Calculating Nuclear Magnetic Resonance Chemical Shifts from Density Functional Theory: A Primer

Gregory J. O. Beran[1, a)]
Department of Chemistry, University of California, Riverside, California 92521, United States

(Dated: 24 February 2019)

Density functional theory (DFT) prediction of nuclear magnetic resonance (NMR) chemical shifts complements NMR experiments. Predicting chemical shifts accurately with DFT requires many different modeling decisions. Intended for novice modelers and non-experts, this article discusses the considerations one should take in selecting a density functional, van der Waals dispersion correction, and basis set. It examines different strategies for handling systems in complex environments such as liquids, biomolecules, and crystals. Strategies include the use of cluster models, electrostatic embedding, continuum representations, periodic boundary conditions, and fragment-based approaches. Finally, approaches for referencing the predicted absolute chemical shieldings for comparison against experimental chemical shifts are discussed.

Keywords: density functional theory, NMR crystallography, chemical shift, basis sets, cluster models, periodic boundary conditions

First-principles quantum chemical calculation of nuclear magnetic resonance (NMR) chemical shifts plays an increasingly important role in interpreting NMR spectroscopy experiments. In the context of NMR crystallography, the agreement between predicted and experimentally observed chemical shifts is used to assess and discriminate among candidate structures. Such interpretations and assessments rely on accurate chemical shift predictions. The challenge of computational chemistry arises in striving for meaningful predictions in real-world systems and assessing the reliability or uncertainties of those predictions. Making reasonable modeling decisions becomes easier with experience, but acquiring basic background knowledge can facilitate the decision-making and help avoid common pitfalls. This review surveys the major choices one must make when predicting NMR chemical shifts and provides a starting point for how to address them in an informed manner.

Before predicting NMR chemical shifts, one must choose the so-called model chemistry. Due to its extreme complexity, one never solves the molecular Schrödinger equation exactly. Instead, a number of different approximations are typically made, starting with the Born-Oppenheimer approximation and including subsequent approximations to the resulting electronic Schrödinger equation. A model chemistry represents a particular set of user-chosen approximations, including the method (e.g. a specific density functional), basis set, and numerical approximations (e.g. **k**-point sampling in periodic systems, integration grids, density fitting approximations, etc).

Next, one must also decide what structure to model. Aspects such as conformation, protonation state, and tautomers can be particularly subtle. Furthermore, the full system may be too large to model directly, in which case it becomes necessary to introduce additional approximations. One might wish to truncate the system somehow, model different parts of the system at different levels of theory, or perhaps mimic the environment via an implicit model. In complex systems with many degrees of freedom, one should also assess whether a single static structure is sufficient, or whether one needs to consider dynamical averaging of the structure.

Once the chemical shieldings have been predicted, they must be referenced against a chemical shielding standard to obtain chemical shifts that can be compared directly against experimental measurements. The following sections provide context and guidance on how one should approach each of these challenges. Selected literature citations are provided, but these are far from exhaustive.

I. ELECTRONIC STRUCTURE TREATMENT

The two most important choices in a model chemistry are the electronic structure model and basis set. The vast majority of chemical shift predictions are currently performed using density functional theory (DFT). Other models such as Hartree-Fock (HF) theory, second-order Møller-Plesset perturbation theory (MP2), and coupled cluster techniques do exist, but they either provide lower accuracy (for HF) or come with much higher computational costs (MP2 & coupled cluster models). DFT provides a pragmatic balance between accuracy and computational efficiency.

A. Selecting a density functional

Novice users of DFT are faced with a baffling array of possible density functionals to choose from. The Hohenburg-Kohn theorems prove the existence of a density functional that maps directly between the three-

a)Electronic mail: gregory.beran@ucr.edu

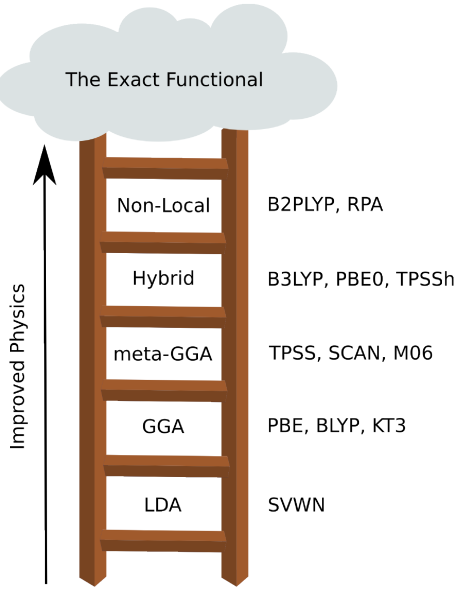

FIG. 1. Density functionals on higher rungs of Jacob's ladder incorporate improved physics and ideally make more accurate predictions. Selected functionals from each rung are listed.

dimensional electron density $\rho(\mathbf{r})$ and the energy, but they do not reveal its mathematical form. Researchers have stepped in to fill this void, producing hundreds of different density functionals over the years. Virtually all current DFT calculations employ the Kohn-Sham formulation, in which the energy is expressed as,

$$E[\rho] = T_s[\{\phi_i\}] + J_{ee}[\rho] + J_{eN}[\rho] + E_{XC}[\rho] \qquad (1)$$

where T_s represents the kinetic energy for a set of non-interacting electrons (expressed in terms of the Kohn-Sham molecular orbitals ϕ_i), the electron-electron Coulomb repulsion J_{ee}, the electron-nuclear Coulomb attraction J_{eN}, and the exchange-correlation functional E_{XC}. All Kohn-Sham density functionals employ the same non-interacting kinetic energy and the Coulombic terms are identical across functionals. They differ in the exchange-correlation functional, which describes the quantum mechanical exchange contributions to the electron-electron repulsion and how the true interacting-electron kinetic energy differs from the non-interacting one ("correlation"). In the presence of a magnetic field, the density functional should actually depend on the current-density as well as the electron density. While some on-going research seeks to develop models that account for this,[1,2] most present-day applications of DFT ignore the current-density contributions.

Density functionals are classified as rungs on "Jacob's ladder" according to the physical features they include, building toward the probably unobtainable exact functional (Figure 1). At the lowest rung is local density approximation (LDA), in which the energy is purely a functional of $\rho(\mathbf{r})$—that is, incremental contributions to the energy depend only on the density locally at position $\mathbf{r}$. LDA performs poorly for most molecular properties, and it should generally be avoided. The next rung, generalized gradient approximation (GGA) functionals, augment the dependence on $\rho(\mathbf{r})$ with contributions that depend on how the density changes in the immediate vicinity via the gradient $\nabla\rho(\mathbf{r})$. GGA functionals perform much better for molecules than LDA ones, and they are widely used both in molecular and solid-state systems. The third rung, meta-GGA functionals, augment GGAs with dependence on the kinetic energy density ($\nabla^2\rho(\mathbf{r})$). Next come hybrid density functionals, in which some "exact" exchange (a.k.a. HF exchange) is mixed into GGA functionals (or less frequently, meta-GGA ones). The amount of exact exchange mixed in is usually determined empirically, with values around 25%. Finally, the fifth rung consists of non-local functionals and other approaches that incorporate virtual orbitals into the functional, such as the random phase approximation, (RPA) and double hybrid density functionals (which augment the functional with MP2-like correlation energy contributions). Figure 1 lists a few representative functionals from each rung.

Once the energy and density are determined with the chosen density functional, the NMR chemical shielding tensor for atom A can be computed as the second derivative of the energy with respect to the magnetic field and the nuclear magnetic moment of nucleus A,

$$\sigma^A_{\alpha\beta} = \frac{\partial^2 E}{\partial B_\alpha \partial \mu^A_\beta} \qquad (2)$$

where α and β refer to the x, y, and z components of the shielding tensor. The isotropic shielding is obtained as the trace of this tensor. GGA and hybrid functionals are most common for NMR chemical shift predictions. GGAs like PBE often perform quite well for chemical shift predictions. The performance of meta-GGA functionals is more controversial,[5,11] but the TPSS meta-GGA functional does not seem to improve significantly upon GGAs. Hybrid functionals like PBE0 and B3LYP do improve appreciably upon their GGA analogs. In molecular crystal benchmarks, switching from GGA to hybrid functionals reduced the root-mean-square errors from ∼2.1 ppm to ∼1.4 ppm for ^{13}C isotropic shifts,[5] and from ∼5.5 ppm to ∼4.2 ppm for ^{15}N.[6] In NMR crystallography, where distinguishing between correct and incorrect structural assignments is essential,[12] such accuracy improvements can meaningfully improve the level of discrimination.[13]

TABLE I. Approximate accuracy of predicted chemical shifts for selected NMR-active nuclei. Representative studies are cited.

