# Accurate $^{13}$C and $^{15}$N molecular crystal chemical shielding tensors from fragment-based electronic structure theory

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

[†]*Department of Chemistry, University of California, Riverside, California 92521 USA*
[‡]*Department of Chemistry, Mt. San Jacinto College, Menifee, California 92584 USA*

E-mail: gregory.beran@ucr.edu

August 1, 2018

## Abstract

Standard nuclear magnetic resonance (NMR) spectroscopy experiments measure isotropic chemical shifts, but measuring the chemical shielding anisotropy (CSA) tensor can provide additional detailed insights into solid state chemical structures. Interpreting the principal components of these tensors is greatly facilitated by first-principles chemical shielding tensor predictions. Here, the ability to predict molecular crystal CSA tensor components for $^{13}$C and $^{15}$N nuclei with fragment-based electronic structure techniques is explored. Similar to what has been found previously for isotropic chemical shifts, the benchmarking demonstrates that fragment-based techniques can accurately reproduce CSA tensor components. The use of hybrid density functionals like PBE0 or B3LYP provide higher accuracy than generalized gradient approximation functionals like PBE. Unlike for planewave density functional techniques, hybrid density functionals can be employed routinely with modest computational cost in fragment approaches. Finally, good consistency between the regression parameters used to map either isotropic shieldings or CSA tensor components is demonstrated, providing further evidence for the quality of the models and highlighting that models trained for isotropic shifts can also be applied to CSA tensor components.

## 1 Introduction

Solid-state nuclear magnetic resonance (NMR) spectroscopy provides detailed local chemical structure information in crystals through the chemical shielding tensor. This 3 × 3 tensor can be defined as the second derivative of the energy with respect to the magnetic field and the nuclear magnetic moment,

$$\sigma^A_{\alpha\beta} = \frac{\partial^2 E}{\partial B_\alpha \partial \mu^A_\beta} \tag{1}$$

In typical NMR experiments, motions of the molecules and/or sample average over various orientations, and one measures the isotropic chemical shift that stems from the trace of the shielding tensor. However, it is also possible to extract the three principal components (eigenvalues) of this chemical shielding anisotropy (CSA) tensor, $\sigma_{11}$, $\sigma_{22}$, and $\sigma_{33}$, from experiments on single crystals, analysis of powder pattern spectra, spinning side band analysis, and other techniques.[1]

Chemical shielding anisotropy arises from distortion of the symmetry of the electron distribution around a nucleus in the presence of a magnetic field, and therefore CSA principal components can provide detailed insights into local structure, geometry, orientation, and even hydrogen bonding. For example, the principal components of the $^{15}$N shielding tensor are sensitive to the hydrogen bonding environment, and measuring their values for backbone amide nitrogens can facilitate protein structure elucidation and refinement.[2]

Relative to the isotropic chemical shift, CSA components are more sensitive to local geometry, and have greater uncertainty in the measured values.[3,4] Therefore, gleaning structural information from the CSA tensors can be challenging. Computational prediction can help by allowing assessment of the agreement between the predicted principal components for hypothetical structures and the experimentally observed ones, or by revealing how structural changes alter the CSA components.[2,5–8]

Of course, the utility of computation CSA tensor prediction depends on its accuracy. The gauge-including projector augmented wave (GIPAW) method, using periodic density functional theory (DFT), has proven highly successful for predicting magnetic properties in extended crystal systems.[9,10] However, practical constraints limit plane wave calculations to GGA-type density functionals. Recent benchmark studies involving $^{1}$H, $^{13}$C, $^{15}$N, and $^{17}$O isotropic shielding predictions indicate that fragment-based calculations using hybrid density functionals offer improved accuracy for isotropic chemical shift prediction.[11] Similar findings have been reported using the symmetry-adapted cluster (SAC) method.[12,13] In turn, higher accuracy in the chemical shifts translates to increased discrimination between different potential structural assignments.[14]

While many studies investigate the accuracy of isotropic chemical shift predictions, fewer have looked at the CSA tensor elements. Existing studies indicate that statistical errors in the $^{13}$C principal components for molecular crystals are typically 2–3-fold higher than those for the isotropic shifts.[13,15–18] In this paper, we confirm that the fragment-based methodology performs well for $^{13}$C and $^{15}$N CSA principal component values using a benchmark test set comprised of 477 individual principal components.

The benchmarks presented here demonstrate comparable accuracy in the predicted $^{13}$C and $^{15}$N CSA tensors for fragment methods and GIPAW using the PBE density functional. In agreement with previous findings, fragment-based calculations employing the PBE0 and B3LYP hybrid density functionals predict the principal components more accurately relative to experiment than those computed from GGA functionals like PBE and BLYP, mimicking the trends found for isotropic chemical shifts in previous work.[11] These trends agree with previous $^{13}$C CSA tensors benchmark studies.[13,15] However, the benchmarks here do not reproduce earlier reports which found favorable performance for the TPSS and TPSSh meta-GGA/hybrid density functionals for $^{15}$N CSA tensors.[13]

In addition, it is common to reference the predicted chemical shielding values to experimentally observable chemical shifts via linear regression.[19] In addition to making it possible to use more accurate hybrid density functionals, fragment methods have the additional benefit of being readily applicable between periodic and non-periodic systems. This means, for example, that one can obtain reliable regression parameters from a training set of well-characterized molecular crystal structures[11,15] and then apply these parameters to biological systems[20,21] whose structures are harder to characterize due to dynamics, solvent, etc.

Here, we also investigate the consistency between the regression parameters obtained from CSA principal components versus those obtained from isotropic chemical shifts. Given the close relationships between the CSA principal components and the isotropic shifts, regression lines fitted to either set ought to be consistent. In practice, the regression models help compensate for systematic errors in the calculations due to finite basis sets or other issues. The level of consistency between the different sets of parameters provides a diagnostic for the quality of the magnetic property modeling. Here, we demonstrate that the regression parameters fitted to CSA tensor components produce predictions that are consistent with those fitted to isotropic shift data, and that regressions trained on isotropic chemical shift data are useful for predicting CSA principal components. This is useful because experimental isotropic chemical shifts can generally be measured much more accurately than CSA tensor components, which makes them better suited for calibrating the first-principles chemical shift models.

## 2 Theory

Fragment-based chemical shielding calculations rely on a many-body expansion for the shielding tensor

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

obtained by differentiating the many-body expansion for the energy with respect to the nuclear magnetic moment and the external magnetic field (Eq. 1). The details of fragment-based chemical shift calculations have been described previously,[11,12,15,22] In brief, the expansion (Eq 2) decomposes the shielding tensor of atom $A$ on molecule $i$ in the unit cell into the tensor on the isolated molecule ($\tilde{\boldsymbol{\sigma}}_i^A$) with corrections to account for interactions of that molecule with other molecules in the crystal.

