Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2017

Supporting Information for

Hydrogen-bond Structure Dynamics in Bulk Water: Insights from *ab initio* Simulations with Coupled Cluster Theory

Jinfeng Liu[1#], Xiao He[2,3#], John Z. H. Zhang[2,3,4], and Lian-Wen Qi[1]*

[1]State Key Laboratory of Natural Medicines, Department of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing, 210009, China

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

[#]Jinfeng Liu and Xiao He contributed equally to this work.

*Corresponding author: Qilw@cpu.edu.cn

1. The EE-GMF scheme

The EE-GMF method was proposed for QM calculation of molecular clusters. For a water cluster, each water molecule is a single fragment. 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 SPCFW force field was utilized to describe the embedding field, while the positions of these point charges were updated during MD simulation to reflect the geometric 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-GMF approach is similar to the electrostatically embedded many-body expansion (EE-MB) method[1-3], except the difference that the distant two-body interaction is treated by classical Coulomb interactions in EE-GMF 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 $E_{\text{water}}^{\text{EE-GMF}}$ 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]:

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

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 $\mathbf{r}_m$. In the case where the *m*th atom appears as a background charge, we calculate the electric field $\vec{\boldsymbol{E}}_k^{\text{bc}}(\mathbf{r}_m)$ at the position $\mathbf{r}_m$. $q(\mathbf{r}_m)$ denotes the point charge of the *m*th atom. In this study, the fixed charge model of SPCFW 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-GMF approach using the following expression:

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

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* molecular dynamics simulation of the liquid water

The initial simulation system consists of 1997 SPCFW[5] water molecules in a truncated octahedron box with edges of 42.6 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 CCD/aug-cc-pVDZ level, while the rest of the system are described by MM. For the QM region, the EE-GMF 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:

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

where $k_{\text{O-H}}$, $r_{\text{O-H}}^{0}$, $k_{\angle\text{HOH}}$, $\theta_{\angle\text{HOH}}^{0}$, q_m, q_n, $A_{m,n}$, $B_{m,n}$ are force field parameters[5]. The electrostatic and van der Waals interactions between QM and MM regions ($E_{\text{water}}^{\text{QM/MM}}$) are calculated using the following expression:

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

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.[6] For each step of the simulation, atomic forces of the QM-water molecules were calculated using the EE-GMF approach and then passed to the MD engine (the Sander module) of Amber14. 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[7] was applied to regulate the temperature with a collision frequency of 2.0 ps^{-1}. The particle mesh Ewald (PME)[8] 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 CCD/aug-cc-pVDZ level using the Gaussian09 package.[9]

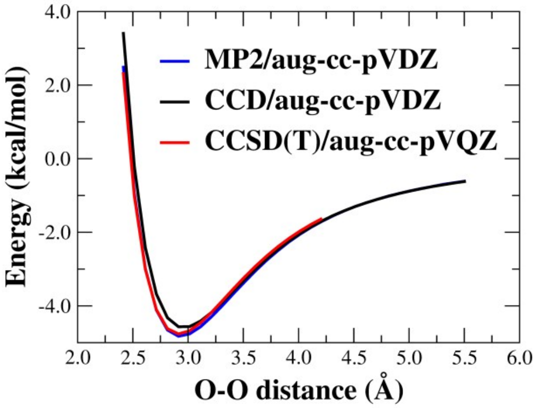

Figure S1. The interaction potential energy curve between two water molecules.

3. The structural and dynamical properties calculation

The last 7.0 ps AIMD simulation trajectory was used for the calculation of the structural and dynamical properties of water. The calculated radial distribution functions (RDFs) 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 oxygen-oxygen-oxygen triplet 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[10] 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[11]:

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

where θ_{ij} is the angle formed by a given water molecule and its nearest neighbors i and j. The angle distribution and q were calculated over the QM-water molecules, whose oxygen atoms are less than or equal to 8 Å away from the center of the simulation box.

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, whose oxygen atoms are less than or equal to 8 Å away from the center of the simulation box. The MSD of the QM-water molecules and the corresponding diffusion coefficient (D) as a function of simulation time are given in Figure S2.

