

PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

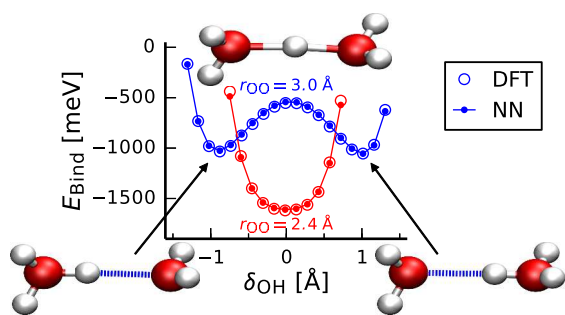
Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

TOC sentence:

We report a reactive neural network potential for protonated water clusters that accurately represents the density-functional theory potential-energy surface.



Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials

Suresh Kondati Natarajan, Tobias Morawietz, and Jörg Behler*

Received Xth XXXXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXXXX 200X

DOI: 10.1039/b000000x

Investigating the properties of protons in water is essential for understanding many chemical processes in aqueous solution. While important insights can in principle be gained by accurate and well-established methods like ab initio molecular dynamics simulations, the computational costs of these techniques are often very high. This prevents studying large systems on long time scales, which is severely limiting the applicability of computer simulations to address a wide range of interesting phenomena. Developing more efficient potentials enabling the simulation of water including dissociation and recombination events with first-principles accuracy is a very challenging task. In particular protonated water clusters have become important model systems to assess the reliability of such potentials, as the presence of the excess proton induces substantial changes in the local hydrogen bond patterns and many energetically similar isomers exist, which are extremely difficult to describe. In recent years it has been demonstrated for a number of systems including neutral water clusters of varying size that neural networks (NNs) can be used to construct potentials with close to first-principles accuracy. Based on density-functional theory (DFT) calculations, here we present a reactive full-dimensional NN potential for protonated water clusters up to the octamer. A detailed investigation of this potential shows that the energetic, structural, and vibrational properties are in excellent agreement with DFT results making the NN approach a very promising candidate for developing a high-quality potential for water. This finding is further supported by first preliminary but very encouraging NN-based simulations of the bulk liquid.

Keywords: Neural networks, density-functional theory, protonated water clusters, potential energy surfaces.

1 Introduction

Water, the most important solvent, is of vital importance for countless chemical systems from biochemistry to large-scale industrial processes. Consequently, it has received considerable attention in numerous experimental as well as in theoretical studies.^{1–5} Although a lot of insights have been gained in its structure, in its physical properties and in its interaction with other molecules,^{6–8} our understanding of water and its role in solvation, which governs most chemical reactions in aqueous solutions, is far from complete. Consequently, further investigations are urgently needed to unravel the role of water at the atomic level for a variety of processes being as different as the stabilization of large biomolecules under physiological conditions and the formation of the famous double layer in electrochemical systems.^{9,10}

Water has two remarkable properties, which are responsible for its unique role in chemistry: First, water forms strong hydrogen bonds, which are the reason why water is a liquid at ambient conditions. Second, water has the ability to disso-

ciate by forming a hydronium and a hydroxide ion. Consequently water is able to act as an acid as well as a base and can provide protons and hydroxide ions in chemical reactions. In combination with the flexible hydrogen bond network, the species emerging in the auto-dissociation process show characteristic mobilities, arising e.g. from the Grotthuss mechanisms of proton transport,^{11,12} while the more complex solvation and mobility of the hydroxide ion is still controversially discussed.^{13,14}

Computer simulations based on molecular dynamics (MD) can provide many details of chemical processes in water at the atomic level. A mandatory condition for obtaining reliable results is an accurate description of the atomic interactions. While the current state-of-the-art in the simulation of liquid water is the application of ab initio MD (AIMD)^{15,16} relying on the determination of the energy and forces by density-functional theory (DFT) on-the-fly, the computational costs of these simulations are enormous, which prevents the investigation of many interesting questions. Consequently, to extend the length and time scales of simulations of water, a vast number of more efficient potentials have been developed, but most of them are rigid body potentials, like TIP3P,¹⁷ TIP4P,¹⁷ SPC,¹⁸ and ST2,¹⁹ which cannot describe the mak-

Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany. Fax: +49 234 2314045; Tel: +49 234 3226749; E-mail: joerg.behler@theochem.ruhr-uni-bochum.de

ing and breaking of chemical bonds and thus the dissociation and recombination of water molecules.⁷ To overcome these limitations, many different types of more advanced potentials for water have been developed. Some potentials built especially for protonated water clusters include empirical valence bond potentials (EVBP),^{21–23} multi-state empirical valence bond (MS-EVB) potentials,²⁴ the Kozack and Jordan potential (KJ),²⁵ fitted to gas phase data and the anisotropic site potential (ASP)²⁶ developed using monomer properties for the long-range interactions and perturbation theory for the short-range terms. Further, also potentials based on very flexible functional forms to represent electronic structure data with very high accuracy have been proposed, like a potential based on permutation invariant polynomials by Bowman and co-workers for the protonated water dimer²⁷ and a correction to DFT calculations to almost coupled cluster quality for water by Gaussian approximation potentials by Csanyi and coworkers.^{28,29} Those approaches rely on a many-body expansion (MBE) of the total energy, usually truncated after the three-body term. Since the chemical composition of the interacting monomers changes during a proton transfer reaction, the extension of MBE-based potentials to larger protonated water clusters is not straightforward. However, very recently a generalization of the MBE method has been suggested that in principle also enables the treatment of chemical reactions.³⁰

In recent years, protonated water clusters have become important model systems to develop potential-energy surfaces (PESs) that are able to include the dissociation of water molecules in MD simulations, and the goal is to reach a quality in the description of the atomic interactions close to electronic structure methods. Considering the dissociation products of water in the construction of these potentials explicitly is very important, as the presence of a proton or hydroxide ion can change the local water structure including the hydrogen-bond network considerably.^{31,32} In the case of small protonated water clusters ($\text{H}^+(\text{H}_2\text{O})_n$ up to $n = 3$), extensive work has already been carried out to understand their structure and vibrational frequencies at various levels of theory.^{33–42,44,45} Hodges and Stone²⁶ proposed that, with increasing cluster size, the Eigen model rules over the Zundel configuration, since the intramolecular free OH bonds become stronger, while the hydrogen bonding network of the whole system gets weaker. For larger clusters, *ab initio* calculation become computationally expensive as the number of possible isomers grows drastically, while empirical models are often not sufficiently accurate.

Apart from the approaches mentioned above, also artificial neural networks (NNs) have been used to address the PES of water. In 1998 Gassner et al. developed a water model based on rigid water molecules to improve the description of ions in solution.⁴³ Also the PESs of the free water molecule⁴⁶ and of the water dimer⁴⁷ have been constructed using NNs, and an extension of the TIP4P model to include polarization

has been published in 2002.⁴⁸ Apart from total energy surfaces, NNs have also been applied very successfully to express atom-centered electrostatic multipoles to generally improve the quality of classical force fields for water.^{49,50}

Recently, we have shown that high-dimensional NN potentials^{51,52} can be constructed for neutral water clusters including all degrees of freedom explicitly and that a very close agreement with reference electronic structure calculations can be achieved.^{53–55} In this work we investigate if NN potentials are able to describe the PES of small protonated water clusters with the accuracy of first-principles. We have chosen DFT as reference method to construct the NN potential, since our long-term goal is to develop a reactive and fully flexible potential for liquid water enabling the dissociation and recombination of water molecules and having the same applicability and accuracy as *ab initio* MD, which is typically DFT-based.¹⁵ On the other hand, due to its efficiency, a NN potential would allow to address much larger systems containing thousands of molecules on long time scales. While other electronic structure approaches like coupled cluster⁵⁶ provide quantitatively more accurate results for water clusters and although they have also been applied to construct PESs of such systems,⁵⁷ currently only DFT will allow us to carry out the required number of reference calculations of periodic systems needed for the construction of a PES for liquid water. Further, selecting DFT as reference method will enable extending the PES step by step by further subsystems, which are difficult or even impossible to study by wavefunction-based methods, like metal surfaces.

Specifically, the aim of the present work is to investigate if NN potentials are able to describe the structural, vibrational, and energetic properties of protonated water clusters with an accuracy comparable to that of electronic structure methods. For this purpose we use DFT employing the generalized gradient approximation (GGA) as reference method, which represents a PES of a very similar complexity like the PES obtained from wavefunction-based methods. Using clusters ranging from a protonated water molecule up to the protonated octamer we demonstrate that all investigated properties are in excellent agreement with the underlying DFT data independent of the system size using a single fit. Apart from the global and local minima of the PESs, also intermediate non-equilibrium structures are very well represented, which is another important prerequisite for future applications to MD simulations. Finally, first preliminary results are given addressing the NN-based simulation of liquid water.

2 High-Dimensional Neural Network Potentials

In recent years a lot of progress has been made in the construction of atomistic potentials based on a combination of electronic structure calculations and machine learning techniques.⁵⁸ Several methods have been developed for the representation of PESs, based, e.g., on the modified Shepard interpolation employing Taylor expansions,⁵⁹ permutation invariant polynomials,⁶⁰ and Gaussian processes.^{28,61} All these methods have in common that they are not relying on a functional form derived from physical considerations. Instead, they use very simple but flexible functional forms to reproduce a given set of electronic structure data as accurately as possible.

Another important class of machine learning potentials, which has first been proposed by Blank et al. in 1995,⁶² is based on artificial neural networks.⁶³ In the following two decades the development of NN potentials has seen enormous progress, and they have now been employed to a wide range of systems,^{64,65} from small molecules in the gas phase to condensed systems. In the present work, we use the NN potential method proposed by Behler and Parrinello,^{51,52,66} which allows to construct high-dimensional PESs of systems containing a large number of atoms with and without periodic boundary conditions. Since this method is applicable to systems of different size, it enables constructing a single PES for protonated water clusters with varying numbers of water molecules. In its most basic form, the total energy expression is given by

$$E_{\text{total}} = \sum_{i=1}^{N_{\text{atom}}} E_i \quad , \quad (1)$$

where the E_i are atomic energy contributions depending on the local chemical environments of the atoms, and the sum is running over all N_{atom} atoms i in the system. The E_i are provided by individual atomic NNs as a function of input vectors consisting of sets of many-body symmetry function values,⁶⁶ which serve as structural fingerprints to describe the geometric environments of the atoms up to a cutoff radius R_c . For each atom, a separate atomic NN is used to construct the functional relation between the chemical environment and the atomic energy contribution. It should be noted that, as atomic energies are no quantum mechanical observables, the parameters of the individual atomic NNs are adjusted to yield the correct total energies as a function of the atomic configuration, while individual reference atomic energies are not required to train the NNs.

