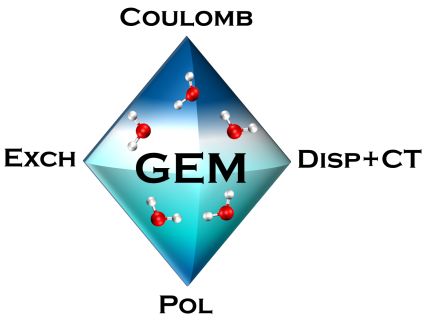
# Improvement of the Gaussian Electrostatic Model by Separate Fitting of Coulomb and Exchange-Repulsion Densities and Implementation of a new Dispersion term

Sehr Naseem-Khan,[1] Jean-Philip Piquemal,[2, a)] and G. Andrés Cisneros*[1]

[1)] *Department of Chemistry, University of North Texas, Denton, Texas, 76201, USA*

[2)] *Laboratoire de Chimie Théorique, Sorbonne Université, UMR 7616 CNRS, 75005, Paris, France*

The description of each separable contribution of the intermolecular interaction is a useful approach to develop polarizable force fields (polFF). The Gaussian Electrostatic Model (GEM) is based on this approach, coupled with the use of density fitting techniques. In this work, we present the implementation and testing of two improvements of GEM: the Coulomb and Exchange-Repulsion energies are now computed with separate frozen molecular densities, and a new dispersion formulation inspired by the SIBFA polFF, which has been implemented to describe the dispersion and charge–transfer interactions. Thanks to the combination of GEM characteristics and these new features, we demonstrate a better agreement of the computed structural and condensed properties for water with experimental results, as well as binding energies in the gas phase with the *ab initio* reference compared with the previous GEM* potential. This work provides further improvements to GEM and the items that remain to be improved, and the importance of the accurate reproduction for each separate contribution.




[a)]Institut Universitaire de France, 75005, Paris, France; Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA

## I. INTRODUCTION

The development of polarizable water models continues to be a field of intense research[1]. As computational studies of biological systems are mostly performed in liquid water when conducting molecular dynamics (MD) simulations, it is then primordial to develop an accurate and robust water model. The latter is therefore used as an foundation for any polFF since water is very challenging to correctly describe with either classical and polarizable potential forms. However, it is widely known that polFF is able to better describe the molecular charge distribution thanks to the use of multipoles instead of point charges, but also anisotropy effect among others.

Several polFF have been developed such as AMOEBA[2], AMOEBA+[3], SIBFA[4], EFP[5], X-Pol[6,7], NEMO[8], or HIPPO[9], that commonly share a main philosophy of development: to reproduce each contribution of the intermolecular energy ($E_{int}$) potential based on Energy Decomposition Analysis methods[10–12] and/or the Symmetry-Adapted Perturbation Theory.[13–21] In those methods, $E_{int}$ is decomposed into several contributions such as the electrostatic (Coulomb), exchange-repulsion, polarization, charge-transfer (charge-delocalization) and dispersion.

The Gaussian Electrostatic Model (GEM)[22] also shares this approach. GEM uses density fitting techniques[23–25] with Hermite Gaussian auxiliary basis sets (ABSs) to reproduce molecular electronic densities.[26] Thanks to these key features, GEM is able to describe with high accuracy the electronic density, and overcomes some known limitations associated with a discrete description of the charge density, such as the charge penetration error and inaccurate anisotropy effects. GEM has also been applied to QM/MM simulations resulting in an accurate approach for the embedding environment in a QM/MM context.[27,28] Additionally, the use of Particle-Mesh Ewald (PME)[29] or Fast Fourier Poisson (FFP)[30] methods enables the fast and efficient evaluation of the required Gaussian integrals.[31]

As previously reported by some of us,[32] the Coulomb and exchange-repulsion contributions calculated with frozen densities in the GEM context, when the densities are fitted with small auxiliary basis sets require separate fitting coefficients. This is due to the fact that small auxiliary basis sets are not capable of describing all the details of the valence molecular electronic densities. Thus, fitting two sets of coefficients allows for a better reproduction of the energies and forces.[32] In this work, we have then implemented the use of separate sets of Hermite coefficients to compute Coulomb and exchange-repulsion energies.

Another challenge for polFF that aim to reproduce each separate intermolecular interaction contribution is the charge-transfer term. In *ab initio* methods, several approaches have been proposed to define the charge-transfer (charge-delocalization).[21,33,34] Similar to the exchange-repulsion, the pair-wise charge–transfer term behaves exponentially, and several methods forms have been proposed such as the analytical form of charge-transfer and its gradients of the SIBFA[35] polFF have been recently implemented in the Tinker-HP package.[36]

Another approach is to neglect or approximate the charge–transfer contribution. Due to its exponential behavior, this term has been included as part of the dispersion contribution in a similar vein as the Lennard-Jones or Halgren potentials. Such as in the case of the previous version of GEM, named GEM$^*$, which is a hybrid polarizable force field defined as:

$$E_{tot}^{GEM^*} = E_{Coulomb}^{GEM} + E_{exch-rep}^{GEM} + E_{pol}^{AMOEBA} + E_{disp+ct}^{modHalgren} + E_{bonded}^{AMOEBA} \qquad (1)$$

where the Coulomb ($E_{Coulomb}$) and exchange-repulsion ($E_{exch-rep}$) terms are described with GEM; the polarization, and bonded ($E_{bonded}$) terms are described with the respective terms from the AMOEBA polarizable force field, and the charge-transfer ($E_{ct}$) and dispersion ($E_{disp}$) are approximated by fitting them together to the modified Halgren term.

In line with this approach, we propose here a modified dispersion functional form, inspired by the SIBFA polFF potential. This work brings then improvements upon GEM* and marks a step forward for the fully separable GEM.

The remainder of the article is composed as follows: we describe first the GEM potential and the new dispersion function. This is followed by the results for water computed with GEM in both gas and condensed phases compared with the previous functional form and parametrization. We finally conclude and discuss on future GEM works.

## II. METHODS

In the original GEM* water model from Duke *et al.*[31], the same fitted molecular densities, $\tilde{\rho}(r)$, expressed as an expansion of primitive Cartesian Hermite Gaussian functions such as:

$$\tilde{\rho}(r) = \sum_k x_k k(r) \tag{2}$$

were used to compute $E_{Coulomb}$ and $E_{exch-rep}$, which they can be expressed, respectively, as follows:

$$E^{GEM}_{Coulomb} = \int \frac{\tilde{\rho}_A(r_A)\tilde{\rho}_B(r_B)}{r_{AB}} dr + \int \frac{Z_A\tilde{\rho}_B(r_B)}{r_{AB}} dr + \int \frac{\tilde{\rho}_A(r_A)Z_B}{r_{AB}} dr + \frac{Z_A Z_B}{r_{AB}} \tag{3}$$

where $Z_{A,B}$ represents the nuclei of atom $A$ and $B$, and $r_{AB}$ is the distance between atoms $A$ and $B$,

$$E^{GEM}_{exch-rep} = K \int \tilde{\rho}_A(r)\tilde{\rho}_B(r) dr \tag{4}$$

where $K$ is a proportionality coefficient.

