Enhanced NMR discrimination of pharmaceutically relevant molecular crystal forms through fragment-based ab initio chemical shift predictions

Joshua D. Hartman,[†] Graeme M. Day,[‡] and Gregory J. O. Beran[*,†]

Department of Chemistry, University of California, Riverside, California 92521 USA, and School of Chemistry, University of Southampton, Highfield, Southampton, UK.

E-mail: gregory.beran@ucr.edu

August 3, 2016

Abstract

Chemical shift prediction plays an important role in the determination or validation of crystal structures with solid state nuclear magnetic resonance (NMR) spectroscopy. The fundamental theoretical challenge lies in discriminating variations in chemical shifts resulting from different crystallographic environments. Fragment-based electronic structure methods provide an alternative to the widely used plane wave gauge-including projector augmented wave (GIPAW) density functional technique for chemical shift prediction. Fragment methods allow hybrid density functionals to be employed routinely in chemical shift prediction, and we have recently demonstrated appreciable

[*]To whom correspondence should be addressed
[†]Department of Chemistry, University of California, Riverside, California 92521 USA
[‡]School of Chemistry, University of Southampton, Highfield, Southampton, UK.

improvements in the accuracy of the predicted shifts when using the hybrid PBE0 functional instead of generalized gradient approximation (GGA) functionals like PBE. Here, we investigate the solid-state ^{13}C and ^{15}N NMR spectra for multiple crystal forms of acetaminophen, phenobarbital, and testosterone. We demonstrate that the use of the hybrid density functional instead of a GGA provides both higher accuracy in the chemical shifts and increased discrimination among the different crystallographic environments. Finally, these results also provide compelling evidence for the transferability of the linear regression parameters mapping predicted chemical shieldings to chemical shifts that were derived in an earlier study.

Introduction

Molecular crystal structure is governed by a delicate balance among intra- and intermolecular interactions, and even small changes in the crystallization process may lead to different crystal packing motifs, or polymorphs. Despite having identical chemical compositions, different polymorphs often manifest significantly altered physical properties. In pharmaceutical applications, changes in crystal packing can significantly impact bioavailability, shelf-life, and even intellectual property protection.[1–3] Modern pharmaceutical development involves extensive polymorph screening, and it is often important to identify the structures of the resulting crystals.

While single-crystal and powder x-ray diffraction remain the primary methods for crystal structure determination and fingerprinting solid forms, solid-state nuclear magnetic resonance (NMR) is an increasingly used alternative. The combination of solid-sate NMR and powder x-ray diffraction has proved particularly potent for solving crystal structures.[4,5] NMR chemical shielding is a function of the local electronic structure, making it sensitive to both the molecular geometry and local crystallographic environment, and is therefore an excellent tool for investigating polymorphism. In addition to solving or confirming crystal structures, pharmaceutical companies increasingly rely upon NMR to monitor crystallization, test sam-

ples and investigate new formulations, especially when crystal structures cannot be obtained easily.

However, translating an NMR spectrum into a 3-D crystal structure is challenging and often requires computational chemical shift predictions to facilitate spectral assignment. Chemical shift prediction is frequently used to confirm structures solved from powder x-ray diffraction,[4,6] to identify structures consistent with the experimental NMR spectrum,[7–16] to assign NMR spectra,[17–21] or even to help refine crystal structures.[21–30] Despite its sensitivity to local crystal packing, the changes in chemical shift across different polymorphs or crystallographic environments can be subtle. In sulfanilamide, for example, key ^{13}C chemical shifts vary by only 1–3 ppm across three polymorphs that differ primarily in their hydrogen bonding networks.[31] In α-testosterone, most of the splittings between the pairs of ^{13}C shifts corresponding to the symmetrically inequivalent molecules in the unit cell are less than 2 ppm.[17]

The ability to discriminate among distinct crystallographic environments is the fundamental theoretical challenge in NMR crystallography. This is especially true when trying to identify the relevant structure(s) from a large set of candidate structures generated via crystal structure prediction techniques.[9–14,32,33] Various strategies to address this discrimination challenge have been advanced in recent years. For example, chemical shifts for ^{1}H, ^{15}N, ^{17}O nuclei can be more sensitive to changes in hydrogen bonding patterns than those for ^{13}C, making those preferable to study in some cases.[11,34–39] Individual chemical shielding tensor components can also provide additional information about the crystallographic environment that is lost in the isotropic shifts.[13,14,28] Here we focus on a different strategy: can we increase discrimination among distinct crystallographic environments by improving the accuracy of the chemical shift prediction?

Since the advent of the gauge-including projector augmented wave (GIPAW) plane wave density functional theory (DFT) method in 2001,[40,41] it has become the *de facto* standard for chemical shift prediction in the solid state. GIPAW provides high-quality chemical shifts and

has been successfully employed in many NMR studies, as discussed in recent reviews.[42–44] Despite its successes, the statistical errors obtained with GIPAW and the commonly used PBE density functional are comparable to the variations in chemical shifts that are often seen across different crystallographic environments or polymorphs. Benchmark tests on GIPAW PBE ^{13}C chemical shifts in molecular crystals typically obtain errors around ∼2 ppm,[10,45–47] for instance, which is on par with the magnitude of the chemical shift resolution needed to distinguish known crystal forms of molecules such as testosterone or sulfanilamide. Reducing the errors in the predicted chemical shifts would improve discrimination in challenging NMR crystallography applications.

Before the widespread use of GIPAW DFT, molecular crystal chemical shift calculations often mimicked the crystalline environment using an individual molecule surrounded by a field of point charges[48,49] or by a few key neighboring molecules. Recently, there has been renewed interest in modernized versions of these sorts of methods. Thanks to increased computer power and efficient algorithms for chemical shift calculations[50–54] within the gauge-including atomic orbital (GIAO) formalism,[55] chemical shift prediction on larger clusters of molecules can now be performed routinely. Clusters consisting of ∼10–15 molecules within a few Angstroms of a molecule in the asymmetric unit mimic the effect of the extended molecular crystal lattice on the chemical shielding of the central molecule very effectively. Various cluster-type models have also been employed in biological systems.[56–69] When using the same density functional, the accuracy of these cluster models is competitive with GIPAW for several different nuclei.[45–47,70] However, the rapidly increasing computational cost with system size becomes a concern when computing shifts on a cluster of large molecules.

Fragment methods are similar to cluster models in that they treat one or more central molecules interacting with other nearby molecules in the crystal. However, they decompose those interactions into sums of contributions from small groups of molecules, or fragments. We have demonstrated that when coupled with electrostatic embedding, fragment methods allow one to predict isotropic ^{1}H, ^{13}C and ^{15}N chemical shieldings with accuracy similar to

both cluster methods and GIPAW, albeit with lower computational cost.[70,71]

The key advantage of these cluster and fragment methods is that they enable the practical use of a wider range of electronic structure methods for chemical shift prediction. In particular, extensive molecular crystal benchmark testing indicates that hybrid density functionals predict appreciably higher-accuracy isotropic chemical shifts for several different nuclei.[47,70] In a plane wave method like GIPAW, hybrid density functionals require at least an order of magnitude more computational effort than generalized gradient approximation (GGA) functionals, which makes them impractical. In contrast, the computational cost premium for switching from a GGA to a hybrid functional when using Gaussian basis sets, as in fragment and cluster models, is only $\sim$50%. In other words, hybrid density functionals can be used routinely with these cluster and fragment methods.

High-accuracy chemical shielding calculations are of particular interest in examining NMR shift differences between polymorphs or solvates and between different independent molecules in a given crystal, i.e. in crystals for which Z' is greater than one. Accordingly, we examine two polymorphic crystals—acetaminophen (a.k.a. paracetamol) and phenobarbital—as well as the monohydrate and neat forms of testosterone. The differences in intramolecular conformation among the different crystal forms of each molecule are small, so much of the variation in the chemical shifts stems from differences in the (intermolecular) crystallographic environments. Thus, these represent challenging cases for NMR discrimination.

We find that fragment and cluster models generally predict the same ^{13}C and ^{15}N spectral assignments as GIPAW. More significantly, we demonstrate that switching from a GGA functional to a hybrid one both provides higher-accuracy chemical shifts and increases discrimination among the different crystallographic environments found in these crystal forms. Finally, chemical shift referencing is always important in both experimental and theoretical studies. We demonstrate that the referencing models fitted in our previous benchmark studies[70] are highly transferable to the crystals studied here, underscoring their suitability for broader use.

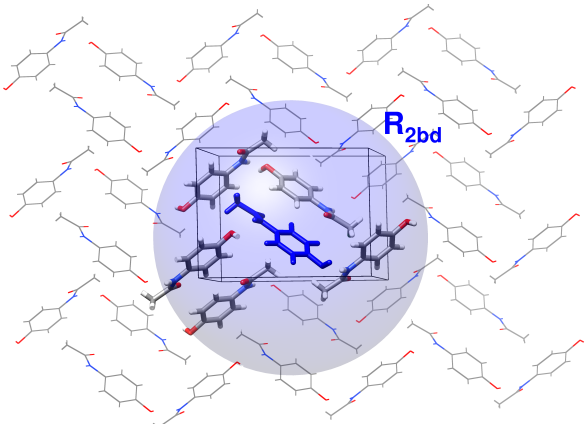


Figure 1: Illustration of the fragment, cluster and combined cluster/fragment models for computing NMR chemical shieldings. Depicted is a cross section from the optimized acetaminophen form I molecular crystal with the crystallographically unique molecule shown in blue. The fragment model includes pairwise interactions between the central molecule and any other molecule whose atoms lie within the distance R_{2bd} (blue sphere). The cluster model uses a cluster of molecules surrounding the central molecule, as represented by a licorice model. The cluster/fragment model combines the cluster with pairwise interactions involving other molecules in the blue sphere that are not present in the cluster. In all cases, electrostatic embedding is employed and extends well-beyond the 2-body cutoff region.

Theory

Fragment-based chemical shift prediction techniques have been described previously.[45–47,70,71] In brief, they rely on a many-body expansion for the shielding tensor,

$$\boldsymbol{\sigma}_i^A = \tilde{\boldsymbol{\sigma}}_i^A + \sum_j \Delta^2\tilde{\boldsymbol{\sigma}}_{ij}^A + \sum_{jk} \Delta^3\tilde{\boldsymbol{\sigma}}_{ijk}^A + \ldots \tag{1}$$

which can be derived by differentiating the many-body expansion for the energy with respect to the nuclear magnetic moment and the external magnetic field. This expansion decomposes the shielding tensor of atom A on molecule i in the unit cell into the shielding tensor on the isolated molecule ($\tilde{\boldsymbol{\sigma}}_i^A$) plus corrections due to the interactions of that molecule with other molecules in the crystal. The leading corrections involve pairwise interactions of molecule i with nearby molecules ($\Delta^2\tilde{\boldsymbol{\sigma}}_{ij}^A$), followed by non-additive three-body (trimer) corrections ($\Delta^3\tilde{\boldsymbol{\sigma}}_{ijk}^A$).[71]

Four different chemical shielding prediction models will be compared here. First, the 2-

Table 1: Root-mean-square errors (in ppm) between predicted and experimental isotropic chemical shifts from benchmarks involving 25 crystals/169 shifts for ^{13}C and 24 crystals/51 shifts for ^{15}N.[70]

	^{13}C		^{15}N	
	PBE	PBE0	PBE	PBE0
2-Body Fragment	2.1	1.5	5.5	4.2
Cluster	2.1	1.5	5.7	3.9
Cluster/Fragment	2.1	1.5	5.8	4.0
GIPAW	2.2	–	5.4	–

body fragment model truncates this many-body expansion by neglecting long-range pairwise interactions and all 3-body and higher terms. The 2-body interactions are computed between the central molecule and any other molecule lying within a user-defined cutoff distance of the central one (typically R_{2bd} =6 Å). Because the neglected long-range and many-body terms involve substantial contributions from polarization, their effects can be approximately captured via electrostatic embedding. Specifically, the 1-body and 2-body terms are calculated in a field of point charges that mimic the extended lattice.

