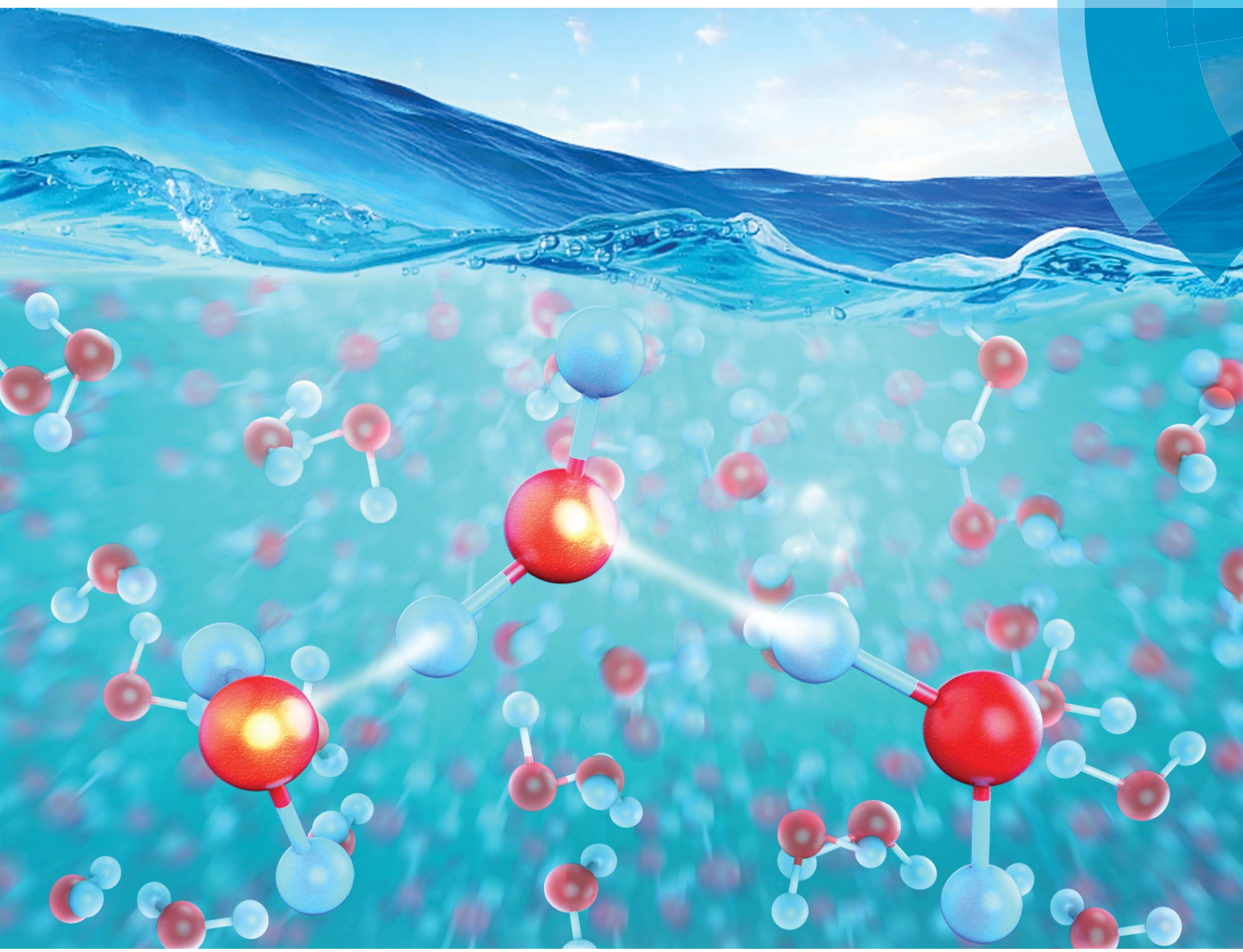
Volume 9 | Number 8 | 28 February 2018 | Pages 2015–2386

Chemical Science

rsc.li/chemical-science



ISSN 2041-6539

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

Lian-Wen Qi *et al.*

Hydrogen-bond structure dynamics in bulk water: insights from *ab initio* simulations with coupled cluster theory

Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2018, **9**, 2065

Hydrogen-bond structure dynamics in bulk water: insights from *ab initio* simulations with coupled cluster theory†

Jinfeng Liu, ‡[a] Xiao He, ‡[bc] John Z. H. Zhang[bcd] and Lian-Wen Qi *[a]

Received 26th September 2017
Accepted 4th December 2017

DOI: 10.1039/c7sc04205a

rsc.li/chemical-science

An accurate and efficient *ab initio* molecular dynamics (AIMD) simulation of liquid water was made possible using the fragment-based approach (J. F. Liu, X. He and J. Z. H. Zhang, *Phys. Chem. Chem. Phys.*, 2017, **19**, 11931–11936). In this study, we advance the AIMD simulations using the fragment-based coupled cluster (CC) theory, more accurately revealing the structural and dynamical properties of liquid water under ambient conditions. The results show that the double-donor hydrogen-bond configurations in liquid water are nearly in balance with the single-donor configurations, with a slight bias towards the former. Our observation is in contrast to the traditional tetrahedral water structure. The hydrogen-bond switching dynamics in liquid water are very fast, with a hydrogen-bond life time of around 0.78 picoseconds, determined using AIMD simulation at the CCD/aug-cc-pVDZ level. This time scale is remarkably shorter than the ~3.0 picoseconds that is commonly obtained from traditional nonpolarized force fields and density functional theory (DFT) based first-principles simulations. Additionally, the obtained radial distribution functions, triplet oxygen angular distribution, diffusion coefficient, and the dipole moment of the water molecule are uniformly in good agreement with the experimental observations. The current high-level AIMD simulation sheds light on the understanding of the structural and dynamical properties of liquid water.

1. Introduction

As the most abundant liquid on Earth, water is essential for life and is involved in nearly all biological, geological, and chemical processes.[1] Despite being extensively studied for several decades, there is still ongoing debate on the dynamical picture of the liquid water structure under ambient conditions.[1–11] For a microscopic understanding of liquid water, one of the most essential questions to address is the structure and dynamics of the hydrogen-bonding network in water which determine the unique water properties. This collective network fluctuation involving hydrogen-bond breaking and forming is of great importance for elucidating the dynamical picture of liquid water.

