Performance of the AMOEBA water model in the vicinity of QM solutes: A diagnosis using energy decomposition analysis

Yuezhi Mao,[†] Yihan Shao,[*,‡] Jacek Dziedzic,[¶,§] Chris-Kriton Skylaris,[¶] Teresa Head-Gordon,[†,‖] and Martin Head-Gordon[*,†,⊥]

† *Kenneth S. Pitzer Center for Theoretical Chemistry, Department of Chemistry, University of California at Berkeley, Berkeley, CA 94720, USA*

‡ *Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73019, USA*

¶ *School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK*

§ *Faculty of Applied Physics and Mathematics, Gdańsk University of Technology, Gdańsk, Poland*

‖ *Department of Bioengineering, Department of Chemical and Biomolecular Engineering, University of California at Berkeley, Berkeley, CA 94720, USA*

⊥ *Chemical Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

E-mail: yihan.shao@ou.edu; mhg@cchem.berkeley.edu

Abstract

The importance of incorporating solvent polarization effects into the modeling of solvation processes has been well-recognized, and therefore a new generation of hybrid quantum mechanics/molecular mechanics (QM/MM) approaches that accounts for this effect is desirable. We present a fully self-consistent, mutually polarizable QM/MM scheme using the AMOEBA force field, in which the total energy of the system is variationally minimized with respect to both the QM electronic density and the MM induced dipoles. This QM/AMOEBA model is implemented through the Q-Chem/LibEFP code interface and then applied to the evaluation of solute-solvent interaction energies for various systems ranging from the water dimer to neutral and ionic solutes (NH_3, NH_4^+, CN^-) surrounded by increasing numbers of water molecules (up to 100). In order to analyze the resulting interaction energies, we also utilize an energy decomposition analysis (EDA) scheme which identifies contributions from permanent electrostatics, polarization and van der Waals (vdW) interaction for the interaction between the QM solute and the solvent molecules described by AMOEBA. This facilitates a component-wise comparison against full QM calculations where the corresponding energy components are obtained via a modified version of the absolutely localized molecular orbitals (ALMO)-EDA. The results show that the present QM/AMOEBA model can yield reasonable solute-solvent interaction energies for neutral and cationic species, while further scrutiny reveals that this accuracy highly relies on the delicate balance between insufficiently favorable permanent electrostatics and softened vdW interaction. For anionic solutes where the charge penetration effect becomes more pronounced, the QM/MM interface turns out to be unbalanced. These results are consistent with and further elucidate our findings in a previous study using a slightly different QM/AMOEBA model (*J. Chem. Phys.* **2016**, *145*, 124106). The implications of these results for further refinement of this model are also discussed.

1 Introduction

The solvation process, where a solute molecule interacts with water or other solvent molecules, plays a key role in many chemical and biochemical systems. In chemical reactions, the success with which inorganic or organic compounds are synthesized can be greatly affected by the choice of the solvent. In macromolecular systems, the rates for enzymatic reactions or protein folding and the strengths of ligand-receptor binding are also dependent upon solvation, because all these processes involve partial exposure of key chemical groups to the solvent.

Given its central role in both basic and applied sciences, molecular solvation has been investigated using *ab initio* quantum chemistry methods in conjunction with implicit or explicit solvent models in numerous theoretical and computational studies. Implicit solvent models,[1] which ignore the molecular resolution of the solvent, have a long history dating back to Born[2] and Onsager.[3] In quantum mechanical (QM) calculations, implicit descriptions for solvent molecules remain in wide use today, mostly in the formulation of generalized Born (GB) models,[4–6] apparent surface charge (ASC) models,[7–18] and models based on direct solution of non-homogeneous Poisson-Boltzmann equations.[19–22] Despite their huge successes — solvation free energies for neutral molecules can be predicted on average within 1.0 kcal/mol[23,24]— implicit solvent models for QM calculations can be inadequate in several situations: (a) Larger errors in the predicted solvation free energies are found for ionic solutes,[23] which can interact strongly with solvent molecules through permanent electrostatics, polarization, and charge transfer; (b) It is rather difficult to describe systems that are partly inhomogeneous, such as ionic liquids (and other mixed solvents) or one solvent at different pH conditions; (c) It is clearly ill-suited for describing completely inhomogeneous environments, such as a "solute" ligand in a partially exposed binding pocket that is accessible to water or other solvent molecules.

Explicit solvent models, where the solvent molecules receive an all-atom or united-atom description, can in principle be employed in such situations that are challenging for implicit solvent models. Ideally, one would like to perform fully *ab initio* QM (such as density functional theory (DFT) or perturbation theory (PT)) calculations on the solute molecule together with a larger number of solvent molecules. While there have been many fully *ab initio* molecular dynamics (AIMD) or Monte Carlo (MC) simulations on liquid water[25–31] and ion-water systems[32–39] reported, their routine use is still beyond the reach of most researchers, due to the substantial cost associated with computing the *ab initio* electronic structure for each configuration, and the enormous number of configurations required to adequately sample the configuration space. Even when feasible, AIMD simulations using DFT are *not always* guaranteed to produce accurate results for bulk liquid or solute-solvent systems as the quality of results depends heavily on the delicate interplay between the density functional and the dispersion correction employed,[40–47] while many PT methods are known to overestimate dispersion even for smaller systems.

Explicit solvent models utilizing hybrid quantum mechanical molecular mechanical (QM/MM) energy functions[48–53] come as a natural compromise between computational efficiency and accuracy. Through treating the solute molecule with *ab initio* QM methods and the solvent molecules with molecular mechanics (MM) force fields,[54–60] it vastly reduces the computational demand compared to full QM calculations, while potentially retaining the QM accuracy for the simulation results. Out of three categories of interactions existing in a solute-solvent system (intramolecular interactions within the solute, solute-solvent interactions, and solvent-solvent interactions), *it is the most crucial to have an accurate description for the solute-solvent interactions.* This is because one can usually find a QM level of theory (such as Kohn-Sham (KS)-DFT[61]) to reliably describe intramolecular interactions within the solute, and because, for solvent-solvent interactions, one can take advantage of the error cancellation within the MM model or reduce the error by averaging over the sampling

ensemble.

In QM/MM based solvation models, the solute-solvent interactions consist of five components: permanent electrostatics, forward (MM $\rightarrow$ QM) polarization, backward (QM $\rightarrow$ MM) polarization, exchange repulsion, and dispersion. (Note: forward and backward polarizations are numerically inseparable if mutual polarization is enabled.) Most of the recent algorithm developments on QM/MM interactions have focused on: a) improving the permanent electrostatics and the forward polarization by adopting a multipolar representation of solvent electrons[62–70] and by introducing damping schemes to account for the spread of solvent valence electron density;[62,63,71–76] and b) adding the backward polarization through employing a polarizable force field for the solvent molecules, including models based on Drude oscillators [77–80], fluctuating charges[81–83] and inducible dipoles.[62–70,84–94]

In QM/MM calculations, exchange repulsion and dispersion interactions are usually combined together in a classical vdW potential using the Lennard-Jones ("12-6") or Halgren ("14-7") formula.[95] To date, there is no fully automated procedure to assign vdW parameters to QM atoms, and there is a lack of systematic ways to improve the description of QM/MM vdW interactions, which potentially limits the accuracy of QM/MM results in general. But vdW interactions are just as important as permanent electrostatics and polarization interactions in QM/MM calculations.[96] In two recent publications on hydration free energies,[97,98] for example, it was reported that the BLYP functional[99,100] (for the QM region) can produce more accurate results than more sophisticated functionals, when the solvent water molecules are described by the TIP3P[101] model. This happens only because the QM/MM permanent electrostatics based on BLYP electron density is the most compatible with the employed classical QM/MM vdW potential. In another recent publication,[70] we also demonstrated that the buffered 14-7 potential of AMOEBA needs to be adjusted to reproduce full QM values for the investigated solute-solvent interaction energies.

In this article, a new implementation of DFT/AMOEBA calculations within the Q-

Chem/LibEFP[102–104] software framework is reported, with a complete derivation of equations for self-consistent field (SCF) calculations. The AMOEBA polarizable force field,[105–107] which places permanent multipoles (up to quadrupoles) and induced dipoles on each solvent atom, is supposed to improve the description of QM/MM permanent electrostatics and forward/backward polarization interactions. Our implementation complements recent efforts within the Gaussian,[69] LICHEM,[68] and ONETEP/TINKER[70] software platforms, and further improves the accessibility to mutually polarizable QM/MM calculations using the AMOEBA force field.

Equally importantly, we propose a scheme to decompose the solute-solvent interaction energy evaluated by mutually polarizable QM/MM into contributions from permanent electrostatics, polarization and vdW interaction. This method has similar objectives to the semi-empirical QM/MM energy decomposition analysis (EDA) scheme suggested by Gao[108] that was utilized to elucidate the role of polarization in solute-solvent interactions, and more recently, a similar approach (but not exactly the same) was also proposed by Hirao et al.[109] to investigate the stabilization effect of the protein environment on the active site in a polarizable QM/MM calculation for cytochrome P450cam, which turns out to be a useful tool for interpreting the results of QM/MM studies of protein modeling. Here, we employ the EDA method to facilitate a component-wise comparison against full QM references obtained via performing the absolutely localized molecular orbitals based energy decomposition analysis (ALMO-EDA).[110–112] This allows us to thoroughly analyze all individual components of the solute-solvent interaction energy, and provides valuable insights for guiding future improvements to the QM/MM modeling of solute-solvent interactions. We note that there have been many other protocols developed for decomposing full QM intermolecular interactions[113] so that the energy components are not uniquely defined. Nevertheless, the choice of the ALMO-EDA scheme (with necessary modifications) has been validated by a recent benchmark study of the AMOEBA force field by several of us.[114] More details about the

QM/MM and full QM EDA schemes employed in this work are provided in Sec. 2.2.

2 Methods

2.1 Fully polarizable QM/AMOEBA SCF calculations

The total energy of the coupled KS-DFT/AMOEBA system can be expressed as

$$\begin{aligned} E_{\text{tot}}^{\text{QM/MM}} = & E_{\text{nuc-nuc}}^{\text{QM}} + E_{\text{val}}^{\text{MM}} + E_{\text{elec}}^{\text{MM}} + E_{\text{vdw}}^{\text{MM}} \\ & + E_{\text{vdw}}^{\text{QM/MM}} + E_{\text{elec}}^{\text{QM(nuc)/MM}} \\ & + E_{\text{el}}^{\text{QM}} + E_{\text{elec}}^{\text{QM(el)/MM}} + E_{\text{pol}}^{\text{MM}}. \end{aligned} \tag{1}$$

The first six terms in Eq. (1) do not depend on the electron density of the QM region. Among them, the first four terms can be evaluated with a QM or MM region in isolation: $E_{\text{nuc-nuc}}^{\text{QM}}$ represents the repulsion between QM nuclei, $E_{\text{val}}^{\text{MM}}$ is the sum of all the valence terms (bond, angle, and Urey-Bradley) in AMOEBA, and $E_{\text{elec}}^{\text{MM}}$ and $E_{\text{vdw}}^{\text{MM}}$ refer to the permanent electrostatic and van der Waals (vdW) interactions between AMOEBA fragments, respectively. For a more detailed introduction to these terms, we refer the readers to the original references of AMOEBA.[105–107]

The next two terms are two components of the interaction crossing the QM and MM boundary, i.e., the vdW interaction between QM and MM atoms and the electrostatic interaction between the QM nuclei and the permanent multipoles (up to quadrupoles) of the AMOEBA water molecules. In our current model, the QM/MM vdW interaction energy is computed at the MM level, i.e., we assign AMOEBA's vdW parameters to each QM atom. And for the evaluation of $E_{\text{elec}}^{\text{QM(nuc)/MM}}$, the same equations for computing charge-charge, charge-dipole and charge-quadrupole interactions in AMOEBA can simply be applied.

The last three terms are electron-density-dependent so they need to be minimized through

an SCF procedure. $E_{\rm el}^{\rm QM}$ is defined as the KS energy associated with the present electron density *without* accounting for its interaction with the MM embedding potential, and we shall denote the corresponding Fock matrix contribution as

$$\mathbf{F}_0 = \partial E_{\rm el}^{\rm QM}/\partial \mathbf{P}, \tag{2}$$

where $\mathbf{P}$ is the density matrix for the QM electrons.

