CrystEngComm

Refining crystal structures using ^{13}C NMR chemical shift tensors as a target function

Journal:	*CrystEngComm*
Manuscript ID	CE-ART-07-2021-000960.R1
Article Type:	Paper
Date Submitted by the Author:	30-Aug-2021
Complete List of Authors:	Wang, Luther; Brigham Young University, Chemistry and Biochemistry Hartman, Joshua D; Mt. San Jacinto College, Chemistry Harper, James; Brigham Young University, Chemistry and Biochemistry

Refining crystal structures using ^{13}C NMR chemical shift tensors as a target function

Luther Wang, James K. Harper*

Brigham Young University, Department of Chemistry and Biochemistry, Provo UT 84602, USA.

Abstract

A two-step process is described for refining crystal structures from any source. This approach employs an initial lattice-including DFT relaxation step followed by a Monte Carlo sampling procedure to create new candidate positions for each atom within a structure. The candidate having the best agreement between experimental and computed ^{13}C NMR shift tensor principal values is selected for further refinement and the Monte Carlo process is repeated until convergence is achieved. This refinement can include all atoms within a structure or can be restricted to only poorly fit sites. This process is shown to improve ^{13}C NMR agreement from 6.1 ppm for a set of benchmark structures obtained from high quality diffraction data and not subjected to any refinement to 1.8 ppm after the two-step refinement. Although changes to atom positions from this refinement process are quite small (usually a few picometers), prior work is summarized to demonstrate that these changes can, in fact, yield new structural insights involving changes in hydrogen bonding, detection of hydrogen tunneling and new insights into protein backbone dynamics. Several non-NMR metrics, examined before and after refinement, indicate that this process does not introduce structural errors.

Introduction

The development of crystallography has profoundly influenced the development of science over the past century. This is partly because crystal structures provide a molecular picture with atom level resolution which allows interactions such as hydrogen bonding and π-π stacking to be more readily understood. Molecular conformation can also be visualized and such information has long been known to facilitate prediction of molecular function.[1] Of equal importance is the fact that the symmetry of the lattice is provided by these studies. This is significant because certain physical properties depend on the longer-range order and symmetry.[2] Over the past century, the vast majority of crystal structure determinations have employed x-ray diffraction using single crystal samples. A much smaller collection of structures has been established from neutron and electron diffraction data. These less common structures represent a valuable resource because they exhibit structural details less readily observed by x-ray methods (e.g. hydrogen positions). Complementing all of these techniques are powerful methods for solving structure from powder diffraction.[3] A challenge that arises from this rich variety of techniques is that structures obtained from these different methods vary considerably in quality, resolution and in the kinds of atoms characterized. For example, it is common for crystal structures of proteins to omit hydrogens, leaving hydrogen bond donor/acceptor designation ambiguous. Likewise, electron diffraction exhibits the greatest diffraction from hydrogens with increasingly weaker interaction as the atomic number increases.[4] This means that electron diffraction structures typically have lower resolution for non-hydrogen atoms than is observed in comparable x-ray diffraction studies. Indeed, in general, electron diffraction structures typically exhibit lower resolution that structure

established by other methods.[4] Another challenge is found in macromolecular structures where it is not uncommon for crystal structures to omit some atoms, including non-hydrogen atoms, for various reasons. Taken together, this variable resolution invites development of techniques for further refining crystal structures.

In recent years, several groups have explored theoretical refinement methods that include lattice fields to further refine crystal structures.[5,6,7,8,9,10] Solid-state NMR (SSNMR) has provided an important complement to these methods because it has been demonstrated that improvement in DFT predicted SSNMR parameters is correlated with greater structural accuracy.[10,11] It is important to note that in these studies, agreement between experimental and computed NMR data *is not explicitly used to guide the DFT refinement*. Rather, it is simply observed that such a lattice-including refinement is strongly correlated with improvement in the DFT calculated NMR parameters.

Several groups have sought to improve upon this tangential use of SSNMR experimental data in crystal structure refinement by including a direct comparison between computed and experimental SSNMR data to guide refinement. This use of a so-called "target function" has been employed using both semi-empirical and higher level theoretical methods. One of the early semi-empirical refinements employed computed shift tensor data from a set of training structures that were selected to be representative of the type of structures to be refined. In this study by Wylie et al.,[12] derivatives of amino acid were used as the training set with the goal of providing methods capable of refining large proteins. This approach is particularly appropriate for protein crystal structure refinement because proteins are too large to be refined using more rigorous lattice-including DFT methods. This refinement utilized only shift tensor information

from C_α carbons. This restriction appears to be necessary because tensors from C_α sites can be accurately calculated for all Ramachandran angles even when solvent or lattice effects are neglected, as was done in the training set. These NMR data were included in the refinement process by introducing a "pseudo-energy" term that measures the agreement between experimental and computed NMR data. This term was included together with more conventional energy terms as a constraint on the refinement process. This methodology has been used to refine the 6 kDa protein GB1.[12] The final refined structure was found to be comparable in quality to a 1 Å x-ray structure.

Another semi-empirical approach known as Bond Polarization Theory (BPT) has been developed by Sternberg and Prieβ.[13] This approach also employs a pseudo-energy term derived from the chemical shift tensor agreement. Because derivatives of chemical shifts can be analytically derived, pseudo-forces can also be obtained and are used in geometry optimizations. The BPT refinement is not restricted only to C_α-positions and can include a more general set of ^{13}C tensors. These refinements can also include ^{15}N amide tensor values. The BPT approach has been utilized to further refine the crystal structures of ubiquitin and gramicidin A.[14] A nearly 10-fold improvement in the agreement between calculated and experimental shift tensors was obtained by these refinements. Although the authors did not put this improvement in context of the degree of structural improvement, prior studies have consistently demonstrated that such improvement in NMR agreement corresponds to an improvement in structural accuracy.[10,11] It is notable that in the case of gramicidin A, the initial NMR based refinement yielded a structure having a single poorly fit valine. The outlier suggested the need for further refinement. A second refinement resulted in unusually good

agreement between experimental and computed ^{15}N shift tensor data and created a structure with significant structural differences from the structure originally reported. Both of the semi-empirical methods described above are best suited to the refinement of protein crystal structures and rely on computed shift tensors predicted in the gas phase.

