# Nanoscale

## PAPER

Check for updates

# A high-accuracy machine-learning water model for exploring water nanocluster structures†

Hao Zhou,[a] Ya-Juan Feng, [ID] *[a] Chao Wang,[b] Teng Huang,[c] Yi-Rong Liu,[a] Shuai Jiang,[a] Chun-Yu Wang[a] and Wei Huang [ID] *[a,c,d]

Water, the most important molecule on the Earth, possesses many essential and unique physical properties that are far from completely understood, partly due to serious difficulties in identifying the precise microscopic structures of water. Hence, identifying the structures of water nanoclusters is a fundamental and challenging issue for studies on the relationship between the macroscopic physical properties of water and its microscopic structures. For large-scale simulations (at the level of nm and ns) of water nanoclusters, a calculation method with simultaneous accuracy at the level of quantum chemistry and efficiency at the level of an empirical potential method is in great demand. Herein, a machine-learning (ML) water model was utilized to explore the microscopic structural features at different length scales for water nanoclusters with a size up to several nm. The ML water model can be employed to efficiently predict the structures of water nanoclusters with a similar accuracy to that of density functional theory and with substantially lower computational resource demands. To validate the low-lying structure search results with experimental spectral results, an ML water model combined with velocity autocorrelation function analysis was used to simulate the vibrational spectra of water nanoclusters with up to thousands of water molecules. By comparing the simulated and experimentally recorded vibrational spectra, the atomic structures determined by a simulation based on the ML water model are all verified. To demonstrate its ability to represent water's structural evolution at large length and time scales, the ML water model was employed to model the structural evolution during the crystal–liquid transition, and the phase transition temperatures of water clusters with different sizes were precisely predicted. The ML water model provides an efficient theoretical calculation tool for exploring the structures and physical properties of water and their relationships, especially for clusters with relatively large sizes and processes with relatively long durations.

## Introduction

Water is the source of life and is crucial for the Earth's overall environment. The unique properties of water, such as the maximum density at 277 K (ref. 1) and the anomalous high proton mobility,[2] determine numerous essential reactions and processes in the lives, nature and industries. These anomalous properties, which significantly differ from those of other simple molecular systems, have attracted intensive attention from researchers for several centuries[1–5] but remain far from being completely understood.

Many studies have focused on the unique local atomic structures of water,[5–14] namely hydrogen bonding networks,[15] because of the relationships between the local structure and the anomalous properties of water, as indicated by many previous studies using water nanoclusters as models. However, the underlying mechanisms are still unclear, partly due to the tremendous difficulties in directly identifying the precise atomic structures of water nanoclusters with experimental techniques.[12–15]

Hence, theoretical methods are widely utilized to explore the atomic structure of water nanoclusters.[16–19] Generally, theoretical studies on the structure of water nanoclusters are conducted in two steps: potential energy surface (PES) searches and validation. The first step is to search for the minima on the PES of water nanoclusters. The second step is to simulate experimentally detectable characteristics (for

[a]*School of Information Science and Technology, University of Science and Technology of China, Hefei, Anhui 230026, China. E-mail: fengyj6@ustc.edu.cn, huangwei6@ustc.edu.cn*
[b]*National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei, Anhui 230026, China*
[c]*Laboratory of Atmospheric Physico-Chemistry, Anhui Institute of Optics & Fine Mechanics, Chinese Academy of Sciences, Hefei, Anhui 230031, China*
[d]*CAS Center for Excellent in Urban Atmospheric Environment, Xiamen, Fujian 361021, China*
†Electronic supplementary information (ESI) available. See DOI: 10.1039/d1nr03128g

example, vibrational features) of the searched structures; then, the PES search results can be validated by comparing the simulated and experimentally measured results.

For PES searches, the potential energy and its gradient of a certain configuration should be calculated to determine the search direction. This step is usually the most time-consuming in the PES search step. In the validation step, the vibrational features of the searched structure can be simulated with molecular dynamics (MD) simulations and compared with the infrared spectroscopy experiment results. For MD simulations, the most time-consuming step is also the calculation of the potential energy and its gradient (namely, atomic forces).

A water model, which is actually a mapping from the structure to the potential energy of water, is used in the calculation of the potential energy and its gradient of water nanoclusters. The accuracy and efficiency of the water model determine its quality and feasibility in the research on water nanocluster structures. *Ab initio* water models based on density functional theory (DFT) or post-Harte-Fork (post-HF) calculations are able to give an accurate description of the intramolecular and intermolecular interactions of water clusters.[20,21] However, these methods are quite time-consuming, especially for relatively large clusters of more than 10 water molecules, owing to their steep scaling behavior. Empirical water models such as TIP4P[22] were constructed based on simplified physical models. Calculations with empirical water models are usually much faster than *ab initio* water models but at the cost of accuracy and universality. For example, the TIP4P model should be separately parameterized for condensed liquid water and ice.[22] Moreover, some empirical water models, such as TIP4P, treat water molecules as many-site rigid;[22] that is, they are not "reactive" and are not suitable for use in some important processes involving bond breaking, such as proton transfer in water. Thus, the efficiencies or accuracies of mainstream water models severely hinder their performance in the simulation of water clusters, especially for larger clusters within >100 atoms. Hence, a water model with both high accuracy and low computational resource consumption is crucial for improving the efficiency of simulations of large water clusters.

Aiming at exploring the structures of water nanoclusters, we adopted a "data-driven" machine learning approach instead of a traditional "physics-based" strategy to develop a reactive water model that achieves *ab initio* accuracy with much lower computational resource demands. This machine-learning (ML) water model is based on a group of highly flexible functions in the form of feedforward neural networks.[23] The neural networks are trained with a database of the potential energies of 300 000 off-equilibrium water clusters containing 1–21 water molecules, which were calculated at the ωB97XD/6-311** level. The power of this ML water model is demonstrated by the successful determination of the low-lying structures and vibrational spectra of water nanoclusters containing up to thousands of water molecules. Using the ML water model, the structures of water nanoclusters of up to thousands of molecules were predicted and validated with significantly improved accuracy and efficiency compared to previous water models.

With the ML water model, the length and time scales of water molecular simulation can be substantially extended. Many important phenomena of water, such as evaporation, melting and crystallization, can be inspected at the molecular level, and the structural evolution and related physical properties of water can be further revealed.

