Reciprocal Polarizable Embedding with a Transferable H_2O Potential Function II: Application to $(H_2O)_n$ Clusters & Liquid Water

Asmus Ougaard Dohn,* Elvar Örn Jónsson, and Hannes Jónsson

Science Institute, University of Iceland, Reykjavík, Iceland

E-mail: asod@hi.is

Abstract

The incorporation of polarization in multiscale quantum-mechanics / molecular-mechanics (QM/MM) simulations is important for a variety of applications, e.g. charge-transfer reactions. A recently developed formalism based on a density functional theory description of the QM region and a potential energy function for H_2O molecules that includes quadrupole as well as dipole polarizability of the MM region is used to simulate liquid water and water clusters. Analysis of the energy, atomic forces, MM polarization, and structure is presented. A quantitative assessment of the QM/MM-MM/MM interaction energy differences of all possible QM/MM configurations of $(H_2O)_n$ clusters shows that the interquartile range of the distributions of the QM/MM binding energies are never more than 20 meV/molecule higher or lower than the binding energies produced with either of the single-model results. Comparing these interaction energy differences with the QM/MM induction differences show that they are not systematically caused by the induced MM moments of our polarizable embedding scheme. Optimized hexamer geometries as well as liquid water structure are shown to be improved in comparison with results obtained using point-charge based embedding models neglecting polarization.

1 Introduction

The multiscale quantum and classical mechanical (QM/MM) approach to simulations of large and complex systems is well-established. Different flavours have been developed with a variety of application areas in mind, such as photochemistry,[1–3] solvation dynamics,[4–7] biology,[8,9] nanostructures,[10] and materials science.[11] In the case of heterogeneous electrochemistry, such a multiscale methodology based on a combination of approaches is needed in order to make simulated systems more realistic, e.g. by representing the aqueous electrolyte as a water phase rather than a bilayer of ice, as has frequently been done.[12–15] Experimentally, electrochemical reactions at stepped / irregular surfaces have been observed to be different from those at flat surfaces[16–18] and ordering of H_2O molecules at such surfaces needs to be coupled with a proper simulation of the liquid phase. Furthermore, analysis of solvated systems requires the determination of free energy based on statistical sampling[19,20] that is too computationally intensive for an electronic structure calculation of the whole system, for example at the level of density functional theory (DFT). A multiscale model that includes electronic structure calculation of the electrode and its near environment, a QM region, while the aqueous electrolyte is otherwise represented with a less computationally intensive MM approach would represent a large step forward in the simulation of heterogeneous electrochemistry. However, before introducing surfaces into the simulations, the accuracy of the coupled methodology must be thoroughly tested.

In most present day QM/MM simulations, the QM calculation incorporates an added Coulomb potential from static point charges of the MM subsystem.[3,21–23] This approximation neglects the reciprocal polarization of the charges in the MM subsystem by the QM subsystem, an effect that was nevertheless included in the inceptive work within the field.[24] Strictly speaking, this reciprocal polarization is still of electrostatic nature, but often the rungs on the 'Jacob's ladder' of multiscale embedding methods are named 'electrostatic embedding' (EE) if the MM point charges remain unchanged during the calculation, and the term 'polarizable embedding' (PE) is used if MM polarization is included.[25–28] This nomen-

clature is thus also adopted here.

Already, quite a few PE-QM/MM methods have been developed. One way of classifying them relies on how they account for polarization and polarizability of the MM subsystem.[29] This includes: induced dipoles,[27,30–34] Drude oscillators,[35] fluctuating charges,[36,37] charge-on-spring models,[38] or atom-centered multipoles.[39–43] Some of the methods have been extended to utilize QM models beyond DFT.[44–46] We have recently developed a globally self-consistent polarizable embedding scheme that is unique in that it couples a polarizable force field for water including both di- and quadrupole induction as well as a non-local exchange repulsion.[47] The force field is based on a single center multipole expansion (SCME) of the electrical potential and its gradients, and as such defines a set of molecular moments that that can be extracted from high-level quantum chemistry calculations.[48] The QM subsystem is represented by DFT and thus the electric field and its derivatives are based on an explicit charge density, evaluated on a grid. Using a multipole model to describe the electrostatics in the MM subsystem is more flexible than e.g. using a point charge model, and the resulting electric field and its derivatives will be more similar to their DFT-counterparts, which reduces the risk of introducing artifacts when coupling the two subsystems.

To date the often computationally costly methodologies for polarizable embedding have impeded their application to problems requiring sampling of the free energy of large systems: Either, the sampling has been carried out on structures generated from classical dynamics based on pure MM description instead of using the QM/MM approach for the dynamics,[27,49] or else the sampling has been done with more intricate schemes than simple Born-Oppenheimer Molecular Dynamics (BOMD).[33]

We present here a comprehensive and critical analysis of the accuracy of the energy, atomic forces, and polarization produced by our recent embedding scheme in the case of H_2O clusters as well as a statistically robust BOMD sampling of liquid water. The hydrogen network produced in the simulation and the way it is affected by the inclusion of polarization in our PE scheme is analyzed by comparison to simulations using the more ubiquitous EE

model.

To assess the accuracy of coupling schemes, one must test if they introduce artifacts in the total energy and atomic forces in the full, coupled system. These artifacts can lead to: (1) The energy of the coupled system being smaller or larger than what is obtained from either the pure QM or pure MM description, and (2) different sub-divisions into QM and the MM regions giving substantially different results.[21] This means that we are not testing the accuracy of the individual single-scale models, but rather the quality of the coupling of the two descriptions. Thus, the density functionals chosen for these test have been selected for their efficiency, allowing for more rapid collection of statistically convergent quantities to assess the coupling scheme, and not for their ability to reproduce accurately all aspects of liquid water and H_2O clusters.

2 Computational Details

The PBE[50] and BLYP[51,52] functionals have been used for testing our polarizable coupling model. A real space grid is used to represent the electron density and the orbitals, as implemented in the GPAW[53,54] software, and a grid spacing of 0.15 Å was used.

To test over- or underbinding of QM/MM systems in relation to the total difference between pure QM and SCME models, we calculated the interaction energy for all possible $n!/(q!(n-q))!$ QM/MM combinations of n-molecule H_2O clusters, of which $q = \{1 \ldots n-1\}$ are described with DFT. For example, in the case of the trimer, $n = 3$ and q can be 1 or 2, for a total of 6 QM/MM combinations for each trimer geometry. For the decamer, the largest clusters analysed here, the number of combinations is 1022. The datasets were produced from the cluster geometries of the Bates & Tschumper hexamers,[55] and the set of $(H_2O)_n$, $n = 3-10$ clusters provided by Temelso and coworkers,[56] for a total of 6214 QM/MM single point energy calculations for each of the two density functionals used.

The hexamer QM/MM RMSD calculations were carried out by obtaining the dataset,[55]

resetting the intramolecular bonds and angles to the SCME geometries, fixing them using RATTLE, and then relaxing each hexamer using SCME exclusively, until the magnitude of the maximum force on any atom drops below 0.02 eV / Å. Then, the relaxations were re-done for each possible QM/MM configuration of 1-5 QM molecules, 496 in total. Every relaxed structure was then re-centered and rotated to optimize the overlap between the SCME and PBE/SCME structure, using the Kabsch algorithm,[57] before the RMSD values were calculated.[58] The results were compared with EE calculations using the TIP4P potental function for H_2O, i.e. PBE/TIP4P calculations, using the same procedure. It has previously been shown that the most accurate QM/MM coupling is achieved by using a basis of linear combinations of atomic orbitals (LCAO).[59] Thus, to ensure the best possible results from this EE model for fair comparisons, GPAW's LCAO mode with tzp basis sets for the QM molecules was used in the simulations.

