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

Albert Hofstetter,[†] Martins Balodis,[†] Federico M. Paruzzo,[†] Cory M. Widdifield,[§] Gabriele Stevanato,[†] Arthur C. Pinon,[†] Peter J. Bygrave,[¥] Graeme M. Day[¥] and Lyndon Emsley[†]

[†]Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

[§]Oakland University, Department of Chemistry, Mathematics and Science Center, 146 Library Drive, Rochester, MI 48309-4479, United States

[¥]School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

ABSTRACT: 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. 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.

1. INTRODUCTION

The 40,000-60,000 crystal structures published every year[1-4] 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.[5] 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.

Recently, solid-state nuclear magnetic resonance (NMR) based crystallography has emerged as a powerful tool to overcome the current limitations in powder structure determination. In contrast to other methods, such as powder XRD[6] or electron diffraction,[7-10] NMR directly probes the local atomic environment, allowing for structural characterization without the need for long-range order. Advances in solid-state NMR together with the development of accurate methods to calculate chemical shifts,[11-13] in particular using plane wave density functional theory (DFT) methods based on the gauge including projected augmented wave (PAW/GIPAW) approach,[14-16] have enabled development of chemical shift based NMR crystallography methods. Recently, machine-learning methods have been introduced, to facilitate the calculation of accurate chemical shifts.[17-18]

The NMR crystallography (NMRX) approach, which often involves crystal structure prediction (CSP) protocols, has been used to determine full *de novo* crystal structures from powders[19-24] with an accuracy at least comparable to single-crystal XRD,[25] as well as to determine elements of structure such as hydrogen bonding, proton positions and stereochemistry,[26-30] or to validate and refine crystal structures of molecular solids and to identify known polymorphs.[27-28, 31-55]

Because CSP can require considerable computational resources, which increases rapidly with the structural degrees of freedom, CSP based NMRX methods (CSP-NMRX) for *de novo* structure determination is currently limited to systems with up to about 10 degrees of torsional freedom within the molecule,[56] and going beyond this requires some prior knowledge or intuition.[23-24] Indeed, in order to circumvent these limitations CSP methods often make assumptions (for example based on space groups or predicted conformational energies) to limit the search space of possible structures. However, this can lead to failure of the CSP-NMRX method to determine the correct crystal structure(s), when the correct structure is excluded from the search space.

A common feature of most the CSP-NMRX methods developed to date is that they exploit geometric 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.

Note that an example has been given in which structural information was included at earlier stages of the CSP by biasing the search using semi-empirical molecule specific pseudo-forces derived from chemical shifts.[44]

Here, 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 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,[18] we correctly determine the crystal structure of powdered ampicillin, for which the usual approach to CSP-NMRX would have failed.

2. METHODS

Figure 1a schematically illustrates the workflow in a successful case for the current CSP-NMRX approaches.[19, 42, 57] 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,[58] 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.[59]

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)[23, 27, 49, 60-62] 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.[25] 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 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,

