Improved electrostatic embedding for fragment-based chemical shift calculations in molecular crystals

Joshua D. Hartman, Ashwin Balaji, and Gregory J. O. Beran*

Department of Chemistry, University of California, Riverside, California 92521 USA

E-mail: gregory.beran@ucr.edu

October 3, 2017

Abstract

Fragment-based methods predict nuclear magnetic resonance (NMR) chemical shielding tensors in molecular crystals with high accuracy and computational efficiency. Such methods typically employ electrostatic embedding to mimic the crystalline environment, and the quality of the results can be sensitive to the embedding treatment. To improve the quality of this embedding environment for fragment-based molecular crystal property calculations, we borrow ideas from the embedded ion method to incorporate self-consistently polarized Madelung field effects. The self-consistent reproduction of the Madelung potential (SCRMP) model developed here constructs an array of point charges that incorporates self-consistent lattice polarization and which reproduces the Madelung potential at all atomic sites involved in the quantum mechanical region of the system. The performance of fragment- and cluster-based ^{1}H, ^{13}C, ^{14}N and ^{17}O chemical shift predictions using SCRMP and density functionals like PBE and PBE0 are assessed. The improved embedding model results in substantial improvements in

the predicted ^{17}O chemical shifts and modest improvements in the ^{15}N ones. Finally, the performance of the model is demonstrated by examining the assignment of the two oxygen chemical shifts in the challenging γ-polymorph of glycine. Overall, the SCRMP-embedded NMR chemical shift predictions are on par with or more accurate than those obtained with the widely used gauge-including projector augmented wave (GIPAW) model.

1 Introduction

Solid-state nuclear magnetic resonance (NMR) spectroscopy is a powerful and versatile tool for obtaining detailed chemical structure information in solids, with many applications in biology, pharmaceuticals, and materials. Mapping from an NMR spectrum to a three-dimensional crystal structure can be challenging, but *ab initio* quantum chemistry techniques can facilitate this process through chemical shift prediction. Indeed, the periodic density functional theory (DFT)-based gauge-including projector augmented wave (GIPAW) method[1,2] has come to play a central role in NMR spectral assignment and structure elucidation of crystalline systems.[3]

Recently, however, modern fragment methods[4] have emerged as a viable alternative to GIPAW for chemical shift prediction in molecular crystals.[5–9] By decomposing a molecular crystal into smaller fragments (often just monomers and dimers), these methods can reduce both the total computational cost of chemical shift prediction and facilitate the highly parallel software implementations which enable chemical shifts of even complicated crystals to be predicted within a few hours of wall time. Even more importantly, fragment methods provide a route to higher-accuracy chemical shift predictions. Whereas GIPAW calculations are limited to generalized gradient approximation (GGA) density functionals in practice, fragment methods can employ hybrid density functionals or perhaps even post-Hartree-Fock correlated wavefunction methods for chemical shift prediction.

Early fragment work predicting chemical shifts in molecular crystals used small, finite

clusters in conjunction with point charge embedding schemes designed to approximate the crystalline environment. For example, electrostatic embedding models such as the surface charge representation of the electrostatic embedding potential (SCREEP)[10] model and the embedded-ion method (EIM)[11,12] surround the molecular/cluster calculations with an array of point charges that have been optimized to reproduce the Madelung potential at each atom site in the crystalline system. Such models capture long-range electrostatic and polarization effects, but using a small cluster can miss important short- or medium-range quantum mechanical interactions.

More recently, advances in computer processing power have made large cluster calculations that include all nearest-neighbors (often 10-15 molecules) tractable. For example, the symmetry-adapted cluster (SAC) method uses symmetry considerations to perform NMR chemical shielding calculations on a large finite cluster, effectively capturing local many-body polarization effects.[13,14] Augmenting a cluster approach with simple electrostatic embedding models further improves accuracy by including long-range electrostatics.[7] Fragment methods which replace the large cluster calculation with many computationally inexpensive monomer and dimer fragment calculations have also proven very effective in predicting chemical shifts at much lower computational cost.[6–9] The dimer fragments capture local quantum mechanical interactions, while the electrostatic embedding helps incorporate long-range and many-body polarization effects arising from the crystalline environment.

Recent benchmark studies involving ^{1}H, ^{13}C, ^{15}N and ^{17}O isotropic shielding predictions have established that when the same density functional is used (e.g. PBE), embedded fragment and cluster models often perform comparably well to the highly successful periodic DFT GIPAW method.[7,8] However, switching to hybrid density functionals like PBE0 or B3LYP instead of GGAs like PBE or BLYP, which is routinely feasible with fragment methods, often enables one to predict chemical shifts with $\sim$20–30% higher accuracy than GIPAW PBE.[7,8] The inclusion of exact exchange in hybrid density functionals improves the description of the electron density,[15,16] which is key to understanding the chemical shielding. This

higher accuracy translates to increased NMR discrimination between correct and incorrect structural assignments in polymorphic crystals or other systems.[9]

Despite numerous successes of the fragment models for chemical shift prediction, limitations have emerged in the existing implementations. The accuracy of ^{17}O chemical shifts, which are particularly sensitive to the electrostatic environment, is somewhat worse with fragment methods compared to other nuclei. In the aforementioned ^{17}O benchmarks, GIPAW PBE outperformed even combined cluster/fragment PBE0 calculations due to its more complete treatment of long-range and many body effects.[8]

Previously, our cluster and fragment electrostatic embedding schemes relied upon atom-centered point charges obtained from Gaussian distributed multipole analysis (GDMA). The GDMA was carried out on isolated monomers and the resulting atom centered charges were placed on all molecules within a 30 Å radius of the asymmetric unit.[6–8] This simple embedding model suffers from three main limitations. First, truncated GDMA schemes do not always converge uniformly toward the complete multipolar expansion.[17] Embedding with GDMA charges and dipoles instead of charges only does not necessarily improve the predicted chemical shieldings.[6] Second, atom-centered charges derived from calculations involving isolated monomers neglect polarization effects in the crystalline environment. Third, a simple 30 Å cutoff radius does not always account for the long-range electrostatics in a formally infinite lattice. This latter point is especially clear in crystals comprised of highly charged monomers or polar unit cells like the γ-polymorph of glycine, which exhibits large errors and poor convergence with regard to both the DFT self-consistent-field iterations and the predicted chemical shifts as a function of the fragment and embedding cutoff parameters.

