

# A Many-Body, Fully Polarizable Approach to QM/MM Simulations

Eleftherios Lambros,<sup>\*,†</sup> Filippo Lipparini,<sup>\*,‡</sup> Gerardo Andrés Cisneros,<sup>\*,¶</sup> and  
Francesco Paesani<sup>\*,†,§,||</sup>

<sup>†</sup>*Department of Chemistry and Biochemistry, University of California San Diego,  
La Jolla, California 92093, United States*

<sup>‡</sup>*Dipartimento di Chimica e Chimica Industriale, University of Pisa,  
via G. Moruzzi 13, 56124, Pisa, Italy,*

<sup>¶</sup>*Department of Chemistry, University of North Texas,  
Denton, Texas 76203, United States*

<sup>§</sup>*Materials Science and Engineering, University of California San Diego,  
La Jolla, California 92093, United States*

<sup>||</sup>*San Diego Supercomputer Center, University of California San Diego,  
La Jolla, California 92093, United States*

E-mail: elambros@ucsd.edu; filippo.lipparini@unipi.it; gerardo.cisneros@unt.edu;  
fpaesani@ucsd.edu

## Abstract

We present a new development in quantum mechanics/molecular mechanics (QM/MM) methods by replacing conventional MM models with data-driven many-body (MB) representations rigorously derived from high-level QM calculations. The new QM/MM approach builds on top of mutually polarizable QM/MM schemes developed for polarizable force fields with inducible

dipoles and uses permutationally invariant polynomials to effectively account for quantum-mechanical contributions (e.g., exchange-repulsion, and charge transfer and penetration) that are difficult to describe by classical expressions adopted by conventional MM models. Using the many-body MB-pol and MB-DFT potential energy functions for water, which include explicit 2-body and 3-body terms fitted to reproduce the corresponding CCSD(T) and PBE0 2-body and 3-body energies for water, we demonstrate a smooth energetic transition as molecules are transferred between QM and MM regions, without the need of a transition layer. By effectively elevating the accuracy of both the MM region and the QM/MM interface to that of the QM region, the new QM/MB-MM approach achieves an accuracy comparable to that obtained with a fully QM treatment of the entire system.

## 1 Introduction

Quantum mechanics/molecular mechanics (QM/MM) approaches that integrate quantum chemistry (QM) with molecular mechanics (MM), have been used to describe systems that cannot be completely modeled by a pure QM or MM approach. These methods, which treat a smaller subsystem (S) via a quantum mechanical framework and the surrounding environment (E) using a force field, were initially described by Warshel and co-workers in 1972, wherein they partitioned the  $\sigma$  and  $\pi$  electrons of a molecule into separate empirical and semi-empirical treatments.<sup>1</sup> The current QM/MM formalism was originally proposed in the seminal 1976 study by Warshel and Levitt, where a protein was split into separate QM and MM subsystems, which interacted via a dielectric coupling.<sup>2</sup>

In their infancy these QM/MM approaches were used to study small chemical reactions, such as halide exchange reactions of the form  $CH_3Cl + Cl^-$  which needed a solvent to be accurately modeled.<sup>3</sup> Following the development of efficient interfaces between popular QM and MM software,<sup>3</sup> along with advances in computational power, it became possible to simulate larger and more complex systems, such as enzymatic and other biomolecular reactions.<sup>4,5</sup> Furthermore, QM/MM saw success in the simulation of spectroscopic experiments of complex molecular systems.<sup>6</sup>

With regards to the formalism, most QM/MM approaches rely on an “effective Hamiltonian” which couples the properties of the electronic density of the subsystem S to a set of point charges representing the environment E. This coupling, known as the electrostatic embedding (EE), is commonly used with non-polarizable force-fields, as all electrostatic interactions between the two regions can be taken into account in a relatively straightforward and cost-effective way. However, in EE only the QM density is polarized by the environment and not vice versa. In particular, this implies that the interaction between the QM subsystem and the environment is the same, independently of the electronic state of the QM subsystem and/or the effects of an external field. The solution to this conundrum, known as a “polarizable embedding” (PE), allows for the self-consistent equilibration of the classical degrees of freedom used to represent the electrostatic distribution of the environment, described with a polarizable force field,<sup>7–10</sup> with the electronic density in the QM region. Early approaches to the PE method involved simple iterations over an electrostatic embedding scheme where the inducible dipoles could be back-polarized by point charges fit to the electronic density until the system converged.<sup>11–14</sup> A related approximation to the polarizable embedding was introduced with the LICHEM interface, which couples generic QM and MM software using a partially self-consistent polarization scheme, compromising some accuracy for a more generalized interface and greater computational efficiency.<sup>15,16</sup>

By solving the QM equations and the MM polarization equations in a fully coupled fashion, some recent approaches adopt a more intimate coupling between the electron density in the QM region and the polarization in the MM region, with the latter being modeled by inducible point dipoles (IPD),<sup>17–25</sup> fluctuating charges,<sup>26–31</sup> Drude oscillators,<sup>32–37</sup> or using a response kernel.<sup>38,39</sup> In this contribution, we limit our discussion to IPD-based approaches, as this is the polarization model adopted by both the MB-pol<sup>40–42</sup> and AMOEBA<sup>43,44</sup> models used in this analysis. We note in passing that an IPD polarization scheme is also present in other advanced QM/MM methods, such as the effective fragment potential method<sup>45</sup> and the X-pol method.<sup>46</sup> Several implementations exist of fully polarizable IPD-based QM/MM schemes.<sup>17,21,23,47–49</sup> Over the years, these implementations have been extended to simulations of various molecular properties, with the

QM subsystem in either the ground or an excited state, hydration properties and, more recently, *ab initio* multiscale molecular dynamics.<sup>22,23,30,31,35–37,50,51</sup>

Beyond the issues associated with coupling the QM and the MM regions, QM/MM faces another methodological roadblock: diffusive breakup of the QM region. During a molecular dynamics simulation in solution, molecules that were originally assigned to the QM region can diffuse apart, which lead to an effective exchange between QM and MM molecules. Such situations can result in low-accuracy MM molecules, which, by construction, cannot undergo any chemical transformation, moving into the original QM region of interest, while high-accuracy QM molecules are dispersed in the original MM region. So-called adaptive QM/MM schemes attempt to overcome this limitation by adaptively partitioning the molecules in the QM and MM regions at each step.<sup>52–55</sup> Although appealing, adaptive QM/MM schemes suffer from some drawbacks associated with the inherent differences between QM and MM descriptions of the energies and corresponding forces. Due to this, when a molecule is switched between QM and MM treatments during re-partitioning, the whole system can experience a large change in energy causing a discontinuity in the force. This can, for example, happen if the minima of the QM and MM energy landscapes are shifted relative to each other, which implies that a molecule located in a minimum energy configuration on the MM energy landscape can be thrust into a high-energy configuration when transitioned to the QM energy landscape.<sup>52,53,56</sup> In simulations, this behavior is manifested as unphysical dynamics along with structural artifacts at the QM/MM boundary.<sup>4,52,53,56,57</sup> In this thread, advances in adaptive QM/MM have focused on suppressing and reducing energetic discontinuities at the QM/MM boundary, often by using a transition layer to smoothly transition atoms between QM and MM treatments.<sup>55</sup> These techniques range from simple smoothing functions to more complex distance-based partitioning approaches which have been shown to conserve energy and smooth forces.<sup>4,53–56,58–71</sup>

Unfortunately, while these adaptive approaches improve energy conservation and alleviate some of the discontinuities in the forces, which, consequently, smooth out boundary artifacts that may occur, they do not solve the underlying issue in that the system is being represented

by two different potential energy functions.<sup>55</sup> Furthermore, adoption of transition layers increases the computational cost of QM/MM simulations, as some schemes can result in up to  $2^N$  additional QM calculations for  $N$  molecules in the transition layer, though more modern partitioning schemes have been shown to significantly reduce that number.<sup>53,55,56,63,67</sup>

Other approaches using advanced representations of the energy landscape have attempted a more rigorous treatment of QM/MM systems. Effective and exchange fragment models (EFP and XFP respectively) partition the system into *ab initio* and effective fragment (EF) components, where the two regions interact via one electron integrals between *ab initio* and EF molecules.<sup>72–76</sup> Recent EFP approaches have further developed this by integrating exchange-repulsion style interaction potentials across the QM/MM boundary in order to account for some quantum effects missing from the core EFP interaction.<sup>77</sup> The Gaussian electrostatic model (GEM) uses a density fitting formalism to represent the MM region using molecular electron densities.<sup>78</sup> Incorporated into a QM/MM approach, GEM allows both the QM and the MM regions to be represented by an electron density, and thus affords the scheme much greater accuracy when evaluating QM/MM interactions.<sup>79,80</sup> Building on these previous QM/MM approaches, we propose a new development, QM/MB-MM, in which the MM region is represented by data-driven models rigorously derived from the many-body expansion of the interaction energy. In this study, we describe the QM/MB-MM scheme using water as a prototypical example where the MB-MM region is described by the MB-pol<sup>40–42</sup> and MB-DFT<sup>81</sup> models. MB-pol and MB-DFT are many-body water models that include explicit terms fitted to reproduce 2-body and 3-body energies calculated at the CCSD(T) and DFT levels of theory, respectively. We demonstrate that QM/MB-MM combined with a polarizable embedding scheme effectively elevate QM/MM calculations to quantum mechanical levels of accuracy, specifically, CCSD(T) for MB-pol and PBE0 for MB-DFT. Importantly, by combining a PBE0 representation of the QM region with a MB-DFT model fitted to PBE0 2-body and 3-body data to represent the MM region, we demonstrate that energy discontinuities at the QM/MM boundary are largely suppressed and reduced to negligible effects.