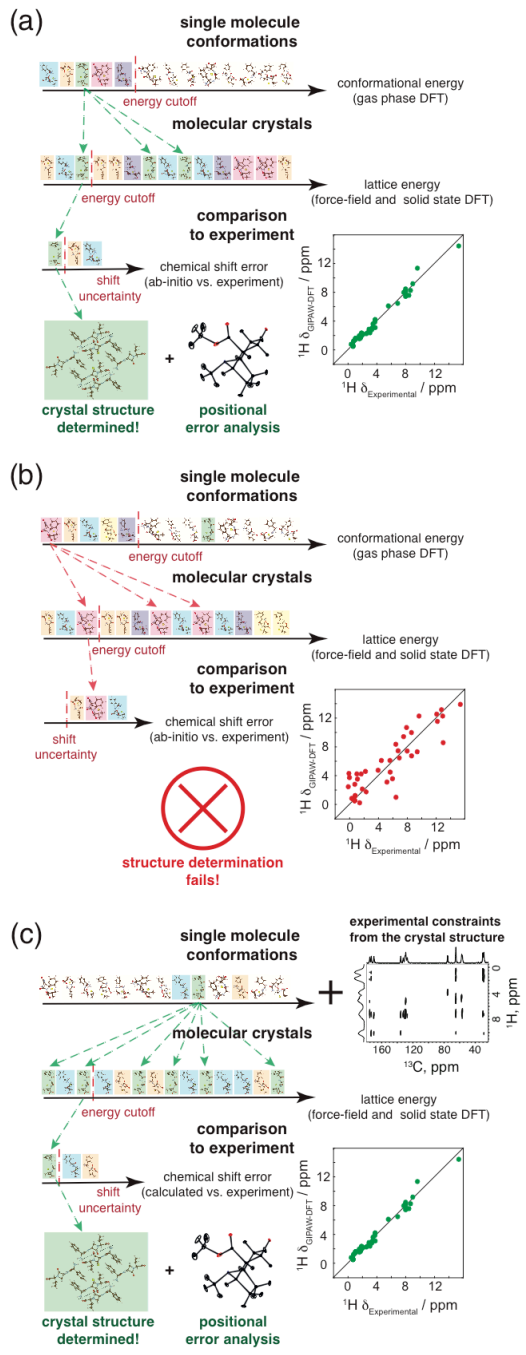

Figure 1. Schematic of current and proposed CSP-NMRX methods. **(a)** an example of a successful structure prediction using the current CSP-NMRX method. **(b)** an example of a failed structure prediction using the current CSP-NMRX method. **(c)** an example of the proposed experimentally constrained 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 ^{1}H chemical shift plotted against the DFT-calculated ^{1}H chemical shift for the trial structure with the lowest error between DFT and experimental chemical shifts.

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 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.

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[21, 63-73] 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.

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**, 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** 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) (**Figures 2c-d**) or from a short inter-molecular interatomic distance (which can be from the "closed" or the "open" conformer) (**Figures 2b-c**). Thus, the presence of a cross peak does not contain unambiguous information about the single molecule conformer, as the fragments in **Figure 2b-d** contain both possible conformations.

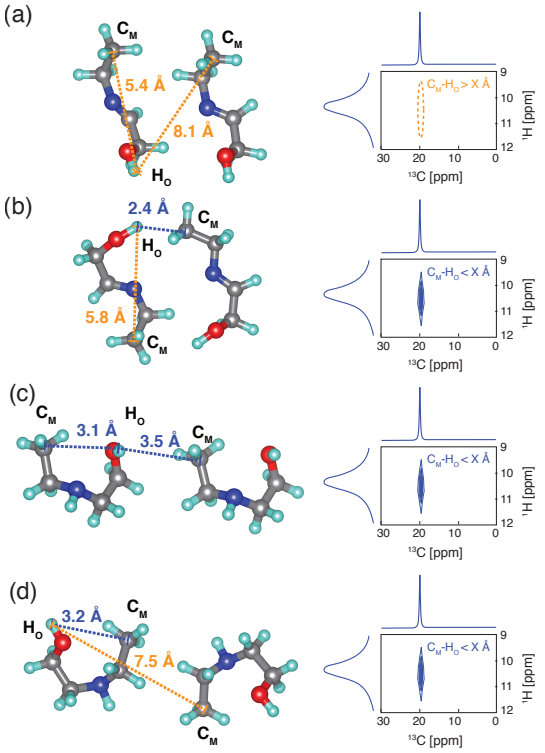

Figure 2a-d. Schematic illustrations of ^{1}H-^{13}C HETCOR spectra (right) for four different structural fragments (left) and the derived constraints. Structures (a) and (b) contain an "open" conformer. Structures (c) and (d) contain a "closed" conformer. Blue dotted lines are sufficiently short C-H 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.

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 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 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.[74] 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 S3a** 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.

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 S1b.**

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.

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 S3b**.

However, **Figure S3b** 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 S5**, and marked with a green dotted-line in **Figure S4a**. This results in a clearer correlation between normalised SNR and the corresponding inter-atomic distances, as shown in **Figure S4b**.

From **Figure S4b** 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.*[42] 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 3a shows the set of parameters for which the selection procedure was successful for all three molecules simultaneously. **Figure S9** shows the set of successful parameters for each molecule individually. The dashed orange line in **Figure 3a** 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.

