CHEMICAL PHYSICS

Bayesian probabilistic assignment of chemical shifts in organic solids

Manuel Cordova[1,2], Martins Balodis[1], Bruno Simões de Almeida[1], Michele Ceriotti[2,3], Lyndon Emsley[1,2]*

A prerequisite for NMR studies of organic materials is assigning each experimental chemical shift to a set of geometrically equivalent nuclei. Obtaining the assignment experimentally can be challenging and typically requires time-consuming multidimensional correlation experiments. An alternative solution for determining the assignment involves statistical analysis of experimental chemical shift databases, but no such database exists for molecular solids. Here, by combining the Cambridge Structural Database with a machine learning model of chemical shifts, we construct a statistical basis for probabilistic chemical shift assignment of organic crystals by calculating shifts for more than 200,000 compounds, enabling the probabilistic assignment of organic crystals directly from their two-dimensional chemical structure. The approach is demonstrated with the ^{13}C and ^{1}H assignment of 11 molecular solids with experimental shifts and benchmarked on 100 crystals using predicted shifts. The correct assignment was found among the two most probable assignments in more than 80% of cases.

INTRODUCTION

Chemical shift assignment is the starting point of any detailed nuclear magnetic resonance (NMR) study (*1*). In organic solids at natural isotopic abundance, this is still a laborious and often challenging process. In particular, ^{13}C resonance assignment typically requires the use of the through-bond ^{13}C-^{13}C double-quantum/single-quantum correlation (INADEQUATE) experiment (*2*, *3*). For materials for which the crystal structure is already known, the assignment can be determined at least partially by comparing the experimental chemical shifts with shifts computed using density functional theory (DFT) in the gauge-invariant projector-augmented wave method (*4*, *5*) or fragment-based methods (*6*, *7*). However, in most applications, the full structure is not known and, in particular, de novo chemical shift–based NMR crystallography relies on chemical shift assignment to confidently identify the crystal structure from among a set of candidates generated, for example, through crystal structure prediction (*8*–*11*).

An illustrative example of the assignment problem is shown in Fig. 1, with the ^{13}C cross-polarization (CP) magic angle spinning (MAS) spectrum of ritonavir. The spectrum contains 32 peaks, corresponding to the 37 magnetically inequivalent carbon atoms in the molecule, and assigning the peaks to the atoms is not at all obvious. Several straightforward experimental methods can be used to simplify the assignment process in organic solids. Heteronuclear correlation (HETCOR) experiments (*12*, *13*) provide pairwise ^{1}H-X (where X = ^{13}C, ^{15}N, etc.) correlations and allow the separation of NMR signals along two dimensions, which simplifies the identification of the bonding environment associated with the observed peaks. In addition, spectral editing (*14*–*18*) can be used to identify the carbon multiplicity (i.e., the number of bonded protons) associated to each observed peak, allowing the reduction of the assignment problem to subsets of peaks and corresponding atomic sites.

Chemical shift assignment of biomolecules such as proteins and RNA can be obtained directly from their sequence through statistical analysis of chemical shifts (*19*–*21*). In addition, simultaneous chemical shift assignment and structure determination can be obtained from matching atomic contacts to nuclear Overhauser effect experiments (*19*). These approaches rely on the existence of a large database of experimental chemical shifts and molecular structures, such as the Biological Magnetic Resonance Data Bank (BMRB) (*22*) and Protein Data Bank (*23*), respectively. For example, the BMRB contains more than 9.4 million instances of experimental chemical shifts for 279 types of proton, carbon, and nitrogen sites in the 20 amino acids that make up proteins, with e.g., more than 89,000 instances of the NH shift in alanine alone. Such large and diverse chemical shift databases however do not exist, to our knowledge, for organic crystals.

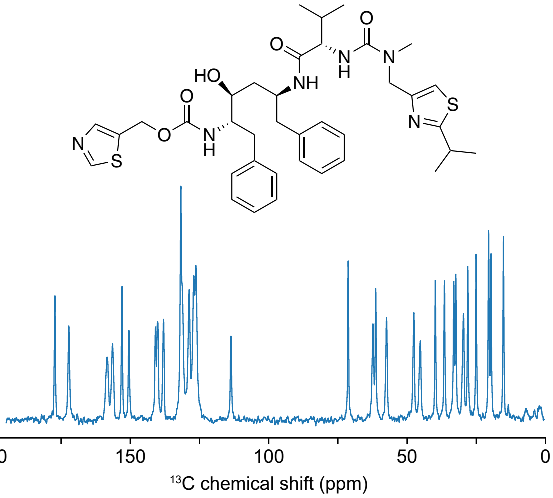


Fig. 1. ^{13}C CPMAS spectrum of ritonavir. Molecular structure of ritonavir and the ^{13}C CPMAS spectrum recorded for a powder sample of ritonavir form II. ppm, parts per million.

[1] Laboratory of Magnetic Resonance, Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. [2] National Centre for Computational Design and Discovery of Novel Materials MARVEL, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. [3] Laboratory of Computational Science and Modelling, Institute of Materials, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.
*Corresponding author. Email: lyndon.emsley@epfl.ch

Recently, ShiftML, a machine learning model able to predict chemical shifts in molecular solids, was introduced (*10*, *24*). This model allows chemical shifts to be obtained directly from the structure of a molecular solid, bypassing the need for an optimized wave function and making the shifts of large ensembles of large structures accessible with DFT accuracy (*11*, *24*).

Here, we show how combining this model with a database of three-dimensional (3D) structures such as the Cambridge Structural Database (CSD) (*25*) enables the probabilistic assignment of organic crystals using chemical shift statistics without any knowledge of the 3D structure. We generate a large database of chemical shifts for organic crystals by predicting shifts using ShiftML on structures extracted from the CSD. By relating the shifts obtained to molecular fragment descriptors, we obtain probabilistic assignments of organic crystals directly from their molecular structure.

RESULTS

The framework presented here was applied to a set of various organic molecules for which the carbon chemical shift assignment was already (at least partially) determined experimentally. The selected set is composed of theophylline (*9*), thymol (*26*), cocaine (*9*), strychnine, AZD5718 (*11*), lisinopril (*27*), ritonavir, the K salt of penicillin G (*28*), β-piroxicam (*29*), decitabine (*30*), and simvastatin (*31*). The experimental spectra used for the assignment of strychnine and ritonavir are shown in figs. S1 and S2. Experimental shifts of lisinopril were obtained from a dihydrate form (*27*). Experimental shifts of ritonavir were obtained from the polymorphic form II.

Graph generation is the starting point of statistical assignment and can be performed directly from the 2D representation of the molecule. Figure 2 (A and B) shows the graphs generated for illustrative carbon atoms in theophylline with a depth $w = 3$. The chemical shift distributions of the carbon labeled 4 in theophylline corresponding to different graph depths are shown in Fig. 2C, together with the corresponding graphs. As expected, the distribution changes as w is increased, until at $w = 3$ and above where they are found to be highly similar, with a width dominated by the uncertainty in the ShiftML prediction. We thus selected a minimum number of 10 instances to construct each probability distribution and used the maximum graph depth that fulfils this requirement for each nucleus.

The prior statistical distribution of chemical shifts for each atom in a molecule can be constructed from the shifts predicted for all atoms in the database that share the same graph. Evaluating the obtained statistical distributions at the observed shifts yields the probability of observing each shift originating from each nucleus in the molecule (Fig. 3, A and B). The possible combinations of individual assignments, based on a Bayesian construction, make it possible to associate a probability to each global assignment of all shifts. After obtaining the probability for each global assignment in the set, marginalization yields individual assignment probabilities (Fig. 3C). In this case, the most probable individual assignment for each carbon, as well as the most probable global assignment, was found to correspond to the experimental assignment of theophylline (black dots in Fig. 3C).

Overlap of the chemical shift distributions can lead to highly ambiguous assignments. A common method to separate overlapping NMR signals consists in spreading them along multiple dimensions. The HETCOR experiment yields high-sensitivity correlated ^{1}H and ^{13}C chemical shifts of dipolar coupled nuclei and can be tuned to obtain a spectrum dominated by one-bond correlations (*12*, *13*). The correlated statistical distributions of chemical shifts corresponding to a simulated HETCOR can be obtained by considering bonded CH pairs in the molecule. This additional dimension often helps separate overlapping 1D statistical distributions and chemical shifts by incorporating the additional information given by the ^{1}H chemical shift. In addition, this can also be used to simultaneously assign ^{13}C and ^{1}H chemical shifts.

