Powder Crystallography of Pharmaceutical Materials by Combined Crystal Structure Prediction and Solid-State ^{1}H NMR Spectroscopy

Maria Baias[a], Cory M. Widdifield[a], Jean-Nicolas Dumez[a], Hugh P. G. Thompson[b], Timothy G. Cooper[b], Elodie Salager[a], Sirena Bassil[a], Robin S. Stein[c], Anne Lesage[a], Graeme M. Day*[d], and Lyndon Emsley*[a]

A protocol for the *ab initio* crystal structure determination of powdered solids at natural isotopic abundance by combining solid-state NMR spectroscopy, crystal structure prediction, and DFT chemical shift calculations was evaluated to determine the crystal structures of four small drug molecules: cocaine, flutamide, flufenamic acid, and theophylline. For cocaine, flutamide and flufenamic acid, we find that the assigned ^{1}H isotropic chemical shifts provide sufficient discrimination to determine the correct structures from a set of predicted structures using the root-mean-square deviation (rmsd) between experimentally determined and calculated chemical shifts. In most cases unassigned shifts could not be used to determine the structures. This method requires no prior knowledge of the crystal structure, and was used to determine the correct crystal structure to within an atomic rmsd of less than 0.12 Å with respect to the known reference structure. For theophylline, the NMR spectra are too simple to allow for unambiguous structure selection.

[a] *Université de Lyon, (CNRS / ENS-Lyon / UCB Lyon 1), Centre de RMN à Très Hauts Champs, 5 rue de la Doua, 69100 Villeurbanne, France. Fax: +33 4 78 89 67 61; E-mail: lyndon.emsley@ens-lyon.fr.*
[b] *Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.*
[c] *now at Bruker Ltd., 555 Steeles Avenue East, Milton, ON L9T 1Y6, Canada.*
[d] *Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. E-mail: G.M.Day@soton.ac.uk.*

1. Introduction

Structural characterization remains one of the key challenges for modern chemistry. Single crystal diffraction methods are capable of characterizing systems as diverse as membrane proteins, whole virus particles, or complex inorganic materials. In contrast, if the sample is a powder, structural characterization represents an enormous challenge and other methods of characterizing powdered solids are required.

In this respect solid-state NMR holds great promise, and several approaches can be envisaged. For example, if the sample can be isotopically labelled using methods available for biological systems, the structure can then be obtained from dipolar recoupling or spin diffusion measurements. For small molecules, however, an approach based on the analysis of chemical shifts would be most attractive. There are today many examples of chemical shifts being combined with density functional theory (DFT) calculations for structure validation in organic molecular compounds with respect to known structures[1-17]. However, there are very few examples of *ab initio* structure determination from powders by NMR without a structural hypothesis.

The development of computational methods for crystal structure prediction (CSP) has been based predominantly on global lattice energy minimisation, applying various methods for locating low energy structures on the crystal energy landscape described by some model of the interactions between atoms or molecules. The scope of these methods has improved in recent years, providing the ability to predict possible stable phases of a wide range of materials. For example, these methods have been used for the successful prediction of single[18-22] and multicomponent[19,23-26] organic molecular crystals[27-30], high-pressure phases of materials[31-33], and other crystalline network structures such as zeolites or carbon polymorphs[34-36]. Progress in this field for organic molecular crystals is regularly assessed in a series of blind tests of structure prediction[37].

Recently, we introduced a method for *ab initio* natural abundance powder NMR crystallography by combining measured NMR chemical shifts and computational structure prediction.[38] The method has so far only been illustrated on a single example, the small molecule thymol. Here, we investigate the feasibility and limitations of this method with four examples of pharmacologically relevant substances: cocaine[39], a dopamine uptake inhibitor drug used in anaesthetics; flutamide[40], a non-steroidal androgen antagonist used for the treatment of prostate cancer; flufenamic acid[41], a non-steroidal anti-inflammatory drug; and theophylline[42], a drug used for the treatment of asthma. We find that this method works well to identify the correct structure from the list of low energy structures generated by CSP methodologies, provided that there are a sufficient number of assigned experimental NMR resonances. We successfully determine the correct structure of cocaine, flutamide and flufenamic acid based on the root-mean-square deviation (rmsd) resulting from the comparison of experimental and DFT calculated ^{1}H chemical shifts. However, in the case of theophylline the ^{1}H NMR spectrum is too simple for the method to unambiguously identify the correct structure based on comparing observed and calculated chemical shifts.

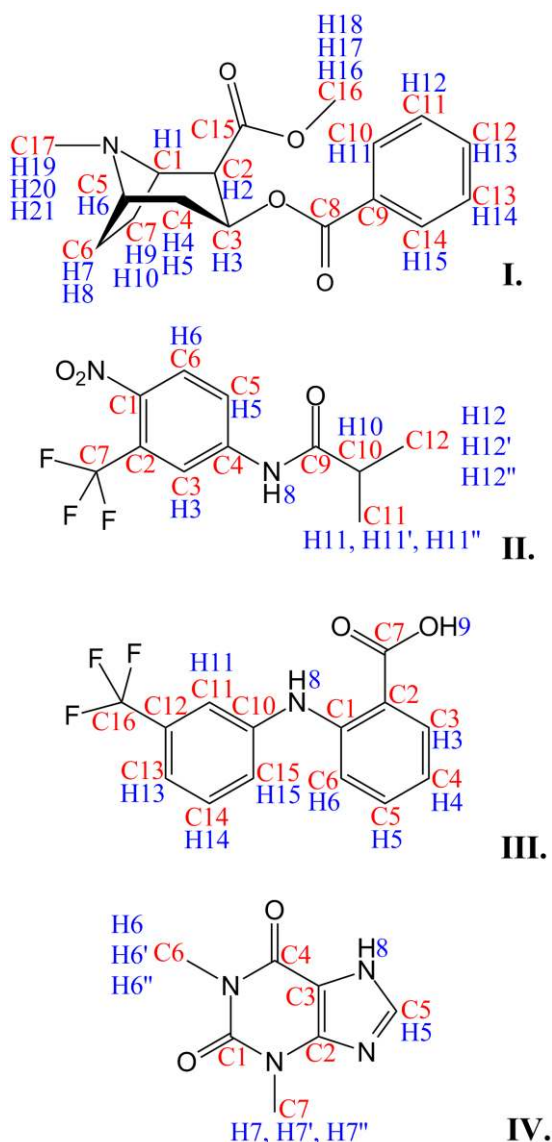# 2. Experimental

2a Samples

Powdered free base cocaine (methyl(1R,2R,3S,5S)-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylate) was purchased from Sigma-Aldrich and used without further purification. The reference crystal structure of cocaine, (CSD entry code: COCAIN10) (Scheme 1-I), was previously determined by single crystal X-ray diffraction (XRD)[39] at room temperature. The structure is comprised of 2 symmetry equivalent molecules in the unit cell, and it belongs to the $P2_1$ monoclinic space group with unit cell dimensions: $a = 10.130(1)$ Å, $b = 9.866(2)$ Å, $c = 8.445(1)$ Å.

Powdered flutamide (2-methyl-N(4-nitro-3-(trifluoromethyl)phenyl)propamide) was purchased from Sigma-Aldrich and used without further purification. The reference crystal structure of flutamide, (CSD entry code: WEZCOT) (Scheme 1-II), was previously determined by single crystal XRD[40] at 294 K. The structure is comprised of 4 symmetry equivalent molecules in the unit cell, and it belongs to the $Pna2_1$ orthorhombic space group with unit cell dimensions: $a = 11.856(2)$ Å, $b = 20.477(3)$ Å, $c = 4.9590(9)$ Å.

Powdered flufenamic acid (2-((3-(trifluoromethyl)phenyl)amino)-benzoic acid) was purchased from Fluka and used without further purification. The reference crystal structure of the corresponding flufenamic acid polymorph, (CSD entry code: FPAMCA11) (Scheme 1-III), was previously determined from single crystal XRD data[41] recorded at room temperature, and was confirmed by powder XRD to be the polymorph studied here. The structure comprises 4 symmetry equivalent molecules in the unit cell, and it belongs to the $P2_1/c$ monoclinic space group with unit cell dimensions: $a = 12.523(4)$ Å, $b = 7.868(6)$ Å, $c = 12.874(3)$ Å, and angle $\beta = 95.2(2)°$.

