Published in final edited form as:
J Comput Chem. 2011 April 15; 32(5): 967–977. doi:10.1002/jcc.21681.

Investigation of Multipole Electrostatics in Hydration Free Energy Calculations

Yue Shi[1], **Chuanjie Wu**[2], **Jay W. Ponder**[2], and **Pengyu Ren**[1,*]

[1]Department of Biomedical Engineering, The University of Texas at Austin, TX 78712

[2]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110

Abstract

Hydration free energy (HFE) is generally used for evaluating molecular solubility, which is an important property for pharmaceutical and chemical engineering processes. Accurately predicting HFE is also recognized as one fundamental capability of molecular mechanics force field. Here we present a systematic investigation on HFE calculations with AMOEBA polarizable force field at various parameterization and simulation conditions. The HFEs of seven small organic molecules have been obtained alchemically using the Bennett Acceptance Ratio (BAR) method. We have compared two approaches to derive the atomic multipoles from quantum mechanical (QM) calculations: one directly from the new distributed multipole analysis (DMA) and the other involving fitting to the electrostatic potential around the molecules. Wave functions solved at the MP2 level with four basis sets (6-311G*, 6-311++G(2d,2p), cc-pVTZ, and aug-cc-pVTZ) are used to derive the atomic multipoles. HFEs from all four basis sets show a reasonable agreement with experimental data (root mean square error 0.63 kcal/mol for aug-ccpVTZ). We conclude that aug-cc-pVTZ gives the best performance when used with AMOEBA, and 6-311++G(2d,2p) is comparable but more efficient for larger systems. The results suggest that the inclusion of diffuse basis functions is important for capturing intermolecular interactions. The effect of long-range correction to van der Waals interaction on the hydration free energies is about 0.1 kcal/mol when the cutoff is 12Å, and increases linearly with the number of atoms in the solute/ligand. In addition, we also discussed the results from a hybrid approach that combines polarizable solute with fixed-charge water in the hydration free energy calculation.

Introduction

Hydration of small molecules is an important phenomenon in many chemical and biochemical processes. The ability to accurately calculate the hydration free energy is critical in the force field development and the application of molecular modeling to molecular design and drug discovery. For example, hydration free energy is one of the components in determining the binding affinity of a ligand to its receptor.[1] Since hydration free energy is a sensitive measure of the interaction between a solute and water, it has been commonly used to assess the accuracy of physical models, such as the quality of partial charges and implicit solvent models, by comparing with the experimental hydration free energies of a wide range of organic molecules.[2–6]

Solvent effects can be computationally investigated with implicit and explicit methods.[7] The implicit solvent approaches including Poisson Boltzmann (PB) and Generalized Born (GB)

*Corresponding Author Tel: (512) 232-1832 Fax: (512)-471-0616 pren@mail.utexas.edu.

methods. Studies with implicit models usually focus on the evaluation of charge parameters, [6] and improvement of the polar/nonpolar solvation models.[7–11] Although implicit-solvent methods are computationally efficient, there are still notable limitations. The continuum approximation ignores finite size effect of water as well as tightly bound individual water molecules.[12] It is unable to distinguish positively and negatively charged molecules of the same size[13] unless they are specially treated.[14] Extensive parameterizations against experimental data and explicit-solvent simulations are necessary.[8,9,15] The alternative in treating solvent is through explicit representation of solvent molecules. With recent computational and methodological advancement, alchemical approaches such as thermodynamic integration (TI)[16,17] and free energy perturbation (FEP)[18,19] are increasingly used to compute precise hydration free energies of amino acid side chain analogs and small molecules in explicit solvent.[4,5,7,20–23] In recent studies that cover a wide range of organic molecules,[4,5,19] the reported root mean square error (RMSE) of predicted hydration free energies from experimental values is slightly over 1 kcal/mol using fixed-charge models, which is somewhat similar to that using implicit solvent approach.[6] It was suggested that treatment of polar and poly-functional molecules need to be improved.[4] Hydration free energy calculations have also been utilized in the parameterization of a Drude oscillator based polarizable force field.[24]

In an effort to develop accurate atomic force field for molecular interactions, polarizable multipole has been introduced in AMOEBA to account for atomic charge distribution. While it is the advantage of a polarizable force field that the charge distribution can be derived from high-level gas-phase QM calculations of model compounds, there are various approaches available to obtain the atomic multipole expansion, such as distributed multipole analysis (DMA),[25,26] atoms-in-molecules (AIM),[27] cumulative atomic multipole moments (CAMM),[28,29] and electrostatic potential fitting.[30] Previously we have shown that DMA derived multipoles perform well in modeling both the inter- and intramolecular electrostatic interactions.[31] In the original DMA the multipole moments of the associated charge distribution are represented by a multipole expansion at a set of atoms (or selected sites). This procedure is exact and efficient, and gives an excellent representation of the molecular charge distribution with up to quadrupole moments at each atom.[26] For large basis sets with diffuse functions, however, DMA yields seemingly “unphysical” atomic multipoles. In addition, the multipole values vary significantly with the size of the basis sets used.[32] Recently, a modified DMA procedure[32] was proposed to overcome these problems by using numerical quadrature for the diffuse functions. While the new DMA produces atomic multipoles converge with improved basis-sets, the magnitudes of monopole, dipole and quadrupoles seem very different from those given by original DMA. At very short range, such as hydrogen-bonding distance, the different distribution of multipoles will lead to different interaction energies, even though the molecular moments are always well reproduced.

In this work, we applied a polarizable multipole-based electrostatic model to calculate hydration free energies of small ligands. The main purpose is to systematically investigate different approaches for derivation of atomic multipoles and the effect of basis sets on the hydration free energy. Beside the new DMA procedure, we evaluated an alternative approach to deal with the diffuse function issue in the original DMA: the atomic multipoles are derived from the original DMA with small basis set and then optimized by fitting to the electrostatic potential around the molecule. In addition, we examined the treatment of long-range correction to vdW interactions with finite cutoffs in the hydration free energy calculations. Finally, we explored the possibility of combining solutes with polarizable multipoles and solvent calculated through traditional molecular mechanics (PM/MM) in a hybrid model.

Computational Methods

Molecular Systems

The free energy of hydration was computed using molecular dynamics and Bennett Acceptance Ratio (BAR) method. Seven organic compounds were investigated: ethylbenzene, *p*-cresol, isopropanol, imidazole, methylethylsulfide, acetic acid and ethanol. This set of molecules represents the common chemical functional groups in bimolecular systems and drug-like compounds, including alkyl, benzyl, phenol, hydroxyl, imidazole, carboxyl, and sulfide groups. The solvent was modeled using both the AMOEBA[33] polarizable water and a TIP3P-like fixed-charge water model that we developed here to use with polarizable solutes.