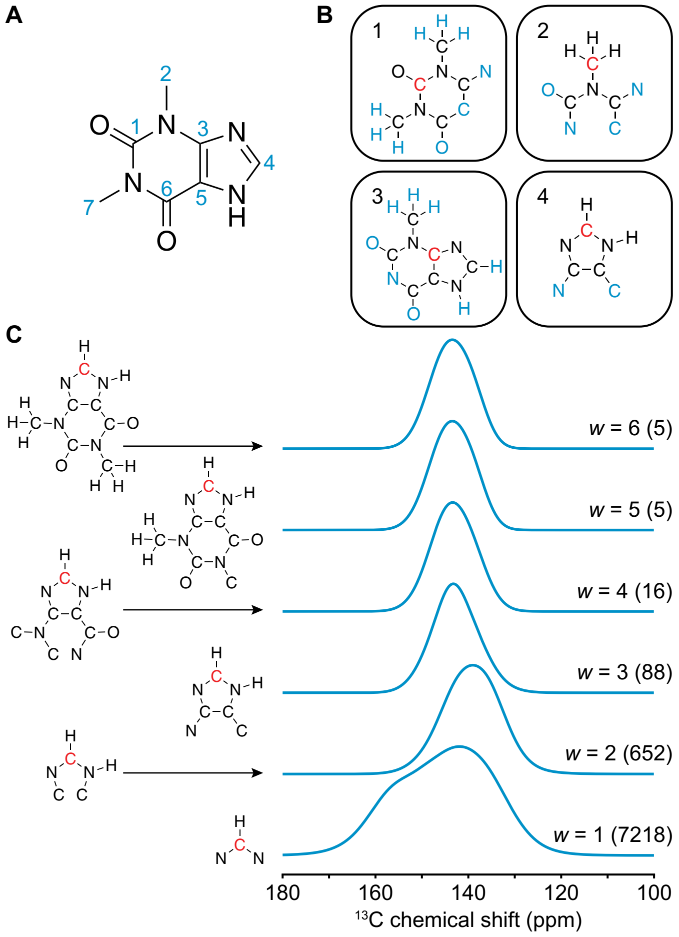


Fig. 2. Graph and statistical distribution. (**A**) 2D structure and carbon labeling scheme of theophylline. (**B**) Graphs of carbons 1, 2, 3, and 4 of theophylline constructed at a depth $w = 3$. In each graph, the red vertex corresponds to the central atom (for which the chemical shift distribution is extracted), and blue vertices indicate the atoms at the maximum shortest path from the central vertex. (**C**) Chemical shift distributions obtained corresponding to the carbon labeled 4, with different graph depths w. The number of instances from the database used to construct each distribution is indicated in parentheses.

Figure 4 depicts the probabilistic assignment of bonded ^{13}C-^{1}H chemical shifts of thymol using 2D correlated statistical shift distributions. The pair of topologically equivalent bonded C-H groups (labeled 9 and 10) was assigned to a pair of experimental shifts in Fig. 3D, as the disambiguation of topologically equivalent nuclei cannot be performed from the 2D representation of a molecule. As seen in Fig. 3B, the assignment of the carbon labeled 8 would have been much more ambiguous using only ^{13}C chemical shifts. The probability of assigning carbon 8 to chemical shift e is 34% using only statistical distributions of ^{13}C chemical shifts (Fig. 4E) and 100%

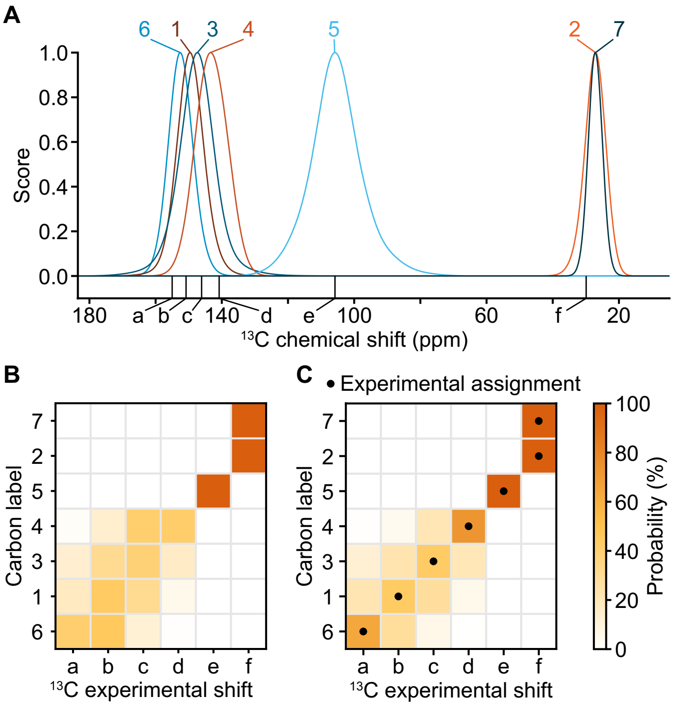


Fig. 3. Probabilistic assignment of theophylline. (**A**) Statistical ^{13}C chemical shift distributions for theophylline (colored lines). The carbon labels follow Fig. 2A. Experimental shifts are indicated by black vertical lines below the distributions and are labeled (a) through (f) in order of decreasing chemical shift (see table S9). (**B**) Probabilities of observing each chemical shift of theophylline for a given carbon nucleus. (**C**) Marginal individual assignment probabilities of the ^{13}C chemical shifts of theophylline after Bayesian inference of the possible global assignments. The dots indicate the experimentally determined correct assignment.