Powdered theophylline (3,7-dihydro-1,3-dimethyl-1H-purine-2,6-dione) was purchased from Acros Organics and used without further purification. The polymorphic form was confirmed to be the most stable orthorhombic polymorph by Ebisuzaki *et al.*[42]. The reference crystal structure of theophylline (CSD entry code: BAPLOT01) (Scheme 1-IV) was previously determined[42] from single crystal X-ray diffraction at room temperature. The structure comprises 4 symmetry equivalent molecules in the unit cell, and it belongs to the $Pna2_1$ orthorhombic space group with unit cell dimensions: $a = 24.612(2)$ Å, $b = 3.8302(4)$ Å, $c = 8.5010(5)$ Å.

For each sample the crystals were carefully ground to give a fine homogeneous (microcrystalline) powder before performing the NMR experiments.

Scheme 1 Molecular structure of cocaine (**I**), flutamide (**II**), flufenamic acid (**III**) and theophylline (**IV**) and the labelling scheme used here.

2b NMR experiments

All NMR experiments were performed at a nominal temperature of 293 K with a Bruker Avance III spectrometer operating at ^{1}H and [13]C Larmor frequencies of 500 MHz and 125 MHz, respectively. 1D ^{1}H MAS spectra were recorded with a 1.3 mm double resonance probe with 60 kHz magic angle spinning (MAS). 1D ^{13}C cross-polarisation MAS (CPMAS) NMR spectra were recorded with 4 mm double or triple resonance probes at 12.5 kHz MAS. [1]H chemical shifts were referenced to the single resonance observed for protons in adamantane at 1.87 ppm with respect to the signal for neat TMS. [13]C chemical shifts were referenced to the CH_2 resonance observed for adamantane at 38.48 ppm with respect to the signal for neat TMS[43].

The 2D refocused ^{13}C-^{13}C INADEQUATE[44] NMR spectra were recorded with a 4 mm triple resonance probe at 12.5 kHz MAS. The SPINAL-64[45] sequence at a proton nutation frequency ν_1 of 80 kHz was used for heteronuclear

decoupling. 256 increments of 512 transients each were acquired with a repetition delay of 4.5 s, resulting in a total experimental time of 7 days for cocaine. 64 increments of 1024 transients each were acquired with a repetition delay of 4 s, resulting in a total experimental time of 3 days for flutamide. The delay in the echo blocks was τ = 4 ms. The acquisition time was set to 22 ms in t_2, and the maximum t_1 delay was t_1^{max} = 2.6 ms. Exponential line broadening of 70 Hz was applied in the direct and indirect dimensions prior to the Fourier transform.

The 2D refocused ^{1}H-^{13}C INEPT[46] NMR spectra were recorded with a 4 mm double resonance probe at 12.5 kHz MAS. The SPINAL-64[45] sequence at a proton ν_1 of 100 kHz was used for heteronuclear decoupling. The eDUMBO-1_{22}[47] sequence at a nutation frequency of 100 kHz was used for proton homonuclear decoupling in the indirect dimension. The ^{1}H axis has been corrected for the experimentally determined homonuclear decoupling scaling factor using a value of 0.6. For cocaine 128 increments of 256 transients each were acquired with a repetition delay of 8 s, resulting in a total experimental time of 3 days. For flutamide 200 increments of 16 transients each were acquired with a repetition delay of 3 s, resulting in a total experimental time of 3 h. The delay in the echo blocks was τ = 1.9 ms. The acquisition time in t_2 was set to 15 ms, and t_1^{max} = 10 ms. Exponential line broadening of 20 Hz was applied in the direct and indirect dimensions prior to Fourier transform.

The 2D [1]H-^{13}C high-resolution HETCOR NMR spectra were recorded with a 4 mm triple resonance probe at 12.5 kHz MAS. The SPINAL-64[45] sequence (ν_1 = 100 kHz) was used for heteronuclear decoupling. The eDUMBO-1_{22}[47] sequence (ν_1 = 100 kHz) was used for proton homonuclear decoupling in the indirect dimension. The ^{1}H axis has been corrected for the experimentally determined homonuclear decoupling scaling factor using a value of 0.49. 178 increments of 48 transients each were acquired for cocaine and 100 increments of 16 transients each were acquired for flufenamic acid, with an acquisition time in t_2 of 27 ms, and a repetition delay of 20 s, resulting in a total experimental time of 19 h for cocaine and 36 h for flufenamic acid. Exponential line broadenings of 50 Hz were used in the direct and indirect dimensions.

The States-TPPI procedure was used for quadrature detection in the indirect dimension for all two-dimensional experiments.

2c Crystal Structure Prediction

Crystal structures were predicted by global lattice energy minimisation, starting from the chemical formulae of each of the molecules investigated here and without any structural hypothesis or any information obtained from the known crystal structure.

Theophylline was treated using a rigid molecular geometry throughout the calculations, using a B3LYP/6-31G(d,p) optimised molecular geometry. Crystal structures were generated with the CrystalPredictor program[48], using quasi-random sampling of unit cell dimensions, molecular positions and orientations within a set of the most commonly observed space groups, all with one molecule in the asymmetric unit (Z´=1). The resulting structures were initially lattice energy

minimised using an empirically parameterised exp-6 atom-atom model of repulsion-dispersion interactions (the FIT potential described by Coombes *et al.*[49]) and electrostatic interactions modelled using atomic partial charges fitted to reproduce the molecular electrostatic potential. The resulting lowest energy crystal structures were re-optimised using the program DMACRYS[50] using the same exp-6 model, but with electrostatics described using atomic multipoles, derived from a distributed multipole analysis[51] of the B3LYP/6-31G(d,p) electron density.

For flufenamic acid and flutamide, we followed the protocol for crystal structure prediction (CSP) of flexible molecules outlined in reference 27 and further tested in references 52 and 18. A set of starting molecular conformations were selected as minima on the conformational energy surfaces of each molecule, which were sampled by systematically varying selected torsion angles: 6 starting conformations were selected for flufenamic acid and 8 for flutamide. Trial crystal structures were generated in common space groups with Z`=1 using the same method as for theophylline. These structures were further optimized (unit cell, molecular positions, and conformations) using a molecular mechanics description of inter- and intramolecular forces. Final energies of the lowest energy structures were calculated as a combination of a DFT (B3LYP/6-31G(d,p)) calculation for the intramolecular contribution and the exp-6 + atomic multipoles description for the intermolecular interactions. The influence of polarisation on the inter- and intra-molecular contributions to the relative crystal energies was approximated by performing the molecular calculations in a continuum dielectric ($\varepsilon = 3$), as we have previously suggested for flexible molecule CSP[53].

Due to the greater flexibility of the cocaine molecule, the CSP method was adapted to include an automated conformer search: conformations were generated using the low-mode search method[54] and the all-atom optimized potentials for liquid simulations (OPLS-aa) force field[55]. The most stable resulting conformations were refined using constrained B3LYP/6-31G(d,p) geometry optimisations and 16 starting conformations were selected for crystal structure generation. Trial structures were generated in common Söhnke space groups with Z'=1 and the lowest energy structures were further optimised using CrystalOptimizer[56], using a quantum mechanical description of the intramolecular forces and an atom-atom (exp-6 + atomic multipole electrostatics) description of the intermolecular forces. The final energy model was the same as that used for flufenamic acid and flutamide. Full details of the computational methods used for all four molecules are provided in the supplementary information.

To remove physically unrealistic structures, only the resulting structures within 8 - 10 kJ/mol of the lowest-energy structure for each molecule were retained for further analysis. Figure S12 shows the relative energies of the predicted structures within 8 - 10 kJ/mol of the minimum for all four molecules. CIF coordinate files for all low energy predicted structures are given in ESI.

2d DFT Calculations

Geometry optimizations and chemical shift calculations were carried out using the DFT program CASTEP[57], which uses a planewave basis set, whose implicit translational symmetry is very well adapted to describing the wavefunctions of crystalline systems. The GIPAW method[58], used with ultrasoft Vanderbilt-type pseudo-potentials[59,60], provides an efficient method to calculate chemical shifts in crystalline solids[61].

The geometry optimizations of the single crystal X-ray reference structures were carried out using the generalized gradient approximation (GGA) functional PBE[62], a plane-wave maximum energy cutoff energy of 700 eV, and a Monkhorst-Pack grid of k-points[63] corresponding to a maximum spacing of 0.05 Å^{-1} in reciprocal space. During the geometry optimizations, the unit cell and all heavy atoms were fixed, and only the hydrogen positions were relaxed.