The Polarizable Multipole Force Field

The molecular dynamics simulations for various systems have been performed using AMOEBA polarizable force field.[31,33–37] In AMOEBA force field, each atom possesses permanent charge, dipole and quadrupole moments. Moreover, the electronic polarization effects are included, using a self-consistent dipole induction procedure.[31] Repulsion-dispersion interactions between pairs of non-bonded atoms are represented by a buffered 14-7 potential.[38] Parameters for all the organic compounds were taken from the AMOEBA force field for small molecules (available in TINKER 5.1 molecular modeling package[39]), except atomic multipoles which have been varied in the current study. New atomic multipole parameters of these molecules were derived from *ab initio* calculations with different basis sets using the GDMA program.[40] Early version of GDMA (v1.x) offers the original DMA while the latest GDMA (v2.2) now implements the modified DMA to address diffuse function and basis set dependence issues. In GMDA v2.2, by setting “Switch 0” and “Radius H 0.65”, one can also access the original DMA procedure. The default behavior is to set “Switch 4”, which invokes the modified DMA protocol. The structure of each molecule was optimized quantum mechanically at the level of HF/6-31G* using Gaussian 03.[41] Single point calculations were subsequently performed on the optimized geometry using MP2 method with four basis sets: (i) 6-311G**, (ii) 6-311++G(2d,2p), (iii) cc-pVTZ, and (iv) aug-cc-pVTZ, for comparison.

We tested two procedures to obtain the atomic multipoles for each basis set. One is to use the original DMA to derive the permanent atomic multipole using the 6-31G* basis set, and then optimize the multipoles to the electrostatic potential (ESP) derived from the above four basis sets. We refer to this procedure as “original-fit” in the discussion below. The optimization was done by using the POTENTIAL program in TINKER; the atomic monopoles were fixed at those from 6-31G* and only dipole and quadrupole moments were allowed to vary. This ESP fitting procedure also gave us a consistent set of multipoles across the basis sets by perturbing the dipole and quadrupole moments from a lower basis set. In the second approach, which we call “DMA2”, the new DMA procedure was used to compute the permanent multipole moments directly at each basis set. Multipoles for the same atom type were then averaged and optimized to the QM electrostatic potential (ESP) from the same basis set. GDMA 2.2 software package was used. The hydrogen atomic radius ratio was set to 0.31 in both original and modified DMA procedures. Also in both procedures, the atomic multipoles were optimized to the QM electrostatic potential until the root mean square gradient difference was smaller than 0.1, the grid offset 1.0 Å from the vdW surface of each atom. For alcohols, the atomic quadrupole components of the hydroxyl group (O and H atoms) were reduced by a constant factor similar to that previously applied to water.[33] This reduction led to better agreement with both *ab initio* water dimer geometry and experimental liquid properties.

Hydration Free Energy Calculation

Alchemical transformation was performed to calculation the hydration free energy of the small organic compounds (Figure 1). The hydration free energy of a compound was calculated from:

$$\Delta A_{\text{hydration}} = -\Delta A_{\text{discharging(aq.)}} - \Delta A_{\text{decoupling(aq.)}} - \Delta A_{\text{recharging(vac.)}} \quad (1)$$

where $\Delta A_{recharging(vac.)}$ is the free energy change due to growing the intramolecular electrostatic interactions in vacuum; $\Delta A_{discharging(aq.)}$ results from annihilating electrostatic interactions both within the solute itself and between solute and solvent. $\Delta A_{decoupling(aq.)}$ represents the free energy change by turning off vdW interactions between the solute and its environment using a soft-core buffered 14-7 potential:

$$U_{ij} = \lambda^5 \varepsilon_{ij} \frac{1.07^7}{\left[0.7(1-\lambda)^2 + \left(\rho_{ij} + 0.07\right)^7\right]} \left(\frac{1.12}{0.7(1-\lambda)^2 + \rho_{ij}^7 + 0.12} - 2 \right) \quad (2)$$

where ε is the well depth and λ is the scaling factor. By varying λ from 1 to 0, the vdW interactions between ligand and its environment are gradually turned off. $\rho_{ij} = R_{ij}/R_{ij}^0$ with R_{ij} as the actual separation between atoms *i* and *j* and R_{ij}^0 the minimum energy distance parameter.

The scaling factor λ is also introduced to the long-range correction (LRC):

$$U_{LRC} = \frac{1}{2} \sum_i \sum_j \left\{ \left(N_i \times N_j\right) - \left(1 - \lambda^5\right) \left[L_i \times N_j + (N_i - L_i) \times L_j \right] \right\} \frac{4\pi}{V} \int_{r_c}^{\infty} U_{ij}(r)\, r^2 dr \quad (3)$$

where r_c is the cutoff radius, *V* is the system volume, *U* is the soft-core potential, N_i and N_j are the total number of atoms of type *i* and *j* in the system, L_i and L_j are the number of soft-core interaction sites of atom type *i* and *j* in the solute being annihilated. In this way, U_{LRC} is scaled to the same extent as soft-core interactions within the cutoff radius. When λ=0, the long-range vdW interactions between ligand and its environment is removed. Polynomial tapering function applied from 0.9 r_c to r_c reduces vdW interactions to zero at the distances beyond r_c and maintains smooth atomic forces. The reduced energy and virial value in the tapering range are also included in the LRC correction.

To compute the free energy changes between neighboring states (λ_i and λ_{i+1}), the Bennett Acceptance Ratio (BAR) method[42] was utilized:

$$\Delta A(j)_{\lambda_i \to \lambda_{i+1}} = ln \frac{<1/[1+exp((U_{\lambda_i} - U_{\lambda_{i+1}} + C)/RT)]>_{\lambda_{i+1}}}{<1/[1+exp((U_{\lambda_{i+1}} - U_{\lambda_i} - C)/RT)]>_{\lambda_i}} + C \quad (4)$$

where C is given by:

$$C = \Delta A(j-1)\, \lambda_i \to \lambda_{i+1} \quad (5)$$

and *j* is the iteration index. Here, $U_{\lambda i}$ is the potential energy of the system evaluated using the parameters from λ_i, and <> is the ensemble average. ΔA is solved iteratively until the value of $(\Delta A(j) - \Delta A(j-1)) < 0.01$ kcal/mol. The statistical error in the free energy change between two steps was computed from Eq. 10 in Reference 42. The total statistical error in the solvation free energy in bulk water was computed as the sum of the errors from individual perturbation steps.