## 2 Theory and Computational Details

All QM/MM calculations were carried out with the PBE0 functional along with the D3 dispersion correction using the aug-cc-pVQZ basis set.<sup>82–85</sup> Gaussian16<sup>86</sup> was used for the QM calculations while an in-house code was used to handle the MB-pol and MB-DFT components incorporated into the QM/MM coupling framework.

### 2.1 Explicit many-body models: MB-pol and MB-DFT

MB-pol and MB-DFT are explicit many-body models based off of the many-body expansion (MBE) which represents the energy of a system as the sum of individual  $n$ -body contributions in a system of  $N$  molecules with  $n \leq N$ .<sup>40–42</sup> This is formally expressed as:<sup>87</sup>

$$E_N(r_1, \dots, r_N) = \sum_{i=1}^N \epsilon_{1B}(r_i) + \sum_{i<j}^N \epsilon_{2B}(r_i, r_j) + \sum_{i<j<k}^N \epsilon_{3B}(r_i, r_j, r_k) + \dots + \epsilon_{NB}(r_1, \dots, r_N) \quad (1)$$

where  $\epsilon_{1B}$  represents the energy required to distort an individual monomer from its equilibrium geometry, and  $\epsilon_{nB}$  with  $n > 1$  are the  $n$ -body energies that are recursively defined as:

$$\epsilon_{nB} = \epsilon_n(1, \dots, n) - \sum_{i=1}^N \epsilon_{1B}(r_i) - \sum_{i<j}^N \epsilon_{2B}(r_i, r_j) - \sum_{i<j<k<\dots<n-1}^N \epsilon_{(n-1)B}(r_i, r_j, \dots, r_{n-1}). \quad (2)$$

For molecules with localized electron density and large band gap, such as water, this expansion converges very quickly, with the first the sum of 2-body and 3-body energies contributing  $\sim 90-95\%$  to the total interaction energy.

The MB-pol<sup>40–42</sup> and MB-DFT<sup>81</sup> models include explicit representations for the first three terms of the MBE, while implicitly accounting for all higher order,  $n > 3$  body terms through a classical polarization term,

$$E_N(r_1, \dots, r_N) = \sum_{i=1}^N \epsilon_{1B}(r_i) + \sum_{i>j}^N \epsilon_{2B}(r_i, r_j) + \sum_{i>j>k}^N \epsilon_{3B}(r_i, r_j, r_k) + E_{pol} \quad (3)$$

In these models, the 1B term is described by the Partridge-Schwenke potential energy surface, which was originally calculated at the configuration interaction level of theory and subsequently refined to quantitatively reproduce the experimental rovibrational transitions.<sup>88</sup> The 2-body term consists of three contributions:

$$\epsilon_{2B} = E_{2B}^{sr} + E_{elec} + E_{disp} \quad (4)$$

where  $E_{2B}^{sr}$  is a 4<sup>th</sup>-degree permutationally invariant polynomial (PIP)<sup>89</sup> representing short-range interactions within a water dimer.  $E_{elec}$  is described by Coulomb interactions between geometry dependent point charges which reproduce the *ab initio* dipole moment surface of a water molecule.<sup>88</sup> Lastly, the 2-body dispersion interaction is expressed as

$$E_{disp} = - \sum_{i,j} f(\delta_{ij}) \frac{C_{6,ij}}{R_{ij}^6} \quad (5)$$

where  $i$  and  $j$  are indices for atoms on two separate water monomers,  $f(\delta_{ij})$  is the Tang-Toennies damping function with a fitted parameter  $\delta_{ij}$ ,<sup>90</sup> and  $C_{6,ij}$  are the dispersion coefficients calculated from the asymptotic reference energies of the water dimer as in CC-pol.<sup>91</sup> The explicit 3-body term,

$$\epsilon_{3B} = E_{3B}^{sr} \quad (6)$$

represents short-range 3-body interactions and is represented by a 4<sup>th</sup>-degree PIP.<sup>89</sup> In MB-pol, these explicit 2B and 3B terms were fitted on top of an the implicit many-body polarizable term ( $E_{pol}$  in Eq. 3) to reproduce 2B and 3B energies calculated at the CCSD(T) level of theory in the complete basis set (CBS) limit.<sup>40,41</sup> In the MB-DFT model, the explicit  $\epsilon_{2B}$  and  $\epsilon_{3B}$  terms were instead fitted to 2B and 3B energies calculated at the PBE0-D3/aug-cc-pVQZ level of theory using the same procedure described in Ref. 81.

## 2.2 The QM/MM interaction with MB-pol and MB-DFT

In this section, we briefly overview the theory of IPD-based polarizable QM/MM at the self-consistent field level of theory, i.e., Hartree-Fock (HF) or DFT. A more detailed overview can be found in a recent review.<sup>92</sup> The MB-pol and MB-DFT polarization energy is variational with respect to the polarization degrees of freedom, which means that the coupled QM/MM equations and energy can be conveniently obtained starting from a variational energy functional<sup>21,24</sup> of the SCF density  $\mathbf{P}$  and the IPDs  $\boldsymbol{\mu}$

$$\mathcal{E}(\mathbf{P}, \boldsymbol{\mu}) = \mathcal{E}^{QM}(\mathbf{P}) + \mathcal{E}^{MM}(\boldsymbol{\mu}) + \mathcal{E}^{Coupl}(\mathbf{P}, \boldsymbol{\mu}) \quad (7)$$

where  $\mathcal{E}^{QM}(\mathbf{P})$  and  $\mathcal{E}^{MM}(\boldsymbol{\mu})$  represent the purely quantum mechanical and classical contributions to the energy, and  $\mathcal{E}^{Coupl}(\mathbf{P}, \boldsymbol{\mu})$  describes the coupling between the two regions. The MM and coupling terms can then be combined into a general “environmental”  $\mathcal{E}^{Env}(\mathbf{P}, \boldsymbol{\mu})$  term:

$$\mathcal{E}(\mathbf{P}, \boldsymbol{\mu}) = \mathcal{E}^{QM}(\mathbf{P}) + \mathcal{E}^{Env}(\mathbf{P}, \boldsymbol{\mu}) \quad (8)$$

which is expressed as a sum of the MM contribution ( $\mathcal{E}^{FF}$ ), the polarization energy ( $\tilde{\mathcal{E}}^{Pol}$ ), and the electrostatic QM/MM coupling ( $\mathcal{E}^{QM/MM}$ ):

$$\mathcal{E}^{Env}(\mathbf{P}, \boldsymbol{\mu}) = \mathcal{E}^{FF} + \tilde{\mathcal{E}}^{Pol}(\mathbf{P}, \boldsymbol{\mu}) + \mathcal{E}^{QM/MM}(\mathbf{P}) \quad (9)$$

The first term in Eq. 9 represents the energy of the MM region as described by a force field, which depends on neither the electron density nor the induced dipoles on the MM atoms, as well as any non-electrostatic interactions between the QM and MM regions, such as van der Waals interactions between molecules in the QM and MM regions:

$$\mathcal{E}^{FF} = E_{non-elec}^{MM} + E_{non-elec}^{QM/MM}. \quad (10)$$



For the MB-pol and MB-DFT models which, besides the standard dispersion energy term, also incorporate 2B and 3B PIPs to account for short-range quantum-mechanical effects (e.g., charge transfer and penetration) missing in standard force fields, the first term of Eq. 10 becomes

$$E_{non-elec}^{MB-pol} = E_{disp} + E_{2b,sr} + E_{3b,sr} \quad (11)$$

Here,  $E_{non-elec}^{MB-pol}$  sums up all the dispersion, and short-range 2B and 3B terms for the MM region alone, as well as all possible cross terms for pairs and triples of molecules split between the QM and MM regions. The 2B and 3B PIPs included in the QM/MM interaction, therefore, act as correction terms, allowing, in principle, the QM/MM energy to achieve the same accuracy as in purely QM calculations. In contrast, when the MM region is described with a standard polarizable force field, such as AMOEBA, the non-electrostatic QM/MM interaction only includes a van der Waals term represented by a force-field specific functional force, such as a standard 12-6 potential, or, for AMOEBA, a buffered 14-7 potential:<sup>15,21,24,43,44</sup>

$$E_{non-elec}^{AMOEBA} = E_{vdW} \quad (12)$$

We note here that in the QM/MB-pol scheme, the non-electrostatic interactions are modeled in a purely classical fashion, i.e., the associated energy contribution does not depend on the QM density, but only on geometrical parameters. This allows one to achieve high accuracy on the energy and structures without complicating the QM equations, but cannot reproduce quantum confinement effects on the QM density, which might be relevant to compute properties and model spectroscopies.<sup>45,93,94</sup> A density-dependent correction for MB-pol and MB-DFT is currently under active investigation.