For $E_{disp+ct}$, GEM* uses only the attractive term of the Halgren potential as GEM* already describes repulsive interactions with molecular densities in Eq. (4). $E_{disp+ct}^{modHalgren}$ is defined as:

$$E_{disp+ct}^{modHalgren} = \varepsilon_{AB} \left[ \frac{1.07\, R_{AB}^*}{R_{AB} + 0.07\, R_{AB}^*} \right]^7 \tag{5}$$

where $\varepsilon_{AB}$ is the depth of the potential well, $R_{AB}$ and $R_{AB}^*$ being respectively the distance and the equilibrium distance between atom $A$ and $B$.

In this work, we have implemented the ability to employ different frozen molecular densities used to compute the Coulomb and exchange-repulsion. In this case, the approximate densities are now fitted to reduce the error for their respective components with respect to their SAPT(DFT) couterparts. That is, the GEM density used for the Coulomb interaction is fitted to minimize the error in Coulomb intermolecular interactions, while the GEM density used for exchange-repulsion is fitted to minimize the exchange-repulsion intermolecular error.[32,37,38]

Additionally, we have implemented a new dispersion formulation inspired by the SIBFA polFF[4] dispersion equation in the gem.pmemd code (AmberTools 21). Following the same approach as done in GEM* with the modified Halgren potential, this new dispersion formulation also describes dispersion and charge-transfer interactions. We denote GEM the polFF developed here which can be defined as:

$$E_{tot}^{GEM} = E_{Coulomb}^{GEM} + E_{exch-rep}^{GEM} + E_{pol}^{AMOEBA} + E_{disp+ct}^{GEM} + E_{bonded}^{AMOEBA} \tag{6}$$

For $E_{Coulomb}$, we have used the previously fitted density obtained from average molecular densities of a selection of 500 water monomers extracted from the dimer surface by Babin *et al.*[39,40] to minimize the Coulomb intermolecular interaction of those same 500 dimers using the SAPT2+3 level of theory as the *ab initio* reference[31]. For $E_{exch-rep}$, we have fitted the molecular densities using the Smith dimers[41] at the SAPT(DFT)/PBE0 level of theory as reference. As demonstrated in Ref[33] where

the SAPT(DFT) method is recommended as the *ab initio* reference, the difference between SAPT(DFT) and SAPT2+3 is negligible for the first-order contributions ($E_{Coulomb}$ and $E_{exch-rep}$) for the water dimer. Thus, the different theories do not result in significant deviations for the reference values.

The choice of using SAPT(DFT) instead of continuing with SAPT2+3 has been also motivated by its ability to decompose the induction energy into polarization and charge-transfer (charge-delocalization), which will be required for future development of GEM. Note that for both $E_{Coulomb}$ and $E_{exch-rep}$, the same auxiliary basis set has been used.

The new $E_{disp+ct}$ is only based on the atom-atom contribution of the SIBFA dispersion equation, which describes short and long range interactions of order $1/R^6$, $1/R^8$ and $1/R^{10}$, also including a damping function, corresponding to:

$$E_{disp+ct}^{GEM} = \sum_{n=6,8,10} C^n O_{AB}\, exp\left(-\alpha(n)\beta_{AB}\left[\frac{r_A + r_B}{R_{AB}^*} - 1\right]\right) / \left(\frac{R_{AB}^*}{2\sqrt{r_A r_B}}\right)^n \qquad (7)$$

where $C^{6,8,10}$ and $\alpha(6,8,10)$ are constant coefficients, $O_{AB}$ and $\beta_{AB}$ are respectively the overlap and damping factors between atom $A$ and $B$. $R_{AB}^*$ is the equilibrium distance, and $r_A$ and $r_B$ represent respectively the effective disp+ct radii of atom $A$ and $B$.

For the $\beta_{AB}$ parameter, we have introduced a dependency according to the nature of the atom-atom interaction. For example, for water, three different values have been parametrized to describe O-O, O-H and H-H interactions. This is not present in the original SIBFA dispersion where only one value of $\beta_{AB}$ is employed (Table S1). However, we have not used the atom/lone-pair and lone-pair/lone-pair interaction contributions that are explicit in the original SIBFA dispersion equation, since GEM does not employ explicit lone-pair positions. Additionally, SIBFA explicitly includes the exchange-dispersion term, which has not been used here, favoring to incorporate this small contribution in the $E_{disp+ct}$ term. It has been shown that

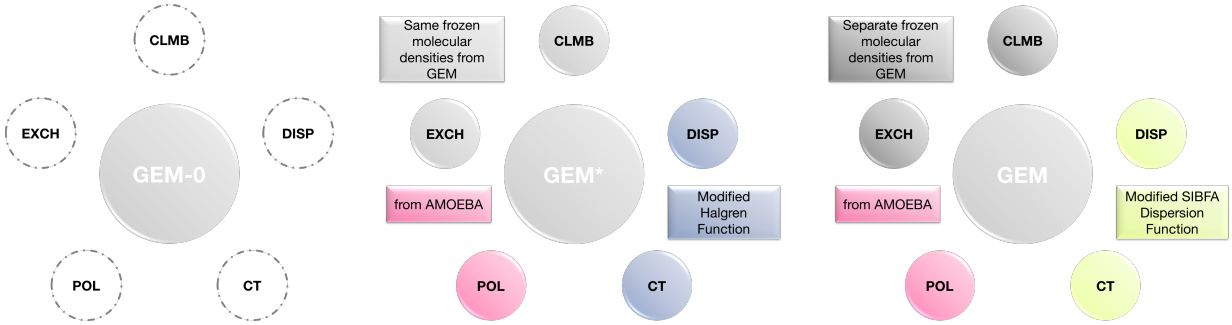
exchange-dispersion and charge-transfer share a similar exponential behavior and magnitude but with opposite sign.[33]. This is a similar feature as the one employed in the Lennard-Jones and Halgren potentials, with the combination of attractive and repulsive terms. However, in this case the exchange-dispersion and charge-transfer terms are about one order of magnitude smaller than the exchange-repulsion and dispersion contribution. Thus, it is expected that the inclusion of the exchange-dispersion term within $E_{disp+ct}$, will yield reasonable results. As shown below, the resulting disp+ct functional form inspired by the SIBFA polFF is able to accurately describe the disp+ct interactions for a variety of oligomers.

For the fitting of the $E^{GEM}_{disp+ct}$ term, we have followed the same strategy as in GEM$^*$, i.e, the *ab initio* reference values are obtained from the difference between of the CCSD(T) binding energy and all the GEM contributions, it can be expressed as:

$$E^{ref}_{disp+ct}[GEM] = E^{[CCSD(T)]}_{bind} - \left(E^{GEM}_{Coulomb} - E^{GEM}_{exch-rep} - E^{GEM}_{pol}\right) \tag{8}$$

In doing so, $E^{GEM}_{disp+ct}$ includes not only dispersion and charge-transfer contributions, but also higher-order many-body effects. Similarly to $E_{exch-rep}$, we have fitted $E^{GEM}_{disp+ct}$ using only the inter-molecular interaction of the Smith dimers as reference. Additionally, we have also slightly adjusted the parameters (effective disp+ct radii of the atom of oxygen and the $\beta_{AB}$ parameter) as a function of the radial distribution function (RDF) and the density, computed using short MD simulations for a water box of 512 water molecules at 300K.