The second density-dependent term, $E_{\rm elec}^{\rm QM(el)/MM}$, refers to the Coulomb interaction energy between the QM electrons and the permanent multipoles on the AMOEBA fragments:

$$E_{\rm elec}^{\rm QM(el)/MM} = \int {\rm d}\mathbf{r}\ \rho_{\rm el}(\mathbf{r}) V_{\text{m-pole}}^{\rm MM}(\mathbf{r}), \tag{3}$$

where $V_{\text{m-pole}}^{\rm MM}(\mathbf{r})$ is the electrostatic potential (ESP) generated by the AMOEBA multipoles. Within the AMOEBA force field, the permanent multipoles ($\{\mathbf{M}_i\}$) are usually located on each atomic site i, and $\mathbf{M}_i = \{q_i, \boldsymbol{\mu}_i, \mathbf{Q}_i\}$. Taking all the atomic sites in the MM region together, we have

$$\begin{aligned} V_{\text{m-pole}}^{\rm MM}(\mathbf{r}) = \sum_{i=1}^{n_{\rm MM}} (& \frac{q_i}{|\mathbf{r}-\mathbf{R}_i|} - \nabla_{\mathbf{r}} \frac{1}{|\mathbf{r}-\mathbf{R}_i|} \cdot \boldsymbol{\mu}_i \\ & + \frac{1}{3}\ \nabla_{\mathbf{r}} \nabla_{\mathbf{r}} \frac{1}{|\mathbf{r}-\mathbf{R}_i|} : \mathbf{Q}_i), \end{aligned} \tag{4}$$

where the point charge, dipole and quadrupole moments on each site are contracted with the electric potential, field and quadrupole field operators, respectively, to give the electrostatic potential at an arbitrary point in the 3-space. Transforming Eq. (3) into the atomic orbital (AO) basis, we obtain

$$E_{\rm elec}^{\rm QM(el)/MM} = {\rm Tr}[\mathbf{P}\mathbf{V}_{\text{m-pole}}^{\rm MM}] \tag{5}$$

and its contribution to the Fock matrix of the coupled QM/AMOEBA system is

$$\mathbf{V}^{\text{MM}}_{\text{m-pole}} = \frac{\partial E^{\text{QM(el)/MM}}_{\text{elec}}}{\partial \mathbf{P}}, \tag{6}$$

where $\mathbf{V}^{\text{MM}}_{\text{m-pole}}$ is the representation of the 3-space ESP (given by Eq. (4)) in the AO basis $\{\omega_\mu(\mathbf{r})\}$:

$$(V^{\text{MM}}_{\text{m-pole}})_{\mu\nu} = \int \text{d}\mathbf{r}\ V^{\text{MM}}_{\text{m-pole}}(\mathbf{r})\omega^*_\mu(\mathbf{r})\omega_\nu(\mathbf{r}). \tag{7}$$

The last term in Eq. (1) denotes the polarization energy of the MM subsystem. (Note: the polarization of the QM system will be incorporated by the SCF minimization procedure implicitly.) In the AMOEBA force field, the polarization effect is described using point induced dipoles distributed onto each MM site. With an isotropic (scalar) polarizability α_i, the induced dipole on MM site i can be expressed as

$$\begin{aligned} \boldsymbol{\mu}^{\text{ind}}_i &= \alpha_i(\boldsymbol{\mathcal{E}}^{\text{perm}}_i + \boldsymbol{\mathcal{E}}^{\text{ind}}_i) \\ &= \alpha_i(\boldsymbol{\mathcal{E}}^{\text{perm}}_i + \sum_{j\neq i}^{n_{\text{MM}}} \tilde{\mathbf{T}}^{\text{d-d}}_{ij}\boldsymbol{\mu}_j) \end{aligned} \tag{8}$$

$\boldsymbol{\mathcal{E}}^{\text{perm}}_i$ is the so-called "permanent" electric field. For the QM/AMOEBA system, it includes the contributions from QM nuclei and electrons, and permanent multipoles on other MM sites:

$$\boldsymbol{\mathcal{E}}^{\text{perm}}_i = \boldsymbol{\mathcal{E}}^{\text{QM(nuc)}}_i + \boldsymbol{\mathcal{E}}^{\text{QM(el)}}_i + \boldsymbol{\mathcal{E}}^{\text{MM(m-pole)}}_i. \tag{9}$$

Note that in the case of AMOEBA water, which is the focus of our present work, the interactions between permanent multipoles on the same fragment are *excluded*. $\boldsymbol{\mathcal{E}}^{\text{ind}}_i$, on the other hand, refers to the electric field generated by induced dipoles on every other MM site ($\tilde{\mathbf{T}}^{\text{d-d}}_{ij}$ is the Thole-damped electrostatic tensor for dipole-dipole interactions), which requires Eq. (8) to be solved self-consistently. Once the induced dipoles are equilibrated, the

polarization energy of the MM subsystem can be computed as

$$E_{\text{pol}}^{\text{MM}} = -\sum_{i=1}^{n_{\text{MM}}} \boldsymbol{\mu}_i^{\text{ind}} \cdot \boldsymbol{\mathcal{E}}_i^{\text{perm}} + \frac{1}{2}\sum_{ij}^{n_{\text{MM}}} \boldsymbol{\mu}_i^{\text{ind}} \cdot \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j^{\text{ind}} \tag{10a}$$

$$= -\frac{1}{2}\sum_{i=1}^{n_{\text{MM}}} \boldsymbol{\mu}_i^{\text{ind}} \cdot \boldsymbol{\mathcal{E}}_i^{\text{perm}} \tag{10b}$$

where $\mathbf{T}_{ij} = \alpha_i^{-1}\boldsymbol{\delta}_{ij} - \tilde{\mathbf{T}}_{i\neq j}^{\text{d-d}}$. The second equality holds as a result of Eq. (8).

In our implementation, we first variationally solve for $\{\boldsymbol{\mu}_i^{\text{ind}}(n)\}$ with the current electron density matrix $\mathbf{P}^{(n)}$ before taking the n+1 SCF step, which guarantees the stationary condition $\partial E_{\text{pol}}^{\text{MM}}/\partial \boldsymbol{\mu}_i^{\text{ind}} = 0$. Note that although the MM polarization energy can be more conveniently evaluated via Eq. (10b), the correct stationary condition $\sum_j \mathbf{T}_{ij}\boldsymbol{\mu}_j^{\text{ind}} = \boldsymbol{\mathcal{E}}_i^{\text{perm}}$ can only be retrieved from Eq. (10a) by differentiating it with respect to $\{\boldsymbol{\mu}_i^{\text{ind}}\}$. Taking advantage of this stationary condition, the contribution from MM polarization to the Fock matrix can be evaluated by differentiating Eq. (10a) with respect to $\mathbf{P}$:

$$\mathbf{F}_{\text{pol}}^{\text{MM}} = \left.\frac{\partial E_{\text{pol}}^{\text{MM}}}{\partial \mathbf{P}}\right|_{\{\boldsymbol{\mu}_i^{\text{ind}}\}}$$

$$= -\sum_{i=1}^{n_{\text{MM}}} \boldsymbol{\mu}_i^{\text{ind}} \cdot \frac{\partial \boldsymbol{\mathcal{E}}_i^{\text{perm}}}{\partial \mathbf{P}} \tag{11}$$

Among the three components on the RHS of Eq. (9), $\boldsymbol{\mathcal{E}}_i^{\text{QM(el)}}$ is the only density-dependent component. At MM site i, we have

$$\boldsymbol{\mathcal{E}}_i^{\text{QM(el)}} = -\nabla_{\mathbf{R}_i}\int \mathrm{d}\mathbf{r}\ \frac{\rho_{\text{el}}(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_i|}$$

$$= -\operatorname{Tr}[\mathbf{P}\mathbf{V}^{\mathbf{R}_i}] \tag{12}$$

where $\mathbf{V}^{\mathbf{R}_i}$ is defined by

$$(V^{\mathbf{R}_i})_{\mu\nu} = \int \mathrm{d}\mathbf{r}\ \omega_\mu^*(\mathbf{r}) \nabla_{\mathbf{R}_i} \frac{1}{|\mathbf{r} - \mathbf{R}_i|} \omega_\nu(\mathbf{r}). \tag{13}$$

Combining Eqs. (11) and (12), we have

$$\mathbf{F}_{\mathrm{pol}}^{\mathrm{MM}} = \sum_{i=1}^{n_{\mathrm{MM}}} \boldsymbol{\mu}_i^{\mathrm{ind}} \cdot \mathbf{V}^{\mathbf{R_i}}, \tag{14}$$

i.e., this contribution to the Fock matrix is simply the ESP generated by the current set of induced dipoles located on all MM sites.

Taking Eqs. (2), (6), and (14) together, the Fock matrix for the coupled QM/AMOEBA system can be expressed as

$$\mathbf{F}^{\mathrm{QM/MM}} = \mathbf{F}_0 + \mathbf{V}_{\text{m-pole}}^{\mathrm{MM}} + \sum_{i=1}^{n_{\mathrm{MM}}} \boldsymbol{\mu}_i^{\mathrm{ind}} \cdot \mathbf{V}^{\mathbf{R_i}}. \tag{15}$$

If we add the latter two terms in Eq. (15) to the core Hamiltonian ($\mathbf{H}_{\mathrm{core}}$), the one-electron energy of the QM electrons in the SCF calculation can be conveniently evaluated using $\mathrm{Tr}[\mathbf{PH}_{\mathrm{core}}]$. However, according to Eq. (10b), this will overcount the MM polarization energy by $-1/2 \sum_i \boldsymbol{\mu}_i^{\mathrm{ind}} \cdot \boldsymbol{\mathcal{E}}_i^{\mathrm{QM(el)}}$, which thus needs to be subtracted out in the end. Therefore, the density-dependent part of the total energy for the QM/AMOEBA system (the last three terms in Eq. (1)) can be computed by

$$\begin{aligned}
&E_{\mathrm{el}}^{\mathrm{QM}} + E_{\mathrm{elec}}^{\mathrm{QM(el)/MM}} + E_{\mathrm{pol}}^{\mathrm{MM}} \\
&= \mathrm{Tr}[\mathbf{PH}_{\mathrm{core}}] + \frac{1}{2}\mathrm{Tr}[\mathbf{PIIP}] + E_{\mathrm{xc}} + \frac{1}{2}\sum_{i=1}^{n_{\mathrm{MM}}} \boldsymbol{\mu}_i^{\mathrm{ind}} \cdot \boldsymbol{\mathcal{E}}_i^{\mathrm{QM(el)}} \\
&- \frac{1}{2}\sum_{i=1}^{n_{\mathrm{MM}}} \boldsymbol{\mu}_i^{\mathrm{ind}} \cdot (\boldsymbol{\mathcal{E}}_i^{\mathrm{QM(nuc)}} + \boldsymbol{\mathcal{E}}_i^{\mathrm{MM(m\text{-}pole)}}),
\end{aligned} \tag{16}$$

where **II** represents the two-electron AO integrals that are used for the construction of Coulomb and exact exchange (if hybrid functionals are employed) matrices, and E_{xc} is the exchange-correlation (XC) energy of KS-DFT. We note that the SCF equations presented above for mutually polarizable QM/AMOEBA calculations are consistent with those reported by Loco et al.[69] in a previous paper and also similar to many other QM/MMpol studies (e.g., the early work by Thompson and Schenter[84]).

With equations for the total energy and the corresponding Fock matrix available, we adopt a double-loop SCF optimization scheme to minimize the total energy of the QM/AMOEBA system, i.e., E_{tot} is optimized with respect to both QM electron density (outer loop) and AMOEBA induced dipoles (inner loop). The entire procedure of this double-loop SCF calculation is as follows:

1. Compute the contribution from AMOEBA permanent multipoles to the core Hamiltonian (Eq. (7)), and the electric field matrix $\mathbf{V}^{\mathbf{R}_i}$ for each inducible MM site i (Eq. (13)).

2. Obtain the initial guess for the QM electron density matrix ($\mathbf{P}$).

3. With the given $\mathbf{P}$, iteratively solve for the induced dipoles in the MM region ($\{\boldsymbol{\mu}_i\}$) within the inner loop.

4. Evaluate the contribution from induced dipoles to the core Hamiltonian (Eq. (14)).

5. Build the Fock matrix for the current (outer loop) iteration according to Eq. (15), and evaluate the SCF energy (Eq. (16)) and the error vector/gradient.

6. Check for convergence:

   - If NOT converged, update the MO coefficients and the density matrix with the employed SCF algorithm, and go back to Step 3.
   - If converged, compute the other energy components in Eq. (1) that are density-independent (e.g. the vdW interaction between QM and MM regions).

2.2 Energy decomposition analysis

As the first assessment of this mutually polarizable and fully self-consistent QM/AMOEBA model, we apply it to the evaluation of solute-solvent interactions (the solute is described by QM and the solvent molecules by AMOEBA) and compare the results to full QM reference calculations. In our previous study (using the ONETEP/TINKER implementation of QM/AMOEBA),[70] it has been revealed that the unmodified QM/AMOEBA model underestimates solute-solvent interaction energies almost consistently across a range of systems, and further softening the buffered 14-7 potential for $E_{\text{vdw}}^{\text{QM/MM}}$ significantly improves the agreement with the full QM reference for most of the investigated systems. Nonetheless, the primary reason for the poor performance of the unmodified QM/AMOEBA model is not completely clear without further analysis. Here, similar to a recent benchmark of ion-water interactions that we performed for the AMOEBA force field (pure MM) against the ALMO-EDA results,[114] we decompose the solute-solvent interactions evaluated by both QM/AMOEBA and full QM into contributions from permanent electrostatics (ELEC), polarization (POL), and vdW interaction (vdW), and the agreement in total interaction energy and in each energy component will be assessed.

Since our QM/AMOEBA model is mutually polarizable, the interaction energy between QM and MM regions ($E_{\text{int}}^{\text{QM/MM}}$) cannot be simply represented by terms in Eq. (1). Instead, it can always be computed through the so-called supermolecular approach, i.e.,

$$E_{\text{int}}^{\text{QM/MM}} = E_{\text{tot}}^{\text{QM/MM}} - E_0^{\text{QM}} - E_0^{\text{MM}}, \tag{17}$$

where E_0^{QM} and E_0^{MM} refer to the total energies of isolated QM and MM subsystems, respectively.

