Molecular Dynamics Simulations of Liquid Water Structure and Diffusivity

Yu-De Chen, Arvin Huang-Te Li, Yi-Siang Wang, and Sheng D. Chao

Institute of Applied Mechanics, National Taiwan University, Taipei 106, Taiwan, Republic of China

(Received December 16, 2012; Revised April 13, 2013)

An ab initio force field has been constructed to study the structure and diffusivity for liquid water using molecular dynamics simulations. The potential energy data were calculated by using both the second order Møller-Plesset (MP2) perturbation theory and the coupled cluster method with single, double, and perturbative triple excitations [CCSD(T)], together with the correlation consistent basis set up to aug-cc-pV5Z. For configuration sampling, we considered 10 dimer orientations, each with about 121 grid points. With an additional “polarization” term, an analytical function similar to that for the well-established TIP4P model was used to represent the ab initio potential data. The simulation results were compared with those using several empirical force fields (SPC/E, TIP4P etc.) or the recent CC-pol series of ab initio force fields and also with available experimental measurements. We obtained good quantitative agreement for the atom-wise radial distribution functions and the self-diffusion coefficients over a wide range of thermodynamic conditions. This new ab initio force field can thus serve as a suitable starting point for predicting liquid water properties without priori empirical knowledge.

DOI: 10.6122/CJP.51.1218

PACS numbers: 34.20.Gj, 31.50.-x, 33.15.-e

I. INTRODUCTION

The important role played by liquid water in almost every scientific discipline can hardly be overemphasized [1–3]. Liquid water exhibits many intricate physical properties, such as the anomalous temperature dependence of density and thermal expansion due to which the whole biosphere operation is possible. It is not surprising that an enormous amount of effort has been devoted to deciphering the mysteries [4–6]. A glimpse at the essentially countless experimental and theoretical studies on this ubiquitous and omnipresent substance might give an impression that a clear understanding of liquid water properties at a molecular level has been well grasped. That there were still debating issues for the bulk water structure in ambient conditions [7] might stun many not directly working on this subject, and only recently a better understanding has been achieved (see the recent review [8]). One of the difficulties in experiments is the inaccessibility of quantitative information from a single specific observation technique. Therefore, techniques such as neutron and X-ray scattering experiments are also required to obtain a more complete picture of liquid water.

*Electronic address: sdchao@spring.iam.ntu.edu.tw
On the theoretical side, molecular simulations have become a useful tool to complement experiments in both predicting verifiable measurements and providing physical insights. Many such simulations on liquid water employed empirical force fields [9–13], which were fitted to available experimental data. However, it is known that the reliability of their predictions may highly depend on the thermodynamic conditions used in the model development. Recently, force fields entirely from first principles calculations without experimental data input a priori, or ab initio force fields, have been actively pursued. For water, there have been numerous ab initio force fields, and their relative performances have been recently reviewed [14–18]. Here we mention two recent ones; i.e., the CC-pol potential by Bukowski et al. [19–21] and the HBB potential by Huang et al. [22–24]. The CC-pol force field is a six-dimension rigid-molecule model calculated at a two-step MP2-CCSD(T) level of theory by sampling 2510 selected grid points of the water dimer. This model reproduces the general gas phase properties, in particular the dimer vibration-rotation-tunneling (VRT) spectra, very accurately. Combined with a symmetry adapted perturbation theory (SAPT) 3-body (and extrapolated N-body) potential, the CC-pol(+NB) model can reproduce the liquid water structural properties in ambient conditions. The HBB force field is, on the other hand, a 12-dimension flexible-molecule model calculated at the CCSD(T)/aVTZ and MP2/aVTZ level of theories for the 1-body and for the 2- and 3-body potentials, respectively. The HBB potential model not only reproduces the monomer infrared (IR) spectra and the dimer VRT spectra [25], but also the structures and energetics of water clusters up to the 22-mer in an unprecedented level of accuracy. These achievements really represent the triumph of current predictive power of first principles calculations [26, 27].

