

De Novo Crystal Structure Determination from Machine Learned Chemical Shifts

Martins Balodis, Manuel Cordova, Albert Hofstetter, Graeme M. Day, and Lyndon Emsley*



Cite This: *J. Am. Chem. Soc.* 2022, 144, 7215–7223



Read Online

ACCESS |



Metrics & More

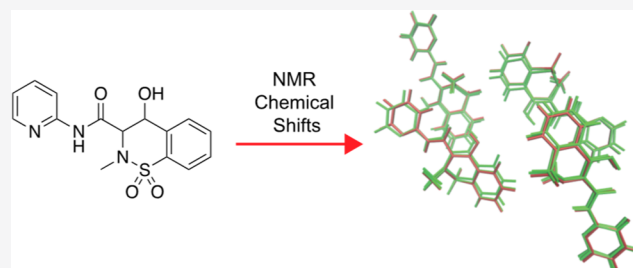


Article Recommendations



Supporting Information

ABSTRACT: Determination of the three-dimensional atomic-level structure of powdered solids is one of the key goals in current chemistry. Solid-state NMR chemical shifts can be used to solve this problem, but they are limited by the high computational cost associated with crystal structure prediction methods and density functional theory chemical shift calculations. Here, we successfully determine the crystal structures of ampicillin, piroxicam, cocaine, and two polymorphs of the drug molecule AZD8329 using on-the-fly generated machine-learned isotropic chemical shifts to directly guide a Monte Carlo-based structure determination process starting from a random gas-phase conformation.



INTRODUCTION

Determination of the atomic-level three-dimensional structure of organic solids is a key step in many areas of chemistry. Many compounds in their final forms are powdered solids, which make structure determination particularly challenging. In the case of powders, one can no longer depend on single-crystal X-ray diffraction, which is the gold standard in the structure determination of periodic solids, and other techniques must be used. These techniques include a combination of powder X-ray diffraction, nuclear magnetic resonance (NMR) spectroscopy, and computational methods.¹ In this respect, methods centered on the use of chemical shifts to determine the structure (often referred to as NMR crystallography) have emerged as being particularly powerful.^{2–8} Since the first *de novo* chemical shift-based structure of a molecular solid solved in 2013,⁹ the technique has been developed and applied to a range of structures from pharmaceuticals⁷ to capping groups on nanoparticle surfaces¹⁰ to the spacer layers in two-dimensional hybrid perovskite materials.¹¹ Striking recent examples include the determination of the structure of a drug molecule in a pharmaceutical formulation,¹² the detailed determination of the structure of active sites in enzyme reaction pathways,¹³ and the precise determination of the disordered structure of an amorphous drug.¹⁴

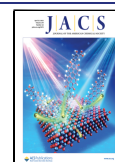
Established approaches to *de novo* structure determination, for example, by single-crystal X-ray diffraction of large molecules or by solution NMR, usually involve an iterative process where a (often random) starting structure is optimized under the combined effect of an (usually empirical) energetic potential and a penalty term that compares the computed observables with the measured values at every step of the optimization.¹⁵ This is a very powerful approach to find the correct structure and is enabled by the fact that the calculation

of observables from any trial structure is very rapid. So far, this has not been possible in chemical shift-based NMR crystallography, with a few notable exceptions where chemical shifts were incorporated and derived from parametrized forcefields.^{16,17} To make this approach general, the calculation of chemical shifts so far would have required highly accurate but very time-consuming electronic structure calculations.^{18–22} This results in *de novo* structure determination currently requiring first the generation of a large ensemble of credible candidate structures, usually done with some form of computational crystal structure prediction (CSP) protocol,^{23–27} followed by density functional theory (DFT) chemical shift calculations for the set of candidates, and only at the end of this process is there a comparison with the experimental shifts to determine which is the correct structure. While powerful, this is a time-consuming and laborious approach whose efficiency could be greatly improved by making use of chemical shift data at an earlier stage of the process. Additionally, if the set of candidates does not contain the correct structure, then the whole process fails.

Here, we show how by using a recently introduced machine learning model to predict chemical shifts, the structure of powdered organic solids can be determined in a manner fully analogous to the methods used in solution NMR or X-ray diffraction by integrating on-the-fly solid-state NMR shift

Received: December 31, 2021

Published: April 13, 2022



calculations into a Monte Carlo-simulated annealing optimization protocol. The approach does not require any structural hypothesis or knowledge of candidate structures (such as those from CSP). The approach is demonstrated to successfully determine five crystal structures for two different polymorphs of the drug molecule AZD8329 (1), ampicillin (2), piroxicam (3), and cocaine (4) (Figure 1).

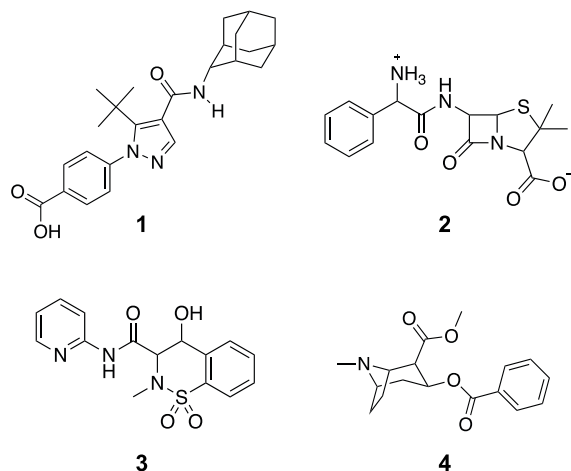


Figure 1. Molecular structures of AZD8329 (1), ampicillin (2), piroxicam (3), and cocaine (4).