With the goal of extracting the energy components of this interaction across the QM/MM boundary, we propose the following approach which has a similar spirit to that of the ALMO-

EDA. First, we perform SCF calculations for isolated QM and MM subsystems, and the resulting QM density matrix ($\mathbf{P}_0$) and MM induced dipoles ($\{\boldsymbol{\mu}_i^{\text{ind}}(0)\}$) are collected and employed as the initial guess for the coupled QM/AMOEBA calculation. Note that the MM induced dipoles and $E_{\text{pol}}^{\text{MM}}$ are both non-zero at this stage unless there is only one single polarization group (equivalent to one single fragment for the AMOEBA water case) in the MM region, whereas the polarization effect arises from the existence of MM multipoles *only*. Then the QM and MM subsystems are coupled together, and before any further relaxation occurs, the permanent electrostatic interaction between QM and AMOEBA fragments can be computed as

$$E_{\text{elec}}^{\text{QM/MM}} = \text{Tr}[\mathbf{P}_0\tilde{\mathbf{V}}_{\text{m-pole}}^{\text{MM}}] + \tilde{E}_{\text{elec}}^{\text{QM(nuc)/MM}}, \tag{18}$$

where the tildes indicate that interaction between the initial set of induced dipoles ($\{\boldsymbol{\mu}_i^{\text{ind}}(0)\}$) and the QM region (electron density and nuclei) is also included. At this point, the induced dipoles are computed *without* coupling with the QM subsystem, so they play the same role as permanent dipoles and can be treated in the same way computationally, i.e., $\tilde{\mathbf{V}}_{\text{m-pole}}^{\text{MM}}$ can be computed based on Eqs. (4) and (7) by simply replacing $\boldsymbol{\mu}_i$ with $\boldsymbol{\mu}_i' = \boldsymbol{\mu}_i + \boldsymbol{\mu}_i^{\text{ind}}(0)$. This definition of permanent electrostatics in mutually polarizable QM/MM calculations, especially the use of $\{\boldsymbol{\mu}_i^{\text{ind}}(0)\}$, is the same as the scheme proposed in Ref. 108.

We then relax the QM/AMOEBA system following the SCF procedure presented above, and the resulting changes in the three terms on the RHS of Eq. (16) define the polarization energy of the whole QM/MM system:

$$E_{\text{pol}}^{\text{QM/MM}} = \Delta E_{\text{el}}^{\text{QM}} + \Delta E_{\text{elec}}^{\text{QM(el)/MM}} + \Delta E_{\text{pol}}^{\text{MM}}. \tag{19}$$

Note that due to the mutual character of polarization here, it is *not* possible to further decompose $E_{\text{pol}}^{\text{QM/MM}}$ into forward and backward contributions while retaining additivity. Once the SCF is converged, the vdW interaction between QM and MM subsystems, $E_{\text{vdw}}^{\text{QM/MM}}$, can

be evaluated at the end since it is fully classical in our current model. Taken together, the interaction across the QM/MM boundary is decomposed into three energy components:

$$E_{\text{int}}^{\text{QM/MM}} = E_{\text{elec}}^{\text{QM/MM}} + E_{\text{pol}}^{\text{QM/MM}} + E_{\text{vdw}}^{\text{QM/MM}}. \tag{20}$$

For the purpose of the benchmark, solute-solvent interaction energies evaluated by full QM also need to be decomposed into the corresponding energy components. This can be achieved by applying a slightly modified version of the ALMO-EDA, which has been introduced in Ref. 114. To briefly recapitulate the modifications, we adopt the "classical" definition for permanent electrostatics, i.e., the classical Coulomb interaction between charge distributions of fragments evaluated in isolation:

$$E_{\text{elec}} = \sum_{A<B} \int_{\mathbf{r}_1} \int_{\mathbf{r}_2} \rho_A^{\text{tot}}(\mathbf{r}_1) \frac{1}{\mathbf{r}_{12}} \rho_B^{\text{tot}}(\mathbf{r}_2) \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2, \tag{21}$$

where $\rho_A^{\text{tot}}(\mathbf{r}) = \rho_A^{\text{el}}(\mathbf{r}) + \rho_A^{\text{nuc}}(\mathbf{r})$. Here ρ_A and ρ_B simply refer to charge distributions of solute and solvent molecules, respectively. Then, with the definition for the dispersion energy (E_{disp}) remaining unchanged (based on Ref. 111), the remainder of the frozen interaction energy (interaction between fragments approaching each other whose molecular orbitals are *not yet* relaxed) is defined as the (modified) Pauli term:

$$E_{\text{pauli}} = E_{\text{frz}} - E_{\text{elec}} - E_{\text{disp}}. \tag{22}$$

The separation of polarization and charge transfer (CT) still utilizes the so-called "SCF for molecular interaction" (SCF-MI) approach. Basis set superposition error (BSSE) evaluated by the counterpoise correction can also be included if desired, which is often combined with the CT term because they both arise from the delocalization effect. Thus the solute-solvent

interaction energy evaluated by full QM can be decomposed as follows:

$$E_{\text{int}} = E_{\text{elec}} + E_{\text{pauli}} + E_{\text{disp}} + E_{\text{pol}} + E_{\text{ct}}. \tag{23}$$

The readers are referred to the original references of ALMO-EDA (Refs. 110–112) for more details about this method.

In order to further simplify the comparison between energy components evaluated by QM/AMOEBA and full QM, in this work we introduce another "coarse-grained" modification to the ALMO-EDA scheme presented above, i.e., we define the "vdW" component of the full QM interaction energy as the sum of Pauli repulsion, dispersion and charge transfer:

$$E_{\text{vdw}} = E_{\text{pauli}} + E_{\text{disp}} + E_{\text{ct}}, \tag{24}$$

and then Eq. (23) becomes

$$E_{\text{int}} = E_{\text{elec}} + E_{\text{pol}} + E_{\text{vdw}}, \tag{25}$$

which has a straightforward correspondence to the decomposition of QM/AMOEBA interaction energy represented by Eq. (20). This choice, especially the addition of CT to the other two energy components that correspond to the vdW interaction physically, is rationalized by our previous investigation of vdW interactions in AMOEBA with the same EDA scheme.[114]

2.3 Implementation through the Q-Chem/LibEFP interface

The QM/AMOEBA model presented above and the energy decomposition analysis scheme for interactions across the QM/MM boundary are implemented through a Q-Chem/LibEFP code structure. LibEFP[103,104] is an open source library for effective fragment potential (EFP)[62,63,115] calculations and has been interfaced with the Q-Chem 4 software package.[102] It can be naturally extended to support the AMOEBA force field thanks to their similarities

in many aspects, e.g., they both use distributed point multipoles to model permanent electrostatics and use distributed inducible dipoles to describe polarization. Therefore, many routines implemented for EFP calculations, such as those that are utilized to compute the interactions between point multipoles, can be directly used for AMOEBA. Nonetheless, several additional functionalities are still required in order to fully support AMOEBA in LibEFP, including

- Parsers for AMOEBA parameter files which contain permanent multipole moments, atomic polarizabilities, vdW parameters, etc.;
- The valence terms in the AMOEBA force field (bond, angle and Urey-Bradley), which were not included in LibEFP since fragments with fixed geometries are usually used in EFP calculations;
- vdW interactions between AMOEBA fragments and between QM and MM atoms, which are both described by the buffered 14-7 potential;
- Routines transforming AMOEBA's permanent multipoles that are in their own local coordinates (as in the parameter file) into the global coordinate, which is a necessary step before including these permanent multipoles into the QM/MM system;
- Thole-damped electric field arising from monopole, dipole, and quadrupole moments, which is required by the evaluation of induced dipoles based on Eq. (8).

Calculations using the AMOEBA force field can be correctly handled when these additions are combined with the original routines in LibEFP. We note that currently we only enabled the use of the AMOEBA water model in LibEFP, although a future extension to more general MM systems described by AMOEBA should be straightforward.

The Q-Chem/LibEFP code interface for mutually polarizable QM/AMOEBA calculations is illustrated in Figure 1. Q-Chem serves as the driver for the entire calculation, which

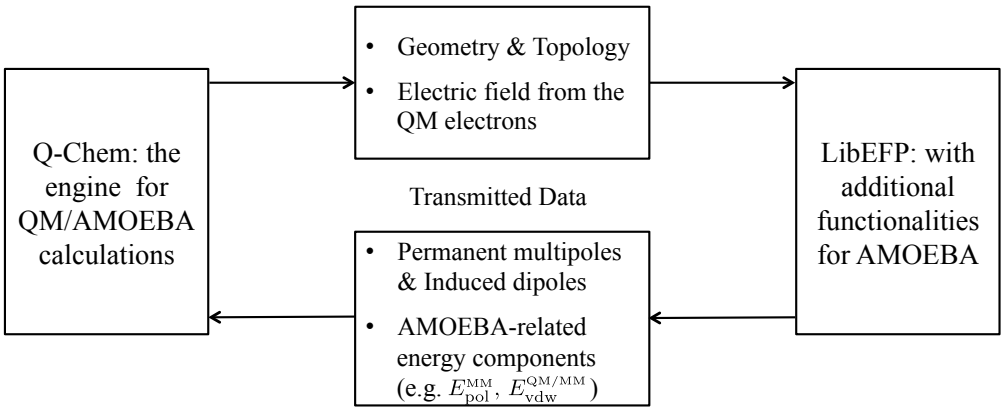

Figure 1: The Q-Chem/LibEFP code interface for mutually polarizable, fully self-consistent QM/AMOEBA calculations.

parses the geometry and topology (connectivity between the MM atoms) and passes the information to LibEFP. LibEFP computes MM energy terms (such as $E_{\rm val}^{\rm MM}$, $E_{\rm elec}^{\rm MM}$) and components of QM/MM interaction energy that are not density-dependent (such as $E_{\rm vdw}^{\rm QM/MM}$ and $E_{\rm elec}^{\rm QM(nuc)/MM}$), and it also passes AMOEBA's permanent multipoles (all transformed into the global coordinate frame) and current set of induced dipoles to Q-Chem. Since routines that evaluate electric potential, field, and field derivative matrices in the AO basis are already available in Q-Chem, only slight modifications to the standard SCF routines are needed to incorporate the MM contributions to the Fock matrix. And standard SCF algorithms, such as the direction inversion of the iterative subspace (DIIS) method,[116,117] can still be employed for minimizing the energy of a coupled DFT/AMOEBA system. In terms of the MM polarization energy ($E_{\rm pol}^{\rm MM}$), the induced dipoles of AMOEBA are self-consistently solved in LibEFP, while the electric field component arising from QM electron density on each inducible site is evaluated by and then passed from Q-Chem.

3 Results

3.1 Computational details

All the QM/AMOEBA and full QM calculations are performed with a locally modified Q-Chem 4.4 software package,[102] which has been interfaced with a locally developed version of LibEFP.[103,104] In this work, "QM" refers to DFT methods *exclusively*, and unless otherwise specified, the ωB97X-V functional[118] is used. ωB97X-V is a range-separated hybrid GGA which employs the VV10 non-local correlation (NLC) functional[119] to describe dispersion, and its accuracy for non-covalent interactions, especially for those involving ionic species, has been shown by several recent studies.[114,120,121] Two Karlsruhe basis sets, def2-SVPD and def2-TZVPPD,[122] are employed in the calculations presented below, which are of augmented double- and triple-ζ quality, respectively.

The full QM and QM/AMOEBA solute-solvent interaction energies are both evaluated with the supermolecular approach, and counterpoise corrections for BSSE are applied to the former. The ALMO-EDA calculations are performed based on the modified scheme introduced in Sec. 2.2, while in this work, the separation of POL and CT is achieved by using the original AO-block based ALMO scheme[110,123] instead of the recently proposed fragment electrical response function (FERF) model.[124] The FERF model gives a well-defined basis set limit to the resulting polarization energy but is computationally more expensive. This choice is made because smaller basis sets that are far from the complete basis set (CBS) limit are used in this work, and it is known that these two models will *not* yield remarkably different results unless very large basis sets (usually beyond augmented triple-ζ) are used.[124] As a validation, we compare the polarization energies obtained by using the ALMO and FERF models with the def2-TZVPPD basis for several systems that are investigated below (the results are shown in Sec. S1 of the Supporting Information), and the differences between them are shown to be insignificant.

The AMOEBA 03 water model[105] is utilized for the solvent molecules in all QM/MM calculations, while the vdW parameters for the solute atoms are taken from the most recently released parameter files of AMOEBA ("amoebapro13" or "amoeba09") in the TINKER 7 molecular modeling package.[125] The induced dipoles are converged to 10^{-10} a.u. in each inner loop, while the convergence criterion for the outer loop SCF is 10^{-8} a.u. We note that for each inducible site, while the polarizing effect of permanent multipoles and other induced dipoles are evaluated with Thole damping, the electric field arising from QM nuclei and electrons are computed *without* applying any damping schemes. Furthermore, the polarization of QM electron density due to the embedding potential of MM is also evaluated *without* any damping. Therefore, the QM/AMOEBA model employed here stands for a preliminary, unmodified coupling between the QM and MM components, and the necessity of applying damping functions for forward or backward polarizations will be examined via EDA calculations.

3.2 The water dimer and the water-Cl^- complex

We start by investigating the performance of QM/AMOEBA on the prototypical system for hydrogen-bonding interaction — the water dimer, where each water molecule can be treated with either QM or AMOEBA. Its dissociation potential energy surfaces (PES) evaluated with full QM, full AMOEBA, and QM/AMOEBA are shown in the top left panel of Figure 2, as functions of the O··O distance between two water molecules. The equilibrium geometry is optimized at the ωB97X-V/def2-QZVPPD level of theory, and the PES scan is performed by modifying the O··O distance only without relaxing other coordinates (the PES scan for the water-Cl^- system below is performed in the same manner). The first thing to note is that the choice of basis set makes a minimal difference to the interaction energies (and their components) computed by both full QM and QM/AMOEBA, which is shown in Figure S3 in the Supporting Information. Therefore, we focus on the results evaluated with the def2-

TZVPPD basis in the following discussion of this system. In the scenario where the H-donor water molecule (water 1) is described by QM and the H-acceptor (water 2) by AMOEBA, the PES yielded by QM/AMOEBA (green curve) is slightly underbound at all ranges measured against the full QM reference, and the difference at the common minimum-energy distance (2.90 Å) is 0.3 kcal/mol. Around the minimum, the full AMOEBA PES (red dashed) shows better agreement with the full QM reference, while in the short range where pure AMOEBA considerably underbinds the water dimer, QM/AMOEBA deviates from the full QM PES less.

