Received: 7 December 2023 | Revised: 15 February 2024 | Accepted: 22 February 2024

DOI: 10.1002/jcc.27338

RESEARCH ARTICLE

Journal of COMPUTATIONAL CHEMISTRY WILEY

Modeling solvent effects and convergence of ^{31}P-NMR shielding calculations with COBRAMM

Francesco Calcagno[1,2] | Boris Maryasin[3,4] | Marco Garavelli[1] | Davide Avagliano[1] | Ivan Rivalta[1,2,5]

[1]Department of Industrial Chemistry "Toso Montanari", University of Bologna, Bologna, Italy

[2]Center for Chemical Catalysis – C3, University of Bologna, Bologna, Italy

[3]Institute of Organic Chemistry, University of Vienna, Vienna, Austria

[4]Institute of Theoretical Chemistry, University of Vienna, Vienna, Austria

[5]ENSL, CNRS, Laboratoire de Chimie UMR 5182, Lyon, France

Correspondence
Marco Garavelli, Davide Avagliano, and Ivan Rivalta, Department of Industrial Chemistry "Toso Montanari", University of Bologna, via Piero Gobetti 85, 40129 Bologna, Italy.
Email: marco.garavelli@unibo.it, davide.avagliano@chimieparistech.psl.eu and i.rivalta@unibo.it

Abstract

Solvent effects on ^{31}P-NMR parameters for triphenylphosphine oxide and triphenylphosphine in chloroform have been extensively investigated by testing different solvation models. The solvent is described implicitly, mixed implicitly/explicitly, and using full explicit models. Polarizable continuum model (PCM), molecular dynamic (MD) simulations, and hybrid quantum mechanics/molecular mechanics (QM/MM) calculations are used to disclose the effects of solute/solvent interactions and, more generally, the role of the embedding in NMR simulations. The results show the beneficial effect of carrying out QM/MM optimizations on top of geometries directly extracted from classical MD simulations, used to ensure representative conformational sampling. The nuclear shielding convergence has been tested against a different number of snapshots and with the inclusion of solvent shells into the QM region. An automated MD//QM/MM//GIAO protocol, implemented in the COBRAMM package, is here proposed and tested on trimethyl phosphite showing that our approach boosts the convergence of nuclear shielding satisfactorily. The present work aims to be a stepping-stone to assess proper QM/MM computational strategies in simulating chemical shifts in non-homogeneous systems like supramolecular and biological systems.

KEYWORDS
automatized computational protocol, density functional theory, molecular dynamics, NMR, QM/MM

1 | INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is one of the most essential characterization and identification techniques in chemistry. Its paramount importance is witnessed in the everyday application to monitor any organic synthesis, but also in the more complex investigations such as unveiling structures and conformations of supramolecular systems,[1] macromolecules,[2] or for medical purposes.[3] NMR spectroscopy is based on the energy gaps between nuclear energy levels under the action of an external magnetic field (B_0), which removes the degeneracy of the ground state. Since each nucleus is surrounded by a different electron density, which shields the magnetic field according to Lenz's law, the effective magnetic field (B_{eff}), that is, the energy gap between states, is different for each chemically non-equivalent nucleus (Figure 1A). This guarantees to uniquely define the chemical nature of the atoms in a species, expressed in terms of chemical shift (δ), which is the relative shift of the resonance frequency with respect to a specific reference for each nucleus. As with many other spectroscopic techniques, interpretation of the experimental spectra alone is often insufficient to resolve structural, conformational, and solvent effects. Computational chemistry comes in support by predicting the chemical shift of the nuclei of interest. Quantum chemical modeling of NMR spectra is based on the calculation of the magnetic shielding tensor, which nowadays can be

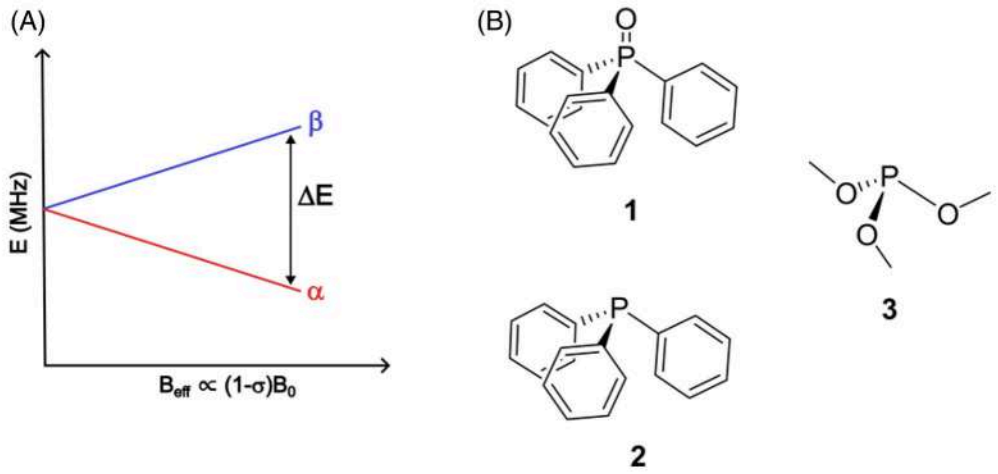

FIGURE 1 (A) Nuclear energy levels (α and β) for an exemplary nucleus with nuclear spin I = 1/2. Applying an external magnetic field (B_0), the energy gap between states (ΔE) will be characteristic of the magnetic shielding σ. (B) Molecular structures of triphenylphosphine oxide (**1**), triphenylphosphine (**2**), and trimethyl phosphite (**3**) studied in this work. **1** is considered as the analyte and **2** as a reference for the calculation of the chemical shift, while **3** is employed for investigations about the convergence behavior of the nuclear shielding value in function of different parameters.

performed *via* various approaches and at different levels of theory.[4,5] Among specific impediments of the magnetic shielding tensor calculations is the gauge-origin problem. Namely, the invariance of the results with respect to the chosen origin of the magnetic field is not automatically guaranteed.[6–8] The well-known gauge-including atomic orbital method (GIAO) is one of the most popular solutions.[9–11]

As simple in theory as challenging in practice, chemical shift prediction is not trivial when solvation or, more in general, environmental effects are considered.[5,12] The solvent deeply modifies the electron density of molecules and its action is strongly dependent on the type of interactions possible for the pair solute/solvent,[13] emerging as a crucial ingredient for the simulations. Both implicit, where the solvent is considered as a continuum medium, and explicit, where solvent molecules are directly included in the calculation, solvation schemes have been applied to NMR calculations,[14] especially for common nuclei such as ^{1}H, ^{13}C, ^{15}N, and ^{17}O.[15–17] Implicit solvation approaches, such as the polarization continuum model (PCM),[18] improve the computations carried out in-vacuum,[15,19] but cannot intrinsically reproduce direct interactions (e.g., hydrogen bonds) between the solute and single solvent molecules, which can drastically alter the chemical shift.[20] The same problem arises with non-homogeneous systems, where the chemical surrounding usually plays a central role in altering the electronics of the system. An intermediate approach for small molecules in solution is based on adding a few explicit molecules of solvent around the sample in combination with the polarizable continuum model (e.g., PCM).[15,21–26] This procedure greatly improves the experiment/theory agreement as compared to the pure PCM approach. However, it is highly system-dependent and not automatable at low computational cost. Concerning explicit solvation, quantum mechanics/molecular mechanics (QM/MM) techniques have been extensively used in NMR calculations.[14,27] Different studies showed the effect of using mechanical, electrostatic, and polarizable embedding to treat the interactions between the QM and the MM regions.[16,17,26,28,29] Hybrid ONIOM schemes,[27,30,31] as well as including solvation shells into the QM region,[16,17,32] were also successfully employed in NMR calculations since a classical description of the first solvation shell does not always reproduce the shielding satisfactorily. This is particularly true for polar molecules and non-homogeneous systems, for which a QM description of the surrounding is needed.[17,32] Ochsenfeld and co-workers showed how a full convergence of the shielding tensor for ^{1}H, ^{13}C, ^{15}N, and ^{17}O is reached when a big radius of the solvent shell (more than 1000 atoms) is included into the QM region,[17] which computational investigation was feasible thanks to a linear scaling density functional theory (DFT) NMR approach.[33–35] Nevertheless, Steinmann et al. showed that shielding convergence for ^{17}O of acetone in water could be alternatively reached, including much fewer solvent molecules if a polarizable embedding scheme is used.[16]

Although less intensively studied than other nuclei, phosphorus is a highly interesting element for the computational investigation of NMR parameters.[36] In fact, it is stable in nature as a 100% abundant isotope (^{31}P), it is widely present in organic,[15] bio-organic[37,38] and organometallic compounds,[39,40] its chemical shift ranges within 1000 ppm ca., and its nuclear spin is I = 1/2, making the experimental recording of spectra very easy and fast. All these characteristics make ^{31}P-NMR one of the most feasible techniques to study small structural changes in chemical processes[41,42] or following catalysts' behavior.[43,44] Despite its extensive applications in chemistry, prediction of the ^{31}P-NMR chemical shifts is not as easily accessible as for ^{1}H and ^{13}C,[42] whose behavior can be well estimated to a first approximation using empirical correlation tables.[45] As already indicated by Ciofini and co-workers,[40] we still lack an extensive validation of the feasibility of DFT in predicting ^{31}P-NMR chemical shifts.[40] Over the years, many studies have been separately reported on investigations of single

effects on the computation of nuclear shielding, especially when solvent effects are employed. Several solvation models have been proposed to better reproduce ^{31}P chemical shift of different compounds comprising both implicit,[40,46–48] explicit, and mixed implicit-explicit models.[15,31,37,49–52] Sklenář co-workers,[38,53] as well as Sychrovský co-workers, have highlighted the importance of sampling and averaging of conformation in phosphorus-containing compounds when explicit solvation is considered, employing classical molecular dynamics (MD)/DFT protocols.[31,37,52] Nevertheless, to the best of our knowledge, an investigation of effects detected by different solvation models, conformational sampling, convergence of the shielding, and especially the potential benefit of QM/MM optimization is not presented in the literature.