In practice, evaluation of the three-body ($\Delta^3 \tilde{\boldsymbol{\sigma}}_{ijk}^A$) and higher order terms in Eq 2 would present a significant computational burden. However, studies have shown that the sum of electrostatically embedded one- ($\boldsymbol{\sigma}_i^{A,emb.}$) and two-body ($\Delta^2 \tilde{\boldsymbol{\sigma}}_{ij}^A$) contributions

$$\tilde{\boldsymbol{\sigma}}_i^A \approx \boldsymbol{\sigma}_i^{A,emb.} + \sum_{ij} \Delta^2 \boldsymbol{\sigma}_{ij}^{A,emb.} \quad (3)$$

effectively reproduces the isotropic chemical shielding to within the accuracy of density functional theory.[11,15,22,23]

In the context of a molecular crystal, a two-body fragment-based calculation is carried out by defining a cutoff radius around the asymmetric unit. The chemical shielding tensor for each atom in the asymmetric unit is approximated by calculating the one-body contribution for each monomer in the asymmetric unit and adding two-body corrections for all dimers lying with the defined cutoff. Each monomer and dimer calculation is performed in an electrostatic embedding environment constructed using the self-consistent reproduction of the Madelung potential (SCRMP) method (see below).[23] Focusing on the asymmetric unit and incorporating space group symmetry provides significant computational savings. Additional computational savings are achieved using locally dense basis sets[24] that employ larger basis sets on the atoms in the asymmetric unit, and smaller basis sets on more distant atoms.

Chemical shielding tensor principal components are particularly sensitive to the local chemical environment.[3] To evaluate the errors associated with truncating the many-body expansion, we compare two-body fragment methods with a combined cluster/fragment approach. The cluster/fragment calculations describe the local interactions (out to 4 Å) with a single point charge-embedded cluster calculation. This approach is equivalent to summing the many-body expansion (Eq. 2) through all orders for the nearest neighboring molecules surrounding the asymmetric unit. Longer-range interactions outside of the cluster region but inside the two-body cutoff are handled in a pairwise fashion.

Treating the local interactions with a cluster is generally more accurate than the pure pairwise fragment approach,[11,15,23] but it also is appreciably more computationally expensive. For a typical molecular crystal studied here, the 4 Å cluster employed in cluster/fragment might contain ~15 molecules surrounding the central molecule in the asymmetric unit. Given the formally cubic scaling of DFT with system size, it can take an order of magnitude more computational effort to evaluate the contributions of those 15 surrounding molecules to the chemical shifts on the central monomer in a large cluster instead of in the pairwise fashion used in the two-body fragment approach. Therefore, assuming it proves sufficiently accurate, one would prefer to use the fragment-only approach over the cluster/fragment one.

The electrostatic embedding environment plays a key role in the accuracy of the two-body fragment approach.[23] Embedding the fragment calculations polarizes the molecules to mimic the crystalline environment, effectively incorporating some of the many-body effects into the one-body and two-body contributions. The SCRMP model[23] used here represents the crystalline environment as a series of point charges which are designed to reproduce the Madelung potential at each atomic center being treated quantum mechanically. The SCRMP approach is very similar to the older embedded ion method,[25] but adapted to the fragment-type calculations used here. Briefly, the SCRMP model first obtains atom-centered CHELPG point charges[26] for each molecule in the unit cell. The CHELPG charges are placed on all atomic centers in the crystal lattice within 30 Å of the asymmetric unit. The CHELPG charges are surrounded with another shell of atom-centered charges which are least-squares fitted to ensure the overall embedding environment approximates the Madelung potential throughout the region enclosed by the two-body cutoff. The CHELPG charges are then re-computed in this environment, and the processes is iterated to self-consistency to capture the many-body polarization effects.

# 3 Computational Methods

## 3.1 Crystal structures

The performance of fragment and cluster/fragment NMR shielding tensor calculations is assessed here using two benchmark sets containing a total of 29 molecular crystals, with 432 $^{13}$C and 45 $^{15}$N experimental principal components. The $^{13}$C test set was taken from ref 15 and the $^{15}$N was obtained from refs 3 and 27. The crystal structures included in the benchmark sets represent a variety of chemical environments with chemical shielding components spanning over 250 ppm for $^{13}$C and 400 ppm for $^{15}$N.

Optimized crystal structures for the species in the $^{13}$C test set were obtained previously.[15] The six experimental nitrogen structures were obtained from the Cambridge Structure Database (CSD): CIMETD03,[28] GLYCIN18,[29] HISTCM01,[30] HXACAN26,[31] THYMIN01,[32] GLCICH01.[33] The atomic positions from the experimental crystal structures were fully relaxed subject to fixed experimental room-temperature lattice parameters using dispersion-corrected density functional theory (DFT) with the PBE functional and D2 dispersion corrections as implemented in the open-source Quantum Espresso[34] software package. Ultrasoft pseudopotentials were employed with a plane wave cutoff of 80 Ry and a 3x3x3 Monkhorst Pack *k*-point grid. The following list of ultrasoft pseudopotentials, obtained from http://www.quantum-espresso.org were employed: H.pbe-rrkjus.UPF, C.pbe-rrkjus.UPF, N.pbe-rrkjus.UPF, and O.pbe-rrkjus.UPF. This protocol is identical to the one previously applied to the crystals in the $^{13}$C test set.[15]

## 3.2 Chemical shielding calculations

The GIPAW PBE shielding tensor components presented here come from previous studies.[3,15,27] Fragment and cluster/fragment NMR chemical shielding tensor calculations were carried out on the relaxed crystal structures. Molecular crystal fragmentation through two-body was performed using our hybrid many-body interaction (HMBI) code.[35] Shielding calculations for individual fragments were carried out using Gaussian 09[36] with the PBE0,[37] PBE,[38] B3LYP,[39] BLYP,[40,41] OPBE, TPSS, and TPSSh density functionals. A large DFT integration grid consisting of 150 radial and 974 Lebedev angular points was used to approach rotational invariance and minimize numerical noise the fragment calculations. Electrostatic embedding was carried out using the SCRMP embedding model.[23] The SCRMP model uses CHELPG atom-centered monopoles calculated with Gaussian 09 at the same density function/basis set combination used in the NMR chemical shielding calculation.

All fragment and cluster/fragment calculations were performed using a locally dense basis set[24] and the gauge-including atomic orbital (GIAO) approximation. The locally dense basis set uses a 6-311+G(2d,p) basis on all atoms in the asymmetric unit and a surrounding 2 Å region, a 6-311G(d,p) basis for all atoms within 2 to 4 Å, and a 6-31G basis on all atoms beyond 4 Å.[11,15] Errors associated with basis set incompleteness are partially compensated for using a linear regression referencing approach described below.

The principal components are obtained by symmetrizing the NMR chemical shielding tensor and then diagonalizing it. The resulting eigenvalues (principal components) are denoted $\sigma_{11}$, $\sigma_{22}$, and $\sigma_{33}$. The isotropic shielding is the average of the principal components.

$$\sigma_{iso} = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) \tag{4}$$

Mapping between absolute shielding ($\sigma_{ii}$) obtained from computation and the experimentally observed chemical shift anisotropy tensor elements ($\delta_{ii}$) was carried out via linear regression,[19]

$$\delta_{ii} = A\sigma_{ii} + B \tag{5}$$

Linear regression models of this form were applied separately to each model/cutoff combination to obtain model-specific regression parameters $A$ and $B$. When fitted to well-curated training sets, the regression approach to chemical shift referencing helps remove systematic errors in the chemical shielding tensor calculations due to basis set incompleteness, imperfect structure modeling, and other modeling errors. Further, the regression parameters from statistically robust training sets are transferable and can be directly applied to compounds not included in the test set.