Higher level approaches for refining crystal structures using SSNMR data as a target function have been proposed by Brouwer[15] and independently by Perras and Bryce.[16] These methods explicitly include lattice fields by using either periodic boundary methods or clusters of molecules. In both cases, new atomic positions are generated for the atoms of interest by sampling the region surrounding selected atoms to create new candidate structures. The initial atom positions are taken from either the x-ray structure[16] or by SSNMR distance constraints from selected atoms.[15] NMR parameters are computed for all new candidate structures using DFT methods and the refinement involves minimizing agreement between experimental and computed data. This approach has been used to adjust the crystal structures of the zeolite Sigma-2 with ^{29}Si tensors[15] and the inorganic structures $Na_2Al_2B_2O_7$, $Na_4P_2O_7$ and $Na_3HP_2O_7{\cdot}H_2O$ using various combinations of ^{31}P shift tensors and/or ^{23}Na, ^{27}Al, ^{17}O and ^{11}B EFG tensors.[16,17] In the case of the zeolite, a comparison to the crystal structure from a single crystal study[18] revealed that the SSNMR refined structure was highly accurate with a mean deviation of only 0.015 Å in atom positions. This level of structural adjustment is particularly noteworthy since it is below the diffraction limit for the Mo Kα radiation employed (0.71073 Å). This means that the structural changes made would not be detectable using x-ray methods but are significant enough to alter the ^{29}Si NMR data to an extent that they were statistically distinguishable. Some of these methods use a constraint similar to the pseudo energy term described above but

differ in that all calculations involved the actual compound of interest rather than relying on an external training set. In addition, a higher level of theory was employed for computing NMR data in both approaches. Each of these methods weight the NMR agreement more heavily than do the semi-empirical approaches. In the case of Bryce et al., an energy term was included to aid in the refinement, but it was found that the SSNMR agreement was the dominant factor and inclusion of energy made a negligible contribution. This dominant influence of SSNMR in guiding the refinement was also observed in the refinement of gramicidin A described above.

A more recent DFT study seeking to include SSNMR data in crystal structure refinement has been reported by Holmes et al.[19] In this work, the dispersion correction term used in the lattice-including geometry refinement step and was empirically adjusted to provide the best agreement between computed and experimental SSNMR parameters. This adjustment significantly improved predicted NMR parameters and the methodology was dubbed "DFT-D2*". At the present time, this approach is, arguably, the most accurate general NMR constrained method for refining crystal structures. Nevertheless, the method proposed represents an *indirect* SSNMR constraint on a geometry refinement rather than being used as a so-called "target function" in which a *direct* comparison to NMR data is employed to guide a refinement.

In the present study, an alternative SSNMR refinement procedure is described. This approach relies on the DFT-D2* of Holmes et al. to initially adjust the diffraction structure. Subsequent refinement follows the approach of Perras and Bryce by make minor adjustments in atom positions. Despite these similarities our approach differs from those described above in three ways as described hereinafter.

Experimental

All experimental NMR shift tensors data was taken from previous studies.[26,27,28,32] These data were selected due to their unusually high accuracy. Although several of these studies reported the six tensor values from the complete tensor, only the three principal values of the diagonalized tensor were employed in this study. This choice reflects the fact that principal values data are more widely availability and thus more likely to be employed in refinement studies. All comparisons of experimental and computed shift tensors involved first converting the tensors data into the icosahedral representation[20] where a more accurate comparison is obtained. The use of three principal values instead the six-component tensor requires a slightly different treatment in the icosahedral representation and a detailed discussion of this point is provided in reference 20. It is notable that the use of principal values requires that one assume that the experimental and computed tensors share the same principal axis system. Since this will only be true for the correct structure (and those lying within the error of this structure), it will be a very poor assumption for many of the candidate structures. In practice, this assumption results in an underestimation of the error in some of the candidate structures. In other words, some structures are erroneously retained in each iteration. However, this tendency to being too cautious when eliminating incorrect structures was considered preferable to the alternative of incorrectly eliminating valid structures. In practice, these erroneously retained structures were never found to have the best agreement with experimental data and thus this ambiguity does not compromise the refinement process.

All DFT-D2* and DFT-D2*/Monte Carlo calculations were performed using periodic boundary methods provided by the program CASTEP.[21] Initial atomic coordinates were taken

from diffraction studies reported elsewhere. The DFT-D2* geometry optimizations employed the RPBE functional with ultrasoft pseudopotentials. In all cases the ultra-fine level was selected and unit cell dimensions were not refined. The geometry optimizations used a relaxation process described elsewhere.[22] Refinements were performed using plane-wave basis set cutoff energies of 550 eV and k-point spacing of 0.05 $Å^{-1}$. Optimizations were considered to be converged when, the change in energy was < $5x10^{-6}$ eV per atom, the maximum force on all atoms was < 0.01 eV $Å^{-1}$, and the maximum displacement was < $5x10^{-4}$ Å for each atom. Dispersion effects were included in the geometry optimization using the D2 method of Grimme[23] with used a non-standard damping value of 5.0. The computation of ^{13}C shielding tensors used the GIPAW[5] approach available in CASTEP together with the RPBE functional and other parameters described above.

The DFT-D2*/Monte Carlo refinement was performed using coordinates created from the DFT-D2* geometry refinement as initial values. The Monte Carlo sampling of atomic positions was performed on the crystallographic asymmetric unit and the symmetry operations were then applied to create the unit cell. In each step 100-200 new structures were created and each was evaluated by computing NMR shift tensors as described above. The quality of each structure was assessed by plotting experimental versus computed shifts and selecting the structure with the highest correlation coefficient that also had a slope near the ideal value of -1.0. This best-fit candidate structure was then selected and its coordinates iteratively subjected to this process until no further improvement could be achieved. This converged structure was retained as the final NMR refined structure. Coordinates for all structures are provided as Supporting Information.

The Python code performing the Monte Carlo sampling procedure is available at http://github.com/starlwe/NMR_Refinement.

Results and discussion

A two-step refinement strategy

The aim of the current work was to utilize an approach that employs lattice-including DFT computational methods rather than semi-empirical calculations. This choice ensures that the methodology will not be restricted to certain classes of materials and will allow highly accurately NMR prediction methods to be included. The proposed methodology involves a two-step process with the first step employing the DFT-D2* refinement process of Holmes et al.[19] with the damping parameter set to 5.0. This lattice-including refinement is a quasi-Newton approach and includes an empirical adjustment to the dispersion term based on optimizing the agreement between computed and experimental NMR data. This step is postulated to create a structure that lies near the global minimum of the time and ensemble averaged structure for most atoms in a structure. A second refinement step is then proposed to further adjust the DFT-D2* coordinates. This step involves generating new atom positions through a Monte Carlo sampling process, then calculating NMR parameters for each new atomic arrangement. Because an initial refinement has already been completed, deviation from atomic positions are usually small (e.g. a few picometers), but large deviations are not prohibited. Positions giving the best agreement between computed and experimental NMR data define the new atom positions and this second step is repeated until no further improvement in NMR agreement can be achieved. In a general sense, this final refinement step mimics that described previously by

Perras and Bryce,[16] however, three modifications are introduced that improve efficiency and guarantee identification of the global best-fit. In addition, the NMR fit is the first to employ ^{13}C shift tensors and this is an advantage because carbons are densely represented throughout the molecule. This high concentration of NMR active atomic sites improves the probability that all regions of a molecule will be adequately refined due to the proximity of most atoms to one or more ^{13}C sites.