Lastly, $E_{pol}$ is computed using the AMOEBA formulation but using the GEM multipoles, which can be derived from the fitted molecular densities.[42,43] Figure 1 illustrates the schematic evolution of the development of the GEM polFF starting from the initial potential fitted only with $l = 0$ Hermite primitives[44] to the current version.




(a) GEM-0 (b) GEM* (c) GEM

FIG. 1: Schematic representation of the development of GEM.

## III. RESULTS AND DISCUSSION

### A. Separability of the frozen molecular densities for the Coulomb and Exchange-Repulsion

As described above, to parametrize $E_{exch-rep}$ and $E_{disp+ct}$, the *ab initio* reference data is based on the ten Smith dimers calculated at the SAPT(DFT)/aug–cc-pVTZ level. These specific dimers are representative of attractive and repulsive forces, which are important to accurately describe intermolecular interactions in both gas and condensed phases. In Figure 2, we can observe that the use of of separate Hermite coefficients is strongly favorable for $E_{exch-rep}$. GEM shows a significantly better agreement for this contribution compared to SAPT(DFT), with a Root Mean Square Error (RMSE) of 0.25 kcal mol$^{-1}$ for GEM vs. 1.11 kcal mol$^{-1}$ for GEM*. For the Coulomb energy, GEM*/GEM employ the same fitted densities and thus the agreement with SAPT(DFT) is very good, with an RMSE of 0.23 kcal mol$^{-1}$.

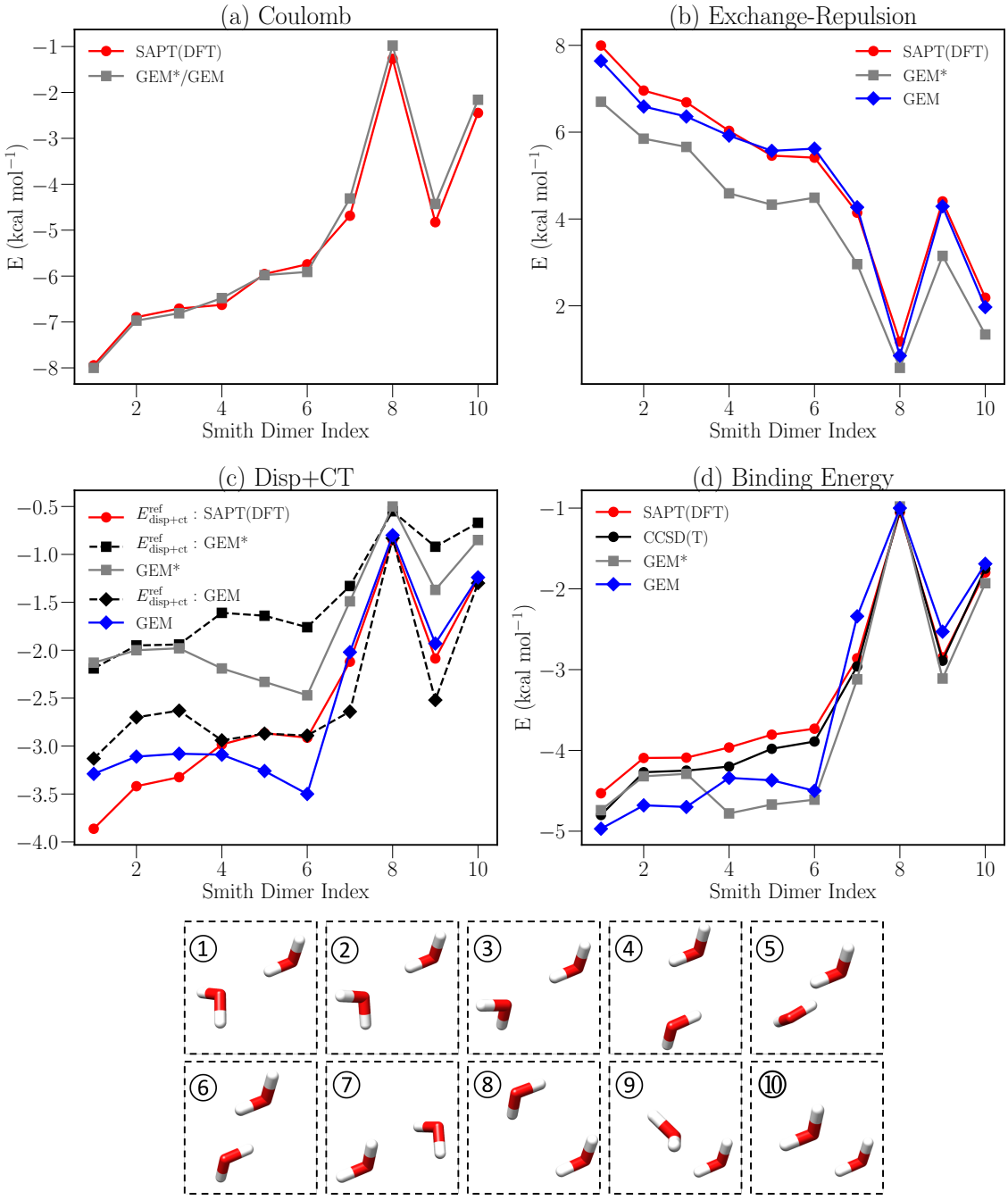



FIG. 2: Comparison of the Coulomb, exchange-repulsion, disp+ct and binding energies computed with GEM* and GEM with SAPT(DFT) for the ten Smith dimers. For computational details of the three disp+ct reference curves see equations (8), (9) and (10) in the main text.

## B. New functional form to describe Dispersion and Charge-Transfer

For the parametrization of the new disp+ct contribution, the reference energy ($E^{ref}_{disp+ct}$) is different between GEM and GEM*, since $E_{exch-rep}$ are computed with distinct sets of parameters (see Eq. (8)). For GEM* the reference energies, $E^{ref}_{disp+ct}$, are calculated by:

$$E^{ref}_{disp+ct}[\mathrm{GEM^*}] = \mathrm{E}^{[\mathrm{CCSD(T)}]}_{\mathrm{bind}} - \left(\mathrm{E}^{\mathrm{GEM^*}}_{\mathrm{Coulomb}} - \mathrm{E}^{\mathrm{GEM^*}}_{\mathrm{exch-rep}} - \mathrm{E}^{\mathrm{GEM^*}}_{\mathrm{pol}}\right) \tag{9}$$

Using these reference values, the RMSE of $E_{disp+ct}$ is around 0.4 kcal mol$^{-1}$ for both GEM models. These calculated $E_{disp+ct}$ results can also be compared with respect to (w.r.t) SAPT(DFT) such as:

$$E^{ref}_{disp+ct}[SAPT(DFT)] = E^{[CCSD(T)]}_{bind} - \left(E^{SAPT(DFT)}_{Coulomb} - E^{SAPT(DFT)}_{exch-rep} - E^{SAPT(DFT)}_{pol}\right) \tag{10}$$

In this case, $E_{disp+ct}$ from GEM is closer to the *ab initio* reference with an RMSE of 0.32 kcal mol$^{-1}$ compared to 0.96 kcal mol$^{-1}$ for GEM*. That is, both $E_{exch-rep}$ and $E_{disp+ct}$ show significant improvement in the agreement with respect to the *ab initio* reference.