Nucleus	Typical Errors	Literature
^{1}H	∼0.2–0.4 ppm	Refs 3–5
^{13}C	∼1.5–3 ppm	Refs 3,4,6
^{15}N	∼3–6 ppm	Refs 6–8
^{17}O	∼7–10 ppm	Refs 6,9,10

Table I lists "typical" errors one can expect for NMR chemical shifts predicted with appropriate methodologies and cites a handful of selected studies upon which these estimates are based. The actual accuracy in a given system will depend on the modeling choices and the type of system.

While one can find numerous benchmark studies arguing in favor of one particular density functional or another for predicting NMR chemical shifts, molecular crystal benchmarks on half a dozen functionals performed in the author's research group found that functionals within a given rung of Jacob's ladder generally perform similarly to one another. This means, for example, that the practical differences between hybrid functionals B3LYP and PBE0 are small—e.g. a few hundredths of a ppm for ^{13}C, or a couple tenths of a ppm for ^{15}N isotropic shifts.[5,6,14]

Calculations employing hybrid density functionals typically require around 50% more computational effort than GGAs when Gaussian basis sets are used. In practice, therefore, this means that one can (and probably should) routinely use hybrid functionals for molecular predictions. For solid-state DFT calculations in a plane wave basis set, however, the computational cost of using hybrid density functionals is an order of magnitude higher than that of GGAs. For this reason, hybrid functionals are rarely used in plane wave DFT calculations. Instead, the vast majority of plane wave DFT calculations employ GGA functionals (most often PBE). Recent developments in fragment- and cluster-based techniques have demonstrated a route to capturing the benefits of hybrid functionals in many solids, as will be discussed in Section II C.

B. Accounting for van der Waals dispersion

Van der Waals dispersion is neglected in virtually all approximate density functionals, including PBE, PBE0, and B3LYP. Dispersion plays a substantial role in both intra- and intermolecular non-covalent interactions, and its omission makes non-covalent interactions artificially repulsive. For example, the molar volume of crystalline benzene is overestimated by ∼45% with B3LYP.[15] Similarly, neglecting dispersion artificially causes the Ace-Ala-Gly-Ala-NMe tetrapeptide to prefer an unfolded β-sheet-like conformation over the correct folded α-helix-like conformation.[16]

Dispersion arises from correlated, non-local electron fluctuations. In the simplest picture, a fluctuation in the electron density on one atom creates an instantaneous dipole that then induces a corresponding fluctuation in another atom, leading to an attraction. Standard local and semi-local functionals incorporate information about the electron density $\rho(\mathbf{r})$ at point $\mathbf{r}$ and perhaps its immediate vicinity (via $\nabla\rho(\mathbf{r})$), but they cannot capture the long-range correlated fluctuations needed for dispersion.

The most pragmatic solution for capturing van der Waals dispersion in DFT is to employ a post-hoc dispersion correction. Evaluating these corrections requires only a small fraction of the computational effort of the underlying DFT calculation. Whereas early dispersion corrections were highly empirical, present-generation dispersion corrections are derived largely from first-principles. Widely used corrections include Grimme's D3 dispersion correction,[17] Becke and Johnson's Exchange-Hole Dipole Moment (XDM) model,[18–20] the Tkatchenko-Scheffler (TS) pairwise correction,[21] and the Tkatchenko and Distasio many-body dispersion (MBD) model.[22,23] The D4 correction, which is a modestly revised version of D3, was reported in 2017,[24] though it has not yet entered widespread use as of this writing. See recent reviews for more details on these and other dispersion corrections.[16,25,26]

In the end, if dispersion plays any role in the system of interest, one should include a dispersion correction during the geometry optimization stage. The additional computational cost incurred by the correction is small, and it generally improves the physics of the model. While theoretical debates continue among proponents of the different models, many of them perform well. Models such as D3, XDM, or MBD all represent good options. Because these post-hoc dispersion corrections do not directly impact the electron density, they are not employed in the final chemical shielding calculations. The effects of self-consistently including dispersion in determining the electron density appear to be small.[27]

C. Choosing a Gaussian basis set

Once a density functional has been chosen, one must decide on the basis set that will be used to represent the orbitals and electronic density. Atom-centered Gaussian basis functions are typically employed in molecular and biomolecular calculations, while plane waves are more commonly used in periodic calculations on liquids and solids. For atom-centered orbitals, gauge invariance is not satisfied in finite basis sets, due to imperfect cancellation between the diamagnetic and paramagnetic terms. Several solutions have been proposed to this problem.[1] At present, the gauge-including atomic orbital (GIAO) approach, which assigns a gauge origin to each basis function, is the most frequently used.

Gaussian basis sets include several different types of basis functions. A minimal basis set contains one atomic orbital (AO) per valence orbital present according to the periodic table. Minimal basis sets are too small and unreliable, however. Double, triple, or quadruple-ζ basis sets increase the flexibility of the basis by including two, three, or four basis functions for each valence orbital. These extra sets of basis functions allow the electron density around an atom to expand or contract, for example. Polarization functions, or basis functions with higher angular momentum (e.g. d orbitals on carbon or p orbitals on hydrogen), are added to allow the electron density to shift off-center from the nucleus, as is required when it

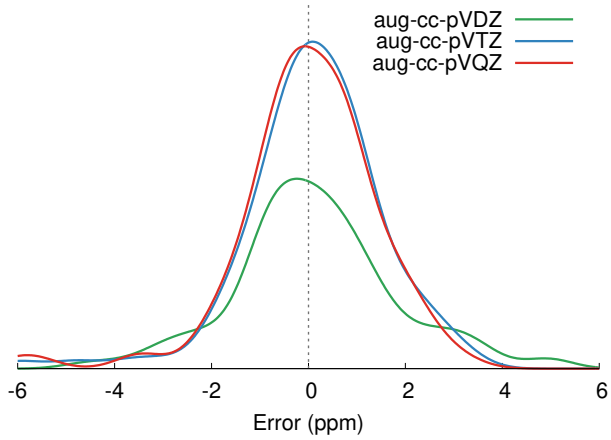

FIG. 2. Distribution of errors between theory and experiment for 110 ^{13}C isotropic chemical shifts from 21 molecular crystals as a function of basis set. The smallest aug-cc-pVDZ basis set gives an rms error of 1.69 ppm, which is reduced to 1.42 ppm for the larger aug-cc-pVTZ and aug-cc-pVQZ basis sets. Shifts were computed with PBE0 and the 2-body fragment approach of Ref 14.