Any electronic structure method can be used to generate the reference data required to determine the NN parameters, and apart from total energies also forces can be used. The latter do not represent independent additional target output nodes of the

NN, but since the NN forces are exact analytic derivatives of the NN total energy expression, the DFT forces can directly be used to complement the training of the PES. In order to ensure permutation symmetry of the potential with respect to the interchange of like atoms, which is very challenging for machine learning potentials, there is one set of NN parameters and one specific NN structure for each element, i.e., in case of water there is an atomic NN for hydrogen and an atomic NN for oxygen. Both NNs are evaluated as many times as there are atoms of the respective elements. More details about the high-dimensional NN approach and the functional forms of the employed many-body symmetry functions can be found elsewhere.^{52,65,66}

High-dimensional NN potentials have the important advantage that they do not contain any system-specific terms and that they are generally able to describe all types of atomic interactions from covalent bonds to metallic bonding and dispersion interactions. This has been demonstrated in many applications.⁶⁷⁻⁷³ Alternatively, it has also been shown that it is possible to construct the total energy as a sum of environment-dependent atom pairs.⁷⁴ Still, in this case the computational costs are higher than for the atom-based energy expression of Eq. 1, as for a given structure the number of pairs is much larger than the number of atoms. The typical errors of high-dimensional NN potentials are in the order of a few meV per atom for total energies and 0.1-0.3 eV/Bohr for forces. Concerning the efficiency, the NN potential enables the calculation of the energies and forces of about 100-200 atoms per compute core and per second, and the method is easy to parallelize as each atom can be computed separately on a different core and only the atomic energies and forces need to be collected to yield the final result.

3 Computational Details

The reference DFT calculations to construct the NN potential for protonated water clusters have been carried out using the program FHI-aims.⁷⁵ FHI-aims is an all-electron code employing atom-centered orbitals as basis functions, which are given numerically on spherical grids. In the present work we have selected the 'tier 2' basis set and a real-space cutoff for the basis functions of 8 Å providing a convergence of a few meV per atom for energy differences. For this type of basis functions the basis set superposition error (BSSE) is usually very small. For the water dimer we have found a BSSE of only 0.64 meV per atom, and consequently no BSSE correction has been required. Another advantage of using atomic orbital basis sets is the possibility to carry out periodic and non-periodic calculations in a consistent way, which will be important for an extension of the present work to liquid water by adding periodic water structures to the training set. The PBE functional by Perdew, Burke and Ernzerhof^{76,77} has been

selected to describe electronic exchange and correlation for all reference calculations.

The selection of the structures to be included in the reference set is a crucial step, because, if important structures are missing, the NN potential will not be reliable for the corresponding atomic configurations. If, on the other hand, the reference set is too large, carrying out the DFT calculations requires a significant amount of CPU time, and the training process to determine the parameters of the NN potential can also become very demanding as a lot of data needs to be processed.

Here we follow an iterative procedure to improve the NN potential consisting of the construction of preliminary potentials and tests of these potentials to identify missing configurations in the training set, which has been suggested by Artrith and Behler.⁶⁹ In essence, based on a first set of available DFT data several initial NN potentials are constructed, which are then used to find structures that are not well described. This is achieved by comparing the NN predictions for energies and forces obtained from the different NN potentials. Only if all NN potentials predict very similar energies and forces, the NN is assumed to be reliable. If different NN potentials predict different energies and forces, then there is not enough information about the respective part of the configuration space in the training data and more reference calculations are needed. This procedure can be repeated with improved NN potentials until a self-consistent potential has been obtained and the predictions are reliable in all relevant situations. Typically, the probe structures are generated by the same type of applications that will also be run with the final potential, e.g. geometry optimizations, MD, replica exchange molecular dynamics (REMD)⁷⁸ or Monte Carlo simulations. In the present work, the training set consists of neutral and protonated water monomers, as well as protonated water clusters up to the octamer. For setting up the first preliminary NN potentials, random clusters, structures from the literature,⁷⁹ as well as randomly distorted configurations of these minimum geometries have been calculated by DFT.

In particular for larger clusters ($N \geq 5$), the identification of energetically favorable random structures is challenging. Therefore, in these cases we have used low-energy structures found in Car-Parrinello molecular dynamics simulations (CPMD)¹⁶ to generate reasonable reference geometries employing the CPMD program.⁸⁰ The resulting CPMD simulations do not require a high level of convergence since we are interested only in the generation of thermally accessible and thus relevant reference configurations but not in their energies. Selected structures from the simulated trajectories have then been recalculated by FHI-aims for their inclusion in the reference set. In the further refinement of the potential, most structures have been obtained from REMD runs using the code TINKER⁸¹ and its replica exchange extension TiReX⁸² based

on preliminary NN potentials.

For a given set of DFT reference data, the NN potential is constructed by first splitting the data set into a training set ($\approx 90\%$), which is used to determine the parameters of NN, and an independent test set ($\approx 10\%$), which is not used for fitting but provides an estimate for the accuracy of the potential for structures not included in the training process. The flexibility of the NN is determined by the number of fitting parameters, which is given by the structure of the NN, i.e., the number of hidden layers and nodes per layer. If the NN is too flexible, overfitting can occur resulting in a poor performance for structures not included in the training set. If a too small NN is used, it will not be able to resolve all fine details of the PES resulting in notable residual errors. The most efficient way to determine the optimum NN size is to simply test different NN structures and to choose the one with the smallest errors of the data in the training and the test set. For each tested NN size, the weight parameters have been determined by minimizing the errors of the energies and force components of the training structures employing the global extended Kalman filter.⁸³ The error function is given by

$$\Gamma = \frac{1}{N_{\text{struct}}} \sum_{i=1}^{N_{\text{struct}}} \left[(E_{\text{NN}}^i - E_{\text{DFT}}^i)^2 + \frac{\alpha}{3N_{\text{atom}}^i} \sum_{j=1}^{3N_{\text{atom}}^i} (F_{j\text{NN}}^i - F_{j\text{DFT}}^i)^2 \right], \quad (2)$$

where N_{struct} is the total number of structures, N_{atom}^i is the number of atoms in structure i , E_{NN}^i and E_{DFT}^i represent the NN and DFT energies, and $F_{j\text{NN}}^i$ and $F_{j\text{DFT}}^i$ are the x , y , and z force components of all atoms, respectively. For each structure in each iteration, the weights are adjusted until the root mean squared error (RMSE) values for the energies and forces in the test set reach a minimum. The relative influence of the energies and forces on the fitting process can be controlled by a parameter α .

The construction of the NN potential has been carried out with our in-house program RuNNer.⁸⁴ The employed symmetry functions, which are described in detail elsewhere,⁶⁶ are given by the parameters in the supplementary information. A cutoff radius of 19.0 Bohr ($\approx 10 \text{ \AA}$) has been used in the present work to define the energetically relevant local atomic environments. For this purpose, convergence tests with different cutoff radii have been carried out. Apart from the short-range atomic energy contributions discussed in Section 2, which are a function of the atomic environments up to the employed cutoff radius, in principle also an electrostatic energy contribution based on environment-dependent atomic charges could be used.^{53,70} In the present work, the electrostatic term has been omitted, because the protonated water clusters are sufficiently small and all atoms in the clusters are located within the cutoff radii of all other atoms. Therefore, all electrostatic energy contributions can be described by the “short-range” atomic energy term.

The global and local minima of the protonated water clusters reported below have been identified in REMD simulations similar to our previous work for neutral water clusters.^{54,55} The relative stabilities of these clusters have been assessed using the binding energy per water molecule E_{bind} ,

$$E_{\text{bind}} [\text{H}^+(\text{H}_2\text{O})_n] = \frac{1}{n} [E [\text{H}^+(\text{H}_2\text{O})_n] - (n-1)E [\text{H}_2\text{O}] - E [\text{H}_3\text{O}^+]] \quad (3)$$

which corresponds to the energy gained by forming the protonated cluster containing n monomers from $n-1$ neutral water molecules and a H_3O^+ ion, and the energy of formation E_{form} , which is defined as the energy change upon adding another water molecule to a protonated cluster already containing $n-1$ monomers,

$$E_{\text{form}} [\text{H}^+(\text{H}_2\text{O})_n] = E [\text{H}^+(\text{H}_2\text{O})_n] - E [\text{H}^+(\text{H}_2\text{O})_{n-1}] - E [\text{H}_2\text{O}] \quad (4)$$

The harmonic normal mode analyses using the NN potential and DFT have been performed by computing the Hessian from finite differences of the atomic forces.

4 Results

4.1 Neural network potential for protonated water clusters

The final DFT training set consists of 29,851 structures including the neutral water monomer, the protonated water monomer as well as a large number of protonated water clusters from the dimer to the octamer. The test set used to check the reliability of the potential for atomic configurations not included in the training set contains 3,231 structures. Various NN architectures have been tested and several initial sets of random weight parameters have been used to find the optimum NN potential to represent these data. We found that many different NNs yield similar errors of the energies and forces. Accordingly, the final potential is not very sensitive to the actual choice of the number of hidden layers and nodes per layer. We have then selected the NN with the smallest errors of the energies and forces in the test set. The best potential we obtained contains a NN with two hidden layers and 80 nodes per layer for each element. Still, the resulting number of NN fitting parameters is not the same for both elements due to the different number of input symmetry functions (cf. supplementary information), they are 8,481 for hydrogen and 8,161 for oxygen. Although the number of NN parameters is thus substantial, this number has to be related to the amount of information that is used to determine the values of these parameters. In Table 1 the numbers of structures for the individual cluster sizes and their RMSE values for the training and the test sets are compiled.

Since we use the total energies and the force components of all atoms to train the NN potential, in total 555,952 pieces of information about the PES are available, which is substantially larger than the number of fitting parameters.