Among these molecules, the structures of AZD8329 forms I and IV,⁹ ampicillin,²⁸ and cocaine³ have been previously found by NMR crystallography. AZD8329 form IV is notable because the structure was not found by X-ray diffraction methods prior to the original NMR crystallography study.⁹ Having a rich polymorphic landscape, it is also an interesting example to test the ability to distinguish between different polymorphs. Ampicillin is notable because CSP methods failed to predict the correct structure until NMR constraints were included to bias the starting conformers.²⁸ Cocaine is one of the first examples in which it was shown that NMR chemical shifts can reliably determine the correct structure among a set of candidate structures.³ The structure of piroxicam so far has not been determined by NMR crystallography, although comparison of calculated and measured chemical shifts was used to validate a structure proposed from powder X-ray diffraction.²⁹

EXPERIMENTAL METHODS

Crystal Structure Determination. Crystal structure generation and optimization were performed using a home-written Python script. The structure determination process follows the scheme shown in Figure 2 and is a version of constrained geometry optimization that is completely analogous to the methods currently used to determine, for example, protein structures from liquid- or solid-state NMR data, adapted to the case of molecular crystals. First, an initial conformation is generated with random torsional angles. The generated conformer is then placed in a randomly generated unit cell with a randomly chosen position and orientation. Details of the structure generation are given in the Supporting Information. After the initial generation of a random crystal structure, 4000 Monte Carlo steps are performed with a linear temperature profile between 2500 and 50 K. The structures are generated in a given space group, and the space group symmetry is conserved during the optimization. In each step, one of the parameters defining the crystal structure (cell length or angle, conformer position or orientation, or conformer dihedral angle) is randomly selected and updated within a given maximum step size. If the change leads to better agreement (as determined by the pseudo-

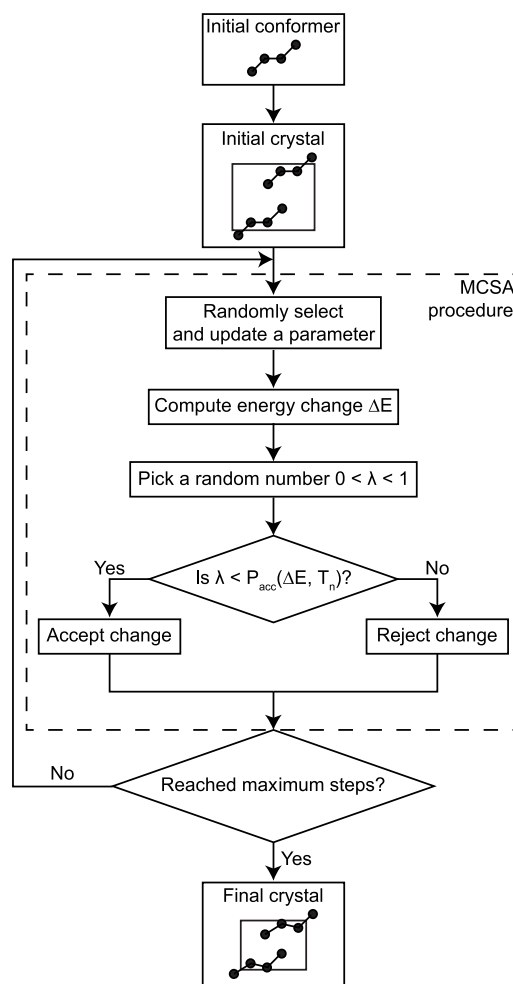


Figure 2. Scheme for crystal structure determination used in this study where $P_{acc} = \exp(-\Delta E/RT)$.

energies discussed below), it is accepted. Otherwise, the step is accepted with a probability $P_{acc} = e^{-\Delta E/RT}$, where ΔE is the change of pseudo-energy induced by the step, R is the gas constant, and T is the temperature. The step size of the updated parameter is doubled if the step is accepted, and halved otherwise (see the Supporting Information for detailed parameters including the step sizes). Every 500 steps, the hydrogen positions were optimized using tight binding DFT (DFTB).

Energy calculations were performed at the semiempirical DFTB3-D3H5 level of theory using the 3ob-3-1 parameter set and the DFTB+ software version 20.1.^{30–33}

The chemical shieldings were predicted using ShiftML version 1.2 (publicly available at <https://shiftml.epfl.ch>).³⁶ Shieldings were converted to chemical shifts via the relation

$$\delta = a + b\sigma \quad (1)$$

where δ is the chemical shift, a and b are the experimentally determined calibration constants (see the Supporting Information for details), and σ is the calculated chemical shielding. Here, we set a to 30.36 and b to -1 . To account for ambiguity when comparing chemical shifts of protons for CH_2 groups, the shifts were compared using the best matching criteria. Shifts which are hard or impossible to distinguish experimentally such as aromatic protons or CH_3 groups were averaged when making the comparison.

Crystal Structure Comparison. The optimized crystal structures were compared using the COMPACT algorithm,³⁷ included in the commercial Cambridge Structural Database (CSD) package,³⁸ which compares interatomic distances and angles within a cluster of molecules taken from the reference and comparison crystal structures.