Second, the cluster model performs a single supermolecular calculation on a central molecule surrounded by nearby molecules (typically those with any atom lying within 4 Å). This cluster calculation captures local pairwise and many-body effects explicitly. However, due to computational constraints that limit the overall size of the cluster, the cluster approach can miss potentially important longer-range interactions. Electrostatic embedding is employed here to capture some of the polarization arising from the extended lattice.

Third, the combined cluster/fragment approach seeks to achieve the best of both models. It describes the local interactions (out to 4 Å) with a cluster calculation, and longer-range interactions in a pairwise fashion (out to 6 Å). Again, electrostatic embedding is employed to approximate longer-range and missing many-body effects. Figure 1 summarizes these first three models. Fourth, fully periodic GIPAW calculations make no such truncation, so should be equivalent to the many-body expansion carried out to all orders.

Benchmark studies on many different molecular crystals have demonstrated that, for a given density functional, the inexpensive 2-body fragment model predicts ^{1}H, ^{13}C, ^{15}N

chemical shifts on par with the cluster, cluster/fragment, and GIPAW approaches.[47,70,71] For ^{17}O, many-body effects are more important, and cluster-type and GIPAW models perform moderately better than the fragment approach. As noted in the Introduction, fragment methods also enable the routine use of hybrid density functionals like PBE0 instead of the GGA PBE, and this provides appreciable improvements in the chemical shifts. Table 1 summarizes the root-mean-square (rms) errors from these benchmark studies for ^{13}C and ^{15}N, which are the two nuclei considered here.

Computationally, the 2-body fragment approach is the least expensive of the four models considered here, especially for large unit cells. The computational effort scales linearly with the number of molecules in the asymmetric unit cell, so calculating the chemical shifts for a polymorph with two independent molecules ($Z' = 2$) requires roughly double the effort of one with $Z' = 1$. On the other hand, the computational cost is independent of the total number of molecules Z in the unit cell. Fragment approaches are also inherently parallel, since each fragment calculation can be performed on a separate group of processors. This allows the calculation to scale efficiently to hundreds of processors, so that chemical shifts can be obtained very quickly when sufficient processors are available (e.g. often within a few hours of wall time for the sorts of crystals described in this work).

Computational Methods

Crystal Structures

The three molecules studied here are depicted in Figure 2, along with their atom numbering. X-ray or neutron diffraction crystal structures for each polymorph were obtained from the Cambridge Structure Database (CSD). Experimental solid-state NMR data under magic angle spinning (MAS) conditions was taken from the literature. CSD reference codes and citations to the experimental data for each crystal is given below:

- Acetaminophen:[27,72] form I (HXACAN26[73]), form II (HXACAN23[74]), and form III

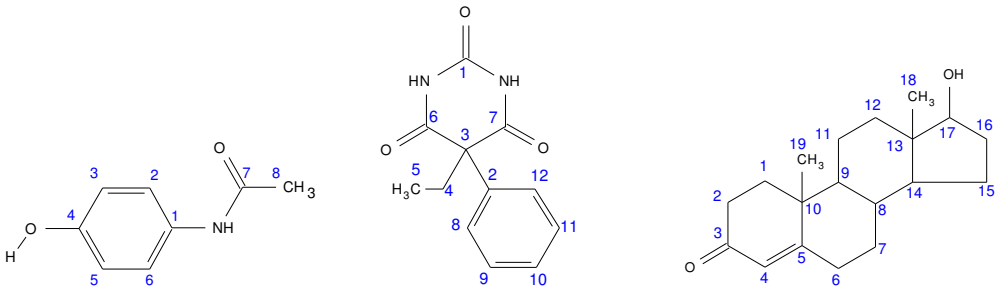


Figure 2: Molecular structures and atom numbering for the three species studied here.

(HXACAN29[75]).

- Phenobarbital:[19] form II (PHBARB06[76]) and form III (PHBARB09[77]).
- Testosterone:[17] α form (TESTON10[78]) and β (monohydrate) form (TESTOM01[79]).

All ^{13}C and ^{15}N chemical shifts reported here are referenced relative to neat TMS and external solid NH_4Cl, respectively.[80]

Computational techniques

Experimental crystal structures were refined using all-atom geometry optimizations with fixed room temperature unit cell parameters. All geometry optimizations were carried out using the freely available, open-source Quantum Espresso software package.[81] The PBE[82] density functional and the D2 dispersion correction,[83] ultrasoft pseudopotentials with a plane wave cut off of 80 Ry, and a 3×3×3 Monkhorst-Pack k-point grid were used for all geometry optimizations. We used the pseudopotentials H.pbe-rrkjus.UPF, C.pbe-rrkjus.UPF, N.pbe-rrkjus.UPF, O.pbe-rrkjus.UPF, S.pbe-n-rrkjus_psl.0.1.UPF from http://www.quantum-espresso.org.

Isotropic chemical shieldings were computed on the relaxed structures using the fragment, cluster, cluster/fragment, and GIPAW techniques. For the fragment-based techniques, crystal structure fragmentation out to a 6 Å 2-body radius from the asymmetric unit was carried out using our hybrid many-body interaction (HMBI) code.[84–86] All cluster-based calculations used a 4 Å cluster to include many-body effects involving nearest neighbor molecules. In-

dividual fragment shielding tensor calculations were performed using Gaussian09[87] with the PBE0[88] and PBE[82] density functionals and numerical integration grid involving 150 radial points and 974 angular Lebedev points.

Atomic point-charges were computed for each crystal using distributed multipole analysis,[89,90] and point charge embedding out to 30 Å was employed in all fragment and/or cluster calculations. Locally dense basis sets[91,92] were used for increased computational efficiency. A 6-311++G(2d,p) basis was used on the molecule(s) of interest, 6-311G(d,p) on all atoms within 4 Å of the central molecule, and 6-31G on more distant atoms. See our previous work for more detailed discussion of these protocols.[70,93]

Gauge-including projector augmented wave (GIPAW) chemical shielding calculations were performed using CASTEP[94] with the PBE functional, ultrasoft pseudopotentials generated on-the-fly and an 850 eV (62.5 Ry) plane wave basis set cut off. Electronic k-points were sampled on a Monkhorst-Pack grid to give a maximum separation between k-points of 0.05 $Å^{-1}$. These basis set cut off and k-point density were chosen based on previous testing[70] to converge relative chemical shifts to better than 0.01 ppm. Full space group symmetry was used in all GIPAW calculations.

Data analysis

Experimentally observed isotropic chemical shifts δ^A are reported relative to a reference compound. Therefore, the computed absolute isotropic shieldings σ^A must be appropriately referenced to compare with the experimental values. Although numerous techniques exist for referencing the predicted shifts,[95] one particularly useful method uses a linear regression model of the form:

$$\delta^A = a\sigma^A + b \tag{2}$$

where a and b are obtained via a linear least-squares fit between calculated shieldings and experimental shifts. In the absence of systematic error, a would take a value of -1, and b

would represent the absolute shielding of the reference compound. Allowing a to deviate from -1 helps compensate for systematic errors in the predicted shieldings. The linear regression parameters used for ^{13}C and ^{15}N here were obtained from our recent benchmark study of 25 crystals/169 ^{13}C isotropic shifts and 24 crystals/51 ^{15}N isotropic shifts.[70] Note that form I acetaminophen was included in the ^{13}C benchmark data set, but none of the other polymorphs studied in the current work were present in the training set upon which the linear regressions were fitted. The specific regression parameters used for each nucleus type and each of the four theoretical models are listed in supporting information (Table S4).

Two statistical metrics are used here to assess the quality of the predicted shifts. The first is the RMS error between the predicted and experimental shifts. The second is a reduced chi-squared analysis, $\tilde{\chi}^2$, which provides a measure of how consistent the errors observed for a given set of predicted shifts are with the expected distribution of errors. The $\tilde{\chi}^2$ is computed as,

$$\tilde{\chi}^2 = \frac{1}{N}\sum_{i}^{N}\frac{\left(\delta_i^{pred} - \delta_i^{expt}\right)^2}{\sigma_{rms}^2} \tag{3}$$

where N is the number of isotropic shifts, δ_i^{pred} is the predicted chemical shift, δ_i^{expt} is the experimentally observed chemical shift, and σ_{rms} is the width of the expected error distribution. Here, the σ_{rms} for each model/nucleus type are given by the RMS errors for the benchmark test sets summarized in Table 1.

Smaller $\tilde{\chi}^2$ values indicate that the errors for a given system are more consistent with the errors one expects based on the benchmark sets. Those benchmark error distributions are roughly Gaussian, albeit with longer tails (i.e. large errors occur more frequently than one would expect for an ideal normal distribution).[70] This means that larger $\tilde{\chi}^2$ values are moderately more probable than one would typically expect, but they still provide a useful metric for comparing different potential assignments.

In many ways, the differences in $\tilde{\chi}^2$ values for different potential assignments are more important than the values themselves. In several cases discussed later in this work, the $\tilde{\chi}^2$

values computed with PBE are smaller than those from PBE0, even though the PBE RMS errors are larger than the PBE0 ones. This simply reflects the larger uncertainty (σ_{rms}) expected for PBE than for PBE0 based on the benchmark sets. Accordingly, larger errors are more likely to occur when using PBE on the systems here, and they therefore do not skew the PBE $\tilde{\chi}^2$ values as much. On the other hand, the differences among $\tilde{\chi}^2$ values computed with a given method for different shift assignments indicates the ability of that method to discriminate between correct and incorrect assignments. This is an important consideration when using NMR for validation of proposed crystal structures. To investigate this, we evaluate $\tilde{\chi}^2$ summed over all crystal forms of each molecule for the correct assignment of measured chemical shifts to the monomers in each structure, as well as for all possible permutations where the observed chemical shifts are assigned to the incorrect monomers. These differences in $\tilde{\chi}^2$ will be the primary focus of the discussion of the results.

Results and Discussion

For each of the three chemical systems examined here, we compare the performance of GIPAW and 2-body fragment with the PBE density functional to check for the similarity between GIPAW and fragment predictions. Then we investigate how the quality of the predictions changes upon switching to the PBE0 functional using the 2-body fragment, cluster, and cluster/fragment models. The ability of each method to discriminate between correct and incorrect assignments of the measured chemical shifts to the known crystal structures is then examined using the $\tilde{\chi}^2$ calculations.