Computational Details

All liquid simulations were performed using PMEMD and SANDER in AMBER 9.[43] TINKER was used to prepare the initial systems and for the gas phase re-charging simulations in vacuum. An automation script to set up the system with all perturbation steps as well as the post free energy analysis procedures is available online.

To carry out the perturbations in bulk water, the solute molecule was placed at the origin of a pre-equilibrated periodic box of solvent containing 800 water molecules in a cube with a 28.78 Å dimension on each side. The system was then equilibrated for 50 ps using NPT ensemble at 298 K. The last frame of the simulation was used as the starting point for all the intermediate states λ_i. The electrostatic interactions were decoupled in 11 steps by scaling down solute atomic multipoles and polarizabilities linearly with λ = (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0). For vdW interactions, we compared different scaling protocols (see Results and Discussion). For each value of λ, 500 ps or 1 ns of constant volume molecular dynamics was performed, with the density fixed at the NPT-average and a time step of 1 fs. The temperature was maintained at 298 K using Berendsen thermostat[44]. The vdW cutoff was set to both 9 Å and 12 Å with and without long-range correction to evaluate the cutoff and LRC effects. Long-range electrostatics for all the systems were treated using Particle Mesh Ewald (PME) summation.[45–47] The PME calculation used a 36 × 36 × 36 grid and fifth-order B-spline interpolation. The induced dipoles were iterated until the root-mean-square change was below 0.01 D/atom. Atomic coordinates of the simulation system were saved every 500 fs. The first 100 ps simulation trajectories were ignored in the free energy analysis. A tighter induced dipole convergence of 10^{-5} D/atom was used in the energy calculation for the free energy analysis.

Gas phase simulations were run on the single solute molecule only for 50 ps with a time step of 0.1 ps. The temperature was maintained at 298 K using stochastic thermostat. The induced dipoles were converged to 10^{-5} D. Atomic coordinates were collected every 100 fs. Post free energy analyses were performed on all 500 configurations. BAR was used to evaluate the free energy change between adjacent states.

Results and Discussion

Hydration Free Energy Calculation Protocol

In alchemical free energy calculations, a sufficient number of “small” perturbation steps would ensure adequate overlaps between adjacent states. However the computational cost is also an important consideration especially given the AMOEBA polarizable potential is more costly than fixed charge models. In the approach adopted in this study, the electrostatic and vdW interactions between solute and solvent were decoupled sequentially. The sampling for vdW decoupling was the most challenging step since solute and solvent molecules begin to overlap. We thus first determined an optimal distribution of intermediate states and simulation length in the vdW decoupling simulations.

Hydration free energy of ethylbenzene has been calculated using different simulation protocols with a different number of intermediate states. Figure 2 shows the hydration free energy results from three vdW decoupling schedules: (a). λ = (1.0, 0.95, 0.9, 0.85, 0.8,

0.775, 0.75, 0.725, 0.7, 0.675, 0.65, 0.625, 0.6, 0.55, 0.5, 0.4, 0.2, 0.0); (b). λ = (1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.5, 0.4, 0.2, 0.0); (c). λ = (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.2, 0.0). Simulations for all intermediate states were performed for 1ns with a 12 Å cutoff, using AMOEBA force field. The results here include the electrostatic component of the hydration free energies and the long-range correction, however both contributions are constants and do not affect the comparisons we made. It can be seen that for the decoupling paths with 18 steps (a) and 11 steps (b), hydration free energies converge after 500 ps. Standard deviations of the hydration free energy value calculated from 500 ps to 1 ns (with a 100 ps interval) simulations are 0.04 kcal/mol for protocol (a), and 0.05 kcal/mol for protocol (b). However in the 9-step (c) vdW decoupling, the hydration free energy does not show a sign of convergence even at 1ns.

We have summarized the vdW decoupling free energies calculated using the three schedules in Table 1. By further examining the free energy change at the intermediate states, it is clear that the difference between the hydration free energy from the 9-step protocol (c) and the results of the other two mostly occurs when λ is varied from 0.8 to 0.6. The free energy change of this perturbation is −4.30 and −4.32 kcal/mol calculated from protocols (a) and (b), but −3.77 kcal/mol using protocol (c). A total of 9 steps (including the boundary states) were used in protocol (a) to transform λ from 0.8 to 0.6, 5 steps in protocol (b), and 3 steps in protocol (c). The more than half a kcal/mol difference indicates the linear schedule of protocol (c) is not sufficient. In addition, the simulation results remain not converged by extending the simulation time from 500 ps to 1 ns. The difficulty in calculating the vdW decoupling free energy at λ near 0.5 is a result of large fluctuation in the soft-core vdW energy. On the other hand, all three protocols gave almost identical free energies when λ is varied from 1.0 to 0.8. The soft-core vdW potential is also greatly “smoothed” at the end point when λ approaches zero, and the contribution between λ=0.4 and 0 is nearly zero. The statistical errors in Figure 2, calculated using Bennett's formula, also show a trend of smaller statistical error with more intermediate states. Note that an excessive number of states could also lead to a large statistical error. At 500 ps, the statistical error of the calculated hydration free energies is ±0.05 kcal/mol, ±0.15 kcal/mol and ±0.25 kcal/mol for protocol (a), (b) and (c), respectively. The error due to the electrostatic decoupling is negligible and again does not affect the comparison here. With both computation expense and precision in consideration, 11 steps of 500 ps simulations were used to decouple the vdW interactions in the remaining hydration free energy calculations.

van der Waals Cutoff and Long-range Correction

Previous works showed that long-range vdW interactions between ligands and proteins can contribute to the binding affinity by more than 1 or 2 kcal/mol, with cutoff at 7.0 to 9.0 Å[48]. The long range vdW interactions beyond a “large” cutoff should be negligible. In the current study, we are computing the hydration free energy values for relatively small solute molecules. To evaluate the effect of vdW cutoffs and long-range correction, we calculated the hydration free energy values of six organic compounds using different cutoffs with and without LRC. All the other setups of the six systems remained the same in the calculations, including box sizes, atomic multipoles (derived at MP2/6-311++G(2d,2p) level using the DMA2 method).