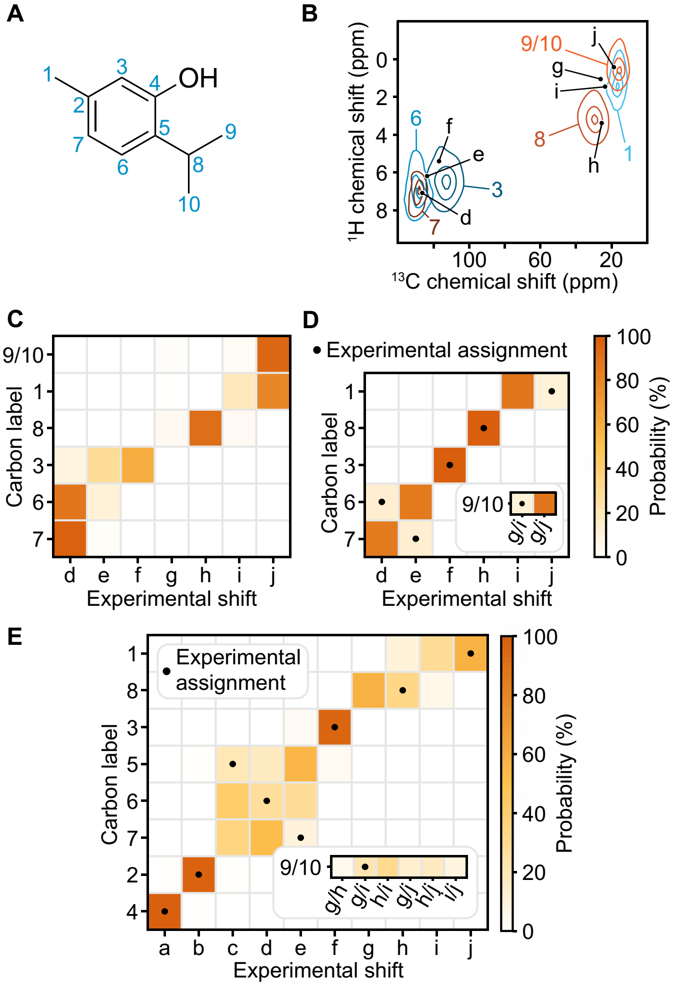


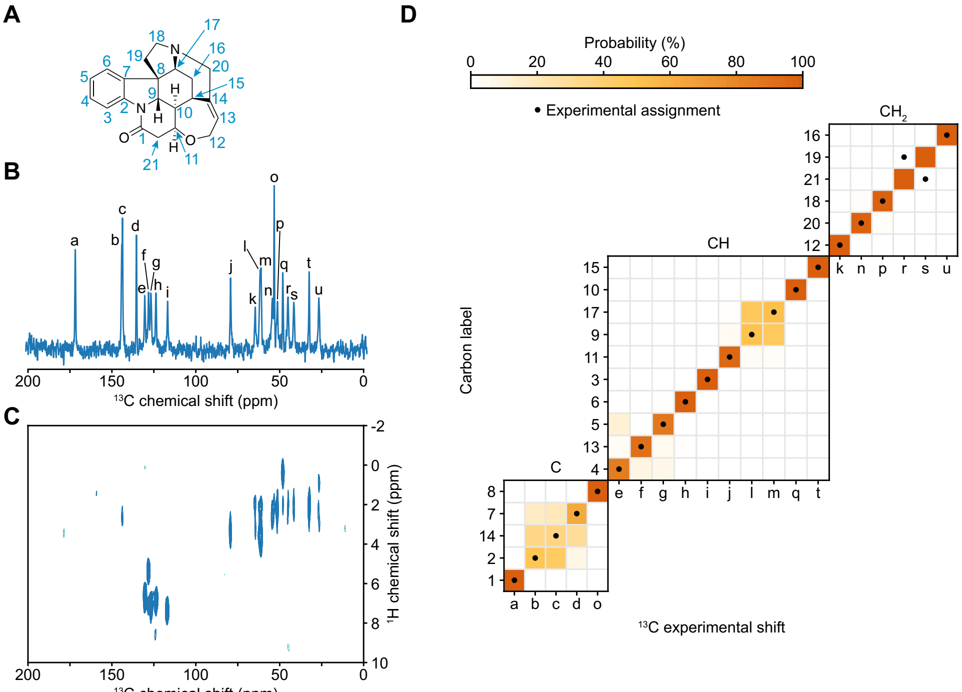
Fig. 4. Probabilistic assignment of thymol using simulated HETCOR. (**A**) Carbon labeling scheme of thymol. (**B**) Contour plot of the correlated statistical chemical shift distributions of bonded ^{13}C-^{1}H in thymol. The carbon labels follow (A). Experimental shifts are indicated by black dots and are labeled alphabetically in order of decreasing ^{13}C chemical shift (see fig. S10). The statistical distributions, normalized such that their maximum is one, are drawn as contour plots at levels 0.1, 0.5, and 0.9. (**C**) Probabilities of observing each ^{13}C-^{1}H shift pair in thymol for a given carbon nucleus. (**D**) Marginal individual assignment probabilities of unique directly bonded CH pairs and of pairs of topologically equivalent CH pairs (insert) in thymol. (**E**) Marginal individual assignment probabilities of unique carbons and of pairs of topologically equivalent carbons (insert) in thymol using only ^{13}C chemical shift distributions. In (D) and (E), the dots indicate the experimentally determined correct assignment.

using correlated statistical distributions of ^{1}H and ^{13}C chemical shifts (Fig. 4D). We note that the most probable assignments of carbons 6 and 7 and of the methyl groups 1, 9, and 10 do not match the experimentally determined ones. We attribute these discrepancies to substantial overlap between the corresponding statistical distributions of chemical shifts that arise because of similar local bonding environments of carbons 6 and 7 and of methyl groups.

In addition to HETCOR, spectral editing methods are also straightforward high-sensitivity experiments that can be performed routinely to aid assignment. These experiments are able to separate ^{13}C chemical shifts according to the number of bonded protons (multiplicity) (*14–17*). The method can thus be directly applied to the statistical assignment framework presented here to break down the statistical assignment problem into smaller subproblems of reduced complexity. This is especially useful when considering molecules yielding substantial overlap of statistical distributions. Knowledge of the multiplicity of ^{13}C chemical shifts can also be used to select a subset of HETCOR peaks to assign.

Figure 5 shows the assignment of ^{13}C and ^{1}H-^{13}C chemical shifts of strychnine using the combination of spectral editing and correlated statistical distributions of chemical shifts. In Fig. 5D, the chemical shifts of carbons without any proton attached were assigned using the 1D ^{13}C chemical shift distributions of the associated nuclei. Carbons with a single bonded proton were assigned using the correlated ^{1}H-^{13}C statistical chemical shift distributions. The carbons with two attached protons were assigned to pairs of correlated ^{1}H-^{13}C chemical shifts, restricting the ^{13}C shift to be unique in each pair.

Figure 5E summarizes the three most probable global assignments of strychnine. For each assignment, the global assignment is broken down into blocks by multiplicity and then potentially into sub-blocks where there is no significant probability of overlap according to a threshold (here, a factor 100 with respect to the highest probability for each nucleus). For each subassignment, there is an associated probability. The most probable assignment of each block was found to match the experimentally determined one, except for the assignment of CH_2 groups, where the assignments of carbons 21 and 19 are swapped compared to the experimentally determined assignment.



E

Unambiguous assignments

Carbon	1	8	12	*P* (%)
a_1	a	o	k	100.0

C

Carbon	2	14	7	*P* (%)
a_1	b	c	d	32.4
a_2	c	b	d	30.7
a_3	b	d	c	16.0

CH$_2$

Carbon	20	18	21	19	16	*P* (%)
a_1	n	p	r	s	u	97.4
a_2	p	n	r	s	u	1.8
a_3	n	p	r	u	s	0.5

CH, aromatic region

Carbon	4	13	5	6	3	*P* (%)
a_1	e	f	g	h	i	64.9
a_2	e	f	h	g	i	16.8
a_3	e	g	f	h	i	9.0

CH, aliphatic region

Carbon	11	9	17	10	15	*P* (%)
a_1	j	l	m	q	t	50.3
a_2	j	m	l	q	t	45.2
a_3	l	j	m	q	t	1.7

Fig. 5. Probabilistic assignment of strychnine using simulated HETCOR and spectral editing. (**A**) Carbon labeling scheme of strychnine. (**B**) 125-MHz ^{13}C CPMAS NMR spectrum of strychnine. (**C**) ^{1}H-^{13}C HETCOR spectrum of strychnine. (**D**) Marginal individual assignment probabilities of the carbon nuclei of strychnine. The carbon multiplicity is indicated above each probability map. The HETCOR shifts were used to assign CH and CH$_2$ carbons. The shifts are labeled alphabetically in order of decreasing chemical shift (see table S11). (**E**) The three most probable global assignments for the different blocks assigned individually along with their probability. The individual assignments making up the global assignments are indicated in blue if they correspond to the experimentally determined assignment and in red otherwise. Carbons 1, 8, and 12 were assigned without ambiguity (*P* = 100%) directly from the evaluation of their statistical distributions of chemical shifts on the observed shifts.

This is due to the large difference between the distribution of chemical shifts and experimental shift of carbon 19 (see fig. S9), which could come from an unusual intermolecular environment of that atomic site in the crystal structure.

We consider that a reliable assignment is difficult to extract from the set of global assignments and associated probabilities, especially in cases with a large number of overlapping distributions and shifts, which yield many possible global assignments. Marginalization helps simplify the analysis of global assignments and identify ambiguities more easily. This can be seen in Fig. 5D, where the assignment of carbon 7 to shift d is favored compared to shifts b and c, which suggests only a pairwise uncertainty between carbon 2 and 14.

In addition to strychnine, shown in Fig. 5, the marginal individual assignment probabilities obtained for a set of nine selected molecules with complete experimental assignments (except for the two phenyl rings of ritonavir) using spectral editing and correlated ^{1}H-^{13}C statistical chemical shift distributions are shown in Fig. 6 and figs. S16 to S18. The assignment of carbon nuclei without any attached proton were obtained from the 1D statistical distributions of ^{13}C chemical shifts. The statistical distributions of chemical shifts for each example are shown in figs. S3 to S10. Notably, the assignment of lisinopril was found to be possible even when omitting the water molecules present in the crystal structure.