Interestingly, the calculated binding energy differences for both GEM models show an RMSE of 0.38 kcal mol$^{-1}$ w.r.t the CCSD(T) reference. However, as noted above, the RMSE for the individual exchange–repulsion and dispersion+charge transfer terms from the original GEM* model are around 1 kcal mol$^{-1}$ wrt SAPT(DFT) and in opposite directions. Thus, GEM* shows a compensation of errors between these two terms. Conversely, GEM is able to accurately describe both attractive and repulsive interactions for the reference Smith dimers with a better agreement for each contribution (see also Figure S1).

## C. From water dimers to water clusters: assessment of the anisotropy and many-body effects with GEM in the gas phase

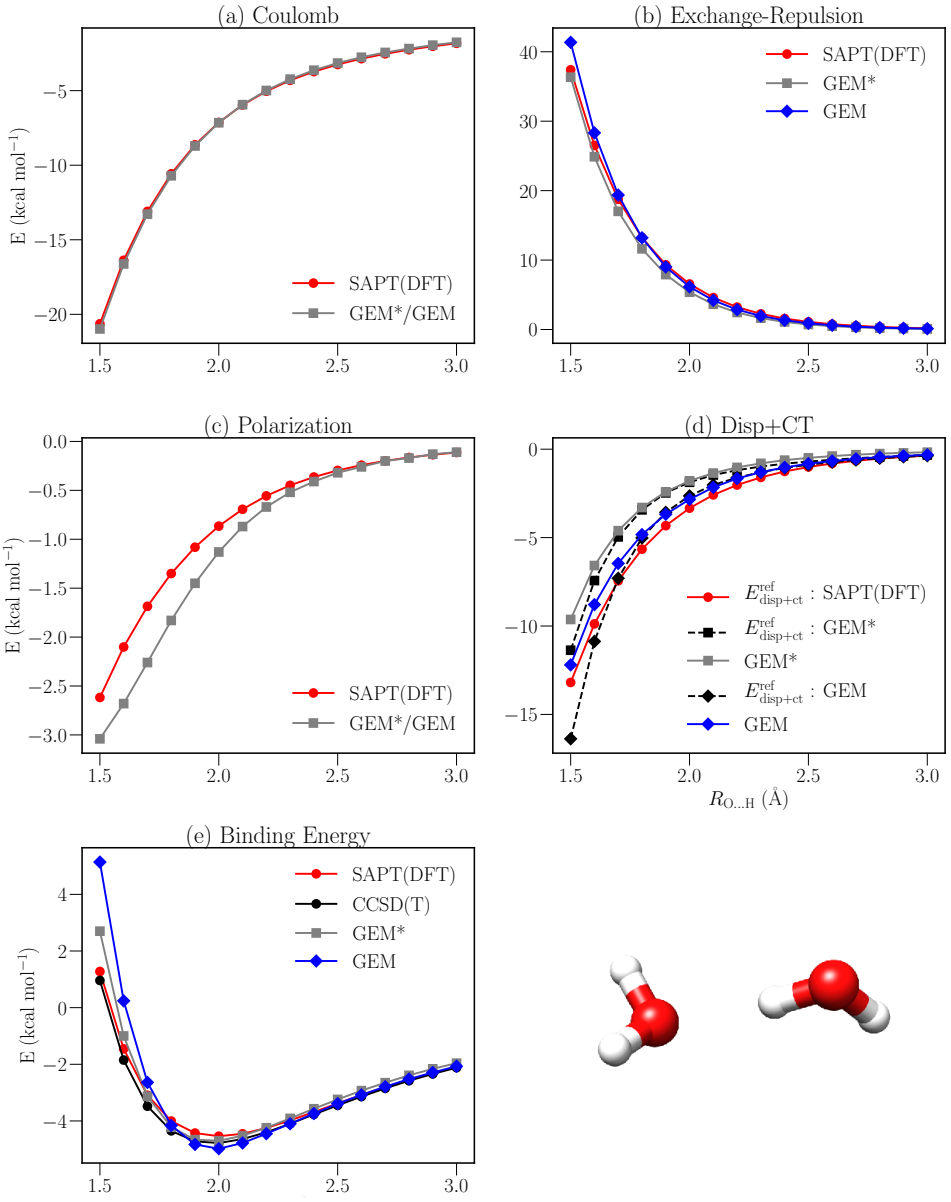



FIG. 3: Energy components computed with GEM* and GEM for the linear scan of the water dimer. For computational details of the three disp+ct reference curves see equations (8), (9) and (10) in the main text.

We next assess the binding energy and individual components on several water dimers to validate the new fitted parameters of GEM. Figure 3 shows the results for the linear scan of the canonical water dimer along the O...H axis. We observe that GEM$^*$/GEM reproduce the Coulomb energy ($E_{Coulomb}$) accurately for all distance of separation from 1.5 to 3.5 Å (RMSE of 0.14 kcal mol$^{-1}$). For $E_{exch-rep}$, GEM slightly overestimates this component at short range, but agrees with SAPT(DFT) at the equilibrium distance (1.9 Å) and at long range, while GEM$^*$ underestimates $E_{exch-rep}$ at these ranges.

Similar to $E_{Coulomb}$, $E_{pol}$ is the same for both GEM$^*$ and GEM with a RMSE of 0.29 kcal mol$^{-1}$ w.r.t SAPT(DFT). We notice that GEM$^*$/GEM overestimates (in absolute value) $E_{pol}$ by about $0.4 - 0.5$ kcal mol$^{-1}$ until 2.0 Å and then agrees asymptotically at long range with SAPT(DFT). It has been demonstrated that adding higher-order effects is essential for a better description of $E_{pol}$ at short range, and reduces error about $0.4 - 1.0$ kcal mol$^{-1}$. Thus, this small overestimation could be considered favorable for the GEM water model.[33]

For $E_{disp+ct}$, we observe that both GEM models agree with their respective $E^{ref}_{disp+ct}$. However, GEM is much closer to SAPT(DFT) than GEM$^*$, with a RMSE of 0.56 versus 1.72 kcal mol$^{-1}$. This shows that the new disp+ct formulation from GEM is able to better describe the *ab initio* reference at short and long range. Finally, GEM is also able to reproduce very well the binding energy from CCSD(T) at the equilibrium and long range.