The chemical shieldings σ_{calc} were calculated using the same functional and parameters as those used for the geometry optimization. They were then converted into calculated chemical shifts δ_{calc} using the relation $\delta_{calc} = \sigma_{ref} - \sigma_{calc}$ with the value of σ_{ref} determined for each molecule by a linear regression between calculated and experimental shifts, imposing a slope of unity. For the experimental chemical shifts that were not assigned, shifts and shielding were simply compared in order. Rotational dynamics were taken under consideration for the chemical shielding calculation for the methyl group protons in cocaine, flutamide and theophylline. They were averaged over the three protons to obtain a single calculated shielding value for each methyl group. Considering the fact that there is no indication from the NMR spectra of any large-scale dynamical processes, no other dynamic effects are considered for the chemical shift calculations.

CASTEP output files are given for all the structures in ESI.

3. Results and discussion

3a Assignment of experimental NMR spectra

Several 1D and 2D NMR experiments are used to assign the ^{1}H and ^{13}C resonances for each compound. The 1D ^{1}H MAS and ^{13}C CPMAS NMR spectra yield the ^{1}H and ^{13}C chemical shifts, which can then be assigned with the help of a 2D refocused INADEQUATE[64] NMR experiment which provides the connectivities between directly connected carbons and a 2D refocused INEPT[46] NMR experiment to correlate the resonances of directly attached protons, thus providing the assignment for the corresponding proton resonances.

Figure 1 shows the ^{1}H MAS NMR spectrum and the ^{13}C CPMAS NMR spectrum of cocaine. The ^{1}H and ^{13}C chemical shifts can be assigned on the basis of an INEPT and an INADEQUATE experiment, for which portions of the spectra are shown in Fig. 2. The INADEQUATE NMR spectrum (Fig. 2a) makes it possible to assign the carbons, but reveals two different possible assignments, and the INEPT NMR spectrum (Fig. 2b) reveals which peaks correspond to protons that are directly bonded to a carbon. Based on the ^{1}H-^{13}C correlations obtained from the INEPT NMR spectrum it becomes clear that only one of the two assignments obtained from the

INADEQUATE NMR experiment is feasible. The ^{1}H and ^{13}C chemical shifts obtained based on this assignment are summarized in Table 1.
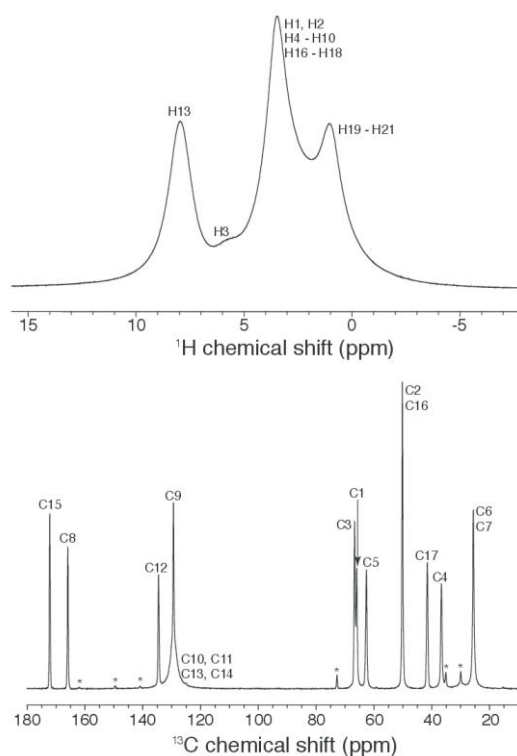
Fig. 1 (upper) ^{1}H (500 MHz) MAS NMR spectrum of cocaine: recorded in 32 scans with a recycle delay of 8 s at 60 kHz MAS and (lower) ^{13}C (125 MHz) CPMAS NMR spectrum. Both spectra were recorded with the bearing gas temperature regulated at 273 K.
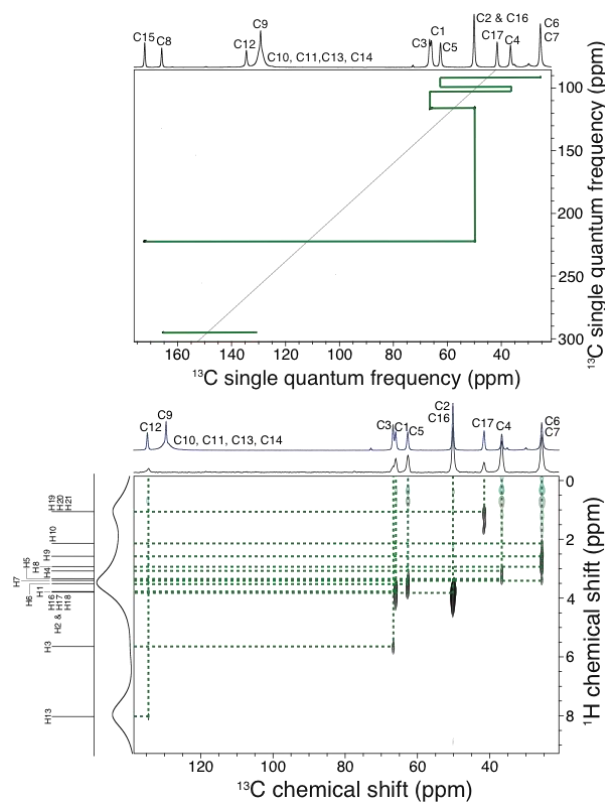
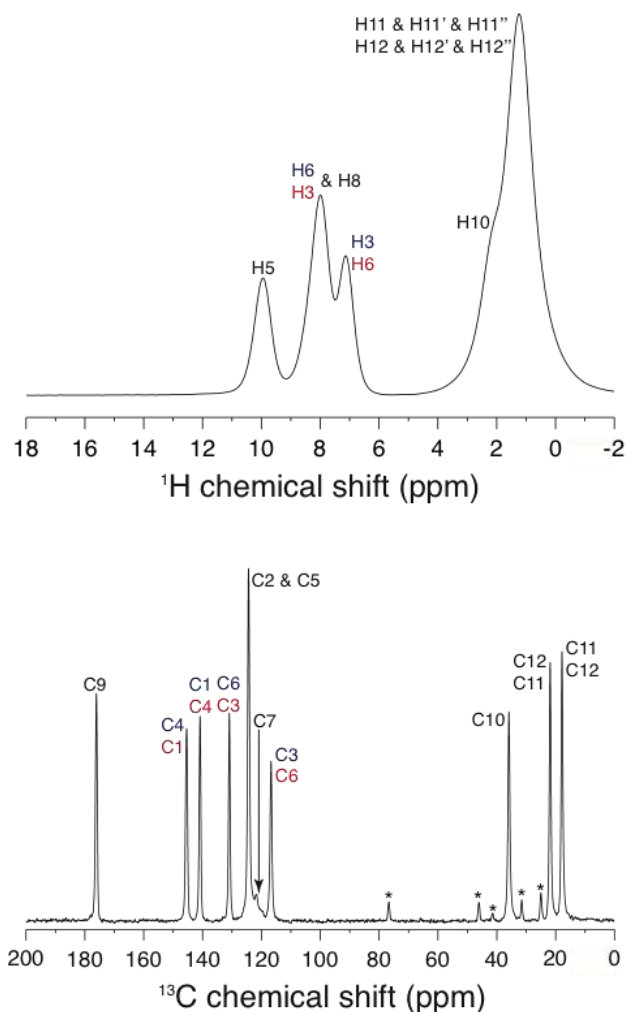**Fig. 2** (upper) The ^{13}C-^{13}C refocused INADEQUATE NMR spectrum and (lower) ^{1}H-^{13}C refocused INEPT NMR spectrum of cocaine. The blue trace above the INEPT direct dimension is the ^{13}C CPMAS NMR spectrum, while the stick plot to the left of the indirect dimension correspond to the ^{1}H signals, as obtained from the INEPT correlations.