The second term of Eq. 9 is the most relevant one for the coupling of the QM and classical subsystems, as it describes the QM/MM polarization:

$$\tilde{\mathcal{E}}^{Pol}(\mathbf{P}, \boldsymbol{\mu}) = \frac{1}{2} \boldsymbol{\mu}^\dagger \mathbf{T} \boldsymbol{\mu} - \boldsymbol{\mu}^\dagger (\mathbf{E}^{MM} + \mathbf{E}^{QM}(\mathbf{P})), \quad (13)$$

where  $\mu$  is a vector that collects all the IPD,  $\mathbf{T}$  describes the Thole-damped dipole interaction tensor, and

$$E_i^{QM}(\mathbf{P}) = \sum_{\alpha} Z_{\alpha} \frac{R_{\alpha} - r_i}{|R_{\alpha} - r_i|^3} - \sum_{\mu\nu} P_{\mu\nu} \int \chi_{\mu}(r) \chi_{\nu}(r) \frac{r - r_i}{|r - r_i|^3} d^3r \quad (14)$$

is the electric field due to the QM density at the  $i$ -th polarizable site.<sup>40,49</sup> We remark that the QM field is the sum of the contributions of the nuclei, with charges  $Z_{\alpha}$  at positions  $R_{\alpha}$ , and of the electronic density, that has been expanded in a basis of atomic orbitals  $\chi$ . Finally, the third term in Eq. 9 represents the interaction of the electron density with the static charges:

$$\mathcal{E}^{QM/MM}(\mathbf{P}) = \mathbf{q}^{\dagger} \mathbf{V}^{QM}(\mathbf{P}), \quad (15)$$

where  $\mathbf{V}^{QM}$  is the QM electrostatic potential and  $\mathbf{q}$  is a vector that collects all the point charges, that represent the permanent electrostatic distribution in MB-pol and MB-DFT. The coupled QM/MM equations are easily obtained by minimizing the variational energy functional in eq. 7 with respect to the QM density and the IPDs, taking into account the constraints on the density. In particular, by differentiating with respect to the elements of the density matrix, we get the QM/MM Fock (Kohn-Sham, KS) matrix:

$$F_{\mu\nu} = F_{\mu\nu}^{\text{vac}} + \mathbf{q}^{\dagger} \mathbf{V}_{\mu\nu} - \mu^{\dagger} \mathbf{E}_{\mu\nu}, \quad (16)$$

where  $F^{\text{vac}}$  is the standard Fock (KS) matrix and  $\mathbf{V}_{\mu\nu}, \mathbf{E}_{\mu\nu}$  are standard potential and electric field one-electron integrals.<sup>21</sup> The polarization equations are obtained by differentiating with respect to the IPD:

$$\mathbf{T}\mu = \mathbf{E}^{\text{MM}} + \mathbf{E}^{\text{QM}}(\mathbf{P}). \quad (17)$$

We note that the QM and polarization equations are coupled, as the Fock (KS) matrix depends on the IPDs, while the right-hand side of the polarization equations depends on the QM density. Within a SCF approach, the coupled equations are solved by computing the IPDs at each SCF cycle by solving eq. 17 either with a direct method or iteratively, and then using them to assemble the environment contribution to the Fock (KS) matrix, which is in turn diagonalized to get a new

density, until convergence is achieved.

For the QM/AMOEBA polarizable QM/MM scheme, a few modifications need to be introduced, as the model is non variational and includes higher order multipolar moments into the permanent electrostatic distribution. A detailed derivation of the electrostatic and polarization interactions can be found in the relevant literature.<sup>21,24</sup>

In our analysis, we also consider a simplified polarizable embedding scheme implemented in the LICHEM software.<sup>15,16</sup> In this scheme, the electrostatic interaction between QM and MM regions is separated into two components. The first component accounts for the polarization of the QM region induced by the point charges in the MM region, while the second component accounts for the back-polarization of the the MM region induced by the polarized QM charges. For multipolar models that contain permanent dipoles and quadrupoles, such as AMOEBA, a set of six point charges are octahedrally distributed about each atom to guarantee that the permanent multipole moments are correctly described.<sup>15</sup>

In this work, the fully self-consistent polarizable QM/MM scheme is referred to as fsc-QM/MB-MM, and the partially self-consistent polarizable QM/MM scheme implemented in LICHEM is referred to as psc-QM/MB-MM.

## 3 Results and discussion

### 3.1 Water dimer

As an initial demonstration, Fig. 1 shows the QM/MB-MM interaction energies of the water dimer were calculated along MP2/aug-cc-pVQZ scans obtained by optimizing the positions of the hydrogen atoms for each fixed oxygen-oxygen distance. In this analysis, both possible QM/MB-MM partitions were examined, with the QM region containing either the hydrogen-bond donor or the hydrogen-bond acceptor. The QM/MB-MM energies, calculated using both the fully self-consistent polarizable (fsc-QM/MB-MM)<sup>21,24,25</sup> and the partially self-consistent polarizable (psc-QM/MB-MM)<sup>15,16</sup> schemes, are compared against the corresponding full MB-MM and QM values

calculated with MB-pol and PBE0-D3, respectively. The kink in the plot occurring at  $\sim 2.7$  Å corresponds to configurations that shift from a “sheared” geometry, where the two water molecules are symmetrically flipped, to a “hydrogen-bonded” geometry, as displayed in the upper left panel of Fig. 1.

When the QM water molecule donates the hydrogen bond, the fsc-QM/MB-pol energies lie close to the MB-pol reference values in the neighborhood of minimum-energy geometry (between 2.7 and 3.2 Å), with the largest deviation of  $\sim 0.1$  kcal/mol occurring near the kink at 2.7 Å. The psc-QM/MB-pol scheme yields comparable results, although it tends to slightly overestimate the interaction strength, giving rise to a larger maximum error of  $\sim 0.25$  kcal/mol in correspondence of the kink. When the partition is reversed, such that the QM water molecule becomes the hydrogen-bond acceptor, a much larger overall error is observed in both QM/MB-MM schemes. In this case, the fsc-QM/MB-pol scheme overestimates the interaction energy by  $\sim 0.3$  kcal/mol at the

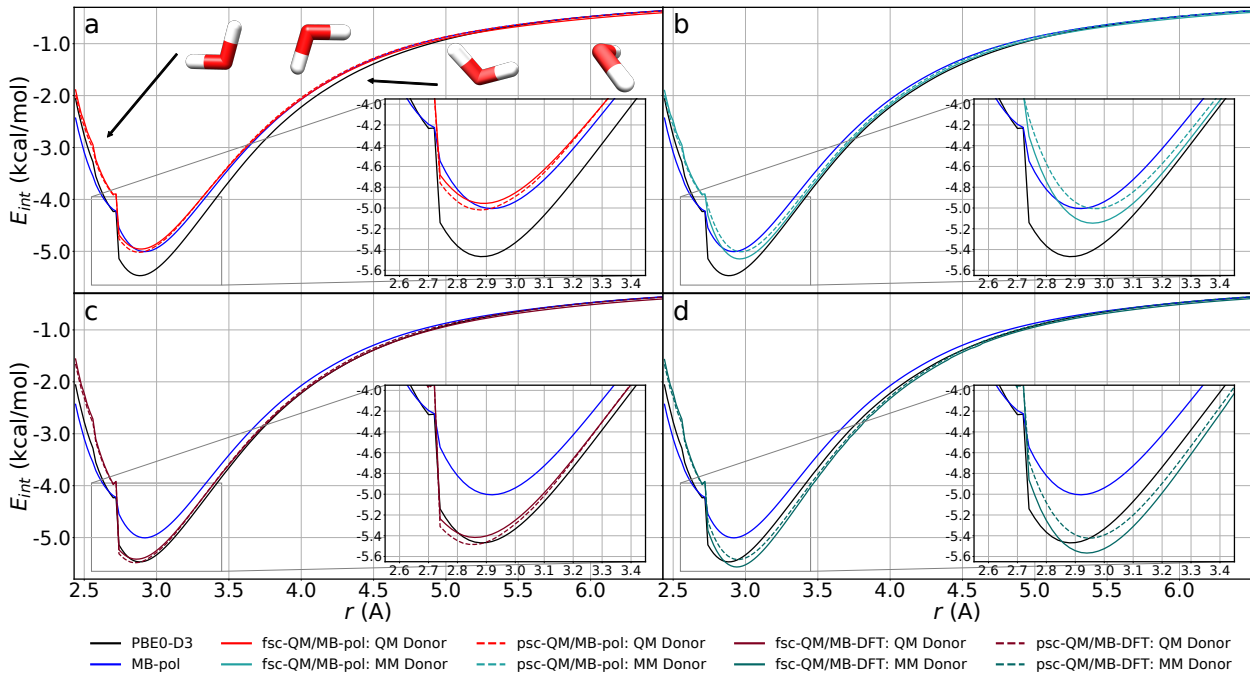


Figure 1: Comparison of the QM/MM interaction energy against the MB-pol and PBE0-d3 values over the MP2/aug-cc-pvqz optimized dimer scan. a) QM/MB-pol results with the QM water molecule as the hydrogen-bond donor. b) QM/MB-pol results with the MM water molecule as the hydrogen-bond donor. c) QM/MB-DFT results with the QM water molecule as the hydrogen-bond donor. d) QM/MB-DFT results with the MM water molecule as the hydrogen-bond donor.

kink, while underestimating the interaction energy at the minimum by about 0.2 kcal/mol. The psc-QM/MB-pol scheme predicts a larger error of  $\sim 0.5$  kcal/mol at the kink but reproduces the minimum energy within 0.05 kcal/mol of the reference MB-pol value.