For the simulations of liquid water, a system of 512 H_2O molecules in cubic box including periodic boundary conditions at ambient density was equilibrated using either SCME or TIP4P for the PE and EE-simulations, respectively. A 1 fs timestep was used in classical dynamics simulations using a Langevin thermostat with a friction coefficient of 0.01. For the QM/MM simulations, one of the H_2O molecules was described with either the PBE or BLYP functional, using all the same parameters as in the previous dynamics calculations. The velocities were re-initialized from a Maxwell-Boltzmann distribution at 300 K and the system re-equilibrated for 3 ps before data was collected. In the PBE/SCME calculation, data was collected over a time interval of 159 ps, while in the BLYP/SCME calculations, the time interval was 132 ps. The total time interval covered by the concatenated PBE/TIP4P simulation after equilibration was 418 ps. The single-model PBE liquid water calculations consisted of 10 systems, each containing 64 molecules, at ambient density, first equilibrated with TIP4P for efficiency, and then re-equilibrated using PBE BOMD for 4 ps each, before sampling data over a time interval of 58 ps, all using the same Langevin thermostat as in the other simulations. The time periods for re-equilibration were tested and found to be

sufficient by extending them further, discarding data, and confirming that this induced no significant changes in the statistical results.

3 Benchmarks

3.1 $(H_2O)_n$ Clusters

Figure 1 shows box/violin plots of distributions of differences in the interaction energy $\Delta\Delta E = \Delta E_{\text{int}}^{\text{QMMM}} - \Delta E_{\text{int}}^{\text{SCME}}$ between every QM/MM combination possible from the set of clusters described in the section on computational details. The white boxes represent 50% of the data points, and the red lines within them show the median of each distribution. The widths of the underlying violin-plots is a visualization of the amount of data points in that region. When $\Delta\Delta E$ is positive, the pure SCME interaction energy is more negative than its QM/MM counterpart, or, equivalently, the QM/MM description underbinds compared to pure SCME. The colored patches represent the biggest pure DFT interaction energy difference with respect to SCME. The left plot in figure 1, shows the PBE/SCME coupling interaction energy differences. The differences between the pure PBE and pure SCME interaction energy is small, which result in correspondingly small differences in the QM/MM interaction energy, that is generally seen to converge towards the absolute maximum single-model results (as represented by the edges of the red/blue patches) with the number of molecules in each subsystem. Previous point-charge based electrostatic embedding models have had issues with too strong QM/MM coupling that could not be removed by tuning the non-electrostatic coupling parameters. This is evident from 'u-shaped' $\Delta\Delta E$-averages, binding strongest when there is the largest possible number of QM/MM interactions in the configuration.[59] From the results presented here, these problems have been eliminated by the more advanced coupling model including the reciprocal polarizability.

For larger ($n > 7$) clusters, the combinations containing only 1 or 2 PBE molecules does show some overbinding on the order of a couple of meV per water molecule. This can be

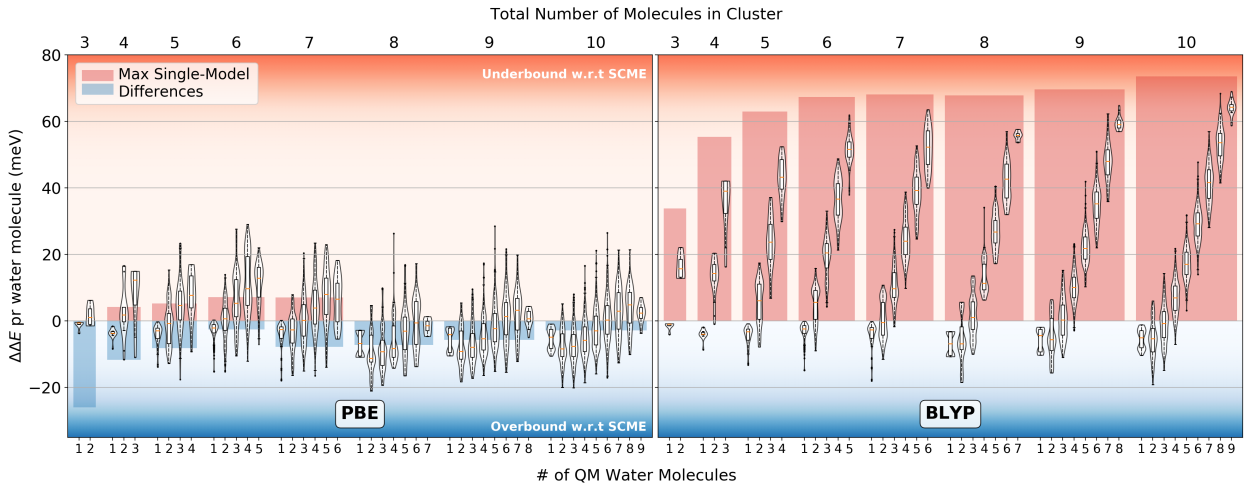

Figure 1: Box/violin plot of the distributions of differences in DFT/SCME interaction energy with respect to the pure SCME interaction energy for water clusters ranging from tri- to decamer. The x-axis groups the results of all QM/MM combinations containing 1,2, $\ldots n-1$ QM molecules together, where n is the total number of molecules in a cluster. The red/blue patches represent the maximum positive/negative differences between the the two single-model results, $\Delta E_{\text{int}}^{\text{QM}} - \Delta E_{\text{int}}^{\text{SCME}}$, delineating the artifact-free region of interaction energy differences.

ascribed to the well known overestimation of polarizability by PBE (and GGA functionals in general, arising from their underestimation of the HOMO-LUMO gap),[60,61] which opens up for later studies using hybrid functionals, since the inclusion of exact exchange is known to improve polarizability,[61] or self-interaction corrected functionals[62] (since the error in the HOMO-LUMO gap in GGA functionals can be traced to the self-interaction error in Kohn-Sham functionals). When the BLYP functional is used instead of the PBE functional (see figure 1, right), the difference between pure DFT and pure SCME interaction energy is generally larger, as shown by the red patches.

It is well known that GGA functionals have several, but different, shortcomings when describing water.[61] This difference can be traced to the different exchange-enhancement functions. The two functionals compared here represent extrema in this regard. BLYP is strongly repulsive in the exchange-overlap region (i.e. the region where electron densities between water molecules overlap) resulting in underbound water clusters when comparing results from pure BLYP to results from pure SCME simulations (red patches in figure 1,

right). On the other hand, PBE is more weakly repulsive in the exchange-overlap region and as a result the pure-PBE water cluster binding energy is closer to the pure SCME result. Both functionals lack the dispersion interaction at long range but it is included in the SCME potential function.

On the basis of these benchmarks it can be seen that the polarizable embedding interface does not introduce any significant errors in the energy outside of the range spanned by the pure DFT and pure SCME results.

3.1.1 Analysis of QM/MM induced multipoles

In the following we will analyse how the QM density affects the polarization of the SCME molecules, compared to the corresponding pure-SCME system. Dipole moments also induced by the QM density will be called 'QM/MM dipoles' for brevity. Similarly to $\Delta\Delta E$, the difference between the norms of each of the final induced dipoles of each molecule in the QM/MM and pure MM description can be defined as

$$\Delta\Delta\mu_\alpha^i = \frac{|\mu_\alpha^{i,\mathrm{QMMM}} + \Delta\mu_\alpha^{i,\mathrm{QMMM}}| - |\mu_\alpha^{i,\mathrm{MMMM}} + \Delta\mu_\alpha^{i,\mathrm{MMMM}}|}{|\mu_\alpha^{i,\mathrm{MMMM}} + \Delta\mu_\alpha^{i,\mathrm{MMMM}}|}. \tag{1}$$

Similarly, using the 2-norm for the quadrupoles, $\Delta\Delta\theta$ can be defined. Lastly, the angle, ΔAngle, between each QM/MM- dipole and its MM equivalent is analysed to characterize the influence of the coupling.