Figure 7 shows the assignment of the K salt of penicillin G. Only the organic ion was considered to construct the graph descriptors used to extract statistical distributions of chemical shifts. As for the presence of the water molecule in the case of lisinopril above, here, the presence of the potassium ion, which is absent from the database, did not lead to a significant decrease in the ability of the model to predict the assignment, highlighting its generality beyond molecules for which chemical shifts can be computed by ShiftML. While ShiftML would not be able to compute shifts for crystals where even only one atom is different from C, H, N, O, and S, this model only requires the molecule to be assigned to only contain these elements to obtain

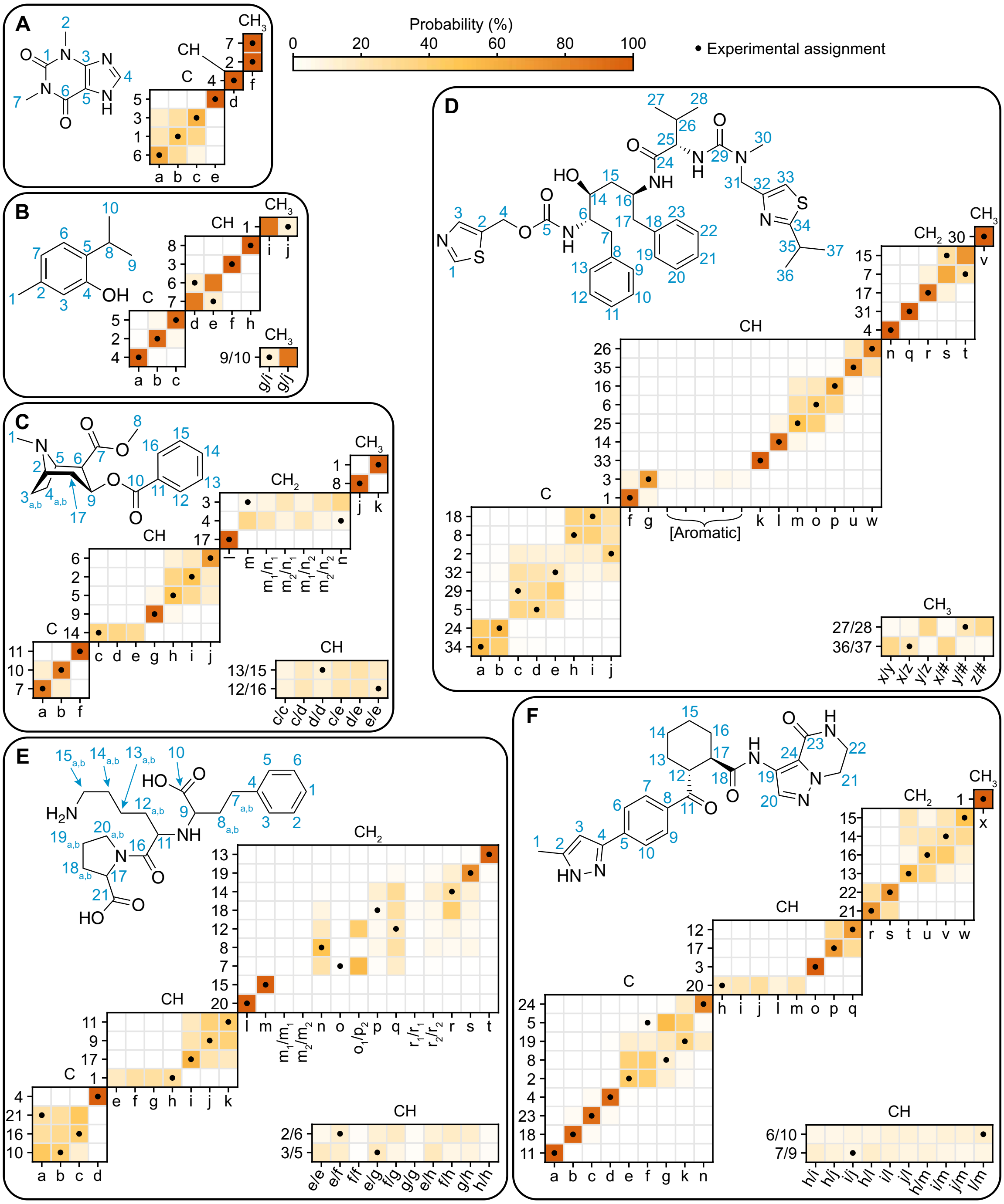


Fig. 6. Probabilistic assignment of six organic crystals. Marginal individual assignment probabilities of ^{13}C chemical shifts of (**A**) theophylline, (**B**) thymol, (**C**) cocaine, (**D**) ritonavir, (**E**) lisinopril, and (**F**) AZD5718 using correlated ^{1}H-^{13}C chemical shift distributions and spectral editing. For each probability map, labels along the vertical axis indicate nuclei, and labels along the horizontal axis denote experimental shifts labeled alphabetically in order of decreasing ^{13}C shift (see tables S9 to S14). The carbon multiplicity is indicated above each marginal assignment probability map. In (D), the assignment of carbons 9 to 13 and 19 to 23 is not shown, as their experimental assignment is ambiguous. Nevertheless, the associated peaks were considered during the assignment process. The assignment probabilities of the aromatic CH groups of ritonavir are shown in fig. S11.

the probabilistic assignment. If the additional component in a salt or a cocrystal were to lead to a very different crystalline environment from those included in the database, then this might lead to poor performance of the probabilistic assignment.

The marginal individual assignment probabilities obtained directly from the 2D representation of the molecules were found to match the experimentally determined assignment in most cases. We observe that assignment ambiguities generally involve pairs or triplets

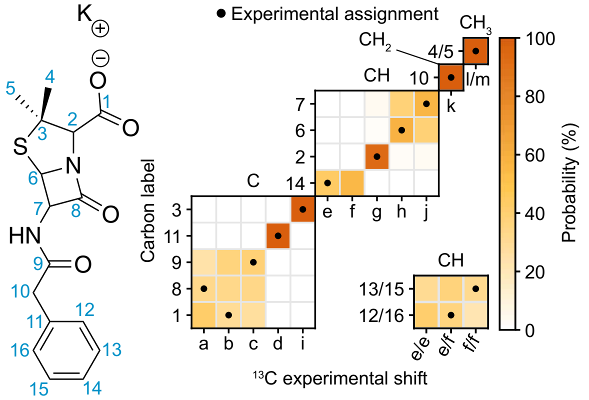


Fig. 7. Probabilistic assignment of the K salt of penicillin G. Carbon labeling scheme and marginal individual assignment probabilities of the K salt of penicillin G. The shifts are labeled (a) through (m) in order of decreasing chemical shift (see table S16).

of nuclei and shifts, leaving only a few possibilities for the NMR spectroscopist to further investigate to obtain the complete chemical shift assignment. Of the 178 experimental individual assignments considered in Figs. 5 to 7 and figs. S16 to S18, only 8 were associated with a probability below 10% and two below 1%. These low probabilities were generally associated with crowded regions in experimental spectra or with statistical outlier shifts compared to the distributions, which could have originated from unusual intermolecular environments.

To validate these results in a statistically significant manner, we evaluated the performance of the framework presented here on a benchmark set of a hundred crystal structures having between 10 and 20 different carbon atoms, randomly selected from the CSD database. In total, this corresponds to 1214 inequivalent carbon atoms. We used the ShiftML-predicted shifts for each atom as the correct assignment and excluded those shifts from the statistical distributions used to assign the molecules. The benchmark set was separated into five subsets containing 20 structures each that were evaluated independently to obtain SDs. Although using shifts predicted by ShiftML may introduce a bias, as the same method was used to construct the database of shifts, we assumed that the Gaussian width used to construct the statistical distributions of chemical shifts and the exclusion of the shifts assigned from the sets of shifts used to construct those distributions mitigate this issue.

Figure 8 summarizes the performance of the probabilistic assignment model on the experimental (Fig. 8A) and synthetic (Fig. 8B) sets of molecules selected. The use of spectral editing and correlated ^{1}H-^{13}C chemical shift distributions was found to improve the ability of the model to correctly assign carbon chemical shifts. Using either 2D statistical distributions of chemical shifts, spectral editing, or combining both led to the experimental assignment being among the two most probable marginal assignments in more than 80% of cases. Overall, the performances on the experimental benchmark set were consistent with the synthetic benchmark set, except when using spectral editing where a slight improvement in the experimental set compared to the synthetic set was observed.