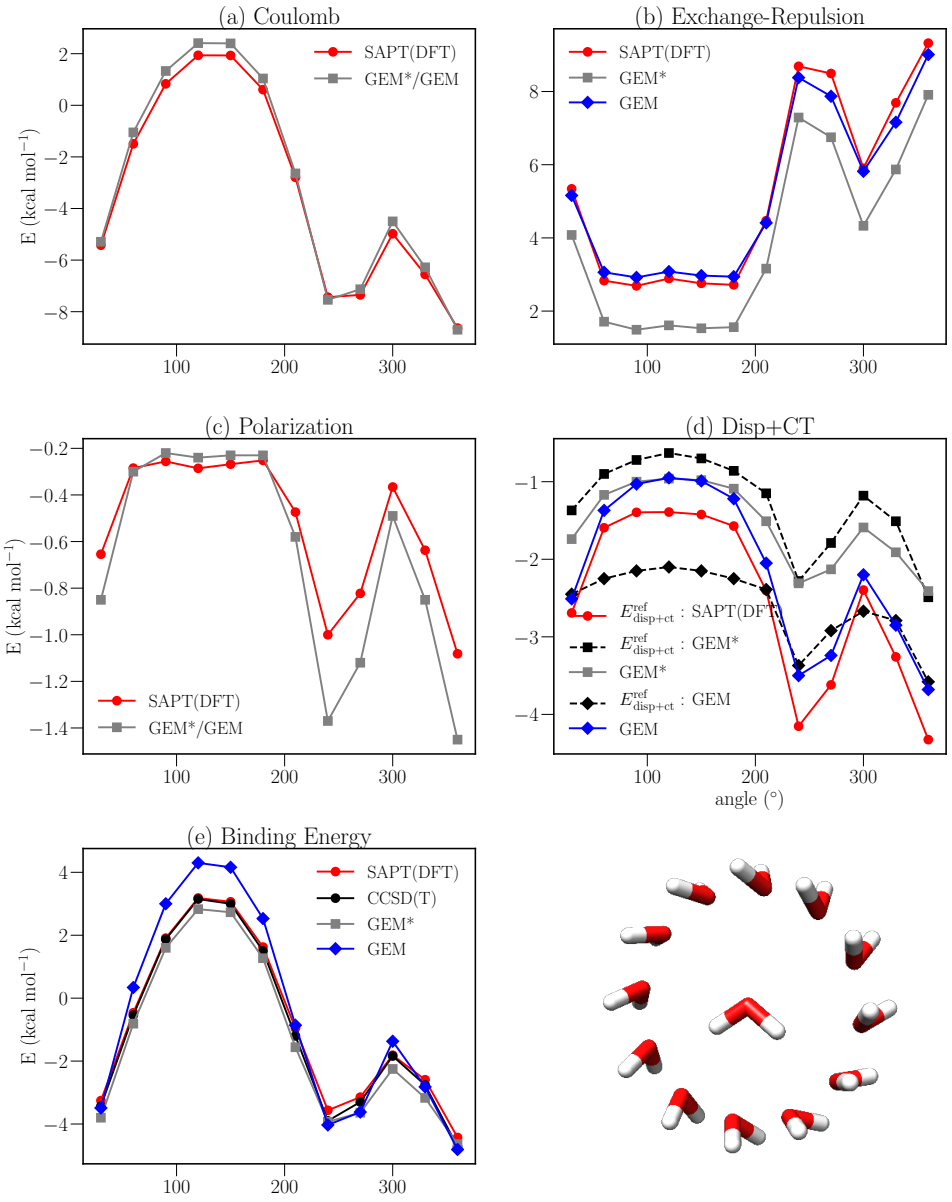



FIG. 4: Energy components computed with GEM* and GEM for the angular scan of the canonical water dimer (Y-axis) at the fixed equilibrium distance O..H of 1.9 Å. For computational details of the three disp+ct reference curves see equations (8), (9) and (10) in the main text.

The new GEM parametrization is further tested with a 3D angular scan of the water dimer. We have also investigated the performance of the GEM parameters around one reference water molecule. In this case, we consider the O...H interacting water dimer at the equilibrium distance (1.9 Å), one water molecule is kept fixed while the other is rotated around the X,Y, and Z axis respectively, from 30 to 360°. Figure 4 shows the results from the angular scan around the Y-axis. It can be seen that $E_{Coulomb}$ and $E_{pol}$ from GEM*/GEM are in good agreement with SAPT(DFT), except for $E_{pol}$ from 230 to 360° with a small error of 0.2 kcal mol$^{-1}$.

For $E_{exch-rep}$, GEM agrees remarkably well with the SAPT(DFT) reference for all angular points, with a RMSE of 0.31 kcal mol$^{-1}$ compared with a significant underestimation of GEM* for all the points, resulting in an RMSE of 1.42 kcal mol$^{-1}$. Conversely, the $E_{exch-rep}$ term calculated with GEM* overestimates the interaction energy, whereas GEM provides a better agreement with SAPT(DFT). Thus, as observed with the Smith dimer training set, GEM* benefits from a compensation of errors between the $E_{exch-rep}$ and $E_{disp+ct}$ terms, while GEM provides a better reproduction of the individual terms.

Regarding $E_{bind}$, the RMSE from GEM (0.72 kcal mol$^{-1}$) is about two-fold greater than the RMSE calculated with with GEM* (0.30 kcal mol$^{-1}$). However, as can be seen from Figure 4 GEM shows good agreement with CCSD(T) for angles $> 200°$ and $< 100°$, with a significant overestimation of $E_{bind}$ for two angles. Thus, the new disp+ct function appears to have issues reproducing the region where the lone pairs of the oxygens from the two waters interact. These errors affect the total RMSE of GEM due to lack of error compensation for certain angles. Overall, GEM separates each contribution and also captures anisotropy effects with better accuracy than GEM* (see also Figures S2 and S3).

We finally verify the transferability of the new GEM parameters on increasing size of water clusters in the gas phase. Figure 5 shows the results for trimers, tetramers,

pentamers and hexamers using the optimized geometries from this work compared to previously reported *ab initio* reference.[45] GEM considerably improves $E_{bind}$ for all water clusters except for the trimers, where the RMSE (1.05 kcal mol$^{-1}$) is slightly larger than the one computed with GEM* (0.56 kcal mol$^{-1}$). However, GEM reduces the error about 2, 3 and 4 kcal mol$^{-1}$ for the tetramers, pentamers and hexamers, respectively. The n-body effects are then better described with the new features of GEM developed in this work.