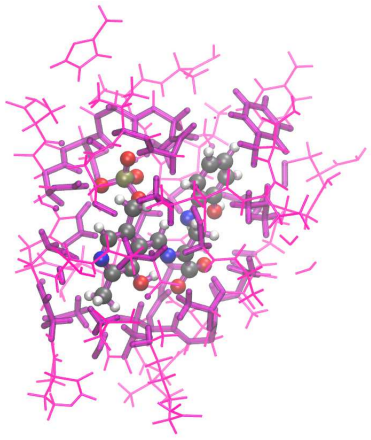

FIG. 3. Example of locally dense basis sets for a cluster of atoms consisting of a substrate plus 7 Å of surrounding protein environment from the β-subunit of tryptophan synthase. The substrate atoms (ball-and-stick model) of interest are represented with the large 6-311++G(d,p) basis, nearby atoms out to 4 Å (stick representation) are treated with the medium 6-311G(d,p) basis set, and the more distant atoms (wireframe) are modeled with the crude 6-31G basis.[32]

is polarized between atoms with different electronegativities. Diffuse functions are very large basis functions designed to, for example, saturate the space between atoms for non-covalent interactions or to allow the substantial expansion of the electron cloud that occurs in anions. Finally, magnetic properties such as nuclear quadrupole coupling constants can be very sensitive to the treatment of the wavefunction near the nucleus, which can make the addition of unusually tight core basis functions important.[28] While most basis sets ignore these tight basis functions, the pcSseg-n basis sets have been specifically designed for computing such properties.[29] The cost of a DFT calculation grows asymptotically as $O(N^4)$ with the number of basis functions. However, not all basis functions are created equal in determining computational cost. Modern electronic structure codes make extensive use of integral estimation and screening to avoid computing integrals that contribute negligibly. Tighter basis functions have far fewer significant interactions with other orbitals than do diffuse functions, and adding more tight basis functions will have a smaller impact on the computational cost than would adding the same number of diffuse functions.

Generally speaking, one should not use basis sets smaller than double-ζ sets with polarization like 6-31G(d), def2-SVP, or pcSseg-1 for DFT. Results obtained with triple-ζ basis sets such as 6-311+G(2d,p), def2-TZVP, or pcSseg-2 are often appreciably improved over double-ζ ones. The gains achieved by increasing to a quadruple-ζ basis set like def2-QZVP or pcSseg-3 are often smaller and may not be worth the additional computational cost. Geometry optimizations frequently require more computer time than NMR chemical shielding calculations. Because geometry is generally less sensitive to basis set, optimizations in double-ζ basis sets are common. Absolute chemical shieldings converge rather slowly with basis set, but chemical shifts relative to a reference compound converge much faster due to consistent error cancellation. In a set of 21 molecular crystals and 110 ^{13}C isotropic shifts, for example, the accuracy of the predicted chemical shifts improves appreciably upon increasing the basis set from double-ζ to triple-ζ (Figure 2). No significant further improvement occurs for the even larger quadruple-ζ basis set.

Locally dense basis sets[30,31] provide an effective strategy for lowering the computational cost of Gaussian basis NMR chemical shielding calculations. Chemical shielding is a highly local phenomenon, and one can often use much smaller basis sets on atoms that are far away from the atom(s) of interest. Specifically, a large basis should be used on the atoms of interest, but only a medium-sized basis is needed on atoms within a few Å, and a small basis can be used for more distant atoms. Physically, this implies that one needs to model the local electronic structure accurately (out to about the size of the van der Waals cavity surrounding the molecule of interest), but that even a relatively simple quantum mechanical representation of the electrostatic effects is sufficient to capture longer-range effects outside the regime where wavefunctions overlap significantly. For slightly longer distances (e.g. beyond ∼6 Å), one can potentially even switch to a point-charge representation.[6,14]

For example, in an investigation of a substrate in an enzyme active site, the combination of a large 6-311++G(d,p) basis for the substrate atoms, 6-311G(d,p) on atoms within 4 Å of the substrate, and 6-31G on more distant atoms (Figure 3) reproduced the carbon atom

shieldings on the substrate to within root-mean-square error of 0.16 ppm compared to the full calculation in the largest basis set.[32] Furthermore, the errors become negligible[14] if a linear regression referencing model is applied. This locally dense basis combination reduces the number of basis functions in this particular example from 19,638 if all atoms employed 6-311++G(d,p) to 5,111 with the locally dense basis set. The same combination, albeit with the slightly faster 6-311+G(2df,p) large basis has also proved effective.[14] Similar analysis has been demonstrated for molecules using the pcS-n basis sets.[33]

Perhaps the biggest practical basis set challenge facing the user lies in determining whether the chosen basis set is adequate and for assessing the degree of uncertainty in the prediction associated with basis set incompleteness. Good agreement between theory and experiment is a poor criterion for assessing the basis set. One commonly finds an intermediately sized basis set which fortuitously predicts some property in good agreement with experiment. However, other properties or species will not necessarily exhibit the same fortuitous error cancellation. A much better approach is to employ empirical sensitivity testing—performing a series of calculations in different basis sets to ascertain: How sensitive is the prediction/interpretation to the chosen basis set? Have the predictions converged sufficiently with respect to basis set size? Using systematically growing basis sets (like the pcSseg-n, cc-pVXZ, or def2-type sets) facilitates such analysis.

D. Plane wave basis sets for periodic systems

When modeling liquids and solids, it is common to represent the system via a plane wave basis set under periodic boundary conditions. Plane waves are well-suited to densely packed condensed-phase systems, and integrals involving plane wave basis functions are inexpensive to evaluate. Plane waves also have the advantage of simplicity in that the basis set size is determined by a single parameter, the plane wave cutoff. The basis set includes all plane waves with periodicity consistent with the unit cell and which have kinetic energies less than the plane wave cutoff. A larger plane wave cutoff (or larger unit cell) increases the size of the basis set. Because they are not atom-centered, the number of plane waves present in the basis set is independent of the actual atoms present in the unit cell.

Evaluating the necessary integrals in a plane wave basis is computationally inexpensive compared to the corresponding integrals involving Gaussian orbitals. On the other hand, plane wave basis sets are typically ∼1–2 orders of magnitude larger than Gaussian basis sets. Describing the sharp changes in the density about the core electrons with plane waves is particularly difficult and would require even larger basis sets. Instead, the core electrons are generally replaced with pseudopotentials that mimic their effects on the valence electrons. Traditionally this was done with norm-conserving pseudopotentials, but in more recent years ultrasoft pseudopotentials have become more widely used because their "softer" shapes allow the use of smaller basis sets and make the calculations more efficient overall.

Pseudopotentials create a challenge for NMR, since many NMR properties depend critically on the behavior of the electron density near the nucleus. The projector augmented wave (PAW) approach reconstructs the all-electron density from pseudopotential schemes. However, this reintroduces a gauge dependence to the model. The gauge-including PAW (GIPAW)[34,35] approach addresses the gauge dependence similarly to how the GIAO model does for Gaussian basis sets.

On a final note, performing periodic DFT calculations requires solving the Kohn-Sham equations at multiple reciprocal space **k**-points throughout the Brillouin zone. A smaller unit cell dimension in real (direct) space requires a larger number of **k**-points along the corresponding axis in reciprocal space. Typically the **k**-points are positioned according to a Monkhorst-Pack grid, and the user has to determine only how many **k**-points to place along each axis (or alternatively, what **k**-point mesh density to use). Best practice is to increase the **k**-point sampling until the desired level of convergence in energy or chemical shielding/shift is achieved.

II. MODELING THE CHEMICAL SYSTEM AND ITS ENVIRONMENT

A. Representing the system

After selecting the level of theory to use, one must decide how to represent the system. In many-cases, the full system will be comprised of a region of interest and surrounded by an environment. For ease of discussion here, "system" refers to the atoms of interest, and "environment" refers to all remaining atoms/molecules. The system and environment may correspond to different molecules, as is the case for a molecule in solution. In an enzyme, the system might represent a substrate and the active site residues, while the environment represents the rest of the protein and the surrounding solvent. Such cases can generally be represented via some sort of cluster model which includes the system plus a portion of the environment.

In other systems, it may make more sense to represent a system as a relatively small unit cell with periodic boundary conditions instead of as a large finite cluster. Molecular crystals, solids, and solutions can all be described readily with periodic boundary conditions. The GIPAW method enables NMR chemical shielding calculations in periodic systems with plane wave DFT and has transformed NMR crystallography.[36]

For biomolecules, one may obtain initial atomic positions from a known crystal structure. For molecules, structures may be constructed from either chemical in-

tuition or crystallographic data. Either way, one should generally perform a geometry optimization to relax the atomic positions. In the case of a cluster extracted from a larger experimental structure (e.g. part of a protein), one might freeze the positions of atoms on the outer edge of the cluster to retain structural features that would otherwise be imposed by the atoms being ignored.

