Molecular Dynamics Simulations Enforcing Nonperiodic Boundary Conditions: New Developments and Application to the Solvent Shifts of Nitroxide Magnetic Parameters

Giordano Mancini,* Marco Fusè, Filippo Lipparini, Michele Nottoli, Giovanni Scalmani, and Vincenzo Barone*

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Multiscale methods combining quantum mechanics and molecular mechanics (QM/MM) have become the most suitable and effective strategies for the investigation of the spectroscopic properties of medium-to-large size chromophores in condensed phases. In this context, we are developing a novel workflow aimed at improving the generality, reliability, and ease of use of the available computational tools. In this paper, we report our latest developments with specific reference to a general protocol based on atomistic simulations, carried out under nonperiodic boundary conditions (NPBC). In particular, we add to our in house MD engine a new efficient treatment of mean field electrostatic contributions to energy and forces, together with the capability of performing the simulations either in the canonical (*NVT*) or in the isothermal–isobaric (*NPT*) ensemble. Next, we provide convincing evidence that the NBPC approach enhanced by specific tweaks for rigid body propagation, allows for the simulation of solute–solvent systems with a minimum number of degrees of freedom and large integration time step. After its validation, this new approach is applied to the challenging case of solvatochromic effects on the electron paramagnetic resonance (EPR) spectrum of a prototypical nitroxide radical. To this end, we propose and validate also an automated protocol to extract and weight simulation snapshots, making use of a continuous description of the strength of solute–solvent hydrogen bridges. While further developments are being worked on, the effectiveness of our approach, even in its present form, is demonstrated by the accuracy of the results obtained through an unsupervised approach characterized by a strongly reduced computational cost as compared to that of conventional QM/MM models, without any appreciable deterioration of accuracy.
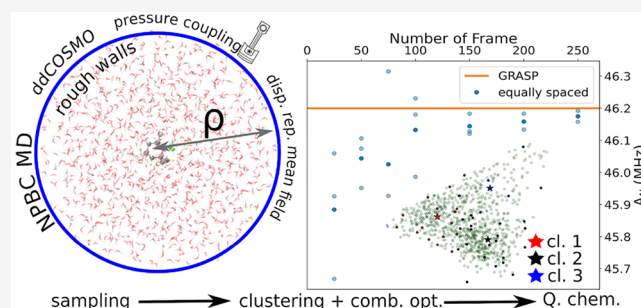
INTRODUCTION

A general, robust, and effective strategy for the computation of spectroscopic parameters in solution can be broken down in a number of successive steps. The first step is the generation of a sufficient number of representative structures by means of atomistic simulations, such as Monte Carlo (MC) or Molecular Dynamics (MD). Next, suitable clustering procedures can be employed to reduce the number of different structures needed to obtain a statistically significant ensemble, able to provide well converged average results. Finally, reliable yet effective multiscale methods combining quantum mechanics and molecular mechanics (QM/MM) can be employed to compute averaged values of the desired properties.[1] While in principle all steps can be performed at the same time by extended Lagrangian[2,3] or Born–Oppenheimer[4,5] *ab initio* molecular dynamics, both the space and time scales which can be covered at present by these approaches are too small to deliver converged results for medium-to-large size chromophores. Furthermore, the accuracy required in the computation of the spectroscopic parameters is usually much higher than

that needed in the generation of the representative snapshots. Alternatively, quite simple molecular mechanics (MM) force fields are now available for describing solvent dynamics and also solute–solvent interactions can be usually modeled quite effectively by well-tested force fields. The situation is however different for the computation of spectroscopic parameters, where a quantum mechanical description of the solute (and possibly its cybotactic region) is unavoidable and solvent polarization can have a non-negligible effect.[6] A companion paper deals with an effective strategy for this last step, exploiting a QM/MM polarizable model based on the floating charge (FQ) approach,[7] including effective recipes for taking

A

into account solvent–solvent charge-transfer,[8] and correcting the wrong coordination numbers of solvent molecules near the outer boundary of the simulation box.[9] Here, we will focus on the initial atomistic simulation and the clustering of the resulting trajectories in the framework of our computational engine for classical molecular dynamics simulations employing a spherical simulation box and nonperiodic boundary conditions (NPBC).[10,11] In particular, we will describe a new quasi-analytical implementation of reaction field effects at the outer boundary of the simulation sphere by means of a series expansion of the polarization density, in the spirit of the domain decomposition conductor-like screening model (ddCOSMO),[12,13] for MD simulations employing either the canonical (NVT) or the isothermal–isobaric (NPT) ensemble. In the latter case, a new thermostat and an effective treatment of the simulation sphere breathing have been introduced. After parametrization and simulation of an effective water model (TIP3P-FB[14]), the overall computational strategy has been applied to a quite demanding case study, namely the tuning of the magnetic properties of a representative nitroxide radical (TEMPO, i.e., (2,2,6,6-tetramethylpiperidin-1-yl)oxidanyl) by different solvents.[15] To this end, we have performed a preliminary benchmark study for the model dimethyl nitroxide radical, taking as references the results obtained by state-of-the-art QM calculations performed specifically for this system. After validating the QM model, the new NPBC molecular dynamics engine is used to generate a statistically significant number of snapshots to take into account solvent modulation effects. Extending the pipeline commonly used to select significant structures or subtrajectories from MD simulations,[16–18] we adopt a two-step procedure: First, we define distinct basins using a clustering[19] procedure in which optimal clusters are found by means of an optimization of validity scores. Next, following a strategy reminiscent of that employed with success for electronic spectra,[11] a representative structure for each basin is treated at the QM/MM level, including in the QM subsystem a few solvent molecules, whereas fluctuations within each basin are taken into account by computing a sufficient number of structures with a cheaper approach in which the QM part is reduced to just the solute. The whole procedure can be cast in terms of a fully unsupervised workflow, so that, together with the intrinsic interest of the specific system, the validation of such protocol paves the route toward the study of several other problems of current technological or biological interest.

METHODS

NPBC Simulations. The framework of our computational strategy is the GLOB model[10] in which a spherical cavity with rigid boundary contains the proper number of solvent molecules needed to enforce an average density equal to the experimental counterpart. Optionally, a solute can be also placed at the center of the cavity, removing the solvent molecules within van der Waals distances from any of its atoms. Next, two soft potentials are added in order to describe the reaction field of the solvent outside the cavity (U_{RF}) and to enforce a nearly constant density in the whole simulation sphere (U_{vdW}).

Reaction Field for a Spherical Cavity. In this section, the main equations that describe the reaction field contribution for a spherical cavity are presented, and their implementation is discussed in some detail. Let Ω be a spherical cavity of radius R centered at C_0 and $\Gamma = \partial\Omega$ its boundary. Given an electrostatic

distribution ρ fully supported inside Ω, we want to solve the Conductor Polarizable Continuum Model[20] (C-PCM) or, equivalently, of the Conductor-like Screening Model[21] (COSMO) electrostatic equation

$$(S\sigma)(s) = \int_\Gamma \frac{\sigma(s')}{|s - s'|}\, ds' = -\Phi(s) \tag{1}$$

where σ is an apparent surface charge (ASC) supported on Γ and Φ is the molecular electrostatic potential generated by the solute's charge distribution (q_k's at positions r_k's)

$$\Phi(s) = \sum_k \frac{q_k}{|C_0 + Rs - r_k|} \tag{2}$$

Given the spherical symmetry of the problem, it is convenient to expand both the potential and the ASC in spherical harmonics, following the procedure introduced in the ddCOSMO algorithm.

$$\sigma(s) = \frac{1}{R^2}\sum_{l=1}^{L_{max}} \sum_{m=-l}^{l} \sigma_l^m Y_l^m(\theta, \phi),$$
$$\Phi(s) = \frac{1}{R}\sum_{lm} \Phi_l^m Y_l^m(\theta, \phi) \tag{3}$$

There we introduced a compact notation for the l, m sum and L_{max} is the angular momentum to which we truncate the expansion. Note that, with respect to previous publications of the ddCOSMO method, we have introduced a slightly different scaling of the potential and ASC with respect to the cavity radius.

Using the spherical harmonics addition theorem, the COSMO integral equation becomes

$$\frac{4\pi}{2l + 1}\sigma_l^m = -\Phi_l^m \tag{4}$$

The solvation energy can then be computed as

$$E_s = \frac{1}{2R}f(\epsilon)\sum_{lm}\sigma_l^m\Phi_l^m = -\frac{1}{2R}f(\epsilon)\sum_{lm}\frac{2l+1}{4\pi}(\Phi_l^m)^2 \tag{5}$$

The spherical harmonics expansion coefficients of the potential needed to solve eq 5 can be easily computed using a quadrature, with the Lebedev–Laikov (LL) ansatz being particularly well suited due to the symmetry of the problem. Let w_n, s_n be the LL weights and points:

$$\Phi_l^m = \frac{1}{R}\int_\Gamma \Phi(s)Y_l^m(s)\, ds \approx R\sum_n w_n\Phi(s_n)Y_l^m(s_n) \tag{6}$$

The contributions to the forces are obtained by straightforward differentiation of eq 5. Let

$$\sigma_n = w_n\sum_{lm}\sigma_l^m Y_l^m(s_n) \tag{7}$$

We get, on atom k,

$$F_k = -\frac{\partial E_s}{\partial r_k} = q_k\sum_n \frac{\sigma_n(r_k - C_0 - Rs_n)}{|r_k - C_0 - Rs_n|^3} = q_k E_k(\sigma) \tag{8}$$

where we can recognize the electric field produced by σ_n distribution at the kth site.

For simulations at constant pressure, the derivative of the solvation energy with respect to the volume is needed. The latter can also be easily computed by differentiating eq 5 with respect to $V = 4/3\pi R^3$:

$$\frac{dE_s}{dV} = \frac{1}{4\pi R^2}\frac{dE_s}{dR}$$

$$= \frac{1}{4\pi R^2}\left(\frac{E_s}{R} - \sum_n \sum_k q_k \sigma_n \frac{(C_0 + Rs_n - r_k)\cdot s_n}{|C_0 + Rs_n - r_k|^3}\right) \quad (9)$$

where we can rewrite the last term by recognizing the electric field produced by the solute's charge distribution at the cavity:

$$\frac{dE_s}{dV} = \frac{1}{4\pi R^2}\left(\frac{E_s}{R} - \sum_n \sigma_n E_n(q)\cdot s_n \right) \quad (10)$$

From a computational point of view, the only quadratic scaling terms in all the equations correspond to the evaluation of the electrostatic potential or field of a charge distribution, which can be performed at a cost that scales linearly with the number of atoms by using a fast summation technique. In our implementation, in particular, we use the Fast Multipole Method.[22,23] The evaluation of Spherical Harmonics at the Lebedev points, which is also required in various parts of the computation flow, is achieved as extensively documented in ref 24. For the sake of brevity, in the following, we will refer to the electrostatic solvation model simply as ddCOSMO, as it indeed corresponds to a ddCOSMO treatment of a single spherical cavity.

