Supporting Information for

Structure of Liquid Water - A Dynamical Mixture of Tetrahedral and "Ring-and-Chain" like Structures

Jinfeng Liu[1,2], Xiao He[1,3], and John Z. H. Zhang[1,3,4*]*

[1]*School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, 200062, China*

[2]*Department of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing, 210009, China*

[3]*NYU-ECNU Center for Computational Chemistry, NYU Shanghai, Shanghai, 200062, China*

[4]*Department of Chemistry, New York University, New York, NY 10003, USA*

*To whom correspondence should be addressed: xiaohe@phy.ecnu.edu.cn, zhzhang@phy.ecnu.edu.cn

1. EE-GMFCC scheme for the water cluster

The EE-GMFCC method was proposed for calculating the total energy of proteins. For a water cluster, each water molecule naturally becomes a single fragment. It does not need to cut chemical bond or add molecular caps. In the EE-GMFCC scheme, the energy of each water molecule and the interaction energy between two water molecules that are spatially in close contact are computed by QM, whereas the interaction energies between two distant water molecules are treated by pairwise charge-charge Coulomb interactions. All QM calculations are embedded in the electrostatic field of the point charges representing the remaining system to account for the environmental effect. In this study, the fixed charge model of the TIP3PF force field was utilized to describe the embedding field, while the positions of these point charges were updated during MD simulation to reflect the geometry change of water molecules in the QM zone. To ensure the convergence of the total energy of water cluster, the distance threshold λ used in this study was set to 5.0 Å. The EE-GMFCC approach for dealing with the water cluster is similar to the electrostatically embedded many-body expansion (EE-MB) method[1-3]. The difference is that in EE-GMFCC the distant two-body interaction is treated by classical Coulomb interactions to achieve linear scale for high-level *ab initio* calculation.

Based on the total energy expression of Eq. 1, the atomic forces can be derived from the differentiation of $E_{\text{water cluster}}^{\text{EE-GMFCC}}$ with respect to atomic coordinates. Because there are both real atoms and background charges in each QM calculation, the *m*th atom, which appears as a real atom in QM calculation of the *k*th fragment, may become a background charge in other QM fragment calculations. Hence, the force on the *m*th atom in the *k*th fragment is computed as follows,[4]

$$\boldsymbol{f}(\mathbf{r}_{\boldsymbol{m}}) = \begin{cases} -\dfrac{\partial \tilde{\boldsymbol{E}}_{\boldsymbol{k}}}{\partial \mathbf{r}_{\boldsymbol{m}}} & \text{(If m is a real atom)} \\ q(\mathbf{r}_{\boldsymbol{m}}) \times \overrightarrow{\boldsymbol{E}}_{\boldsymbol{k}}^{\text{bc}}(\mathbf{r}_{\boldsymbol{m}}) & \text{(If m is a background charge)} \end{cases} \qquad \text{(S1)}$$

If the *m*th atom appears as a real atom, we can obtain the derivative of $\tilde{E}_k$ (the self-energy of the *k*th fragment along with the interaction between the fragment and its background charges) with respect to $\mathbf{r}_{\boldsymbol{m}}$. In the case where the *m*th atom appears as a background charge, we calculate the electric field $\overrightarrow{\boldsymbol{E}}_{\boldsymbol{k}}^{\text{bc}}(\mathbf{r}_{\boldsymbol{m}})$ at the position $\mathbf{r}_m$. $q(\mathbf{r}_{\boldsymbol{m}})$ denotes the point charge of the *m*th atom. In this study, the fixed charge model of TIP3PF was utilized to describe the embedding field.

