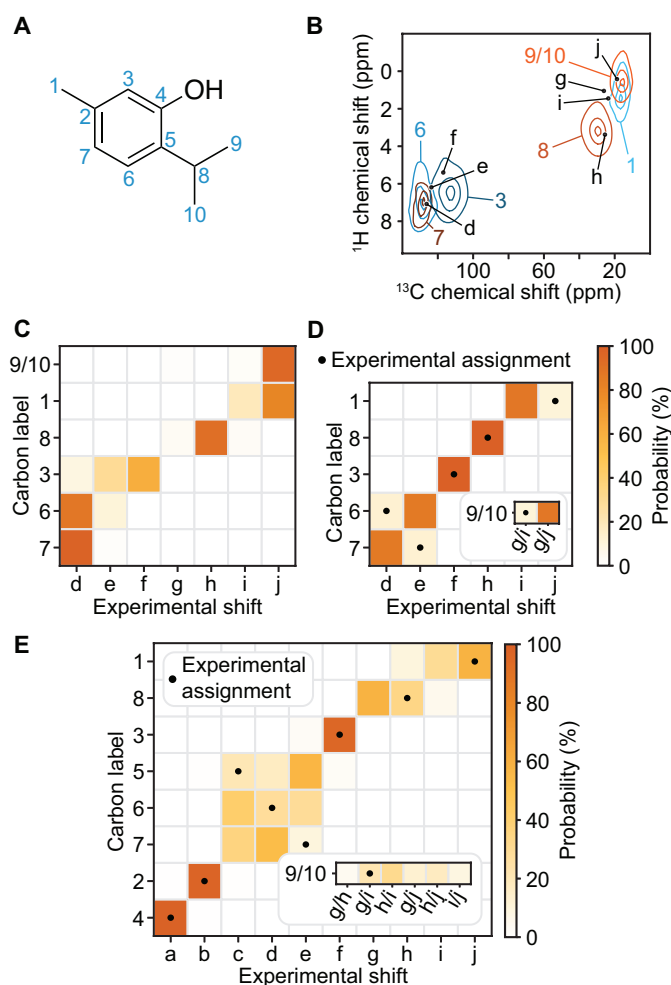
Figure 3-3 Probabilistic assignment of thymol using simulated HETCOR. (A) Carbon labelling scheme of thymol. **(B)** Contour plot of the correlated statistical chemical shift distributions of bonded ^{13}C-^{1}H in thymol. The carbon labels follow **A**. Experimental shifts are indicated by black dots and are labelled alphabetically in order of decreasing ^{13}C chemical shift (see fig. S10). The statistical distributions, normalized such that their maximum is one, are drawn as contour plots at levels 0.1, 0.5 and 0.9. **(C)** Probabilities of observing each ^{13}C-^{1}H shift pair in thymol for a given carbon nucleus. **(D)** Marginal individual assignment probabilities of unique directly bonded CH pairs and of pairs of topologically equivalent CH pairs (insert) in thymol. **(E)** Marginal individual assignment probabilities of unique carbons and of pairs of topologically equivalent carbons (insert) in thymol using only ^{13}C chemical shift distributions. In **(D)** and **(E)**, the dots indicate the experimentally determined correct assignment.

Figure 3-3 depicts the probabilistic assignment of bonded ^{13}C-^{1}H chemical shifts of thymol using two-dimensional correlated statistical shift distributions. The pair of topologically equivalent bonded C-H groups (labelled 9 and 10) was assigned to a pair of experimental shifts in **Figure 3-3D** as the disambiguation of topologically equivalent nuclei cannot be performed from the two-dimensional representation of a molecule. As seen in **Figure 3-3B**, the assignment of the carbon labelled 8 would have been much more ambiguous using only ^{13}C chemical shifts. Indeed, the probability of assigning carbon 8 to chemical shift *e* is 34% using only statistical distributions of ^{13}C chemical shifts (**Figure 3-3E**), and 100% using correlated statistical distributions of ^{1}H and ^{13}C chemical shifts (**Figure 3-3D**). We note that the most probable assignments of carbons 6 and 7 and of the methyl groups 1, 9 and 10 do not match

the experimentally determined ones. We attribute these discrepancies to substantial overlap between the corresponding statistical distributions of chemical shifts, that arise because of similar local bonding environments of carbons 6 and 7, and of methyl groups.

In addition to HETCOR, spectral editing methods are also straightforward high-sensitivity experiments that can be performed routinely to aid assignment. Such experiments are able to separate ^{13}C chemical shifts according to the number of bonded protons (multiplicity) [293-296]. The method can thus be directly applied to the statistical assignment framework presented here in order to break down the statistical assignment problem into smaller sub-problems of reduced complexity. This is especially useful when considering molecules yielding substantial overlap of statistical distributions. Knowledge of the multiplicity of ^{13}C chemical shifts can also be used to select a subset of HETCOR peaks to assign.
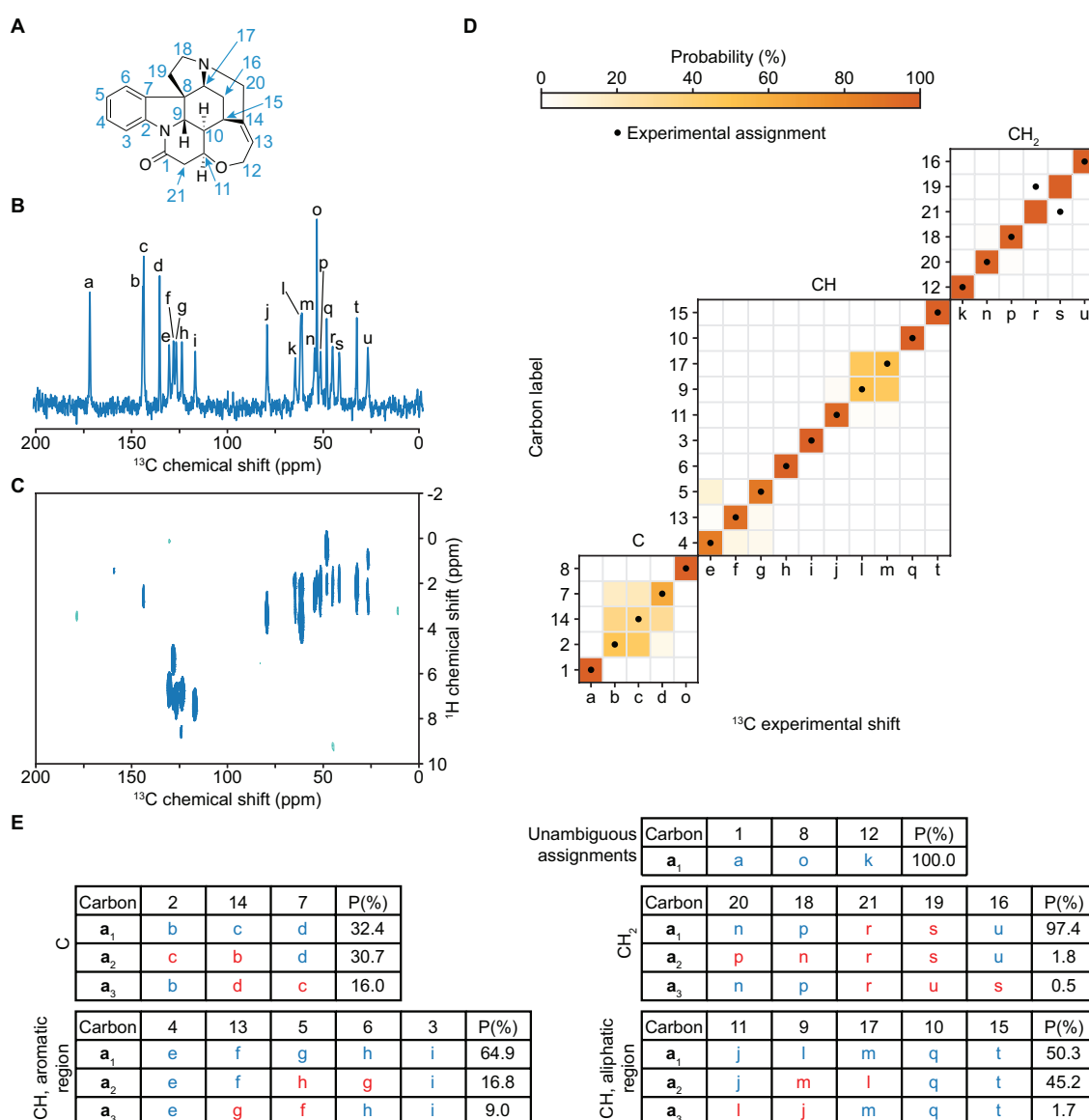
Figure 3-4 Probabilistic assignment of strychnine using simulated HETCOR and spectral editing. (A) Carbon labelling scheme of strychnine. **(B)** 125 MHz ^{13}C CPMAS NMR spectrum of strychnine. **(C)** ^{1}H-^{13}C HETCOR spectrum of strychnine. **(D)** Marginal individual assignment probabilities of the carbon nuclei of strychnine. The carbon multiplicity is indicated above

each probability map. The HETCOR shifts were used to assign CH and CH_2 carbons. The shifts are labelled alphabetically in order of decreasing chemical shift. The three most probable global assignments for the different blocks assigned individually along with their probability. The individual assignments making up the global assignments are indicated in blue if they correspond to the experimentally determined assignment, and in red otherwise. Carbons 1, 8 and 12 were assigned without ambiguity (P = 100%) directly from the evaluation of their statistical distributions of chemical shifts on the observed shifts.

Figure 3-4 shows the assignment of ^{13}C and ^{1}H-^{13}C chemical shifts of strychnine using the combination of spectral editing and correlated statistical distributions of chemical shifts. In **Figure 3-4D**, the chemical shifts of carbons without any proton attached were assigned using the one-dimensional ^{13}C chemical shift distributions of the associated nuclei. Carbons with a single bonded proton were assigned using the correlated ^{1}H-^{13}C statistical chemical shift distributions. The carbons with two attached protons were assigned to pairs of correlated ^{1}H-^{13}C chemical shifts, restricting the ^{13}C shift to be unique in each pair.

Figure 3-4E summarises the three most probable global assignments of strychnine. For each assignment, the global assignment is broken down into blocks by multiplicity, and then potentially into sub-blocks where there is no significant probability of overlap according to a threshold (here a factor 100 with respect to the highest probability for each nucleus). For each sub-assignment there is an associated probability. The most probable assignment of each block was found to match the experimentally determined one, except for the assignment of CH_2 groups, where the assignments of carbons 21 and 19 are swapped compared to the experimentally determined assignment. This is due to the large difference between the distribution of chemical shifts and experimental shift of carbon 19, which could come from an unusual intermolecular environment of that atomic site in the crystal structure.

We consider that a reliable assignment is difficult to extract from the set of global assignments and associated probabilities, especially in cases with a large number of overlapping distributions and shifts, which yield a lot of possible global assignments. Marginalization helps simplify the analysis of global assignments and identify ambiguities more easily. This can be seen in **Figure 3-4D**, where the assignment of carbon 7 to shift d is favoured compared to shifts b and c, which suggests only a pairwise uncertainty between carbon 2 and 14.

In addition to strychnine, shown in **Figure 3-4**, the marginal individual assignment probabilities obtained for a set of 9 selected molecules with complete experimental assignments (except for the two phenyl rings of ritonavir) using spectral editing and correlated ^{1}H-^{13}C statistical chemical shift distributions are shown in **Figure 3-5**. The assignment of carbon nuclei without any attached proton were obtained from the one-dimensional statistical distributions of ^{13}C chemical shifts. Notably, the assignment of lisinopril was found to be possible even when omitting the water molecules present in the crystal structure.
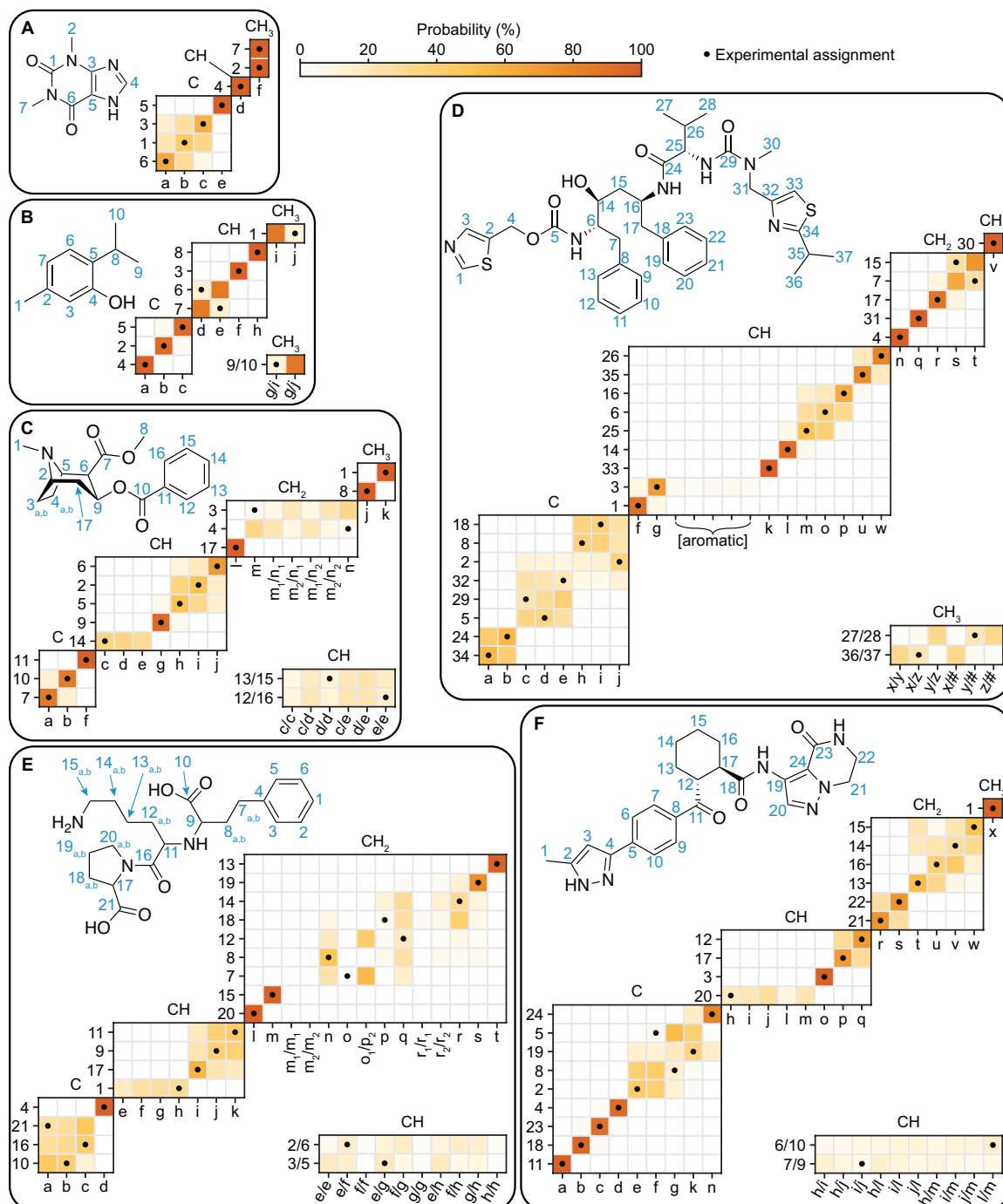
Figure 3-5 Probabilistic assignment of six organic crystals. Marginal individual assignment probabilities of ^{13}C chemical shifts of **(A)** Theophylline, **(B)** thymol, **(C)** cocaine, **(D)** ritonavir, **(E)** lisinopril and **(F)** AZD5718 using correlated ^{1}H-^{13}C chemical shift distributions and spectral editing. For each probability map, labels along the vertical axis indicate nuclei, and labels along the horizontal axis denote experimental shifts labelled alphabetically in order of decreasing ^{13}C shift. The carbon multiplicity is indicated above each marginal assignment probability map. In **(D)**, the assignment of carbons 9-13 and 19-23 is not shown as their experimental assignment is ambiguous. Nevertheless, the associated peaks were considered during the assignment process.
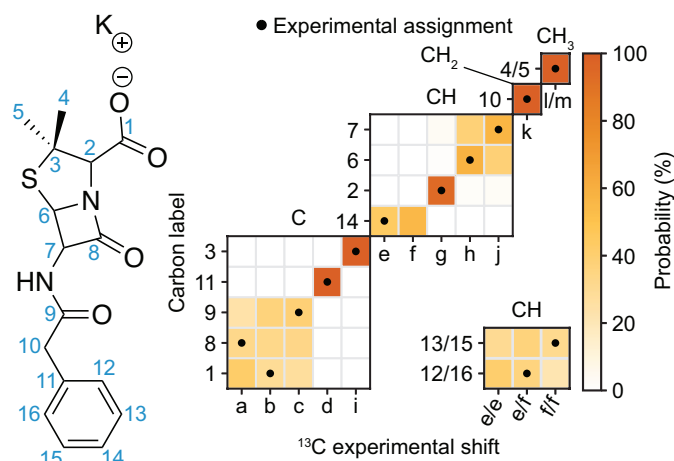
Figure 3-6 Probabilistic assignment of the K salt of penicillin G. Carbon labelling scheme and marginal individual assignment probabilities of **t**he K salt of penicillin G. The shifts are labelled *a* through *m* in order of decreasing chemical shift.

Figure 3-6 shows the assignment of the K salt of penicillin G. Only the organic ion was considered to construct the graph descriptors used to extract statistical distributions of chemical shifts. As for the presence of the water molecule in the case of lisinopril above, here the presence of the potassium ion, which is absent from the database, did not lead to a significant decrease in the ability of the model to predict the assignment, highlighting its generality beyond molecules for which chemical shifts can be computed by ShiftML. While ShiftML would not be able to compute shifts for crystals where even only one atom is different from C, H, N, O and S, this model only requires the molecule to be assigned to only contain these elements in order to obtain the probabilistic assignment. Of course, if the additional component in a salt or a co-crystal were to lead to a very different crystalline environment from those included in the database, this might lead to poor performance of the probabilistic assignment.

The marginal individual assignment probabilities obtained directly from the two-dimensional representation of the molecules were found to match the experimentally determined assignment in most cases. We observe that assignment ambiguities generally involve pairs or triplets of nuclei and shifts, leaving only a few possibilities for the NMR spectroscopist to further investigate in order to obtain the complete chemical shift assignment. Out of the 178 experimental individual assignments considered in **Figure 3-4**, **Figure 3-5**, **Figure 3-6** and **S16-S18**, only eight were associated with a probability below 10%, and two below 1%. These low probabilities were generally associated with crowded regions in experimental spectra, or with statistical outlier shifts compared to the distributions, which could have originated from unusual intermolecular environments.

In order to validate these results in a statistically significant manner, we evaluated the performance of the framework presented here on a benchmark set of a hundred crystal structures having between 10 and 20 different carbon atoms, randomly selected from the CSD database. In total, this corresponds to 1214 inequivalent carbon atoms. We used the ShiftML predicted shifts for each atom as the correct assignment, and excluded those shifts from the statistical distributions used to assign the molecules. The benchmark set was separated into five subsets containing 20 structures each that were evaluated independently

in order to obtain standard deviations. Although using shifts predicted by ShiftML may introduce a bias as the same method was used to construct the database of shifts, we assumed that the Gaussian width used to construct the statistical distributions of chemical shifts as well as the exclusion of the shifts assigned from the sets of shifts used to construct those distributions mitigate this issue.

Figure 3-7 summarizes the performance of the probabilistic assignment model on the experimental (**Figure 3-7A**) and synthetic (**Figure 3-7B**) sets of molecules selected. The use of spectral editing and correlated ^{1}H-^{13}C chemical shift distributions was found to improve the ability of the model to correctly assign carbon chemical shifts. Using either two-dimensional statistical distributions of chemical shifts, spectral editing, or combining both led to the experimental assignment being among the two most probable marginal assignments in over 80% of cases. Overall, the performances on the experimental benchmark set were consistent with the synthetic benchmark set, except when using spectral editing where a slight improvement on the experimental set compared to the synthetic set was observed.
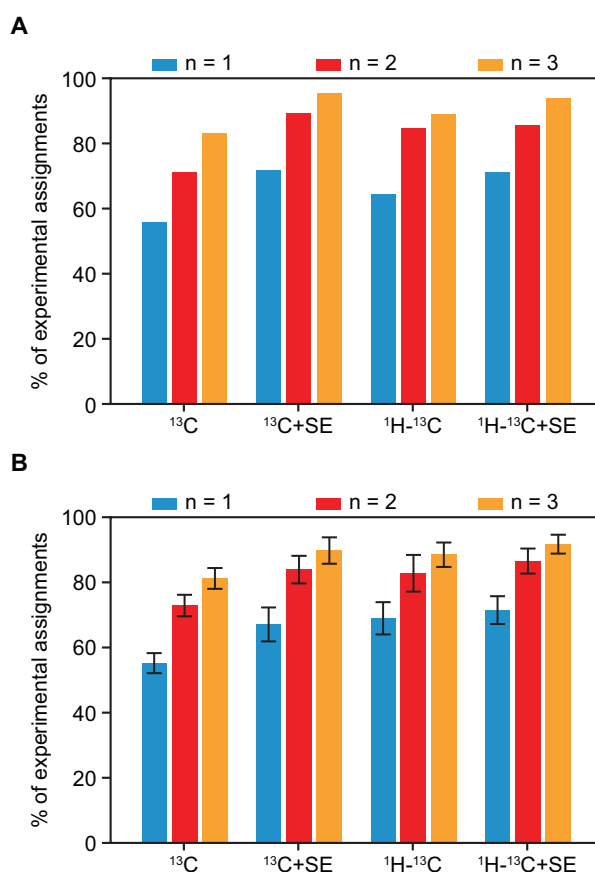
Figure 3-7 Model performances. Comparison of probabilistic assignment performances using one-dimensional (^{13}C) or two-dimensional (^{1}H-^{13}C) statistical distributions, and including spectral editing (SE). Proportion of the experimental assignments being within the n (n = 1, 2, 3) most probable marginal individual assignments in the **(A)** experimental and **(B)** synthetic benchmark sets of molecules. Error bars indicate the standard deviation over the five subsets making up the synthetic benchmark set.

3.1.3 Discussion

The framework presented here allows chemical shift assignment of organic crystals directly from their two-dimensional structure. This was achieved through the chemical shift prediction for over 200,000 organic crystal structures, which yields statistical distributions of chemical shifts for given covalent environments. A Bayesian framework was then used to obtain probabilistic marginal assignments of individual nuclei from the probabilities of the set of global assignments generated. Overall, using correlated ^{1}H-^{13}C chemical shift distributions in tandem with spectral editing, the method was found to include the experimental assignment among the two most probable marginal assignments in more than 80% of cases.

Furthermore, in most cases any ambiguity is found in small subgroups of shifts. This is highlighted in lisinopril in for example the CH_2 carbons because of significant overlap between the corresponding statistical distributions of chemical shifts and due to similar experimental shifts.

In summary, the approach presented here can provide marginal assignments based only on the two-dimensional molecular structure, where typically most of the resonances will be assigned with high probabilities, and only a few resonances will show ambiguities among doubles or triples that can then be the subject of targeted experiments for disambiguation if needed, or left unresolved and assigned such that the error is minimized when compared with computed shifts for model structures (e.g. when performing NMR-driven crystal structure determination). This can greatly accelerate the assignment process. In particular the method is shown to provide assignments for molecules such as strychnine, lisinopril, AZD5718 and ritonavir, which have crowded ^{13}C spectra with between 20 and 40 distinct carbons, and which would have been previously completely unaddressable without resorting to natural abundance ^{13}C-^{13}C correlations. For example, in strychnine, of the 21 carbons, 14 are correctly assigned with more than 75% confidence. The model was also successfully applied to the assignment of a hydrate and an organic salt, with no significant performance loss compared to the benchmark set. We expect that a more accurate model of chemical shifts could lead to improved probabilistic assignment through the framework presented here.

The method shown here is not restricted to ^{1}H and ^{13}C, and can be used to assign the isotropic shifts of any NMR-active isotope of hydrogen, carbon, nitrogen and oxygen in principle.

The code is publicly available at https://github.com/manucordova/ProbAsn and a user guide is available in the supplementary text as well as on the Github webpage. A suggested workflow to assign an organic solid is also described in the supplementary text.

3.1.4 Materials and Methods

NMR spectroscopy

The sample of strychnine and ritonavir form II were purchased from Sigma-Aldrich and Tokyo Chemical Industry, respectively. Experiments were performed on Bruker Ascend 400 and Ascend 500 wide-bore Avance III, and 900 US^2 wide-bore Avance Neo NMR spectrometers. The spectrometers operate at ^{1}H Larmor frequencies of 400, 500 and 900 MHz respectively, and are equipped with H/X/Y 3.2 mm, H/X/Y 4.0 mm, H/C/N/D 1.3 mm and H/C/N 0.7 mm CPMAS probes.

1D ^{1}H MAS NMR spectra were recorded at a temperature of 298 K using rotor spinning rates (v_r) up to 111 kHz. 1D ^{13}C cross-polarization (CP) [261] MAS NMR spectra were acquired at 298 K with v_r of 12.5 and 22 kHz for ritonavir and strychnine respectively. During the signal acquisition SPINAL-64 decoupling [29] was applied with a ^{1}H rf field amplitude of 100 kHz. For ritonavir spectral editing experiments were used to distinguish carbons with different numbers of protons attached to them. To selectively remove quaternary carbons a 1D version of MAS-J-HSQC [297] was used, to remove quaternary and primary carbons a double quantum filter was added to the MAS-J-HSQC [297] sequence and to remove primary and secondary carbons a simple CP experiment with an inserted delay of 0.5 ms before acquisition and after the CP pulse was applied [293].

2D ^{1}H-^{13}C HETCOR experiments were carried out at 298 K using v_r = 22 kHz. During t_1 100 kHz eDUMBO-1$_{22}$ was applied to decouple the ^{1}H-^{1}H dipolar coupling [23], and during t_2 100 kHz SPINAL-64 decoupling was applied.

The natural abundance 2D ^{13}C-^{13}C refocused INADEQUATE [18, 268] spectra required for the direct experimental assignment for ritonavir and strychnine were acquired using a Bruker 400 MHz Ascend NMR spectrometer. The probe was configured into ^{1}H/^{13}C double resonance mode. Variable amplitude cross-polarization [262] was used to transfer polarization from ^{1}H to ^{13}C. SPINAL-64 [29] heteronuclear ^{1}H decoupling with RF fields of 100 kHz was applied in all cases. The temperature of the sample for ritonavir was 250 K and a 4 mm rotor was used with a spinning frequency of 12.5 kHz. 2 x 120h experiments were acquired and combined in post processing to obtain the final spectrum (total time: 10 days). For strychnine DNP was used [298]. The sample was impregnated with 10 mM AMUPOL dissolved in 60:30:10 glycerol-d8:D$_2$O:H$_2$O. The spectrometer is equipped with a low temperature magic angle spinning (LTMAS) 3.2 mm probe and connected through a corrugated waveguide to a 263 GHz gyrotron capable of outputting ca. 5-10 W of continuous wave microwaves [155]. The sweep coil of the main magnetic field was optimized so that the microwave irradiation gave the maximum positive proton DNP enhancement with binitroxide cross effect-based polarizing agents (e.g. AMUPOL, TEKPOL). The temperature of the sample for ritonavir was 92 K and a 3.2 mm rotor was used with a spinning frequency of 12.5 kHz. A DNP enhancement of 36 was determined based on the ratio of the area of the spectra acquired with and without microwave irradiation. The DNP enhanced natural abundance 2D ^{13}C-^{13}C refocused INADEQUATE experiment [298] was run for 45 hours.

All chemical shifts were referenced via alanine.

Selection of crystal structures

The structures used to construct the chemical shift database were obtained from the CSD [299]. Only the organic crystal structures suitable for chemical shift predictions were selected. The corresponding selection criteria were that every structure must only contain C, H, N, O and S atoms, and that the disorder is resolvable. Missing protons were added automatically using the tool built into the CSD Python API. In total, 205,069 valid structures were selected.

Relaxation and chemical shift prediction

Because proton positions in published single-crystal X-ray diffraction structures may not correspond to the actual hydrogen positions in the crystals, they have to be optimized. Due to the large number of structures selected, DFT relaxation would be prohibitively costly. The semiempirical DFTB method [248] was thus chosen to relax proton positions in all structures. The structures were optimized at the DFTB3-D3H5 level of theory [250, 252] using the 3ob-3-1 parameter set [251, 300]. Further computational details are given in the supplementary text. Instances where the structure relaxation failed were discarded. 203,303 structures were successfully relaxed and considered for chemical shift prediction.

All chemical shift predictions were performed using ShiftML version 1.2 (publicly available at https://shiftml.epfl.ch) [275, 301]. Conversions of predicted shieldings to chemical shifts were performed by least squares fitting of the shieldings obtained for benchmark sets of DFTB-relaxed structures to their experimental chemical shifts, fixing the slope to a value of -1. The offsets obtained were found to be 30.96 ppm for ^{1}H, 168.64 ppm for ^{13}C, 185.99 for ^{15}N and 205.08 for ^{17}O. This corresponds to ^{1}H and ^{13}C shifts relative to TMS, ^{15}N shifts relative to NH_4Cl, and ^{17}O shifts relative to liquid H_2O. The sets of structures and isotropic chemical shifts used to determine shielding-to-shift conversions are described in **Tables S5-S8**. We note that chemical shieldings are stored in the database, and converted to chemical shifts on-the-fly during the construction of chemical shift distributions. In total, the database contains 5,243,129 unique ^{1}H, 4,847,864 unique ^{13}C, 466,370 unique ^{15}N and 867,446 unique ^{17}O chemical shifts, respectively.

Molecular fragment descriptors

For assignment of the spectrum of a molecule of unknown structure, classification of the predicted shifts should be done such that a statistical distribution of chemical shifts can be obtained for any nucleus from the two-dimensional representation of a molecule. The molecular fragment descriptor should thus not contain any information about conformation or molecular packing in the crystal structures. Among the topological atom-centered descriptors that fit these requirements [302-304], we chose to represent topological atomic environments by graphs where vertices represent atoms and edges represent covalent connectivities. The vertices were labelled by element, and the edges were kept unlabelled. Graphs were cut to a maximum depth w of 6, defined as the maximum shortest path between the central vertex (for which the chemical shift is predicted) and any other vertex in the path.

Conversion of the three-dimensional crystal structures to their corresponding graphs was performed by identifying atom pairs as covalently bonded when the distance between the atoms in the pair is less than 1.1 times the sum of the covalent radii of the atoms involved.

Database construction and search

A given topological atomic environment can be searched by identifying which graphs in the database match the graph of the selected atomic environment. However, there is no known algorithm able to solve the graph isomorphism problem required for each database entry in polynomial time [305-306]. Thus, the search was simplified by using the Weisfeiler-Lehman hash [307] as a unique graph identifier. If the number of instances of a given atomic environment identified in the database was deemed too small to produce statistically significant chemical

shift distributions, the atomic environment was searched again after reducing the graph depth. For this work, we chose a minimum number of instances of 10.

Construction of probability distributions

We use a notation and a conceptual framework extending the Bayesian selection of crystal-structure prediction candidate structures compatible with measured shifts [275]. From the set of chemical shifts and uncertainties $\{y_k, \sigma_k\}$ predicted by ShiftML for the CSD structures that share the same graph G_i as the atom i in the molecule of interest, we define the probability of observing a chemical shift y for that atom as proportional to the sum of Gaussian functions centered on each predicted shift y_k and with a width σ_k given by its prediction uncertainty.

$$p_i(y) \propto \sum_{k \in G_i} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{(y - y_k)^2}{2\sigma_k^2}\right] \qquad 3.1$$

Similarly, we define the probability of observing a cross-peak $(y^{(1)}, y^{(2)})$ for a pair of bonded atoms (i, j) in a molecule as proportional to the sum of uncorrelated two-dimensional Gaussian functions,

$$p_{ij}(y^{(1)}, y^{(2)}) \propto \sum_{k \in G_{ij}} \frac{1}{2\pi\sigma_k^{(1)}\sigma_k^{(2)}} \exp\left[-\frac{\left(y^{(1)} - y_k^{(1)}\right)^2}{2\sigma_k^{(1)^2}} - \frac{\left(y^{(2)} - y_k^{(2)}\right)^2}{2\sigma_k^{(2)^2}}\right] \qquad 3.2$$

where $\{y_k^{(1)}, \sigma_k^{(1)}\}$ and $\{y_k^{(2)}, \sigma_k^{(2)}\}$ are the sets of chemical shifts and predicted uncertainties computed for all the bonded atoms in the reference dataset that share the same graph G_{ij} as the pair being considered.

