PCCP

Improving the accuracy of solid-state nuclear magnetic resonance chemical shift prediction with a simple molecular correction

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-03-2019-001666.R1
Article Type:	Paper
Date Submitted by the Author:	31-May-2019
Complete List of Authors:	Dracinsky, Martin; Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Czech Repu, NMR spectroscopy Unzueta, Pablo; University of California Riverside, Chemistry Beran, Gregory; University of California Riverside, Chemistry

PAPER

Improving the accuracy of solid-state nuclear magnetic resonance chemical shift prediction with a simple molecular correction

Received 00th January 2019,
Accepted 00th January 2019

DOI: 10.1039/x0xx00000x

Martin Dračínský,*[a] Pablo Unzueta[b] and Gregory J. O. Beran[b]

A fast, straightforward method for computing NMR chemical shieldings of crystalline solids is proposed. The method combines the advantages of both conventional approaches: periodic calculations using plane-wave basis sets and molecular computational approaches. The periodic calculations capture the periodic nature of crystalline solids, but the computational level of the electronic structure calculation is limited to general-gradient-approximation (GGA) density functionals. It is demonstrated that a correction to the GGA result calculated on an isolated molecule at a higher level of theory significantly improves the correlations between experimental and calculated chemical shifts while adding almost no additional computational cost. Corrections calculated with a hybrid density functional improved the accuracy of ^{13}C, ^{15}N and ^{17}O chemical shift predictions significantly and allowed identifying errors in previously published experimental data. Applications of the approach to crystalline isocytosine, methacrylamide, and testosterone are presented.

Introduction

Solid-state NMR spectroscopy (SS-NMR) plays an indispensable role in the characterization of solids. In past two decades, the progress of SS-NMR methods has led to the development of NMR crystallography, which combines experimental SS-NMR data with theoretical simulations to obtain otherwise inaccessible insights into the structure and dynamics of solid materials. The recent rapid development of NMR crystallography has been greatly facilitated by the availability of fast and reliable computational methods that enable direct linking between structure and NMR observables.

Two main approaches are used to predict NMR parameters (and other properties) of crystalline solids. First, solids can be modelled as infinite crystals using periodic boundary conditions (PBC) that ensure that a basic structural element (typically a crystallographic unit cell) is periodically repeated in all three dimensions. Second, a small part of an infinite crystal can be modelled as a molecular cluster or using small fragments. Both computational approaches have certain advantages and limitations.

The PBC approach exploits the translational repetition in crystals. Inherently periodic plane waves are used to form a basis set, instead of the local atomic orbital basis sets typically employed in molecular calculations. Because rapid variations in electron density are difficult to describe with plane waves, effective-core pseudopotentials are used to describe interactions close to the nuclei. Almost two decades ago, the gauge-including projector-augmented wave (GIPAW) procedure was developed for the prediction of the magnetic-resonance parameters in crystalline materials.[1] The method has been implemented in several density functional theory (DFT) software packages and it has been used successfully in many applications.[2-4] Unfortunately, hybrid density functionals are prohibitively demanding computationally for plane-wave calculations, and therefore, the GIPAW method has been used with the general-gradient-approximation (GGA) family of density functionals. However, many studies have demonstrated that going beyond the GGA level improves the accuracy of the predicted NMR parameters.[5-9]

On the other hand, in the cluster approach, neighboring molecules or fragments are considered explicitly during the NMR calculations and traditional molecule-based software packages may be used for the calculations.[5, 9-16] Although there is no fundamental limitation on the level of theory that can be used to compute the chemical shieldings in the fragments or cluster, the choice of the cluster size may be limiting, as the calculations must be maintained at a manageable size.[17] NMR parameters are generally mostly sensitive to the local environment. However, there are effects, such as electrostatic effects and ring currents, where long-range interactions are significant. It has been demonstrated that relatively large clusters have to be used for accurate predictions of NMR parameters.

Fragment methods reduce the computational costs of cluster calculations by replacing a large, many-molecule cluster with a series of electrostatically monomer and dimer calculations.[18] Drawing inspiration from the earlier embedded-ion model and related approaches, the self-consistent reproduction of the Madelung potential (SCRMP) model[7] embeds these fragments electrostatically in a field of point

[a] Institute of Organic Chemistry and Biochemistry, AS CR, Flemingovo 2, Prague, Czech Republic. E-mail: dracinsky@uochb.cas.cz
[b] Department of Chemistry, University of California, Riverside, California 92521, USA.

†Electronic Supplementary Information (ESI) available: All crystal structures, experimental NMR spectra and correlations between experimental and predicted chemical shifts. See DOI: 10.1039/x0xx00000x

charges designed to mimic the crystalline environment. Benchmark calculations on both isotropic shifts[7] and the principal components of the chemical shielding anisotropy (CSA) tensor[8] demonstrate very good performance of these fragment methods when hybrid density functionals are used, especially for ^{13}C and ^{15}N NMR parameters. For ^{17}O, these fragment methods exhibit a moderate degradation due to the high sensitivity of that nucleus to the electrostatic environment.

Here, we propose a fast, straightforward method for computing NMR chemical shieldings that combines the advantages of both plane-wave and molecular computational approaches, capturing the fully periodic nature of the crystal while also obtaining the higher accuracy associated with computational models beyond GGA DFT functionals. This simple method performs a standard periodic GIPAW GGA calculation and then corrects it based on single, non-embedded gas-phase molecule calculations at any higher level of theory. This approach has roots in the incremental methods pioneered by Stoll and others decades ago.[19, 20] Recently, Boese and co-workers have presented a similar strategy for molecular crystal energies based on periodic DFT or density functional tight binding corrected with higher-level monomer and dimer corrections.[21, 22] We demonstrate that the new method significantly improves the correlations between experimental and calculated chemical shifts while adding almost no additional computational cost.