The character of the QM/AMOEBA PES can be elucidated with the EDA results. According to the top right panel of Figure 2, the permanent electrostatic interaction between water 1 (QM) and 2 (AMOEBA) is marginally less attractive compared to that in the full AMOEBA case. As we have shown in Ref. 114, the charge penetration error (CPE), which can be interpreted as the difference between AMOEBA's multipole-based permanent electrostatics and the full QM value, is significant for the water dimer at the equilibrium distance, and the good agreement in the total energy around equilibrium relies on the cancellation of errors through the softened vdW potential in AMOEBA, which is shown in the bottom right panel of Figure 2. This argument applies to the QM/AMOEBA case as well since it has a similar profile for permanent electrostatics, and the special characteristics of QM/AMOEBA's total PES can be attributed to the polarization component. At the ranges around and beyond equilibrium, the polarization energies given by pure AMOEBA and ALMO-EDA agree almost perfectly, whereas that given by QM/AMOEBA is slightly underestimated, which appears to be the primary reason for the underbinding of QM/AMOEBA in this region. However, in the short range the polarization energy of full AMOEBA becomes unphysical due to the onset of Thole damping and that results in the overly repulsive total interaction energy, while the profile given by QM/AMOEBA shows a more proper behavior despite still being insufficiently favorable, which contributes to its better agreement with full QM in this

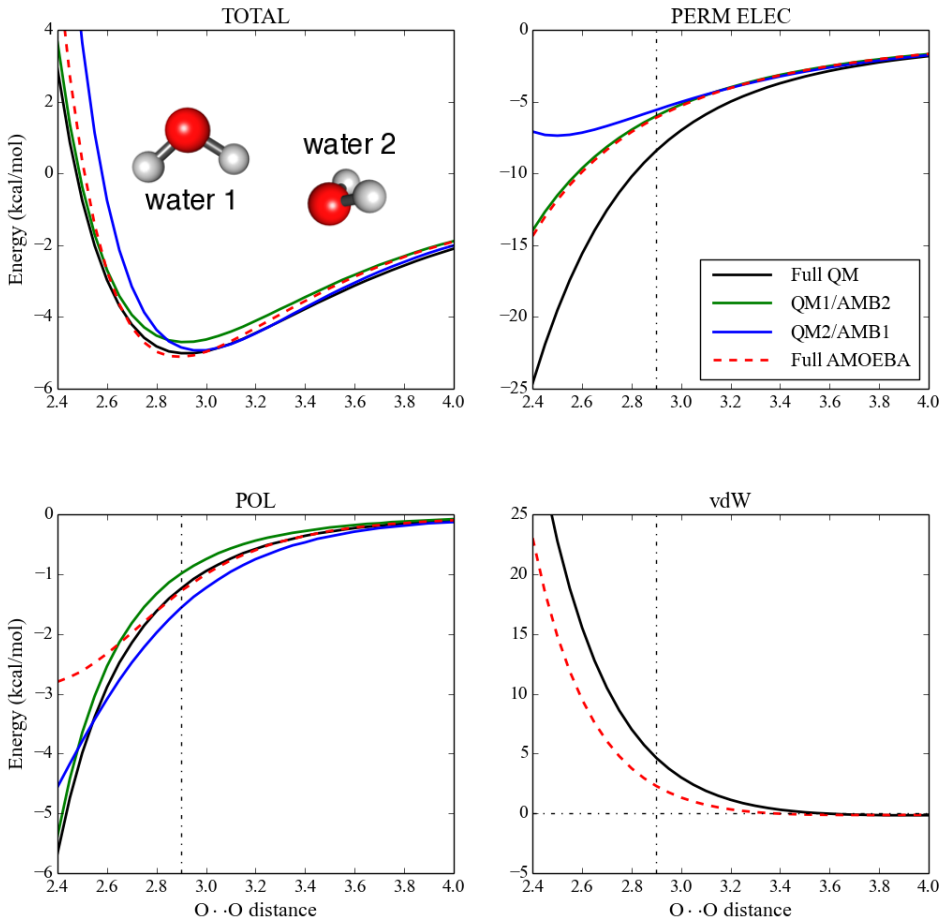

Figure 2: Total interaction energies and their components (in kcal/mol) evaluated with full QM, full AMOEBA, and QM/AMOEBA for the water dimer as functions of the O··O distance (the interval between two neighboring data points is 0.05 Å). The def2-TZVPPD basis set is employed for the QM and QM/AMOEBA calculations. Water 1 and 2 refer to the proton donor and acceptor, respectively, whose geometries are unrelaxed during the PES scan. "QM1/AMB2" refers to the scenario where water 1 is described by QM and water 2 by AMOEBA, and "QM2/AMB1" is defined in the same way. The vdW interactions in QM/AMOEBA calculations are identical to those computed by pure AMOEBA so they are not plotted in the bottom right panel.

region in terms of total interaction energy.

It is intriguing to examine the performance of QM/AMOEBA after switching the QM and MM regions, i.e., we describe the H-acceptor (water 2) with QM and the H-donor (water 1) with AMOEBA instead. This, referred to as the "reversed" QM/AMOEBA model for the water dimer in the discussions below, yields a rather skewed total potential energy profile (blue curve in Figure 2), where the equilibrium O··O distance is shifted to 2.95–3 Å and the repulsive wall becomes even harder than the pure AMOEBA case. Turning to the energy components, the first striking result is that the permanent electrostatics deviates from the values given by pure AMOEBA and the former QM/AMOEBA model enormously after entering the compressed region (r(O··O) $<$ 2.9 Å). At r(O··O) = 2.5 Å, it differs from pure AMOEBA's permanent electrostatic interaction by over 4 kcal/mol, and below that distance it becomes less attractive with decreasing intermolecular distances, which is physically incorrect and further enlarges the gap. Recalling that the vdW potential employed here is parameterized with the CPE in the pure AMOEBA case taken into account, it will thus be insufficient to address an even larger CPE here, and this is the primary reason for the excessively hard repulsive wall given by the "reversed" QM/AMOEBA model. Its polarization energy, contrary to that given by the former QM/AMOEBA model, is more favorable than the full QM reference except at the most compressed distances (r(O··O) $\leq$ 2.5 Å). Nevertheless, the difference in polarization energy has a small magnitude so that it barely affects the error dominated by CPE.

In order to see if this represents the typical situation when AMOEBA water serves as the proton donor in H-bonding systems, we also investigate the performance of QM/AMOEBA for the rigid dissociation PES of the water-Cl^- complex (the chloride anion is treated with QM), which is closely related to the "reversed" water dimer case discussed above. Its total energy and EDA results are shown in Figure 3. Note that although the choice of basis set has a slightly larger effect on this system (which can be seen from the "POL" panel

of Figure S4 in the Supporting Information), it brings no qualitative changes so that we can still focus on the def2-TZVPPD results. Indeed, when the (QM) H-acceptor water is replaced by Cl^- which interacts with the AMOEBA water more strongly, the issues revealed by the investigation of the water dimer above are further exacerbated, which gives rise to an enormously shifted PES. In contrast to the pure AMOEBA case, QM/AMOEBA underbinds the $H_2O{\cdot\cdot}Cl^-$ complex by 3.6 kcal/mol at the QM minimum (3.10 Å), and the error further increases in the compressed region. Meanwhile, the equilibrium Cl··O distance given by QM/AMOEBA (3.35 Å) is also substantially elongated compared to the full QM result.

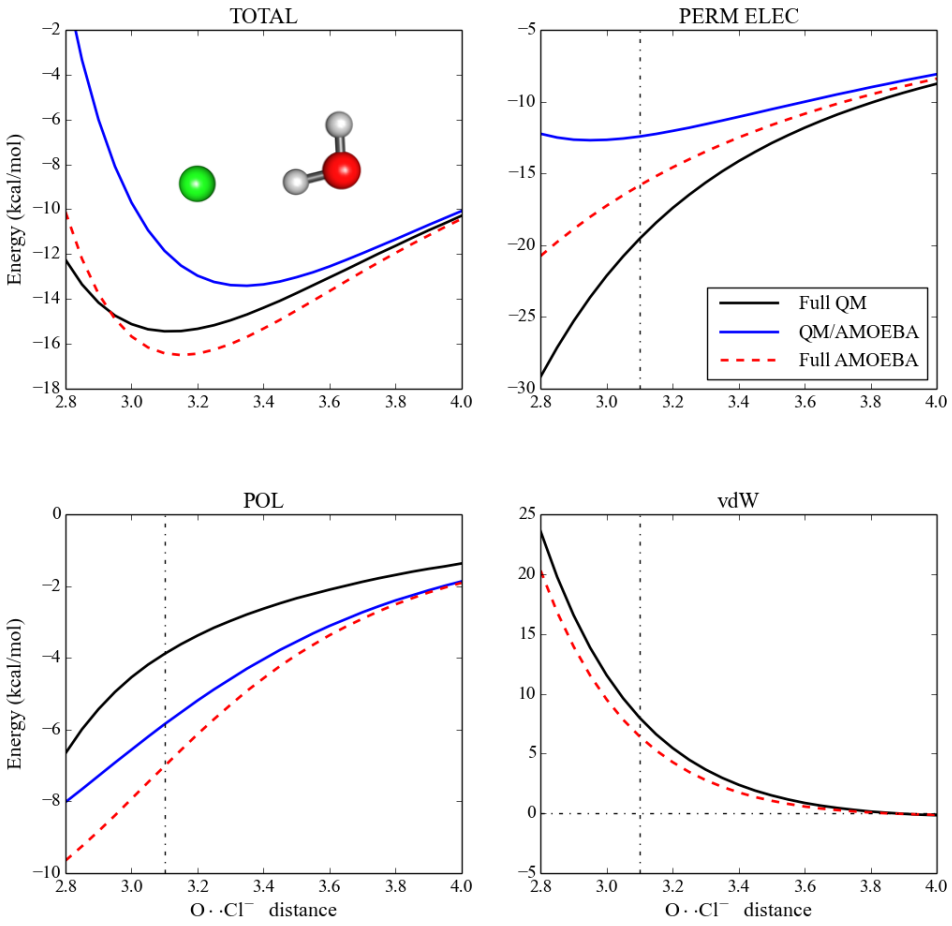

Figure 3: Total interaction energies and their components (in kcal/mol) for the water-Cl^- complex as functions of the O··Cl^- distance. In QM/AMOEBA calculations, Cl^- is treated with QM and the H_2O molecule is described by AMOEBA. The other computational and plotting details are the same as in Figure 2.

Turning to the energy components, the permanent electrostatics of QM/AMOEBA shows a similar but more pronounced feature as in the "reversed" water dimer case, as it starts to deviate from the pure AMOEBA curve considerably at an even longer distance, and the difference at the QM equilibrium distance is already as large as 7 kcal/mol. This is evidently the main culprit for the significant underbinding of QM/AMOEBA at all ranges. Although the polarization energy and vdW interaction given by QM/AMOEBA are both more favorable than their full QM counterparts, they are far from being sufficient to cancel the enormous error in permanent electrostatics. It should be noted that the undamped QM/AMOEBA polarization energy lies in between the full QM and full AMOEBA results across the board, i.e., it still overestimates the polarization energy relative to the full QM reference but less severely than pure AMOEBA, even though Thole damping is consistently applied to the latter.

Taking these two examples (the "reversed" water dimer and the water-Cl^- complex) together, it is clearly revealed that QM/AMOEBA can suffer more from CPE than pure AMOEBA, and that the resulting profile of permanent electrostatics, instead of lying in between the full QM and full AMOEBA ones, can be *less attractive than both*. This is slightly counterintuitive and will be further discussed in Sec. 3.5.

3.3 Interaction with solvent molecules in the first solvation shell

As the first step from gas-phase dimers to clusters in condensed phase, we turn to the interaction between solutes and solvent (water) molecules in their first solvation shells. Three solutes are considered in this study: H_2O, Na^+, and Cl^-, which are representative of neutral, cationic, and anionic species, respectively. The configurations are taken from MD simulations of one solute molecule solvated in a box of 215 H_2O molecules using the AMOEBA force field, and the details regarding the equilibration and production steps of the MD simulation are the same as in the previous work by Grossfield et al.[126] For each species, the number of water

molecules in the first solvation shell (coordination number) is determined by an integration over the first peak of the resulting radial distribution function (RDF) until the position of the first minimum, and the average (closest integer value) turns out to be 4 for H_2O, 6 for Na^+, and 8 for Cl^-. The resulting coordination numbers for Na^+ and Cl^- are consistent with the values reported in Ref. 127 (note: the coordination numbers reported in Ref. 126 were not calculated from the RDFs), while they might differ from those obtained from AIMD simulations performed by others.[32,37,38] Nonetheless, as our focus here is to benchmark solute-solvent interaction energies rather than the solvent structure, the configurations prepared by classical MD should still be able to fulfill this purpose. To simplify the discussions below, for each solute species, the same number of nearest water molecules (the average integer coordination numbers above) is applied to all its configurations. Therefore, these clusters are also referred to as $(H_2O)_5$ (water pentamer), $Na^+(H_2O)_6$ and $Cl^-(H_2O)_8$ in the discussions below.

Table 1: Maximum errors (MAX), mean absolute errors (MAE), and mean signed errors (MSE) of QM/AMOEBA (in kcal/mol) for interactions between three solutes (H_2O, Na^+, and Cl^-) and the H_2O molecules in their first solvation shells. The statistical errors in total energies and energy components are evaluated relative to the full QM references over 100 samples (snapshots) for each solute.