The first modification over prior atom selection methods involves how new positions are generated. Because prior studies involved adjustment of only a few atoms, manual adjustments were feasible. However, when treating larger structures this process soon becomes unwieldy and it was necessary to create an automated process for generating these positions. For this process, a program was created that performs Monte Carlo sampling around each atom in a spherical shape and having uniform point density. This program allows different atom types to be treated separately. For example, hydrogen atoms usually have a larger uncertainty than non-hydrogen sites in structures obtained from x-ray diffraction. Thus, it is reasonable to sample a larger sphere diameter when selecting candidate sites for hydrogen atoms. Likewise, it has been demonstrated that the thermal ellipsoids (i.e. anisotropic displacement parameters, ADPs) from NMR determined structures usually are smaller than those obtained from diffraction studies.[24,25] This suggests that using the ADP magnitudes from published diffraction structures to estimate the sampling diameter of the sphere will usually retain all valid points and thus can provide a reasonable starting point. When such ADPs are not available for a crystal structure of interest or for a specific atom, representative ADPs from

other crystal structures may be utilized as upper bounds. The Python program used to generate new atom positions is provided at http://github.com/starlwe/NMR_Refinement.

The second difference from prior work is that *each* of the individual candidate structures created by the Monte Carlo sampling is evaluated for agreement of experimental and computed shifts. This differs from procedures that attempt to select an optimal path that minimizes the number of calculations needed (e. g. steepest descent). This choice recognizes that such a path selection process has already occurred in the quasi-Newton DFT-D2* initial refinement. By performing a simple grid search as the final refinement step, the probability of locating the global minimum in coordinates is greatly improved. This choice significantly increases the number of structures that must be considered, but for this preliminary study the ability to guarantee location of the global minimum for benchmark data is a high priority. Addition of a step that selects a path direction can be added in subsequent studies if desired.

A third difference between the proposed approach and those described previously is that lattice energy considerations were often employed in prior studies together with NMR fit. In this study, the consideration of lattice energy is omitted from the final refinement step. This omission was based on results from prior work where it was observed that inclusion of energy made a negligible effect in selecting the best fit structure. For example, in the refinement of $Na_2Al_2B_2O_7$, the inclusion of an energy term in the NMR refinement process created a structure that was statistically indistinguishable from the one created only from NMR data.[17] Thus, the NMR data appears to dominate the refinement. A similar observation was made in Sternberg and Witter's refinement of the protein ubiquitin.[14] In our proposed methodology, the omission of energy is further justified because a DFT-D2* initial "rough" refinement step has already

been performed with the aim of locating an approximate global minimum and this step considers energy. It is also notable that in previous studies where energy was included as a refinement criterion, it was necessary for the user to scale how much the energy term was "turned on". Thus, in most cases energy considerations were used as the dominant factor in early refinement steps and the NMR term was only scaled up in later steps. Thus, even when energy and NMR are both included, they are not equally weighted in all steps. For the reasons outline above, we have elected to rely solely on NMR agreement as a convergence criterion in the final refinement step.

Refinement of benchmark organic structures with ^{13}C tensors

A group of five organic structures were selected to evaluated the proposed refinement methodology. The structures included sucrose, methyl β–D-xylopyranoside, methyl α-D-glucopyranoside, naphthylene and acetaminophen (Figure 1). These compounds provide 114 shift tensor principal values and cover a shift range of roughly 250 ppm. The majority of the experimental NMR data were obtained from large single crystals and were selected because such studies are highly accurate with estimated experimental errors of less than 0.7 ppm.[26,27,28] Of equal importance, both aromatic and aliphatic sites are represented in the dataset. This is important because there has been debate on the ability of a single approach to accurately model both sp^3 and sp^2 carbons.[29,30,31] A summary of the source of the diffraction and NMR data together with other relevant parameters is given in Table 1.

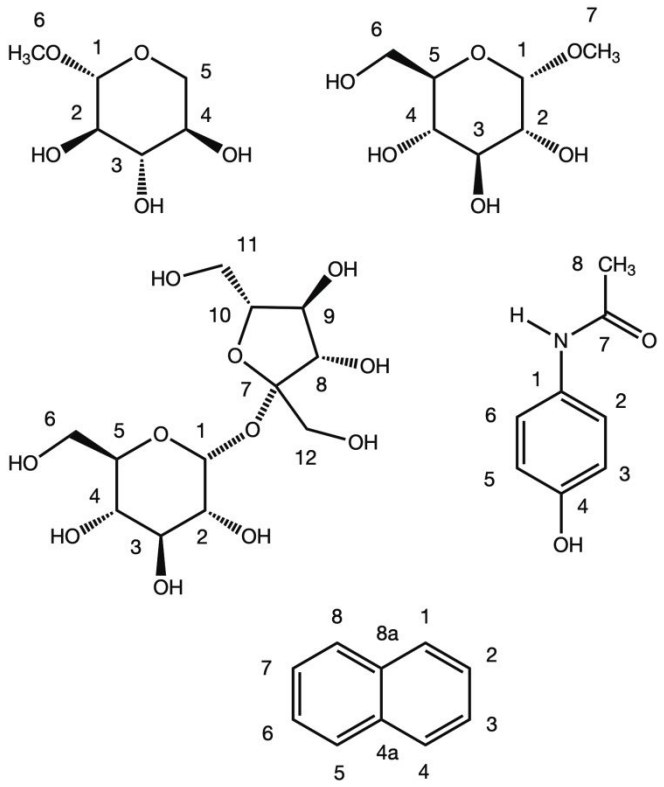

Figure 1. Structures of benchmark compounds showing numbering. At top are Methyl β-D-xylopyranoside (left) and Methyl α-D-glucopyranoside (right). Middle row includes sucrose (left) and acetaminophen (right) with naphthalene at bottom.

Table 1. The benchmark structures evaluated herein and relevant figures-of-merit for the diffraction and NMR data.

Structure	Diffraction data type[a]	R-value (%)	NMR method	Exp. NMR Error (ppm)
Sucrose	Neutron	3.3	Single crystal	0.29[26]
Methyl α-D-glucopyranoside	Neutron	4.5	Single crystal	0.27[27]
Methyl β-D-xylopyranoside	Neutron	4.5	Single crystal	0.71[27]
Naphthalene	X-ray	3.8	Single crystal	0.54[28]
Acetaminophen	X-ray	7.2	powder[b]	0.9[32,c]

[a]All diffraction studies employed single crystals.
[b]Date acquisition employed the FIREMAT technique.
[c]Estimated error from similar analyses.[30]

The error in computed tensors before any refinement of the five model diffraction structures was 6.1 ppm. Refinement of the diffraction structures using the two-step process described above decreased the uncertainty to 3.1 ppm, a value nearly identical to the error reported recently from a DFT-D2* refinement study using the same functional, but having a larger and more diverse group of model structures.[31] The final Monte Carlo refinement step further reduced the error to 1.8 ppm. An F-test indicates that the final DFT-D2*/Monte Carlo structure has an NMR agreement that is statistically different from those arising from DFT-D2* at the $p > 0.00001$ level. Thus, from the viewpoint of the NMR uncertainty, the DFT-D2* structure can be said to be differ by more than the error from the DFT-D2*/Monte Carlo coordinates. Plots showing the agreement between experimental and computed ^{13}C tensor values are given in Figure 2. A slight improvement is also observed in the R^2-value with a value of 0.998 observed after Monte Carlo refinement versus 0.994 from the DFT-D2* step. A summary of all metrics from these plots is provided in Table 2.