The overall energy RMSEs of the training and the test set are only 0.97 meV/atom and 1.08 meV/atom. The corresponding values for the force RMSEs are 40.11 and 40.95 meV/Bohr. The negligible differences between the training and the test set errors indicate that no significant overfitting is present, which would give rise to a much higher error for structures not used for training the NNs. While the energy errors remain in the order of about 1 meV/atom for all clusters, as can be seen in Table 1 there is a slight increase in the force errors with growing cluster size. Still, even the force RMSE of the octamer is only about 50 meV/Bohr. As will be discussed below, this small force error does not affect the global and local minima found using the NN potential.

For each cluster size, the distribution of the DFT binding energies of the reference geometries according to Eq. 3 and their corresponding energy RMSE values are plotted in Fig. 1. As the size of the clusters increases, more reference structures are included because of the growing complexity of the system. It can also be seen that the most stable structures with the lowest binding energies have smaller errors than high-energy structures, which is an advantage for applications like geometry optimizations. Still, even the high-energy structures have small errors well below 2 meV per atom.

4.2 Structural and energetic properties of protonated water clusters

4.2.1 Global minima and vibrational frequencies The first test of the NN potential has been the determination of the global minima for the protonated water clusters up to the octamer by REMD simulations. The results are summarized in Table 2, which also includes figures of the NN optimized geometries. For comparison these structures have also been reoptimized by DFT. For all clusters, the binding energy difference between DFT and the NN potential is very small, the largest deviation of 2 meV per molecule occurred for the dimer, which in contrast to the larger clusters exhibits a symmetric hydrogen bond corresponding to a Zundel cation.

Several quantities have been used to characterize the structural properties of the clusters: the average length of the free OH bonds $\bar{R}_f(\text{OH})$, the average length of the shorter OH distance in the hydrogen bonds $\bar{R}_t(\text{OH})$, the average oxygen-oxygen distance in hydrogen bonds $\bar{R}(\text{OO})$, and the average difference of both OH distances in the hydrogen bonds $\bar{\delta}(\text{OH})$. We find (cf. Table 2) that $\bar{R}_f(\text{OH})$ reduces slightly with increasing cluster size, and we will see later that this trend correlates with the trend in their stretch frequencies. Similarly, $\bar{R}_t(\text{OH})$ decreases with cluster size due an increase in the num-

ber of hydrogen bonds in the system. While H_5O_2^+ is a Zundel cation, the larger clusters from H_7O_3^+ to $\text{H}_{17}\text{O}_8^+$ have 2, 3, 5, 6, 9, and 11 hydrogen bonds, respectively. $\bar{\delta}(\text{OH})$ increases with cluster size indicating a more pronounced asymmetry of the hydrogen bond. The weakening of the hydrogen bond with increasing cluster size is also evident in the trend of $\bar{R}(\text{OO})$. For all clusters, the non-averaged structural properties $R(\text{OO})$, $R_f(\text{OH})$, $\delta(\text{OH})$ are plotted and analyzed in Fig. 2.

The trend in the binding energy per water molecule of the global minima with increasing cluster size is shown in Fig. 3a. The strongest binding per water molecule of about $-880 \text{ meV}/\text{H}_2\text{O}$ is found for the protonated water trimer. H_5O_2^+ exhibits a weaker binding energy per molecule compared to that of the H_7O_3^+ cluster, since it is a Zundel ion where the central hydrogen atom is shared by both water molecules. From H_7O_3^+ to $\text{H}_{17}\text{O}_8^+$, the binding energy decreases continuously. The trend in the formation energies in Fig. 3b indicates that the largest energy gain is obtained when forming the H_5O_2^+ ion from H_3O^+ and a water molecule. For larger clusters the energy gain when adding an additional water molecule decreases to about 600 meV .

For the global minimum structures of the protonated water clusters a harmonic normal mode analysis has been carried out. In Fig. 4a the NN and DFT values of all harmonic normal mode frequencies of the minimum geometries of the protonated water clusters are compared, and we find excellent agreement. This plot also illustrates different regions of the spectrum, like the high frequency OH stretches between 3500 and 4000 cm^{-1} , hydrogen bond stretches between 3000 and 3500 cm^{-1} , H_3O^+ stretches between 2000 and 3000 cm^{-1} and low frequency bending motions below 2000 cm^{-1} . In all regions, the NN frequencies are very close to the DFT results, which confirms an accurate representation of the PES by the NN potential. Some important stretch frequencies are also listed in the supplementary information and compared to experimental values.

The NN predicted average red shift of the OH stretch frequencies $\bar{\nu}_f(\text{OH})$ of the protonated water clusters with respect to the neutral water molecule is compared to DFT and available experimental results in Fig. 4b. When a H^+ ion is added to a neutral water molecule, the oxygen atom attracts the H^+ and forms an Eigen cation H_3O^+ resulting in an increase in its $\bar{R}_f(\text{OH})$ as shown in Fig. 4b. This is followed by increased OH stretch amplitude and thus red shift of the respective frequencies by about 230 cm^{-1} .

As a water molecule is added to the hydronium ion, a Zundel ion is formed accompanied by a considerable red shift of its stretch frequency. All other hydrogen atoms are bonded to the respective oxygen atom and act as free OH bonds characterized by a shorter bond length (Fig. 4c). Thus, $\bar{\nu}_f(\text{OH})$ of H_5O_2^+ is blue shifted compared to the protonated water monomer. As the size of the clusters increases, more neutral

water molecules are added and more OH bonds are available in the system. The deviation observed between the experimental and DFT shifts is because of the anharmonicity and also because of the limitations of the employed exchange correlation functional. In order to stabilize the cluster, several of the OH bonds participate in hydrogen bonding, which allows the free OH bonds to become shorter and vibrate much faster than that of the smaller clusters. This phenomenon is described in Fig. 4b and Fig. 4c. The slight deviation in the NN prediction for $\bar{R}_f(\text{OH})$ in case of $\text{H}_{15}\text{O}_7^+$ and $\text{H}_{17}\text{O}_8^+$ is correlated to a similar deviation in their corresponding $\bar{\nu}_f(\text{OH})$ as seen in Fig. 4b, and $\bar{\nu}_f(\text{OH})$ approaches the neutral water OH stretch frequency with increasing cluster size. Fig. 4d shows a direct correlation between the increase in $\bar{R}_f(\text{OH})$ of the protonated water clusters and the shift in their corresponding stretch frequencies.

4.2.2 Local minima, potential scans, and non-equilibrium structures Apart from a description of the global minima and their structural and vibrational properties, also an accurate representation of non-equilibrium structures is of crucial importance for obtaining a reliable PES for protonated water clusters. In Fig. 5 the global minima and some local minima that we have identified in the REMD simulations are shown. We have reoptimized these structures by DFT to determine their DFT binding energies, which are shown along with the NN potential-based binding energies in Fig. 6. In spite of the very small energy difference between the different isomers, the energetic sequence of most structures is very well described by the NN potential with only few exceptions for e.g. structures B and C of the tetramer, which are energetically very similar, and it should be noted that also in case of the reference DFT calculations the selected PBE functional does not allow for an unambiguous assessment of the energetic order of these structures.

In Fig. 7 the one-dimensional energy landscape corresponding to the inversion symmetry of H_3O^+ is presented. The dihedral angle between the plane of the three hydrogen atoms and one of the three equivalent hydrogen-oxygen bonds has been varied. The inversion pathway has been generated by the PATH program of the TINKER⁸¹ package. This cut through the PES represents the low frequency umbrella normal mode and the NN potential predicts the energies in close agreement with DFT. A second one-dimensional cut through the PES of H_3O^+ in Fig. 8 describes the dissociation of a hydrogen atom from the hydronium ion. It should be noted that the reference data set used to train the NN does not contain structures with any OH distance longer than 1.2 \AA . Still, the NN potential describes the dissociation PES accurately up to much longer bond lengths, which underlines at least limited extrapolation capabilities of the potential. In general, whenever such an extrapolation situation occurs, the NN program issues a warning,

which can be used to further extend the reference set systematically.

Fig. 9 describes the energy barrier for the proton transfer in H_5O_2^+ for several fixed oxygen-oxygen distances R_{OO} . The structures have been optimized using the NN potential as well as DFT by fixing the two oxygen atoms and the central hydrogen atom. For all investigated values of R_{OO} there is an excellent agreement between the NN potential and DFT. For the equilibrium distance there is a single minimum with $\delta(\text{OH}) = 0$ corresponding to the Zundel cation, while for larger oxygen-oxygen separations a double well potential emerges. It is evident that the barrier for proton transfer increases with R_{OO} . At larger separation, the preferred Zundel structure changes to an Eigen model with two minima.

In the case of protonated water trimer, three distinct minima have been identified differing in their hydrogen-bond pattern. The global minimum structure A in Fig. 10a is of C_1 symmetry. Another local minimum labeled B has C_s symmetry with a mirror plane along the free OH axis of the central Eigen complex is found within 0.15 meV from the global minimum. A high energy local minimum marked C with C_s symmetry is about 3.07 meV higher in energy than the global minimum. Fig. 10b describes the interconversion pathway between the three minima. Fig. 10a shows the corresponding pathway from A, where one water molecule bound to the central hydronium ion is stationary, while the other rearranges to form C. From there, both the terminal water molecules reorient to form B. This pathway is consistent with earlier publications where the DFT BLYP/aug-cc-pVDZ level of theory has been used.³⁷ Fig. 10b compares the NN and corresponding single point DFT reference energies of the pathway.

As the complexity of the energy landscape increases rapidly with the size of the clusters, the largest cluster we have investigated in more detail is the protonated tetramer. The global minimum structure A in Fig. 11a is an extended Eigen complex with a central hydronium ion hydrogen bonded to three dangling water molecules in its periphery. The structure with the next lowest energy is an extended Zundel complex B with two water molecules hydrogen bonded to the Zundel ion in *trans* fashion. Slightly higher in energy is the third local minimum structure C, which is again an extended Zundel complex with two water molecules hydrogen bonded to the Zundel ion in *cis* alignment. The high energy minimum D is a cyclic structure formed by four hydrogen bonded chain of water molecules with a central Eigen type hydronium ion. A disconnectivity graph⁸⁵ is used to visualize the total PES of H_6O_4^+ in the form of super basins containing the minima. All the minima are connected in the order of their transition state energies. From the resulting graph, we can ascertain the relative ease in reaching different minima from any starting structure.⁸⁶ We find three super basins where the super basin I in Fig. 11b consist of minima B and C since the energy barrier

between these two minima is very low. Super basin II includes minima B, C and D. Although minimum D has the highest energy, it enters super basin II before A since the transition state connecting B, C and D is lower in energy than the transition state between minima B, C, and A. The global minimum together with the other three local minima forms the final super basin III.