The present work adapts ideas from the SCREEP and EIM models to improve the electrostatic embedding environment employed in fragment-based chemical shielding calculations. First, to avoid errors associated with truncated GDMA multipolar expansions, the GDMA point charges for the individual molecules are replaced with charges from electrostatic potentials using a grid-based method (CHELPG).[18] Second, Ewald summation[19] over the lattice

of CHELPG charges is employed to capture long-range electrostatic effects on the potential. Third, the key quantum mechanical (QM) fragment/cluster region is embedded with a set of point charges. Similar to our previous work, charges close to the central region are set to the CHELPG values. However, in the new model, more distant charges are fitted such that the Madelung potential is accurately reproduced throughout the QM region. Finally, a self-consistent embedding procedure[20] is introduced wherein the atom-centered CHELPG charges are re-computed in this crystalline embedding environment. The entire procedure repeats until the embedding charges converge. This self-consistent reproduction of the Madelung potential (SCRMP) model augments the high-quality QM treatment of local effects afforded by fragment-based models with an improved treatment of lattice electrostatics and polarization akin to that found in the SCREEP and EIM methods.

We first discuss the SCRMP embedding model before examining its performance in fragment and cluster-based isotropic chemical shift calculations. In particular, we examine the impact of the improved embedding model on the prediction of ^{1}H, ^{13}C, ^{15}N and ^{17}O experimental chemical shifts in molecular crystals. Although the effects of the improved SCRMP embedding model are minimal for ^{1}H and ^{13}C shifts and modest for ^{15}N chemical shifts, it significantly increases the statistical accuracy of the ^{17}O chemical shifts versus experiment. Finally, the benchmark data also provides a set of linear regression parameters for mapping predicted chemical shieldings to the experimentally determined chemical shifts. We apply SCRMP embedded fragment and cluster calculations along with the linear regression parameters to predict the ^{17}O chemical shifts for the challenging γ-polymorph of glycine. The combined cluster/fragment SCRMP methods improve upon both the pure fragment and GIPAW results.

2 Theory

2.1 *Ab initio* fragment-based shielding calculations

Fragment-based chemical shift prediction methods rely upon a many-body expansion for the chemical shielding tensor, which can be derived via differentiation of the many-body expansion for the energy with respect to the nuclear magnetic moment and the external magnetic field,[8]

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

where $\tilde{\boldsymbol{\sigma}}_i^A$ is the chemical shielding tensor for atom A on molecule i in the full system environment. Equation 1 decomposes $\tilde{\boldsymbol{\sigma}}_i^A$ into contributions from the shielding tensor for the isolated molecule ($\boldsymbol{\sigma}_i^A$), pair-wise contributions ($\Delta^2\boldsymbol{\sigma}_{ij}^A$) from each dimer, three-body contributions $\Delta^3\boldsymbol{\sigma}_{ijk}^A$ from each trimer, and contributions from higher-order terms.

The two-body fragment approach achieves low computational cost by truncating Eq. 1 to neglect long-range two-body terms and all three-body and higher terms.[6–8] An electrostatic embedding environment is introduced to account for polarization/electrostatic effects that are neglected upon truncation of the expansion:

$$\tilde{\boldsymbol{\sigma}}_i^A \approx \boldsymbol{\sigma}_i^{A,emb.} + \sum_j \Delta^2\boldsymbol{\sigma}_{ij}^{A,emb.} \tag{2}$$

where $\boldsymbol{\sigma}_i^{A,emb.}$ and $\Delta^2\boldsymbol{\sigma}_{ij}^{A,emb.}$ are the one and two-body contributions with each monomer and dimer calculation carried out in an electrostatic embedding environment (discussed in Section 2.2). In the context of a molecular crystal, a two-body fragment-based calculation defines a cutoff radius R_{QM} around the asymmetric unit, as illustrated in Figure 1. The chemical shielding tensor for each atom in the asymmetric unit is then approximated by calculating the monomer and two-body contributions for all dimers involving that monomer and any other molecule whose atoms lie within R_{QM} (typically 6 Å in practice).

Errors in the fragment approach may arise from the limitations of the embedding model

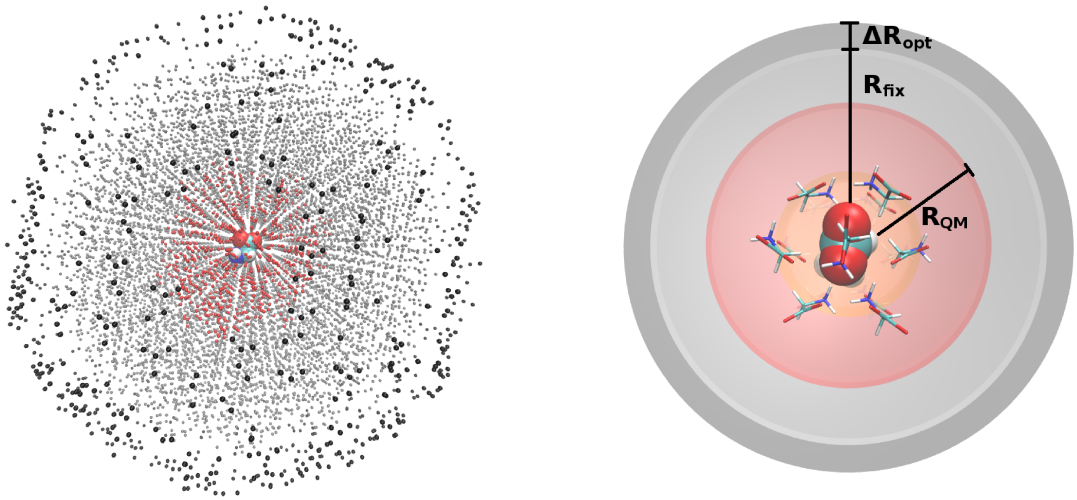


Figure 1: Schematic illustrating the SCRMP method in the context of fragment and cluster-based calculations. The asymmetric unit is depicted using a space filling model. All molecules used in fragment and/or cluster calculations comprise the QM region (R_{QM}). Combined cluster/fragment calculations use a central cluster, depicted using a licorice model (right), and two-body contributions involving all other molecules outside the cluster but inside R_{QM}. Atom-centered CHELPG are used on all red and gray atom centers within the fixed point charge region, R_{fix}. The black atom centered charges lying within R_{fix} and $R_{fix+opt}$ are optimized such that V_{MP} is reproduced at all red atom centers within R_{QM}.

for treating longer-range pairwise interactions and many-body effects. To capture the local many-body effects explicitly in the quantum mechanical treatment, a hybrid cluster/fragment model can be used. The cluster/fragment approach typically combines a ∼15-molecule cluster composed of all molecules with atoms lying within 4 Å of the central monomer, with longer-range interactions out to 6 Å evaluated in a pairwise fashion. Those particular cutoffs were adopted based on previous studies of the convergence behaviors of the chemical shifts.[7,8,21] Because many-body contributions beyond 4 Å tend to be small, combining the modestly-sized cluster with longer-range pairwise contributions balances between including important many-body effects while avoiding the excessive computational cost associated with very large clusters.