A closer look at the subsequent model validations [28, 29] reveals that although the CC-pol model performs well for the gas phase and dilute vapor water, it does not perform as equally well in reproducing the liquid phase atom-wise RDFs and the transport properties at low temperatures. Probably part of the reasons can be attributed to data sampling and analytical representation [20, 28]. On the other hand, the ability of the HBB model to reproduce liquid properties has not been demonstrated, probably due to the inconvenience of using the complicated functional form in molecular simulations. From the practical point of view, a less sophisticated ab initio water model which can compete with commonly used empirical force fields is desired. Therefore, the purpose here is to construct a new ab initio water potential model which employs extensive configuration sampling, has an easy-to-use functional form for molecular simulations, and can reproduce both the structural and dynamical properties of liquid water.

In this paper we develop an ab initio 2-body force field using a high level of theory, up to the CCSD(T)/aug-cc-pV5Z level of theory. To avoid biased sampling, we consider both hydrogen-bond and non-hydrogen-bond configurations without imposing molecular symmetry constraints. An analytical potential functional form similar to the commonly used TIP4P model has been used to fit the potential data. Our MD simulations using this force field compare favorably to the available experiments, thus giving confidence to the validity of this force field.
II. COMPUTATIONAL DETAILS

The simulations were performed in a similar way as in our previous studies [30–32]. Briefly, quantum chemistry calculations were performed using the Gaussian 09 program package [33]. The water monomer was optimized at the MP2/aug-cc-pVQZ level of theory. The bond length for O-H is 0.9586 (0.9572) Å, and the bond angle for $\angle H-O-H$ is 104.2906° (104.52°), with the experimental data [34] shown in the parentheses. The supermolecular counterpoise corrected interaction energies [35, 36] were calculated by using both the MP2 [37] theory and the CCSD(T) [38] method with the aug-cc-pVXZ, X=D, T, Q, and 5, basis functions [39]. The 10 configurations [40] for the water dimer are depicted in Fig. 1. The distance points between the two oxygen atoms were sampled for the range of $2 \sim 14$ Å, with 121 configuration points for each orientation. During the scan we fixed the monomer geometry (rigid monomer assumption), which results in an energy uncertainty from the fully relaxed conformation by about 0.04 kcal/mol. The interaction energies at the complete basis set (CBS) limit were estimated using the well established extrapolation methods [41–45]. In MD simulations, a system of 256 water molecules in a cubic box with periodic boundary condition was simulated in a canonical (NVT) ensemble. Newton's equations of motion for the center of mass motions and Euler equations for the rotational motions were integrated. The box size is determined by the density and temperature used in the simulation. At 298 K, the cut-off radius $r_{\text{cut}}$ was varied from 7.95 angstroms to half box size. We have carefully checked the convergence on the simulation results. For example, using the $r_{\text{cut}} = 9.275$ angstroms, the diffusion constant increases to 2.556 ($10^{-9}$ m²/s), which is merely 1% increase from using $r_{\text{cut}} = 7.95$ angstroms. To avoid long range artificial correlations on calculated properties, the box size has been extended to contain 512 water molecules. No significant differences in the calculated properties were observed. The temperature control is achieved by rescaling the velocity every 1000 time steps. The equations of motion were solved by the velocity Verlet algorithm. The rigid body constraint has been taken into account as we solved the Euler equation for rotation motion. The time step is 1 fs and the total simulation time is up to 200 ps with the first 20 ps for equilibration. For better convergence on the transport properties, we extended the simulation time to 1 ns. All the calculated diffusion constants were checked to be converged and the numerical errors were limited to be less than 5%. For dense (dilute) liquid, the density and temperature ranges cover $1.000 \sim 0.990$ (0.0385 $\sim 0.065$) g/cm³ and $274 \sim 318$ (400) K [46], respectively, where the experimental data are available. Because we used the second order propagators in the solution phase, the simulation wall-time scaled as the square of the system size. For the current simulation scale, it took about 3 days for a single run on a standard table top PC, excluding the post treatment time.