Probabilistic assignment

Considering the vector of observed shifts **y**, the probability that one of its elements y_j originates from atom i is obtained by evaluating equation 3.1 (or 3.2) for all elements in **y**.

$$p(y_j|i) = \frac{p_i(y_j)}{\sum_k p_i(y_k)} \qquad 3.3$$

For a given assignment **a** (defined as the vector mapping atoms in the molecule to experimental shifts such that $a_i = j$ if atom i is assigned to shift j) the probability of observing a vector of chemical shifts **y** is given by

$$p(\mathbf{y}|\mathbf{a}) = \prod_i p(y_{a_i}|i) \qquad 3.4$$

Applying Bayes theorem on equation 3.4 yields the probability of an assignment **a** given the observed vector of shifts **y**.

$$p(\mathbf{a}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{a})p(\mathbf{a})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{a})p(\mathbf{a})}{\sum_{\mathbf{a}'} p(\mathbf{y}|\mathbf{a}')\,p(\mathbf{a}')} \qquad\qquad 3.5$$

In equation 3.5, we assume that $p(\mathbf{a})$ is a non-zero constant if the assignment is valid (i.e., if all nuclei are assigned to only one chemical shift, and if all observed shifts are assigned at least one nucleus), and zero otherwise. Whenever some of the assignments can be made according to experimental data or heuristic arguments, such prior information can be incorporated in the definition through $p(\mathbf{a})$. By combining individual assignments, the complete set of possible global assignments can be generated. Because of the combinatorial complexity of generating all possible global assignments, several procedures were implemented to reduce the global assignment generation cost while ensuring that the most probable assignments are generated, and these are described in the supplementary text. Note that if the probability of any shift originating from a given nucleus is lower by a set threshold (typically a factor of 100) than the maximum probability for that nucleus then it is discarded. This results in some nuclei being assigned unambiguously independently of the rest of the global assignment (e.g. shift "e" in **Figure 3-2**).

Eq. 3.5 assigns a distinct probability to each possible assignment of the entries of the measured shifts vector **y** to *all* the environments. It is the correct probabilistic metric to compare two assignments but is hard to interpret. A more compact indicator is given by the marginal probability that atom *i* is assigned to shift *j*, which can be extracted from the set of generated assignments by considering only the vectors **a** containing that particular individual assignment. This is shown in equation 3.6 by the Kronecker delta δ_{a_ij} which selects the assignments for which $a_i = j$

$$p(a_i = j|\mathbf{y}) = \frac{\sum_{\mathbf{a}} \delta_{a_ij}\, p(\mathbf{a}|\mathbf{y})}{\sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{y})} \qquad\qquad 3.6$$

For topologically equivalent nuclei, which have identical graphs and probability distributions, tuples of nuclei were assigned to tuples of experimental shifts (which can be partly or entirely identical).

Synthetic benchmark set

A set of 100 randomly selected crystal structures from the database were selected to benchmark the probabilistic assignment. The selection was restricted to crystals having between 10 and 20 unique carbon atoms. The selected structures are listed in the supplementary text. The ShiftML predicted shifts associated to each nucleus were used as ground-truth assignment. The structure to assign was systematically excluded from the database search performed to construct statistical distributions of chemical shifts. The synthetic benchmark set was separated into five sets containing 20 crystals each and 241, 260, 212, 259 and 242 unique carbon atoms, respectively.

3.1.5 Supplementary
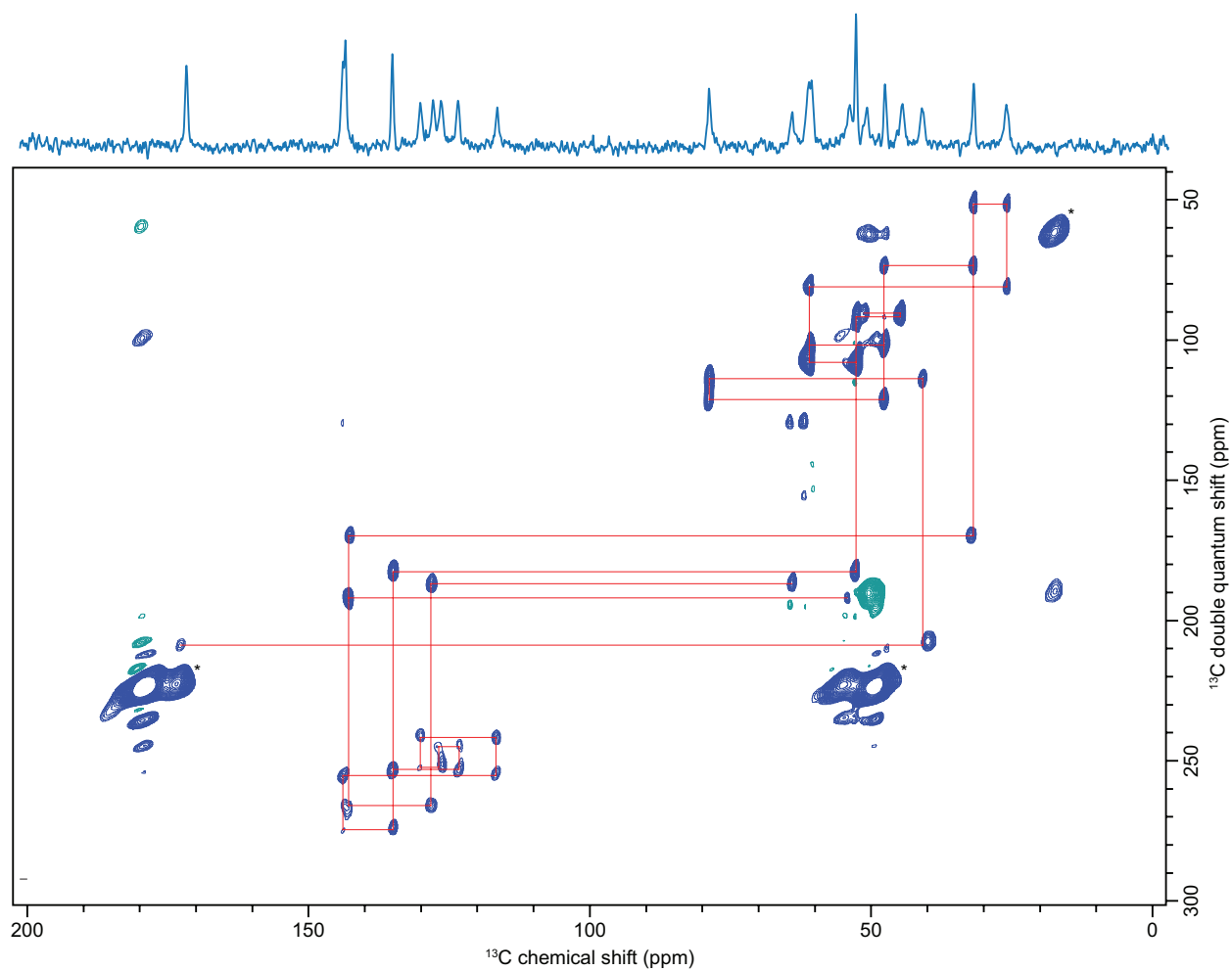
Figure 3-8 DNP-enhanced ^{13}C-^{13}C refocused INADEQUATE spectrum of strychnine. The peaks indicated with a * are assigned to impurities introduced during sample preparation for DNP NMR.
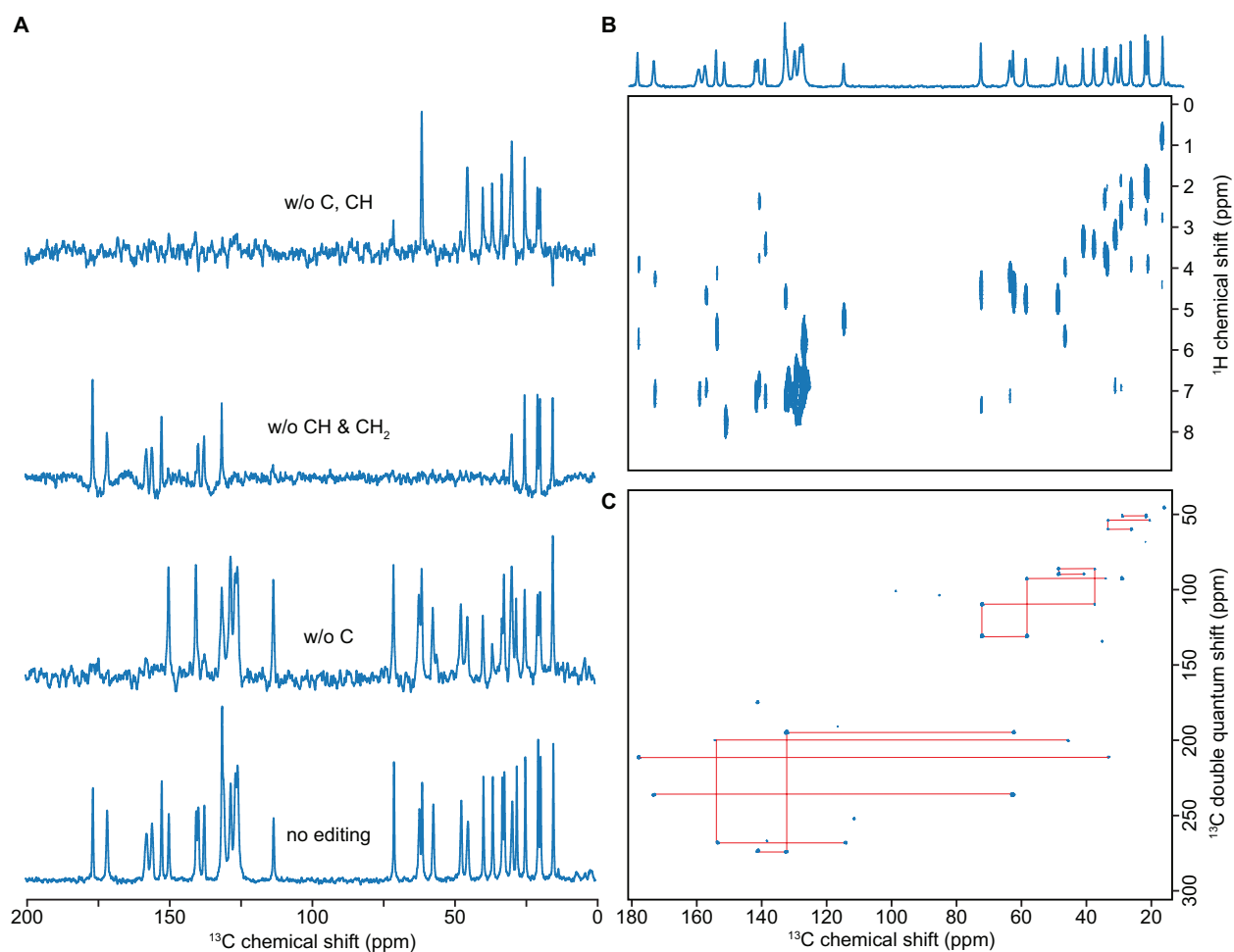
Figure 3-9 (A) Spectral editing of the ^{13}C CPMAS spectrum of ritonavir. **(B)** ^{1}H-^{13}C HETCOR spectrum of ritonavir. **(C)** ^{13}C-^{13}C refocused INADEQUATE spectrum of ritonavir.

	^{1}H	^{13}C	^{13}C w/o primary and secondary	^{13}C w/o quartenary	^{13}C w/o quartenary and primary
MAS rate	22 kHz	12.5 kHz	12.5 kHz	12.5 kHz	12.5 kHz
Recycle delay (d_1)	5 s	2 s	2 s	3s	3s
^{1}H to X CP					
Spin lock duration	-	3 ms	3 ms	3 ms	3 ms
Delay after acquisition for primary and secondary carbon filtering	-	-	0.5 ms	-	-
Total acquisition time	4 ms	30 ms	30 ms	30 ms	30 ms
Dwell time	1 µs	5 µs	5 µs	5 µs	5 µs

Number of points	4096	6144	6144	6144	6144
Number of scans	4	512	64	512	12288
Acquisition mode	qsim	qsim	qsim	qsim	qsim

Table 3-1 Experimental parameters for 1D experiments on ritonavir.

	^{1}H-^{13}C HETCOR	^{13}C-^{13}C INADEQUATE
MAS rate	22 kHz	12.5 kHz
Recycle delay (d_1)	2.7 s	2.15 s
^{1}H to X CP		
Spin lock duration	0.1 and 1.0 ms	3.5 ms
Acquisition in the indirect dimension (t_1)		
Total acquisition time	4.6 ms	2.6 ms
Dwell time	96 µs	20 µs
Number of points	96	256
Acquisition in the direct dimension (t_2)		
Total acquisition time	40.5 ms	25 ms
Dwell time	13.2 µs	5 µs
Number of points	256	2494
Number of scans per increment	16	1'536
Acquisition mode	States-TPPI	States-TPPI
Delay t	-	3.6 ms

Table 3-2 Experimental parameters for 2D experiments on ritonavir.

	^{1}H	^{13}C
MAS rate	111 kHz	22 kHz
Recycle delay (d_1)	26 s	120 s
^{1}H to X CP		
Spin lock duration	-	1 ms
Total acquisition time	8.4 ms	30 ms
Dwell time	2.8 µs	13.2 µs
Number of points	2998	2268
Number of scans	4	32
Acquisition mode	DQD	qsim

Table 3-3 Experimental parameters for 1D experiments on strychnine.

	^{1}H-^{13}C HETCOR	^{13}C-^{13}C INADEQUATE
MAS rate	22 kHz	12.5 kHz
Recycle delay (d$_1$)	120 s	5 s
^{1}H to X CP		
Spin lock duration	0.1 ms	3 ms
Acquisition in the indirect dimension (t$_1$)		
Total acquisition time	2.8 ms	2.6 ms
Dwell time	88 μs	20 μs
Number of points	64	256
Acquisition in the direct dimension (t$_2$)		
Total acquisition time	30 ms	15 ms
Dwell time	13.2 μs	5 μs
Number of points	2268	2988
Number of scans per increment	34	128
Acquisition mode	States-TPPI	States-TPPI
Delay t	-	4 ms

Table 3-4 Experimental parameters for 2D experiments on strychnine.

3.2 One database to rule them all

3.2.1 Introduction

With the advent of large-scale calculations there is a growing need for a solid-state NMR database for small organic molecules. While many other NMR fields, such as protein or solution state NMR already have large databases for many years, to our knowledge none exist that could be used by the NMR crystallography community. Therefore, here we establish a basis for a solid-state NMR database. We provide a structure and the initial concepts of the design which are crucial to start making the database itself.

3.2.2 Results

All of the results in this section are still a work in progress and reflect the current status of the database project.

Structure. The database has been built using MySQL 8.0.26. **Table 3-5** shows the created tables and their relationships to each other.
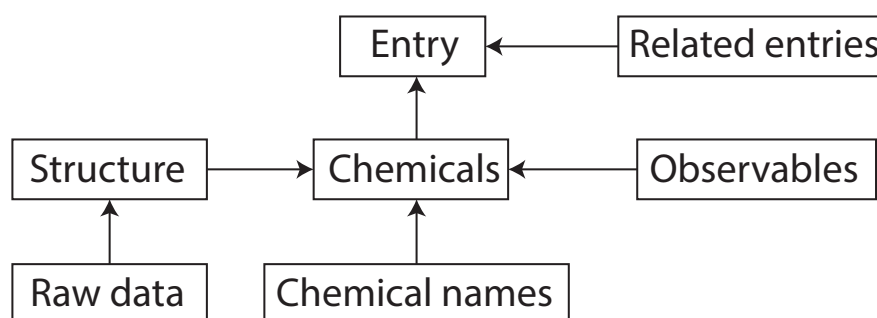
Table 3-5 The backbone structure of the database where each rectangle symbolizes a table with the table name. The arrows show how each of the tables are related to each other. For example, an arrow coming from "Related entries" table to "Entry" table means that each element in "Related entries" table will have a reference leading to a corresponding entry in the "Entry" table.

The structure of the Entry table and the corresponding Related entries tables are shown in **Table 3-6**. **Entry_id** is the unique number assigned to each entry. **Flag** is a string that can be used to warn if the entry has become outdated after a better-quality data has been uploaded. **Related entries** is a helper table containing references to similar entries.

Entry

Entry_id	Int
Flag	Varchar (default Null)

Related entries	
related_entry	Int
id_entry	Int

Table 3-6 The structure of the Entry and Related entries tables.

The structure of the **Chemicals** and **Chemical names** tables is shown in **Table 3-7**. The rows contain **Chemical Formula**, **Molecular Mass** and **Preferred IUPAC name**, being the main searchable parameters, including Chemical name that is in the **Chemical names** table and links to the **Chemicals** table. The **Chemicals** table links to the **Entry** table therefore allowing different chemicals to be associated with one entry.

Chemicals	
Chemical formula	Varchar
Molecular mass	Float
PIN (preferred IUPAC name)	Varchar
id_entry	Int

Chemical names	
Chemical_name	Int
id_chemical	Int

Table 3-7 The structure of the Chemicals and Chemical names tables.

The structure of the **Observables** table is shown in **Table 3-8**. **Observable_type** is a string describing the experiment from which the observables come from, e.g., 1D ^{13}C CP experiment, 2D HETCOR. **Nucleus_X**, where X = 1, 2 ,3 ,4 tells the nucleus that one is depositing in each of the dimensions, with 1D being required up to 4D. **Labels_X**, where X = 1, 2 ,3 ,4 are for the label input that will correspond to the **Observables_X**, where X = 1, 2 ,3 ,4 which is where one puts the corresponding observables, e.g. isotropic chemical shifts. **Temperature** is meant for the temperature at the time of the experiment. This table is linked to the **Chemicals** table.

Observables	
Observable_type	Varchar

Nucleus_1	Varchar
Nucleus_2 / 3 / 4 (multiple rows)	Varchar, default Null
Labels_1	Varchar
Labels_2 / 3 / 4 (multiple rows)	Varchar, default Null
Observables (e.g. isotropic shifts)	Varchar
Observables_2 / 3 / 4 (multiple rows)	Varchar, default Null
Temperature	Float
id_chemical	Int

Table 3-8 The structure of the Observables table.

The structure of the **Structure** table is shown in **Table 3-9**. This table is dedicated to the chemical structure data associated with the chemical. **Structure_type** describes the structure that is being upload, mol or cif formats being examples. **Publication** links to the publication associated with the structure if available. **Path** contains the link to the structure. **Refcode** is the code assigned by the Cambridge Structural Database if there is one. **Details** include any relevant information such as "this structure has been determined by X-ray / neutron diffraction". **Temperature** explicitly asks for the temperature where the structure has been determined if there is one.

Structure	
Structure_type	Varchar
Publication	Varchar, default Null
Path	Varchar, default Null
Refcode	Varchar, default Null
Details	Varchar, default Null
Temperature	Float, default Null
id_chemical	Int

Table 3-9 The structure of the Structure table

The structure of the **Raw_data** table is shown in **Table 3-10**. **Raw_data_type** describes the type of the data that is being uploaded, such as HETCOR or ^{13}C CP 1D experiment. **Publication** holds the doi of the publication. **Data_path** contains the link to the data. A note about the data - the data has been converted to the Jcamp-dx format[308] to make it accessible for everyone. **Details** contain any relevant information about the data, e.g., the strength of the

magnetic field at the moment of the acquisition, referencing, spinning rate. **Quality** is a short descriptor of the quality of the data given by us after inspecting the data. The table is linked to the **Observables** table.

Raw_data	
Raw_data_type	Varchar
Publication	Varchar, default Null
Data_path	Varchar
Details	Varchar
Quality	Varchar
id_observables	Int

Table 3-10 The structure of the **Raw_data** table.

Interface. With the underlying structure done, conceptual web page templates were created. The starting page is shown in the **Figure 3-10**. The main elements there are the **Search** and **Upload a structure** possibilities. To keep the design simple only the most important elements are included.

Figure 3-10 The starting page.

The basic search page is shown in **Figure 3-11**. It includes one theoretical search result and shows the common elements that associated with a structure, such as the its formula, names and also associated data.

Figure 3-11 The basic **Search** page.

The advanced search page is shown in **Figure 3-12**. It gives the user the possibility to search by different metrics separately or together and also by simply drawing the structure.

Figure 3-12 The **Advanced search** page

The page that shows up after clicking on one of the found results is shown in **Figure 3-13**. It includes options to look at all the experiments that are available for the particular upload, shows the NMR spectrum and the chemical structure, labeling, observables and the references to the publications if there are any.

Figure 3-13 The **Results**

The last part is the **Submit your experiment** pages that are shown in **Figure 3-14** and **Figure 3-15**.

Figure 3-14 The first **Submit Your Experiment** page

Figure 3-15 The second **Submit Your Experiment** page

3.2.3 Conclusions

A skeleton of a solid-state NMR database was created. While the shown structure and the design choices may change significantly over time while everything is implemented, we believe that what has been developed provides a strong foundation to build upon. We have made a significant step in the direction of the first solid-state NMR database for organic molecules.

3.2.4 Supplementary

Before building the internet database an extensive search of the literature has been done by F. Pickard and F. Racine as part of their BSc and MSc projects under the supervision of F. M. Paruzzo and I. Data shown in **Appendix B** is a compilation of most of the organic molecular crystals found in the current literature that have ^{13}C and/or ^{1}H chemical shifts assigned. It contains the name of the chemical, ^{13}C and/or ^{1}H chemical shifts if they have been published, labels and the reference to the publication reporting the data.

Chapter 4. Investigation of carbonate speciation by dissolution DNP

4.1.1 Introduction.

Formation of calcium carbonate (Ca_2CO_3) is of great importance in biomineralization, geosciences and industry. Ca_2CO_3 is one of the most prevalent minerals in the Earth's crust and it strongly affects the chemistry of ocean water.[309] Ca_2CO_3 is the primary precursor for cements (4 billion tonnes/year), and it is used in industrial applications such as a filler for polymers.[310] More topically, global warming is becoming more and more relevant[311] and CO_2 capture and conversion back to Ca_2CO_3, or other methods of conversion, is a topic of great current interest.[312-313] Understanding the mechanisms underlying the formation of Ca_2CO_3 is of key importance, as it would help to control and understand the formation of different polymorphs, particle sizes and morphologies.

The crystalline phases of Ca_2CO_3 are calcite, aragonite and vaterite. Two hydrated phases of $CaCO_3$, monohydrate ($CaCO_3 \cdot H_2O$) and hexahydrate ($CaCO_3 \cdot 6H_2O$) are also possible. In addition, amorphous calcium carbonate (ACC), which normally exists as monohydrate calcium carbonate[314], has been identified as a source for the development of calcium carbonate-based shells and can act as a precursor of the crystalline phases[315].

While the main reactions leading to the formation of Ca_2CO_3 look simple (eq. 4.1-4.3),

$$CO_2 + OH^- \leftrightharpoons HCO_3^- \qquad\qquad 4.1$$
$$HCO_3^- + OH^- \leftrightharpoons CO_3^{2-} + H_2O \qquad\qquad 4.2$$
$$CO_3^{2-} + Ca^{2+} \leftrightharpoons CaCO_{3(s)} \qquad\qquad 4.3$$
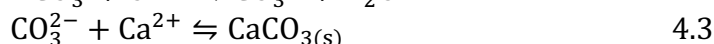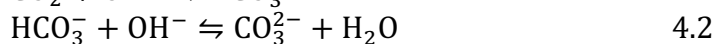
the nucleation processes are not fully understood even after 100 years of research. By definition nucleation is the local emergence of a distinct thermodynamic phase at the nanoscale that macroscopically grows in size with the attachment of growth units. Noteworthy is that the phase changes happen at a length scale of 10^{-10} m and time scales equivalent 10^{-13} s thus making the studies of nucleation very complicated.[316] Recently stable so-called prenucleation clusters (PNC)[317] were observed in the formation of Ca_2CO_3 (**Figure 4-1**). This challenged the classical nucleation and growth theory (CNT) that described the formation of solid Ca_2CO_3. During the formation of solid particles, the established classical models consider only monomeric building blocks such as atoms, ions or molecules. The potential role of stable oligomeric or polymeric species is mostly neglected.[318] That raised a question if an alternative theory is needed to explain the existence PNCs. Even after a substantial amount of work done both computationally and experimentally[318-323] the consensus is still not established and some authors even argue that there is no need for a new theory and the formation of PNCs doesn't contradict CNT.[324-327] But what is evident is that the formation and the subsequent transformation of PNCs are intimately connected with the forming of the final stable phase.
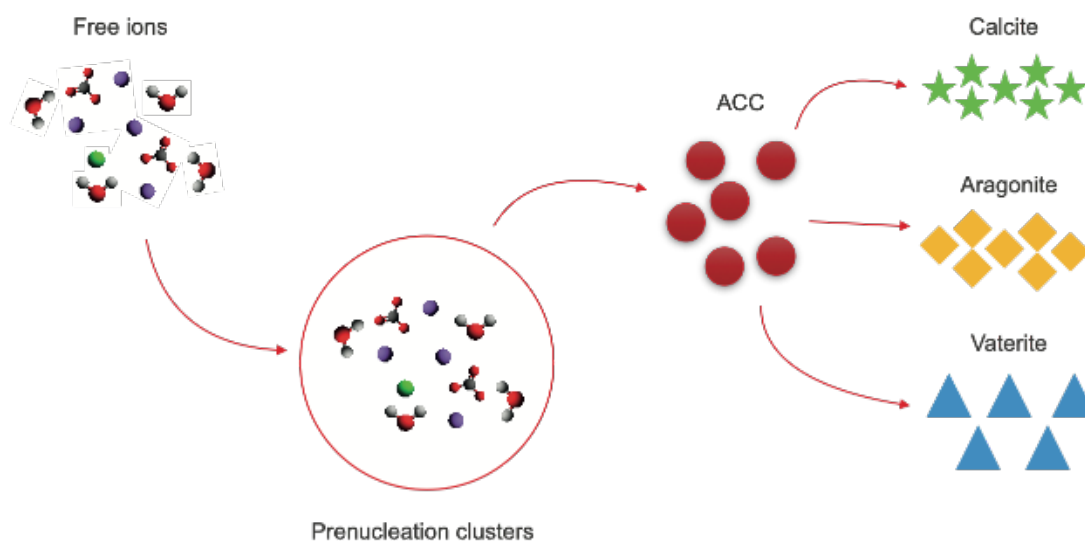
Figure 4-1 A scheme of calcium carbonate forming prenucleation clusters which then turn into amorphous calcium carbonate which is the precursor for other solid-state phases like calcite, aragonite and vaterite.

In this investigation our goal is not to give a clear answer on which school of thought the formation of calcium carbonate follows, but instead give a clearer vision of what happens in the first seconds of calcium carbonate formation. These species are present at low concentrations that are unreachable by traditional NMR techniques. For that we apply dissolution DNP as a method of choice to substantially improve the signal to noise ratio. We observe a complicated peak pattern in our spectra, that we attribute to PNCs, and we support this conclusion by the difference in the peak decay both for the peaks attributed to PNCs and the peak attributed to free carbonate ions.

4.1.2 Materials and methods.

The experimental setup. The home-built DNP setup closely follows the method and instrumentation developed by the Golman group.[328] See **Figure 4-2** for the scheme of the dissolution DNP system. A continuous flow He-4 cryostat **A** is inserted in a 6.7 T superconducting magnet **B** and the system is connected to a 100 L dewar **C** to reach the operating temperatures of 4 K and 1.2 K. The cryostat is connected to a pump **D** that creates a very low pressure in the outer chamber (<10^{-5} mbar) to keep the system cold. The dewar contains liquid helium at ambient pressure and is connected to the system via isolated transfer line **E**. The transfer line goes into a phase separator **F** that is connected to a pump **G**, a helium recuperation line and two capillary tubes which in turn are connected to the sample space **H**. The capillary tubes are operated via needle valves. When a needle valve is opened the liquid helium can reach the sample space. For that to happen the sample space is managed by a two-pump system that can create vacuum in the sample space <1 mbar. The sample space is also connected to a helium recovery line.

A probe is immersed in the sample space and a sample can be delivered to the end of it with a plastic tube. The upper part of the probe is connected to two tuning and matching boxes for each channel and a microwave source Elva VCOM-10/0.5/94/400 **J**. The microwave source can provide microwaves at a power up to 400 mW operating at a frequency 94 GHz. Between the microwave source and the probe is a frequency doubling device (Virginia Diode D200) that converts the frequency to 188 Ghz.

During the transfer between the DNP polariser and the NMR spectrometer **K** the hyperpolarized liquid will experience drastic losses of polarisation if exposed only to very low magnetic fields like the Earth's magnetic field. To preserve the polarisation a magnetic tunnel **L** connects the DNP polariser and the NMR spectrometer.[329] The tunnel consists of an assembly of permanent neodymium boron magnets secured by a several home-built aluminium structures. A 2.5 mm PTFE tube runs through it to carry the hyperpolarized fluid. The magnetic field at the center of the tunnel is greater than 0.9 T.

The hyperpolarized liquid is carried to the 11.7 Tesla high-resolution NMR magnet where the spectral acquisition happens.
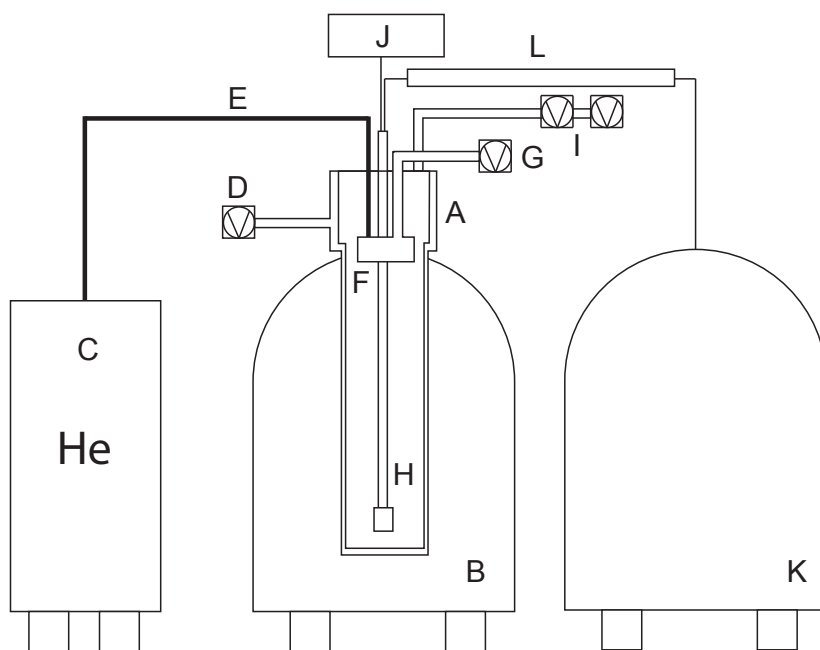
Figure 4-2. A simplified scheme of the dissolution DNP system. **A** is the cryostat, **B** is the 6.7 T superconducting magnet, **C** is the 100 L helium dewar, **D** is the pump connected to the outer vacuum chamber, **E** is the helium transfer line, **F** is the separator, **G** is the pump to transfer helium from to the separator, **H** is the sample space, **I** are the two pumps regulating pressure

in the sample space, **J** is the microwave source, **K** is the magnet where spectral acquisition happen, **L** is the magnetic tunnel.

The sample preparation. Carbon-13 labelled Na_2CO_3 was used. It was dissolved in 1:3:6 $H_2O:D_2O:$d8-glycerol solution with the final concentration of 1M Na_2CO_3. The solution contained 50 mM TEMPOL radical as the hyperpolarization agent. The buffer solutions used in the work were di-sodium tetraborate decahydrate (pH = 9.2), a mixture of glycine, sodium chloride, sodium carbonate (pH = 10.0) and diisopropylamine (pH = 11.0).

The sample hyperpolarization and dissolution. 100 μL of the sample was inserted in a sample holder made from PEEK (polyether ether ketone) and the sample holder in the cryostat and directly polarised for 2-3 hours while simultaneously following the polarisation by applying small tip angle pulses of 1 degree each 5 minutes (supplementary, **Figure 4-7**). To access higher polarisation the temperature was cooled down to 1.2K by lowering the internal pressure. Microwaves were provided by the microwave source at the output power of 400 mW. The working frequency was 188.250 MHz optimised by doing a frequency sweep. Before the dissolution a solution containing buffer (9.2 / 10.0 / 11 pH) and D_2O in 1:1 v/v proportion was superheated in a Teflon holder until the internal vapour pressure reached 10 bar. Then the dissolution was done by inserting a home-made dissolution stick inside the cryostat so that the end of it reaches the sample holder and the superheated solution was pushed through the sample holder and then to the 11.7 T magnet via the magnetic tunnel. At the end of the dissolution line was a tube containing 50 μL of the buffer solution with dissolved $CaCl_2$ with the calcium ion concentration being proportional to the carbonate ion concentration in the final solution. The used proportions were 0.5x, 1.0x and 2.0x the carbonate ion concentration.