Understanding the nature of the dynamical hydrogen-bonding network in liquid water under ambient conditions is a challenge for both experimental and theoretical researchers. According to a series of spectral experiments and theoretical simulations,[12–17] liquid water is usually regarded as a tetrahedral structure involving minimally distorted hydrogen-bonding configurations. However, based on X-ray absorption and Raman scattering experiments, Wernet *et al.* concluded that the first coordination shell around a water molecule in liquid water has two hydrogen-bonding partners on average with one donor and one acceptor.[18,19] It favors a "ring-and-chain"-like structure in water, which is in contrast to the conventionally accepted tetrahedral structure of liquid water. It remains a topic of intense debate as to whether water has a tetrahedral or "ring-and-chain"-like structure.[20]

Another important issue is the hydrogen-bond switching dynamics in water. The dynamics of this network occur over a wide range of time scales, from femtosecond fluctuations that involve a few molecules to picosecond diffusive motions.[21,22] Specifically, Bakker *et al.* observed that orientational relaxation of the HDO molecules dissolved in D_2O occurred on either a very slow or a very fast time scale, with corresponding time constants of 13.0 and 0.7 picoseconds, respectively.[23–26] Fecko *et al.* showed that the hydrogen-bond vibrational correlations decay with a period of 0.17 and 1.2 picoseconds due to the underdamped oscillation of hydrogen-bond and collective

[a]State Key Laboratory of Natural Medicines, Department of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing, 210009, China. E-mail: Qilw@cpu.edu.cn

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

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

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

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7sc04205a

‡ These two authors contributed equally to this work.

structural reorganizations, respectively.[22] There are many other related experimental and theoretical investigations[10,27–29] demonstrating large variations in the time constants of the hydrogen-bond relaxation.

With experimental methods, it is normally difficult to yield detailed information on the hydrogen-bond dynamics in water.[30] Understanding the specific molecular structure and dynamics of liquid water at the atomistic level depends on molecular dynamics (MD) simulations. MD with empirical force fields have already provided fundamental insights into the microstructural and dynamical properties of water, with continuous improvement of the traditional force fields.[31–38] *Ab initio* molecular dynamics (AIMD) simulations, with the molecular potentials described by the first-principles electronic structure theory, are a significant improvement over the empirical force fields.[39–45] Density functional theory (DFT) has been the most widely used electronic structure method in AIMD simulations with reasonable accuracy and moderate computational cost.[46] However, the performance of DFT-based AIMD simulations is dramatically affected by the choice of density functionals.[46,47] High-level wavefunction theories, such as the second order Møller–Plesset perturbation (MP2) method, have also been utilized to obtain an accurate description of the structural and dynamical properties of liquid water.[43–45,48] In our previous work, a MP2-based AIMD simulation of liquid water was carried out on a large body of water molecules (~140 water molecules) through the fragmentation quantum mechanical (QM) method,[49–53] and the simulated water structural and dynamical properties were uniformly in good agreement with experimental observations.[53]

It is of great importance and necessity to apply high-level wavefunction theories for an accurate simulation of the microstructural and dynamical properties of liquid water. To the best of our knowledge, the coupled cluster (CC) theory has not been utilized to treat a large body of water molecules for an AIMD simulation of liquid water, owing to the high scaling of the CC method with tremendous computational cost. Here we report substantial progress toward an accurate liquid water dynamics simulation using the fragment-based CC method. The nuclear quantum effect (NQE) is not included in this study. The NQE of a hydrogen atom in liquid water is mainly associated with O–H vibration-related properties, and most of the statistical properties of water such as diffusion or the O–O radial distribution function (RDF) are not much affected by the NQE as demonstrated in recent work by Marsalek and Markland.[54] The good agreement between the experimental observations and the calculated properties of water without explicit inclusion of the NQE in previous studies[46,53] implied that the influence of the NQE on hydrogen atoms has subtle effects on most of the statistical properties of liquid water. This study sheds light on a detailed understanding of the water structure and hydrogen-bonding dynamics in liquid water from AIMD simulations using the fragment-based CC method.

2. Results and discussion

2.1. Water properties

2.1.1. RDF. In order to characterize the structure of water, we first calculated its intermolecular oxygen–oxygen (g_{OO}), oxygen–hydrogen (g_{OH}), and hydrogen–hydrogen (g_{HH}) radial distribution functions (RDFs). The obtained data were compared to the experimental observations,[55,56] as well as the results obtained from the authors' previous study performed at the MP2/aug-cc-pVDZ level.[53] As shown in Fig. 1, the simulated g_{OO} curve obtained using the CCD method is in excellent agreement with the experimental results for both the positions and intensities of the first two peaks. The g_{OO} curve obtained using the MP2 method is also in good agreement with the experimental results, except that the trough between the first two peaks is slightly lower compared to the CCD and experimental results. Moreover, the MP2 RDF curve is not as smooth as that given by CCD owing to the relatively shorter simulation time at the MP2 level in the authors' previous study. For the g_{OH} and g_{HH} curves, both of the intensities of the first peaks obtained from the MP2 simulation are overestimated in comparison with the experimental results, which is mainly due to the lack of the NQE in the simulation. In comparison, the CCD results significantly reduce the intensities of the first peaks for the g_{OH} and g_{HH} curves, although they are still overestimated in reference to the experimental results. For both the MP2 and CCD results, the second peaks of g_{OH} and g_{HH} slightly shift to higher radial positions in comparison with the experimental values, while the positions of the third peaks for g_{OH} and g_{HH} are both in good agreement with the experimental results. Overall, our AIMD simulated results using the fragment-based CCD/aug-cc-pVDZ method in this study are in good accordance with the experimental observations.

Recently, Hirata and co-workers suggested that with such a small basis set as aug-cc-pVDZ, RDFs cannot generally be reproduced accurately.[45] Furthermore, their subsequent study revealed a precise mechanism by which the adjustments of density, temperature, and/or pressure would lead to correct RDFs in MP2 and DFT water simulations.[57] According to their study, the MP2 simulated water tends to be denser (ρ > 1.00 g cm^{-3}) under ambient conditions due to the overestimated dispersion interaction, and thus requires a lower temperature and negative pressure to generate the experimentally observed RDF of liquid water. In this study, the initial density of the system was equilibrated to 1.002 g cm^{-3}. During the AIMD simulation, the average water density of the QM region was 1.008 g cm^{-3}, which was slightly larger than the experimental value under ambient conditions. There are some approximations made in this study, and the simulated water properties may result from a variety of factors, such as the utilized canonical NVT ensemble for the whole system, Langevin thermostat for adjusting the system temperature and the mechanical embedding scheme for treating the QM/MM coupling. The AIMD simulation details are summarized in the theory and computation section and in the ESI.† The agreement of the results with the experimental results is partially due to the