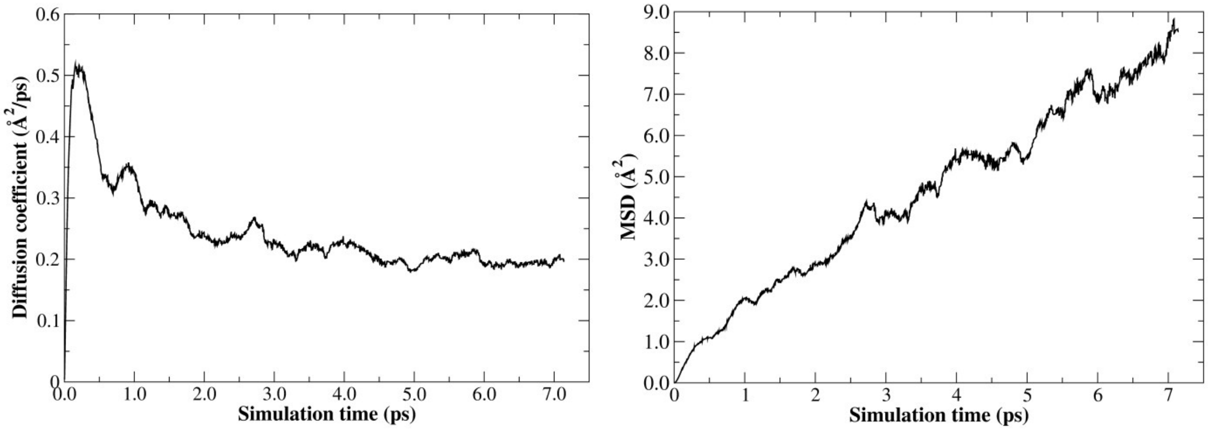

Figure S2. The MSD of the QM-water molecules and the corresponding diffusion coefficient (D) calculated from the EE-GMF-based AIMD simulation at the CCD/aug-cc-pVDZ level.

In principle, the QM water cluster size is required to be as large as possible for more accurate estimation of the statistical properties. In this simulation, the QM region is with a radius of 10 Å (i.e. water molecules whose oxygen atoms are less than or equal to 10 Å away) from the center of the simulation box. To reduce the QM/MM boundary effect, we utilized a reduced QM region with a radius of 8 Å (from the center of the simulation box) for calculation of water properties (we also used the same strategy in our previous study). To test the sensitivity of the estimated water properties to the QM size, we calculated the water diffusion coefficient as a function of the QM size. As shown in Figure S3, the obtained diffusion coefficient is close to converge when the radius of the QM size is larger than 7.5 Å. Therefore, the QM region with a radius of 8 Å for the prediction of the statistical water properties in this study is reasonable and acceptable.

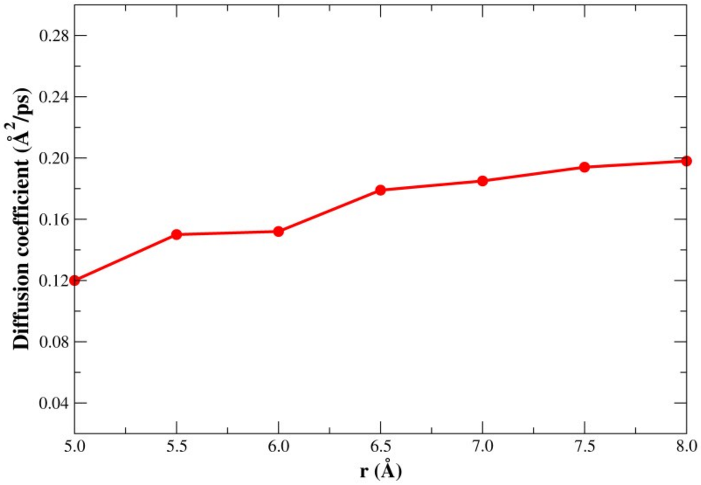

Figure S3. The calculated diffusion coefficient at the CCD/aug-cc-pVDZ level as a function of the radius of the QM water cluster size.

The dipole moments for the central water molecule and its first solvation shell in the QM region are calculated in the electrostatically embedding field for the plot of the molecular dipole distribution in Figure S4.

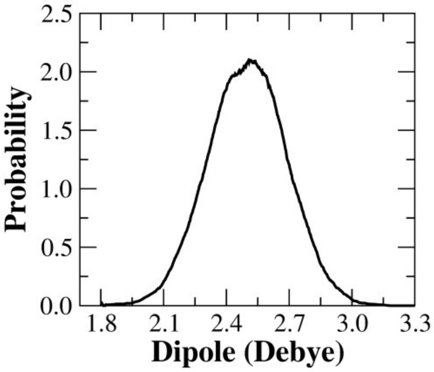

Figure S4. The distribution of dipole moments of the calculated water molecules.