Acetaminophen

Acetaminophen (a.k.a. paracetamol) is a widely used, over-the-counter pain reliever and fever reducing agent which adopts three known polymorphic crystal forms (forms I, II, and III). The solid state versatility of acetaminophen serves as an excellent example of how crystal

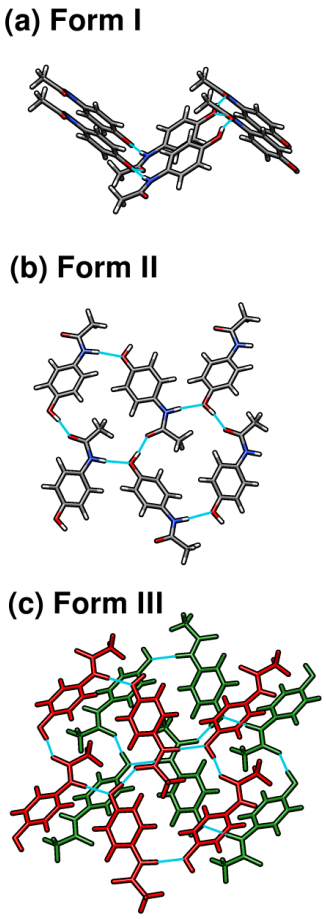


Figure 3: Optimized acetaminophen crystal structures for (a) form I showing puckered hydrogen-bonded sheets, (b) form II showing flat hydrogen-bonded sheets, and (c) form III with two independent flat hydrogen-bonded sheets (colored red and green) which are similar to the sheets found in form II.

packing impacts the drug manufacturing process. Form I is easily isolated and is characterized by puckered hydrogen-bonded sheets (see Figure 3a). This packing configuration inhibits shearing, which in turn impacts the compressibility of the solid and the tabletting process. On the other hand, form II is more difficult to isolate and consists of flat hydrogen-bonded sheets which are more amenable to direct compression (Figure 3b).[72,96–101] Form III, with two molecules in the asymmetric unit, consists of alternating layers of symmetrically independent, flat two-dimensional sheets (Figure 3c). Historically, form III has been much more difficult to obtain,[101,102] and its structure was solved only in 2009.[75] The intramolecular acetaminophen geometries are very similar across all four symmetrically unique molecules

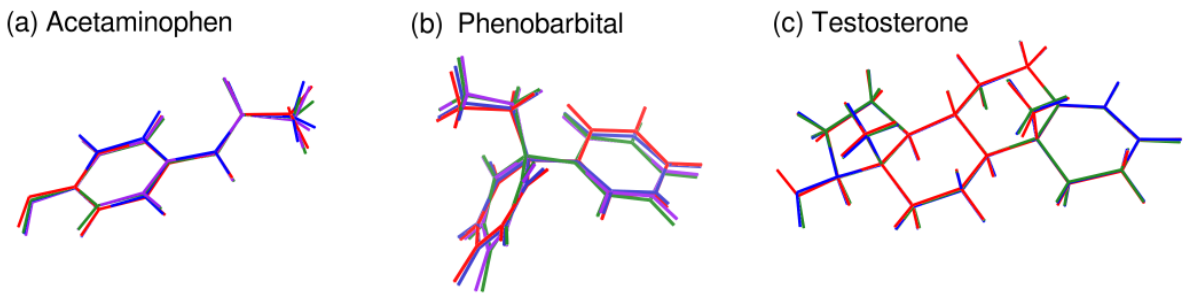


Figure 4: Monomer overlays for the symmetrically unique monomers in each set of crystals. (a) Acetaminophen: forms I (red), II (blue), IIIa (green), and IIIb (purple). (b) Phenobarbital: forms IIa (red), IIb (blue), IIc (green), and III (purple). (c) Testosterone: αu (red), αv (blue), and β (green).

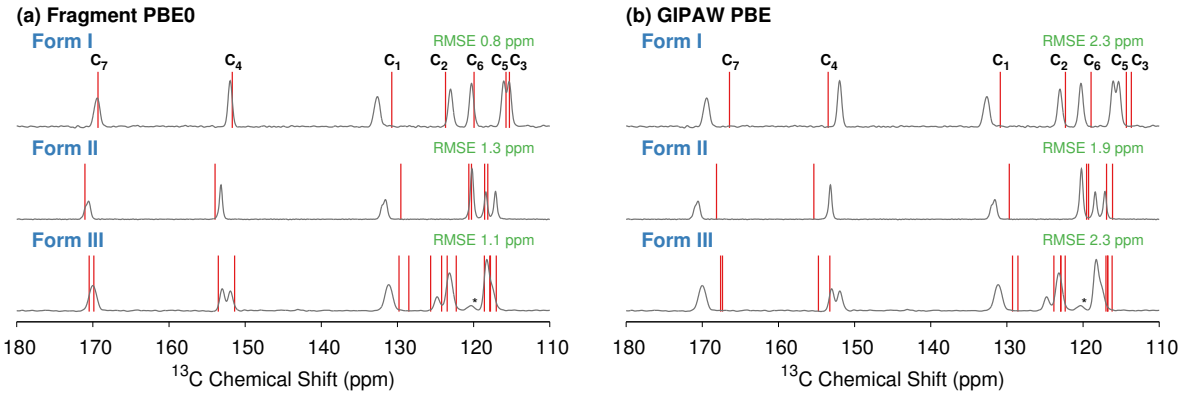


Figure 5: Overlay of experimental acetaminophen spectra[72] and predicted shifts (in red) for (a) the two-body fragment PBE0 calculations and (b) GIPAW PBE calculations. The C8 methyl peak at 23 ppm is not shown.

which occur in the three polymorphs, with only subtle variations in the hydroxyl and methyl groups. Structure overlays among the four monomers (Figure 4a) find root-mean-square deviations (rmsd) of only 0.07–0.19 Å for the non-hydrogen atoms. The two monomers in form III are the most similar (rmsd 0.07 Å), followed by form I and form II (rmsd 0.09 Å). Accordingly, much of the variation in the observed NMR chemical shifts across the different polymorphs stems from differences in the crystallographic environment, rather than changes in the intramolecular geometry.

The literature contains multiple sets of ^{13}C shifts for form I.[27,72,103] Data from Ref 72 is used here because it includes ^{13}C and ^{15}N chemical shifts for all three forms. With revised

referencing, the form I ^{13}C shifts from Ref 72 are consistently 0.3–0.4 ppm upfield of the more recent ones reported by Harper et al.[27] Similarly, the re-referenced values for the form I ^{15}N isotropic shift from Ref 72 agree with a more recent study to within ∼0.1 ppm.[28] See supporting information for referencing details.

Figure 5 compares the experimental and predicted ^{13}C isotropic chemical shifts using (a) fragment PBE0 or (b) GIPAW PBE for the three acetaminophen polymorphs. Both models predict chemical shifts in good agreement with the experimental spectra. To our knowledge, the experimental structure of form III solved via crystal structure prediction and powder x-ray diffraction[75] has not been validated against the NMR spectrum. The consistency between the predicted and experimental spectra here provides additional support for this structure.

Because they arise largely from differences in intermolecular packing, the variations in experimental ^{13}C shifts among the three polymorphs are often subtle. For instance, the minor differences in the hydrogen bonding to oxygen or nitrogen atoms between polymorphs lead to differences in ^{13}C chemical shifts (C1, C4, and C7) of only ∼1 ppm. Nevertheless, the individual peak assignments suggested by both 2-body PBE0 and GIPAW PBE are very similar. Both reproduce the general trends in peak positions across the three polymorphs. The main difference between the two models occurs for the C3/C5 peaks near ∼115–118 ppm in all three structures. The splitting of these peaks is ∼1 ppm or less in each polymorph. Interestingly, both 2-body PBE and 2-body PBE0 predict C5 downfield of C3, while the cluster, cluster/fragment, and GIPAW results predict the opposite ordering (see Table S1). In other words, the ordering of these two resonances appears to be sensitive to many-body effects.

A recent ^{1}H/^{13}C heteronuclear correlation and ^{13}C shielding tensor study on form I suggests the GIPAW and cluster-based model assignment is correct for form I,[27] and it is used throughout this work. For the sake of computing RMS errors below, we assume the GIPAW and cluster-based model assignment of C3/C5 is correct for form II as well, though

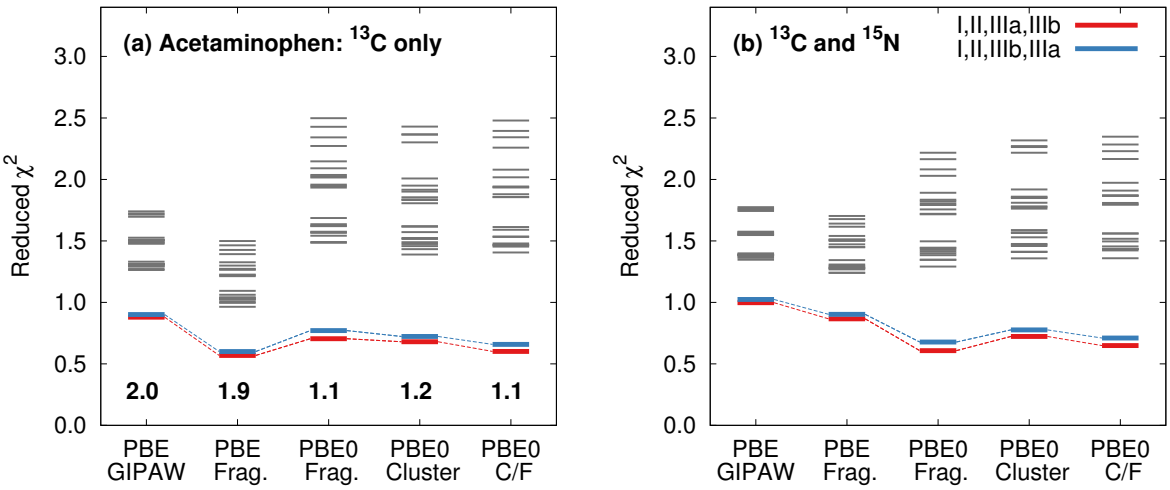


Figure 6: Reduced χ^2 analysis using (a) ^{13}C or (b) ^{13}C and ^{15}N isotropic shifts for the 24 possible acetaminophen polymorph assignments using fragment, cluster and cluster/fragment methods. The best assignment (I, II, IIIa, IIIb) is shown in red, while the blue line indicate the assignment that swaps the two monomers in form III. The ^{13}C RMS errors (in ppm) for the best assignment with each model are reported near the bottom of (a).

this is uncertain. In form III, C3 and C5 are unresolved experimentally.

The ordering of C2 and C6 in form II shows similar disagreement between the 2-body models versus the cluster and GIPAW results. In this case, the predicted splitting is even smaller, at ∼0.5 ppm or less, while the experimental splitting is unresolved. In all of these cases where the models disagree on the shift ordering, the splittings are small compared to the benchmark ^{13}C RMS errors (Table 1). In other words, such differences probably cannot be resolved with confidence with either GIPAW or fragment methods. Overall, the RMS error across all 24 ^{13}C chemical shifts is 2.0 ppm for GIPAW/PBE and 1.9 ppm for 2-body PBE. Fragment, cluster, and cluster/fragment methods using the PBE0 hybrid functional nearly halve the overall RMS error to 1.1–1.2 ppm (Figure 6).