The isotropic shifts and principal components of the NMR chemical shielding tensor reported here are referenced relative to TMS under magic angle spinning (MAS) conditions for all $^{13}$C nuclei and the $^{15}$N resonance of crystalline glycine at 33.4 ppm for all $^{15}$N nuclei.

# 4 Results and Discussion

The results here focus on both $^{13}$C and $^{15}$N chemical shielding tensors. For carbon, 432 experimental principal components have been gathered from 23 different crystal structures and spanning a ∼250 ppm range. Reported experimental uncertainties ranging from ∼0.4 ppm to more than 2 ppm are typical in the literature.[17,18,42–46]

Compared to the $^{13}$C ones, $^{15}$N shieldings present additional challenges from both an experimental and computational perspective. Experimental uncertainties in $^{15}$N shielding tensor measurements are often larger than for $^{13}$C due to the much lower sensitivity of $^{15}$N, especially at natural abundance. For the L-histidine monohydrochloride monohydrate crystal, for example, the most accurately known $^{15}$N tensor values from single crystal experiments are estimated to exhibit ±5.5 ppm rms uncertainty.[47] Computationally, $^{15}$N shielding tensors are more sensitive to the subtleties of the electrostatic environment (both long-range and many-body effects).

To reduce the influence of confounding variables on the $^{15}$N tensor data, we test the models on an experimentally consistent set of 45 individual $^{15}$N shielding tensor principal components measured using the accurate FIREMAT approach.[3,27] For the aforementioned histidine·HCl·$H_2O$ crystal, the FIREMAT data lies within rms ±2.8 ppm of the single crystal values (i.e. well-within the experimental uncertainty of those values). The 45 principal components in this set originate from systems with well-defined crystal structures and span a range of over 400 ppm. On the other hand, the smaller size of this test set and the non-uniform distribution of principal component shift values limits the transferability of the chemical shift tensor regression models derived from it.

## 4.1 Two-body cutoff convergence and many-body contributions

Fragment methods for predicting chemical shifts in molecular crystals are successful due to the inherently local nature of the chemical shifts. To predict the shifts correctly on a molecule in the asymmetric unit, one needs to treat interactions with nearby atoms explicitly. However, the longer-range contributions arising

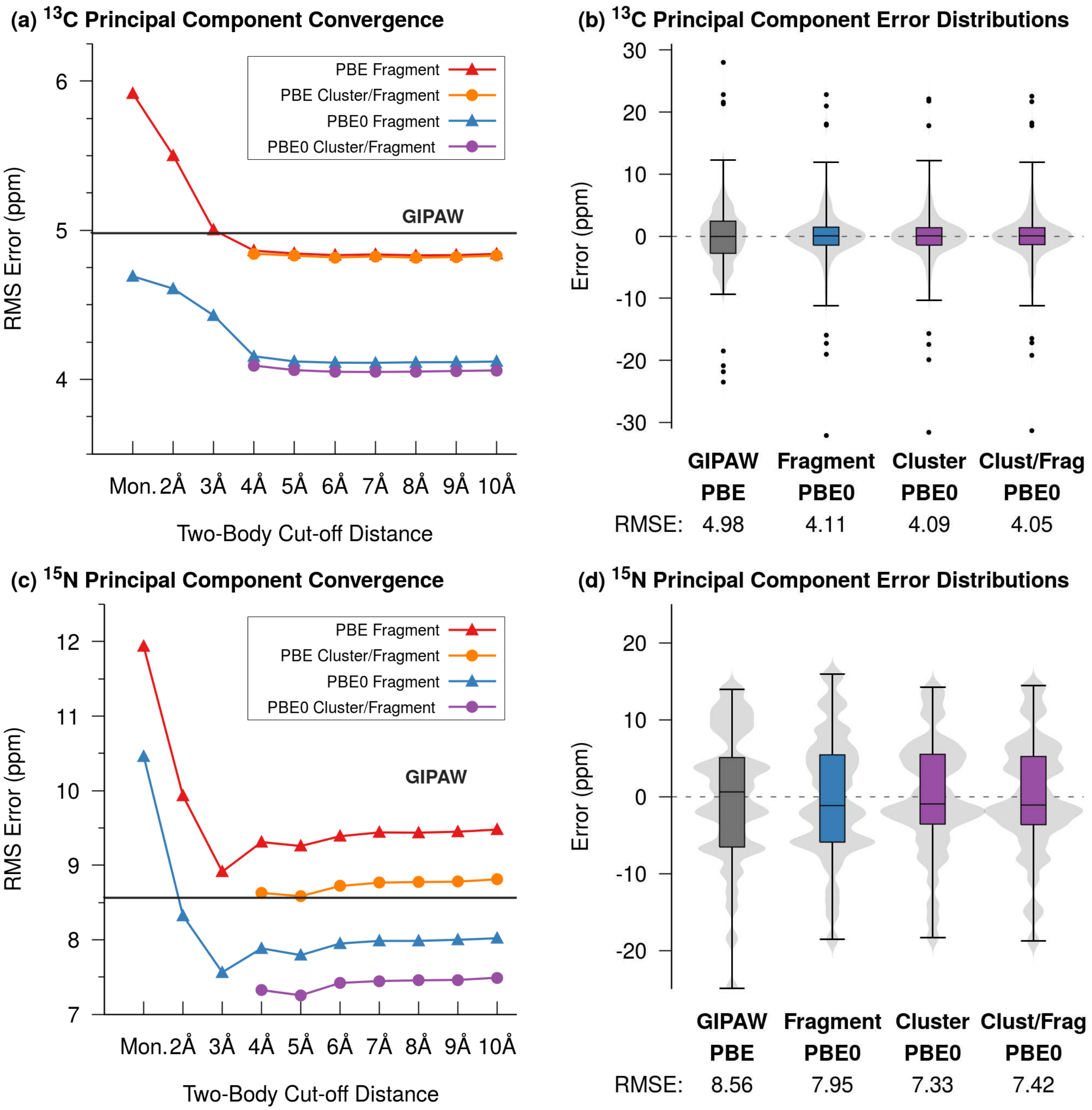


**Figure 1:** Convergence of the predicted principal components for the (a) $^{13}$C and (c) $^{15}$N benchmark sets with respect to the two-body cutoff distance. Comparison of the principal component error distributions for GIPAW PBE, fragment PBE0, cluster PBE0, and cluster/fragment PBE0 models. 4 Å clusters were used, and for (b) and (d), fragment contributions out to 6 Å were included.

from the crystal lattice can be approximated readily via electrostatic embedding. In the fragment approach, interactions of the molecule(s) in the asymmetric unit with nearby ones are evaluated quantum mechanically if any atoms on the nearby molecules lie within the two-body cutoff radius of the central molecule. Outside that radius, the contributions are included only via the SCRMP point charge embedding.

Previous studies examining isotropic shielding calculations using fragment and cluster approaches and the SCRMP embedding model have shown rapid convergence with respect to the two-body cutoff distance, with interactions beyond 6 Å being well-captured by the embedding model.[23] Here we examine the impact of local many-body effects and two-body cutoff distance on the accuracy of fragment and cluster-based calculations involving the three principal components of the chemical shielding tensor rather than the isotropic shielding. Figures 1a and 1c illustrate the root-mean-square errors in the calculated principal components of the magnetic shielding tensor relative to experiment for all $^{13}$C and $^{15}$N nuclei, respectively. Separate linear regression models (Eq 5) were applied to each model/cut-off combination.