2.2 The SCRMP model

The accuracy of Eq. 2 depends on the effective reproduction of the crystalline electrostatic environment. Our initial efforts[6–8] employed a straightforward electrostatic embedding model based upon fixed GDMA point charges obtained for isolated monomers in the asymmetric unit.[22–24] The monomer-derived GDMA charges were positioned at each atomic center for every monomer within a user-defined charge embedding radius of molecule i in the central unit cell (R_{fix} in Figure 1).

The SCRMP method adapts ideas from the embedded ion method to fragment-based electronic structure calculations. It seeks to construct a point charge array which more accurately mimics the many-body electrostatic environment in the infinite crystal. To improve upon the sometimes poor representation of the electrostatic potential offered by GDMA monopoles,[17] it replaces GDMA charges with CHELPG ones. The CHELPG charges are directly placed on crystal lattice atomic centers to construct the short- and medium-range electrostatic embedding environment out to a distance R_{fix}. At long-range (beyond R_{fix}), SCRMP surrounds these CHELPG charges with another shell of atom-centered charges (shell thickness ΔR_{opt}) which are least-squares fitted such that the entire point-charge embedding environment mimics the Madelung potential from the infinite lattice throughout the QM region. This two-layer charge embedding scheme ensures that the QM fragments are surrounded by physically appropriate charges for locally important interactions, while also experiencing an overall electrostatic environment that closely approximates the infinite lattice. Finally, the entire set of charges is computed self-consistently to incorporate the many-body polarization effects that occur in the crystal lattice. More detailed description of the procedure follows:

1. Compute CHELPG atom-centered charges for each isolated monomer in the unit cell. In principle, one need only compute charges on monomers in the asymmetric unit and then replicate those charges via the space group symmetry operators. However, the current software implementation of the SCRMP construction only exploits translational

symmetry between periodic images of the cell, and all monomer charges within the cell are computed explicitly.

2. Use the atom-centered CHELPG charges to compute the Madelung potential ($V_{MP}(\mathbf{r}_i)$) at N probe points within the probe region R_{QM} according to the Ewald formalism[19] under tin-foil boundary conditions:

$$V_{MP}(\mathbf{r}_i) = \sum_{\mathbf{n}} \sum_{\mathbf{j}}^{N} q_j \frac{erfc(\sqrt{\kappa}|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|} + \sum_{\mathbf{k}\neq 0} \sum_{j}^{N} q_j \Big(\frac{4\pi}{Ck^2}\Big) e^{\mathbf{ik}\cdot(\mathbf{r}_i - \mathbf{r}_j)} e^{-k^2/4\kappa} \tag{3}$$

where q_j is the partial charge of the jth atom located at $\mathbf{r}_j$, C is the volume of the unit cell, $\mathbf{n}$ is the real space vector and $\mathbf{k}$ represents the reciprocal space vector. Probe points are placed at each atom center. Further increasing the number of probe points via icosahedral grids surrounding the nuclei did not appreciably impact the subsequent reproduction of the Madelung potential or the predicted chemical shifts.

3. Construct a two-layered array of point charges to reproduce the Madelung potential at every probe point within R_{QM}. In the inner layer, CHELPG point charges are placed on all atoms within R_{fix} of the asymmetric unit (typically 30 Å and indicated by the red and gray points in Figure 1). For the outer layer, an additional shell consisting of N point charges is placed on a shell ΔR_{opt} of the N closest atoms beyond R_{fix} (black charges in Figure 1). These outer N charges are optimized such that the sum of the potentials from the fixed charges (V_{fix}) and the optimized charges (V_{opt}) approximates the Madelung potential (V_{MP}) at every probe point $\mathbf{r}_i$:

$$V_{MP}(\mathbf{r}_i) \approx V_{fix}(\mathbf{r}_i) + V_{opt}(\mathbf{r}_i) \tag{4}$$

The charge fitting in the outer shell is performed via least-squares minimization of the objective function,

$$\left|\left| V_{MP}(\mathbf{r}_i) - \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} q_j \right|\right|^2 = ||b - Ax||^2 \tag{5}$$

where $\mathbf{r}_j$ is the position of charge q_j in the outer shell. Numerical aspects of this fitting are discussed below.

4. New CHELPG calculations are carried out for each molecule in the unit cell in the presence of the optimized charge array obtained in Step 3. Evaluation of these charges within the embedding potential incorporates polarization due to the crystal lattice. The resulting embedded monomer atomic charges are compared to those used in Step 2.

Steps 2–4 are repeated until self-consistency is achieved and the new charges differ from the previous ones by no more than root-mean-square (RMS) deviations of 10^{-3} atomic charge units. This procedure typically converges in about 5 iterations.

The SCRMP approach is very similar to the embedded ion method.[11,12] The two key differences lie in the size of the QM region and the method of least squares fitting of the outer charges. With the 2-body fragment or cluster/fragment methods, one can readily treat comparatively larger QM regions with low computational cost. Adapting to fragment calculations requires consideration of how to treat the atoms in the QM region that are not present in a particular fragment contribution. In addition to surrounding the QM region with the SCRMP embedding charges, the self-consistent CHELPG point charges are also placed on all atoms within the QM region that are not explicitly included in a given QM fragment.