The three polymorphs contain four unique ^{15}N chemical shifts. Figure 7 compares the experimental and predicted shifts. All the theoretical methods predict the shifts to occur too far downfield compared to experiment. The largest errors versus experiment occur with GIPAW PBE (RMSE 7.5 ppm), while the 2-body fragment PBE0 model gives the best agreement (RMSE 3.6 ppm). Interestingly, all the models tested here predict that the form

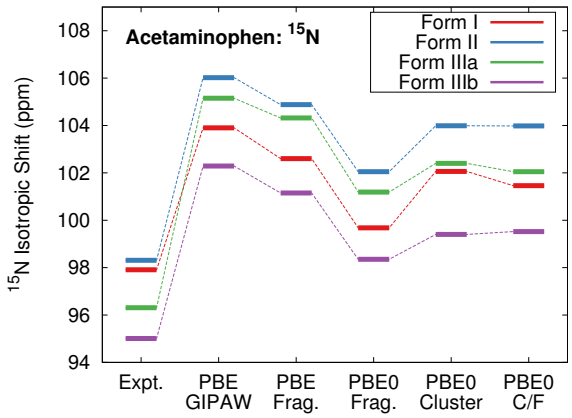


Figure 7: Comparison of the predicted and experimental ^{15}N isotropic chemical shifts for the three polymorphs of acetaminophen.

I nitrogen shift should lie upfield of the form IIIa one, contrary to the experimental results. Of course, all four experimental shifts span only 3.3 ppm, which is small relative to the ∼4–6 ppm RMS errors found in the nitrogen benchmarks (Table 1).[70]

While the good agreement between predicted and experimentally observed chemical shifts is valuable, a more challenging question is to what extent can the chemical shift predictions discriminate among the different crystallographic environments of the acetaminophen molecules found in the various polymorphs? The three polymorphs provide a total of four crystallographically unique molecules: forms I and II each contribute one and form III contributes two. There are 24 possible ways one can assign the predicted shifts from these four crystallographically unique molecules to the experimental shifts (assuming the carbon atom assignments within the monomer are fixed according to the discussion above). Figure 6 illustrates the reduced χ^2 values for each of the 24 possible assignments using either the ^{13}C isotropic shifts or both the ^{13}C and ^{15}N isotropic shifts as computed with the various different models.

Several features stand out in Figure 6. First, the experimental assignment of the two inequivalent monomers in form III is unknown, but all five models considered predict the same assignment to have the smallest $\tilde{\chi}^2$ values (I, II, IIIa, IIIb in Figure 6). However, the methods cannot clearly distinguish this monomer assignment from the one that swaps the

two monomers in form III. The difficulty distinguishing these crystallographically unique monomers in form III is unsurprising, since the two molecules adopt nearly identical intramolecular geometries (rmsd 0.07 Å) and are in very similar crystallographic environments, which results in overlap of most of the ^{13}C shifts. Only the pair of peaks corresponding to C4 are clearly resolved in the experimental spectrum, and those two resonances are separated by only ∼1 ppm. We also note that, despite the geometric similarity of the hydrogen bonded layers in polymorphs II and III, the methods all effectively discriminate between the molecular environments in those two polymorphs: $\tilde{\chi}^2$ increases significantly if the set of shifts for either monomer in form III is swapped with those of form II. Although the molecules in forms I and II adopt very similar intramolecular conformations, their ^{13}C spectra are comparatively easy to distinguish (particularly near 120 ppm) due to the appreciable differences in the crystallographic environments.

Second, although the RMS errors with PBE are almost twice as large as those from PBE0, the PBE $\tilde{\chi}^2$ values tend to be smaller than those from PBE0. This reflects the fact PBE exhibits broader error distributions than PBE0 in the benchmark sets,[70] which leads to a larger denominator in Eq 3. A more useful interpretation of these $\tilde{\chi}^2$ plots focuses on the resolution between different crystallographic environments. The hybrid PBE0 functional provides increased separation among the $\tilde{\chi}^2$ values for different potential assignments, which corresponds to increased discrimination among the different crystallographic environments. Interestingly, the more expensive combined cluster/fragment PBE0 mode exhibits marginally smaller RMS errors than the fragment PBE0 model, but it provides no additional discrimination among the monomer assignments.

Third, combining both ^{13}C and ^{15}N shift predictions (Figure 6b) does not appreciably alter the discrimination among the different potential assignments relative to the ^{13}C-only data (Figure 6a). As noted above, the ∼3 ppm variations among the four unique ^{15}N shifts are smaller than the ∼4–6 ppm errors expected from the ^{15}N benchmark set, so the $\tilde{\chi}^2$ analysis does not readily discriminate among the different potential ^{15}N shift assignments in

acetaminophen. The low discriminatory power of ^{15}N compared to ^{13}C shifts is unsurprising given the nature of the structural differences between the three polymorphs: the hydrogen bonding environment around the nitrogen atom is each polymorph is nearly the same in each structure, while the differences in the geometry of the layers and the way that these layers are packed has a much stronger influence on the crystalline environments of the carbon atoms (and hence the ^{13}C shifts).

Phenobarbital

Phenobarbital (5-ethyl-5-phenyl-2,4,6(1H,3H,5H)-pyrimidine-trione) is a widely administered barbiturate and has been in clinical use for close to a century. With twelve polymorphs reported in the literature to date,[104] only a few of which have been structurally characterized,[104,105] phenobarbital makes a particularly interesting test-case for assessing the accuracy of NMR chemical shielding calculations.[19] In the present work, we examine polymorph II ($Z' = 3$) and III ($Z' = 1$), for which experimental x-ray crystal structures, ^{13}C, and ^{15}N NMR isotropic chemical shifts have been reported. The experimental shifts have previously been assigned with the help of GIPAW.[19]

Phenobarbital is more flexible than acetaminophen, which leads to two slightly different intramolecular conformations among the four symmetrically unique monomers found in these two polymorphs. Figure 4b overlays the four monomer structures. The conformation of molecules IIa and IIb is very similar (rmsd 0.12 Å), as is the conformation for molecules IIc and III (rmsd 0.08 Å). However, these two pairs differ from each other by 0.28–0.36 Å in rmsd, due primarily to differences in the dihedral angles among the moieties surrounding the quaternary carbon.

The intermolecular hydrogen bonding patterns in forms II and III phenobarbital (Figure 8) are also interesting. The molecules in polymorph III form linear chains consisting of pairs of N-H...O hydrogen bonds (Figure 8b). The same chains are formed by monomers *a* and *c* in polymorph II (blue and red molecules in Figure 8a). The third crystallographically

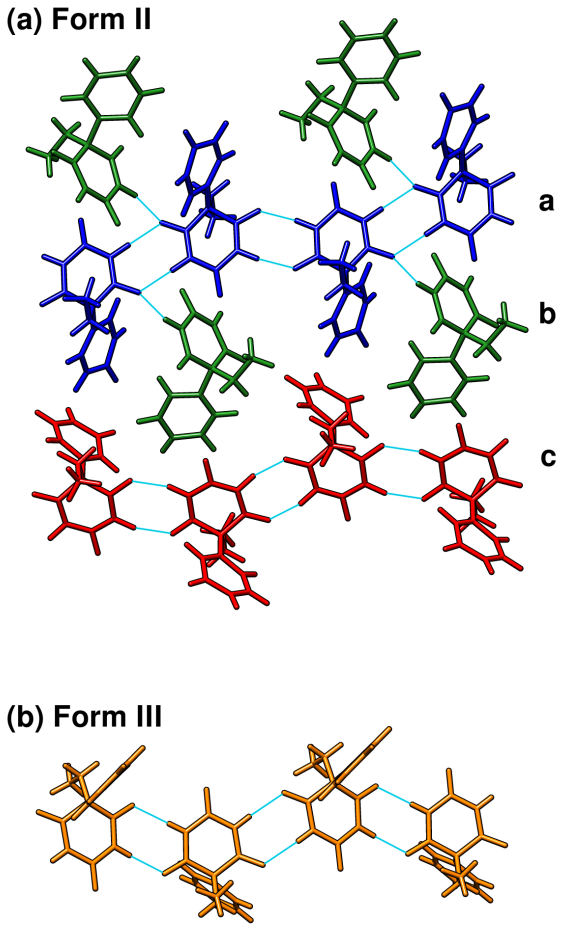


Figure 8: (a) The two hydrogen bonding motifs in phenobarbital form II colored according to symmetry equivalency. (b) Depiction of the linear chain hydrogen-bonding motif in form III.

unique molecule b in form II hydrogen bonds to monomer a, connecting the chains into a two-dimensional motif.

The similarities between the linear hydrogen-bonded chains and intramolecular conformation manifests in similar experimental isotropic ^{13}C chemical shieldings, provided in Table 2. The experimental chemical shifts for the carbonyl carbons (C6 and C7) which are involved in hydrogen bonding in form II monomer c and form III differ by 1 ppm or less, while the same shifts on form II monomers a and b differ from the c ones by $\sim$3 ppm. Furthermore, the observed chemical shifts correlate nicely with the hydrogen bonding patterns. The C4/C6 signal for form II molecule a exhibits the highest frequency chemical shift and displays the greatest degree of hydrogen bonding, having both strong hydrogen bonds (blue linear chain

in Figure 8) and weaker hydrogen bonds to molecule *b*. Molecule *b*, on the other hand, which has no hydrogen bonds at C4 or C6, exhibits the lowest-frequency chemical shifts.

Table 2 provides the predicted isotropic ^{13}C chemical shifts using a two-body fragment model with PBE0 for each crystallographically unique molecule in phenobarbital forms II and III. The predicted shieldings for the other computational methods used in this study are provided in Table S2. The carbon atoms in Table 2 are labeled using the same convention used by Abraham et al[19] designating ring carbons which are in closer proximity to the adjacent ring as "edge." This data clearly shows that both trends regarding the impact of hydrogen bonding on the chemical shifts mentioned above are faithfully reproduced at the fragment level. The overall RMS errors for fragment PBE0 are a couple tenths of a ppm smaller than those from either fragment PBE or GIPAW PBE (1.5 vs 1.7 ppm).

Furthermore, the GIPAW/PBE calculations performed in this study, the fragment PBE0 calculations, and the cluster/fragment PBE0 calculations support the previously reported GIPAW PBE assignments.[19] The only exception is the assignment of the 132.81 ppm and 130.18 ppm peaks to the meta carbons of form IIa and IIc, respectively. While our GIPAW and cluster/fragment calculations agree with the earlier assignment, two-body fragment calculations suggest the reverse assignment. However, in every model tested, the two predicted peaks lie within half a ppm or less of one another, making the significance of the orderings questionable.