		MAX	MAE	MSE
H_2O	INT	5.21	1.35	1.29
	ELEC	21.46	11.79	11.79
	POL	1.79	0.37	0.36
	vdW	-21.40	10.86	-10.86
Na^+	INT	3.78	1.45	1.44
	ELEC	6.55	3.79	3.79
	POL	1.89	0.65	0.25
	vdW	-4.99	2.60	-2.60
Cl^-	INT	19.51	10.94	10.94
	ELEC	37.32	27.17	27.17
	POL	-5.94	4.91	-4.91
	vdW	-18.25	11.33	-11.33

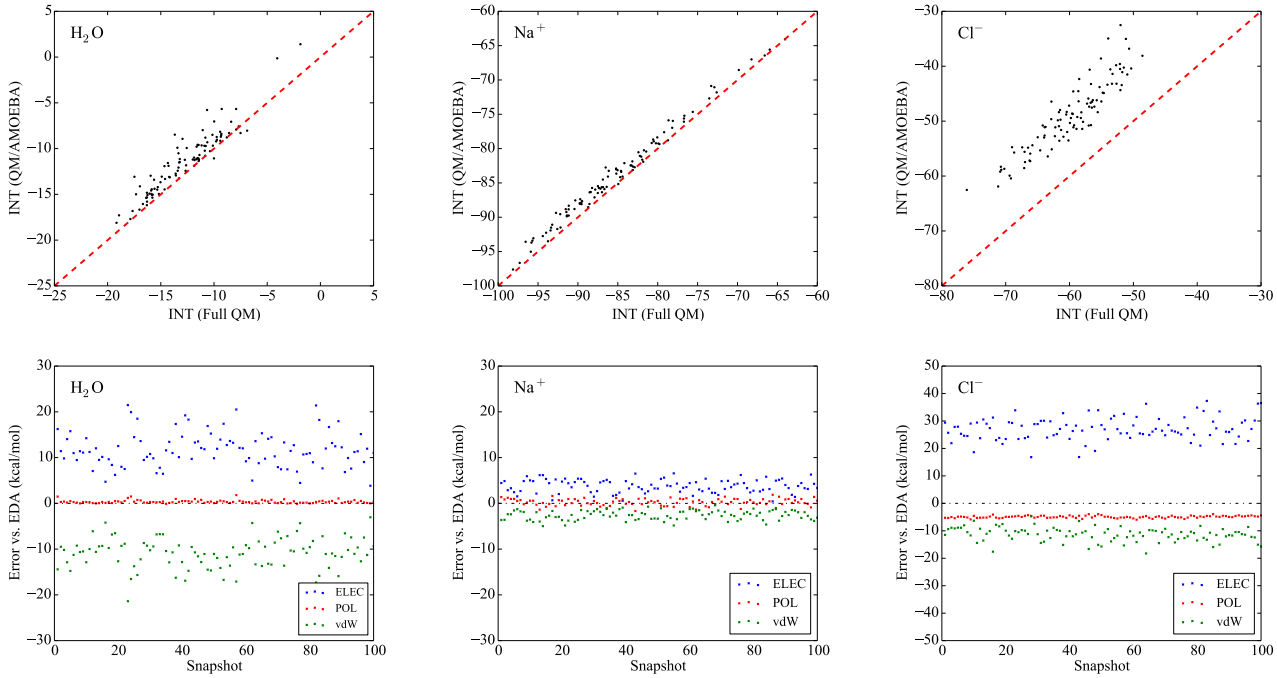

Figure 4: First row: interaction energies (in kcal/mol) between three solutes (H_2O, Na^+, Cl^-) and water molecules in their first solvation shells evaluated with full QM (x) and QM/AMOEBA (y); second row: errors (in kcal/mol) for the three components of QM/AMOEBA interaction energies (ELEC, POL, vdW) measured against the reference values given by ALMO-EDA. 100 snapshots are calculated and then plotted for each solute-solvent system.

The solute-solvent interaction energies evaluated with QM/AMOEBA are plotted with respect to the full QM results in the top row of Figure 4. For the H_2O and Na^+ solutes, the QM/AMOEBA interaction energies agree with the full QM results reasonably but are slightly underbound, where the mean signed errors (MSEs) are +1.29 kcal/mol and +1.44 kcal/mol, respectively (the statistical errors for each solute are shown in Table 1). Note that the "solute" water molecule can play the role of H-donor and H-acceptor simultaneously, and according to our previous results for the water dimer (Figure 2), QM/AMOEBA underestimates the binding energy near equilibrium in either case, which seems to be the case here where a QM water molecule is interacting with four AMOEBA waters in the vicinity. On the other hand, the agreement between QM/AMOEBA and full QM for Cl^- is poor, where we see a substantial underbinding by QM/AMOEBA (the MSE over 100 snapshots is +10.94

kcal/mol). This result is also consistent with what we have found for the dissociation PES of the water-Cl^- complex, where QM/AMOEBA underestimates the interaction energy more significantly than in the water dimer case at their individual equilibrium distances.

The components of QM/AMOEBA interaction energies are compared against their full QM counterparts as well, and the errors are shown in the bottom row of Figure 4. Distinct patterns emerge for each of these species. For the cationic solute (Na^+), the mutual polarization of QM/AMOEBA (which is dominated by solvent polarization in this case) matches that given by ALMO-EDA closely, where the mean absolute error (MAE) is only 0.65 kcal/mol. The errors in permanent electrostatics and vdW interaction are exclusively above and below zero, respectively, which largely cancel each other but not completely, yielding the slightly underbound results compared to full QM. Turning to the neutral solute H_2O, while the small deviation from ALMO-EDA's polarization energy is retained, the errors in permanent electrostatics and vdW become considerably larger (see the statistics in Table 1). The errors also span much larger ranges than in the former Na^+ case, which can be seen from the contrast between the first two bottom panels in Figure 4. Nevertheless, thanks to the rather similar magnitude of the positive and negative errors, the resulting total interaction energies are brought to a reasonable level. This, however, does not hold for the anionic solute Cl^-, and the features demonstrated by EDA are in line with our expectations based on the results for the water-Cl^- complex. The major source of error is clearly the permanent electrostatics, which on average is less attractive than its full QM counterpart by 27.17 kcal/mol, and the vdW potential of AMOEBA is evidently not sufficiently softened to compensate for such a large error. Also, as opposed to the other two solutes, the polarization energies given by QM/AMOEBA for $Cl^-(H_2O)_8$ clusters are systematically overestimated (on average it is more favorable by 4.9 kcal/mol than the full QM references), which is also consistent with the trend demonstrated in Figure 3.

The results above indicate that the performance of this QM/AMOEBA model in de-

scribing the interaction between solute and solvent molecules in its vicinity strongly depends on the balance of the QM/MM interface, i.e., the error cancellation between individual components of the interaction crossing the boundary of QM and MM regions. Since the error in polarization energy usually has the smallest magnitude, the balance between inadequately attractive electrostatic interaction and softened, fully empirical vdW potential has a decisive influence on the accuracy of solute-solvent interaction energies computed with QM/AMOEBA.

3.4 Two case studies: $F^-(H_2O)_{10}$ and $CH_4(H_2O)_{20}$ clusters

In order to verify the generality of our findings above, we investigate two other solute-water clusters, $F^-(H_2O)_{10}$ (including ten isomers) and $CH_4(H_2O)_{20}$ (methane in a dodecahedral cage formed by water molecules), whose geometries are taken from Ref. 128 and 129, respectively. These systems have been previously studied by Lao et al. with a broad range of quantum chemistry methods.[121] It should be noted that all ten water molecules in $F^-(H_2O)_{10}$ are treated as one single monomer here so that the binding energies reported below will not be comparable to those in Ref. 121, where each H_2O was treated as an individual monomer.

The total interaction and energy decomposition results for 10 isomers of $F^-(H_2O)_{10}$ are shown in Figure 5 (the original data are provided in Table S3 in the Supporting Information), and the manifested features are similar to those of the $Cl^-(H_2O)_8$ systems studied above. The plot for the total interaction energies given by QM/AMOEBA is almost parallel with that for the full QM values, and according to the original data, QM/AMOEBA reproduces the correct ordering of these isomers except that it computes the binding energy for Isomer 8 too low relative to Isomer 1 and 2. However, there exists an over 20 kcal/mol gap between the absolute binding energies given by QM/AMOEBA and full QM, due to the highly unbalanced QM/MM interface. Figure 5 shows the tremendous discrepancy between their permanent electrostatics, which is about 60 kcal/mol, whereas the vdW interactions given

by AMOEBA's 14-7 potential are only about 30 kcal/mol below the full QM values. Also, like other H-bonding systems where AMOEBA water donates its proton to the QM region, the polarization energies for these isomers are overestimated by about 7 kcal/mol, which is slightly larger than the MAE for the $Cl^-(H_2O)_8$ clusters.

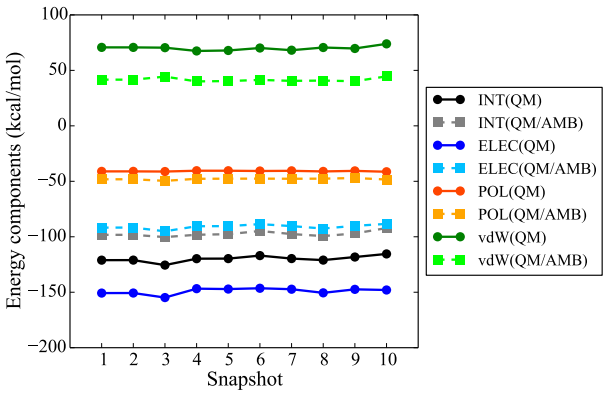

Figure 5: Total interaction energies and energy components (in kcal/mol) computed with full QM and QM/AMOEBA for F^- interacting with 10 surrounding water molecules in 10 isomers of the $F^-(H_2O)_{10}$ cluster.

Methane, on the other hand, is a non-polar solute so that the contributions from permanent electrostatics and polarization to the interaction of CH_4 with the dodecahedral water cage are expected to be weaker. The total interaction energy given by QM/AMOEBA is -5.09 kcal/mol, which is about 0.8 kcal/mol above the full QM result. Nevertheless, according to Table 2, the components of full QM and QM/AMOEBA interactions manifest sharp discrepancies. With full QM, permanent electrostatics and polarization make non-trivial contributions to binding, especially the former which accounts for roughly 50% of the stabilization, while in the QM/AMOEBA case, their contributions appear to be trivial and the whole system is almost entirely bound by QM/MM vdW interaction which is described at the AMOEBA level. We note that contrary to the former $F^-(H_2O)_{10}$ system, the AMOEBA waters are no longer proton donors here and the resulting QM/AMOEBA polarization energy is largely underestimated. It is thus evident that the electrostatic interaction (permanent

and induced) between a neutral, non-polar solute and solvent water molecules cannot be correctly described by the present QM/AMOEBA model which has no explicit treatment for the charge penetration effect, and the quality of the resulting total interaction energy is entirely controlled by the empirical 14-7 potential.

Table 2: Total interaction energies and their components (in kcal/mol) for methane interacting with a dodecahedral water cage ($CH_4(H_2O)_{20}$) evaluated with full QM and QM/AMOEBA.

	INT	ELEC	POL	VDW
Full QM	-5.90	-2.90	-0.56	-2.44
QM/AMOEBA	-5.09	-0.08	-0.03	-4.98

3.5 More detailed investigations on electrostatics and polarization in H-bonding systems

According to the examples investigated above, we have noticed that the most challenging systems for QM/AMOEBA are H-bonding complexes where AMOEBA water serves as the proton donor. For these systems, one of the most notable features of QM/AMOEBA is that the resulting permanent electrostatics is even less favorable than that given by pure AMOEBA. In order to understand this better, we perform a further analysis on the electrostatic interaction between Cl^- and H_2O at the equilibrium geometry of the water-Cl^- complex, where the contributions from nuclei and electrons are separated. The results are shown in Table 3. Note that for atomic site i in the MM region, the "electron" part includes a point charge of value $q_i - N_i$ (q_i is the original AMOEBA monopole for site i and N_i is the corresponding nuclear charge) and all the higher-order multipoles. Such a definition ensures that the "nuclei" part in QM and MM regions are treated equivalently. According to Table 3, when the QM water is replaced by AMOEBA, the interactions between the electron part of H_2O with the nucleus ($+17e$) and electron density of Cl^- both become less attractive, and the substantial deviation of QM/AMOEBA's permanent electrostatics relative to the

full QM reference primarily arises from the overly unfavorable electron-electron component. Turning to the pure AMOEBA case, where the electron density of Cl^- collapses onto the same position as its nucleus and reduces to a point charge (-18e), the electron-electron part becomes even more repulsive. However, the attractive interaction between the electrons of Cl^- and the nuclei of H_2O is also enhanced upon this change, which compensates for the overly repulsive electron-electron component to a large extent and gives rise to its more favorable electrostatic interaction than that of QM/AMOEBA. Here we see that the charge penetration effect, as reflected by the lack of attractiveness in permanent electrostatics, involves the interplay of three distinct energy components.

Table 3: Components of permanent electrostatic interaction (in kcal/mol) for the $H_2O{\cdot\cdot}Cl^-$ complex at its equilibrium geometry, as computed with full QM, full AMOEBA, and QM/AMOEBA. "n" refers to nuclei, "e" represents electrons, and the numbers reported in parentheses are errors with regard to the full QM values. The nuclei-nuclei component is identical in all three calculations (18808.74 kcal/mol) so it is not listed in the table.