Figure 2 shows how these values are affected by the QM/MM partitioning. The underlying idea is the same as for the box graphs of the previous section, but here the top graph shows the differences in norm between every possible QM/MM-induced SCME dipole and its pure SCME-based equivalent. The bottom graph shows the angle between each pure SCME dipole and its QM/MM equivalent. The lines represent the median values, while the underlying heatmaps are created from histograms of the values, since violin- or boxplots would be too cluttered to facilitate an easy visual comparison between results from the two energy

functionals. Both functionals introduce virtually the same $\Delta\Delta\mu$ in SCME, consistent with the fact that GGA functionals generally overestimate induction,[61] resulting in overpolarized QM densities, which in turn produces larger fields that polarizes SCME than in pure-SCME clusters.

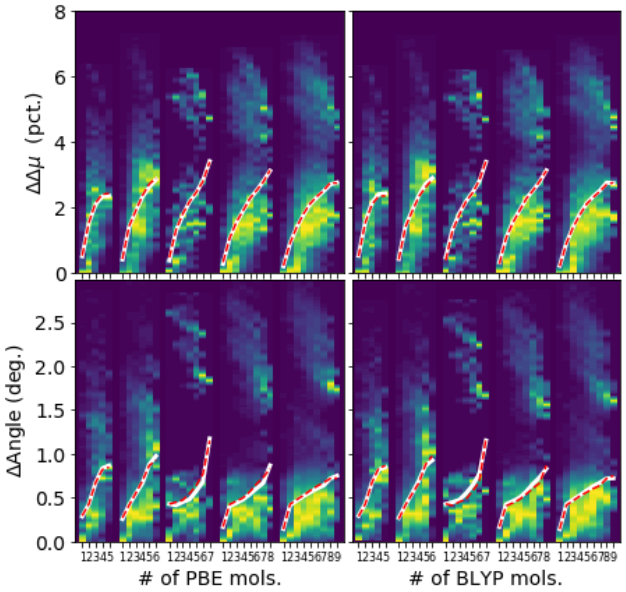

Figure 2: Top left: Median dipole norm differences (equation 1) between each SCME dipole of every possible PBE/SCME configuration of hexamer- and larger clusters, and its pure SCME equivalent. Top right: The same for BLYP/SCME. Bottom left: Median angle between aforementioned PBE/SCME dipole pairs. Bottom right: The same for BLYP/SCME. The red, dashed lines show the median of the other functional for comparison, i.e, on the left (PBE/SCME) graphs the red line is the BLYP/SCME median, and vice versa.

In order to investigate how the distance between QM and MM molecules affects the overall polarization and energy, figure 3 correlates $\Delta\Delta E$ (top graphs) and the induced dipole norm difference with respect to pure SCME (middle graphs) to the distance of the nearest QM molecule. In both cases, we observe that the differences from pure SCME simulations can only reach the largest values if there is at least one neighboring QM molecule, i.e. within 3.25 Å. The bottom graphs correlate the two values directly to each other, revealing no apparent correlation between the induction difference and the $\Delta\Delta E$.

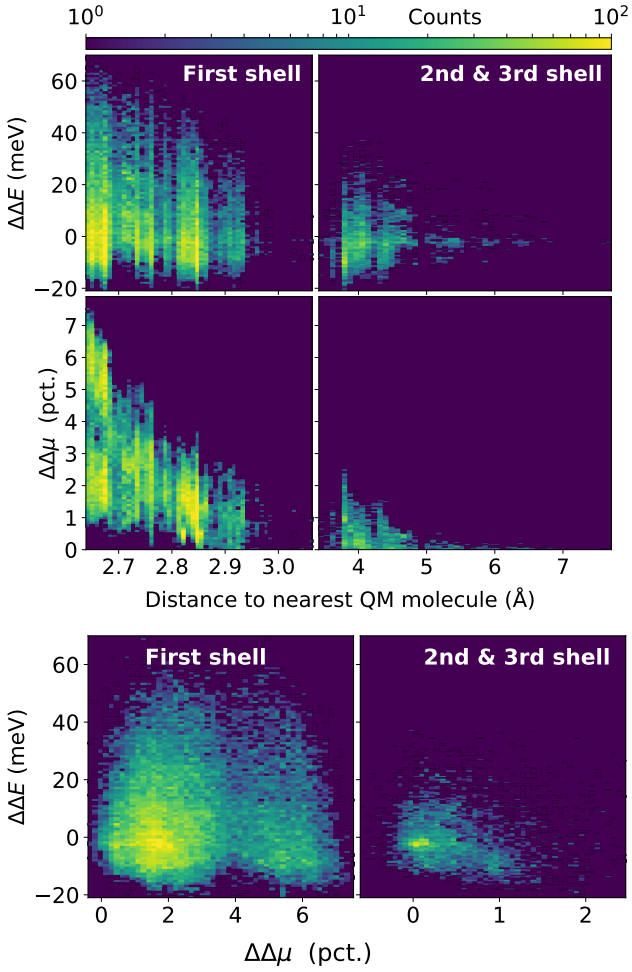

Figure 3: Correlation plots showing the impact of being close to a QM molecule. Top: The $\Delta\Delta E$-values from figure 1, repeated for each QM/MM induced multipole, and histogrammed versus the distance to the nearest QM water molecule from the multipole. The horizontal axis have been split up in two regions for easier assesment of the short-range. Middle: QM/MM Induction difference to the pure-SCME equivalent. Bottom: Correlation between the induction difference and $\Delta\Delta E$.

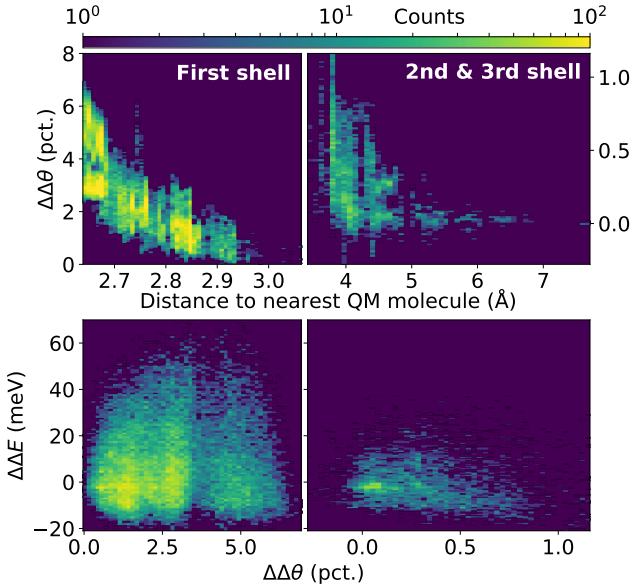

Figure 4: Correlation between the difference in norms of the QM/MM-induced quadrupoles and the pure-SCME quadrupoles (top graph), and $\Delta\Delta E$ (bottom graph).