Noteworthy in the field, Maryasin and Zipse extensively investigated mixed implicit/explicit solvation models to reproduce ^{31}P-NMR chemical shifts. They showed that the addition of only one solvent molecule at specific user-selected positions in combination with PCM leads to a good reproduction of the chemical shift of triphenylphosphine oxide (**1**, Figure 1B) in chloroform ($CHCl_3$) taking triphenylphosphine (**2**, Figure 1B) as reference.[15] Prompted by these results and the lack of more detailed investigations of other solvation models on these systems, in the current work we present a systematic study of ^{31}P-NMR nuclear shielding values and chemical shifts of **1** as a case study. We test different solvation approaches along with MD and QM/MM computational protocols. We also test the efficacy and performance of these protocols to converge to a given result, according to the size of the solvent shell treated at QM level, using a QM-PCM scheme,[16,17] or the number of snapshots, when using MD simulation describing the solute/solvent interactions. We pay particular attention to the conformational sampling procedure and evaluate the effect of geometric and electrostatic factors in calculating magnetic shielding tensors. Finally, we propose our new competitive protocol based on electrostatic embedding to quickly reach convergence of computed nuclear shielding values, reducing the computational cost without using either a polarizable embedding scheme or linear scaling software, showing its performances on trimethyl phosphite (**3**, Figure 1B). The protocol is now automatized and available in the COBRAMM package.[54,55] Our study aims to be a stepping-stone for further investigations of more complicated non-homogeneous systems where the environmental effects and their modeling are not negligible. We aim to provide a comprehensive overview of what to expect from the different protocols when simulating ^{31}P-NMR in the presence of an environment, paving the way for simulations of complex, extended, and chemically relevant systems such as supramolecular systems, biological compounds, and macromolecules, where environmental effects must be considered balancing accuracy and computational cost.

2 | THEORY AND METHODS

2.1 | Experimental

NMR spectra of triphenylphosphine oxide (**1**, Figure S1), and triphenylphosphine (**2**, Figure S2) were acquired at 298 K with a Varian Mercury Plus VX 400 (^{31}P{^{1}H}, 161.8 MHz; {^{13}C}, 100.6 MHz) spectrometer and employing commercial compounds. Chemical shifts were referred to an external standard mixture of Ph_2PH and Ph_3P in chloroform-d (see the Supporting Information).

2.2 | NMR chemical shift calculations

The nuclear shielding tensor is defined as the mixed second derivative of the energy with respect to the external magnetic field, B, and the magnetic moment of considered nucleus N, m_N.

$$\sigma_N^{ji} = \frac{\partial^2 E}{\partial B^i \partial m_N^j}, \tag{1}$$

where j and i are the components of the induced magnetic moment and external magnetic field, respectively.[4] One of the most commonly used approaches to solve the gauge-origin problem while computing the nuclear shielding tensor is the GIAO-self consistent field (GIAO-SCF) method.[9–11] Within this approach, the tensor at a Kohn-Sham density functional theory (DFT) level is defined as:

$$\sigma_N^{ji} = \sum_{\mu\nu} P_{\mu\nu} \frac{\partial^2 h_{\mu\nu}}{\partial B^i \partial m_N^j} + \sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial B^i} \frac{\partial h_{\mu\nu}}{\partial m_N^j}, \tag{2}$$

where $P_{\mu\nu}$ is the one-particle density matrix with the basis μ and ν, $h_{\mu\nu}$ is the one-electron matrix, B_i is the i-th component of the external magnetic field component and $m_{Nj}{}^j$ is the j-th component of the magnetic moment of nucleus N.

Since for samples in liquids only isotropic shielding, σ_{iso}, can be experimentally measured, all next chemical shifts value will be reported to the isotropic term of the shielding tensor (Equation (3)).

$$\sigma_{iso} = \frac{1}{3} Tr(\sigma) = \frac{1}{3} Tr \begin{pmatrix} \sigma_{XX} & \sigma_{YX} & \sigma_{ZX} \\ \sigma_{XY} & \sigma_{YY} & \sigma_{ZY} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_{ZZ} \end{pmatrix}. \tag{3}$$

From now, we will use σ_{iso} and σ as synonyms.

The error for σ calculations is computed as standard deviation (s), standard error of the mean (SEM) for the set of snapshots (Equation (4))[16] and bootstrapping the value over 100 samples of 100 random snapshots each.

$$\mathrm{SEM} = \frac{s(\sigma_N)}{\sqrt{N_{snapshots}}}. \tag{4}$$

For direct comparison with experimental results, it is necessary to compute the chemical shift (δ) of the nuclei under study, since this is the experimental quantity used to extrapolate structural information about the molecules from NMR spectra. In this work, we calculate the chemical shift of the analyte (**1**, Figure 1B), by comparing its calculated magnetic shielding with that of the chosen reference molecule (**2**, Figure 1B) and correcting with the experimental chemical shift of

the reference that we have recorded (Table S1).[15] Thus, we apply the intermediate reference approach using

$$\delta_{\text{Comp.}}(\mathbf{1}) = \sigma(\mathbf{2}) - \sigma(\mathbf{1}) + \delta_{\text{Exp.}}(\mathbf{2}), \quad (5)$$

where $\delta_{\text{Exp.}}(\mathbf{2})$ is the experimental chemical shift of the species **2**.

2.3 | QM/mm scheme

The QM/MM approach allows computing energy and molecular properties of a molecule of interest with the accuracy of a QM method while accounting for the effect of the environment, which is treated at a lower level of theory, mainly by means of a classical force field (FF). The theory and the applications of such a method are well known,[56] and we report here only the details needed to understand our workflow. In COBRAMM,[54,55] a subtractive scheme for the calculation of energies and properties is applied. In this scheme, the QM/MM Hamiltonian ($\widehat{H}_{\text{QM/MM}}$) is given by

$$\widehat{H}_{\text{QM/MM}} = \widehat{H}_{\text{QM}} + \widehat{H}_{(\text{QM+MM})} - \widehat{H}_{\text{QM[MM]}} + \widehat{H}_{\text{emb}}, \quad (6)$$

where the Hamiltonian of the QM part $\left(\widehat{H}_{\text{QM}}\right)$ is summed to the full system one, calculated at the MM level $\left(\widehat{H}_{(\text{QM+MM})}\right)$, and from which the QM one, but now calculated at the MM level $\left(\widehat{H}_{\text{QM[MM]}}\right)$, is subtracted. In order to go beyond the purely classical description of the QM-MM interaction, an extra term can be added to describe the embedding interaction $\left(\widehat{H}_{\text{emb}}\right)$. In COBRAMM, the electrostatic embedding scheme is used, where $\widehat{H}_{\text{QM}}$ includes the point charges of the environment that polarize the QM density. QM/MM calculations in COBRAMM are performed by using a finite-system approach. In detail, a spherical droplet is cut around the QM region, which composes the so-called High layer (H). The environment is partitioned into two regions: a medium layer (M), where the molecules of the MM region are allowed to move, and a low layer (L), where the energy of the molecules is computed, but those are kept frozen to ensure a constant and stable potential around the H-M regions. The link-atom approach is used when a covalent bond connects the two regions.[57] In the case of nuclear shielding calculations, only the QM part is considered, with the environment solely interacting by the inclusion of the point charges in the QM Hamiltonian.

2.4 | Computational protocol

DFT was used to describe the QM region. QM calculations were carried out with the Gaussian16 software package,[58] using the MPW1K functional[59] as suggested by previous studies,[15] and different basis sets according to the required task. Magnetic shielding tensors were computed at the DFT level and using the GIAO method as implemented in Gaussian16. MD and MM calculations were done using the AMBER suite,[60] and QM/MM computations were performed using COBRAMM.[54,55] Cluster analyses were done using the principal component analysis (PCA) and the KernelPCA (KPCA) with rbf kernel, as implemented into the scikit-learn (version 0.24.2) Python3 library.[61] Related code is available on GitLab (https://gitlab.com/cobrammgroup/cobramm/-/tree/cobramm-NMR).

2.4.1 | Gas and implicit solvation calculation

Geometries of all conformers of **1** and **2**, as reported in ref.,[15] were optimized in the gas phase using the 6-31G(d) basis set[62–66] for all atoms. Each structure was characterized as a minimum by computing analytical frequencies at the same level of theory. NMR parameters and refined energies for thermodynamics were obtained from single point calculations using a larger basis set, that is, 6–311++G(2d,2p).[67–70] Thermal corrections were computed at 298.15 K and, when present, solvent corrections were added. Solvation was computed by adding a polarizable continuum model (PCM)[18] for chloroform, both for analytic frequencies and NMR computations, using united atom radii computed at the Hartree-Fock (HF/6-31G(d)) level of theory (UAHF) and a dielectric constant of 4.7113 for $CHCl_3$.

2.4.2 | Conformational sampling

Prior to computing the NMR chemical shifts, appropriate conformational space exploration is necessary for highly flexible solute/solvent clusters within the explicit solvation approach. The sampling was obtained by means of classical MD. Simulations followed the recently automatized procedure in COBRAMM.[71] The analytes were solvated with chloroform molecules filling a truncated octahedral box till a distance of 16 Å between the solute and the edges of the box. The MD protocol was divided into three steps. First, the system was minimized for 500 steps with the steepest descendent algorithm,[72] and subsequently for 1000 steps using the conjugated gradient algorithm.[73] After that, the system was heated to 300 K, using a random seed to generate the initial velocities. After 25 ps, the system was stable at the required temperature, and pressure and volume were equilibrated for 130 ps. Temperature was controlled by a Langevin thermostat,[74] and pressure by a Berendsen barostat.[75] Periodic boundary conditions were used and the Particle Meshed Ewald method[76] was used to calculate the electrostatic interactions, with a cutoff of 10 Å. General Amber Force Field (GAFF)[77] was used for the solute, with the atomic charges calculated with the automatized tool in COBRAMM[71] either using the AM1 Hamiltonian (when solute is kept frozen), as implemented in AMBER,[78] or at the B3LYP[79–83]/6-31G(d) level of theory (when solute is free to move), and the FF implemented in AMBER was used for the chloroform molecules.[84] All-atoms unrestrained molecular dynamics was run for 100 ns and 100 uncorrelated snapshots were evenly extracted after 1 ns. These geometries were

used for single point and optimization QM/MM calculations. In the case of solvent thermal sampling, a bigger box is used (i.e., 30 Å) to surround the species as reported in Reference 15. The solvent was firstly minimized for 1500 steps, heated to 300 K and equilibrated for 130 ps keeping the solute frozen, as already described. The last geometries from equilibrations were taken for further computations.