Based on Eq. S1, the overall force on the *m*th atom can be obtained analytically through the EE-GMFCC approach using the following expression,

$$\mathbf{F}^{\text{EE-GMFCC}}(\mathbf{r}_m) = \sum_{i=1}^{N} \boldsymbol{f}_i(\mathbf{r}_m) + \sum_{i=1}^{N-1} \sum_{\substack{j=i+1 \\ |\mathbf{R}_{\text{O}i} - \mathbf{R}_{\text{O}j}| \le \lambda}}^{N} \left(\boldsymbol{f}_{ij}(\mathbf{r}_m) - \boldsymbol{f}_i(\mathbf{r}_m) - \boldsymbol{f}_j(\mathbf{r}_m) \right) - \mathbf{F}_{\text{MM}}(\mathbf{r}_m) \qquad \text{(S2)}$$

where the last term is the derivative for the double counting long-range Coulomb interactions (arising from the last term of Eq. 1).[4]

2. *Ab initio* liquid water dynamics simulation

The initial simulation system consists of 2002 TIP3PF[5] water molecules in a truncated octahedron box with edges of 42.05 Å under periodic boundary conditions. After 100 ps equilibration run using classical force field in NPT ensemble, the equilibrated structure was used as the initial configuration for AIMD simulation.

To improve the computational efficiency, a QM/MM scheme is used in this study. The water molecules, whose oxygen atoms are less than or equal to 10 Å away from the center of the simulation box (approximately 140 water molecules in total), are treated by QM at the MP2/aug-cc-pVDZ level, while the rest of the system are described by MM. For the QM part, the EE-GMFCC approach is utilized to calculate the total energy and atomic forces, while the coupling between QM and MM regions is treated using the mechanical embedding scheme. Therefore, the total energy of the water system can be expressed as the sum of QM-water energy, QM-MM interaction energy, and MM-water energy as follows,

$$E_{\text{total}} = E_{\text{water}}^{\text{EE-GMFCC}} + E_{\text{water}}^{\text{QM/MM}} + E_{\text{water}}^{\text{MM}} \tag{S3}$$

The self-energy of MM-water molecules $E_{\text{water}}^{\text{MM}}$ is expressed as the summation of the intra-molecular bond energies and nonbonded electrostatic and van der Waals interactions using the TIP3PF model as follows,

$$E_{\text{water}}^{\text{MM}} = \sum_{i}^{N_{\text{MM-water}}} k_{\text{bond},i}(r_{\text{bond},i} - r_{\text{bond},i}^{0})^2 + \sum_{\substack{\text{non-bonded} \\ \text{atom pairs } m,n}} \left(\frac{1}{4\pi\varepsilon_0}\frac{q_m q_n}{r_{m,n}} + \frac{A_{m,n}}{r_{m,n}^{12}} - \frac{B_{m,n}}{r_{m,n}^{6}}\right) \tag{S4}$$

where $k_{\text{bond},i}$, $r_{\text{bond},i}^{0}$, q_m, q_n, $A_{m,n}$, $B_{m,n}$ are Amber force field parameters[6]. The electrostatic and van der Waals interactions between QM and MM regions ($E_{\text{water}}^{\text{QM/MM}}$) are calculated using the following expression,

$$E_{\text{water}}^{\text{QM/MM}} = \sum_{m}^{3N_{\text{QM-water}}} \sum_{n}^{3N_{\text{MM-water}}} \left(\frac{1}{4\pi\varepsilon_0}\frac{q_m q_n}{r_{m,n}} + \frac{A_{m,n}}{r_{m,n}^{12}} - \frac{B_{m,n}}{r_{m,n}^{6}}\right) \tag{S5}$$

Then, the atomic force of an given atom *m* in the system can be expressed as,

$$\mathbf{F}_m = -\nabla_m E_{\text{total}} \tag{S6}$$

In this study, AIMD simulation of liquid water was performed using a modified version of Amber14 package.[7] For each step of the simulation, atomic forces of the QM-water molecules were calculated using the EE-GMFCC approach and then passed to the MD engine (the Sander module) of Amber. The simulation was carried out in the canonical (NVT) ensemble under classical periodic boundary conditions at 300 K with a time step of 1 fs. The Langevin dynamics[8] was applied to regulate the temperature with a collision frequency of 2.0 ps^{-1}. The particle mesh Ewald (PME)[9] was employed to treat the long-range electrostatic interactions, and a 10 Å cutoff for the long-range electrostatic and van der Waals interactions was adopted. During the evolution of AIMD simulation, the water molecules in the QM and MM regions could exchange with each other. All coordinates written to the trajectory were wrapped into the primary simulation box. Fragment QM calculations were performed in parallel at the MP2/aug-cc-pVDZ level using the Gaussian09 package.[10]