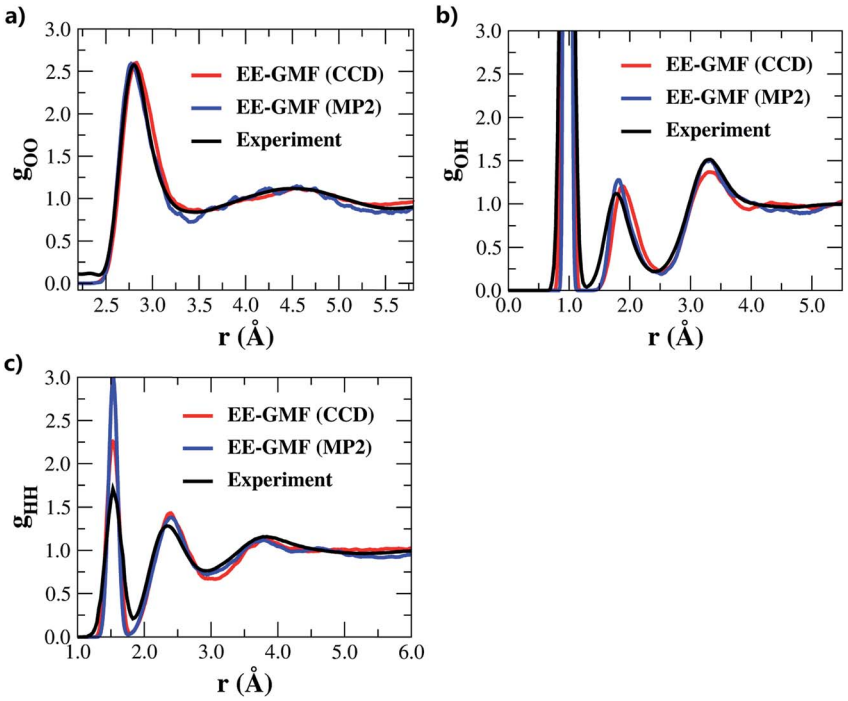


Fig. 1 (a) Oxygen–oxygen, (b) oxygen–hydrogen and (c) hydrogen–hydrogen radial distribution functions (RDF) of liquid water under ambient conditions obtained using the fragment-based AIMD simulations at the MP2/aug-cc-pVDZ[53] and CCD/aug-cc-pVDZ levels, respectively.

fortuitously accurate intermolecular interaction potential described at the MP2/aug-cc-pVDZ or CCD/aug-cc-pVDZ level, with reference to the potential calculated at the CCSD(T)/aug-cc-pVQZ level (see Fig. S1 of the ESI†).

2.1.2. Oxygen–oxygen–oxygen triplet angles. To further analyze the local arrangement of the water molecules in the liquid phase, we calculated the distribution of oxygen–oxygen–oxygen triplet angles within the first coordination shell and tetrahedral order parameter q for the QM water molecules in simulation. The CCD result obtained in this study and the MP2 simulated result from the authors' previous study[53] are compared in Fig. 2 along with the experimental curve. Three oxygen atoms were considered as a given triplet if two of the oxygen atoms are within a predefined distance cutoff from the third. The computational details are given in the ESI.† The experimental angular distribution shows a shoulder at around 60° and a broad and strong peak at around 100°.[56] The CCD simulated angular distribution curve is in good agreement with the experimental curve, with one distinct shoulder at around 70° and a strong peak at around 100°, and the intensity of the strong peak exactly matches the experimental observation. In contrast, for the MP2 simulated results, the intensity of the strong peak is underestimated and the shoulder is overestimated. The tetrahedral order parameter q, an index to describe the similarity of the simulated water structure with the perfect tetrahedral structure, would yield a value of 1 for the perfect tetrahedral structure and 0 for the ideal gas. The tetrahedral order parameter q is 0.515 based on the fragment-based CCD/aug-cc-pVDZ simulation, close to our previous MP2 simulated value of 0.520 and the experimental value of 0.570. The current result is more accurate than 0.670, which was obtained from a DFT-based Car–Parrinello molecular dynamics (CPMD) simulation using PBE0 + TS-vdW(SC).[47]

2.1.3. Diffusion coefficient. The dynamical properties of liquid water can be measured from the diffusion coefficient. The convergence of the diffusion coefficient as a function of the simulation time and the QM water cluster size are shown in

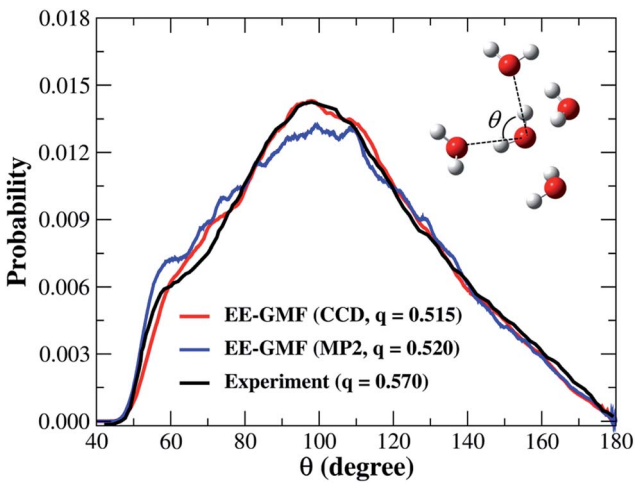


Fig. 2 The oxygen–oxygen–oxygen triplet angular distribution and tetrahedral order parameter of liquid water obtained from the fragment-based AIMD simulation at the MP2/aug-cc-pVDZ[53] and CCD/aug-cc-pVDZ levels, respectively. The water molecules whose oxygen atoms are less than or equal to 8 Å away from the center of the water box were used for this analysis.

Fig. S2 and S3 of the ESI.† The calculated diffusion coefficient at 300 K converged to 0.20 $Å^2$ ps^{-1} in the last 7.0 ps AIMD simulation, as compared to the experimental value of 0.23 $Å^2$ ps^{-1}.[58] This result is consistent with our previous study.[53] The empirical force fields usually predict a rather larger diffusion coefficient (>0.26 $Å^2$ ps^{-1}),[59] while the DFT-based AIMD simulations generally give a relatively smaller value (<0.20 $Å^2$ ps^{-1}).[46,47]

2.1.4. Dipole moment. The dipole moment of a water molecule in the liquid phase has been under debate for a long time, partially because of the uncertainty on its value (2.6–3.0 Debye) measured from the X-ray scattering form factors.[60–62] The current AIMD simulation also reveals the distribution of the molecular dipole of water as shown in Fig. S4 of the ESI.† The molecular dipole is broadly distributed from 1.8 to 3.2 Debye with an average value of 2.53 Debye, reflecting the diversity of the electrostatic fields of the environment and the local geometry fluctuation of the water molecules. The predicted dipole moment from the DFT-based CPMD simulation is relatively larger (around 2.90 Debye),[47] and it also heavily depends on the choice of density functionals.

Hence, for these basic properties of liquid water, the fragment-based AIMD simulation at the CCD/aug-cc-pVDZ level gives reasonable results that are uniformly in good agreement with the experimental observations. Based on these results, we further investigated the detailed microstructure and hydrogen-bond dynamics of liquid water from the AIMD simulated trajectory.