2.4.3 | QM/MM calculations

For each of the geometries, a droplet was obtained by cutting a 15 Å radius solvent shell around the phosphines. QM/MM optimizations were run for each of these geometries until convergence was reached, as implemented in COBRAMM. The QM region was treated at the DFT level, using the MPW1K functional and the 6-31G(d) basis set. The M- and the L-layers were treated at the MM level, as for the MD sampling. Different sizes of the M-layer were used, as specified for each calculation. Single point NMR calculations were computed using the 6–311++G(2d,2p) basis set or the pcS-1[85] one, with the last one chosen because it is optimized for calculating NMR shielding constants basis sets. pcS-1 basis set was used as reported by the basis set exchange (BSE) database.[86–88] When indicated, the solvent effect was computed using PCM[18] for $CHCl_3$ and employing the UAHF radii for all the atoms.

2.4.4 | Implementations in COBRAMM

Following our recent implementations characterized by special efforts to improve user-friendliness in the COBRAMM package,[71] we first updated the COBRAMM/Gaussian16 interface to include the GIAO calculations of NMR parameters. Then, we wrote a new auxiliary program for COBRAMM to assist non-expert users with the calculations. This code is able to read a topology/coordinate file in COBRAMM(AMBER) format after an MD run or QM/MM optimization. Certain parameters (e.g., the size of the droplet) can be adjusted by the user, and it is possible to include solvent shells in the QM region. The MPW1K functional and the 6–311++G (2p,2d) basis set were chosen as the default parameters. This code complements the software's existing library of tools that allow system preparation (solvation, FFs parametrization, minimization-heating-equilibration procedure), classical MD, QM/MM optimization and vibrational frequencies, and now NMR calculations with a minimal effort of the user, but always leaving the freedom to select crucial parameters (e.g., the size of the H-layer). This is an essential tool to increase the number of tests and applications for computing NMR spectra of not only ^{31}P but also other nuclei in various environments (e.g., solvation, supramolecular systems, macromolecules, molecular aggregate, etc.). The new implementation is open-source, like all the code, and is already available by downloading COBRAMM from its official release webpage (https://gitlab.com/cobrammgroup/cobramm/-/tree/cobramm-NMR).

3 | RESULTS AND DISCUSSION

In the following, we discuss and compare different solution models (SolMs, see Figure 1) for **1** to simulate its chemical shift, according to Equation (5). The triphenylphosphine oxide **1** is taken as a case study for the chemical shifts computation since it is a common compound in organic chemistry. As the degradation product of triphenylphosphine **2**, which is widely used as a ligand or organocatalyst, **1** is a side product in various processes.[15] The first calculations are performed in the absence of solvent (gas phase, SolM0), then δ is calculated in the presence of solvent *via* implicit solvation approach PCM (SolM1), and with PCM in conjunction with a microsolvation model (SolM2), involving a single molecule of explicit solvent, similarly to the procedure recommended in Reference 15, but with a new full DFT protocol instead of a mixed MP2//DFT one. Next, we systematically investigate the effects of explicit solvation on the chemical shift of **1** using a QM/MM scheme (SolM3) considering (i) different sizes of the M-layer, (ii) the correlation between the QM/MM optimized structure and the computed shielding, (iii) the different number of MD snapshots and (iv) size of the H-layer (Figure 2). Finally, we propose a protocol implemented in COBRAMM (version 2.3) to quickly compute magnetic shielding and apply it to trimethyl phosphite (**3**, Figure 1).

3.1 | Predicting chemical shifts with a static approach

Initially, we calculated the chemical shifts of **1** with SolM0, SolM1, and SolM2. It was previously reported that the number of conformers for **1** and **2** is different and varies depending on the applied solution model: for SolM0 and SolM1, there are two conformers for **1** and one for **2**, while with SolM2, the number of conformers is reversed, that is, one conformer for **1** and two conformers for **2**, due to the addition of an explicit solvent molecule (see the Supporting Information).[15] Since the number of nondegenerate conformers varies in different SolMs, all computed shielding values reported here are averaged according to the Boltzmann probabilities of the conformers at room temperature (see the Supporting Information for details).[15]

In agreement with ref.,[15] neglecting solvation leads to an underestimation of the chemical shift of **1** by 6.9 ppm with respect to the experimental value ($\delta_{SolM0} = 23.1$ ppm, $\delta_{Exp.} = 29.0$). Including the solvent in the model improves the computed results, as shown in Figure S3. Indeed, in SolM1, only applying implicit PCM solvation, the computed chemical shift is increased by 3.6 ppm ($\delta_{SolM1} = 26.7$ ppm), showing that the solvent effect is not negligible in NMR chemical shift computation of **1**. The implicit solvation significantly affects the magnetic shielding value of the analyte, which varies by more than 3 ppm (σ_{SolM0}(**1**) = 308.1 ppm, σ_{SolM1}(**1**) = 304.8 ppm), while the reference compound is not really affected by PCM (σ_{SolM0}(**2**) = 336.7 ppm, σ_{SolM1}(**2**) = 336.8 ppm, Table 1). Moreover, only the addition of an explicit solvent molecule in combination with PCM (SolM2) allows reaching a good agreement with the experiment ($\delta_{SolM2} = 28.5$ ppm), in line with the previous work by Maryasin and Zipse.[15] One of the

reasons for this improvement is that, in the case of SolM2, the presence of an explicit chloroform molecule ensures better modeling of the solute/solvent interaction, which is not merely a polarization of the solute. In fact, chloroform can interact with **1** by means of a P═O···$HCCl_3$ hydrogen bond (1.918 Å). This is reflected in an increase of the P═O bond length from 1.487 Å in SolM0 to 1.492 Å in SolM2, and a decrease in the magnetic shielding value, from $\sigma_{SolM0}(\mathbf{1}) = 308.1$ to $\sigma_{SolM2}(\mathbf{1}) = 301.8$ ppm. Thus, the nuclear magnetic shielding is highly sensitive to the length of the bonds directly linking the nucleus under investigation (P atom in our case), which determines, to a first approximation, the variation of the electron density around that nucleus. It is particularly noteworthy, as shown in Table 1, that the influence of the solvent on the computed ^{31}P-NMR magnetic shielding values is not the same for both compounds but is stronger for **1** than for **2** in all SolMs. The computed ^{31}P-NMR magnetic shielding value of **1** varies from 308.1 ppm (SolM0) to 301.8 ppm (SolM2), while **2** is almost unaffected by the solution model since its ^{31}P-NMR magnetic shielding value does not change by more than 1.1 ppm.

Besides the SolM2 features good prediction of ^{31}P-NMR chemical shift of **1** (using **2** as reference), this model still significantly approximates the solute/solvent interactions, lacking a comprehensive picture of both the geometric and electronic influences of the surrounding. Such effects play central role in non-homogeneous systems, for example, supramolecular and biological compounds, and cannot be neglected or roughly approximated when attempting to simulate the ^{31}P-NMR chemical shifts in such systems. Therefore, we now question on the performances of a more complex solution model (e.g., SolM3), which accounts for more realistic embedding and improved conformational sampling of both solute and solvent.

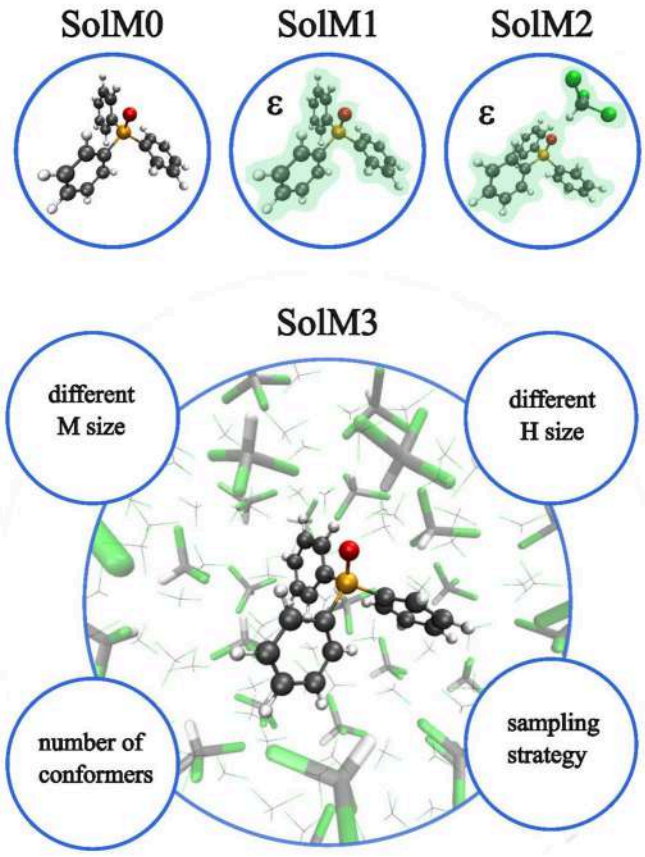

FIGURE 2 Summary of investigated solvation models. Solution model 0 (SolM0) does not include any solvation, solution model 1 (SolM1) involves implicit solvation, while in solution model 2 (SolM2) microsolvation (i.e., one molecule of solvent) is added to implicit solvation. Solution model 3 (SolM3) is represented by explicit solvation, with a representative snapshot shown in the 3D inset, depicting an example of High (balls and sticks), Medium (thick sticks), and Low (thin ticks) layers assignment in the QM/MM scheme of the COBRAMM code.