## Methods

### ML water model based on artificial neural networks

In ESI Fig. 1,† a schematic representation of the construction and a flowchart of the ML water model are presented. Firstly, the structural features are transformed to symmetry functions $G_i^2$ and $G_i^4$ to extract the local chemical environment of each atom. Secondly, the vectors for the local chemical environments of each atom were input into the atomic neural networks for hydrogen and oxygen elements. Generally, the ML water model consists of two feedforward neural networks trained to represent the "atomic potential energy" of hydrogen and oxygen atoms in water clusters. Thirdly, each atomic energy was obtained by atomic neural networks. Finally, the atomic energies were summed up for the total energy of the water clusters. The training and deployment of the ML water model were based on the TensorFlow software library.[24]

**Descriptor.** One of the crucial issues for the development of the ML water model is to construct a descriptor for the local chemical environments of atoms to be fed into the neural networks. The original local chemical environments in the form of Cartesian coordinates are not a suitable choice for the descriptor because translation and rotation operation would cause the Cartesian coordinates to vary; moreover, the atomic potential energy should remain unchanged. Hence, the descriptor should be invariant with respect to translations and rotations.[25,26] In addition, the descriptor should be highly sensitive to represent the structural difference between the off-equilibrium configurations in the training data. Several well-known descriptors were tested for the accuracy of the prediction of the potential energy and atomic forces. Symmetry functions[26] with the best performance were chosen in this research. Symmetry functions $G_i^2$ and $G_i^4$ were used to extract the features of the local chemical environment of each atom from the Cartesian coordinates.

$$G_i^2 = \sum_{j \neq i} e^{-\eta(r_{ij}-r_s)^2} f_c(r_{ij})$$

$$G_i^4 = 2^{1-\xi} \sum_{j \neq i} \sum_{k \neq i,j} \left[ (1 + \lambda \cos(\alpha_{ijk}))^\xi e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)} f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk}) \right]$$

where $r_{ij}$ is the distance between the *i*th and *j*th atoms, $\alpha_{ijk}$ is the angle centered at the *i*th atom and formed with the *j*th and *k*th atoms, and $f_c$ is the cutoff function. The parameters of the symmetry functions can be chosen according to the local chemical environment (such as the radial distribution function) of each element.

**Feedforward neural networks.** The ML water model consists of two feedforward neural networks that represent the "atomic potential energy" of hydrogen and oxygen atoms. In the developed ML water model, each of the networks contains an input layer, two hidden layers with 32 nodes, and an output layer. The functional form of the neural networks is

$$E_{\text{atom}} = f_{\text{activation}}(f_{\text{activation}}(G_i \times W_0 + b_0) \times W_1 + b_1) \times W_2 + b_2$$

where $f_{\text{activation}}$ is an activation function that improves the feasibility of the model to represent the complicated full-dimensional PES of a water cluster by introducing nonlinear transformations. In this study, we used a hyperbolic tangent function as the activation function. $G_i$ is the local chemical environment of atoms in the form of symmetry functions. $W$ and $b$ are the weight matrix and bias matrix, respectively. The weight and bias matrices were optimized with the Adam optimization algorithm during model training to minimize the mean squared error (MSE) between the predicted potential energies and the referenced *ab initio* potential energies. The loss function $\sum_i \left(E_i^{\text{predicted}} - E_i^{\text{reference}}\right)^2 + L_2$ was used to evaluate the difference between the predicted and referenced energies, where $L_2$ is the L2 regularization term used to prevent overfitting, which results in a low-quality prediction of potential energy gradients.

**Reference dataset.** The whole dataset contains ~300 000 structures of $(H_2O)_{1-21}$ clusters and the corresponding potential energies and atomic forces calculated at the ωB97XD/6-311** level. The cluster structures were generated with NVT molecular dynamics and *meta*-dynamics simulations. 90% of these data are used for training the machine learning water model, while the remaining 10% are used for validating the accuracy of the water model. The distributions of the potential energies in the whole dataset are presented in ESI Fig. 2.†

## Geometry optimization on the PES

Nanocluster structures with global minima on the PES were obtained by the previously established Basin-Hopping (BH) method[27–29] and the compressed sampling strategy[30] coupled with the ML water model.

## MD simulations

In MD simulations, the atomic positions were propagated using the velocity-Verlet algorithm, with a time step of 1 fs, in an NVT ensemble. The forces used to update the atomic positions were calculated with the developed ML water model for machine-learning molecular dynamics (MLMD) simulations and with the ωB97XD/6-311** level, as implemented in Gaussian 09 for AIMD simulations.

For vibrational spectrum simulations, the optimized low-lying geometries were used as the starting structures for MD simulations. Each trajectory was propagated for approximately 30 ps, and the last 20 ps were used for analysis.

For the phase transition simulation, 9 trajectories were propagated at 190–270 K for 30 ps, and the last 20 ps were used for structure analysis (local tetrahedral order parameter) and vibrational spectrum simulations.

## Simulations of vibrational spectra

The vibrational spectra of water clusters were calculated using the Fourier transform (FT) of the velocity autocorrelation function (VACF),

$$I_{\text{IR}}(\omega) \propto \int_0^\infty C_V(t)e^{-i\omega t}dt$$

where $C_V(t) = \dfrac{1}{N}\sum_{i=1}^N \langle v_i(0) \cdot v_i(t)\rangle$, $N$ is the total number of atoms within the clusters and $v_i(t)$ is the velocity of the $i$th atom at time $t$. The VACF can be obtained from the MD simulation trajectory at a specific temperature. With this approach, the total vibrational spectrum can be divided into contributions from each atom, and the origin of some vibrational bands can be identified. The scale factors for the fundamental frequencies calculated with the MLMD, AIMD and static quantum chemistry calculations at the ωB97XD/6-311** level were estimated by linear fitting with the previously reported experimental fundamental frequencies of water.[31]

## Structural identification of water

The liquid phase was distinguished from crystal ice by the local tetrahedral order parameter method.[32]

$$q_t(k) = 1 - \frac{3}{8}\sum_{i=1}^3 \sum_{j=i+1}^4 \left(\cos\theta_{ikj} + \frac{1}{3}\right)^2$$

where $\theta_{ikj}\theta_{ikj}$ is the angle between the central water molecule $k$ and two ($i$ and $j$) of the four closest neighbors.

## Radii of water nanoclusters

The radii of the water nanoclusters were obtained by a previously reported ellipsoidal approximation.[33,34]

# Results and discussion

The energies and atomic forces of ~30 000 $(H_2O)_n$ ($n$ = 1–21) cluster configurations in the test dataset are predicted to evaluate the accuracy of the trained ML water model. The absolute error distributions of $(H_2O)_n$ ($n$ = 1–21) are shown in ESI Fig. 3–7.† Compared to the reference potential energies and atomic forces calculated by quantum chemistry at the ωB97XD/6-311** level, the mean absolute errors (MAEs) of the energies and atomic forces predicted by the ML water model are ~2.0 meV per $H_2O$ and ~170 meV $\text{Å}^{-1}$, respectively (ESI Fig. 3–7†). The errors are comparable to or smaller than the intrinsic uncertainties from the exchange–correlation functional of the DFT method.[23] The MAE of the predicted atomic forces of a single $H_2O$ molecule is less than 100 meV $\text{Å}^{-1}$, as shown in Fig. 1a. For larger clusters, the MAEs of atomic