Table 2 shows the hydration free energy results using either a 9 Å or a 12 Å cutoff for vdW interactions, with and without LRC in the vdW decoupling. Note that the LRC is a constant in the NVT simulations and thus contribute no forces on each atom at any given time. Therefore the LRC has no effect on the simulation trajectories or the electrostatic decupling free energy at the same vdW cutoff, but it does affect the vdW decoupling free energy when the solute atoms are “disappeared” in the solvent. The root mean square difference (RMSD) in the hydration free energies between the 9 Å and 12 Å cutoff simulations is 0.33 kcal/mol

without the LRC. When the LRC is included in the vdW decoupling free energy calculation, the difference is reduced to 0.17 kcal/mol. Since the electrostatic component of hydration free energy is not affected by the LRC, this error reduction was exclusively in the vdW component. For smaller molecules such as acetic acid, the results from 9 or 12 Å cutoff are essentially identical with the LRC. LRC is clearly making the results at different cutoffs more consistent. However a 9 Å with LRC can still lead to non-negligible errors for large molecules such as p-cresol. With regard to the computation cost, simulations with 12 Å cutoff are about 1.25 times slower than those using 9 Å cutoff. In our study, we chose the 12 Å cutoff, to get the most rigorous hydration free energy results.

When a 12 Å cutoff is used for the vdW interaction, the inclusion of LRC lowered the hydration free energy by 0.14 kcal/mol (RMSD). For simulations using 9 Å cutoffs the contribution of LRC was −0.33 kcal/mol. The contribution of LRC to hydration free energy is more negative than those without LRC as the vdW interactions beyond the cutoff radii are always favorable. In the remaining study, a 12 Å cutoff has been applied to the vdW interactions, with LRC included. As the LRC to hydration free energy is already comparable to the magnitude of the statistical error in our hydration free energy calculations at the 12 Å cutoff, we believe the error (comparing to an infinite cutoff) should be negligible. However, it should be noted that the LRC contribution is roughly proportional to number of the atoms of the solute molecule. In the system studied here, the average atomic LRC contribution during vdW decoupling is ~0.01 kcal/mol per atom at a 12 Å cutoff; for a molecule with 100 atoms, there would be an error of ~1 kcal/mol if the LRC decoupling is ignored. Similarly, for simulations with a 9 Å cutoff, every 37 atoms in a drug molecule will lead to an error around 1 kcal/mol if the LRC decoupling is ignored. Therefore, when a 9 Å cutoff, which is frequently used in molecular simulations, is employed, LRC decoupling term is highly recommended. In addition, it should be kept in mind that the correction is based on the assumption of isotropic environment which is appropriate for solvation in a homogenous solvent.

Electrostatic Multipoles from Different Methods and Basis Sets

We investigated different approaches to derive atomic multipoles and the effect of basis sets on hydration free energy calculations. The vdW and remaining parameters are transferred from AMOEBA. All the simulations in this section employed a 12 Å cutoff for vdW plus LRC, so that the vdW contributions to the hydration free energy are exactly the same in all comparisons.

Atomic multipoles for the small molecules were derived from QM single point calculations using the original DMA with ESP fitting (original-fit) or the new DMA2 method (see Computational Methods). Table 3 compares the hydration free energy of p-cresol computed from atomic multipoles using four different basis sets (6-311G**, 6-311++G(2d,2p), cc-pVTZ, and aug-cc-pVTZ) at the MP2 level. It can be seen that with 6-311++G(2d,2p) and cc-pVTZ basis sets, hydration free energy difference between using original-fit and DMA2 is within 0.1 kcal/mol. With aug-cc-pVTZ and 6-311G**, the difference is around 0.4 kcal/mol. The overall RMSEs are 0.69 and 0.66 kcal/mol for original-fit and DMA2 respectively. Thus for small molecules, the atomic multipoles from the new DMA2 procedure and old ESP fitting give very similar intermolecular interaction energies as reflected in the hydration free energy.

Consistent with our previous study, we found that it is necessary to scale down the atomic quadrupole moments of the hydroxyl group for alcohols. According to the work by Ren and Ponder,[33] the quadrupole moments of AMOEBA water molecules were reduced to 73% of the QM DMA values, which led to a reduction of the water-water flap angle to 57° and better reproduced a series of *ab initio* and experimental properties. From Table 4, we can see

that hydration free energy values of isopropanol without scaling poorly agree with experiment (RMSE of 1.89 kcal/mol). All basis sets overestimate the hydration free energy values by more than 1 kcal/mol. However when quadrupoles of hydroxyl groups are scaled to 70% of the DMA2 values, the RMSE significantly reduces to 0.62 kcal/mol. Same quadrupole parameters obtained with original-fit needs to be scaled to 60% to achieve similar RMSE (0.58 kcal/mol). This is due to the different distribution of multipole moments given by the new and original DMA methods, even though both theoretically produce the same molecule moments and ESP. To verify the transferability, the same scaling factors were applied to the hydroxyl atomic quadrupoles in ethanol. Indeed, with the scaling, the hydration free energies of ethanol show a satisfactory agreement with experimental data using DMA2 (RMSE=0.61 kcal/mol) and original-fit (RMSE=0.70 kcal/mol). Acetic acid and p-cresol also have the hydroxyl group, but no scaling is required since their hydroxyl groups are considered part of the larger functional groups and the scaling of the quadrupoles have little effect on the gas-phase dimer or hydration properties.

Table 5 illustrates the effects of basis sets on the hydration free energies for seven compounds. The atomic multipoles were optimized using the DMA2 approach. Overall Comparisons show that hydration free energies calculated with cc-pVTZ and 6-311G** are similar to each other. The RMSD in the hydration free energy between these two basis sets is 0.23 kcal/mol. Also these two basis sets underestimate the experimental data in most cases, both with RMSEs around 1.0 kcal/mol. On the contrary, 6-311++G(2d,2p) overestimates the hydration free energies in all cases except acetic acid, RMSE 0.77 kcal/mol. Acetic acid is a special case, hydration free energies calculated with all the four methods are underestimated by 1~2 kcal/mol. In general, 6-311++G(2d,2p) gives more favorable hydration free energies for all the small molecules than the calculated results with all the other basis sets. Compared with cc-pVTZ and 6-311G**, hydration free energies calculated from 6-311++G(2d,2p) are closer to that from aug-cc-pVTZ, with a RMSD of 0.58 kcal/mol. The aug-cc-pVTZ set does not show a systematical under- or overestimation.

Overall, the agreement with experimental hydration free energy results improve with the size of the basis set used in the *ab initio* calculations to derive the atomic multipoles. The inclusion of diffuse function the QM basis set has a large effect, as seen in the comparison between the results from aug-cc-pVTZ and 6-311++G(2d,2p) with those of 6-311G** and cc-pVTZ. Obviously, aug-cc-pVTZ gives the best hydration free energy result, and 6-311++G(2d,2p) is also comparable. Plus 6-311++G(2d,2p) is much more affordable since it employs approximately half Gaussian basis functions as the aug-cc-pVTZ basis set, it is recommended as the optimal basis set to MP2 to perform hydration free energy calculations for large molecules. Furthermore, since aug-cc-pVTZ and 6-311++G(2d,2p) have better performances than the other two basis sets, we conclude that it is important to use basis sets with diffusion functions to capture intermolecular interactions.