Theory and Methods

The greatest advantage of GIPAW calculations is that they inherently contain long-range interactions in crystals. On the other hand, the advantage of cluster calculations is that any computational level, such as hybrid DFT functionals or post-Hartree-Fock methods can be used. The idea behind the newly proposed method is that the inaccuracy of GGA functionals for NMR shielding calculations is mostly limited to close (intramolecular) neighborhood of the nucleus of interest and long-range effects are well-approximated by the GGA-based GIPAW method. Therefore, we add a correction to the GIPAW calculated shieldings that is calculated as the difference between the shielding calculated at a higher computational level and at the GGA-level employed in the GIPAW calculation. These corrections are calculated for a single isolated molecule in the geometry taken from the crystal structure. The corrected shielding for a given atom (σ_{corr}) is calculated, for example, according to equation (1), where hybrid PBE0 functional is applied to correct PBE-GIPAW shieldings.

$$\sigma_{corr} = \sigma(\text{GIPAW,cryst.}) - \sigma(\text{PBE,mol.}) + \sigma(\text{PBE0,mol.}) \quad (1)$$

The proposed method consists of three basic steps: 1) Geometry optimization of the crystal structure obtained by X-ray or neutron diffraction experiment and calculation of NMR chemical shieldings using the GIPAW method. 2) Calculation of NMR shieldings for a single molecule taken from the geometry-optimized structure obtained in step 1). The calculations are performed at the same level as the GIPAW calculation (typically the PBE functional) and at a higher computational level (typically a hybrid functional, such as PBE0). 3) Evaluation of the corrected shieldings according to equation 1.

Separate benchmark sets of molecular crystal structures were used to evaluate the effect of the proposed method on the agreement with experimental data of carbon, nitrogen and oxygen nuclei. All benchmark sets are based on benchmark sets used in previous studies of fragment-based chemical shift predictions in molecular crystals.[5] The benchmarks here consist of 21 structures with 132 chemical shifts in the carbon set, 16 structures and 37 shifts in the nitrogen set and 15 structures and 28 shieldings in the oxygen set. The chemical structures of all systems studied are shown in the Electronic Supplementary Information (ESI).

The NMR shieldings of the studied structures were calculated by the CASTEP program,[23] version 17.2, which is a DFT-based code that uses pseudopotentials to model the effects of core electrons, and plane waves to describe the valence electrons. Positions of all atoms were optimized prior to the NMR calculation; the unit cell parameters were fixed. Electron-correlation effects were modeled using the generalized gradient approximation of Perdew, Burke, and Ernzerhof.[24] A plane wave basis set energy cutoff of 600 eV, default 'on-the-fly generation' pseudopotentials, and a *k*-point spacing of 0.05 $Å^{-1}$ over the Brillouin zone via a Monkhorst-Pack grid[25] was used. The NMR calculations were performed using the GIPAW approach.[1, 26] For comparison, the structures in the carbon set were also optimized using empirical dispersion correction, but the resulting calculated chemical shifts and corrected chemical shifts were almost identical to those obtained without the correction.[27, 28] The use of the fixed experimental unit cell parameters compensates for the artificially repulsive nature of the uncorrected density functionals. Finite temperature effects[29, 30] were not included in the calculations. However, constraining the lattice parameters to their experimental room-temperature values effectively captures the thermal expansion that occurs upon heating the crystal to room temperature.[31]

DFT NMR shieldings for the isolated molecules (in vacuum) were calculated by the Gaussian16 program.[32] For co-crystals, solvates, or salts, the molecular correction was performed only on the molecule whose shielding was of interest, without the other coformer species. The gas-phase molecule input geometries were taken from the periodic DFT geometry-optimized structures and were not further optimized. To explore how the results depend on the choice of the Gaussian basis set employed, the 6-31G(d), 6-311+G(2d,p), and pcSseg-n (n=1-3) were selected as representative basis sets for NMR shielding calculations. The pcSseg-n basis sets were obtained from basis set exchange website (https://bse.pnl.gov/bse/portal).[33] NMR shieldings at the coupled cluster singles and doubles (CCSD) level and 6-311+g(2d,p) basis set were calculated with CFOUR program package, which is suitable for performing high-level quantum chemical calculations on atoms and molecules.[34, 35]

Corrected shieldings were obtained using equation (1). The correlation between the corrected shieldings and experimental

chemical shifts was fitted to a straight line, $\delta = A + B\sigma$, where σ is the computed chemical shielding and δ corresponds to the experimentally observable chemical shift. The *A* and *B* parameters of this linear correlation were used for the calculations of chemical shifts, which were then compared with experimental data. The slope *B* of the shielding-shift correlation should equal –1 in an ideal case, but it has been shown previously that nuclear quantum effects,[36] incomplete basis sets, and other systematic errors in the DFT calculations can lead to deviations from this ideal value.

Experimental chemical shifts were re-measured for a few crystals in the test sets to correct issues with the earlier experiments. Solution-state NMR spectra of adenosine in DMSO-d_6 were recorded on Bruker Avance 500 (^{1}H at 500 MHz, ^{13}C at 125.8 MHz) spectrometer. The spectra were referenced to the residual solvent signal (2.50 ppm for ^{1}H and 39.7 ppm for ^{13}C). A combination of 1D and 2D experiments (H,H-COSY, H,C-HSQC and H,C-HMBC) was used to assign all proton and carbon signals.