Isothermal Isobaric Ensemble and Weak Coupling. In this study, we have adopted the simplest available pressure control algorithm, namely the so-called Berendsen barostat[25] with the purpose of assessing the stability of the MD trajectory in conjunction with the ddCOSMO model and a breathing cavity. The pressure tensor $\mathbf{P}$ can be calculated using Clausius' virial theorem as[25]

$$\mathbf{P} = \frac{2}{V}(\mathbf{E}_{kin} - \Xi) \quad (11)$$

where V is the box volume, $\mathbf{E}_{kin}$ is the kinetic energy, and Ξ is the virial tensor:

$$\Xi = -\frac{1}{2}\sum_{i<j} \vec{r}_{ij}\vec{F}_{ij} \quad (12)$$

with $\vec{r}$ and $\vec{F}$ being the coordinates and forces, respectively. In the framework of spherical NPBC, the isotropic pressure is

$$P = Tr(\mathbf{P})/3 \quad (13)$$

Since in a *NPT* simulation E_{kin} is coupled to the thermostat, the pressure can be controlled by changing the virial, i.e., by scaling the particle positions and, in the present case, the sphere radius R. This is achieved by weakly coupling the system to an external piston, i.e., by letting the pressure relax according to

$$\frac{dP}{dt} = \frac{P_{ref} - P}{\tau_P} \quad (14)$$

where τ_P is the time constant for the coupling. The relationship between the pressure and the isothermal compressibility β (incorporating the factors for different box types in the α parameter) is

$$\frac{dP}{dt} = -\frac{1}{\beta V}\frac{dV}{dt} = -\frac{\alpha}{\beta} \quad (15)$$

which can be used to calculate the scaling factor μ needed to scale coordinates and radius from $\vec{r}$ to $\mu\vec{r}$ (and from R to μR) in a single time step δt:

$$\mu = 1 - \frac{\beta\delta t}{\tau_P}(P_{ref} - P) \quad (16)$$

Computational Details. *MD Simulations.* The MD simulations presented in this work include different solvents (CH_3CN, CH_3OH, and two different water models, namely SPC[26] and TIP3P-FB[14]) and, possibly, the TEMPO solute. We choose to use the TIP3P-FB water model for solute−solvent simulations due to its improved bulk properties with respect to the SPC counterpart, including dipole moment (2.43 vs 2.2 D), density (0.995 vs 0.960 g/cm^3) and static dielectric constant (81.3 vs 65). The topology of the TEMPO solute was adapted from the study of Stendardo et al.,[27] discarding intramolecular interactions since we kept the molecular structure frozen in all the simulations. In this model, two virtual sites are used to describe the lone-pairs of the nitroxide oxygen. The topology and force field parameters of the nonaqueous solvents were the same used in previous studies,[10] whereas the details of the different water models are given in the original publications quoted above.

All MD simulations were run with a locally modified version of the Gaussian[28] suite of programs using the rotational velocity Verlet (RVV1) integrator[29] with an $\epsilon = 10^{-9}$ convergence criterion for the calculation of quaternion derivatives, with the exception of CH_3OH which was treated as a flexible molecule and simulated employing the standard Velocity-Verlet integrator and enforcing holonomic constraints by means of the RATTLE method.[30] The Bussi−Donadio−Parrinello[31] thermostat was used, and a weak coupling scheme was employed in *NPT* simulations. Reaction field effects at the outer boundary of the simulation sphere were taken into account by the ddCOSMO model. Mechanical restraints were imposed instead by "rough walls", such that whenever a solvent molecule goes outside the simulation sphere, its center of mass (COM) velocity is aligned along a new direction sampled with uniform probability in a unit sphere, with a new magnitude corresponding to the average velocity at the thermostat temperature, and the angular momentum is reset in the same way, generating a new angular velocity.

All the starting structures were initially brought to a low-energy configuration with a conjugate gradient minimization. Runs were started sampling initial COM velocities and angular momenta at 298.15 K and carried out for 5 ns (pure water), 30 ns (mean field optimization), 2.5 ns (*NPT* simulations), and 12.5 ns (TEMPO simulations). The first 1 ns of each trajectory (2.5 ns for TEMPO/solvent) was considered equilibration and was not used in the following analysis.

The complete simulation list includes:

- Pure water with the SPC model (*NVT* with and without ddCOSMO) with increasing box radii: 8.5, 10.5, 12,5, 15.5, 18.0, and 20.0 Å, with the largest box employed also for mean field optimization and TEMPO simulations. The integration time step was 2.0 fs when the radius was 8.5 Å, 4.0 fs otherwise. The thermostat coupling constant was 0.4 ps.

- *NPT* ddCOSMO simulations with two different box radii (15.5 and 20.0 Å), a time step of 2.0 fs, and coupling constants for thermostat and barostat of 0.4 and 4.0 ps, respectively.
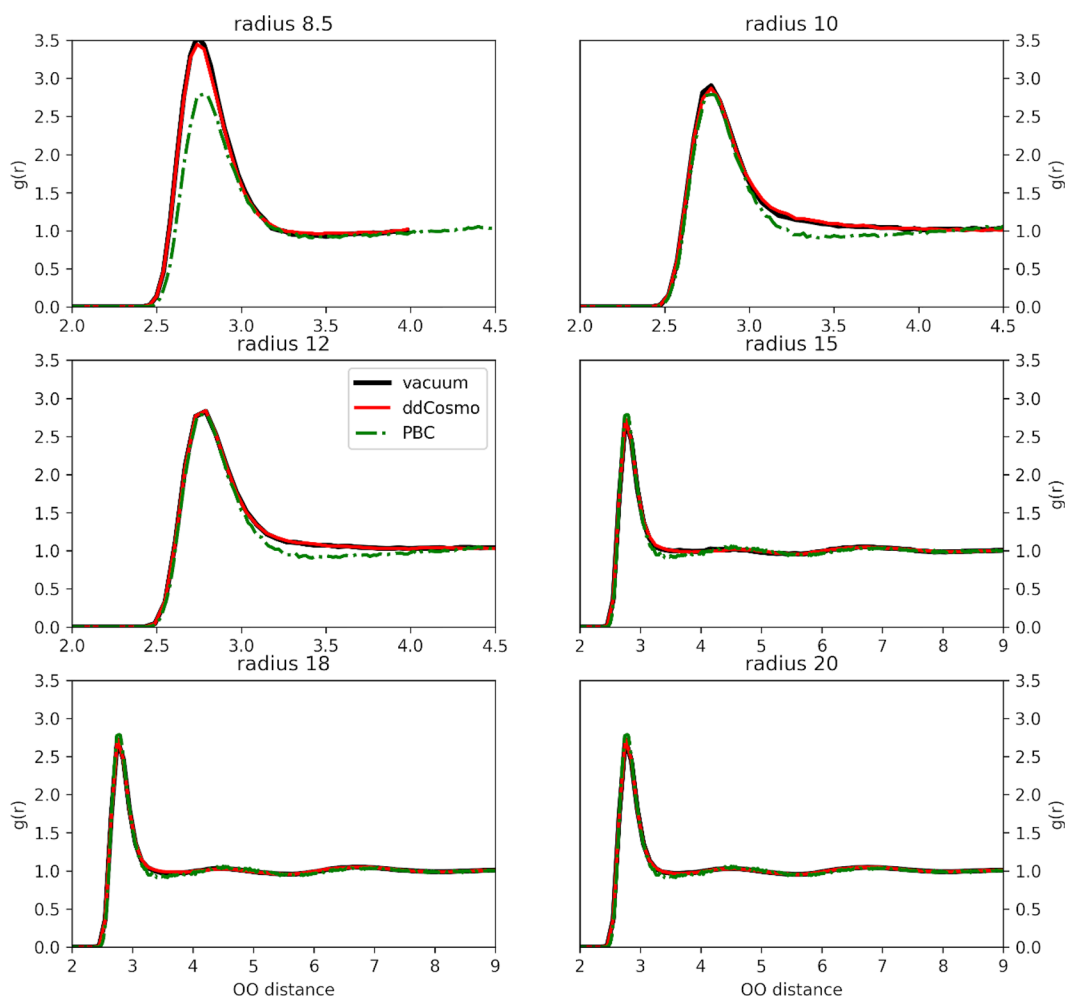
Figure 1. Oxygen−oxygen radial distribution functions of *NVT* pure SPC water simulations under NPBC−vacuum (black solid line), NPBC−ddCOSMO (red solid line), and PBC (green dot dashed line).

- Optimization of a new U_{vdW} potential for TIP3P-FB water, performing the required MD runs with the same parameters used for the NVT simulations of SPC water.

- Production runs of TEMPO with CH_3CN, CH_3OH and TIP3P-FB water solvents employing the same parameters as above with the exception of CH_3OH (time step of 2.0 ps and thermostat coupling constant of 0.2 ps).

Optimization and Fitting of the U_{vdW} Potential. The optimization of the U_{vdW} potential to be employed in the GLOB model[32] proceeds as follows: the simulation sphere is divided in concentric shells and, at a given interval, a tiny repulsive or attractive term localized in the shell is added, depending on past density fluctuations so that at the end these will be damped.[10] Here we used a total of 80 constant radius shells, updating the density profile every 50000 simulation steps. Other parameters of the GLOB model were set as in previous publications.[10] When the density fluctuations are small enough the potential energy profile is saved. Since U_{vdW} acts mainly near the outer boundary of the NPBC sphere, prior to the fitting the profile is truncated once its value is below 0.1 kJ/mol. The potential energy points are split in training and test subsets of the same size, and the profile is finally fitted by a polynomial expression:

$$U_{vdW} = \sum_{j=1}^{n} a_j r^j$$

(17)

The degree of the polynomial is determined running the corresponding ridge regressions. We tested degrees from 0 to 12 and for each degree of the polynomial the shrinking factor value was optimized with a λ, μ Evolutionary Algorithm $(EA)^{[33]}$ with a population size of 100, mutation rates of 0.3 and 0.5 for parents and children, and a crossover rate of 0.5 for 500 iterations. The degree of the polynomial was finally selected by choosing the best outcome of the corresponding learning curves for the root mean square error (RMSE) and R^2 values.