Finally, in order to unravel the quality of the representation of arbitrary configurations, we have investigated a large number of structures emerging in the REMD simulations. In Fig. 12 we show a comparison of the NN and DFT energies for a number of structures of the $\text{H}_{15}\text{O}_7^+$ cluster. Like for all other cluster sizes that we studied, the NN predictions are very close to DFT.

4.3 Outlook: Towards liquid water

High-dimensional neural networks are not limited to neutral and protonated water clusters but they are also applicable to the condensed phase. Currently,⁸⁷ we are working on the development of full-dimensional and reactive NN potentials for bulk water, which will be applied to study condensed water under various conditions and also allow the description of proton-transfer reactions in the liquid. Several NN potentials are under construction based on different approximations of the exchange-correlation functional and trained to a large set of periodic water configurations in the liquid and crystalline state.

As a first test of a PBE-based NN potential we show here the radial distribution function of liquid water obtained from an NN simulation of 128 water molecules in the *NVT* ensemble averaged over a 1 ns trajectory (cf. Fig. 13). For comparison, results from AIMD simulations for a simulation containing 64 molecules are shown.⁸⁸ Both simulations lead to an over-structured and undercooled liquid compared to experiment, which is a well-known feature of the underlying PBE PES. Due to the high computational costs, the simulation time of the AIMD simulations is limited to the picosecond timescale, which might not be sufficient to properly equilibrate the undercooled liquid. With NN potentials however, simulation times of several nanoseconds can be routinely accessed, which ensures converged results and also allows to compute thermodynamic properties that are beyond reach of on-the-fly AIMD simulations.

To illustrate the capability of NN potentials to simulate proton transfer in the condensed phase, Fig. 14 shows the free-energy profile (potential of mean force) along the proton transfer coordinate of an excess proton in a periodic box with 32 water molecules obtained from NN simulations. Despite the fact that the potential was only parametrized using distorted configurations of neutral water molecules without explicit charged defects it is able to describe the structural

diffusion of the H^+ ion in the liquid. The proton transfer barrier is below $k_{\text{B}}T$ and about half as large as the one obtained from AIMD simulations based on a different density-functional (BLYP, Ref. 89). It will be interesting to further refine the NN potential and to study the influence of density-functional, system size and nuclear quantum effects on the structure and dynamics of H^+ and OH^- ions in liquid water. This work is currently in progress.⁸⁷

5 Summary

A high-dimensional neural network potential for protonated water clusters up to the protonated octamer has been constructed and analyzed using DFT calculations as reference method employing the PBE functional. For all investigated energetic, structural and vibrational properties we find excellent agreement between the NN predictions and DFT data, which have been obtained for comparison. There is no significant difference in the accuracy for global and local minima as well as for non-equilibrium structures along transition pathways or structures extracted from REMD simulations. In all cases we find very small energy errors of only a few meV per atom and forces, which deviate from the corresponding DFT values typically by less than 50 meV/Bohr.

Although the construction of the NN potential requires a substantial number of reference data from electronic structure calculations to ensure that the right topology of the PES is obtained, the evaluation of the final potential is many orders of magnitude more efficient than DFT enabling significantly faster simulations of a comparable quality.

One of the major aims of the present work has been to investigate if NN potentials are able to describe the properties of protonated water clusters with the same accuracy that has been achieved in previous work for neutral water clusters. Although the structure of free protonated water clusters is very different from the solvation pattern of hydronium ions in the bulk, still the capability of the NN to capture the structural features of protonated water clusters varying strongly with cluster size is a stringent test for the general applicability of NN potentials. Our findings, along with first preliminary but very encouraging results for liquid water, demonstrate that NN potentials are a very promising tool to construct highly accurate PESs for bulk water, which necessarily has to provide a reliable description of the dissociation and recombination of water molecules close to first-principles accuracy. Such work on a reactive NN PES for liquid water employing our high-dimensional NN method is currently in progress.⁸⁷ As NNs have been shown to provide reliable energy landscapes for other materials like metals⁶⁹ and oxides,⁷⁰ also an extension to the solid-liquid interface is expected to be an interesting application of NN PESs.

Acknowledgements

This work was supported by the Cluster of Excellence RESOLV (EXC 1069) funded by the Deutsche Forschungsgemeinschaft and the DFG (Emmy Noether program Be3264/3-1, Heisenberg program Be3264/6-1 and project Be3264/5-1). T.M. thanks the Studienstiftung des Deutschen Volkes for a PhD fellowship. Discussions with Daniel Muñoz-Santiburcio are gratefully acknowledged.

References

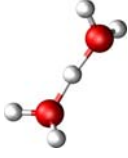
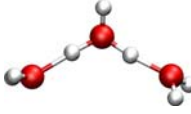
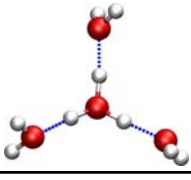
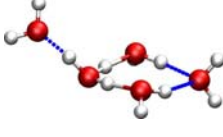
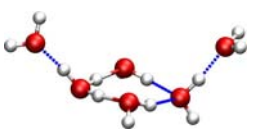
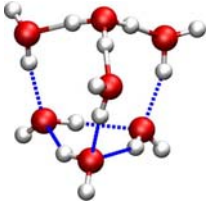
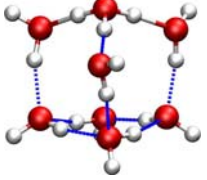
- 1 P. Ball, *Life's matrix: a biography of water*, Univ. of California Press, 2001.
- 2 R. Ludwig, *Angew. Chem. Int. Ed.*, 2001, **40**, 1808.
- 3 N. Agmon, *Acc. Chem. Res.*, 2012, **45**, 63.
- 4 R. Bukowski, K. Szalewicz, G. C. Groenenboom and A. van der Avoird, *Science*, 2007, **315**, 1249.
- 5 B. Guillot, *J. Mol. Liq.*, 2002, **101**, 219.
- 6 A. Tokmakoff, *Science*, 2007, **317**, 54.
- 7 Y. Zubavicus and M. Grunze, *Science*, 2004, **304**, 974.
- 8 P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odellius, H. Ogasawara, L. Å. Näslund, T. K. Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson and A. Nilsson, *Science*, 2004, **304**, 995.
- 9 H. Helmholtz, *Ann. Phys. Chem.*, 1853, **LXXXIX**, 211.
- 10 W. Schmickler, *Chem. Rev.*, 1996, **96**, 3177.
- 11 N. Agmon, *Chem. Phys. Lett.*, 1995, **244**, 456.
- 12 D. Marx, *ChemPhysChem*, 2006, **7**, 1848.
- 13 M. E. Tuckerman, A. Chandra and D. Marx, *Acc. Chem. Res.*, 2006, **39**, 151.
- 14 D. Marx, A. Chandra and M. E. Tuckerman, *Chem. Rev.*, 2010, **110**, 2174.
- 15 D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, 2009.
- 16 R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471.
- 17 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
- 18 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, *Intermolecular Forces*, Reidel, Dordrecht, 1981, p. 331.
- 19 F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, 1974, **60**, 1545.
- 20 C. Vega and J. L. F. Abascal, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19663.
- 21 T. James and D. J. Wales, *J. Chem. Phys.*, 2005, **122**, 134306.
- 22 R. Kumar, R. A. Christie and K. D. Jordan, *J. Phys. Chem. B*, 2009, **113**, 4111.
- 23 U. W. Schmitt and G. A. Voth, *J. Phys. Chem. B*, 1998, **102**, 5547.
- 24 C. Knight and G. A. Voth, *Acc. Chem. Res.*, 2012, **45**, 101.
- 25 R. E. Kozack and P. C. Jordan, *J. Chem. Phys.*, 1992, **96**, 3131.
- 26 M. P. Hodges and A. J. Stone, *J. Chem. Phys.*, 1999, **110**, 6766.
- 27 X. C. Huang, B. J. Braams and J. M. Bowman, *J. Chem. Phys.*, 2005, **122**, 044308.
- 28 A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Phys. Rev. Lett.*, 2010, **104**, 136403.
- 29 A. P. Bartók, M. J. Gillan, F. R. Manby and G. Csányi, *Phys. Rev. B*, 2013, **88**, 054104.
- 30 P. Pinski and G. Csányi, *J. Chem. Theory Comput.*, 2014, **10**, 68–75.
- 31 T. S. Zwier, *Science*, 2004, **304**, 1119.
- 32 D. Marx, M. E. Tuckerman, J. Hutter and M. Parrinello, *Nature*, 1999, **397**, 601.
- 33 D. H. Zhang, M. A. Collins and S. Y. Lee, *Science*, 2000, **290**, 961.