3.2 | QM/MM optimization and chemical shifts: The solvent relaxation effect

Macrosolvation MD-based methods could provide a reliable way to model sophisticated solute/solvent interactions, overcoming the limitations associated with user-controlled positioning of solvent molecules in microsolvation approaches, such as in SolM2. To test the applicability of the explicit macrosolovation approach (SolM3) for the computation of ^{31}P-NMR chemical shifts, we first investigated the influence of MM solvation on a single representative structure. Each of the conformers previously reported in vacuum, for both **1** and **2**,[15] was placed in a box of explicit solvent (chloroform) molecules (see Section 2.4.2 for details) using the recently developed tools in COBRAMM.[71] The solvent was equilibrated around the phosphines, which were kept frozen. Then, geometrical relaxations were performed using the QM/MM hybrid method on the last geometry extracted from MD simulations, treating only the solute (**1** or **2**) at the QM level and cutting out a fixed-size solvent droplet (see Section 2.4.3 for details). QM/MM optimizations were performed by choosing different thresholds for the mobile M-layer (from 3 to 8 Å) to evaluate how much different extent of relaxation influences the calculation of the NMR shielding value. Surprisingly, the nuclear shielding of **1** is less affected by the size of the M-layer than **2**, as shown in Table 2, contrary to what was observed in the previous section, where **2** showed

TABLE 1 Computed Boltzmann-averaged ^{31}P-NMR magnetic shielding values and corresponding chemical shifts obtained at the MPW1K/6-311++G (2d,2p) level of theory.

	σ(1) ppm	σ(2) ppm	δ(1) ppm	Solvation model
SolM0	308.1	336.7	23.1	Vacuum
SolM1	304.8	336.8	26.7	PCM
SolM2	301.8	335.7	28.5	One explicit molecule and PCM
Experimental			29.0	

TABLE 2 Nuclear shielding values for the two conformers of **1** and **2** characterized in the gas phase, after QM/MM optimization on single structures considering different sizes of M-layer.

	QM/MM			Only H-layer			Solvent effect		
	σ(1) (ppm)		σ(2) (ppm)	σ(1) (ppm)		σ(2) (ppm)	$\sigma^{QM/MM}(1) - \sigma^{vacuum}(1)$		$\sigma^{QM/MM}(2) - \sigma^{vacuum}(2)$
M-Layer (Å)	conf1	conf2		conf1	conf2		conf1	conf2	
3	300.9	304.1	356.9	305.3	305.2	339.3	−4.4	−1.1	17.6
4	300.3	302.6	329.3	305.2	303.8	335.7	−4.9	−1.2	−6.4
5	304.3	301.2	325.8	306.5	303.1	333.3	−2.2	−1.9	−7.5
6	302.9	302.1	338.4	306.6	304.3	334.9	−3.7	−2.2	3.5
7	300.8	305.5	336.0	305.4	307.9	332.6	−4.6	−2.4	3.5
8	305.1	307.7	330.1	306.9	309.1	347.2	−1.8	−1.4	−17.0

Note: The shielding values were computed both considering explicit MM solvation and in a vacuum at the MPW1K/6–311++G(2d,2p) level of theory. For each M layer chosen, the difference between the shielding value computed at the QM/MM level and at the "Only H-layer" level, that is, considering the QM/MM optimized geometry of the H layer in vacuum, is presented.

stable values of the shielding according to different solution models (Table 1). In fact, the nuclear shielding of **2** fluctuates in a spectral window of 31.1 ppm against a maximum of 6.5 ppm deviation for **1**. Remarkably, for none of the compounds, a convergence value of shielding is reached when the threshold values for the radius of the M-layer shell are increased since the nuclear shielding is strongly influenced by small geometrical changes and minimal reorganization of the solvent, but for **2** a trend can be observed. When the M-layer is set to 3 Å and the solute/solvent reorganization is limited in a highly constrained region, the shielding value of **2** (σ_{SolM3}(**2**) = 356.9 ppm) is very far from that observed for SolM2 (σ_{SolM2}(**2**) = 335.7 ppm). However, as the size of the M-layer increases, the nuclear shielding stabilizes at around 330 ppm (Table 2). These outcomes indicate that MD can provide an adequate embedding of the solvent, but QM/MM relaxation of a large enough shell of solvent is needed to properly mitigate the effect of non-optimal, fully classical solute/solvent geometries. In fact, small reorganizations of point charges lead to large fluctuations in nuclear shielding, and the presence of mobile MM molecules notably affects the geometries during the optimization, as reported for **1** in the Supporting Information (see Figure S4). Since, to a first approximation, the total nuclear shielding σ can be imagined as the sum of a rovibrational component σ_0 and a solvent correction σ_s,[12] the effect of the solvent can be easily evaluated by removing the point charges around the relaxed phosphines and recomputing the NMR parameters ("Only H-layer," Table 2). The results showed that the solvent effect on **1** fluctuates between −1.1 ppm and −4.9 ppm, similar to what was observed for SolM1 and SolM2 with respect to SolM0, and follows the same decreasing trend ($\sigma_{s,SolM1}$(**1**) = −3.4 ppm and $\sigma_{s,SolM2}$(**1**) = −6.3 ppm, Table 1). On the other hand, σ_s for **2** oscillated up to ca. 35 ppm. According to what was previously discussed, it is not surprising that σ_s(**2**) = 17.6 ppm when the M-layer is set to 3 Å, or that σ_s(**2**) = −17.0 ppm when the M-layer is 8 Å. In fact, when the solvent is not able to properly reorganize around the solute (M-layer: 3 Å), the rovibrational shielding of **2** is close to the one in the gas phase (σ_{SolM0}(**2**) = 336.7; σ_0(**2**) = 339.3 ppm) and the presence of point charges leads to outlier NMR parameters (σ_s(**2**) = 17.6 ppm) as their effect is not compensated by proper geometrical relaxation. When enough surrounding molecules are free to move during the optimization (M-layer: 8 Å), even if the rovibrational component of the nuclear shielding of **2** is far from the corresponding in the gas phase (σ_{SolM0}(**2**) = 336.7; σ_0(**2**) = 347.2 ppm), the proper reorganization of point charges mitigates its distortion (σ_s(**2**) = −17.0 ppm).

Overall, QM/MM relaxation is shown to provide a better description of solute/solvent structural and electrostatic interactions, leading to a reduction in outlier values for individual structures tested. However, a better sampling of the σ_0 component could improve the results, considering that a single geometry is not representative of a real dynamical system, which is especially true for spectroscopy like NMR, where the sample is able to explore its conformational space during the time of the acquisition of the signal. Therefore, in the next section, we will discuss how dynamic effects affect the computation of NMR parameters.

3.3 | The ensemble effect on chemical shifts and the role of QM/MM geometry optimization

3.3.1 | Nuclear shielding ensemble average

To better describe rovibrational shielding σ_0 and provide a more complete description of the embedding than in SolM2 (see Section 3.2), solute/solvent conformations were sampled by running all-atoms unrestrained MD simulations after re-parametrizing the FF of solutes with their atomic charges computed at the B3LYP/6-31G(d) level of theory. Nuclear shielding values were averaged over 100 uncorrelated geometries and calculated before and after QM/MM optimization, which was performed with an M-layer of 5 Å (see Section 2.4 for more details). The size of the layer was chosen as a good compromise between the freedom of the solvent to properly reorganize around the solute and the computational cost (see Section 3.2). Remarkable differences are found in the computation of nuclear shielding for

both species depending on whether the NMR calculation with GIAO method is performed directly on top of the MD geometries (MD//GIAO protocol) or after QM/MM geometry relaxation of MD snapshots (MD//QM/MM//GIAO protocol) (Tables 3 and 4). The standard deviation of the ^{31}P-NMR nuclear shielding values computed as the mean over 100 geometries strongly decreases after QM/MM relaxation for both **1** (from 8.61 to 3.42 ppm) and **2** (from 14.32 to 4.87 ppm). Analogously, the maximum variation of the ^{31}P-NMR shielding values is reduced from 47.6 ppm to 16.5 ppm for **1** and from 67.7 ppm to 26.9 ppm for **2** (Tables 3 and 4). These results—similar to those already presented in Section 3.2, but in contrast to the observations for SolM0, SolM1, and SolM2—prove that the nuclear shielding value of **2** is much more strongly influenced by the specific solute/solvent cluster conformation and, thus, by the relaxation than that of **1**. This results into the need to sample the solute/solvent for bunch of conformations of **2** as much representative as possible to provide a reliable picture of shielding variations.

The beneficial trend, according to which QM/MM relaxation on top of geometries directly extracted from MD simulation mitigates the fluctuations of the shielding values, is shared by both **1** and **2** compounds, also leading to mean values (computed as averages along the MD trajectory) closer to those computed in SolM2 (**1**: $\sigma_{SolM2} = 301.8$ ppm, $\sigma_{MD//GIAO} = 300.5$ ppm, $\sigma_{MD//QM/MM//GIAO} = 304.0$ ppm; **2**: $\sigma_{SolM2} = 335.7$ ppm, $\sigma_{MD//GIAO} = 323.4$ ppm, $\sigma_{MD//QM/MM//GIAO} = 332.2$ ppm). These outcomes confirm what has already been presented in Section 3.2—that is, that QM/MM relaxation of the solute/solvent clusters improves the prediction of the nuclear shielding value, acting both on σ_0 and σ_s—and disclose that proper conformational sampling is needed. In order to uncover the geometric and electronic reasons for these behaviors, further analyses were performed on **1** (Figure 3). First, the conformational distribution of the phosphaneoxide was investigated by performing KernelPCA analysis over the six O-P-C_x-C_y dihedrals governing the phenyl groups rotations. The results confirmed that the QM/MM relaxation on top of geometries extracted from MD acts as conformational filter grouping conformers (Figure 3A,B). According to the two principal components (PCs), the geometric variability is greater in the snapshots directly extracted from MD simulation when at least seven conformational regions can be identified (Figure 3A) rather than after solute/solvent relaxation when three well-separated bands are reported (Figure 3B). These findings are in perfect agreement with the decreasing standard deviation of the nuclear shielding values, which mirrors a smaller variability of the shielding values, that is,

TABLE 3 Mean value, standard deviation (*s*), standard error of the mean, maximum difference, and bootstrapped standard deviation are presented for shielding values of **1** and **2**, each computed over 100 snapshots of 100 ns long MD simulation, according to the MD//GIAO and the MD//QM/MM//GIAO protocols and the MPW1K/6-311++G(2d,2p) level of theory.