Spectral acquisition.

A pseudo 2D ^{13}C detection experiment was used to acquire the spectra. 30-degree ^{13}C pulses were used with AQ of 0.3 sec and repetition delay of 0.2 s. The acquisition was started before the dissolution so it was possible to follow the reaction from the very start. Glycerol was present in all samples and used as an internal reference with the rightmost peak set to 63.0 ppm.

4.1.3 Results and discussion

We hyperpolarized a sample of Na_2CO_3 in a polarisation matrix and when sufficient polarisation was achieved, transferred it to a spectrometer. ^{13}C signal was detected every 0.5 s before and after the mixing to follow the reaction from the very start. **Figure 4-3** shows an example of acquiring an NMR spectrum of hyperpolarized Na_2CO_3 after dissolution. Peaks below 80 ppm belong to glycerol that is in the hyperpolarisation matrix. Peaks at 160.7 and 160.2 ppm belong to a glycerol-carbonate adduct that has been found in previous work by Santos.[330] The peak at 167.8 ppm belongs to free carbonate. The expansion of the relevant carbonate region shows an expected signal decay due to combined effect of spin relaxation and signal depletion of successive NMR experiments.
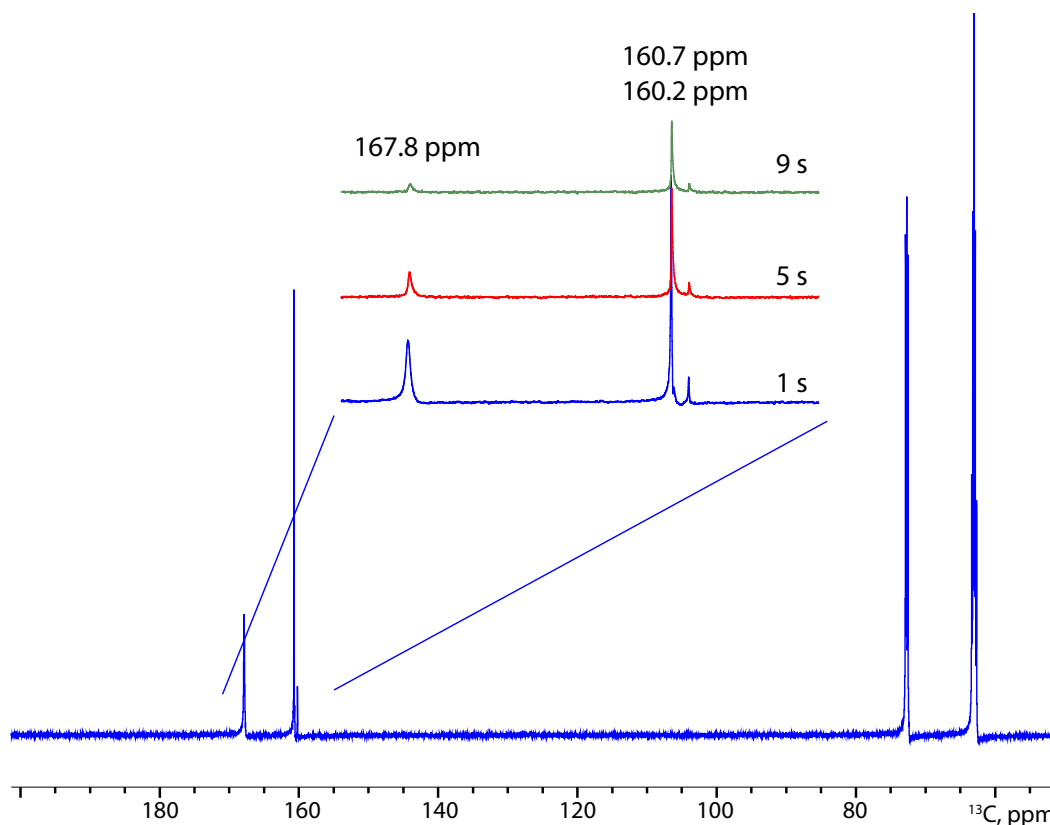
Figure 4-3 The spectrum of hyperpolarized 20 mM Na_2CO_3 solution in 9.2 pH buffer solution.

To observe the formation of the prenucleation clusters we added dissolved $CaCl_2$ in the tube that receives the dissolved Na_2CO_3 solution. We did three different ratios of the two species where the final concentration of Na_2CO_3 always stayed the same and the concentration of $CaCl_2$ was varied. The ratios were 0.5, 1.0 and 2.0 of the amount of Ca^{2+} ions in respect to CO_3^{2-} ions. Because pH can create more or less favourable effect when trying to observe the prenucleation clusters we tried three different conditions with pH 9.2, 10.0 and 11.0 done in a buffer.

We did not see any prenucleation cluster formation at pH = 9.2, but at pH = 10.0 and 11.0 we saw a set of new broad peaks that we ascribe to prenucleation clusters. **Figure 4-4b** shows a result of an experiment done at pH = 10.0 at the ratio of both reactants 1:1. We see a broad peak (peaks) in the region of 167.0-169.0 ppm that we ascribe to prenucleation clusters. The other set of peaks is in the region 166.2-165.5 ppm, which corresponds to the free carbonate ions. To support our assignment, we plotted the integrated peak area of both regions as a function of time, **Figure 4-4a**. After normalization it is clearly seen that the 166.2-165.5 ppm intensities decay rapidly, that we ascribe to the carbonate ion consumption in the reaction. We note that the peak positions slightly vary over time (**Figure 4-4b**) as a result of the mixing that is happening in the first few seconds of the reaction when the solutions come into contact.
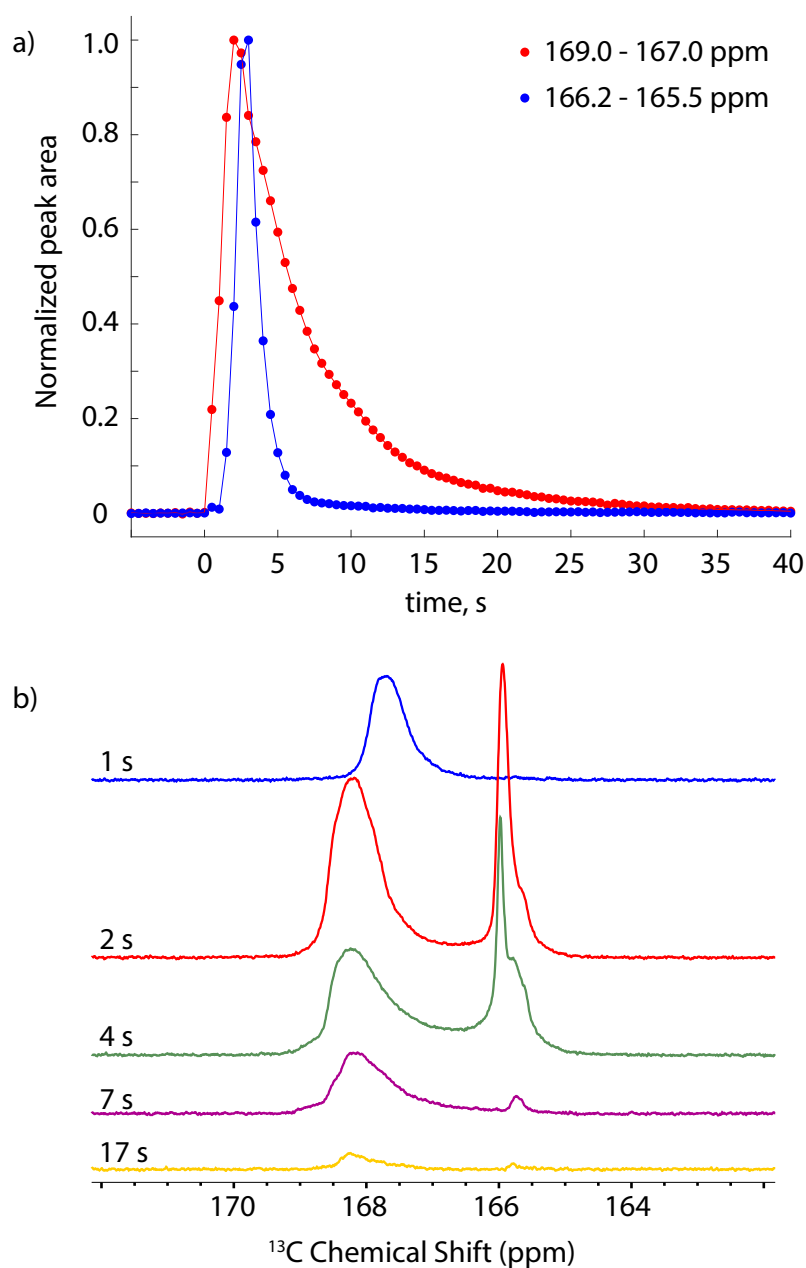
Figure 4-4 A dissolution experiment with Ca^{2+} and CO_3^{2-} concentrations in the reacting solutions being equal at pH = 10. a) Normalized intensity of the two regions measured at each scan. b) Evolution of the peaks over time at 1, 2, 4, 7 and 17 seconds.

Figure 4-5 and **Figure 4-6** show the rest of the experiments and the normalized intensities for other ratios of reactants at pH 10 and 11. As mentioned previously, the peak positions and shapes tend to vary from experiment to experiment as a result of the mixing processes but in all cases two separate peak regions are seen.
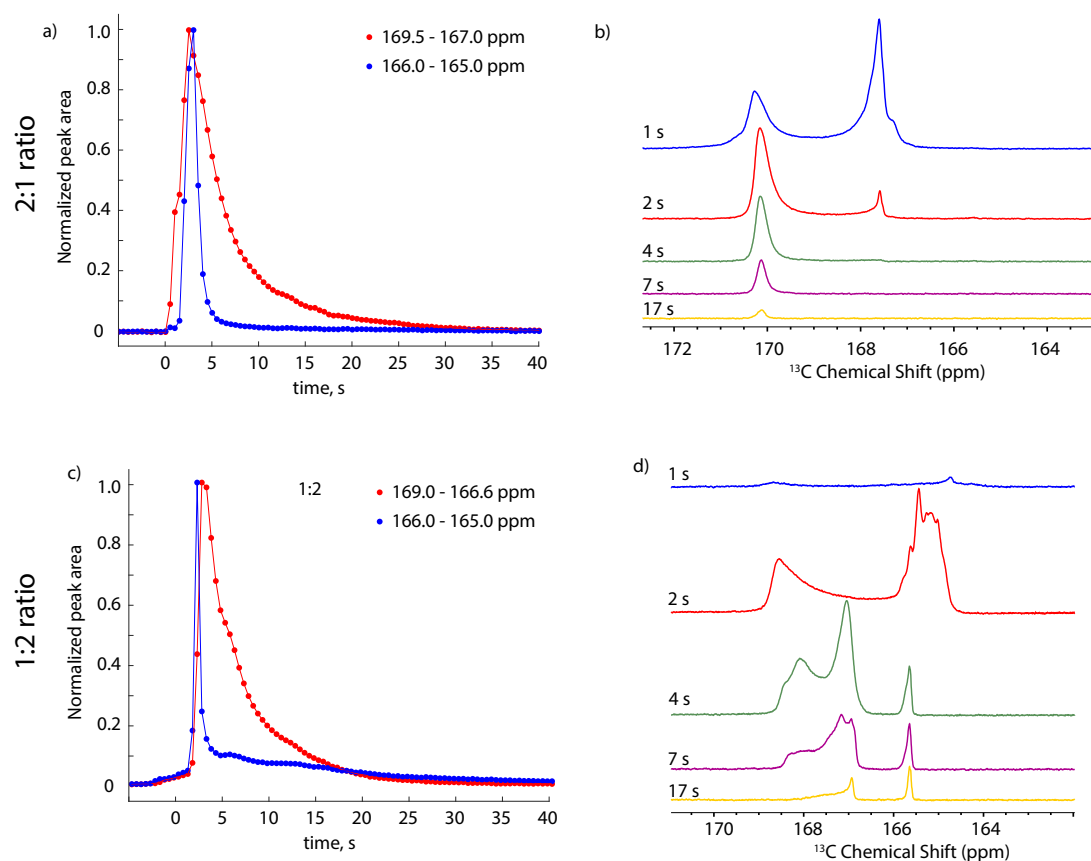
Figure 4-5 A dissolution experiment at ph = 10. Ca^{2+} and CO$_3^{2-}$ concentrations in the reacting solutions at ratios 2:1 (a and b) and 1:2 (c and d). a and c are the normalized intensities of two regions in the spectra shown in b and d.
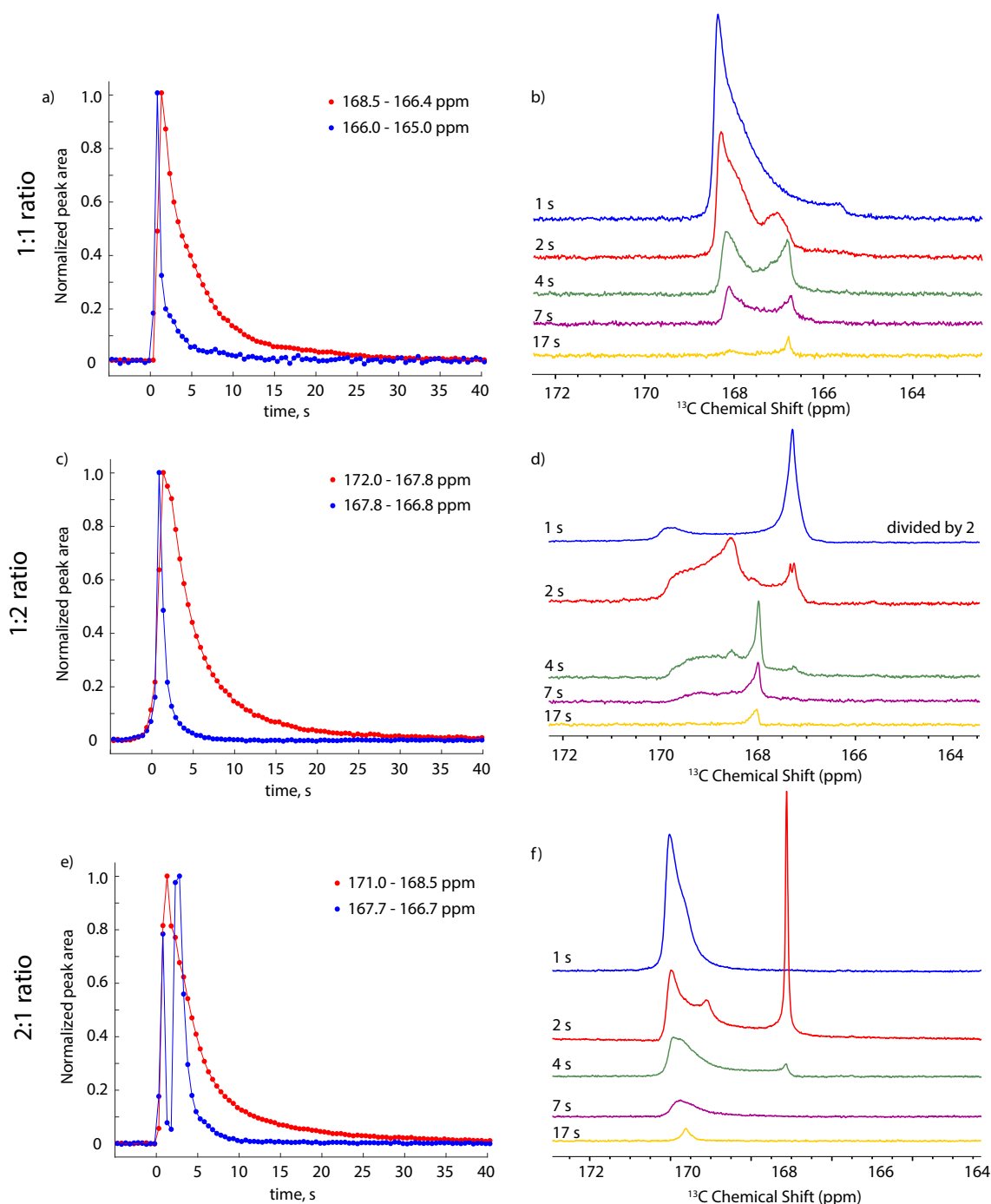
Figure 4-6 A dissolution experiment at ph = 11. Ca^{2+} and CO_3^{2-} concentrations in the reacting solutions at ratios 1:1 (a and b), 2:1 (c and d) and 1:2 (e and f). a, c and e are the normalized intensities of two regions in the spectra shown in b and d and f.

Technical issues

Unfortunately, many technical issues were experienced while working with the dissolution system that significantly delayed the scientific part of the work to the point where most of the time was spent fixing the system. This was influenced also by the nature of the system. To fix a problem the system needs to be brought to room temperature which takes at least a

few hours and to test it one needs to cool it down to 4K which usually takes a whole working day.

Here is a list of just a few major issues concerning the instrumentation that happened during my thesis. All four of the pumps connected to the system broke at one point and needed to be repaired or changed. The microwave source broke two times. The needle valves that connect the sample space and the separator broke a few times. The outer vacuum chamber experienced a micro fissure at the point of soldering resulting in problems with vacuum that took a long time to find and solve. The injection system experienced multiple issues that needed to be addressed. Fortunately, even with these issues we managed to acquire a series of experiments and deepen our understanding of both the dissolution DNP system and the chemistry of carbonates.

4.1.4 Conclusions

We have presented evidence of the existence for prenucleation clusters after mixing of sodium carbonate and calcium chloride solutions. This shows the possibility to investigate this and similar systems by using the dissolution DNP setup. We note that this would not have been possible by using other NMR methods due to the small concentration of the investigated constituents and the fast timescale of the reaction.

4.1.5 Supplementary
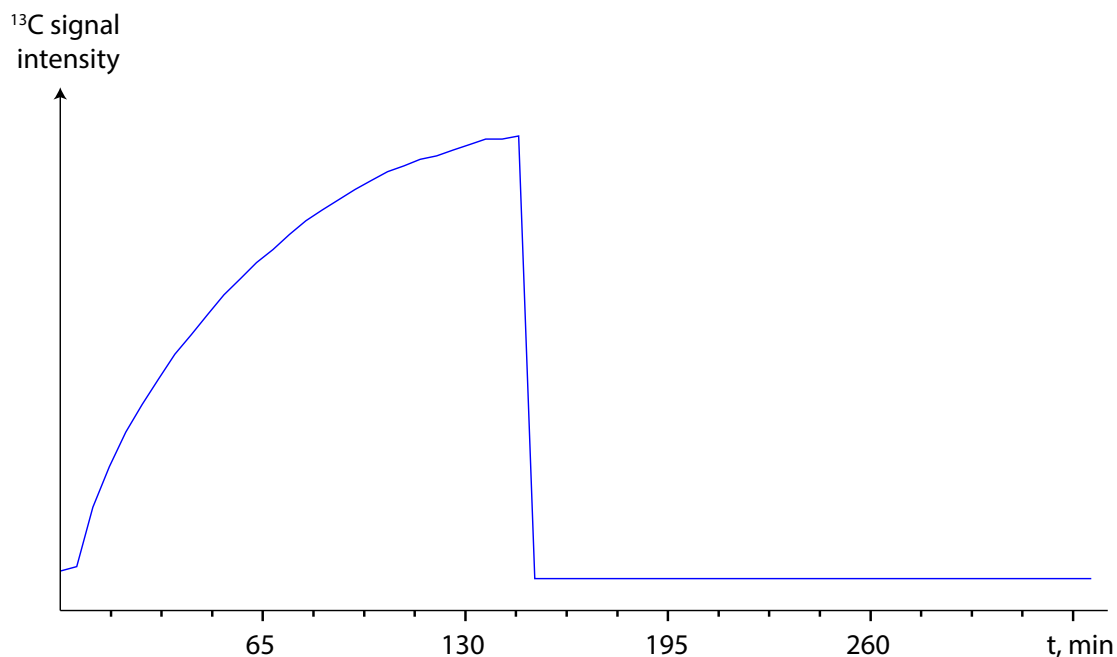
Figure 4-7 A polarisation build-up curve of 1M Na_2CO_3 in 1:3:6 H_2O:D_2O:d8-glycerol solution containing 50 mM TEMPOL radical. Temperature 1.3 K.

Chapter 5.   Conclusions

In summary, we have shown that the inclusion of NMR based constraints in crystal structure determination leads to faster and more precise structural solutions. We have demonstrated that by using NMR to select amongst conformers in the first steps of the crystal structure demonstration to significantly accelerate the structure determination. Machine learned NMR shift calculations have been a big breakthrough in the NMR community and we have demonstrated that they can be added as an additional pseudo-energy term in structural determination protocols. We also showed that they can be used to investigate other systems such as amorphous compounds, where large scale chemical shift calculations are necessary. We note that such calculations were impossible before due to the lack of fast and precise methods to calculate chemical shifts. Additionally, we have shown how dissolution DNP can be used to investigate the first steps of crystalline or amorphous structure formation by applying it to calcium carbonate prenuclear cluster formation.

Appendix

Appendix A. The core part of the script for crystal structure determination using machine learned chemical shifts.

```
### Import libraries
import numpy as np
import os
import sys
import copy
import pickle as pk
import time
import ase.io

### Path to the source files
sys.path.insert(0, "../src")

### Import source files
import new_mc_functions as mc
import new_ml_functions as ml
import new_energy_functions as en
import new_generate_crystals as cr

### Select space group, symmetry and number of molecules in the unit cell

sg = 14
n_mol = 4
sym = "M"

### Parameters for the chemical system under study

molecule = "azd2"
experiment = 0 #sys.argv[1]
comment = ""

# Directory for results
out_dir = "../data/test/"

if molecule == "cocaine":
    # Path to initial file
    initial_structure_original                                    =
'../input_structures/cocaine/COCAIN10_H_relaxed_out_cell_vc-
relax_DFTBplus_gas.xyz'

    # Number of atoms in one molecule
    n_atoms = 43

    # Number of dihedral angles
    n_conf = 5

    # Atoms and masks of the dihedral angles
    conf_params = {}
    conf_params["a1"] = [13, 19, 7, 2, 1]
    conf_params["a2"] = [8, 7, 18, 1, 14]
    conf_params["a3"] = [7, 18, 2, 14, 21]
    conf_params["a4"] = [18, 2, 3, 20, 15]
    conf_params["mask"] = [[9, 10, 11, 12, 13, 32, 33, 34, 35, 36], [8, 9,
10, 11, 12, 13, 19, 32, 33, 34, 35, 36],
```

```
                                    [7, 8, 9, 10, 11, 12, 13, 19, 32, 33, 34, 35,
36], [15, 20, 21, 37, 38, 39],
                                    [15, 37, 38, 39]]

if molecule == "azd":
    initial_structure_original                                        =
'../input_structures/azd8329/azd8329_reference_relaxed_again_dftb.xyz'
    n_atoms = 62
    n_conf = 5  # amino fixed
    conf_params = {}
    conf_params["a1"] = [41, 40, 55, 31, 59]
    conf_params["a2"] = [58, 39, 52, 36, 37]
    conf_params["a3"] = [42, 41, 38, 56, 33]
    conf_params["a4"] = [43, 61, 39, 57, 32]
    conf_params["mask"] = [
        [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 43, 44,
45, 46, 47, 48, 49, 50, 51],
        [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 42, 43,
44, 45, 46, 47, 48, 49, 50, 51, 58, 6, 61],
        [53, 54, 55, 28, 29, 30, 23, 22, 24, 25, 26, 27], [4, 60, 37, 59,
33, 32, 31, 35, 34, 1, 0, 2, 3], [4, 60, 59]]

if molecule == "azd2":
    initial_structure_original                                        =
'../input_structures/azd8329/AZD8329_Form1_dftb_vc_relax.xyz'
    n_atoms = 62
    n_conf = 5  # amino fixed
    conf_params = {}
    conf_params["a1"] = [17, 4, 0, 38, 25]
    conf_params["a2"] = [7, 6, 22, 8, 9]
    conf_params["a3"] = [1, 0, 13, 25, 20]
    conf_params["a4"] = [11, 14, 58, 43, 15]
    conf_params["mask"] = [[3, 2, 17], [3, 2, 7, 17, 1, 11, 31, 18, 4, 12,
32, 19, 5],
                                    [58, 59, 60, 61, 51, 52, 53, 54, 44, 45, 46, 47],
                                    [3, 2, 7, 17, 1, 11, 31, 18, 4, 12, 32, 19, 5, 6,
0, 14, 38, 39, 22, 13, 58, 59, 60, 61, 51,
                                     52, 53, 54, 44, 45, 46, 47],
                                    [3, 2, 7, 17, 1, 11, 31, 18, 4, 12, 32, 19, 5, 6,
0, 14, 38, 39, 22, 13, 58, 59, 60, 61, 51,
                                     52, 53, 54, 44, 45, 46, 47, 8, 25, 43, 10]]

### Parameters for the simulation

# Set of parameters to change
parameter_set = ["a", "b", "c", "beta", "trans", "rot", "conf"]

# Whether random choices are between all parameters (True) or between groups
of parameters (False, like cell lengths, cell angles, conformers)
weighted = True

# Initial step size of each parameter: unit cell length, unit cell angle,
translation (fraction of unit cell length), rotation, dihedral angle
init_step_size = [2., 20., 0.05, 30., 40.]

# Maximum absolute step size of each parameter: unit cell length, unit cell
angle, translation (fraction of unit cell length), rotation, dihedral angle
max_step_size = [20., 360., 1., 360., 360.]

# How much does the amplitude of a step change after each step (divide by
this value when rejected, multiply when accepted)
step_factor = 2.
```

```python
# Maximum volume (ratio w.r.t. sum of VDW spheres in the molecule * number
of molecules in unit cell) for the initial unit cell
max_V_factor = 2.0

# Use smart cell generation
smart_cell = True

# Minimum and maximum values of unit cell lengths and angles
cell_params_lims = [1., 50., 45., 135.]

# Number of structures to run
n_structures = 999

# Number of Monte-Carlo loops (oly valid if the criterion for stopping the
MC run is not variable, otherwise maximum number of loops)
n_loops = 4

# Criterion for stopping the MC run. T=temperature
stopping_criterion = "T"
T_start = 2500.
T_stop = 50.
T_profile = "linear"

# Threshold for stopping the simulation
A_thresh = 1e-3

# Gas constant (kJ/(mol*K))
gas_cst = 8.314e-3

# Perform an optimization of the parameters at the end of the MC run
opt = False

# Maximum number of iterations of the optimization (Powell-like method)
n_opt = 10

# Perform hydrogen position relaxation every n_dftb step (-1 for no
relaxation)
n_dftb = 500

# Maximum number of DFTB relaxation steps
n_opt_dftb = 50

# Form of the cost function. E=Energy, H=1H RMSE, C=13C RMSE
cost_function = ["E", "H"]

# Factors and options for each part of the cost function

# (For energy: factor = Ha -> kJ/mol)
cost_factors = {"E": 2625.50, "H": 200.}
cost_options = {"E": {}, "H": {"cutoff": 0.1, "slope": -1., "offset": 30.36}}

# Available nuclei for ShiftML
symb2z = {"H": 1, "C": 6, "N": 7, "O": 8}
z2symb = {1: "H", 6: "C", 7: "N", 8: "O"}

### Experimental shifts

exp_shifts = {}
equivalent = {}
ambiguous = {}
```

```
# 1H

if molecule == "cocaine":
    # Experimental chemical shifts: the shifts should be in the order of the
input structure
    # cost_options["H"]["exp_shifts"] = [3.76, 3.78, 5.63, 3.06, 3.32, 3.49,
3.38, 2.91, 2.56, 2.12, 8.01, 8.01, 8.01, 8.01, 8.01, 3.78, 3.78, 3.78, 1.04,
1.04, 1.04]
    # TEMPORARY: Use ShiftML predicted shifts instead of experimental ones
(COCAIN10_H_relaxed_out_cell_vc-relax_DFTBplus.pdb)
    cost_options["H"]["exp_shifts"]          =          [3.9780189569673787,
3.636220267098391, 5.50111356170509, 1.7017083744995354,
                                             2.97372302702669,
3.6334961628397053, 3.4831881678204475, 1.7656191040126927,
                                             2.051699996346425,
1.6921747247077903, 7.867125539838806, 7.461451020956069,
                                             7.57521557453952, 7.566502423701898,
7.354128212230904, 4.212966601516133,
                                             3.3785319903128546,
3.695539645312305, 0.513272767915403, 2.5564618319438317,
                                             1.3211352909471152]

    # Equivalent shifts, where computed shifts should be averaged
    cost_options["H"]["equivalent"] = [[10, 11, 12, 13, 14], [15, 16, 17],
[18, 19, 20]]

    # Ambiguous shifts, where best matching criterion should be used
    cost_options["H"]["ambiguous"] = [[3, 4], [6, 7], [8, 9]]

if molecule == "azd":
    # TEMPORARY: Use ShiftML predicted shifts instead of experimental ones
(azd8329_reference_relaxed_again_dftb.pdb)
    cost_options["H"]["exp_shifts"]          =          [7.1847614337427395,
8.019417536236109, 8.981483537078102, 9.127007011316138,
                                             10.972039557925008,
7.590504887014717, 8.809498191575582, 3.0254898357411406,
                                             1.2099446262222138,
1.8715797502761902, 2.2008161233765904, 1.7999889924498333,
                                             0.7589794590651451,
0.5982792421141738, 2.332944264321835, 1.9756070695322805,
                                             1.9591698333539043,
0.7327033381833026, 0.888684149575198, 1.1242121185541656,
                                             1.3936565407686743,
0.42664731956235613, -1.56819281256708, 0.644642563258877,
                                             -0.43414459379003034,
0.8934892628513076, 0.48016379574931634,
                                             0.06088604186567537,            -
1.093036817638783, -0.7779541416382578,
                                             -0.8168341504087238]
    #                   cost_options["H"]["exp_shifts"]                    =
[6.63,8.28,9.12,8.27,17.1,7.65,10.4,2.72,1.58,1.61,2.46,1.57,0.84,0.55,1.60
,2.23,1.80,0.83,-0.09,1.41,1.63,0.94,-0.25,1.09,0.42,0.62,0.84,0.71,-0.51,-
0.36,-0.51]

    # Equivalent shifts, where computed shifts should be averaged
    cost_options["H"]["equivalent"] = [[22, 23, 24], [25, 26, 27], [28, 29,
30]]

    # Ambiguous shifts, where best matching criterion should be used
    cost_options["H"]["ambiguous"] = [[9, 10], [20, 21], [12, 13], [17, 18]]

if molecule == "azd2":
```

```python
    cost_options["H"]["exp_shifts"]                     =           [9.204499364822265,
7.621297852503048, 4.693370241015021, 7.892352372852706,
                                       1.410986703871579, 8.04781786216055,
4.130864655660265, 1.7950741200761122,
                                       1.5908689663372186,
0.18552860342953537, 0.9140897889972557, 6.535734200276945,
                                       1.2442325173457967,
2.589213934025505, -0.006275014798418965, 9.041262366438989,
                                       1.5797344773520443,
0.9591675704675175, 1.6359860773621193, 1.6359860773621193,
                                       1.6359860773621193,
0.5252608043051872, 2.0623891790050486, -0.23114035197289695,
                                       -0.23114035197289695,                     -
0.23114035197289695, -0.0757847725061005,
                                       -0.11428546533861805,                     -
0.6095694461503349, -0.6095694461503349,
                                       -0.6095694461503349]

    cost_options["H"]["equivalent"] = [[18, 19, 20], [23, 24, 25], [28, 29,
30]]
    cost_options["H"]["ambiguous"] = [[8,9],[13,14],[16,17],[26,27]]

### Parameters for the system

# Name of the DFTB+ program
dftb_pgm_name = "dftb+"

# Is the simulation running on the cluster?
cluster = False

# Set a seed for random number generation
seed = None

# Write crystal structure at each MC step
write_intermediates = False

# Verbosity
verbose = True

# To follow how parameters change in each step
print_param_change = False

###############################################################################
################

### Initialization

###############################################################################
###############

print("Initializing the system...")

# Finding dftb+ path on the local machine
if "E" in cost_function:
    cost_options["E"]["dftb_path"] = en.which(dftb_pgm_name)

# If a seed is set:
if seed:
    np.random.seed(seed)

if stopping_criterion == "T":
```

```python
    T_list     =     mc.generate_T_profile(T_start,    T_stop,    n_loops,
profile=T_profile)

# Create output directory
if not os.path.exists(out_dir):
    os.mkdir(out_dir)

# Generate name of experiment
if stopping_criterion == "T":
    name = "{}_loops_".format(n_loops)
    for c in cost_function:
        name += "{}_factor_{}_".format(c, cost_factors[c])
    name += "{}_{}".format(comment, experiment)
    if opt:
        name += "_opt"

else:
    raise ValueError("Experiment name not implemented yet for the selected
criterion ().".format(stopping_criterion))

if not os.path.exists(out_dir + name):
    os.mkdir(out_dir + name)

if "E" in cost_function:
    if cluster:
        cost_options["E"]["directory"] = "/dev/shm/"
    else:
        cost_options["E"]["directory"] = out_dir + name + "/"
print(cost_options["E"])

# Load ShiftML kernels

krr = {}
representation = {}
trainsoaps = {}
model_numbers = {}
zeta = {}

for c in cost_function:
    if c in symb2z.keys():
        cost_options[c]["krr"],                    cost_options[c]["rep"],
cost_options[c]["tr"], cost_options[c]["m_num"], \
        cost_options[c]["zeta"] = ml.load_kernel(symb2z[c])

print("Done!")

###############################################################################
###############

### Loading initial structure

###############################################################################
###############

print("Loading initial structure...")

# Generate random number to set as name of the structure
random_number = np.random.random()

# Load initial structure
initial_structure = ase.io.read(initial_structure_original)
```

```python
# Generate gas-phase molecule
cr.write_gas_phase_conformer(initial_structure, n_atoms,
                             out_dir          +          name          +
"/{}_single_molecule.xyz".format(random_number))
print("Done!")

### For all structures

for k in range(n_structures):


###########################################################################
###############

    ### Monte-Carlo


###########################################################################
##############

    # Load the starting structure
    starting_structure = copy.deepcopy(initial_structure)

    # Generate a trial crystal structure
    best_crystal,       lat,       trans,       R,       conf_angles       =
cr.generate_crystal(starting_structure, n_atoms, n_mol, sg,

parameter_set, cell_params_lims, n_conf=n_conf,

conf_params=conf_params, smart_cell=smart_cell,

max_V_factor=max_V_factor, verbose=verbose)

    # Write the initial trial crystal structure
    ase.io.write(out_dir   +   name   +   "/init_crystal_{}.cif".format(k),
best_crystal)

    ## Initialize arrays for monitoring

    # Parameter monitoring

    all_params = {}
    sel_params = {}
    acc_params = {}
    Amp = {}

    for p in parameter_set:
        all_params[p]    =    [mc.current_parameter(p,    lat,    trans,    R,
conf_angles)]
        sel_params[p] = 0
        acc_params[p] = 0
        Amp[p] = 1.

    # Cost function monitoring
    all_costs = {}
    all_costs["Tot"] = []

    for c in cost_function:
        all_costs[c] = []

    # Compute initial cost function
```

```python
    old_cost = mc.compute_cost(best_crystal, cost_function, cost_factors,
cost_options)
    new_cost = copy.deepcopy(old_cost)
    new_cost["Tot"] = 1.
    all_costs["Tot"].append(old_cost["Tot"])

    for c in cost_function:
        all_costs[c].append(old_cost[c])

    if stopping_criterion == "T":

        # Initialize time monitoring
        start = time.time()

        # Temperature-based profile
        for i, T in enumerate(T_list):
            if i > 0 and n_dftb >= 0 and i % n_dftb == 0:
                print("Optimizing      proton      positions      for      {}
steps...".format(n_opt_dftb))
                tmp_structure                                              =
en.dftb_relax(cost_options["E"]["directory"], best_crystal,

cost_options["E"]["dftb_path"], n_opt)
                tmp_mol   =   cr.retrieve_initial_structure(tmp_structure,
n_atoms, starting_structure, n_conf=n_conf,

conf_params=conf_params)
                tmp_crystal, clash = cr.create_crystal(tmp_mol, lat, trans,
R, conf_angles, sg, n_atoms,

conf_params=conf_params)
                tmp_cost   =   mc.compute_cost(tmp_crystal,   cost_function,
cost_factors, cost_options)
                if tmp_cost["Tot"] < old_cost["Tot"]:
                    print("Cost           decreased           by          {:.2f}
kJ/mol".format(old_cost["Tot"] - tmp_cost["Tot"]))
                    old_cost = copy.deepcopy(tmp_cost)
                    starting_structure = copy.deepcopy(tmp_mol)
                    best_crystal = copy.deepcopy(tmp_crystal)

                else:
                    print("Optimization  did  not  yield  a  lower  energy
structure.")

            if verbose:
                print("Loop {}/{}, T = {} K".format(i + 1, n_loops, T))

            # Select a random parameter to change

            if weighted:
                weighted_parameter_set = copy.deepcopy(parameter_set)
                if "conf" in parameter_set:
                    for _ in range(len(conf_angles) - 1):
                        weighted_parameter_set.append("conf")
                param_to_change = np.random.choice(weighted_parameter_set)

            else:
                param_to_change = np.random.choice(parameter_set)
            sel_params[param_to_change] += 1

            # Update the selected parameter
```

```python
            trial_lat,     trial_trans,     trial_R,     trial_conf     =
mc.randomize(param_to_change, lat, trans, R, conf_angles,

init_step_size, A=Amp[param_to_change])

            clash = False

            # Check if the cell parameters are within the bounds, otherwise
reject the step
            if     np.min(trial_lat[:3])     <     cell_params_lims[0]     or
np.max(trial_lat[:3]) > cell_params_lims[1] or np.min(
                    trial_lat[3:])     <     cell_params_lims[2]     or
np.max(trial_lat[3:]) > cell_params_lims[3]:
                clash = True

            if not clash:
                # Generate the new trial crystal
                trial_crystal,                 clash                 =
cr.create_crystal(starting_structure, trial_lat, trial_trans, trial_R,
                                                trial_conf,    sg,
n_atoms, conf_params=conf_params)

            accept = False

            # Check the structure for clashes
            if not clash:
                # Check intramolecular clashes
                clash  =  cr.check_clash(trial_crystal,  n_atoms,  pbc=True,
clash_type="intra", factor=0.85)
                if not clash:
                    # Check intermolecular clashes
                    clash = cr.check_clash(trial_crystal, n_atoms, pbc=True,
clash_type="inter", factor=0.85)

            if not clash:
                # Compute the new cost function
                new_cost  =  mc.compute_cost(trial_crystal,  cost_function,
cost_factors, cost_options)
                # If energy decreases, accept the step, otherwise accept it
with Boltzmann probability
                if new_cost["Tot"] < old_cost["Tot"]:
                    accept = True
                elif   np.random.random()   <   np.exp((old_cost["Tot"]   -
new_cost["Tot"]) / (gas_cst * T)):
                    accept = True
            # If the step is accepted, update the cost function, parameters,
and unit cell

            if verbose:
                stop = time.time()
                dt = stop - start
                eta = dt / (i + 1) * (len(T_list) - i - 1)

            if accept:
                Amp[param_to_change] *= step_factor
                # If the amplitude makes a step go over the maximum step
size, reduce it
                Amp     =     mc.normalize_amplitude(Amp,     param_to_change,
init_step_size, max_step_size, verbose=verbose)
                acc_params[param_to_change] += 1

                if verbose:
```

```python
                    print(
                        "parameter changed: {}, energy difference: {}
kJ/mol, step accepted, new amplitude: {}, time elapsed: {:.2f} s, ETA {:.2f}
s.".format(
                            param_to_change,      new_cost["Tot"]      -
old_cost["Tot"], Amp[param_to_change], dt, eta))
                old_cost = copy.deepcopy(new_cost)
                best_crystal = copy.deepcopy(trial_crystal)
                lat = copy.deepcopy(trial_lat)
                trans = copy.deepcopy(trial_trans)
                R = copy.deepcopy(trial_R)
                conf_angles = copy.deepcopy(trial_conf)

            else:
                Amp[param_to_change] /= step_factor
                # If the amplitude makes a step go over the maximum step
size, reduce it
                Amp    =    mc.normalize_amplitude(Amp,    param_to_change,
init_step_size, max_step_size, verbose=verbose)
                if verbose:
                    print(
                        "parameter changed: {}, energy difference: {}
kJ/mol, step rejected, new amplitude: {}, time elapsed: {:.2f} s, ETA {:.2f}
s.".format(
                            param_to_change,      new_cost["Tot"]      -
old_cost["Tot"], Amp[param_to_change], dt, eta))

            # Append the cost function and parameters for monitoring
            all_costs["Tot"].append(old_cost["Tot"])
            for c in cost_function:
                all_costs[c].append(old_cost[c])
            for p in parameter_set:
                all_params[p].append(mc.current_parameter(p, lat, trans, R,
conf_angles))
            # Stop the MC simulation if the maximum amplitude is lower than
the threshold
            for p in parameter_set:
                converged = True
                if Amp[p] > A_thresh:
                    converged = False
            if converged:
                print("Convergence reached!")
                break

            if write_intermediates:
                ase.io.write(out_dir       +       name       +
"/intermediate_crystal_{}.cif".format(i), trial_crystal)
            if print_param_change:
                print("a:   "  +  str(all_params["a"][-1]),  "b:  "  +
str(all_params["b"][-1]),
                        "c:   "  +  str(all_params["c"][-1]),  "alpha:  "  +
str(all_params["alpha"][-1]),
                        "beta: " + str(all_params["beta"][-1]), "gamma: " +
str(all_params["gamma"][-1]),
                        "trans: " + str(all_params["trans"][-1]), "rot: " +
str(all_params["rot"][-1]),
                        "conf: " + str(all_params["conf"][-1]))

        # Save parameter list
        if verbose:
            with        open(out_dir        +        name        +
"/params_crystal_{}.pickle".format(k), "wb") as f:
```

```python
                pk.dump([all_costs, all_params, sel_params, acc_params], f)
        # Write the final crystal structure
        ase.io.write(out_dir  +  name  +  "/final_crystal_{}.cif".format(k),
best_crystal)

    if verbose:
        for p in parameter_set:
            print(p, all_params[p][-1])
        for p in all_costs.keys():
            print(p, all_costs[p][-1])
```