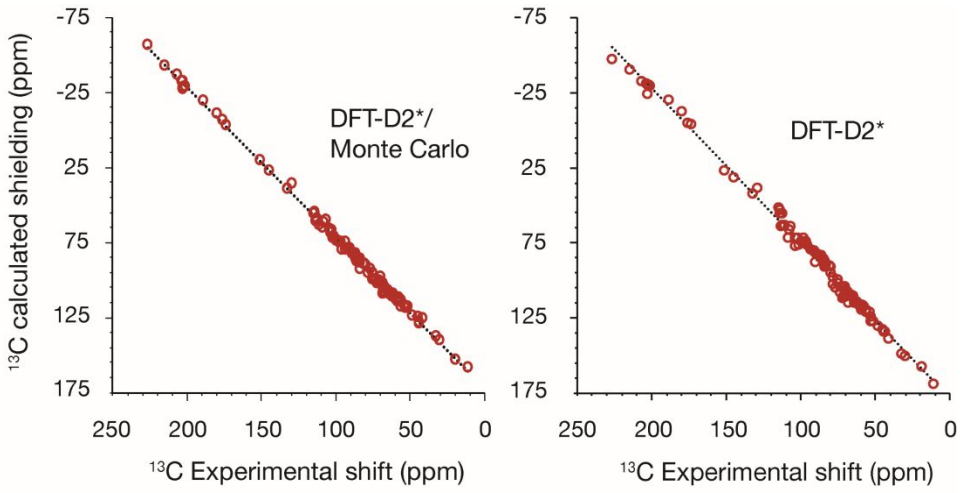

Figure 2. A comparison of the 114 computed and experimental shift tensor values for the five benchmark structures. Agreement between calculated shifts and experimental data improves by 42 % after DFT-D2*/Monte Carlo refinement.

Table 2. The errors and other metrics from a least-squares fitting of ^{13}C shift tensor data.

Treatment	Error (ppm)	Slope	Intercept	R^2
No refinement	6.1	-1.090	185.33	0.9834
DFT-D2*[,a]	3.1	-1.038	179.03	0.9941
DFT-D2*/Monte Carlo[a]	1.8	-1.004	172.08	0.9981

[a]All DFT-D2* refinements employed a non-standard damping value of 5.0

The comparison described above, provides the *overall* agreement between computed and experimental tensor principal values. It is also important to evaluate the agreement for *individual* compounds to identify any structural features or functional groups that are not well treated by this approach. Accordingly, agreement for each of the benchmark compounds was computed and is given in Table 3. While most of the computed data is highly accurate with errors of 2.0 ppm or less, the error for acetaminophen is significantly larger. This is likely due to the fact the nitrogen is present as ^{14}N with a 99.6% natural abundance and will thus experience dipolar coupling to nearby ^{13}C sites with particularly strong coupling to C1 and C7 which are directly bonded to ^{14}N. Weaker dipole coupling will also be influence C2, C6 and C8, with

sideband patterns for these sites including contributions from the shift tensor and dipolar coupling. Since there was no attempt to fit these dipolar couplings, a larger error at these site is to be expected. The fact that the DFT-D2*/Monte Carlo adjustment did not impose unreasonable distortions in an attempt to compensate for this complication, indicates that overfitting of the NMR data is not occurring.

Table 3. The errors and other metrics from a least-squares fitting of ^{13}C shift tensor data.

	NMR Error (ppm)		
Structure[a]	Unrefined	DFT-D2*	DFT-D2*/ Monte Carlo
Sucrose	4.7	2.7	2.0
Methyl α-D-glucopyranoside	4.2	2.9	1.2
Methyl β-D-xylopyranoside	4.6	3.1	1.2
Naphthalene	8.4	3.4	1.8
Acetaminophen	8.8	4.6	2.9

In each step of the refinement, an improvement was observed in the NMR agreement for all structures considered herein. Numerous prior studies have also postulated that such improvements are evidence of structural improvement. In support of this conclusion, these studies have included several types of non-NMR evidence to verify that no structural errors were introduced by the refinement. To provide consistency with these earlier studies, other metrics are evaluated here to verify that the DFT-D2*/Monte Carlo refinements do not create structural errors.

Other metrics supporting structural improvement

The outcome of the Monte Carlo refinement requires thoughtful evaluation because improvement is presumed based solely on improvement of the NMR agreement. Consideration of additional figures-of-merit is therefore important to ensure that structural errors have not been introduced by the proposed refinement. Here we consider, magnitude of movement of atoms from their original diffraction positions, changes in bond lengths, variations in the predicted powder diffraction patterns and changes in energy from the refinement.

A comparison of the structures obtained from diffraction data versus those ultimately obtained from the two-step refinement described above shows only minor differences in atom positions. An overlay illustrating the differences is given in Figure 3. A more quantitative comparison is provided in Table 4 and shows that although some adjustment occurred at most sites, the amount of movement is negligible with rms deviations ranging from 0.028–0.146 Å in heavy atoms. In order to place this observation in context, we note that prior studies have found that when the same crystalline phase has been independently solved by multiple researchers under similar conditions (e.g. temperature), the average difference in atom positions between the two structures lie in the range of 0.01 to 0.1 Å.[33,34,35] Thus, the changes created by the two-step refinement are comparable to the error observed between two high quality single crystal diffraction studies performed on the same structure.

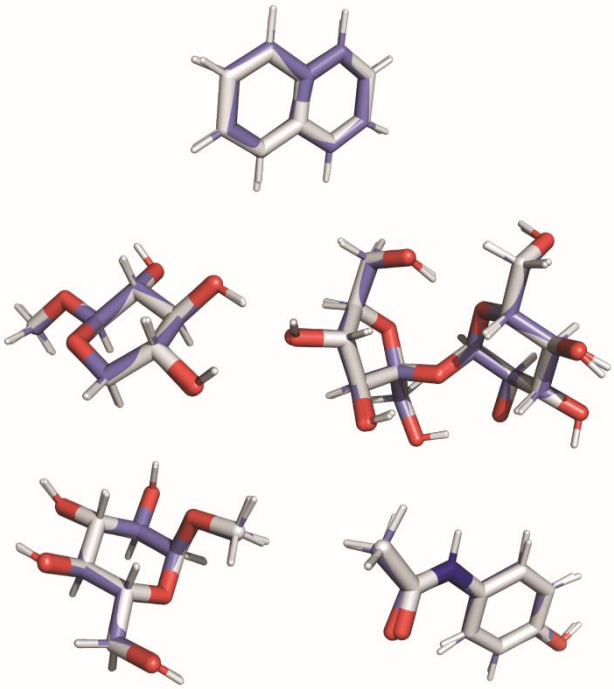

Figure 3. Superimposed structures of all benchmark compounds showing the original diffraction coordinates (grey) and coordinates after DFT-D2*/Monte Carlo refinement (purple).