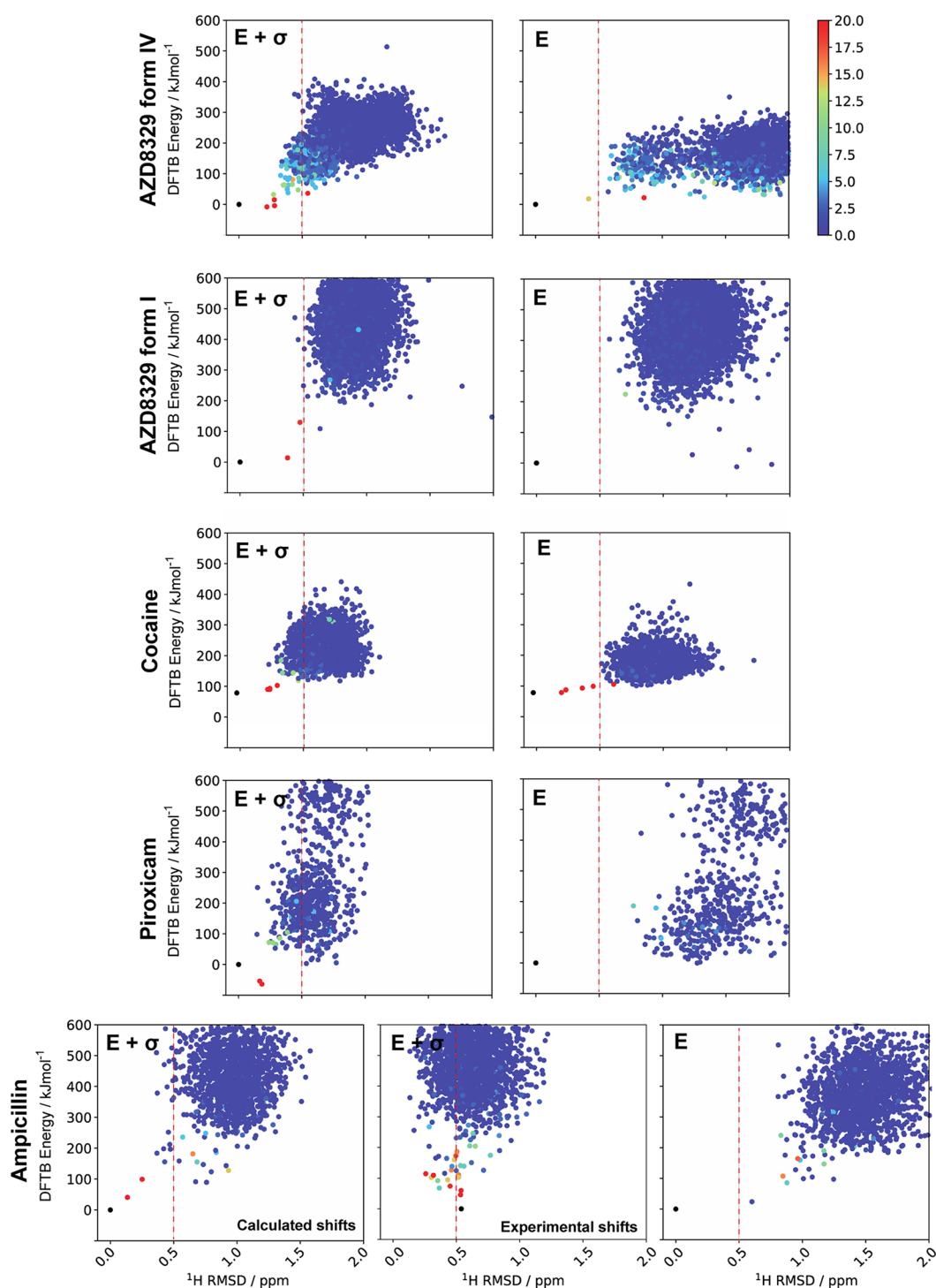


Figure 3. Plots of DFTB energy versus ^1H chemical shift rmsd for the results of 10 000 simulated annealing runs on AZD8329 form IV, 10 000 runs on AZD8329 form I, 2500 runs for ampicillin, 1000 runs for piroxicam, and 2500 runs on cocaine. The left column shows the optimizations done using both chemical shift and energy, while the right column shows the optimizations done using only energy. For ampicillin, the results are shown for both where ^1H shifts calculated from the known reference structure were used and where the experimental ^1H shifts were used as targets for the optimization. Each point represents a structure optimized as described in the methods section. The vertical axis shows DFTB energies and the horizontal axis ^1H shift rmsd values with respect to the shifts calculated for the known experimental structure which is set to 0 and is colored black. The color of each point reflects the similarity between each of the calculated structures and the reference structure, according to the scale on the right and as described in the methods section. The red vertical dashed line shows the cutoff value of 0.5 ppm for the ^1H rmsd. For piroxicam, unconstrained optimization of the experimental structure leads to a large deviation in the structure, so the reference energy is the energy of the experimental structure with only hydrogen atom positions optimized.

A cluster of 20 molecules were used for comparison in this work. Before the comparison, physically unrealistic structures were removed, for example, structures where neighboring molecules are too close in

space or where the density is unrealistically low. Most of the physically unrealistic structures are easily spotted due to their high energy or shift root-mean-square deviation (rmsd). The known reference