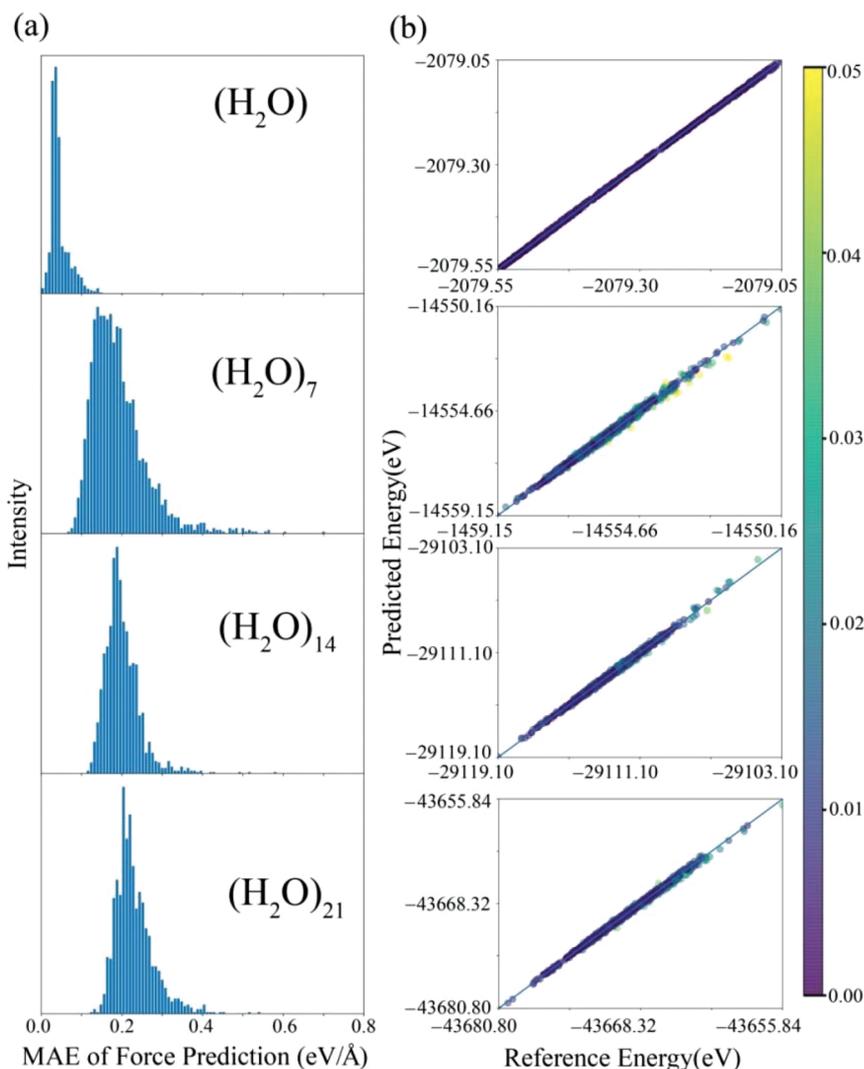## ■ The Accuracy and Efficiency of the ML Water Model



**Fig. 1** (a) The MAE of the force errors of $(H_2O)_{1,7,14,21}$, and (b) the comparisons of the energies of $(H_2O)_{1,7,14,21}$ clusters calculated using the ML water model and *ab initio* calculations.

forces are approximately 200 meV Å⁻¹. These test results confirmed not only the ability of the ML water model to predict the potential energies of water molecules and clusters, but also its power to estimate the potential energy gradients (or atomic forces), although the potential energy gradients of water clusters in the training dataset were not used in neural network training.

The gradients of the PES are important for geometry optimization, MD simulations and other molecular simulations. Hence, the gradients of the covalent bonding energy and intermolecular H-bonding energy of the ML water model were compared with the *ab initio* values. The potential energy profiles obtained using the ML water model and DFT calculation at the ωB97XD/6-311** level are nearly identical, as shown in Fig. 2a–c. Fig. 2a shows a plot of the potential

energy profiles of the O–H stretching vibrations of a single $H_2O$ molecule. The O–H bond length predicted by both the ML water model and quantum chemistry is 0.96 Å, only 0.01 Å shorter than the experimental value of 0.97 Å.[35] The potential energy profiles of the H–O–H bending vibration of a single $H_2O$ molecule shown in Fig. 2b also exhibit good agreement between the ML water model and the *ab initio* results. The predicted H–O–H bond angle is ~104.5°, in good accordance with the experimental value of 105.1°.[35] Fig. 2c presents the potential energy profiles of the H-bond of the $(H_2O)_2$ dimer. The ML water model predicted a potential energy profile that was highly similar to that predicted by quantum chemistry. For the ML water model, the length of the H-bond of the $(H_2O)_2$ dimer is 2.1 Å, which is consistent with the value calculated by quantum chemistry and the
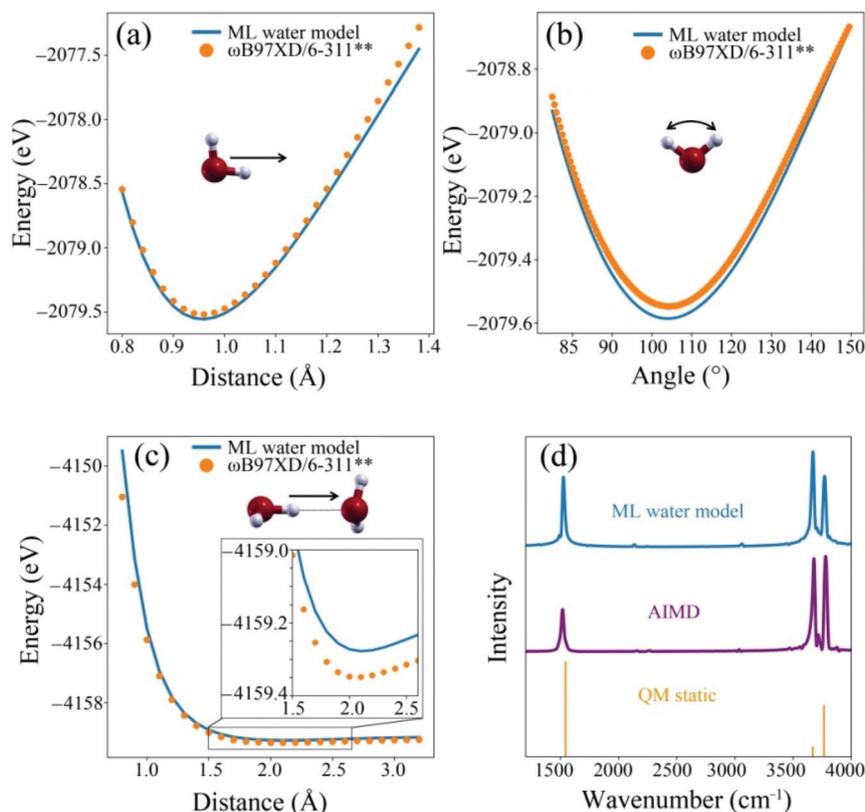
**Fig. 2** (a) The PES of a single $H_2O$ molecule with one hydrogen atom pulled along the O−H bond. (b) The PES of a single $H_2O$ molecule with a changing value of ∠HOH. (c) The PES of a $(H_2O)_2$ dimer with one water molecule pulled along the H-bond. (d) The vibrational spectra of a $H_2O$ molecule calculated *via* MLMD simulation coupled with the FT of VACFs compared with the AIMD simulation coupled with the FT of VACFs and the quantum chemistry spectrum calculated at the ωB97XD/6-311** level.