In flexible molecules, the optimal conformation may not be obvious. One may need to explore the conformational energy landscape via a series of geometry optimizations from different initial guess structures to identify the most stable one(s). Care should be taken in that the optimal conformation of the system may vary depending on the environment. Flexible molecules that "fold" and/or adopt intramolecular hydrogen bonds in the gas phase may elongate or form intermolecular hydrogen bonds in condensed phases. Finally, in some cases, the protonation state of ionizable sites or tautomer may not be obvious *a priori*. Assessment of the energetic stability and comparison between predicted and observed NMR chemical shifts can help resolve such ambiguities.

As the number of flexible degrees of freedom increases, it becomes increasingly likely that a system will not be well-represented by a single static structure. Instead, dynamical averaging over many structure snapshots from a molecular dynamics (MD) simulation may be needed to achieve good agreement between theory and experimental NMR parameters.[37–39] Dynamical averaging generally becomes more important as the number of degrees of freedom in the system increases, especially for "softer" motions like low-frequency bond torsions or solvent dynamics. Dynamical averaging proved key to achieving atomic-level understanding of HIV-1 Capsid, for example.[40] Unfortunately, chemical shifts converge relatively slowly with the number of snapshots, with dozens of snapshots being needed in small molecules, and perhaps hundreds or thousands in large biomolecules[37,38] The use of classical force fields to generate the snapshots greatly reduces the computational cost of the molecular dynamics, but classical force fields may or may not reproduce the correct ensemble of structures compared to *ab initio* MD.[39] For hydrogen bonds, quantum mechanical treatment of the hydrogen nucleus via path integral MD simulations can significantly alter the hydrogen probability distribution and therefore its NMR parameters.[41]

Modeling well-ordered solids with periodic boundary conditions can readily be done using the crystallographic unit cell, assuming the structure is known experimentally. The biggest question is which degrees of freedom to relax via geometry optimization prior to computing the chemical shifts. Relaxation of the experimental crystallographic atom positions typically improves agreement between the theoretical and experimental chemical shifts.[36] This has been found both hydrogen and heavier atoms. The decision of whether or not to optimize the unit cell parameters is more subtle. A fully optimized unit cell with an accurate electronic structure model will typically underestimate the molar volume due to the neglect of zero-point vibrational energy and thermal expansion.[42,43] Thermal expansion occurs more readily in molecular crystals with relatively weak intermolecular interactions along certain directions (e.g. van der Waals dispersion instead of hydrogen bonding). If available, constraining the lattice parameters at their room-temperature experimental values provides a simple means of maintaining a crystal structure close to the room-temperature one. On the other hand, if experimental structural data is limited or of questionable quality, it may be better to optimize the cell.

If the structure is not known experimentally, one can potentially combine crystal structure prediction together with NMR chemical shift prediction to solve the structure.[12,44] Random structure generation followed by DFT relaxation has proved an effective combination for smaller unit cells comprised of fairly rigid molecules.[45] Crystal structure prediction becomes increasingly challenging as the number of flexible degrees of freedom or the number of molecules in the asymmetric unit increases, though many advances have been made in recent years.[46–49]

Disordered materials are considerably more challenging to model than ordered ones.[50] Any defects, substitutions, or other forms of disorder introduced into the unit cell will be replicated with a periodicity/defect density determined by the unit cell dimensions. The defect density can be reduced via use of a supercell comprised of multiple crystallographic unit cells, but the computational costs increase rapidly with cell size. Constructing a good single model of the disordered solid that is tractable with DFT is difficult. A better approach can be to consider a set of many different representative structures. Certain disorders, such as positional disorder, may be representable with a fairly small number of structures. For more complex cases, one might consider the distribution of predicted chemical shifts obtained over a large number of randomly generated and DFT-optimized structures.[50] Such an approach was used to help interpret the spectrum of the mineral wadsleyite, for instance.[51]

B. Addressing the size challenge in non-periodic systems

Handling large system sizes presents one of the major challenges of *ab initio* chemical shift predictions. The cost of DFT calculations grows steeply with system size. While linear-scaling DFT algorithms have been developed,[52] the onset of linear scaling only occurs for very large systems and/or small basis sets. Cubic or quartic scaling with system size are observed more typically. As a result, biomolecules are generally too large to model quantum mechanically in their entirety. The environment often plays a significant role in solution- and solid-phase systems as well, which complicates the modeling. One generally has several options for modeling such complex systems:

1. Neglect part or all of the environment entirely.

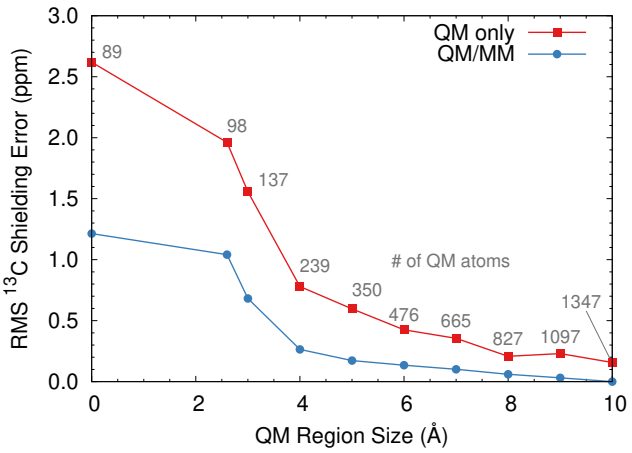

FIG. 4. Errors in the predicted aminopyrazole ^{13}C chemical shieldings in explicit water solvent. The electrostatically embedded QM/MM chemical shielding simulations converge more with size of the QM region compared to purely QM calculations. Figure adapted from Ref 54.

2. Model the environment at lower level of theory than the system of interest, either using explicit or implicit treatments.

3. Employ periodic boundary conditions to represent a large system/environment with a smaller unit cell.

4. Partition the entire system into many smaller systems via a fragment approach.

Neglecting part or all the environment is the most straightforward strategy, and it can be sufficient in simple cases (e.g. perhaps for a small molecule in solution with weak solute-solvent interactions and little ambiguity in the chemical shift assignment). The impact of neglecting the environment can be lessened by treating a large enough group of atoms/molecules that the portion of the environment that is being neglected does not strongly interact with the system. In an enzyme where one is interested primarily in the active site, for example, one might include several Å of the surrounding protein in the calculations. Unfortunately, the size of the system that needs to be included is quite large—often a 6+ Å radius,[53,54] which can translate 500 atoms or more (Figure 4). Such calculations can be made more tractable via the locally dense basis set approximations described in Section I C. The symmetry adapted clusters approach for modeling molecular crystal chemical shifts represents another example of this.[11,55] Computational savings can also be obtained by predicting the NMR chemical shifts only for selected nuclei,[56] though such algorithms are not widely available in standard quantum chemistry software packages.

Alternatively, one might represent part or all of the environment at a lower level of theory, either explicitly or implicitly. In an explicit representation, one might employ hybrid quantum mechanical/molecular mechanics (QM/MM) representations, where for example the enzyme active site and surrounding atoms are modeled with DFT, while more distant atoms are approximated classically. A simple version of this would be electrostatic embedding via atom-centered point charges for the environment. Early models such as the embedded ion method (EIM)[57] and the surface charge representation of the electrostatic embedding potential (SCREEP)[58,59] can be considered examples of this approach. The farther away the point charges lie from the atoms of interest, the better. Such hybrid QM/MM approaches can reduce the radius of the system that needs to be treated with QM from 6+ Å with no MM embedding down to 4–6 Å (Figure 4).[54] In practice, that decrease in QM cluster size can potentially reduce the number of QM atoms in half. Even greater reductions in the size of the QM region may be possible via the use of polarizable MM environments.[60]

Instead of modeling the environment explicitly, one can model it implicitly using a polarizable continuum model (PCM) or related model.[61] A typical PCM embeds the calculation within a cavity in a polarizable dielectric. Because they omit all local structuring of the environment, polarizable continuum environment approximations work best when the coupling between the system and environment is weak, such as when there are no direct hydrogen bonds or other strong specific interactions between the system and environment.