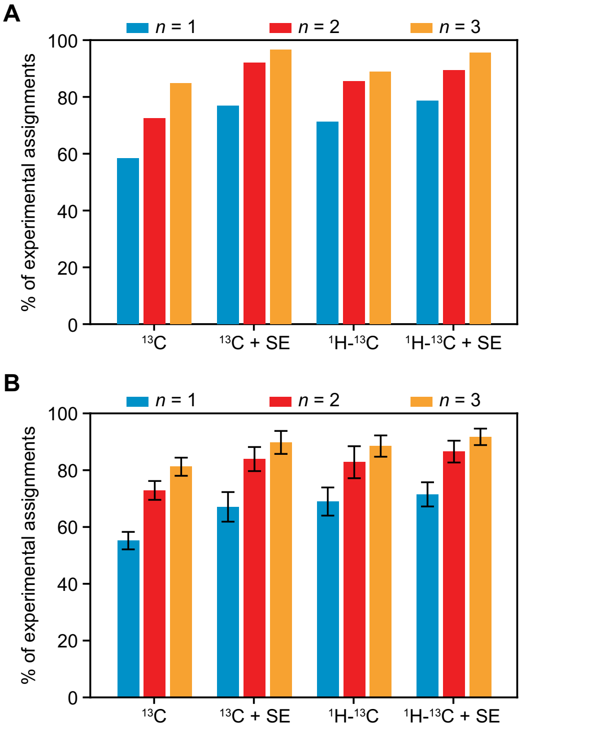


Fig. 8. Model performances. Comparison of probabilistic assignment performances using 1D (^{13}C) or 2D (^{1}H-^{13}C) statistical distributions and including spectral editing (SE). Proportion of the experimental assignments being within the *n* (*n* = 1, 2, and 3) most probable marginal individual assignments in the (**A**) experimental and (**B**) synthetic benchmark sets of molecules. Error bars indicate the SD over the five subsets making up the synthetic benchmark set.

DISCUSSION

The framework presented here allows chemical shift assignment of organic crystals directly from their 2D structure. This was achieved through the chemical shift prediction for more than 200,000 organic crystal structures, which yields statistical distributions of chemical shifts for given covalent environments. A Bayesian framework was then used to obtain probabilistic marginal assignments of individual nuclei from the probabilities of the set of global assignments generated. Overall, using correlated ^{1}H-^{13}C chemical shift distributions in tandem with spectral editing, the method was found to include the experimental assignment among the two most probable marginal assignments in more than 80% of cases.

Furthermore, in most cases, any ambiguity is found in small subgroups of shifts. This is highlighted in lisinopril in, for example, the CH_2 carbons because of significant overlap between the corresponding statistical distributions of chemical shifts and because of similar experimental shifts (see fig. S7).

In summary, the approach presented here can provide marginal assignments based only on the 2D molecular structure, where typically most of the resonances will be assigned with high probabilities, and only a few resonances will show ambiguities among doubles or triples that can then be the subject of targeted experiments for disambiguation, if needed, or left unresolved and assigned such that the error is minimized when compared with computed shifts for model structures (e.g., when performing NMR-driven crystal structure determination). This can greatly accelerate the assignment process. In particular, the method is shown to provide assignments for molecules such as strychnine, lisinopril, AZD5718, and ritonavir, which have crowded ^{13}C spectra with between 20 and 40 distinct carbons and which would have been previously completely unaddressable without resorting to natural abundance ^{13}C-^{13}C correlations. For example, in strychnine, of the 21 carbons, 14 are correctly assigned with more than 75% confidence. The model was also successfully applied to the assignment of a hydrate and an organic salt, with no significant performance loss compared to the benchmark set. We expect that a more accurate model of chemical shifts could lead to improved probabilistic assignment through the framework presented here.

The method shown here is not restricted to ^{1}H and ^{13}C and can be used to assign the isotropic shifts of any NMR-active isotope of hydrogen, carbon, nitrogen, and oxygen in principle. To illustrate that, fig. S19 and Supplementary Text describe the probabilistic assignment of the ^{15}N shifts of AZD5718.

The code is publicly available at https://github.com/manucordova/ProbAsn, and a user guide is available in Supplementary Text and on the GitHub webpage. A suggested workflow to assign an organic solid is also described in Supplementary Text.

MATERIALS AND METHODS

NMR spectroscopy

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

1D ^{1}H MAS NMR spectra were recorded at a temperature of 298 K using rotor spinning rates (ν_r) up to 111 kHz. 1D ^{13}C CPMAS (*32*) NMR spectra were acquired at 298 K with ν_r of 12.5 and 22 kHz for ritonavir and strychnine, respectively. During the signal acquisition SPINAL-64 decoupling (*33*) was applied with a ^{1}H radio frequency (rf) field amplitude of 100 kHz. For ritonavir spectral editing, experiments were used to distinguish carbons with different numbers of protons attached to them. To selectively remove quaternary carbons, a 1D version of MAS-J-HSQC (*18*) was used; to remove quaternary and primary carbons, a double quantum filter was added to the MAS-J-HSQC (*18*) sequence, and to remove primary and secondary carbons, a simple CP experiment with an inserted delay of 0.5 ms before acquisition and after the CP pulse was applied (*14*).

2D ^{1}H-^{13}C HETCOR experiments were carried out at 298 K using ν_r = 22 kHz. During t_1, 100-kHz eDUMBO-1_{22} was applied to decouple the ^{1}H-^{1}H dipolar coupling (*34*), and during t_2, 100-kHz SPINAL-64 decoupling was applied.

The natural abundance 2D ^{13}C-^{13}C refocused INADEQUATE (*2*, *35*) spectra required for the direct experimental assignment for ritonavir and strychnine were acquired using a Bruker 400-MHz Ascend NMR spectrometer. The probe was configured into $^{1}H/^{13}C$ double resonance mode. Variable amplitude CP (*36*) was used to transfer polarization from ^{1}H to ^{13}C. SPINAL-64 (*33*) heteronuclear ^{1}H decoupling with rf fields of 100 kHz was applied in all cases. The temperature of the sample for ritonavir was 250 K, and a 4-mm rotor was used with a spinning frequency of 12.5 kHz. Experiments (2 × 120 hours) were acquired and combined in postprocessing to obtain the final spectrum (total time, 10 days). For strychnine, dynamic nuclear polarization (DNP) was used (*37*). The sample was impregnated with 10 mM AMUPol dissolved in 60:30:10 glycerol-d8:D_2O:H_2O. The spectrometer is equipped with a low-temperature magic angle spinning 3.2-mm probe and connected through a corrugated waveguide to a 263-GHz gyrotron capable of outputting ca. 5 to 10 W of continuous-wave microwaves (*38*). The sweep coil of the main magnetic field was optimized so that the microwave irradiation gave the maximum positive proton DNP enhancement with binitroxide cross-effect–based polarizing agents (e.g., AMUPol and TEKPol). The temperature of the sample for ritonavir was 92 K, and a 3.2-mm rotor was used with a spinning frequency of 12.5 kHz. A DNP enhancement of 36 was determined on the basis of the ratio of the area of the spectra acquired with and without microwave irradiation. The DNP-enhanced natural abundance 2D ^{13}C-^{13}C refocused INADEQUATE experiment (*37*) was run for 45 hours.

All chemical shifts were referenced via alanine. The full set of acquisition parameters is given in tables S1 to S4.

Selection of crystal structures

The structures used to construct the chemical shift database were obtained from the CSD (*25*). Only the organic crystal structures suitable for chemical shift predictions were selected. The corresponding selection criteria were that every structure must only contain C, H, N, O, and S atoms and that the disorder is resolvable. Missing protons were added automatically using the tool built into the CSD Python API. In total, 205,069 valid structures were selected.

Relaxation and chemical shift prediction

Because proton positions in published single-crystal x-ray diffraction structures may not correspond to the actual hydrogen positions in the crystals, they have to be optimized. Because of the large number of structures selected, DFT relaxation would be prohibitively costly. The semiempirical density functional tight binding (DFTB) method (*39*) was thus chosen to relax proton positions in all structures. The structures were optimized at the DFTB3-D3H5 level of theory (*40*, *41*) using the 3ob-3-1 parameter set (*42*, *43*). Further computational details are given in Supplementary Text. Instances where the structure relaxation failed were discarded. A total of 203,303 structures were successfully relaxed and considered for chemical shift prediction.