	MD//GIAO					MD//QM/MM//GIAO				
Species	**Mean (σ)**	***s***	**SEM**	**Max (σ)−Min (σ)**	***s* (Bootstrapping)**	**Mean (σ)**	***s***	**SEM**	**Max (σ)−Min (σ)**	***s* (Bootstrapping)**
Vacuum										
1	301.9	8.60	0.860	47.3	0.84	306.3	2.85	0.285	14.0	0.26
2	323.4	14.38	1.438	68.5	1.3	332.0	5.04	0.504	26.4	0.45
PCM										
1	298.3	8.63	0.863	47.9	0.85	303.0	2.92	0.292	14.4	0.28
2	323.9	13.83	1.383	65.7	1.3	332.3	4.74	0.474	25.5	0.52
Explicit MM										
1	300.5	8.61	0.861	47.6	0.84	304.0	3.42	0.342	16.5	0.33
2	323.7	14.32	1.432	67.7	1.3	332.2	4.87	0.487	26.9	0.49

Note: Three different environments were considered: (i) vacuum, after removing solvent molecules, (ii) with PCM after removing solvent molecules, and (iii) explicit solvation with solvent molecules at the MM level.

TABLE 4 Summary of the computed chemical shift for species **1** according to different protocols and different embedding models at the MPW1K/6–311++G(2d,2p) level of theory.

	MD//GIAO				
Embedding	**Gas phase**		**MM**	**PCM**	
δ, ppm	16.1		17.8	20.2	
	MD//QM/MM//GIAO			**MD//QM/MM//GIAO$_{EH}$**	**MD//QM/MM//GIAO$_{EH}$-PCM**
Embedding	**Gas phase**	**MM**	**PCM**	**3 Å QM_2 + MM**	**3 Å QM_2 + PCM**
δ, ppm	20.3	22.8	23.9	24.9	26.2

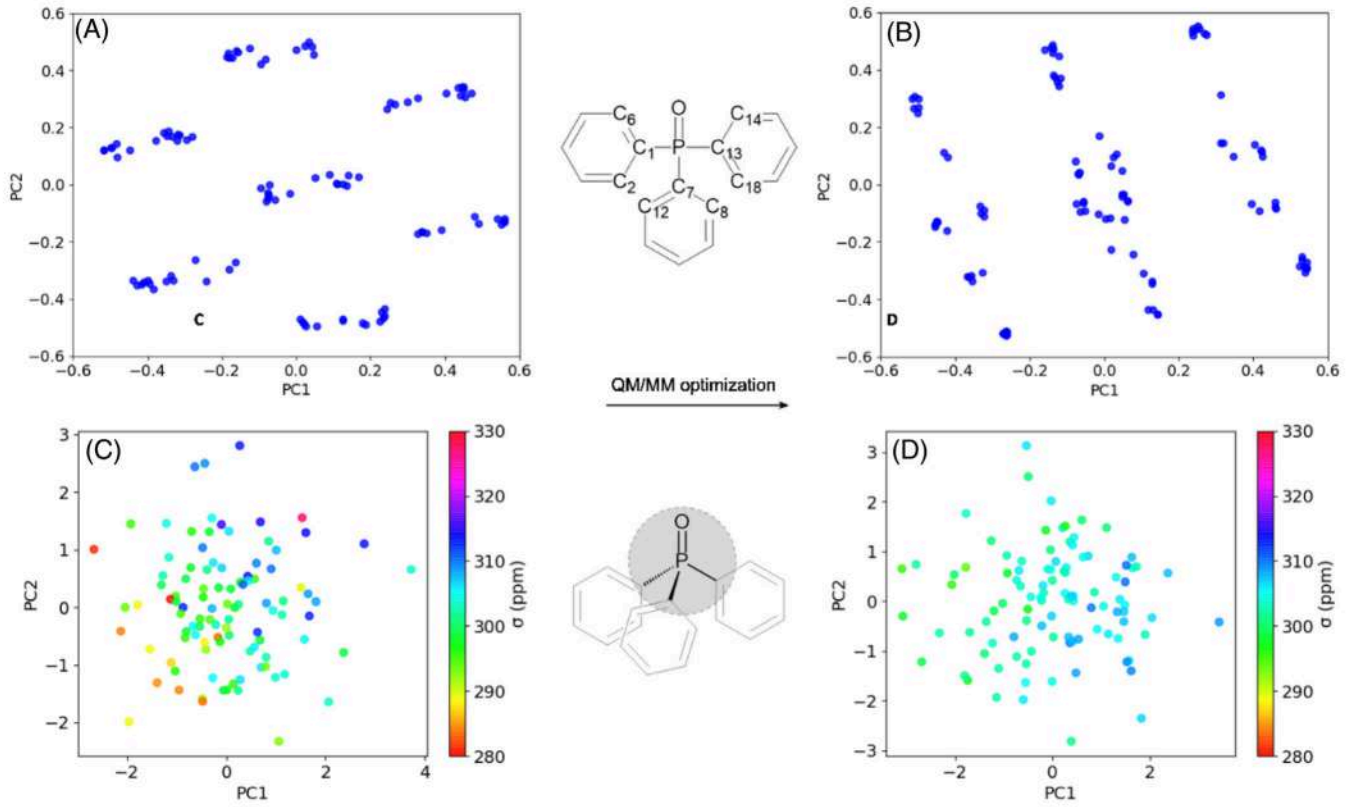

FIGURE 3 Cluster analysis of 100 snapshots of solute (**1**) conformations with kernel PCA considering all six O-P-C_x-C_y dihedral angles ($x = 1$, 7, 13 and $y = 2$, 6, 8, 12, 14, 18) before (A) and after (B) QM/MM optimization. Correlation between isotropic shielding values computed at the MPW1K/6-311++G(2d,2p) level of theory and geometrical descriptors along the first two principal components defined by the four bond lengths of the P atom, before (C) and after (D) QM/MM optimization.

rovibrational fluctuations. The direct correlation between nuclear shielding values and rovibrations was proved by performing a PCA of the lengths of the four nuclei directly bound to phosphorus, and shielding values were reported as a function of the two PCs of each geometry for both the MD//GIAO and the MD//QM/MM//GIAO protocols (Figure 3C,D). In the case of the MD//GIAO protocol, a correlation can be observed as the shielding values increase progressively along the diagonal of the two PCs (Figure 3C). After relaxation, a correlation between the geometry parameters and the shielding values is also recorded but only along the PC1 (Figure 3D). Nevertheless, the correlation is not as strong as for the MD//GIAO protocol: the low and high shielding regions partly overlap, suggesting that the influence of further parameter(s) is now greater and other effects interplay in nuclear shielding calculations. According to what reported above (see Section 3.2), this can be due to the solvent playing a leading role in QM/MM relaxation affecting geometries. This hypothesis is in line with the increasing weight of the solvent contribution σ_s on **1** for the rovibrational shielding value σ_0 when going from the MD//GIAO protocol ($\sigma_s = 1.4$ ppm) to the MD//QM/MM//GIAO one ($\sigma_s = 2.3$ ppm), evaluated by computing the mean shielding values over all the geometries for both protocols, after removing the contribution of explicit solvent molecules (Tables 3 and 4).

3.3.2 | QM/MM relaxation smooths the ensemble average convergence procedure

MD-based conformational sampling provides a statistical ensemble of geometries from which the shielding value can be computed as the mean value along the MD trajectory, showing convergence over the number of (sequential) snapshots considered, as shown in Figure 4. The large oscillations observed in the mean shielding value, calculated directly on top of geometries extracted from MD simulations (i.e., before running QM/MM optimization, Figure 4, blue line), show that the value converges around 300 ppm, with a SEM of ca. 0.85, only after considering 60 uncorrelated geometries. However, when these geometries are further optimized at the QM/MM level, an ensemble of already 30 geometries reproduced a converged shielding value of 304 ppm with a SEM of ca. 0.30 (Figure 4, red line). The QM/MM optimization greatly reduces the number of snapshots required to fully converge to a shielding value over the ensemble and improves the expectation value, as already discussed in Section 3.3.1. However, QM/MM optimization, while beneficial, is a computationally expensive procedure. Indeed, the optimization step becomes more than 9 times computationally more expensive than the computation of NMR parameters (Table S4). As the size of the system increases, the cost of long MD simulations and NMR computations without

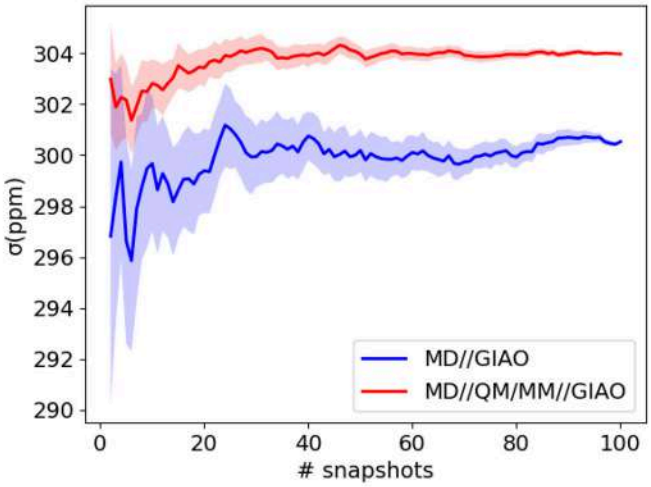

FIGURE 4 Running average of the shielding value (computed at the MPW1K/6-311++G(2d,2p) level of theory) for species **1** over 100 snapshots (every 1 ns) with (red solid line) and without (blue solid line) QM/MM geometry optimizations. Bootstrapped standard deviations over 100 samples are reported in blurred for each simulation.

linear scaling algorithms[33–35] becomes prohibitive and it would be advantageous to carry out QM/MM optimizations for a smaller ensemble.