One notable difference between different PCM models lies in how the size and shape of the solute cavity is chosen. The spherical cavities employed in early models can be problematic for molecules whose shapes are far from spherical. More recent models employ cavities constructed from the union of spheres derived from each atom's van der Waals radius or by rolling a virtual probe around the molecule to determine the solvent accessible surface. Generally the user need only select the dielectric of the desired solvent. PCMs can also potentially be used to represent more complex environments, such as a protein environment. Defining an appropriate dielectric for those environments can be challenging, however.[62–65]

Overall, implicit and explicit models each have advantages and disadvantages. Explicit models capture key, specific interactions (e.g. hydrogen bonding) that are missed by implicit models. On the other hand, implicit models can be simpler to use (since they don't require atomic-level structural knowledge of the environment being approximated), and they provide a natural means for reaching the bulk limit. The reduction in degrees of freedom in the PCM models can also reduce the need for configurational sampling/dynamical averaging. Note that one need not pick explicit or implicit models exclusively. It can be very productive to model local interactions such as key hydrogen bonding partners or other nearby atoms explicitly, and then surround those models with an implicit environment to capture the longer-range contributions. Additional discussion of the merits of different strategies and examples can be found in Ref 66.

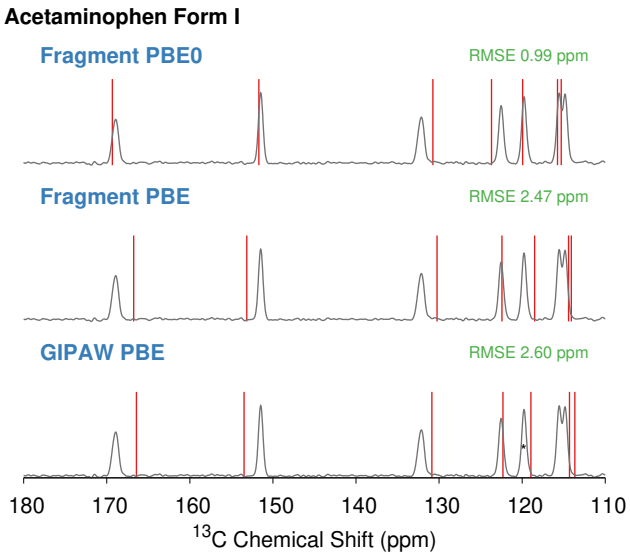

FIG. 5. Comparison of predicted ^{13}C isotropic chemical shifts (sticks) against experiment for crystalline form I acetaminophen using 2-body fragment PBE0, fragment PBE, and GIPAW PBE.[13] The fragment and GIPAW PBE models perform similarly, but the fragment PBE0 shifts are significantly more accurate.

In all of these approaches, care should be taken at the interface/edge regions, especially when covalent bonds must be cut. Any "dangling bonds" at those cuts must be handled carefully. First, not all cut points are equally good. One should ideally make cuts on single bonds with atoms of similar electronegativity (e.g. C-C bonds) well away from the region of interest. Second, they must be terminated with hydrogen atoms, link atoms, or other schemes.[67] Other subtleties can arise when electrostatic embedding is employed. For example, the placement of an embedding charge from a deleted atom very close to the hydrogen or link atom used to saturate the bond can lead to problems. Strategies for addressing these challenges in QM/MM simulations are discussed in Ref 67 and elsewhere.

C. Reducing computational costs via fragment techniques

Fragmentation provides a different strategy for reducing the computational costs in large systems. Instead of performing a single chemical shielding calculation on the entire system, one performs many smaller chemical shielding calculations on portions of the system. These individual fragment calculations are then combined in some fashion to obtain results for the whole system.

Fragmentation offers a number of advantages. Partitioning the system into many smaller fragments generally decreases the overall computational costs. It may also be possible to perform more accurate chemical shielding calculations on the fragments. Furthermore, the individual fragment calculations that contribute to the whole can typically be run independently and in parallel, making them amenable to current distributed computer architectures. Finally, the same fragment model can be applied to both periodic and non-periodic systems. As discussed in Section III below, well-ordered and characterized molecular crystals provide a nice data set for developing chemical shift referencing models that can then be applied to more complex systems such as biomolecules. Of course, fragment methods also have the downside of approximating the treatment of the interactions between fragments. This makes them most suitable for systems with localized electronic structure (e.g. organic species, biomolecules, insulators). Capturing the important electronic structure of a metal or semi-conductor with a fragment approach would be difficult. Furthermore, partitioning non-covalent interactions between molecules is easier than fragmenting across a covalent bond within a molecule.

Quite a few different successful fragment approaches have been developed. The specific formulation often depends on the type of systems being targeted. In molecular crystals, for example, an electrostatically embedded fragment model has proved quite effective.[5,13,14,68] This model computes the shielding for a key monomer of interest and then adds a series of corrections to it based on pairwise interactions between it and nearby molecules (e.g. out to 6 Å). To approximate overall polarization effects from the crystalline lattice, these monomer and dimer calculations are embedded in a set of self-consistently polarized charges designed to reproduce the Madelung potential,[6] similar to the embedding used in the embedded ion model.[57] Even better treatment of the lattice polarization effects can be obtained by a model that combines a cluster calculation with a series of longer-range pairwise fragment corrections.[5,6] Fragment methods and cluster representations[11,55] of periodic systems have much in common.

Because it employs Gaussian basis functions, this electrostatically embedded fragment approach allows routine application of hybrid density functionals instead of GGAs to molecular crystals. Compared to either GIPAW or fragment calculations with the PBE functional, switching to PBE0 via the fragment approach reduces ^{13}C chemical shift errors in organic crystals by about a third,[5] from ∼2.1 ppm to ∼1.4 ppm. Figure 5 compares fragment and GIPAW chemical shift calculations for form I crystalline acetaminophen (paracetamol).[13] In principle, fragment methods can be used to perform chemical shift calculations at very high levels of theory (e.g. coupled cluster techniques[69]), but this has not been widely explored in practice. One challenge is that properties predicted from correlated wave function methods generally converge more slowly with basis set than do ones from DFT. So not only are the models more expensive for a given basis set, but they require larger basis sets as well.

A number of fragment methods have been designed to

address the challenge of cutting across covalent bonds as well. One strategy is to perform a series of cluster calculations, where shieldings typically are taken only from the central atoms in the cluster. In the automated fragmentation QM/MM approach in proteins,[70–72] for example, the system is divided into non-overlapping residues, called core regions. To compute the shieldings on each residue, these core regions are surrounded by a buffer region of nearby residues and embedded with point charges corresponding to the remainder of the protein. This larger embedded cluster provides a more realistic environment for the evaluation of the chemical shieldings in the core region. In the end, this amounts to performing a series of overlapping embedded cluster calculations to compute the shifts throughout the system, keeping only the shifts obtained from the core of each cluster. Additional ways of combining the fragments exist as well.[73–75]

Other approaches employ quantum mechanical embedding via the fragment molecular orbital method[76,77] or adjustable density matrix assembly.[78,79] A third strategy is to perform low-level calculations on a large system together with higher-level ones on small subsystems, as in the molecules-in-molecules approach.[80] Capping atoms are frequently required to terminate dangling bonds at the cut locations, and cuts should be made at appropriate bond types and sufficiently far from the atoms whose shieldings are of interest in a given fragment.

III. REFERENCING PREDICTED CHEMICAL SHIELDINGS

A. Approaches for chemical shift referencing

Quantum chemistry calculations predict absolute chemical shieldings σ_i, while NMR experiments measure chemical shifts δ_i. The chemical shift is the difference between the chemical shielding of the target nucleus σ_i and that of a reference nucleus σ_{ref}. To compare with experiment, the predicted chemical shieldings must be converted to chemical shifts via appropriate referencing,

$$\delta_i = \sigma_{ref} - \sigma_i \tag{3}$$

There are several ways to perform this referencing. The most straightforward would be to compute the chemical shielding of the reference compound at the exact same model chemistry as the system of interest. Common chemical shift referencing scales include neat tetramethylsilane (TMS) for ^{1}H and ^{13}C, solid ammonium chloride or liquid ammonia for ^{15}N, and liquid water for ^{17}O. Unfortunately, computing the chemical shielding of the reference compound is not always straightforward. For example, properly modeling liquid ammonia would require computing the ensemble-averaged chemical shielding over many liquid-state configurations in a fairly large, periodic simulation cell.