Figures 1a and 1c show marked reduction in root-mean-square (rms) errors once the two-body cutoff is sufficiently large to include all hydrogen-bonding partners (2–3 Å two-body cutoff), and two-body contributions beyond ∼4 Å impact the rms errors only modestly. The rapid and stable convergence with respect to two-body cutoff stems in part from the accurate SCRMP model electrostatic embedding, and the behaviors parallel those previously observed in our analogous studies on isotropic chemical shifts.[23] While we generally adopt a conservative 6 Å cutoff, a 4 Å two-body cutoff is viable for computationally demanding applications.

The chemical shift tensor components are sensitive to the local chemical environment.[3] Differences in the relative performance of the fragment and cluster/fragment methods reflect the influence of local many-body effects on the calculated principal components. Figure 1a shows a negligible difference in accuracy between fragment and cluster/fragment calculations for $^{13}$C principal components. On the other hand, the $^{15}$N principal components of the shielding tensor prove slightly more sensitive to many-body effects, with the rms errors reducing by 7% upon switching from a pure fragment model to a cluster/fragment one (Figure 1c). However, the difference in $^{15}$N rms errors between fragment and cluster/fragment models is small relative to the experimental uncertainties expected in the measured principal components. Taken together these results highlight the efficacy of the combined fragment/SCRMP model in terms of capturing the local many-body effects necessary to reproduce the chemical shift anisotropy.

For additional insight, Figures 1b and 1d compare the error distributions for GIPAW PBE, fragment PBE0, cluster PBE0, and cluster/fragment PBE0. These plots combine box and violin plots. The box plots include 50% of the data within the boxes, 98% of the data within the whiskers, and any remaining outliers are indicated as points. The violin plots show kernel density estimates for the same error distributions, providing a slightly more nuanced visualization of the data. The cluster calculations employ only a 4 Å cluster, while the cluster/fragment ones augment the 4 Å cluster with pairwise contributions out to 6 Å.[15]

For $^{13}$C, the various fragment and cluster models perform very similarly, with error distributions that are more sharply peaked about zero and rms errors that are 18% smaller than those for GIPAW PBE. The differences between the cluster and cluster/fragment model are small, again highlighting how effectively the electrostatic embedding captures long-range contributions to the shielding tensors. Figure 1a also indicates that the fragment models with the PBE functional give results quite similar to GIPAW PBE, indicating that the improvement here stems from the choice of functional rather than the fragment versus planewave modeling approach. Notably, each model in Figure 1b exhibits a handful of outlier tensor components, particularly for glutamine (CSD RefCode GLUTAM01). These outliers are consistent across the different methods, suggesting that these errors likely stem either from some issue in the DFT-optimized crystal structure or from the experimental measurements of the CSA tensors.

For $^{15}$N, once again the fragment and cluster approaches with the hybrid PBE0 functional improve upon GIPAW PBE by 7–13%. Interestingly, the $^{15}$N error distributions are less normally distributed for all methods, though this might simply reflect the smaller size of the data set. In any case, the error distributions in Figure 1d generally shift density toward zero as the model switches from PBE to PBE0 and as explicit many-body effects are incorporated via the cluster and cluster/fragment approaches. Still, even the pure fragment model performs reasonably well for $^{15}$N principal components. Five of the seven largest principal component errors in the $^{15}$N set occur for cimetidine. However, given the smaller size of the test set, it is unclear whether these are outliers or simply reflect the typical errors associated with the models.

## 4.2 Comparisons among different density functionals

The choice of density functional can have an appreciable impact on the accuracy of computed NMR parameters.[11,13,15] Numerous recent studies examining isotropic chemical shielding for $^{1}$H, $^{13}$C, $^{14}$N and $^{17}$O nuclei show improved performance for hybrid density functionals relative to their GGA counterparts.[11,12,15,23] The previous section already showed ∼15% improvements in the quality of the predicted chemical shielding principal components from using the hybrid PBE0 functional instead of PBE. Here, we investigate several additional functionals to bolster the observation that hybrid functionals perform better for the principal components as well as the isotropic shifts.

Holmes et al[13] examined the influence of density functional choice on 393 principal components of chemical shielding tensors from $^{13}$C, $^{15}$N, $^{19}$F, and $^{31}$P nuclei. In agreement with previous isotropic shielding results,[11,12,15] that study found that hybrid density functionals improve the computed results over GGA functionals for $^{13}$C, $^{19}$F, and $^{31}$P nuclei. Interestingly, $^{15}$N demonstrated anomalous behavior in that work, with hybrid density functionals performing worse than their GGA counterparts.[13] Additionally, the meta-GGA density functionals TPSS and TPSSh outperformed both the hybrid and pure density functionals for $^{15}$N. Here we explore the relative performance of a variety of commonly used density functionals on the accuracy of calculated principal components for 144 $^{13}$C nuclei and 15 $^{15}$N nuclei and compare our findings to these previous studies.

The accuracy of fragment and cluster-based *ab initio* principal component calculations for all 432 $^{13}$C principal components used here were reported previously.[15] However, the fragment and cluster/fragment calculations in that earlier work employed a simpler electrostatic embedding model based on atom-centered Gaussian distributed multipole analysis (GDMA) point charges[48,49] instead of the self-consistently polarized SCRMP model ones used here.

Figure 2 reports error distributions and root-mean-square errors for the $^{13}$C isotropic shifts and principal components using both fragment and cluster/fragment calculations for half a dozen different density functionals and for the widely used GIPAW PBE approach. The isotropic results in Figures 2a and 2b extend our previous analysis of the SCRMP embedding model[23] to include the TPSS, BLYP, OPBE, TPSSh and B3LYP density functionals. In agreement with the results from our previous benchmark studies,[11,15] the GGA functionals show relatively uniform performance with rms errors around 2 ppm. OPBE performs ∼0.2 ppm better than the other GGAs, as seen previously.[15] The hybrid density functionals uniformly out-perform their GGA counterparts with B3LYP and PBE0 giving rms errors under 1.3 ppm for both fragment and cluster/fragment calculations. Although the meta-hybrid density functional TPSSh out performs its GGA counterpart, the rms errors are ∼0.3 ppm greater relative to both PBE0 and B3LYP.