III. RESULTS AND DISCUSSIONS

The Hartree-Fock (HF) and the MP2 interaction potentials for the 10 configurations are shown in Fig. 2. We see that the correlation energies play a dominant role in accurately
determining the intermolecular interaction energy. The minimum-energy conformation corresponds to the configuration 1, which is similar to that found in previous works [40]. In Fig. 3 we show the basis set dependence of the CCSD(T) potential curves for this conformer. The convergence pattern shows that the uncertainty level of the calculated interaction energies is below 0.05 kcal/mol. In Table I we present the oxygen-oxygen distance $R_0$, where the potential energy is zero, the equilibrium bond length $R_m$, and the binding energy $E_b$ extracted from the MP2 and CCSD(T) potential curves for several basis sets. The MP2 and CCSD(T) binding energies converge to about 0.088 kcal/mol and 0.039 kcal/mol at the aug-cc-pVQZ basis set, respectively. The MP2/CBS(CCSD(T)/CBS) binding energies using the methods of Martin [41], Helgaker et al. [42], Feller [43], the numerical method [44] and the “focal-point” method of Jureccka et al. [45] are $-5.000 (-5.000)$, $-5.021 (-5.009)$, $-5.046 (-4.980)$, $-5.081 (-4.981)$, and $-5.094$ in kcal/mol, respectively. We see that these estimations are consistent with each other, indicating the basis set saturation. Therefore, we believed from the calculations that the potential data have achieved a high (ca. 0.05 kcal/mol) precision. With this precision level, the MP2 interaction energies are essentially equivalent to the CCSD(T) counterparts numerically, which verifies the conclusion by Xantheas and coworkers [47] in terms of direct numerical calculation.

An analytical 4-site model similar to the TIP4P model [48], dubbed as TIP4P(M),
FIG. 2: The HF (a) and the MP2 (b) potentials for the 10 configurations using the aug-cc-pVQZ basis set.

was used to fit the ab initio data. The interaction sites are associated with the oxygen atoms, the hydrogen atoms, and a pseudo-atom M located at the $\angle$H-O-H bisector with a distance 0.15 Å from the oxygen towards the hydrogens. The site-site interaction is represented by the Lennard-Jones (L-J) function, Coulomb potential, and an additional term with an inverse fifth power dependence on the interatomic distance.

$$U_{ij} = 4\varepsilon_{OO} \left[ \left( \frac{\sigma_{OO}}{r_{OO}} \right)^{12} - \left( \frac{\sigma_{OO}}{r_{OO}} \right)^{6} \right] + \sum_{ij} \left( \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}} + \left( \frac{b_{ij}}{r_{ij}} \right)^5 \right)$$

where the indices $i$ and $j$ denote the interaction sites in separated water monomers, respec-
FIG. 3: The CCSD(T) potentials for configuration 1 using the aug-cc-pVXZ, X = D, T, Q, 5, basis sets.