Isotropic ^{15}N chemical shifts are also available for the two polymorphs. Assignment of the ^{15}N spectra is unambiguous with all of the different models (see Table S2). However, Table 3 illustrates a significant improvement in the agreement between the predicted and experimental shifts when using the hybrid PBE0 functional instead of PBE. The RMS errors associated with both GIPAW and fragment-based calculations using PBE are ∼9–9.5 ppm, compared to ∼5.5–6 ppm with the PBE0 models. Comparison of individual ^{15}N shifts computed with GIPAW PBE and fragment PBE0 is plotted in Figure 9. The fragment PBE0 model provides better agreement with the specific shift values and the relative shift

Table 2: Experimental and calculated ^{13}C isotropic chemical shifts (in ppm) for phenobarbital forms II and III. Predicted shieldings are reported for two-body fragment calculations using PBE0.

	Form IIa		*Form IIb*		*Form IIc*		*Form III*	
Carbon	Expt.	Calc.	Expt.	Calc.	Expt.	Calc.	Expt.	Calc.
C1 – Carbonyl	147.15	147.53	148.91	149.43	147.15	147.80	149.01	149.00
C2 – Ispo	136.00	133.02	137.17	137.54	137.17	137.17	137.56	136.94
C3 – Quaternary	61.68	63.50	61.00	62.15	62.37	63.46	62.27	63.27
C4 – Methylene	30.35	33.78	32.21	36.10	27.22	28.74	27.12	28.63
C5 – Methyl	6.86	8.67	7.93	9.53	8.91	10.47	11.36	12.98
C6 – CO (edge)	177.41	179.70	169.87	171.61	173.20	175.19	174.20	176.89
C7 – CO	177.41	178.30	173.20	174.44	174.96	175.79	174.20	175.89
C8 – Ortho (edge)	125.76	125.02	127.02	127.10	125.40	124.93	127.57	128.12
C9 – Meta (edge)	131.39	131.12	130.18	130.72	133.74	134.39	130.70	130.29
C10 – Para	132.41	133.08	129.30	129.32	130.18	131.60	129.53	129.18
C11 – Meta	132.81	130.19	127.02	126.90	130.18	130.57	129.92	130.93
C12 – Ortho	129.70	131.72	127.02	128.20	125.76	126.52	127.57	128.26
RMSE:		1.95		1.46		1.09		1.24

Table 3: RMS errors in the predicted ^{15}N isotropic shieldings for phenobarbital.

Method	RMSE (ppm)
GIPAW PBE	9.5
Fragment PBE	8.8
Fragment PBE0	5.5
Cluster PBE0	5.8
Cluster/Fragment PBE0	5.9

splittings.

Finally, Figure 10 illustrates the $\tilde{\chi}^2$ for each of the possible permutations for assigning the four phenobarbital monomers in these two polymorphs to the four sets of chemical shifts. The correct assignment (shown in red) has the lowest $\tilde{\chi}^2$ for every model. The most difficult discrimination (shown in blue) involves interchanging the molecules IIc and III, which have very similar intramolecular conformations and the same intermolecular hydrogen bonding motif. Each model identifies this mis-assignment as having a larger $\tilde{\chi}^2$, but the magnitude of the resolution is greater using the PBE0 fragment approach (see separation between red and blue lines in Figure 10a). In addition, the overall ^{13}C RMS error (bottom of Figure 10a)

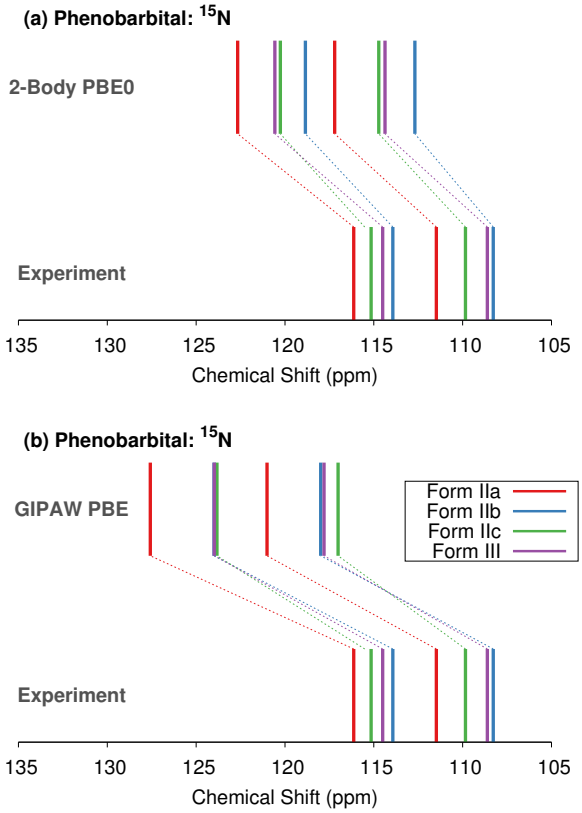


Figure 9: Comparison between experimental and predicted ^{15}N chemical shifts for forms II and III phenobarbital using either the (a) 2-body fragment PBE0 or (b) GIPAW PBE models.

is close to ∼0.2 ppm smaller using the hybrid density functional. All other mis-assignments interchange molecules with different hydrogen bonding and lead to a larger increase in $\tilde{\chi}^2$. For example, molecule IIa forms the same hydrogen bonded chains, but it also hydrogen bonds to IIb and has a slightly different intramolecular conformation, making it easier to distinguish.

Like the acetaminophen case, including ^{15}N isotropic shielding data in the $\tilde{\chi}^2$ analysis (Figure 10b) here does not significantly improve the resolution between the different possible assignments. Assignment of the two nitrogens within a given phenobarbital monomer is unambiguous from the calculations (see Figure 9). Mis-assignments between the polymorphs introduce errors of only ∼2–3 ppm, which is small relative to the errors expected for ^{15}N. Accordingly, the nitrogen shifts contribute little to the overall $\tilde{\chi}^2$ values. Nevertheless, with or without including the ^{15}N shift data, the fragment PBE0 shows roughly a two-fold

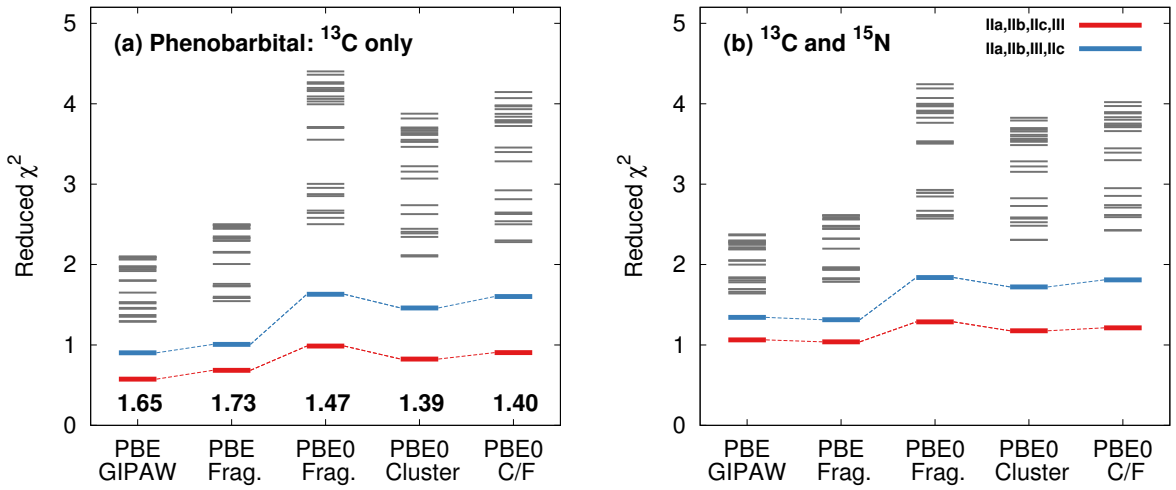


Figure 10: Reduced χ^2 analysis for phenobarbital using either (a) ^{13}C isotropic shifts or (b) both ^{13}C and ^{15}N isotropic chemical shifts. The red lines correspond to the correct assignment, while the blue lines swap the assignment of monomer c in form II with the form III monomer.

improvement in the $\tilde{\chi}^2$ resolution between the correct and incorrect structures over GIPAW PBE.

Testosterone

The results for acetaminophen and phenobarbital demonstrate that the chemical shift calculations are able to distinguish correct and incorrect assignments of ^{13}C spectra, despite strong similarities in hydrogen bonding arrangements between polymorphs. Another important situation is the distinction between neat and hydrate crystal forms, which we illustrate with the crystal forms of testosterone.

Two crystal forms of testosterone have been investigated using both cross-polarization MAS and two-dimensional carbon-carbon NMR experiments.[17] The α form contains two symmetrically inequivalent molecules in the unit cell (denoted u and v), while the β monohydrate contains one testosterone molecule and one water in the asymmetric unit (Figure 11). The intramolecular conformation is very similar in all three symmetrically unique molecules, with molecular structure overlay root-mean-square deviations of 0.05–0.06 Å for the non-hydrogen atoms (Figure 4c). The only major conformational difference is seen in

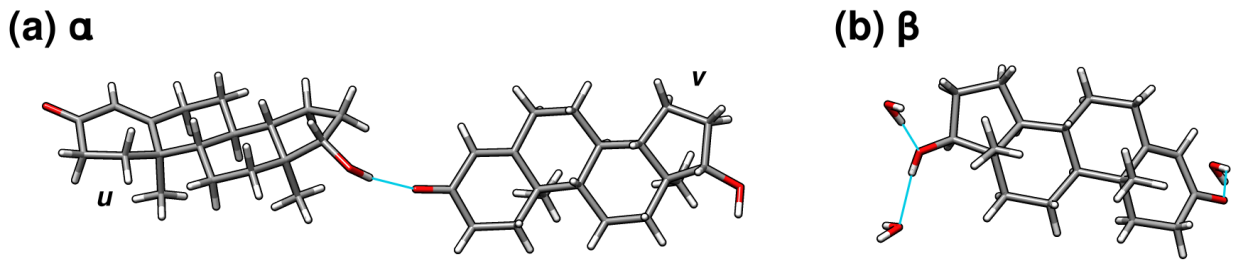


Figure 11: Structures showing the hydrogen bonding patters in the (a) α and (b) β forms of testosterone.

the hydroxyl orientation between αu and the conformation in αv and β, which affects the environment of C17. Accordingly, variations in the chemical shifts stem primarily from differences in the crystallographic environments of the three monomers. The presence of 19 unique carbon atoms in each testosterone molecule in the α form gives rise to a congested solid-state ^{13}C NMR spectra with numerous closely spaced peaks, especially in the ∼30–40 ppm region (see Figure 12). The assignment of each doublet of peaks to molecules u and v is particularly challenging.

Harris et al were able to assign most of the ^{13}C resonances through the use of two dimensional INADEQUATE carbon-carbon correlation experiments.[17] However, the experiments alone did not allow unambiguous assignment of the shifts for C3 and C4 to molecules u and v. Instead, they used GIPAW PBE calculations on the experimentally determined crystal structures (with only hydrogen atom positions relaxed) to predict the chemical shifts for this system and make tentative assignments for each of these two sets of carbon shifts. In addition, the u/v assignments of C7, C13, and C16 are suggested by the experiments, although not definitive. Finally, the experimental splitting in C15 is extremely small and might be interchanged. Table S3 in the supporting information lists the carbon assignments from Ref 17.