2.2. Water microstructure

Controversy regarding whether liquid water has a tetrahedral or "ring-and-chain"-like microstructure still persists.[19,20] According to Wernet *et al.*,[18] water molecules with single hydrogen-bond donors (SD) are dominant in liquid water. However, nearly all of the theoretical studies based on force fields or DFT overwhelmingly favor the tetrahedral structure, in which water molecules with the double hydrogen-bond donor (DD) configurations are dominant in the liquid phase.[47] From our previous fragment-based AIMD study at the MP2/aug-cc-pVDZ level,[53] we found that both of these two configurations (SD and DD) were almost equally distributed in the liquid phase, with a slight bias towards the tetrahedral structure.

In this study, we made quantitative statistics of the number and types of hydrogen-bond configurations in liquid water by utilizing the popular hydrogen-bond definition proposed by Luzar and Chandler.[63] The defining parameters for hydrogen bonds include both distance and angular criteria, namely, r_{OO} < 3.5 Å and $\theta_{\angle OA\cdots OD-HD}$ < 30° (Fig. 3a). Using this definition, we first analyzed the difference among the hydrogen-bonds in CCD simulated liquid water using the 2D weighted histogram analysis method (WHAM-2D)[64] (see Fig. 3b). We also compared it with the MP2 result from the authors' previous study[53] (Fig. 3c). The free energy landscape clearly shows that the most probable region corresponding to the ideal hydrogen-bond in CCD simulated liquid water is with the distance r_{OO} in the range of 2.8–3.0 Å and the angle θ around 8–15° (not 0°), respectively. The other regions under the hydrogen-bond definition represent the non-ideal or bent hydrogen-bond configurations. In comparison with the CCD result, the hydrogen-bonds formed in the MP2 simulated liquid water show a very similar pattern except that the MP2 simulated ideal hydrogen-bond has a relatively shorter r_{OO} distance (in the range of 2.7–2.9 Å), which is caused by the relatively stronger intermolecular interaction calculated using MP2/aug-cc-pVDZ (see Fig. S1†). The ideal hydrogen-bonds are usually more stable than those non-ideal or bent hydrogen-bonds. The deviation of the distance r_{OO} and angle θ from their most probable region would decrease the stability of the hydrogen bonds in liquid water. Hence, these strong and weak hydrogen-bonds in liquid water would lead to different hydrogen-bond dynamics.

The fractions of double-donor (DD), single-donor (SD), and non-donor (ND) configurations from the last 7 ps AIMD simulation are given in Table 1. According to Wernet *et al.*,[18] around 80% of water molecules in the liquid phase donate only one hydrogen to form a hydrogen bond with their neighbors, whereas DD and ND configurations only account for 15% and 5%, respectively. However, the DFT-based CPMD simulation[47] predicted the opposite trend, *i.e.*, 79% of water molecules are in the DD configuration and only 20% of them are in the SD configuration. In addition, the result obtained from the empirical force field (SPCFW) also shows a similar pattern, with dominant 70% DD configurations and only 27% SD configurations. The present fragment-based AIMD simulation at the CCD/aug-cc-pVDZ level predicts a more balanced picture of the hydrogen-bond structure in liquid water. Specifically, our result shows that the DD configuration accounts for just 50% of the overall population, which is closely followed by 42% SD configurations (Table 1). This result is also in accordance with the predicted tetrahedral order parameter of 0.515, which means that the structure of water presents a mixture of two configurations with a slight bias toward the DD configuration (tetrahedral structure) over the SD configuration ("ring-and-chain"-like structure). Our previous fragment-based AIMD simulation at the MP2/aug-cc-pVDZ level[53] also reached a similar conclusion.

To provide more insight into the microstructure of liquid water, some representative hydrogen-bond network structures were extracted from our AIMD simulated trajectory as shown in Fig. 4 and S5.† Both the tetrahedral and "ring-and-chain"-like structures of water are present in this simulation. As shown in Fig. 4a, the traditional picture of a water molecule, accepting and donating two hydrogen bonds to form a tetrahedral structure, partly exists in our simulation. All three inner water molecules form tetrahedral structures in the snapshot shown in Fig. 4a. On the other hand, in Fig. 4b, there are ring-like structures which are formed by a group of DD and SD configurations of the water molecules. There are also unclosed ring-like (or chain-like) structures (see Fig. 4c), in which the central water molecule does not donate any hydrogen bond (in the ND configuration) at the end of the chain. The ND configuration may arise from local coordination of the hydrogen-bond network to form a temporarily meta-stable state. All of these hydrogen-bonding structures change dynamically throughout the entire AIMD simulation. The tetrahedral, ring-like and

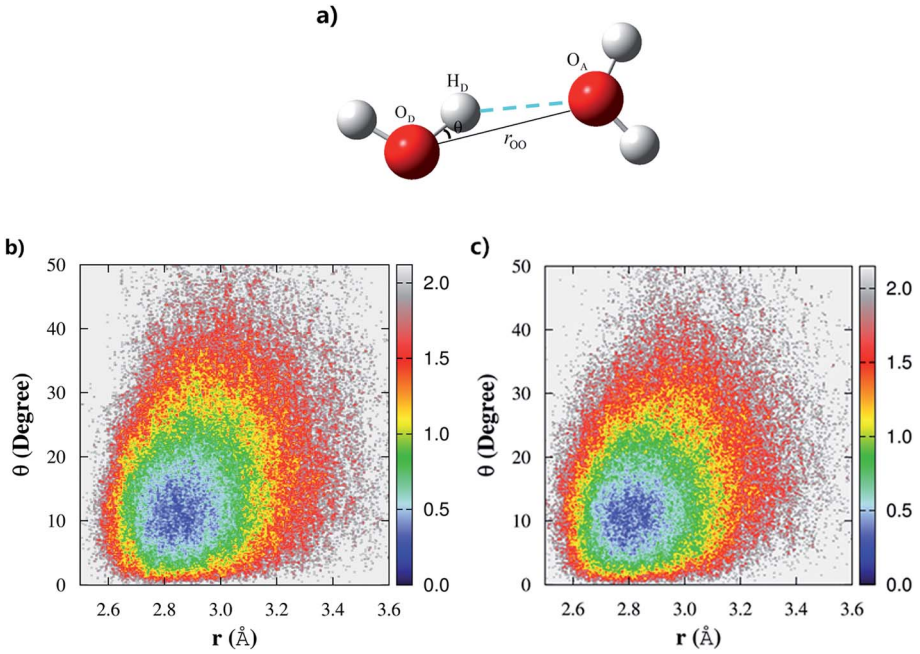


Fig. 3 (a) The definition of a hydrogen-bond between two water molecules and the free energy landscape of the hydrogen-bonds in the (b) CCD/aug-cc-pVDZ and (c) MP2/aug-cc-pVDZ simulated liquid water. The definition for a hydrogen-bond is that the distance between two oxygen atoms r_{OO} < 3.5 Å and $\theta_{\angle OA\cdots OD-HD}$ < 30°.