Table 5 also lists previously reported hydration free energies calculated with some of the fixed charge models GAFF, CHARMm-MSI and OPLS_2005. Although some of the calculated values are close to experimental data, some have erroes as large as 5 kcal/mol. The MUEs and RMSEs of the available hydration free energy valuesare between 1 kcal/mol and 3 kcal/mol. Recent studies over a much wider range of organic molecules using fixed-charge model[4,5,19] also reported root mean square error (RMSE) greater than 1 kcal/mol, consistent with the current finding. Our results suggest that the inclusion of polarization effect and/or the use of atomic multipoles moments offer better performance. It is interesting to note that the gas-phase electrostatic potentials of the atomic multipoles derived from the four basis sets are very similar for all the molecules (Table 6). However, the final hydration free energy values can be as different as more than 1 kcal/mol in some cases. For example, the RMSD of the average magnitude of gas phase potentials is 0.14 kcal/mol between results

from aug-cc-pVTZ and 6-311G**; while the RMSD of the hydration free energies between these two basis sets is increased to 0.62 kcal/mol. The hydration free energies with atomic multipoles derived from gas phase show a favorable agreement with the experimental measurements, suggesting that the polarization effect between solute and solvent environment are well captured by the explicit atomic dipole induction model in AMOEBA. Note that the RMS gas phase potentials for isopropanol at 6-311++G(2d,2p) is very different from the other three basis sets because of the quadrupole scaling on the hydroxyl group, leading to the hydration free energy difference of more than 1 kcal/mol from the other basis sets.

The overall hydration free energy results using the polarizable atomic multipoles are encouraging compared those with fixed atomic charge force fields. It was suggested[4] that the fundamental limit of fixed-charge force fields is roughly 1 kcal/mol (mean unsigned error (MUE)) for hydration free energy. Our work here shows hydration free energy can be calculated within an accuracy of 0.41 kcal/mol (MUE) using gas-phase atomic multipoles from MP2/aug-cc-pVTZ, with the polarization modeled via induced atomic dipoles. Previously we showed that the polarization enhances the solvation of benzamidinium ion in water, and the contribution of polarization to hydration free energy is about 10% of the total electrostatic contributions. Therefore it is expected that even a fixed charge model can lead to a reasonable hydration free energy when the atomic charges are systematically increased to implicitly account for the polarization effect in water environment. The challenge however is that the polarization in the protein-like environment had an opposite effect which actually weakens the interaction between the small molecule and the protein.

Effect of Water Models

We further computed the hydration free energies using polarizable solutes but with a fixed-charge water model (PM/MM). As the AMOEBA vdW functional form (buffered 14-7) is different from the common Lennard-Jones 12-6 utilized by TIP3P and other popular simple water models, we have developed a TIP3P-like model by removing the atomic polarizability, atomic dipole and quadrupole moments in AMOEBA water model. The monopole and vdW parameters were adjusted in this TIP3P-like model to match the liquid properties of TIP3P and experiments. The parameters of TIP3P-like water model are summarized in Table 7. The density and heat of vaporization calculated using a box of 216 water molecules at 1 atom and 298 K are given in Table 8, compared to those of TIP3P, AMOEBA, and experimental measurements. The calculated radial distribution function (RDF) of O···O (Figure 3) also shows a good agreement with the result given by the original TIP3P model[49]. The O···O RDF from both models show a first peak at 2.77 Å, in agreement with the experimental measurement[50]. The second peak of TIP3P-like water is somewhat shifted to the right compared with the original TIP3P model, although the second peaks of both water models are too flat. With this TIP3P-like model, we were able to construct a hybrid system, in which the solute molecule possesses explicitly polarizable atomic multipoles, while the TIP3P-like model is used for solvent where the polarization in bulk environment is implicitly included in charge parameters. The parameters for the six small molecule solutes are the same as above (Table 5) with the multipoles derived at the MP2/6-311++G(2d,2p). The reduction in computational cost by making the water non-polarizable is about 50%, not as significant as one may hope, due to the need to converge the few induced-dipole moments on the solute molecule.

The hydration free energy results from AMOEBA water model and TIP3P-like water model are compared in Table 9. The average error from simulations using the TIP3P-like water is 1.16 kcal/mol (MUE) or 1.96 kcal/mol (RMSE). The magnitude of the errors seems to be consistent with other studies.[5,20] Especially in the case of the hydration of ethylbenzene, a system likely dominated by repulsion and dispersion, the fixed-charge water introduces little

alteration in the hydration free energy compared to the polarizable water model. For the other molecules, the TIP3P-like water systematically overestimates the hydration free energies, resulting in too favorable solvation. It however suggests that it may be possible to modify the fixed-charge model to improve the accuracy of calculated hydration free energy. It is conceivable that many combinations of atomic charges and vdW parameters can perform similarly well for bulk water properties, but some may give better solvation free energy. An exhaustive search of parameter space is beyond the scope of the current study. In addition, better water models such as TIP4P[19] and TI4P-EW[51] would be worth considering. This preliminary study however demonstrates that it is possible to combine polarizable atomic multipole and fixed-charge models to describe different molecules in a hybrid setting, which potentially offers the attractive combination of physical accuracy and computational efficiency in molecular simulations.

Conclusion

In this study, we computed the hydration free energies for several organic molecules using AMOEBA polarizable force field. We first evaluated the effect of vdW cutoff length and importance of decoupling of the long-range correction by performing a series hydration free energy simulations. We used an alchemical approach in which the electrostatic and then vdW interaction between the solute and solvent molecules were turned off in several steps. We showed that a long vdW cutoff (12 Å) with LRC and 11 perturbation steps in vdW annihilation are necessary for high-precision calculations of hydration free energy. Extra steps in the middle of vdW annihilation are needed as the system energy fluctuation is rather significant when solvent and solute molecules begin to penetrate each other.