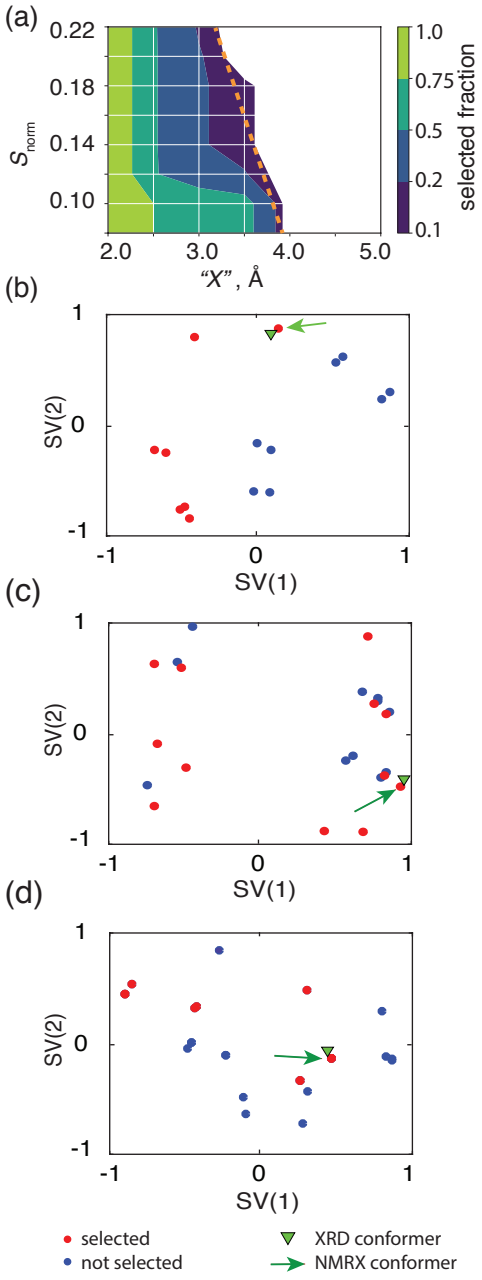

Figure 3 (a) Grid search results of the threshold distance "X" and S_{norm} cut-off values for flutamide, cocaine and flufenamic acid. The colour-map shows the percentage of selected structures from within the conformer ensemble. The white area indicates the region where the correct conformer is not 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.

To aid our interpretation of the selection procedure we apply a sketch-map[75-78] 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 the SI in **Figures S6**, **S7** and **S8**.

Flutamide. The initial gas-phase ensemble of flutamide conformers generated in the first step of CSP contains 15 conformers,[42] of which 7 are in the *trans* and 8 are in the *cis* conformation with respect to the amide group (**Figure S7**). All the absent cross-peaks in a series of ^{1}H-^{13}C HETCOR spectra (**Figure 4a**) were used to generate the conformational constraints shown in **Figure 4a**. **Figure 3b** 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,[42] 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 rans form in all 7 conformers (see **Figure S10**).

Flufenamic acid. The gas-phase conformer ensemble for flufenamic acid contains 26 molecular conformations.[42] **Figure 3c** 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 4c**. 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 **Figures 4c** and **S10**) 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.

Cocaine. The initial CSP conformer ensemble for cocaine contains 27 conformers.[42] **Figure 3d** 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 4b**). 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 S11** 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 5b** and **S3** 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.

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 was generated (for details, see SI) and the ensemble was then sorted according to the isolated molecule conformational energy. **Figure 5b** and **S5** 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 5b also shows that the single molecule conformation closest to the crystal conformer is one of the highest energy gas phase conformers, nearly 100 kJ · mol^{-1} higher in energy than the lowest energy single molecule conformer. In the normal CSP method a cut-off of around 20-25 kJ · mol^{-1} would typically be applied to the conformational ensemble[42, 58] to limit the number of conformers that must be considered during the time-consuming crystal packing search. The correct conformer falls well outside this energy range and, thus, would be eliminated at this stage, preventing the successful generation of the observed crystal structure. To successfully determine the correct crystal structure, the subsequent CSP steps would have had to proceed without applying any energetic cutoff on the single-molecule conformers. This would be possible for the 16 conformers of ampicillin and use of large scale computing to perform the searches in parallel, but is problematic as a general method, as the conformational space of even moderately flexible molecules can often include hundreds of individual conformers.[58]