experimentally measured results.[36] The above simulation results indicate that the ML water model can correctly represent the gradient of both the covalent bonding interactions and H-bonding interactions of water. To assess the quality of the vibrational spectrum simulated with MLMD, the vibrational spectra of the $H_2O$ molecule at 20 K were calculated with the FT of the VACFs for the MLMD and AIMD simulations. The spectra are shown in Fig. 2d, together with the vibrational spectrum obtained using static quantum chemistry calculations. For all the simulated vibrational spectra, three pronounced peaks can be identified. The peak located at ~1530 cm$^{-1}$ can be assigned to the H–O–H bending vibrations. The peaks at approximately 3650 and 3770 cm$^{-1}$ can be ascribed to symmetric and asymmetric O–H stretching vibrations, respectively.[37] Both the frequencies and amplitudes calculated with MLMD simulations are very close to those obtained by AIMD, confirming the quality of the vibrational spectra simulated with MLMD.

As shown in ESI Fig. 8,† we compared the computational efficiency for energies and forces of water clusters using the ML water model, DFT and the empirical water model of TIP4P. The calculation speed of the potential energies and the energy gradients of the $(H_2O)_{10-60}$ clusters is significantly improved by

~10 000 times compared to that of DFT, and is close to that of TIP4P. All the calculation efficiency tests were carried out in a node with an Intel(R) Xeon(R) Gold 6130 CPU (2.10 GHz). Calculations were performed in 8 parallel processes if the multiprocessing mode is available.

Since the established ML water model is proven to be capable of representing the potential energy and potential energy gradient with DFT, the low-lying geometry configurations of $(H_2O)_n$ ($n$ = 2–21) clusters were optimized *via* the ML water model coupled with the BH algorithm. The optimization results are shown in Fig. 3 and ESI Fig. 9–13,† with labels in the form of "$n–i$". In this notation, "$n$" denotes the number of water molecules and the index "$i$" is used to index the isomers. $\Delta E_{QM}$ and $\Delta E_{ML}$ represent the relative potential energies calculated by quantum chemistry at the ωB97XD/6-311** level and the developed ML water model, respectively. The lowest-energy structures for $(H_2O)_{3-5}$ are the planar cyclic configurations shown in Fig. 3. For $(H_2O)_{6-10}$, the lowest-energy structures have predominantly three-dimensional configurations (ESI Fig. 10†). For instance, the potential energy ordering for $(H_2O)_6$ is trigonal prism < cage < ring, which is in agreement with the CCSD(T) results.[38] The lowest-energy structures predicted by the ML water model and *ab initio* calculations are

## ■ $(H_2O)_n$ (n = 2 ~ 21): from Planar to Cage-like Structure



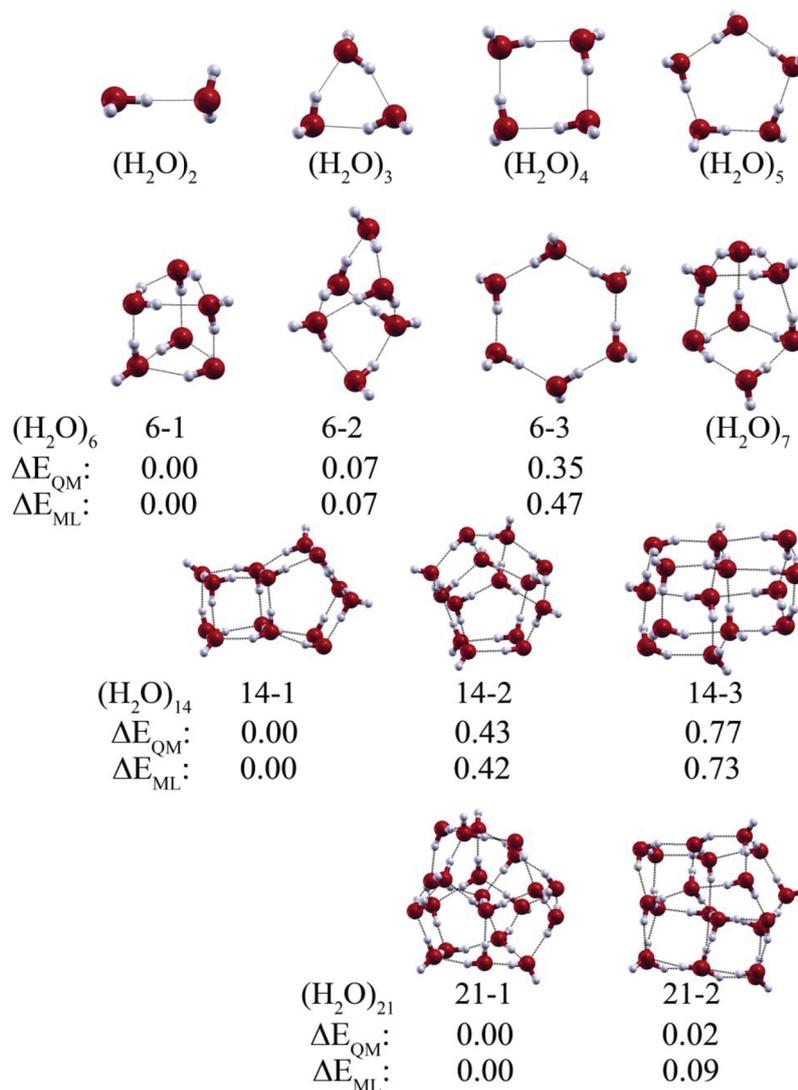**Fig. 3** The low-lying geometry configurations of $(H_2O)_n$ (n = 2–7, 14, 21) calculated *via* the ML water model and quantum chemistry, as well as the relative energies (in eV).