High-resolution ^{13}C solid-state NMR spectrum of adenosine, L-cysteine, L-glutamine, L-threonine and L-tyrosine were obtained using a JEOL ECZ600R spectrometer operating at 150.9 MHz for ^{13}C and 600.2 MHz for ^{1}H and. Samples were packed into 3.2 mm magic angle spinning rotors (MAS) and measurements taken at MAS rate of 18 kHz using cross polarization (CP). The chemical shifts were referenced to crystalline α-glycine as a secondary reference (δ_{st} = 176 ppm for the carbonyl carbon). The ramped amplitude shape pulse was used during the cross-polarization. The contact time for CP was 5 ms and the relaxation delays were estimated from ^{1}H saturation recovery experiments and ranged from 3 s for L-threonine to 200 s for adenosine. The assignment of the signals was done with the help of a CPMAS experiment with a short contact time (50 μs), where the signals of quaternary carbons are suppressed. Furthermore, a C,H-HETCOR experiment was done with the L-glutamine sample to assign unequivocally the two C=O carbon signals.

Experimental chemical shifts of other systems were taken from refs.[5, 10, 37] and references therein.

Results and Discussion

Carbon isotropic shifts

At first sight, the chemical shifts obtained from uncorrected GIPAW shieldings correlate well with the experimental data (Figure S4 in the ESI) and the mean absolute error (MAE, Table 1) of 1.6 ppm looks also reasonable. However, a closer inspection of the data shows that 26% of the differences between experimental and calculated chemical shifts are larger than 2 ppm, 14% are larger than 3 ppm and the maximal error of 6.8 ppm is quite large.

Correcting the chemical shieldings with molecular PBE0/6-311+G(2d,p) calculations according to the newly proposed method improves agreement with experiment considerably; the MAE drops to 0.8 ppm and the maximal error is 3.9 ppm. Only one out of the 132 (0.8 %) calculated carbon chemical shifts differs by more than 3 ppm from the experimental shift and eight (6.1 %) shifts differ by 2–3 ppm. All the remaining shifts (93 %) are predicted with accuracy better than 2 ppm. The violin plots in Figure 1 visualize how adding the PBE0 molecular correction tightens the error distribution about zero error. The corrected GIPAW results have the same mean absolute and maximum errors as the PBE0 results obtained using the self-consistent charge embedded fragment approach SCRMP,[7] as seen in Figure 1.

One might wonder if the combination of plane-wave GIPAW and Gaussian basis set molecular calculations here could conceivably prove problematic due to differing degrees of basis set completeness in the two calculations. To investigate this possibility, the monomer correction to the chemical shielding was also evaluated with several additional Gaussian basis sets. For each possible basis set, a new linear regression was fitted on the data to convert the shieldings to chemical shifts. As shown in Figure 1, the quality of the molecular correction is quite insensitive to basis set. Even the small and computationally inexpensive 6-31G* basis gives results of nearly equal quality, with a MAE of 0.9 ppm and a maximum error of 3.6 ppm. The systematically growing pcSseg-*n* basis sets were also tested for *n*=1-3, and all three gave similar mean absolute errors of 0.9 ppm and maximum errors ranging 3.7-4.0 ppm.

Recently, Hartman and Beran used the SCRMP method to predict the three principal components (σ_{11}, σ_{22}, and σ_{33}) of the chemical shielding anisotropy (CSA) tensor.[8] Using the experimental data collected there for the crystals used in the present study, Figure 2 compares the errors of each method for reproducing each experimental principal component. Employing the monomer hybrid density functional correction to GIPAW PBE CSA tensors significantly improves their accuracy, with mean absolute errors reducing from 3.2 ppm to 2.3 ppm, and giving accuracy very similar to that obtained with PBE0 using the SCRMP fragment model. Using the same computed and experimental data, the error distributions were also evaluated for the chemical shielding anisotropy and asymmetry (Haeberlen convention), as shown in Figures S6 and S7 of the ESI. The behavior observed for the anisotropy, mimics that seen for the principal components in Figure 2: GIPAW PBE performs well (MAE 4.5 ppm), but the SCRMP and corrected GIPAW results perform appreciably better (MAE 3.0-3.3 ppm). On the other hand, no significant difference is observed among GIPAW PBE, SCRMP PBE0, and the corrected GIPAW models for the asymmetry. All methods tested give MAE of 0.08-0.09, and maximum errors of ~0.4 ppm.

The high accuracy of the corrected GIPAW approach actually helped us identify errors in the experimental data for several of the systems in the test set. When comparing the experimental and calculated carbon chemical shifts, we noticed particularly large errors for adenosine, L-cysteine, L-glutamine, L-threonine and L-tyrosine systems. Therefore, we re-examined the experimental data of these systems.

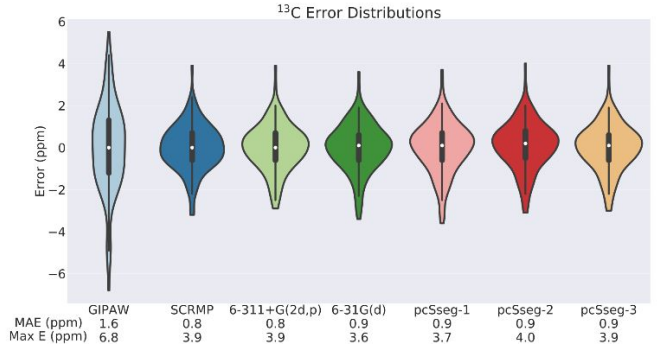

Figure 1. Errors in the ^{13}C chemical shift predictions from GIPAW PBE against those with the gas-phase monomer PBE0 corrections computed in various basis Gaussian sets. Violin plots indicate the kernel density estimate of the error distributions. Boxplots in the interior of each violin indicate the median (white dot), middle two quartiles (black box), and outer quartile data (within a factor of 1.5 times the inner quartile range).