To address this problem, we apply experimental constraints extracted from ^{1}H-^{13}C HETCOR NMR spectra at different contact times (i.e. 0.1, 0.3, 0.5, 0.75, 1.0, 1.25,

1.5, 1.75, 2.0 and 2.25 ms, detailed in the SI). **Figure 4d** shows the assigned HETCOR NMR spectra of ampicillin at a 1.5 ms contact time together with the labelled 2D structure. Following the protocol established for cocaine, flutamide and flufenamic acid, the SNR was then normalised over all experimental setups as described above. As we did for the other three molecules, we only consider cross-peaks resulting from *terminal*-protons, see **Figure S4a**. Using an S_{norm} of 0.14, and an X of 3.5 Å, that were parametrized on the reference compounds above, the extracted constraints are circled in orange and are shown on three example conformers below the spectra. **Figure 5a** shows the sub-ensembles with no violations (0 out of 1 total constraint). **Figures 4d** and **S13** show that only conformers without an intra-molecular hydrogen bond are selected. Also, from **Figure 5b** it is clear that the energetically high conformers are preferentially selected. Note, that in a classical CSP-NMRX approach these conformers would have not been selected.

For the next step in the CSP procedure, only 7 out of the original 16 structures were considered. This reduces the computational cost by approximately 55%.