For the least-squares charge fitting, the matrix A in Eq 5 is typically highly rank-deficient, with condition numbers on the order of 10^{15} for systems we have investigated. Therefore, the least-squares problem does not have a unique solution in the absence of additional constraints. The ill-conditioned least-squares fitting frequently leads to excessively large, unphysical charges. The embedded ion method constrains the solution space by minimizing the norm of the solution vector.[11,12] Here, we solve the least-squares problem via singular value decomposition and remove the linear dependencies via a low-rank approximation for A. Specifically, we effectively limit the condition number to 10^6 by neglecting all singular

values λ_i that are less then $10^{-6}\lambda_{max}$. Using this approach, the SCRMP model charge array reproduces the Madelung potential throughout the QM region to within a root-mean-square error of $\sim$0.001 kcal mol^{-1} au^{-1}, which is well-below the threshold for influencing chemical shift calculations. The resulting fitted charges have magnitudes of no more than a few atomic charge units.

3 Computational Methods

3.1 Chemical shielding calculations

The performance of the models is assessed on four test sets containing a total of 63 molecular crystals, with 169 ^{13}C, 80 ^{1}H, 51 ^{15}N, and 28 experimental ^{17}O isotropic chemical shifts. Optimized crystal structures for the ^{1}H, ^{13}C, ^{15}N and ^{17}O test sets were taken from Ref 8. The experimental crystal structure for the γ polymorph of glycine was obtained from the Cambridge Structure Database (CSD), reference code GLYCIN34.[25] The lattice parameters were fixed at their experimental values and all atoms in the unit cell were relaxed using the freely available, open-source Quantum Espresso software package.[26] Geometry optimization was carried out using the PBE[27] density functional and the D2 dispersion correction,[28] ultrasoft pseudopotentials with a plane wave cutoff of 80 Ry, and a 5×5×5 Monkhorst-Pack k-point grid. We used the pseudopotentials H.pbe-rrkjus.UPF, C.pbe-rrkjus.UPF, N.pbe-rrkjus.UPF and O.pbe-rrkjus.UPF from http://www.quantum-espresso.org. This protocol is identical to the one previously applied to all the crystals in the test sets.[8]

Molecular crystal fragmentation through two-body was performed using our hybrid many-body interaction (HMBI) code.[29–31] Shielding calculations for individual fragments were carried out using Gaussian 09[32] with the B3LYP, BLYP, PBE0 and PBE density functionals. PBE is very widely used in GIPAW. The other functionals were selected based on our previous benchmark studies that found that hybrid functionals PBE0 and B3LYP perform similarly to one another and appreciably better than their related GGA functionals PBE and BLYP.[7,8]

Calculations were performed using a locally dense basis set[33,34] and the gauge-including atomic orbital (GIAO) approximation.[35] All calculations used a 6-311+G(2d,p) basis[36–39] for atoms in the asymmetric unit, a 6-311G(d,p) basis for neighboring atoms out to 4 Å, and a 6-31G[40,41] basis for all atoms beyond 4 Å. This locally dense basis set combination, referred to here as our "mixed basis," has proved effective in previous studies. The use of smaller basis sets on distant atoms has minimal impact on the predicted chemical shieldings for the central atoms of interest,[21] while the errors introduced by basis set incompleteness in the shieldings on the atoms of interest is compensated for by the chemical shift referencing strategy described below.[7] A large DFT integration grid consisting of 150 radial and 974 Lebedev angular points was used to approach rotational invariance and mitigate numerical noise in the fragment calculations, as described previously.[6]

The embedding environment was constructed using atom-centered monopoles computed using CHELPG as implemented in Gaussian 09. Atom-centered charges were calculated using the same density functional and 6-311+G(2d,p) basis set as the chemical shielding calculation. Point charges were placed on all molecules lying within $R_{fix} = 30$ Å of any atom in the asymmetric unit. In the SCRMP implementation, an additional set of N point charges were placed at atom centers directly surrounding the 30 Å shell and optimized according to the procedure outlined in Section 2.2. Results which employ GDMA point charge embedding were taken from earlier studies.[7,8]

Some results which include self-consistent embedding (SCE) but omit the Madelung potential portions of the SCRMP model are reported for comparison with the SCRMP results. In that case, CHELPG charges are computed for each monomer in the asymmetric unit, and then replicated on all other monomers out to 30 Å (akin to the GDMA embedding used previously[8]). The CHELPG charges are then recomputed in that embedding environment, and the process is repeated until the CHELPG charges converge.

3.2 Chemical shift referencing

Chemical shifts reported here are referenced relative to neat TMS under magic angle spinning (MAS) conditions for ^{1}H and ^{13}C, solid NH_4Cl under MAS for ^{15}N, and liquid water for ^{17}O. Mapping between computed absolute shieldings σ_i and the experimentally observed chemical shift δ_i was carried out using the linear regression scheme[42] shown in Eq. 6.

$$\delta_i = A\sigma_i + B \tag{6}$$

Obtaining regression parameters A and B using carefully constructed test sets[8] helps mitigate systematic errors inherent in the chemical shielding calculations. For example, for the oxygen test set and the basis sets used here, basis set superposition error (BSSE) typically causes isotropic ^{17}O shieldings on atoms in the asymmetric unit to be deshielded by a little more than half a ppm on average for the crystals considered here. However, because the BSSE is systematic, the linear regression compensates for a significant fraction of this effect. In the end, compensating for BSSE would alter the shifts by root-mean-square 0.3 ppm, and the overall root-mean-square error versus experiment by a negligible 0.01 ppm.

Linear regression models of the form presented in Eq. 6 were applied separately to each model/cutoff combination using the experimental isotropic shifts for the respective test sets. This approach provides transferable regression parameters which can be used to scale predicted shifts obtained for compounds not included in the test set. In Section 4.3, the ^{17}O chemical shift regression parameters are applied to γ-glycine, which was not included in the ^{17}O test set used to fit Eq 6.