All chemical shift predictions were performed using ShiftML version 1.2 (publicly available at https://shiftml.epfl.ch) (*10*, *24*). Conversions of predicted shieldings to chemical shifts were performed by least squares fitting of the shieldings obtained for benchmark sets of DFTB-relaxed structures to their experimental chemical shifts, fixing the slope to a value of −1. The offsets obtained were found to be 30.96 parts per million (ppm) for ^{1}H, 168.64 ppm for ^{13}C, 185.99 ppm for ^{15}N, and 205.08 ppm for ^{17}O. This corresponds to ^{1}H and ^{13}C shifts relative to tetramethylsilane (TMS), ^{15}N shifts relative to NH_4Cl, and ^{17}O shifts relative to liquid H_2O. The sets of structures and isotropic chemical shifts used to determine shielding-to-shift conversions are described in tables S5 to S8. We note that chemical shieldings are stored in the database and converted to chemical shifts on the fly during the construction of chemical shift distributions. In total, the database contains 5,243,129 unique ^{1}H, 4,847,864 unique ^{13}C, 466,370 unique ^{15}N, and 867,446 unique ^{17}O chemical shifts, respectively.

Molecular fragment descriptors

For assignment of the spectrum of a molecule of unknown structure, classification of the predicted shifts should be done such that a statistical distribution of chemical shifts can be obtained for any nucleus from the 2D representation of a molecule. The molecular fragment descriptor should thus not contain any information about conformation or molecular packing in the crystal structures. Among the topological atom-centered descriptors that fit these requirements (*44*–*46*), we chose to represent topological atomic environments by graphs where vertices represent atoms and edges represent covalent connectivities. The vertices were labeled by element, and the edges were kept unlabeled. Graphs were cut to a maximum depth w of 6, defined as the maximum shortest path between the central vertex (for which the chemical shift is predicted) and any other vertex in the path. Conversion of the 3D crystal structures to their corresponding graphs was performed by identifying atom pairs as covalently bonded when the distance between the atoms in the pair is less than 1.1 times the sum of the covalent radii of the atoms involved.

Database construction and search

A given topological atomic environment can be searched by identifying which graphs in the database match the graph of the selected atomic environment. However, there is no known algorithm able to solve the graph isomorphism problem required for each database entry in polynomial time (*47*, *48*). Thus, the search was simplified by using the Weisfeiler-Lehman hash (*49*) as a unique graph identifier. If the number of instances of a given atomic environment identified in the database was deemed too small to produce statistically significant chemical shift distributions, then the atomic environment was searched again after reducing the graph depth. For this work, we chose a minimum number of instances of 10. Further details concerning the database architecture and search can be found in Supplementary Text.

Construction of probability distributions

We use a notation and a conceptual framework extending the Bayesian selection of crystal structure prediction candidate structures compatible with measured shifts (*10*). From the set of chemical shifts and uncertainties $\{y_k, \sigma_k\}$ predicted by ShiftML for the CSD structures that share the same graph G_i as the atom i in the molecule of interest, we define the probability of observing a chemical shift y for that atom as proportional to the sum of Gaussian functions centered on each predicted shift y_k and with a width σ_k given by its prediction uncertainty

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

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

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

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

Probabilistic assignment

Considering the vector of observed shifts $\mathbf{y}$, the probability that one of its elements y_j originates from atom i is obtained by evaluating Eq. 1 (or Eq. 2) for all elements in $\mathbf{y}$

$$p(y_j \mid i) = \frac{p_i(y_j)}{\sum_k p_i(y_k)} \tag{3}$$

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

$$p(\mathbf{y} \mid \mathbf{a}) = \prod_i p(y_{a_i} \mid i) \tag{4}$$

Applying Bayes theorem on Eq. 4 yields the probability of an assignment $\mathbf{a}$ given the observed vector of shifts $\mathbf{y}$

$$p(\mathbf{a} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{a})\,p(\mathbf{a})}{p(\mathbf{y})} = \frac{p(\mathbf{y} \mid \mathbf{a})\,p(\mathbf{a})}{\sum_{\mathbf{a}'} p(\mathbf{y} \mid \mathbf{a}')\,p(\mathbf{a}')} \tag{5}$$

In Eq. 5, we assume that $p(\mathbf{a})$ is a nonzero constant if the assignment is valid (i.e., if all nuclei are assigned to only one chemical shift and if all observed shifts are assigned at least one nucleus) and zero otherwise. Whenever some of the assignments can be made according to experimental data or heuristic arguments, such prior information can be incorporated in the definition through $p(\mathbf{a})$. By combining individual assignments, the complete set of possible global assignments can be generated. Because of the combinatorial complexity of generating all possible global assignments, several procedures were implemented to reduce the global assignment generation cost while ensuring that the most probable assignments are generated, and these are described in Supplementary Text. Note that if the probability of any shift originating from a given nucleus is lower by a set threshold (typically a factor of 100) than the maximum probability for that nucleus, then it is discarded. This results in some nuclei being assigned unambiguously independently of the rest of the global assignment (e.g., shift "e" in Fig. 3).

Equation 5 assigns a distinct probability to each possible assignment of the entries of the measured shifts vector $\mathbf{y}$ to all the environments. It is the correct probabilistic metric to compare two assignments but is hard to interpret. A more compact indicator is given by the marginal probability that atom i is assigned to shift j, which can be extracted from the set of generated assignments by considering only the vectors $\mathbf{a}$ containing that particular individual assignment. This is shown in Eq. 6 by the Kronecker delta $\delta_{a_i j}$, which selects the assignments for which $a_i = j$

$$p(a_i \;=\; j \mid \mathbf{y}) = \frac{\sum_{\mathbf{a}} \delta_{a_i j}\, p(\mathbf{a} \mid \mathbf{y})}{\sum_{\mathbf{a}} p(\mathbf{a} \mid \mathbf{y})} \tag{6}$$