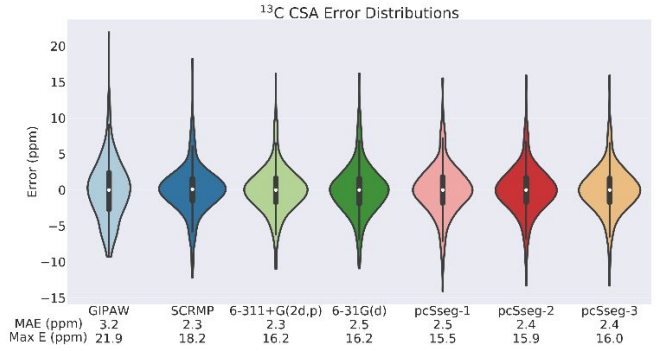

Figure 2. Errors in the principal components of the ^{13}C chemical shift anisotropy tensor predictions from GIPAW PBE against those with the gas-phase monomer PBE0 corrections computed in various basis Gaussian sets.

Table 1. Mean absolute errors (*MAE*) and maximal errors (ppm) of the predicted chemical shifts (in comparison with experiment) obtained for the conventional GIPAW method (PBE functional), corrected GIPAW (PBE0 correction, 6-311+g(2d,p) basis set) and for previously proposed SCRMP fragment method.[7]

Nucleus	Method	*MAE*	Max. error
^{13}C	GIPAW	1.6	6.8
	GIPAW-corrected	0.8	3.9
	SCRMP 1+2-body	0.8	3.9
^{15}N	GIPAW	4.1	10.6
	GIPAW-corrected	2.8	8.3
	SCRMP 1+2-body	2.8	7.7
^{17}O	GIPAW	5.2	11.6
	GIPAW-corrected	4.3	10.4
	SCRMP 1+2-body	5.9	14.1

The experimental ^{13}C SS-NMR chemical shifts of adenosine were taken from ref.,[38] where the assignment of the signals was based on a comparison of the SS-NMR spectrum with solution-state spectrum. However, in the correlation of these experimental data with NMR shieldings calculated with the newly proposed method, one can notice that the assignment of carbon atoms C2' and C3' seems to be interchanged (Figure S9 in the ESI). We re-measured adenosine in solution and using a combination of 1D and 2D NMR experiments, we unambiguously assigned all carbon signals (details in the SI). These experiments revealed that, indeed, that chemical shifts of C2' and C3' were wrongly assigned in the original report.

The experimental ^{13}C SS-NMR chemical shifts of L-cysteine used in previous studies for comparison with calculated data were taken from ref.,[39] where chemical shifts and CSAs of 20 amino acids were reported. However, the authors of the paper admit that they measured SS-NMR spectra of purchased amino acids without any recrystallization or crystal-structure determination, and that some of the amino acids were racemates. The calculated chemical shifts of L-cysteine were far from these experimental values (Figure 3). Therefore, we measured ^{13}C SS-NMR spectrum of enantiomerically pure crystalline L-cysteine and the obtained spectrum is very close to the predicted one.

L-Glutamine spectrum contains two signals of carbonyl carbons (COO and CON) at 173.0 and 176.5 ppm. Our calculations predicted the opposite assignment of these signals than that proposed in ref.[39] Therefore, we performed a C,H-HETCOR experiment, which confirmed our prediction. A cross-peak between the signal of the hydrogen atom in position α has a strong correlation with one of the carbonyl signals (173.0 ppm), which confirms that this signal is the COO carbon adjacent and Cα (see Figure S19 in the ESI).

The new experiments with L-tyrosine and L-threonine did not change the previously published assignment of the signals, but they provided slightly different carbon chemical shifts after careful referencing of the spectra. The newly determined and assigned carbon chemical shifts are used in the experiment-prediction correlations (Table 1).

These examples demonstrate that the proposed method improves the reliability of GIPAW chemical shift predictions, which allows finding previously unnoticed signal or structure mis-assignment.

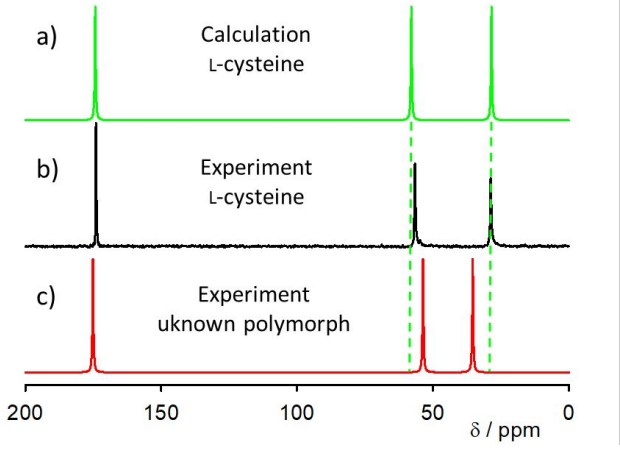

Figure 3. a) Calculated ^{13}C spectrum of solid L-cysteine (corrected GIPAW-PBE0); spectrum simulated using line broadening of 50 Hz. b) Experimental CP-MAS spectrum of crystalline L-cysteine. c) Simulated CP-MAS spectrum of cysteine using experimental chemical shifts from ref.[39] and line broadening of 50 Hz.

The largest error in the GIPAW calculations of 6.8 ppm is found for the anomeric carbon atom C2 of β-D-fructopyranose; with the molecular PBE0 correction, this error drops to 2.2 ppm. Interestingly, all other saccharide anomeric carbons in this set of structures have also very large deviations of GIPAW-calculated carbon chemical shifts (4.4–5.2 ppm), the only exception is glucopyranose carbon C1 in sacharose with the error of 2.7 ppm. Apparently, the PBE functional is not reliable in the chemical shift calculations of anomeric carbon atoms, which are attached to two electronegative oxygen atoms. The errors of the corrected GIPAW chemical shifts are substantially smaller for all anomeric carbons (0.2–1.7 ppm).