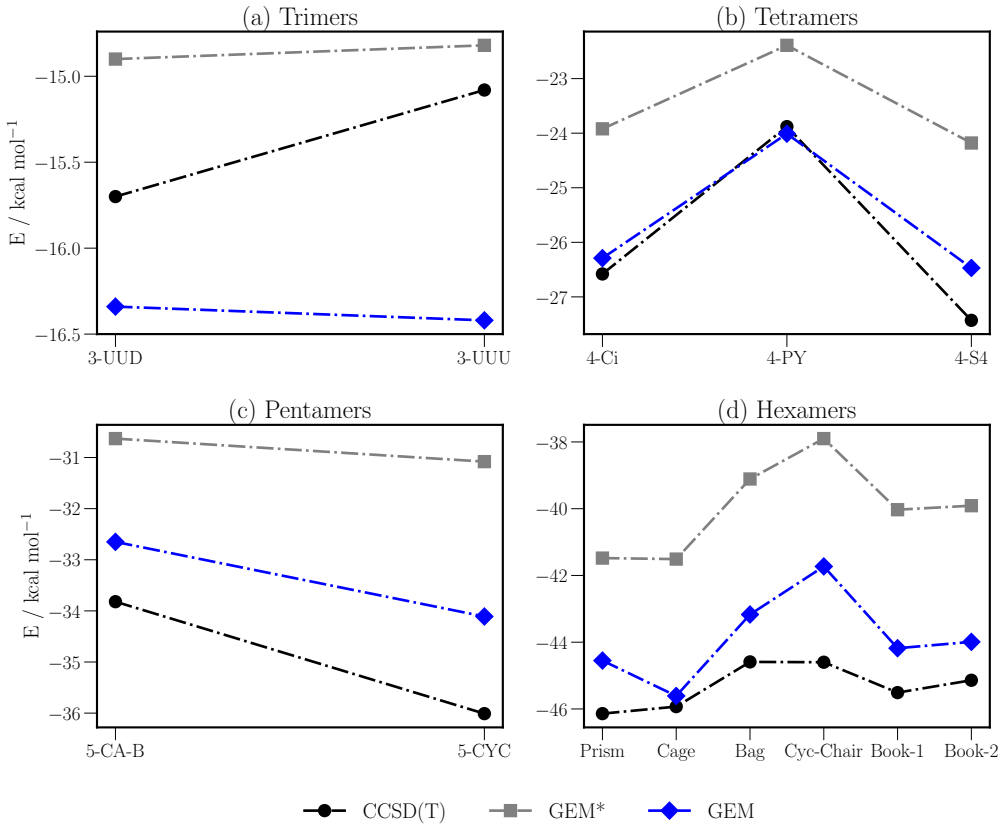
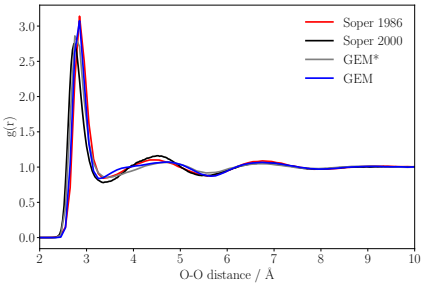
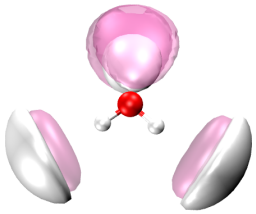



FIG. 5: Comparison of binding energies computed with GEM* and GEM with the CCSD(T)/CBS[45] for the water clusters (trimers, tetramers, pentamers and hexamers).

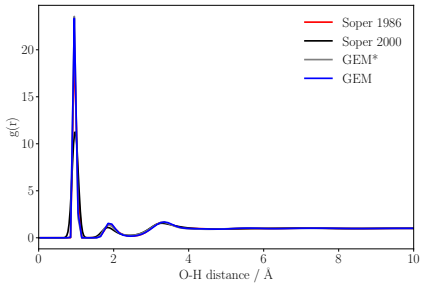
## D. Better water condensed phase properties computed with GEM




(a) RDF (O-O)

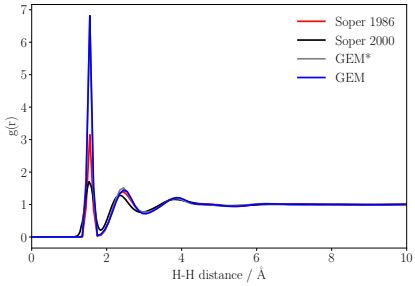


(b) SDF (xy)

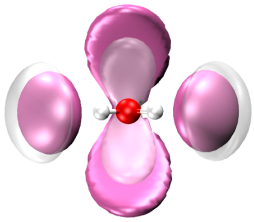



(c) RDF (O-H)

(d) SDF (xz)



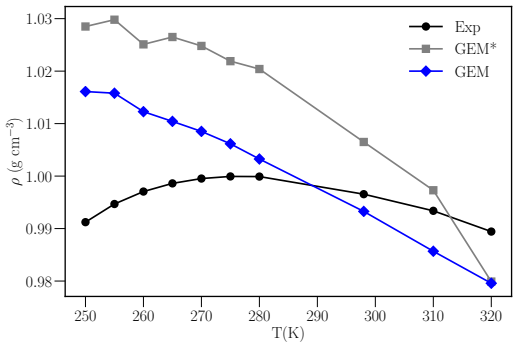

(e) RDF (H-H)



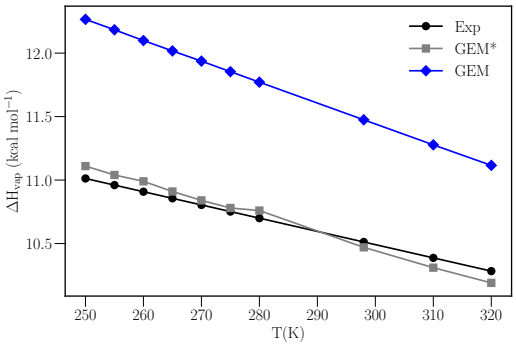
(f) SDF (yz)

FIG. 6: Comparison of the Radial Distribution Functions (RDF) with experimental data[46,47] at 300K (left panel). And the Spatial Distribution Functions (SDF) in three directions computed with GEM. The pink isosurfaces represent the oxygen atom and its lone pairs while the white isosurfaces represents the hydrogen atoms (right panel).

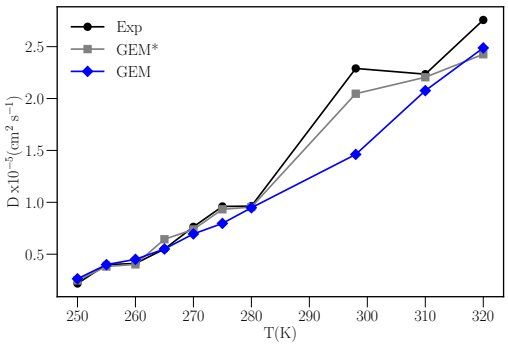
Lastly, we performed MD simulations of a box composed of 512 water molecules for 20ns at a range of temperatures. Figure 6 represents the structural properties computed with GEM at 300K. The radial distribution function (RDF) for the O-O pair atom is better described with GEM than GEM$^*$, with the first and second shells being closer to the experiment. For the two other pairs of atom, RDF(O-H) and RDF(H-H), GEM is in very good agreement with experiment. We have also displayed the spatial distribution functions in three different orientations in Figure 6. The O...H and H...O interactions between two water molecules are respectively represented by pink and white isosurfaces. We observe that one water molecule interacts effectively with four other molecules. GEM reproduces then correctly the water arrangement in the liquid phase.