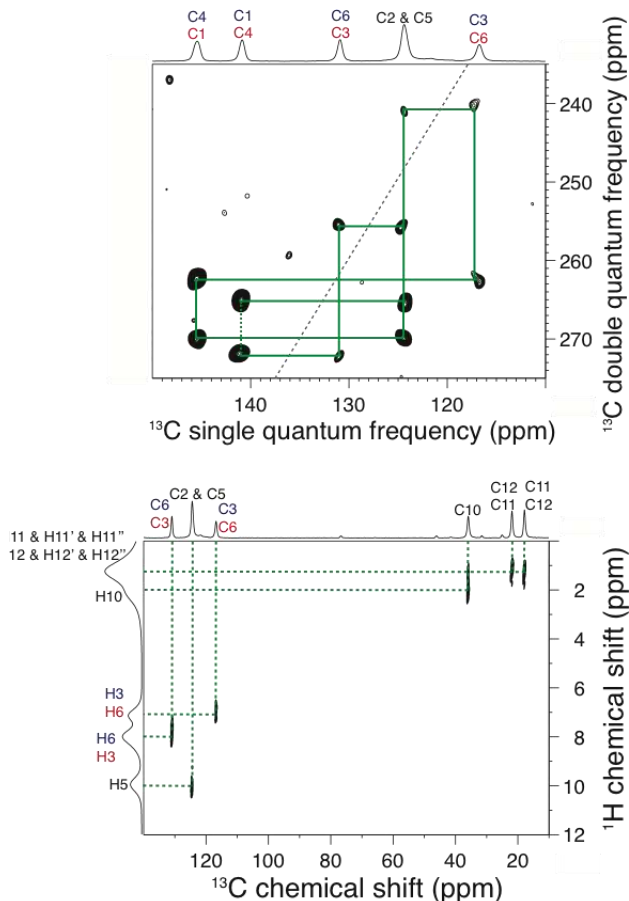**Table 1** Experimentally measured chemical shifts for cocaine. Labels are as given in Scheme 1-I.

Atom label	^{13}C Chemical shift (ppm)	^{1}H Chemical shift (ppm)
1	65.95	3.76
2	50.16	3.78
3	66.70	5.63
4	36.66	3.06
5	62.63	3.32
6	25.62	3.49
7	25.62	3.38
8	165.94	2.91
9	129.37	2.56
10	131.50	2.12
11	133.50	8.01
12	134.53	8.01
13	133.50	8.01
14	131.50	8.01
15	172.18	8.01
16	50.16	3.78
17	41.52	3.78
18	—	3.78
19	—	1.04
20	—	1.04
21	—	1.04

Fig. 3 (upper) ^{1}H (500 MHz) MAS NMR spectrum of flutamide: recorded in 8 scans with a recycle delay of 30 s at 60 kHz MAS and (lower) ^{13}C (125 MHz) CPMAS NMR spectrum. Both spectra were recorded with the bearing gas temperature regulated at 293 K. The two possible assignments for ^{1}H and ^{13}C are indicated in red and blue.

Figure 3 shows the ^{1}H MAS NMR spectrum and the ^{13}C CPMAS NMR spectrum of flutamide. The ^{1}H and ^{13}C chemical shifts can be partially assigned on the basis of an INEPT and an INADEQUATE experiment, for which portions of the spectra are shown in Fig. 4. In the aromatic region, the INEPT NMR spectrum reveals which peaks correspond to carbons that are directly bonded to a hydrogen, and the INADEQUATE NMR spectrum makes it possible to identify a circular chain of carbons. When these two pieces of information are combined, two assignments are possible for the aromatic carbons, and correspondingly for the protons bonded to an aromatic carbon. The assignment of the two methyl groups is also ambiguous. The possible assignments are summarized in Table 2. For the comparison with the calculated chemical shifts, both assignments were taken under consideration and the assignment with the lowest rmsd is shown in the rmsd plots (Section 3c).

Fig. 4 (upper) Extracts from the aromatic region of the ^{13}C-^{13}C refocused INADEQUATE NMR spectrum and (lower) ^{1}H-^{13}C refocused INEPT NMR spectrum of flutamide. The two possible assignments for ^{1}H and ^{13}C are indicated in red and blue.

Table 2 Experimentally measured chemical shifts for flutamide. The two assignments compatible with the experimental NMR data are shown in the left and right columns for each nucleus. Labels are as given in Scheme 1-II. The possible permutations of the assignments are discussed in the text.

Atom label	^{13}C Chemical shift (ppm)	^{1}H Chemical shift (ppm)
1	140.9 or 145.4	—
2	124.4	—
3	116.7 or 130.9	7.1 or 8.0
4	145.4 or 140.9	—
5	124.4	9.9
6	130.9 or 116.7	8.0 or 7.1
7	~122	—
8	—	8.0
9	176.1	—
10	35.9	2.0
11	17.8 or 21.8	1.2
12	21.8 or 17.8	1.2

were assigned by comparison with the assigned ^{13}C shifts measured in a solution of CDCl$_3$ (Figure S19). Using this traditional method, and the identification of quaternary carbons from the 2D ^{1}H-^{13}C HETCOR NMR spectrum (Figure 6), we could clearly assign all except the C4, C13 and C11 peaks which are sufficiently close in the two spectra as to be ambiguous. The ^{1}H chemical shifts were then assigned based on the ^{1}H-^{13}C connectivities obtained from the HETCOR NMR spectrum. The assignments of the experimental peaks obtained in this way for ^{1}H and ^{13}C are shown in Table 3. Note that, in practice, a long T_1 is indeed a key problem for assignments at natural abundance for compounds of unknown structure. Recently introduced impregnation DNP approaches may alleviate this issue in the future[65].
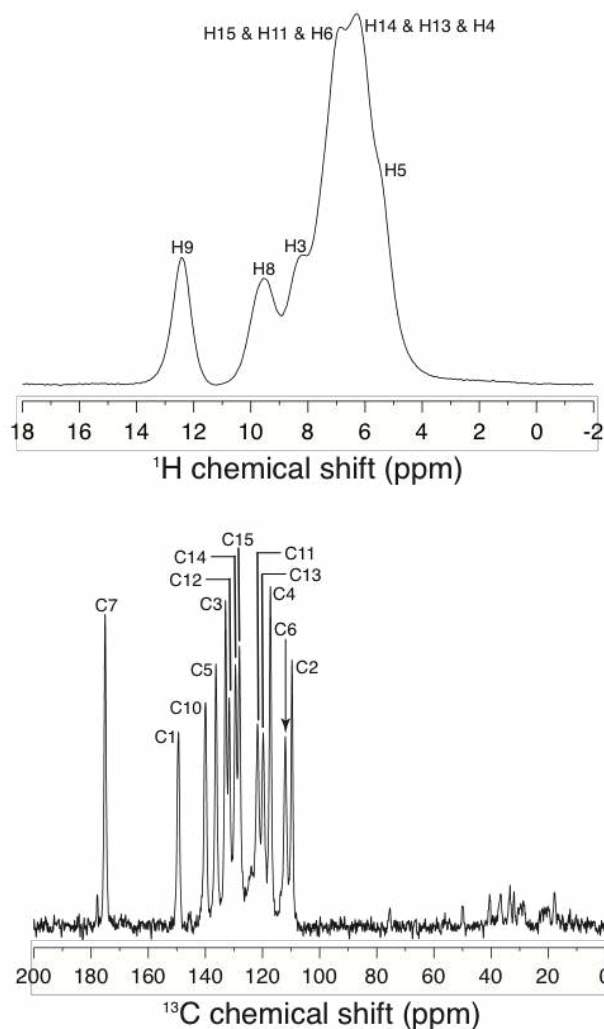
Fig. 5 (upper) ^{1}H (500 MHz) MAS NMR spectrum of flufenamic acid and (lower) ^{13}C (125 MHz) CPMAS NMR spectrum.
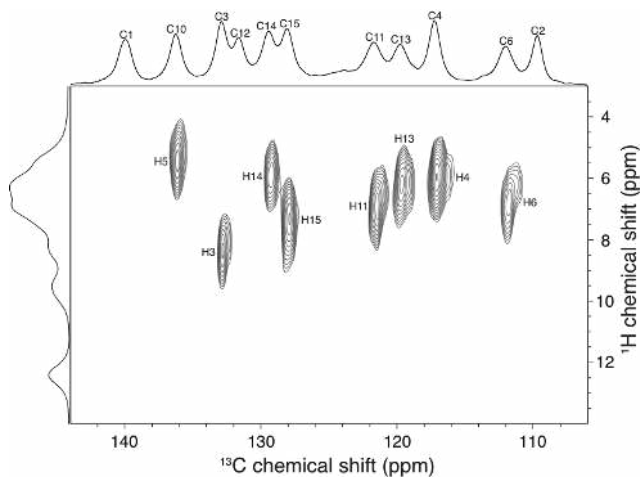
Figure 5 shows the ^{1}H MAS NMR spectrum and the ^{13}C CPMAS NMR spectrum of flufenamic acid. Due to a very long ^{1}H longitudinal relaxation time, it was not possible to record an INADEQUATE NMR spectrum to obtain the ^{13}C-^{13}C correlations. As a result, the carbon-13 chemical shifts

Fig. 6 Aromatic part of the ^{1}H-^{13}C HETCOR NMR spectrum of flufenamic acid. The trace above the direct dimension corresponds to the ^{13}C CPMAS NMR spectrum.