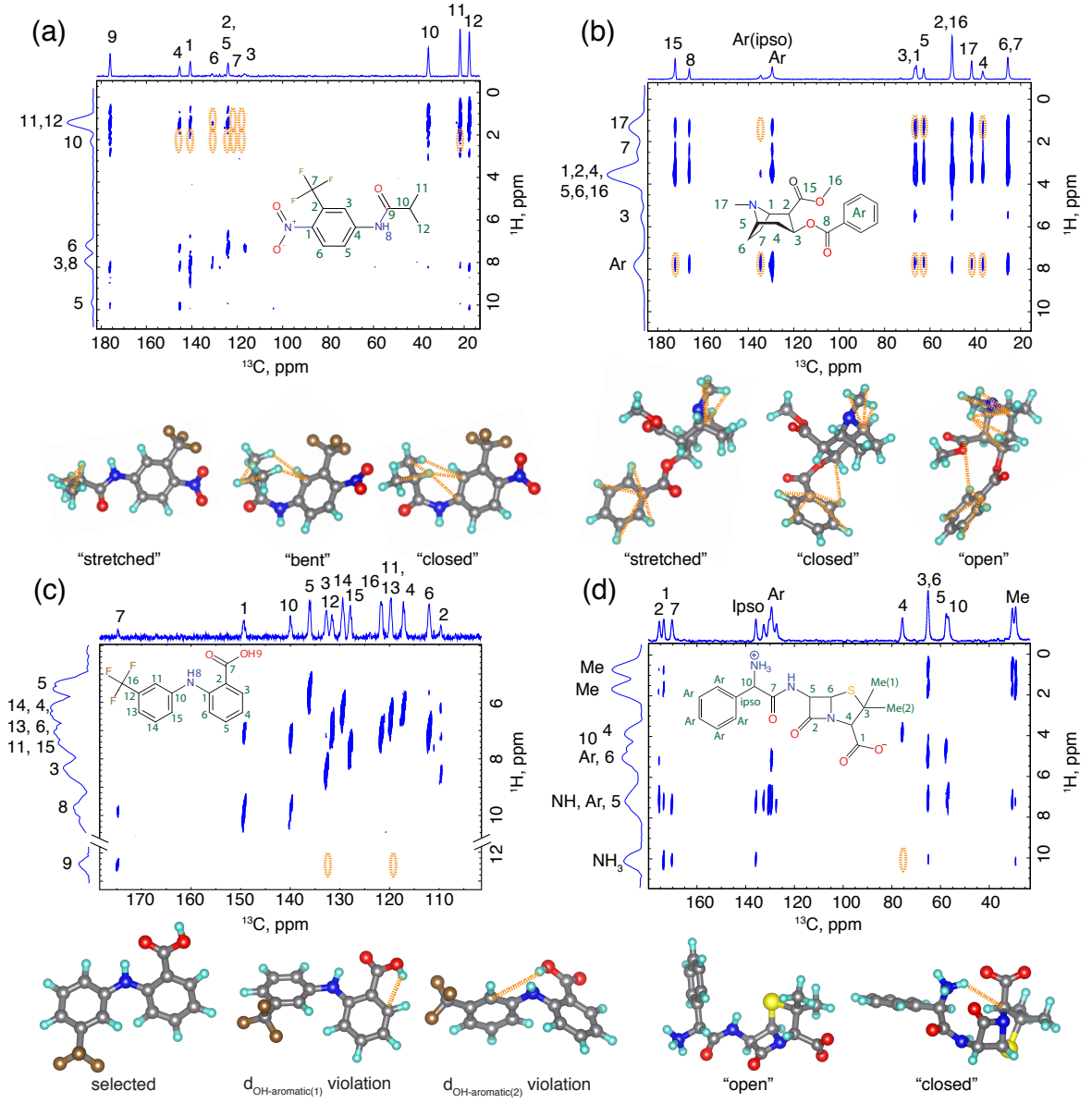

Figure 4 The top part in each panel shows the ^{1}H-^{13}C HETCOR spectrum of: flutamide with a 1.25 ms ^{1}H-^{13}C cross-polarization contact time **(a)**, cocaine with a 1.0 ms contact time **(b)**, flufenamic acid with a 1.5 ms contact time **(c)** and ampicillin with a 1.5 ms contact time **(d)** (further details and raw data in SI). ^{13}C peaks are assigned based on the literature[79] and ^{1}H peaks are assigned from HETCOR spectra and DFT chemical shift calculations (see SI). The cross-peaks from the terminal protons (**Figure S4a**) 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.

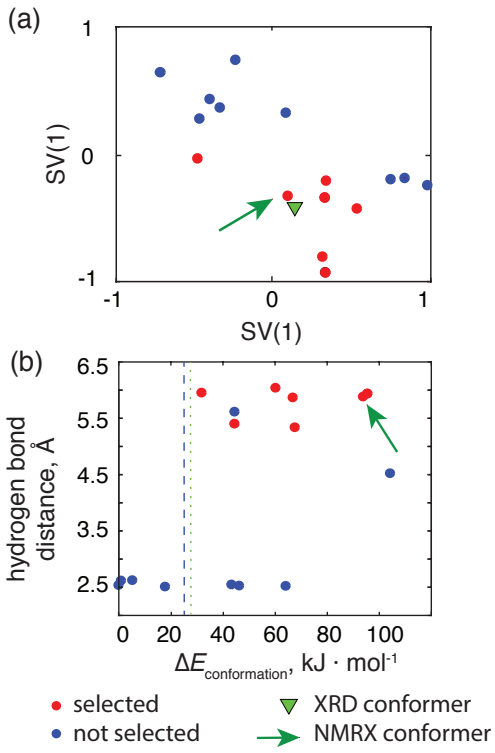

Figure 5. Conformer selection for ampicillin. **(a)** The panel shows the sketch-map projections of the gas-phase conformer ensemble. Red dots represent the structures which are selected using a threshold distance "X" of 3.5 Å and a S_{norm} of 0.14. The green triangle denotes the conformer found in 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[80] 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,[81] 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 (Section X).

^{1}H chemical shift values were then calculated with GIPAW DFT and a machine learned method (ShiftML)[18] for each candidate structure and compared to the experimental chemical shifts (details are given in the SI). **Figure 6** 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 (±0.16 ppm) or lower.[39] This is indicated as the grey zone in **Figure 6**. 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 S15 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 (as described in the SI, **Table S8**) we expect a correct structure to have a ^{1}H RMSE of 0.346 ppm (±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 6**).

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 (**Figures 6, S15**). 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)[82] above the best candidate structure. Thus, our final structure selection relies on both the chemical shifts and calculated lattice energies.