3.3.3 | QM/MM relaxation and the role of the QM region size

Besides the convergence over the number of selected snapshots, the QM/MM optimization improves the prediction of the chemical shift for **1** by 5.0 ppm (from $\delta_{MD//GIAO} = 17.8$ ppm to $\delta_{MD//QM/MM//GIAO} = 22.8$ ppm, Table 4), definitely showing that geometrical relaxation at the QM level improves the computation of the shielding values. Nevertheless, the chemical shift computed according to the MD//QM/MM//GIAO protocol ($\delta_{MD//QM/MM//GIAO} = 22.8$ ppm, Table 4) is still far from both the experimental value ($\delta_{Exp.} = 29.0$ ppm) and the chemical shift computed *via* the SolM2 approach ($\delta_{SolM2} = 28.5$ ppm, Table 1), pointing out that a better description of the environment is needed. A better (but not satisfactory) chemical shift can already be achieved by replacing solvent molecules with PCM ($\delta_{MD//QM/MM//GIAO\text{-}PCM} = 23.9$ ppm, Table 4), which averages the solvent effect without neglecting the rovibrational sampling in the H-layer, which is guaranteed by MD sampling and QM/MM optimization. As already proved by Ochsenfeld and co-workers, long-range effects computed at QM level can strongly influence the nuclear shielding values and are not well approximated by the MM treatment in hybrid QM/MM methods.[17] As shown above for SolM2 computations, when the addition of an explicit molecule of chloroform at the QM level in close proximity to the solute led to an outcome that, with a difference of 1.8 ppm with respect to SolM1, was in good agreement with the experimental chemical shift (see Section 3.1). To disclose how much such effects can influence the chemical shift of **1**, nuclear shielding values for both the solute and the reference were recomputed according to the MD//QM/MM//GIAO protocol, but now including the first solvent shell (i.e., 3 Å radius) in the H-layer, i.e. at the QM level, namely QM2. The results of enlarging the H-layer (enlarged H-layer, EH) with solvent molecules (MD//QM/MM//$GIAO_{EH}$ protocol) show that there is an improvement in chemical shifts both when the rest of the solvent is described as point charges ($\delta = 24.9$ ppm, Table 4) and when it is replaced by PCM (MD//QM/MM//$GIAO_{EH}$-PCM protocol, $\delta = 26.2$ ppm, Table 4), but this leads to a dramatic increase in computational time. Indeed, the NMR calculations that include 3 Å of solvent at the QM level are 5.4 times more expensive than analogous calculations that only consider the solute in the H-layer (Table S4), making it prohibitive to further increase the size of the H-layer with the chosen basis set. Nevertheless, due to the importance of such a result, in the following we have investigated the effects of the size of the solvent shell included in the QM region more in detail.

3.4 | Convergence of the NMR shielding calculation increasing the size of QM region

The systematic study of the convergence of the ^{31}P-NMR shielding values as a function of the size of the H-layer was performed removing the solvent molecules not included in the QM region and replacing them with PCM. Since the computational cost of the QM calculations increases drastically with the size of the H-layer, a different basis set was used to calculate the magnetic shielding values in this case. We chose the pcS-1 basis set proposed by Jensen, which has been shown to be reliable and efficient in various studies.[31,89] Indeed, the pcS-1 basis set leads to faster computations than the 6-311++G(2d,2p) one: the computational time required for an NMR calculation of **1** without the addition of any explicit molecule of solvent in the H-layer is 0.17 times the CPU time required for the same calculation done with the 6-311++G(2d,2p) basis set (Table S4). The computed value of chemical shift with pcS-1 is consistent with those adopting a larger basis set: for example, the chemical shift computed without adding any solvent molecule in the H-layer (i.e., H-layer of 0 Å), is 20.3 ppm with the 6-311++G(2d,2p) basis set (Table 4) and 22.0 ppm with the pcS-1 basis set. The use of this basis set makes our investigations of convergence of the nuclear shielding value as a function of the size of the H-layer feasible also without using linear scaling methods.

As shown in Table S3, the convergence of the nuclear shielding values is achieved by adding a few solvent molecules (i.e., H-layer of 4 Å) both for **1** and **2**, with the magnitudes close to those calculated with all solvent molecules at the MM level: $\sigma_{0Å}(\mathbf{1}) = 329.8$ ppm, $\sigma_{4Å}(\mathbf{1}) = 328.0$ ppm, $\sigma_{0Å}(\mathbf{2}) = 357.2$ ppm, $\sigma_{4Å}(\mathbf{2}) = 358.6$ ppm. Even though the nuclear shielding does not vary more than 1.8 ppm during the convergence procedure and only small changes in the nuclear parameters are observed, a non-negligible improvement in the target chemical shift is achieved as the H-layer size increases, moving

from 22.0 to 25.2 ppm and approaching the experimental value ($\delta_{Exp.}$ = 29.0 ppm), confirming the importance of treating a sufficiently large solvent shell at the QM level.

Since it is known that the convergence of nuclear shielding values is generally achieved not as quickly as observed for the chemical shifts of compounds **1** and **2**, but large fluctuations in shielding values can be observed as the size of the H-layer increases,[16,17] we have extended the study to another system, that is, trimethyl phosphite (compound **3**, Figure 1B) in $CHCl_3$ solution. The trimethyl phosphite molecule **3** is small and conformationally very flexible,[15] with the phosphorus directly bonded to three equivalent oxygen atoms exposed to solvent molecules that can interact *via* both $CCl_3H\cdots O$ and $CCl_3H\cdots P$ hydrogen bonds. The simultaneous presence of flexibility and intermolecular interactions, which can strongly influence the electron density and thus the nuclear shielding values, makes **3** an excellent candidate for a complementary study of the shielding convergence. As a first step, we applied the MD//QM/MM protocol to sample solute/solvent cluster conformations and to perform the geometries relaxations. Convergence studies in terms of the number of snapshots, which do not require involvement of large QM regions and thus small basis sets, before and after QM/MM optimization were first performed using the 6-311++G(2d,2p) basis set and electrostatic embedding. The results show that without relaxation, no converged value could be achieved despite considering 100 uncorrelated geometries (Figure S5, blue line). However, upon QM/MM geometry optimization, as few as 30 optimized geometries are sufficient to converge in a stable magnetic shielding value (Figure S5, red line), showing great improvement in computational cost since the number of geometries required gets reduced by 70%. On top of optimized geometries, the NMR parameters were then calculated using the NMR-PCM_{EH} protocol, i.e., the MM-level explicit solvent was replaced by PCM, and the H-layer size was gradually increased employing the pcS-1 basis set. As expected, the convergence of nuclear shielding over 100 averaged geometries is reached when a consistent number of solvent molecules is included in the H-layer, that is, a solvent shell of a 4 Å radius. Furthermore, the converged value for an H-layer with a radius of 4 Å is 3.7 ppm larger than the value calculated without including any chloroform molecule (Table 5), confirming the non-negligible influence of the long-range QM interactions between **3** and the solvent. It is worth to highlight that, even though the pcS-1 basis set speeds up calculations, performing many NMR computations on models involving large QM regions is still very expensive, as long as linear scaling models are not employed.[33–35] Therefore, reducing the number of geometries to be considered significantly reduces the number of NMR computations and the overall computational cost. As discussed above, also for **3** the QM/MM optimization of the first shell of the solvent limits not only the fluctuations but also the convergence of the screening value as a function of the number of snapshots considered. Table 5 shows the fluctuations of the magnetic shielding value of **3** as a function of the number of uncorrelated relaxed geometries considered and the size of the solvent shell included in the H-layer. The results show that after QM/MM relaxation, the number of uncorrelated geometries does not have a large influence on the magnetic shielding value, which fluctuates in a spectral window of ca. 1 ppm. On the other hand, the size of the layer has a large influence on the absolute value of the shielding of up to 3.8 ppm (from 0 to 5 Å considering 100 geometries, Table 5). Overall, the total spectral shift is 4.9 ppm from a value averaged over 25 uncorrelated geometries without solvent molecules in the H-layer to the value calculated over 100 geometries with 5 Å solvent shell (Table 5). The optimal combination is, thus, 50 uncorrelated MD geometries and a 4 Å solvent shell at QM level, pointing out the relevant reduction of the number of MD snapshots selection when QM/MM optimization is performed.

TABLE 5 Nuclear shielding values (reported in ppm) for species **3** as a function of an increasing number of snapshots (added in sequence) and size of the H-layer after QM/MM relaxation.

	# Snapshots			
QM_2 (Å)	25	50	75	100
0	185.4	186.2	186.0	186.5
3	187.1	187.9	187.7	188.2
4	188.9	189.9	189.7	190.2
5	189.0	190.1	189.8	190.3

Note: Shielding values were computed at the MPW1K/pcS-1 level of theory.

3.5 | Summary and discussion

The automatized protocols described in the present work show the comparison of different computational approaches to simulate the ^{31}P-NMR shielding and chemical shifts. The thorough performance analyses presented, which highlight the strengths and weaknesses of the protocols considered, allow users to consciously choose the best compromise between accuracy and computational cost (Figure 5).