The large number of chemical shifts in a small region of the spectrum make testosterone an excellent system for assessing the performance of cluster and fragment-based NMR chemical shielding calculations. Figure 12 compares the predicted and experimental ^{13}C chemical

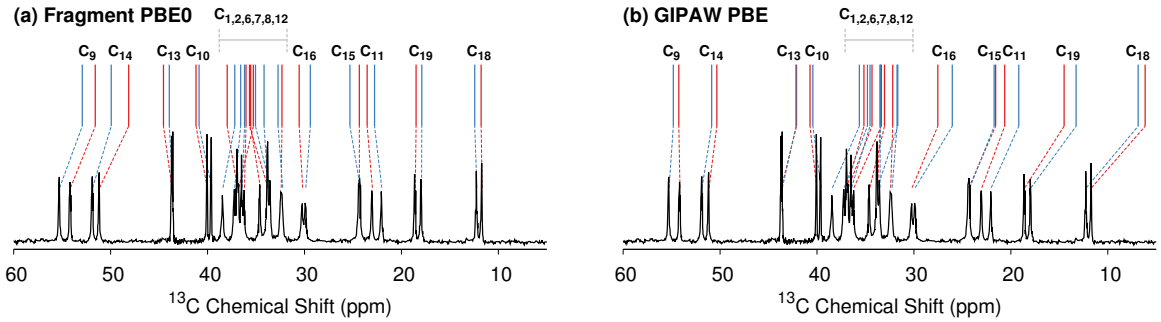


Figure 12: Comparison of experimental[17] and predicted ^{13}C chemical shifts in the low-frequency region for α-testosterone using either (a) 2-body fragment PBE0 or (b) GIPAW PBE. Shifts from molecule u and v are indicated in red or blue, respectively.

shifts for α-testosterone in the low-frequency region for two-body fragment PBE0 and GIPAW PBE. Qualitatively, the GIPAW PBE calculations place many of the chemical shifts in this region too far upfield, while the fragment PBE0 shifts show improved agreement with experiment.

Overall, the PBE0 fragment-based models reduce the RMS error for the α and β forms to around 1.9–2.1 ppm, compared to 3.3 ppm with PBE (Figure 14). Most individual ^{13}C chemical shifts are reproduced to within a few ppm (Table S3), with the notable exception of C5. In the earlier GIPAW PBE work by Harris et al,[17] this shift was overestimated by ∼11–14 ppm, which they attributed to a possible artifact of having relaxed only the hydrogen atoms in the crystal structure. However, even with our structures in which all atomic coordinates were relaxed with PBE-D2, C5 still exhibits errors of 11 ppm in the α form and 7.4 ppm in the β form at the GIPAW PBE level (Table 4). Similarly large errors occur for fragment PBE. Switching to the PBE0 functional improves the results moderately, but the errors remain ∼5–8.5 ppm. The reasons for these unusually large errors are unclear, but they may indicate the importance of dynamics or other issue with the crystal structure.

Consider next the assignments in α-testosterone that were experimentally ambiguous. Figure 13 plots the splittings between the shifts for molecules u and v. GIPAW PBE, fragment PBE0, and cluster/fragment PBE0 are generally in qualitative agreement with experiment. All three models predict the same assignments for the doublets ascribed to C3, C4,

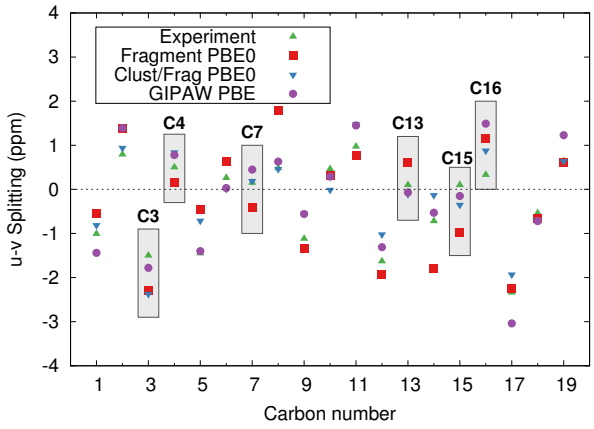


Figure 13: Comparison between the predicted and experimental splittings for the chemical shifts on molecules u and v in α-testosterone. Atoms discussed in the text are highlighted.

Table 4: Predicted ^{13}C isotropic shieldings for the C5 carbon in each of the testosterone polymorphs.

	α form				β form	
	Molecule u		Molecule v			
	Shift	Error	Shift	Error	Shift	Error
Experiment	170.64		172.09		173.75	
PBE0:						
2-Body Fragment	177.87	7.23	178.32	6.23	182.29	8.54
Cluster	175.98	5.34	176.90	4.81	182.33	8.58
Cluster/Fragment	176.24	5.60	176.95	4.86	182.08	8.33
PBE:						
2-Body Fragment	180.62	9.98	180.93	8.84	185.20	11.45
GIPAW	181.69	11.05	183.09	11.00	181.11	7.36

and C16. For C7 and C13, fragment PBE0 predicts the opposite sign for the splitting than either GIPAW PBE or cluster/fragment PBE0. However, in both cases, the magnitude of the experimental splittings is only ∼0.1–0.2 ppm, and the discrepancies among the predictions are less than 1 ppm. For C15, all the theoretical predictions suggest the opposite sign of splitting from the one inferred experimentally, which suggests that perhaps the experimental assignment should be reversed. Of course, the small splittings associated with all of these discrepancies between the models are probably below the threshold of significance, based on the 1.5–2.2 ppm errors expected for these models from earlier benchmark tests.

Finally, we investigate the ability of the models to discriminate among the different crys-

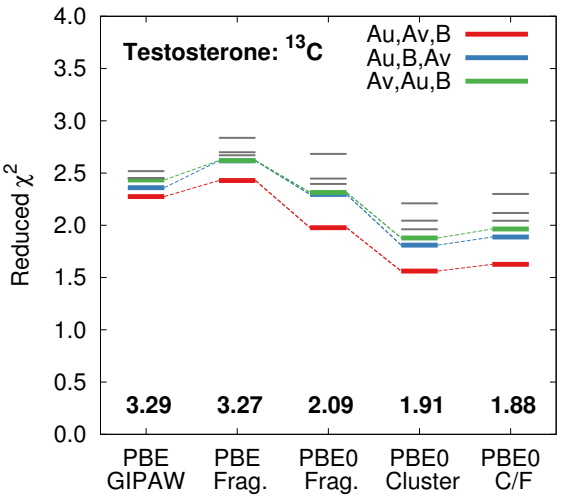


Figure 14: Reduced χ^2 analysis for the monomer assignments α and β forms of testosterone. The correct assignment is shown in red. The blue line swaps molecules αu and β, while the green line swaps molecules αu and αv

tallographic environments for the three unique monomers in the two crystal structures (using the atom assignments from Ref. 17). Six different ways of assigning the three structures to the three sets of isotropic shifts exist. Figure 14 demonstrates that all four computational methods produce the smallest RMS errors and reduced χ^2 for the assignment that is consistent with the earlier work by Harris et al,[17] despite the strong similarities in the intramolecular testosterone geometries (see Figure 11c). Due to the anomalously large errors for C5, the smallest $\tilde{\chi}^2$ values for testosterone are roughly double the corresponding values for acetaminophen and phenobarbital. Nevertheless, using the hybrid PBE0 functional instead of PBE both lowers the RMS error from 3.3 ppm to 1.9–2.1 ppm and modestly increases the resolution between the correct and incorrect monomer environment assignments (e.g. compare fragment PBE0 vs. fragment PBE or GIPAW PBE).

Accuracy of the chemical shielding regression models

The work here is predicated on the transferability of previously determined[70] linear regression models for scaling the chemical shieldings computed in the systems here. As mentioned

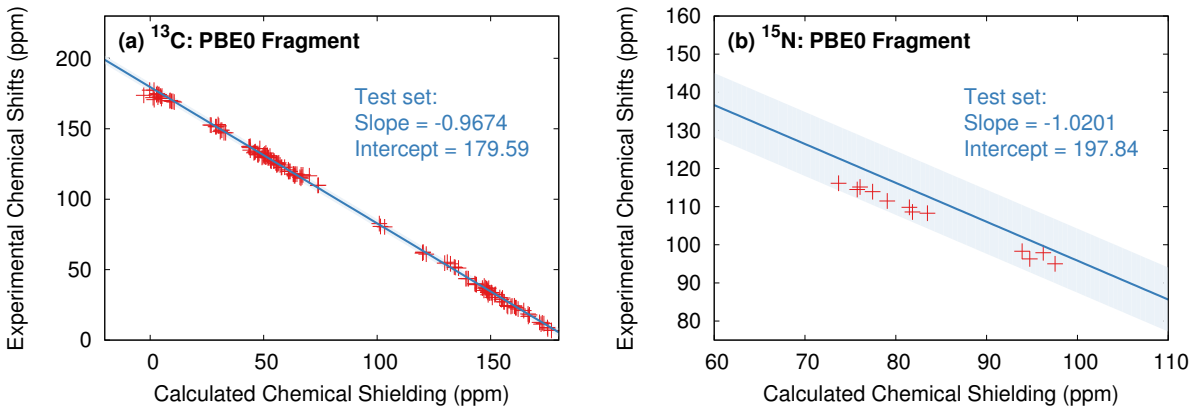


Figure 15: Comparison between the shielding regression line from Ref 70 (in blue) and the chemical shifts predicted here (in red) for (a) the ^{13}C isotropic shieldings and (b) the ^{15}N isotropic shieldings. The blue shaded region indicates plus or minus two standard deviations in the errors from the Ref 70 benchmarks.

above, the earlier ^{13}C regression model was fitted to data from 25 crystals and 169 isotropic shifts, while the nitrogen line was fitted to 24 crystals and 51 isotropic shifts. The resulting linear regression parameters were previously shown to be highly robust with respect to the specific composition of the test set through statistical cross-validation.[70]

To test this robustness further, we investigate the consistency between the chemical shieldings computed for the systems here and those from our previous work. As an example, Figure 15 illustrates the previously published fragment PBE0 regression lines (in blue) along with the predicted ^{13}C and ^{15}N isotropic shieldings for the crystals in the present study (shown in red). Note that acetaminophen form I was included in the original ^{13}C regression (8 out of the 169 shifts), but none of the other crystal forms considered here were used in fitting the regression models. Figure 15 demonstrates that the predicted ^{13}C shieldings for each of the polymorphs in this study are in excellent agreement with the earlier regression model. The ^{15}N regression line overestimates the twelve experimental ^{15}N shifts chemical shifts considered here, particularly for the 8 phenobarbital ones (calculated shieldings in the 70–85 ppm range). Nevertheless, the predictions still fall within two standard deviations (shaded region in Figure 15) of the data from the original benchmark set.