Table 4. Magnitude of the change in the mean atom positions before and after DFT-D2*/Monte Carlo refinement.

	rms deviations (Å)	
Structure	Non-hydrogen	All atoms
Naphthalene	0.068	0.158
Sucrose	0.057	0.089
Xylopyranoside	0.048	0.071
Glucopyranoside	0.037	0.065
acetaminophen	0.072	0.180

Another metric that has been useful in assessing the influence of refinement is changes to bond lengths. Prior work has reported that DFT-D2* refinement invariably decreases bond lengths involving two non-hydrogen atoms.[19] An evaluation of bond lengths in the benchmark compounds after our two-step refinement process finds that in all cases, bond lengths between non-hydrogen atoms also decreased, albeit by smaller amounts in most cases than was

observed from DFT-D2* alone. The difference in bond lengths for C–C bonds before and after DFT-D2*/Monte Carlo refinement was smaller than the estimated error reported in the diffraction study. For C–O bonds the adjustment was larger than the estimated error in the diffraction data and the NMR constrained bond lengths were statistically indistinguishable from those obtained from DFT –D2*. A comparison of the average change in bond length from each refinement process is illustrated in Figure 4.

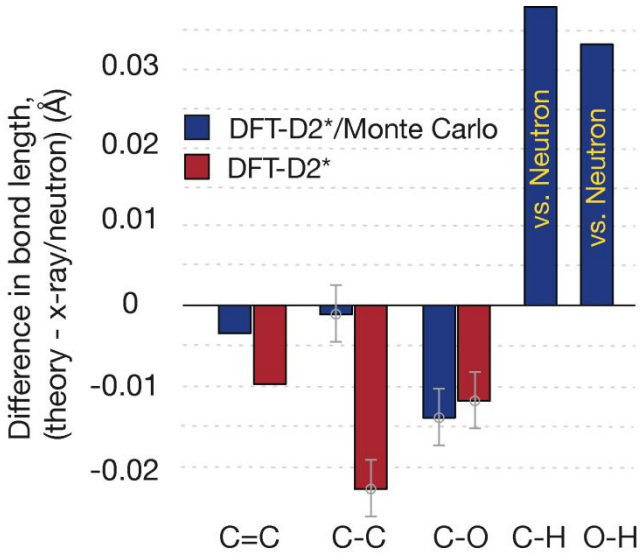

Figure 4. A comparison of average bond length changes from DFT-D2* (red) and DFT-D2*/Monte Carlo (blue). Where error bars are indicated, they represent the reported error from a diffraction structure containing the bond type. Bonds involving hydrogen include only comparison to bond lengths from neutron diffraction studies since these are experimentally determined.

A second comparison of bond lengths was made involving X–H bonds (X = C or O). This analysis differs from prior work[19] by only including data from neutron diffraction studies involving the carbohydrates. This was done because the reported bond lengths from x-ray diffraction differ significantly in how they are determined and include cases where bond lengths

are simply fixing at predetermined values. In contrast, bond lengths from neutron diffraction data are experimentally determined and presumed to be of comparable accuracy to non-hydrogen atoms. Prior DFT-D2* work has reported that X–H bonds invariably increase in length.[19] This is likely due to their strong reliance on x-ray diffraction data where X–H lengths are typically underestimated. Somewhat surprisingly, DFT-D2*/Monte Carlo adjustment to X–H bond lengths produced a much larger change than that observed for bonds between non-hydrogen atoms. Specifically, average X–H bond length increased by 0.034 to 0.039 Å while the absolute change to bond lengths involving non-hydrogen atoms ranged from 0.001 to 0.013 Å. Despite these large differences, the adjustments to X–H bonds in the present study are still 3–4 times smaller than those created by DFT-D2*. Overall, it can be concluded that the DFT-D2*/Monte Carlo adjustments to bond lengths are typically smaller than those obtained from DFT-D2* alone, but occur in the expected directions (i.e. increase or decrease).

A summary of all bond lengths before and after DFT-D2*/Monte Carlo refinement together with the standard deviation is given in Table 5. One notable feature of the refined bond lengths is that the variation in bond lengths from the DFT-D2*/Monte Carlo process is two to three time larger than the variation observed in the original diffraction structures. The largest and smallest bond lengths found for each data set are included in Table 5. Typical uncertainty in experimental diffraction data have been reported to be ± 0.004 for C–O, and C–C,[10] thus the refined C–C bond lengths lie within the diffraction error while the refined C–O bond lengths can be said to deviate from diffraction values. In contrast, based on the improvement in the NMR agreement, the refined structures can each be classified as statistically distinguishable structures.

Table 5. Bond lengths (Å) obtained from diffraction, DFT-D2* and DFT-D2*/Monte Carlo refinement.

Bond type	Source	Average	St. dev.	Max.	Min.
C–C	Diffraction	1.523	0.008	1.540	1.508
	DFT-D2*	1.521	0.010	1.539	1.498
	DFT-D2*/M.C.	1.522	0.026	1.578	1.483
C–O	Diffraction	1.418	0.012	1.445	1.381
	DFT-D2*	1.424	0.013	1.447	1.383
	DFT-D2*/M.C.	1.405	0.024	1.449	1.359
C=C	Diffraction	1.397	0.018	1.425	1.375
	DFT-D2*	1.396	0.015	1.426	1.372
	DFT-D2*/M.C.	1.394	0.048	1.472	1.335
C–H[a]	Diffraction	1.087	0.020	1.108	1.026
	DFT-D2*	1.078	0.003	1.085	1.071
	DFT-D2*/M.C.	1.126	0.035	1.208	1.049
O–H[a]	Diffraction	0.962	0.020	0.985	0.912
	DFT-D2*	0.967	0.006	0.976	0.955
	DFT-D2*/M.C.	0.996	0.037	1.051	0.916

[a]Includes only bond lengths from the three carbohydrate structures where neutron diffraction data were reported.

Another measurable that can be evaluated to assess the influence of the refinement process is differences in the powder diffraction patterns. To make this comparison, the difference between the powder pattern from the original single crystal diffraction study was subtracted from the comparable simulated pattern obtained from the DFT-D2*/Monte Carlo process. This difference is referred to as the “residual” and in the case of no change from the refinement results in a flat line (i.e. zero at all points). Herein residuals are defined as the peaks intensities of the experimental data minus that those obtained from DFT-D2*/Mote Carlo refinement. Such a comparison for each of the five model structures is illustrated in Figure 5.