Table 1 The percentage (%) of hydrogen-bond configurations: double-donor (DD), single-donor (SD), and non-donor (ND) configurations in liquid water from different methods

		Method			
Type	Exp[a]	CCD[b]	MP2[c]	SPCFW	CPMD[d]
DD	15 ± 25	50	53	70	79
SD	80 ± 20	42	40	27	20
ND	5 ± 5	8	7	3	1

[a] The experimentally fitted percentage of hydrogen-bond configurations in liquid water from ref. 18. [b] The EE-GMF approach at the CCD/aug-cc-pVDZ level from this study. [c] The EE-GMF approach at the MP2/aug-cc-pVDZ level from ref. 53. [d] CPMD simulation results (using PBE0+TS-vdW(SC)) from ref. 47, which utilized the same definition of hydrogen bonds as this study.

chain-like structures interchange with each other from time to time, resulting in a dynamical picture of a mixture of tetrahedral and "ring-and-chain"-like structures in liquid water as indicated from the current AIMD simulation.

2.3. Hydrogen-bond dynamics

The hydrogen-bond dynamics in liquid water take place in a very fast manner.[10] To quantitatively describe the dynamical properties of water, we first computed the residence time correlation function $S(t)$ for the water molecules around the central water molecule and their first coordination shell in the QM region using the following equation,[65]

$$S(\tau_k) = \sum_{j=1}^{N_{t_{max}}-k} \sum_{i=1}^{N_{H_2O}} v_i(t_j)\, v_i(t_j + \tau_k) \quad (1)$$

where τ_k is the kth time interval ($\tau_k = k\Delta\tau$; $\Delta\tau = 1$ fs; $k = 0, 1, \ldots N_{t_{max}} - 1$), v_i(t) becomes 1 or 0 depending on the water molecule i within or outside the shell defined by a radius around a tagged water molecule at time t, and N_{H_2O} is the number of water molecules in the QM region. The $S(t)$ measures the average number of water molecules which continuously stay around the tagged water molecule. In this simulation, $S(t)$ decays in an exponential fashion, and the time constant associated with this decay gives a measure of the residence time of water around a specific site. Fig. 5a shows the calculated $S(t)$ along with the exponential fit $S(t) = \exp(-t/\tau)$ using a radius cutoff of 3.5 Å from the last 10 ps AIMD simulation. The residence time τ of a water molecule obtained from the exponential fit is 2.0 ps given by the fragment-based AIMD simulation at the CCD/aug-cc-pVDZ level, indicating that the diffusion of the water molecules in the liquid phase is very fast.

Furthermore, we investigated the hydrogen-bond dynamics by introducing the hydrogen-bond correlation function[66]

$$C(t) = \frac{\langle h(0)h(t)\rangle}{\langle h\rangle} \quad (2)$$

where $h(t)$ is a hydrogen-bond population descriptor, which is unity when a tagged pair of molecules is hydrogen-bonded at time t and is zero otherwise. This correlation function is calculated for all QM water molecules defined to be hydrogen-bonded at both times 0 and t, which describes the probability of the tagged pair of molecules being hydrogen-bonded at time t given that the pair was hydrogen-bonded at time 0. The calculated hydrogen-bond correlation function is shown in Fig. 5b, which is best described by a series of two exponential decay functions. Two time constants obtained from the exponential fit

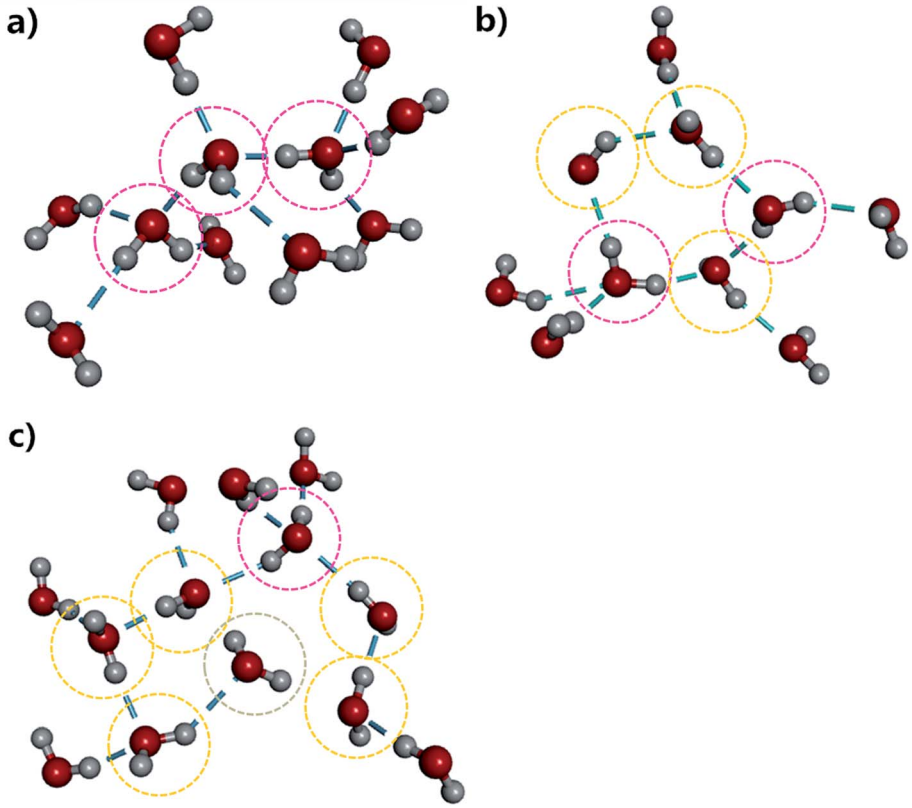


Fig. 4 The representative hydrogen-bond structures in the simulated liquid water. The dashed line denotes the hydrogen-bond. The pink, yellow and grey circles label the double-donor (DD), single-donor (SD) and none-donor (ND) configurations of the hydrogen-bonded water molecules, respectively.

are identified as the average hydrogen bond lifetime,[2] and are 0.78 and 5.11 ps, respectively. This result indicates that the hydrogen-bond dynamics in liquid water occur on either a very fast or a slow time scale. This result is in agreement with Bakker and coworkers' experiment,[23] where they measured that the hydrogen-bonds in liquid water decay on either a very fast or a slow time scale (0.7 and 13.0 ps for HDO dissolved in D_2O), corresponding to the weak and strong hydrogen bonds in liquid water, respectively. However, the hydrogen-bond life time obtained from the empirical force fields and previous DFT-based CPMD simulations are normally overestimated, with the fast and slow time scales around 3.0 ps and 18.0 ps, respectively.[2] Therefore, we conclude that the fragment-based AIMD simulation at the CCD/aug-cc-pVDZ level can well describe the ultrafast hydrogen-bond dynamics in liquid water, which would help to better understand the dynamic properties of liquid water.