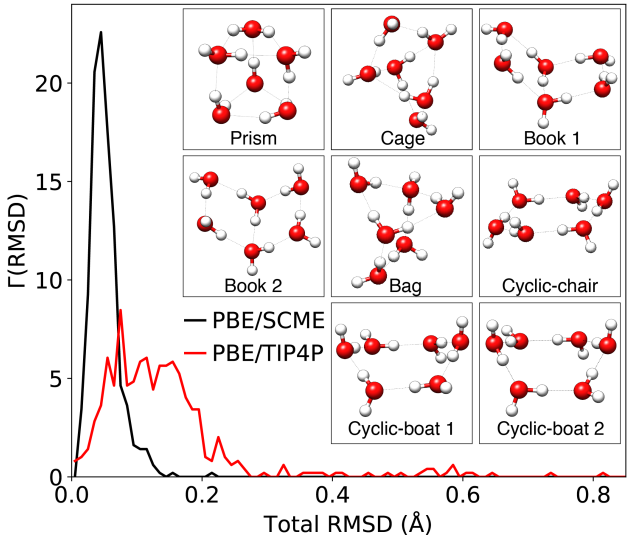

Figure 5: Geometry optimizations of all 496 possible QM/MM combinations of eight of the most energetically favorable hexamer conformations.[55] The histograms show the total RMS differences between all atomic positions of each of the QM/MM relaxed structures, and their pure-MM-relaxed counterpart. The insets depict the used hexamer geometries.

3.1.2 Hexamer structure

The hexamer is the smallest cluster size that can be taken to represent the hydrogen-bonding network in ice, with some of the minimum energy-geometries governed by a fully realized

three dimensional hydrogen bonding network. We therefore use it to test relaxed geometries obtained from the QM/MM description of the atomic forces. Figure 5 shows root mean square displacements (RMSD) obtained after relaxations of all possible QM/MM partitions of the 8 lowest-energy hexamer geometries, starting from the Bates & Tschumper structures,[55] and using either a PE scheme with PBE/SCME or an EE scheme with PBE/TIP4P. Comparing the EE and the PE method reveals a highly pronounced decrease in both the average RMSD and the variance of the distribution when using the PE scheme, which also eliminates almost all total structure-differences above 0.2 Å.

3.2 Liquid Water

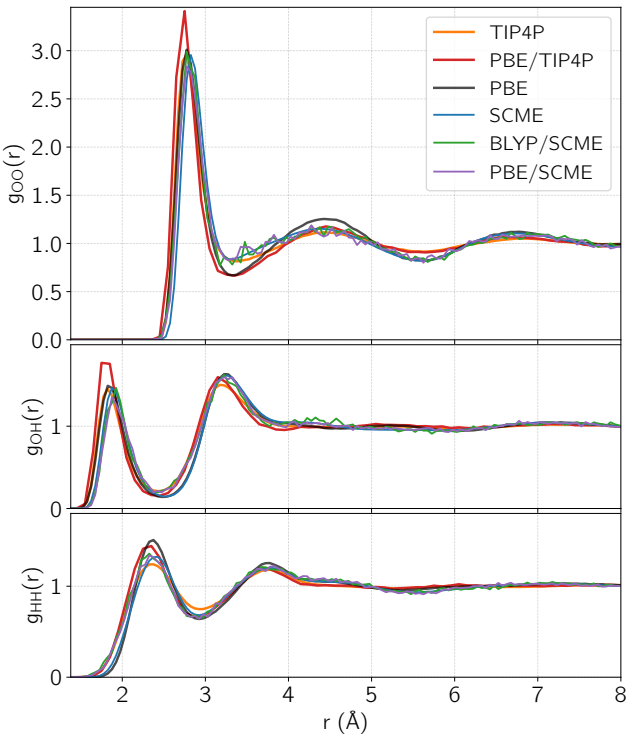

Figure 6: Radial distribution functions (RDFs) of liquid water for the O-O (top), O-H (middle), and H-H (bottom) correlations. The RDFs produced from coupled potentials are sampled from correlations between a single QM molecule to the rest of the 511 MM molecules. The single-model RDFs have been sampled using all possible pairwise correlations. For the PE-QMMM simulations, the total simulation times sampled were 159 ps and 132 ps for the PBE/SCME and the BLYP/SCME production runs, respectively. For PBE/TIP4P we sampled the RDFs over a total of 418 ps, while the pure PBE RDFs were sampled for 58 ps.

Figure 6 compares the pairwise radial distribution functions (RDFs) from simulating liquid water using single-model PBE (black), SCME (blue) to PE-coupled potentials with a single molecule being described using either PBE (purple) or BLYP (green), and the rest with SCME. RDFs from EE-coupled PBE/TIP4P (red) are also shown.

It is well known that classical dynamics and Monte Carlo simulations based DFT with GGA functionals produce RDFs for liquid water that are over-structured as compared with experimental measurements (see e.g.[61] and references therein). This is to some extent because quantum delocalization of the atomic nuclei is not taken into account,[63] but to a larger extent it is the result of too large polarizability of the H_2O molecules due to self-interaction error. We compare here the RDFs obtained classical trajectories using SCME, PBE or QM/MM descriptions with either PBE or BLYP functionals for describing the QM region and either SCME or TIP4P description of the MM molecules. The O-O RDF is indeed over-structured in the PBE and PBE/TIP4P simulations. The largest O-O RDF discrepancies between coupled and single-model results are found at the first minimum of the curve, where the difference between the pure PBE and pure SCME results is also largest. However, also in the liquid phase, the coupled structure is within the two extremes of the single model results. The largest outlier is the PBE/TIP4P RDF produced using our implementation of the more ubiquitous EE scheme.[59] Since the peak of the PBE/TIP4P RDF is higher than both single-models, some coupling-introduced artifacts seem unavoidable in electrostatic embedding. These artifacts are eliminated by the PE model as is evident from the PBE/SCME and BLYP/SCME results.

The O-H RDF is also of importance since the first peak is directly connected to the properties of the H-bond between the donor H-atom and the acceptor O-atom. Mirroring the situation of the O-O correlations, the overstructuring artifact from electrostatic embedding is eliminated by the PE using SCME.

For the intermolecular H-H correlations, the PBE/SCME structure seems to follow the pure SCME structure closer than the pure-PBE, but still within the two single-model limits.