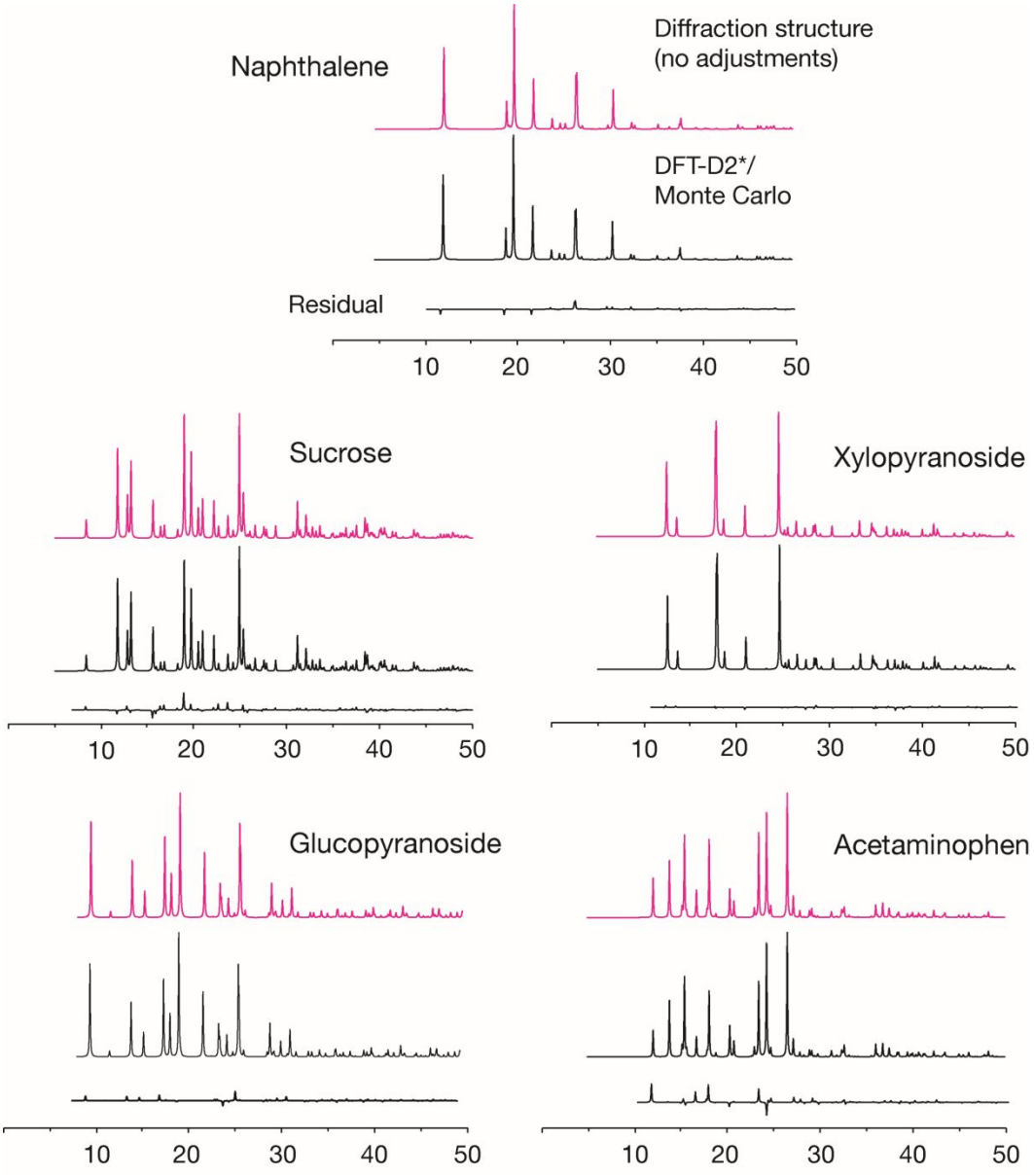

Figure 5. A comparison of the simulated powder diffraction patterns from the original diffraction study before any relaxation (red) and from the DFT-D2*/Monte Carlo refinement (black). At the bottom of each plot the residuals are shown to provide a more quantitative comparison.

Figure 5 shows that in all cases, the DFT-D2* refinement creates almost no discernable change in the simulated powder pattern. It can thus be concluded that the powder patterns

support the conclusion that no structural errors have been introduced by our two-step refinement process.

A final evaluation compared the energy of the final DFT-D2*/Monte Carlo structure against energies of both the unrefined diffraction structure and the DFT-D2* relaxed coordinates. In all cases the DFT-D2* refinement of the diffraction structure decreased energy, sometimes by as much as - 0.25 % of the total energy. The subsequent DFT-D2*/Monte Carlo adjustment caused the energy to increase, albeit by less than + 0.06% of the total energy for each structure. This tendency of the NMR refinement to slightly increase energy has been previously observed[14] and has been explained as a process of pushing the system into a new local minimum within the crystal. Figure 6 shows a plot of the energy changes for all structures.

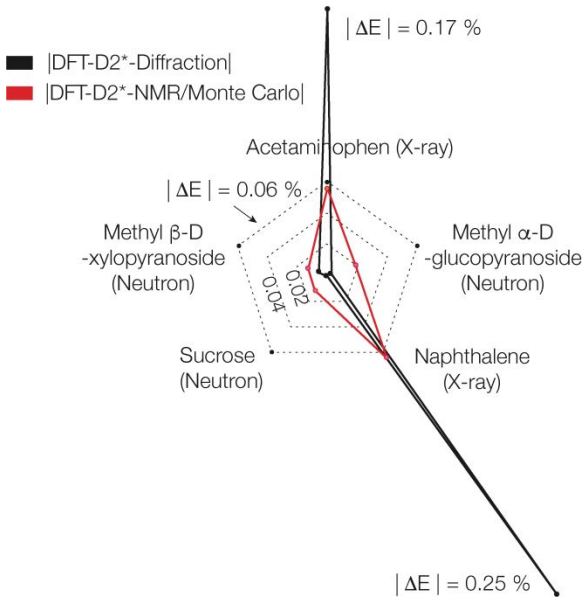

Figure 6. A comparison of the energy changes from the DFT-D2* refinement (black) and the DFT-D2*/Monte Carlo adjustment (red). In all cases the energy changes from the DFT-D2*/Monte Carlo step lie between the energies of structures obtained from x-ray and neutron diffraction with the closest similarity to energies of unrefined neutron diffraction structures.

It is notable that the changes in energy from the DFT-D2* refinement of the diffraction coordinates vary greatly with large changes observed when the initial structure is derived from x-ray diffraction data and much smaller changes found when the structure comes from neutron diffraction. This variation likely reflects the difference in X–H bond lengths (X = C, O or N) found in these two data types because it is known that x-ray and neutron diffraction provide nearly indistinguishable non-hydrogen position.[10] Since the Monte Carlo step creates X–H bond lengths quite similar to those from neutron diffraction, it may be expected that the energies are similar to the neutron diffraction structures. Indeed, this similarity of Monte Carlo refined and neutron structures can be observed in Figure 6. Overall, the refinement process creates structures that all lie intermediate between the energies of the unrefined x-ray and unrefined neutron diffraction structures. In cases where a direct comparison can be made, the Monte Carlo refined structures lie within + 0.02% of the initial neutron diffraction structures.

For the structures studied here, final energies from the Monte Carlo refinement are consistently higher than those obtained from DFT-D2* refinement. We postulate that at least some of this difference come from that fact that unit cell parameters were not varied. Because the Monte Carlo refinement consistently creates X–H bond lengths that are longer than those from DFT-D2*, hydrogen rich structures will occupy a large molecular volume after the Monte Carlo process. Accordingly, unit cell dimensions may need to be adjusted to accommodate this increase. Variations in unit cell parameters were not investigated here and represent a path for future study.