At present, there are still no consistent conclusions on the impact of the NQE on the structure and dynamics of water hydrogen-bonds. According to a series of studies,[67–69] the competition between the intra- and intermolecular quantum

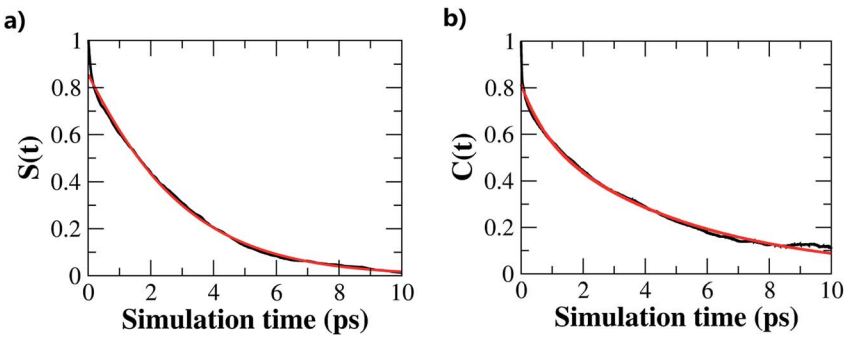


Fig. 5 (a) Residence time correlation function $S(t)$ for the water molecules around the central water molecule and their first coordination shell in the QM region (normalized to $S(t_{min}) = 1$). (b) Hydrogen-bond correlation function $C(t)$ for the water molecules in the reduced QM region. The red curve is the corresponding exponential fit.

effects in water further reduces the impact of neglecting the NQEs on the average statistical properties of liquid water. Li *et al.* carried out systematic examination of a wide range of hydrogen-bonded systems through the *ab initio* path integral molecular dynamics, and showed that NQEs weaken weak hydrogen bonds but strengthen relatively strong ones.[70] Furthermore, from their study, the NQEs would accelerate the hydrogen dynamics for the weak hydrogen-bonds. On the other hand, for the strong hydrogen-bonds, the NQEs would slow down their hydrogen dynamics. More recently, Wilkins *et al.* combined classical and ring polymer molecular dynamics simulations to provide a molecular description of the NQEs on water reorientation and hydrogen-bond dynamics in liquid water. They showed that NQEs lead to a moderate acceleration (~13%) of the hydrogen-bond dynamics as compared to a classical description.[71]

3. Conclusions

In this study, we investigated the structural and dynamical properties of liquid water under ambient conditions using fragment-based AIMD simulation at the CCD/aug-cc-pVDZ level. The water properties, such as the RDFs, diffusion coefficient, molecular dipole and oxygen–oxygen–oxygen triplet angular distribution, are uniformly in excellent agreement with the experimental observations, which demonstrates the robustness of the EE-GMF method for AIMD simulation with high-level wavefunction theories.

Furthermore, we investigated the hydrogen-bond structures and dynamics in liquid water and gave a clear dynamic picture of liquid water. We found that the tetrahedral and "ring-and-chain"-like structures were almost equally distributed in liquid water, with around 50% of water molecules donating two hydrogen-bond donors to their neighbors (DD configuration) and 42% donating just one hydrogen-bond donor (SD configuration). In addition, the hydrogen-bonds in liquid water decay on either a very fast or a slow time scale, which correspond to short and long hydrogen-bond life times of 0.78 and 5.11 ps, respectively.

This study suggests that, using the aug-cc-pVDZ basis set, the overall performance of CCD is just slightly better than MP2 in describing the structural and dynamical properties of liquid water. The larger basis sets (such as aug-cc-pVTZ and aug-cc-pVQZ) may be needed to systematically investigate the relative merits between MP2 and CCD. The EE-GMF approach could in principle implement a systematic series of diagrammatic many-body theories (even for the CCSD and CCSD(T) methods) in conjunction with also systematic basis sets, together converging towards the exact limits for liquid water simulation. It is general in the choice of electronic structure theory for monomers and dimers, and most importantly exhibits linear scaling in the computational cost with respect to the system size. However, for EE-GMF calculations using high-level *ab initio* theories, the larger basis sets would significantly increase the computational cost on each fragment QM calculation. In future studies, we will utilize more efficient techniques, such as the density fitting algorithm and other local methods in fragment QM calculations, to accelerate such high-level AIMD simulations.

The current approach also has limitations. First, the transition of potentials at the QM/MM boundary is not smooth when water molecules leave or enter the QM region. This may cause some artificial effects on the structural and dynamical properties of the simulated liquid water near the QM/MM boundary, although we used a reduced QM region for analyzing the liquid water properties to alleviate the boundary effect. Second, the NQE is not included in this work, which may have an impact on the hydrogen-bond dynamics simulation. Improvements on the fragment-based AIMD simulation along these lines are currently underway in our laboratory.

4. Theory and computation

In this study, we employed the electrostatically embedded generalized molecular fractionation (EE-GMF)[51,52] to perform the *ab initio* calculations of a large body of water molecules. According to the EE-GMF scheme, the total energy of a water cluster can be expressed as follows,

$$E^{\text{EE-GMF}}_{\text{water cluster}} = \sum_{i=1}^{N} \tilde{E}_i + \sum_{i=1}^{N-1} \sum_{\substack{j=i+1 \\ |R_{ij}| \le \lambda}}^{N} \left(\tilde{E}_{ij} - \tilde{E}_i - \tilde{E}_j\right) - \sum_{i=1}^{N-1} \sum_{\substack{j=i+1 \\ |R_{ij}| > \lambda}}^{N} \times \sum_{m\in i} \sum_{n\in j} \frac{q_{m(i)} q_{n(j)}}{R_{m(i)n(j)}} \quad (3)$$

where N is the number of water molecules. The QM energy calculations of monomer i (one water molecule) and dimer ij (two water molecules) are performed in the embedded electrostatic fields of the rest of the water molecules, represented by the Coulomb field of atomic charges, to account for the electronic polarization effect from the surrounding environment. $\tilde{E}$ in eqn (3) denotes the sum of the self-energy of the fragment along with the interaction energy between the fragment and background charges of the remaining system. $\tilde{E}_i$ and $\tilde{E}_{ij}$ are the energies of the water monomer i and dimer ij, respectively. The first term of eqn (3) is the summation of one-body energies. The second term is the local two-body QM interaction when the distance R_{ij} between any two oxygen atoms from monomers i and j is less than or equal to a predefined distance threshold λ.[53] Otherwise, the interaction energy between two water molecules ($R_{ij} > \lambda$) is described by the classical Coulomb interaction. Therefore, the monomer energy and local pairwise interactions are explicitly treated by QM, while the higher-ordered many-body electrostatic interactions are implicitly incorporated using the electrostatic embedding scheme. Because of the charge embedding scheme utilized in fragment QM calculations, the long range interactions are included in each fragment QM calculations. The doubly counted electrostatic interactions between distant water pairs (beyond the distance threshold λ) need to be deducted using the last term of eqn (3) through the classical Coulomb interactions, where $q_{m(i)}$ denotes the atomic charge of the mth atom in the ith water molecule. The atomic charges from the SPCFW[72] water model were utilized as the

embedding field in this study. The distance threshold λ was set to 5.0 Å for ensuring the convergence of the total EE-GMF energy of the water cluster, while achieving an efficient computation. The detailed description of the EE-GMF method can be found in a series of recent publications[51–53] and the ESI.†