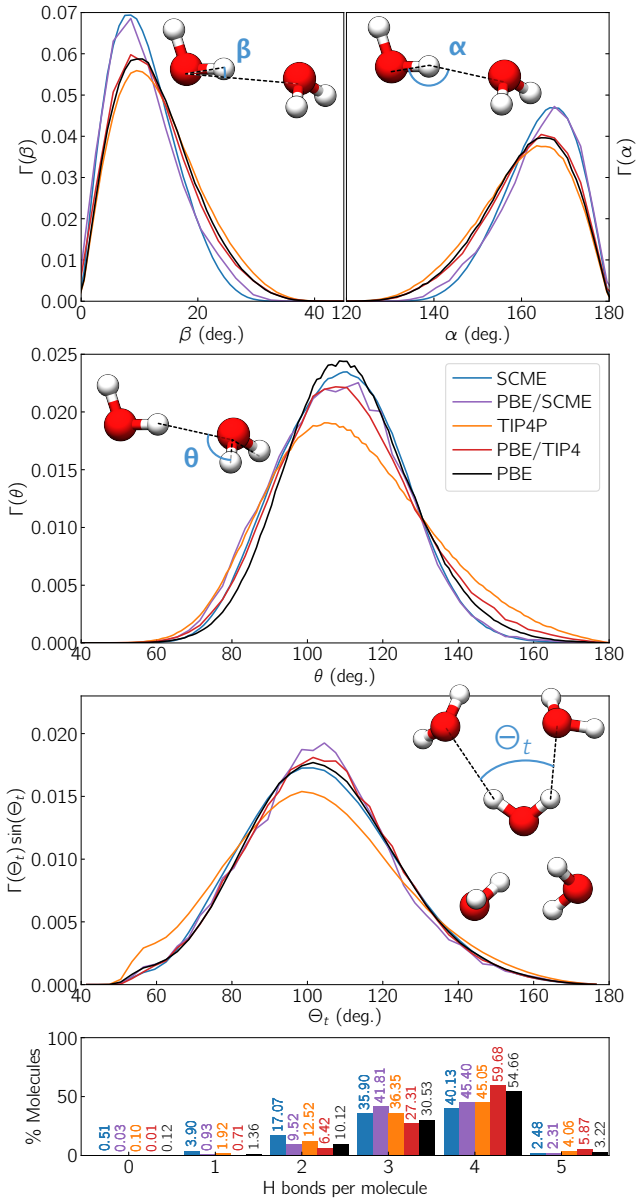

Figure 7: Top three graphs: Probability distributions of the most important angles and dihedrals that define the structure of liquid water. Top left: $\beta = \angle(\mathrm{H_D}, \mathrm{O_D}, \mathrm{O_A})$. Top right: donor angle $\alpha = \angle(\mathrm{O_D}, \mathrm{H_D}, \mathrm{O_A})$. Upper middle: acceptor angle $\theta = \angle(\mathrm{H_D}, \mathrm{O_A}, \mathrm{H_A})$. Lower middle: tetrahedrality angle $\Theta_T = \angle([\mathrm{O_A}, \mathrm{H_D}], [\mathrm{H_{D'}}, \mathrm{O'_A}])$. Bottom: Histograms of the average number of hydrogen bonds per water molecule, sampled via a geometric definition.[64]

Going beyond distance-distributions, figure 7 shows the results of sampling the angles often used in the analysis of liquid water structure.[64–66] The sampling is carried for the first solvation shell as defined by the location of the first minimum of the O-O RDF in figure 6 and the geometric H-bond criterion of Wernet *et al.*[64] is used for counting the hydrogen bonds. Many definitions exist in literature, but for this analysis, the important thing is not so much which convention is used, as long as its the same throughout. For the β-angle, the PBE/SCME results end up being closer to the pure SCME result, than the pure PBE result, indicating that a single PBE molecule embedded in SCME ends up assuming the (angular) structure of the SCME potential energy function. This pattern repeats itself for the PBE/SCME α-angle, but interestingly not for the PBE/TIP4P, which is closer to the pure PBE result.

For the acceptor angle, θ, the distribution from the PE-model is again similar to the pure SCME result, while the peak of the EE-coupled results is significantly higher, and closer to the rest of the results than to TIP4P. For point-charge based force fields, the geometry of the hydrogen-bonded pair dictates that rotating the accepting molecule has some effect on the position of the hydrogen-charges in relation to the donating molecule, but the distances between the oxygen-attributed charge on the hydrogen-accepting molecule and the charges on the donating molecule are only altered very slightly, which could explain the less structured TIP4P distribution. In EE-coupled geometries with a PBE-modeled H-acceptor molecule, the whole electronic density of the hydrogen-accepting molecule has to be taken into consideration, which would re-introduce the structuring in this configuration, but not necessarily in the reverse configuration, thus possibly causing coupling-introduced artifacts. The PE-results do not show this discrepancy.

Following DiStasio *et al.*,[66] we have sampled the distribution of three-body oxygen-oxygen-oxygen angles, $\Gamma(\Theta_t)$, which gives insight into the tetrahedrality of the water structure, going beyond pairwise correlations. For each frame of the simulations, the angle was sampled if two oxygen atoms were within the distance corresponding to an O-O coordina-

tion number of 4. For this three-body interaction, the PBE/SCME peak is slightly higher than both single-scale models, but all three curves show more similarity than there is found between TIP4P, PBE/TIP4P and PBE. For PBE/TIP4P, moving the oxygen-related point charges radially around the QM H-atoms forming the hydrogen bonds to change the tetrahedrality angle does not greatly affect the distances between the point charges and the grid points close to the QM H-atoms. These distances determine the strength of the EE external potential and it is thus not sensitive to this angle. This most likely cancels out some of the other overbound features of the EE water structure.

The number of H-bonds per molecule is shown in the bottom graph of figure 7. Comparing SCME to PBE/SCME, the main change is a $\sim$6% increase in the number of molecules with 3 hydrogen bonds, whereas the for EE the amount of triple-bonded water molecules is reduced by almost 9%. The PBE/TIP4P shows even larger differences from the pure TIP4P sampling: a $\sim$15% increase in tetrahedrally bonded molecules. The percentage of PE-coupled tetrahedrally coordinated water molecules is within the two extremes spanned by the single-scale models, whereas PBE/TIP4P is not. Both coupled models fall slightly outside their respective single-model limits in the 3 hydrogen bond case.

During the dynamics simulations, we sampled the induced SCME- di- and quadrupoles at each timestep to profile the magnitude of the induced moments in the liquid phase. Figure 8 shows these profiles by radially averaging the pole norms outward from the central molecule in the simulations. The first peak rises abruptly when the O-O distance gets larger than the minimum sampled O-O distance in the liquid (at the grow-in of the O-O RDF in figure 6), as the closest water molecules will experience the most polarization. The effect of the higher-than-bulk-density of the second solvation shell is also observed, with a less pronounced peak growing in after approximately 3.2 Å. Both di- and quadrupole profiles of the SCME and PBE/SCME runs are similar with the main difference being a roughly 7% higher peak in the pure SCME quadrupole profile. These short distance averages have the least sampling statistics, since molecules rarely come this close. The first bin is based on an average of

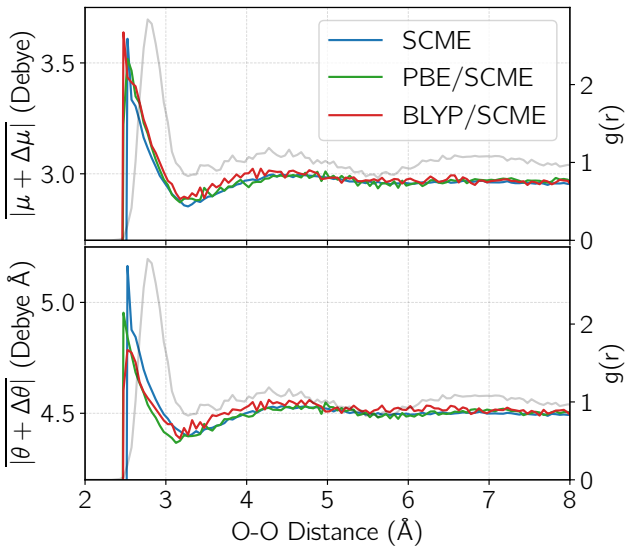

Figure 8: Average norms of the di- and quadrupole moments (top and bottom graphs respectively), sampled radially outward from the central water molecule in the simulations. The radial bins are 0.1 Å wide. The O-O RDF of PBE/SCME is overlaid on both plots, to indicate the probability of encountering an SCME-water molecule at the various distances, compared to the bulk liquid density.

only 15 samples and are thus much more prone to statistical noise, compared to the bulk-sampling counts which are in the tens of thousands in total. All in all, the conclusion is that the SCME-modeled liquid environment responds similarly to induction from the QM electronic density as the induction from other SCME multipoles.