Similarly, analyzing the predicted CSA tensor components for the largest errors indicated that the experimental reference for L-glutamine was incorrect. The glutamine carbonyls were swapped and the HETCOR experimental spectrum in the ESI shows the correct assignment. Furthermore, the C6 carbon of adenosine yields the largest error across all methods ranging in deviations of ~15 ppm. Although the experimental chemical shifts from the reference have been validated, the consistent errors indicate that adenosine CSA principal values should be remeasured.

Nitrogen and oxygen isotropic shifts

Molecular PBE0 corrections to GIPAW chemical shifts of nitrogen ^{15}N lead also to significant improvement of the agreement with experimental data (Table 1, MAE decreases from 4.1 to 2.8 ppm). This MAE is identical to that of the SCRMP PBE0 model, albeit with a slightly larger maximum error (7.7 ppm for SCRMP vs 8.3 ppm for the corrected-GIPAW result). The improvement of the chemical-shift prediction of oxygen nuclei is also considerable (MAE decreases from 5.2 to 4.3 ppm). Oxygen chemical shifts are highly sensitive to their electronic environment of the nucleus, making them the most difficult to predict correctly with the fragment-based SCRMP approach. Here, the monomer-corrected GIPAW approach significantly out-performs the 5.9 ppm MAE obtained with SCRMP. Somewhat smaller SCRMP errors would be obtained if a cluster-based approach were used instead of just 1-body and 2-body (monomer and dimer) contributions,[7] but that requires appreciably higher computational cost. These oxygen results truly highlight the advantage of combining the complete treatment of the crystalline lattice with the local higher-level correction.

For co-crystals, salts, and solvates, one might conceivably perform the gas-phase correction on the entire asymmetric unit instead of just the molecule of interest. For the two such species in the carbon test set, L-asparagine monohydrate (ASPARM03) and L-serine monohydrate (LSERMH10), the mean absolute difference in the ^{13}C monomer shielding correction obtained on the asymmetric unit versus the amino acid molecule only is a mere 0.02 ppm, with a max error of 0.08 ppm. Even for the CSA tensors, the mean and maximum differences to the shielding correction are only 0.04 and 0.16 ppm, respectively. In other words, the choice of the "monomer" used for the correction is rather unimportant.

On the other hand, the effect of the monomer definition is much more significant for nitrogen and oxygen chemical shieldings. For the five multi-component crystals in the ^{15}N set (GEHHEH, TEJWAG, FUSVAQ, LTYRHC10, and CYSCLM; four of them are salts, one is a trihydrate), the mean absolute change in the shielding correction between using the full asymmetric unit instead of just the molecule of interest is 1.8 ppm, with a maximum change of 6.9 ppm. For these five crystals, computing the correction using only the single molecule of interest gives a slightly better MAE relative to experiment compared to using the full asymmetric unit (3.1 vs 3.4 ppm).

The impact of the monomer choice on the gas-phase correction is similar for the ^{17}O chemical shifts. Nine of the fifteen crystals contained in the oxygen set are amino acid hydrochloride salts. DFT suffers from delocalization error, which causes problems with charge transfer[40] and can artificially stabilize salt forms of co-crystals.[41] The MAE versus experiment for the oxygen atoms for the nine HCl salts is 3.3 ppm (max 6.3 ppm) when the correction is obtained for just the protonated amino acid, versus 4.0 ppm (max 11.4 ppm) when the full asymmetric unit is employed. Taken together, this evidence indicates that the gas-phase correction should be evaluated using only the molecule of interest.

Once again, the dependence of the results on the basis set used to compute the correction is found to be fairly small (Figures 4 and 5). For nitrogen, the MAE values range 2.7-2.8 ppm across the different basis sets. Larger basis set dependence is observed for the ^{17}O set, where the MAE ranges from 4.0 to 4.5 ppm, and the maximal error from 8.1 to 10.9 ppm. As before, the small-basis 6-31G* results are similar to those from larger basis sets. Interestingly, however, all basis sets except pcSseg-3 predict a large ~10 ppm error for the oxygen in cytosine (CSD refcode CYTSIN). In pcSseg-3, this error drops to less than 1 ppm. So while one generally can use small basis sets to evaluate the monomer correction, the computational cost is low enough that it is probably worthwhile to use relatively large ones in most cases.

Finally, it should be noted that the nitrogen and oxygen test sets are substantially smaller and exhibit less chemical variety than the carbon test set. Further validation of the proposed method on a wider variety of systems would be appropriate. Indeed, the small test set size is probably also what causes the skewed and/or bimodal error distributions observed for most models in the ^{17}O results. Note also that experimental determination of isotropic shifts of ^{17}O, which is a spin 5/2 nucleus with large electric quadrupole moment, is substantially more difficult than the measurement of ^{13}C and ^{15}N shifts.

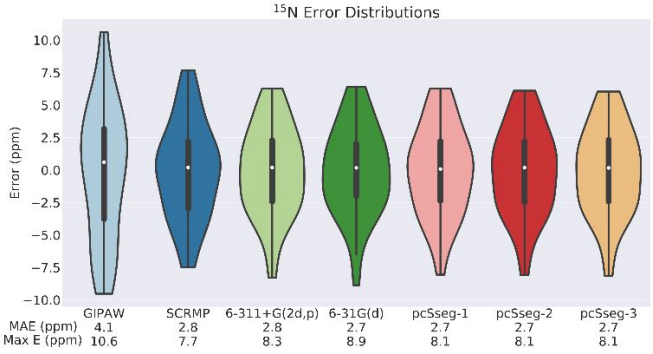

Figure 4. Errors in the ^{15}N chemical shift predictions from GIPAW PBE against those with the gas-phase monomer PBE0 corrections computed in various basis Gaussian sets.