Analysis of Trajectories. The calculation of structural properties was carried out with standard procedures, but taking into account proper normalization for NPBCs. The strength of the hydrogen bonds was calculated with a continuous function[34] referred to in the following as F_{HB}.[72]

Solvent effects on the solute properties were evaluated for a selection of simulation frames using unsupervised learning methods. We proceeded as follows: first we built a set of coordinates to be used to compare MD frames (the feature space); next, the number of dimensions was reduced by a Principal Component Analysis (PCA) under the constraint of a (projected variance)/(total variance) ratio larger than 0.9. Finally, we used the projected feature space and a clustering
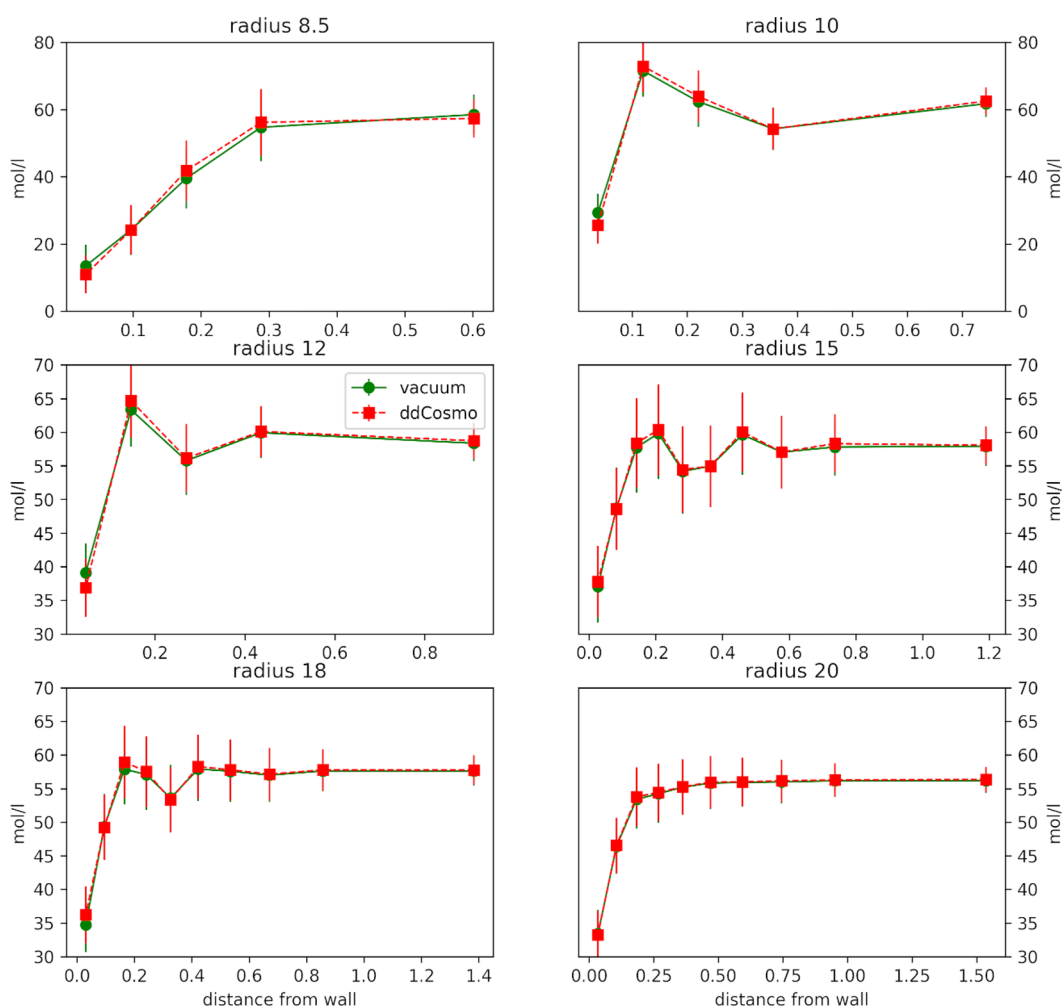
Figure 2. Mean density and standard deviations (error bars) in five (radius 8.5−12 Å) or 10 (radius ≥12 Å) constant volume concentric shells for *NVT* SPC MD simulations.

procedure to separate the trajectories in different basins, each described by its centroid. Once a centroid was obtained, we sampled from each cluster a number of points proportional to the cluster size in order to obtain a fixed number of simulation frames employing a combinatorial optimization method[35] to cover both high and low density regions in the projected feature space region of each cluster. The details of each step are described below.

As a feature space for the snapshot selection of TEMPO trajectories we used the first and second neighbor distances for all the N atom1, N atom2, LP_1-atom1, LP_1-atom2, LP_2-atom1, and LP_2-atom2, where atom1 and atom2 are two properly chosen atoms for each solvent species and LP1, LP2 are the virtual sites of the nitroxide oxygen. We took a frame every 10 ps to build the feature space in order to avoid short time correlations in the clustering. Simulation frames were clusterized by the Partition Around Medoids (PAM) algorithm.[36] Since cluster analysis without robust validation is meaningless,[17,19,37,38] the best number of clusters (k) was determined from the analysis of PAM runs for values from 2 to 20 in terms of the internal validation criteria of the Silhouette Score (Supporting Information), Dunn Index (DI) and Calinski−Harabasz score (pSF).[36,39] In addition, we looked for a break-even point in the Within Sum of Squares error (WSS). SI, DI, and pSF should have a maximum

corresponding to the parameter set (just the value of k in this case) that yields the best clustering while for WSS one looks for a change in the slope. The best value of k was obtained from the consensus of three criteria over four.[40,41]

Representative frames within clusters were selected by an in house implementation of the Greedy Randomized Adaptive Search Procedure (GRASP),[42] which is a procedure to solve combinatorial optimization problems. Given a problem defined by a finite set $E = \{1, 2, ..., n\}$ (here the MD frames[43]) and objective function $g: 2^E \rightarrow \mathbb{R}$, which measures the total dissimilarity (DS) of the selected snapshots taking into account density, we want to find a set of feasible solutions $F \subseteq 2^E$ (possible subset of frames) which maximizes the objective; i.e., it gives a description as complete as possible of the cluster. Here, we have used an implementation of GRASP made up of two phases,[42] namely the construction of a feasible solution followed by its refinement employing a local search method. In the first phase, a Restricted Candidate List (RCL) is built adding elements based on their contribution to the total dissimilarity:[44] a candidate c is added if $DS_{min} \leq DS(c) \leq DS_{min} + \alpha \cdot (DS_{max} - DS_{min})$; i.e., α tunes the algorithm from a greedy to a completely random solution (here we used $\alpha = 0.1$), and a random element is then chosen from the candidates to build the solution. In the second phase, the built solution is refined selecting elements in the neighborhood of each

candidate and checking if the swap gives some improvement. The swap is tried starting from the nearest neighbor and proceeding toward all the points near a specific candidate or for a fixed fraction of the data set (here 0.05). The whole procedure is repeated for a predefined number of times (here 100) selecting the best overall outcome.

■ RESULTS AND DISCUSSION

Pure Solvent *NVT* Simulations. *Boundary Properties of NPBC/ddCOSMO Simulations.* Figure 1 shows the OO radial
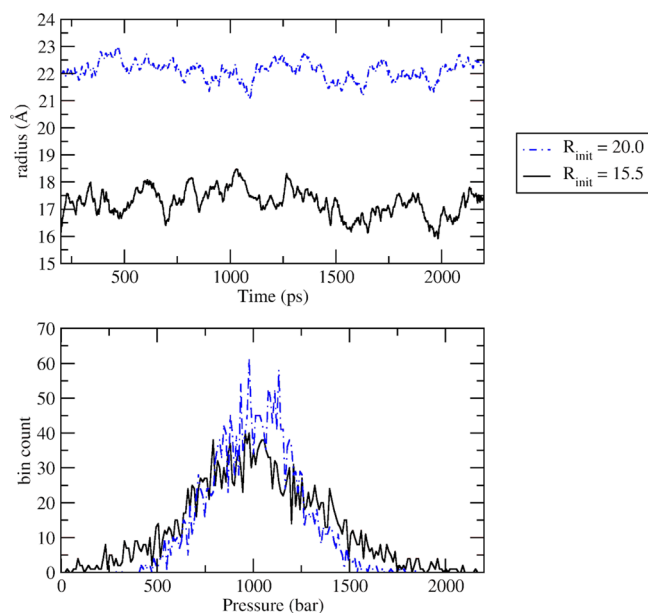
Figure 3. *NPT* simulations, skipping a calibration time of 200 ps. Upper panel: radius of simulations starting at 15.5 and 20.0 Å. Lower panel: histogram of sampled pressure values.
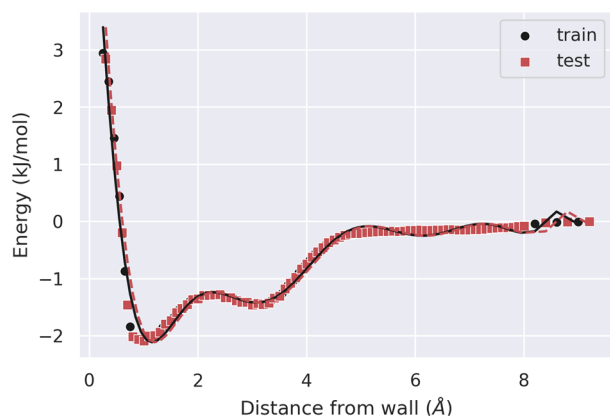

Figure 4. TIP3P-FB mean field potential energy profile. The data used in the fitting and the corresponding fitted polynomial are drawn in black, whereas the test points are drawn in red.

distribution function $(g(r))$ obtained from the different pure water simulations. For the smaller simulation spheres (radii from 8.5 to 12 Å) the $g(r)$ was calculated considering the distribution for all the oxygen atoms within 4.5 Å and looking at neighbor atoms up to 4.0, 5.5, and 7.5 Å, respectively; for the larger systems, we took into account atoms within 8.0 or 10.0 Å and searched for neighbor ones up to 7.0, 8.0, and 10.0 Å, respectively.

Table 1. U_{vdW} Linear Fit Parameters for TIP3P-FB water

parameter	TIP3P-FB
a_0	6.958446e+00
a_1	−1.381768e+01
a_2	−7.095241e+00
a_3	2.678593e+01
a_4	−2.224122e+01
a_5	9.410601e+00
a_6	−2.343013e+00
a_7	3.571387e−01
a_8	−3.280733e−02
a_9	1.669372e−03
a_{10}	−3.616741e−05
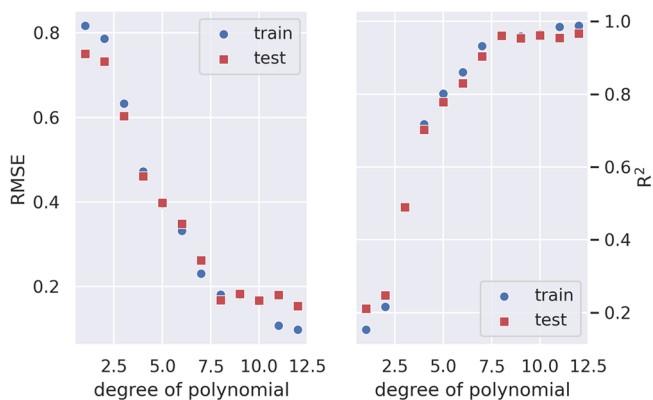
Figure 5. Value of RMSE and R^2 as a function of the degree of the polynomial. Train and test values are superimposed.
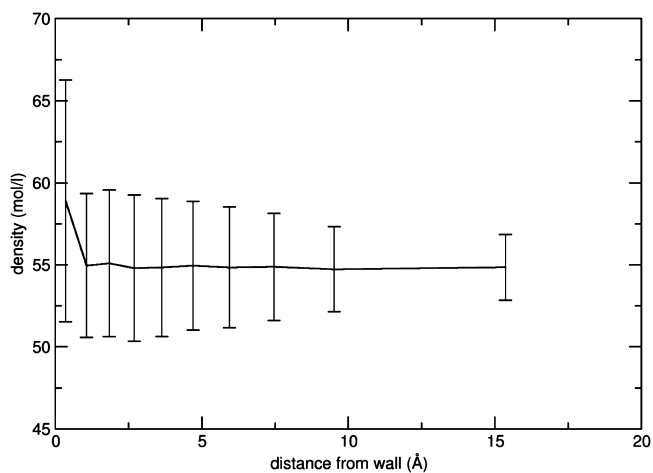
Figure 6. Average density and corresponding standard deviations in 10 concentric shells of constant volume for NPBC simulations of TIP3P-FB water.