The linear regression referencing approach circumvents these difficulties. In this approach, Eq 3 is re-written as,

$$\delta_i = A + B\sigma_i \tag{4}$$

where A and B are empirical parameters. One then performs linear regression between a set of predicted shieldings and experimentally measured chemical shifts (Figure 6). Ideally, intercept A would correspond to the chemical shielding of the reference nucleus, and slope B would equal -1, thereby reducing Eq 4 back to Eq 3. In fact, some authors constrain $B = -1$ and fit only the intercept A. Allowing B to vary partially compensates for systematic errors in the computational model chemistry (e.g. incomplete basis sets or density functional errors). In well-behaved model chemistries, the deviations in B should typically be no more than ±5–10%, and ideally less. Large deviations from unity are indicative of problems in the modeling or the experimental chemical shift assignment.

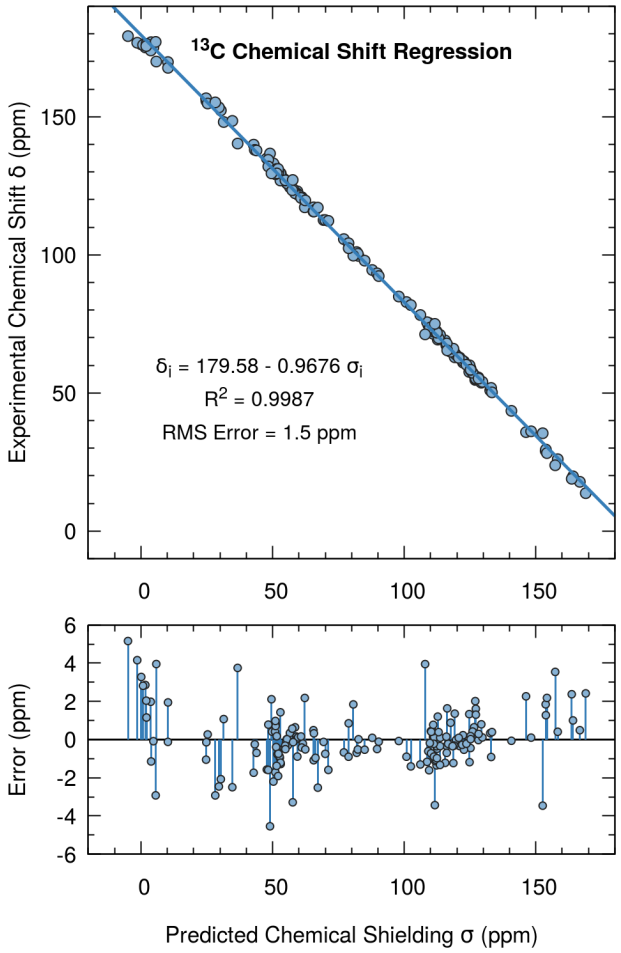

FIG. 6. Top: Sample linear regression of 169 predicted ^{13}C isotropic chemical shieldings and experimental chemical shifts.[5] Bottom: Errors in the predicted shifts resulting from the linear regression model.

B. Choosing a referencing model

What makes a good set of data for training such regression lines? First, the structures of the training set species should be well-characterized, and the experimental shifts need to be measured accurately and assigned correctly. Second, to ensure a useful regression line, the chemical shifts used in training the regression should exhibit a balanced distribution over an appropriately broad range of ppm that allows the referencing model to be interpolated, rather than extrapolated. In practice, this typically means including shifts that span a wide range of functional groups/molecules.

Third, the best regression models will be trained on data from well-characterized systems that is entirely distinct from the "unknown" systems for which new predictions are being made. Though one frequently finds studies in the literature where the regression is line is fitted directly on the system of interest, this reduces the discriminating power of the model. Least-squared minimization inherently minimizes the discrepancies between theory and experiment for any candidate structure, potentially leading to artificial agreement between theory and experiment if performed on the unknown system. Employing a regression line trained externally provides greater discrimination and can help establish higher confidence in the NMR-based structural assignment. For example, the use of an external referencing model revealed a systematic error between the predicted and measured ^{15}N isotropic chemical shifts for the three polymorphs of acetaminophen (paracetamol) that was eventually traced to an error in the original experimental referencing.[13]

The use of a large external training set for the regression model also provides information about the distribution of errors between theory and experiment for a given model chemistry in known systems, which is useful when assessing the level of agreement between theory and experiment in an unknown system. For example, an NMR crystallography study which solved the structure 9-*tert*-butyl anthracene ester photodimer crystals considered eight possible structures that were consistent with the powder x-ray diffraction data.[81] As shown in Figure 7, six of those structures exhibited very similar root-mean-square (rms) errors in the ^{13}C isotropic chemical shifts, ranging from 1.7–2.1 ppm. The other two structures exhibited appreciably larger errors in the 3.1–3.3 ppm range. The referencing model was trained on 25 molecular crystals and 169 chemical shifts, for which the predictions reproduced experiment with a ∼1.5 ppm rms error. Statistical χ^2 testing indicated that while the NMR calculations could not meaningfully discriminate among the six-lowest rms error candidates, those six structures were more likely candidates than the larger error *Pcc*2 and $P2_1ca$ ones. Further analysis revealed high similarity in the crystal packing of the six lowest structures, which can likely interconvert dynamically at room temperature. In other words, the true structure is likely an ensemble average over all six. On the other hand, the other two

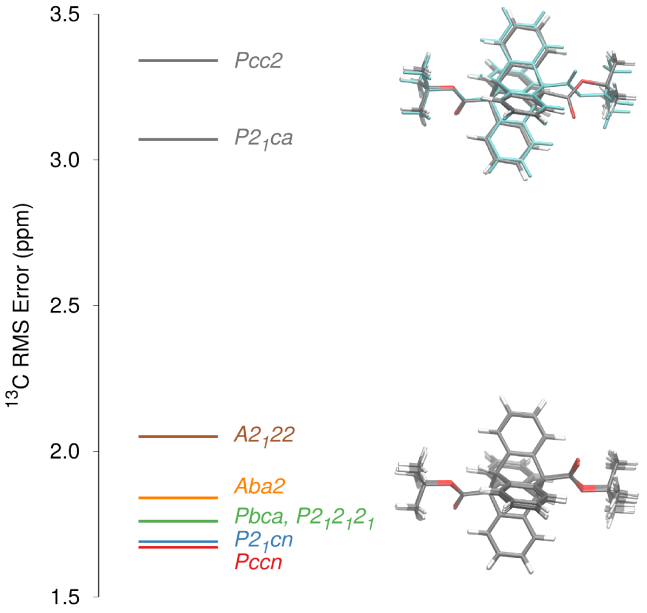

FIG. 7. Root mean square errors for eight candidate structures of the 9-*tert*-butyl anthracene ester photodimer crystals.[81] The true structure is believed to be a dynamic average over the six low-error structures whose crystal packings are very similar (as suggested by the lower molecular overlay). The *Pcc*2 and $P2_1ca$ crystal structures differ more appreciably, including having a different intramolecular conformation. The upper structure overlay compares those structures in gray against the *Pccn* structure in blue.

structures exhibit considerably different intramolecular conformations and crystal packing. Subsequent analysis of the polymorph energetics supports the structural assignment.[82]

Well-devised regression models can be quite transferable. For example, regression lines fitted to molecular crystal data have proved effective in solid-state NMR crystallographic studies of enzymes.[83,84] Unlike enzymes, molecular crystals have well-known structures with long-range order and exhibit relatively little dynamics, making them an excellent source of training data for how chemical shifts vary with the local structure and the explicit surrounding environment. A number of regression models trained on solid state NMR in molecular crystals can be found in the literature.[5,11,55,68] Many other general-purpose regressions have been produced for solution-phase data, such as those in the Chemical Shift Repository (`http://cheshirenmr.info/`)[4] and elsewhere.[3]

C. Further considerations for chemical shift referencing

When adopting a specific set of linear regression parameters from the literature, several aspects need to be considered. First, the user must compute the shieldings of the compounds of interest with exactly the same model chemistry as that upon which the regression model was

trained. This includes factors such as the density functional, basis set, exchange-correlation integration grid, solvation model, etc. Predicted chemical shieldings are rarely converged with respect to model parameters in absolute terms. Instead, the accuracy of the chemical shift predictions relies heavily on the cancellation of finite basis set error and other approximations between the target and reference nuclei.