With the most appropriate simulation protocol determined, we have investigated various models for intermolecular electrostatic interactions between the organic molecules and water environment via the calculation of hydration free energies. By putting the solute with gas-phase atomic multipoles, the polarization effect between solute and water environment was well captured by the explicit atomic dipole induction model in AMOEBA, as evident by the good agreement between the calculated and experimental hydration free energies for the seven molecules. We tested two methods to derive the gas-phase atomic multipoles from *ab initio* calculations for the solute molecules. One is ESP fitting based on the original DMA method, and the other utilizes the new DMA procedure by Stone. The two methods gave very similar hydration free energies for p-cresol in our test, with the difference comparable to the statistical error of the simulation. We recommend the original-fit procedure to derive atomic multipoles based on our experience. Also this approach allows fitting to the ESP of multi-conformations of flexible molecules simultaneously to derive “conformation-independent” atomic multipoles and allows atoms in symmetry to share the same atom type (e.g. the three H atoms in a methyl group). We subsequently compared the electrostatic parameters (atomic multipoles) derived from MP2/6-311G*, MP2/6-311++G(2d,2p), MP2/cc-pVTZ, and MP2/aug-cc-pVTZ. The hydration free energy results based on the four basis sets were in good agreement with experimental data, with RMSEs from 0.63 to 1.09 kcal/mol. Among these four *ab initio* methods, the aug-ccpVTZ basis set gave the best hydration free energy results while the 6-311++G(2d,2p) also performed well and was computationally more affordable. It is encouraging that the overall accuracy increases with the larger basis sets and the inclusion of diffuse functions in the *ab initio* basis set is highly recommended in deriving the atomic multipoles for modeling intermolecular interactions.

Furthermore, we investigated a hybrid PM/MM approach where the solute molecule was modeled with polarizable atomic multipoles and solvent was represented by fixed-charge TIP3P-like water molecules. The saving in computational cost is about 50% when compared to a fully polarizable model. The calculated hydration free energy values were within 1.16

kcal/mol (MUE) of the experimental measurements. While the error is about twice of that from simulations using the polarizable AMOEBA water, there was a systematic overestimation for five out of the six molecules compared. It is likely that a better water model or careful parameterization can further improve the accuracy.

Acknowledgments

This research is supported by grants from the National Institute of General Medical Sciences (R01GM079686). The authors also acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. The authors thank Dian Jiao for his help with the automation script.

Bibliography

1. Shoichet BK, Leach AR, Kuntz ID. Proteins: Structure, Function, and Genetics. 1999; 34(1):4–16.
2. Jiang H, Jordan KD, Taylor CE. Journal of Physical Chemistry B. 2007; 111(23):6486–6492.
3. Kaminski G, Duffy EM, Matsui T, Jorgensen WL. Journal of Physical Chemistry. 1994; 98(49): 13077–13082.
4. Mobley DL, Dumont E, Chodera JD, Dill KA. Journal of Physical Chemistry B. 2007; 111(9):2242–2254.
5. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. Journal of Chemical Theory and Computation. 2009; 5(2):350–358. [PubMed: 20150953]
6. Rizzo RC, Aynechi T, Case DA, Kuntz ID. Journal of Chemical Theory and Computation. 2006; 2(1):128–139.
7. Shivakumar D, Deng YQ, Roux B. Journal of Chemical Theory and Computation. 2009; 5(4):919–930.
8. Gallicchio E, Levy RM. Journal of Computational Chemistry. 2004; 25(4):479–499. [PubMed: 14735568]
9. Gallicchio E, Paris K, Levy RM. Journal of Chemical Theory and Computation. 2009; 5(9):2544–2564. [PubMed: 20419084]
10. Tan C, Tan YH, Luo R. Journal of Physical Chemistry B. 2007; 111(42):12263–12274.
11. Wagoner JA, Baker NA. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(22):8331–8336. [PubMed: 16709675]
12. Onufriev A. Annual Reports in Computational Chemistry. 2008; 4:125–137.
13. Mobley DL, Barber AE, Fennell CJ, Dill KA. Journal of Physical Chemistry B. 2008; 112(8): 2405–2414.
14. Purisima EO, Sulea T. Journal of Physical Chemistry B. 2009; 113(4):8206–8209.
15. Mongan J, Simmerling C, McCammon JA, Case DA, Onufriev A. Journal of Chemical Theory and Computation. 2007; 3(1):156–169. [PubMed: 21072141]
16. Straatsma TP, Berendsen HJC. Journal of Chemical Physics. 1988; 89(9):5876–5886.
17. Frenkel D, Ladd AJC. Journal of Chemical Physics. 1984; 81(7):3188–3193.
18. Kollman P. Chemical Reviews. 1993; 93(7):2395–2417.
19. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Journal of Chemical Physics. 1983; 79(2):926–935.
20. Shivakumar D, Williams J, Wu YJ, Damm W, Shelley J, Sherman W. Journal of Chemical Theory and Computation. 2010; 6(5):1509–1519.
21. Shirts MR, Pitera JW, Swope WC, Pande VS. Journal of Chemical Physics. 2003; 119(11):5740–5761.
22. Chipot C. Journal of Computational Chemistry. 2003; 24(4):409–415. [PubMed: 12594783]
23. Hess B, van der Vegt NFA. Journal of Physical Chemistry B. 2006; 110(35):17616–17626.
24. Baker CM, Lopes PEM, Zhu X, Roux B, MacKerell AD. Journal of Chemical Theory and Computation. 2010; 6(4):1181–1198. [PubMed: 20401166]