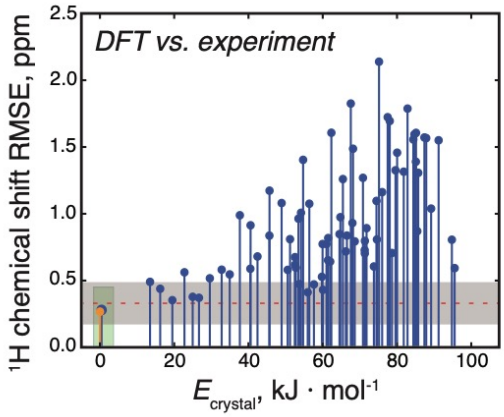

Figure 6. Comparison of crystal structure candidates. The structures are sorted according to their relative lattice energy, as specified on the horizontal axis. The vertical axis shows ^{1}H chemical shift RMSE between DFT calculated and experimental chemical shifts. The orange marker shows the ^{1}H chemical shift RMSE for the single-crystal XRD structure. The red line shows the mean of the current difference between experimental and DFT calculated ^{1}H chemical shifts with the distinguishability limits (at the 1σ level) indicated as grey shaded zone, as described in the main text.

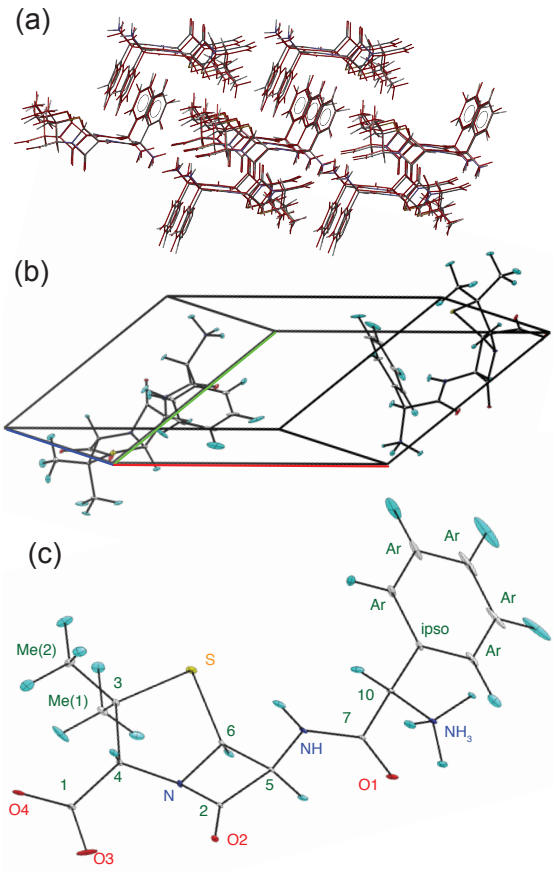

Figure 7 (a) Comparison between the structure of ampicillin as determined by the constrained powder ^{1}H CSP-NMRX and the single crystal XRD determined structure.[83] **(b-c)** ORTEP plot of the ampicillin crystal **(b)** and single molecule **(c)** structure drawn at the 90% probability level. The 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. (d)

The structure determined here agrees very well with the known reference structure determined by single-crystal XRD,[83] as illustrated in **Figure 7a.** 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 (see **Table S9**). 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.

Finally, we proceed with a positional error analysis that leads to the fully determined structure shown in **Figures 7c-d** and **S17**. The positional error analysis is performed using the DFT-calculated ^{1}H chemical shifts following the procedure outlined by Hofstetter and Emsley[25] 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.[83] 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.

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.

CORRESPONDING AUTHOR

E-mail: lyndon.emsley@epfl.ch,

ACKNOWLEDGEMENTS

We thank Edgar A. Engel, Andrea Anelli, Felix Musil and Prof. Michele Ceriotti (EPFL, Lausanne) for developing and parametrizing the version of ShiftML used here, details of which are subject to another publication.

We are grateful for financial support from Swiss National Science Foundation Grant No. 200021_160112. PJB and GMD thank the European Research Council for funding under grant ERC-StG-2012-ANGLE-307358, as well as the EPSRC via grant EP/J01110X/1. We acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton. We also acknowledge the ARCHER UK National Supercomputing Service which was accessed via the UK's HPC Materials Chemistry Consortium membership, which is funded by the EPSRC (grant EP/L000202).

REFERENCES

1. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge structural database. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2016,** *72* (2).

2. Grazulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quiros, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A., Crystallography Open Database - an open-access collection of crystal structures. *J Appl Crystallogr* **2009,** *42*, 726-729.

3. 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 Res* **2012,** *40* (D1), D420-D427.

4. Brown, I. D.; Bergerhoff, G., Databases of Inorganic Crystal-Structures. *Abstr Pap Am Chem S* **1979,** (Sep), 36-36.