Second, the user should note that two different conventions for reporting the regression parameters can be found in the literature. Some authors use the convention in Eq 4, in which the predicted shielding is the independent variable and the experimental shift is the dependent one.[5] However, other authors swap the dependent and independent variables, $\sigma_i = A' + B'\delta_i$, in which case the chemical shift is computed as:

$$\delta_i = \frac{A' - \sigma_i}{-B'} \tag{5}$$

This latter convention is used on the Chemical Shift Repository website, for example. Both conventions are equally valid, and the parameter values can be interconverted trivially.

Finally, while plots such as Figure 6 are visually appealing, they should be interpreted with some caution. Because the chemical shift range is ∼1–2 orders of magnitude larger than the typical errors in the chemical shift predictions, these regression plots can visually obscure chemically significant errors. Differences in the coefficients of determination (R^2 values) are also often quite small. For example, switching the density functional from the PBE0 hybrid functional to the GGA functional PBE increases the rms error between the predicted and experimental shifts from 1.5 ppm to 2.1 ppm,[5] but the R^2 value only decreases from 0.9987 to 0.9975. Analysis of the errors between the predicted and experimental chemical shifts (e.g. lower part of Figure 6) provides a much clearer gauge of how well a given model is performing.

IV. SUMMARY RECOMMENDATIONS AND FUTURE DIRECTIONS

In summary, when predicting chemical shifts from DFT, one should typically employ hybrid or GGA density functionals and pair them with a dispersion correction, if appropriate. Basis sets of triple-ζ quality provide a good balance between accuracy and computational cost. Locally dense basis set strategies can reduce computational costs without impacting the quality of the predicted shifts significantly. Special basis sets may be required for selected properties, such as electric field gradients. Atomic positions should generally be optimized with quantum chemistry prior to computing the chemical shifts, though finite truncations of the system (e.g. cluster models) may necessitate holding some atoms fixed at their crystallographic positions.

Modeling of the environment surrounding the atoms/molecules of interest is important. In some cases, this can be done with a simple continuum model. Otherwise, explicit treatments of the environment should be used, such as a large 4–6 Å cluster augmented with electrostatic embedding, or periodic boundary conditions. Fragmentation strategies can reduce the computational costs associated with predicting chemical shifts in large systems. It is always a good idea to assess the sensitivity of one's predictions to both the model chemistry (functional/basis set) and the atomic representation (e.g. size of the QM region). Finally, effective chemical shift referencing can be performed via the linear regression approach, preferably using a regression model trained on an external and well-characterized data set.

During the next several years, several trends in chemical shift prediction are likely to emerge. Increasing computer power makes it easier to treat larger portions of system with explicit QM methods. Machine-learning techniques are likely to gain importance for rapid predictions of chemical shifts across a wide range of chemical species. They have already been useful in protein systems for several years,[85] but new research is showing that they can perform well for more general organic molecules and crystals.[86,87] While the current accuracy of these general-purpose machine-learning models is somewhat worse than true DFT calculations (e.g. ∼4 ppm for ^{13}C versus ∼1.5–2 ppm from DFT), the calculations can be performed in seconds instead of hours.

The treatment of dynamics and its impact on NMR parameters is likely to improve and become more commonplace as well. Better potentials (machine-learned or otherwise) are likely to make performing *ab initio*-quality molecular dynamics simulations more affordable, from which one will be able to compute ensemble-averaged chemical shifts. Fast, accurate chemical shift prediction models will also enable averaging over the hundreds or thousands of configurational snapshots required to converge the chemical shieldings in large systems. These advances will all combine to improve and accelerate our ability to analyze and solve structures via NMR spectroscopy.

ACKNOWLEDGMENTS

Funding for this work from the National Science Foundation (CHE-1665212) is gratefully acknowledged.

BIOGRAPHICAL INFORMATION

Gregory Beran studied chemistry at the University of California (UC) San Diego (B.S. 2000) and went on to earn a Ph.D. with Martin Head-Gordon at UC Berkeley in 2005 developing new electronic structure methods. From 2005–2007, he performed postdoctoral research in

heterogeneous catalysis with William H. Green at the Massachusetts Institute of Technology. In 2007, he became a faculty member at UC Riverside. His research focuses on the development of efficient electronic structure models for systems where non-covalent interactions are important and the application of these methods to molecular crystal polymorphism.