A truncated octahedron box with edges of 42.6 Å (containing a total of 1997 flexible SPCFW[72] water molecules) under periodic boundary conditions was constructed for simulation in this study. Before the AIMD simulation, the density of the system was equilibrated to 1.002 g cm^{-3} using the classical MD simulation. To improve the computational efficiency, a QM/MM scheme is utilized (the same as our previous study[53]), in which water molecules with oxygen atoms that are less than or equal to 10 Å away from the center of the simulation box are treated by QM, while the rest of the water molecules are described by MM. The EE-GMF approach is employed to calculate the total energy (eqn (3)) and atomic forces (see the ESI†) of the QM region (approximately 140 water molecules). The coupling between the QM and MM regions is treated by mechanical embedding. For all water molecules, the intramolecular bonds are fully flexible. The technical details of the AIMD simulation are given in the ESI.† To reduce the boundary effect, a buffer zone of about 2 Å from the QM/MM boundary was established. We used a reduced QM region (the radius of which is 8.0 Å from the center of the simulation box) for analyzing the physicochemical properties of liquid water.

The CCD/aug-cc-pVDZ method was employed for the AIMD simulation of liquid water under ambient conditions. The wall clock time for each MD step was around 3.5 minutes on 30 computer nodes with the Intel Xeon E5-4640 2.4 GHz processor (28 cores per node). The interaction potential energy between two water molecules calculated at the CCD/aug-cc-pVDZ level is shown in Fig. S1 of the ESI.† Prior to the production run of AIMD, the initial water structure was pre-equilibrated by classical MD simulation using the SPCFW force field. Subsequently, a total of 15.0 ps AIMD simulation was carried out. This simulation time is found to be sufficient in converging the predicted water properties as shown in the ESI.† During the AIMD simulation, the transition of the water molecules between the QM and MM regions, which is quantified as the number of the water molecules in the QM region, is shown in Fig. S6 of the ESI.†

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

This work was supported by the National Key R&D Program of China (Grant No. 2016YFA0501700), National Natural Science Foundation of China (No. 21703289, 91639115, 21673074 and 21433004), Youth Top-Notch Talent Support Program of Shanghai, and NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai Putuo District (Grant 2014-A-02), and NYU Global Seed Grant for Collaborative Research. We thank the Supercomputer Center of East China Normal University for providing us with computational time. We also thank Dr Raphael N. Alolga for helpful comments on the manuscript.

References

1 L. G. M. Pettersson, R. H. Henchman and A. Nilsson, *Chem. Rev.*, 2016, **116**, 7459–7462.
2 Y. Ding, A. A. Hassanali and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 3310–3315.
3 D. C. Clary, *Science*, 2016, **351**, 1268–1269.
4 G. Stirnemann, E. Wernersson, P. Jungwirth and D. Laage, *J. Am. Chem. Soc.*, 2013, **135**, 11824–11831.
5 A. Hassanali, F. Giberti, J. Cuny, T. D. Kuhne and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 13723–13728.
6 Y. Marcus, *Chem. Rev.*, 2009, **109**, 1346–1370.
7 C. A. Angell, *Science*, 2008, **319**, 582–587.
8 J. R. Errington and P. G. Debenedetti, *Nature*, 2001, **409**, 318–321.
9 H. J. Bakker and J. L. Skinner, *Chem. Rev.*, 2010, **110**, 1498–1517.
10 D. Laage and J. T. Hynes, *Science*, 2006, **311**, 832–835.
11 A. Tokmakoff, *Science*, 2007, **317**, 54–55.
12 J. D. Smith, C. D. Cappa, K. R. Wilson, B. M. Messer, R. C. Cohen and R. J. Saykally, *Science*, 2004, **306**, 851–853.
13 T. D. Kuhne, M. Krack and M. Parrinello, *J. Chem. Theory Comput.*, 2009, **5**, 235–241.
14 M. V. Fernandez-Serra and E. Artacho, *Phys. Rev. Lett.*, 2006, **96**, 016404.
15 D. Prendergast and G. Galli, *Phys. Rev. Lett.*, 2006, **96**, 215502.
16 F. Perakis, L. De Marco, A. Shalit, F. J. Tang, Z. R. Kann, T. D. Kuhne, R. Torre, M. Bonn and Y. Nagata, *Chem. Rev.*, 2016, **116**, 7590–7607.
17 P. Gasparotto, A. A. Hassanali and M. Ceriotti, *J. Chem. Theory Comput.*, 2016, **12**, 1953–1964.
18 P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. A. Naslund, T. K. Hirsch, L. Ojamae, P. Glatzel, L. G. M. Pettersson and A. Nilsson, *Science*, 2004, **304**, 995–999.
19 A. Nilsson and L. G. M. Pettersson, *Nat. Commun.*, 2015, **6**, 8998.
20 T. Head-Gordon and M. E. Johnson, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 7973–7977.
21 C. P. Lawrence and J. L. Skinner, *J. Chem. Phys.*, 2003, **118**, 264–272.
22 C. J. Fecko, J. D. Eaves, J. J. Loparo, A. Tokmakoff and P. L. Geissler, *Science*, 2003, **301**, 1698–1702.
23 S. Woutersen, U. Emmerichs and H. J. Bakker, *Science*, 1997, **278**, 658–660.
24 H. J. Bakker, S. Woutersen and H. K. Nienhuys, *Chem. Phys.*, 2000, **258**, 233–245.
25 S. T. van der Post and H. J. Bakker, *J. Phys. Chem. B*, 2014, **118**, 8179–8189.
26 S. T. van der Post, C. S. Hsieh, M. Okuno, Y. Nagata, H. J. Bakker, M. Bonn and J. Hunger, *Nat. Commun.*, 2015, **6**, 8384.