tively, and $r_{ij}$ represents the site($i$)–site($j$) distance. In most previous empirical force fields, only higher (than the sixth) inverse power terms were included to account for the polarization effect. This is of course well justified because the (averaged) orientation, induction, and dispersion interactions for dipole-dipole bound systems all start from the inverse sixth power law. However, for the water system (or in general hydrogen bond systems), the nature of the hydrogen bond is more like a charge-dipole interaction. In addition, hydrogen bonds are highly anisotropic, so a better trial functional form would be to allow lower inverse power terms in the regression [49]. In this model $\varepsilon_{OO}$, $\sigma_{OO}$, $q_i$, and $b_{ij}$ are the potential parameters to be determined in the numerical regression. The fitting parameters we obtained are $\sigma_{OO} = 2.65$ Å, $\varepsilon_{OO} = 0.48$ kcal/mol, $q_H = 0.58e$, $q_M = -1.16e$, $b_{OO} = 2.91$, $b_{OH} = 2.32$, and $b_{HH} = -0.48$ with energy in kcal/mol and distance in Å. The fitting curves and the \textit{ab initio} data are compared in Fig. 4. We see that the potential data are well represented by the analytical fit. The root mean square error of the fitting is about 0.20 kcal/mol. A closer look at the fitting results reveals an interesting interplay among different potential terms, as compared to the TIP4P model. First we notice that $b_{OO}$ and $b_{OH}$ are positive while $b_{HH}$ is negative. This means the additional term takes into account interactions other than induction. Second, our fitting partial charges are larger than those in TIP4P model, resulting in about 10% increase of the monomer dipole moment. Thus, a significant part of the polarization effect has been incorporated through the enhanced partial charges. The combination of the Coulomb and the additional terms effectively accounts for the charge penetration contributions. Because of this, the simulated static dielectric constant using
TABLE I: The basis set dependence of some potential parameters of the MP2 and the CCSD(T) potentials (shown in the parentheses) for the configuration 1.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>Number of basis function</th>
<th>( R_0 (\text{Å}) )</th>
<th>( R_m (\text{Å}) )</th>
<th>( E_b (\text{kcal/mol}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ</td>
<td>48</td>
<td>2.58</td>
<td>3.06</td>
<td>−4.215</td>
</tr>
<tr>
<td>6-311++G**</td>
<td>72</td>
<td>2.55</td>
<td>3.04</td>
<td>−4.588</td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td>82</td>
<td>2.55</td>
<td>3.01</td>
<td>−4.430</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.54)*</td>
<td>(3.00)</td>
<td>(−4.408)</td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>116</td>
<td>2.54</td>
<td>3.00</td>
<td>−4.449</td>
</tr>
<tr>
<td>6-311+++G(2df,2pd)</td>
<td>128</td>
<td>2.53</td>
<td>2.99</td>
<td>−4.536</td>
</tr>
<tr>
<td>6-311+++G(3df,3pd)</td>
<td>150</td>
<td>2.52</td>
<td>2.98</td>
<td>−4.600</td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td>184</td>
<td>2.51</td>
<td>2.96</td>
<td>−4.687</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.50)</td>
<td>(2.94)</td>
<td>(−4.760)</td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>230</td>
<td>2.52</td>
<td>2.96</td>
<td>−4.697</td>
</tr>
<tr>
<td>aug-cc-pVQZ</td>
<td>344</td>
<td>2.50</td>
<td>2.94</td>
<td>−4.841</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.51)</td>
<td>(2.93)</td>
<td>(−4.929)</td>
</tr>
<tr>
<td>cc-pV5Z</td>
<td>402</td>
<td>2.50</td>
<td>2.94</td>
<td>−4.833</td>
</tr>
<tr>
<td>aug-cc-pV5Z</td>
<td>574</td>
<td>2.47</td>
<td>2.91</td>
<td>−4.929</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.50)</td>
<td>(2.92)</td>
<td>(−4.968)</td>
</tr>
</tbody>
</table>

* Numbers in parentheses are calculated by the CCSD(T) method.

our model is about 60, which is closer to the experiment value (78.5) than TIP4P (50).

The atom-wise radial distribution functions (RDFs) are calculated by the usual definition [50] using the histogram method. Fig. 5 presents the simulated atom-atom radial distribution functions of water for temperature \( T = 298.15 \text{ K} \) and density \( \rho = 0.997 \text{ g/cm}^3 \). As we can see from the comparison of the calculated peak and valley positions with the experimental data [51], the agreement is generally good. Moreover, we compare our simulation results with those using the CC-pol(+NB) potential model. The present model performs equally well for the first peak region of \( g_{OO} [16] \), but much better for the regions beyond the first peak, where the CC-pol(+NB) potential was found to be insufficiently accurate. In particular, the coordination number we obtained using the present model is 4.9, which, as compared to the experimental value 4.8, is much closer to the experimental value than the value 5.6 using CC-pol(+NB). We also compare the RDFs obtained from several commonly used (not exhaustive) empirical force fields in Fig. 6. It can be seen that the present model can even compete with these potential models which used the experimental data in the fitting.