(a) Density




(b) Enthalpy of vaporization




(c) Self-Diffusion Coefficient

FIG. 7: Condensed phases properties computed with GEM* and GEM compared to experiment.[48]

Moreover, we have computed three condensed phase properties such as the density ($\rho$), enthalpy of vaporization ($\Delta H_{\text{vap}}$) and the self-diffusion coefficient ($D$) at different temperatures from 250K to 320K (Figure 7). At 298K, the density computed with GEM (0.993) is closer to the experimental value (0.997) than GEM$^*$ (1.0065). The density at lower temperatures is slightly overestimated (about 0.04 g cm$^{-3}$), which could be explained to the absence of nuclear quantum effects (NQE).[49] Despite the small underestimation at higher temperatures (above 300K), GEM reduces the error of the calculated densities and is in closer agreement with the experiment than GEM$^*$.

However, GEM overestimates the enthalpy of vaporization by about 1 kcal mol$^{-1}$ while GEM$^*$ is closer to experiment. Thus, GEM is then more affected by the NQE than GEM$^*$ for $\Delta H_{\text{vap}}$. We have discussed above that the repulsion and disp+ct contributions from GEM$^*$ are compensating each other, which could explain the better agreement with the experiment. Or, it has been shown that computed $\Delta H_{\text{vap}}$ with polarizable force fields is higher (by about 0.5 to 1 kcal mol$^{-1}$) than the experiment.[50,51] The explicit inclusion of NQE is required to reproduce experimental data. GEM, which lacks of error compensation and NQE, is reproducing then the expected computed values of $\Delta H_{\text{vap}}$. Finally, at 298K, the self-diffusion coefficient from GEM is 0.5 x 10$^{-5}$ cm$^2$ s$^{-1}$ smaller than GEM$^*$, but overall GEM shows very good agreement with experiment for the full temperature range.

## IV. CONCLUSION

We have shown that the separation of the Coulomb and exchange-repulsion contributions using distinct frozen molecular densities, and the implementation of a modified and simpler (i.e no use of explicit description of lone pairs) dispersion functional inspired from the SIBFA polFF in the gem.pmemd code have significantly improved the description of attractive and repulsive interactions in the GEM water model. We

have tested this new parametrization on a set of different water dimers showing that each contribution from GEM reproduces accurately the SAPT(DFT) results, as well as the CCSD(T) binding energy. Whereas, GEM* agrees with CCSD(T) thanks to a compensation of error between the exchange-repulsion and disp+ct contributions. GEM is also able to better describe the anisotropy and n-body effects as seen with the 3D angle scans of the water dimer and selected water oligomers. The densities and diffusion coefficients at different temperatures calculated with GEM are in better agreement with the experiment compared with GEM*, whereas an overestimation of the enthalpy of vaporization is observed likely due to the lack of explicit NQE. We have shown here that the inclusion of charge-transfer and higher n-body effects in the proposed disp+ct functional form is able to correctly compute energies in the gas phase as well as the structural and thermodynamic properties in the condensed phase with acceptable errors. Finally, this work paves the way to reach the full separability of GEM, the charge-transfer will be implemented following the procedure as done in this work i.e modifying the charge-transfer formulation from SIBFA. Doing so, we expect even better accuracy for the condensed phase properties.[35]

## ACKNOWLEDGMENTS

The authors thank Prof. Nohad Gresh (Sorbonne Université) for the initial parameter set derived from SAPT(DFT) for $E_{disp+ct}^{GEM}$. This work was funded by NIH R01GM108583. Computational time was provided by the University of North Texas CASCaMs CRUNTCh3 high-performance cluster partially supported by NSF grant CHE-1531468 and XSEDE supported by project TG-CHE160044.

## AUTHOR DECLARATIONS

The authors have no conflicts to disclose. The data that support the findings of this study are available from the corresponding author upon reasonable request.

## SUPPORTING INFORMATION

Parametrization details, parameters, individual component energy and binding energy results for dimers, average density, energy and temperature plots for the bulk simulations are provided in the SI.

## REFERENCES

[1]G. A. Cisneros, K. T. Wikfeldt, L. Ojamae, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero, and F. Paesani, "Modeling molecular interactions in water: From pairwise to many-body potential energy functions," Chem. Rev. **116**, 7501–7528 (2016).

[2]P. Ren and J. W. Ponder, "Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation," J. Phys. Chem. B **107**, 5933–5947 (2003).

[3]C. Liu, J.-P. Piquemal, and P. Ren, "Amoeba+ classical potential for modeling molecular interactions," J. Chem. Theo. Comput. **15**, 4122–4139 (2019).

[4]N. Gresh, G. A. Cisneros, T. A. Darden, and J.-P. Piquemal, "Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy," J. Chem. Theory Comput. **3**, 1960–1986 (2007).

[5]P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, "An effective fragment method for modeling solvent effects in quantum mechanical calculations," The Journal of chemical physics **105**, 1968–1986 (1996).

[6]W. Xie and J. Gao, "Design of a next generation force field: the x-pol potential," Journal of chemical theory and computation **3**, 1890–1900 (2007).

[7]W. Xie, M. Orozco, D. G. Truhlar, and J. Gao, "X-pol potential: An electronic structure-based force field for molecular dynamics simulation of a solvated protein in water," Journal of chemical theory and computation **5**, 459–467 (2009).

[8]J. M. Hermida-Ramón, S. Brdarski, G. Karlström, and U. Berg, "Inter-and intramolecular potential for the n-formylglycinamide-water system. a comparison between theoretical modeling and empirical force fields," Journal of computational chemistry **24**, 161–176 (2003).

[9]J. A. Rackers, R. R. Silva, Z. Wang, and J. W. Ponder, "A polarizable water potential derived from a model electron density," (2021), arXiv:2106.13116 [physics.chem-ph].

[10]K. Kitaura and K. Morokuma, "A new energy decomposition scheme for molecular interactions within the hartree-fock approximation," International Journal of Quantum Chemistry **10**, 325–340 (1976), https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560100211.

[11]W. J. Stevens and W. H. Fink, "Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to the water dimer," Chem. Phys. Lett. **139**, 15–22 (1987).

[12]P. S. Bagus, K. Hermann, and C. W. Bauschlicher, "A new analysis of charge transfer and polarization for ligand-metal bonding: Model studies of Al4CO and Al4NH3," J. Chem. Phys. **80**, 4378–4386 (1984).

[13]B. Jeziorski, R. Moszynski, and K. Szalewicz, "Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes." Chem. Rev. **94**, 1887–1930 (1994).

[14]T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill, "Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies," J. Chem. Phys. **140** (2014), 10.1063/1.4867135.

[15]A. Heßelmann and G. Jansen, "First-order intermolecular interaction energies from Kohn–Sham orbitals," Chem. Phys. Lett. **357**, 464–470 (2002).

[16]A. Heßelmann and G. Jansen, "Intermolecular induction and exchange-induction energies from coupled-perturbed Kohn–Sham density functional theory," Chem. Phys. Lett. **362**, 319–325 (2002).