4 Results and Discussion

Because ^{17}O chemical shifts are sensitive to electrostatic embedding, we first assess the performance of the different electrostatic embedding models on the isotropic shifts in the

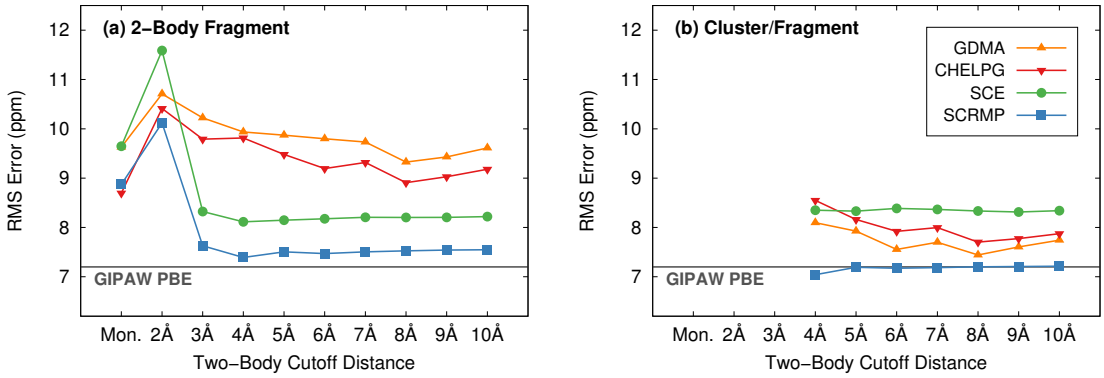


Figure 2: Performance of the (a) 2-body fragment and (b) cluster/fragment models for computing ^{17}O isotropic chemical shifts as a function of the 2-body cutoff distance with fixed GDMA charges, fixed CHELPG charges, self-consistent embedding (SCE) CHELPG charges, and the SCRMP model. Shifts calculated at the PBE0/mixed basis level, with a 4 Å cluster for cluster/fragment. The gray line indicates the performance of GIPAW PBE for the same test set.[8]

^{17}O test set. We then compare the SCRMP model results against experiment and the widely used GIPAW method for ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shifts. Linear regression parameters across all four nuclei and several commonly used density functionals are provided. Finally, we apply fragment/cluster methods and the linear regression parameters to predict ^{17}O isotropic shifts in the challenging case of γ-glycine.

4.1 Improved accuracy ^{17}O isotropic shielding predictions

Figure 2 illustrates the RMS errors in the ^{17}O isotropic chemical shifts relative to experiment with different embedding models as a function of two-body cutoff distance for both fragment and cluster/fragment models. For each model/cutoff combination, linear regression models of the form presented in Eq 6 were fitted separately against the experimental oxygen test set. For ease of comparison, previously reported values for GDMA fragment, GDMA cluster/fragment (orange) and GIPAW calculations (gray line) are included.[8]

Similar to previous findings,[8] a substantial reduction in the ^{17}O isotropic shift RMS error is observed in Figure 2a once the two-body cutoff is large enough to capture all nearest-neighbor molecules (∼4 Å). Nevertheless, both the fragment (Figure 2a) and cluster/fragment (Figure 2b) results with simple fixed GDMA or CHELPG point-charge em-

bedding exhibit distinct, somewhat oscillatory variation in the RMS error as the 2-body cutoff is increased. In contrast, the chemical shift calculations employing self-consistent embedding (SCE) without the Madelung potential fits or the full SCRMP embedding show much less sensitivity to the 2-body cutoff, with negligible variations once the cutoff exceeds 4–5 Å. Furthermore, because the SCE and SCRMP models incorporate lattice polarization, they improve the treatment of many-body effects and reduce the discrepancies between the fragment and cluster/fragment results.

Comparison of the SCE and SCRMP models highlights the benefit of accurately treating long-range electrostatic effects in ^{17}O isotropic chemical shift calculations. Both fragment and cluster/fragment methods show a ∼1 ppm reduction in RMS error versus experiment when Ewald techniques are employed using the SCRMP model. Overall, the SCRMP embedding model clearly improves the accuracy of fragment-based chemical shielding calculations. A computationally inexpensive PBE0 two-body fragment calculation using a 4 Å two-body cutoff yields well-converged ^{17}O isotropic chemical shift predictions with an accuracy comparable to both cluster PBE0 and GIPAW PBE methods. The statistical differences among the three methods are within the ∼0.5 ppm or larger uncertainties that are typical for experimental ^{17}O chemical shifts.

The box plots in Figure 3a–c compare the distribution of errors across the entire oxygen test set for fragment, cluster, cluster/fragment and GIPAW methods. These error distributions closely mirror trends in the RMS errors reported in Figure 2. Examining the range of errors of both the middle 50% of the data (the colored boxes) and the largest errors (box plot whiskers) shows improved accuracy for all models when SCRMP embedding is used. The largest improvement stemming from the SCRMP embedding treatment occurs for two-body fragment methods (Figure 3a), once again highlighting the better treatment of local many-body effects in SCRMP. Similar to previous findings,[7–9] Figure 3d shows that hybrid density functionals (PBE0 and B3LYP) provide a noticeable improvement over their GGA-based counterparts (PBE and BLYP).