25. Stone AJ. Chemical Physics Letters. 1981; 83(2):233–239.
26. Stone AJ, Alderton M. Molecular Physics. 2002; 100(1):221–233.
27. Bader, RFW. Atoms in Molecules—A Quantum Theory. Clarendon Press; Oxford: 1990.
28. Sokalski WA, Shibata M, Ornstein RL, Rein R. Theoretica Chimica Acta. 1993; 85(1–3):209–216. [PubMed: 11539836]
29. Sokalski WA, Poirier RA. Chemical Physics Letters. 1983; 98(1):86–92.
30. Williams DE. Journal of Computational Chemistry. 1988; 9(7):745–763.
31. Ren PY, Ponder JW. Journal of Computational Chemistry. 2002; 23(16):1497–1506. [PubMed: 12395419]
32. Stone AJ. Journal of Chemical Theory and Computation. 2005; 1(6):1128–1132.
33. Ren PY, Ponder JW. Journal of Physical Chemistry B. 2003; 107(24):5933–5947.
34. Jiao D, Golubkov PA, Darden TA, Ren P. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(17):6290–6295. [PubMed: 18427113]
35. Jiao D, Zhang JJ, Duke RE, Li GH, Schnieders MJ, Ren PY. Journal of Computational Chemistry. 2009; 30(11):1701–1711. [PubMed: 19399779]
36. Shi Y, Jiao D, Schnieders M, Ren P. IEEE Engineering in Medicine and Biology Society. 2009; (3–6):2328–2331.
37. Ren PY, Ponder JW. Journal of Physical Chemistry B. 2004; 108(35):13427–13437.
38. Halgren TA. Journal of the American Chemical Society. 1992; 114(20):7827–7843.
39. Ponder, JW. Washington University Medical School. Version 5.1. 2009. Available at:http://dasher.wustl.edu/tinker/
40. Stone, AJ. University of Cambridge. version 2.2. 2005. 02
41. Frisch, M. J. e. a. Gaussian Inc; Pittsburgh, PA: 2003.
42. Bennett CH. Journal of Computational Physics. 1976; 22(2):245–268.
43. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. Journal of Computational Chemistry. 2005; 26(16):1668–1688. [PubMed: 16200636]
44. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Journal of Chemical Physics. 1984; 81(8):3684–3690.
45. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. Journal of Chemical Physics. 1995; 103(19):8577–8593.
46. Darden T, York D, Pedersen L. Journal of Chemical Physics. 1993; 98(12):10089–10092.
47. Sagui C, Darden TA. Annual Review of Biophysics and Biomolecular Structure. 1999; 28:155–179.
48. Shirts MR, Mobley DL, Chodera JD, Pande VS. Journal of Physical Chemistry B. 2007; 111(45): 13052–13063.
49. Mark P, Nilsson L. Journal of Physical Chemistry A. 2001; 105(43):9954–9960.
50. Soper AK. Chemical Physics. 2000; 258(2–3):121–137.
51. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Journal of Chemical Physics. 2004; 120(20):9665–9678. [PubMed: 15267980]
52. Cabani S, Gianni P, Mollica V, Lepori L. Journal of Solution Chemistry. 1981; 10:563–595.
53. Wolfenden R, Liang Y, M M, R W. Journal of the American Chemical Society. 1987; 109:463–466.
54. Mahoney MW, Jorgensen WL. Journal of Chemical Physics. 2000; 112(20):8910–8922.

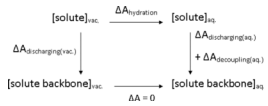

Figure 1.

Thermodynamic cycle of HFE calculation in explicit water MD simulations. Potential energy of the solute backbone includes valence interactions and vdW interactions within the solute itself.

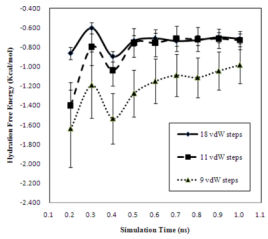

Figure 2.

Convergence of HFE of ethylbenzene at different simulation time with three vdW decoupling protocols. Hydration free energies with 18 vdW steps (a) are in solid line with diamond, 11 vdW steps (b) in dashed line with square, and 9 vdW steps (c) in dotted line triangle. The bars are the statistical errors computed using Bennett's formula

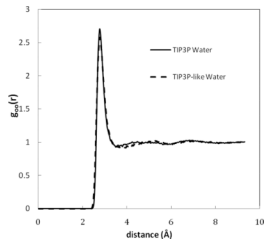

Figure 3.
O··O radial distribution function for TIP3P liquid water and TIP3P–like liquid water at 298 K.

Table 1

vdW contributions of ethylbenzene hydration free energies from intermediate decoupling steps. All units are kcal/mol. Results are calculated using the dynamics simulations trajectories up to 500 ps. For schedule (a) and (b), which uses more intermediate steps, the results have been combined for comparison.

Steps[1]		Intermediate Free Energy Changes								Total
		1.0–0.9	0.9–0.8	0.8–0.7	0.7–0.6	0.6–0.5	0.5–0.4	0.4–0.2	0.2–0.0	
(a)	18	5.127	2.281	−0.241	−4.037	−4.848	−0.896	0.039	0.048	−2.53
(b)	11	5.131	2.304	−0.237	−4.059	−4.851	−0.892	0.040	0.048	−2.52
(c)	9	5.135	2.304	−0.169	−3.605	−4.851	−0.898	0.040	0.048	−2.00

[1]Three vdW decoupling schedules:

(a) λ = (1.0, 0.95, 0.9, 0.85, 0.8, 0.775, 0.75, 0.725, 0.7, 0.675, 0.65, 0.625, 0.6, 0.55, 0.5, 0.4, 0.2, 0.0);

(b) λ = (1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.5, 0.4, 0.2, 0.0);

(c) λ = (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.2, 0.0).

Table 2

The effect of vdW cutoff length and LRC on the AMOEBA hydration free energies. Hydration free energies were estimated using both 9 Å and 12 Å with and without LRC decoupling. All the hydration free energies were decomposed into the electrostatic and vdW components. Statistical errors of the total HFE are given in the parenthesis. All values are in kcal/mol.

Cutoff(Å)		Ethylbenze		p-cresol		Isopropanol		imidazole		methylethyl sulfide		acetic acid	
		ele[1]	vdW[2]	ele	vdW	ele	vdW	ele	vdW	ele	vdW	ele	vdW
w/LRC	12.0	−3.35	2.65	−9.37	2.04	−8.28	2.64	−11.38	1.31	−4.42	2.55	−7.89	2.36
		(0.15)		(0.10)		(0.11)		(0.03)		(0.09)		(0.04)	
	9.0	−3.37	2.74	−8.75	1.78	−7.99	2.58	−11.43	1.31	−4.38	2.45	−7.93	2.36
		(0.11)		(0.10)		(0.11)		(0.01)		(0.09)		(0.04)	
w/o LRC	12.0	−3.35	2.88	−9.37	2.23	−8.28	2.58	−11.38	1.34	−4.42	2.69	−7.89	2.44
		(0.15)		(0.10)		(0.11)		(0.03)		(0.09)		(0.04)	
	9.0	−3.37	2.95	−8.75	2.24	−7.99	2.68	−11.43	1.66	−4.38	2.88	−7.93	2.61
		(0.11)		(0.10)		(0.11)		(0.03)		(0.09)		(0.04)	

[1] "ele" refers to the free energy contributions of $-\Delta A_{\text{discharging(aq.)}}+\Delta A_{\text{discharging(aq.)}}$ in equation (1).

[2] "vdW" represents the contribtion of $-\Delta A_{\text{decoupling(aq.)}}$ in the same equation.

Table 3

Comparison of hydration free energies between original-fit and DMA2 methods for p-cresol. All units are kcal/mol. Statistical errors are given in the parenthesis.