[1] C. van Wüllen, *Calculation of NMR and EPR Parameters* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, 2004), pp. 83–100.
[2] S. Reimann, A. Borgoo, E. I. Tellgren, A. M. Teale, and T. Helgaker, Journal of Chemical Theory and Computation **13**, 4089 (2017).
[3] I. A. Konstantinov and L. J. Broadbelt, J. Phys. Chem. A **115**, 12364 (2011).
[4] M. W. Lodewyk, M. R. Siebert, and D. J. Tantillo, Chem. Rev. **112**, 1839 (2012).
[5] J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller, and G. J. O. Beran, Phys. Chem. Chem. Phys. **18**, 21686 (2016).
[6] J. D. Hartman, A. Balaji, and G. J. O. Beran, J. Chem. Theory Comput. **13**, 6043 (2017).
[7] F. Blanco, I. Alkorta, and J. Elguero, Magn. Reson. Chem. **45**, 797 (2007).
[8] D. O. Samultsev, V. A. Semenov, and L. B. Krivdin, Magn. Reson. Chem. **52**, 222 (2014).
[9] A. A. Auer, J. Chem. Phys. **131**, 024116 (2009).
[10] A. M. Teale, O. B. Lutnaes, T. Helgaker, D. J. Tozer, and J. Gauss, J. Chem. Phys. **138**, 024111 (2013).
[11] S. T. Holmes, R. J. Iuliucci, K. T. Mueller, and C. Dybowski, J. Chem. Theory Comput. **11**, 5229 (2015).
[12] M. Baias, J.-N. Dumez, P. H. Svensson, S. Schantz, G. M. Day, and L. Emsley, J. Am. Chem. Soc. **135**, 17501 (2013).
[13] J. D. Hartman, G. M. Day, and G. J. O. Beran, Cryst. Growth Des. **16**, 6479 (2016).
[14] J. D. Hartman, S. Monaco, B. Schatschneider, and G. J. O. Beran, J. Chem. Phys. **143**, 102809 (2015).
[15] B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano, and P. Ugliengo, CrystEngComm **10**, 405 (2008).
[16] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, Chem. Rev. **116**, 5105 (2016).
[17] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[18] A. D. Becke and E. R. Johnson, J. Chem. Phys. **124**, 014104 (2006).
[19] A. D. Becke and E. R. Johnson, J. Chem. Phys. **127**, 154108 (2007).
[20] E. R. Johnson, J. Chem. Phys. **135**, 234109 (2011).
[21] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).
[22] A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, 236402 (2012).
[23] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, and A. Tkatchenko, J. Chem. Phys. **140**, 18A508 (2014).
[24] E. Caldeweyher, C. Bannwarth, and S. Grimme, J. Chem. Phys. **147**, 034112 (2017).
[25] J. Hermann, R. A. DiStasio, and A. Tkatchenko, Chem. Rev. **117**, 4714 (2017).
[26] G. J. O. Beran, Chem. Rev. **116**, 5567 (2016).
[27] J. Kong, Z. Gan, E. Proynov, M. Freindorf, and T. R. Furlani, Phys. Rev. A **79**, 042510 (2009).
[28] G. S. Harbison, J. Magn. Res. **257**, 24 (2015).
[29] F. Jensen, J. Chem. Theory Comput. **11**, 132 (2015).
[30] D. B. Chesnut and K. D. Moore, J. Comp. Chem. **10**, 648 (1989).
[31] D. B. Chesnut, B. E. Rusiloski, K. D. Moore, and D. A. Egolfs, J. Comp. Chem. **14**, 1364 (1993).
[32] J. D. Hartman, T. J. Neubauer, B. G. Caulkins, L. J. Mueller, and G. J. O. Beran, J. Biomol. NMR **62**, 327 (2015).
[33] D. M. Reid, R. Kobayashi, and M. A. Collins, J. Chem. Theory Comput. **10**, 146 (2014).
[34] C. Pickard and F. Mauri, Phys. Rev. B **63**, 245101 (2001).
[35] C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azaïs, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri, and C. J. Pickard, Chem. Rev. **112**, 5733 (2012).
[36] S. E. Ashbrook and D. McKay, Chem. Commun. **52**, 7186 (2016).
[37] T. E. Exner, A. Frank, I. Onila, and H. M. Möller, J. Chem. Theory Comput. **8**, 4818 (2012).
[38] M. Dračínský and P. Hodgkinson, CrystEngComm **15**, 8705 (2013).
[39] M. Dračínský, H. M. Möller, and T. E. Exner, J. Chem. Theory Comput. **9**, 3806 (2013).
[40] H. Zhang, G. Hou, M. Lu, J. Ahn, I.-J. L. Byeon, C. J. Langmead, J. R. Perilla, I. Hung, P. L. Gor'kov, Z. Gan, W. W. Brey, D. A. Case, K. Schulten, A. M. Gronenborn, and T. Polenova, J. Am. Chem. Soc. **138**, 14066 (2016).
[41] M. Dračínský, P. Bouř, and P. Hodgkinson, J. Chem. Theory Comput. **12**, 968 (2016).
[42] Y. N. Heit and G. J. O. Beran, Acta Cryst. B **72**, 514 (2016).
[43] J. L. McKinley and G. J. O. Beran, Faraday Disc. **211**, 181 (2018).
[44] E. Salager, G. M. Day, R. S. Stein, C. J. Pickard, B. Elena, and L. Emsley, J. Am. Chem. Soc. **132**, 2564 (2010).
[45] C. J. Pickard and R. J. Needs, J. Phys. Condens. Mat. **23**, 053201 (2011).
[46] S. L. Price, Chem. Soc. Rev. **43**, 2098 (2014).
[47] S. L. Price, D. E. Braun, and S. M. Reutzel-Edens, Chem. Commun. **52**, 7065 (2016).
[48] S. L. Price, Faraday Disc. **211**, 9 (2018).
[49] A. R. Oganov, Faraday Disc. **211**, 643 (2018).
[50] R. F. Moran, D. M. Dawson, and S. E. Ashbrook, Int. Rev. Phys. Chem. **36**, 39 (2017).
[51] R. F. Moran, D. McKay, C. J. Pickard, A. J. Berry, J. M. Griffin, and S. E. Ashbrook, Phys. Chem. Chem. Phys. **18**, 10173 (2016).
[52] J. Kussmann, M. Beer, and C. Ochsenfeld, WIRES: Comput. Mol. Sci. **3**, 614 (2013).
[53] E. R. Johnson and G. A. DiLabio, J. Mol. Struct. (THEOCHEM) **898**, 56 (2009).
[54] D. Flaig, M. Beer, and C. Ochsenfeld, J. Chem. Theory Comput. **8**, 2260 (2012).
[55] S. T. Holmes, R. J. Iuliucci, K. T. Mueller, and C. Dybowski, J. Chem. Phys. **141**, 164121 (2014).
[56] M. Beer, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **134**, 074102 (2011).
[57] D. Stueber, Conc. Magn. Reson. A **28**, 347 (2006).
[58] E. V. Stefanovich and T. N. Truong, J. Phys. Chem. B **102**, 3018 (1998).
[59] M. B. Ferraro and J. C. Facelli, J. Mol. Struct. **603**, 159 (2002).
[60] C. Steinmann, J. M. H. Olsen, and J. Kongsted, J. Chem. Theory Comput. **10**, 981 (2014).
[61] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105**, 2999 (2005).
[62] C. N. Schutz and A. Warshel, Proteins **44**, 400 (2001).
[63] L. Li, C. Li, Z. Zhang, and E. Alexov, J. Chem. Theory Comput. **9**, 2126 (2013).
[64] P. Kukic, D. Farrell, L. P. McIntosh, B. García-Moreno E, K. S. Jensen, Z. Toleikis, K. Teilum, and J. E. Nielsen, J. Am. Chem. Soc. **135**, 16968 (2013).
[65] L. An, Y. Wang, N. Zhang, S. Yan, A. Bax, and L. Yao, J. Am. Chem. Soc. **136**, 12816 (2014).
[66] B. Mennucci, Phys. Chem. Chem. Phys. **15**, 6583 (2013).
[67] H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2006).
[68] J. D. Hartman and G. J. O. Beran, Solid State Nucl. Magn. Reson. **96**, 10 (2018).
[69] D. M. Reid and M. A. Collins, J. Chem. Theory Comput. **11**, 5177 (2015).
[70] X. He, B. Wang, and K. M. Merz, J. Phys. Chem. B **113**, 10380 (2009).

[71]X. He, T. Zhu, X. Wang, J. Liu, and J. Z. H. Zhang, Acc. Chem. Res. **47**, 2748 (2014).
[72]J. Swails, T. Zhu, X. He, and D. A. Case, J. Biomol. NMR **63**, 125 (2015).
[73]A. M. Lee and R. P. A. Bettens, J. Phys. Chem. A **111**, 5111 (2007).
[74]H.-J. Tan and R. P. A. Bettens, Phys. Chem. Chem. Phys. **15**, 7541 (2013).
[75]D. M. Reid and M. A. Collins, Phys. Chem. Chem. Phys. **17**, 5314 (2015).
[76]Q. Gao, S. Yokojima, T. Kohno, T. Ishida, D. G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura, Chem. Phys. Lett. **445**, 331 (2007).
[77]Q. Gao, S. Yokojima, D. G. Fedorov, K. Kitaura, S. M, and N. S, J. Chem. Theory Comput. **6**, 1428 (2010).
[78]A. Frank, I. Onila, H. M. Möller, and T. E. Exner, Proteins **79**, 2189 (2011).
[79]A. Victora, H. M. Möller, and T. E. Exner, Nucl. Acids Res. **42**, 1 (2014).
[80]K. V. J. Jose and K. Raghavachari, J. Chem. Theory Comput. **13**, 1147 (2017).
[81]C. Yang, L. Zhu, R. A. Kudla, J. D. Hartman, R. O. Al-Kaysi, S. Monaco, B. Schatschneider, A. Magalhaes, G. J. O. Beran, C. J. Bardeen, and L. J. Mueller, CrystEngComm **18**, 7319 (2016).
[82]G. J. O. Beran, CrystEngComm **21**, 758 (2018).
[83]B. G. Caulkins, R. P. Young, R. A. Kudla, C. Yang, T. J. Bittbauer, B. Bastin, E. Hilario, L. Fan, M. J. Marsella, M. F. Dunn, and L. J. Mueller, J. Am. Chem. Soc. **138**, 15214 (2016).
[84]R. P. Young, B. G. Caulkins, D. Borchardt, D. N. Bulloch, C. K. Larive, M. F. Dunn, and L. J. Mueller, Angew. Chem. Int. Ed. **55**, 1350 (2016).
[85]D. S. Wishart, Prog. NMR. Spectrosc. **58**, 62 (2011).
[86]M. Rupp, R. Ramakrishnan, and O. A. von Lilienfeld, J. Phys. Chem. Lett. **6**, 3309 (2015).
[87]F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, and L. Emsley, Nature Commun. **9**, 4501 (2018).