Appendix B A compilation of assignment for ^{13}C and ^{1}H for organic crystals in the literature.

1 Brucine

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C10, C11, C12, C13, C14, C15, C16, C17, C18, C20, C21, C22, C23, C24 | 103.18, 146.04, 149.38, 101.59, 136.68, 124.74, 53.06, 59.55, 168.33, 42.39, 76.51, 50.05, 30.75, 28.12, 62.69, 42.47, 50.89, 55.31, 141.40, 126.69, 64.41, 55.43.[331]

2 Cortisone acetate (form I)

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23 | 33.7, 33.7, 202.7, 125.9, 169.6, 33.7, 33.7, 36.4, 62.7, 39.9, 210.6, 53, 49.6, 50.3, 25.7, 33.7, 88.6, 15.6, 17.6, 204.7, 68.1, 175.4, 21.1.[332]

3 (S,S)-ethambutol·HCl (form I)

^{13}C shifts: C2, C1, C5, C3, C4 | 58.9, 54.2, 39.6–40.2, 22.7, 10.8[333]

^{1}H shifts: H4A, H4B, H4C, H3B, H3A, H5A, H5B, H1A, H1B, H2, H1, H1Y, H1X | 1.09, 1.09, 1.09, 1.66, 2.06, 3.48, 3.48, 3.76, 3.76, 3.76, 5.13, 8.90, 9.43[333]

4 Fluoxetine·HCla

^{13}C shifts: C10, C2, C1, C3, C17, C5, C9, C7, C14, C6, C12, C15, C8, C16, C13, C11, C4 | 31.3, 32.47, 44.8, 79.12, 115.76, 115.76, 121.38, 121.38, 125.81, 127.11, 127.11, 128.7, 128.7, 128.7, 129.13, 142.54, 160.92[333]

^{1}H shifts: H1, H3, H12, H4, H2, H5, H19, H19, H19, H6, H8, H18, H7, H15, H9, H11, H11 | 0.37, 0.37, 2.73, 2.73, 2.73, 3.5, 6.47, 6.47, 6.47, 7.19, 7.19, 7.19, 7.19, 8.23, 8.23, 10.19, 10.19[333]

5 Succinylsulfathiazole·H2O

^{13}C shifts: C2, C3, C12, C6, C13, C10, C7, C9, C8, C5, C11, C4, C1 | 27.37, 30.05, 107.56, 114.02, 120.30, 122.70, 125.63, 129.87, 135.60, 141.59, 169.36, 172.99, 175.85[333]

6 α-testosterone

^{13}C shifts: C32, C34, C33, C28, C38, C25, C24, C2, C23, C1, C37, C31, C35, C36, C27, C26, C12, C30, C29, C5, C7, C6, C20, C11, C17, C16, C13, C15, C3, C9, C4, C8, C9, C19, C18, C14, C22, C21 | 201.22, 170.64, 125.67, 80.35, 54.14, 51.14, 43.67, 40.03, 36.78, 36.19, 36.92, 34.55, 33.76, 32.40, 30.19, 24.33, 22.99, 18.56, 11.67, 202.72, 172.09, 125.17, 82.69, 55.26, 51.86, 41.71, 39.49, 37.40, 35.63, 34.18, 30.94, 33.85, 32.35, 26.38, 22.49, 20.91, 14.20, 7.80[334]

7 β-testosterone

^{13}C shifts: C1, C12, C13, C14, C15, C16, C17, C18, C19, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11 | 35.25, 35.25, 200.15, 124.65, 173.75, 33.45, 33.45, 35.25, 39.35, 39.35, 20.95, 35.25, 43.55 , 51.55, 23.95, 28.75, 80.65, 12.55, 16.85[334]

8 L-Alanine

^{13}C shifts: C1, C2, C3 | 176.8, 50, 19.8[335]

9 4-Aminobenzenesulfonamide (α polymorph)

^{13}C shifts: C4, C3, C6, C1, C2, C5 | 153.7, 113.1/115.3, 128.3, 128.0, 128.3, 113.1/115.3[336]

10 L-Asparagine monohydrate

^{13}C shifts: C1, C2, C3, C4 | 176.4, 51.8, 36.1, 177.1[337]

11 L-Cysteine

^{13}C shifts: C1, C2, C3 | 175.1, 53.7, 35.4[338]

12 1,2-Dihydro-acenaphthylene

^{13}C shifts: C1, C2, C3, C4, C41, C42, C5 | 148.1, 120.3, 129.4, 122.3, 131.9, 139.9, 29.5[339]

13 Flufenamic acid (Form I)

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C1 | 149.3, 109.7, 133, 117.2–121.7–119.8, 136.3, 112, 175, 139.9, 121.7–119.8–117.2, 131.7, 119.8–117.2–121.7, 129.5, 128.1, 124.1, 149.3[118]

^{1}H shifts: H1, H2, H3, H4, H6, H5, H7, H8, H9, H10 | 8.3, 6.0–6.9–6.2, 5.4, 6.8, 9.6, 12.4, 6.9–6.2–6.0, 6.2–6.0–6.9, 5.9, 7.3[118]

14 β-D-fructopyranose

^{13}C shifts: C1, C2, C3, C4, C5, C6 | 65.4, 99.7, 67.2, 69.0, 71.4, 64.9[340]

15 Methyl β-D-galactopyranoside

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7 | 100.4, 67.6, 72.6, 70.0, 72.9, 61.4, 55.2[341]

16 Methyl α-D-Galactopyranoside Monohydrate

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7 | 105.7, 71.2, 72.1, 69.3, 75.6, 62.8, 57.6[341]

17 Methyl α-D-glucopyranoside

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7 | 101.0, 72.3, 74.6, 72.5, 75.3, 63.8, 56.5[341]

18 Methyl α-D-mannopyranoside

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7 | 99.6, 71.3, 71.7, 64.8, 71.9, 58.9, 54.9[341]

19 Methyl β-D-xylopyranoside

^{13}C shifts: C2, C1, C5, C4B, C3B, C4, C5B, C1B, C2B, C3B, c6 | 129.3, 126.0, 125.4, 129.9, 134.9, 129.3, 126.0, 125.4, 129.9, 134.9, missing[341]

20 Naphthalene

^{13}C shifts: C1, C2, C3, C4, C5, C6 | 104.2, 72.2, 78.2, 69.5, 66.9, 57.3[342]

21 Paracetamol (form I)a

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8 | 133.1, 123.4, 115.7, 152.3, 116.4, 120.6, 169.8, 23.8[343]

^{1}H shifts: H7, H8, H9, H2, H5, H3, H4, H6, H1, H7 | 0.99, 0.99, 0.99, 5.68, 6.61, 6.61, 7.74, 8.8, 9.1, 0.99[333]

22 Pentaerythritol

^{13}C shifts: C1, C2 | 50.2, 58.4[344]

23 α-L-Rhamnose

^{13}C shifts: C1, C2, C3, C4, C5, C6 | 94.5, 72.2, 71.0, 72.5, 69.8, 17.8[340]

24 L-Serine

^{13}C shifts: C1, C2, C3 | 175.1, 55.6, 62.9[338]

25 L-Serine monohydrate

^{13}C shifts: C1, C2, C3 | 175.6, 58.3, 61.8[337]

26 Sucrose

^{13}C shifts: C1, C3, C5, C7, C9, C11, C12, C10, C8, C6, C4, C2 | 93.06, 73.59, 72.65, 67.82, 73.59, 59.72, 65.79, 102.04, 82.69, 71.64, 81.58, 60.98[345]

27 L-Threonine

^{13}C shifts: C1, C2, C3, C4 | 170.0, 60.2, 65.4, 18.9[346]

28 Triphenylene

^{13}C shifts: C15, C1, C10, C5, C4, C16, C14, C11, C9, C6, C3, C17, C13, C12, C8, C7, C2, C18 | 126.4, 129.5, 124.5, 125.9, 127.5, 122.3, 130.2, 129.5, 120.9, 125.9, 121.7, 129.5, 129.5, 122.3, 126.9, 126.9, 123.8, 129.8[347]

29 Amoxicillin trihydrate

^{13}C shifts: C7, C6, C8, C1, C2, C3, C4, C5, C9, C10, C11, C16, C15, C14, C13, C12 | 27.6, 28.6, 172.8, 65.2, 72.4, 65.2, 53.9, 174.0, 170.5, 56.4, 123.8, 124.5, 118.5, 158.3, 120.1, 128.1[179]

30 Ampicillin trihydrate

^{13}C shifts: C18, C20, C28, C11, C29, C30, C31, C32, C33, C12, C14, C16, C19, C24, C15, C26 | 30.2, 173.3, 135.5, 64.8, 130.0, 129.2, 132.2, 129.2, 127.1, 73.4, 65.3, 175.2, 28.9, 170.2, 57, 57[348]

31 Aspirine

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9 | 122.5, 153.3, 125.8, 139.1, 128.3, 135.1, 171.2, 172.5, 20.7[349]

32 Cortisone acetate (form II)

^{13}C shifts: C3, C13, C17, C15, C19, C18, C9, C14, C23, C5, C4, C6, C2, C22, C8, C12, C20, C1, C21, C7, C16, C11, C10 | 209.5, 206.0, 198.7, 170.8, 175.0, 123.5, 89.4, 70.4, 63.5, 52.4, 49.9, 49.9, 38.7, 37.8, 36.6, 34.7, 33.6, 33.2, 31.6, 24.5, 20.5, 16.6, 16.3[350]

33 Cimetidine hydrochloride

^{13}C shifts: C2, C4, C5, C6, C7, C8, C9, C11, C13, C16 | 133.3, 133.3, 128.7, 23.0, 10.3, 28.6, 42.4, 160.9, 126.2, 28.6[72]

34 Finasteride (form I°)

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23 | 146.9, 126.9, 164.3, 60.6, 26.8, 30.6, 36.8, 48.0, 40.2, 21.9, 38.8, 44.3, 56.1, 26.0, 24.3, 57.6, 15.2, 14.7, 169.3, 51.9, 29.3, 29.3, 29.3[351]

35 Finasteride (form II)

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C27, C29, C31, C32, C33, C34, C36, C39, C40, C41, C42, C44, C45, C46 | 150.1, 123, 169.4, 59.2, 25.9, 30.8, 35.3, 48.5, 39.9, 21, 37.4, 44, 56.7, 24.8, 23.6, 58, 13.1, 12.1, 165.1, 50.8, 27.9, 27.9, 27.9, 151.6, 60.4, 31.4, 48.8, 40.3, 21.5, 38.4, 57.8, 58.5, 13.6, 12.5, 164.4, 28.2, 28.2, 28.2[352]

36 [Leu5]Enkephalin monohydrate

^{13}C shifts: C2, C1, C3, C4, C6, C8, C9, C7, C5, C11, C10, C13, C12, C15, C14, C16, C17, C18, C20, C22, C21, C19, C24, C23, C25, C26, C28, C27 | 169.2, 54.6, 43.8, 124.4, 129.6, 114.2, 156.4-154.7, 114.2, 129.6, 166.2, 54.6, 168.0, 43.8, 171.3, 54.6, 43.8, 137.2–136.0, 127.2, 127.2, 127.2, 127.2, 127.2, 174.7, 54.6, 43.8, 38.0, 25.8, 23.8[353]

37 [Leu5]Enkephalin dihydrate dimethylformamide solvate

^{13}C shifts: C2, C1, C11, C12, C17, C16, C15, C14, C13, C4, C3, C6, C5, C8, C7, C18, C19, C24, C23, C22, C21, C20, C10, C9, C25, C26, C28, C27, C30, C29, C39, C45, C44, C43, C42, C41, C40, C32, C31, C34, C33, C36, C35, C46, C47, C48, C49, C50, C51, C52, C38, C37, C53, C54, C56, C55, C58, C57, C67, C73, C72, C71, C70, C69, C68, C60, C59, C62, C61, C64, C63, C74, C75, C76, C77, C78, C79, C80, C66, C65, C81, C82, C84, C83, C86, C85, C95, C101, C100, C99, C98, C97, C96, C88, C87, C90, C89, C92, C91, C102, C103, C104, C105, C106, C107, C108, C94, C93, C109, C110, C112, C111 | 169.8, 52.3, 36.4, 126.0, 131.8, 115.7, 155.5, 115.7, 131.8, 167.8, 42.2, 167.8, 42.2, 176.2, 54.0, 36.4, 137.9, 130.8, 127.4, 126.0, 127.0, 129.4, 178.6, 54.0, 24.3, 21.7, 19.7, 18.4, 167.8, 52.3, 36.4, 126.0, 131.8, 115.7, 155.5, 115.7, 131.8, 169.8, 42.2, 169.8, 42.2, 174.1, 54.0, 36.4, 136.3, 130.8, 127.4, 126.0, 127.0, 129.4, 178.6, 54.0, 24.3, 21.7, 19.7, 18.4, 167.8, 52.3, 36.4, 126.0, 131.8, 121.3, 155.5, 115.7, 131.8, 169.8, 42.2, 167.8, 42.2, 176.2, 54.0, 36.4, 137.9, 130.8, 127.4, 126.0, 127.0, 129.4, 177.9, 54.0, 24.3, 21.7, 19.7, 18.4, 169.8, 52.3, 36.4, 126.0, 131.8, 115.7, 158.3, 121.3, 131.8, 167.8, 42.2, 169.8, 42.2, 174.1, 54.0, 36.4, 137.9, 130.8, 127.4, 126.0, 127.0, 129.4, 178.6, 54.0, 24.3, 21.7, 19.7, 18.4[353]

38 [Met]Enkephalin hydrate

^{13}C shifts: C1, C2, C3, C8, C5, C6, C9, C7, C4, C10, C11, C12, C13, C14, C15, C16, C21, C18, C20, C22, C19, C17, C23, C24, C25, C27, C26, C28, C29, C30, C35, C32, C33, C36, C34, C31, C37, C28, C39, C40, C41, C42, C43, C48, C45, C47, C49, C46, C44, C50, C51, C52, C54, C53 | 4.0, 115.4, 135.3, 43.6, 36.5, 54.1, 11.5, 55.3, 37.6, 1.6, 124.8, -0.1, 124.0, 1.1, 112.7, 132.0, 30.6, 34.2, 38.7, 37.1, 34.9, 41.4, -14.8, 117.0, 135.2, 140.2, 158.1, -5.3, 115.1, 130.2, 36.4, 33.9, 56.3, 10.8, 52.4, 43.1, -2.3, 129.2, 2.4, 128.6, -3.6, 114.7, 129.4, 31.8, 36.5, 40.8, 39.4, 41.8, 39.0, -11.6, 113.3, 140.2, 141.1, 155.7[353]

39 N-formyl-L-methionyl-L-leucyl-L-phenylalanine-OMe (MLF)

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C1, C2, C3, C4, C5, C6, C7, C8 | 162.1, 53.0, 41.0, 32.8, 16.4, 175.8, 59.4, 41.0, 26.9, 20.9, 177.6, 55.7, 37.2, 139.0, 132.0, 130.4, 129.8, 131.3, 134.0, 175.0, 55.7, 162.1, 53.0, 41.0, 32.8, 16.4, 175.8, 59.4, 41.0[354]

40 Nicotinamide palmitic acid (1_1 cocrystal)

^{13}C shifts: C(17), C(18), C(19), C(20), C(21), C(22), C(1), C(2), C(3), C(4), C(5), C(6) | 147.5, 128.5, 135.8, 125.8, 154.4, 172.7, 175.4, 35.0, 25.4, 34.0, 32.6, 34.0[355]

<superscript>1</superscript>H shifts: H(1), H(37), H(38), H(33), H(34), H(35), H(36), H(2), H(3), H(4), H(5), H(6), H(7), H(8)
, H(9), H(10), H(11), H(12), H(13), H(14), H(15), H(16), H(17), H(18), H(19), H(20), H(21), H(22)
, H(23), H(24), H(25), H(26), H(27), H(28), H(29), H(30), H(31) | 13.6, 7.5–9.0, 7.5–9.0, 8.7, 6.
3, 6.7, 8.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5
–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0,
0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0, 0.5–2.0[355]

41 Nifedipine (modification I)

<superscript>13</superscript>C shifts: C4, C8, C17, C1, C10, C12, C14, C13, C15, C16, C7, C3, C5, C9, C6, C11, C2 | 168.1–
169.9, 168.1–169.9, 148.4, 149.4–144.8, 149.94–144.8, 143.9, 136.6, 135.7, 126.7, 124.1, 10
3.1–101.6, 103.1–101.6, 51.1–49.3, 51.1–49.3, 35.4, 19.2, 19.2[356]

42 Paclitaxel

<superscript>13</superscript>C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19
, C20, C26, C28, C35, C37, C41, C43, C45, C47, C55, C57, C48, C101, C102, C103, C104, C105,
C106, C107, C108, C109, C110, C111, C112, C113, C114, C115, C116, C117, C118, C119, C120
, C126, C128, C135, C137, C141, C143, C145, C147, C155, C157, C148 | 79.5, 75.5, 48.4, 82.9,
84.8, 37.2, 73.8, 59.7, 204.0, 76.3, 132.6, 139.8, 72.4, 35.2, 43.9, 23.0, 26.2, 14.1, 11.5, 77.4,
167.1, 128.9, 171.5, 20.7, 23.1, 170.7, 72.3, 55.8, 166.1, 131.7, 142.1, 79.5, 75.5, 48.4, 82.9,
84.8, 37.2, 73.8, 59.7, 204.0, 76.3, 132.6, 139.8, 72.4, 35.2, 43.9, 23.0, 26.2, 14.1, 11.5, 77.4,
167.1, 128.9, 171.5, 20.7, 23.1, 170.7, 72.3, 55.8, 166.1, 55.8, 166.1[357]

43 Parthenolide

<superscript>13</superscript>C shifts: C26, C25, C24, C22, C21, C19, C18, C30, C29, C27, C16, C17, C20, C23, C28, C11, C1
0, C9, C7, C6, C4, C3, C15, C14, C12, C1, C2, C5, C8, C13 | 123.3, 25.3, 36.5, 62.7, 67.9, 82.4,
49.8, 30.5, 42.0, 134.7, 170.4, 140.1, 121.4, 20.0, 18.1, 126.0, 25.0, 36.5, 61.7, 65.1, 84.5, 46.
5, 32.4, 43.9, 137.7, 171.4, 140.4, 123.3, 17.8, 17.6[358]

44 Tolbutamide (form I)

<superscript>13</superscript>C shifts: C10, C5, C9, C4, C3, C6, C8, C2, C1, C7, C11, C12 | 21.8, 145.8, 126.2, 126.2, 131.8,
131.8, 138.1, 155.0, 40.1, 31.8, 19.4, 13.7[359]

45 Tolbutamide (form IV)

<superscript>13</superscript>C shifts: C29, C37, C36, C26, C11, C19, C27, C4, C8, C40, C44, C34 | 22.1, 144.8, 127.2, 126.
2, 130.4, 130.4, 137.5, 154.0, 41.5, 33.8, 20.7, 16.1[359]

46 Vitamin-D3

<superscript>13</superscript>C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19
, C20, C21, C22, C23, C24, C25, C26, C27, C28, C29, C30, C31, C32, C33, C34, C35, C36, C37, C
38, C39, C40, C41, C42, C43, C44, C45, C46, C47, C48, C49, C50, C51, C52, C53, C54 | 33.65, 3
8.90, 70.21, 44.64, 137.59, 119.69, 120.29, 139.16, 28.50, 145.84, 24.55, 41.34, 45.57, 56.99

, 23.89, 28.98, 58.23, 13.19, 111.60, 38.81, 19.05, 37.49, 26.02, 40.11, 26.02, 23.10, 22.90, 31.59, 35.26, 66.82, 44.04, 135.57, 121.62, 117.05, 140.14, 31.69, 147.62, 25.05, 41.86, 46.64, 57.43, 23.31, 28.53, 53.75, 12.11, 111.83, 35.71, 19.87, 33.80, 19.05, 41.10, 26.57, 23.79, 21.65[360]

47 dG(C10)2 (polymorph I)

[13]C shifts: C3, C5, C6, C1, C8, C16, C18, C21, C23, C26, C30, C32, C35, C38, C41, C44, C47, C50, C53, C56, C61, C63, C66, C69, C72, C75, C78, C81, C84, C87 | 153.6, 152.4, 117.8, 158.5, 140.4, 86.7, 39.0, 77.0, 84.9, 64.9, 172.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 15.0, 174.3, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 23.2-34.6, 14.6[38]

[1]H shifts: H11, H15, H13, H14, H17, H19, H20, H22, H24, H27, H28, H33, H34, H36, H37, H39, H40, H42, H43, H45, H46, H48, H49, H51, H52, H54, H55, H57, H58, H59, H64, H65, H67, H68, H70, H71, H73, H74, H76, H77, H79, H80, H82, H83, H85, H86, H88, H89, H90 | 10.5, 8.7, 8.5, 6.8, 6.6, 4.0¬2.2, 4.0¬-2.2, 6.0, 3.6, 4.3, 4.3, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.5, 0.5, 0.5, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.8, 0.5, 0.5, 0.5[361]

48 Dethoxyprone (form I)

[13]C shifts: C19, C18, C25, C16, C15, C6, C21, C4, C7, C8, C1, C12, C22, C10, C5, C13, C23, C14, C9, C11, C17, C24, C3, C2, C20 | 14.4, 17.2, 17.5, 23.6, 26.1, 29.4, 30.5, 33.2, 34.1, 35.9, 37.3, 38.3, 39.3, 39.3, 39.3, 43.3, 45.2, 56.8, 58.8, 62.0, 64.5, 64.5, 64.5, 80.1, 206.2[362]

49 Dethoxyprone (form II)

[13]C shifts: C19, C18, C25, C16, C15, C6, C21, C4, C7, C8, C1, C12, C22, C10, C5, C13, C23, C14, C9, C11, C17, C24, C3, C2, C20 | 12.4, 15.8, 16.6, 22.1, 23.7, 29.6, 32.9, 31.8, 32.9, 36.4, 37.1, 36.4, 39.1, 38.2, 39.1, 44.0, 46.3, 58.2, 58.2, 63.6, 64.4, 65.8, 66.7, 79.8, 213.7[362]

50 Biphenyl

[13]C shifts: C1, C1B, C2, C2B, C6, C6B, C3, C3B, C5, C5B, C4, C4B | 137.9, 137.9, 125.5, 125.5, 125.5, 125.5, 129.8, 129.8, 129.8, 129.8, 127.2, 127.2[363]

51 β-aspartylalanine

[13]C shifts: C1, C2, C3, C4, C5, C6, C7 | 52.36, 39.9, 170.98, 175.35, 47.04, 16.45, 175.35[333]

[1]H shifts: H10, H11, H12, H7/H8, H6, H9, H2, H3, H4, H5, H1 | 0.86, 0.86, 0.86, 2.2/2.74, 4.05, 4.99, 7.59, 7.59, 7.59, 8, 12.85[23]