Table 3 Experimentally measured chemical shifts for flufenamic acid. Here the assignments are made based on the comparison between experimental and calculated chemical shifts using the flufenamic acid crystal structure obtained from the literature. Labels are as given in Scheme 1-III.

Atom label	^{13}C Chemical shift (ppm)	^{1}H Chemical shift (ppm)
1	149.3	—
2	109.7	—
3	133	8.3
4	117.2, 121.7 or 119.8	6.0, 6.9, or 6.2
5	136.3	5.4
6	112	6.8
7	175	—
8	—	9.6
9	—	12.4
10	139.9	—
11	121.7, 119.8 or 117.2,	6.9, 6.2 or 6.0
12	131.7	—
13	119.8, 117.2 or 121.7	6.2, 6.0 or 6.9
14	129.5	5.9
15	128.1	7.3
16	124.1	—

Figure 7 shows the ^{1}H MAS NMR spectrum and the ^{13}C CPMAS NMR spectrum for theophylline. The ^{1}H NMR spectrum of theophylline consists of only three resonance frequencies, the assignment of which is unambiguous. A tentative assignment of the ^{13}C NMR spectrum, based on solution-state NMR data, can be found in the literature[66]. However, no attempt was made to directly determine the assignment here, as this information was not helpful in the case of theophylline for the method studied here, as will be discussed below.
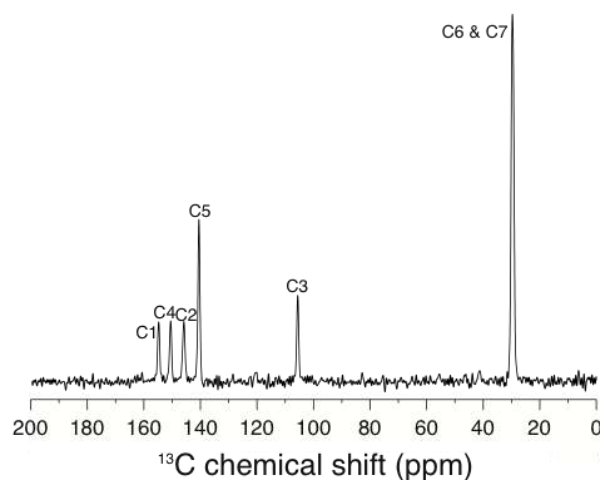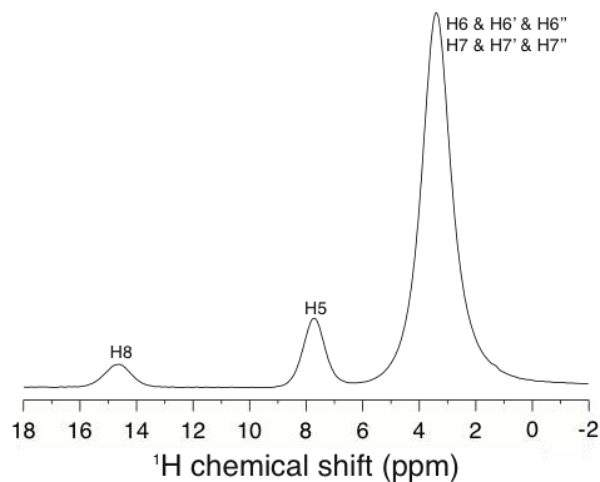
Fig. 7 (upper) ^{1}H (500 MHz) MAS NMR spectrum of theophylline and (lower) ^{13}C (125 MHz) CPMAS NMR spectrum.

**Table 4** Proposed chemical shift assignments for theophylline. Labels are as given in Scheme 1-IV.

| Atom label	^{13}C Chemical shift (ppm)	^{1}H Chemical shift (ppm)
1	150.8	—
2	146.1	—
3	105.8	—
4	155.0	—
5	140.8	7.7
6	29.9	3.4
7	29.9	3.4
8	—	14.6

3b Cocaine Structure Selection

Figure 8 shows the ^{1}H rms deviations between experimental and calculated chemical shifts for the set of the 30 lowest energy predicted crystal structures of cocaine (all structures are within 10 kJ mol^{-1} of the global lattice energy minimum). Importantly, as observed previously for thymol[38], the agreement between calculated and experimental chemical shifts is not correlated with the predicted energy (structures are ordered by ascending predicted energy). Thus,

experimental isotropic chemical shifts contain information complementary to that which is contained in the energy models used in the CSP protocol. Based on the agreement between calculated and experimental chemical shifts, we can determine structure 1 to be the correct crystal structure of the sample used in the present study. This is validated when comparing the calculated chemical shifts for the previously known reference single crystal X-ray structure with the experimentally recorded chemical shifts, as observed in Fig. 8. Note that to estimate the uncertainty in the calculated values, chemical shift calculations for 15 organic compounds (X-ray structures with CASTEP optimized hydrogen positions) were carried out, finding a mean rms error of 0.33 ppm (±0.16 ppm) from the experimental values for ^{1}H and 1.9 ppm (±0.4 ppm) for [13]C[38]. These mean errors are indicated as a dotted horizontal line in Figure 8 for ^{1}H and Figures S1-S2 for ^{13}C, for example, with the limits indicated by the grey zones in the figures. Differences smaller than these average values can thus be considered insignificant here. Furthermore, we should not be concerned that predicted structures lead to chemical shift rmsd values which lie just outside the expected range, as small structural deviations could lead to larger differences between chemical shifts calculated from predicted structures than from experimentally determined structures.
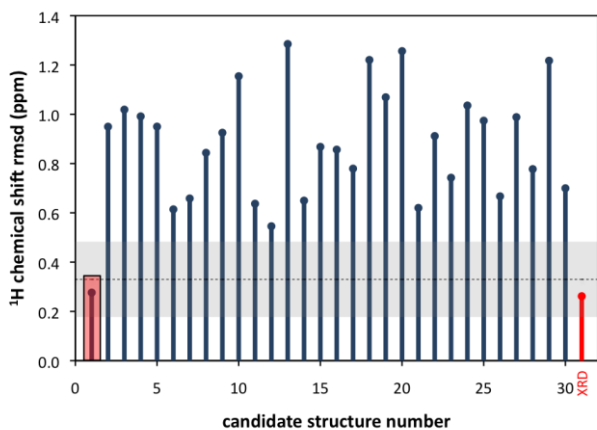
Fig. 8 Comparison between experimental and calculated ^{1}H chemical shifts for cocaine. The comparison is made using assigned experimental chemical shifts. Predicted structures are ordered by increasing calculated lattice energies (decreasing predicted stability). The comparison with the crystal structure determined from single crystal XRD is shown on the far right. The dotted horizontal black line shows the mean rmsd error as described in the main text and the horizontal grey shaded zone indicates the expected limits of the rmsd in chemical shift.
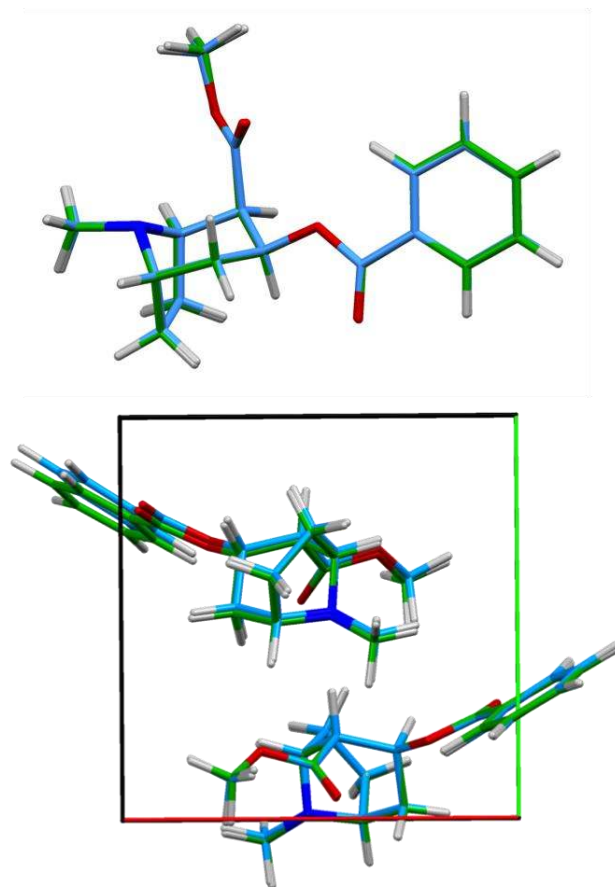

Fig. 9 Comparison between the structure of cocaine free base determined by powder ^{1}H NMR and computational modelling here and the single crystal XRD determined structure.