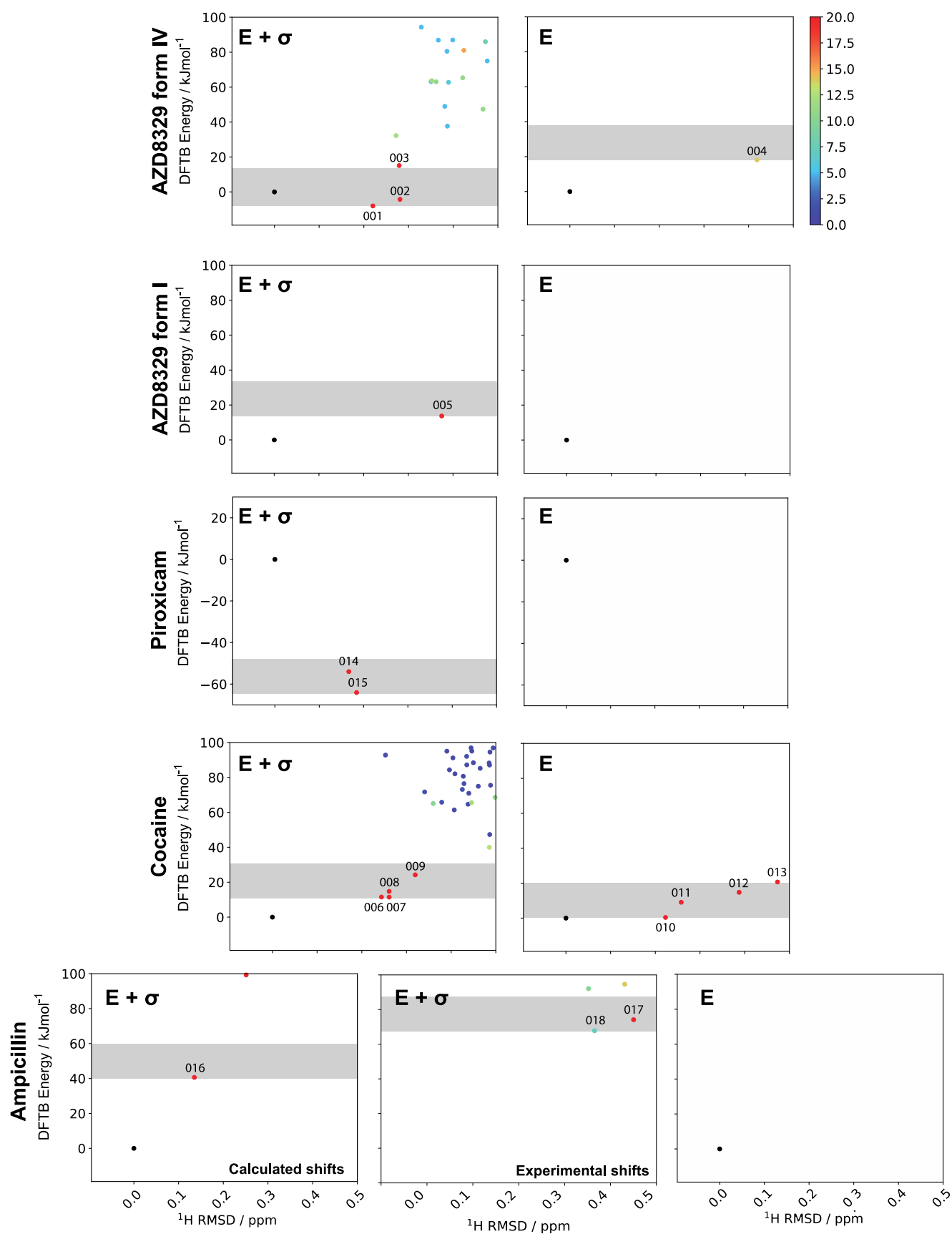


Figure 4. Plots of DFTB energy versus ^1H chemical shift rmsd, as shown in Figure 3, expanded to include a range of 100 kJ/mol and up to 0.5 ppm ^1H rmsd. The gray areas represent the area within 20 kJ/mol of the lowest energy structure found in the optimization. Labels refer to the structures as defined in Table S1. For piroxicam, unconstrained optimization of the experimental structure leads to a large deviation in the structure, so the reference energy is the energy of the experimental structure with only hydrogen atom positions optimized.

structures used are given in the [Supporting Information](#), together with the CSD codes where available.

RESULTS AND DISCUSSION

The optimization scheme introduced here is summarized in [Figure 2](#).

In the first step, a viable conformation of the single molecule is generated, and bond angles and lengths are optimized using, here, DFTB3-D3H5 which provides a good compromise between accuracy and computational cost (on the same timescale as ShiftML chemical shift calculations) (see the [Supporting Information](#) for details). Then, for each run, a random conformation is generated by randomizing the flexible torsion angles, and a starting crystal structure is generated by randomly selecting cell parameters in a given space group (cell lengths, cell angles, and position and orientation of the molecule). Between 1000 and 10 000 trial structures were generated for each system. Each structure was then optimized by a Monte Carlo-simulated annealing process described in the [Methods](#) section, where in each step, one of the parameters defining the crystal structure (i.e., a single torsion angle or cell parameter) was randomly changed, and chemical shifts and the DFTB system energy were calculated following the change.

Here, to enable the possibility to calculate shifts at each step, the ShiftML prediction algorithm was used.³⁶ ShiftML is a fast and accurate method to compute chemical shifts in a matter of seconds even for the largest of molecular crystals. It was recently developed using DFT optimized structures derived from CSD as a training set for a machine learning framework. The current version can predict chemical shifts for molecules containing H, C, N, O, or S atoms.

The cost function used in the Monte Carlo process is

$$E_{\text{tot}} = E_{\text{DFTB}} + cE_{\text{cs}} \quad (2)$$

where

$$E_{\text{cs}} = \sqrt{\frac{\sum_{i=1}^n (\delta_{i,\text{trg}} - \delta_{i,\text{shiftML}})^2}{n}} \quad (3)$$

where $\delta_{i,\text{trg}}$ is the target chemical shift of the i th nucleus in the molecule containing n nuclei and $\delta_{i,\text{shiftML}}$ is the corresponding shift computed using the ShiftML model. c is an empirically adjusted constant (in kJ/mol) that weights the relative contribution of the internal energy and the agreement with the experiment in the cost function. (Note that the values of E_{cs} are independent of the size of the molecule but will change from one type of nucleus to another, and E_{DFTB} will depend on the size of the molecule. In the examples here, satisfactory results were found with values of c such that $\Delta E_{\text{DFTB}} \sim \Delta E_{\text{cs}}$, where ΔE is the difference observed between two Monte Carlo steps at the end of the optimization process.) In the following, for the proof of principle demonstration here, we use shifts calculated with ShiftML from the known structure as the $\delta_{i,\text{trg}}$ target set in E_{cs} . This reduces any bias due to experimental variability between compounds in the comparisons below and makes the process fully self-consistent. We note that the estimated errors on ShiftML shifts are in any case similar to or larger than the error ranges in the experimental shifts.

The other parameters in the simulated annealing process are given in the [methods](#) section and [Supporting Information](#).