52 Campho[2,3-c]pyrazole

¹³C shifts: C23, C25, C33, C32, C30, C27, C26, C29, C24, C31, C28, C23, C25, C33, C32, C30, C27, C12, C14, C22, C21, C19, C16, C15, C18, C13, C20, C17, C34, C36, C44, C43, C41, C38, C37, C40, C35, C42, C39, C1, C3, C11, C10, C8, C5, C4, C7, C2, C9, C6, C56, C58, C66, C65, C63, C60, C59, C62, C59, C64, C61, C45, C47, C55, C54, C52, C49, C48, C51, C46, C49, C50 | 122.2, 48.1, 20.2, 18.9, 13.0, 33.6, 28.7, 165.0, 124.9, 61.5, 50.8, 122.2, 48.1, 20.2, 18.9, 13.0, 33.6, 121.5, 48.1, 20.2, 18.2, 12.0, 33.2, 28.2, 166.9, 125.4, 62.2, 51.0, 118.8, 46.9, 23.4, 20.9, 10.1, 32.2, 28.6, 165.5, 125.6, 62.9, 50.4, 118.8, 47.4, 21.3, 21.3, 10.8, 35.3, 29.6, 166.0, 125.0, 60.8, 50.6, 120.5, 46.9, 21.3, 20.7, 12.1, 34.9, 27.4, 165.3, 125.3, 62.4, 49.7, 118.8, 47.4, 20.0, 20.9, 11.8, 36.2, 28.9, 165.2, 126.3, 61.0, 49.8[364]

¹H shifts: H34, H35, H46, H47, H48, H43, H44, H45, H40, H41, H42, H38, H39, H36, H37, H33, H18, H19, H30, H31, H32, H27, H28, H29, H24, H25, H26, H22, H23, H20, H21, H17, H82, H83, H94, H95, H96, H91, H92, H93, H88, H89, H90, H86, H87, H84, H85, H81, H50, H51, H62, H63, H64, H59, H60, H61, H56, H57, H58, H54, H55, H52, H53, H49, H2, H3, H14, H15, H16, H11, H12, H13, H8, H9, H10, H6, H7, H4, H5, H1, H66, H67, H78, H79, H80, H75, H76, H77, H72, H73, H74, H70, H71, H68, H69, H65 | 7.4, 2.1, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 1.2, 1.2, 1.2, 1.2-1.8, 1.2-1.8, 1.6-2.4, 1.6-2.4, 13.6, 7.0, 2.5, 0.3, 0.3, 0.3, -0.2, -0.2, -0.2, 0.5, 0.5, 0.5, 1.4-1.6, 1.4-1.6, 1.1-1.5, 1.1-1.5, 13.6, 6.6, 2.7, 0.0, 0.0, 0.0, 0.4, 0.4, 0.4, 1.1, 1.1, 1.1, 1.3-1.6, 1.3-1.6, 1.6-2.4, 1.6-2.4, 13.6, 6.9, 2.5, 0.0, 0.0, 0.0, 0.3, 0.3, 0.3, 0.2, 0.2, 0.2, 0.9-1.1, 0.9-1.1, 0.9-2.3, 0.9-2.3, 13.6, 6.9, 2.6, 0.0, 0.0, 0.0, 0.4, 0.4, 0.4, 0.8, 0.8, 0.8, 0.7-1.6, 0.7-1.6, 0.9-1.5, 0.9-1.5, 13.6, 7.0, 1.9, -0.2, -0.2, -0.2, -0.3, -0.3, -0.3, 0.6, 0.6, 0.6, 0.5-0.7, 0.5-0.7, 0.4-1.9, 0.4-1.9, 13.6[364]

53 Adenosine

¹³C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10 | 152.4, 149.0, 119.3, 156.2, 139.9, 87.9, 73.4, 70.6, 85.9, 61.6[365]

¹H shifts: H3, H4, H1, H2, H5, H6, H7, H8, H9, H10, H11, H12, H13 | 8.13, 8.34, 7.33, 7.33, 5.87, 4.61, 4.14, 3.96, 3.67/3.55, 3.55/3.67, 5.43, 5.17, 5.41[365]

54 Anthranilic acid

¹³C shifts: C7, C6, C1, C2, C3, C4, C5 | 171.5, 126.2, 131.0, 122.5, 132.9, 129.6, 131.6[366]

¹H shifts: H7, H5, H6, H1, H2, H3, H4 | 16.5, 9.0, 9.0, 5.8-6.8, 5.8-6.8, 5.8-6.8, 5.8-6.8[366]

55 Cimetidine

¹³C shifts: C4, C5, C9, C6, C7, C3, C2, C10, C1, C8 | 10.6, 23.3, 27.9, 28.4, 41.2, 119.2, 120.8, 135.0, 135.9, 161.1[367]

¹H shifts: H5, H6, H7, H14, H15, H16, H10, H11, H12, H13, H8, H9, H2, H3, H4, H1 | 2.0, 2.0, 2.0, 2.0, 2.0, 2.0, 2.0–4.0, 2.0–4.0, 2.5–3.4, 2.5–3.4, 4.0–4.5, 4.0–4.5, 7.4, 8.2, 9.7, 11.6[368]

56 Flutamide

¹³C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C1, C2 | 140.9, 124.4, 116.7, 145.4, 124.4, 130.9–116.7, 122, 176.1, 35.9, 17.8–21.8, 21.8–17.8, 140.9, 124.4[118]

1H shifts: H3, H2, H1, H4, H5, H9, H10, H11, H6, H7, H8 | 7.1–8.0, 9.9, 8.0–7.1, 8.0, 2.0, 1.2, 1.2, 1.2, 1.2, 1.2, 1.2[118]

57 α-L-Glycine

13C shifts: C1, C2 | 176.2, 43.5[335]

1H shifts: H3, H4, H5, H1, H2 | 8.0, 8.0, 8.0, 3.8-2.3, 3.8-2.3[369]

58 Phenobarbital (form II)

¹³C shifts: Ca2, Ca9, Ca5, Ca7, Ca8, Ca6, Ca4, Ca14, Ca13, Ca12, Ca11, Ca10, Cb2, Cb9, Cb5, Cb7, Cb8, Cb6, Cb4, Cb14, Cb13, Cb12, Cb11, Cb10, Cc2, Cc9, Cc5, Cc7, Cc8, Cc6, Cc4, Cc14, Cc13, Cc12, Cc11, Cc10 | 147.15, 136.00, 61.68, 30.35, 6.86, 177.41, 177.41, 125.76, 131.39, 132.41, 132.81, 129.70, 148.71, 137.17, 61.00, 32.21, 7.93, 169.87, 173.20, 127.02, 130.18, 129.30, 127.02, 127.02, 147.15, 137.17, 62.37, 27.22, 8.91, 173.20, 174.96, 125.40, 133.74, 130.18, 130.18, 125.76[370]

¹H shifts: Ha1, Ha3, Ha10, Ha11, Ha12, Ha13, Ha14, Ha7a, Ha7b, Ha8a, Ha8b, Ha8c, Hb1, Hb3, Hb10, Hb11, Hb12, Hb13, Hb14, Hb7a, Hb7b, Hb8a, Hb8b, Hb8c, Hc1, Hc3, Hc10, Hc11, Hc12, Hc13, Hc14, Hc7a, Hc7b, Hc8a, Hc8b, Hc8c | 10.3, 10.3, 6.9, 6.9, 6.9, 6.9, 6.9, 2.7-1.7, 2.7-1.7, 0.6, 0.6, 0.6, 8.1, 8.1, 6.9, 6.9, 6.9, 6.9, 6.9, 2.7-1.7, 2.7-1.7, 0.6, 0.6, 0.6, 10.3, 10.3, 6.9, 6.9, 6.9, 6.9, 6.9, 2.7-1.7, 2.7-1.7, 0.6, 0.6, 0.6[370]

59 Terbutaline hemisulfate

¹³C shifts: C10A, C11A, C12A, C8A, C9A, C7A, C4A, C2A, C6A, C1A, C3A, C5A | 25.5, 25.5, 25.5, 46.9–49.8, 58.5, 70.6-72.0, 105.4–104.2, 108.6–106.4, 109.7-107.8, 143.5-141.5, 156.6, 158.4-157.8[371]

¹H shifts: H0A1, H0A2, H0A3, H1A1, H1A2, H1A3, H2A1, H2A2, H2A3, H8A1, H8A2, H7A, H2A, H6A, H4A, H5OA, H7OA, H1NA, H2NA, H3OA | 1.3, 1.3, 1.3, 1.3, 1.3, 1.3, 1.3, 1.3, 1.3, 2.8–4.0, 2.8–4.0, 4.7, 5.9–6.8, 6.8, 6.8, 7.6, 7.6, 7.6, 7.6, 10.0–11.5[371]

60 AZD7624

¹³C shifts: C127, C126, C125, C124, C123, C122, C121, C120, C119, C116, C117, C118, C113, C112, C115, C114, C109, C108, C111, C107, C106, C105, C104, C103, C102, C101 | 37.65, 51.35, 67.52, 159.13, 110.10, 130.74, 121.62, 129.98, 130.41, 33.84, 12.11–15.67, 12.11–15.67, 152.93, 120.17, 123.30, 141.12, 125.39, 11.50, 160.2, 113.46, 136.41, 122.15, 166.85, 24.65, 5.66–3.18, 5.66–3.18[69]

¹H shifts: H12E, H12F, H12G, H105, H12C, H12D, H12A, H12B, H123, H122, H121, H120, H11G, H11F, H11E, H11D, H115, H114, H11A, H11B, H11C, H106, H110, H101, H103, H10D, H10

139

C, H10B, H10A | 1.32, 1.32, 1.32, -0.60, 1.77, 1.77, 3.35, 3.35, 6.55, 8.10, 7.84, 8.21, 0.78-1.39, 0.78-1.39, 0.78-1.39, 0.78-1.39, 7.85, 6.99, 7.45, 1.49, 1.49, 6.63, 7.40, 9.12, 1.74, -1.03–-1.31, -1.03–-1.31, -1.03–-1.31, -1.03–-1.31[69]

61 α-Maltose

^{13}C shifts: C12, C6, C4, C8, C11, C5, C9, C2, C3, C10, C7, C1 | 62.2, 63,9, 69.8, 69.8, 70.8, 70.8, 71.7, 72.4, 75.8, 82.2, 92.6, 101.9[372]

^{1}H shifts: H14, H15, H6, H7, H4, H10, H13, H5, H11, H2, H3, H12, H(C7), H1 | 3.7, 3.7, 4.0, 4.0, 3.2, 3.2, 3.8, 3.8, 3.7, 3.0, 3.2, 3.5, 4.5, 4.7[372]

62 Gace_0_5H2O

^{1}H shifts: H11, H12, H13, H14, H15, H16, H4, H5, H6, H7, H8, H9, H10, H1, H2, H3, H27, H28, H29, H30, H31, H32, H20, H21, H22, H23, H24, H25, H26, H17, H18, H19 | 1.1, 1.1, 1.1, 1.1, 1.1, 1.1, 4.9, 4.7, 4.6, 4.7, 3.4, 2.7-3.8, 2.7-3.8, 11.6, 5.8-6.3, 5.8-6.3, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 6.1, 5.7, 4.8, 4.9, 3.6, 4.0-2.3, 4.0-2.3, 12.3, 5.7-6.4, 5.7-6.4[373]

63 Cocaine

^{13}C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17 | 65.95, 50.16, 66.7, 36.66, 62.63, 25.62, 25.62, 165.94, 129.37, 131.5, 131.5, 134.53, 133.5, 131.5, 172.18, 50.16, 41.52[118]

^{1}H shifts: H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, H17, H18, H19, H20, H21 | 3.76, 3.78, 5.63, 3.06, 3.32, 3.49, 3.38, 2.91, 2.56, 2.12, 8.01, 8.01, 8.01, 8.01, 8.01, 3.78, 3.78, 3.78, 1.04, 1.04, 1.04[118]

64 Hydroflumethiazide

^{13}C shifts: C(1), C(2), C(7), C(8), C(4), C(5), C(3), C(6) | 56.24, 117.91, 119.10, 121, 123.89, 128.36, 129.77, 146.23[333]

^{1}H shifts: H(2), H(3), H(8), H(5), H(4), H(6), H(7), H(8) | 5.5, 5.7, 7.1, 7.1, 7.1, 7.1, 7.1, 7.8[333]

65 Metformin·HCl (form A)

^{13}C shifts: C3, C4, C1, C2 | 37.68, 39.73, 156.63, 159.95[374]

^{1}H shifts: H9, H8, H7, H10, H11, H12, H2, H3, H5, H4, H6, H1 | 2.9, 2.9, 2.9, 3.2, 3.2, 3.2, 6.16, 6.85, 7.85, 7.85, 7.85, 8.35[374]

66 Paracetamol (form II)

^{13}C shifts: C8, C3, C5, C6, C2, C1, C4, C7 | 24.9, 117, 118.3, 120., 120., 131.5, 153., 170.5[375]

¹H shifts: H6, H8, H9, H3, H2, H5, H1, H4, H7 | 1.8, 1.8, 1.8, 4.7, 4.7, 6.4, 7.3, 9.4, 9.7[375]

67 Penicillin G (K salt)

¹³C shifts: C6, C7, C10, C4, C1, C3, C2, C13, C15, C12, C16, C14, C11, C9, C8, C5 | 27.1, 37.6, 43.3, 60.5, 65.2, 68.2, 74.9, 128.7, 128.7, 128.7, 130.6, 130.6, 136.4, 172.3, 172.9, 176.1[288]

¹H shifts: H7, H8, H9, H4, H5, H6, H12, H1, H11, H3, H10, H2, H14, H13, H15, H17, H16 | 0.9, 0.9, 0.9, 1.7, 1.7, 1.7, 3.9, 4.1, 4.7, 5.7, 6.2, 6.4, 7.1, 7.1, 7.1, 7.1, 7.1[288]

68 Salicylic acid

¹³C shifts: C1, C2, C3, C4, C5, C6, C7 | 111.83, 161.91, 118.14, 138.35, 120.93, 133, 175.92[333]

¹H shifts: H1, H2, H3, H4, H5, H6 | 12.26, 9.7, 6.3, 7.55, 7.55, 7.55[333]

69 L-Tyrosine·HCl

¹³C shifts: C1, C3, C6, C2, C4, C7, C8, C2, C5 | 172.04, 56.42, 129.20, 133.07, 36.88, 114.41, 117.98, 151.68, 126.13[376]

¹H shifts: H2, H3, H4, H6, H7, H8, H10, H11, H12, H1, H5, H9 | 10.65, 5.01, 3.01, 13.24, 5.85, 68.05, 7.10, 5.70, 5.10, 8.20, 8.20, 8.20[376]

70 4,5-Dimethylimidazole

¹³C shifts: C9, C11, C12, C14, C13 | 132, 132, 122.1, 10.6, 7.9[377]

¹H shifts: H91, H131, H132, H133, H141, H142, H143, H81 | 4.8, 1.4, 1.4, 1.4, 0.7, 0.7, 0.7, 13.0 [377]

71 3,5-Dimethylpyrazole

¹³C shifts: C1E, C2, C1, C3E, C3 | 147.2, 103, 140.7, 14.1, 10.6[377]

¹H shifts: H2, H3E, H4E, H5E, H3, H4, H5, H1 | 5.2, 1.5, 1.5, 1.5, 1.4, 1.4, 1.4, 15[377]

72 Flurbiprofen

¹³C shifts: C13, C12, C9, C1, C6, C8, C10, C2, C14, C15, C3, C4, C11, C5, C8 | 136.6, 128.4, 140.6, 160.0, 184.0, 117.1, 123.7, 129.0, 129.0, 129.0, 129.0, 129.0, 129.0, 46.6, 16.1[373]

¹H shifts: H6, H7, H8, H5, H9, H10, H11, H12, H4, H3, H13 | 0.9, 0.9, 0.9, 2.9, 6.7, 6.7, 6.7, 6.7, 6.7, 6.7, 6.7[373]

73 L-Histidine

¹³C shifts: C6, C5, C4, C3, C1, C2 | 175.6, 57.0, 27.0, 137.7, 135.3, 113.6[378]

¹H shifts: H31N, H32N, H33N, H5, H4A, H4B, H2N, H2, H1 | 9.0, 9.0, 9.0, 4.3, 2.7, 2.7, 13.7, 4.9, 6.1[379]

74 γ-Indomethacin

¹³C shifts: C11, C13, C14, C15, C16, C17, C18, C19, C9, C10, C12, C8, C1, C2, C7, C6, C5, C4, C3 | 134.5, 112.7, 132.0, 97.7, 156.7, 112.4, 115.5, 131.1, 28.1, 179.0, 55.1, 13.5, 167.7, 136.7, 131.8, 126.9, 140.1, 129.8, 131.8[380]

¹H shifts: H4, H6, H7, H9, H11, H12, H15, H16, H18, H19, OH | 5.8, 6.1, 5.8, 1.7, 2.2, 1.8, 7.3, 5.7, 7.2, 7.3, 12.7[380]

75 Naproxen

¹³C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14 | 124.0, 134.9, 129.1, 130.6, 119.2, 158.1, 104.3, 129.1, 47.0, 179.0, 17.5, 53.2, 134.9, 134.9[381]

¹H shifts: H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14 | 7.0, 6.1, 3.8, 4.5, 4.1, 5.9, 3.2, 1.8, 1.8, 1.8, 2.3, 2.3, 2.3, 11.5[381]

76 Phenobarbital (form III)

¹³C shifts: C4, C6, C2, C9, C13, C11, C12, C14, C10, C5, C7, C8, C4 | 147.20, 147.20, 149.01, 137.56, 130.70, 129.92, 129.53, 127.57, 127.57, 62.27, 27.12, 11.36, 147.20[370]

¹H shifts: H1, H3, H10, H11, H12, H13, H14, H7A, H7B, H8A, H8B, H8C | 10.0, 10.0, 7.4, 7.4, 7.4, 7.4, 7.4, 0.8, 0.8, 0.0, 0.0, 0.0[370]

77 Theophylline

¹³C shifts: C1, C2, C3, C4, C5, C6, C7 | 150.8, 146.1, 105.8, 155.0, 140.8, 29.9, 29.9[118]

¹H shifts: H8, H10a, H10b, H10c, H12a, H12b, H12c, H7 | 7.7, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 154.9[118]

78 Thymol

¹³C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10 | 18.7, 23.6, 25.5, 26.1, 116.9, 123.6, 126.3, 131.7, 138.4, 150.2[116]

¹H shifts: H1, H2, H3, H4, H8, H9, H10, H5, H6, H7, H11, H12, H13, H14 | 0.42, 1.05, 1.45, 3.38, 5.40, 5.40, 5.40, 6.19, 6.19, 6.19, 7.08, 7.08, 7.08, 9.99[116]

79 L-Tyrosine

^{13}C shifts: C1, C4, C8, C9, C5, C6, C8, C3, C2 | 175.8, 56.7, 37, 123.9, 131, 118.3, 155.8, 118.3, 131[335]

^{1}H shifts: H1, H6, H10, H2, H3, H7, H4, H5, H8, H9, H11 | 8.70, 8.70, 8.70, 9.90, 3.05, 3.05, 6.40, 4.65, 6.50, 4.25, 4.80[376]

80 Uracil

^{13}C shifts: C2, C1, C4, C3 | 170.9, 151.7, 147.0, 99.9[382]

^{1}H shifts: H3, H2, H1, H4 | 11.2, 10.8, 7.5, 6.0[382]

81 Anhydrous ciprofloxacin

^{13}C shifts: C1, C2, C3, C5, C10, C17, C16, C15, C14, C13, C4, C6, C7, C8, C9, C11, C12 | 35.6, 7.6, 8.2, 143.4, 115.6, 106.4, 47.9, 42.9, 42.1, 46.0, 138.8, 120.2, 172.7, 173.1, 123.3, 151.6, 141.9[383]

^{1}H shifts: H11, H21, H22, H32, H31, H51, H101, H171, H161, H162, H151, H152, H142, H141, H131, H132, H311, H312 | 4.0, 1.6, 0.2, 0.0, -1.9, 7.1, 7.8, 6.3, 3.5, 3.5, 3.3, 3.3, 4.6, 3.3, 2.3, 3.5, 7.2, 14.4[383]

82 Indomethacin nicotinamide

^{13}C shifts: C11, C10, C9, C7, C6, C4, C2, C1, C22, C25, C18, C12, C29, C31, C32, C34, C36, C38, C40, C43, C42, C50, C48, C46, C52 | 133.5, 112.6, 130.8, 103.6, 156.3, 106.5, 113.1, 128.8, 30.4, 176.0, 55.2, 12.9, 167.7, 130.8, 128.8, 130.8, 146.0, 127.9, 130.8, 147.0, 130.8, 139.5, 125.8, 149.7, 167.7[384]

^{1}H shifts: H8, H5, H3, H23, H24, H19, H20, H21, H13, H14, H15, H33, H35, H39, H41, H28, H44, H51, H49, H47, H55, H56 | 6.8, 5.5, 7.3, 3.4, 3.4, 2.9, 2.9, 2.9, 0.9, 0.9, 0.9, 6.0, 7.3, 6.0, 6.4, 16.3, 9.8, 7.7, 8.3, 9.8, 9.0, 7.3[384]

83 Phenylphosphonic acid

^{13}C shifts: C1, C2, C6, C4, C3, C5 | 128.0, 129.7, n.a., 135.0, 132, n.a.[335]

^{1}H shifts: H1, H2, H4, H2, H3, H5, H6, H1, H2, H4, H2, H3, H5 | 12.7, 11.5, 8.5, 6.5, 6,8, 7.1, 6.2, 12.7, 11.5, 8.5, 6.5, 6,8, 7.1[335]

84 Sulfathiazole oxalate

^{13}C shifts: C4, C3, C5, C2, C6, C1, C7, C9, C8, C10, C11, C6, C1, C7 | 138.6, 124.7-126.8, 124.7–126.8, 130.6, 130.6, 142.8, 170.0, 124.7, 107.7, 162.6–168.8, 162.6–168.8, 130.6, 142.8, 170.0[385]

¹H shifts: H1A, H1B, H3A, H5, H2, H6, H3, H8, H9, H4, H1C | 9.9, 9.9, 7.0, 7.0, 8.1, 8.1, 12.7, 5.9, 3.9, 9.9, 9.9[385]

85 Diethylcarbamazine citrate salt (phase I)

¹³C shifts: C10A, C9A, C7A, C8A, C6A, C4A, C3A, C5A, C2A, C1A, C6C, C3C, C2C, C1C, C4C, C5C, C10A, C9A, C7A, C8A, C6A, C4A | 14.2, 42.1, 40.7, 13.6, 164.9, 46.8, 51.3, 44.8, 53.2, 44.5, 177.6, 72.8, 48.2, 177.3, 42.4, 177.3, 14.2, 42.1, 40.7, 13.6, 164.9, 46.8[386]

¹H shifts: H10A1, H10A2, H10A3, H9A1, H9A2, H7A1, H7A2, H8A1, H8A2, H8A3, H4A1, H4A2, H3A1, H3A2, H5A1, H5A2, H5A3, H2A1, H2A2, H1A1, H1A2, H2A, H2C1, H2C2, H4C1, H4C2, H6C, H3C, H7C | 0.5, 0.5, 0.5, 2.5, 2.5, 2.3, 2.3, 0.5, 0.5, 0.5, 2.7, 2.7, 2.8, 2.8, 2.3, 2.3, 2.3, 2.7, 2.7, 2.5, 2.5, 9.1, 2.1, 2.1, 2.4, 2.4, 16.2, 4.8, 12.8[386]

86 Dithianon pyrimethanil (1_1 co-crystal)

¹³C shifts: C35, C34, C31, C2, C19, C3, C17, C23, C27, C13, C10, C26, C16, C9, C25, C23, C24, C14, C15, C22, C5, C28, C32, C30, C7, C11 | 23.9, 25.7, 112.6, 114.4, 114.4, 114.4, 114.4, 119.4, 120.3, 125.7, 125.7, 125.7, 129.8, 129.8, 130.2, 131.1, 131.2, 133.9, 136.8, 141.5, 141.4, 160.1, 168.2, 168.2, 176.5, 178.2[75]

¹H shifts: H351, H352, H353, H341, H342, H343, H311, H231, H271, H131, H261, H351, H251, H241, H141, H151 | 1.9, 1.9, 1.9, 2.0, 2.0, 2.0, 4.0, 9.1, 8.0, 7.4, 7.4, 8.2, 7.8, 7.7, 6.2, 7.7[75]

87 γ-L-glutamyl-L-cysteinyl-glycine

¹³C shifts: C1, C2, C3, C4, C5, C6, C7, C8, C9, C10 | 51.70, 32.10, 32.50, 178.60, 174.30, 56.10, 27.00, 170.70, 41.30, 173.10[387]

¹H shifts: H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, H17 | 3.00, 13.85, 9.10, 9.10, 9.10, 9.00, 8.60, 3.80, 2.50, 2.00, 2.40, 2.70, 4.65, 3.15, 2.85, 3.75, 4.20[387]

88 Terbutaline hemisulphate acetic acid solvate

¹³C shifts: C2, C4, C6, C1, C5, C3, C7, C9, C8, C10, C11, C12, C18, C16, C14, C13, C15, C14, C19, C21, C20, C22, C23, C24, C26A, C25, C27, C28 | 158.7, 156.4, 142.2, 108.3, 106.1, 101.3, 71.4, 59.2, 49.0, 25.8, 25.8, 25.8, 157.9, 157.2, 141.4, 108.0, 107.6, 104.3, 71.4, 57.6, 47.6, 24.7, 24.7, 24.7, 172.2, 19.9, 176.2, 22.2[388]

¹H shifts: H1, H5, H3, H7, H81, H82, H101, H102, H103, H111, H112, H113, H121, H122, H123, H15, H13, H17, H19, H201, H202, H221, H222, H223, H231, H232, H233, H241, H242, H243, H251, H252, H253, H271, H272, H273 | 4.4, 6.7, 5.6, 3.9, 2.3, 2.3, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 5.8, 6.7, 5.6, 3.9, 2.3, 2.3, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 1.1, 1.1, 1.1, 1.0, 1.0, 1.0[388]

References

1.	Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P., A New Method of Measuring Nuclear Magnetic Moment. *Physical Review* **1938,** *53* (4), 318-318.
2.	Bloch, F.; Hansen, W. W.; Packard, M., Nuclear Induction. *Physical Review* **1946,** *69* (3-4), 127-127.
3.	Purcell, E. M.; Torrey, H. C.; Pound, R. V., Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review* **1946,** *69* (1-2), 37-38.
4.	Hahn, E. L., Spin Echoes. *Physical Review* **1950,** *80* (4), 580-594.
5.	Dickinson, W. C., Dependence of the F19 Nuclear Resonance Position on Chemical Compound. *Physical Review* **1950,** *77* (5), 736-737.
6.	Proctor, W. G.; Yu, F. C., The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound. *Physical Review* **1950,** *77* (5), 717-717.
7.	Emsley, L.; Laws, D. D.; Pines, A. In *Lectures on pulsed NMR (3rd edition)*, International School of Physics Enrico Fermi on Magnetic Resonance and Brain Function: Approaches from Physics, Varenna Lake Como, Italy, Jun 23-Jul 03; Varenna Lake Como, Italy, 1998; pp 45-210.
8.	Lowe, I. J., Free Induction Decays of Rotating Solids. *Physical Review Letters* **1959,** *2* (7), 285-287.
9.	Andrew, E. R.; Bradbury, A.; Eades, R. G., Nuclear Magnetic Resonance Spectra from a Crystal rotated at High Speed. *Nature* **1958,** *182* (4650), 1659-1659.
10.	Lee, M.; Goldburg, W. I., Nuclear-Magnetic-Resonance Line Narrowing by a Rotating rf Field. *Physical Review* **1965,** *140* (4A), A1261-A1271.
11.	Waugh, J. S.; Huber, L. M.; Haeberlen, U., Approach to High-Resolution nmr in Solids. *Physical Review Letters* **1968,** *20* (5), 180-182.
12.	Leskes, M.; Steuernagel, S.; Schneider, D.; Madhu, P. K.; Vega, S., Homonuclear dipolar decoupling at magic-angle spinning frequencies up to 65kHz in solid-state nuclear magnetic resonance. *Chem Phys Lett* **2008,** *466* (1), 95-99.
13.	Salager, E.; Stein, R. S.; Steuernagel, S.; Lesage, A.; Elena, B.; Emsley, L., Enhanced sensitivity in high-resolution 1H solid-state NMR spectroscopy with DUMBO dipolar decoupling under ultra-fast MAS. *Chem Phys Lett* **2009,** *469* (4), 336-341.
14.	Salager, E.; Dumez, J.-N.; Stein, R. S.; Steuernagel, S.; Lesage, A.; Elena-Herrmann, B.; Emsley, L., Homonuclear dipolar decoupling with very large scaling factors for high-resolution ultrafast magic angle spinning 1H solid-state NMR spectroscopy. *Chem Phys Lett* **2010,** *498* (1), 214-220.
15.	Gan, Z.; Madhu, P. K.; Amoureux, J.-P.; Trébosc, J.; Lafon, O., A tunable homonuclear dipolar decoupling scheme for high-resolution proton NMR of solids from slow to fast magic-angle spinning. *Chem Phys Lett* **2011,** *503* (1), 167-170.
16.	Nishiyama, Y.; Lu, X.; Trébosc, J.; Lafon, O.; Gan, Z.; Madhu, P. K.; Amoureux, J.-P., Practical choice of 1H–1H decoupling schemes in through-bond 1H–{X} HMQC experiments at ultra-fast MAS. *J Magn Reson* **2012,** *214*, 151-158.
17.	Burum, D. P.; Rhim, W. K., Analysis of multiple pulse NMR in solids. III. *The Journal of Chemical Physics* **1979,** *71* (2), 944-956.
18.	Bax, A.; Freeman, R.; Frenkiel, T. A., An Nmr Technique for Tracing out the Carbon Skeleton of an Organic-Molecule. *J Am Chem Soc* **1981,** *103* (8), 2102-2104.

19. Bielecki, A.; Kolbert, A. C.; Levitt, M. H., Frequency-switched pulse sequences: Homonuclear decoupling and dilute spin NMR in solids. *Chem Phys Lett* **1989,** *155* (4), 341-346.

20. Vinogradov, E.; Madhu, P. K.; Vega, S., High-resolution proton solid-state NMR spectroscopy by phase-modulated Lee–Goldburg experiment. *Chem Phys Lett* **1999,** *314* (5), 443-450.

21. Halse, M. E.; Emsley, L., Improved Phase-Modulated Homonuclear Dipolar Decoupling for Solid-State NMR Spectroscopy from Symmetry Considerations. *The Journal of Physical Chemistry A* **2013,** *117* (25), 5280-5290.

22. Levitt, M. H., Symmetry in the design of NMR multiple-pulse sequences. *The Journal of Chemical Physics* **2008,** *128* (5), 052205.

23. Elena, B.; de Paepe, G.; Emsley, L., Direct spectral optimisation of proton-proton homonuclear dipolar decoupling in solid-state NMR. *Chem Phys Lett* **2004,** *398* (4-6), 532-538.

24. Paruzzo, F. M.; Emsley, L., High-resolution 1H NMR of powdered solids by homonuclear dipolar decoupling. *J Magn Reson* **2019,** *309*, 106598.

25. Bloom, A. L.; Shoolery, J. N., Effects of Perturbing Radiofrequency Fields on Nuclear Spin Coupling. *Physical Review* **1955,** *97* (5), 1261-1265.

26. Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G., Heteronuclear decoupling in rotating solids. *The Journal of Chemical Physics* **1995,** *103* (16), 6951-6958.

27. Vinther, J. M.; Nielsen, A. B.; Bjerring, M.; Eck, E. R. H. v.; Kentgens, A. P. M.; Khaneja, N.; Nielsen, N. C., Refocused continuous-wave decoupling: A new approach to heteronuclear dipolar decoupling in solid-state NMR spectroscopy. *The Journal of Chemical Physics* **2012,** *137* (21), 214202.

28. Thakur, R. S.; Kurur, N. D.; Madhu, P. K., Swept-frequency two-pulse phase modulation for heteronuclear dipolar decoupling in solid-state NMR. *Chem Phys Lett* **2006,** *426* (4), 459-463.

29. Fung, B. M.; Khitrin, A. K.; Ermolaev, K., An improved broadband decoupling sequence for liquid crystals and solids. *J Magn Reson* **2000,** *142* (1), 97-101.

30. Weingarth, M.; Tekely, P.; Bodenhausen, G., Efficient heteronuclear decoupling by quenching rotary resonance in solid-state NMR. *Chem Phys Lett* **2008,** *466* (4), 247-251.

31. Equbal, A.; Bjerring, M.; Madhu, P. K.; Nielsen, N. C., A unified heteronuclear decoupling strategy for magic-angle-spinning solid-state NMR spectroscopy. *The Journal of Chemical Physics* **2015,** *142* (18), 184201.