MP2 Basis Set	cc-pVTZ	6-311G**	6-311++G(2d,2p)	aug-cc-pVTZ	Experiment[1]
Original-fit	−6.56 (0.10)	−6.47 (0.10)	−7.26 (0.11)	−7.38 (0.10)	−6.10 −6.60
DMA2	−6.47 (0.11)	−6.89 (0.10)	−7.33 (0.10)	−6.98 (0.16)	

[1] Experimental values are reported in ref 52.

Table 4

Hydration free energies with scaling on the quadrupoles of hydroxyl groups in alcohol molecules, using both original-fit and DMA2. Isopropanol and ethanol were selected for comparison. All units are kcal/mol. Statistical errors are given in the parenthesis.

MP2 Basis Set	Isopropanol			Ethanol	
	Original-fit unscaled	Original-fit scaled to 0.6	DMA2 scaled to 0.7	Original-fit scaled to 0.6	DMA2 scaled to 0.7
cc-pVTZ	−5.43 (0.11)	−4.36 (0.18)	−4.04 (0.11)	−3.58 (0.19)	−4.19 (0.07)
6-311G**	−5.60 (0.10)	−4.16 (0.15)	−4.34 (0.10)	−4.84 (0.20)	−4.51 (0.09)
6-311++G(2d,2p)	−7.60 (0.11)	−5.59 (0.18)	−5.58 (0.12)	−5.12 (0.19)	−5.67 (0.09)
aug-cc-pVTZ	−6.98 (0.16)	−4.36 (0.28)	−4.32 (0.11)	−4.53 (0.22)	−5.04 (0.09)
Experiment[1]		−4.70, −4.80		−4.90	

[1] Experimental values are reported in ref 52.

Table 5

Comparison of hydration free energies of small molecules with four different basis sets: (i) 6-311G**, (ii) 6-311++G(2d,2p), (iii) cc-pVTZ, and (iv) aug-cc-pVTZ. Hydration free energies calculated with fixed-charge models were also listed for comparison. All units are kcal/mol. Statistical errors are given in the parenthesis.

	Ethylbenze	p-cresol	Isopropanol	Imidazole	Methylethyl sulfide	Acetic acid	Ethanol	MUE	RMSE
cc-pvTZ	−0.68 (0.14)	−6.56 (0.11)	−4.04 (0.11)	−9.07 (0.03)	−1.61 (0.09)	−4.57 (0.04)	−4.19 (0.19)	0.63	1.09
6-311g**	−0.53 (0.14)	−6.47 (0.10)	−4.34 (0.10)	−8.70 (0.04)	−1.53 (0.10)	−4.45 (0.05)	−4.51 (0.19)	0.61	1.15
6-311++g(2d,2p)	−0.70 (0.14)	−7.26 (0.14)	−5.58 (0.12)	−10.11 (0.03)	−1.87 (0.09)	−5.53 (0.04)	−5.67 (0.20)	0.64	0.77
aug-cc-pvTZ	−0.44 (0.12)	−7.05 (0.11)	−4.32 (0.11)	−9.72 (0.03)	−1.56 (0.10)	−5.48 (0.04)	−5.04 (0.22)	0.41	0.63
GAFF/AM1[1]	0.8	−5.1	-	-	−0.9	-	−3.9	1.09	1.14
GAFF/AM1 (SSBP)[2]	−0.56	-	−2.88	-	-	−8.31	−2.88	1.41	1.60
MSI/AM1 (SSBP)[2]	0.03	-	−1.42	-	-	−12.52	−1.94	3.21	3.68
OPLS_2005[3]	−0.46	-	−4.15	-	-	−5.44	−1.94	1.26	1.64
Experiment[4]	−0.70	−6.10	−4.70	−9.63	−1.50	−6.70	−4.90		
		−6.60	−4.80						

[1]Calculated hydration free energies from ref 4. Solute molecules used GAFF, along with the semiempirical AM1-BCC charge model. Waters were represented by TIP3P model.

[2]Calculated hydration free energies from ref 7. Solute molecules used GAFF or CHARMm-MSI force fields, along with the semiempirical AM1-BCC charge model. Water molecules close to the solute were represented by TIP3P model. SSBP were incorporated as the remaining bulk.

[3]Calculated hydration free energies from ref 20. Solute molecules used OPLS_2005 force field. Waters were represented by TIP3P model.

[4]Experimental values are reported in ref 52, except for imidazole taken from ref 53.

Table 6

The root mean square difference of gas phase potentials of the small molecules with multipole parameters generated from four different basis sets. All units are kcal/mol.

	Ethylbenze	p-cresol	Isopropanol	Imidazole	Methylethyl Sulfide	Acetic acid	Ethanol
cc-pVTZ	0.13	0.25	0.00	0.20	0.16	0.32	0.29
6-311g**	0.21	0.34	0.16	0.27	0.07	0.67	0.33
6-311++g(2d,2p)	0.07	0.11	0.41	0.16	0.07	0.12	0.14
aug-cc-pVTZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 7

Force Field Parameters of TIP3P-like water model. Only non-zero parameters are listed.

O-H bond	b_0(Å)	K_b (kcal/Å^2/mol)
	0.9572	529.6
H-O-H angle	θ_0 (deg)	K_0 (kcal/deg^2/mol)
	108.5	34.05
Urey-Bradley	l_0 (Å)	K_l (kcal/Å^2/mol)
	1.5537	38.25
van der Waals	R^0 (Å)	ε (kcal/mol)
O	3.505	0.1100
H	2.655	0.0040
$H_{\text{reduction}}$	91%	
Charge	(a.u.)	
O	−0.89200	
H	0.44600	

Table 8

Calculated and experimental properties for liquid water at 298 K and 1 atm.

	AMOEBA[1]	TIP3P[2]	TIP3P-like	Experiment[2]
ρ (g/cm3)	1.000	1.018	1.002	0.997
ΔH_{vap} (kcal/mol)	10.51	10.54	10.41	10.51

[1]Reference 33

[2]Reference 54

Table 9

Experimental and computed hydration free energies for small molecules with AMOEBA water model and “TIP3P-like” model. All the hydration free energies are shown in kcal/mol.

	ethylbenzene	p-cresol	isopropanol	imidazole	methylethyl sulfide	acetic acid	MUE	RMSE
AMOEBA	−0.73 (0.14)	−7.26 (0.10)	−5.58 (0.12)	−10.11 (0.03)	−1.78 (0.09)	−5.69 (0.04)	0.58	0.68
TIP3P-like	−0.89 (0.11)	−10.72 (0.07)	−5.29 (0.11)	−10.87 (0.03)	−2.94 (0.09)	−7.46 (0.07)	1.16	1.96
experiment	−0.70	−6.10	−4.70	−9.6	−1.50	−6.70	-	-
		−6.60	−4.80					