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

Synthetic benchmark set

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

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abk2341

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. B. Reif, S. E. Ashbrook, L. Emsley, M. Hong, Solid-state NMR spectroscopy. *Nat. Rev. Methods Primers* **1**, 2 (2021).
2. A. Lesage, C. Auger, S. Caldarelli, L. Emsley, Determination of through-bond carbon–carbon connectivities in solid-state NMR using the INADEQUATE experiment. *J. Am. Chem. Soc.* **119**, 7867–7868 (1997).
3. A. Lesage, M. Bardet, L. Emsley, Through-bond carbon–carbon connectivities in disordered solids by NMR. *J. Am. Chem. Soc.* **121**, 10987–10993 (1999).
4. C. J. Pickard, F. Mauri, All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys. Rev. B* **63**, 245101 (2001).
5. R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates, V. Zorin, Chemical shift computations on a crystallographic basis: Some reflections and comments. *Magn. Reson. Chem.* **45**, S174–S186 (2007).
6. J. D. Hartman, G. J. O. Beran, Fragment-based electronic structure approach for computing nuclear magnetic resonance chemical shifts in molecular crystals. *J. Chem. Theory Comput.* **10**, 4862–4872 (2014).
7. J. D. Hartman, S. Monaco, B. Schatschneider, G. J. O. Beran, Fragment-based ^{13}C nuclear magnetic resonance chemical shift predictions in molecular crystals: An alternative to planewave methods. *J. Chem. Phys.* **143**, 102809 (2015).
8. E. Salager, G. M. Day, R. S. Stein, C. J. Pickard, B. Elena, L. Emsley, Powder crystallography by combined crystal structure prediction and high-resolution ^{1}H solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 2564–2566 (2010).
9. M. Baias, C. M. Widdifield, J. N. Dumez, H. P. G. Thompson, T. G. Cooper, E. Salager, S. Bassil, R. S. Stein, A. Lesage, G. M. Day, L. Emsley, Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state ^{1}H NMR spectroscopy. *Phys. Chem. Chem. Phys.* **15**, 8069–8080 (2013).
10. E. A. Engel, A. Anelli, A. Hofstetter, F. Paruzzo, L. Emsley, M. Ceriotti, A Bayesian approach to NMR crystal structure determination. *Phys. Chem. Chem. Phys.* **21**, 23385–23400 (2019).
11. M. Cordova, M. Balodis, A. Hofstetter, F. Paruzzo, S. O. Nilsson Lill, E. S. E. Eriksson, P. Berruyer, B. Simões de Almeida, M. J. Quayle, S. T. Norberg, A. Svensk Ankarberg, S. Schantz, L. Emsley, Structure determination of an amorphous drug through large-scale NMR predictions. *Nat. Commun.* **12**, 2964 (2021).
12. P. Caravatti, G. Bodenhausen, R. R. Ernst, Heteronuclear solid-state correlation spectroscopy. *Chem. Phys. Lett.* **89**, 363–367 (1982).
13. P. Caravatti, L. Braunschweiler, R. R. Ernst, Heteronuclear correlation spectroscopy in rotating solids. *Chem. Phys. Lett.* **100**, 305–310 (1983).
14. S. J. Opella, M. H. Frey, Selection of nonprotonated carbon resonances in solid-state nuclear magnetic resonance. *J. Am. Chem. Soc.* **101**, 5854–5856 (1979).
15. X. L. Wu, K. W. Zilm, Complete spectral editing in CPMAS NMR. *J. Magn. Reson. Ser. A* **102**, 205–213 (1993).
16. X. L. Wu, S. T. Burns, K. W. Zilm, Spectral editing in CPMAS NMR. Generating subspectra based on proton multiplicities. *J. Magn. Reson. Ser. A* **111**, 29–36 (1994).
17. A. Lesage, S. Steuernagel, L. Emsley, Carbon-13 spectral editing in solid-state NMR using heteronuclear scalar couplings. *J. Am. Chem. Soc.* **120**, 7095–7100 (1998).
18. A. Lesage, D. Sakellariou, S. Steuernagel, L. Emsley, Carbon–proton chemical shift correlation in solid-state NMR by through-bond multiple-quantum spectroscopy. *J. Am. Chem. Soc.* **120**, 13194–13201 (1998).
19. P. Guerry, T. Herrmann, Advances in automated NMR protein structure determination. *Q. Rev. Biophys.* **44**, 257–309 (2011).
20. E. Schmidt, P. Güntert, A new algorithm for reliable and general NMR resonance assignment. *J. Am. Chem. Soc.* **134**, 12817–12829 (2012).
21. T. Aeschbacher, E. Schmidt, M. Blatter, C. Maris, O. Duss, F. H. T. Allain, P. Güntert, M. Schubert, Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.* **41**, e172–e172 (2013).
22. E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Y. Yao, J. L. Markley, BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408 (2008).
23. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
24. F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, L. Emsley, Chemical shifts in molecular solids by machine learning. *Nat. Commun.* **9**, 4501 (2018).
25. C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, The Cambridge Structural Database. *Acta Crystallogr. B* **72**, 171–179 (2016).
26. E. Salager, R. S. Stein, C. J. Pickard, B. Elena, L. Emsley, Powder NMR crystallography of thymol. *Phys. Chem. Chem. Phys.* **11**, 2610–2621 (2009).
27. M. Miclaus, I.-G. Grosu, X. Filip, C. Tripon, C. Filip, Optimizing structure determination from powders of crystalline organic solids with high molecular flexibility: The case of lisinopril dihydrate. *CrstEngComm* **16**, 299–303 (2014).
28. N. Mifsud, B. Elena, C. J. Pickard, A. Lesage, L. Emsley, Assigning powders to crystal structures by high-resolution ^{1}H–^{1}H double quantum and ^{1}H–^{13}C J-INEPT solid-state NMR spectroscopy and first principles computation. A case study of penicillin G. *Phys. Chem. Chem. Phys.* **8**, 3418–3422 (2006).
29. A. S. Tatton, H. Blade, S. P. Brown, P. Hodgkinson, L. P. Hughes, S. O. N. Lill, J. R. Yates, Improving confidence in crystal structure solutions using NMR crystallography: The case of β-piroxicam. *Cryst. Growth Des.* **18**, 3339–3351 (2018).
30. J. Brus, J. Czernek, L. Kobera, M. Urbanova, S. Abbrent, M. Husak, Predicting the crystal structure of decitabine by powder NMR crystallography: Influence of long-range molecular packing symmetry on NMR parameters. *Cryst. Growth Des.* **16**, 7102–7111 (2016).
31. J. Brus, A. Jegorov, Through-bonds and through-space solid-state NMR correlations at natural isotopic abundance: Signal assignment and structural study of simvastatin. *J. Phys. Chem. A* **108**, 3955–3964 (2004).
32. A. Pines, M. G. Gibby, J. S. Waugh, Proton-enhanced NMR of dilute spins in solids. *J. Chem. Phys.* **59**, 569–590 (1973).
33. B. M. Fung, A. K. Khitrin, K. Ermolaev, An improved broadband decoupling sequence for liquid crystals and solids. *J. Magn. Reson.* **142**, 97–101 (2000).
34. B. Elena, G. de Paepe, L. Emsley, Direct spectral optimisation of proton-proton homonuclear dipolar decoupling in solid-state NMR. *Chem. Phys. Lett.* **398**, 532–538 (2004).
35. A. Bax, R. Freeman, T. A. Frenkiel, An NMR technique for tracing out the carbon skeleton of an organic-molecule. *J. Am. Chem. Soc.* **103**, 2102–2104 (1981).
36. O. B. Peersen, X. L. Wu, I. Kustanovich, S. O. Smith, Variable-amplitude cross-polarization MAS NMR. *J. Magn. Reson. Ser. A* **104**, 334–339 (1993).
37. A. J. Rossini, A. Zagdoun, F. Hegner, M. Schwarzwalder, D. Gajan, C. Coperet, A. Lesage, L. Emsley, Dynamic nuclear polarization NMR spectroscopy of microcrystalline solids. *J. Am. Chem. Soc.* **134**, 16899–16908 (2012).
38. M. Rosay, L. Tometich, S. Pawsey, R. Bader, R. Schauwecker, M. Blank, P. M. Borchard, S. R. Cauffman, K. L. Felch, R. T. Weber, R. J. Temkin, R. G. Griffin, W. E. Maas, Solid-state dynamic nuclear polarization at 263 GHz: Spectrometer design and experimental results. *Phys. Chem. Chem. Phys.* **12**, 5850–5860 (2010).
39. M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58**, 7260–7268 (1998).
40. M. Gaus, Q. Cui, M. Elstner, DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* **7**, 931–948 (2011).
41. J. Rezac, Empirical self-consistent correction for the description of hydrogen bonds in DFTB3. *J. Chem. Theory Comput.* **13**, 4804–4817 (2017).
42. M. Gaus, A. Goez, M. Elstner, Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.* **9**, 338–354 (2013).
43. M. Gaus, X. Lu, M. Elstner, Q. Cui, Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications. *J. Chem. Theory Comput.* **10**, 1518–1537 (2014).
44. D. Rogers, M. Hahn, Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
45. R. Garcia-Domenech, J. Galvez, J. V. de Julian-Ortiz, L. Pogliani, Some new trends in chemical graph theory. *Chem. Rev.* **108**, 1127–1169 (2008).
46. Danishuddin, A. U. Khan, Descriptors and their selection methods in QSAR analysis: Paradigm for drug design. *Drug Discov. Today* **21**, 1291–1302 (2016).
47. L. Babai, Graph isomorphism in quasipolynomial time [extended abstract]. *Acm. S. Theory Comput.* 684–697 (2016).
48. M. Grohe, P. Schweitzer, The graph isomorphism problem. *Commun. Acm* **63**, 128–134 (2020).
49. B. Y. Weisfeiler, A. A. Leman, The reduction of a graph to canonical form and the algebra which appears therein. *NTI Series* **2**, 12–16 (1968).
50. H. J. Monkhorst, J. D. Pack, Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).