Next, we calculated the self-diffusion coefficients \( D \) using the Green-Kubo formula [52–
In Fig. 7, we present the temperature dependence of the self-diffusion coefficients from 274K to 318K. The calculated results are compared with experimental data from Mills [55]. We see that the results are generally in good agreement with the experiment. In particular at 25 °C and 1 atm, the diffusion constant we obtained, in units of \(10^{-9} \text{ m}^2/\text{s}\), is 2.521, as compared favorably with the experimental value of 2.30. As a further comparison, in Table II we list the calculated D values from the TIP3P, TIP4P, TIP5P, TIP4P-Ew, and SPC/E water models [56], respectively. Clearly, the present model can even compete with these empirical force fields in reproducing the liquid diffusivity. To examine the performance of the present force field for low-density water, we also calculated the self-diffusion coefficients D at a high temperature 400 °C. The D values obtained at
FIG. 5: The calculated partial atom-wise radial distribution functions ($g_{OO}$, $g_{OH}$, and $g_{HH}$) at $T = 298$ K and $\rho = 0.997$ g/cm$^3$, as compared to the simulation data using the CC-pol(+NB) model and the experimental data [51].

FIG. 6: The calculated partial atom-wise radial distribution functions ($g_{OO}$, $g_{OH}$, and $g_{HH}$) at $T = 298$ K and $\rho = 0.997$ g/cm$^3$, as compared to the simulation data from the empirical force fields and the experimental data [51].
density $\rho = 0.065 \text{ g/cm}^3$, 0.0564 g/cm$^3$, 0.051 g/cm$^3$, and 0.0385 g/cm$^3$ are 0.553 (0.585), 0.595 (0.595), 0.659 (0.654), and 0.777 (0.900), in units of $10^{-6}$ m$^2$/s, with the experimental data [57, 58] shown in the parentheses. We see that to a very wide range of thermodynamic conditions, MD simulations with the \textit{ab initio} force field are capable of reproducing the experimental data within the estimated experimental uncertainties.

### TABLE II: Comparison of the calculated D values at 25° and 1 atm from the TIP4P(M) model and selective empirical force fields with experiment.

<table>
<thead>
<tr>
<th>Force Field</th>
<th>Value (10$^{-6}$) m$^2$/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIP3P</td>
<td>5.06</td>
</tr>
<tr>
<td>TIP4P</td>
<td>3.29</td>
</tr>
<tr>
<td>TIP5P</td>
<td>2.62</td>
</tr>
<tr>
<td>TIP4P-Ew</td>
<td>2.40</td>
</tr>
<tr>
<td>SPC/E</td>
<td>2.49</td>
</tr>
<tr>
<td>TIP4P(M)</td>
<td>2.52</td>
</tr>
<tr>
<td>Exp.</td>
<td>2.30</td>
</tr>
</tbody>
</table>

### IV. CONCLUSION

In this paper we achieve our original objective to construct a water potential model from the state of the art quantum chemistry methods, with a simple analytical functional form and with good predictive power. The potential energy data are calculated at the MP2 and CCSD(T) level of theory and analytically represented by a modified TIP4P potential function. Our MD simulation results compare favorably with the experiments for a very wide range of thermodynamic conditions. The present model competes with the standard empirical force fields, as well as the recent CC-pol \textit{ab initio} force field. Therefore, this \textit{ab initio} force field can serve as a suitable starting point to predict liquid water properties without \textit{priori} empirical knowledge.

### Acknowledgements

This work was supported by the CQSE of National Taiwan University (10R80914). We acknowledge the National Center for High-performance Computing (NCHC) for providing computing resources.

### References