As can be easily observed, the simulation with the smallest box (8.5 Å radius) not including the reaction field contribution features a first peak of the radial distribution function which is higher and slightly left shifted with respect to the PBC one, indicating a loss of structural disorder. This effect is partially overcome by using ddCOSMO as illustrated by the values of the half weight at half-maximum (HWHM) of 0.188, 0.216, and 0.209 Å calculated at 1.725, 1.518, and 1.385 Å, respectively). This effect already disappears for boxes with radii larger than 10 Å; the second and third plot show a lack of the depletion zone which separates the first and second
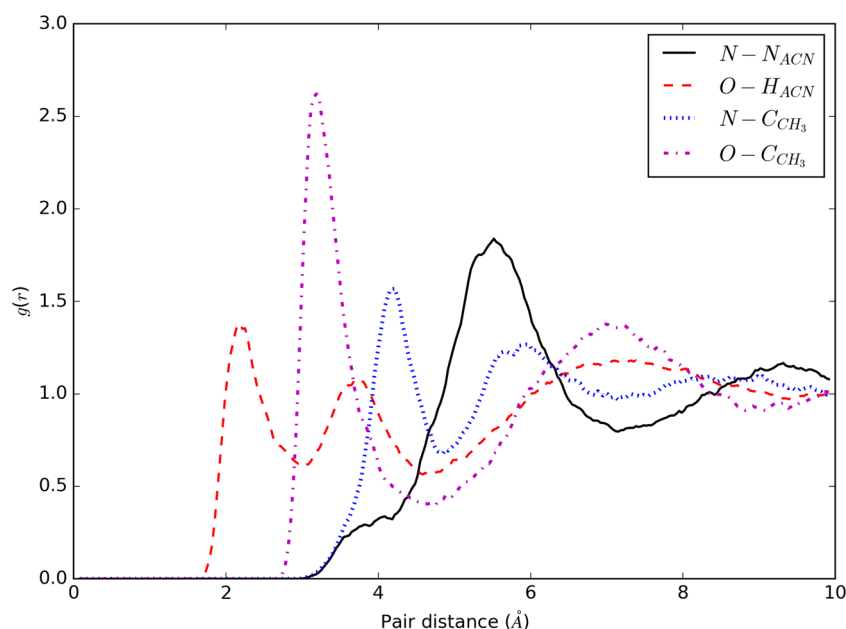
Figure 7. TEMPO−acetonitrile simulation: radial distribution functions.
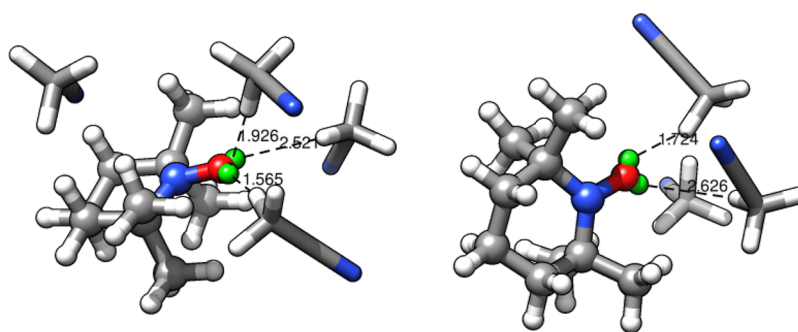

Figure 8. Cluster centroids of the TEMPO−CH₃CN simulation. The TEMPO molecule (ball and stick) is shown with the solvent molecules (licorice) within 4 Å of the nitroxide group atoms (including the virtual sites). Some solute−solvent distances are indicated as well.
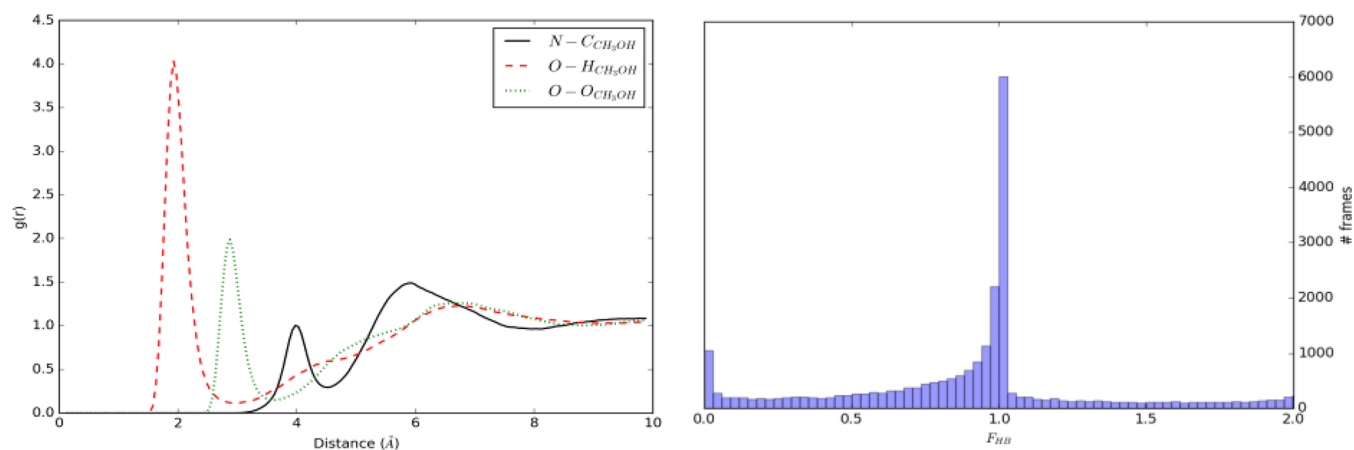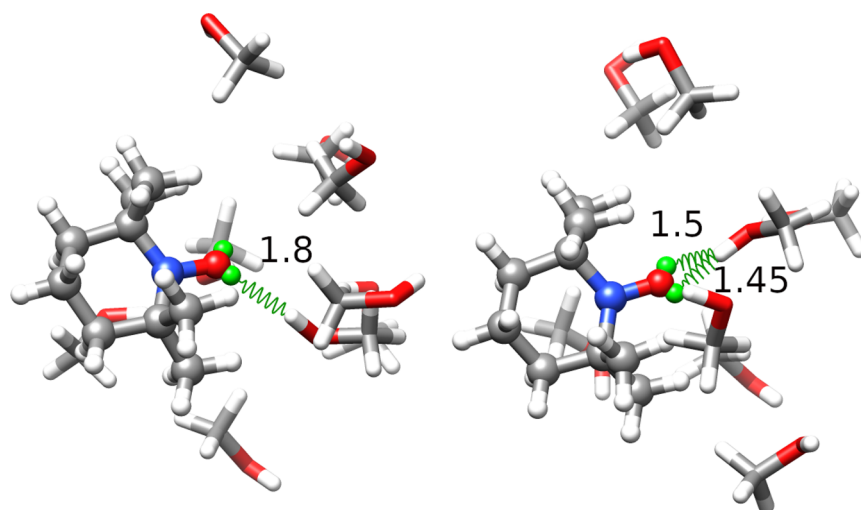

Figure 9. TEMPO−CH₃OH simulation. Left panel: radial distribution functions for N−C, O−H, and O−O atom pairs. Right panel: histogram of the F_{HB} function.

coordination shell, but even this effect vanishes for radii larger than 15 Å. From a different point of view, the same trend is observed when looking at density fluctuations in concentric shells of constant volume (see Figure 2)

The shortcomings of the SPC model are well-known, and the green curve is not a really good model for the structure of liquid water; however, under the same simulation conditions, the agreement between the liquid structure provided by PBC and NPBC simulations gives a convincing indication of adequate system size and absence of additional artifacts.

Figure 10. Cluster centroids of the TEMPO−CH₃OH simulation. H-bonds are depicted as green springs.

Figure 11. TEMPO−H₂O simulation. Left panel: radial distribution functions for N−O, O−H, and O−O atom pairs. Right panel: histogram of the F_{HB} function.
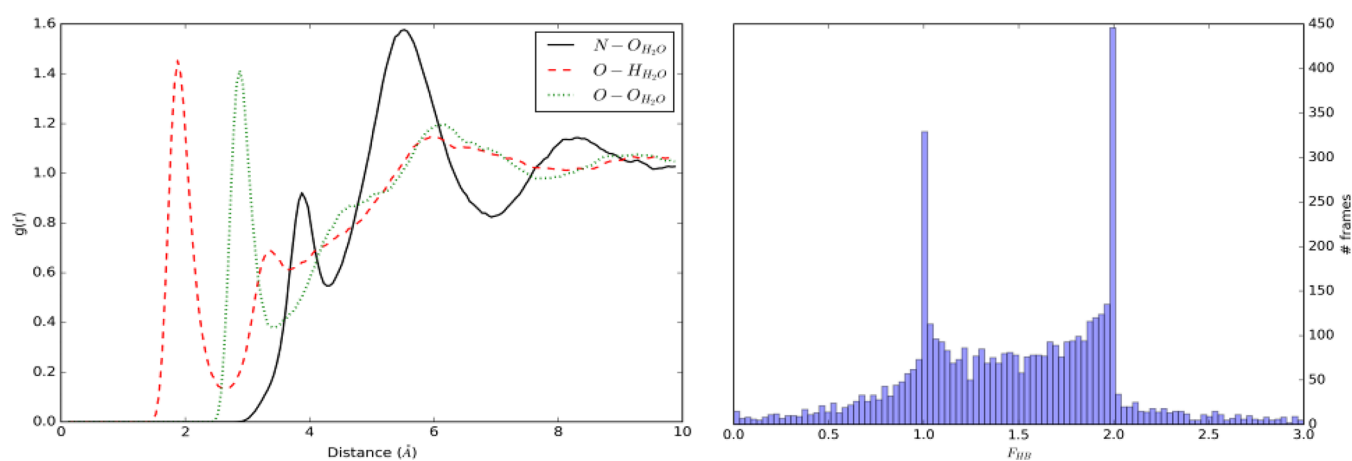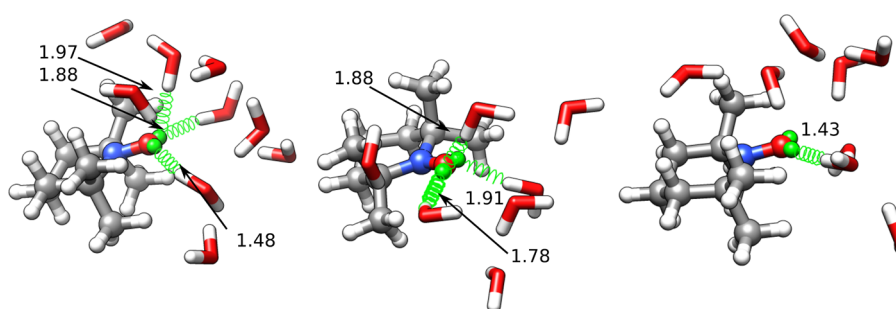

Figure 12. Cluster centroids of the TEMPO−H₂O simulation. The same graphical conventions of Figure 10 have been used.