both the prism isomer 6-1. The second-lowest-energy structure (isomer 6-2) is a cage structure. The $\Delta E_{QM}$ and $\Delta E_{ML}$ values of isomer 6-2 are both approximately 0.07 eV higher than those of isomer 6-1. The third-lowest-energy structure is isomer 6-3, with a cyclic structure, and its $\Delta E_{QM}$ and $\Delta E_{ML}$ values are approximately 0.35 eV and 0.47 eV higher than those of isomer 6-1, respectively. For $(H_2O)_{11-18}$, the lowest-energy structures consist of two cage-like structures, as shown in ESI Fig. 11 and 12.† For instance, the lowest-energy structure of $(H_2O)_{14}$ (isomer 14-1) consists of a cube and a pentagonal prism sharing one face. The second-lowest-energy structure is isomer 14-2, with a cage-like structure consisting of circular quadrangles and pentagons, and the $\Delta E_{QM}$ and $\Delta E_{ML}$ values of isomer 14-2 are approximately 0.43 eV and 0.42 eV higher than those

of isomer 14-1, respectively. The rest of the low-lying structure is isomer 14-3, consisting of a distorted cube and a pentagonal prism sharing one face, and the $\Delta E_{QM}$ and $\Delta E_{ML}$ values of isomer 14-3 are approximately 0.77 eV and 0.73 eV higher than those of isomer 14-1, respectively. For $(H_2O)_{19-21}$, the predominant lowest-energy structures are high-dimensional cages consisting of circular quadrangles and pentagons (ESI Fig. 13†). As shown in Fig. 3, the lowest-energy structure of $(H_2O)_{21}$ is a high-dimensional cage (isomer 21-1). Similarly, the second-lowest-energy structure is also a high-dimensional cage, and the $\Delta E_{QM}$ and $\Delta E_{ML}$ values of isomer 21-2 are approximately 0.77 eV and 0.73 eV higher than those of isomer 21-1, respectively. The above analysis leads to a conclusion that the lowest-energy geometries and energy ordering for $(H_2O)_n$ clusters cal-

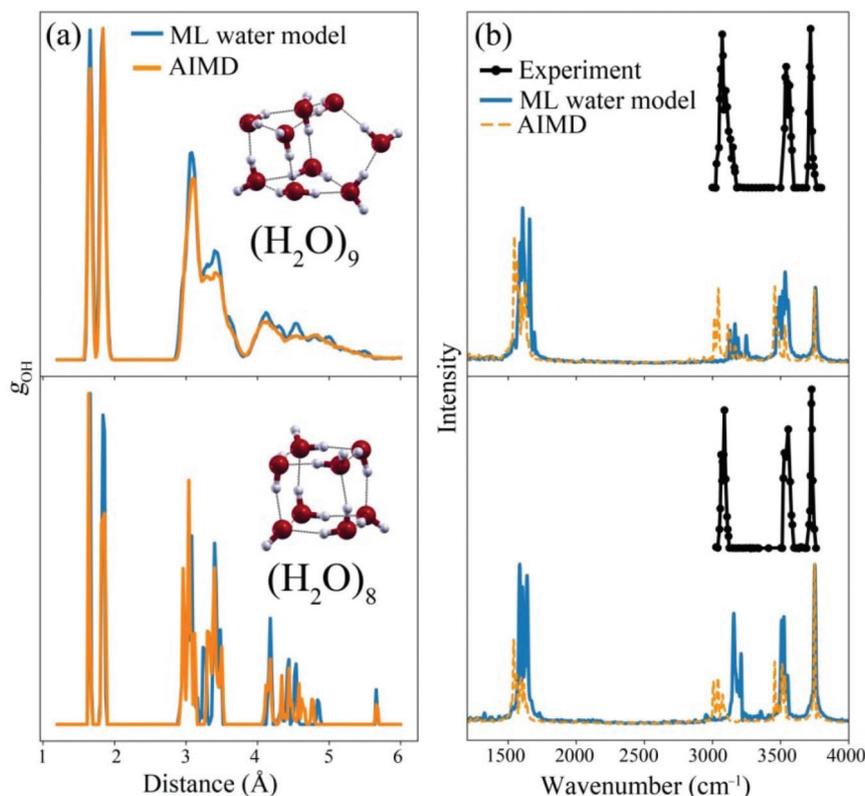**Fig. 4** (a) Comparisons of the radial distribution functions of the lowest-energy structures of $(H_2O)_{8,9}$ calculated by MLMD and AIMD simulations. (b) Vibrational spectra of the lowest-energy structures of $(H_2O)_{8,9}$ calculated by MLMD simulation coupled with the FT of the VACFs compared with the AIMD and experimental results.[39]

culated by the ML water model are in agreement with the high-level quantum chemistry results.

To validate the above low-lying structure search results, the ML water model combined with velocity autocorrelation function analysis was used to simulate the vibrational spectra of $(H_2O)_n$ clusters (Fig. 4). The comparisons of the radial distribution functions (RDFs) of $g_{OH}$ calculated *via* the MLMD and AIMD simulations are shown in Fig. 4a, which show perfect consistency in the range of 1.6–1.8 Å, indicative of similar local structure relaxation.

The vibrational spectra of $(H_2O)_{8,9}$ clusters calculated by MLMD simulations coupled with the FT of the VACFs are compared with the AIMD and experimental results,[39] as shown in Fig. 4b. Three vibrational features in the range of 3000–3800 cm$^{-1}$ in the experimental spectra are well reproduced by the MLMD and AIMD simulations. The complex vibrational features in the range of 1500–2000 cm$^{-1}$ in the AIMD spectra and MLMD spectra are also similar. The agreement between the predicted vibrational spectra and the experimental results provides experimental support for the lowest-energy structures predicted by the ML water model coupled with the BH search.

Next, the vibrational spectra simulated with MLMD were further used to explore the structure of larger water clusters. The molecular densities of larger $(H_2O)_n$ ($n$ = 32, 48 and 113)

clusters were calculated by MLMD trajectories, as shown in Fig. 5a. The radii of the $(H_2O)_n$ ($n$ = 32, 48 and 113) clusters are approximately 6, 8 and 12 Å, respectively, and the water molecules in the core and surface of clusters were separated, as shown in Fig. 5a.

The comparisons of the vibrational spectra of $(H_2O)_n$ ($n$ = 32, 48 and 113) calculated using the MLMD simulations and the experimental spectra[40,41] are shown in Fig. 5b. The experimental spectra[40,41] of $(H_2O)_n$ ($n$ = 32, 48 and 113) are well reproduced by the MLMD simulations. Based on the comparisons of the vibrational spectra calculated using the MLMD simulations and the experimental spectra, the predominant features at approximately 3400 cm$^{-1}$ can be ascribed to the typical amorphous behavior, such as that of liquid water.[40] However, the dominating vibrational absorptions at around 3200 cm$^{-1}$ of the crystalline structure[12] are not observed. The results predicted by MLMD simulations are consistent with the previous experimental results[40,41] for $(H_2O)_n$ ($n$ = 32, 48 and 113).