Are meaningful structural changes obtained from refinements with NMR as a target function?

An interesting fact about most NMR constrained refinements that have now been reported is that they usually result in only small changes to atom positions. In most cases the refined positions differ from those obtained from diffraction by less than ± 0.1 Å. Accordingly, it is justifiable to ask if these refinements have the ability to provide new structural insights or if they are simply done to obtain improved agreement between experimental and computed NMR parameters. Despite the relatively small number of refinements that have now been reported using NMR as a target function, a few studies do, in fact, describe meaningful structure changes. At the present time, three types of changes have been described and each of these is summarized here.

The first kind of structural change involves the observation that some NMR constrained refinements create new hydrogen bonding arrangements that are more consistent with the NMR data while preserving all the general features of diffraction experiments. The first such refinement involved a correction to the crystal structure of cellulose I_{α}.[36] The published neutron diffraction study was unable to distinguish between two possible hydrogen bonding arrangements and speculated that a dynamic interchange between the two candidates may be present. A ^{13}C chemical shift constrained geometry optimization identified only one of these arrangements as being consistent with NMR shifts (see Figure 7). This refinement also demonstrated that the dynamic interchange between the two hydrogen bonding arrangements was infeasible.

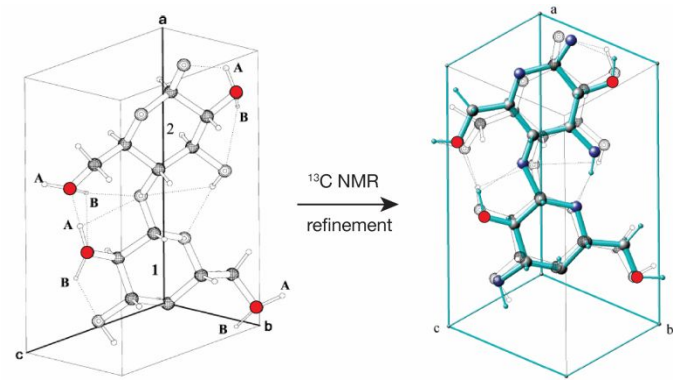

Figure 7. A comparison of the crystal structure of cellulose I_α from diffraction measurements (left) and from a NMR guided refinement (right, blue-green bonds). The NMR refinement clarifies hydrogen bonding at the four oxygen sites shown in red. Adapted with permission from Witter, R.; Sternberg, U.; Hesse, S.; Kondo, T.; Koch, F. –T.; Ulrich, A. S.; Macromolecules, 2006, 39, 6125 – 6132. Copyright 2006 American Chemical Society.

In a related NMR refinement study, ^{15}N chemical shift tensors were used as a target function in the semi-empirical refinement of the protein ubiquitin.[14] Although the overall changes in the backbone geometry was small (i. e. $\pm$ 0.07 Å), seven new hydrogen bonds were identified. This observation is particularly relevant because hydrogen bonds are one of the primary forces driving protein folding.[37]

A final example of an NMR study altering a hydrogen bonding arrangement was a recent study of the pharmaceutical furosemide.[38] In this analysis, the consideration of an alternative hydrogen bonding arrangement at the COOH in furosemide significantly improved agreement between computed and experimental ^{1}H shifts at the COOH. This information was employed to distinguish two phases of the microcrystalline solid that that differed only in the placement of a single hydrogen, with all other atoms having an rms difference of only 0.015 Å. Taken together these three studies demonstrate that NMR based refinements can significantly alter hydrogen bonding schemes in molecules. This change can even be observed in structures that are considered to be well refined.

A second type of consequential structural adjustment from an NMR based refinement, involves the adjustment of COOH hydrogen positions in certain *n*-alkyl fatty acids. In these ^{13}C studies lauric[39] and palmitic acid[25] were evaluated, with Monte Carlo sampling employed to locate COOH hydrogens. It was reported that this refinement was able to distinguish COOH hydrogens that experience tunneling between hydrogen bonded sites from those COOH sites that have localized hydrogens. Tunneling hydrogens were found to have a characteristic agreement between computed and experimental NMR data that resembles a double well minimum analogous to those found[40] from energy considerations. Those COOH sites that experience tunneling were also shown to have markedly different $^{13}COOH$ shift tensors than COOH sites with a localized hydrogen. This tunneling was difficult or impossible to characterize by traditional diffraction techniques.[41] Because tunneling creates stronger hydrogen bonds, this study provides a practical way to identify unexpectedly strong hydrogen bonds based on a simple refinement involving primarily the COOH hydrogen. The authors of these studies used the results to explain anomalous melting behavior known for over 140 years[42] to occur in *n*-alkyl fatty acids.

A third area where NMR constrained refinements have been found to yield new structural insight is in describing molecular dynamics. In a 2019 study, the protein ubiquitin was refined using ^{15}N chemical shift tensors from backbone amides as constraints.[14] Although a 1998 study proposed a group of feasible motions in ubiquitin that were consistent with a ^{13}C and ^{15}N relaxation measurements,[43] the authors demonstrated that only two of these motional modes were consistent with the structure refined using ^{15}N constraints.

Taken together, the results from these previous studies on structural refinement indicate that although NMR based refinements usually involved small structural changes, in several cases they have been demonstrated to yield new structural insights while remaining consistent with prior diffraction studies.

Evaluating the choice of functional on the NMR refinement

All work described herein was all performed using the RPBE functional and it is justifiable to ask what structural differences would be obtained if another functional were employed. It is likely that another functional would select a different structure having the best agreement at each step of the process. However, for shift tensors, those functionals providing the best agreement to experimental data are known.[44] We therefore anticipate that functionals having similar uncertainties in benchmark studies will create structures that lie very close to one another. To test this assumption, the crystal structure of naphthalene was refined using PW91, a functional that is reported to have uncertainties similar to those from PBE.[10] A comparison of the structure refined using PW91 against that obtained from RPBE showed rms differences in non-hydrogen atom positions of 0.006 Å and of 0.010 Å when all atoms are compared. A visual comparison of these differences is given in Figure 8 showing the entire unit cell. These changes are 6–12 times smaller than those reported in Table 4 support the conclusion that functionals of comparable accuracy will provide similar NMR refined structures. Recently, methods have recently been proposed for quantifying such SSNMR defined errors in atom positions in terms of the so-called anisotropic displacement parameters (i.e. thermal ellipsoids).[24,25] We posit the difference between PW91 and RPBE lies within the uncertainty of

the NMR refinements for all atoms. Admittedly, this result is only for a single compound and further work is needed to verify that this result can be achieved in a more diverse group of structures.