32. Sharma, K.; Equbal, A.; Nielsen, N. C.; Madhu, P. K., A unified heteronuclear decoupling picture in solid-state NMR under low radio-frequency amplitude and fast magic-angle-spinning frequency regime. *The Journal of Chemical Physics* **2019,** *150* (14), 144201.

33. Gerstein, B. C.; Pembleton, R. G.; Wilson, R. C.; Ryan, L. M., High resolution NMR in randomly oriented solids with homonuclear dipolar broadening: Combined multiple pulse NMR and magic angle spinning. *The Journal of Chemical Physics* **1977,** *66* (1), 361-362.

34. Lesage, A.; Auger, C.; Caldarelli, S.; Emsley, L., Determination of through-bond carbon-carbon connectivities in solid-state NMR using the INADEQUATE experiment. *J Am Chem Soc* **1997,** *119* (33), 7867-7868.

35. Cadars, S.; Sein, J.; Duma, L.; Lesage, A.; Pham, T. N.; Baltisberger, J. H.; Brown, S. P.; Emsley, L., The refocused INADEQUATE MAS NMR experiment in multiple spin-systems: Interpreting observed correlation peaks and optimising lineshapes. *J Magn Reson* **2007,** *188* (1), 24-34.

36.   Brown, S. P.; Pérez-Torralba, M.; Sanz, D.; Claramunt, R. M.; Emsley, L., Determining hydrogen-bond strengths in the solid state by NMR: the quantitative measurement of homonuclear J couplings. *Chem Commun* **2002,** (17), 1852-1853.

37.   Benedict, H.; Shenderovich, I. G.; Malkina, O. L.; Malkin, V. G.; Denisov, G. S.; Golubev, N. S.; Limbach, H.-H., Nuclear Scalar Spin–Spin Couplings and Geometries of Hydrogen Bonds. *J Am Chem Soc* **2000,** *122* (9), 1979-1988.

38.   Pham, T. N.; Griffin, J. M.; Masiero, S.; Lena, S.; Gottarelli, G.; Hodgkinson, P.; Filip, C.; Brown, S. P., Quantifying hydrogen-bonding strength: the measurement of 2hJNN couplings in self-assembled guanosines by solid-state 15N spin-echo MAS NMR. *Physical Chemistry Chemical Physics* **2007,** *9* (26), 3416-3423.

39.   Zhang, W.; Yoon, M.-K.; Meredith, R. J.; Zajicek, J.; Oliver, A. G.; Hadad, M.; Frey, M. H.; Carmichael, I.; Serianni, A. S., 13C–13C spin-coupling constants in crystalline 13C-labeled saccharides: conformational effects interrogated by solid-state 13C NMR spectroscopy. *Physical Chemistry Chemical Physics* **2019,** *21* (42), 23576-23588.

40.   Martineau, C.; Fayon, F.; Legein, C.; Buzaré, J.-Y.; Silly, G.; Massiot, D., Accurate heteronuclear J-coupling measurements in dilute spin systems using the multiple-quantum filtered J-resolved experiment. *Chem Commun* **2007,** (26), 2720-2722.

41.   Wegner, S.; van Wüllen, L.; Tricot, G., The structure of phosphate and borosilicate glasses and their structural evolution at high temperatures as studied with solid state NMR spectroscopy: Phase separation, crystallisation and dynamic species exchange. *Solid State Sciences* **2010,** *12* (4), 428-439.

42.   Pake, G. E., Nuclear resonance absorption in hydrated crystals: Fine structure of the proton line. *The Journal of Chemical Physics* **1948,** *16* (4), 327-336.

43.   Hodgkinson, P., NMR crystallography of molecular organics. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2020,** *118-119*, 10-53.

44.   Bryce, D., NMR crystallography: structure and properties of materials from solid-state nuclear magnetic resonance observables. *IUCrJ* **2017,** *4* (4), 350-359.

45.   Martineau-Corcos, C.; Senker, J.; Taulelle, F., NMR Crystallography. *Annual Reports on NMR Spectroscopy* **2013,** *82*.

46.   Ashbrook, S. E.; McKay, D., Combining solid-state NMR spectroscopy with first-principles calculations – a guide to NMR crystallography. *Chem Commun* **2016,** *52* (45), 7186-7204.

47.   Vogt, F. G., Evolution of solid-state NMR in pharmaceutical analysis. *Future Med Chem* **2010,** *2* (6), 915-21.

48.   Dudek, M. K.; Kazmierski, S.; Kostrzewa, M.; Potrzebowski, M. J., Chapter One - Solid-State NMR Studies of Molecular Crystals. In *Annual Reports on NMR Spectroscopy*, Webb, G. A., Ed. Academic Press: 2018; Vol. 95, pp 1-81.

49.   Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J., Ritonavir: An Extraordinary Example of Conformational Polymorphism. *Pharmaceut Res* **2001,** *18* (6), 859-866.

50.   Rietveld, I. B.; Céolin, R., Rotigotine: Unexpected Polymorphism with Predictable Overall Monotropic Behavior. *J Pharm Sci-Us* **2015,** *104* (12), 4117-4122.

51.   Borisov, S. V.; Podberezskaya, N. V., X-ray diffraction analysis: A brief history and achievements of the first century. *Journal of Structural Chemistry* **2012,** *53* (1), 1-3.

52.   Sumner, T., Dazzling History. *Science* **2014,** *343* (6175), 1092-1093.

53.   Widdifield, C. M.; Robson, H.; Hodgkinson, P., Furosemide's one little hydrogen atom: NMR crystallography structure verification of powdered molecular organics. *Chem Commun* **2016,** *52* (40), 6685-6688.

54.   Jastrzebska, I.; Pawlak, T.; Arcos-Ramos, R.; Florez-Lopez, E.; Farfán, N.; Czajkowska-Szczykowska, D.; Maj, J.; Santillan, R.; Morzycki, J. W.; Potrzebowski, M. J., Synthesis, Structure, and Local Molecular Dynamics for Crystalline Rotors Based on Hecogenin/Botogenin Steroidal Frameworks. *Cryst Growth Des* **2016,** *16* (10), 5698-5709.

55.   Braga, D.; Chelazzi, L.; Grepioni, F.; Dichiarante, E.; Chierotti, M. R.; Gobetto, R., Molecular Salts of Anesthetic Lidocaine with Dicarboxylic Acids: Solid-State Properties and a Combined Structural and Spectroscopic Study. *Cryst Growth Des* **2013,** *13* (6), 2564-2572.

56.   Vogt, F. G.; Katrincic, L. M.; Long, S. T.; Mueller, R. L.; Carlton, R. A.; Sun, Y. T.; Johnson, M. N.; Copley, R. C. B.; Light, M. E., Enantiotropically-related polymorphs of {4-(4-chloro-3-fluorophenyl)-2-[4-(methyloxy)phenyl]-1,3-thiazol-5-yl} acetic acid: Crystal structures and multinuclear solid-state NMR. *J Pharm Sci-Us* **2008,** *97* (11), 4756-4782.

57.   Leclaire, J.; Poisson, G.; Ziarelli, F.; Pepe, G.; Fotiadu, F.; Paruzzo, F. M.; Rossini, A. J.; Dumez, J.-N.; Elena-Herrmann, B.; Emsley, L., Structure elucidation of a complex CO2-based organic framework material by NMR crystallography. *Chemical Science* **2016,** *7* (7), 4379-4390.

58.   Busi, B.; Yarava, J. R.; Hofstetter, A.; Salvi, N.; Cala-De Paepe, D.; Lewandowski, J. R.; Blackledge, M.; Emsley, L., Probing Protein Dynamics Using Multifield Variable Temperature NMR Relaxation and Molecular Dynamics Simulation. *The Journal of Physical Chemistry B* **2018,** *122* (42), 9697-9702.

59.   Busi, B.; Yarava, J. R.; Bertarello, A.; Freymond, F.; Adamski, W.; Maurin, D.; Hiller, M.; Oschkinat, H.; Blackledge, M.; Emsley, L., Similarities and Differences among Protein Dynamics Studied by Variable Temperature Nuclear Magnetic Resonance Relaxation. *The Journal of Physical Chemistry B* **2021,** *125* (9), 2212-2221.

60.   Gorelik, T. E.; Czech, C.; Hammer, S. M.; Schmidt, M. U., Crystal structure of disordered nanocrystalline αII-quinacridone determined by electron diffraction. *CrystEngComm* **2016,** *18* (4), 529-535.

61.   Das, P. P.; Mugnaioli, E.; Nicolopoulos, S.; Tossi, C.; Gemmi, M.; Galanis, A.; Borodi, G.; Pop, M. M., Crystal Structures of Two Important Pharmaceuticals Solved by 3D Precession Electron Diffraction Tomography. *Organic Process Research & Development* **2018,** *22* (10), 1365-1372.

62.   Gruene, T.; Wennmacher, J. T. C.; Zaubitzer, C.; Holstein, J. J.; Heidler, J.; Fecteau-Lefebvre, A.; De Carlo, S.; Müller, E.; Goldie, K. N.; Regeni, I.; Li, T.; Santiso-Quinones, G.; Steinfeld, G.; Handschin, S.; van Genderen, E.; van Bokhoven, J. A.; Clever, G. H.; Pantelic, R., Rapid Structure Determination of Microcrystalline Molecular Compounds Using Electron Diffraction. *Angewandte Chemie International Edition* **2018,** *57* (50), 16313-16317.

63.   Jones, C. G.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H. M.; Gonen, T., The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Central Science* **2018,** *4* (11), 1587-1592.

64.   Mugnaioli, E.; Gorelik, T.; Kolb, U., "Ab initio" structure solution from electron diffraction data obtained by a combination of automated diffraction tomography and precession technique. *Ultramicroscopy* **2009,** *109* (6), 758-765.

65.   Gorelik, T.; Matveeva, G.; Kolb, U.; Schleuß, T.; Kilbinger, A. F. M.; van de Streek, J.; Bohle, A.; Brunklaus, G., H-bonding schemes of di- and tri-p-benzamides assessed by a

combination of electron diffraction, X-ray powder diffraction and solid-state NMR. *CrystEngComm* **2010,** *12* (6), 1824-1832.

66.   Fletton, R. A.; Harris, R. K.; Kenwright, A. M.; Lancaster, R. W.; Packer, K. J.; Sheppard, N., A comparative spectroscopic investigation of three pseudopolymorphs of testosterone using solid-state i.r. and high-resolution solid-state NMR. *Spectrochimica Acta Part A: Molecular Spectroscopy* **1987,** *43* (9), 1111-1120.

67.   Fletton, R. A.; Lancaster, R. W.; Harris, R. K.; Kenwright, A. M.; Packer, K. J.; Waters, D. N.; Yeadon, A., A comparative spectroscopic investigation of two polymorphs of 4'-methyl-2'-nitroacetanilide using solid-state infrared and high-resolution solid-state nuclear magnetic resonance spectroscopy. *Journal of the Chemical Society, Perkin Transactions 2* **1986,** (11), 1705-1709.

68.   Harris, R. K.; Say, B. J.; Yeung, R. R.; Fletton, R. A.; Lancaster, R. W., Cross-polarization/magic-angle spinning NMR studies of polymorphism: Androstanolone. *Spectrochimica Acta Part A: Molecular Spectroscopy* **1989,** *45* (4), 465-469.

69.   Widdifield, C. M.; Nilsson Lill, S. O.; Broo, A.; Lindkvist, M.; Pettersen, A.; Svensk Ankarberg, A.; Aldred, P.; Schantz, S.; Emsley, L., Does Z' equal 1 or 2? Enhanced powder NMR crystallography verification of a disordered room temperature crystal structure of a p38 inhibitor for chronic obstructive pulmonary disease. *Physical Chemistry Chemical Physics* **2017,** *19* (25), 16650-16661.

70.   Fernandes, J. A.; Sardo, M.; Mafra, L.; Choquesillo-Lazarte, D.; Masciocchi, N., X-ray and NMR Crystallography Studies of Novel Theophylline Cocrystals Prepared by Liquid Assisted Grinding. *Cryst Growth Des* **2015,** *15* (8), 3674-3683.

71.   Li, P.; Chu, Y.; Wang, L.; Wenslow, R. M.; Yu, K.; Zhang, H.; Deng, Z., Structure determination of the theophylline–nicotinamide cocrystal: a combined powder XRD, 1D solid-state NMR, and theoretical calculation study. *CrystEngComm* **2014,** *16* (15), 3141-3147.

72.   Watts, A. E.; Maruyoshi, K.; Hughes, C. E.; Brown, S. P.; Harris, K. D. M., Combining the Advantages of Powder X-ray Diffraction and NMR Crystallography in Structure Determination of the Pharmaceutical Material Cimetidine Hydrochloride. *Cryst Growth Des* **2016,** *16* (4), 1798-1804.

73.   Hušák, M.; Jegorov, A.; Rohlíček, J.; Fitch, A.; Czernek, J.; Kobera, L.; Brus, J., Determining the Crystal Structures of Peptide Analogs of Boronic Acid in the Absence of Single Crystals: Intricate Motifs of Ixazomib Citrate Revealed by XRPD Guided by ss-NMR. *Cryst Growth Des* **2018,** *18* (6), 3616-3625.

74.   Seyfarth, L.; Sehnert, J.; El-Gamel, N. E. A.; Milius, W.; Kroke, E.; Breu, J.; Senker, J., Structure elucidation of cyameluric acid by combining solid-state NMR spectroscopy, molecular modeling and direct-space methods. *Journal of Molecular Structure* **2008,** *889* (1), 217-228.

75.   Poppler, A. C.; Corlett, E. K.; Pearce, H.; Seymour, M. P.; Reid, M.; Montgomery, M. G.; Brown, S. P., Single-crystal X-ray diffraction and NMR crystallography of a 1:1 cocrystal of dithianon and pyrimethanil. *ACTA CRYSTALLOGRAPHICA SECTION C-STRUCTURAL CHEMISTRY* **2017,** *73*, 149-+.

76.   Spinelli, F.; Dichiarante, E.; Curzi, M.; Giaffreda, S. L.; Chierotti, M. R.; Gobetto, R.; Rossi, F.; Chelazzi, L.; Braga, D.; Grepioni, F., Molecular Salts of the Antidepressant Venlafaxine: An Effective Route to Solubility Properties Modifications. *Cryst Growth Des* **2017,** *17* (8), 4270-4279.

77.   Harris, R. K., Applications of solid-state NMR to pharmaceutical polymorphism and related matters*. *Journal of Pharmacy and Pharmacology* **2010,** *59* (2), 225-239.

78.   Tremayne, M.; Kariuki, B. M.; Harris, K. D. M., Structure Determination of a Complex Organic Solid from X-Ray Powder Diffraction Data by a Generalized Monte Carlo Method: The Crystal Structure of Red Fluorescein. *Angewandte Chemie International Edition in English* **1997,** *36* (7), 770-772.

79.   Middleton, D. A.; Peng, X.; Saunders, D.; Shankland, K.; David, W. I. F.; Markvardsen, A. J., Conformational analysis by solid-state NMR and its application to restrained structure determination from powder diffraction data. *Chem Commun* **2002,** (17), 1976-1977.

80.   Harris, K. D. M.; Xu, M., Crystal Structure Determination by the Combined Analysis of NMR and Powder Diffraction Data. In *eMagRes*.

81.   Harris, R. K.; Ghi, P. Y.; Hammond, R. B.; Ma, C. Y.; Roberts, K. J.; Yates, J. R.; Pickard, C. J., Solid-state NMR and computational studies of 4-methyl-2-nitroacetanilide. *Magn Reson Chem* **2006,** *44* (3), 325-33.

82.   Paluch, P.; Pawlak, T.; Oszajca, M.; Lasocha, W.; Potrzebowski, M. J., Fine refinement of solid state structure of racemic form of phospho-tyrosine employing NMR Crystallography approach. *Solid State Nucl Mag* **2015,** *65*, 2-11.

83.   Nilsson Lill, S. O.; Widdifield, C. M.; Pettersen, A.; Svensk Ankarberg, A.; Lindkvist, M.; Aldred, P.; Gracin, S.; Shankland, N.; Shankland, K.; Schantz, S.; Emsley, L., Elucidating an Amorphous Form Stabilization Mechanism for Tenapanor Hydrochloride: Crystal Structure Analysis Using X-ray Diffraction, NMR Crystallography, and Molecular Modeling. *Molecular Pharmaceutics* **2018,** *15* (4), 1476-1487.

84.   Pickard, C. J.; Mauri, F., All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys Rev B* **2001,** *63* (24).

85.   Bonhomme, C.; Gervais, C.; Babonneau, F.; Coelho, C.; Pourpoint, F.; Azaïs, T.; Ashbrook, S. E.; Griffin, J. M.; Yates, J. R.; Mauri, F.; Pickard, C. J., First-Principles Calculation of NMR Parameters Using the Gauge Including Projector Augmented Wave Method: A Chemist's Point of View. *Chemical Reviews* **2012,** *112* (11), 5733-5779.

86.   Charpentier, T., The PAW/GIPAW approach for computing NMR parameters: A new dimension added to NMR study of solids. *Solid State Nucl Mag* **2011,** *40* (1), 1-20.

87.   Vaara, J., Theory and computation of nuclear magnetic resonance parameters. *Physical Chemistry Chemical Physics* **2007,** *9* (40), 5399-5418.

88.   Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I. J.; Refson, K.; Payne, M. C., First principles methods using CASTEP. *Zeitschrift für Kristallographie - Crystalline Materials* **2005,** *220* (5-6), 567-570.

89.   Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M., QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter* **2009,** *21* (39), 395502.

90.   Lejaeghere, K.; Bihlmayer, G.; Björkman, T.; Blaha, P.; Blügel, S.; Blum, V.; Caliste, D.; Castelli, I. E.; Clark, S. J.; Dal Corso, A.; de Gironcoli, S.; Deutsch, T.; Dewhurst, J. K.; Di Marco, I.; Draxl, C.; Dułak, M.; Eriksson, O.; Flores-Livas, J. A.; Garrity, K. F.; Genovese, L.; Giannozzi, P.; Giantomassi, M.; Goedecker, S.; Gonze, X.; Grånäs, O.; Gross, E. K. U.; Gulans, A.; Gygi, F.; Hamann, D. R.; Hasnip, P. J.; Holzwarth, N. A. W.; Iuşan, D.; Jochym, D. B.; Jollet, F.; Jones, D.; Kresse, G.; Koepernik, K.; Küçükbenli, E.; Kvashnin, Y. O.; Locht, I. L. M.; Lubeck, S.; Marsman, M.; Marzari, N.; Nitzsche, U.; Nordström, L.; Ozaki, T.; Paulatto, L.; Pickard, C. J.; Poelmans,

W.; Probert, M. I. J.; Refson, K.; Richter, M.; Rignanese, G.-M.; Saha, S.; Scheffler, M.; Schlipf, M.; Schwarz, K.; Sharma, S.; Tavazza, F.; Thunström, P.; Tkatchenko, A.; Torrent, M.; Vanderbilt, D.; van Setten, M. J.; Van Speybroeck, V.; Wills, J. M.; Yates, J. R.; Zhang, G.-X.; Cottenier, S., Reproducibility in density functional theory calculations of solids. *Science* **2016,** *351* (6280), aad3000.

91.   Badri, Z.; Bouzková, K.; Foroutan-Nejad, C.; Marek, R., Origin of the Thermodynamic Stability of the Polymorph IV of Crystalline Barbituric Acid: Evidence from Solid-State NMR and Electron Density Analyses. *Cryst Growth Des* **2014,** *14* (6), 2763-2772.

92.   Raleigh, D. P.; Levitt, M. H.; Griffin, R. G., Rotational resonance in solid state NMR. *Chem Phys Lett* **1988,** *146* (1), 71-76.

93.   Yannoni, C. S.; Clarke, T. C., Molecular Geometry of cis- and trans-Polyacetylene by Nutation NMR Spectroscopy. *Physical Review Letters* **1983,** *51* (13), 1191-1193.

94.   Gullion, T.; Schaefer, J., Rotational-echo double-resonance NMR. *Journal of Magnetic Resonance (1969)* **1989,** *81* (1), 196-200.

95.   Bak, M.; Rasmussen, J. T.; Nielsen, N. C., SIMPSON: A general simulation program for solid-state NMR spectroscopy. *J Magn Reson* **2011,** *213* (2), 366-400.

96.   Potnuru, L. R.; Duong, N. T.; Ahlawat, S.; Raran-Kurussi, S.; Ernst, M.; Nishiyama, Y.; Agarwal, V., Accuracy of 1H–1H distances measured using frequency selective recoupling and fast magic-angle spinning. *The Journal of Chemical Physics* **2020,** *153* (8), 084202.

97.   Marshall, G. R.; Beusen, D. D.; Kociolek, K.; Redlinski, A. S.; Leplawy, M. T.; Pan, Y.; Schaefer, J., Determination of a precise interatomic distance in a helical peptide by REDOR NMR. *J Am Chem Soc* **1990,** *112* (3), 963-966.

98.   Ishii, Y.; Terao, T., Determination of Interheteronuclear Distances by Observation of the Pake-Doublet Patterns Using the MLEV-8 Sequences with Composite Pulses. *Journal of Magnetic Resonance, Series A* **1995,** *115* (1), 116-118.

99.   Tycko, R.; Dabbagh, G., Measurement of nuclear magnetic dipole—dipole couplings in magic angle spinning NMR. *Chem Phys Lett* **1990,** *173* (5), 461-465.

100.   Takegoshi, K.; Nakamura, S.; Terao, T., 13C–1H dipolar-driven 13C–13C recoupling without 13C rf irradiation in nuclear magnetic resonance of rotating solids. *The Journal of Chemical Physics* **2003,** *118* (5), 2325-2341.

101.   Nomura, K.; Takegoshi, K.; Terao, T.; Uchida, K.; Kainosho, M., Determination of the Complete Structure of a Uniformly Labeled Molecule by Rotational Resonance Solid-State NMR in the Tilted Rotating Frame. *J Am Chem Soc* **1999,** *121* (16), 4064-4065.

102.   Jaroniec, C. P.; Filip, C.; Griffin, R. G., 3D TEDOR NMR Experiments for the Simultaneous Measurement of Multiple Carbon–Nitrogen Distances in Uniformly 13C,15N-Labeled Solids. *J Am Chem Soc* **2002,** *124* (36), 10728-10742.

103.   Nakai, T.; Ashida, J.; Terao, T., Measurements of two-dimensional NMR powder patterns in rotating solids. *The Journal of Chemical Physics* **1988,** *88* (10), 6049-6058.

104.   Takegoshi, K.; Ito, M.; Terao, T., Deuteron 2D exchange sample-turning NMR: determination of interbond angles. *Chem Phys Lett* **1996,** *260* (1), 159-165.

105.   Malfait, W. J.; Halter, W. E.; Verel, R., Si-29 NMR spectroscopy of silica glass: T-1 relaxation and constraints on the Si-O-Si bond angle distribution. *Chemical Geology* **2008,** *256* (3-4), 269-277.

106.   Trease, N. M.; Clark, T. M.; Grandinetti, P. J.; Stebbins, J. F.; Sen, S., Bond length-bond angle correlation in densified silica—Results from 17O NMR spectroscopy. *The Journal of Chemical Physics* **2017,** *146* (18), 184505.

107.   Ashida, J.; Ohgo, K.; Asakura, T., Determination of the Torsion Angles of Alanine and Glycine Residues of Bombyx Mori Silk Fibroin and the Model Peptides in the Silk I and Silk II Forms Using 2D Spin Diffusion Solid-State NMR under Off Magic Angle Spinning. *The Journal of Physical Chemistry B* **2002,** *106* (36), 9434-9439.

108.   Ishii, Y.; Hirao, K.; Terao, T.; Terauchi, T.; Oba, M.; Nishiyama, K.; Kainosho, M., Determination of peptide phi angles in solids by relayed anisotropy correlation NMR. *Solid State Nucl Magn Reson* **1998,** *11* (3-4), 169-75.

109.   Jadeja, Y. S.; Kapadiya, K. M.; Jebaliya, H. J.; Shah, A. K.; Khunt, R. C., Dihedral angle study in Hesperidin using NMR Spectroscopy. *Magnetic Resonance in Chemistry* **2017,** *55* (6), 589-594.

110.   Paëpe, G. D., Dipolar Recoupling in Magic Angle Spinning Solid-State Nuclear Magnetic Resonance. *Annual Review of Physical Chemistry* **2012,** *63* (1), 661-684.

111.   Ladizhansky, V., Homonuclear dipolar recoupling techniques for structure determination in uniformly 13C-labeled proteins. *Solid State Nucl Mag* **2009,** *36* (3), 119-128.

112.   Dumez, J.-N.; Halse, M. E.; Butler, M. C.; Emsley, L., A first-principles description of proton-driven spin diffusion. *Physical Chemistry Chemical Physics* **2012,** *14* (1), 86-89.

113.   Elena, B.; Emsley, L., Powder crystallography by proton solid-state NMR spectroscopy. *J Am Chem Soc* **2005,** *127* (25), 9140-6.

114.   Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L., Molecular Structure Determination in Powders by NMR Crystallography from Proton Spin Diffusion. *J Am Chem Soc* **2006,** *128* (29), 9555-9560.

115.   Pickard, C. J.; Salager, E.; Pintacuda, G.; Elena, B.; Emsley, L., Resolving Structures from Powders by NMR Crystallography Using Combined Proton Spin Diffusion and Plane Wave DFT Calculations. *J Am Chem Soc* **2007,** *129* (29), 8932-8933.

116.   Salager, E.; Day, G. M.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L., Powder Crystallography by Combined Crystal Structure Prediction and High-Resolution 1H Solid-State NMR Spectroscopy. *J Am Chem Soc* **2010,** *132* (8), 2564-2566.

117.   Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M. A.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A. K.; Schweizer, B., Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. *Acta Crystallographica Section B* **2009,** *65* (2), 107-125.

118.   Baias, M.; Widdifield, C. M.; Dumez, J. N.; Thompson, H. P. G.; Cooper, T. G.; Salager, E.; Bassil, S.; Stein, R. S.; Lesage, A.; Day, G. M.; Emsley, L., Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state H-1 NMR spectroscopy. *Physical Chemistry Chemical Physics* **2013,** *15* (21), 8069-8080.

119.   Baias, M.; Dumez, J. N.; Svensson, P. H.; Schantz, S.; Day, G. M.; Emsley, L., De Novo Determination of the Crystal Structure of a Large Drug Molecule by Crystal Structure Prediction-Based Powder NMR Crystallography. *J Am Chem Soc* **2013,** *135* (46), 17501-17507.

120.   Mollica, G.; Dekhil, M.; Ziarelli, F.; Thureau, P.; Viel, S., Probing crystal packing of uniformly 13C-enriched powder samples using homonuclear dipolar coupling measurements. *Solid State Nucl Mag* **2015,** *65*, 114-121.

121.   Mollica, G.; Dekhil, M.; Ziarelli, F.; Thureau, P.; Viel, S., Quantitative Structural Constraints for Organic Powders at Natural Isotopic Abundance Using Dynamic Nuclear

Polarization Solid-State NMR Spectroscopy. *Angewandte Chemie International Edition* **2015,** *54* (20), 6028-6031.

122. Thureau, P.; Sturniolo, S.; Zilka, M.; Ziarelli, F.; Viel, S.; Yates, J. R.; Mollica, G., Reducing the computational cost of NMR crystallography of organic powders at natural isotopic abundance with the help of 13C-13C dipolar couplings. *Magnetic Resonance in Chemistry* **2019,** *57* (5), 256-264.

123. Harper, J. K.; Grant, D. M., Enhancing crystal-structure prediction with NMR tensor data. *Cryst Growth Des* **2006,** *6* (10), 2315-2321.

124. Kalakewich, K.; Iuliucci, R.; Harper, J. K., Establishing Accurate High-Resolution Crystal Structures in the Absence of Diffraction Data and Single Crystals—An NMR Approach. *Cryst Growth Des* **2013,** *13* (12), 5391-5396.

125. Selent, M.; Nyman, J.; Roukala, J.; Ilczyszyn, M.; Oilunkaniemi, R.; Bygrave, P. J.; Laitinen, R.; Jokisaari, J.; Day, G. M.; Lantto, P., Clathrate Structure Determination by Combining Crystal Structure Prediction with Computational and Experimental 129Xe NMR Spectroscopy. *Chemistry – A European Journal* **2017,** *23* (22), 5258-5269.

126. Brus, J.; Czernek, J.; Kobera, L.; Urbanova, M.; Abbrent, S.; Husak, M., Predicting the Crystal Structure of Decitabine by Powder NMR Crystallography: Influence of Long-Range Molecular Packing Symmetry on NMR Parameters. *Cryst Growth Des* **2016,** *16* (12), 7102-7111.

127. Brus, J.; Czernek, J.; Hruby, M.; Svec, P.; Kobera, L.; Abbrent, S.; Urbanova, M., Efficient Strategy for Determining the Atomic-Resolution Structure of Micro- and Nanocrystalline Solids within Polymeric Microbeads: Domain-Edited NMR Crystallography. *Macromolecules* **2018,** *51* (14), 5364-5374.

128. Mueller, L. J.; Dunn, M. F., NMR Crystallography of Enzyme Active Sites: Probing Chemically Detailed, Three-Dimensional Structure in Tryptophan Synthase. *Accounts of Chemical Research* **2013,** *46* (9), 2008-2017.

129. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583-589.

130. Paruzzo, F. M.; Hofstetter, A.; Musil, F.; De, S.; Ceriotti, M.; Emsley, L., Chemical shifts in molecular solids by machine learning. *Nat Commun* **2018,** *9*.

131. Bartók, A. P.; Kondor, R.; Csányi, G., On representing chemical environments. *Phys Rev B* **2013,** *87* (18), 184115.

132. De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M., Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics* **2016,** *18* (20), 13754-13769.

133. Rasmussen, C. E.; Williams, C. K., *Gaussian Processes for Machine Learnin*. MIT Press: Cambridge, 2006; Vol. 1.

134. Musil, F.; Willatt, M. J.; Langovoy, M. A.; Ceriotti, M., Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. *Journal of Chemical Theory and Computation* **2019,** *15* (2), 906-915.

135. Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.;

van Eijck, B. P.; Verwer, P.; Williams, D. E., A test of crystal structure prediction of small organic molecules. *Acta Crystallographica Section B* **2000,** *56* (4), 697-714.

136.   Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E., Crystal structure prediction of small organic molecules: a second blind test. *Acta Crystallographica Section B* **2002,** *58* (4), 647-661.

137.   Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P., A third blind test of crystal structure prediction. *Acta Crystallographica Section B* **2005,** *61* (5), 511-527.

138.   Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X. M.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F. J. J.; Maleev, A. V.; Misquitta, A. J.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K., Towards crystal structure prediction of complex organic compounds - a report on the fifth blind test. *Acta Crystallographica Section B* **2011,** *67* (6), 535-551.

139.   Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R., Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B* **2016,** *72* (4), 439-459.

140.   Czernek, J.; Urbanova, M.; Brus, J., NMR Crystallography of the Polymorphs of Metergoline. *Crystals* **2018,** *8* (10), 378.

141.   Musil, F.; De, S.; Yang, J.; Campbell, J. E.; Day, G. M.; Ceriotti, M., Machine learning for the structure–energy–property landscapes of molecular crystals. *Chemical Science* **2018,** *9* (5), 1289-1300.

142.   McDonagh, D.; Skylaris, C.-K.; Day, G. M., Machine-Learned Fragment-Based Energies for Crystal Structure Prediction. *Journal of Chemical Theory and Computation* **2019,** *15* (4), 2743-2758.

143. Egorova, O.; Hafizi, R.; Woods, D. C.; Day, G. M., Multifidelity Statistical Machine Learning for Molecular Crystal Structure Prediction. *The Journal of Physical Chemistry A* **2020,** *124* (39), 8065-8078.

144. Duckett, S. B.; Mewis, R. E., Application of Parahydrogen Induced Polarization Techniques in NMR Spectroscopy and Imaging. *Accounts Chem Res* **2012,** *45* (8), 1247-1257.

145. Lawler, R. G., Chemically Induced Dynamic Nuclear Polarization (Cidnp) .2. Radical-Pair Model. *Accounts Chem Res* **1972,** *5* (1), 25-&.

146. Adams, R. W.; Aguilar, J. A.; Atkinson, K. D.; Cowley, M. J.; Elliott, P. I. P.; Duckett, S. B.; Green, G. G. R.; Khazal, I. G.; Lopez-Serrano, J.; Williamson, D. C., Reversible Interactions with para-Hydrogen Enhance NMR Sensitivity by Polarization Transfer. *Science* **2009,** *323* (5922), 1708-1711.

147. Albert, M. S.; Cates, G. D.; Driehuys, B.; Happer, W.; Saam, B.; Springer, C. S.; Wishnia, A., Biological Magnetic-Resonance-Imaging Using Laser Polarized Xe-129. *Nature* **1994,** *370* (6486), 199-201.