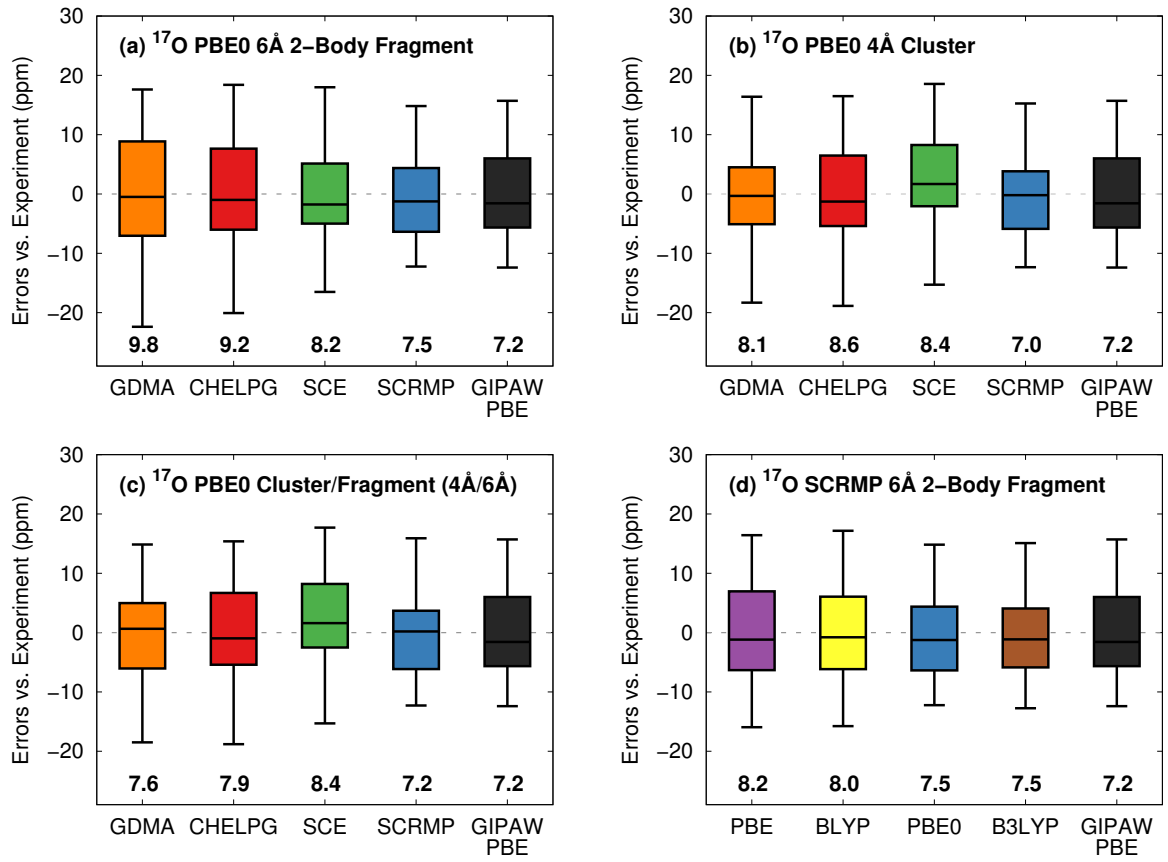


Figure 3: Distribution of errors relative to experiment for 28 predicted ^{17}O isotropic shifts and different embedding schemes with (a) a two-body fragment model, (b) an embedded cluster model (c) a combined cluster/fragment model. (d) Compares results for several different density functionals with the two-body fragment model and SCRMP embedding. 4 Å cluster and 6 Å 2-body cutoffs were employed. RMS errors (ppm) versus experiment are reported in bold.

The fragment approaches employing SCRMP are competitive with planewave GIPAW PBE. Two-body fragment PBE with SCRMP exhibits slightly larger ^{17}O isotropic chemical shifts errors (RMS 7.9 ppm) than GIPAW PBE (RMS 7.2 ppm). The difference between these is comparable in magnitude to the ∼0.5 ppm experimental uncertainties in ^{17}O chemical shifts. Switching to the PBE0 functional reduces the RMS errors to 7.5 (2-body only) and 7.2 ppm (cluster/fragment).

Table 1: Root-mean-square errors and recommended linear regression parameters (Eq 6) for the ^{1}H, ^{13}C, ^{15}N and ^{17}O test sets using either a two-body fragment model (6 Å cutoff) or a cluster/fragment model (4 Å cluster + 6 Å 2-body cutoff). Calculations employ the mixed basis and either the earlier GDMA[7,8] or current SCRMP electrostatic embedding models.

		Two-body Fragment			Cluster/Fragment		
Nucleus	**Functional**	**RMSE**	**Slope**	**Intercept**	**RMSE**	**Slope**	**Intercept**
^{1}H	PBE/GDMA	0.34	-0.9335	29.05	0.36	-0.9260	28.78
	PBE/SCRMP	0.34	-0.9311	29.00	0.34	-0.9274	28.84
	PBE0/GDMA	0.34	-0.9169	28.69	0.35	-0.9111	28.49
	PBE0/SCRMP	0.33	-0.9154	28.66	0.34	-0.9114	28.50
^{13}C	PBE/GDMA	2.09	-1.0273	180.43	2.12	-1.0240	180.25
	PBE/SCRMP	2.03	-1.0258	180.35	2.08	-1.0235	180.20
	PBE0/GDMA	1.48	-0.9676	179.58	1.47	-0.9661	179.49
	PBE0/SCRMP	1.46	-0.9666	179.53	1.46	-0.9658	179.44
^{15}N	PBE/GDMA	5.48	-1.0808	197.53	5.83	-1.0609	197.72
	PBE/SCRMP	5.18	-1.0607	196.78	5.53	-1.0541	197.40
	PBE0/GDMA	4.20	-1.0201	197.84	4.06	-0.9997	197.15
	PBE0/SCRMP	3.86	-1.0106	197.46	4.22	-0.9992	196.71
^{17}O	PBE/GDMA	11.56	-1.1440	262.61	8.79	-1.1291	264.15
	PBE/SCRMP	8.24	-1.1134	261.05	7.87	-1.1140	263.78
	PBE0/GDMA	9.80	-1.0607	270.18	7.55	-1.0502	271.60
	PBE0/SCRMP	7.47	-1.0320	268.50	7.17	-1.0361	271.11

4.2 Performance for other nuclei and linear regression parameters

To examine the performance of SCRMP more broadly, Table 1 presents benchmark data on all 63 molecular crystals and 328 isotropic ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shifts from the four test sets.[7,8] These shifts span a reasonably broad range of chemical shifts typical for organic molecular crystals. In addition to assessing the performance of different electrostatic embedding models, these benchmark studies provide linear regression models that convert predicted shieldings to chemical shifts for each nucleus under investigation. Such linear regression models partially correct for systematic errors in the chemical shift prediction and streamline the referencing of chemical shieldings computed for other crystals. Because GIAO-based fragment and cluster calculations are equally applicable to both periodic and non-periodic systems, these regression models are transferable to a wide variety of problems.

Linear regression parameters from our previous benchmark studies[7,8] have subsequently been used to solve the crystal structure of photodimerized 9-tertbutyl anthracene ester,[43] to distinguish crystal polymorphs,[9] and to elucidate features of the catalytic mechanism in the enzyme tryptophan synthase.[44,45]

From Table 1, we observe the original GDMA and new SCRMP embedding models perform similarly for both ^{1}H and ^{13}C nuclei, with negligible differences in the root-mean-square errors and regression parameters for a given functional. The fragment and cluster/fragment results also perform similarly. Errors in the predicted ^{15}N isotropic shifts vary up to a few tenths of a ppm between the two embedding models. Surprisingly, the ^{15}N cluster/fragment model results are actually often a few tenths of a ppm worse than those from the 2-body fragment model. Nevertheless, ^{17}O nuclei clearly exhibit the most pronounced sensitivity to the embedding environment, with 2–3 ppm reduction in errors for the SCRMP embedded models and substantially smaller discrepancies between the fragment and cluster/fragment model performances.