The bottom panels of Fig. 1 compare the reference MB-pol and PBE0-D3 interaction energies to the corresponding QM/MB-DFT values where, as described in Section 2, the MB-DFT model was fitted to reproduce 2B and 3B PBE0-D3/aug-cc-pVQZ energies. Similar to the QM/MB-pol case, when the QM water molecule is the hydrogen-bond donor, the fsc-QM/MB-DFT scheme underestimates the PBE0-D3 interaction energies, with the largest error of about 0.1 kcal/mol found at the kink. The psc-QM/MB-DFT scheme predicts an error of  $\sim 0.2$  kcal/mol for the same configuration. When the QM/MM partition is switched and the QM water molecule becomes the hydrogen-bond acceptor, both the fully and partially self-consistent schemes overestimate the reference PBE0-D3 interaction energy at the kink, with the errors being  $\sim 0.3$  and  $\sim 0.5$  kcal/mol, respectively. The fsc-QM/MB-DFT scheme underestimates the interaction energy at the minimum by  $\sim 0.1$  kcal/mol, while the psc-QM/MB-DFT scheme overestimates it by  $\sim 0.1$  kcal/mol.

Because QM/MB-pol and QM/MB-DFT only differ in the identity of the 2B and 3B PIPs (along with the Tang-Toennies damped dispersion), the shift in the interaction energies from QM/MB-pol, which closely reproduces the MB-pol values, to QM/MB-DFT, which instead closely reproduces the PBE0-D3 values, is due to the MB-DFT PIPs correctly compensating the core QM/MM electrostatic interaction. If the QM/MM electrostatic interaction is sufficiently similar to the core MM electrostatic interaction, these results suggest that the 2B and 3B corrections used in MB-pol and MB-DFT can be successfully integrated into a QM/MM scheme to correctly account for the deficiencies in the core QM/MM interaction. In both QM/MB-pol and QM/MB-DFT, smaller overall deviations from the corresponding MB-pol and PBE0-D3 reference values are observed when the QM water molecule donates the hydrogen bond compared to those found when the QM water is the hydrogen-bond acceptor. Furthermore, between these two partitions, the QM hydrogen-bond donor case predicts a minimum that is slightly shifted to a shorter distance, indicating a slightly more attractive interaction between water molecules. On the other hand, in the QM hydrogen-

bond acceptor case, the minimum is shifted significantly to a larger distance (by  $\sim 0.15$  Å), indicating that both QM/MB-pol and QM/MB-DFT overall predict a more repulsive interaction relative to the MB-pol and PBE0-D3 references, respectively. This “repulsive shift” observed in the QM hydrogen-bond acceptor case carries out to the region beyond the minimum, where both QM/MB-pol and QM/MB-DFT consistently underestimate the interaction energies as the two water molecules are taken apart. This indicates that, depending on which molecule (QM or MM) is the hydrogen-bond donor, the difference between the QM/MB-MM and MB-MM descriptions of the underlying electrostatic interactions may become significant and possibly incompatible with the corrections provided by the PIPs at short range.

In the short-range region, where the two water molecules are closer than 2.7 Å, they adopt a symmetrically flipped “sheared” geometry. Due to the symmetry of these dimer configurations, the two possible QM/MM partitions are virtually identical and, consequently, predict the same QM/MM energies. The close vicinity of the water molecules in these configurations makes the distinction of the QM and MM regions somewhat blurred, which implies that the QM/MM electrostatic interaction become more quantum mechanical in nature, with the point charges and dipoles of the MM molecule penetrating the electron density of the QM molecule. The consistently higher error of  $\sim 0.5$  kcal/mol observed for these configurations can thus be attributed to the expected incompatibility of the short-range PIPs, which were optimized for a purely MB-MM description of the interaction energies, with the QM/MM electrostatics.

### 3.2 Water 2-body energies

With the previous section evaluating the performance of QM/MB-pol and QM/MB-DFT for an optimized water dimer scan where the geometries of the two water molecules lied close to the corresponding minimum-energy conformation, this section explores the performance of QM/MB-pol and QM/MB-DFT for a set of 1000 dimers with more distorted molecular geometries randomly extracted from the original MB-pol 2B training set.<sup>40</sup> Fig. 2 shows correlations between 2-body energies calculated with the QM/MB-pol, QM/MB-DFT, and QM/AMOEBA14 schemes, and the

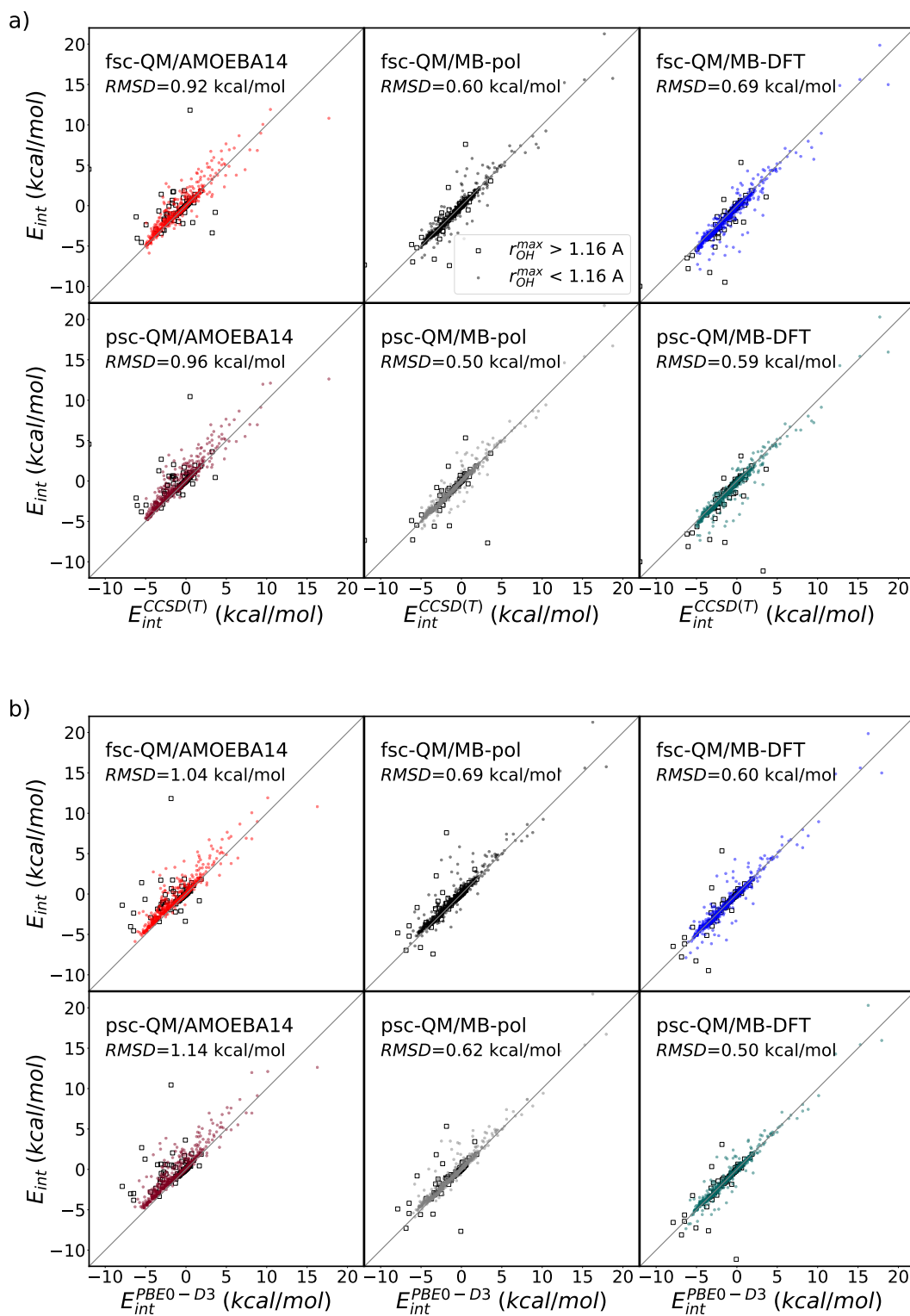


Figure 2: Correlations between the QM/MM interaction energies of 1000 water dimers calculated with AMOEBA14, MB-pol, and MB-PBE0 representing the MM region, and PBE0-D3 representing the QM region, and the corresponding CCSD(T) (panels a) and PBE0-D3 (panels b) values. Each RMSD is calculated for configurations whose maximum intramolecular OH bond length is shorter than 1.6 Å.

corresponding CCSD(T) and PBE0-D3 reference values for this set of distorted dimers. For this analysis, the CCSD(T) 2B energies in the complete basis set limit were taken from the original MB-pol reference<sup>40</sup> while the corresponding PBE0-D3 values were calculated with Gaussian16 using the aug-cc-pVQZ basis set. In order to exclude from this analysis overly distorted dimers for which, as mentioned above, the separation between QM and MM regions becomes blurred, the RMSDs reported in Fig. 2 are calculated considering only configurations whose maximum intramolecular OH bond length is shorter than 1.16 Å, corresponding to the 90<sup>th</sup> percentile of the entire set.