- 34 K. R. Asmis, N. L. Pivonka, G. Santambrogio, M. Bruemmer, C. Kaposta, D. M. Neumark and L. Woeste, *Science*, 2003, **299**, 1375.
- 35 M. H. Begemann, C. S. Gudeman, J. Pfaff and R. J. Saykally, *Phys. Rev. Lett.*, 1983, **51**, 554.
- 36 D. J. Wales, *J. Chem. Phys.*, 1999, **110**, 10403.
- 37 D. J. Wales, *J. Chem. Phys.*, 1999, **111**, 8429.
- 38 M. Jieli and M. Aida, *J. Phys. Chem. A*, 2009, **113**, 1586.
- 39 J. Dai, Z. Bacic, X. Huang, S. Carter and J. M. Bowman, *J. Chem. Phys.*, 2003, **119**, 6571.
- 40 Y. Xie, R. B. Remington and H. F. Schaefer III, *J. Chem. Phys.*, 1994, **101**, 4878.
- 41 D. Wei and D. R. Salahub, *J. Chem. Phys.*, 1994, **101**, 7633.
- 42 S. W. Charoensak Lao-ngam, P. Asawakun and K. Sagarik, *Phys. Chem. Chem. Phys.*, 2011, **13**, 4562.
- 43 H. Gassner, M. Probst, A. Lauenstein and K. Hermansson, *J. Phys. Chem. A*, 1998, **102**, 4596.
- 44 G. M. Chaban, J. O. Jung and R. B. Gerber, *J. Phys. Chem. A*, 2000, **104**, 2772.
- 45 J. M. Headrick, E. G. Diken, R. S. Walters, N. I. Hammer, R. A. Christie, J. Cui, E. M. Myshakin, M. A. Duncan, M. A. Johnson and K. D. Jordan, *Science*, 2005, **308**, 1765.
- 46 S. Manzhos, X. Wang, R. Dawes and T. Carrington, *J. Phys. Chem. A*, 2006, **110**, 5295.
- 47 K. T. No, B. H. Chang, S. Y. Kim, M. S. Jhon and H. A. Scheraga, *Chem. Phys. Lett.*, 1997, **271**, 152.
- 48 K. W. Cho, K. T. No and H. A. Scheraga, *J. Mol. Struct.*, 2002, **641**, 77.
- 49 C. M. Handley and P. L. A. Popelier, *J. Chem. Theory Comput.*, 2009, **5**, 1474.
- 50 C. M. Handley, G. I. Hawe, D. B. Kell and P. L. A. Popelier, *Phys. Chem. Chem. Phys.*, 2009, **11**, 6365.
- 51 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.
- 52 J. Behler, *J. Phys.: Condens. Matter*, 2014, **26**, 183001.
- 53 T. Morawietz, V. Sharma and J. Behler, *J. Chem. Phys.*, 2012, **136**, 064103.
- 54 T. Morawietz and J. Behler, *J. Phys. Chem. A*, 2013, **117**, 7356.
- 55 T. Morawietz and J. Behler, *Z. Phys. Chem.*, 2013, **227**, 1559.
- 56 R. J. Bartlett and M. Musial, *Rev. Mod. Phys.*, 2007, **79**, 291.
- 57 Y. Wang and J. M. Bowman, *Chem. Phys. Lett.*, 2010, **491**, 1.
- 58 C. M. Handley and J. Behler, *Eur. Phys. J. B*, 2014, **87**, 152.
- 59 J. Ischtwan and M. A. Collins, *J. Chem. Phys.*, 1994, **100**, 8080.
- 60 A. Brown, B. J. Braams, K. Christoffel, Z. Jin and J. M. Bowman, *J. Chem. Phys.*, 2003, **119**, 8790.
- 61 M. Rupp, A. Tkatchenko, K. R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.
- 62 T. B. Blank, S. D. Brown, A. W. Calhoun and D. J. Doren, *J. Chem. Phys.*, 1995, **103**, 4129.
- 63 C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1996.
- 64 C. M. Handley and P. L. A. Popelier, *J. Phys. Chem. A*, 2010, **114**, 3371.
- 65 J. Behler, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17930.
- 66 J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.
- 67 J. Behler, R. Martoňák, D. Donadio and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 185501.
- 68 N. Artrith, B. Hiller and J. Behler, *Phys. Status Solidi B*, 2013, **250**, 1191.
- 69 N. Artrith and J. Behler, *Phys. Rev. B*, 2012, **85**, 045439.
- 70 N. Artrith, T. Morawietz and J. Behler, *Phys. Rev. B*, 2011, **83**, 153101.
- 71 G. C. Sosso, G. Miceli, S. Caravati, J. Behler and M. Bernasconi, *Phys. Rev. B*, 2012, **85**, 174103.
- 72 H. Eshet, R. Z. Khaliullin, T. D. Kühne, J. Behler and M. Parrinello, *Phys. Rev. B*, 2010, **81**, 184107.
- 73 R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler and M. Parrinello, *Phys. Rev. B*, 2010, **81**, 100103.
- 74 Jovan Jose K.V., N. Artrith and J. Behler, *J. Chem. Phys.*, 2012, **136**, 194111.
- 75 V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput. Phys. Commun.*, 2009, **180**, 2175.
- 76 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
- 77 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1997, **78**, 1396.
- 78 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141.
- 79 M. P. Hodges and D. J. Wales, *Chem. Phys. Lett.*, 2000, **324**, 279.
- 80 CPMD, <http://www.cpmc.org/>, Copyright IBM Corp 1990-2008, Copyright MPI für Festkörperforschung Stuttgart 1997-2001.
- 81 J. W. Ponder and F. M. Richards, *J. Comp. Chem.*, 1987, **8**, 1016.
- 82 E. S. Penev, S. Lampoudi and J. E. Shea, *Comput. Phys. Commun.*, 2009, **180**, 2013.
- 83 T. B. Blank and S. D. Brown, *J. Chemometrics*, 1994, **8**, 391.
- 84 J. Behler, *RuNNer – A Software Package for Constructing High-Dimensional Neural Network Potentials*. Ruhr-Universität Bochum 2007-2014.
- 85 O. M. Becker and M. Karplus, *J. Chem. Phys.*, 1997, **106**, 1495.
- 86 D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, Cambridge University Press, 2004.
- 87 T. Morawietz et al., (*in preparation*).
- 88 E. Schwegler, J. C. Grossman, F. Gygi and G. Galli, *J. Chem. Phys.*, 2004, **121**, 5400.
- 89 D. Marx, M. E. Tuckerman and M. Parrinello, *J. Phys.: Condens. Matter*, 2000, **12**, A153–A159.

Table 1 Root mean squared errors (RMSEs) of the energies normalized per atom and of the forces for the full data set (“all”) and for each individual cluster size. The values refer to the training set, while the numbers in parentheses correspond to the structures in the independent test set not used for fitting the neural network parameters.

data set	number of structures	RMSE energies [meV / atom]	RMSE forces [meV / Bohr]
All	29851 (3231)	0.97 (1.08)	40.11 (40.95)
H ₂ O	796 (92)	0.31 (0.14)	2.32 (2.26)
H ₃ O ⁺	1273 (143)	0.3 (0.28)	5.28 (4.99)
H ₅ O ₂ ⁺	1321 (145)	1.06 (1.09)	17.2 (16.41)
H ₇ O ₃ ⁺	2408 (296)	0.73 (0.76)	24.77 (17.28)
H ₉ O ₄ ⁺	3426 (373)	0.71 (0.80)	19.52 (19.37)
H ₁₁ O ₅ ⁺	3196 (348)	0.69 (0.68)	23.33 (24.41)
H ₁₃ O ₆ ⁺	4700 (509)	1.12 (1.28)	40.37 (40.01)
H ₁₅ O ₇ ⁺	5985 (639)	1.07 (1.10)	41.47 (39.81)
H ₁₇ O ₈ ⁺	6746 (686)	1.17 (1.42)	49.43 (54.02)

Table 2 Global minimum geometries and binding energies E_{bind} of the protonated water clusters H_5O_2^+ to $\text{H}_{17}\text{O}_8^+$ optimized with the neural network (NN) potential and with density-functional theory (DFT). M is the number of water monomers in the cluster. $\bar{R}_f(\text{OH})$ is the average length of the free OH bonds not participating in hydrogen bonds, $\bar{R}_t(\text{OH})$ is the average length of the shorter OH distances in the hydrogen bonds, $\bar{R}(\text{OO})$ is the average distance between the oxygen atoms of neighboring molecules, and $\bar{\delta}(\text{OH})$ denotes the average absolute difference of both OH distances in a hydrogen bond. $|\text{DFT} - \text{NN}|$ is the absolute deviation between the DFT and the NN values.

M	cluster	method	E_{bind} [meV / H_2O]	$\bar{R}_f(\text{OH})$ [Å]	$\bar{R}_t(\text{OH})$ [Å]	$\bar{R}(\text{OO})$ [Å]	$\bar{\delta}(\text{OH})$ [Å]
2		DFT	-804	0.976	1.211	2.417	0.000
		NN	-806	0.976	1.210	2.415	0.000
		$ \text{DFT} - \text{NN} $	2	0.000	0.001	0.002	0.000
3		DFT	-880	0.972	1.064	2.491	0.367
		NN	-881	0.972	1.062	2.491	0.370
		$ \text{DFT} - \text{NN} $	1	0.000	0.002	0.000	0.003
4		DFT	-871	0.971	1.030	2.557	0.500
		NN	-870	0.972	1.029	2.556	0.501
		$ \text{DFT} - \text{NN} $	1	0.001	0.001	0.001	0.001
5		DFT	-821	0.971	0.997	2.673	0.672
		NN	-822	0.972	0.996	2.674	0.675
		$ \text{DFT} - \text{NN} $	1	0.001	0.001	0.001	0.003
6		DFT	-783	0.970	1.015	2.641	0.626
		NN	-784	0.970	1.018	2.645	0.625
		$ \text{DFT} - \text{NN} $	1	0.000	0.003	0.004	0.001
7		DFT	-755	0.970	1.003	2.719	0.765
		NN	-756	0.972	1.002	2.745	0.795
		$ \text{DFT} - \text{NN} $	1	0.002	0.001	0.026	0.030
8		DFT	-741	0.969	0.997	2.694	0.701
		NN	-741	0.971	1.009	2.681	0.684
		$ \text{DFT} - \text{NN} $	0	0.002	0.012	0.013	0.017