Trends for the errors associated with the principal component calculations mirror those of the isotropic shieldings (Figures 2c and 2d), except the errors are approximately three times larger. Specifically, hybrid density function-

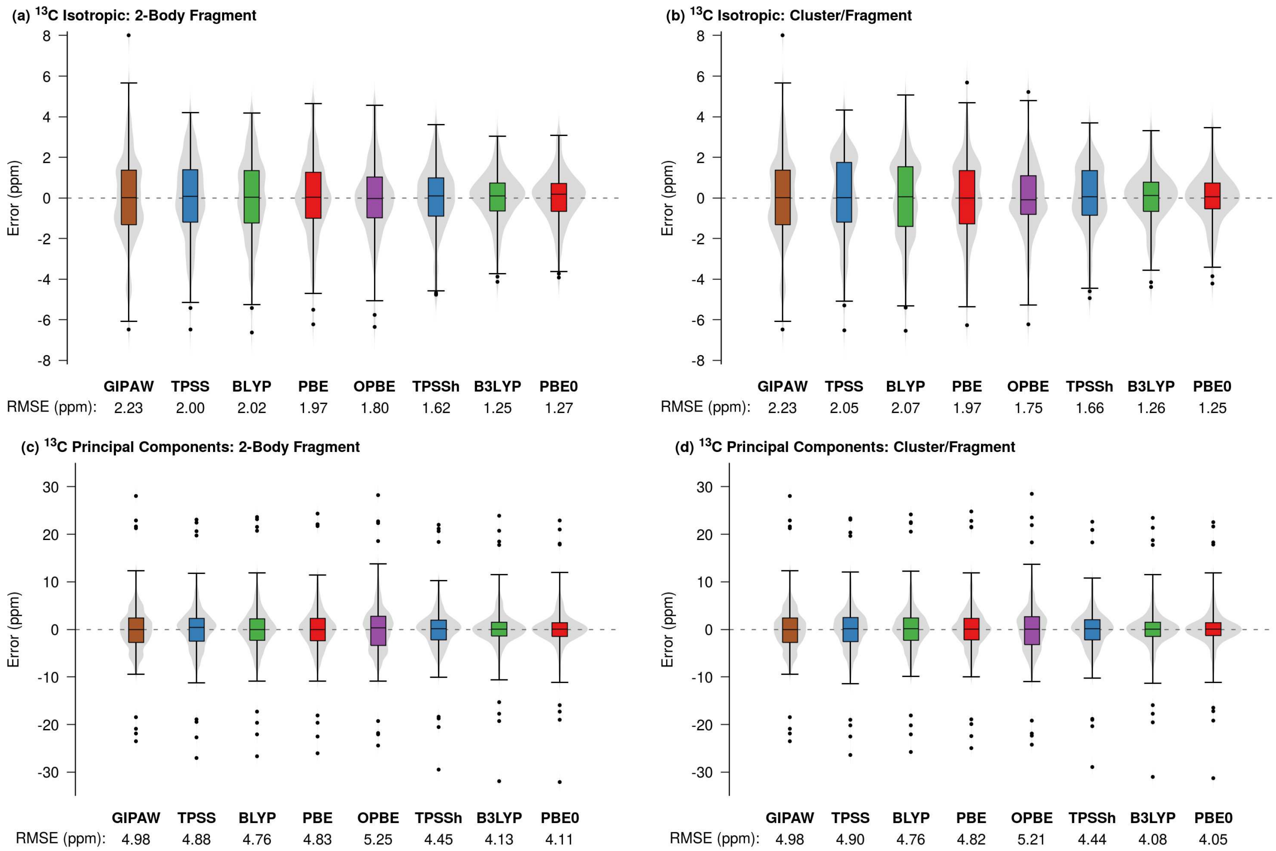
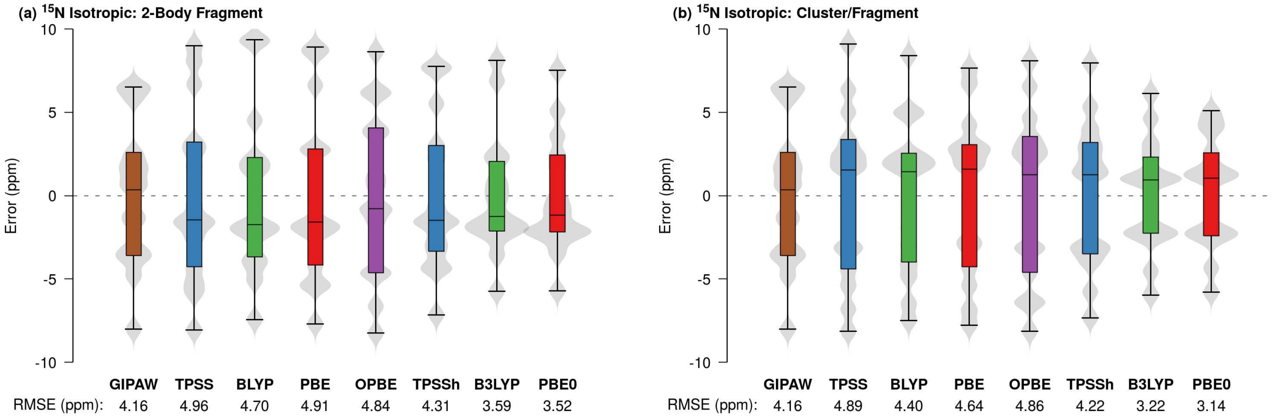
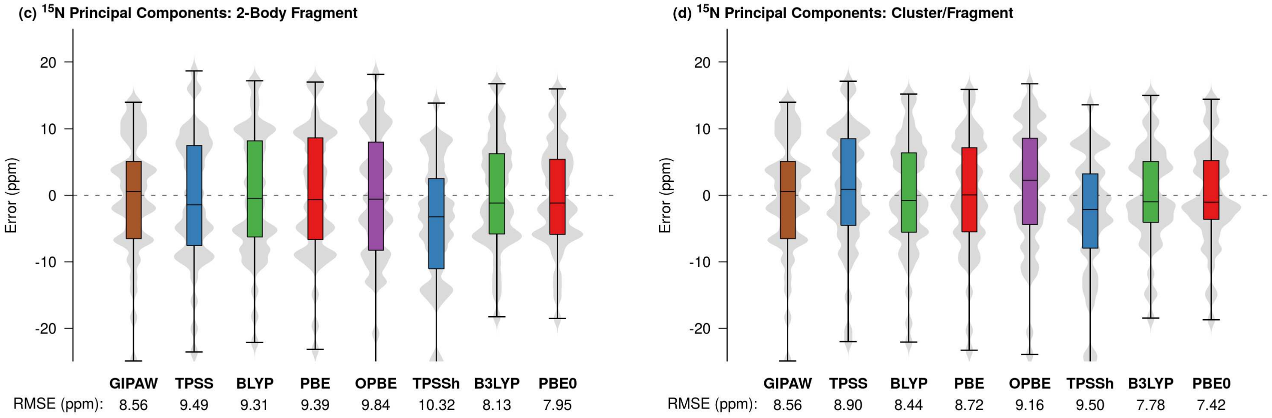


**Figure 2:** Comparison of error distributions for the $^{13}$C isotropic shifts (a and b) and principal components (c and d) across different density functionals using the fragment and cluster/fragment models. Planewave GIPAW PBE results are included for comparison.