The all-atom rmsd between the molecular geometry of the structure determined here and the previously known structure of cocaine (CSD entry code: COCAIN10) is found to be 0.069 Å, and Figure 9 shows the two structures superimposed. The unit cell dimensions all agree to within 2.3% and the volume difference between the two structures is 0.8% (3.29 Å^3 per molecule).

3c Flutamide Structure Selection

Figure 10 shows the ^{1}H rmsds between experimental and calculated chemical shifts for the set of the 21 lowest energy predicted crystal structures of flutamide (all structures within 10 kJ mol^{-1} of the global lattice energy minimum).
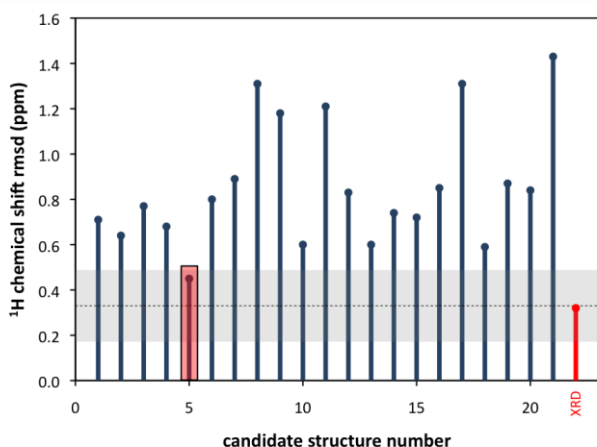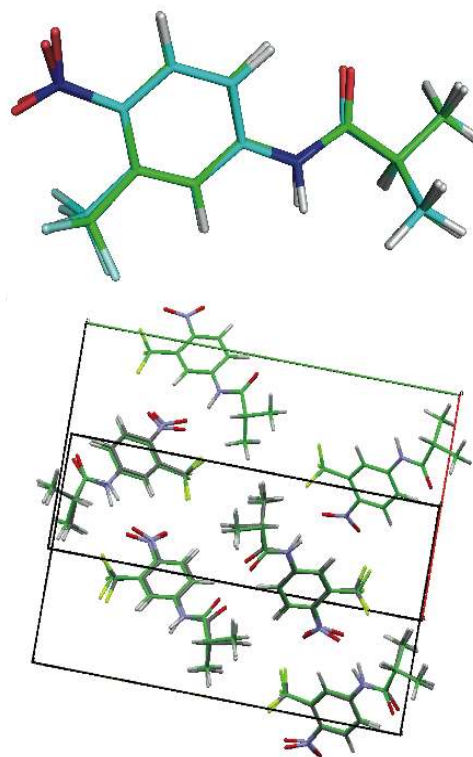
Fig. 10 Comparison between experimental and calculated ^{1}H chemical shifts for flutamide. The comparison is made using assigned experimental chemical shifts. Predicted structures are ordered by increasing calculated lattice energies (decreasing predicted stability). The display parameters are as for Figure 8. For the rmsd calculation, both experimental assignments were taken into account and the assignment that gave the smallest rmsds is presented in the Figure.

For flutamide, the ^{1}H rmsds were first calculated by assuming that the experimental chemical shifts were not assigned. For such a comparison, the experimental and calculated chemical shifts are simply ordered by increasing value. This case is presented in Fig. S3 (see ESI). The rms deviation does not provide a sufficient criterion to unambiguously select a computationally generated structure since many predicted structures result in similar rms deviations. When the assignment of the experimental ^{1}H chemical shifts is taken into account (Fig. 10), the rms deviations change significantly for some predicted structures, with structures that had low rms deviations of unassigned chemical shifts having much larger rmsds once assigned, and therefore are no longer compatible with the NMR data to within the estimated tolerances.

However, we note that even with the (ambiguous) assignment there are four predicted structures that are in good agreement with the unassigned experimental ^{13}C shifts (Figure S3). In contrast, we note that the relative variation in the proton shifts is much larger, and only one structure (ranked number 5 by lattice energy) is in good agreement. Structure 5 is thus selected here for the powder sample. The structure found here agrees with the structure of flutamide previously determined by single crystal XRD[40] as shown in Figure 11. There is a 0.097Å all-atom rmsd between the molecular geometries in the structure determined here and the reference structure. The unit cell lengths are all in agreement to within 1.55% or better, and the volume agrees to within 0.6% (1.71 Å^3 per molecule).

Note that this result contrasts with the case of thymol[38], for which unassigned ^{1}H chemical shifts were sufficient to determine the correct structure. For flutamide, the ^{13}C chemical shifts do not identify the correct structure and assigned ^{1}H chemical shifts are required to rule out several of the higher energy predicted structures (see Figure S3 for a rmsd plot using unassigned ^{1}H shifts). Here, the significant change in the rms deviation for assigned and unassigned ^{1}H

chemical shifts is due in particular to the large variability of the intermolecular contributions to the chemical shifts for the aromatic protons of flutamide. The peak-by-peak comparison of calculated shifts for the structure determined here and the measured values is shown in Figure S10.

Fig. 11 Comparison between the structure of flutamide determined by powder ^{1}H NMR and computational modelling here and the reference single crystal XRD determined structure.
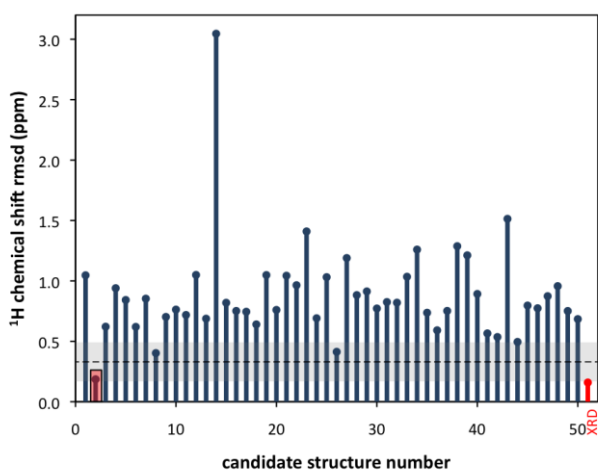
3d Flufenamic acid Structure Selection

Figure 12 shows the rmsds between the assigned experimental and calculated ^{1}H chemical shifts for the set of the 50 lowest energy predicted structures of flufenamic acid (spanning just over 8 kJ mol^{-1} from the global lattice energy minimum). Here, unassigned shifts are again insufficient to identify the correct crystal structure (Figure S5) and it is seen even more clearly than for flutamide that neither unassigned or assigned ^{13}C shifts discriminate strongly (Figures S5 and S6); in fact, all of the structures lead to calculated ^{13}C chemical shifts that are in poor agreement with the data. This is probably because of the relatively limited range in chemical shifts covered by the carbon-13 NMR spectrum.

In contrast, we again see that the assigned ^{1}H shifts provide a much stronger discrimination (Figure 12). In this case only four structures are in agreement to within our estimated tolerances, with structure 2 giving by far the best agreement.

Note that the ^{1}H rmsd comparison shown in Figure 12 excludes the OH and NH chemical shifts since these two shifts show considerable temperature dependence (Figure S20). Including these peaks for the rmsd calculation introduces errors due to the fact that the experiments are recorded at 293 K, while DFT calculations are temperature-

free (temperature only enters predicted crystal structures via the empirical parameterisation of the interatomic model potential). However, even when using all the ^{1}H chemical shifts for the comparison, predicted structure 2 remains in very good relative agreement with the observed ^{1}H chemical shifts (Fig. S7), albeit with an overall higher rmsd than in the case where the NH and OH shifts are not used. The peak-by-peak comparison of calculated shifts for the structure determined here and the measured values is shown in Figure S11.