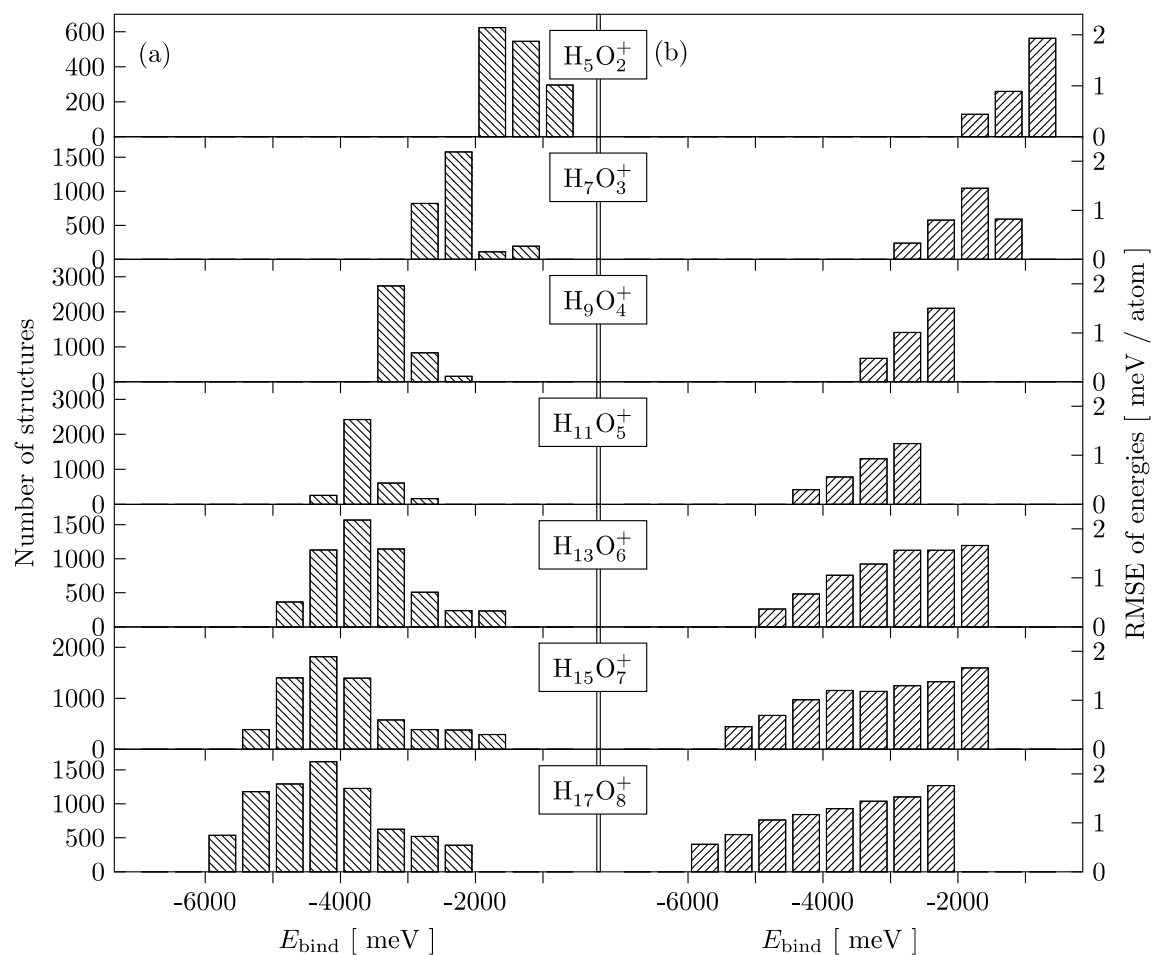


Fig. 1 Number of reference geometries utilized to generate the neural network (NN) potential for each cluster size as a function of the binding energy is shown in (a). The NN potential has been constructed using the combined data for all cluster sizes. In (b) the root mean squared errors (RMSEs) of the energies, which have been normalized per atom, are given for each group of structures. The most stable structures of each cluster size have the lowest errors.

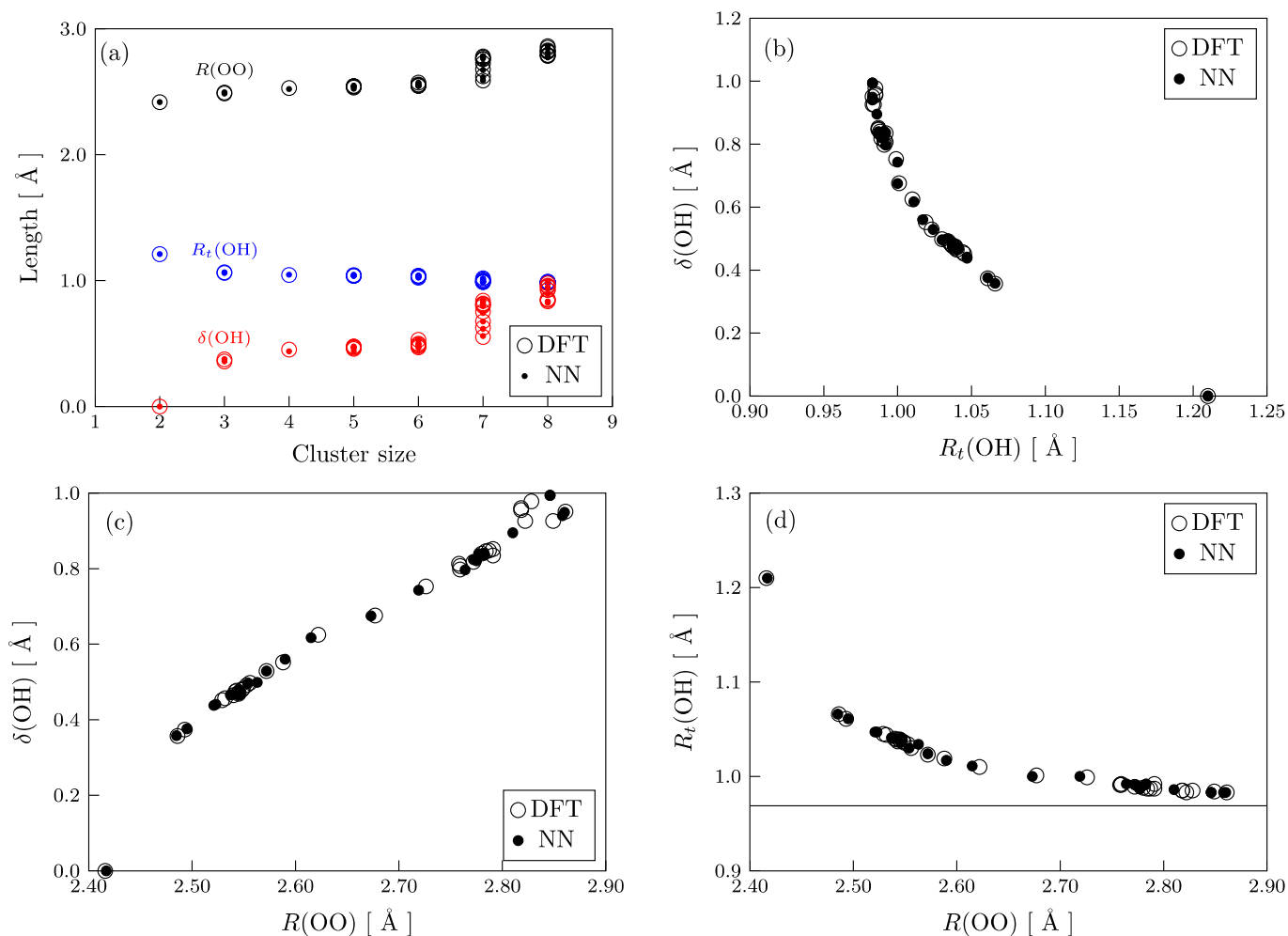


Fig. 2 Analysis of the structural trends in the global minimum geometries of the protonated water clusters. In panel (a) all inter-atomic distances of the protonated water clusters are shown. With the exception of the protonated water dimer, all structures have an Eigen-type central hydronium ion, which can be confirmed from the $\delta(OH)$ values representing the difference of the two OH bond lengths of the hydrogen bond. $R(OO)$ are the oxygen-oxygen distances between hydrogen-bonded molecules, and $R_t(OH)$ are the shorter oxygen-hydrogen distances in hydrogen bonds. In panel (b) it is demonstrated that $\delta(OH)$ decreases with growing $R_t(OH)$ as the excess proton moves towards an equidistant position from both oxygen atoms for smaller clusters. Panel (c) shows how $\delta(OH)$ increases with $R(OO)$ since the excess proton localizes in the vicinity of one of the two nearest water molecules and forms an Eigen ion if the distance between the monomers increases. Panel (d) shows how $R_t(OH)$ decreases with growing $R(OO)$ indicating as well a more asymmetric hydrogen bond for extended molecular separations. The line represents the OH bond length in a free neutral water molecule.

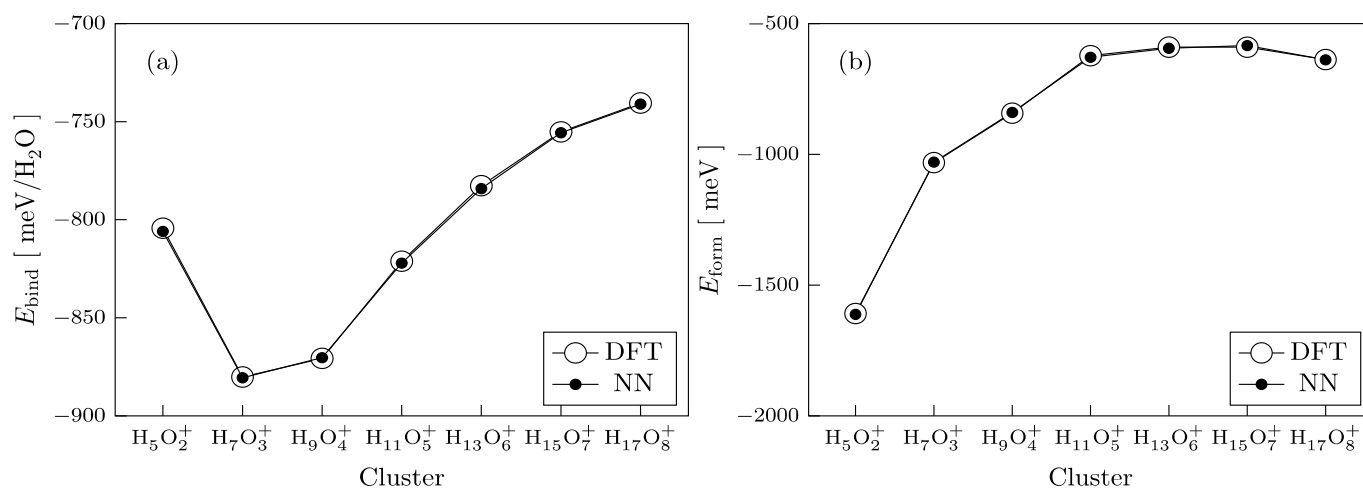


Fig. 3 Binding energy per water monomer as a function of the cluster size is shown in (a) (cf. Table 2). The protonated water trimer is found to have the highest energetic stability. The plot of the formation energy in (b) shows that the energy gain per additional water molecule in the protonated water clusters decreases with system size. In the special case of the protonated dimer, the binding and the formation energy are related by $E_{\text{form}}[H_5O_2^+] = 2 \cdot E_{\text{bind}}[H_5O_2^+]$.