5. 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.

6. Rietveld, H. M., A Profile Refinement Method for Nuclear and Magnetic Structures. *J Appl Crystallogr* **1969,** *2*, 65-&.

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

8. 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.

9. Gruene, T.; Wennmacher, J. T. C.; Zaubitzer, C.; Holstein, J. J.; Heidler, J.; Fecteau-Lefebvre, A.; De Carlo, S.; Muller, 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. *Angew Chem Int Edit* **2018,** *57* (50), 16313-16317.

10. G Jones, C.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H.; Gonen, T., The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. **2018**.

11. Dedios, A. C.; Pearson, J. G.; Oldfield, E., Secondary and Tertiary Structural Effects on Protein Nmr Chemical-Shifts - an Abinitio Approach. *Science* **1993,** *260* (5113), 1491-1496.

12. Facelli, J. C.; Grant, D. M., Determination of molecular symmetry in crystalline naphthalene using solid-state NMR. *Nature* **1993,** *365* (6444), 325-7.

13. Sebastiani, D.; Parrinello, M., A new ab-initio approach for NMR chemical shifts in periodic systems. *J Phys Chem A* **2001,** *105* (10), 1951-1958.

14. Blochl, P. E., Projector augmented-wave method. *Phys Rev B Condens Matter* **1994,** *50* (24), 17953-17979.

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

16. Yates, J. R.; Pickard, C. J.; Mauri, F., Calculation of NMR chemical shifts for extended systems using ultrasoft pseudopotentials. *Phys Rev B* **2007,** *76* (2), 024401.

17. Cuny, J.; Xie, Y.; Pickard, C. J.; Hassanali, A. A., Ab Initio Quality NMR Parameters in Solid-State Materials Using

a High-Dimensional Neural-Network Representation. *Journal of Chemical Theory and Computation* **2016,** *12* (2), 765-773.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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 (129) Xe NMR Spectroscopy. *Chemistry* **2017,** *23* (22), 5258-5269.

23. 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.

24. 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. *Mol Pharm* **2018,** *15* (4), 1476-1487.

25. Hofstetter, A.; Emsley, L., Positional Variance in NMR Crystallography. *J Am Chem Soc* **2017,** *139* (7), 2573-2576.

26. Harper, J. K.; Mulgrew, A. E.; Li, J. Y.; Barich, D. H.; Strobel, G. A.; Grant, D. M., Characterization of stereochemistry and molecular conformation using solid-state NMR tensors. *Journal of the American Chemical Society* **2001,** *123* (40), 9837-9842.

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

28. Heider, E. M.; Harper, J. K.; Grant, D. M., Structural characterization of an anhydrous polymorph of paclitaxel by solid-state NMR. *Phys Chem Chem Phys* **2007,** *9* (46), 6083-97.

29. Witter, R.; Priess, W.; Sternberg, U., Chemical shift driven geometry optimization. *J Comput Chem* **2002,** *23* (2), 298-305.

30. 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.

31. Ochsenfeld, C.; Brown, S. P.; Schnell, I.; Gauss, J.; Spiess, H. W., Structure assignment in the solid state by the coupling of quantum chemical calculations with NMR experiments: a columnar hexabenzocoronene derivative. *J Am Chem Soc* **2001,** *123* (11), 2597-606.

32. Goward, G. R.; Sebastiani, D.; Schnell, I.; Spiess, H. W.; Kim, H. D.; Ishida, H., Benzoxazine oligomers: evidence for a helical structure from solid-state NMR spectroscopy and DFT-based dynamics and chemical shift calculations. *J Am Chem Soc* **2003,** *125* (19), 5792-800.

33. Harris, R. K., NMR crystallography: the use of chemical shifts. *Solid State Sciences* **2004,** *6* (10), 1025-1037.

34. Harris, R. K., NMR studies of organic polymorphs and solvates. *Analyst* **2006,** *131* (3), 351-73.

35. Mifsud, N.; Elena, B.; Pickard, C. J.; Lesage, A.; Emsley, L., Assigning powders to crystal structures by high-resolution (1)H-(1)H double quantum and (1)H-(13)C J-INEPT solid-state NMR spectroscopy and first principles computation. A case study of penicillin G. *Phys Chem Chem Phys* **2006,** *8* (29), 3418-22.