Applications

In this section, the new method is applied to three specific examples beyond the basic benchmarks described above. To test the limits of the proposed method, "difficult" examples were selected purposefully. All three systems have previously been studied by SS-NMR and DFT calculations and the limited accuracy of the GIPAW approach was stressed.

Isocytosine

Isocytosine is a constitutional isomer of cytosine with interesting biological activities. Isocytosine crystallizes as a 1:1 mixture of two tautomers, which form hydrogen bonded pairs similar to pairs of guanine and cytosine in nucleic acids (Figure 6). It has been shown recently that a combination of experimental and simulated chemical shifts of isocytosine may serve as a probe of proton transfer reactions and hence, rare tautomer formation.[42, 43]

The presence of two non-equivalent isocytosine molecules in the crystal structure enables direct comparison of their experimental chemical shift differences against the predicted values. Table 2 summarizes the experimental and calculated ^{13}C and ^{15}N chemical shift differences between the two non-equivalent isocytosine molecules. Once again, applying the PBE0 correction to GIPAW predictions improves the agreement with experiment significantly.

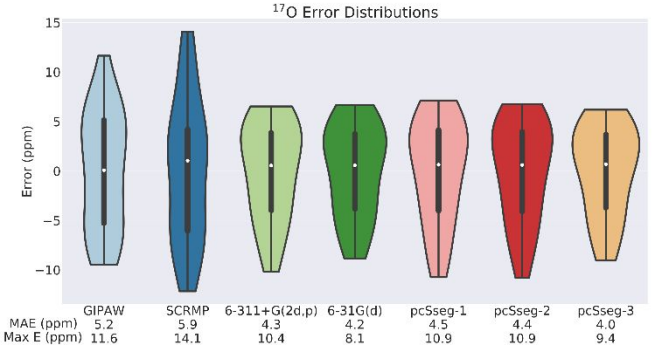

Figure 5. Errors in the ^{17}O chemical shift predictions from GIPAW PBE against those with the gas-phase monomer PBE0 corrections computed in various basis Gaussian sets.

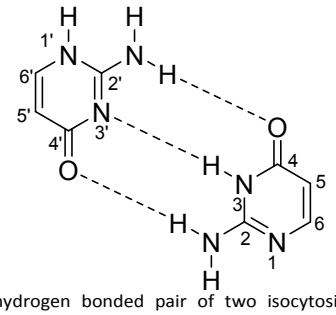

Figure 6. The hydrogen bonded pair of two isocytosine tautomers in solid isocytosine.

Methacrylamide

In the pharmaceutical industry, solid-state NMR is commonly used for the identification of polymorphic crystal structures. SS-NMR can detect polymorphic impurities and characterize polymorphic forms of active pharmaceutical ingredients (APIs) in formulated drug products and drug carriers.[44, 45] The industrially important compound methacrylamide has two known polymorphs; the monoclinic form contains only the *s-cis* molecules (Figure 7), whereas the orthorhombic polymorph is exclusively formed by the *s-trans* conformer.[46] Carbon chemical shift differences between the two forms of methacrylamide are very small (Table 3) and, therefore, may be used as a stringent test of chemical shift predictions.

The methacrylamide molecule is small enough to allow high-level *ab initio* calculations of its NMR shieldings. Table 3 summarizes the predicted chemical shift differences between the two methacrylamide forms calculated at the GIPAW level and at the GIPAW level corrected with PBE0, MP2 or CCSD monomer calculations. Surprisingly, applying the PBE0 correction slightly deteriorates the agreement with experiment, and the *MAE* calculated for MP2-corrected GIPAW result is almost identical to the uncorrected GIPAW one. The CCSD correction improves the *MAE* value only slightly. All four models reproduce the experimental shifts to within a ppm or better.

Table 2. Experimental[10] and calculated chemical shift differences (ppm) in solid isocytosine. Mean absolute errors (*MAE*) obtained for the conventional GIPAW method (PBE functional) and for corrected GIPAW (PBE0 correction, 6-311+G(2d,p) basis set). Atom numbering is depicted in Figure 6.

	Experiment	GIPAW	GIPAW-corrected
C2–C2'	0.0	-2.77	-2.29
C4–C4'	6.1	6.11	6.11
C5–C5'	4.0	3.24	4.50
C6–C6'	-19.3	-18.54	-19.44
MAE		1.1	0.7
N1–N1'	-73.4	-74.91	-74.34
N2–N2'	49.9	54.42	52.78
N3–N3'	-3.4	-5.17	-4.56
MAE		2.6	1.7

These particularly subtle differences in the chemical shifts between the two polymorphs probably represent the limit of what can be achieved by corrections computed for a single, isolated molecule. Chemical shift differences between

polymorphs are mostly governed by molecular packing and intermolecular interactions in the crystals; these intermolecular interactions are modelled with the GGA level of theory only and are not included in the molecular correction proposed here.

Testosterone

Two crystal forms of testosterone have been studied by SS-NMR and most of the carbon signals have been assigned using INADEQUATE carbon–carbon experiment.[47] The α form contains two crystallographically non-equivalent molecules in the asymmetric unit, while the β form is a monohydrate. The conformation is almost identical in all three crystallographically unique molecules.[6]

Carbon chemical shifts of solid testosterone have also been calculated using the GIPAW and cluster/fragment approach.[6, 47] Most individual chemical shifts were reproduced to within a few ppm, with the notable exception of C5, which was significantly overestimated (Table 4) by both methods.