The RMSDs relative to the CCSD(T) 2B energies are 0.92, 0.60, and 0.69 kcal/mol for the fsc-QM/AMOEBA14, QM/MB-pol, and QM/MB-DFT schemes, respectively. A similar trend is found for the RMSDs calculated relative to the PBE0-D3 2B energies, with values of 1.04, 0.69, and 0.60 kcal/mol for QM/AMOEBA14, QM/MB-pol, and QM/MB-DFT, respectively. Slightly smaller values are obtained using the partially self-consistent scheme, with QM/AMOEBA14, QM/MB-pol, and QM/MB-DFT being associated with RMSDs of 0.96, 0.50, and 0.59 kcal/mol, respectively, relative to the CCSD(T) 2B energies, and 1.14, 0.62, and 0.50 kcal/mol, respectively, relative to the PBE0-D3 2B energies.

The overall decrease in the RMSD values observed for the QM/MB-pol and QM/MB-DFT schemes relative to QM/AMOEBA14 is attributed to the ability of the PIPs adopted by the explicit many-body models to account for short-range quantum-mechanical effects (e.g., charge transfer and penetration) which cannot be quantitatively recovered by the purely classical expressions used in the QM/AMOEBA14 scheme. QM/MB-pol, which employs PIPs specifically fitted to account for the difference in 2B energies between the 2B CCSD(T) reference energies and the classical components of the 2B MB-pol energies, shows the smallest RMSD relative to CCSD(T). Correspondingly, QM/MB-DFT, which uses the same PIPs as MB-pol but fitted to PBE0-D3 2B reference energies, shows the smallest QM/MM RMSD relative to PBE0-D3. These results, which hold for both the fully and partially self-consistent schemes, suggest that, relative to their references, the PIPs fitted to CCSD(T) and PBE0-D3 are able to compensate, at least in part, for the differ-



ences between the actual quantum-mechanical 2B energies ( $\epsilon_{2B}$  in Eq. 4) and the corresponding classical 2B terms (i.e.,  $E_{elec}$  and  $E_{disp}$  in Eq. 4), even for distorted dimer configurations.

It is particularly interesting to note that the psc-QM/MB-pol and QM/MB-DFT exhibit lower RMSDs than their fully-consistent counterparts. This apparently counterintuitive result can be understood by considering that the partially self-consistent QM/MM scheme may benefit from a more “MM-like” representation of the QM/MM electrostatics, as part of the QM/MM polarization is evaluated using the MM functional form. Specifically, in the psc-QM/MB-MM, the MB-MM region gets back-polarized by the polarized QM point charges, instead of the electron density as in the fully self-consistent QM/MB-MM scheme, which makes part of the QM/MM polarization more similar to that of the MM model. As a result, this more MM-like treatment of the QM/MM electrostatics within the psc-QM/MM scheme is, by construction, more compatible with the PIPs of the MB-pol and MB-DFT models which were fitted to the differences between the 2B CCSD(T) and PBE0-D3 energies, respectively, and the classical electrostatic and dispersion terms.

### 3.3 QM/MM partitions for the water hexamer

With the the two previous sections focusing on dimer configurations, which report at most on 2-body contributions to the interaction energies, in this section we examine the ability of the QM/MB-MM scheme to reproduce the energetics of the water hexamer isomers. While the analysis described in the following will specifically focus on the prism isomer, which is the lowest-energy isomer of the water hexamer,<sup>95</sup> analogous analyses for other low-lying isomers are reported in the Supporting Information. Starting with the cluster containing one water molecule in the QM region, water molecules are successively included in the QM region until a total of five water molecules are placed in the QM region. As water molecules are added, one at a time, all possible QM/MM partitions are examined. This analysis thus examines the effects of adding (subtracting) water molecules to (from) the QM region, as it might happen during adaptive QM/MM simulations in solution. The interaction energies calculated for the different QM/MM partitions as a function of number of water molecules in the QM region are shown in Fig. 3 where they are compared with the

corresponding MB-pol and PBE0-D3 values. Within each column of Fig. 3, the QM/MM partitions keep the number of QM and MM molecules constant but permute the QM and MM indices, which allow us to assess the performance of the QM/MB-pol and QM/MB-DFT schemes depending on all possible spatial arrangements of the QM and MM water molecules within the prism cluster.

With one water in the QM region, the interaction energies calculated for the different QM/MM partitions using both the fully and partially self-consistent QM/MB-pol schemes lie, on aver-

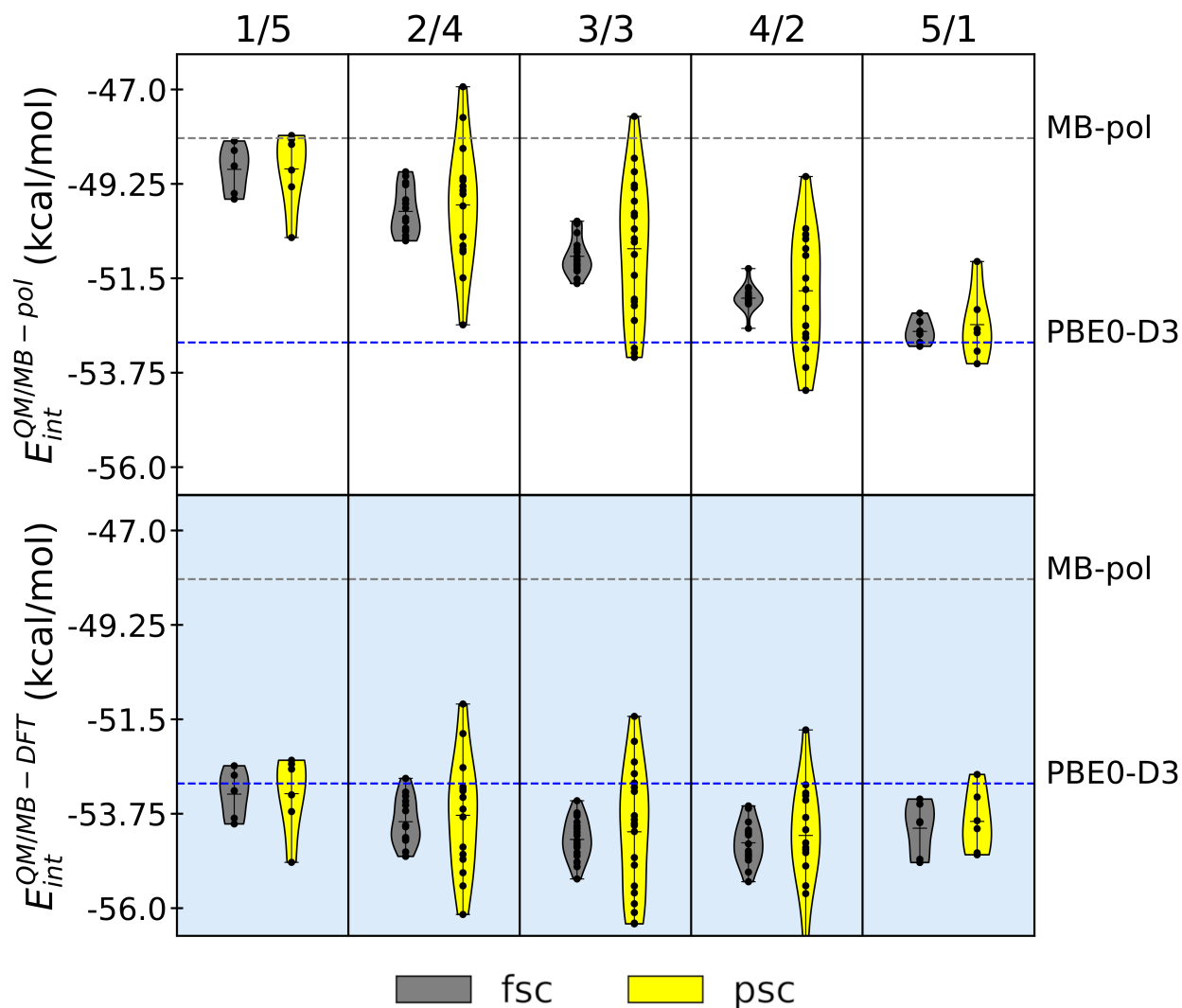


Figure 3: Interaction energies of the prism isomer (isomer 1 in Ref. 96) of the water hexamer calculated using QM/MB-pol and QM/MB-DFT for all possible QM/MM partitions involving 1 to 5 water molecules in the QM region. The grey line is the MB-pol reference energy and the blue line is the PBE0 reference energy. The numbers at the top of each column determine the number of water molecules in the QM and MM regions (num QM / num MM).