The calculated radial distribution functions (RDF) are based on the central water molecule with the molecules in its first coordination shell in the simulation box. In order to calculate the distribution of triplet oxygen-oxygen-oxygen angles within the first coordination shell for the QM-water molecules in MD simulation, three oxygen atoms were considered as a triplet if two of the oxygen atoms were within a prescribed cutoff distance from the third, and this cutoff (3.25 Å[11] was used in this study) was chosen to yield an average oxygen-oxygen coordination number of around 4. The tetrahedral order parameter q is calculated by,[12]

$$q = 1 - \frac{3}{8}\sum_{i=1}^{3}\sum_{j=i+1}^{4}\left(\cos(\theta_{ij}) + \frac{1}{3}\right)^2 \qquad \text{(S7)}$$

where θ_{ij} is the angle formed by a given water molecule and its nearest neighbors i and j. The diffusion coefficient (D) is related to the mean square displacement (MSD) by Einstein's diffusion equation,

$$D = \frac{MSD}{6t} = \frac{1}{6t}\left\langle \left|\mathbf{r}(t) - \mathbf{r}(0)\right|^2 \right\rangle \qquad \text{(S8)}$$

The MSD was calculated from the squared relative displacement of the oxygen atoms and averaged over the QM-water molecules from all snapshots saved in the trajectory. The QM-water molecules, whose oxygen atoms are less than or equal to 8 Å away from the center of the simulation box are used to calculate the triplet oxygen-oxygen-oxygen angles, tetrahedral order parameter and diffusion coefficient.

3. The oxygen-hydrogen and hydrogen-hydrogen RDFs

The simulated oxygen-hydrogen RDF (g_{OH}) and hydrogen-hydrogen RDF (g_{HH}) (see Figure S1) of liquid water from EE-GMFCC are also in better agreement with the experimental observations[13] than the CPMD results using DFT. In both of the cases, the width and intensity of the first peaks obtained from EE-GMFCC-based AIMD and CPMD deviate significantly from the experiments, which is mainly due to the lack of nuclear quantum effect for describing the O-H covalent bond.[11] The classical treatment of the nuclear degrees of freedom in current AIMD simulations is not sufficient to capture the quantum nature of the lighter hydrogen atoms. However, the positions of the first peaks of g_{OH} and g_{HH} obtained from EE-GMFCC-based AIMD are more consistent with the experiment than those obtained from CPMD using DFT. Furthermore, for the second and third peaks of g_{OH} and g_{HH}, EE-GMFCC-based AIMD's results are in much better agreement with the experimental results than CPMD for both the positions of the peaks and their intensities.