27 Y. S. Lin, P. A. Pieniazek, M. Yang and J. L. Skinner, *J. Chem. Phys.*, 2010, **132**, 174505.
28 F. Paesani, S. Yoo, H. J. Bakker and S. S. Xantheas, *J. Phys. Chem. Lett.*, 2010, **1**, 2316–2321.
29 R. A. Nicodemus, S. A. Corcelli, J. L. Skinner and A. Tokmakoff, *J. Phys. Chem. B*, 2011, **115**, 5604–5616.
30 V. P. Voloshin and Y. I. Naberukhin, *J. Struct. Chem.*, 2009, **50**, 78–89.
31 M. Kohagen, P. E. Mason and P. Jungwirth, *J. Phys. Chem. B*, 2016, **120**, 1454–1460.
32 G. R. Medders, V. Babin and F. Paesani, *J. Chem. Theory Comput.*, 2014, **10**, 2906–2910.
33 M. L. Laury, L. P. Wang, V. S. Pande, T. Head-Gordon and J. W. Ponder, *J. Phys. Chem. B*, 2015, **119**, 9423–9437.
34 L. P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez and V. S. Pande, *J. Phys. Chem. B*, 2013, **117**, 9956–9972.
35 H. C. Liu, Y. M. Wang and J. M. Bowman, *J. Chem. Phys.*, 2015, **142**, 194502.
36 G. A. Cisneros, K. T. Wikfeldt, L. Ojamae, J. B. Lu, Y. Xu, H. Torabifard, A. P. Bartok, G. Csanyi, V. Molinero and F. Paesani, *Chem. Rev.*, 2016, **116**, 7501–7528.
37 F. Paesani, *Acc. Chem. Res.*, 2016, **49**, 1844–1851.
38 S. K. Reddy, S. C. Straight, P. Bajaj, C. H. Pham, M. Riera, D. R. Moberg, M. A. Morales, C. Knight, A. W. Gotz and F. Paesani, *J. Chem. Phys.*, 2016, **145**, 194504.
39 J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik and M. Parrinello, *J. Chem. Phys.*, 2005, **122**, 014515.
40 K. Laasonen, M. Sprik, M. Parrinello and R. Car, *J. Chem. Phys.*, 1993, **99**, 9080–9089.
41 S. Yoo and S. S. Xantheas, *J. Chem. Phys.*, 2011, **134**, 121105.
42 M. D. Baer, C. J. Mundy, M. J. McGrath, I. F. W. Kuo, J. I. Siepmann and D. J. Tobias, *J. Chem. Phys.*, 2011, **135**, 124712.
43 M. Del Ben, M. Schonherr, J. Hutter and J. VandeVondele, *J. Phys. Chem. Lett.*, 2013, **4**, 3753–3759.
44 A. Zen, Y. Luo, G. Mazzola, L. Guidoni and S. Sorella, *J. Chem. Phys.*, 2015, **142**, 144111.
45 S. Y. Willow, M. A. Salim, K. S. Kim and S. Hirata, *Sci. Rep.*, 2015, **5**, 14358.
46 L. R. Pestana, N. Mardirossian, M. Head-Gordon and T. Head-Gordon, *Chem. Sci.*, 2017, **8**, 3554–3565.
47 R. A. DiStasio, B. Santra, Z. F. Li, X. F. Wu and R. Car, *J. Chem. Phys.*, 2014, **141**, 084502.
48 M. Del Ben, J. Hutter and J. VandeVondele, *J. Chem. Phys.*, 2015, **143**, 054506.
49 X. He, T. Zhu, X. W. Wang, J. F. Liu and J. Z. H. Zhang, *Acc. Chem. Res.*, 2014, **47**, 2748–2757.
50 X. W. Wang, J. F. Liu, J. Z. H. Zhang and X. He, *J. Phys. Chem. A*, 2013, **117**, 7149–7161.
51 J. F. Liu and X. He, *Phys. Chem. Chem. Phys.*, 2017, **19**, 20657–20666.
52 J. F. Liu, L. W. Qi, J. Z. H. Zhang and X. He, *J. Chem. Theory Comput.*, 2017, **13**, 2021–2034.
53 J. F. Liu, X. He and J. Z. H. Zhang, *Phys. Chem. Chem. Phys.*, 2017, **19**, 11931–11936.
54 O. Marsalek and T. E. Markland, *J. Chem. Phys.*, 2016, **144**, 054112.
55 L. B. Skinner, C. C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nilsson and C. J. Benmore, *J. Chem. Phys.*, 2013, **138**, 074506.
56 A. K. Soper and C. J. Benmore, *Phys. Rev. Lett.*, 2008, **101**, 065502.
57 S. Y. Willow, X. C. Zeng, S. S. Xantheas, K. S. Kim and S. Hirata, *J. Phys. Chem. Lett.*, 2016, **7**, 680–684.
58 M. Holz, S. R. Heil and A. Sacco, *Phys. Chem. Chem. Phys.*, 2000, **2**, 4740–4742.
59 D. J. Price and C. L. Brooks, *J. Chem. Phys.*, 2004, **121**, 10096–10103.
60 Y. S. Badyal, M. L. Saboungi, D. L. Price, S. D. Shastri, D. R. Haeffner and A. K. Soper, *J. Chem. Phys.*, 2000, **112**, 9206–9208.
61 P. L. Silvestrelli and M. Parrinello, *Phys. Rev. Lett.*, 1999, **82**, 3308–3311.
62 J. K. Gregory, D. C. Clary, K. Liu, M. G. Brown and R. J. Saykally, *Science*, 1997, **275**, 814–817.
63 A. Luzar and D. Chandler, *Phys. Rev. Lett.*, 1996, **76**, 928–931.
64 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 1339–1350.
65 V. A. Makarov, B. K. Andrews, P. E. Smith and B. M. Pettitt, *Biophys. J.*, 2000, **79**, 2966–2974.
66 H. S. Lee and M. E. Tuckerman, *J. Chem. Phys.*, 2007, **126**, 164501.
67 S. Habershon, T. E. Markland and D. E. Manolopoulos, *J. Chem. Phys.*, 2009, **131**, 024501.
68 T. E. Markland and B. J. Berne, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 7988–7991.
69 J. Liu, R. S. Andino, C. M. Miller, X. Chen, D. M. Wilkins, M. Ceriotti and D. E. Manolopoulos, *J. Phys. Chem. C*, 2013, **117**, 2944–2951.
70 X. Z. Li, B. Walker and A. Michaelides, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 6369–6373.
71 D. M. Wilkins, D. E. Manolopoulos, S. Pipolo, D. Laage and J. T. Hynes, *J. Phys. Chem. Lett.*, 2017, **8**, 2602–2607.
72 Y. J. Wu, H. L. Tepper and G. A. Voth, *J. Chem. Phys.*, 2006, **124**, 024503.