We calculated carbon C5 chemical shift at the GIPAW level and, indeed, the agreement with experiment is surprisingly poor. The molecular PBE0 corrections improve the agreement by ca. 3 ppm, but the shifts are still overestimated by ~10 ppm. Therefore, we calculated a CCSD correction for a partial fragment of the testosterone molecule (because CCSD chemical shielding calculations on the full testosterone would be very expensive). Starting from the GIPAW-optimized structure of the β-form of testosterone, this partial fragment consists of the C4-C5 double bond and three carbon atoms directly attached to the double bond (C3, C6 and C10, see Figure 8); missing hydrogen atoms were added to saturate the dangling bonds on the terminal carbon atoms. NMR shieldings of this fragment were then calculated at the PBE and CCSD levels of theory. To allow comparison of the calculated shieldings with the chemical shifts of testosterone, we calculated NMR shieldings of α-glycine, a commonly used reference compound, at the same levels of theory. The chemical shift of C5 in the molecular fragment calculated at the PBE level is by 22.3 ppm lower than the shift of glycine carbonyl. On the other hand, the CCSD calculation predicts the C5 chemical shift lower by 34.2 than that of glycine, i.e. CCSD level of theory predicts that the shift of C5 in the molecular fragment is by 11.9 ppm lower than the value predicted by PBE. If we transfer this correction to the whole β-testosterone molecule, the GIPAW-predicted chemical shift (186.9) drops to 175 ppm, which is reasonably close to the experimental value (173.8 ppm).

It is not clear, why DFT with both the PBE and PBE0 functionals fail to predict this particular carbon atom chemical shift correctly. However, this example demonstrates that high-level *ab initio* corrections may be calculated for molecular fragments and these corrections may be used to improve the agreement of predicted shifts with experiment.

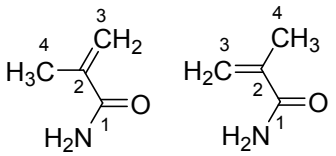

Figure 7. *S-cis* (left) and *s-trans* (right) conformers of methacrylamide found in monoclinic and orthorhombic polymorphs, respectively.

Table 3. Experimental[37] and calculated chemical shift differences (ppm) between the monoclinic and orthorhombic polymorphs of methacrylamide. Mean absolute errors (*MAE*) obtained for the conventional GIPAW method (PBE functional) and for corrected GIPAW (PBE0, MP2 and CCSD correction, 6-311+G(2d,p) basis set). Atom numbering is depicted in Figure 7.

Atom	Experiment	GIPAW	GIPAW PBE0-corrected	GIPAW MP2-corrected	GIPAW CCSD-corrected
C1	0.15	1.50	1.23	0.89	0.87
C2	0.27	0.68	1.46	0.14	0.85
C3	-0.73	-1.37	-1.95	0.97	0.28
C4	0.06	0.38	0.49	0.17	0.10
MAE		0.68	0.98	0.67	0.59

Conclusions

In conclusion, this study has demonstrated a very simple strategy for improving the quality of GGA-based GIPAW NMR chemical shielding calculations in molecular crystals by evaluating a correction to the shielding computed at a higher level of theory on an isolated molecule. The new approach achieves accuracy rivaling or beating that of fragment-based methods. The correction is quite insensitive to the basis set used for the monomer calculation, ensuring that the cost of evaluating the correction is minimal. Typically one would employ a hybrid density functional for the higher level of theory. However, as some of the applications demonstrate, it is also possible to consider the use of higher-level chemical shielding calculations, such as CCSD. The CCSD correction proved essential to predicting the chemical shift of carbon C5 in β-testosterone, for example.

Table 4. Experimental and predicted chemical shifts of carbon C5 in solid testosterone.

	α form		β form
	molecule *u*	molecule *v*	
Experiment[47]	170.6	172.1	173.8
GIPAW	182.6	184.1	186.9
GIPAW PBE0-corrected	179.9	181.4	183.9
Cluster/fragment[6]	176.2	177.0	182.1
GIPAW CCSD-corrected			175.0

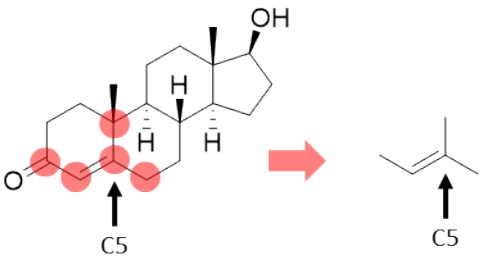

Figure 8. The structure of testosterone and its fragment used for the calculation of CCSD corrections.

Finally, while the work here focused on molecular organic systems, it would be interesting to explore the application of the technique to inorganic systems with localized electronic structure as well. The testosterone example demonstrates how even a calculation on a small local "cluster" of atoms cut out from a larger covalent network may be enough to achieve a meaningful correction to the GGA-level chemical shifts.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The work has been supported by the Czech Science Foundation (Grant No. 18-11851S). G.B. and P.U. gratefully acknowledge funding from the U.S. National Science Foundation (CHE-1665212) and supercomputer time from XSEDE (TG-CHE110064).