The three-dimensional arrangement of the neighbours of a molecule in the liquid is rotationally averaged out in the calculation of the RDF, $g(r)$, and the angle-distributions do not provide simultaneous information about the distance to the neighbours. However, both quantities can be effectively captured by the average oxygen and hydrogen pairwise spatial distribution functions (SDFs),[67,68] $G_{\mathrm{OX}}(\mathbf{r})$ between oxygen (O) and hydrogen or other oxygens (X). The relationship between $G_{\mathrm{OX}}(\mathbf{r})$ and $g_{\mathrm{OX}}(r)$ is such that $g_{\mathrm{OX}}(r)$ can be obtained by radially integrating the SDF:

$$g_{\mathrm{OX}}(r) = \frac{1}{4\pi r^2} \int G_{\mathrm{OX}}(\mathbf{r}')\delta(|\mathbf{r}'| - r)d\mathbf{r}'. \tag{2}$$

Therefore, the integral, when multiplied with the bulk density ρ_0 must sum up to the number of water molecules in the simulation N, minus the central molecule from which the sampling takes place:

$$\rho_0 \int G_{\mathrm{OX}}(\mathbf{r}) d\mathbf{r} = N - 1, \tag{3}$$

and can thus be numerically binned from aligned (BO)MD trajectories, and normalized using the above. This will make the isovalues represent the ratio with which the density is larger than in the bulk liquid. Since SDFs due to their lack of rotational averaging require a lot more data to reduce sampling noise, we have employed the same hydrogen-bond criterion[64] as used for counting the hydrogen bonds in figure 7, before sampling, to reduce the noise and allow production of SDFs from the shorter QM/MM simulations. The SDFs where sampled in cubic bins of 0.3 Å sides, and later interpolated to 0.1 Å using a cubic spline. The first two rows of figure 9 show the (hydrogen bond-filtered) SDFs $G_{\mathrm{OO}}(\mathbf{r}) = 4$ (red) and $G_{\mathrm{OH}}(\mathbf{r}) = 2$ (white) for (PBE/)TIP4P and for (PBE/)SCME. The row below shows the difference in densities $G_{\mathrm{OO}}^{\mathrm{QMMM}}(\mathbf{r}) - G_{\mathrm{OO}}^{\mathrm{MM}}(\mathbf{r})$. In both cases, the only systematic differences in the densities are increases. In the graph, the difference densities have been renormalized to 1 so isovalues corresponding to the same percentage of each total increase can be compared visually. The graph shows the isovalue corresponding to a 15% increase in density going from the MM density to the QM/MM density. Both coupling-models show increase in the coupled densities for the hydrogen-accepting O-atoms at the hydrogen-ends of the central molecule, although the polarizable coupling density difference 15% isovalue is more localized. However, most noticeable is the additional density increases in the electrostatic coupling using the TIP4P force field, where we observe new difference-volumes correlated with the locations of the hydrogen-accepting lone-pairs on the central molecule that are not present in the PE. This finding is consistent with the shortcoming of the coupling model that is based on static point charges in describing the lone pairs on the acceptor molecule, as illustrated

via the ω-angle dimer potential energy curve in the bottom row of figure 9. In the case of the dimer, the EE coupling fails entirely to reproduce the rotation-barrier as the acceptor molecule is modeled with TIP4P. The results in figure 9 show how this is carried over into the structure of liquid water and how it is amended by our PE coupling scheme.

Conclusions

A thorough benchmark of the accuracy of our newly formulated PE scheme is presented, on its energies, atomic forces, and on the way the electrostatic potential from the QM subsystem and its gradients induce polarization in the SCME force field. For the quantitative energetic benchmark on tri- to decamer water clusters, we see that the QM/MM scheme is accurate in that the QM/MM binding energy systematically converges to the respective single-scale limits, QM or MM, as should be the case for a well-behaved interface. Analysing the polarization of the di- and quadrupoles from in the QM/MM systems reveals that the two GGA functionals produce very similar polarization characters of the MM subsystem, and thus, the differences between the $\Delta\Delta E$-distributions of the two functionals must be mostly due to the different semi-local approximations. This is further supported by the finding of no direct, linear correlation between $\Delta\Delta E$ and $\Delta\Delta\mu$, while the largest $\Delta\Delta E$- and $\Delta\Delta\mu$ values are mostly found for the smallest QM-MM distances.

The tests of the hexamer structures show that the PE scheme greatly reduces both the average RMSD and the width of its distribution of QM/MM optimised structures as compared to the EE scheme. Similarly, tests of the structure of liquid water show that the overstructuring of the RDFs obtained with EE is eliminated by the PE scheme. For the angular distributions, the PE scheme produces QM/MM angles that follow the SCME angular structure more closely than PBE, and shows a small increase in the tetrahedrality angle, outside of the two single-model limits, which, perhaps is not seen in EE due to fortuitous error cancellation from the simple point-charge model for the lone pairs on the two TIP4P

acceptor molecules that form the angle with the H-donating QM water molecule. However, the EE model produces too many tetrahedrally coordinated water molecules overall, whereas the PE model does not. Lastly, analysing angles and distances simultaneously, by looking at the 3D spatial distribution functions, we observe how this improved model for the water lone pairs eliminates structural artifacts in the coupled water 3D structure.

All in all, we have shown that our PE coupling method using the SCME potential energy function increases the coupling accuracy and greatly reduces artifacts, thereby opening new doors for application and further development. One example is in semi-adaptive QM/MM schemes suhc as FIRES,[35] which is very efficient in that it does not require more QM calculations per step as do many other schemes, but is unfortunately hindered by structural issues for coupled potentials.[69] Perhaps with PE, the artifacts have been reduced enough to revitalise interest into FIRES and similar methods. Work in this direction is ongoing. Another example involves the new, more advanced handle on the (changes in the) electronic density as modelled by the SCME multipoles. This paves the way for dynamically updating non-electrostatic coupling potentials, describing exchange-repulsion in a more general and flexible manner, while still avoiding the need for e.g. explicit density-fitting or other expensive methods.