	Cl^-(n)$\cdot\cdot H_2O$(e)	Cl^-(e)$\cdot\cdot H_2O$(n)	Cl^-(e)$\cdot\cdot H_2O$(e)
Full QM	-18544.86	-19901.80	19618.83
Full AMOEBA	-18544.12 (+0.74)	-19915.14 (-13.34)	19634.95 (+16.12)
QM/AMOEBA	-18544.12 (+0.74)	-19901.80 (0.00)	19624.81 (+5.98)

Another notable feature of QM/AMOEBA for these H-bonding systems is that it yields overly favorable polarization energies. To shed some light on that, we revisit the water dimer and the water-halide (F^-, Cl^-) complexes. Three different types of calculations are performed: (i) mutually polarizable; (ii) allowing the polarization of H-acceptor only; (iii) allowing the polarization of H-donor only, and their results are collected in Table 4. Here excluding the polarization of a certain fragment is realized by freezing the associated molecular orbitals or induced dipoles in an SCF calculation. (e.g., for an "H-donor only" ALMO-EDA (SCF-MI) calculation, the orbital rotations on the fragment corresponding to the H-acceptor are suppressed.) According to these results, when AMOEBA water serves as the H-donor, it

consistently *under-polarizes* the H-acceptor in terms of the resulting polarization energy once its own polarization is forbidden, which should be related to its too weak electrostatic attraction with the electrons of the H-acceptor, as indicated in Table 3. It is the *over-polarization of AMOEBA water* as an H-donor (the so-called backward polarization) and the *exaggerated mutual polarization effect* that contribute to the overestimated polarization energies given by QM/AMOEBA for these systems. The most illustrative example is the water-F^- complex. Considering two unidirectional direct polarizations in QM/AMOEBA, the QM region (the H-acceptor F^-) is significantly *under-polarized* compared to the full QM reference based on the energetic criterion, while the AMOEBA water is only *marginally over-polarized* by F^-. However, the mutual polarization effect (whose magnitude can be measured by the difference between the total polarization energy and the sum of two unidirectional polarization energies) in QM/AMOEBA is found to be substantially larger than that in full QM. We think that such an exaggerated mutual polarization effect is related to the lack of explicit treatment of Pauli repulsion that would otherwise curb excessive electric polarization in our QM/AMOEBA model. When the H-acceptor is also described by AMOEBA, the under-polarization of H-acceptor and over-polarization of H-donor both become more pronounced, and the resulting total polarization energy can be either less (water dimer, water-F^-) or more (water-Cl^-) favorable than that of QM/AMOEBA. Although for some systems such as the water dimer, the pure AMOEBA polarization energy is in good agreement with its QM counterpart, the underlying physical pictures are not in line with each other, as indicated by the relative strength of "H-donor only" and "H-acceptor only" polarization energies for the water dimer. Here we see that the AMOEBA water is excessively prone to electric polarization in general as a proton donor, and that the improper strength of mutual polarization effect due to the lack of explicit modeling of Pauli repulsion in the present QM/AMOEBA model further exacerbates the over-polarization problem.

Table 4: Polarization energies (in kcal/mol) for the water dimer and water-Cl^-, F^- complexes computed with ALMO-EDA and QM/AMOEBA, by allowing (i) mutual polarization, (ii) polarization of the H-acceptor only and (iii) polarization of the H-donor only. "D" (H-donor) and "A" (H-acceptor) are used to specify the regions described by QM or AMOEBA.

		mutual	H-acceptor only	H-donor only
water dimer	ALMO-EDA	-1.17	-0.73	-0.35
	AMOEBA	-1.23	-0.42	-0.64
	QM(A)/AMB(D)	-1.49	-0.56	-0.59
water-Cl^-	ALMO-EDA	-3.77	-0.99	-2.35
	AMOEBA	-6.83	-0.59	-5.52
	QM(A)/AMB(D)	-5.72	-0.85	-3.53
water-F^-	ALMO-EDA	-16.88	-5.33	-10.12
	AMOEBA	-19.08	-0.59	-16.67
	QM(A)/AMB(D)	-19.68	-2.35	-10.23

3.6 Convergence of the errors with the size of MM region

The last aspect that we investigate in this work is the convergence behavior of the errors demonstrated above with increasing sizes of the MM region, i.e., the number of AMOEBA water molecules surrounding the QM solute. Here we revisit three solutes that have been previously studied by us in Ref. 70: NH_3, NH_4^+ and CN^-, which form another set of representatives of neutral, cationic and anionic species. For each solute, we choose one single snapshot from MD simulation (see Ref. 70 for the simulation details) and vary the number of solvent water molecules. By starting from the 10 water molecules that are closest to the solute (the position of the latter is marked by the center of the simulation sphere), we include 10 more water molecules that are the next closest at a time until the number of solvent molecules reaches 100, which is the maximum size of the MM region in this work. In order to make the full QM benchmarks for these systems computationally less demanding, we switch to the B97M-V functional[130] (which is a semi-local meta-GGA with the VV10 NLC) and the smaller def2-SVPD basis set. The choice of this basis has been validated by a benchmark study on systems containing 10–30 water molecules, where def2-SVPD and def2-TZVPPD yield similar full QM interaction energies once counterpoise corrections are

applied (see Table S1 in the Supporting Information). Correspondingly, the same density functional and basis set are applied to the QM region in the QM/AMOEBA calculations for these systems.

With respect to the full QM references, the errors of QM/AMOEBA in total interaction energies and their components for three solutes surrounded by varying numbers of water molecules are shown in Figure 6 (the original data for interaction energies and EDA are provided in Table S4 in the Supporting Information). Note that the solute-solvent interaction energies (and the energy components) are not guaranteed to vary monotonically with the number of solvent molecules since those distant solvent molecules may not be aligned favorably relative to the solute. For each system, the error in total interaction energy converges to a certain value with increasing sizes of the MM region, and so do the errors in three energy components. This is reasonable because both polarization and vdW interaction (including exchange-repulsion, dispersion and other short-range effects such as charge transfer) decay fairly rapidly with respect to distance, and the multipole moments on those distant solvent molecules should be able to give a proper description for their long-range electrostatic interaction with the QM region.

Turning to the results for each individual solute, NH_3 and NH_4^+ show reasonably balanced QM/MM interfaces, for which the errors in permanent electrostatics are largely compensated by the sufficiently softened vdW potentials such that the deviations in total interaction energy are small in magnitude. Contrary to most of the systems investigated above, QM/AMOEBA *overbinds* NH_4^+ with the solvent water molecules, and the absolute error fortuitously decreases with the addition of more water shells until the number of H_2O molecules reaches 50, thanks to the faster-increasing error in permanent electrostatics that makes the interaction energy less overbound. For NH_3, QM/AMOEBA only slightly underestimates its interaction with the first 10 water molecules (by 0.15 kcal/mol), and the error increases to about 1.8 kcal/mol when the number of water molecules reaches 100. Also, with the addi-

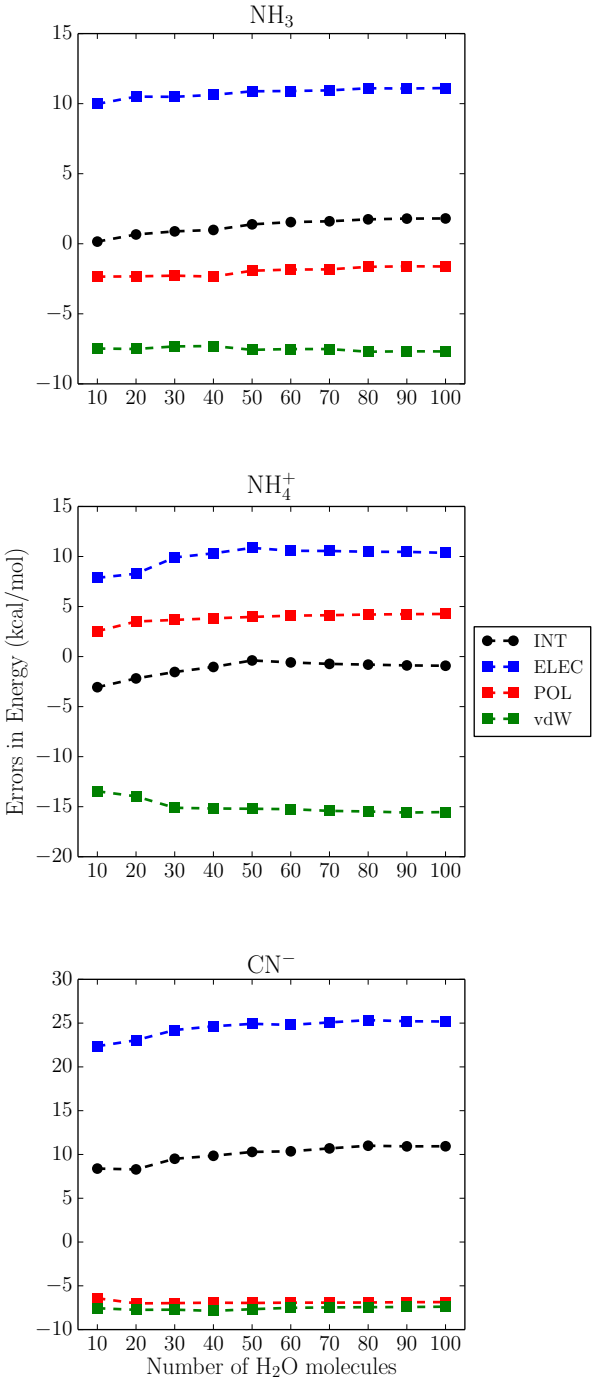

Figure 6: Errors of QM/AMOEBA (in kcal/mol) with respect to the full QM references for the solute-solvent interaction energies and their components when three solutes (NH_3, NH_4^+, and CN^-) are solvated with increasing numbers of water molecules. Here "QM" stands for the B97M-V/def2-SVPD level of theory.

tion of the first several water shells, we notice that the magnitude of errors in permanent electrostatics (and vdW interaction) varies more in the NH_4^+ case, indicating the difference between neutral and ionic solutes. CN^-, on the other hand, manifests the typical behavior of an anionic solute. The unbalanced errors in permanent electrostatics and vdW interactions result in the systematic underestimation of total interaction energy. It is also unusual that the errors in vdW interaction almost align with those in polarization energy. While a negative error in polarization energy whose magnitude is slightly larger than 5 kcal/mol is typical for anionic solutes (*vide supra*), the vdW potential here is undoubtedly too hard considering the significant charge penetration effect associated with CN^-. Nevertheless, in terms of the convergence behavior of the total interaction energy and its components, CN^- is *not* qualitatively different from the other two solutes.

4 Discussion and outlook

The results shown in Figure 6 indicate that aside from the fortuitous error cancellation between different energy components, the errors in solute-solvent interaction energies given by the present QM/AMOEBA model primarily arise from the interaction with solvent molecules in the close vicinity, while the long-range portion of the QM/MM interaction seems to be properly described. For instance, considering the solute-solvent electrostatic interactions for the three species studied in Sec. 3.6, the interactions with the first 10 water molecules account for 75–90% of the errors evaluated at the bulk limit (which can be estimated by the nearly converged errors with increasing numbers of water shells). Therefore, in order to reduce the errors accumulated in the short range and thus improve the accuracy of the resulting solute-solvent interaction energies, it seems necessary to treat the solvent molecules in close proximity to the solute with more sophisticated models.

The conceptually simplest approach is to absorb these solvent molecules into the QM

region, which, however, might significantly increase the computational demand. Meanwhile, it is often challenging to choose an appropriate size for the QM region *a priori*, as the convergence of QM/MM results with respect to the size of the QM region has been shown to be rather slow and not as monotonic as one might expect.[131–137] (In these studies, the size of the QM region required to reach the asymptotic limits for various properties often goes up to hundreds of atoms and sometimes exceeds the size of the largest model system in this work.) With the three systems discussed in Sec. 3.6, we evaluated the solute-solvent interaction energies with a set of QM/AMOEBA calculations where the first 10 closest water molecules are also included into the QM region. The results indicate that for a given solute-solvent system, including some solvent molecules into the QM region based on proximity does not consistently improve the resulting solute-solvent interaction energy. We refer the readers to Sec. S5 of the Supporting Information for a more detailed discussion.

Another plausible approach based on Figure 6 is to add a buffer layer in the middle of QM and MM regions wherein an improved AMOEBA model with modified functional forms is applied to describe the solvent molecules when we consider their interactions with the QM solute, while the interactions between solvent molecules (including those in the buffer layer) remain unchanged (still described with the original AMOEBA model). This buffer region is devised to reduce the errors in short-range solute-solvent interactions while providing a smooth transition from QM to AMOEBA, and methods such as adaptive QM/MM[138,139] can be used to ensure the smoothness of the PES when solvent molecules move in and out of the buffer region. We note that a buffer layer similar to this has been used by Olsen et al. in the formulation of their polarizable density embedding (PDE) model,[87] wherein the molecules have both frozen molecular orbitals (QM-like) and inducible dipoles (MM-like).

As a roadmap to a modified AMOEBA model whose interfacing with QM is improved, the first obvious goal is the proper treatment of charge penetration effect. There have been many efforts made aiming to incorporate this effect in various MM and QM/MM

models,[62,71–76,140–152] and under the framework of the AMOEBA force field, appropriate functional forms that account for charge penetration in the context of pure MM calculations have also been suggested in recent works.[153,154] Many of these methods are similar in spirit: on each MM site, the point charge or point multipoles are first separated into nuclear and electron contributions (similar to what has been done for the data in Table 3), and then the latter is replaced by a continuous distribution (e.g. Slater or Gaussian function) with a certain spatial extent, or alternatively, damping functions are applied to the components of electrostatic interactions that involve electrons. The development of these models has provided many options that can be potentially adopted in an improved QM/AMOEBA model.