Optimization Using Computed Target Shifts. [Figure 3](#) shows the results for AZD8329 form I, AZD8329 form IV, ampicillin, piroxicam, and cocaine. In order to demonstrate

that the chemical shifts are indeed the driving force for structure determination, for each case, optimization was performed with the penalty function that includes both the DFTB energy and chemical shift differences and, for comparison, using only the DFTB energy. [Figure 4](#) shows expansions of the regions below 100 kJ/mol and 0.5 ppm.

We expect correct structures to occur in the region of low chemical shift rmsd and low calculated energy. For ^1H shift rmsd, we use a cutoff of 0.5 ppm, taken from Engel et al. where they determined the expected error of the ShiftML model for ^1H to be 0.48 ppm.³⁹ Nyman and Day showed that with accurate calculations, most polymorphs are separated by less than 7.2 kJ/mol,⁴⁰ which can be treated as the most relevant energy range on CSP landscapes. In this study we use DFTB, whose energies are less accurate and have been shown to place observed crystal structures over a much wider energy range in CSP studies.⁴¹ To account for this larger spread, we use a cutoff for the accepted structures of up to 20 kJ/mol from the lowest energy structure. Indeed, the spread of predicted energies decreases significantly when the structures that are within 20 kJ/mol and 0.5 ppm rmsd are further optimized using DFT, as illustrated in [Figure S5](#) (and [Table S2](#)). Typically, after optimization, the predicted DFT energy difference between the structures is less than ~ 2 kJ/mol.

For all compounds, we note that the majority of Monte Carlo runs do not yield any results with either low DFTB energy or with a low chemical shift rmsd to experiment. Indeed, if we define a region of acceptable structures to have simultaneously a DFTB energy within 20 kJ/mol of the lowest energy structure in the Monte Carlo set and a chemical shift rmsd to experiment below 0.5 ppm, then the pure Monte Carlo approach using only DFTB energy as the driving force does not find any structures that match the rmsd_{20} criteria for either form of AZD8329. This is completely in line with expectations since this simple semiempirical type approach is not expected to easily find crystalline polymorphs.

Including chemical shifts in the penalty function yields three structures for form IV (001-003) within the acceptable ranges and one structure for form I (005).

These structures for both forms are shown in [Figure 5](#), superimposed on the known structures, and we see that they are in excellent agreement with the correct structures as previously determined by X-ray diffraction or NMR.

Ampicillin is another interesting example as noted in the introduction because it is a case where current CSP methods fail since the conformer present in the crystal structure has a relatively high energy in the gas phase.²⁸ As a result, chemical shift-driven structure determination based on prior generation of candidates fails. In contrast, Monte Carlo runs for ampicillin including DFTB energy and chemical shifts produced two structures that perfectly match with the known crystal structure, with one of them (016) being selected by our criteria. The structure determined by our criteria is superimposed on the known crystal structure in [Figure 6](#). Runs using only DFTB energy did not produce any matching structures either in the acceptable region or outside it.

Similar to ampicillin, runs for piroxicam produced structures (014 and 015) matching with the known crystal structure, both of which are in the acceptable region. Again, no matching structures were found for the runs using only energy in the penalty function. Overlay of the structures determined here with the known crystal structure is shown in [Figure 5](#). From [Figure 5](#), it is seen that both of the structures found are

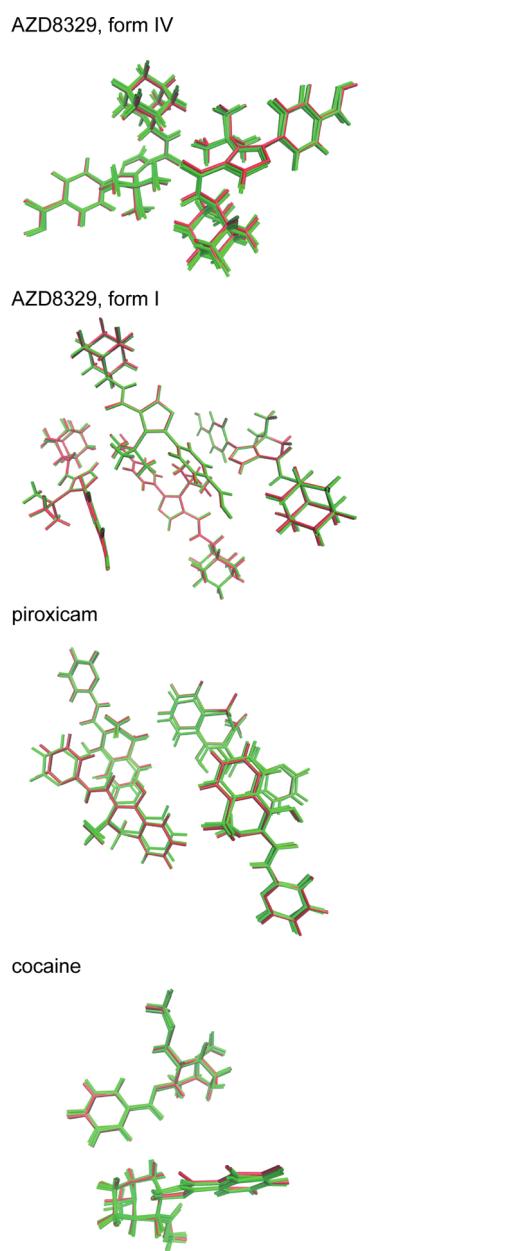


Figure 5. Overlay of the asymmetric unit for the structures determined here for AZD8329 form IV, AZD8329 form I, piroxicam, and cocaine. For AZD8329 form IV, there are three structures (Figure 4), one for form I, 2 for piroxicam, and 4 for cocaine. The red structures are the known structures, and the green structures are the structures determined here that are less than 20 kJ/mol from the lowest energy determined structure and 0.5 ppm ^1H rmsd compared to the target shifts.