36. Harris, R. K., Applications of solid-state NMR to pharmaceutical polymorphism and related matters. *J Pharm Pharmacol* **2007,** *59* (2), 225-39.

37. Othman, A.; Evans, J. S.; Evans, I. R.; Harris, R. K.; Hodgkinson, P., Structural study of polymorphs and solvates of finasteride. *J Pharm Sci-Us* **2007,** *96* (5), 1380-97.

38. Salager, E.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L., Powder NMR crystallography of thymol. *Phys Chem Chem Phys* **2009,** *11* (15), 2610-21.

39. 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.

40. 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. *J Phys Chem A* **2010,** *114* (38), 10435-42.

41. Dudenko, D.; Kiersnowski, A.; Shu, J.; Pisula, W.; Sebastiani, D.; Spiess, H. W.; Hansen, M. R., A strategy for revealing the packing in semicrystalline pi-conjugated polymers: crystal structure of bulk poly-3-hexyl-thiophene (P3HT). *Angew Chem Int Ed Engl* **2012,** *51* (44), 11068-72.

42. 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.

43. Pawlak, T.; Jaworska, M.; Potrzebowski, M. J., NMR crystallography of alpha-poly(L-lactide). *Phys Chem Chem Phys* **2013,** *15* (9), 3137-45.

44. Santos, S. M.; Rocha, J.; Mafra, L., NMR crystallography: Toward chemical shift-driven crystal structure determination of the β-lactam antibiotic amoxicillin trihydrate. *Crystal Growth and Design* **2013,** *13* (6).

45. 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.

46. Ludeker, D.; Brunklaus, G., NMR crystallography of ezetimibe co-crystals. *Solid State Nucl Mag* **2015,** *65*, 29-40.

47. 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.

48. Pinon, A. C.; Rossini, A. J.; Widdifield, C. M.; Gajan, D.; Emsley, L., Polymorphs of Theophylline Characterized by DNP Enhanced Solid-State NMR. *Mol Pharm* **2015,** *12* (11), 4146-53.

49. 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).

50. 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.

51. 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), 6685-8.

52. Mali, G., Ab initio crystal structure prediction of magnesium (poly)sulfides and calculation of their NMR parameters. *Acta Crystallogr C Struct Chem* **2017,** *73* (Pt 3), 229-233.

53. Hartman, J. D.; Day, G. M.; Beran, G. J. O., Enhanced NMR Discrimination of Pharmaceutically Relevant Molecular Crystal Forms through Fragment-Based Ab Initio Chemical Shift Predictions. *Crystal Growth and Design* **2016,** *16* (11).

54. 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.

55. Hartman, J. D.; Beran, G. J. O., Accurate 13-C and 15-N molecular crystal chemical shielding tensors from fragment-based electronic structure theory. *Solid State Nucl Mag* **2018,** *96*, 10-18.

56. Neumann, M. A.; van de Streek, J.; Fabbiani, F. P.; Hidber, P.; Grassmann, O., Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nat Commun* **2015,** *6*, 7793.

57. 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.

58. 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.

59. 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.

60. Chierotti, M. R.; Gobetto, R., NMR crystallography: the use of dipolar interactions in polymorph and co-crystal investigation. *Crystengcomm* **2013,** *15* (43).

61. 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.

62. 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.

63. 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.

64. 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.

65. Raleigh, D.; Levitt, M.; Griffin, R., Rotational resonance in solid state NMR. *Chemical Physics Letters* **1988,** *146* (1-2), 71-76.

66. 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.

67. 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.

68. 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.

69. Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L., Molecular structure determination in powders by NMR crystallography from proton spin diffusion. *Journal of the American Chemical Society* **2006,** *128* (29), 9555-9560.

70. 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.

71. 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.

72. 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.

73. 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**.

74. 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.

75. M. Cerrioti; S. De; R. H. Meissner; Tribello, G. A. Sketch map package. https://github.com/cosmo-epfl/sketchmap/.

76. 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.

77. 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.

78. 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*.

79. 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.

80. 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).

81. 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.

82. Nyman, J.; Day, G. M., Static and lattice vibrational energy differences between polymorphs. *Crystengcomm* **2015,** *17* (28), 5154-5165.

83. 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.

TOC Graphic