(a) $^{15}$N Isotropic: 2-Body Fragment

(b) $^{15}$N Isotropic: Cluster/Fragment



(c) $^{15}$N Principal Components: 2-Body Fragment

(d) $^{15}$N Principal Components: Cluster/Fragment

**Figure 3:** Comparison of error distributions for the $^{15}$N isotropic shifts (a and b) and principal components (c and d).

als uniformly outperform their GGA counterparts, and both PBE0 and B3LYP outperform TPSSh. The $^{13}$C principal components are only modestly sensitive to long-range electrostatic effects: The rms errors in the principal components from the SCRMP embedding calculations reported in Figure 2c are about 0.3 ppm smaller than those previously reported with the simpler GDMA-based embedding scheme.[15] Similarly, including local many-body effects explicitly via the cluster/fragment approach has minimal impact (compare Figures 2c and 2d). Comparing to our previously reported results with the simpler GDMA electrostatic embedding, the benefits of the cluster/fragment model are even smaller here. Taken together, the $^{13}$C data presented in Figure 2 agrees with both our previous findings and with the analysis of Holmes et al.[13]

Next, consider the errors for the $^{15}$N isotropic chemical shifts and principal components in Figure 3 presents a detailed comparison of the relative performance of GIPAW, fragment and cluster/fragment models. Two trends are readily apparent. First, just as for the $^{13}$C data, the hybrid functionals (particularly PBE0 and B3LYP) out-perform the GGAs. The performances of the meta-GGA TPSS and its hybrid variant TPSSh are unexceptional, just as for the $^{13}$C results above and contrary to the aforementioned previously reported $^{15}$N data.[13]

Second, the difference between the fragment and cluster/fragment models is appreciably larger for $^{15}$N than it is for $^{13}$C (regardless of the specific density functional choice), as can be seen by comparing the left and right hand sides of Figure 3. For PBE0, for instance, the cluster/fragment model reduces the rms errors from the fragment model by ∼0.4 ppm, or about 11% in the isotropic shifts and 7% for the principal components. This reflects the greater sensitivity of $^{15}$N chemical shieldings to their chemical environments. Further evidence for this is provided by comparing GIPAW to the fragment and cluster/fragment results. The cluster/fragment PBE results are closer to the GIPAW PBE ones than the two-body fragment PBE ones. On the other hand, the reduction in error provided by switching from PBE to the hybrid PBE0 functional is sizable enough that both the fragment and cluster/fragment models exhibit errors smaller than those of GIPAW PBE.

Table 1 summarizes the linear regression parameters and rms errors for both the $^{13}$C and $^{15}$N principal components used to map the predicted CSA principal components of the chemical shielding tensor for each functional. The best-performing density functionals (PBE0 and B3LYP) have slopes which deviate from the ideal value of -1 by only a few percent. Such deviations are expected and partially compensate for systematic errors in the calculations (e.g. basis set incompleteness). Although the meta-GGA density functionals under-perform hybrid functionals in terms of the error distributions and rms errors, the linear regression parameters in Table 1 do indicate a lower degree of systematic error, in agreement with Holmes et al.[13]

Given the modest size of the nitrogen test set, caution should be exercised when interpreting the CSA regression parameters for nitrogen in Table 1. For example, including the six CSA tensor elements from glycylglycine hydrochloride monohydrate (GLCICH01) alters the rms error for fragment PBE0 calculations by ∼0.4ppm, or 5%. We do not recommend $^{15}$N CSA regression parameters be used for applications outside this particular test set.

## 5 Consistency between the principal components and isotropic shifts

Because the isotropic shielding is simply the average of the three principal components of the chemical shielding tensor, the regression parameters obtained from fitting the principal components could ideally be used to scale absolute isotropic shieldings (or vice-versa). In practice, the experimental uncertainties are generally larger for the principal components, and they are harder to predict accurately from computation as well.

Nevertheless, it is interesting to assess how consistent the isotropic and CSA data sets are.

**Table 1:** Linear regression parameters and rms errors for the principal components of the NMR shielding tensors. Results are reported for both the $^{13}$C and $^{15}$N test sets using the two-body fragment, cluster/fragment (with a 4 Å cluster), and GIPAW calculations. All cluster and fragment calculations employ the mixed basis, electrostatic embedding and a two-body fragment cutoff of 6 Å. $^{\dagger}$CSA tensor values for $^{15}$N reported for comparing relative performance of the density functionals.

| | | **Two-body Fragment** | | | **Cluster/Fragment** | | |
|---|---|---|---|---|---|---|---|
| **Atom** | **Functional** | **RMSE** | **Slope** | **Intercept** | **RMSE** | **Slope** | **Intercept** |
| $^{13}$C | OPBE | 4.11 | -1.0373 | 192.54 | 4.05 | -1.0366 | 192.39 |
| | TPSS | 4.88 | -1.0227 | 183.91 | 4.90 | -1.0211 | 183.73 |
| | BLYP | 4.76 | -0.9932 | 172.15 | 4.76 | -0.9922 | 172.02 |
| | PBE | 4.83 | -1.0022 | 178.26 | 4.82 | -1.0003 | 178.12 |
| | TPSSh | 4.45 | -1.0026 | 183.51 | 4.44 | -1.0013 | 183.36 |
| | B3LYP | 4.13 | -0.9559 | 172.69 | 4.08 | -0.9552 | 172.59 |
| | PBE0 | 4.11 | -0.9579 | 178.77 | 4.05 | -0.9569 | 178.66 |
| | GIPAW PBE | 4.98 | -0.9582 | 165.54 | (plane wave results) | | |
| $^{15}$N$^{\dagger}$ | OPBE | 9.84 | -1.0178 | 242.02 | 9.16 | -1.0120 | 242.84 |
| | TPSS | 9.49 | -1.0050 | 234.81 | 8.90 | -0.9993 | 235.53 |
| | BLYP | 9.31 | -0.9778 | 221.39 | 8.44 | -0.9706 | 222.18 |
| | PBE | 9.39 | -0.9876 | 227.45 | 8.72 | -0.9809 | 228.10 |
| | TPSSh | 10.32 | -0.9877 | 234.77 | 9.50 | -0.9826 | 235.45 |
| | B3LYP | 8.13 | -0.9451 | 222.76 | 7.43 | -0.9391 | 223.38 |
| | PBE0 | 7.95 | -0.9487 | 229.04 | 7.42 | -0.9440 | 229.58 |
| | GIPAW PBE[a] | 8.56 | -0.9438 | 217.06 | (plane wave results) | | |

[a] Raw $^{15}$N GIPAW principal component shielding values obtained from ref 3.

Figure 4 plots the raw data and regression lines obtained from fitting to either isotropic shifts or the principal components. To the eye, the $^{13}$C regression lines are virtually identical across the entire 250 ppm range. For nitrogen, the regression lines predict similar shifts over much of the chemical shift range, though the lines deviate somewhat at more extreme values (particularly near 400 ppm). Given the relatively small size of the $^{15}$N set and the presence of only a single principal component value above 325 ppm, one should perhaps not attach too much significance to these differences.

To gauge the transferability of the principal component regression lines in more detail, Figure 5 considers the different combinations of fitting and test data. For example, if one fits the $^{13}$C regression line to the isotropic shift data, how accurately do the resulting parameters line scale the predicted the principal component values? Conversely, how accurately does a regression fitted on the principal components reproduce the experimentally observed isotropic shifts?

For $^{13}$C with the fragment and cluster/fragment PBE0 models, the rms errors for the isotropic shifts and principal components effectively identical regardless of whether one fits to the isotropic or principal component data. This provides strong evidence that the systematic errors being eliminated via the regression model slope are uniform across the two data types. Interestingly, for GIPAW PBE, fitting to the principal components and testing on the isotropic data (or vice versa) leads to a slight ∼0.2–0.3 ppm increase in errors.