Furthermore, the vibrational features of core and surface water molecules in clusters were calculated by the MLMD method. For instance, the broad features of the $(H_2O)_{32}$ cluster in the range of 2800–3600 cm$^{-1}$ can be ascribed to the O–H stretching vibrations of the core water molecules. The sharp vibrational features in the range of 3600–3800 cm$^{-1}$ can be
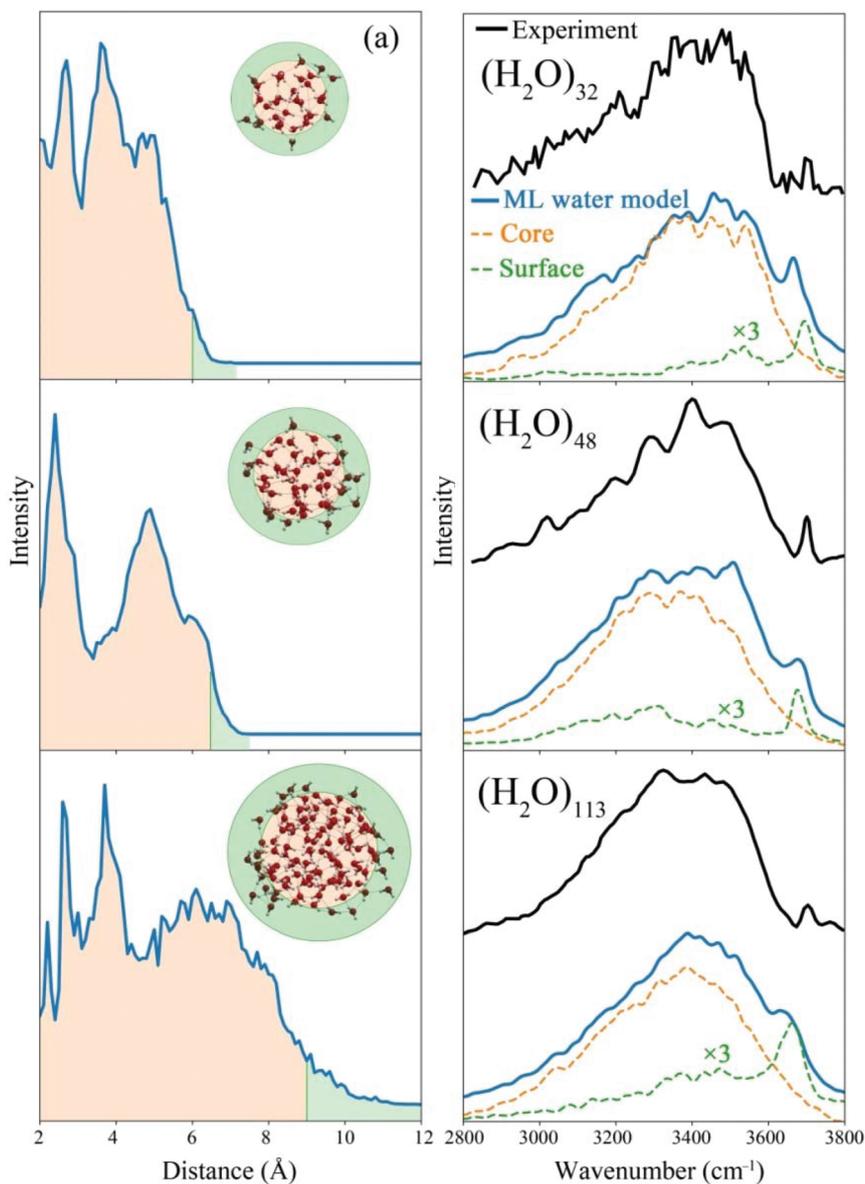
## ■ (H₂O)ₙ (n =32, 48, 113): Core-Shell-like Structure



**Fig. 5**  (a) Molecular densities of (H₂O)ₙ (n = 32, 48 and 113). (b) Vibrational spectra of (H₂O)ₙ (n = 32, 48 and 113) calculated by MLMD simulation coupled with the FT of the VACFs compared with the experimental spectra.[40,41]

attributed to the O–H stretching vibrations of surface molecules with dangling H or dangling O atoms. Similar to (H₂O)₃₂, the contributions of the core and surface water molecules of the (H₂O)₄₈,₁₁₃ clusters are well identified.

To demonstrate its universality, the ML water model was employed to simulate the crystal–liquid transition of water droplets, a focus of studies on water.[42–46] Three water nanoclusters consisting of 275, 475 and 900 water molecules were chosen as their crystalline structures have been well characterized experimentally.[40] Using MD simulations coupled with the ML water model, three nanoclusters consisting of 275, 475 and 900 water molecules were applied to model the crystal–amor-

phous transition of water nanoclusters. The temperature dependence of the structural evolution for the (H₂O)₄₇₅ nanocluster was taken to estimate the phase transition (Fig. 6a). The local tetrahedral ordering parameter $q_t$ was used to distinguish between different phases of water.[32] The water molecules in the (H₂O)₄₇₅ nanocluster with $q_t$ above 0.95 are regarded as crystal molecules and are represented by blue lines in Fig. 6a. It is obvious that the (H₂O)₄₇₅ nanoclusters at 100 K are dominated by tetrahedrally configured water molecules. As the temperature increases to 230 K, the number of crystal molecules significantly decreases with an increasing number of disordered molecules. When the temperature

■ **(H₂O)ₙ (n = 275, 475 and 900): from Tetrahedral to Amorphous Structure**



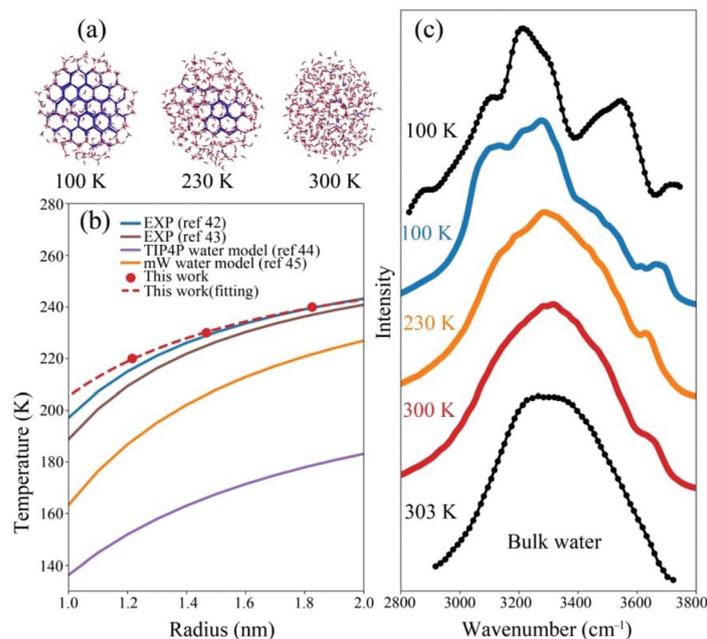**Fig. 6** (a) Temperature dependence of the structural evolution for the (H₂O)₄₇₅ nanocluster. (b) Temperature–radius phase diagram simulated with the Gibbs–Thomson equation and parameters from previous experiments[42,43] and simulations.[44,45] (c) Temperature dependence of the vibrational spectra of the (H₂O)₄₇₅ nanocluster calculated by MLMD simulation coupled with the FT of the VACFs compared with the experimental IR spectra of the (H₂O)₄₇₅ nanocluster at 100 K (ref. 40) and bulk water at 303 K.[12]