For the ^{17}O and ^{15}N sets in particular, the fitted slopes for the SCRMP results are closer to the ideal value of -1 than with the older fixed-charge GDMA model. With the GDMA model, constraining the slope to -1 in the ^{17}O regression increases the root-mean-square errors from 9.8 (2-body) and 7.6 ppm (cluster/fragment) to 11.0 and 8.6 ppm, respectively. With the SCRMP model on the other hand, the errors only increase marginally from 7.5 (2-body) and 7.2 (cluster/fragment) ppm to 7.9 and 7.8 ppm, respectively. This data indicates that the SCRMP model is providing a more robust description of the chemical shielding with less systematic error that the regression needs to compensate for.

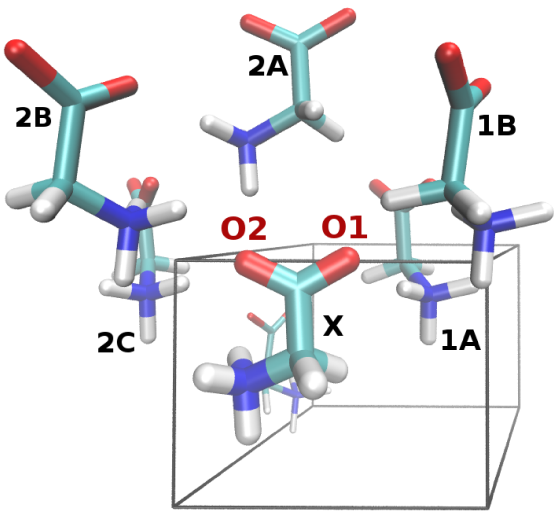
Although employing the SCRMP model for ^{1}H, ^{13}C, and ^{15}N nuclei does not have substantial effects on the statistical accuracy of the resulting chemical shifts, the additional computational costs it introduces are modest, and it does provide an improved physical description of the crystalline environment, especially for polar unit cells. We therefore recommend its use for all nuclei.

4.3 γ-Glycine ^{17}O chemical shielding predictions

Finally, we apply the SCRMP model and the linear regression parameters from Section 4.2 to predict the ^{17}O isotropic chemical shifts in γ-glycine. This polymorph exhibits zwitterionic monomers and a non-zero net dipole moment in the unit cell, making careful treatment of the electrostatic environment essential to accurate shift prediction. As shown in Figure 4, the two oxygens occupy different hydrogen bonding environments: O1 has two nearby ammonium groups (labeled **1A** and **1B**), while O2 has three (**2A**, **2B**, and **2C**). The experimentally measured isotropic shifts for O1 and O2 differ by about 8 ppm (280 and 272 ppm, respectively), with the smaller isotropic shift for O2 being consistent with its stronger hydrogen bonding.[46] Earlier B3LYP calculations on a 12-molecule cluster with no electrostatic embedding overestimated the isotropic shifts by 30–40 ppm (B3LYP/Cluster in Table 2).[46] Those calculations also failed to distinguish the two hydrogen bonding environments, predicting only a 2 ppm separation between the shifts and reversing the ordering of the assignments.

We revisit this system with the fragment approaches and GIPAW PBE. The SCRMP embedding is essential in the fragment models due to the polar unit cell. Fragment calculations without SCRMP converge poorly with respect to both the DFT self-consistent field iterations and the fragment model cutoffs. The PBE-D2 optimized atomic positions in the unit cell agree excellently with experiment, with the key hydrogen bond distances matching neutron diffraction data[25] to within $\sim$0.04 Å. Table 2 lists the predicted ^{17}O isotropic shifts relative to experiment for both O1 and O2. The fragment and cluster/fragment shieldings were translated to chemical shifts via the SCRMP-derived linear regression parameters from Table 1, while the GIPAW PBE results use the regression parameters from our earlier work.[8]

Compared to the earlier simple B3LYP/Cluster calculations, the embedded two-body fragment model with either PBE or PBE0 reduces the errors in the chemical shifts significantly. Still, O2 appears particularly problematic, with a 29 ppm error for PBE. The hybrid PBE0 functional helps, reducing that error to 19 ppm (and the O1 error down to only 2 ppm). Nevertheless, neither set of two-body fragment model predictions orders the two



	$\Delta^3\sigma_{i,j,k}$(ppm)	
Trimer	**Site O1**	**Site O2**
X,1A,1B	2.2	-0.2
X,2A,2B	-0.4	5.0
X,2A,2C	-0.1	3.0
X,2B,2C	-0.4	3.4

Figure 4: Molecules surrounding the two oxygen atoms of the monomer **X** in the asymmetric unit of γ-glycine. Several important 3-body PBE0 contributions to the chemical shielding of atoms O1 and O2 are listed.

Table 2: Predicted and experimental ^{17}O isotropic chemical shifts for γ-glycine, in ppm. Entries in bold order the two shifts correctly.

	Site O1		**Site O2**	
	δ_{iso}	$\lvert\delta_{pred} - \delta_{expt}\rvert$	δ_{iso}	$\lvert\delta_{pred} - \delta_{expt}\rvert$
B3LYP/Cluster (Ref 46)	310.2	30.2	312.5	40.4
PBE/Fragment	285.2	5.2	301.3	29.3
PBE/Cluster/Fragment	**286.8**	**6.8**	**282.3**	**10.3**
PBE/GIPAW (Ref 47)	**285.7**	**5.7**	**278.3**	**6.3**
PBE0/Fragment	282.1	2.1	291.2	19.2
PBE0/Cluster/Fragment	**284**	**4.0**	**276.5**	**4.5**
Experiment (Ref 46)	280(1)		272(1)	

chemical shifts correctly.

Switching to the cluster/fragment model improves the situation. Now both the PBE and PBE0 functionals order the two shifts correctly, and the errors are more uniform between the two oxygen atoms. GIPAW PBE also orders the shifts correctly, with smaller errors than cluster/fragment PBE, and slightly larger errors than cluster/fragment PBE0. These methods allow direct assignment of the two experimental oxygen shifts based on calculation alone.

The substantial difference between the fragment and cluster/fragment results suggests

particularly strong many-body effects beyond what are captured in the embedding model, especially for O2. To investigate these effects, the three-body contributions to the absolute isotropic shielding were computed for the monomers that interact most strongly with O1 and O2 (Figure 4). The three-body contribution to O1 and its two nearby monomers **1A** and **1B** is a fairly strong 2.2 ppm. However, the three-body contributions to O2 and its neighbors **2A**–**2C** are much stronger at 3–5 ppm each.