51. J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller, G. J. O. Beran, Benchmark fragment-based ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shift predictions in molecular crystals. *Phys. Chem. Chem. Phys.* **18**, 21686–21709 (2016).
52. A. S. Tatton, T. N. Pham, F. G. Vogt, D. Iuga, A. J. Edwards, S. P. Brown, Probing intermolecular interactions and nitrogen protonation in pharmaceuticals by novel ^{15}N-edited and 2D ^{14}N-^{1}H solid-state NMR. *CrstEngComm* **14**, 2654–2659 (2012).
53. A.-C. Uldry, J. M. Griffin, J. R. Yates, M. Pérez-Torralba, M. Dolores Santa Maria, A. L. Webber, M. L. L. Beaumont, A. Samoson, R. M. Claramunt, C. J. Pickard, S. P. Brown, Quantifying weak hydrogen bonding in uracil and 4-cyano-4'-ethynylbiphenyl: A combined computational and experimental investigation of NMR chemical shifts in the solid state. *J. Am. Chem. Soc.* **130**, 945–954 (2008).
54. R. K. Harris, P. Jackson, High-resolution ^{1}H and ^{13}C NMR of solid 2-aminobenzoic acid. *J. Phys. Chem. Solid* **48**, 813–818 (1987).
55. E. Carignani, S. Borsacchi, J. P. Bradley, S. P. Brown, M. Geppi, Strong intermolecular ring current influence on ^{1}H chemical shifts in two crystalline forms of naproxen: A combined solid-state NMR and DFT study. *J. Phys. Chem. C* **117**, 17731–17740 (2013).
56. R. K. Harris, P. Hodgkinson, V. Zorin, J. N. Dumez, B. Elena-Herrmann, L. Emsley, E. Salager, R. S. Stein, Computation and NMR crystallography of terbutaline sulfate. *Magn. Reson. Chem.* **48**, S103–S112 (2010).
57. F. Liu, C. G. Phung, D. W. Alderman, D. M. Grant, Carbon-13 chemical shift tensors in methyl glycosides, comparing diffraction and optimized structures with single-crystal NMR. *J. Am. Chem. Soc.* **118**, 10629–10634 (1996).
58. M. H. Sherwood, D. W. Alderman, D. M. Grant, Assignment of carbon-13 chemical-shift tensors in single-crystal sucrose. *J. Magn. Reson. Ser. A* **104**, 132–145 (1993).
59. F. Liu, C. G. Phung, D. W. Alderman, D. M. Grant, Analyzing and assigning carbon-13 chemical-shift tensors in α-L-rhamnose monohydrate single crystals. *J. Magn. Reson. Ser. A* **120**, 242–248 (1996).
60. F. Liu, C. G. Phung, D. W. Alderman, D. M. Grant, Analyzing and assigning carbon-13 chemical-shift tensors in fructose, sorbose, and xylose single crystals. *J. Magn. Reson. Ser. A* **120**, 231–241 (1996).
61. V. G. Malkin, O. L. Malkina, D. R. Salahub, Influence of intermolecular interactions on the ^{13}C NMR shielding tensor in solid α-glycine. *J. Am. Chem. Soc.* **117**, 3294–3295 (1995).
62. A. Naito, S. Ganapathy, K. Akasaka, C. A. Mcdowell, Chemical shielding tensor and ^{13}C–^{14}N dipolar splitting in single crystals of L-alanine. *J. Chem. Phys.* **74**, 3190–3197 (1981).
63. X. Chen, C.-G. Zhan, First-principles studies of C-13 NMR chemical shift tensors of amino acids in crystal state. *J. Mol. Struct.* **682**, 73–82 (2004).
64. A. Naito, S. Ganapathy, P. Raghunathan, C. A. Mcdowell, Determination of the ^{14}N quadrupole coupling tensor and the ^{13}C chemical shielding tensors in a single crystal of L-serine monohydrate. *J. Chem. Phys.* **79**, 4173–4182 (1983).
65. A. Naito, C. A. Mcdowell, Determination of the ^{14}N quadrupole coupling tensors and the ^{13}C chemical shielding tensors in a single crystal of *L*-asparagine monohydrate. *J. Chem. Phys.* **81**, 4795–4803 (1984).
66. N. Janes, S. Ganapathy, E. Oldfield, Carbon-13 chemical shielding tensors in L-threonine. *J. Magn. Reson.* **54**, 111–121 (1983).
67. M. H. Sherwood, J. C. Facelli, D. W. Alderman, D. M. Grant, C-13 chemical-shift tensors in polycyclic aromatic-compounds. 2. Single-crystal study of naphthalene. *J. Am. Chem. Soc.* **113**, 750–753 (1991).
68. R. J. Iuliucci, J. C. Facelli, D. W. Alderman, D. M. Grant, C-13 chemical-shift tensors in polycyclic aromatic-compounds. 5. Single-crystal study of acenaphthene. *J. Am. Chem. Soc.* **117**, 2336–2343 (1995).
69. J. K. Harper, R. Iuliucci, M. Gruber, K. Kalakewich, Refining crystal structures with experimental ^{13}C NMR shift tensors and lattice-including electronic structure methods. *CrstEngComm* **15**, 8693–8704 (2013).
70. A. Portieri, R. K. Harris, R. A. Fletton, R. W. Lancaster, T. L. Threlfall, Effects of polymorphic differences for sulfanilamide, as seen through ^{13}C and ^{15}N solid-state NMR, together with shielding calculations. *Magn. Reson. Chem.* **42**, 313–320 (2004).
71. D. Stueber, D. M. Grant, ^{13}C and ^{15}N chemical shift tensors in adenosine, guanosine dihydrate, 2'-deoxythymidine, and cytidine. *J. Am. Chem. Soc.* **124**, 10539–10551 (2002).
72. S. Sharif, D. Schagen, M. D. Toney, H. H. Limbach, Coupling of functional hydrogen bonds in pyridoxal-5'-phosphate–enzyme model systems observed by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 4440–4455 (2007).
73. Y. F. Wei, A. C. de Dios, A. E. McDermott, Solid-state ^{15}N NMR chemical shift anisotropy of histidines: Experimental and theoretical studies of hydrogen bonding. *J. Am. Chem. Soc.* **121**, 10389–10394 (1999).
74. C. Gervais, R. Dupree, K. J. Pike, C. Bonhomme, M. Profeta, C. J. Pickard, F. Mauri, Combined first-principles computational and experimental multinuclear solid-state NMR investigation of amino acids. *J. Phys. Chem. A* **109**, 6960–6969 (2005).
75. J. Z. Hu, J. C. Facelli, D. W. Alderman, R. J. Pugmire, D. M. Grant, ^{15}N chemical shift tensors in nucleic acid bases. *J. Am. Chem. Soc.* **120**, 9863–9869 (1998).
76. E. D. L. Smith, R. B. Hammond, M. J. Jones, K. J. Roberts, J. B. O. Mitchell, S. L. Price, R. K. Harris, D. C. Apperley, J. C. Cherryman, R. Docherty, The determination of the crystal structure of anhydrous theophylline by X-ray powder diffraction with a systematic search algorithm, lattice energy calculations, and ^{13}C and ^{15}N solid-state NMR: A question of polymorphism in a given unit cell. *J. Phys. Chem. B* **105**, 5818–5826 (2001).
77. K. Yamada, S. Dong, G. Wu, Solid-state ^{17}O NMR investigation of the carbonyl oxygen electric-field-gradient tensor and chemical shielding tensor in amides. *J. Am. Chem. Soc.* **122**, 11602–11609 (2000).
78. S. Dong, K. Yamada, G. Wu, Oxygen-17 nuclear magnetic resonance of organic solids. *Z. Naturforsch. A* **55**, 21–28 (2000).

Acknowledgment

Funding: This work was supported by Swiss National Science Foundation grant 200020_178860 and NCCR MARVEL. **Author contributions:** Conceptualization: M.Co. and L.E. Methodology: M.Co., M.B., M.Ce., and L.E. Investigation: M.Co., M.B., and B.S.d.A. Visualization: M.Co. and L.E. Supervision: L.E. Writing (original draft): M.Co. and writing (review and editing): M.Co., M.B., B.S.A., M.Ce., and LE. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data used in this study are freely available in the Materials Cloud repository (https://doi.org/10.24435/materialscloud:vp-ft). The code is available on GitHub (https://github.com/manucordova/ProbAsn) and is also archived in the Materials Cloud repository.

Submitted 30 June 2021
Accepted 8 October 2021
Published 26 November 2021
10.1126/sciadv.abk2341