significantly lower in DFTB energy than the known structure. We note that to compare our determined structures and the known reference structures, we systematically relaxed the atom positions and the cell parameters for the experimental reference structures using DFTB. While the result of the relaxation was fairly similar to the starting structures for most of the reference structures, this was not the case for the structure of piroxicam. Full DFTB relaxation of piroxicam changed the structure to a point where its space group changed. To avoid this, we relaxed only ^1H positions with DFTB, and we suspect that this is why the energy of the

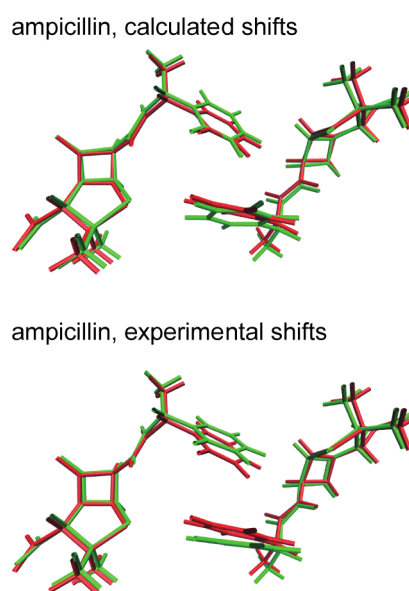


Figure 6. Overlay of the asymmetric unit for the structures determined here for ampicillin with calculated (top, structure 016) or with experimental (bottom, structure 017) chemical shifts. The red structures are the known structures, and the green structures are the structures determined here.

reference structure appears higher than expected. When both the determined structures and the known structure were optimized with DFT, the (DFT) energy difference between them was reduced to 0.4 kJ/mol for the best matching structure.

Cocaine is an interesting example since it is significantly less flexible than AZD8329. In this case, the Monte Carlo approach with energy alone does already produce four structures in the acceptable region (010-013). Adding chemical shifts did not improve the result, and the same number of structures were found in the acceptable region (006-009). The four structures optimized using shifts are shown in Figure 5 superimposed on the known structure for cocaine, and we again see that they are in good agreement with the correct structure. We explain this as cocaine having a relatively simple energy landscape with few competing structures: the results of the Monte Carlo search using only energy direct the search efficiently toward the known crystal structure of the only known polymorph, suggesting that there are few competing, “false” structures. It is in cases where there are many energetically competing structures, which is the norm, that adding the chemical shift to the fitness function is expected to increase the effectiveness of the search at locating the correct structure. The other compounds studied here, on the other hand, have much richer energy landscapes with at least four anhydrous polymorphs known for AZD8329 for example,⁹ and by using the chemical shifts of two different forms as targets, we were able to successfully determine both structures here. Figures 5 and 6 show the overlay of the asymmetric unit of the crystal structures determined here for each compound (green) with the known reference structures (red).

When comparing against the known reference crystal structures all atom rmsd₂₀ values are given in Table 1. The highest rmsd₂₀ value is 0.51 Å for ampicillin, meaning that all of the optimized structures correspond very well to the experimental reference crystal structure. In comparison, in the current latest CSP blind test (sixth) the highest rmsd₂₀ value

Table 1. Reduced Unit Cell Parameters and Atom rmsd_{20} Values for the Determined Structures Using Chemical Shifts and DFTB without Subsequent DFT Relaxation^a

name	$a/\text{\AA}$	$b/\text{\AA}$	$c/\text{\AA}$	$\alpha/^\circ$	$\beta/^\circ$	$\gamma/^\circ$	$\text{rmsd}_{20}/\text{\AA}$
AZ8329, form IV (3)	9.5 ± 0.1 (9.9)	11.0 ± 0.1 (10.8)	11.8 ± 0.3 (11.6)	65.3 ± 1.7 (65.7)	75.9 ± 2.2 (75.0)	75.5 ± 3.4 (74.0)	0.44 ± 0.15
AZ8329, form I (1)	11.3 (11.4)	13.2 (13.1)	15.1 (15.0)	114.2 (113.0)	90 (90)	90 (90)	0.14
piroxicam (2)	6.9 ± 0.1 (6.8)	13.3 ± 0.2 (13.9)	15.12 ± 0.1 (15.1)	90 (90)	90 (90)	93.2 ± 1.0 (97.3)	0.40 ± 0.13
cocaine (4)	8.1 ± 0.1 (8.1)	9.2 ± 0.1 (9.0)	10.1 ± 0.2 (10.0)	90 (90)	105.8 ± 1.0 (106.0)	90 (90)	0.28 ± 0.02
ampicillin calculated (1)	5.8 (5.8)	12.3 (11.4)	12.5 (12.3)	116.4 (113.6)	90 (90)	90 (90)	0.51
ampicillin experimental (1)	5.8 (5.8)	11.3 (11.4)	12.3 (12.3)	117.2 (113.6)	90 (90)	90 (90)	0.43

^aThe number in the brackets after the name of the compound is the number of structures found. Standard deviation is given where more than one structure is found. The number of brackets after the mean value of the cell parameters is the value for the known experimental structure.

was 0.81 Å, which, while considered high, was still considered acceptable.⁴² In the examples here, after the DFT optimization, the highest value decreased to 0.49 Å and the lowest to 0.05 Å. Table 1 also gives the distribution of the unit cell dimensions for the optimized structures which are very close to the experimental values. Individual rmsd_{20} values and the cell parameters for all best matching structures are given in Supporting Information, Table S1.