reaches 300 K, disordered molecules dominate in the water nanocluster. Based on the above results, it is important to note that the structural evolution from a tetrahedrally coordinated structure to a disordered structure is a direct examination of the crystal–liquid phase transition. The curves of the temperature dependence of the potential energy for the (H₂O)₂₇₅, (H₂O)₄₇₅, and (H₂O)₉₀₀ nanoclusters were taken to reveal the relationship between the melting temperature and nanocluster radius (ESI Fig. 14†). The melting temperatures of (H₂O)₂₇₅, (H₂O)₄₇₅ and (H₂O)₉₀₀ nanoclusters determined using potential energies *vs.* the temperature curves are 220, 230 and 240 K, respectively. The dependence of the melting temperature on the particle radius can be described by the Gibbs–Thomson equation determined by experiments[42,43] and simulations.[44,45] We find that the melting temperatures predicted by the ML water model match well with those from experiments,[42,43] with an extraordinarily better accuracy than those from simulations with the TIP4P water model[44] and the mW water model[45] (Fig. 6b).

The structural evolution characteristics during the crystal–amorphous transition of the (H₂O)₄₇₅ nanocluster were further validated by comparing the MLMD predicted vibrational spectra with the experimental spectra of the (H₂O)₄₇₅ nanocluster[40] and bulk water[12] (Fig. 6c). In the experimental spectra of the (H₂O)₄₇₅ nanocluster at about 100 K,[40] there are four pronounced features. The shoulder at approximately 3100 cm⁻¹ can be ascribed to an ice-like structure.[12] The dominant peaks at approximately 3200 and 3300 cm⁻¹ are typical of

four-coordinated structures and can be ascribed to the crystalline structure and distorted tetrahedral structure, respectively.[40] The peaks at approximately 3400–3650 cm⁻¹ are associated with water molecules whose coordinated number is smaller than that of the tetrahedral configurations.[12] We find that all four predominant features in the above experimental spectrum were well represented in the MLMD predicted spectra at 100 K, confirming that the structural characteristics of (H₂O)₄₇₅ nanoclusters at 100 K were dominated by tetrahedral configurations. In addition, the experimental spectrum of bulk water at 303 K (ref. 12) consists of only one broad feature at approximately 3350 cm⁻¹, which is also present in the MLMD predicted spectrum of the (H₂O)₄₇₅ nanocluster at 300 K, implying the amorphous structure of the (H₂O)₄₇₅ nanocluster at 300 K. However, it was important to determine the structure of the (H₂O)₄₇₅ nanocluster at the melting temperature. To this end, the vibrational spectra of the (H₂O)₄₇₅ nanoclusters at 230 K were calculated by MLMD simulations coupled with the FT of the VACFs. The configuration analysis suggests that there is a mixture of tetrahedral and amorphous stacking at the melting temperature. By comparing the MLMD simulated vibrational spectra and the experimental IR spectra, the coexistence of vibrational features from normal tetrahedral, distorted tetrahedral and amorphous configurations can be observed. These results indicate that the melting of ice nanoclusters is a process in which the tetrahedral structures are distorted, and this structural evolution can be tracked with IR spectroscopy techniques.

## Conclusions

To explore the microscopic structures and physical properties of water nanoclusters and their relationships with high accuracy and efficiency, an ML water model based on multiple feed-forward neural networks was developed to provide rapid predictions of the potential energies and energy gradients of water nanoclusters. The ML water model is proven to be able to represent the potential fields and potential field gradients of water nanoclusters at DFT accuracy but with a substantially lower computational resource and time demands. Coupled with the BH global optimization technique and VACF analysis, the ML water model successfully predicted the atomic structures and vibrational spectra of water nanoclusters with up to thousands of water molecules, and they are in good agreement with previous DFT studies and experimental measurements. Vibrational spectral simulations also elucidated the core–shell-like structures of $(H_2O)_{32-113}$ nanoclusters. The universality and scalability of the ML water model were demonstrated by modeling the liquid phase and crystal phase of water nanoclusters or bulk water. The structural evolution during the phase transition (from a tetrahedral coordination structure to an amorphous structure) could be precisely described with the ML water model. The applications of the ML water model can be further extended by coupling it with other molecular simulation techniques, providing a powerful theoretical calculation tool for predicting the structures and physical properties of water nanoclusters and their relationships. In conclusion, by introducing the ML water model with high performance both in accuracy and efficiency, the structural features of water nanoclusters can be well predicted and validated, based on which the structure–property relationship of water can be further theoretically studied.

## Author contributions

W. H. and Y. J. F. conceived and managed the project. C. W. and Y. J. F. developed the ML water model and codes. Y. J. F. and H. Z. implemented the MLMD simulations and molecular vibrational spectral simulations, and analyzed the results. All authors jointly analyzed the simulated results. Y. J. F. and H. Z. wrote the manuscript and prepared the figures.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

## References

1 O. Mishima and H. E. Stanley, The relationship between liquid, supercooled and glassy water, *Nature*, 1998, **396**, 329–335.

2 J. W. Shin, *et al.*, Infrared signature of structures associated with the $H^+(H_2O)_n$ (n = 6 to 27) clusters, *Science*, 2004, **304**, 1137–1140.

3 O. Mishima, L. D. Calvert and E. Whalley, 'Melting ice' I at 77 K and 10 kbar: a new method of making amorphous solids, *Nature*, 1984, **310**, 393–395.

4 E. B. Moore and V. Molinero, Structural transformation in supercooled water controls the crystallization rate of ice, *Nature*, 2011, **479**, 506–508.

5 K. H. Kim, *et al.*, Maxima in the thermodynamic response and correlation functions of deeply supercooled water, *Science*, 2017, **358**, 1589–1593.

6 H. J. Bakker and J. L. Skinner, Vibrational spectroscopy as a probe of structure and dynamics in liquid water, *Chem. Rev.*, 2010, **110**, 1498–1517.

7 U. Buck and F. Huisken, Infrared spectroscopy of size-selected water and methanol clusters, *Chem. Rev.*, 2000, **100**, 3863–3890.

8 K. Amann-Winkel, *et al.*, X-ray and neutron scattering of water, *Chem. Rev.*, 2016, **116**, 7570–7589.

9 C. A. Tulk, *et al.*, Structural studies of several distinct metastable forms of amorphous ice, *Science*, 2002, **297**, 1320–1323.

10 J. A. Sellberg, *et al.*, Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature, *Nature*, 2014, **510**, 381–384.

11 F. Mallamace, C. Branca, M. Broccio, C. Corsaro, C. Y. Mou and S. H. Chen, The anomalous behavior of the density of water in the range 30 K < T < 373 K, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 18387–18391.

12 F. Mallamace, *et al.*, Evidence of the existence of the low-density liquid phase in supercooled, confined water, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 424–428.