***NPT* Simulations.** Figure 3 shows the results of both the *NPT* simulations carried out for pure water spherical boxes setting the barostat at 1000 bar.

The starting coordinates and box radii were the same of the corresponding largest *NVT* simulations, i.e., 15.5 and 20.0 Å. The upper panel shows the evolution of the box radius as a function of time after the equilibration time of 200 ps. It is apparent that the average radii of the two simulation spheres (17.15 and 22.10 Å, respectively) are about 2.0 Å larger than the starting values and that they oscillate around the average values. The lower panel shows the histograms of the sampled

pressure values where it is apparent that the instantaneous pressure distribution is centered on the barostat value.

In our opinion, these results provide a satisfactory demonstration of the robustness and stability of the approach we implemented.

Optimization of the U_{vdW} Potential for TIP3P-FB Water. Having assessed the stability of the new features for the SPC water model, we ran a 40 ns *NVT* simulation to obtain an optimized U_{vdW} potential energy profile for TIP3P-FB H₂O.

The resulting profile is shown in Figure 4 (without the truncated portion) together with the fitted polynomial, whose
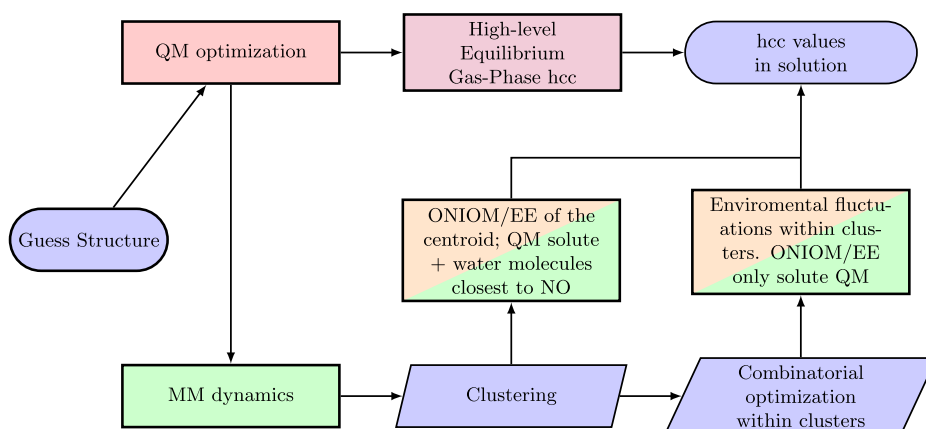
Figure 13. Flowchart of the computational strategy employed for the evaluation of TEMPO A_N in different solvents. Green boxes represent MM steps of the procedure, and orange boxes represent simulations typically performed with DFT and hybrid functionals, whereas the red ones are performed with double-hybrid functional. The purple box represents the step evaluated at the highest affordable level of theory (CC).

Table 2. Geometrical Parameters of the Dimethyl Nitroxide Template Molecule at Different Levels of Theory[a]

| level of theory | basis sets | minimum | | | | transition state | | | |
		rNO	rCN	CNC	CCNO	rNO	rCN	CNC	ΔE (cm^{-1})
B3LYP	SNSD	1.2818	1.4611	118.82	155.2	1.2808	1.4557	119.59	270.49
	julDZ	1.2818	1.4604	118.85	154.3	1.2817	1.4546	119.46	293.43
PBE0	SNSD	1.2683	1.4494	118.48	154.6	1.2688	1.4452	119.24	265.61
	julDZ	1.2700	1.4498	118.55	154.5	1.2702	1.4454	119.36	277.21
PW6B95D3	SNSD	1.2720	1.4506	118.75	155.9	1.2713	1.4458	119.63	264.70
	julDZ	1.2720	1.4495	118.82	154.6	1.2723	1.4452	119.67	271.85
rDSD-D3	junTZ	1.2754	1.4548	118.59	152.9	1.2765	1.4502	119.47	339.04
	EPR(III)	1.2743	1.4540	118.51	152.7	1.2752	1.4494	119.43	361.42
CCSD(T)-F12	DZ-F12	1.2773	1.4554	118.32	150.8	1.2774	1.4491	119.64	404.60
CCSD(T)-F12	TZ-F12	1.2762	1.4534	118.41	151.1	1.2764	1.4485	119.62	387.90
CCSD(T)-F12	TZ-F12+CV	1.2754	1.4517	118.41	151.2	1.2757	1.4469	119.60	371.45

[a]At all levels of theory the CCNO angle in the transitions state is 180.0°, and therefore, it was omitted from the table.
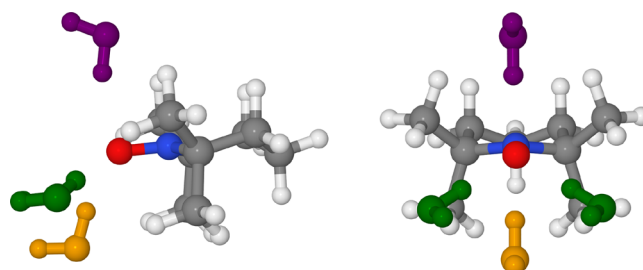
Table 3. Nitrogen Hyperfine Couplings in MHz for the Dimethyl Nitroxide Template Molecules Computed at Different Levels (Conversion Factors au MHz = 323.13 and au Gauss = 115.3)

level of theory	basis sets	minimum	TS
PBE0	SNSD	37.3	28.9
B3LYP	SNSD	41.7	28.5
PW6B95	SNSD	43.3	29.7
rDSD	EPR(III)	44.6	31.1
CCSD(T)	cc-pwCVTZ	40.6	28.2
CCSD(T)	aCV(T,Q)[a]	44.9	30.9
CCSD(T)	aCV(T,Q)/EPR(III)[a]	44.9	31.0

[a]aug-cc-pwCVQZ for s,p and aug-cc-pwCVTZ for d,f functions.

Figure 14. Representation of the relative positions of a water molecule with respect to the TEMPO NO group: **t**, orange; **p**, green; **o**, purple.

coefficients are shown in Table 1). Both training and test sets contain 80 points with a resolution of 0.1 Å, after truncation at 10 and 8 Å, respectively. Looking at the learning curves in Figure 5 for RMSE and R^2, we selected a tenth-degree polynomial. Note that the H_2O profile shows a double well, which explains the higher degree of the fitting polynomial needed for water with respect to organic solvents, as already reported for the SPC water model.[10,32]

At the end of the procedure, we ran an additional 10 ns simulation for each solvent in order to check the performance of the U_{vdW} term; Figure 6 clearly shows that the density

fluctuations are small and are nearly constant within the whole simulation sphere.

Application: TEMPO Radical in Different Solvents. NPBC Simulations. The first simulation of TEMPO was carried out in acetonitrile, a polar non-hydrogen-bonding solvent. The resulting radial distribution functions for the oxygen and nitrogen distances from the methyl carbon (Figure 7) show two peaks with maxima located at 3.15 and 4.2 Å, respectively, which are indicative of an alignment of the methyl–cyanide dipole roughly antiparallel to the N–O one of TEMPO, which is confirmed by the position (2.2 Å) of the first peak maximum for the O–H_{CH3CN} pair.

Table 4. Hyperfine Coupling Constants at the PW6B95/SNSD Level of Theory (Water Charges: O, −0.8; H, 0.4)

system	a_N(QM)	a_N(QM/MM)[a]	a_N(QM+C-PCM)	Δ(QM)[b]	Δ(C-PCM)[c]
TEMPO	40.96	40.96	42.48	0.00	1.52
o	41.58	41.84 (0.26)	43.12	0.62	1.54
t	42.85	42.82 (−0.03)	44.76	1.90	1.91
p	42.37	42.09 (−0.28)	44.32	1.41	1.95
2·p	43.62	43.89 (0.27)	45.08	2.66	1.46
o+t	43.58	43.16 (0.42)	44.49	2.62	0.91
2·p+o	43.84	44.19 (0.35)	44.78	2.88	0.94
2·p+t	44.81	44.23 (−0.58)	46.21	3.85	1.40
p+o+t	44.17	44.54 (0.37)	45.08	3.21	0.91
2·p+o+t	45.68	45.11 (−0.57)	45.60	4.72	−0.08

[a]In parentheses is given the difference between QM and QM/MM. The mean absolute error and maximum error are 0.25 and 0.42 MHz, respectively. [b]Differences between the different rows of column 2 and its first row (QM calculation for the bare TEMPO radical). [c]Difference between the different rows of column 4 and the corresponding rows of column 2.
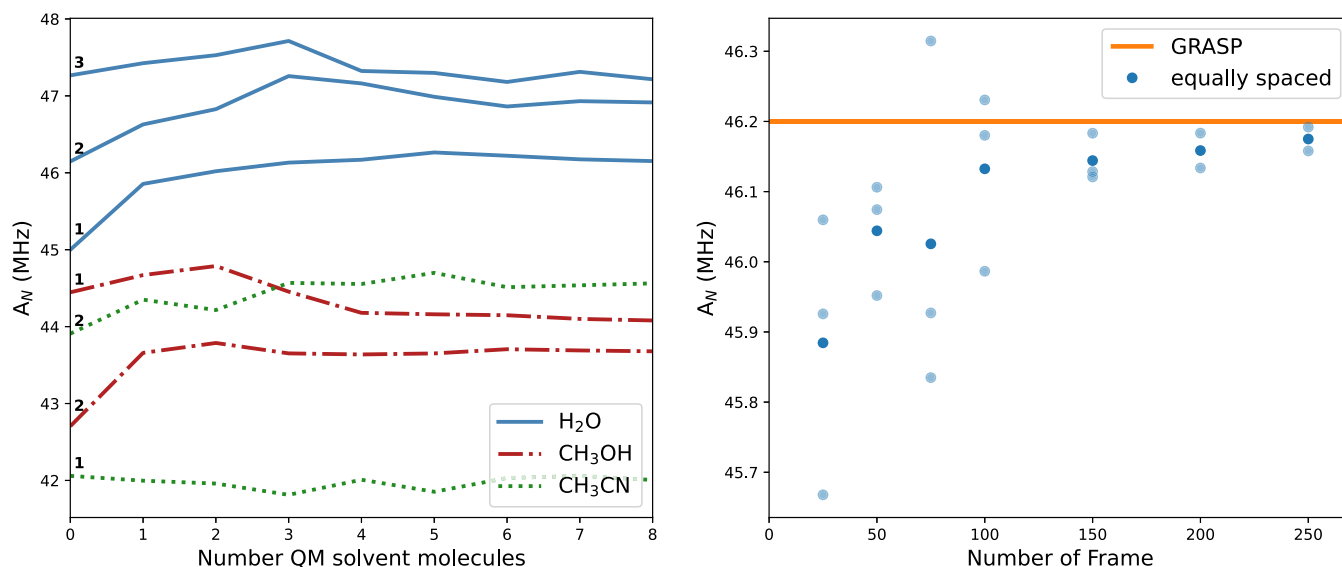
Figure 15. Convergence of TEMPO A_N as a function of the number of QM solvent molecules in different solvents (left panel, bold numbers identify the centroid) and of the number of equally spaced snapshots in aqueous solution with an MM (TIP3P-FB) representation of all the water molecules (right panel). In the latter case, the orange line corresponds to the value issued from the GRASP procedure (75 frames divided in three clusters) and the fading dots represent multiple evaluations starting from random points of the trajectory, whereas the bold dots are the mean values.