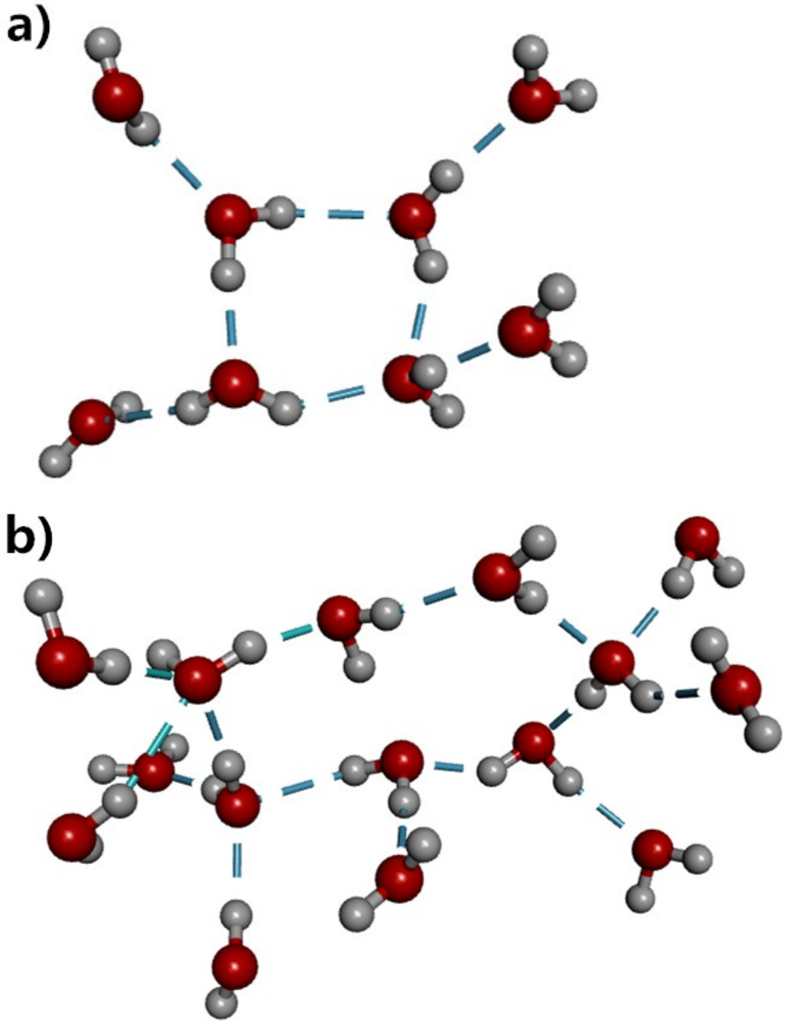

Figure S5. The representative hydrogen-bonded structures in the simulated liquid water.

Because the coupling between the QM and MM regions was treated through the mechanical embedding scheme, we did not observe imbalance in the water-water interactions between QM and MM regions during AIMD simulations in this and previous studies. This can be checked through the number of the water molecules in the QM region (see Figure S6). In this study, the QM region was treated by CCD/aug-cc-pVDZ, and we used the SPCFW water model for the MM region. In our previous study, the QM region was treated by MP2/aug-cc-pVDZ and the TIP3PF water model was used for the MM region. As can be seen from Fig. S6, the number of the water molecules in the QM region fluctuates within a certain range. For either SPCFW or TIP3PF force fields utilized for the MM region, there was no pronounced imbalance between QM and MM regions observed during MD simulation.

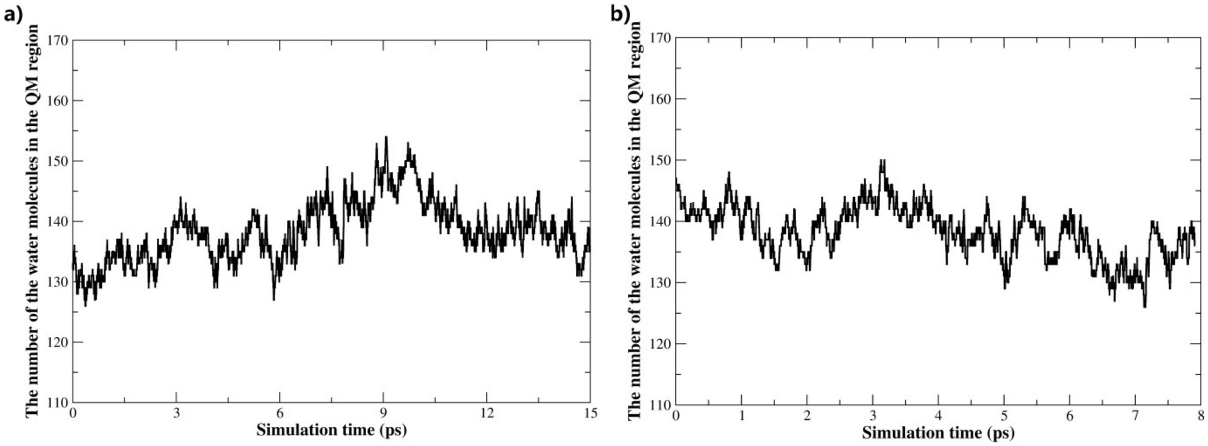

Figure S6. The number of the water molecules in the QM region during the *ab initio* molecular dynamics simulation at the a) CCD/aug-cc-pVDZ level with the MM region using the SPCFW water model and b) MP2/aug-cc-pVDZ level with the MM region using the TIP3PF water model, respectively.

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 Y. J. Wu, H. L. Tepper and G. A. Voth, *J. Chem. Phys.*, 2006, **124**, 024503.

6 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.

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

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

9 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*, Gaussian Inc., Wallingford, CT, US.

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

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