[17]A. Heßelmann and G. Jansen, "Intermolecular dispersion energies from time-dependent density functional theory," Chem. Phys. Lett. **367**, 778–784 (2003).

[18]A. J. Misquitta and K. Szalewicz, "Intermolecular forces from asymptotically corrected density functional description of monomers," Chem. Phys. Lett. **357**, 301–306 (2002).

[19]A. J. Misquitta, B. Jeziorski, and K. Szalewicz, "Dispersion Energy from Density-Functional Theory Description of Monomers," Phys. Rev. Lett. **91**, 033201 (2003).

[20]A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, "Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations," J. Chem. Phys. **123**, 214103

(2005).

[21]A. J. Misquitta, "Charge-transfer from regularized symmetry-adapted perturbation theory," J. Chem. Theory Comput. **9**, 5313–5326 (2013).

[22]G. A. Cisneros, J. P. Piquemal, and T. A. Darden, "Intermolecular electrostatic energies using density fitting," J. Chem. Phys. **123**, 184101 (2005).

[23]S. Boys and I. Shavit, "A fundamental calculation of the energy surface for the system of three hydrogen atoms," NTIS, Springfield, VA, AD212985 (1959).

[24]B. Dunlap, J. Connolly, and J. Sabin, "On first-row diatomic molecules and local density models," The Journal of Chemical Physics **71**, 4993–4999 (1979).

[25]A. M. Köster, P. Calaminici, Z. Gómez, and U. Reveles, "Density functional theory calculation of transition metal clusters," in *Reviews of Modern Quantum Chemistry* (World Scientific, 2002) pp. 1439–1475.

[26]G. A. Cisneros, D. M. Elking, J.-P. Piquemal, and T. A. Darden, "Numerical fitting of molecular properties to hermite gaussians," J. Phys. Chem. A **111**, 12049–12056 (2007).

[27]H. Gökcan, E. Kratz, T. A. Darden, J.-P. Piquemal, and G. A. Cisneros, "QM/MM Simulations with the Gaussian Electrostatic Model: A Density-based Polarizable Potential," J. Phys. Chem. Lett. **9**, 3062–3067 (2018).

[28]J. Nochebuena, S. Naseem-Khan, and G. A. Cisneros, "Development and application of quantum mechanics/molecular mechanics methods with advanced polarizable potentials," Wiley Interdiscip. Rev. Comput. Mol. Sci. **11**, 1–27 (2021).

[29]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. Pedersen, "A smooth particle mesh ewald: an n· log (n) method for ewald sums in large systems," J. Chem. Phys **103**, 8577–8593 (1995).

[30]D. York and W. Yang, "The fast fourier poisson method for calculating ewald sums," The Journal of Chemical Physics **101**, 3298–3300 (1994).

[31]R. E. Duke and G. A. Cisneros, "Ewald-based methods for Gaussian integral evaluation: application to a new parameterization of GEM*," J. Mol. Model. **25**, 307 (2019).

[32]G. A. Cisneros, J.-P. Piquemal, and T. A. Darden, "Generalization of the Gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods," J. Chem. Phys. **125**, 184101 (2006).

[33]S. Naseem-Khan, N. Gresh, A. J. Misquitta, and J. P. Piquemal, "An assessment of SAPT and supermolecular EDAs approaches for the development of separable and polarizable force fields," J. Chem. Theory Comput. **17**, 2759–2774 (2021), 2008.01436.

[34]S. P. Veccham, J. Lee, Y. Mao, P. R. Horn, and M. Head-Gordon, "A non-perturbative pairwise-additive analysis of charge transfer contributions to intermolecular interaction energies," Phys. Chem. Chem. Phys. **23**, 928–943 (2021), arXiv:2011.04918.

[35]S. Naseem-Khan, L. Lagardère, G. A. Cisneros, P. Ren, N. Gresh, and J.-P. Piquemal, "Molecular dynamics with the sibfa many-body polarizable force field: From symmetry adapted perturbation theory to condensed phase properties," To be published.

[36]L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder, and J.-P. Piquemal, "Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields," Chem. Sci. **9**, 956–972 (2018).

[37]R. E. Duke, O. N. Starovoytov, J.-P. Piquemal, G. Andrés, and A. Cisneros, "GEM*: A Molecular Electronic Density-Based Force Field for Molecular Dynamics Simulations," J. Chem. Theory Comput **10**, 1361–1365 (2014).

[38]J.-P. Piquemal and G. A. Cisneros, *Many-Body Eff. Electrost. Biomol.*, Vol. 7 (Pan Standford Publishing, 2015) pp. 978–981.

[39]V. Babin, G. R. Medders, and F. Paesani, "Toward a universal water model: First principles simulations from the dimer to the liquid phase," J. Phys. Chem. Lett. **3**, 3765–3769 (2012), arXiv:1210.7022.

[40]V. Babin, C. Leforestier, and F. Paesani, "Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient," J. Chem. Theory Comput. **9**, 5395–5403 (2013).

[41]B. J. Smith, D. J. Swanton, J. A. Pople, H. F. Schaefer, and L. Radom, "Transition structures for the interchange of hydrogen atoms within the water dimer," J. Chem. Phys. **92**, 1240–1247 (1990).

[42]G. A. Cisneros, "Application of gaussian electrostatic model (GEM) distributed multipoles in the AMOEBA force field," J. Chem. Theory Comput. **8**, 5072–5080 (2012).

[43]H. Torabifard, O. N. Starovoytov, P. Ren, and G. A. Cisneros, "Development of an AMOEBA water model using GEM distributed multipoles," Theor. Chem. Acc. **134**, 1–10 (2015).

[44]J. P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh, and T. A. Darden, "Towards a force field based on density fitting," J. Chem. Phys. **124** (2006), 10.1063/1.2173256, arXiv:NIHMS150003.

[45]B. Temelso, K. A. Archer, and G. C. Shields, "Benchmark structures and binding energies of small water clusters with anharmonicity corrections," J. Phys. Chem. A **115**, 12034–12046 (2011).

[46]A. K. Soper and M. G. Phillips, "A new determination of the structure of water at 25°C," Chem. Phys. **107**, 47–60 (1986).

[47]A. K. Soper, "The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa," Chem. Phys. **258**, 121–137 (2000).

[48]W. Wagner and A. Pruß, "The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use," J. Phys. Chem. Ref. Data **31**, 387–535 (2002).

[49]S. K. Reddy, S. C. Straight, P. Bajaj, C. Huy Pham, M. Riera, D. R. Moberg, M. A. Morales, C. Knight, A. W. Götz, and F. Paesani, "On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice," J. Chem. Phys. **145** (2016), 10.1063/1.4967719, arXiv:1609.02884.

[50]G. S. Fanourgakis, G. K. Schenter, and S. S. Xantheas, "A quantitative account of quantum effects in liquid water," J. Chem. Phys **125**, 141102 (2006).

[51]F. Paesani, S. Iuchi, and G. A. Voth, "Quantum effects in liquid water from an ab initio -based polarizable force field," J. Chem. Phys. **127**, 074506 (2007).