age, within  $\sim 1.5$  kcal/mol of the reference MB-pol energies, with only one partition of the psc-QM/MM-pol scheme displaying a deviation of  $\sim 2.3$  kcal/mol. Successively including one water to the QM region linearly shifts the distribution of QM/MM-pol energies towards the QM value. When the QM/MM partition contains five water molecules in the QM region, the QM/MM-pol energies lie within 0.6 and 1.5 kcal/mol of the PBE0-D3 value for the fsc-QM/MM-pol and psc-QM/MM-pol schemes, respectively. On the other hand, fsc-QM/MM-DFT and psc-QM/MM-DFT calculations for the partition with one water molecule in the QM region predict interaction energies that are within 1.0 and 2.0 kcal/mol, respectively, of the reference PBE0-D3 value. The successive addition of water molecules results in fsc-QM/MM-DFT and psc-QM/MM-DFT interaction energies that are within 1.5 and 1.3 kcal/mol, respectively, of the PBE0-D3 reference value for the partition containing five water molecules in the QM region. The largest deviations (up to 2.25 kcal/mol for fsc-QM/MM-DFT and up to 3.54 kcal/mol for psc-QM/MM-DFT) relative to the PBE0-D3 value are found for the symmetric partition with three water molecules in both the QM and MM regions. Furthermore, independently of using the partially or fully self-consistent schemes, the largest deviations in the distributions of both QM/MM-pol and QM/MM-DFT interaction energies are associated with configurations in which the QM water molecule is a double hydrogen-bond acceptor from two MM water molecules. This is consistent with the results in Section 3.1 where the case of the MM donor (QM acceptor) displays the largest error in the scan of the water dimer potential energy surface.

Compared to the QM/MM-pol interaction energies, which are increasingly weighted towards the PBE0-D3 reference value as the number of water molecules in the QM region increases, the QM/MM-DFT results remain relatively close to the PBE0-D3 reference value, independently of how the water molecules are partitioned in the QM and MM regions. The different trend found for interaction energies calculated with QM/MM-pol and QM/MM-DFT as a function of the number of water molecules in the QM region can be explained by considering that the only differences between the MB-pol and MB-DFT potential energy functions derive from the differences in the corresponding 2B and 3B PIPs. Since the MB-pol PIPs were fitted to CCSD(T)/CBS 2B and 3B

energies, this implies that, when the QM/MM partition includes only one water molecule in the QM region, all possible 2B and 3B energies within the hexamer contains contributions from the corresponding MB-pol PIPs. In the opposite limit, when five water molecules are placed in the QM region, only 6 out of 15 2B energies and 10 out of 20 3B energies include contributions from the corresponding MB-pol PIPs. As a result, the QM/MB-pol interaction energy approaches the PBE0-D3 reference value as the number of water molecules in the QM region increases because the contributions from the MB-pol PIPs become progressively smaller. In contrast, since the MB-DFT PIPs were fitted to PBE0-D3 2B and 3B energies, all possible 2B and 3B contributions to the QM/MB-DFT interaction energy, independently of the QM/MM partition, are effectively described at the PBE0-D3 level (either through actual PBE0-D3 calculations in the QM region or through the MB-DFT PIPs in the MM region). As a result, the QM/MB-DFT interaction energies remain relatively close to the PBE0-D3 reference value for all different partitions, that the QM/MB-DFT scheme enables smooth transitions between QM regions described by an arbitrary DFT model and MM regions described by the corresponding MB-DFT potential energy function, which is particularly appealing for QM/MM simulations in solution.

Fig. 3 shows that, while the psc-QM/MB-pol and psc-QM/MB-DFT scheme displayed smaller RMSDs than the corresponding fully self-consistent schemes for energetics of water dimers (Fig. 2), they predict wider spread than the corresponding fully self-consistent schemes the in the interaction energies of the prism hexamer, depending on the spatial arrangements of the water molecules in the QM and MM regions. In particular, the psc-QM/MB-pol and psc-QM/MB-DFT calculations predict the largest spread of 4.5 kcal/mol for the symmetric QM/MM partition, with 3 water molecules in both QM and MM regions, which must be compared with a spread of  $\sim 1.5$  kcal/mol predicted by the corresponding fully self-consistent schemes. This symmetric QM/MM partition contains the largest number of different spatial arrangements of the QM and MM molecules, resulting in 20 total permutations, as well as the largest number of 2B (9) and 3B (18) QM/MM energies. The large number of 2B and 3B QM/MM interactions serves to amplify possible inaccuracies of the 2B and 3B PIPs as even relatively small errors in the description of individual 2B and

3B energies add up, resulting in larger total errors and, consequently, wider spreads for interaction energies calculated for different QM/MM partitions.

The analysis reported in Fig. 3 underscores the importance of many-body polarization effects in QM/MM calculations. In particular, the relatively wider spreads associated, on average, with interaction energies of the water hexamer calculated with the psc-QM/MB-pol and psc-QM/MB-DFT schemes are due to the neglect of fully self-consistent polarization effects between the QM and MM regions, which are included in the corresponding fsc-QM/MB-pol and fsc-QM/MB-DFT schemes. These findings are further underscored by observing that both fully and partially self-consistent polarization schemes predict similar spreads of the interaction energies for the two limiting QM/MM partitions, with 1 and 5 water molecules in the QM region, respectively, which correspond to the two QM/MM partitions with the smallest contributions from mutual polarization between the QM and MM regions.

## 4 Conclusions

We presented here a new development in QM/MM, integrating the MB-pol and MB-DFT models into a polarizable embedding, allowing the dipoles in the MM region to self-consistently equilibrate with the density in the QM region. On top of this embedding, the PIPs used in MB-pol and MB-DFT, which are fitted to recover the CCSD(T) and PBE0-D3 energies at the 2B and 3B levels, are incorporated into the QM/MM interaction. We show that use of both QM/MB-pol and QM/MB-DFT increases the overall accuracy of the QM/MM interaction, as the PIPs account for the leading order dispersion and repulsion energies. Using both partially and fully self consistent mutual polarization between the QM and MM region, we show that a fully self consistent scheme is required to fully capture the polarization effects in larger and more complex water clusters. More importantly, we demonstrate a consistent treatment between QM and MM regions. Using MB-DFT, the MM region can be represented at the same level of accuracy as an arbitrary DFT functional used in the QM region. In our case, using PBE0-D3 as the functional of choice, we can

consistently represent an entire QM/MM system, with the pure MM, QM, and QM/MM interactions at the PBE0-D3 level of theory. As demonstrated in our analysis of the water hexamer, this consistent scheme mitigates the change in energy when a molecule is added to or removed from the QM region, as might happen during an adaptive QM/MM simulation. The QM/MB-DFT scheme offers the advantage, that since both the pure QM and MM components, as well as the QM/MM interaction are treated at the same level of accuracy, the need for a transition layer is reduced or eliminated. In more succinct terms, this approach can be thought of as treating the entire system under the same "effective" Hamiltonian, with different, but numerically equivalent representations for the QM and MM components.

## Acknowledgement

E.L. thanks Kartik Lakshmi Rallapalli, Kelly Hunter, Colin Egan, Marc Riera, and Jay Hu for stimulating discussion about QM/MM. All figures in this paper were generated with UCSF Chimera and Matplotlib.<sup>97,98</sup> This research was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Science, through grant No. DE-SC0019490 (F.P.), and NIGMS/NIH through grant No. R01GM108583 (G.A.C.). F.L. acknowledges support from Gaussian, inc. The authors are grateful to the MoLECoLab group in Pisa for granting access to the LIFETimeS cluster, that was used to perform all the fsc-QM/MM calculations. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231, as well as the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputer Center (SDSC).

## Supporting Information Available

Analyses of the interaction energies of eight low-lying isomers of the water hexamer calculated using different QM/MM partitions.