For $^{15}$N, all three models exhibit larger errors when the fitting set differs from the test set, with error increases of ∼0.6–0.8 ppm for GIPAW PBE and cluster/fragment PBE0, and a slightly larger ∼1-1.3 ppm for fragment PBE0. This larger change for fragment PBE0 is consistent with the other results demonstrating the importance of local many-body effects on the $^{15}$N chemical shielding and suggests that the regression models are compensating for those effects somewhat. Access to a larger data set of $^{15}$N chemical shield tensors would likely facilitate the comparison.

In an earlier study,[11] we benchmarked a larger and more chemically diverse set of $^{15}$N isotropic chemical shifts—24 crystal structures and 51 isotropic shifts. Regression parameters from that set have proved transferable to other systems outside the training set.[11,14] If one employs those regression parameters to scale the principal components in the current study, one finds rms errors of 10.5 ppm for 2-body fragment PBE0, 9.8 ppm for cluster/fragment PBE0, and 11.2 for GIPAW PBE. These errors are roughly 2 ppm larger than those obtained with the isotropic regression line derived from the smaller test set here, and ∼3 ppm larger than the results fitted directly to the principal components. Given the magnitude of the experimental uncertainties for the CSA principal components, we recommend using regression models fitted to that earlier $^{15}$N set if transferable scaling parameters are needed.

# 6 Conclusion

Overall, the results here demonstrate that the two-body fragment and cluster/fragment models are capable of accurately predicting components of the chemical shielding tensor in addition to the isotropic chemical shift. In fact, the results here closely mimic those found in earlier studies for the isotropic shift: (1) The predicted $^{13}$C and $^{15}$N chemical shielding tensor components are well-converged already with two-body cutoff distances of 4–6 Å. (2) Hybrid density functionals like PBE0 and B3LYP outperform the corresponding GGA functionals by ∼0.6–0.8 ppm for $^{13}$C and by up to 1 ppm or more for $^{15}$N in root-mean-square error. (3) The cluster/fragment approach provides negligible improvement for the $^{13}$C chemical shielding tensors, and only modest improvement for $^{15}$N ones. This result is particularly encouraging, given the appreciable additional computational cost associated with the cluster fragment approach.

Finally, the regression models used to map between predicted chemical shielding and observed chemical shift exhibit good consistency between the isotropic and CSA tensor data.

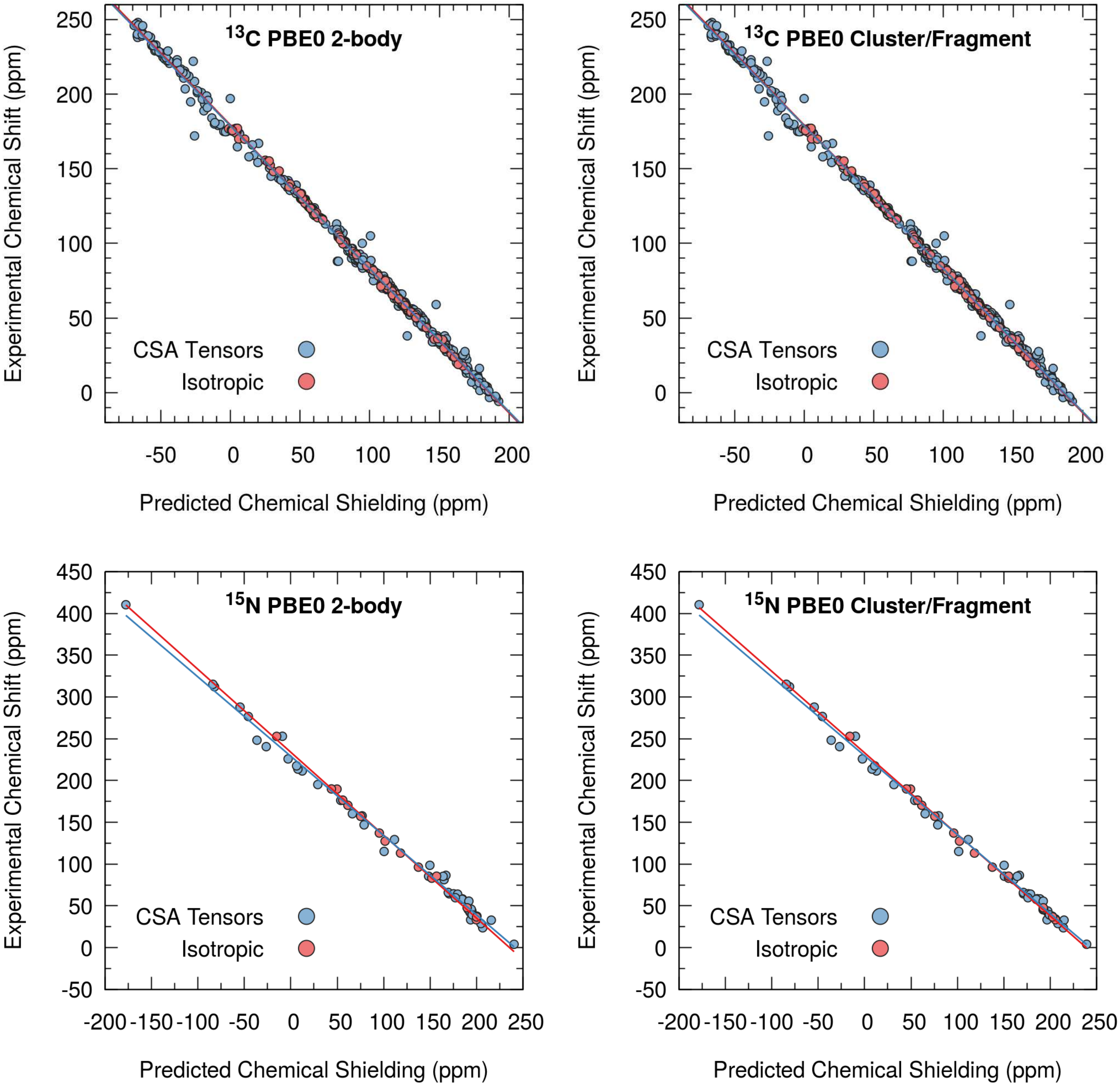


**Figure 4:** Plot of experimental vs. calculated shifts for the $^{13}$C and $^{15}$N using either fragment or cluster/fragment PBE0. The principal components and isotropic shifts are plotted on the same scale to demonstrate the level of consistency between the two data sets.