However, we find that although applying a scheme similar to that in Ref. 153 to our QM/AMOEBA model is able to improve the description of permanent electrostatics (we refer the readers to the left panel of Figure S5 in the Supporting Information and Figure 2 in Ref. 155 for preliminary results), it might exacerbate the so-called "electron-spill" effect,[156] i.e., the QM region is enormously over-polarized so that the electrons are pulled out of the QM region, and energetically it results in vastly overestimated polarization energies (see Figure S6 and the right panel of Figure S5 in the Supporting Information). This is related to the aforementioned splitting of multipole moments which effectively places point charges of large magnitude (nuclei of MM atoms) near the QM/MM boundary, while the essential deficiency, nonetheless, is the purely empirical treatment of the repulsive vdW interaction (also termed exchange/Pauli/non-electrostatic repulsion), which fails to preclude the over-polarization of QM density by the electrostatic potential of MM. A model for Pauli repulsion that depends on electron density explicitly, despite being challenging because of the pure quantum nature of this interaction, is highly desirable not only for the sake of a more balanced and physically pertinent QM/MM interface, but also for the correct modeling of molecular properties that rely on the actual electronic structure, such as optical

excitation energies.[156,157] Popular methods for evaluating or incorporating Pauli repulsion in QM/MM (or other embedding) calculations include placing effective potentials on MM atoms[62,63,158–161] and other QM-derived approaches (e.g. the use of non-additive kinetic energy functionals, projection operators, etc.) that make direct use of pre-computed electron densities/MOs on embedding fragments.[87,162–169] On the other hand, overlap-based models have also been proposed to mimic the effect of Pauli repulsion in pure MM[145,170–178] and charge-dependent QM/MM[179,180] models. Further investigation is required to seek for a suitable approach that explicitly accounts for the effect of Pauli repulsion between QM electrons and AMOEBA fragments with moderate computational costs.

5 Conclusion

In this work, we have presented a QM/polarizable MM model which employs modern density functionals and the AMOEBA force field. The total energy of the coupled QM/AMOEBA system is variationally minimized with respect to both the QM electron density and the AMOEBA induced dipoles following the procedure introduced in Sec. 2.1 so that the mutual polarization between QM and MM regions is treated in a fully self-consistent fashion. The implementation of this model is achieved through the Q-Chem/LibEFP code interface, where Q-Chem serves as the driver for the whole QM/AMOEBA calculation and LibEFP is modified to accommodate the additional functionalities for the support of AMOEBA. We note that LibEFP is a portable library so it can be readily interfaced with other quantum chemistry software packages.

The proposed QM/AMOEBA model is employed for the evaluation of interaction energies between several simple solutes (including neutral and ionic species) and various numbers of solvent water molecules, which are equivalent to the interactions between QM and MM regions. With the goal of investigating the source of errors in the resulting solute-solvent

interaction energies, an EDA scheme is proposed to separate the total interaction energy crossing the QM/MM interface into contributions from permanent electrostatics, polarization and vdW interaction. This allows us to ascertain the agreement of each energy component with its counterpart in full QM calculations obtained by using the modified ALMO-EDA scheme. In general, the present QM/AMOEBA model yields reasonable total solute-solvent interaction energies for investigated neutral (H_2O, CH_4, and NH_3) and cationic (Na^+ and NH_4^+) species (often *not* for the correct reason, *vide infra*), but significantly underestimates the interactions for anionic solutes (Cl^-, F^-, and CN^-). Looking at the energy components more closely by means of EDA, the following points emerge:

- The permanent electrostatic interaction given by the current QM/AMOEBA model is always not sufficiently attractive, and it can suffer from even more significant CPE than in the pure AMOEBA scenario. This is most pronounced when the QM solute serves as the acceptor of protons from the MM region, as in systems like anion-water clusters.

- The vdW potential of AMOEBA is usually softened relative to its counterpart in full QM interaction energy. However, bearing in mind that the associated parameters are fitted together with pure AMOEBA's permanent electrostatics and polarization, they might no longer be suitable for QM/AMOEBA whose other two energy components are both shifted (see Figures 2 and 3).

- The error in polarization energy is usually smaller compared to the discrepancies in other two energy components, and the sign of the error turns out to be system-dependent. According to our tests, QM/AMOEBA overestimates the polarization energy for H-bonding systems where AMOEBA water plays the role of a proton donor (which are roughly the same systems that have the largest CPEs), and it can substantially underestimate the polarization energy for other systems (such as $CH_4(H_2O)_{20}$) as well.

Taken together, as we have summarized at the end of Sec. 3.3, the quality of solute-

solvent interaction energies given by the present QM/AMOEBA model strongly relies on the error cancellation amongst three components, especially the delicate balance between insufficiently favorable permanent electrostatics and vdW potential with extra softness. For several investigated solute-solvent systems, the "reasonable" total interaction energy masks the incorrectness of the underlying physics (one typical example is the $CH_4(H_2O)_{20}$ cluster), and such a "brittle" balance can break down once the permanent electrostatic and polarization components of QM/AMOEBA interactions are largely discrepant from their pure AMOEBA counterparts. It has also been shown that the errors in individual energy components are mostly rooted in the interaction with the solvent molecules in proximity to the QM region, and they converge rather rapidly with the addition of solvent shells, which implies that this model can be systematically improved by properly addressing the short-range discrepancy in each individual energy component. Future work will be devoted to this aspect based on the discussions in Sec. 4.

6 Associated Content

- **Supporting Information** Comparison of full QM polarization energies computed with the ALMO and the FERF approaches (for the systems in Secs. 3.2 and 3.3); basis set dependence of the full QM and QM/AMOEBA results for the systems in Secs. 3.2 and 3.6; original total interaction and EDA data for the $F^-(H_2O)_{10}$ cluster and the three systems discussed in Sec. 3.6; preliminary results of QM/AMOEBA using a modified AMOEBA water model that employs Gaussian-blurred monopoles; discussion of solute-solvent interaction energies given by QM/AMOEBA with 10 water molecules closest to the solute included in the QM region.

7 Acknowledgments

Y.M., T.H.G., and M.H.G. acknowledge the support by Grant No. CHE-1365731 from the U.S. National Science Foundation (NSF). Y.M. and M.H.G. also acknowledge the support by NSF Grant No. CHE-1363342. Y.S. acknowledges financial support by NIH grant No. GM096678-02, DOE grant No. DE-SC0011297, and the University of Oklahoma startup fund. J.D. and C.K.S. acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) Grant Nos. EP/J015059/1 and EP/K039156/1. The authors thank Tianyi Liu and Omar Demerdash (from the T.H.G. group) for providing the configurations of solutes interacting with their first solvation shells that are used in Sec. 3.3. Y.M. and Y.S. thank Lyudmila Slipchenko and Ilya Kaliman for the support on LibEFP development.

References

(1) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.

(2) Born, M. *Z. Phys.* **1920**, *1*, 45–48.

(3) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

(4) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011–2033.

(5) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2008**, *112*, 8651–8655.

(6) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 609–620.

(7) Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perkin Trans. 2* **1993**, *1993*, 799–805.

(8) Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1995**, *240*, 253–260.

(9) Mennucci, B.; Cancès, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506–10517.

(10) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.

(11) Truong, T. N. *J. Phys. Chem. B* **1998**, *102*, 7877–7881.

(12) Tomasi, J.; Mennucci, B.; Cancès, E. *J. Mol. Struct. THEOCHEM* **1999**, *464*, 211–226.

(13) Chipman, D. M. *J. Chem. Phys.* **2000**, *112*, 5558.

(14) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

(15) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(16) Lange, A. W.; Herbert, J. M. *J. Phys. Chem. Lett.* **2010**, *1*, 556–561.

(17) Lange, A. W.; Herbert, J. M. *J. Chem. Phys.* **2010**, *133*, 244111.

(18) You, Z.-Q.; Herbert, J. M. *J. Chem. Theory Comput.* **2016**, *12*, 4338–4346.

(19) Fattebert, J.-L.; Gygi, F. *J. Comput. Chem.* **2002**, *23*, 662–666.

(20) Fattebert, J.-L.; Gygi, F. *Int. J. Quant. Chem.* **2003**, *93*, 139–147.

(21) Scherlis, D. A.; Fattebert, J.-L.; Gygi, F.; Cococcioni, M.; Marzari, N. *J. Chem. Phys.* **2006**, *124*, 074103.

(22) Dziedzic, J.; Helal, H. H.; Skylaris, C.-K.; Mostofi, A. A.; Payne, M. C. *Europhys. Lett.* **2011**, *95*, 43001.

(23) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760–768.

(24) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. *Acc. Chem. Res.* **2009**, *42*, 489–492.

(25) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9080.

(26) Sprik, M.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1996**, *105*, 1142–1152.

(27) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 3572–3580.

(28) McGrath, M. J.; Siepmann, J. I.; Kuo, I. F. W.; Mundy, C. J.; Vandevondele, J.; Hutter, J.; Mohamed, F.; Krack, M. *ChemPhysChem* **2005**, *6*, 1894–1901.

(29) Zhang, C.; Donadio, D.; Gygi, F.; Galli, G. *J. Chem. Theory Comput.* **2011**, *7*, 1443–1449.

(30) Del Ben, M.; Schönherr, M.; Hutter, J.; VandeVondele, J. *J. Phys. Chem. Lett.* **2013**, *4*, 3753–3759.

(31) Gaiduk, A. P.; Gygi, F.; Galli, G. *J. Phys. Chem. Lett.* **2015**, *6*, 2902–2908.

(32) White, J. A.; Schwegler, E.; Galli, G.; Gygi, F. *J. Chem. Phys.* **2000**, *113*, 4668–4673.

(33) Lightstone, F. C.; Schwegler, E.; Hood, R. Q.; Gygi, F.; Galli, G. *Chem. Phys. Lett.* **2001**, *343*, 549–555.

(34) Lightstone, F. C.; Schwegler, E.; Allesch, M.; Gygi, F.; Galli, G. *ChemPhysChem* **2005**, *6*, 1745–1749.

(35) Ikeda, T.; Boero, M.; Terakura, K. *J. Chem. Phys.* **2007**, *126*, 01B611.

(36) Schmidt, D. A.; Scipioni, R.; Boero, M. *J. Phys. Chem. A* **2009**, *113*, 7725–7729.

(37) Zhang, C.; Pham, T. A.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2013**, *138*, 181102.

(38) Gaiduk, A. P.; Zhang, C.; Gygi, F.; Galli, G. *Chem. Phys. Lett.* **2014**, *604*, 89–96.

(39) Gaiduk, A. P.; Galli, G. *J. Phys. Chem. Lett.* **2017**, *8*, 1496–1502.

(40) Schmidt, J.; Vandevondele, J.; William Kuo, I. F.; Sebastiani, D.; Ilja Siepmann, J.; Hutter, J.; Mundy, C. J. *J. Phys. Chem. B* **2009**, *113*, 11959–11964.

(41) Wang, J.; Roman-Perez, G.; Soler, J. M.; Artacho, E.; Fernandez-Serra, M.-V. *J. Chem. Phys.* **2011**, *134*, 024516.

(42) Lin, I.-C.; Seitsonen, A. P.; Tavernelli, I.; Rothlisberger, U. *J. Chem. Theory Comput.* **2012**, *8*, 3902–3910.

(43) Ma, Z.; Zhang, Y.; Tuckerman, M. E. *J. Chem. Phys.* **2012**, *137*, 044506.

(44) Forster-Tonigold, K.; Groß, A. *J. Chem. Phys.* **2014**, *141*, 064501.

(45) DiStasio, R. A.; Santra, B.; Li, Z.; Wu, X.; Car, R. *J. Chem. Phys.* **2014**, *141*, 084502.

(46) Bankura, A.; Santra, B.; DiStasio, R. A.; Swartz, C. W.; Klein, M. L.; Wu, X. *Mol. Phys.* **2015**, *113*, 2842–2854.

(47) Gillan, M. J.; Alfè, D.; Michaelides, A. *J. Chem. Phys.* **2016**, *144*, 130901.

(48) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

(49) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718–730.

(50) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.

(51) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298–305.

(52) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(53) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.

(54) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(55) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wirkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(56) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(57) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(58) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(59) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(60) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(61) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

(62) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968–1986.

(63) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.

(64) Gurunathan, P. K.; Acharya, A.; Ghosh, D.; Kosenkov, D.; Kaliman, I.; Shao, Y.; Krylov, A. I.; Slipchenko, L. V. *J. Phys. Chem. B* **2016**, *120*, 6562–6574.

(65) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(66) Steindal, A. H.; Ruud, K.; Frediani, L.; Aidas, K.; Kongsted, J. *J. Phys. Chem. B* **2011**, *115*, 3027–3037.

(67) Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. *J. Chem. Theory Comput.* **2011**, *7*, 2209–2217.

(68) Kratz, E. G.; Walker, A. R.; Lagardère, L.; Lipparini, F.; Piquemal, J. P.; Cisneros, G. A. *J. Comput. Chem.* **2016**, *37*, 1019–1029.

(69) Loco, D.; Polack, E.; Caprasecca, S.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Mennucci, B. *J. Chem. Theory Comput.* **2016**, *12*, 3654–3661.

(70) Dziedzic, J.; Mao, Y.; Shao, Y.; Ponder, J. W.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K. *J. Chem. Phys.* **2016**, *145*, 124106.

(71) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534–10547.

(72) Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A. *J. Chem. Phys. B* **2006**, *110*, 13682–13684.

(73) Cisneros, G. A.; Tholander, S.; Parisel, O.; Darden, T.; Elking, D.; Perera, L.; Piquemal, J.-P. *Int. J. Quant. Chem.* **2008**, *108*, 1905–1912.