References

1. C. J. Pickard and F. Mauri, *Phys. Rev. B*, 2001, **6324**, 245101.
2. C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azais, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri and C. J. Pickard, *Chem. Rev.*, 2012, **112**, 5733-5779.
3. S. E. Ashbrook and D. McKay, *Chem. Commun.*, 2016, **52**, 7186-7204.
4. T. Charpentier, *Solid State Nucl. Magn. Reson.*, 2011, **40**, 1-20.
5. J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller and G. J. O. Beran, *Phys. Chem. Chem. Phys.*, 2016, **18**, 21686-21709.
6. J. D. Hartman, G. M. Day and G. J. O. Beran, *Cryst. Growth Des.*, 2016, **16**, 6479-6493.
7. J. D. Hartman, A. Balaji and G. J. O. Beran, *J. Chem. Theory Comput.*, 2017, **13**, 6043-6051.
8. J. D. Hartman and G. J. O. Beran, *Solid State Nucl. Mag.*, 2018, **96**, 10-18.
9. O. Socha, P. Hodgkinson, C. M. Widdifield, J. R. Yates and M. Dračínský, *J. Phys. Chem. A*, 2017, **121**, 4103-4113.
10. M. Dračínský, P. Jansa, K. Ahonen and M. Buděšínský, *Eur. J. Org. Chem.*, 2011, 1544-1551.
11. K. Maliňáková, L. Novosadová, M. Pipíška and R. Marek, *ChemPhysChem*, 2011, **12**, 379-388.
12. M. Mirzaei and N. L. Hadipour, *J. Phys. Chem. A*, 2006, **110**, 4833-4838.
13. M. Babinský, K. Bouzková, M. Pipíška, L. Novosadová and R. Marek, *J. Phys. Chem. A*, 2013, **117**, 497-503.
14. K. Bouzková, M. Babinský, L. Novosadová and R. Marek, *J. Chem. Theory Comput.*, 2013, **9**, 2629-2638.
15. S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Phys.*, 2014, **141**, 164121.
16. S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Theory Comput.*, 2015, **11**, 5229-5241.
17. R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates and V. Zorin, *Magn. Reson. Chem.*, 2007, **45**, S174-S186.
18. G. J. O. Beran, J. D. Hartman and Y. N. Heit, *Acc. Chem. Res.*, 2016, **49**, 2501-2508.
19. H. Stoll, *Chem. Phys. Lett.*, 1992, **191**, 548-552.
20. B. Paulus, *Phys. Rep.*, 2006, **428**, 1-52.
21. A. D. Boese and J. Sauer, *Cryst. Growth Des.*, 2017, **17**, 1636-1646.
22. G. A. Dolgonos, O. A. Loboda and A. D. Boese, *J. Phys. Chem. A*, 2018, **122**, 708-713.
23. S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson and M. C. Payne, *Z. Kristallogr.*, 2005, **220**, 567-570.
24. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865-3868.
25. H. J. Monkhorst and J. D. Pack, *Phys. Rev. B*, 1976, **13**, 5188-5192.
26. J. R. Yates, C. J. Pickard and F. Mauri, *Phys. Rev. B*, 2007, **76**, 024401.
27. E. R. McNellis, J. Meyer and K. Reuter, *Phys. Rev. B*, 2009, **80**, 205414.
28. A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.
29. I. De Gortari, G. Portella, X. Salvatella, V. S. Bajaj, P. C. A. van der Wel, J. R. Yates, M. D. Segall, C. J. Pickard, M. C. Payne and M. Vendruscolo, *J. Am. Chem. Soc.*, 2010, **132**, 5993-6000.
30. M. Dračínský, P. Bouř and P. Hodgkinson, *J. Chem. Theory Comput.*, 2016, **12**, 968-973.
31. J. L. McKinley and G. J. O. Beran, "Improving Predicted Nuclear Magnetic Resonance Chemical Shifts Using the Quasi-Harmonic Approximation." submitted (2019).
32. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, X. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. A. , J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Journal*, 2016.
33. K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. S. Sun, V. Gurumoorthi, J. Chase, J. Li and T. L. Windus, *J. Chem. Inf. Model.*, 2007, **47**, 1045-1052.
34. A. A. Auer and J. Gauss, *J. Chem. Phys.*, 2001, **115**, 1619-1622.
35. CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package by J.F. Stanton, J. Gauss, L. Cheng, M.E. Harding, D.A. Matthews, P.G. Szalay with contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.
36. M. Dračínský and P. Hodgkinson, *Chem. Eur. J.*, 2014, **20**, 2201-2207.
37. M. Dračínský, E. Procházková, J. Kessler, J. Šebestík, P. Matějka and P. Bouř, *J. Phys. Chem. B*, 2013, **117**, 7297-7307.
38. D. Stueber and D. M. Grant, *J. Am. Chem. Soc.*, 2002, **124**, 10539-10551.
39. C. H. Ye, R. Q. Fu, J. Z. Hu, L. Hou and S. W. Ding, *Magn. Reson. Chem.*, 1993, **31**, 699-704.
40. A. J. Cohen, P. Mori-Sanchez and W. T. Yang, *Chem. Rev.*, 2012, **112**, 289-320.

41. L. M. LeBlanc, S. G. Dale, C. R. Taylor, A. D. Becke, G. M. Day and E. R. Johnson, *Angew. Chem. Int. Ed.*, 2018, **57**, 14906-14910.
42. R. Pohl, O. Socha, M. Šála, D. Rejman and M. Dračínský, *Eur. J. Org. Chem.*, 2018, 5128-5135.
43. R. Pohl, O. Socha, P. Slavíček, M. Šála, P. Hodgkinson and M. Dračínský, *Faraday Discuss*, 2018, **212**, 331-344.
44. E. Skorupska, S. Kazmierski and M. J. Potrzebowski, *Mol. Pharmaceut.*, 2017, **14**, 1800-1810.
45. X. Yang, T. C. Ong, V. K. Michaelis, S. Heng, J. Huang, R. G. Griffin and A. S. Myerson, *CrystEngComm*, 2014, **16**, 9345-9352.
46. C. Y. Guo, M. B. Hickey, E. R. Guggenheim, V. Enkelmann and B. M. Foxman, *Chem. Commun.*, 2005, 2220-2222.
47. R. K. Harris, S. A. Joyce, C. J. Pickard, S. Cadars and L. Emsley, *Phys. Chem. Chem. Phys.*, 2006, **8**, 137-143.