The position of the N_{TEMPO}–N_{CH3CN} first maximum at about 5.5 Å is indicative of little direct interaction above and below the average molecular plane of the solute. Anyhow, to include all possible contributions, we built the feature space using the dummy atom–methyl hydrogen and N–N first and second neighbor distances.

A subsequent PCA showed that using four coordinates yielded 91.1% of the total variance. The values of first and second neighbor distances are shown in Figure S1 of the Supporting Information. Two clusters of comparable size (389 and 552 members, respectively) were obtained by inspection of validation scores (Figure S2 of the Supporting Information), whose centroids are shown in Figure 8 and differ for the number of solvent molecules close to the nitroxide moiety.

The left panel in Figure 9 shows the radial distribution functions calculated from the TEMPO - CH$_3$OH simulation. In this case, the g(r) shows sharp well-defined peaks and depletion zones for the atom involved in the hydrogen bond between methanol and TEMPO; the first maxima and minima are located at 1.93 and 2.88 Å, respectively. The F_{HB} histogram

shows that one stable hydrogen bond is formed between TEMPO and methanol ($\langle F_{HB} \rangle$ = 0.896).

A PCA of the N−O, LP$_1$−H, LP$_2$−H, and O−O distances shows that 93.1% of the total variance is accounted for when using the first three components, which were then used for clustering (see Figure S3 in the Supporting Information). Inspection of the validation results (Figure S4 in the Supporting Information) indicates the presence of two clusters (three scores out of four) made up of 707 and 548 points, respectively. These were then passed to the GRASP step. The results of the PAM+GRASP procedure (see Figure S5 in the Supporting Information) show that the distributions sampled from the two clusters span the complete (irregular) shape of the feature space projected on its first two eigenvectors, whose corresponding eigenvalues are 0.78 and 0.11, respectively.

Figure 10 shows a ball and stick representation of the two centroids including the solute and the solvent molecules within 4.0 Å of the atoms used to build the feature space, i.e., N, O, LP1, and LP2.

It is apparent that the left centroid, which corresponds to the most populated cluster, forms stable hydrogen bonds oriented

Table 5. Comparison between Computed and Experimental A_N for TEMPO Radical in Different Solvents

	CH$_3$CN solution		CH$_3$OH solution		H$_2$O solution		
dielectric constant (ϵ)	35.688		32.613		78.3553		
C-PCM	42.4		42.4		42.5		
cluster number	1	2	1	2	1	2	3
cluster size	432	463	707	548	173	425	343
weight of the cluster	0.48	0.52	0.56	0.44	0.18	0.45	0.37
GRASP points	12	13	36	28	9	23	19
averaged a_N(MM) in the cluster	42.8	43.1	43.8	43.5	46.7	46.0	46.3
a_N(MM) at the centroid	42.1	43.9	44.4	42.7	47.3	46.1	45.0
a_N(QM) at the centroid	42.0	44.6	44.2	43.8	47.3	47.1	46.2
weighted value	43.2		44.0		47.1		
best estimate[a]	44.2		45.0		48.1		
experiment	43.8[b]		45.4[b]		48.4[c]		

[a]weighted value + hybrid/double-hybrid difference. [b]Reference 67. [c]Reference 65.

along one of the virtual site-oxygen bonds. The average value of the F_{HB} function in this cluster is 0.89, very close to the value for the whole trajectory. In the right centroid, the nearest methanol molecule lies between the two solute virtual sites. In this cluster, $\langle F_{HB} \rangle = 0.67$, with this suggesting that the included frames account for solvent exchange events.

The radial distribution functions obtained from the TEMPO–TIP3P-FB water trajectory are shown in Figure 11.

This trajectory shows stronger and more directional solute–solvent interactions, which are well represented by the positions of the $g(r)$ first peaks, located at 3.88, 2.88, and 1.93 Å for N–O, O–H, and O–O, respectively, and the well-defined depletion zones separating the first and second coordination shell for O_{TEMPO}–$H_{TIP 3p-FB}$ and O_{TEMPO}–$O_{TIP3P-FB}$. The F_{HB} histogram shows a bimodal shape with peaks corresponding to one and two hydrogen bonds, respectively ($\langle F_{HB} \rangle = 1.467$). It is noteworthy how the intermediate zone in the interval $F_{HB} = [1, 2]$ is densely populated as well, indicating a continuous transition between the two most likely states.

The feature space employed for the frame selection was the same as that used for the TEMPO–methanol simulation, and the corresponding graphs are shown in Figure S6 of the Supporting Information. The filtered trajectory used to build the feature space included 941 points, and according to the corresponding PCA, five components are necessary to account for 90% of the total variance. The consensus of three scores out of four (see Figure S7 in the Supporting Information) led to the selection of three clusters containing 343, 425, and 173 points, respectively. Figure S8 in the Supporting Information is analogous to Figure S5, and the shape of the point distribution in the feature space is again well reproduced by GRASP. In this case, the eigenvalues of the first two principal components are 0.51 and 0.22, respectively.

Figure 12 shows a ball and stick representation of the three centroids. In analogy with the methanol solution, the solvent molecules located within 4 Å of the atoms whose two nearest neighbor distances were used to build the feature space are

shown together with TEMPO. The distances between water hydrogens and lone pairs of the nitroxide oxygen are indicated as well. It is apparent that the first two clusters involve water molecules close to both sides of the nitroxide moiety, at distances compatible with the formation of hydrogen bonds. The only exception is the molecule located at 1.48 Å in the leftmost panel, which is indicative of the same exchange phenomenon observed for methanol. Accordingly, the first two clusters show $\langle F_{HB} \rangle$ values of 1.55 and 1.51, respectively. The third (and smallest) cluster shows a different structure with just one water molecule near a virtual site, and the average number of hydrogen bonds is reduced to 1.36.

The results of this subsection show that, as expected, the competition between solvent–solvent and solute–solvent interactions tunes the structure of the cybotactic region of TEMPO. In particular, one or two intermolecular hydrogen bonds can be formed by the lone pairs of the nitroxide oxygen, whose relative populations and interconversion rates depend on the nature of the solvent. Roughly speaking, water forms two stable hydrogen bonds, whereas methanol prefers a single hydrogen bond, possibly due to steric constraints by the methyl groups belonging to the solute and the solvent. The lack of solute–solvent hydrogen bonds in acetonitrile solution leads to a less rigid cybotactic region, although some order is apparent also in this case due to a preferential alignment of the solvent dipole moment roughly antiparallel to the NO one of TEMPO. Noted is that all these trends are in agreement with the available experimental evidence (see ref 45 and references therein).

Solvatochromism of Nitrogen Isotropic Hyperfine Couplings. The computation of nitrogen isotropic hyperfine couplings (A_N) for TEMPO in different solvents will be based on a validation of the QM model followed by an integrated strategy employing the results of the trajectory analyses described in the preceding section. The overall computational protocol is sketched in Figure 13, whose different steps will be analyzed in the following.

In order to define the most effective QC method to be employed in the present context, a small reference system (dimethyl nitroxide) was investigated by different hybrid functionals. We considered B3LYP,[46,47] PBE0,[48] and PW6B95,[49] in conjunction with the double-ζ jul-cc-pVDZ[50] and SNSD[51] basis sets, and we also applied a last-generation double-hybrid functional rev-DSDPBEP86-D3[52] (hereafter rDSD) with two triple-ζ basis sets, jun-cc-pVTZ[50] and EPR(III).[53] Reference geometries of both the pyramidal equilibrium structure and the planar transition state ruling the inversion have been obtained at the CCSD(T)-F12b(3C/FIX) level[54] (hereafter referred to simply as CCSD(T)-F12) using the cc-pVTZ-F12 basis set,[55] which is able to deliver results comparable with those obtainable by conventional CCSD(T) calculations employing much larger basis sets.[56] Core–valence (CV) correlation effects, often mandatory for obtaining accurate geometrical parameters,[57] were taken into account by adding to the results the difference between geometrical parameters optimized correlating all the electrons (AE) or freezing the core electrons (FC) using the cc-pCVDZ-F12 basis set.[58] All the DFT computations were performed with the Gaussian suite of programs,[28] employing, when needed, analytical first and second energy derivatives, whereas CCSD(T)-F12 computations were performed with the MOLPRO package,[59] employing, when needed, numerical gradients.[60]

The results collected in Table 2 show small, but non-negligible variations with respect to previous CCSD(T)-F12/cc-pVDZ-F12 results.[61] Furthermore, both PW6B95 and rDSD results are quite accurate and represent significant improvements with respect to B3LYP and PBE0 values. This remark applies also to the inversion barrier, which is particularly important in connection with magnetic properties, especially in the case of the double-hybrid rDSD functional. It is noteworthy that the geometrical parameters obtained using basis sets purposely tailored for the computation of electron paramagnetic resonance (EPR) parameters (SNSD for hybrid functionals and EPR(III) for double-hybrid functionals) are close to those provided by the basis sets jul-cc-pVDZ and jun-cc-pVTZ for hybrid- and double-hybrid functionals, respectively.[62] Consequently, only SNSD and EPR(III) basis sets will be employed in the following in conjunction with PW6B95 and rDSD functional, respectively.