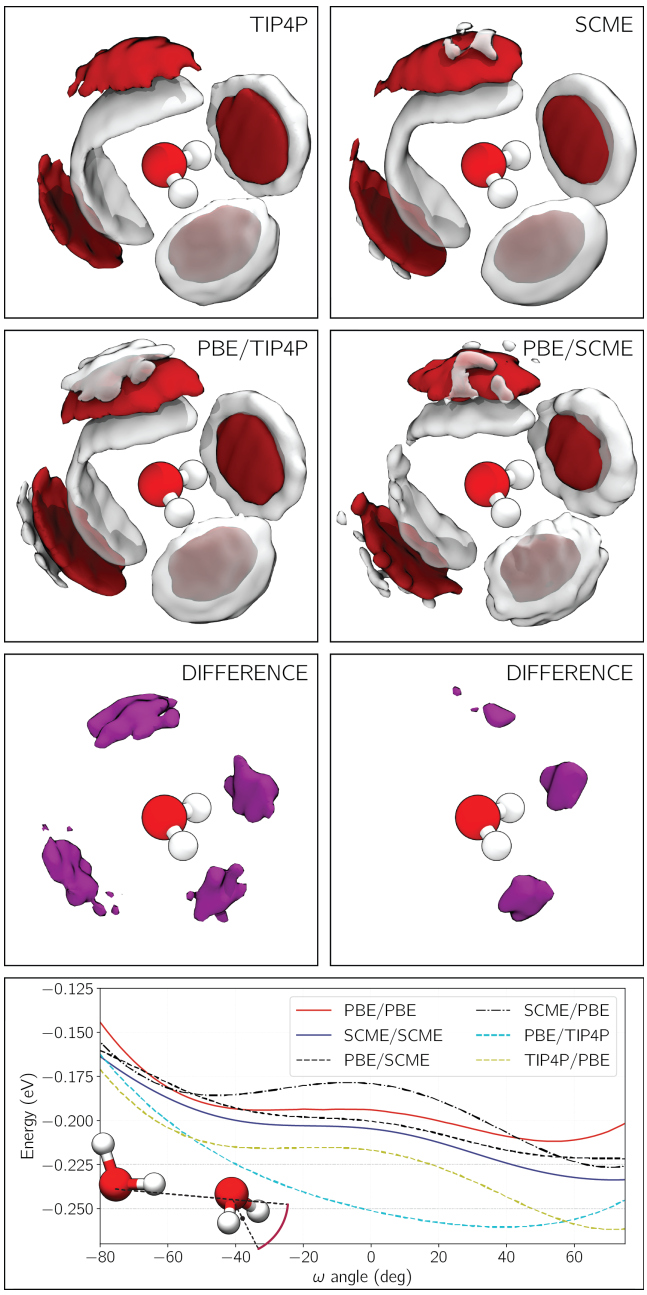

Figure 9: Top two rows: Spatial distribution functions of water molecules at isovalues of 4 and 2 times the average bulk density, for O (red) and H (white), respectively, sampled from aligned frames that are filtered using the previously employed hydrogen-bonding criterion,[64] to reduce sampling noise of the shorter QM/MM trajectories. Third row from top: Difference densities created by subtracting each pure MM O-density from its coupled counterpart, and renormalizing to show the isosurface corresponding to a 15% increase. Bottom: Dimer potential energy curve scanning the hydrogen-accepting lone-pairs, showing the qualitative breakdown in point-charge-coupled models that cause the differences in the EE-coupled liquid structure.

Acknowledgement

The authors thank K. T. Wikfeldt, E. R. Batista and J. Öström for helpful discussions. Funding was provided by the Icelandic Research Fund, Grants 174082-051, 174244-051, and 141080-051. The calculations were carried out at the Icelandic High-Performance Computing Center. Furthermore, E.O.J acknowledges additional computer access from Delingr. A.O.D. acknowledges support from Villum Fonden.

References

(1) Knorr, J.; Sokkar, P.; Schott, S.; Costa, P.; Thiel, W.; Sander, W.; Sanchez-Garcia, E.; Nuernberger, P. Competitive solvent-molecule interactions govern primary processes of diphenylcarbene in solvent mixtures. *Nat. Comm.* **2016**, *7*, 12968.

(2) Nogueira, J. J.; González, L. Computational Photophysics in the Presence of an Environment. *Annu. Rev. Phys. Chem.* **2018**, *69*, 473–497.

(3) Morzan, U. N.; de Armiño, D. J. A.; Foglia, N. O.; Ramírez, F.; Lebrero, M. C. G.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QMMM Simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.

(4) Pham, V.-T.; Penfold, T. J.; van der Veen, R. M.; Lima, F.; Nahhas, A. E.; Johnson, S. L.; Beaud, P.; Abela, R.; Bressler, C.; Tavernelli, I.; Milne, C. J.; Chergui, M. Probing the Transition from Hydrophilic to Hydrophobic Solvation with Atomic Scale Resolution. *J. Am. Chem. Soc.* **2011**, *133*, 12740–12748.

(5) Dohn, A. O.; Jónsson, E. O.; Kjær, K. S.; B. van Driel, T.; Nielsen, M. M.; Jacobsen, K. W.; Henriksen, N. E.; Møller, K. B. Direct Dynamics Studies of a Binuclear Metal Complex in Solution: The Interplay Between Vibrational Relaxation, Coherence, and Solvent Effects. *J. Phys. Chem. Lett.* **2014**, *5*, 2414–2418.

(6) Dohn, A. O.; Kjær, K. S.; Harlang, T. B.; Canton, S. E.; Nielsen, M. M.; Møller, K. B. Electron Transfer and Solvent-Mediated Electronic Localization in Molecular Photocatalysis. *Inorg. Chem.* **2016**, *55*, 10637–10644.

(7) van Driel, T. B.; Kjær, K. S.; Hartsock, R. W.; Dohn, A. O.; Harlang, T.; Chollet, M.; Christensen, M.; Gawelda, W.; Henriksen, N. E.; Kim, J. G.; Haldrup, K.; Kim, K. H.; Ihee, H.; Kim, J.; Lemke, H.; Sun, Z.; Sundström, V.; Zhang, W.; Zhu, D.; Møller, K. B.; Nielsen, M. M.; Gaffney, K. J. Atomistic characterization of the active-site solvation dynamics of a model photocatalyst. *Nat. Commun.* **2016**, *7*, 13678.

(8) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem.* **2009**, *48*, 1198–1229.

(9) van der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52*, 2708–28.

(10) Dohn, A. O.; Selli, D.; Fazio, G.; Ferraro, L.; Mortensen, J.; Civalleri, B.; Valentin, C. D. Interfacing CRYSTAL/AMBER to Optimize QM/MM Lennard–Jones Parameters for Water and to Study Solvation of TiO2 Nanoparticles. *Molecules* **2018**, *23*, 2958.

(11) Zhang, Y.-J.; Khorshidi, A.; Kastlunger, G.; Peterson, A. A. The potential for machine learning in hybrid QM/MM calculations. *J. Chem. Phys.* **2018**, *148*, 241740.

(12) Nielsen, M.; Björketun, M. E.; Hansen, M. H.; Rossmeisl, J. Towards first principles modeling of electrochemical electrode–electrolyte interfaces. *Surf. Sci* **2015**, *631*, 2–7.

(13) Skúlason, E. Modeling electrochemical reactions at the solid-liquid interface using density functional calculations. *Procedia Comput Sci* **2015**, *51*, 1887–1896.

(14) Skúlason, E.; Jónsson, H. Atomic scale simulations of heterogeneous electrocatalysis: recent advances. *Adv. Phys. X* **2017**, *2*, 481–495.

(15) Islam, M. S.; Fisher, C. A. Lithium and sodium battery cathode materials: computational insights into voltage, diffusion and nanostructural properties. *Chem. Soc. Rev.* **2014**, *43*, 185–204.

(16) García, G.; Koper, M. T. Stripping voltammetry of carbon monoxide oxidation on stepped platinum single-crystal electrodes in alkaline solution. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3802–3811.

(17) Skúlason, E.; Bligaard, T.; Gudmundsdóttir, S.; Studt, F.; Rossmeisl, J.; Abild-Pedersen, F.; Vegge, T.; Jónsson, H.; Nørskov, J. K. A theoretical evaluation of possible transition metal electro-catalysts for N 2 reduction. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1235–1245.

(18) Del Colle, V.; Berna, A.; Tremiliosi-Filho, G.; Herrero, E.; Feliu, J. Ethanol electrooxidation onto stepped surfaces modified by Ru deposition: electrochemical and spectroscopic studies. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3766–3773.

(19) Spohr, E. Some recent trends in computer simulations of aqueous double layers. *Electrochim. Acta* **2003**, *49*, 23–27.

(20) Hartnig, C.; Koper, M. T. Molecular dynamics simulation of solvent reorganization in ion transfer reactions near a smooth and corrugated surface. *J. Phys. Chem. B* **2004**, *108*, 3824–3827.

(21) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2006**, *117*, 185.

(22) Pezeshki, S.; Lin, H. Recent Advances in the Molecular Simulation of Chemical Reactions. *Mol. Sim.* **2015**, *41*, 168–189.

(23) Sneskov, K.; Schwabe, T.; Christiansen, O.; Kongsted, J. Scrutinizing the effects of

polarization in QM/MM excited state calculations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18551–18560.

(24) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(25) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed. (English)* **2009**, *48*, 1198–229.