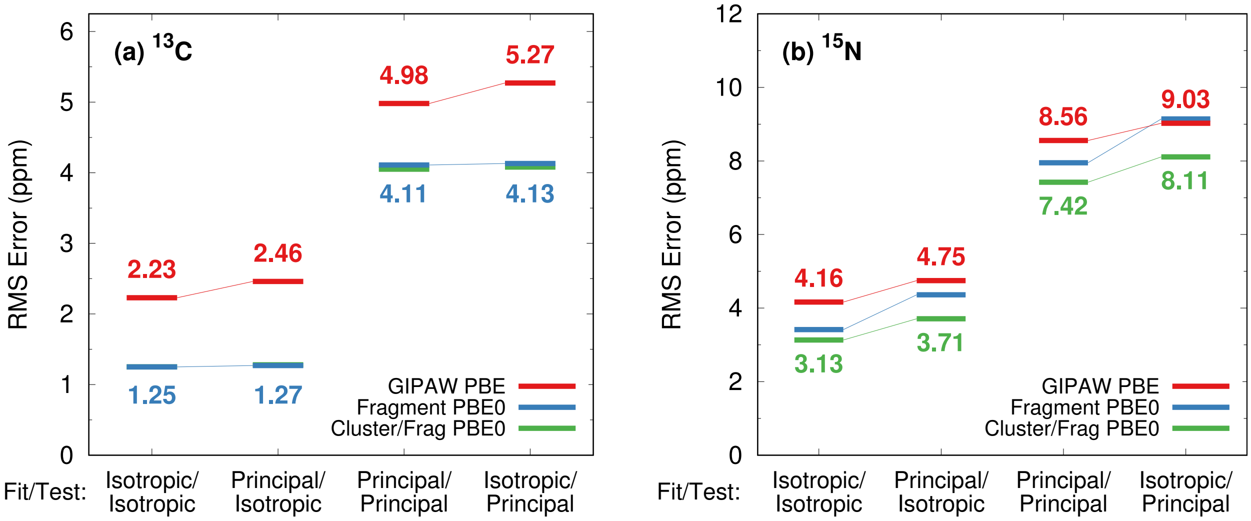


**Figure 5:** Root-mean-square errors for various models that were fit against one data set and tested on another. The notation Principal/Isotropic, for example, indicates that the regression was fitted on principal component data and tested on isotropic shifts, while Isotropic/Isotropic indicates that it was fit and tested on the same isotropic data set.

Given the generally larger uncertainties associated with measuring CSA tensor data and smaller amounts of data available, the results here suggest it is reasonable to train the regression models on suitably large isotropic chemical shift data sets and then apply those models to predicted principal components (maintaining consistent choices in the computational model chemistry). Regression parameters fitted to the large $^{13}$C data set used here have already proved transferable to a variety of systems. On the other hand, the $^{15}$N set here is probably too small to allow broader use of the regression model outside the training set. For $^{15}$N, we recommend using regression models based on our earlier, larger set of isotropic shifts.[11]

## 7 Acknowledgments

Funding for this work from the National Science Foundation (CHE-1665212) and supercomputer time from XSEDE (TG-CHE110064) are gratefully acknowledged. We thank Prof. James Harper for sharing his GIPAW data on the $^{15}$N data set and for helpful conversations.

**Supporting Information Available:** Experimental and predicted 2-body fragment PBE0 $^{13}$C and $^{15}$N experimental and 2-body fragment PBE0 shielding tensor data is provided. Optimized structures for the $^{15}$N set crystals are also provided. The $^{13}$C structures have been reported previously.[11] This material is available free of charge via the Internet at `http://pubs.acs.org/`.

## References

(1) Saitô, H.; Ando, I.; Ramamoorthy, A. *Prog. NMR Spectrosc.* **2010**, *57*, 181–228.

(2) Paramasivam, S.; Gronenborn, A. M.; Polenova, T. *Solid State Nucl. Magn. Reson.* **2018**, *92*, 1–6.

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

(4) Duncan, T. *Principal Components of Chemical Shift Tensors: A Compilation*; Farragut Press, 1997.

(5) Munoz-Castro, A.; Caimanque-Aguilar, W.; Morales-Verdejo, C. *J. Phys. Chem. A* **2017**, *121*, 2698–2703.

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

(7) Powell, J.; Kalakewich, K.; Uribe-Romo, F. J.; Harper, J. K. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12541–12549.

(8) 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.

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

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

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

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

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

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

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

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

(17) Harper, J. K.; Iuliucci, R.; Gruber, M.; Kalakewich, K. *CrystEngComm* **2013**, *15*, 8693.

(18) Orendt, A. M.; Facelli, J. C.; Bai, S.; Rai, A.; Gossett, M.; Scott, L. T.; Boerio-Goates, J.; Pugmire, R. J.; Grant, D. M. *J. Phys. Chem. A* **2000**, *104*, 149–155.

(19) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839–1862.

(20) Young, R.; Caulkins, B.; Borchardt, D.; Bulloch, D.; Larive, C.; Dunn, M.; Mueller, L. *Angew. Chem. Int. Ed.* **2016**, *55*, 1350.

(21) 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. Soc.* **2016**, *138*, 15214–15226.

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

(23) Hartman, J. D.; Balaji, A.; Beran, G. J. O. *J. Chem. Theory Comput.* **2017**, *13*, 6043–6051.

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

(25) Stueber, D. *Conc. Magn. Reson. A* **2006**, *28*, 347–368.

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

(27) Soss, S.; Flynn, P.; Iuliucci, R. *Chem. Phys. Chem.* **2017**, *18*, 2225–2232.

(28) Cernik, R. J.; A. K. Cheetham, C. K. P.; Watkin, D. J.; Wilkinson, A. P.; Willis, B. T. M. *J. Appl. Crystallogr.* **1991**, *24*, 222.

(29) Simon, L. J. W.; Lahav, M.; Leiserowitz, L. *New J. Chem.* **1986**, *10*, 723.

(30) Oda, K.; Koyama, H. *Acta Crystallogr. Sect. B: Struct. Crystallogr. Cryst. Chem.* **1972**, *28*, 639.

(31) Stone, K. H.; Lapids, S.; Stephens, P. W. *J. Appl. Crystallogr.* **2009**, *42*, 385.

(32) Portalone, G.; Bencivenni, L.; Colapietro, M.; Pieretti, A.; Ramondo, F. *Acta Chem. Scand.* **1999**, *53*, 57.

(33) Koetzle, T.; Hamilton, W.; Parthasarathy, R. *Acta Cryst. Sect. B* **1972**, *28*, 2083.

(34) 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).

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

(36) 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.

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

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

(39) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(40) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(41) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(42) Holmes, S. T.; Dybowski, C. *Solid State Nucl. Magn. Reson.* **2015**, *72*, 90–95.

(43) Ye, C.; Fu, R.; Hu, J.; Hou, L.; Ding, S. *Magn. Reson. Chem.* **1993**, *31*, 699–704.

(44) Liu, F.; Orendt, A. M.; Alderman, D. W.; Grant, D. M. *J. Am. Chem. Soc.* **1997**, *119*, 8981–8984.

(45) Stueber, D.; Grant, D. M. *J. Am. Chem. Soc.* **2002**, *124*, 10539–10551.

(46) Sherwood, M. H.; Facelli, J. C.; Alderman, D. W.; Grant, D. M. *J. Am. Chem. Soc.* **1991**, *113*, 750–753.

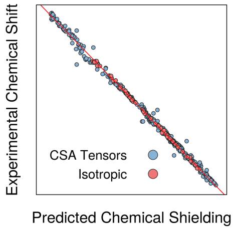
(47) Raya, B.; Bechinger, B. *Solid State Nucl. Magn. Reson.* **2014**, *15*, 61–62.

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

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

# Graphical TOC Entry

**Consistent prediction of Isotropic Shifts & CSA Tensors**



**$^{13}$C CSA Errors**
GIPAW PBE: 5.0 ppm
Fragment PBE0: 4.1 ppm

**$^{15}$N CSA Errors**
GIPAW PBE: 8.6 ppm
Fragment PBE0: 8.0 ppm