13 R. Z. Ma, *et al.*, Atomic imaging of the edge structure and growth of a two-dimensional hexagonal ice, *Nature*, 2020, **577**, 60–63.

14 G. Y. Bai, D. Gao, Z. Liu, X. Zhou and J. J. Wang, Probing the critical nucleus size for ice formation with graphene oxide nanosheets, *Nature*, 2019, **576**, 437–441.

15 A. Nilsson and L. G. M. Pettersson, The structural origin of anomalous properties of liquid water, *Nat. Commun.*, 2015, **6**, 8998.

16 T. K. Roy and R. B. Gerber, Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications, *Phys. Chem. Chem. Phys.*, 2013, **15**, 9468–9492.

17 G. M. Chaban, J. O. Jung and R. B. Gerber, Ab initio calculation of anharmonic vibrational states of polyatomic

systems: Electronic structure combined with vibrational self-consistent field, *J. Chem. Phys.*, 1999, **111**, 1823–1829.

18 M. Thomas, M. Brehm, R. Fligg, P. Vohringer and B. Kirchner, Computing vibrational spectra from ab initio molecular dynamics, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6608–6622.

19 M. Brehm and B. Kirchner, TRAVIS - A free analyzer and visualizer for monte carlo and molecular dynamics trajectories, *J. Chem. Inf. Model.*, 2011, **51**, 2007–2023.

20 V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin and W. A. Goddard, Evaluation of B3LYP, X3LYP, and M06-class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters, *J. Chem. Theor. Comput.*, 2009, **5**, 1016–1026.

21 M. J. Gillan, D. Alfe and A. Michaelides, Perspective: How good is DFT for water?, *J. Chem. Phys.*, 2016, **144**, 130901.

22 J. L. F. Abascal, E. Sanz, R. G. Fernandez and C. Vega, A potential model for the study of ices and amorphous water: TIP4P/Ice, *J. Chem. Phys.*, 2005, **122**, 234511.

23 T. Morawietz, A. Singraber, C. Dellago and J. Behler, How van der Waals interactions determine the unique properties of water, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 8368–8373.

24 M. Abadi, *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, http://tensorflow.org/.

25 M. Rupp, A. Tkatchenko, K. R. Muller and O. A. von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, *Phys. Rev. Lett.*, 2012, **108**, 058301.

26 J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, *J. Chem. Phys.*, 2011, **134**, 074106.

27 W. Huang, A. P. Sergeeva, H. J. Zhai, B. B. Averkiev, L. S. Wang and A. I. Boldyrev, A concentric planar doubly π-aromatic $B_{19}^-$ cluster, *Nat. Chem.*, 2010, **2**, 202–206.

28 W. Huang and L. S. Wang, Probing the 2D to 3D Structural Transition in Gold Cluster Anions Using Argon Tagging, *Phys. Rev. Lett.*, 2009, **102**, 153401.

29 W. Huang, H. J. Zhai and L. S. Wang, Probing the interactions of $O_2$ with small gold cluster anions ($Au_n^-$, n=1-7): Chemisorption vs physisorption, *J. Am. Chem. Soc.*, 2010, **132**, 4344–4351.

30 Y. R. Liu, *et al.*, Structural exploration of water, nitrate/water, and oxalate/water clusters with Basin-Hopping method using a compressed sampling technique, *J. Phys. Chem. A*, 2014, **118**, 508–516.

31 I. M. Alecu, J. Zheng, Y. Zhao and D. G. Truhlar, Computational Thermochemistry: Scale factor databases and scale factors for vibrational frequencies obtained from electronic model chemistries, *J. Chem. Theor. Comput.*, 2010, **6**, 2872–2887.

32 J. R. Errington and P. G. Debenedetti, Relationship between structural order and the anomalies of liquid water, *Nature*, 2001, **409**, 318–321.

33 F. Trudu, D. Donadio and M. Parrinello, Freezing of a Lennard-Jones fluid: From nucleation to spinodal regime, *Phys. Rev. Lett.*, 2006, **97**, 105701.

34 H. Wang, H. Gould and W. Klein, Homogeneous and heterogeneous nucleation of Lennard-Jones liquids, *Phys. Rev. E*, 2007, **76**, 031604.

35 S. Habershon, T. E. Markland and D. E. Manolopoulos, Competing quantum effects in the dynamics of a flexible water model, *J. Chem. Phys.*, 2009, **131**, 024501.

36 A. K. Soper, The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa, *Chem. Phys.*, 2000, **258**, 121–137.

37 P. E. Fraley and K. N. Rao, High resolution infrared spectra of water vapor: V1 and V3 bands of $H_2^{16}O$, *J. Mol. Spectrosc.*, 1969, **29**, 348–364.

38 E. E. Dahlke, R. M. Olson, H. R. Leverentz and D. G. Truhlar, Assessment of the accuracy of density functionals for prediction of relative energies and geometries of low-lying isomers of water hexamers, *J. Phys. Chem. A*, 2008, **112**, 3976–3984.

39 U. Buck, I. Ettischer, M. Melzer, V. Buch and J. Sadlej, Structure and spectra of three-dimensional $(H_2O)_n$ clusters, n = 8, 9, 10, *Phys. Rev. Lett.*, 1998, **80**, 2578–2581.

40 C. C. Pradzynski, R. M. Forck, T. Zeuch, P. Slavicek and U. Buck, A fully size-resolved perspective on the crystallization of water clusters, *Science*, 2012, **337**, 1529–1532.

41 U. Buck, C. C. Pradzynski, T. Zeuch, J. M. Dieterich and B. Hartke, A size resolved investigation of large water clusters, *Phys. Chem. Chem. Phys.*, 2014, **16**, 6859–6871.

42 R. Schmidt, E. W. Hansen, M. Stocker, D. Akporiaye and O. H. Ellestad, Pore size determination of MCM-51 mesoporous materials by means of $^1H$ NMR spectroscopy, N2 adsorption, and HREM. A preliminary study, *J. Am. Chem. Soc.*, 1995, **117**, 4049–4056.

43 A. Schreiber, I. Ketelsen and G. H. Findenegg, Melting and freezing of water in ordered mesoporous silica materials, *Phys. Chem. Chem. Phys.*, 2001, **3**, 1185–1195.

44 D. Pan, L. M. Liu, B. Slater, A. Michaelides and E. Wang, Melting the ice: On the relation between melting temperature and size for nanoscale ice crystals, *ACS Nano*, 2011, **5**, 4562–4569.

45 J. C. Johnston and V. Molinero, Crystallization, melting, and structure of water nanoparticles at atmospherically relevant temperatures, *J. Am. Chem. Soc.*, 2012, **134**, 6650–6659.

46 C. Hock, *et al.*, Calorimetric observation of the melting of free water nanoparticles at cryogenic temperatures, *Phys. Rev. Lett.*, 2009, **103**, 073401.