(74) Biswas, P.; Gogonea, V. *J. Chem. Phys.* **2008**, *129*, 154108.

(75) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 3330–3342.

(76) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 1989–1998.

(77) Woodcock, H. L.; Hodošček, M.; Gilbert, A. T.; Gill, P. M.; Schaefer, H. F.; Brooks, B. R. *J. Comput. Chem.* **2007**, *28*, 1485–1502.

(78) Lu, Z.; Zhang, Y. *J. Chem. Theory Comput.* **2008**, *4*, 1237–1248.

(79) Boulanger, E.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 4527–4538.

(80) Boulanger, E.; Thiel, W. *J. Chem. Theory Comput.* **2014**, *10*, 1795–1809.

(81) Bryce, R. A.; Buesnel, R.; Hillier, I. H.; Burton, N. A. *Chem. Phys. Lett.* **1997**, *279*, 367–371.

(82) Lipparini, F.; Cappelli, C.; Barone, V. *J. Chem. Theory Comput.* **2012**, *8*, 4153–4165.

(83) Lipparini, F.; Cappelli, C.; Barone, V. *J. Chem. Phys.* **2013**, *138*, 234108.

(84) Thompson, M. A.; Schenter, G. K. *J. Chem. Phys.* **1995**, *99*, 6374–6386.

(85) Aida, M.; Yamataka, H.; Dupuis, M. *Int. J. Quant. Chem.* **2000**, *77*, 199–210.

(86) Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J. *J. Chem. Phys.* **2007**, *126*, 154112.

(87) Olsen, J. M. H.; Steinmann, C.; Ruud, K.; Kongsted, J. *J. Phys. Chem. A* **2015**, *119*, 5344–5355.

(88) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

(89) Thellamurege, N. M.; Si, D.; Cui, F.; Zhu, H.; Lai, R.; Li, H. *J. Comput. Chem.* **2013**, *34*, 2816–2833.

(90) Schwörer, M.; Breitenfeld, B.; Tröster, P.; Bauer, S.; Lorenzen, K.; Tavan, P.; Mathias, G. *J. Chem. Phys.* **2013**, *138*, 244103.

(91) Schwörer, M.; Wichmann, C.; Tavan, P. *J. Chem. Phys.* **2016**, *144*, 114504.

(92) Caprasecca, S.; Curutchet, C.; Mennucci, B. *J. Chem. Theory Comput.* **2012**, *8*, 4462–4473.

(93) Caprasecca, S.; Jurinovich, S.; Viani, L.; Curutchet, C.; Mennucci, B. *J. Chem. Theory Comput.* **2014**, *10*, 1588–1598.

(94) Caprasecca, S.; Jurinovich, S.; Lagardere, L.; Stamm, B.; Lipparini, F. *J. Chem. Theory Comput.* **2015**, *11*, 694–704.

(95) Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.

(96) Shaw, K. E.; Woods, C. J.; Mulholland, A. J. *J. Phys. Chem. Lett.* **2010**, *1*, 219–223.

(97) König, G.; Mei, Y.; Pickard, F. C.; Simmonett, A. C.; Miller, B. T.; Herbert, J. M.; Woodcock, H. L.; Brooks, B. R.; Shao, Y. *J. Chem. Theory Comput.* **2016**, *12*, 332–344.

(98) Pickard, F. C.; König, G.; Simmonett, A. C.; Shao, Y.; Brooks, B. R. *Bioorg. Med. Chem.* **2016**, *24*, 4988–4997.

(99) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(100) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(101) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(102) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Ku, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; III, H. L. W.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; Jr., R. A. D.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; ONeill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stck, D.; Su, Y.-C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J.; III, W. A. G.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; III, H. F. S.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Voorhis, T. V.; Herbert, J. M.; Krylov, A. I.;

Gill, P. M.; Head-Gordon, M. *Mol. Phys.* **2015**, *113*, 184–215.

(103) Kaliman, I. A.; Slipchenko, L. V. *J. Comput. Chem.* **2013**, *34*, 2284–2292.

(104) Kaliman, I. A.; Slipchenko, L. V. *J. Comput. Chem.* **2015**, *36*, 129–135.

(105) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(106) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(107) Ren, P.; Wu, C.; Ponder, J. W. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.

(108) Gao, J. *J. Comput. Chem.* **1997**, *18*, 1061–1071.

(109) Thellamurege, N. M.; Hirao, H. *J. Phys. Chem. B* **2014**, *118*, 2084–2092.

(110) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. *J. Phys. Chem. A* **2007**, *111*, 8753–8765.

(111) Horn, P. R.; Mao, Y.; Head-Gordon, M. *J. Chem. Phys.* **2016**, *144*, 114107.

(112) Horn, P. R.; Mao, Y.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2016**, *18*, 23067–23079.

(113) Phipps, M. J. S.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. *Chem. Soc. Rev.* **2015**, *44*, 3177–3211.

(114) Mao, Y.; Demerdash, O.; Head-Gordon, M.; Head-Gordon, T. *J. Chem. Theory Comput.* **2016**, *12*, 5422–5437.

(115) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. *Annual reports in computational chemistry*; Elsevier, 2007; Vol. 3; pp 177–193.

(116) Pulay, P. *Chem. Phys. Lett.* **1980**, *73*, 393–398.

(117) Pulay, P. *J. Comput. Chem.* **1982**, *3*, 556–560.

(118) Mardirossian, N.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904–9924.

(119) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*, 244103.

(120) Lao, K. U.; Schaffer, R.; Jansen, G.; Herbert, J. M. *J. Chem. Theory Comput.* **2015**, *11*, 2473–2486.

(121) Lao, K. U.; Herbert, J. M. *J. Phys. Chem. A* **2015**, *119*, 235–252.

(122) Rappoport, D.; Furche, F. *J. Chem. Phys.* **2010**, *133*, 134105.

(123) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. *J. Chem. Phys.* **2006**, *124*, 204105.

(124) Horn, P. R.; Head-Gordon, M. *J. Chem. Phys.* **2015**, *143*, 114111.

(125) Ponder, J. W. *TINKER: Software Tools for Molecular Design*, version 7.1 ed.; Washington University: St. Louis, MO, 2015.

(126) Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.

(127) Grossfield, A. *J. Chem. Phys.* **2005**, *122*, 024506.

(128) Lao, K. U.; Herbert, J. M. *J. Chem. Phys.* **2013**, *139*, 034107.

(129) Deible, M. J.; Tuguldur, O.; Jordan, K. D. *J. Phys. Chem. B* **2014**, *118*, 8257–8263.

(130) Mardirossian, N.; Head-Gordon, M. *J. Chem. Phys.* **2015**, *142*, 074111.

(131) Fox, S. J.; Pittock, C.; Fox, T.; Tautermann, C. S.; Malcolm, N.; Skylaris, C.-K. *J. Chem. Phys.* **2011**, *135*, 224107.

(132) Flaig, D.; Beer, M.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2012**, *8*, 2260–2271.

(133) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. *J. Chem. Theory Comput.* **2012**, *8*, 5092.

(134) Liao, R.-Z.; Thiel, W. *J. Comput. Chem.* **2013**, *34*, 2389–2397.

(135) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martinez, T. J. *J. Phys. Chem. B* **2016**, *120*, 11381.

(136) Roßbach, S.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2017**, *13*, 1102–1107.

(137) Karelina, M.; Kulik, H. J. *J. Chem. Theory Comput.* **2017**, *13*, 563–576.

(138) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. *J. Chem. Theory Comput.* **2009**, *5*, 2212–2221.

(139) Nielsen, S. O.; Bulo, R. E.; Moore, P. B.; Ensing, B. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12401–12414.

(140) Wheatley, R. J. *Mol. Phys.* **1993**, *79*, 597–610.

(141) Wheatley, R. J.; Mitchell, J. B. *J. Comput. Chem.* **1994**, *15*, 1187–1198.

(142) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300–7306.

(143) Slipchenko, L.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276–291.

(144) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353–10359.

(145) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.

(146) Elking, D. M.; Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190–202.

(147) Cisneros, G. A. *J. Chem. Theory Comput.* **2012**, *8*, 5072–5080.

(148) Spackman, M. A. *Chem. Phys. Lett.* **2006**, *418*, 158–162.

(149) Stone, A. J. *J. Phys. Chem. A* **2011**, *115*, 7017–7027.

(150) Tafipolsky, M.; Engels, B. *J. Chem. Theory Comput.* **2011**, *7*, 1791–1803.

(151) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2014**, *10*, 4480–4487.

(152) Narth, C.; Lagardere, L.; Polack, E.; Gresh, N.; Wang, Q.; Bell, D. R.; Rackers, J. A.; Ponder, J. W.; Ren, P. Y.; Piquemal, J.-P. *J. Comput. Chem.* **2016**, *37*, 494–506.

(153) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. *J. Chem. Theory Comput.* **2015**, *11*, 2609–2618.

(154) Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. *Phys. Chem. Chem. Phys.* **2017**, *19*, 276–291.

(155) Albaugh, A.; Boateng, H. A.; Bradshaw, R. T.; Demerdash, O. N.; Dziedzic, J.; Mao, Y.; Margul, D. T.; Swails, J. M.; Zeng, Q.; Case, D. A.; Kenneth, P.; Essex, J. W.; Head-Gordon, M.; Pande, V. S.; Ponder, J. W.; Shao, Y.; Skylaris, C.-K.; Todorov, I. T.; Tuckerman, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2016**, *120*, 9811–9832.

(156) Nâbo, L. J.; Olsen, J. M. H.; Holmgaard List, N.; Solanko, L. M.; Wüstner, D.; Kongsted, J. *J. Chem. Phys.* **2016**, *145*.

(157) Fradelos, G.; Wesołowski, T. A. *J. Phys. Chem. A* **2011**, *115*, 10018–10026.

(158) Vaidehi, N.; Wesolowski, T. A.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 4264–4271.

(159) Chalmet, S.; Ruiz-López, M. F. *Chem. Phys. Lett.* **2000**, *329*, 154–159.

(160) Sebastiani, D.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 2807–2815.

(161) Jin, Y.; Johnson, E. R.; Hu, X.; Yang, W.; Hu, H. *J. Comput. Chem.* **2013**, *34*, 2380–2388.

(162) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050–8053.

(163) Wesolowski, T. A.; Weber, J. *Chem. Phys. Lett.* **1996**, *248*, 71–76.

(164) Wesolowski, T. A. *Computational chemistry: reviews of current trends*; World Scientific: Singapore, 2006; Vol. 10; pp 1–82.

(165) Barandiarán, Z.; Seijo, L. *J. Chem. Phys.* **1988**, *89*, 5739–5746.

(166) Swerts, B.; Chibotaru, L. F.; Lindh, R.; Seijo, L.; Barandiaran, Z.; Clima, S.; Pierloot, K.; Hendrickx, M. F. *J. Chem. Theory Comput.* **2008**, *4*, 586–594.

(167) Ben-Nun, M.; Martınez, T. J. *Chem. Phys. Lett.* **1998**, *290*, 289–295.

(168) Öhrn, A.; Karlström, G. *Mol. Phys.* **2006**, *104*, 3087–3099.

(169) Söderhjelm, P.; Öhrn, A. *Chem. Phys. Lett.* **2009**, *468*, 94–99.

(170) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quant. Chem.* **1986**, *29*, 101–118.

(171) Gresh, N.; Piquemal, J.-P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113–1130.

(172) Wheatley, R. J.; Price, S. L. *Mol. Phys.* **1990**, *69*, 507–533.

(173) Jensen, J. H.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313–1325.

(174) Jensen, J. H.; Gordon, M. S. *J. Chem. Phys.* **1998**, *108*, 4772.

(175) Gavezzotti, A. *J. Phys. Chem. B* **2003**, *107*, 2344–2353.

(176) Söderhjelm, P.; Karlström, G.; Ryde, U. *J. Chem. Phys.* **2006**, *124*, 244101.

(177) Tafipolsky, M.; Ansorg, K. *J. Chem. Theory Comput.* **2016**, *12*, 1267–1279.

(178) Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. *J. Chem. Theory Comput.* **2016**, *12*, 3851–3870.

(179) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2007**, *127*, 194101.

(180) Kuechler, E. R.; Giese, T. J.; York, D. M. *J. Chem. Phys.* **2015**, *143*, 234111.

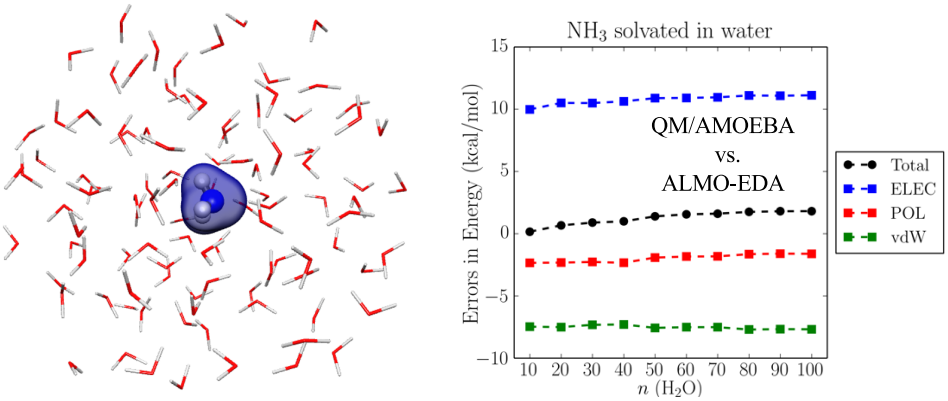

Figure 7: For Table of Contents (TOC) only