Next, the reliability of the computed isotropic hyperfine coupling of ^{14}N (hereafter A_N) has been checked using again dimethyl nitroxide and reference values obtained at the all-electron CCSD(T) level in conjunction with a large basis set including proper treatment of core electrons, namely aug-cc-pwCVQZ for s, p orbitals and aug-cc-pwCVTZ[63] for the polarization functions. Test computations for H_2NO showed that this basis set delivers nitrogen hyperfine couplings within 0.05 MHz from those obtained by the full aug-cc-pwCVQZ basis set. Furthermore, replacement of this basis set by the cheaper EPR-(III) counterpart for C and H atoms leads to very accurate results with a significant reduction of the computational cost. Note that the bare cc-pwCVTZ basis set delivers unsatisfactory results. All these computations have been performed with the CFOUR package.[64]

The results collected in Table 3 show that PBE0 and B3LYP functionals significantly underestimate the CCSD(T) reference values especially for pyramidal structures. The PW6B95 functional, which includes a larger amount of Hartree–Fock exchange, reduces the errors to 1.3 MHz for both planar and pyramidal structures, thus suggesting that environmental effects can be safely computed at this level. Along the same line, rDSD/EPR(III) results are in quantitative agreement with the reference values, with the difference (0.3 MHz) being within the experimental error bar for both planar and pyramidal structures. As far as the TEMPO radical is concerned, the A_N computed at the PW6B95/SNSD level is 41.0 MHz employing either PW6B95 or rDSD equilibrium structures, confirming that PW6B95 geometries can be confidently used for computing solvent effects. The difference of 1.3 MHz between the computed value of 41.4 MHz including a bulk solvent contribution estimated at the C-PCM level, and the experimental counterpart in cyclohexane solution (42.7 MHz[65]) is identical with its estimate obtained from the corresponding difference between *in vacuo* PW6B95 and rDSD results for TEMPO (1.0 MHz) plus the difference between rDSD and CCSD(T)/aug-cc-PwCV(Q,T)Z results for the pyramidal structure of dimethyl nitroxide (0.3 MHz). Taking into account the experimental error bar (0.3 MHz) and the uncertainties of the computational estimate, we will adopt in the following a correction of 1 MHz for all the PW6B95 results.

A first analysis of the role of solute–solvent interactions was performed by considering different clusters containing TEMPO and an increasing number of water molecules, according to the scheme shown in Figure 14

The results reported in Table 4 show that non negligible modifications of hyperfine coupling constants are induced by up to four water molecules near the NO moiety, with QM and MM representations of these water molecules leading to comparable results.

Next, a more systematic study of the difference between QM and MM (point charges) representations of solvent molecules has been performed for the centroids of the most populated clusters of TEMPO in acetonitrile, methanol, and water.

Figure 15 shows that converged results are always obtained using four solvent molecules described at the QM level. Furthermore, the right panel of the same figure confirms the efficiency of the clustering procedure, which allows one to reach the converged results, even for the most demanding solvent (i.e., water), with about one-third of the computational cost required when employing equally spaced snapshots.

Those findings suggest the following strategy based on the clustering of the trajectories. In the first step, the A_N of each cluster centroid is evaluated including the four solvent molecules closest to the NO moiety in the QM system embedded in the environment represented by the other MM water molecules. Next, fluctuations within each cluster are evaluated reducing the QM system to just the TEMPO radical, and the final value is increased by 1.0 MHz, as explained above. The results obtained for the different solvents are collected in Table 5. Note that CH_3CN and CH_3OH have similar dielectric constants, so that any bulk solvent model (here C-PCM[66]) gives close results and that saturation is already seen for dielectric constants above 35, leading to a very similar value also for water.

On the contrary, the proposed computational procedure is able to disentangle the role of different interactions in the cybotactic region providing the correct trend of A_N. The final results are indeed satisfactory also from a quantitative point of view when recalling that the estimated error of the experimental results is of the order of 0.3 MHz.

■ CONCLUSIONS

In this paper, we have extended a general molecular dynamics tool capable of enforcing nonperiodic boundary conditions, by adding a very effective treatment (ddCOSMO) of reaction field effects from solvent bulk, and an isothermal isobaric ensemble. Furthermore, robust clustering procedures have been implemented in order to reduce the number of QM/MM computations to be performed to achieve well converged spectroscopic parameters. After a comprehensive test of the new tool for different conditions and optimization of the parameters for a new solvent, we have studied the solvatochromic shifts on the EPR spectra of a prototypical nitroxide radical (TEMPO). To this end, we have tuned the different computational parameters with reference to very accurate reference values for a small model system, showing that last-generation hybrid (PW6B95-D3) and double-hybrid (rev-DSDPBEP86-D3) functionals, in conjunction with purposely tailored basis sets (SNSD and EPR(III), respectively), deliver accurate results for both geometric and spectroscopic parameters. On these grounds, we have computed the nitrogen hyperfine coupling constant of a representative nitroxide radical (TEMPO) in different solvents. The good agreement of the results with the available experimental data confirms that we have at our disposal a robust and accurate tool for the study of spectroscopic parameters of medium-to-large size systems in condensed

phases. Other fields of applications of the new tool can be also envisaged, including, for instance, fluxional atomic metal clusters or noble gas nanodroplets. In the first case, the recent development of synthetic techniques allowing for the production of monodisperse metal clusters composed by a few atoms[68] paves the route toward bio-oriented applications.[69] In the latter case, spectroscopic parameters (e.g., photoabsorption spectra) of molecules solvated in helium nanodroplets[70] can be computed in the framework of a QM/MM model employing the recently developed renormalized He−He potentials to account for the fluid nature of the solvent.[71] Work along these and related lines is ongoing in our laboratories.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00046.

> Figures that describe the feature space and the results of clustering for the TEMPO−CH$_3$CN, TEMPO−CH$_3$OH, and TEMPO−H$_2$O simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Giordano Mancini − *Scuola Normale Superiore, 56126 Pisa, Italy; Istituto Nazionale di Fisica Nucleare (INFN) sezione di Pisa, 56127 Pisa, Italy;* ⓘ orcid.org/0000-0002-1327-7303; Email: giordano.mancini@sns.it

Vincenzo Barone − *Scuola Normale Superiore, 56126 Pisa, Italy; Istituto Nazionale di Fisica Nucleare (INFN) sezione di Pisa, 56127 Pisa, Italy;* ⓘ orcid.org/0000-0001-6420-4107; Email: vincenzo.barone@sns.it

Authors

Marco Fusè − *Scuola Normale Superiore, 56126 Pisa, Italy;* ⓘ orcid.org/0000-0003-0130-5175

Filippo Lipparini − *Dipartimento di Chimica e Chimica Industriale, Universitaá di Pisa, 56124 Pisa, Italy;* ⓘ orcid.org/0000-0002-4947-3912

Michele Nottoli − *Dipartimento di Chimica e Chimica Industriale, Universitaá di Pisa, 56124 Pisa, Italy*

Giovanni Scalmani − *Gaussian, Inc., Wallingford, Connecticut 06492, United States;* ⓘ orcid.org/0000-0002-4597-7195

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.2c00046

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Cao, L.; Ryde, U. On the Difference Between Additive and Subtractive QM/MM Calculations. *Front. Chem.* **2018**, *6*, 89.

(2) Car, R.; Parrinello, M. Unified Approach for Molecular-Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.

(3) Schlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. III. Comparison with Born-Oppenheimer dynamics. *J. Chem. Phys.* **2002**, *117*, 8694−8704.

(4) Helgaker, T.; Uggerud, E.; Jensen, J. A. Integration of the Classical Equations of Motion on ab initio Molecular-Potential Energy Surfaces Using Gradients and Hessians − Application to Translational Energy-Release Upon Fragmentation. *Chem. Phys. Lett.* **1990**, *173*, 145−150.

(5) Millam, J. M.; Bakken, V.; Chen, W.; Hase, W. L.; Schlegel, H. B. Ab initio classical trajectories on the Born-Oppenheimer Surface: Hessian-Based Integrators using Fifth Order Polynomial and Rational Function Fits. *J. Chem. Phys.* **1999**, *111*, 3800−3805.

(6) Morzan, U. N.; Alonso de Armiño, D. J.; Foglia, N. O.; Ramirez, F.; González Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QM−MM Simulations. *Chem. Rev.* **2018**, *118*, 4071−4113.

(7) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **1994**, *101*, 6141.

(8) Lee, A. J.; Rick, S. W. The effect of charge transfer on the properties of liquid water. *J. Chem. Phys.* **2011**, *134*, 184507.

(9) Carnimeo, I.; Cappelli, C.; Barone, V. Analytical Gradients for MP2, Double Hybrid Functionals, and TD-DFT with Polarizable Embedding Described by Fluctuating Charges. *J. Comput. Chem.* **2015**, *36*, 2271−2290.

(10) Mancini, G.; Brancato, G.; Chandramouli, B.; Barone, V. Organic solvents simulations under non periodic boundary conditions: a library of effective potentials for the GLOB model. *Chem. Phys. Lett.* **2015**, *625*, 186−192.

(11) Del Galdo, S.; Chandramouli, B.; Mancini, G.; Barone, V. Assessment of multiscale approaches for computing UV-vis spectra in condensed phases: towards an effective yet reliable integration of variational and perturbative QM/MM approaches. *J. Chem. Theory Comput.* **2019**, *15*, 3170−3184.

(12) Cancès, E.; Maday, Y.; Stamm, B. Domain decomposition for implicit solvation models. *J. Chem. Phys.* **2013**, *139*, 054111.

(13) Stamm, B.; Lagardère, L.; Scalmani, G.; Gatto, P.; Cancès, E.; Piquemal, J.-P.; Maday, Y.; Mennucci, B.; Lipparini, F. How to make continuum solvation incredibly fast in a few simple steps: A practical guide to the domain decomposition paradigm for the conductor-like screening model. *Int. J. Quantum Chem.* **2019**, *119*, No. e25669.

(14) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885−1891.

(15) Barone, V.; Bencini, A.; Cossi, M.; Di Matteo, M.; Mattesini, A.; Totti, F. Assessment of a combined QM/MM approach for the study of large nitroxide systems in vacuo and in condensed phases. *J. Am. Chem. Soc.* **1998**, *120*, 7069−7078.

(16) Rosa, M.; Micciarelli, M.; Laio, A.; Baroni, S. Sampling Molecular Conformers in Solution with Quantum Mechanical Accuracy at a Nearly Molecular-Mechanics Cost. *J. Chem. Theory Comput.* **2016**, *12*, 4385−4389.

(17) Licari, D.; Fusè, M.; Salvadori, A.; Tasinato, N.; Mendolicchio, M.; Mancini, G.; Barone, V. Towards the SMART workflow system for computational spectroscopy. *Phys. Chem. Chem. Phys.* **2018**, *20*, 26034−26052.

(18) Barone, V.; Puzzarini, C.; Mancini, G. Integration of theory, simulation, artificial intelligence and virtual reality: a four-pillar approach for reconciling accuracy and interpretability in computational spectroscopy. *Phys. Chem. Chem. Phys.* **2021**, *23*, 17079−17096.

(19) Downs, G. M.; Barnard, J. M. Clustering Methods and Their Uses in Computational Chemistry. *Reviews in Computational Chemistry* **2002**, *18*, 1−40.