## References

- (1) Warshel, A.; Karplus, M. Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I. Formulation and Parametrization. *J. Am. Chem. Soc.* **1972**, *94*, 5612–5625.
- (2) 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.
- (3) Singh, U. C.; Kollman, P. A. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the  $\text{CH}_3\text{Cl} + \text{Cl}^-$  Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718–730.
- (4) Duster, A. W.; Garza, C. M.; Aydintug, B. O.; Negussie, M. B.; Lin, H. Adaptive Partitioning QM/MM for Molecular Dynamics Simulations: 6. Proton Transport Through a Biological Channel. *J. Chem. Theory Comput.* **2019**, *15*, 892–905.
- (5) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- (6) Morzan, U. N.; Alonso de Armino, D. J.; Foglia, N. O.; Ramirez, F.; Gonzalez Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QM–MM Simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.
- (7) Halgren, T. A.; Damm, W. Polarizable Force Fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242.
- (8) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.

- (9) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **2019**, *48*, 371–394.
- (10) Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
- (11) Gao, J. Energy Components of Aqueous Solution: Insight From Hybrid QM/MM Simulations Using a Polarizable Solvent Model. *J. Comput. Chem.* **1997**, *18*, 1061–1071.
- (12) Gooding, S. R.; Winn, P. J.; Maurer, R. I.; Ferenczy, G. G.; Miller, J. R.; Harris, J. E.; Griffiths, D. V.; Reynolds, C. A. Fully Polarizable QM/MM Calculations: An Application To the Nonbonded Iodine–Oxygen Interaction in Dimethyl-2-Iodobenzoylphosphonate. *J. Comput. Chem.* **2000**, *21*, 478–482.
- (13) Illingworth, C. J. R.; Gooding, S. R.; Winn, P. J.; Jones, G. A.; Ferenczy, G. G.; Reynolds, C. A. Classical Polarization in Hybrid QM/MM Methods. *J. Phys. Chem. A* **2006**, *110*, 6487–6497.
- (14) Illingworth, C.; Parkes, K.; Snell, C.; Marti, S.; Moliner, V.; Reynolds, C. The Effect of MM Polarization on the QM/MM Transition State Stabilization: Application to Chorismate Mutase. *Mol. Phys.* **2008**, *106*, 1511–1515.
- (15) Kratz, E. G.; Walker, A. R.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Andrés Cisneros, G. LICHEM: A QM/MM Program for Simulations with Multipolar and Polarizable Force Fields. *J. Comput. Chem.* **2016**, *37*, 1019–1029.
- (16) Gökcan, H.; Vázquez-Montelongo, E. A.; Cisneros, G. A. LICHEM 1.1: Recent Improvements and New Capabilities. *J. Chem. Theory Comput.* **2019**, *15*, 3056–3065.
- (17) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. Elec-



- tronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.
- (18) Caprasecca, S.; Jurinovich, S.; Viani, L.; Curutchet, C.; Mennucci, B. Geometry Optimization in Polarizable QM/MM Models: The Induced Dipole Formulation. *J. Chem. Theory Comput.* **2014**, *10*, 1588–1598.
- (19) Dziedzic, J.; Mao, Y.; Shao, Y.; Ponder, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K. TINKTEP: A Fully Self-Consistent, Mutually Polarizable QM/MM Approach Based on the AMOEBA Force Field. *J. Chem. Phys.* **2016**, *145*, 124106.
- (20) Dziedzic, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K. Mutually Polarizable QM/MM Model With in Situ Optimized Localized Basis Functions. *J. Chem. Phys.* **2019**, *150*, 074103.
- (21) Loco, D.; Polack, E.; 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.
- (22) Olsen, J. M. H.; Kongsted, J. In *Advances in Quantum Chemistry*; Sabin, J. R., Brändas, E., Eds.; Elsevier, 2011; pp 107–143.
- (23) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.
- (24) 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.
- (25) Loco, D.; Jurinovich, S.; Cupellini, L.; Menger, M. F.; Mennucci, B. The Modeling of the Absorption Lineshape for Embedded Molecules Through a Polarizable QM/MM Approach. *Photochem. Photobiol. Sci.* **2018**, *17*, 552–560.

- (26) Bryce, R. A.; Buesnel, R.; Hillier, I. H.; Burton, N. A. A Solvation Model Using a Hybrid Quantum Mechanical/Molecular Mechanical Potential With Fluctuating Solvent Charges. *Chem. Phys. Lett.* **1997**, 279, 367 – 371.
- (27) Lipparini, F.; Cappelli, C.; Scalmani, G.; De Mitri, N.; Barone, V. Analytical First and Second Derivatives for a Fully Polarizable QM/Classical Hamiltonian. *J. Chem. Theory Comput.* **2012**, 8, 4270–4278.
- (28) 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.
- (29) Giovannini, T.; Puglisi, A.; Ambrosetti, M.; Cappelli, C. Polarizable QM/MM Approach with Fluctuating Charges and Fluctuating Dipoles: The QM/FQF $\mu$  Model. *J. Chem. Theory Comput.* **2019**, 15, 2233–2245.
- (30) Lipparini, F.; Cappelli, C.; Barone, V. A Gauge Invariant Multiscale Approach to Magnetic Spectroscopies in Condensed Phase: General Three-Layer Model, Computational Implementation and Pilot Applications. *J. Chem. Phys.* **2013**, 138, 234108.
- (31) Cappelli, C. Integrated QM/Polarizable MM/Continuum Approaches to Model Chiroptical Properties of Strongly Interacting Solute–Solvent Systems. *Int. J. Quantum Chem.* **2016**, 116, 1532–1542.
- (32) Lu, Z.; Zhang, Y. Interfacing ab Initio Quantum Mechanical Method with Classical Drude Oscillator Polarizable Model for Molecular Dynamics Simulation of Chemical Reactions. *J. Chem. Theory Comput.* **2008**, 4, 1237–1248.
- (33) Ganguly, A.; Boulanger, E.; Thiel, W. Importance of MM Polarization in QM/MM Studies of Enzymatic Reactions: Assessment of the QM/MM Drude Oscillator Model. *J. Chem. Theory Comput.* **2017**, 13, 2954–2961.

- (34) König, G.; Pickard, F. C.; Huang, J.; Thiel, W.; MacKerell, A. D.; Brooks, B. R.; York, D. M. A Comparison of QM/MM Simulations With and Without the Drude Oscillator Model Based on Hydration Free Energies of Simple Solutes. *Mol.* **2018**, *23*, 2695.
- (35) Riahi, S.; Rowley, C. N. The CHARMM–TURBOMOLE Interface for Efficient and Accurate QM/MM Molecular Dynamics, Free Energies, and Excited State Properties. *J. Comput. Chem.* **2014**, *35*, 2076–2086.
- (36) Rowley, C. N.; Roux, B. The Solvation Structure of Na<sup>+</sup> and K<sup>+</sup> in Liquid Water Determined from High Level ab Initio Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 3526–3535.
- (37) Sahoo, S. K.; Nair, N. N. Interfacing the Core-Shell or the Drude Polarizable Force Field With Car-Parrinello Molecular Dynamics for QM/MM Simulations. *Front. Chem.* **2018**, *6*, 275.
- (38) Hu, H.; Lu, Z.; Yang, W. QM/MM Minimum Free-Energy Path: Methodology and Application to Triosephosphate Isomerase. *J. Chem. Theory Comput.* **2007**, *3*, 390–406.
- (39) Nakano, H.; Yamamoto, T. Variational Calculation of Quantum Mechanical/Molecular Mechanical Free Energy With Electronic Polarization of Solvent. *J. Chem. Phys.* **2012**, *136*, 134107.
- (40) Babin, V.; Leforestier, C.; Paesani, F. Development of a “First Principles” Water Potential With Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.* **2013**, *9*, 5395–5403.
- (41) Babin, V.; Medders, G. R.; Paesani, F. Development of a “First Principles” Water Potential With Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. *J. Chem. Theory Comput.* **2014**, *10*, 1599–1607.

- (42) Medders, G. R.; Babin, V.; Paesani, F. Development of a “First Principles” Water Potential With Flexible Monomers. III. Liquid Phase Properties. *J. Chem. Theory Comput.* **2014**, *10*, 2906–2910.
- (43) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
- (44) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B* **2015**, *119*, 9423–9437.
- (45) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route To Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632–672.
- (46) Gao, J.; Truhlar, D. G.; Wang, Y.; Mazack, M. J. M.; Löffler, P.; Provorse, M. R.; Rehak, P. Explicit Polarization: A Quantum Mechanical Framework for Developing Next Generation Force Fields. *Acc. Chem. Res.* **2014**, *47*, 2837–2845.
- (47) Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J. Density Functional Self-Consistent Quantum Mechanics/Molecular Mechanics Theory for Linear and Nonlinear Molecular Properties: Applications to Solvated Water and Formaldehyde. *J. Chem. Phys.* **2007**, *126*, 154112.
- (48) Caprasecca, S.; Jurinovich, S.; Lagardère, L.; Stamm, B.; Lipparini, F. Achieving Linear Scaling in Computational Cost for a Fully Polarizable MM/Continuum Embedding. *J. Chem. Theory Comput.* **2015**, *11*, 694–704.
- (49) Lipparini, F. General Linear Scaling Implementation of Polarizable Embedding Schemes. *J. Chem. Theory Comput.* **2019**, *15*, 4312–4317.
- (50) Bondanza, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. The Multiple Roles of the Protein in the Photoactivation of Orange Carotenoid Protein. *Chem* **2020**, *6*, 187–203.