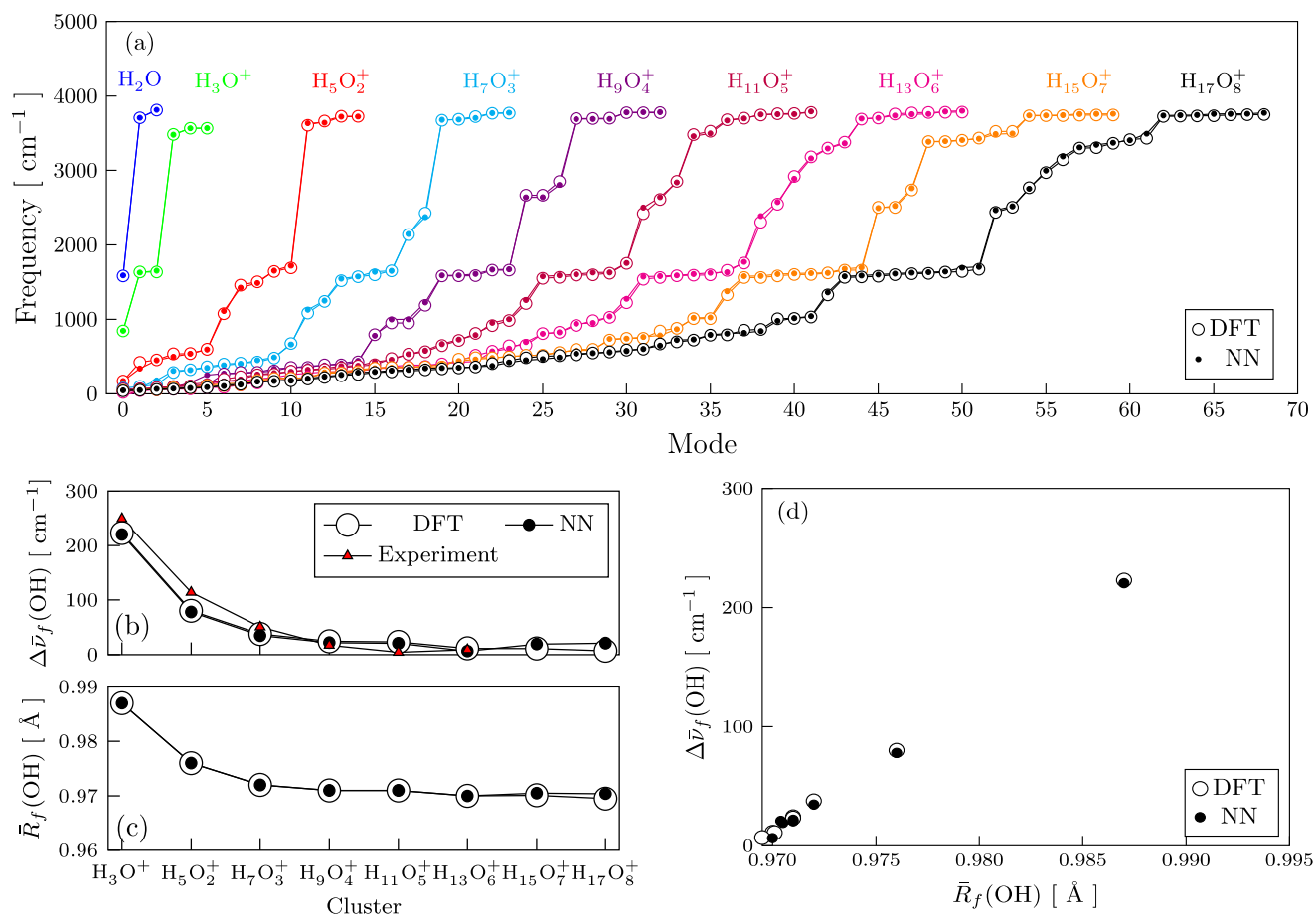


Fig. 4 The neural network (NN)-predicted harmonic frequencies of the optimized structures of the protonated water clusters are in excellent agreement with the density-functional theory (DFT) data for all clusters as can be seen in panel (a). An isolated neutral H_2O molecule has been used as reference and the shifts in the averaged stretch frequencies $\Delta\bar{\nu}_f(\text{OH})$ and averaged free OH bond lengths $\bar{R}_f(\text{OH})$ for the individual clusters are shown in panels (b) and (c), respectively. The experimental shifts of the vibrational frequencies have been taken from Ref. 44 for the neutral water molecule and the hydronium ion and from Ref. 45 for the protonated dimer and larger clusters. Correlation between the shift in the averaged stretch frequencies $\Delta\bar{\nu}_f(\text{OH})$ and the average length of the free OH bonds not participating in hydrogen bonds $\bar{R}_f(\text{OH})$ is seen in panel (d).

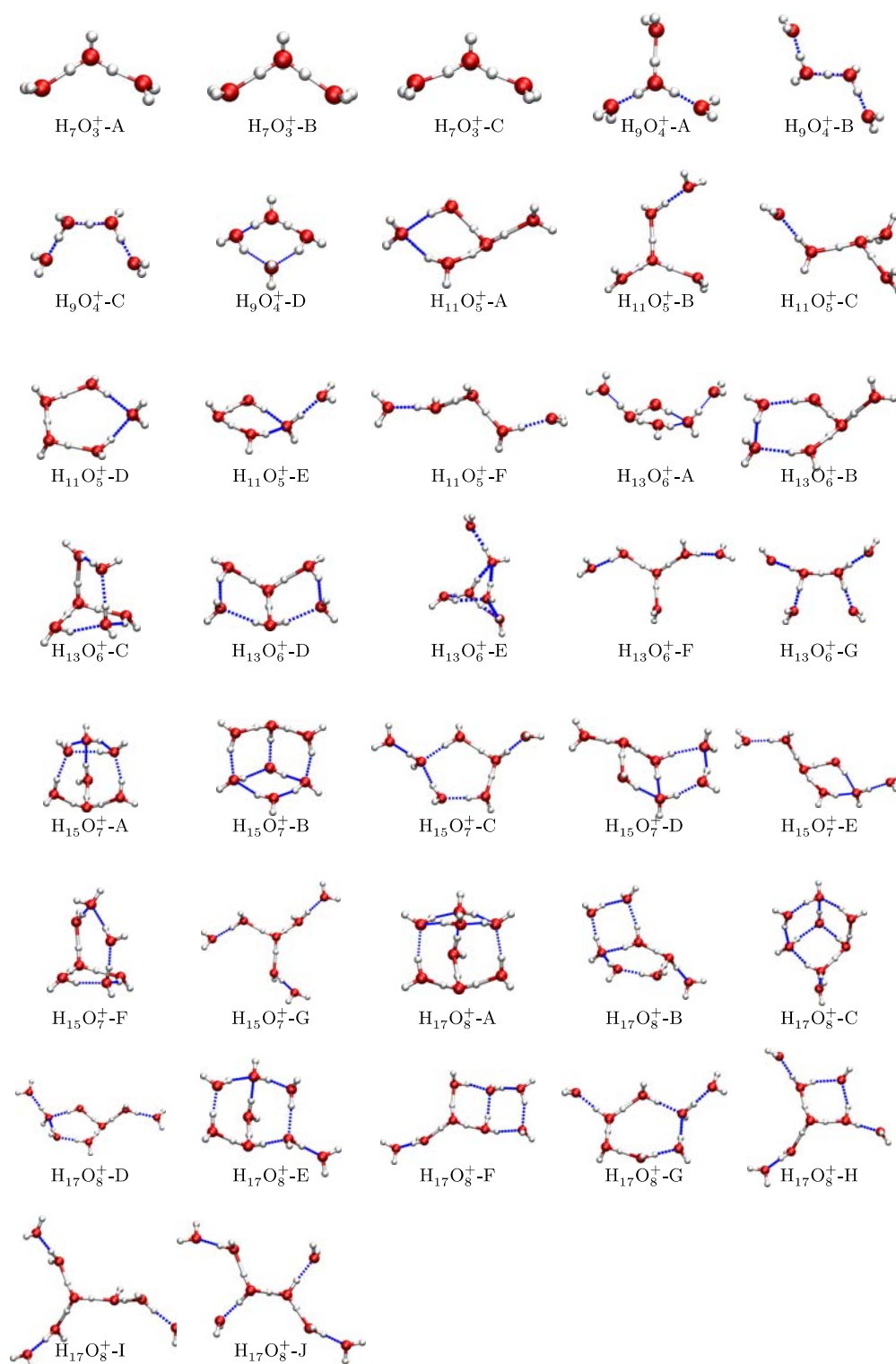


Fig. 5 Local minima of protonated water clusters optimized employing the neural network (NN) potential. The geometries have been confirmed by subsequent density-functional theory (DFT) calculations. In case of the larger clusters, from $H_{13}O_6^+$ onwards, a large number of local minima exists and thus only selected structures are presented for these clusters.