The 2-body fragment model omits these and other 3-body contributions, which explains the particularly large errors for O2. In fact, augmenting the 2-body fragment model with just these four three-body terms would be sufficient to order the two oxygen chemical shifts correctly. Unlike the 2-body model, both GIPAW and the cluster/fragment model capture those contributions naturally, which explains the more uniform-quality treatment of O1 and O2 in those models. In practice, we believe the cluster/fragment approach is preferable to a model that includes local three-body terms. The cluster implicitly sums the local interactions up through all n bodies in the cluster, without any of the potential numerical precision errors that can arise when summing many small, higher-order contributions.[48] The additional computational cost for a cluster calculation compared to a three-body model at the DFT level is modest for a typical 4 Å cluster containing ∼15-molecules.

This strong many-body contribution outside what is captured by the embedding model appears to be somewhat unusual, probably due to the high concentration of charged functional groups in spatial proximity. In γ-glycine, the difference between the 2-body fragment and cluster/fragment shifts is 15 ppm. In contrast, the root-mean-square difference in isotropic shift between the 2-body fragment and cluster/fragment models for the ^{17}O test set discussed in Section 4.1 is only 1.5 ppm, with a max difference of 3.9 ppm.

5 Conclusion

In summary, an improved, self-consistent reproduction of the Madelung potential embedding model has been developed for fragment-based NMR chemical shift prediction. This model combines an inner layer of explicit atomic charges on molecules close to the QM region with a second layer of more distant point charges optimized to reproduce the overall Madelung potential in the QM region. To incorporate polarization in the crystal lattice, the point charges are determined self-consistently.

This SCRMP embedding model is particularly important for treating ^{17}O (and ^{15}N to a lesser extent) chemical shifts and for crystals with polar unit cells. Employing the SCRMP model substantially narrows the performance gap between 2-body fragment and cluster-type models, which means that inexpensive 2-body fragment models can be used more often. The γ-polymorph of glycine demonstrates that many-body effects missed by the embedded 2-body fragment model can still be important, but such cases appear somewhat atypical.

Overall, combining SCRMP-embedded fragment or cluster/fragment calculations with the hybrid PBE0 functional allows one to predict chemical shifts with accuracy that matches (^{17}O) or exceeds (^{1}H, ^{13}C, and ^{15}N) the accuracy of the widely used GIPAW PBE model. The linear regression parameters that map predicted chemical shieldings to observable isotropic shifts reported here should be generally applicable to molecular crystals and non-periodic systems modeled with the same fragment approaches. Future work will examine the treatment of chemical shielding anisotropies and other magnetic properties with these techniques.

6 Acknowledgments

Funding for this work from the National Science Foundation (CHE-1362465 and CHE-1665212) and supercomputer time from XSEDE (TG-CHE110064) are gratefully acknowledged. We thank Prof. Graeme Day for providing GIPAW PBE results for the γ-polymorph of glycine.

Supporting Information Available

Tables of SCRMP PBE0 chemical shieldings and shifts for the ^{1}H, ^{13}C, ^{15}N, and ^{17}O test sets, the predicted shieldings and shifts for the γ-polymorph of glycine, and the coordinates for the optimized γ-glycine structure are provided. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

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

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

(3) 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.

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

(5) Beran, G. J. O.; Hartman, J. D.; Heit, Y. N. *Acc. Chem. Res.* **2016**, *49*, 2501–2508.

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

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

(8) Hartman, J. D.; Kudla, R. A.; Day, G. M.; Mueller, L. J.; Beran, G. J. O. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21686–21709.

(9) Hartman, J. D.; Day, G. M.; Beran, G. J. O. *Cryst. Growth Des.* **2016**, *16*, 6479–6493.

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

(11) Stueber, D.; Guenneau, F. N.; Grant, D. M. *J. Chem. Phys.* **2001**, *114*, 9236–9243.

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

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

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

(15) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. *Science* **2017**, *355*, 49–52.

(16) Hammes-Schiffer, S. *Science* **2017**, *355*, 28–29.

(17) Misquitta, A.; Stone, A.; Fazeli, F. *J. Chem. Theory Comput.* **2014**, *10*, 5405–5418.

(18) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.

(19) Ewald, P. P. *Annalen der Physik* **1921**, *369*, 253–287.

(20) Hirata, S. *J. Chem. Phys.* **2008**, *129*, 204104.

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

(22) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233–239.

(23) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047–1064.

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

(25) Boldyreva, E.; Drebushchak, T.; Shutova, E.; Kristallogr, Z. *Cryst. Mater.* **2003**, *218*, 366.

(26) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.;

Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. *J. Phys. Cond. Mat.* **2009**, *21*, 395502 (19pp).

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

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

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

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

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

(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision E.01. 2009; Gaussian Inc. Wallingford CT.

(33) Chesnut, D. B.; Moore, K. D. *J. Comp. Chem.* **1989**, *10*, 648–659.

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

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

(36) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.

(37) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639–5648.

(38) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265–3269.

(39) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; v. R. Schleyer, P. *J. Comp. Chem.* **1983**, *4*, 294–301.

(40) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(41) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.

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

(43) Yang, C.; Zhu, L.; Kudla, R. A.; Hartman, J. D.; Al-Kaysi, R. O.; Monaco, S.; Schatschneider, B.; Magalhaes, A.; Beran, G. J. O.; Bardeen, C. J.; Mueller, L. J. *CrystEngComm* **2016**, *18*, 7319–7329.

(44) Young, R. P.; Caulkins, B. G.; Borchardt, D.; Bulloch, D. N.; Larive, C. K.; Dunn, M. F.; Mueller, L. J. *Angew. Chem. Int. Ed.* **2016**, *55*, 1350–1354.

(45) Caulkins, B. G.; Young, R. P.; Kudla, R. A.; Yang, C.; Bittbauer, T. J.; Bastin, B.; Hilario, E.; Fan, L.; Marsella, M. J.; Dunn, M. F.; Mueller, L. J. *J. Am. Chem. S* **2016**, *138*, 15214–15226.

(46) Yamada, K.; Honda, H.; Yamazaki, T.; Yoshida, M. *Sol. State NMR* **2006**, *30*, 162–170.

(47) Day, G. M. 2017; private communication.

(48) Richard, R. M.; Lao, K. U.; Herbert, J. M. *Acc. Chem. Res.* **2014**, *47*, 2828–36.

Graphical TOC Entry