Optimization Using Experimental Target Shifts. As noted above, we use ¹H chemical shifts calculated for the known crystal structures as the target for optimization here. This allows us to explore the method without any biases introduced by any possible errors in chemical assignments and to make the analysis self-consistent. Of course, it is most important that the method also works using experimental shifts. This is demonstrated in Figures 3 and 4 where we also show the results of optimization against experimental ¹H shifts for ampicillin. The experimental shifts were taken from Hofstetter et al.²⁸ In this case, two structures (017 and 018) matched the selection criteria. One structure (017) yielded a very good rmsd_{20} of 0.41 Å with respect to the known structure, as illustrated in Figure 6. It is interesting to note that the other structure (018) at first glance matches less well, but on further examination, we see that the cell parameters match very well (see Table S1), and the main difference is a slight change in the orientation of the aromatic ring position. An overlay of the unit cell of the known structure and structure 018 is shown in Figure S4. After optimization with DFT, the relative (DFT) energy for the structures converged to -0.4 and 9.4 kJ/mol for (017) and (018), respectively, with respect to the known structure (see Table S2), and the ¹H rmsd to DFT calculated shifts was 0.13 and 0.41 ppm, suggesting that the optimized structure 017 is in better agreement with the experiment.

This is the first example of a molecular crystal structure determined directly from experimentally measured chemical shifts in contrast to earlier approaches where chemical shifts were used to select from a predetermined set of predicted crystal structures.

CONCLUSIONS

We have shown that crystal structures can be directly determined from chemical shifts, without any prior structural hypothesis and without any knowledge from candidate structures (such as from CSP), through the use of machine learned chemical shifts which enable on-the-fly calculation of shifts at each step of a simulated annealing structure determination protocol. We have illustrated this for the

structures of ampicillin, piroxicam, and cocaine, as well as for AZD8329 where the inclusion of machine learned chemical shifts allows the determination of the correct structures for two different polymorphic forms. We note that the AZD8329 case is a particularly important illustration since it clearly shows how the chemical shifts can drive the optimization toward two very different structures for the same molecule.

Here, we chose to use a Monte Carlo-simulated annealing algorithm due to its relatively straightforward nature, but in principle, machine learned chemical shifts can be incorporated into other optimization methods as they are easy to add as an additional pseudo-energy term, and we believe there is significant room for further development and increased efficiency of this approach to chemical shift-based structure determination in molecular solids. Finally, we note that the method presented here no longer relies on a purely energy-driven computational candidate crystal structure generation step. By driving the structure determination directly from chemical shifts, integrated through the entire optimization procedure, the method is applicable even in cases where CSP is extremely challenging, such as the example of ampicillin here.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.1c13733>.

Link to the Python codes (ZIP)

Experimental details, details and coordinates of the determined structures, plots of energy for the initially generated structures, energy-density plots for the optimized structures, and plots of energy during optimization (PDF)

AUTHOR INFORMATION

Corresponding Author

Lyndon Emsley – Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland; National Centre for Computational Design and Discovery of Novel Materials MARVEL, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland; orcid.org/0000-0003-1360-2572; Email: lyndon.emsley@epfl.ch

Authors

Martins Balodis – Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Manuel Cordova – Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland; National Centre for Computational Design and Discovery of Novel Materials MARVEL, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland; orcid.org/0000-0002-8722-6541

Albert Hofstetter – Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland; Present Address: Current address: Laboratory of Physical Chemistry, ETH, Zurich, CH-8093 Zurich, Switzerland

Graeme M. Day – School of Chemistry, University of Southampton, Highfield SO17 1BJ Southampton, United Kingdom; orcid.org/0000-0001-8396-2771

Complete contact information is available at:

<https://pubs.acs.org/10.1021/jacs.1c13733>

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support from the Swiss National Science Foundation grant no. 200020_178860 and the NCCR MARVEL is acknowledged.

REFERENCES

- (1) Reif, B.; Ashbrook, S. E.; Emsley, L.; Hong, M. Solid-state NMR spectroscopy. *Nat. Rev. Methods Primers* **2021**, *1*, 2.
- (2) 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*, 2564–2566.
- (3) Baias, M.; Widdifield, C. M.; Dumez, J.-N.; Thompson, H. P. G.; Cooper, T. G.; Salager, E.; Bassil, S.; Stein, R. S.; Lesage, A.; Day, G. M.; Emsley, L. Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state 1H NMR spectroscopy. *Phys. Chem. Chem. Phys.* **2013**, *15*, 8069–8080.
- (4) 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*, 3674–3683.
- (5) 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*, 4379–4390.
- (6) Selent, M.; Nyman, J.; Roukala, J.; Ilczyszyn, M.; Oilunkaniemi, R.; Bygrave, P. J.; Laitinen, R.; Jokisaari, J.; Day, G. M.; Lantto, P. Clathrate Structure Determination by Combining Crystal Structure Prediction with Computational and Experimental ¹²⁹Xe NMR Spectroscopy. *Chem.—Eur. J.* **2017**, *23*, 5258–5269.
- (7) 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*, 16650–16661.
- (8) 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*, 1476–1487.

(9) 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*, 17501–17507.

(10) Al-Johani, H.; Abou-Hamad, E.; Jedidi, A.; Widdifield, C. M.; Viger-Gravel, J.; Sangaru, S. S.; Gajan, D.; Anjum, D. H.; Ould-Chikh, S.; Hedhili, M. N.; Gurinov, A.; Kelly, M. J.; El Eter, M.; Cavallo, L.; Emsley, L.; Basset, J.-M. The structure and binding mode of citrate in the stabilization of gold nanoparticles. *Nat. Chem.* **2017**, *9*, 890–895.

(11) Hope, M. A.; Nakamura, T.; Ahlawat, P.; Mishra, A.; Cordova, M.; Jahanbakhshi, F.; Mladenović, M.; Runjhun, R.; Merten, L.; Hinderhofer, A.; Carlsen, B. I.; Kubicki, D. J.; Gershoni-Poranne, R.; Schneeberger, T.; Carbone, L. C.; Liu, Y.; Zakeeruddin, S. M.; Lewinski, J.; Hagfeldt, A.; Schreiber, F.; Rothlisberger, U.; Grätzel, M.; Milić, J. V.; Emsley, L. Nanoscale Phase Segregation in Supramolecular π -Templating for Hybrid Perovskite Photovoltaics from NMR Crystallography. *J. Am. Chem. Soc.* **2021**, *143*, 1529–1538.