Next, Table 5 compares both the linear regression parameters and the RMS errors for

three different choices in fitting set: (1) Regression parameters from the previously reported ^{13}C test set with the acetaminophen form I shifts removed (24 crystals/161 shifts); (2) Regression parameters fitted to only the ^{13}C isotropic shift data from the polymorphic crystals studied in the present work (7 crystals/137 shifts); and (3) Regression parameters for the union of both data sets (31 crystals/298 shifts). For the fragment-based PBE0 model, the overall RMS errors on the polymorphic crystals considered here vary by less than ∼0.2 ppm across the three different sets, and the linear regression parameters vary only slightly (Table 5). The GIPAW RMS errors and regression parameters exhibit slightly larger variability across the three different training sets (e.g. RMS error variations of ∼0.4 ppm), but such deviations are well within the larger ∼2–2.5 ppm errors inherent in the model. These results provide strong support for the transferability of the regression parameters developed in Ref 70 previously and used here.

The same three fitting scenarios were considered for ^{15}N. Fitting directly to the 12 ^{15}N shifts here reduces the errors dramatically, down to 0.7 ppm with fragment PBE0. However, this fitting produces a slope that deviates substantially from unity. This could suggest problems in describing the nitrogen environments present in these molecules that are corrected in the direct fit, or perhaps some issues in the experimental shift referencing (particularly for phenobarbital). Because the ^{15}N benchmark set contains only 51 shifts, adding the dozen more shifts with a systematic error from this study does modestly alter the regression line and improves the errors for these twelve shifts by about 1 ppm at the fragment PBE0 level. Nevertheless, because including so many nitrogen shifts from only two systems might overly bias the overall test set, we continue to advocate use of the original regression line from Ref 70. Qualitatively similar behavior is observed for ^{15}N with GIPAW PBE, but the regression lines and rms errors exhibit even higher sensitivity to the composition of the fitting set.

Table 5: Impact of training set on linear regression parameters for the ^{13}C and ^{15}N isotropic shieldings. Reported slopes and intercepts were fitted using (1) only the previously published benchmark test set shifts (excluding form I acetaminophen from the ^{13}C test set), (2) only the polymorph shifts from this work, or (3) the combination of both the benchmark and polymorph shifts. The RMS errors are reported for the polymorphic crystal ^{13}C and ^{15}N isotropic shifts in the polymorphic crystals studied in this work.

Training Set	**Slope**	**Intercept**	**RMSE**	**Slope**	**Intercept**	**RMSE**
	^{13}C: Fragment PBE0			*^{15}N: Fragment PBE0*		
1. Benchmark Set	-0.9661	179.45	1.74	-1.0201	197.84	4.93
2. Polymorph Set	-0.9633	178.50	1.59	-0.8987	182.86	0.73
3. Combined Set	-0.9648	179.01	1.63	-1.0209	197.00	4.05
	^{13}C: GIPAW PBE			*^{15}N: GIPAW PBE*		
1. Benchmark Set	-0.9893	169.05	2.46	-1.0165	184.98	9.19
2. Polymorph Set	-0.9663	167.62	2.02	-0.8496	164.87	2.38
3. Combined Set	-0.9746	168.13	2.08	-1.0126	183.30	7.84

Conclusion

In conclusion, the relative performance of GIPAW, fragment, cluster and combined cluster/fragment models for predicted ^{13}C and ^{15}N isotropic chemical shieldings has been assessed in several different polymorphic crystal systems. Consistent with our recent benchmark studies,[47,70] the hybrid PBE0 density functional provides higher-accuracy chemical shifts than the GGA functional PBE. More important, this improved accuracy stemming from the use of a hybrid density functional provides increased discrimination among different crystallographic environments, which is an essential ingredient in NMR crystallography studies. The chemical shift prediction methods are able to reliably distinguish between the monomers in different crystal structures, even when the strong intermolecular interactions are nearly identical in the various structures.

The 2-body and GIPAW methods generally agree on the assignments for individual isotropic ^{13}C chemical shift assignments. In the handful of cases where the two models disagreed, the differences in chemical shifts were $\sim$1 ppm or less, which is below the resolving power of the models. In those cases, switching to a cluster or cluster/fragment model

produced assignments that are fully consistent with the GIPAW ones, suggesting that these minor discrepancies result from the simplified description of many-body effects in the 2-body fragment model. So, despite the successes of the 2-body fragment model demonstrated here, there may be room for refining the electrostatic embedding treatment further to improve the approximate treatment of many-body effects.

Finally, we demonstrated that the linear regression parameters for scaling chemical shieldings to chemical shifts which were developed in our previous benchmark study[70] are clearly transferable to the systems studied here, which bodes well for their widespread application. In the examples studied here, ^{13}C isotropic chemical shifts were sufficient to discriminate among the different crystal structures, but this is not always the case. In the future, it will be interesting to apply these fragment chemical shift prediction techniques to structure determination problems through their combination with crystal structure prediction techniques, where PBE GIPAW ^{13}C shifts alone have previously proved insufficient for identifying the correct structures.[11]

Acknowledgments

Funding for this work from the National Science Foundation (CHE-1362465 for JH and GB) and supercomputer time from XSEDE (TG-CHE110064) are gratefully acknowledged. GD thanks the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC through grant agreement n. 307358 (ERC-stG-2012-ANGLE), and acknowledges the ARCHER UK National Supercomputing Service via UK's HEC Materials Chemistry Consortium membership, which is funded by the EPSRC (EP/L000202). We thank Prof. Jonathan Burley (U. Nottingham) for sharing their experimental acetaminophen NMR spectra and for help with referencing the chemical shifts, and Prof. Leonard Mueller (U.C. Riverside) for numerous helpful discussions.

References

(1) Singhal, D.; Curatolo, W. *Adv. Drug Deliv. Rev.* **2004**, *56*, 335–47.

(2) Bernstein, J. *Cryst. Growth Des.* **2011**, *11*, 632–650.

(3) Censi, R.; Di Martino, P. *Molecules* **2015**, *20*, 18759–18776.

(4) Dudenko, D. V.; Williams, P. A.; Hughes, C. E.; Antzutkin, O. N.; Velaga, S. P.; Brown, S. P.; Harris, K. D. M. *J. Phys. Chem. C* **2013**, *117*, 12258–12265.

(5) Martineau, C.; Senker, J.; Taulelle, F. *Ann. Rep. NMR Spectros.* **2014**, *82*, 1–57.

(6) Filip, X.; Borodi, G.; Filip, C. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17978–17986.

(7) Elena, B.; Emsley, L. *J. Am Chem. Soc.* **2005**, *127*, 9140–9146.

(8) Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L. *J. Am. Chem. Soc.* **2006**, *128*, 9555–9560.

(9) Salager, E.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2610.

(10) Salager, E.; Day, G. M.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. *J. Am. Chem. Soc.* **2010**, *132*, 2564–2566.

(11) Baias, M.; Widdifield, C. M.; Dumez, J.-N.; Thompson, H. P. G.; Cooper, T. G.; Salager, E.; Bassil, S.; Stein, R. S.; Lesage, A.; Day, G. M.; Emsley, L. *Phys. Chem. Chem. Phys.* **2013**, *15*, 8069–8080.

(12) Baias, M.; Dumez, J.-N.; Svensson, P. H.; Schantz, S.; Day, G. M.; Emsley, L. *J. Am. Chem. Soc.* **2013**, *135*, 17501–7.

(13) Harper, J. K.; Grant, D. M. *Cryst. Growth Des.* **2006**, *6*, 2315–2321.

(14) Kalakewich, K.; Iuliucci, R.; Harper, J. K. *Cryst. Growth Des.* **2013**, *13*, 5391–5396.

(15) Harper, J. K.; Tishler, D.; Richardson, D.; Lokvam, J.; Pendrill, R.; Widmalm, G. *J. Phys. Chem. A* **2013**, *117*, 5534–5541.

(16) Santos, S. M.; Rocha, J.; Mafra, L. *Cryst. Growth Des.* **2013**, *13*, 2390–2395.

(17) Harris, R. K.; Joyce, S. A.; Pickard, C. J.; Cadars, S.; Emsley, L. *Phys. Chem. Chem. Phys.* **2006**, *8*, 137–43.

(18) Webber, A. L.; Emsley, L.; Claramunt, R. M.; Brown, S. P. *J. Phys. Chem. A* **2010**, *114*, 10435–10442.

(19) Abraham, A.; Apperley, D. C.; Gelbrich, T.; Harris, R. K.; Griesser, U. J. *Can. J. Chem.* **2011**, *89*, 770–778.

(20) Kucukbenli, E.; Sonkar, K.; Sinha, N.; Gironcoli, S. D. *J. Phys. Chem. A* **2012**, *116*, 3765–3769.

(21) Apperley, D. C.; Batsanov, A. S.; Clark, S. J.; Harris, R. K.; Hodgkinson, P.; Jochym, D. B. *J. Mol. Struct.* **2012**, *1015*, 192–201.

(22) Olsen, R. A.; Struppe, J.; Elliott, D. W.; Thomas, R. J.; Mueller, L. J. *J. Am. Chem. Soc.* **2003**, *125*, 11784–11785.

(23) Witter, R.; Sternberg, U.; Hesse, S.; Kondo, T.; Koch, F. T.; Ulrich, A. S. *Macromol.* **2006**, *39*, 6125–6132.

(24) Johnston, J. C.; Iuliucci, R. J.; Facelli, J. C.; Fitzgerald, G.; Mueller, K. T. *J. Chem. Phys.* **2009**, *131*, 144503.

(25) Harris, R. K.; Hodgkinson, P.; Zorin, V.; Dumez, J. N.; Herrmann, B. E.; Emsley, L.; Salager, E.; Stein, R. S. *Magn. Reason. Chem.* **2010**, *48*, S103–S112.

(26) Carignani, E.; Borsacchi, S.; Bradley, J. P.; Brown, S. P.; Geppi, M. *J. Phys. Chem. C* **2013**, *117*, 17731–17740.

(27) Harper, J. K.; Iuliucci, R.; Gruber, M.; Kalakewich, K. *Cryst. Eng. Comm.* **2013**, *15*, 8693–8704.

(28) Kalakewich, K.; Iuliucci, R.; Mueller, K. T.; Eloranta, H.; Harper, J. K. *J. Chem. Phys.* **2015**, *143*, 1–10.

(29) Paluch, P.; Pawlak, T.; Oszajca, M.; Lasocha, W.; Potrzebowski, M. J. *Solid State Nuc. Magn. Reson.* **2015**, *65*, 2–11.

(30) Pindelska, E.; Szeleszczuk, L.; Pisklak, D. M.; Majka, Z.; Kolodziejski, W. *J. Pharm. Sci.* **2015**, *104*, 2285–2292.

(31) Portieri, A.; Harris, R. K.; Fletton, R. A.; Lancaster, R. W.; Threlfall, T. L. *Magn. Reson. Chem.* **2004**, *42*, 313–20.

(32) Braun, D. E.; McMahon, J. A.; Koztecki, L. H.; Price, S. L.; Reutzel-Edens, S. M. *Cryst. Growth Des.* **2014**, *14*, 2056–2072.

(33) Braun, D. E.; Koztecki, L. H.; McMahon, J. a.; Price, S. L.; Reutzel-Edens, S. M. *Molec. Pharm.* **2015**, *12*, 3069–3088.

(34) Brown, S. P. *Solid State Nuc. Magn. Reson.* **2012**, *41*, 1–27.

(35) Dračínský, M.; Hodgkinson, P. *RSC Adv.* **2015**, *5*, 12300–12310.