(26) Zheng, M.; Waller, M. P. Adaptive quantum mechanics/molecular mechanics methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2016**, *6*, 369–385.

(27) Loco, D.; Polack, É.; Caprasecca, S.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Mennucci, B. A QM/MM Approach Using the AMOEBA Polarizable Embedding: From Ground State Energies to Electronic Excitations. *J. Chem. Theory Comput.* **2016**, *12*, 3654–3661.

(28) Mennucci, B.; Corni, S. Multiscale modelling of photoinduced processes in composite systems. *Nat. Rev. Chem.* **2019**, *3*, 315–330.

(29) Giovannini, T.; Lafiosca, P.; Cappelli, C. A General Route to Include Pauli Repulsion and Quantum Dispersion Effects in QM/MM Approaches. *J. Chem. Theory Comput.* **2017**, *13*, 4854–4870.

(30) Jensen, L.; van Duijnen, P. T.; Snijders, J. G. A discrete solvent reaction field model within density functional theory. *J. Chem. Phys.* **2003**, *118*, 514–521.

(31) Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M. S. Solvent effects on optical properties of molecules: A combined time-dependent density functional theory/effective fragment potential approach. *J. Chem. Phys.* **2008**, *129*, 144112.

(32) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

(33) Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Hybrid QM/MM molecular dynamics with AMOEBA polarizable embedding. *J. Chem. Theory Comput.* **2017**, *13*, 4025–4033.

(34) Menger, M. F. S. J.; Caprasecca, S.; Mennucci, B. Excited-State Gradients in Polarizable QM/MM Models: An Induced Dipole Formulation. *J. Chem. Theory Comput.* **2017**, *13*, 3778–3786.

(35) Rowley, C. N.; Roux, B. The Solvation Structure of Na+ and K+ in Liquid Water Determined from High Level ab Initio Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 3526–3535.

(36) Lipparini, F.; Barone, V. Polarizable Force Fields and Polarizable Continuum Model: A Fluctuating Charges/PCM Approach. 1. Theory and Implementation. *J. Chem. Theory Comput.* **2011**, *7*, 3711–3724.

(37) Lipparini, F.; Cappelli, C.; Barone, V. Linear Response Theory and Electronic Transition Energies for a Fully Polarizable QM/Classical Hamiltonian. *J. Chem. Theory Comput.* **2012**, *8*, 4153–4165.

(38) Visscher, K.; Swope, W.; Geerke, D. A QM/MM Derived Polarizable Water Model for Molecular Simulation. *Molecules* **2018**, *23*, 3131.

(39) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* **1996**, *105*, 1968–1986.

(40) Söderhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U. Protein influence on electronic spectra modeled by multipoles and polarizabilities. *J. Chem. Theory Comput.* **2009**, *5*, 649–658.

(41) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(42) Sneskov, K.; Schwabe, T.; Christiansen, O.; Kongsted, J. Scrutinizing the effects of polarization in QM/MM excited state calculations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18551.

(43) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234–20250.

(44) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. The QM/MM approach for wavefunctions, energies and response functions within self-consistent field and coupled cluster theories. *Mol. Phys.* **2002**, *100*, 1813–1828.

(45) Hrak, D.; Olsen, J. M. H.; Kongsted, J. Polarizable Density Embedding Coupled Cluster Method. *J. Chem. Theory Comput.* **2018**, *14*, 1351–1360.

(46) Liu, H.; Jenkins, A. J.; Wildman, A.; Frisch, M. J.; Lipparini, F.; Mennucci, B.; Li, X. Time-Dependent Complete Active Space Embedded in a Polarizable Force Field. *J. Chem. Theory Comput.* **2019**, *15*, 1633–1641.

(47) Jónsson, E. O.; Dohn, A. O.; Jónsson, H. Reciprocal Polarizable Embedding with a Transferable H_2O Potential Function I: Formulation & Tests on Dimer. *Submitted to JCTC simultaneously with the current manuscript* **2019**,

(48) Wikfeldt, K. T.; Batista, E. R.; Vila, F. D.; Jónsson, H. A Transferable H_2O Interaction Potential Based on a Single Center Multipole Expansion: SCME. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16542.

(49) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234–20250.

(50) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(51) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098.

(52) 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.

(53) Mortensen, J.; Hansen, L.; Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **2005**, *71*, 035109.

(54) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dulak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; Kristoffersen, H. H.; Kuisma, M.; Larsen, A. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. A.; Stausholm-Møller, J.; Strange, M.; Tritsaris, G. A.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys. Condens. Matter* **2010**, *22*, 253202.

(55) Bates, D. M.; Tschumper, G. S. CCSD(T) Complete Basis Set Limit Relative Energies for Low-Lying Water Hexamer Structures. *J. Phys. Chem. A* **2009**, *113*, 3555–3559.

(56) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.

(57) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **1976**, *32*, 922–923.

(58) Kromann, J. C. Calculate Root-mean-square deviation (RMSD) of Two Molecules Using Rotation, GitHub, version 1.2.7. 2018; `http://github.com/charnley/rmsd`.

(59) Dohn, A. O.; Jónsson, E. Ö.; Levi, G.; Mortensen, J. J.; Lopez-Acevedo, O.; Thygesen, K. S.; Jacobsen, K. W.; Ulstrup, J.; Henriksen, N. E.; Møller, K. B.; Jónsson, H. Grid-Based Projector Augmented Wave (GPAW) Implementation of Quantum Mechanics/Molecular Mechanics (QM/MM) Electrostatic Embedding and Application to a Solvated Diplatinum Complex. *J. Chem. Theory Comput.* **2017**, *13*, 6010–6022.

(60) Calaminici, P.; Jug, K.; Kster, A. M. Density functional calculations of molecular polarizabilities and hyperpolarizabilities. *J. Chem. Phys.* **1998**, *109*, 7756–7763.

(61) Gillan, M. J.; Alfè, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, 130901.

(62) Lehtola, S.; Jónsson, H. Variational, Self-Consistent Implementation of the Perdew–Zunger Self-Interaction Correction with Complex Optimal Orbitals. *J. Chem. Theory Comput.* **2014**, *10*, 5324–5337.

(63) Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E. Nuclear Quantum Effects in Water and Aqueous Systems: Experiment, Theory, and Current Challenges. *Chem. Rev.* **2016**, *116*, 7529–7550.

(64) Wernet, P.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Näslund, L. A.; Hirsh, T. K.; Ojamäe, L.; Glatzel, P.; Petterson, L. G. M.; Nilsson, A. The Structure of the First Coordination Shell in Liquid Water. *Science* **2011**, *304*, 995–999.

(65) Lee, H.-S.; Tuckerman, M. E. Structure of liquid water at ambient temperature from ab initio molecular dynamics performed in the complete basis set limit. *J. Chem. Phys* **2006**, *125*, 154507.

(66) DiStasio, R. a.; Santra, B.; Li, Z.; Wu, X.; Car, R. The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *J. Chem. Phys.* **2014**, *141*, 084502.

(67) Kusalik, P. G.; Svishchev, I. M. The Spatial Structure in Liquid Water. *Science* **1994**, *265*, 1219–1221.

(68) Mantz, Y. A.; Chen, B.; Martyna, G. J. Structural correlations and motifs in liquid water at selected temperatures: Ab initio and empirical model predictions. *J. Phys. Chem. B* **2006**, *110*, 3540–3554.

(69) Bulo, R. E.; Michel, C.; Fleurat-Lessard, P.; Sautet, P. Multiscale Modeling of Chemistry in Water: Are We There Yet? *J. Chem. Theory Comput.* **2013**, *9*, 5567–5577.

Graphical TOC Entry