148. Walker, T. G., Fundamentals of Spin-Exchange Optical Pumping. *J Phys Conf Ser* **2011,** *294*.

149. Overhauser, A. W., Polarization of Nuclei in Metals. *Physical Review* **1953,** *92* (2), 411-415.

150. Carver, T. R.; Slichter, C. P., Polarization of Nuclear Spins in Metals. *Physical Review* **1953,** *92* (1), 212-213.

151. Becerra, L. R.; Gerfen, G. J.; Temkin, R. J.; Singel, D. J.; Griffin, R. G., Dynamic nuclear polarization with a cyclotron resonance maser at 5 T. *Physical Review Letters* **1993,** *71* (21), 3561-3564.

152. Becerra, L. R.; Gerfen, G. J.; Bellew, B. F.; Bryant, J. A.; Hall, D. A.; Inati, S. J.; Weber, R. T.; Un, S.; Prisner, T. F.; McDermott, A. E.; Fishbein, K. W.; Kreischer, K. E.; Temkin, R. J.; Singel, D. J.; Griffin, R. G., A Spectrometer for Dynamic Nuclear Polarization and Electron Paramagnetic Resonance at High Frequencies. *Journal of Magnetic Resonance, Series A* **1995,** *117* (1), 28-40.

153. Gerfen, G. J.; Becerra, L. R.; Hall, D. A.; Griffin, R. G.; Temkin, R. J.; Singel, D. J., High frequency (140 GHz) dynamic nuclear polarization: Polarization transfer to a solute in frozen aqueous solution. *The Journal of Chemical Physics* **1995,** *102* (24), 9494-9497.

154. Rosay, M.; Blank, M.; Engelke, F., Instrumentation for solid-state dynamic nuclear polarization with magic angle spinning NMR. *J Magn Reson* **2016,** *264*, 88-98.

155. Rosay, M.; Tometich, L.; Pawsey, S.; Bader, R.; Schauwecker, R.; Blank, M.; Borchard, P. M.; Cauffman, S. R.; Felch, K. L.; Weber, R. T.; Temkin, R. J.; Griffin, R. G.; Maas, W. E., Solid-state dynamic nuclear polarization at 263 GHz: spectrometer design and experimental results. *Physical Chemistry Chemical Physics* **2010,** *12* (22), 5850-5860.

156. Denysenkov, V.; Prandolini, M. J.; Gafurov, M.; Sezer, D.; Endeward, B.; Prisner, T. F., Liquid state DNP using a 260 GHz high power gyrotron. *Physical Chemistry Chemical Physics* **2010,** *12* (22), 5786-5790.

157. Ardenkjaer-Larsen, J. H.; Fridlund, B.; Gram, A.; Hansson, G.; Hansson, L.; Lerche, M. H.; Servin, R.; Thaning, M.; Golman, K., Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. *P Natl Acad Sci USA* **2003,** *100* (18), 10158-10163.

158. Golman, K.; in't Zandt, R.; Thaning, M., Real-time metabolic imaging. *P Natl Acad Sci USA* **2006,** *103* (30), 11270-11275.

159. Zhang, G. N.; Hilty, C., Applications of dissolution dynamic nuclear polarization in chemistry and biochemistry. *Magn Reson Chem* **2018,** *56* (7), 566-582.

160. Kovtunov, K. V.; Pokochueva, E. V.; Salnikov, O. G.; Cousin, S. F.; Kurzbach, D.; Vuichoud, B.; Jannin, S.; Chekmenev, E. Y.; Goodson, B. M.; Barskiy, D. A.; Koptyug, I. V., Hyperpolarized NMR Spectroscopy: d-DNP, PHIP, and SABRE Techniques. *Chemistry – An Asian Journal* **2018,** *13* (15), 1857-1871.

161. Brindle, K. M.; Bohndiek, S. E.; Gallagher, F. A.; Kettunen, M. I., Tumor imaging using hyperpolarized 13C magnetic resonance spectroscopy. *Magnetic Resonance in Medicine* **2011,** *66* (2), 505-519.

162. Jannin, S.; Dumez, J.-N.; Giraudeau, P.; Kurzbach, D., Application and methodology of dissolution dynamic nuclear polarization in physical, chemical and biological contexts. *J Magn Reson* **2019,** *305*, 41-50.

163. Köckenberger, W., Dissolution Dynamic Nuclear Polarization. In *eMagRes*, pp 161-170.

164. Akbey, Ü.; Franks, W. T.; Linden, A.; Orwick-Rydmark, M.; Lange, S.; Oschkinat, H., Dynamic Nuclear Polarization Enhanced NMR in the Solid-State. In *Hyperpolarization Methods in NMR Spectroscopy*, Kuhn, L. T., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 181-228.

165. van Bentum, J.; van Meerten, B.; Sharma, M.; Kentgens, A., Perspectives on DNP-enhanced NMR spectroscopy in solutions. *J Magn Reson* **2016,** *264*, 59-67.

166. Keshari, K. R.; Wilson, D. M., Chemistry and biochemistry of 13C hyperpolarized magnetic resonance using dynamic nuclear polarization. *Chemical Society Reviews* **2014,** *43* (5), 1627-1659.

167. Berthault, P.; Boutin, C.; Martineau-Corcos, C.; Carret, G., Use of dissolved hyperpolarized species in NMR: Practical considerations. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2020,** *118-119*, 74-90.

168. Miéville, P.; Jannin, S.; Helm, L.; Bodenhausen, G., Kinetics of Yttrium–Ligand Complexation Monitored Using Hyperpolarized 89Y as a Model for Gadolinium in Contrast Agents. *J Am Chem Soc* **2010,** *132* (14), 5006-5007.

169. Leggett, J.; Hunter, R.; Granwehr, J.; Panek, R.; Perez-Linde, A. J.; Horsewill, A. J.; McMaster, J.; Smith, G.; Köckenberger, W., A dedicated spectrometer for dissolution DNP NMR spectroscopy. *Physical Chemistry Chemical Physics* **2010,** *12* (22), 5883-5892.

170. Zeng, H.; Bowen, S.; Hilty, C., Sequentially acquired two-dimensional NMR spectra from hyperpolarized sample. *Journal of Magnetic Resonance* **2009,** *199* (2), 159-165.

171. Giraudeau, P.; Frydman, L., Ultrafast 2D NMR: An Emerging Tool in Analytical Spectroscopy. *Annu Rev Anal Chem* **2014,** *7*, 129-161.

172. Swisher, C. L.; Koelsch, B.; Sukumar, S.; Sriram, R.; Delos Santos, R.; Wang, Z. J.; Kurhanewicz, J.; Vigneron, D.; Larson, P., Dynamic UltraFast 2D EXchange SpectroscopY (UF-EXSY) of hyperpolarized substrates. *Journal of Magnetic Resonance* **2015,** *257*, 102-109.

173. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge Structural Database. *Acta Crystallographica Section B* **2016,** *72* (2), 171-179.

174. Grazulis, S.; Daskevic, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quiros, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A., Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research* **2012,** *40* (D1), D420-D427.

175. Brown, I. D.; Bergerhoff, G., Databases of Inorganic Crystal-Structures. *Abstracts of Papers of the American Chemical Society* **1979,** (Sep), 36-36.

176. Batisai, E.; Ayamine, A.; Kilinkissa, O. E. Y.; Bathori, N. B., Melting point-solubility-structure correlations in multicomponent crystals containing fumaric or adipic acid. *Crystengcomm* **2014,** *16* (43), 9992-9998.

177.   Paruzzo, F. M.; Hofstetter, A.; Musil, F.; De, S.; Ceriotti, M.; Emsley, L., Chemical shifts in molecular solids by machine learning. *Nat Commun* **2018**, *9* (1), 4501.

178.   Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M., Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences* **2007**, *104* (23), 9615.

179.   Santos, S. M.; Rocha, J.; Mafra, L., NMR Crystallography: Toward Chemical Shift-Driven Crystal Structure Determination of the β-Lactam Antibiotic Amoxicillin Trihydrate. *Cryst Growth Des* **2013**, *13* (6), 2390-2395.

180.   Sternberg, U.; Koch, F.-T.; Prieß, W.; Witter, R., Crystal Structure Refinements of Cellulose Polymorphs using Solid State 13C Chemical Shifts. *Cellulose* **2003**, *10* (3), 189-199.

181.   Curtis, F.; Li, X.; Rose, T.; Vázquez-Mayagoitia, Á.; Bhattacharya, S.; Ghiringhelli, L. M.; Marom, N., GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *Journal of Chemical Theory and Computation* **2018**, *14* (4), 2246-2264.

182.   Wang, Y.; Lv, J.; Zhu, L.; Ma, Y., Crystal structure prediction via particle-swarm optimization. *Phys Rev B* **2010**, *82* (9), 094116.

183.   Yang, S.; Day, G. M., Exploration and Optimization in Crystal Structure Prediction: Combining Basin Hopping with Quasi-Random Sampling. *Journal of Chemical Theory and Computation* **2021**, *17* (3), 1988-1999.

184.   Karfunkel, H. R.; Gdanitz, R. J., Ab Initio prediction of possible crystal structures for general organic molecules. *Journal of Computational Chemistry* **1992**, *13* (10), 1171-1183.

185.   Zilka, M.; Dudenko, D. V.; Hughes, C. E.; Williams, P. A.; Sturniolo, S.; Franks, W. T.; Pickard, C. J.; Yates, J. R.; Harris, K. D. M.; Brown, S. P., Ab initio random structure searching of organic molecular solids: assessment and validation against experimental data. *Physical Chemistry Chemical Physics* **2017**, *19* (38), 25949-25960.

186.   Bazterra, V. E.; Ferraro, M. B.; Facelli, J. C., Modified genetic algorithm to model crystal structures. I. Benzene, naphthalene and anthracene. *The Journal of Chemical Physics* **2002**, *116* (14), 5984-5991.

187.   Zhu, Q.; Oganov, A. R.; Glass, C. W.; Stokes, H. T., Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystallographica Section B* **2012**, *68* (3), 215-226.

188.   Brown, S. P., Applications of high-resolution 1H solid-state NMR. *Solid State Nucl. Magn. Reson.* **2012**, *41*, 1-27.

189.   Rossini, A. J.; Widdifield, C. M.; Zagdoun, A.; Lelli, M.; Schwarzwälder, M.; Copéret, C.; Lesage, A.; Emsley, L., Dynamic Nuclear Polarization Enhanced NMR Spectroscopy for Pharmaceutical Formulations. *J. Am. Chem. Soc.* **2014**, *136* (6), 2324-2334.

190.   Li, M.; Meng, F.; Tsutsumi, Y.; Amoureux, J.-P.; Xu, W.; Lu, X.; Zhang, F.; Su, Y., Understanding Molecular Interactions in Rafoxanide–Povidone Amorphous Solid Dispersions from Ultrafast Magic Angle Spinning NMR. *Mol. Pharm.* **2020**, *17* (6), 2196-2207.

191.   Xiang, T. X.; Anderson, B. D., Molecular Dynamics Simulation of Amorphous Indomethacin. *Mol. Pharm.* **2013**, *10* (1), 102-114.

192.   Xiang, T.-X.; Anderson, B. D., Molecular Dynamics Simulation of Amorphous Indomethacin-Poly(Vinylpyrrolidone) Glasses: Solubility and Hydrogen Bonding Interactions. *J. Pharm. Sci.* **2013**, *102* (3), 876-891.

193.   Yuan, X.; Xiang, T.-X.; Anderson, B. D.; Munson, E. J., Hydrogen Bonding Interactions in Amorphous Indomethacin and Its Amorphous Solid Dispersions with Poly(vinylpyrrolidone) and Poly(vinylpyrrolidone-co-vinyl acetate) Studied Using 13C Solid-State NMR. *Mol. Pharm.* **2015**, *12* (12), 4518-4528.

194.   Engel, E. A.; Anelli, A.; Hofstetter, A.; Paruzzo, F.; Emsley, L.; Ceriotti, M., A Bayesian approach to NMR crystal structure determination. *Phys Chem Chem Phys* **2019,** *21* (42), 23385-23400.

195.   Baias, M.; Dumez, J. N.; Svensson, P. H.; Schantz, S.; Day, G. M.; Emsley, L., De novo determination of the crystal structure of a large drug molecule by crystal structure prediction-based powder NMR crystallography. *J Am Chem Soc* **2013,** *135* (46), 17501-7.

196.   Baias, M.; Widdifield, C. M.; Dumez, J. N.; Thompson, H. P.; Cooper, T. G.; Salager, E.; Bassil, S.; Stein, R. S.; Lesage, A.; Day, G. M.; Emsley, L., Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state 1H NMR spectroscopy. *Phys Chem Chem Phys* **2013,** *15* (21), 8069-80.

197.   Baias, M.; Lesage, A.; Aguado, S.; Canivet, J.; Moizan-Basle, V.; Audebrand, N.; Farrusseng, D.; Emsley, L., Superstructure of a substituted zeolitic imidazolate metal-organic framework determined by combining proton solid-state NMR spectroscopy and DFT calculations. *Angewandte Chemie - International Edition* **2015,** *54* (20), 5971-5976.

198.   Thompson, H. P. G.; Day, G. M., Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape. *Chem Sci* **2014,** *5* (8), 3173-3182.

199.   Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio, R. A.; Dzyabchenko, A.; Van Eijck, B. P.; Elking, D. M.; Van Den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C. A.; Gee, T. S.; De Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; De Jong, D. l. T.; Kendrick, J.; De Klerk, N. J. J.; Ko, H. Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. m. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; De Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R., Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2016,** *72* (4), 439-459.

200.   Chierotti, M. R.; Gobetto, R., NMR crystallography: the use of dipolar interactions in polymorph and co-crystal investigation. *Crystengcomm* **2013,** *15* (43).

201.   Widdifield, C. M.; Nilsson Lill, S. O.; Broo, A.; Lindkvist, M.; Pettersen, A.; Svensk Ankarberg, A.; Aldred, P.; Schantz, S.; Emsley, L., Does Z' equal 1 or 2? Enhanced powder NMR crystallography verification of a disordered room temperature crystal structure of a p38 inhibitor for chronic obstructive pulmonary disease. *Phys Chem Chem Phys* **2017,** *19* (25), 16650-16661.

202.   Harper, J. K.; Barich, D. H.; Hu, J. Z.; Strobel, G. A.; Grant, D. M., Stereochemical analysis by solid-state NMR: Structural predictions in ambuic acid. *Journal of Organic Chemistry* **2003,** *68* (12), 4609-4614.

203.   Jacob, P.; Kalakewich, K.; Uribe-Romo, F. J.; Harper, J. K., Solid-state NMR and DFT predictions of differences in COOH hydrogen bonding in odd and even numbered n-alkyl fatty acids. *Physical Chemistry Chemical Physics* **2016,** *18*, 12541-12549.

204.   Kalakewich, K.; Iuliucci, R.; Mueller, K. T.; Eloranta, H.; Harper, J. K., Monitoring the refinement of crystal structures with N-15 solid-state NMR shift tensor data. *J Chem Phys* **2015,** *143* (19).

205.   Hofstetter, A.; Emsley, L., Positional Variance in NMR Crystallography. *J Am Chem Soc* **2017,** *139* (7), 2573-2576.

206.   Roberts, J. E.; Harbison, G. S.; Munowitz, M. G.; Herzfeld, J.; Griffin, R. G., Measurement of heteronuclear bond distances in polycrystalline solids by solid-state NMR techniques. *Journal of the American Chemical Society* **1987,** *109* (14), 4163-4169.

207.   Colombo, M.; Meier, B.; Ernst, R., Rotor-driven spin diffusion in natural-abundance 13C spin systems. *Chemical physics letters* **1988,** *146* (3-4), 189-196.

208.   Raleigh, D.; Levitt, M.; Griffin, R., Rotational resonance in solid state NMR. *Chemical Physics Letters* **1988,** *146* (1-2), 71-76.

209.   Van Rossum, B.-J.; De Groot, C.; Ladizhansky, V.; Vega, S.; De Groot, H., A method for measuring heteronuclear (1H− 13C) distances in high speed MAS NMR. *Journal of the American Chemical Society* **2000,** *122* (14), 3465-3472.

210.   Castellani, F.; Rossum van, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Hartmut, O., Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **2002,** *420* (November), 98-102.

211.   Seidel, K.; Etzkorn, M.; Sonnenberg, L.; Griesinger, C.; Sebald, A.; Baldus, M., Studying molecular 3D structure and dynamics by high-resolution solid-state NMR: Application to l-tyrosine-ethylester. *The Journal of Physical Chemistry A* **2005,** *109* (11), 2436-2442.

212.   Mollica, G.; Dekhil, M.; Ziarelli, F.; Thureau, P.; Viel, S. p. S., Quantitative structural constraints for organic powders at natural isotopic abundance using dynamic nuclear polarization solid-state NMR spectroscopy. *Angewandte Chemie - International Edition* **2015,** *54* (20), 6028-6031.

213.   Dekhil, M.; Mollica, G.; Bonniot, T. T.; Ziarelli, F.; Thureau, P.; Viel, S., Determining carbon-carbon connectivities in natural abundance organic powders using dipolar couplings. *Chemical Communications* **2016,** *52* (55), 8565-8568.

214.   Leclaire, J.; Poisson, G.; Ziarelli, F.; Pepe, G.; Fotiadu, F.; Paruzzo, F. M.; Rossini, A. J.; Dumez, J. N.; Elena-Herrmann, B.; Emsley, L., Structure elucidation of a complex CO2-based organic framework material by NMR crystallography. *Chem Sci* **2016,** *7* (7), 4379-4390.

215.   Märker, K.; Paul, S.; Fernández-de-Alba, C.; Lee, D.; Mouesca, J.-M.; Hediger, S.; De Paëpe, G., Welcoming natural isotopic abundance in solid-state NMR: probing π-stacking and supramolecular structure of organic nanoassemblies using DNP. *Chem Sci* **2017,** *8* (2), 974-987.

216.   Thureau, P.; Sturniolo, S.; Zilka, M.; Ziarelli, F.; Viel, S.; Yates, J.; Mollica, G., Reducing the computational cost of NMR crystallography of organic powders at natural isotopic abundance with the help of 13C-13C dipolar couplings. *Magnetic Resonance in Chemistry* **2019**.

217.   Gu, Z.; Ridenour, C. F.; Bronnimann, C. E.; Iwashita, T.; McDermott, A., Hydrogen Bonding and Distance Studies of Amino Acids and Peptides Using Solid State 2D 1H–13C Heteronuclear Correlation Spectra. *Journal of the American Chemical Society* **1996,** *118* (4), 822-829.

218.   M. Cerrioti; S. De; R. H. Meissner; Tribello, G. A. Sketch map package. https://github.com/cosmo-epfl/sketchmap/.

219.   Ceriotti, M.; Tribello, G. A.; Parrinello, M., From the Cover: Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences* **2011,** *108* (32), 13023-13028.

220.   De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M., Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016,** *18* (20), 13754-13769.

221.   De, S.; Musil, F.; Ingram, T.; Baldauf, C.; Ceriotti, M., Mapping and classifying molecules from a high-throughput structural database. *JOURNAL OF CHEMINFORMATICS* **2017,** *9*.

222.   Antzutkin, O. N.; Lee, Y. K.; Levitt, M. H., 13C and15N—Chemical Shift Anisotropy of Ampicillin and Penicillin-V Studied by 2D-PASS and CP/MAS NMR. *Journal of Magnetic Resonance* **1998,** *135* (1), 144-155.

223.   Case, D. H.; Campbell, J. E.; Bygrave, P. J.; Day, G. M., Convergence Properties of Crystal Structure Prediction by Quasi-Random Sampling. *Journal of Chemical Theory and Computation* **2016,** *12* (2).

224.   Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M., Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Physical Chemistry Chemical Physics* **2010,** *12* (30), 8478-8490.

225.   Salager, E.; Day, G. M.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L., Powder crystallography by combined crystal structure prediction and high-resolution 1H solid-state NMR spectroscopy. *J Am Chem Soc* **2010,** *132* (8), 2564-6.

226.   Nyman, J.; Day, G. M., Static and lattice vibrational energy differences between polymorphs. *Crystengcomm* **2015,** *17* (28), 5154-5165.

227.   Boles, M. O.; Girven, R. J., The structures of ampicillin: a comparison of the anhydrate and trihydrate forms. *Acta Crystallographica Section B* **1976,** *32* (8), 2279-2284.

228.   Cense, J. M.; Agafonov, V.; Ceolin, R.; Ladure, P.; Rodier, N., Crystal and Molecular-Structure Analysis of Flutamide - Bifurcated Helicoidal C-H ... O Hydrogen-Bonds. *Struct Chem* **1994,** *5* (2), 79-84.

229.   Hrynchuk, R. J.; Barton, R. J.; Robertson, B. E., The Crystal-Structure of Free Base Cocaine, C17h21no4. *Can J Chem* **1983,** *61* (3), 481-487.

230.   Murthy, H. M. K.; Bhat, T. N.; Vijayan, M., Structural Studies of Analgesics and Their Interactions .9. Structure of a New Crystal Form of 2-((3-(Trifluoromethyl)Phenyl)Amino)Benzoic Acid (Flufenamic Acid). *Acta Crystallogr B* **1982,** *38* (Jan), 315-317.

231.   Hayashi, S.; Hayamizu, K., Chemical Shift Standards in High-Resolution Solid-State NMR (1) 13C, 29Si, and 1H Nuclei. *Bulletin of the Chemical Society of Japan* **1991,** *64* (2), 685-687.

232.   Baias, M.; Widdifield, C. M.; Dumez, J.-N.; Thompson, H. P. G.; Cooper, T. G.; Salager, E.; Bassil, S.; Stein, R. S.; Lesage, A.; Day, G. M.; Emsley, L., Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state 1H NMR spectroscopy. *Physical Chemistry Chemical Physics* **2013,** *15* (21), 8069-8069.

233.   Clayden, N. J.; Dobson, C. M.; Lian, L.-Y.; Twyman, J. M., A solid-state 13C nuclear magnetic resonance study of the conformational states of penicillins. *Journal of the Chemical Society, Perkin Transactions 2* **1986,** (12), 1933-1940.

234.   Kolossvary, I.; Guida, W. C., Low mode search. An efficient, automated computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. *Journal of the American Chemical Society* **1996,** *118* (21), 5011-5019.

235.   Kolossváry, I.; Guida, W. C., Low-mode conformational search elucidated: Application to C39H80 and flexible docking of 9-deazaguanine inhibitors into PNP. *Journal of Computational Chemistry* **1999,** *20* (15), 1671-1684.

236.   *MacroModel*, V9.0; Schrödinger LLC: New York, NY, 2011.

237.   Harder, E.; Damm, W.; Maple, J.; Wu, C. J.; Reboul, M.; Xiang, J. Y.; Wang, L. L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation* **2016,** *12* (1), 281-296.

238.   Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *Journal of Computational Chemistry* **2011,** *32* (7), 1456-1465.

239.   Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitrică, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; Heide, T. v. d.; Hermann, J.; Irle, S.; Kranz, J. J.; Köhler, C.; Kowalczyk, T.; Kubař, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A. M. N.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Řezáč, J.; Sánchez, C. G.; Sternberg, M.; Stöhr, M.; Stuckenberg, F.; Tkatchenko, A.; Yu, V. W.-z.; Frauenheim, T., DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of Chemical Physics* **2020,** *152* (12), 124101.

240.   Motherwell, S.; Chisholm, J. A., COMPACK: a program for identifying crystal structure similarity using distances. *Journal of Applied Crystallography* **2005,** *38* (1), 228-231.

241.   Hartman, J. D.; Kudla, R. A.; Day, G. M.; Mueller, L. J.; Beran, G. J. O., Benchmark fragment-based 1H, 13C, 15N and 17O chemical shift predictions in molecular crystals. *Physical Chemistry Chemical Physics* **2016,** *18* (31), 21686-21709.

242.   Hukushima, K.; Nemoto, K., Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **1996,** *65* (6), 1604-1608.

243.   Hartman, J. D.; Beran, G. J., Fragment-Based Electronic Structure Approach for Computing Nuclear Magnetic Resonance Chemical Shifts in Molecular Crystals. *J Chem Theory Comput* **2014,** *10* (11), 4862-72.

244.   Hartman, J. D.; Kudla, R. A.; Day, G. M.; Mueller, L. J.; Beran, G. J., Benchmark fragment-based (1)H, (13)C, (15)N and (17)O chemical shift predictions in molecular crystals. *Phys Chem Chem Phys* **2016,** *18* (31), 21686-709.

245.   Hartman, J. D.; Monaco, S.; Schatschneider, B.; Beran, G. J., Fragment-based (13)C nuclear magnetic resonance chemical shift predictions in molecular crystals: An alternative to planewave methods. *J Chem Phys* **2015,** *143* (10), 102809.

246.   Hofstetter, A.; Emsley, L., Positional Variance in NMR Crystallography. *J. Am. Chem. Soc.* **2017,** *139* (7), 2573-2576.

247.   Steiner, T., The hydrogen bond in the solid state. *Angew Chem Int Edit* **2002,** *41* (1), 48-76.

248.   Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G., Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* **1998,** *58* (11), 7260-7268.

249.   Aradi, B.; Hourahine, B.; Frauenheim, T., DFTB+, a sparse matrix-based implementation of the DFTB method. *J. Phys. Chem. A* **2007,** *111* (26), 5678-5684.

250.   Gaus, M.; Cui, Q. A.; Elstner, M., DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J Chem Theory Comput* **2011,** *7* (4), 931-948.

251.   Gaus, M.; Goez, A.; Elstner, M., Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013,** *9* (1), 338-354.

252.   Rezac, J., Empirical Self-Consistent Correction for the Description of Hydrogen Bonds in DFTB3. *J. Chem. Theory Comput.* **2017,** *13* (10), 4804-4817.

253.   Hourahine, B.; Aradi, B.; Blum, V.; Bonafe, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitrica, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; van der Heide, T.; Hermann, J.; Irle, S.; Kranz, J. J.; Kohler, C.; Kowalczyk, T.; Kubar, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A. M. N.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Rezac, J.; Sanchez, C. G.; Sternberg, M.; Stohr, M.; Stuckenberg, F.; Tkatchenko, A.; Yu, V. W. Z.; Frauenheim, T., DFTB plus , a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020,** *152* (12), 124101.

254.   Brown, S. P.; Spiess, H. W., Advanced solid-state NMR methods for the elucidation of structure and dynamics of molecular, macromolecular, and supramolecular systems. *Chem Rev* **2001,** *101* (12), 4125-56.

255.   Melnyk, A.; Junk, M. J. N.; McGehee, M. D.; Chmelka, B. F.; Hansen, M. R.; Andrienko, D., Macroscopic Structural Compositions of π-Conjugated Polymers: Combined Insights from Solid-State NMR and Molecular Dynamics Simulations. *The Journal of Physical Chemistry Letters* **2017,** *8* (17), 4155-4160.

256.   Schahl, A.; Gerber, I. C.; Réat, V.; Jolibois, F., Diversity of the Hydrogen Bond Network and Its Impact on NMR Parameters of Amylose B Polymorph: A Study Using Molecular Dynamics and DFT Calculations Within Periodic Boundary Conditions. *The Journal of Physical Chemistry B* **2020,** *125* (1), 158-168.

257.   Lesage, A.; Lelli, M.; Gajan, D.; Caporini, M. A.; Vitzthum, V.; Mieville, P.; Alauzun, J.; Roussey, A.; Thieuleux, C.; Mehdi, A.; Bodenhausen, G.; Coperet, C.; Emsley, L., Surface Enhanced NMR Spectroscopy by Dynamic Nuclear Polarization. *J Am Chem Soc* **2010,** *132* (44), 15459-15461.

258.   Zagdoun, A.; Rossini, A. J.; Gajan, D.; Bourdolle, A.; Ouari, O.; Rosay, M.; Maas, W. E.; Tordo, P.; Lelli, M.; Emsley, L.; Lesage, A.; Coperet, C., Non-aqueous solvents for DNP surface enhanced NMR spectroscopy. *Chem Commun* **2012,** *48* (5), 654-656.

259.   Lelli, M.; Chaudhari, S. R.; Gajan, D.; Casano, G.; Rossini, A. J.; Ouari, O.; Tordo, P.; Lesage, A.; Emsley, L., Solid-State Dynamic Nuclear Polarization at 9.4 and 18.8 T from 100 K to Room Temperature. *J Am Chem Soc* **2015,** *137* (46), 14558-14561.

260.   Ong, T. C.; Mak-Jurkauskas, M. L.; Walish, J. J.; Michaelis, V. K.; Corzilius, B.; Smith, A. A.; Clausen, A. M.; Cheetham, J. C.; Swager, T. M.; Griffin, R. G., Solvent-Free Dynamic Nuclear Polarization of Amorphous and Crystalline ortho-Terphenyl. *J Phys Chem B* **2013,** *117* (10), 3040-3046.

261.   Pines, A.; Gibby, M. G.; Waugh, J. S., Proton-Enhanced Nmr of Dilute Spins in Solids. *Journal of Chemical Physics* **1973,** *59* (2), 569-590.

262.   Peersen, O. B.; Wu, X. L.; Kustanovich, I.; Smith, S. O., Variable-Amplitude Cross-Polarization Mas Nmr. *J Magn Reson Ser A* **1993,** *104* (3), 334-339.

263.   States, D. J.; Haberkorn, R. A.; Ruben, D. J., A Two-Dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in 4 Quadrants. *J. Magn. Reson.* **1982,** *48* (2), 286-292.

264.   Rossini, A. J.; Zagdoun, A.; Lelli, M.; Lesage, A.; Coperet, C.; Emsley, L., Dynamic Nuclear Polarization Surface Enhanced NMR Spectroscopy. *Acc. Chem. Res.* **2013,** *46* (9), 1942-1951.

265.   Yarava, J. R.; Chaudhari, S. R.; Rossini, A. J.; Lesage, A.; Emsley, L., Solvent suppression in DNP enhanced solid state NMR. *J Magn Reson* **2017,** *277*, 149-153.

266.   Marion, D.; Ikura, M.; Tschudin, R.; Bax, A., Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **1989,** *85* (2), 393-399.

267.   Reif, B.; Ashbrook, S. E.; Emsley, L.; Hong, M., Solid-state NMR spectroscopy. *Nature Reviews Methods Primers* **2021,** *1* (1), 2.

268.   Lesage, A.; Auger, C.; Caldarelli, S.; Emsley, L., Determination of Through-Bond Carbon–Carbon Connectivities in Solid-State NMR Using the INADEQUATE Experiment. *J. Am. Chem. Soc.* **1997,** *119* (33), 7867-7868.

269.   Lesage, A.; Bardet, M.; Emsley, L., Through-Bond Carbon–Carbon Connectivities in Disordered Solids by NMR. *J. Am. Chem. Soc.* **1999,** *121* (47), 10987-10993.

270.   Pickard, C. J.; Mauri, F., All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys Rev B* **2001,** *63* (24), 245101.

271.   Harris, R. K.; Hodgkinson, P.; Pickard, C. J.; Yates, J. R.; Zorin, V., Chemical shift computations on a crystallographic basis: some reflections and comments. *Magn. Reson. Chem.* **2007,** *45* (S1), S174-S186.

272.   Hartman, J. D.; Beran, G. J. O., Fragment-Based Electronic Structure Approach for Computing Nuclear Magnetic Resonance Chemical Shifts in Molecular Crystals. *J. Chem. Theory Comput.* **2014,** *10* (11), 4862-4872.

273.   Hartman, J. D.; Monaco, S.; Schatschneider, B.; Beran, G. J. O., Fragment-based 13C nuclear magnetic resonance chemical shift predictions in molecular crystals: An alternative to planewave methods. *J. Chem. Phys.* **2015,** *143* (10), 102809.

274.   Salager, E.; Day, G. M.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L., Powder Crystallography by Combined Crystal Structure Prediction and High-Resolution H-1 Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **2010,** *132* (8), 2564-2566.

275.   Engel, E. A.; Anelli, A.; Hofstetter, A.; Paruzzo, F.; Emsley, L.; Ceriotti, M., A Bayesian approach to NMR crystal structure determination. *Phys. Chem. Chem. Phys.* **2019,** *21* (42), 23385-23400.

276.   Cordova, M.; Balodis, M.; Hofstetter, A.; Paruzzo, F.; Nilsson Lill, S. O.; Eriksson, E. S. E.; Berruyer, P.; Simões de Almeida, B.; Quayle, M. J.; Norberg, S. T.; Svensk Ankarberg, A.; Schantz, S.; Emsley, L., Structure determination of an amorphous drug through large-scale NMR predictions. *Nat. Commun.* **2021,** *12* (1), 2964.

277.   Guerry, P.; Herrmann, T., Advances in automated NMR protein structure determination. *Q. Rev. Biophys.* **2011,** *44* (3), 257-309.