(36) Wei, Y.; de Dios, A. C.; McDermott, A. E. *J. Am. Chem. Soc.* **1999**, *121*, 10389–10394.

(37) Lemaitre, V.; Smith, M. E.; Watts, A. *Soild State Nuc. Magn. Reas.* **2004**, *26*, 215–235.

(38) Gervais, C.; Dupree, R.; Pike, K. J.; Bonhomme, C.; Profeta, M.; Pickard, C. J.; Mauri, F. *J. Phys. Chem. A* **2005**, *109*, 6960–6969.

(39) Wong, A.; Pike, K. J.; Jenkins, R.; Clarkson, G. J.; Anupold, T.; Howes, A. P.; Crout, D. H.; Samoson, A.; Dupree, R.; Smith, M. *J. Phys. Chem. A* **2006**, *110*, 1824–1835.

(40) Pickard, C.; Mauri, F. *Phys. Rev. B* **2001**, *63*, 245101.

(41) Yates, J. R.; Pickard, C. J.; Mauri, F. *Phys. Rev. B* **2007**, *76*, 024401.

(42) Bonhomme, C.; Gervais, C.; Babonneau, F.; Coelho, C.; Pourpoint, F.; Azais, T.; Ashbrook, S. E.; Griffin, J. M.; Yates, J. R.; Mauri, F.; Pickard, C. J. *Chem. Rev.* **2012**, *112*, 5733–5779.

(43) Beran, G. J. O. *Chem. Rev.* **2016**, *116*, 5567–5613.

(44) Ashbrook, S. E.; McKay, D. *Chem. Commun.* **2016**, *52*, 7186–7204.

(45) Holmes, S. T.; Iuliucci, R. J.; Mueller, K. T.; Dybowski, C. *J. Chem. Phys.* **2014**, *141*, 164121.

(46) Holmes, S. T.; Iuliucci, R. J.; Mueller, K.; Dybowski, C. *J. Chem. Theory Comput.* **2015**, *11*, 5229–5241.

(47) Hartman, J. D.; Monaco, S.; Schatschneider, B.; Beran, G. J. O. *J. Chem. Phys.* **2015**, *143*, 102809.

(48) Ferraro, M. B.; Facelli, J. C. *J. Mol. Struct.* **2002**, *603*, 159–164.

(49) Stueber, D. *Concepts Magn. Reson. A* **2006**, *28*, 347–368.

(50) Ochsenfeld, C.; Kussmann, J.; Koziol, F. *Angew. Chem. Int. Ed.* **2004**, *43*, 4485–4489.

(51) Zienau, J.; Kussmann, J.; Ochsenfeld, C. *Mol. Phys.* **2010**, *108*, 333–342.

(52) Maurer, M.; Ochsenfeld, C. *J. Chem. Phys.* **2013**, *138*, 174104.

(53) Loibl, S.; Manby, F. R.; Schütz, M. *Mol. Phys.* **2010**, *108*, 477–485.

(54) Loibl, S.; Schütz, M. *J. Chem. Phys.* **2012**, *137*, 084107.

(55) Ditchfield, R. *Mol Phys.* **1974**, *27*, 789.

(56) Zheng, A.; Yang, M.; Yue, Y.; Ye, C.; Deng, F. *Chem. Phys. Lett.* **2004**, *399*, 172–176.

(57) Chen, X.; Zhan, C.-G. *J. Mol. Struct. (THEOCHEM)* **2004**, *682*, 73–82.

(58) Zheng, A.; Chen, L.; Yang, J.; Yue, Y.; Ye, C.; Lu, X.; Deng, F. *Chem. Commun.* **2005**, 2474.

(59) Vila, J. A.; Scheraga, H. A. *Acc. Chem. Res.* **2009**, *42*, 1545–53.

(60) Gao, Q.; Yokojima, S.; Kohno, T.; Ishida, T.; Fedorov, D. G.; Kitaura, K.; Fujihira, M.; Nakamura, S. *Chem. Phys. Lett.* **2007**, *445*, 331–339.

(61) Gao, Q.; Yokojima, S.; Fedorov, D. G.; Kitaura, K.; M, S.; S, N. *J. Chem. Theory Comput.* **2010**, *6*, 1428–1444.

(62) Frank, A.; Onila, I.; Möller, H. M.; Exner, T. E. *Proteins* **2011**, *79*, 2189–202.

(63) Tang, S.; Case, D. A. *J. Biomol. NMR* **2011**, *51*, 303–12.

(64) Frank, A.; Moller, H. M.; Exner, T. E. *J. Chem. Theory Comput.* **2012**, *8*, 1480–1492.

(65) He, X.; Wang, B.; Merz, K. M. *J. Phys. Chem. B* **2009**, *113*, 10380–8.

(66) Zhu, T.; He, X.; Zhang, J. Z. H. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7837–45.

(67) He, X.; Zhu, T.; Wang, X.; Liu, J.; Zhang, J. Z. H. *Acc. Chem. Res.* **2014**, *47*, 2748–57.

(68) Flaig, D.; Beer, M.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2012**, *8*, 2260–2271.

(69) Reid, D. M.; Collins, M. A. *Phys. Chem. Chem. Phys.* **2015**, *17*, 5314–5320.

(70) Hartman, J. D.; Kudla, R. A.; Day, G. M.; Mueller, L. J.; Beran, G. J. O. *Phys. Chem. Chem. Phys.* **2016**, *in press*.

(71) Hartman, J. D.; Beran, G. J. O. *J. Chem. Theory Comput.* **2014**, *10*, 4862–4872.

(72) Burley, J.; Duer, M.; Stein, R.; Vrcelj, R. *European J. of Pharm. Sci.* **2007**, *31*, 271–276.

(73) Stone, K.; Lapidus, S.; Stephens, P. *J. Appl. Crystallogr.* **2009**, *42*, 385.

(74) Drebushchak, T.; Boldyerva, E. *Z. Kristallogr.* **2004**, *219*, 506.

(75) Perrin, M.; Neumann, M.; Elmaleh, H.; Zaske, L. *Chem. Commun.* **2009**, 3181–3183.

(76) Platteau, C.; Lefebvre, J.; Hemon, S.; Baehtz, C.; Florence, D.; Prevost, D. *Acta Crystallogr.,Sect.B:Struct.Sci.* **2005**, *61*, 80.

(77) Zencirci, N.; Gelbrich, T.; Apperley, D. C.; Harris, R. K.; Kahlenberg, V.; Griesser, U. J. *Cryst. Growth Des.* **2010**, *10*, 302.

(78) Roberts, P.; Pettersen, R.; Sheldrick, G.; Isaacs, N.; Kennard, O. *J. Chem. Soc.* **1973**,

(79) Precigoux, G.; Hospital, M.; Bosche, G. *Cryst. Struc. Commun.* **173**, *2*, 435.

(80) Harris, R. K.; Becker, E. D.; Cabral de Menezes, S. M.; Granger, P.; Hoffman, R. E.; Zilm, K. W. *Pure Appl. Chem.* **2008**, *80*, 59–84.

(81) Giannozzi, P. et al. *J. Phys. Cond. Mat.* **2009**, *21*, 395502 (19pp).

(82) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(83) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293.

(84) Beran, G. J. O. *J. Chem. Phys.* **2009**, *130*, 164115.

(85) Beran, G. J. O.; Nanda, K. *J. Phys. Chem. Lett.* **2010**, *1*, 3480–3487.

(86) Wen, S.; Beran, G. J. O. *J. Chem. Theory Comput.* **2011**, *7*, 3733–3742.

(87) Frisch, M. J. et al. Gaussian 09 Revision A.1. 2009; Gaussian Inc. Wallingford CT.

(88) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

(89) GDMA, Distributed Multipole Analysis of Gaussian Wavefunctions, version 2.2.09, A. J. Stone. `http://www-stone.ch.cam.ac.uk/pub/gdma/`. Accessed May 28, 2014.

(90) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.

(91) Chesnut, D. B.; Moore, K. D. *Journal of Computational Chemistry* **1989**, *10*, 648–659.

(92) Chesnut, D. B.; Rusiloski, B. E.; Moore, K. D.; Egolfs, D. A. *J. Comp. Chem.* **1993**, *14*, 1364–1375.

(93) Hartman, J. D.; Neubauer, T. J.; Caulkins, B. G.; Mueller, L. J.; Beran, G. J. O. *J. Biomol. NMR* **2015**, *62*, 327–340.

(94) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I.; Refson, K.; Payne, M. C. *Z. Kristallogr.* **2005**, *220*, 567–570.

(95) Lodewyk, M.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839–62.

(96) Haisa, M.; Kashino, S.; Maeda, H. *Acta Crystallogr.* **1974**, *B30*, 2510–2512.

(97) Al-Zoubi, N.; Kachrimanis, K.; Malamataris, S. *Euro. J. Pharm. Sci.* **2002**, *17*, 13–21.

(98) Al-Zoubi, N.; Malamataris, S. *Int. J. Pharm.* **2003**, *260*, 123–135.

(99) Mikhailenko, M. *J. Cryst. Growth* **2004**, *265*, 616–618.

(100) Nichols, G.; Frampton, C. *J. Pharm. Sci.* **1998**, *87*, 684–693.

(101) Peterson, M.; Morissette, S.; McNulty, C.; Goldsweig, A.; Shaw, P.; LeQuesne, M.; Monagle, J.; Encina, N.; Marchionna, J.; Johnson, A.; Gonsalez-Zugasti, J.; Lemmo, A.; Ellis, S.; Cima, M.; Almarsson, . *J. Am. Chem. Soc.* **2002**, *124*, 10958–10959.

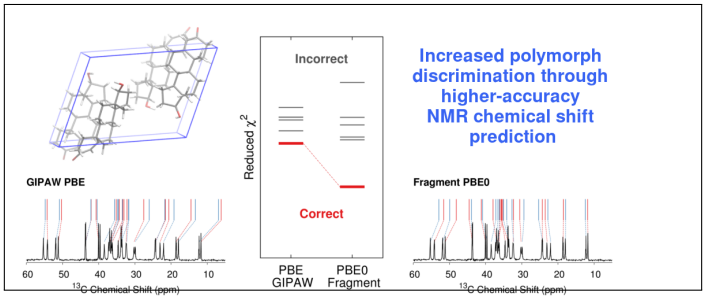
(102) Martino, P. D.; Guyot-Hermann, A.; Conflant, P.; Drache, M.; Guyot, J. *J. Therm. Anal. Calorim.* **1997**, *48*, 447–458.

(103) Jagannathan, N. R. *Curr. Sci.* **1987**, *56*, 827–830.

(104) Roy, S.; Goud, N. R.; Matzger, A. J. *Chem. Commun.* **2016**, *52*, 4389–4392.

(105) Gelbrich, T.; Braun, D. E.; Griesser, U. J. *Chem. Cent. J.* **2016**, *10*, 1–21.

Graphical TOC Entry



Synopsis: Fragment-based electronic structure methods allow nuclear magnetic resonance chemical shifts to be predicted with higher accuracy and lower computational cost. Higher-accuracy chemical shifts leads to increased discrimination between correct and incorrect structural assignments in the NMR crystallography of polymorphic molecular crystals.