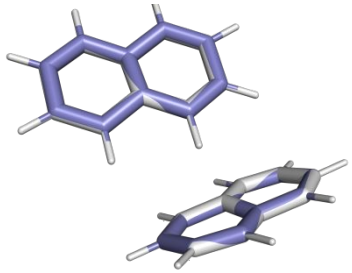

Figure 8. A comparison of atom positions within the unit cell of naphthalene after an NMR refinement of the crystal structure using the functional RPBE (purple) versus PW91 (grey).\

Conclusions

This study describes a two-step DFT-D2*/Monte Carlo refinement process that improves agreement between experimental and computed SSNMR ^{13}C shift tensors by a factor of 3.4. The structural changes are small with average changes to atom positions of less than 0.08 Å and modifications to bond lengths of less than 0.01 Å. Although these changes are near or below the diffraction limit for the radiation employed, the improvement in the NMR agreement is statistically significant and results in structures that are distinguishable from the DFT-D2* coordinates. Other metrics have been evaluated and support the conclusion that structural errors have not been introduced by the refinement.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the National Science Foundation under CHE-2016185 to J.K.H.

Footnotes

1. O. M. Becker, Y. Levy and O. Ravitz, *J. Phys. Chem. B*, 2000, **104**, 2123–2135.

2. S. Bhagavantam, *Crystal symmetry and physical properties*, Academic Press, London, 1966.

3. W. I. F. David, K. Shankland, L. M. McCusker and C. Baerlocher, *Structure Determination from Powder Diffraction Data*, Oxford, New York, 2002.

4. D. L. Dorset, *Structural electron Crystallography*, Plenum press, New York, 1995.

5. C. J. Pickard and F. Mauri, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2001, **63**, 245101.

6. S. E. Ashbrook, M. Cutajar, C. J. Pickard, R. I. Walton and S. Wimperis, *Phys. Chem. Chem. Phys.*, 2008, **10**, 5754–5764.

7. D. H. Brouwer, I. L. Moudrakovski, R. J. Darton and R. E. Morris, *Magn. Reson. Chem.*, 2010, **48**, S113–S121.

8. E. Salager, R. S. Stein, C. J. Pickard, B. Elena and L. Emsley, *Phys. Chem. Chem. Phys.*, 2009, **11**, 2610–2621.

9. R. A. Olsen, J. Struppe, D. W. Elliott, R. S. Thomas and L. J. Mueller, *J. Am. Chem. Soc.*, 2003, **125**, 11784–11785.

10. J. K. Harper, R. Iuliucci, M. Gruber and K. Kalakewich, *CrystEngComm*, 2013, **15**, 8693–8704.

11. K. Kalakewich, R. J. Iuliucci and J. K. Harper, *Cryst. Growth Des.*, 2013, **13**, 5391–5396.

12. B. J. Wylie, C. D. Schwieters, E. Oldfield and C. M. Rienstra, *J. Am. Chem. Soc.*, 2009, **131**, 985–992.

13. U. Sternberg and W. Prieβ, *J. Magn. Reson.*, 1997, **125**, 8–19.

14. U. Sternberg and R. Witter, *J. Biomol. NMR,* 2019, **73**, 727–741.

15. D. H. Brouwer, *J. Am. Chem. Soc.*, 2008, **130**, 6306–6307.

16. F. A. Perras and D. L. Bryce, *J. Phys. Chem. C*, 2012, **116**, 19472–19482.

17. F. A. Perras, I. Korobkov and D. L. Bryce, *CrystEngComm*, 2013, **15**, 8727–8738.

18. D. H. Brouwer and G. Enright, *J. Am. Chem. Soc.*, 2008, **130**, 3095–3105.

19. S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Phys.*, 2017, **146**, 064201.

20. D. W. Alderman, M. H. Sherwood and D. M. Grant, *J. Magn. Reson. Ser. A*, 1993, **101**, 188–197.

21. M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark and M. C. Payne, *J. Phys.: Condens. Matter*, 2002, **14**, 2717.

22. B. G. Pfrommer, M. Côté, S. G. Louie and M. L. Cohen, *J. Comput. Phys.*, 1997, **131**, 233–240.

23. S. J. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.

24. A. Hofstetter and L. Emsley, *J. Am. Chem. Soc.*, 2017, **139**, 2573–2576.

25. L. Wang, F. J. Uribe-Romo, L. J. Mueller and J. K. Harper, *Phys. Chem. Chem. Phys.*, 2018, **20**, 8475–8487.

26. M. H. Sherwood, D. W. Alderman and D. M. Grant, *J. Magn. Reson.*, 1989, **84**, 466–489.

27. F. Liu, C. G. Phung, D. W. Alderman and D. M. Grant, *J. Am. Chem. Soc.*, 1996, **118**, 10629–10634.

28. J. C. Facelli and D. M. Grant, *Nature*, 1993, **365**, 325–327.

29. J. K. Harper, D. H. Barich, J. Z. Hu, G. A. Strobel and D. M. Grant, *J. Org. Chem.* 2003, **68**, 4609–4614.

30. J. K. Harper, J. A. Doebbler, E. Jacques, D. M. Grant and R. B. J. Von Dreele, *J. Am. Chem. Soc.*, 2010, **132**, 2928–2937.

31. S. T. Holmes, O. G. Engl, M. N. Srnec, J. D. Madura, R. Quiñones, J. K. Harper, R. W. Schurko and R. J. Iuliucci, *J. Phys. Chem. A*, 2020, **124**, 3109–3119.

32. K. Kalakewich, R. Iuliucci, K. T. Mueller, H. Eloranta and J. K. Harper, *J. Chem. Phys.*, 2015, **143**, 194702.

33. L. J. Fitzgerald and R. E. Gerkin, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 1993, **49**, 1952–1958.

34. F. R. Fronczek, G. Gannoch, W. L. Mattice, F. L. Tobiason, J. L. Broeker and R. W. Hemingway, *J. Chem. Soc., Perkin Trans. 2*, 1984, 1611–1616

35. A. L. Spek, B. Kojic-Prodic and R. P. Labadie, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1984, **40**, 2068–2071.

36. R. Witter, U. Sternberg, S. Hesse, T. Kondo, F –T. Koch and A. S. Ulrich, *Macromolec.*, 2006, **39**, 6125–6132.

37. S. Lee, C. Wang, H. Liu, J. Xiong, R. Jiji, X. Hong, X. Yan, Z. Chen, M. Hammel, Y. Wang, S. Dai, J. Wang, C. Jiang and G. Zhang, *Acta Crystallogr. Sect. D*, 2017, **73**, 955–969.

38. C. M. Widdifield, H. Robson and P. Hodgkinson, *Chem. Commun.* 2016, **52**, 6685–6688.

39. J. Powell, K. Kalakewich, F. J. Uribe-Romo and J. K. Harper, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12541–12549.

40. M. Garcia-Viloca, A. Gonzalez-Lafont and J. M. Lluch, *J. Phys. Chem. A*, 1997, **101**, 3880–3886.

41. A. D. Bond, *New J. Chem.*, 2004, **28**, 104–114.

42. A. W. Ralston, *Fatty Acids and Their Derivatives*, Wiley, New York, 1948.

43. S. F. Leinin, T. Bremi, R. Bruschweiler and R. R. Ernst, *J. Am. Chem. Soc.*, 1998, **120**, 9870–9879.

44. J. D. Hartman and G. J. O. Beran, *Solid State Nucl. Magn. Reson.*, 2018, **96**, 10–18.