- (51) Loco, D.; Lagardère, L.; Cisneros, G. A.; Scalmani, G.; Frisch, M.; Lipparini, F.; Men-  
nucci, B.; Piquemal, J.-P. Towards Large Scale Hybrid QM/MM Dynamics of Complex Sys-  
tems With Advanced Point Dipole Polarizable Embeddings. *Chem. Sci.* **2019**, *10*, 7200–7211.
- (52) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. A QM/MM Simulation Method Applied To the  
Solution of  $\text{Li}^+$  in Liquid Ammonia. *Chem. Phys.* **1996**, *211*, 313–323.
- (53) Buló, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. Toward a Practical Method for Adaptive  
QM/MM Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2212–2221.
- (54) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. The Number-Adaptive Multiscale  
QM/MM Molecular Dynamics Simulation: Application To Liquid Water. *Chem. Phys. Lett.*  
**2012**, *524*, 56–61.
- (55) Duster, A. W.; Wang, C.-H.; Garza, C. M.; Miller, D. E.; Lin, H. Adaptive Quan-  
tum/Molecular Mechanics: What Have We Learned, Where Are We, and Where Do We  
Go From Here? *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, e1310.
- (56) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: An Extension of the ONIOM Method for  
Molecular Simulation in Condensed Phase. *Chem. Phys. Lett.* **2002**, *355*, 257–262.
- (57) Bernstein, N.; Várnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.;  
Csányi, G. QM/MM Simulation of Liquid Water With an Adaptive Quantum Region. *Phys.*  
*Chem. Chem. Phys.* **2012**, *14*, 646–656.
- (58) Heyden, A.; Lin, H.; Truhlar, D. G. Adaptive Partitioning in Combined Quantum Mechan-  
ical and Molecular Mechanical Calculations of Potential Energy Functions for Multiscale  
Simulations. *J. Phys. Chem. B.* **2007**, *111*, 2231–2241.
- (59) Park, K.; Götz, A. W.; Walker, R. C.; Paesani, F. Application of Adaptive QM/MM Methods  
to Molecular Dynamics Simulations of Aqueous Systems. *J. Chem. Theory Comput.* **2012**, *8*,  
2868–2877.

- (60) Pezeshki, S.; Lin, H. Adaptive-Partitioning Redistributed Charge and Dipole Schemes for QM/MM Dynamics Simulations: On-the-Fly Relocation of Boundaries That Pass Through Covalent Bonds. *J. Chem. Theory Comput.* **2011**, *7*, 3625–3634.
- (61) Buló, 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.
- (62) Waller, M. P.; Kumbhar, S.; Yang, J. A Density-Based Adaptive Quantum Mechanical/Molecular Mechanical Method. *ChemPhysChem* **2014**, *15*, 3218–3225.
- (63) Watanabe, H. C.; Kubař, T.; Elstner, M. Size-Consistent Multipartitioning QM/MM: A Stable and Efficient Adaptive QM/MM Method. *J. Chem. Theory Comput.* **2014**, *10*, 4242–4252.
- (64) Boereboom, J. M.; Potestio, R.; Donadio, D.; Buló, R. E. Toward Hamiltonian Adaptive QM/MM: Accurate Solvent Structures Using Many-Body Potentials. *J. Chem. Theory Comput.* **2016**, *12*, 3441–3448.
- (65) Zheng, M.; Waller, M. P. Adaptive Quantum Mechanics/Molecular Mechanics Methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2016**, *6*, 369–385.
- (66) Dohm, S.; Spohr, E.; Korth, M. Developing Adaptive QM/MM Computer Simulations for Electrochemistry. *J. Comput. Chem.* **2017**, *38*, 51–58.
- (67) Field, M. J. An Algorithm for Adaptive QC/MM Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 2342–2351.
- (68) Boereboom, J. M.; Fleurat-Lessard, P.; Buló, R. E. Explicit Solvation Matters: Performance of QM/MM Solvation Models in Nucleophilic Addition. *J. Chem. Theory Comput.* **2018**, *14*, 1841–1852.
- (69) Duster, A. W.; Wang, C.-H.; Lin, H. Adaptive QM/MM for Molecular Dynamics Simulations: 5. On the Energy-Conserved Permuted Adaptive-Partitioning Schemes. *Molecules* **2018**, *23*, 2170.

- (70) Watanabe, H. C.; Cui, Q. Quantitative Analysis of QM/MM Boundary Artifacts and Correction in Adaptive QM/MM Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 3917–3928.
- (71) Zheng, M.; Waller, M. P. Yoink: An Interaction-Based Partitioning API. *J. Comput. Chem.* **2018**, *39*, 799–806.
- (72) 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.
- (73) Chen, W.; Gordon, M. S. The Effective Fragment Model for Solvation: Internal Rotation in Formamide. *J. Chem. Phys.* **1996**, *105*, 11081–11090.
- (74) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach To Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- (75) Ghosh, D.; Kosenkov, D.; Vanovschi, V.; Flick, J.; Kaliman, I.; Shao, Y.; Gilbert, A. T.; Krylov, A. I.; Slipchenko, L. V. Effective Fragment Potential Method in Q-CHEM: A Guide for Users and Developers. *J. Comput. Chem.* **2013**, *34*, 1060–1070.
- (76) Chen, X.; Gao, J. Fragment Exchange Potential for Realizing Pauli Deformation of Interfragment Interactions. *J. Phys. Chem. Lett.* **2020**, *11*, 4008–4016.
- (77) Viquez Rojas, C. I.; Slipchenko, L. V. Exchange Repulsion in Quantum Mechanical/Effective Fragment Potential Excitation Energies: Beyond Polarizable Embedding. *J. Chem. Theory Comput.* **2020**, DOI:10.1021/acs.jctc.9b01156.
- (78) Piquemal, J.-P.; Cisneros, G. In *Many-Body Effects and Electrostatics in Multi-Scale Computations of Biomolecules*; Cui, Q., Ren, P., Meuwly, M., Eds.; Pan Stanford Publishing, 2015; pp 269–300.

- (79) Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A. Quantum Mechanics/Molecular Mechanics Electrostatic Embedding with Continuous and Discrete Functions. *J. Phys. Chem. B* **2006**, *110*, 13682–13684.
- (80) Gökcan, H.; Kratz, E.; Darden, T. A.; Piquemal, J.-P.; Cisneros, G. A. QM/MM Simulations With the Gaussian Electrostatic Model: A Density-Based Polarizable Potential. *J. Phys. Chem. Lett.* **2018**, *9*, 3062–3067.
- (81) Riera, M.; Lambros, E.; Nguyen, T. T.; Götz, A. W.; Paesani, F. Low-Order Many-Body Interactions Determine the Local Structure of Liquid Water. *Chem. Sci.* **2019**, *10*, 8211–8218.
- (82) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (83) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (84) Dunning Jr, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (85) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (86) 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 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

- (87) Hankins, D.; Moskowitz, J.; Stillinger, F. Water Molecule Interactions. *J. Chem. Phys.* **1970**, *53*, 4544–4554.
- (88) Partridge, H.; Schwenke, D. W. The Determination of an Accurate Isotope Dependent Potential Energy Surface for Water From Extensive Ab Initio Calculations and Experimental Data. *J. Chem. Phys.* **1997**, *106*, 4618–4639.
- (89) Braams, B. J.; Bowman, J. M. Permutationally Invariant Potential Energy Surfaces in High Dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.
- (90) Tang, K.; Toennies, J. P. An Improved Simple Model for the Van der Waals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *J. Chem. Phys.* **1984**, *80*, 3726–3741.
- (91) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; Van der Avoird, A. Predictions of the Properties of Water From First Principles. *Science* **2007**, *315*, 1249–1252.
- (92) Bondanza, M.; Nottoli, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. Polarizable Embedding QM/MM: The Future Gold Standard for Complex (Bio)systems? *Phys. Chem. Chem. Phys.* **2020**, *22*, 14433–14448.
- (93) Curutchet, C.; Cupellini, L.; Kongsted, J.; Corni, S.; Frediani, L.; Steindal, A. H.; Guido, C. A.; Scalmani, G.; Mennucci, B. Density-Dependent Formulation of Dispersion–Repulsion Interactions in Hybrid Multiscale Quantum/Molecular Mechanics (QM/MM) Models. *J. Chem. Theory Comput.* **2018**, *14*, 1671–1681.

- (94) 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.
- (95) Brown, S. E.; Götz, A. W.; Cheng, X.; Steele, R. P.; Mandelshtam, V. A.; Paesani, F. Monitoring Water Clusters “Melt” Through Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2017**, *139*, 7082–7088.
- (96) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the Accuracy of the MB-pol Many-Body Potential for Water: Interaction Energies, Vibrational Frequencies, and Classical Thermodynamic and Dynamical Properties From Clusters To Liquid Water and Ice. *J. Chem. Phys.* **2016**, *145*, 194504.
- (97) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- (98) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

TOC Figure