Using triphenylphosphine oxide **1** as a case study, we showed that one of the key elements to predict the ^{31}P-NMR chemical shifts is to carefully sample the rovibrations and solute/solvent clusters conformations *via* MD simulations. Since the NMR spectra result from the average of different experimental conformations of the analyte, sampling different solute/solvent clusters is a key ingredient to return a more accurate description of the phenomenon. The MD//GIAO protocol (Section 3.3.1) is computationally not very expensive in itself but requires the sampling of many geometries to achieve converged values of the magnetic shielding values, and the computed chemical shifts are not very accurate ($\delta_{Exp.}$ = 29.0 ppm vs. $\delta_{MD//GIAO\text{-}PCM}$ = 20.2 ppm, Table 4). Significant improvements in both the reduction of number of sampled geometries needed for reaching convergence and the chemical shift value ($\delta_{MD//QM/MM//GIAO\text{-}PCM}$ = 24.8 ppm, Table 4) can be achieved if the QM/MM optimization is performed on geometries extracted from MD simulations (MD//QM/MM//GIAO protocol, see Sections 3.3.2 and 3.3.3). The QM/MM optimization is, thus, a strategy to 'refine' the MD

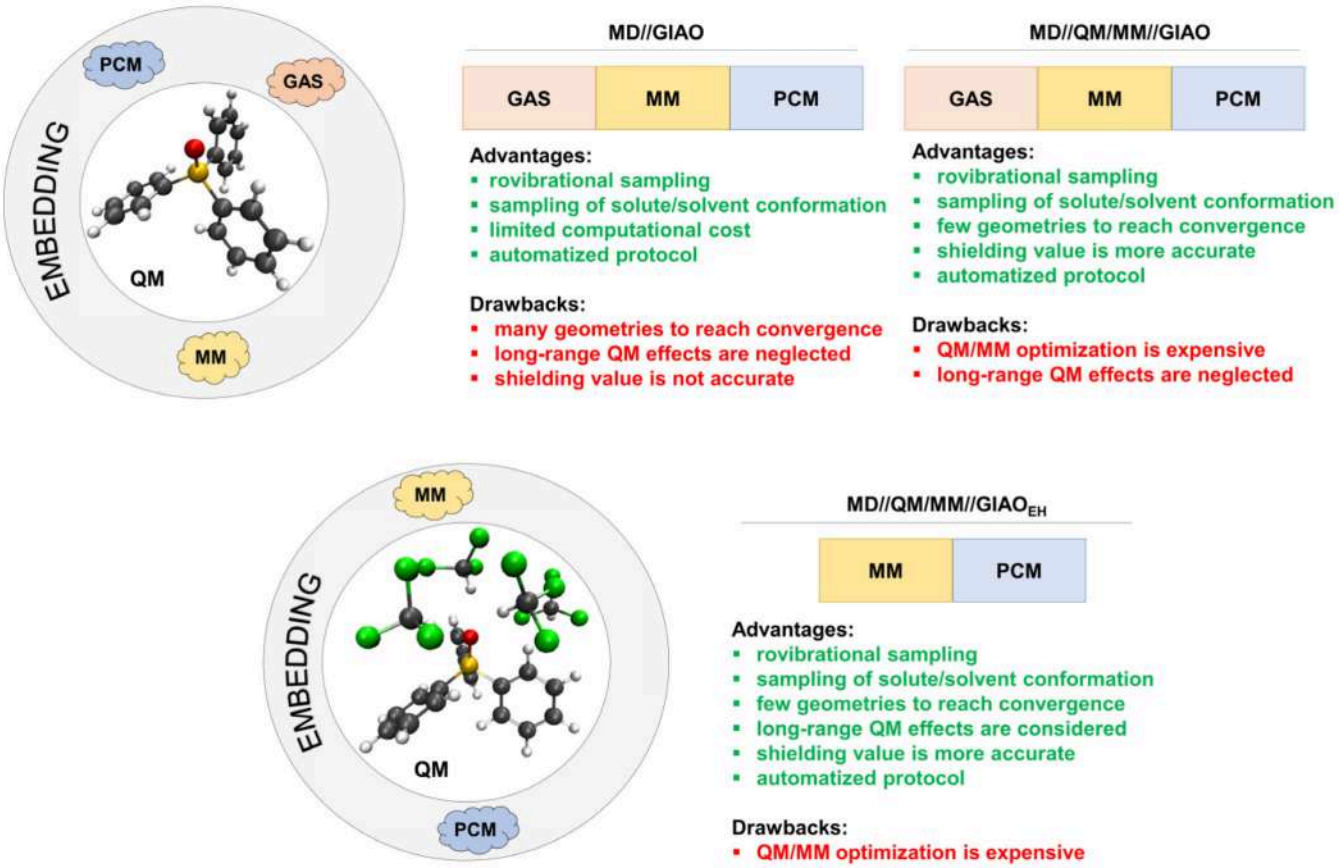

FIGURE 5 Summary of improvements in computational prediction of ^{31}P-NMR chemical shifts with increasing solvent description level and comparison of MD//GIAO, MD//QM/MM//GIAO, and MD//QM/MM//GIAO$_{EH}$ protocols.

sampling. In fact, when the solute/solvent clusters are relaxed at a higher level of theory, the quality of the sampled geometries is improved by returning a more reliable picture of the interactions occurring between the analyte and the surrounding. The MD//QM/MM//GIAO protocol leads to a better result but with an increased computational cost: the costs for the optimization are more than 9 times higher than that of a single-point NMR calculation (Table S4).

The way the embedding is described also influences the computed chemical shift value, improving the accuracy, with the trend being: gas-phase < MM < PCM (Table 4), for all protocols (Figure 5). However, to obtain a better description of the long-range QM effects between the solute and the solvent, a sufficiently large solvent shell should be treated at the QM level, and the environment could be considered as point charges ($\delta = 24.9$ ppm, Table 4) or with a PCM ($\delta = 26.2$ ppm, Table 4). Increasing the size of the H-layer is also a computationally expensive procedure (i.e., a single point NMR calculation is 5.4 times more expensive if a 3 Å shell of the solvent is included, Table S4). The good accuracy is achieved at a high computational cost (Figure 5).

To summarize, the improvement towards experimental values associated with the MD//QM/MM//GIAO$_{EH}$-PCM approach comprises different ingredients: (i) sampling many solute/solvent clusters through MD simulations averages the shielding value over many rovibrational conformations, similarly to what happens during a real experiment; (ii) the quality of the geometry of these samples is refined using QM/MM optimizations, which improve the description of the solute/solvent interactions; (iii) the long-range QM effects from the environment on the shielding value of the nucleus of interest are added by including an increasing number of solvent molecules until reaching the convergence; and finally, (iv) the bulk effect of the solvent on the solute/solvent clusters is approximated with an implicit solvent model.

4 | CONCLUSIONS

In this work, we systematically investigated four different levels of environmental modeling (vacuum, implicit solvation with PCM, explicit microsolvation plus PCM, and macrosolvation using both QM/MM and QM-PCM schemes) for computing the ^{31}P-NMR chemical shift of triphenylphosphine oxide (**1**), using triphenylphosphine (**2**) as reference. We showed that classical MD simulations are crucial for capturing a sufficiently big ensemble of solute/solvent conformations, mimicking the variety of possible electrostatic and van der Waals interactions in solute/solvent clusters, but classical MD alone does not lead to reliable geometries for the calculation of ^{31}P-NMR isotropic

shielding values. The latter can be substantially improved if QM/MM optimizations (with electrostatic embedding) are carried out on geometries extracted from the MD trajectories and a sufficiently large solvent shell is treated at the QM level (as the molecule of interest), while the rest of the solvent is treated with the PCM implicit model ($GIAO_{EH}$-PCM). Furthermore, we found that QM/MM geometry optimizations improve the convergence of the shielding values, reducing the number of geometries extracted from MD trajectories to be considered in NMR computations. This was further demonstrated by applying our MD//QM/MM//$GIAO_{EH}$-PCM protocol on the trimethyl phosphite molecule (**3**), a relevant test case featuring high flexibility and multiple intermolecular interactions with solvent molecules. Our automatic protocol for the calculation of nuclear shielding values, is implemented and freely available in the COBRAMM software package. Still, considering the limitation of this work, given the small set of molecules considered, the focus on one nucleus, that is, ^{31}P, and the specific choice of the DFT methodology employed, the reported results are really valuable but, certainly, they could not provide a universal trend. Notably, it is worth mentioning that, here, we selected a single QM level of theory in order to focus on the environmental modeling effects but future studies should compare different QM levels and confirm the role of QM/MM optimizations highlighted in this study. Nevertheless, the outcome of this work opens to new (more extensive) investigations on the prediction of the ^{31}P-NMR chemical shifts (and potentially those of other nuclei) with an approach that reduces the computational time. In particular, it paves the way towards the simulation of supramolecular systems, macromolecules, and biological systems, where intricate environmental effects call for the use of QM/MM hybrid methods while requiring a reasonable computational cost.

ACKNOWLEDGMENTS

IR gratefully acknowledges the use of HPC resources of the "Pôle Scientifique de Modélisation Numérique" (PSMN) of the ENS-Lyon, France. FC and IR thank Prof. Cristiana Cesari from University of Bologna for the help in recording NMR experimental spectra. BM expresses gratitude to the University of Vienna for generously funding his research program. Prof. Nuno Maulide and Prof. Leticia González from the University of Vienna are gratefully acknowledged for their support and valuable discussions.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY STATMENT

New implementations in COBRAMM and codes about principal component analysis are available on GitLab at the following address: https://gitlab.com/cobrammgroup/cobramm/-/tree/cobramm-NMR.