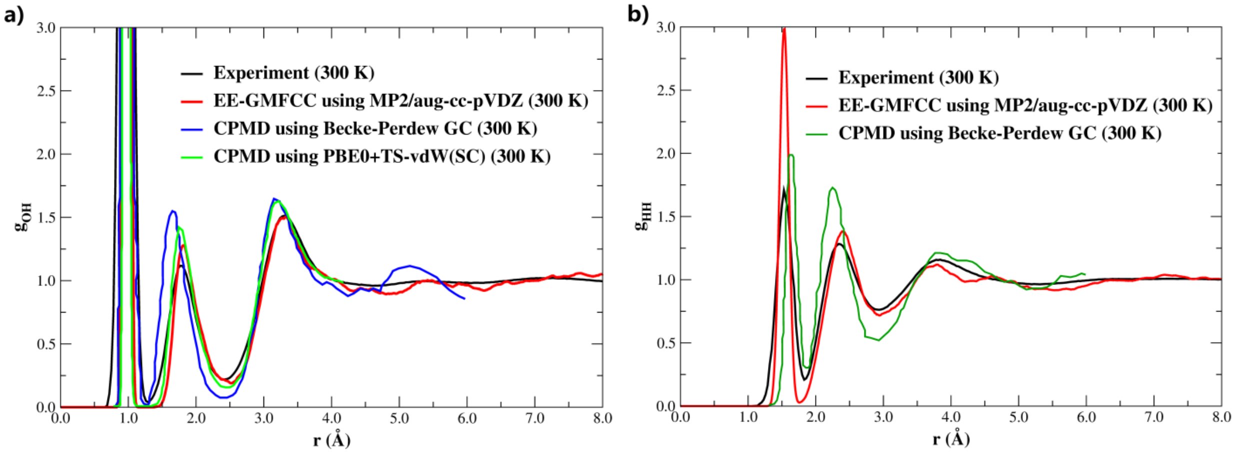

Figure S1. Oxygen-hydrogen (g_{OH}) and hydrogen-hydrogen (g_{HH}) radial distribution functions (RDFs) of liquid water under ambient conditions obtained from EE-GMFCC-based AIMD simulation and the CPMD results using different density functionals.[11, 14] The experimental curves are obtained from Ref. [13].

4. The diffusion coefficient

The MSD was calculated from the squared relative displacement of the oxygen atoms. The MSD of the QM-water molecules and the corresponding diffusion coefficient (*D*) as a function of simulation time are given in Figure S2.

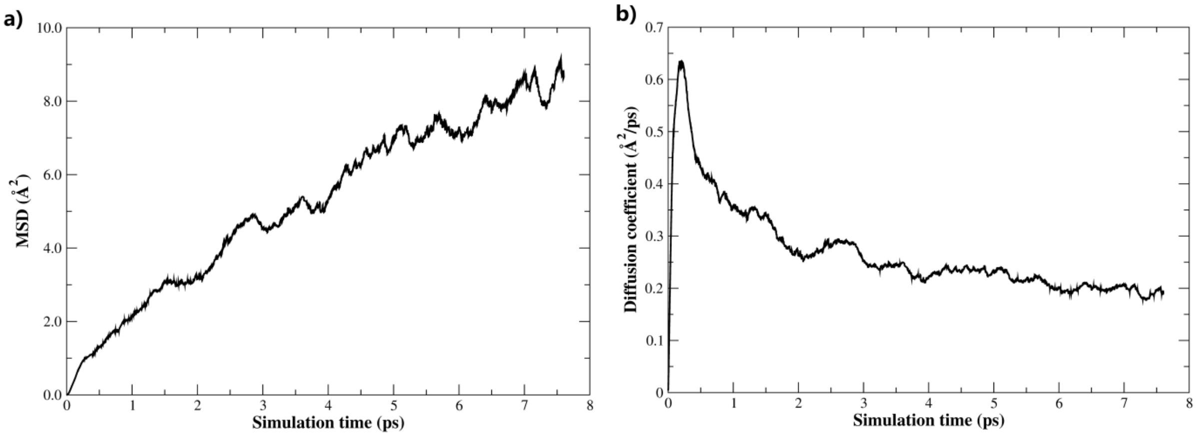

Figure S2. The MSD of the QM-water molecules and the corresponding diffusion coefficient (*D*) calculated from the EE-GMFCC-based AIMD simulation.

References:

1 E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 46-53.

2 E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 1342-1348.

3 E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.*, 2008, **4**, 1-6.

4 J. F. Liu, J. Z. H. Zhang and X. He, *Phys. Chem. Chem. Phys.*, 2016, **18**, 1864-1875.

5 D. J. Price and C. L. Brooks, *J. Chem. Phys.*, 2004, **121**, 10096-10103.

6 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696-3713.

7 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688.

8 R. W. Pastor, B. R. Brooks and A. Szabo, *Mol. Phys.*, 1988, **65**, 1409−1419.

9 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.

10 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V.

N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford, CT, USA.

11 R. A. DiStasio, B. Santra, Z. F. Li, X. F. Wu and R. Car, *J. Chem. Phys.*, 2014, **141**, 084502.

12 J. R. Errington and P. G. Debenedetti, *Nature*, 2001, **409**, 318-321.

13 A. K. Soper and C. J. Benmore, *Phys. Rev. Lett.*, 2008, **101**, 065502.

14 K. Laasonen, M. Sprik, M. Parrinello and R. Car, *J. Chem. Phys.*, 1993, **99**, 9080-9089.