(20) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(21) Klamt, A.; Schuurmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.

(22) Greengard, L.; Rokhlin, V. A fast algorithm for particle simulations. *J. Comput. Phys.* **1987**, *73*, 325−348.

(23) Lipparini, F. General Linear Scaling Implementation of Polarizable Embedding Schemes. *J. Chem. Theory Comput.* **2019**, *15*, 4312−4317.

(24) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci, B. Fast Domain Decomposition Algorithm for Continuum Solvation Models: Energy and First Derivatives. *J. Chem. Theory Comput.* **2013**, *9*, 3637−3648.

(25) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(26) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. *Intermol. Forces* **1981**, *14*, 331−342.

(27) Stendardo, E.; Pedone, A.; Cimino, P.; Cristina Menziani, M.; Crescenzi, O.; Barone, V. Extension of the AMBER force-field for the study of large nitroxides in condensed phases: an ab initio parameterization. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11697−11709.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Rev. C.01.; Gaussian Inc.: Wallingford, CT, 2016.

(29) Mancini, G.; Del Galdo, S.; Chandramouli, B.; Pagliai, M.; Barone, V. Computational Spectroscopy in Solution by Integration of Variational and Perturbative Approaches on Top of Clusterized Molecular Dynamics. *J. Chem. Theory Comput.* **2020**, *16*, 5747−5761.

(30) Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24−34.

(31) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(32) Mancini, G.; Brancato, G.; Barone, V. Combining the Fluctuating Charge Method, Non-periodic Boundary Conditions and Meta-dynamics: Aqua Ions as Case Studies. *J. Chem. Theory Comput.* **2014**, *10*, 1150−1163.

(33) Whitley, D. A genetic algorithm tutorial. *Stat. Comput.* **1994**, *4*, 65−85.

(34) Lazzari, F.; Salvadori, A.; Mancini, G.; Barone, V. Molecular Perception for Visualization and Computation: The Proxima Library. *J. Chem. Inf. Model.* **2020**, *60*, 2668−2672.

(35) Schrijver, A. *Combinatorial optimization: polyhedra and efficiency*; Algorithms and combinatorics 24; Springer: Berlin and New York, 2003.

(36) Aggarwal, C. C., Reddy, C. K., Eds. *Data clustering: algorithms and applications*; Chapman & Hall/CRC data mining and knowledge discovery series; Chapman and Hall/CRC.

(37) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312−2334.

(38) Mancini, G.; Zazza, C. F429 Regulation of Tunnels in Cytochrome P450 2B4: A Top Down Study of Multiple Molecular Dynamics Simulations. *PLoS ONE* **2015**, *10*, e0137075.

(39) Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1−27.

(40) Macchiagodena, M.; Mancini, G.; Pagliai, M.; Cardini, G.; Barone, V. New atomistic model of pyrrole with improved liquid state properties and structure. *Int. J. Quantum Chem.* **2018**, *118*, No. e25554.

(41) Mancini, G.; Fusè, M.; Lazzari, F.; Chandramouli, B.; Barone, V. Unsupervised search of low-lying conformers with spectroscopic accuracy: A two-step algorithm rooted into the island model evolutionary algorithm. *J. Chem. Phys.* **2020**, *153*, 124110.

(42) Feo, T. A.; Resende, M. G. C. Greedy Randomized Adaptive Search Procedures. *J. Glob. Opt.* **1995**, *6*, 109−133.

(43) Fracchia, F.; Del Frate, G.; Mancini, G.; Rocchia, W.; Barone, V. Force Field Parametrization of Metal Ions From Statistical Learning Techniques. *J. Chem. Theory Comput.* **2018**, *14*, 255−273.

(44) Hart, J. P.; Shogan, A. W. Semi-greedy heuristics: an empirical study. *Oper. Res. Lett.* **1987**, *6*, 107−114.

(45) Savitsky, A.; Nalepa, A.; Petrenko, T.; Plato, M.; Mobius, K.; Lubitz, W. Hydrogen-bonded complexes of neutral nitroxide radicals with 2-propanol studied by multifrequency EPR/ENDOR. *Appl. Magn. Reson.* **2021**, DOI: 10.1007/s00723-021-01442-y.

(46) Becke, A. D. Density-functional thermochemistry. iii. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(47) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.

(48) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.

(49) Zhao, Y.; Truhlar, D. G. Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *J. Phys. Chem. A* **2005**, *109*, 5656−5667.

(50) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027−3034.

(51) Barone, V.; Biczysko, M.; Bloino, J. Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759−1787.

(52) Santra, G.; Sylvetsky, N.; Martin, J. M. L. Minimally Empirical Double-Hybrid Functionals Trained against the GMTKN55 Database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4. *J. Phys. Chem. A* **2019**, *123*, 5129−5143.

(53) Barone, V. Structure, magnetic properties and reactivity of open-shell species from density functional and self-consistent hybrid methods. *Recent Advances in Density Functional Methods* **1995**, *1*, 287−334.

(54) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 methods: theory and benchmarks. *J. Chem. Phys.* **2009**, *130*, 054104.

(55) Peterson, K. A.; Adler, T. B.; Werner, H.-J. Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B−Ne, and Al−Ar. *J. Chem. Phys.* **2008**, *128*, 084102.

(56) Spackman, P. R.; Jayatilaka, D.; Karton, A. Basis set convergence of CCSD(T) equilibrium geometries using a large and diverse set of molecular structures. *J. Chem. Phys.* **2016**, *145*, 104101.

(57) Puzzarini, C.; Bloino, J.; Tasinato, N.; Barone, V. Accuracy and Interpretability: The Devil and the Holy Grail. New Routes Across Old Boundaries in Computational Spectroscopy. *Chem. Rev.* **2019**, *119*, 8131−8191.

(58) Hill, J. G.; Mazumder, S.; Peterson, K. A. Correlation consistent basis sets for molecular core-valence effects with explicitly correlated wave functions: The atoms B−Ne and Al−Ar. *J. Chem. Phys.* **2010**, *132*, 054108.

(59) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Black, J. A.; Doll, K.; Heßelmann, A.; Kats, D.; Köhn, A.; Korona, T.; Kreplin, D. A.;

Ma, Q.; Miller, T. F.; Mitrushchenkov, A.; Peterson, K. A.; Polyak, I.; Rauhut, G.; Sibaev, M. The Molpro quantum chemistry package. *J. Chem. Phys.* **2020**, *152*, 144107.

(60) Adler, T. B.; Knizia, G.; Werner, H.-J. A simple and efficient CCSD(T)-F12 approximation. *J. Chem. Phys.* **2007**, *127*, 221106.

(61) Barone, V.; Fusè, M.; Vieira Pinto, M.; Tasinato, N. A computational journey across nitroxide radicals: from structure to spectroscopic properties and beyond. *Molecules* **2021**, *26*, 7404.

(62) Ceselin, G.; Barone, V.; Tasinato, N. Accurate biomolecular structures by the nano-LEGO approach: pick the bricks and build your geometry. *J. Chem. Theory Comput.* **2021**, *17*, 7290−7311.

(63) Peterson, K. A.; Dunning, T. H. Accurate correlation consistent basis sets for molecular core−valence correlation effects: The second row atoms Al−Ar, and the first row atoms B−Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548−10560.

(64) Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G. CFOUR. A quantum chemical program package. 2016; with contributions from Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Christiansen, O.; Engel, F.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T.-C.; Jonsson, D., J.; Jusélius; Klein, K.; Lauderdale, W. J.; Lipparini, F.; Matthews, D.; Metzroth, T.; Mück, L. A.; O'Neill, D. P.; Price, D. R.; Prochnow, E.; Puzzarini, C.; Ruud, K.; Schiffmann, F.; Schwalbach, W.; Stopkowicz, S.; Tajti, A.; Vázquez, J.; Wang, F.; Watts, J. D., and the integral packages MOLECULE (J., Almlöf; Taylor, P. R.), PROPS (Taylor, P. R.), ABACUS (Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.), and ECP routines by Mitin, A. V.; C., van Wüllen For the current version, see http://www.cfour.de (accessed 23 February 2022).

(65) Aurich, H. G.; Hahn, K.; Stork, K.; Weiss, W. Aminyloxide (Nitroxide) - XXIV. Empirische ermittlung der spindichteverteilung in aminyloxiden. *Tetrahedron* **1977**, *33*, 969−975.

(66) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures and electronic properties of molecules in solution by the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669−681.

(67) Windle, J. J. Hyperfine coupling constants for nitroxide spin probes in water and carbon tetrachloride. *J. Magn. Reson.* **1981**, *45*, 432−439.
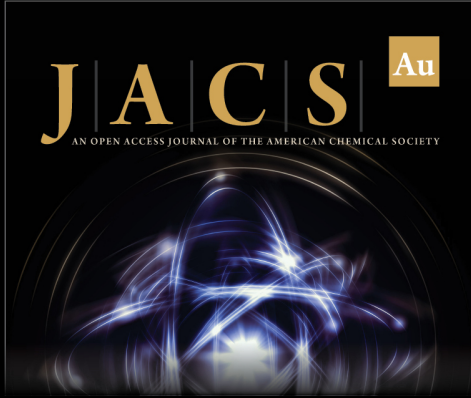
(68) Huseyinova, S.; Blanco, J.; Requejo, F. G.; Ramallo-Lòpez, J. M.; Blanco, M. C.; Buceta, D.; Lòpez-Quintela, M. A. Synthesis of Highly Stable Surfactant-free Cu5 Clusters in Water. *J. Phys. Chem. C* **2016**, *120*, 15902−15908.

(69) Mitrushchenkov, A. O.; Zanchet, A.; Hauser, A. W.; de Lara-Castells, M. P. Nonadiabatic Effects in the Molecular Oxidation of Subnanometric Cu5 Clusters. *J. Phys. Chem. C* **2021**, *125*, 9143−9150.

(70) Martini, P.; Albertini, S.; Laimer, F.; Meyer, M.; Gatchell, M.; Echt, O.; Zappa, F.; Scheier, P. Splashing of Large Helium Nanodroplets upon Surface Collisions. *Phys. Rev. Lett.* **2021**, *127*, 263401.

(71) Fernández-Perea, R.; Gómez, L. F.; Cabrillo, C.; Pi, M.; Mitrushchenkov, A. O.; Vilesov, A. F.; de Lara-Castells, M. P. Helium Droplet-Mediated Deposition and Aggregation of Nanoscale Silver Clusters on Carbon Surfaces. *J. Phys. Chem. C* **2017**, *121*, 22248−22257.

(72) A Python code for the calculation of $g(r)$ and other structural properties under NPBCs is available on GitHub at https://github.com/tuthmose/NPBCpy (accessed 23 February 2022). Scripts for clustering are available at https://github.com/tuthmose/Clustering (accessed 23 February 2022). The alpha version of the Proxima library is downloadable at: https://bitbucket.org/sns-smartlab/proxima/src/master/ (accessed 23 February 2022)