The all-atom rmsd between the molecular geometry in structure number 2 determined here for the powder with the reference structure of flufenamic acid (CSD entry code: FPAMCA11) is found to be 0.117 Å, and Figure 13 shows the two structures superimposed. The unit cell dimensions all agree to within 5% and the volume difference between the two structures is 1.4% (4.58 Å^3 per molecule).
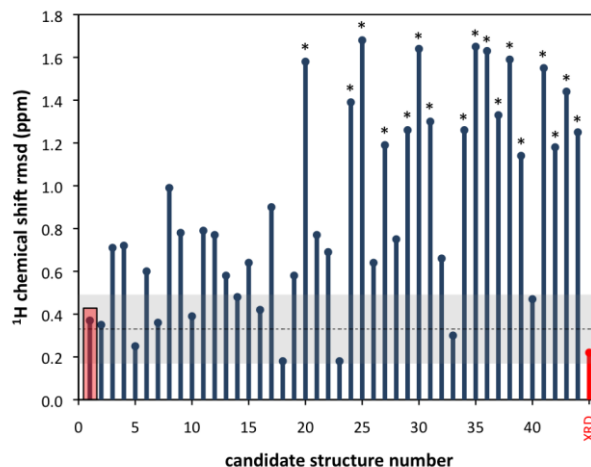

Fig. 12 Comparison between assigned experimental and calculated chemical shifts for flufenamic acid for [1]H (excluding the OH and NH chemical shifts). Predicted structures are ordered by decreasing stability. The display parameters are as for Figures 8 and 10.
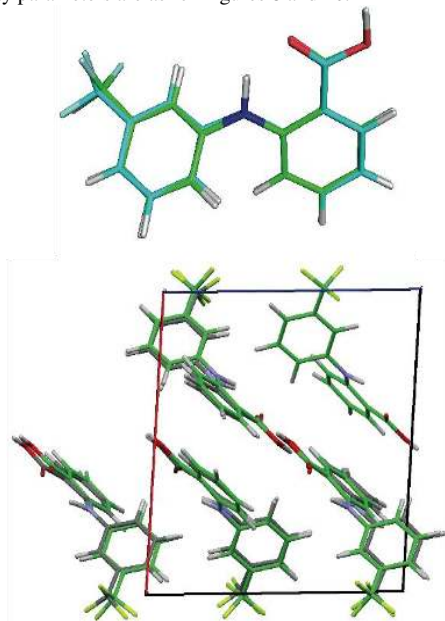
Fig. 13 Comparison between the reference *P*2$_1$/*c* crystal structure of flufenamic acid and the structure determined here.

3e Theophylline Structure Selection

Figure 14 shows the rmsds between the experimental and calculated ^{1}H chemical shifts for a set of the 44 lowest energy predicted structures of theophylline (all structures are within 10 kJ mol^{-1} of the global lattice energy minimum). In this case, the results remain unchanged when unassigned instead of assigned chemical shifts are used (see Figure S8 in the supplementary information).

Fig. 14 Comparison between ^{1}H experimental and calculated chemical shifts for theophylline. Predicted structures are ordered by decreasing stability. The display parameters are as for Figures 8, 10, and 12. Structures marked with * contain N-H…O hydrogen bonding.

For ^{13}C chemical shifts, again the agreement between calculated and experimental chemical shifts (regardless if they are assigned or unassigned) does not vary significantly from one predicted structure to another as can be seen in Figures S8 and S9. However, in this case, the same is true for the ^{1}H chemical shifts: although some structures lead to a large rms deviation from observed values, there are many structures that produce comparable values in agreement with the observed data. In this case, while structure 1 (the global minimum in lattice energy) corresponds to the known crystal structure of the theophylline polymorph under investigation, neither set of chemical shifts is sufficient to identify this structure from the list of predictions. This observation is not surprising, given the fact that the ^{1}H NMR spectrum of theophylline (Figure 7) consists of only three peaks, and illustrates one of the limits of the method.

Despite providing insufficient information to distinguish between many of the predicted structures of theophylline, we do observe that the ^{1}H chemical shifts correctly determine the hydrogen bonding topology in the theophylline polymorph studied here: all predicted structures with rms deviations of ^{1}H chemical shifts within our expected tolerances display hydrogen bonding of the NH to the nitrogen atom in the five-membered ring. In contrast, all predicted structures in which one of the oxygen atoms acts as the hydrogen bond acceptor lead to very high deviations in chemical shifts; these structures are labelled in Figure 14. The ability to distinguish between possible hydrogen bonding patterns is important for theophylline, whose polymorphs differ in which hydrogen bond acceptors are used[67,68].

4. Conclusion

We have further investigated a protocol for natural abundance powder NMR crystallography, based on combining theoretical crystal structure prediction with experimental solid-state NMR measurements and density functional theory calculations of NMR chemical shifts. The method has been applied to four powdered pharmaceutical compounds: cocaine, flutamide, theophylline and flufenamic acid.

The study reveals two features of the method. Firstly, we find that for both flutamide and flufenamic acid that the ^{1}H NMR spectra need to be at least partially assigned for the method to be robust, although the method has been proven to be successful without assignment for the previous case of thymol, and for cocaine. The partial requirement for assignment is not surprising since given chemical shifts can vary between one predicted crystal structure and another, leading to changes in the order of the peaks in the spectrum. This can make an unassigned comparison less sensitive, i.e. an incorrect structure can yield matching chemical shifts. Secondly, and this is possibly the most interesting feature of this study, we confirm that proton chemical shifts are more sensitive than carbon shifts to structural changes. For cocaine, flutamide, and flufenamic acid, we find that proton chemical shifts lead to a clear structure determination by comparison with the ensemble of structure predictions, whereas in all these cases the carbon-13 chemical shifts are not sufficiently sensitive to lead to structure determination. This observation suggests that proton NMR will play an increasingly important role in NMR crystallography.

We believe that the method is robust, but currently some precautions do need to be taken. For example, spectra should be recorded as a function of temperature to determine if there are any peaks (often involved in H-bonding, for example) that vary significantly with temperature. Since the DFT chemical shift calculations do not currently take temperature into account, such peaks may need to be excluded from the analysis (as was the case here for the NH and OH protons of flufenamic acid). Alternatively, temperature should be included in the crystal structure prediction calculations, by optimising structures on the free energy surface in place of the lattice energy minimisation.

Finally, we note that the method did not succeed for theophylline, since the NMR data are too sparse to be strongly discriminant, though we could identify the correct H-bonding pattern. We do note that a method based on combining DFT calculations with powder XRD and ^{13}C and ^{15}N solid-state NMR has been previously reported to successfully identify the correct polymorph of theophylline.[69] In general it is likely that the combination of the NMR methods here with powder XRD will always improve the quality of the procedure (though the objective of the present work was to evaluate the performance of a method which uses the NMR chemical shifts in isolation).

We note in conclusion that the powders studied here were not subjected to any modification prior to the experiments, and that they were investigated at natural isotopic abundance. The method should be of widespread interest in many areas,

and particularly in pharmaceutical materials science. The accurate prediction of the cocaine crystal structure included here demonstrates recent advances that have been made in CSP methodologies for large, flexible molecules and, as developments continue, the structure prediction methodologies could lead to solving the crystal structure of organic molecules up to 1000 g/mol or larger. This would cover most pharmaceutically relevant systems, and could also open up the possibility of predicting more complex materials such as hybrid organic-inorganic materials.

Notes and references

1 J. Yates, S. Dobbins, C. Pickard, F. Mauri, P. Ghi and R. Harris, *Phys. Chem. Chem. Phys.*, 2005, **7**, 1402.

2 A. L. Webber, L. Emsley, R. M. Claramunt and S. P. Brown, *J. Phys. Chem. A*, 2010, **114**, 10435.

3 A. L. Webber, B. Elena, J. M. Griffin, J. R. Yates, T. N. Pham, F. Mauri, C. J. Pickard, A. M. Gil, R. Stein, A. Lesage, L. Emsley and S. P. Brown, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6970.

4 R. Harris, *Analyst*, 2006, **131**, 351.

5 A. Abraham, D. C. Apperley, T. Gelbrich, R. K. Harris and U. J. Griesser, *Can. J. Chem.*, 2011, **89**, 770.

6 R. K. Harris, P. Hodgkinson, V. Zorin, J.-N. Dumez, B. Elena-Herrmann, L. Emsley, E. Salager and R. S. Stein, *Magn. Reson. Chem.*, 2010, **48**, S103.

7 E. Salager, R. S. Stein, C. J. Pickard, B. Elena and L. Emsley, *Phys Chem Chem Phys*, 2009, **11**, 2610.

8 A. L. Webber, S. Masiero, S. Pieraccini, J. C. Burey, A. S. Tatton, D. Iuga, T. N. Pham, G. P. Spada and S. P. Brown, *J. Am. Chem. Soc.*, 2011, **133**, 19777.

9 R. K. Harris, S. A. Joyce, C. J. Pickard, S. Cadars and L. Emsley, *Phys. Chem. Chem. Phys.*, 2006, **8**, 137.

10 R. K. Harris, S. Cadars, L. Emsley, J. R. Yates, C. J. Pickard, R. K. R. Jetti and U. J. Griesser, *Phys. Chem. Chem. Phys.*, 2007, **9**, 360.