ORCID

Francesco Calcagno https://orcid.org/0000-0002-0986-4425
Marco Garavelli https://orcid.org/0000-0002-0796-289X
Davide Avagliano https://orcid.org/0000-0001-5539-9731
Ivan Rivalta https://orcid.org/0000-0002-1208-602X

REFERENCES

[1] M. Pons, *NMR in supramolecular chemistry*, Springer Science & Business Media, Berlin **1999**.
[2] I. Bertini, C. Luchinat, G. Parigi, *Coord. Chem. Rev.* **2011**, *255*, 649.
[3] M. P. M. Letertre, P. Giraudeau, P. de Tullio, *Front. Mol. Biosci.* **2021**, *8*, 8.
[4] J. R. Cheeseman, G. W. Trucks, T. A. Keith, M. J. Frisch, *J. Chem. Phys.* **1996**, *104*, 5497.
[5] J. C. Facelli, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 176.
[6] J. Gauss, *Mod. Methods Algorithms Quant. Chem.* **2000**, *3*, 541.
[7] J. Gauss, *Ber. Bunsenges. Phys. Chem.* **1995**, *99*, 1001.
[8] T. Helgaker, M. Jaszuński, K. Ruud, *Chem. Rev.* **1999**, *99*, 293.
[9] F. London, *J. Phys. Radium* **1937**, *8*, 397.
[10] R. Ditchfield, *Mol. Phys.* **1974**, *27*, 789.
[11] K. Wolinski, J. F. Hinton, P. Pulay, *J. Am. Chem. Soc.* **1990**, *112*, 8251.
[12] A. Bagno, F. Rastrelli, G. Saielli, *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *47*, 41.
[13] G. E. Maciel, G. C. Ruben, *J. Am. Chem. Soc.* **2002**, *85*, 3903.
[14] F. V. Toukach, V. P. Ananikov, *Chem. Soc. Rev.* **2013**, *42*, 8376.
[15] B. Maryasin, H. Zipse, *Phys. Chem. Chem. Phys.* **2011**, *13*, 5150.
[16] C. Steinmann, J. Magnus Haugaard Olsen, J. Kongsted, *J. Chem. Theory Comput.* **2014**, *10*, 981.
[17] D. Flaig, M. Beer, C. Ochsenfeld, *J. Chem. Theory Comput.* **2012**, *8*, 2260.
[18] G. Scalmani, M. J. Frisch, *J. Chem. Phys.* **2010**, *132*, 114110.
[19] I. Ciofini, *Calculation of NMR and EPR Parameters*, John Wiley & Sons, Ltd, **2004**, Ch. 12. https://onlinelibrary.wiley.com/doi/abs/10.1002/3527601678.ch12
[20] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.
[21] H. C. Da Silva, W. B. De Almeida, *Chem. Phys.* **2020**, *528*, 110479.
[22] I. Ciofini, *Magn. Reson. Chem.* **2004**, *42*, S48.
[23] C. Adamo, M. Cossi, V. Barone, *J. Mol. Struct. Theochem.* **1999**, *493*, 145.
[24] B. Mennucci, J. M. Martínez, *J. Phys. Chem. B* **2005**, *109*, 9830.
[25] M. Cossi, O. Crescenzi, *J. Chem. Phys.* **2003**, *118*, 8863.
[26] A. Baggioli, O. Crescenzi, M. J. Field, F. Castiglione, G. Raos, *Phys. Chem. Chem. Phys.* **2013**, *15*, 1130.
[27] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding, K. Morokuma, *Chem. Rev.* **2015**, *115*, 5678.
[28] S. Caprasecca, L. Cupellini, S. Jurinovich, D. Loco, F. Lipparini, B. Mennucci, *Theor. Chem. Acc.* **2018**, *137*, 84.
[29] B. Mennucci, *Phys. Chem. Chem. Phys.* **2013**, *15*, 6583.
[30] B. Mennucci, J. M. Martínez, J. Tomasi, *J. Phys. Chem. A* **2001**, *105*, 7287.
[31] J. Fukal, O. Páv, M. Buděšínský, I. Rosenberg, J. Šebera, V. Sychrovský, *Phys. Chem. Chem. Phys.* **2019**, *21*, 9924.
[32] J. C. B. Dietschreit, A. Wagner, T. A. Le, P. Klein, H. Schindelin, T. Opatz, B. Engels, U. A. Hellmich, C. Ochsenfeld, *Angew. Chem. Int. Ed.* **2020**, *59*, 12669.
[33] C. Ochsenfeld, J. Kussmann, F. Koziol, *Angew. Chem. Int. Ed.* **2004**, *43*, 4485.
[34] J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2007**, *127*, 54103.
[35] M. Beer, C. Ochsenfeld, *J. Chem. Phys.* **2008**, *128*, 221102.
[36] L. B. Krivdin, *Magn. Reson. Chem.* **2020**, *58*, 478.
[37] J. Fukal, M. Buděšínský, O. Páv, P. Jurečka, M. Zgarbová, J. Šebera, V. Sychrovský, *J. Comput. Chem.* **2022**, *43*, 132.
[38] J. Přecechtělová, M. L. Munzarová, J. Vaara, J. Novotný, M. Dračínský, V. Sklenář, *J. Chem. Theory Comput.* **2013**, *9*, 1641.
[39] S. K. Latypov, S. A. Kondrashova, F. M. Polyancev, O. G. Sinyashin, *Organometallics* **2020**, *39*, 1413.

[40] P.-A. Payard, L. Alessandro Perego, L. Grimaud, I. Ciofini, *Organometallics* **2020**, *39*, 3121.
[41] P. S. Pregosin, R. W. Kunz, *31P and ^{13}C NMR of Transition Metal Phosphine Complexes*, Springer Science & Business Media, Berlin **2012**.
[42] O. Kühl, *Phosphorus-31 NMR Spectroscopy: A Concise Introduction for the Synthetic Organic and Organometallic Chemist*, Springer Science & Business Media, Berlin **2008**.
[43] S. Fu, Z. Shao, Y. Wang, Q. Liu, *J. Am. Chem. Soc.* **2017**, *139*, 11941.
[44] D. S. Glueck, *Coord. Chem. Rev.* **2011**, *255*, 356.
[45] M. Hesse, H. Meier, B. Zeeh, *Spectroscopic Methods in Organic Chemistry*, Georg Thieme Verlag, New York **2014**.
[46] S. K. Latypov, F. M. Polyancev, D. G. Yakhvarov, O. G. Sinyashin, *Phys. Chem. Chem. Phys.* **2015**, *17*, 6976.
[47] K. A. Chernyshev, L. B. Krivdin, *Russ. J. Org. Chem.* **2011**, *47*, 355.
[48] K. A. Chernyshev, L. I. Larina, E. A. Chirkina, V. G. Rozinov, L. B. Krivdin, *Russ. J. Org. Chem.* **2011**, *47*, 1865.
[49] K. A. Chernyshev, L. I. Larina, E. A. Chirkina, L. B. Krivdin, *Magn. Reson. Chem.* **2012**, *50*, 120.
[50] G. Begimova, E. Y. Tupikina, V. K. Yu, G. S. Denisov, M. Bodensteiner, I. G. Shenderovich, *J. Phys. Chem. C* **2016**, *120*, 8717.
[51] K. A. Chernyshev, L. I. Larina, E. A. Chirkina, V. G. Rozinov, L. B. Krivdin, *Russ. J. Org. Chem.* **2011**, *47*, 1859.
[52] J. Fukal, O. Páv, M. Buděšínský, J. Šebera, V. Sychrovský, *Phys. Chem. Chem. Phys.* **2017**, *19*, 31830.
[53] J. Přecechtělová, P. Novák, M. L. Munzarová, M. Kaupp, V. Sklenář, *J. Am. Chem. Soc.* **2010**, *132*, 17139.
[54] P. Altoè, M. Stenta, A. Bottoni, M. Garavelli, *Theor. Chem. Acc.* **2007**, *118*, 219.
[55] O. Weingart, A. Nenov, P. Altoè, I. Rivalta, J. Segarra-Martí, I. Dokukina, M. Garavelli, *J. Mol. Model.* **2018**, *24*, 271.
[56] H. Lin, D. G. Truhlar, *Theor. Chem. Acc.* **2007**, *117*, 185.
[57] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.
[58] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian, Inc., Wallingford CT. **2016**.
[59] B. J. Lynch, P. L. Fast, M. Harris, D. G. Truhlar, *J. Phys. Chem. A* **2000**, *104*, 4811.
[60] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668.
[61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.
[62] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724.
[63] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, *J. Chem. Phys.* **1982**, *77*, 3654.
[64] P. C. Hariharan, J. A. Pople, *Mol. Phys.* **1974**, *27*, 209.
[65] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213.
[66] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.
[67] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294.
[68] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **1984**, *80*, 3265.
[69] A. D. McLean, G. S. Chandler, *J. Chem. Phys.* **1980**, *72*, 5639.
[70] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.
[71] D. Avagliano, M. Bonfanti, A. Nenov, M. Garavelli, *J. Comput. Chem.* **2022**, *43*, 1641.
[72] H. B. Curry, *Q. Appl. Math.* **1944**, *2*, 258.
[73] E. Stiefel, *J. Res. Nat. Bur. Stand.* **1952**, *49*, 409.
[74] J. Liu, D. Li, X. Liu, *J. Chem. Phys.* **2016**, *145*, 24103.
[75] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.
[76] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577.
[77] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
[78] R. C. Walker, M. F. Crowley, D. A. Case, *J. Comput. Chem.* **2008**, *29*, 1019.
[79] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[80] A. D. Becke, *J. Chem. Phys.* **1992**, *97*, 9173.
[81] A. D. Becke, *J. Chem. Phys.* **1992**, *96*, 2155.
[82] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
[83] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[84] P. Cieplak, J. Caldwell, P. Kollman, *J. Comput. Chem.* **2001**, *22*, 1048.
[85] F. Jensen, *J. Chem. Theory Comput.* **2008**, *4*, 719.
[86] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, T. L. Windus, *J. Chem. Inf. Model.* **2007**, *47*, 1045.
[87] D. Feller, *J. Comput. Chem.* **1996**, *17*, 1571.
[88] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, T. L. Windus, *J. Chem. Inf. Model.* **2019**, *59*, 4814.
[89] D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville, C. Ochsenfeld, *J. Chem. Theory Comput.* **2014**, *10*, 572.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: F. Calcagno, B. Maryasin, M. Garavelli, D. Avagliano, I. Rivalta, *J. Comput. Chem.* **2024**, *45*(18), 1562. https://doi.org/10.1002/jcc.27338