(12) Ni, Q. Z.; Yang, F.; Can, T. V.; Sergeev, I. V.; D'Addio, S. M.; Jawla, S. K.; Li, Y.; Lipert, M. P.; Xu, W.; Williamson, R. T.; Leone, A.; Griffin, R. G.; Su, Y. In Situ Characterization of Pharmaceutical Formulations by Dynamic Nuclear Polarization Enhanced MAS NMR. *J. Phys. Chem. B* **2017**, *121*, 8132–8141.

(13) Mueller, L. J.; Dunn, M. F. NMR Crystallography of Enzyme Active Sites: Probing Chemically Detailed, Three-Dimensional Structure in Tryptophan Synthase. *Acc. Chem. Res.* **2013**, *46*, 2008–2017.

(14) Cordova, M.; Balodis, M.; Hofstetter, A.; Paruzzo, F.; Nilsson Lill, S. O. N.; Eriksson, E. S. E.; Berruyer, P.; Simões de Almeida, B.; Quayle, M. J.; Norberg, S. T.; Svensk Ankarberg, A.; Schantz, S.; Emsley, L. Structure determination of an amorphous drug through large-scale NMR predictions. *Nat. Commun.* **2021**, *12*, 2964.

(15) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615.

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

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

(18) Pickard, C. J.; Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2001**, *63*, 245101.

(19) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 094116.

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

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

(22) Curtis, F.; Li, X.; Rose, T.; Vázquez-Mayagoitia, Á.; Bhattacharya, S.; Ghiringhelli, L. M.; Marom, N. Gator: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 2246–2264.

(23) Karfunkel, H. R.; Gdanitz, R. J. Ab Initio prediction of possible crystal structures for general organic molecules. *J. Comput. Chem.* **1992**, *13*, 1171–1183.

(24) Bazterra, V. E.; Ferraro, M. B.; Facelli, J. C. Modified genetic algorithm to model crystal structures. I. Benzene, naphthalene and anthracene. *J. Chem. Phys.* **2002**, *116*, 5984–5991.

- (25) Zhu, Q.; Oganov, A. R.; Glass, C. W.; Stokes, H. T. Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystallogr., Sect. B: Struct. Sci.* **2012**, *68*, 215–226.
- (26) Zilka, M.; Dudenko, D. V.; Hughes, C. E.; Williams, P. A.; Sturniolo, S.; Franks, W. T.; Pickard, C. J.; Yates, J. R.; Harris, K. D. M.; Brown, S. P. Ab initio random structure searching of organic molecular solids: assessment and validation against experimental data. *Phys. Chem. Chem. Phys.* **2017**, *19*, 25949–25960.
- (27) Yang, S.; Day, G. M. Exploration and Optimization in Crystal Structure Prediction: Combining Basin Hopping with Quasi-Random Sampling. *J. Chem. Theory Comput.* **2021**, *17*, 1988–1999.
- (28) Hofstetter, A.; Balodis, M.; Paruzzo, F. M.; Widdifield, C. M.; Stevanato, G.; Pinon, A. C.; Bygrave, P. J.; Day, G. M.; Emsley, L. Rapid Structure Determination of Molecular Solids Using Chemical Shifts Directed by Unambiguous Prior Constraints. *J. Am. Chem. Soc.* **2019**, *141*, 16624–16634.
- (29) Tatton, A. S.; Blade, H.; Brown, S. P.; Hodgkinson, P.; Hughes, L. P.; Lill, S. O. N.; Yates, J. R. Improving Confidence in Crystal Structure Solutions Using NMR Crystallography: The Case of β -Piroxicam. *Cryst. Growth Des.* **2018**, *18*, 3339–3351.
- (30) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7260–7268.
- (31) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.
- (32) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- (33) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.
- (34) Rezáč, J. Empirical Self-Consistent Correction for the Description of Hydrogen Bonds in DFTB3. *J. Chem. Theory Comput.* **2017**, *13*, 4804–4817.
- (35) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshayé, M. Y.; Dumitrică, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; van der Heide, T.; Hermann, J.; Irle, S.; Kranz, J. J.; Köhler, C.; Kowalczyk, T.; Kubař, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A. M. N.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Rezáč, J.; Sánchez, C. G.; Sternberg, M.; Stöhr, M.; Stuckenberg, F.; Tkatchenko, A.; Yu, V. W.-z.; Frauenheim, T. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020**, *152*, 124101.
- (36) Paruzzo, F. M.; Hofstetter, A.; Musil, F.; De, S.; Ceriotti, M.; Emsley, L. Chemical shifts in molecular solids by machine learning. *Nat. Commun.* **2018**, *9*, 4501.
- (37) Motherwell, S.; Chisholm, J. A. COMPACT: a program for identifying crystal structure similarity using distances. *J. Appl. Crystallogr.* **2005**, *38*, 228–231.
- (38) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci.* **2016**, *72*, 171–179.
- (39) Engel, E. A.; Anelli, A.; Hofstetter, A.; Paruzzo, F.; Emsley, L.; Ceriotti, M. A Bayesian approach to NMR crystal structure determination. *Phys. Chem. Chem. Phys.* **2019**, *21*, 23385–23400.
- (40) Nyman, J.; Day, G. M. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **2015**, *17*, 5154–5165.
- (41) Iuzzolino, L.; McCabe, P.; Price, S. L.; Brandenburg, J. G. Crystal structure prediction of flexible pharmaceutical-like molecules: density functional tight-binding as an intermediate optimisation method and for free energy estimation. *Faraday Discuss.* **2018**, *211*, 275–296.
- (42) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylisma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A., Jr; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meeke, 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.; Groom, C. R. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr., Sect. B: Struct. Sci.* **2016**, *72*, 439–459.