11 N. Mifsud, B. Elena, C. J. Pickard, A. Lesage and L. Emsley, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3418.

12 D. Sebastiani, G. Goward, I. Schnell and H. W. Spiess, *J. Mol. Struct. (THEOCHEM)*, 2003, **625**, 283.

13 D. H. Brouwer, K. P. Langendoen and Q. Ferrant, *Can. J. Chem.*, 2011, **89**, 737.

14 G. R. Goward, D. Sebastiani, I. Schnell, H. W. Spiess, H.-D. Kim and H. Ishida, *J. Am. Chem. Soc.*, 2003, **125**, 5792.

15 S. P. Brown, T. Schaller, U. P. Seelbach, F. Koziol, C. Ochsenfeld, F. G. Klarner and H. W. Spiess, *Angew. Chem.-Int. Edit.*, 2001, **40**, 717.

16 C. Ochsenfeld, S. P. Brown, I. Schnell, J. Gauss and H. W. Spiess, *J. Am. Chem. Soc.*, 2001, **123**, 2597.

17 E. M. Heider, J. K. Harper and D. M. Grant, *Phys. Chem. Chem. Phys.*, 2007, **9**, 6083.

18 A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, S. L. Price, P. T. A. Galek, G. M. Day and A. J. Cruz-Cabeza, *Int. J. Pharm.*, 2011, **418**, 168.

19 J. T. A. Jones, T. Hasell, X. Wu, J. Bacsa, K. E. Jelfs, M. Schmidtmann, S. Y. Chong, D. J. Adams, A. Trewin, F. Schiffman, F. Cora, B. Slater, A. Steiner, G. M. Day and A. I. Cooper, *Nature*, 2011, **474**, 367.

20 M. A. Neumann, F. J. J. Leusen and J. Kendrick, *Angew. Chem. Int. Ed.*, 2008, **47**, 2427.

21 W. T. M. Mooij, B. P. v. Eijck and J. Kroon, *J. Am. Chem. Soc.*, 2000, **122**, 3500.

22 D. E. Braun, M. Ardid-Candel, E. D'Oria, P. G. Karamertzanis, J.-B. Arlin, A. J. Florence, A. G. Jones and S. L. Price, *Cryst. Growth Des.*, 2011, **11**, 5659.

23 P. G. Karamertzanis, A. V. Kazantsev, N. Issa, G. W. A. Welch, C. S. Adjiman, C. C. Pantelides and S. L. Price, *J. Chem. Theory Comput.*, 2009, **5**, 1432.

24 C. H. Görbitz, B. Dalhus and G. M. Day, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8466.

25 A. J. Cruz-Cabeza, G. M. Day and W. Jones, *Chem. Eur. J.*, 2008, **14**, 8830.

26 B. P. v. Eijck and J. Kroon, *Acta Crystallographica B*, 2000, **56**, 535.

27 G. M. Day, W. D. S. Motherwell and W. Jones, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1693.

28 G. M. Day, J. van de Streek, A. Bonnet, J. C. Burley, W. Jones and W. D. S. Motherwell, *Cryst. Growth Des.*, 2006, **6**, 2301.

29 A. J. C. Cabeza, G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2006, **6**, 1858.

30 G. M. Day and T. G. Cooper, *Crystengcomm*, 2010, **12**, 2443.

31 C. Pickard and R. Needs, *Phys. Rev. Lett.*, 2006, **97**, 045504.

32 C. Mailhiot, L. H. Yang and A. K. McMahan, *Phys. Rev. B*, 1992, **46**, 14419.

33 I. D. H. Oswald, D. R. Allan, G. M. Day, W. D. S. Motherwell and S. Parsons, *Cryst. Growth Des.*, 2005, **5**, 1055.

34 O. D. Friedrichs, A. W. M. Dress, D. H. Huson, J. Klinowski and A. L. Mackay, *Nature*, 1999, **400**, 644.

35 A. L. Mackay and H. Terrones, *Nature*, 1991, **352**, 762.

36 B. Winkler, C. J. Pickard, V. Milman and G. Thimm, *Chem. Phys. Lett.*, 2001, **337**, 36.

37 D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. D. Valle, G. R. Desiraju, B. P. v. Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. v. d. Streek, T. S.

Thakur, S. Tiwari, E. Venuti and I. K. Zhitkov, *Acta Cryst. B*, 2011, **67**, 535.

38 E. Salager, G. M. Day, R. S. Stein, C. J. Pickard, B. Elena and L. Emsley, *J. Am. Chem. Soc.*, 2010, **132**, 2564.

39 R. J. Hrynchuk, R. J. Barton and B. E. Robertson, *Can. J. Chem.*, 1983, **5**, 481.

40 J. M. Cense, V. Agafonov, R. Ceolin, P. Ladure and N. Rodier, *Struct. Chem.*, 1994, **5**, 79.

41 H. Krishna Murthy, T. Bhat and M. Vijayan, *Acta Cryst.*, 1982, **B38**, 315.

42 Y. Ebisuzaki, P. Boyle and J. Smith, *Acta Cryst. C*, 1997, **53**, 777.

43 C. Morcombe and K. Zilm, *J. Magn. Reson.*, 2003, **162**, 479.

44 A. Lesage, M. Bardet and L. Emsley, *J. Am. Chem. Soc.*, 1999, **121**, 10987.

45 B. Fung, A. Khitrin and K. Ermolaev, *J. Magn. Reson.*, 2000, **142**, 97.

46 B. Elena, A. Lesage, S. Steuernagel, A. Bockmann and L. Emsley, *J. Am. Chem. Soc.*, 2005, **127**, 17296.

47 B. Elena, G. de Paepe and L. Emsley, *Chem. Phys. Lett.*, 2004, **398**, 532.

48 P. G. Karamertzanis and C. C. Pantelides, *J. Comput. Chem.*, 2004, **26**, 304.

49 D. S. Coombes, S. L. Price, D. J. Willock and M. Leslie, *J. Phys. Chem.*, 1996, 7352.

50 S. L. Price, M. Leslie, G. W. A. Welch, M. Habgood, L. S. Price, P. G. Karamertzanis and G. M. Day, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8478.

51 A. J. Stone and M. Alderton, *Mol. Phys.*, 1985, **56**, 1047.

52 G. M. Day and T. G. Cooper, *CrystEngComm*, 2010, **12**, 2443.

53 T. G. Cooper, K. E. Hejczyk, W. Jones and G. M. Day, *J. Chem. Theory Comput.*, 2008, **4**, 1795.

54 I. Kolossváry and W. C. Guida, *J. Am. Chem. Soc.*, 1996, **118**, 5011.

55 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, **118**, 11225.

56 A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman and C. C. Pantelides, *J. Chem. Theory Comput.*, 2011, **7**, 1998.

57 S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson and M. C. Payne, *Z. Kristallogr.*, 2005, **220**, 567.

58 C. Pickard and F. Mauri, *Phys. Rev. B*, 2001, **63**, 245101.

59 K. Laasonen, R. Car, C. Lee and D. Vanderbilt, *Phys. Rev. B: Condens. Matter*, 1991, **43**, 6796.

60 D. Vanderbilt, *Phys. Rev. B: Condens. Matter*, 1990, **41**, 7892.

61 J. R. Yates, C. J. Pickard and F. Mauri, *Phys. Rev. B*, 2007, **76**, 024401.

62 J. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

63 H. Monkhorst and J. Pack, *Phys. Rev. B: Condens. Matter*, 1976, **13**, 5188.

64 A. Lesage, P. Charmont, S. Steuernagel and L. Emsley, *J. Am. Chem. Soc.*, 2000, **122**, 9739.

65 A. J. Rossini, A. Zagdoun, F. Hegner, M. Schwarzwälder, D. Gajan, C. Copéret, A. Lesage and L. Emsley, *J. Am. Chem. Soc.*, 2012, **134**, 16899.

66 C. Nicolau and K. Hildenbrand, *Z Naturforsch C*, 1974, **29**, 475.

67 K. Fucke, G. J. McIntyre, C. Wilkinson, M. Henry, J. A. K. Howard and J. W. Steed, *Cryst. Growth Des.*, **12**, 1395.

68 D. Khamar, I. J. Bradshaw, G. A. Hutcheon and L. Seton, *Cryst. Growth Des.*, **12**, 109.

69 E. Smith, R. Hammond, M. Jones, K. Roberts, J. Mitchell, S. Price, R. Harris, D. C. Apperley, J. Cherryman and R. Docherty, *J. Phys. Chem. B*, 2001, **105**, 5818.