278.   Schmidt, E.; Güntert, P., A New Algorithm for Reliable and General NMR Resonance Assignment. *J. Am. Chem. Soc.* **2012,** *134* (30), 12817-12829.

279.   Aeschbacher, T.; Schmidt, E.; Blatter, M.; Maris, C.; Duss, O.; Allain, F. H. T.; Güntert, P.; Schubert, M., Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.* **2013,** *41* (18), e172-e172.

280.   Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L., BioMagResBank. *Nucleic Acids Res* **2008,** *36* (Database issue), D402-8.

281.   Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28* (1), 235-242.

282.   SDBSWeb. https://sdbs.db.aist.go.jp (accessed 10.09.2021).

283.   Kuhn, S.; Schlörer, N. E., Facilitating quality control for spectra assignments of small organic molecules: nmrshiftdb2 – a free in-house NMR database with integrated LIMS for academic service laboratories. *Magnetic Resonance in Chemistry* **2015,** *53* (8), 582-589.

284.   Kang, X.; Zhao, W.; Dickwella Widanage, M. C.; Kirui, A.; Ozdenvar, U.; Wang, T., CCMRD: a solid-state NMR database for complex carbohydrates. *Journal of Biomolecular NMR* **2020,** *74* (4), 239-245.

285.   John Wiley & Sons, Inc. SpectraBase. https://spectrabase.com/ (accessed 10/09/2021).

286.   Salager, E.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L., Powder NMR crystallography of thymol. *Physical Chemistry Chemical Physics* **2009,** *11* (15), 2610-2621.

287.   Miclaus, M.; Grosu, I.-G.; Filip, X.; Tripon, C.; Filip, C., Optimizing structure determination from powders of crystalline organic solids with high molecular flexibility: the case of lisinopril dihydrate. *Crystengcomm* **2014,** *16* (3), 299-303.

288.   Mifsud, N.; Elena, B.; Pickard, C. J.; Lesage, A.; Emsley, L., Assigning powders to crystal structures by high-resolution H-1-H-1 double quantum and H-1-C-13 J-INEPT solid-state NMR spectroscopy and first principles computation. A case study of penicillin G. *Physical Chemistry Chemical Physics* **2006,** *8* (29), 3418-3422.

289.   Tatton, A. S.; Blade, H.; Brown, S. P.; Hodgkinson, P.; Hughes, L. P.; Lill, S. O. N.; Yates, J. R., Improving Confidence in Crystal Structure Solutions Using NMR Crystallography: The Case of β-Piroxicam. *Cryst. Growth Des.* **2018,** *18* (6), 3339-3351.

290.   Brus, J.; Jegorov, A., Through-Bonds and Through-Space Solid-State NMR Correlations at Natural Isotopic Abundance: Signal Assignment and Structural Study of Simvastatin. *J. Phys. Chem. A* **2004,** *108* (18), 3955-3964.

291.   Caravatti, P.; Bodenhausen, G.; Ernst, R. R., Heteronuclear solid-state correlation spectroscopy. *Chem. Phys. Lett.* **1982,** *89* (5), 363-367.

292.   Caravatti, P.; Braunschweiler, L.; Ernst, R. R., Heteronuclear correlation spectroscopy in rotating solids. *Chem. Phys. Lett.* **1983,** *100* (4), 305-310.

293.   Opella, S. J.; Frey, M. H., Selection of nonprotonated carbon resonances in solid-state nuclear magnetic resonance. *J. Am. Chem. Soc.* **1979,** *101* (19), 5854-5856.

294.   Wu, X. L.; Zilm, K. W., Complete Spectral Editing in CPMAS NMR. *J. Magn. Reson. Ser. A* **1993,** *102* (2), 205-213.

295.   Wu, X. L.; Burns, S. T.; Zilm, K. W., Spectral Editing in CPMAS NMR. Generating Subspectra Based on Proton Multiplicities. *J. Magn. Reson. Ser. A* **1994,** *111* (1), 29-36.

296.   Lesage, A.; Steuernagel, S.; Emsley, L., Carbon-13 Spectral Editing in Solid-State NMR Using Heteronuclear Scalar Couplings. *J. Am. Chem. Soc.* **1998,** *120* (28), 7095-7100.

297.   Lesage, A.; Sakellariou, D.; Steuernagel, S.; Emsley, L., Carbon–Proton Chemical Shift Correlation in Solid-State NMR by Through-Bond Multiple-Quantum Spectroscopy. *J Am Chem Soc* **1998,** *120* (50), 13194-13201.

298.   Rossini, A. J.; Zagdoun, A.; Hegner, F.; Schwarzwalder, M.; Gajan, D.; Coperet, C.; Lesage, A.; Emsley, L., Dynamic Nuclear Polarization NMR Spectroscopy of Microcrystalline Solids. *J. Am. Chem. Soc.* **2012,** *134* (40), 16899-16908.

299.   Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge Structural Database. *Acta Crystallogr. B* **2016,** *72* (2), 171-179.

300.   Gaus, M.; Lu, X.; Elstner, M.; Cui, Q., Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* **2014,** *10* (4), 1518-1537.

301.   Paruzzo, F. M.; Hofstetter, A.; Musil, F.; De, S.; Ceriotti, M.; Emsley, L., Chemical shifts in molecular solids by machine learning. *Nat. Commun.* **2018,** *9*, 4501.

302.   Rogers, D.; Hahn, M., Extended-Connectivity Fingerprints. *J. Chem. Inf. Model* **2010,** *50* (5), 742-754.

303.   Garcia-Domenech, R.; Galvez, J.; de Julian-Ortiz, J. V.; Pogliani, L., Some new trends in chemical graph theory. *Chem. Rev.* **2008,** *108* (3), 1127-1169.

304.   Danishuddin; Khan, A. U., Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discov. Today* **2016,** *21* (8), 1291-1302.

305.   Babai, L., Graph Isomorphism in Quasipolynomial Time [Extended Abstract]. *Acm. S. Theory Comput.* **2016**, 684-697.

306.   Grohe, M.; Schweitzer, P., The Graph Isomorphism Problem. *Commun. Acm* **2020,** *63* (11), 128-134.

307.   Weisfeiler, B. Y.; Leman, A. A., The Reduction of a Graph to Canonical Form and the Algebra which Appears Therein. *NTI Series* **1968,** *2* (9), 12-16.

308.   Davies, A. N.; Lampen, P., JCAMP-DX for NMR. *Appl. Spectrosc.* **1993,** *47* (8), 1093-1099.

309.   Zeebe, R. E.; Zachos, J. C.; Caldeira, K.; Tyrrell, T., Carbon Emissions and Acidification. *Science* **2008,** *321* (5885), 51-52.

310.   Gorna, K.; Hund, M.; Vučak, M.; Gröhn, F.; Wegner, G., Amorphous calcium carbonate in form of spherical nanosized particles and its application as fillers for polymers. *Materials Science and Engineering: A* **2008,** *477* (1), 217-225.

311.   Kerr, R. A., Global Warming Is Changing the World. *Science* **2007,** *316* (5822), 188-190.

312.   Li, Y.; Duan, X.; Song, W.; Ma, L.; Jow, J., Reaction mechanisms of carbon dioxide capture by amino acid salt and desorption by heat or mineralization. *Chemical Engineering Journal* **2021,** *405*, 126938.

313.   Butler, J. W.; Lim, C. J.; Grace, J. R., CO2 capture capacity of CaO in long series of pressure swing sorption cycles. *Chemical Engineering Research and Design* **2011,** *89* (9), 1794-1804.

314.   Goodwin, A. L.; Michel, F. M.; Phillips, B. L.; Keen, D. A.; Dove, M. T.; Reeder, R. J., Nanoporous Structure and Medium-Range Order in Synthetic Amorphous Calcium Carbonate. *Chem Mater* **2010,** *22* (10), 3197-3205.

315.   Nebel, H.; Neumann, M.; Mayer, C.; Epple, M., On the structure of amorphous calcium carbonate - A detailed study by solid-state NMR spectroscopy. *Inorg Chem* **2008,** *47* (17), 7874-7879.

316.   Karthika, S.; Radhakrishnan, T. K.; Kalaichelvi, P., A Review of Classical and Nonclassical Nucleation Theories. *Cryst Growth Des* **2016,** *16* (11), 6663-6681.

317.   Gebauer, D.; Volkel, A.; Colfen, H., Stable Prenucleation Calcium Carbonate Clusters. *Science* **2008,** *322* (5909), 1819-1822.

318.   Gebauer, D.; Wolf, S. E., Designing Solid Materials from Their Solute State: A Shift in Paradigms toward a Holistic Approach in Functional Materials Chemistry. *J Am Chem Soc* **2019,** *141* (11), 4490-4504.

319.   Demichelis, R.; Raiteri, P.; Gale, J. D.; Quigley, D.; Gebauer, D., Stable prenucleation mineral clusters are liquid-like ionic polymers. *Nat Commun* **2011,** *2* (1), 590.

320.   Habraken, W. J. E. M.; Tao, J.; Brylka, L. J.; Friedrich, H.; Bertinetti, L.; Schenk, A. S.; Verch, A.; Dmitrovic, V.; Bomans, P. H. H.; Frederik, P. M.; Laven, J.; van der Schoot, P.; Aichmayer, B.; de With, G.; DeYoreo, J. J.; Sommerdijk, N. A. J. M., Ion-association complexes

unite classical and non-classical theories for the biomimetic nucleation of calcium phosphate. *Nat Commun* **2013,** *4* (1), 1507.

321.   Dorvee, J. R.; Veis, A., Water in the formation of biogenic minerals: Peeling away the hydration layers. *Journal of Structural Biology* **2013,** *183* (2), 278-303.

322.   Zahn, D., Thermodynamics and Kinetics of Prenucleation Clusters, Classical and Non-Classical Nucleation. *ChemPhysChem* **2015,** *16* (10), 2069-2075.

323.   Scheck, J.; Wu, B.; Drechsler, M.; Rosenberg, R.; Van Driessche, A. E. S.; Stawski, T. M.; Gebauer, D., The Molecular Mechanism of Iron(III) Oxide Nucleation. *The Journal of Physical Chemistry Letters* **2016,** *7* (16), 3123-3130.

324.   Carino, A.; Testino, A.; Andalibi, M. R.; Pilger, F.; Bowen, P.; Ludwig, C., Thermodynamic-Kinetic Precipitation Modeling. A Case Study: The Amorphous Calcium Carbonate (ACC) Precipitation Pathway Unravelled. *Cryst Growth Des* **2017,** *17* (4), 2006-2015.

325.   Carino, A.; Ludwig, C.; Cervellino, A.; Müller, E.; Testino, A., Formation and transformation of calcium phosphate phases under biologically relevant conditions: Experiments and modelling. *Acta Biomaterialia* **2018,** *74*, 478-488.

326.   Andalibi, M. R.; Wokaun, A.; Bowen, P.; Testino, A., Kinetics and Mechanism of Metal Nanoparticle Growth via Optical Extinction Spectroscopy and Computational Modeling: The Curious Case of Colloidal Gold. *ACS Nano* **2019,** *13* (10), 11510-11521.

327.   Mohammed, A. S. A.; Carino, A.; Testino, A.; Andalibi, M. R.; Cervellino, A., Precipitation Pathways: In Situ Liquid SAXS Studies on the Early Stage of Calcium Carbonate Formation (Part. Part. Syst. Charact. 6/2019). *Particle & Particle Systems Characterization* **2019,** *36* (6), 1970014.

328.   Wolber, J.; Ellner, F.; Fridlund, B.; Gram, A.; Jóhannesson, H.; Hansson, G.; Hansson, L. H.; Lerche, M. H.; Månsson, S.; Servin, R.; Thaning, M.; Golman, K.; Ardenkjær-Larsen, J. H., Generating highly polarized nuclear spins in solution using dynamic nuclear polarization. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **2004,** *526* (1), 173-181.

329.   Milani, J.; Vuichoud, B.; Bornet, A.; Miéville, P.; Mottier, R.; Jannin, S.; Bodenhausen, G., A magnetic tunnel to shelter hyperpolarized fluids. *Review of Scientific Instruments* **2015,** *86* (2), 024101.

330.   dos Santos, V. B.; Vidal, D. T. R.; Francisco, K. J. M.; Ducati, L. C.; do Lago, C. L., Formation of isomers of anionic hemiesters of sugars and carbonic acid in aqueous medium. *Carbohydrate Research* **2016,** *428*, 18-22.

331.   De Paëpe, G.; Lesage, A.; Steuernagel, S.; Emsley, L., Transverse Dephasing Optimised NMR Spectroscopy in Solids: Natural-Abundance 13C Correlation Spectra. *ChemPhysChem* **2004,** *5* (6), 869-875.

332.   Christopher, E. A.; Harris, R. K.; Fletton, R. A., Assignments of solid-state 13C resonances for polymorphs of cortisone acetate using shielding tensor components. *Solid State Nucl Magn Reson* **1992,** *1* (2), 93-101.

333.   Internal Lyon data.

334.   Harris, R. K.; Joyce, S. A.; Pickard, C. J.; Cadars, S.; Emsley, L., Assigning carbon-13 NMR spectra to crystal structures by the INADEQUATE pulse sequence and first principles computation: a case study of two forms of testosterone. *Phys Chem Chem Phys* **2006,** *8* (1), 137-43.

335.   Gervais, C.; Dupree, R.; Pike, K. J.; Bonhomme, C.; Profeta, M.; Pickard, C. J.; Mauri, F., Combined First-Principles Computational and Experimental Multinuclear Solid-State NMR Investigation of Amino Acids. *The Journal of Physical Chemistry A* **2005,** *109* (31), 6960-6969.

336. Portieri, A.; Harris, R. K.; Fletton, R. A.; Lancaster, R. W.; Threlfall, T. L., Effects of polymorphic differences for sulfanilamide, as seen through 13C and 15N solid-state NMR, together with shielding calculations. *Magn Reson Chem* **2004,** *42* (3), 313-20.

337. Naito, A.; McDowell, C. A., Determination of the 14N quadrupole coupling tensors and the 13C chemical shielding tensors in a single crystal of L-asparagine monohydrate. *The Journal of Chemical Physics* **1984,** *81* (11), 4795-4803.

338. Chen, X.; Zhan, C.-G., First-principles studies of C-13 NMR chemical shift tensors of amino acids in crystal state. *Journal of Molecular Structure: THEOCHEM* **2004,** *682* (1), 73-82.

339. Iuliucci, R. J.; Facelli, J. C.; Alderman, D. W.; Grant, D. M., Carbon-13 Chemical Shift Tensors in Polycyclic Aromatic Compounds. 5. Single-Crystal Study of Acenaphthene. *J Am Chem Soc* **1995,** *117* (8), 2336-2343.

340. Liu, F.; Phung, C. G.; Alderman, D. W.; Grant, D. M., *J. Magn. Reson* **1996,** (120), 231–241.

341. Liu, F.; Phung, C. G.; Alderman, D. W.; Grant, D. M., Carbon-13 Chemical Shift Tensors in Methyl Glycosides, Comparing Diffraction and Optimized Structures with Single-Crystal NMR. *J Am Chem Soc* **1996,** *118* (43), 10629-10634.

342. Sherwood, M. H.; Facelli, J. C.; Alderman, D. W.; Grant, D. M., Carbon-13 chemical shift tensors in polycyclic aromatic compounds. 2. Single-crystal study of naphthalene. *J Am Chem Soc* **1991,** *113* (3), 750-753.

343. Harper, J. K.; Iuliucci, R.; Gruber, M.; Kalakewich, K., Refining crystal structures with experimental 13C NMR shift tensors and lattice-including electronic structure methods. *CrystEngComm* **2013,** *15* (43), 8693-8704.

344. Liu, F.; Orendt, A. M.; Alderman, D. W.; Grant, D. M., Carbon-13 Chemical Shift Tensors in Pentaerythritol. *J Am Chem Soc* **1997,** *119* (38), 8981-8984.

345. Pisklak, D. M.; Zielińska-Pisklak, M. A.; Szeleszczuk, Ł.; Wawer, I., ^{13}C solid-state NMR analysis of the most common pharmaceutical excipients used in solid drug formulations, Part I: Chemical shifts assignment. *J Pharm Biomed Anal* **2016,** *122*, 81-9.

346. Janes, *J. Magn. Reson* **1983,** *1* (54), 111.

347. Iuliucci, R. J.; Phung, C. G.; Facelli, J. C.; Grant, D. M., Carbon-13 Chemical Shift Tensors in Polycyclic Aromatic Compounds. 7.1 Symmetry Augmented Chemical Shift–Chemical Shift Correlation Spectroscopy and Single Crystal Study of Triphenylene. *J Am Chem Soc* **1998,** *120* (36), 9305-9311.

348. Antzutkin, O. N.; Lee, Y. K.; Levitt, M. H., 13C and 15N-chemical shift anisotropy of ampicillin and penicillin-V studied by 2D-PASS and CP/MAS NMR. *J Magn Reson* **1998,** *135* (1), 144-55.

349. Chang, C.-J.; Díaz, L. E.; Morin, F.; Grant, D. M., Solid-state 13C NMR study of drugs: Aspirin. *Magnetic Resonance in Chemistry* **1986,** *24* (9), 768-771.

350. Harris, R. K.; Kenwright, A. M.; Say, B. J.; Yeung, R. R.; Fletton, R. A.; Lancaster, R. W.; Hardgrove, G. L., Cross-polarization/magic-angle spinning NMR studies of polymorphism: Cortisone acetate. *Spectrochimica Acta Part A: Molecular Spectroscopy* **1990,** *46* (6), 927-935.

351. Othman, A.; Evans, J. S.; Evans, I. R.; Harris, R. K.; Hodgkinson, P., Structural study of polymorphs and solvates of finasteride. *J Pharm Sci* **2007,** *96* (5), 1380-97.

352. Wenslow, R. M.; Baum, M. W.; Ball, R. G.; McCauley, J. A.; Varsolona, R. J., A spectroscopic and crystallographic study of polymorphism in an aza-steroid. *J Pharm Sci-Us* **2000,** *89* (10), 1271-1285.

353.   Pawlak, T.; Potrzebowski, M. J., Fine Refinement of Solid-State Molecular Structures of Leu- and Met-Enkephalins by NMR Crystallography. *The Journal of Physical Chemistry B* **2014,** *118* (12), 3298-3309.

354.   De Gortari, I.; Portella, G.; Salvatella, X.; Bajaj, V. S.; van der Wel, P. C.; Yates, J. R.; Segall, M. D.; Pickard, C. J.; Payne, M. C.; Vendruscolo, M., Time averaging of NMR chemical shifts in the MLF peptide in the solid state. *J Am Chem Soc* **2010,** *132* (17), 5993-6000.

355.   Tatton, A. S.; Pham, T. N.; Vogt, F. G.; Iuga, D.; Edwards, A. J.; Brown, S. P., Probing Hydrogen Bonding in Cocrystals and Amorphous Dispersions Using 14N–1H HMQC Solid-State NMR. *Molecular Pharmaceutics* **2013,** *10* (3), 999-1007.

356.   Foumier, R. PhD thesis. Durham university, 2006.

357.   Heider, E. M.; Harper, J. K.; Grant, D. M., Structural characterization of an anhydrous polymorph of paclitaxel by solid-state NMR. *Physical Chemistry Chemical Physics* **2007,** *9* (46), 6083-6097.

358.   Harper, J. K.; McGeorge, G.; Grant, D. M., Solid-State 13C Chemical Shift Tensors in Terpenes. 2. NMR Characterization of Distinct Molecules in the Asymmetric Unit and Steric Influences on Shift in Parthenolide. *J Am Chem Soc* **1999,** *121* (27), 6488-6496.

359.   Kimura, K.; Hirayama, F.; Uekama, K., Characterization of tolbutamide polymorphs (burger's forms II and IV) and polymorphic transition behavior. *J Pharm Sci-Us* **1999,** *88* (4), 385-391.

360.   Olsen, R. A.; Struppe, J.; Elliott, D. W.; Thomas, R. J.; Mueller, L. J., Through-Bond 13C–13C Correlation at the Natural Abundance Level: Refining Dynamic Regions in the Crystal Structure of Vitamin-D3 with Solid-State NMR. *J Am Chem Soc* **2003,** *125* (39), 11784-11785.

361.   Webber, A. L.; Masiero, S.; Pieraccini, S.; Burley, J. C.; Tatton, A. S.; Iuga, D.; Pham, T. N.; Spada, G. P.; Brown, S. P., Identifying guanosine self assembly at natural isotopic abundance by high-resolution 1H and 13C solid-state NMR spectroscopy. *J Am Chem Soc* **2011,** *133* (49), 19777-95.

362.   Harris, R. K.; Kenwright, A. M.; Fletton, R. A.; Lancaster, R. W., NMR and polymorphism of a steroidal anaesthetic. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1998,** *54* (12), 1837-1847.

363.   Barich, D. H.; Pugmire, R. J.; Grant, D. M.; Iuliucci, R. J., Investigation of the Structural Conformation of Biphenyl by Solid State 13C NMR and Quantum Chemical NMR Shift Calculations. *The Journal of Physical Chemistry A* **2001,** *105* (28), 6780-6784.

364.   Webber, A. L.; Emsley, L.; Claramunt, R. M.; Brown, S. P., NMR Crystallography of Campho[2,3-c]pyrazole (Z′ = 6): Combining High-Resolution 1H-13C Solid-State MAS NMR Spectroscopy and GIPAW Chemical-Shift Calculations. *The Journal of Physical Chemistry A* **2010,** *114* (38), 10435-10442.

365.   Ciuffreda, P.; Casati, S.; Manzocchi, A., Complete (1)H and (13)C NMR spectral assignment of alpha- and beta-adenosine, 2'-deoxyadenosine and their acetate derivatives. *Magn Reson Chem* **2007,** *45* (9), 781-4.

366.   Harris, R. K.; Jackson, P., High-resolution 1H and 13C NMR of solid 2-Aminobenzoic acid. *Journal of Physics and Chemistry of Solids* **1987,** *48* (9), 813-818.

367.   Pacilio, J. E.; Tokarski, J. T.; Quiñones, R.; Iuliucci, R. J., High-Resolution Solid-State NMR Spectroscopy: Characterization of Polymorphism in Cimetidine, a Pharmaceutical Compound. *Journal of Chemical Education* **2014,** *91* (8), 1236-1239.

368.   Tatton, A. S.; Pham, T. N.; Vogt, F. G.; Iuga, D.; Edwards, A. J.; Brown, S. P., Probing intermolecular interactions and nitrogen protonation in pharmaceuticals by novel 15N-edited and 2D 14N-1H solid-state NMR. *CrystEngComm* **2012,** *14* (8), 2654-2659.

369.   Stievano, L.; Tielens, F.; Lopes, I.; Folliet, N.; Gervais, C.; Costa, D.; Lambert, J.-F., Density Functional Theory Modeling and Calculation of NMR Parameters: An ab Initio Study of the Polymorphs of Bulk Glycine. *Cryst Growth Des* **2010,** *10* (8), 3657-3667.

370.   AbrahamAnuji; C., A.; GelbrichThomas; K., H.; J., G., NMR crystallography — Three polymorphs of phenobarbital. *Canadian Journal of Chemistry* **2011,** *89* (7), 770-778.

371.   Harris, R. K.; Hodgkinson, P.; Zorin, V.; Dumez, J.-N.; Elena-Herrmann, B.; Emsley, L.; Salager, E.; Stein, R. S., Computation and NMR crystallography of terbutaline sulfate. *Magnetic Resonance in Chemistry* **2010,** *48* (S1), S103-S112.

372.   Yates, J. R.; Pham, T. N.; Pickard, C. J.; Mauri, F.; Amado, A. M.; Gil, A. M.; Brown, S. P., An Investigation of Weak CH···O Hydrogen Bonds in Maltose Anomers by a Combination of Calculation and Experimental Solid-State NMR Spectroscopy. *J Am Chem Soc* **2005,** *127* (29), 10216-10220.

373.   Reddy, G. N.; Cook, D. S.; Iuga, D.; Walton, R. I.; Marsh, A.; Brown, S. P., An NMR crystallography study of the hemihydrate of 2', 3'-O-isopropylidineguanosine. *Solid State Nucl Magn Reson* **2015,** *65*, 41-8.

374.   Internal data from Robin Harris.

375.   Data measured at Cambridge by Robin Stein.

376.   Czernek, J.; Brus, J., The covariance of the differences between experimental and theoretical chemical shifts as an aid for assigning two-dimensional heteronuclear correlation solid-state NMR spectra. *Chem Phys Lett* **2014,** *608*, 334-339.

377.   Sardo, M.; Santos, S. M.; Babaryk, A. A.; López, C.; Alkorta, I.; Elguero, J.; Claramunt, R. M.; Mafra, L., Diazole-based powdered cocrystal featuring a helical hydrogen-bonded network: structure determination from PXRD, solid-state NMR and computer modeling. *Solid State Nucl Magn Reson* **2015,** *65*, 49-63.

378.   Cookney, J.; Light, M. E.; Benes, N. E.; Fila, V., CCDC 1483947: Experimental crystal structure determination. The Cambridge Crystallographic Data Centre (CCDC): 2016.

379.   Li, S.; Hong, M., Protonation, Tautomerization, and Rotameric Structure of Histidine: A Comprehensive Study by Magic-Angle-Spinning Solid-State NMR. *J Am Chem Soc* **2011,** *133* (5), 1534-1544.

380.   Bradley, J. P.; Velaga, S. P.; Antzutkin, O. N.; Brown, S. P., Probing Intermolecular Crystal Packing in γ-Indomethacin by High-Resolution 1H Solid-State NMR Spectroscopy. *Cryst Growth Des* **2011,** *11* (8), 3463-3471.

381.   Carignani, E.; Borsacchi, S.; Bradley, J. P.; Brown, S. P.; Geppi, M., Strong Intermolecular Ring Current Influence on 1H Chemical Shifts in Two Crystalline Forms of Naproxen: a Combined Solid-State NMR and DFT Study. *The Journal of Physical Chemistry C* **2013,** *117* (34), 17731-17740.

382.   Uldry, A.-C.; Griffin, J. M.; Yates, J. R.; Pérez-Torralba, M.; Santa María, M. D.; Webber, A. L.; Beaumont, M. L. L.; Samoson, A.; Claramunt, R. M.; Pickard, C. J.; Brown, S. P., Quantifying Weak Hydrogen Bonding in Uracil and 4-Cyano-4'-ethynylbiphenyl: A Combined Computational and Experimental Investigation of NMR Chemical Shifts in the Solid State. *J Am Chem Soc* **2008,** *130* (3), 945-954.

383.   Mafra, L.; Santos, S. M.; Siegel, R.; Alves, I.; Paz, F. A.; Dudenko, D.; Spiess, H. W., Packing interactions in hydrated and anhydrous forms of the antibiotic Ciprofloxacin: a solid-state NMR, X-ray diffraction, and computer simulation study. *J Am Chem Soc* **2012,** *134* (1), 71-4.

384.   Dudenko, D. V.; Williams, P. A.; Hughes, C. E.; Antzutkin, O. N.; Velaga, S. P.; Brown, S. P.; Harris, K. D. M., Exploiting the Synergy of Powder X-ray Diffraction and Solid-State NMR

Spectroscopy in Structure Determination of Organic Molecular Solids. *The Journal of Physical Chemistry C* **2013,** *117* (23), 12258-12265.

385.   Koike, R.; Higashi, K.; Liu, N.; Limwikrant, W.; Yamamoto, K.; Moribe, K., Structural Determination of a Novel Polymorph of Sulfathiazole–Oxalic Acid Complex in Powder Form by Solid-State NMR Spectroscopy on the Basis of Crystallographic Structure of Another Polymorph. *Cryst Growth Des* **2014,** *14* (9), 4510-4518.

386.   Venâncio, T.; Oliveira, L. M.; Ellena, J.; Boechat, N.; Brown, S. P., Probing intermolecular interactions in a diethylcarbamazine citrate salt by fast MAS 1H solid-state NMR spectroscopy and GIPAW calculations. *Solid State Nucl Mag* **2017,** *87*, 73-79.

387.   Sardo, M.; Siegel, R.; Santos, S. M.; Rocha, J.; Gomes, J. R. B.; Mafra, L., Combining Multinuclear High-Resolution Solid-State MAS NMR and Computational Methods for Resonance Assignment of Glutathione Tripeptide. *The Journal of Physical Chemistry A* **2012,** *116* (25), 6711-6719.

388.   Apperley, D. C.; Batsanov, A. S.; Clark, S. J.; Harris, R. K.; Hodgkinson, P.; Jochym, D. B., Computation of magnetic shielding to simultaneously validate a crystal structure and assign a solid-state NMR spectrum. *Journal of Molecular Structure* **2012,** *1015*, 192-201.

Curriculum Vitae

First, last name	Martins Balodis
Date of birth	January 2, 1988
Place of birth	Riga, Latvia
Nationality	Latvian
Address	Chemin de la colline 20, Crissier, 1023, Switzerland
Email	martins.balodis@epfl.ch
	balodis.lu@gmail.com
Phone number	+41 78 674 96 69, +371 26452734

Professional experience

2017-2021   Doctoral Assistant, Laboratory of Magnetic Resonance, full time
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Supervisor: Prof. Lyndon Emsley. Work done on experimental and computational aspects of nuclear magnetic resonance (NMR) crystallography and experimental aspects of dissolution dynamic nuclear polarization (d-DNP). Teaching assistant for master and bachelor students.

2016-2017   Scientific assistant, full time
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Supervisor: Prof. Lyndon Emsley. Work done on experimental and computational aspects of nuclear magnetic resonance (NMR) crystallography

2016   Scientific assistant, full time
Durham University, Durham, UK
Supervisor: Prof. Paul Hodgkinson. Work done on experimental and computational aspects of nuclear magnetic resonance (NMR) crystallography

2013-2015   Scientific assistant, part time
University of Latvia, Riga, Latvia
Scientific supervisor: Andris Actins. Work done on studying molecular crystals with emphasis on co-crystals with x-ray diffraction and thermal analysis methods. Teaching assistant for bachelor students.

2011-2013   Synthetic chemist, part time
Latvian Institute of Organic Synthesis, Riga, Latvia
Planning and realization of synthesis of organic molecules used as drug candidates.

Education

2017 – 2021   Ph. D. in chemistry
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Thesis title: New methods for structure determination and speciation by NMR crystallography. Supervisor: Prof. Lyndon Emsley.

2013 – 2015 **M. Sc. in chemistry, awarded with distinction (average mark 99.7%).**
University of Latvia, Riga, Latvia
Thesis title: Investigation of some organic reactions with the help of molecular modelling methods. Supervisor: Prof. Ödön Farkas.

2010 – 2013 **B. Sc. in chemistry, awarded with distinction (average mark 97.1%).**
University of Latvia, Riga, Latvia
Cocrystals of 4-(acetilamino)benzoic acid and nitrogen containing heterocyclic compounds. Supervisor: Dr. Liana Orola

List of publications

1. Balodis, M.; Cordova, M.; Hofstetter, A.; Ceriotti, M.; Day, G. M.; Emsley, L., Crystal structure determination via machine learned chemical shifts. *In preparation.*

2. Cordova, M.; Balodis, M.; Simoes de Almeida, B.; Ceriotti, M.; Emsley, L., Bayesian Probabilistic Assignment of Chemical Shifts in Organic Solids. *Submitted and in Revision,* **2021**

3. Cordova, M.; Balodis, M.; Hofstetter, A.; Paruzzo, F.; Lill, S. O. N.; Eriksson, E. S. E.; Berruyer, P.; de Almeida, B. S.; Quayle, M. J.; Norberg, S. T.; Ankarberg, A. S.; Schantz, S.; Emsley, L., Structure determination of an amorphous drug through large-scale NMR predictions. Nat Commun **2021**, 12 (1).

4. Hofstetter, A.; Balodis, M.; Paruzzo, F. M.; Widdifield, C. M.; Stevanato, G.; Pinon, A. C.; Bygrave, P. J.; Day, G. M.; Emsley, L., Rapid Structure Determination of Molecular Solids Using Chemical Shifts Directed by Unambiguous Prior Constraints. J Am Chem Soc **2019**, 141 (42), 16624-16634.

5. Rekis, T.; Beerziņš, A.; Sarcevica, I.; Kons, A.; Balodis, M.; Orola, L.; Lorenz, H.; Actins, A., A Maze of Solid Solutions of Pimobendan Enantiomers: An Extraordinary Case of Polymorph and Solvate Diversity. Cryst Growth Des **2018**, 18 (1), 264-273.

Languages

Latvian   First language
English   Advanced
Russian   Intermediate
French   Beginner

Awards and grants

2016-2019   Swiss government excellence scholarship
2013-2014   Mundheim family scholarship of excellence, University of Latvia
2012-2013   Morberga scholarship, University of Latvia