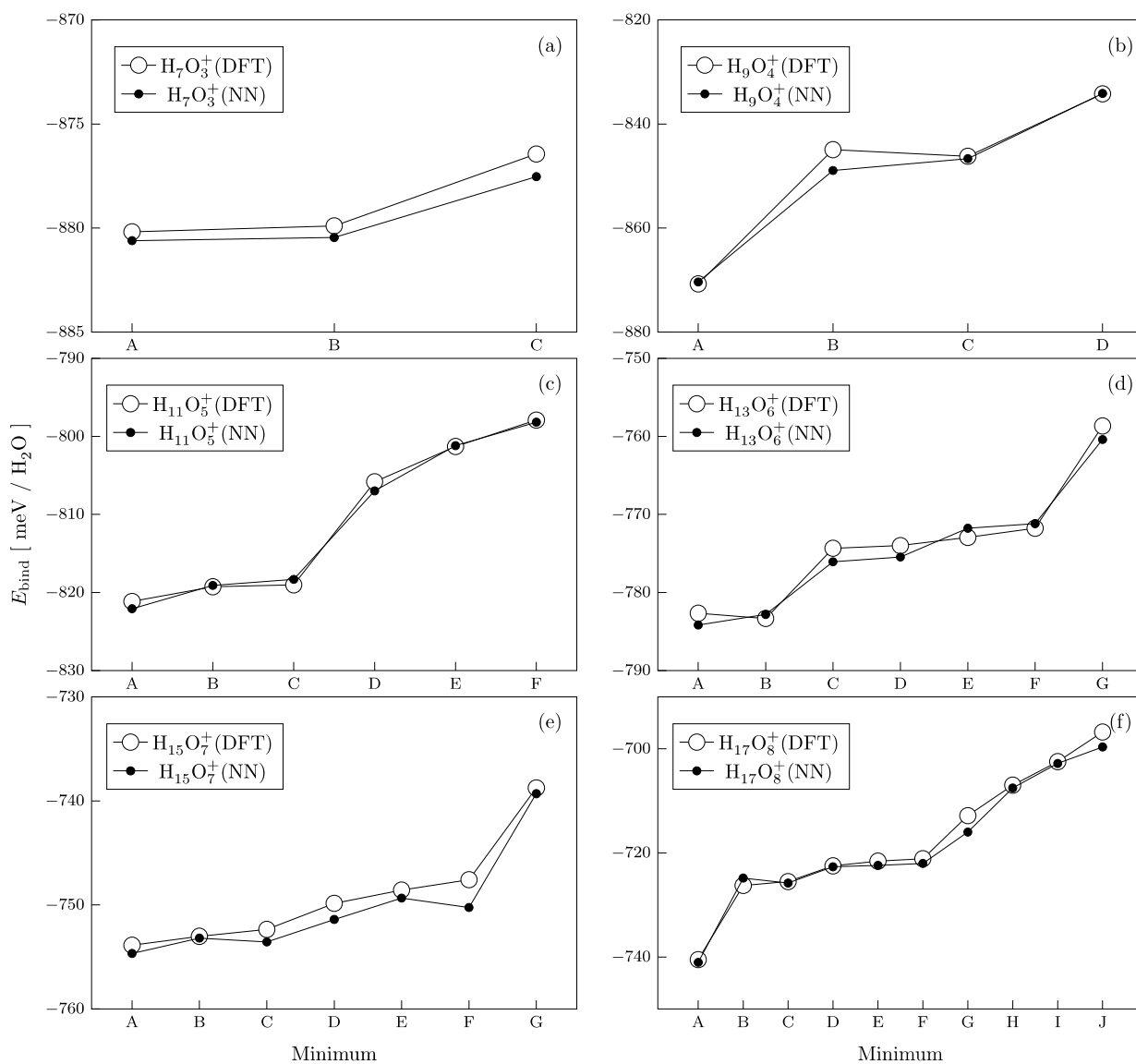


Fig. 6 Binding energies per water monomer for the local minima from the trimer to the octamer structures as shown in Fig. 5. Only in a few cases there are differences in the energetic ordering between density-functional theory (DFT) and the neural network (NN) potential. The relative stabilities of different isomers vary only by a few meV per molecule and the uncertainty of the NN predictions is in the order of the RMSE of the NN potential.

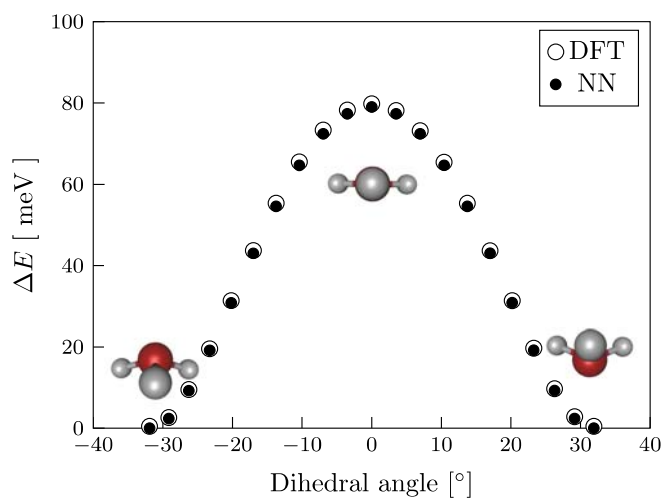


Fig. 7 One-dimensional potential-energy surfaces describing the inversion of the H_3O^+ ion obtained from density-functional theory (DFT) and the neural network (NN) potential. The energy difference ΔE refers to the energy of the optimized structure.

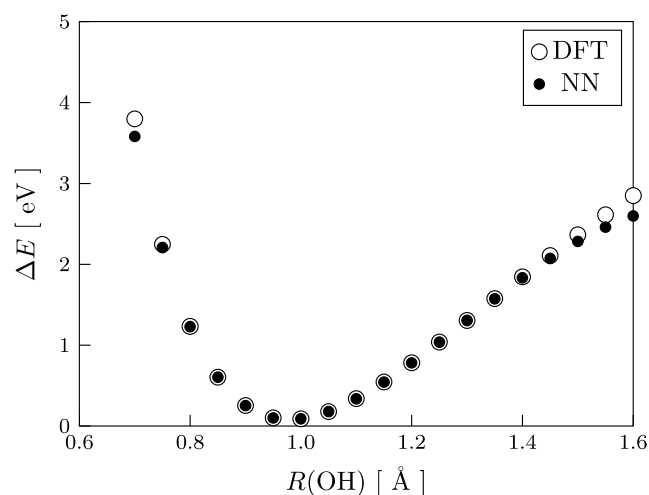


Fig. 8 One-dimensional potential-energy surface describing the detachment of a hydrogen atom/proton from the H_3O^+ ion obtained from density-functional theory (DFT) and the neural network (NN) potential. The constrained OH distance is given by $R(\text{OH})$ and the positions of all other H atoms have been relaxed for each structure. The dotted line indicates the maximum OH bond length that has been used to train the NN potential. Even for larger separations the NN potential is still able to predict energies close to DFT and only beyond 1.4 Å the deviations become larger than the overall error of the NN potential. The equilibrium OH bond length in the hydronium ion corresponding to the minimum of the curve is 0.987 Å.

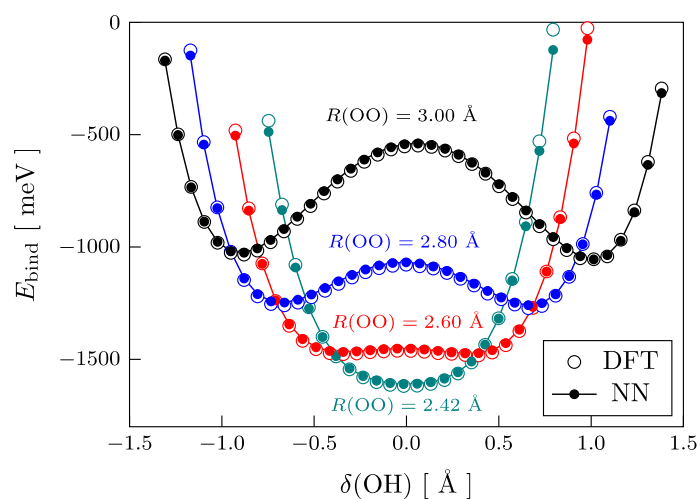


Fig. 9 One-dimensional binding energy curves for a proton transfer within the protonated water dimer for different fixed oxygen-oxygen distances R_{OO} . The binding energies obtained with density-functional theory (DFT) and the neural network (NN) are in very close agreement in all investigated cases. While for a short intermolecular distance there is only one minimum corresponding to a Zundel ion, for larger separations of both water monomers there are two equal minima with asymmetric OH distances.

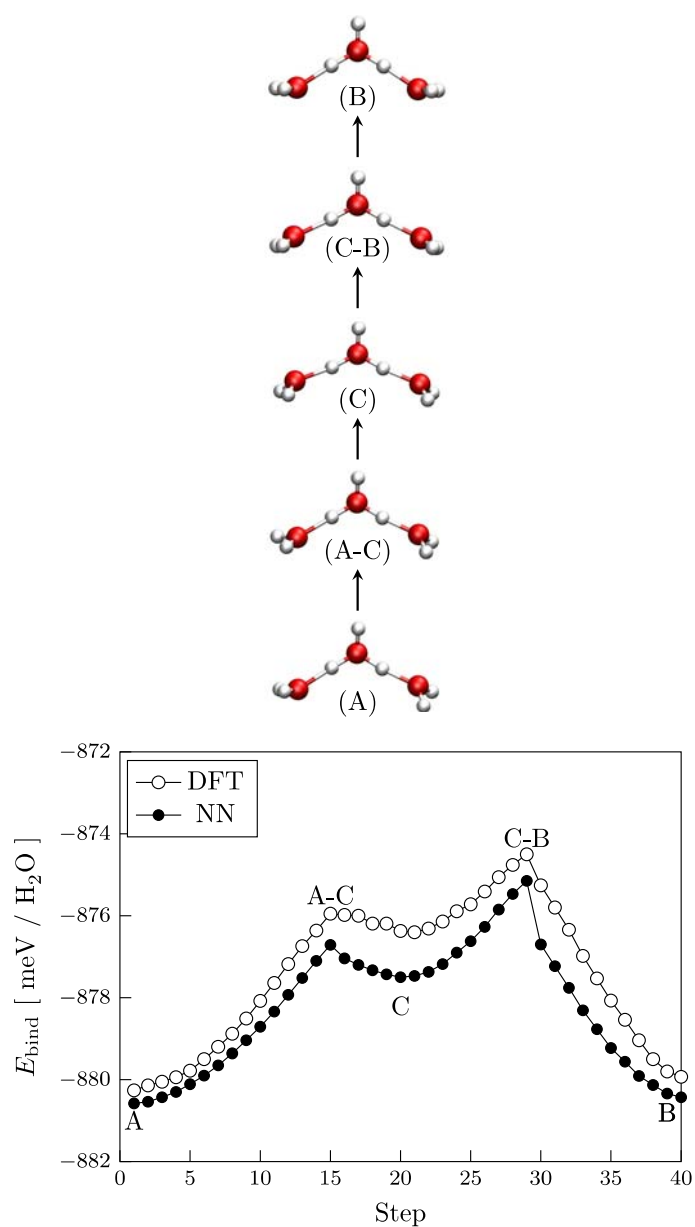


Fig. 10 Structures along the interconversion pathway between minima A, B, and C of the protonated water trimer are shown in (a). Binding energy along the interconversion pathway connecting these three minima are given in (b).

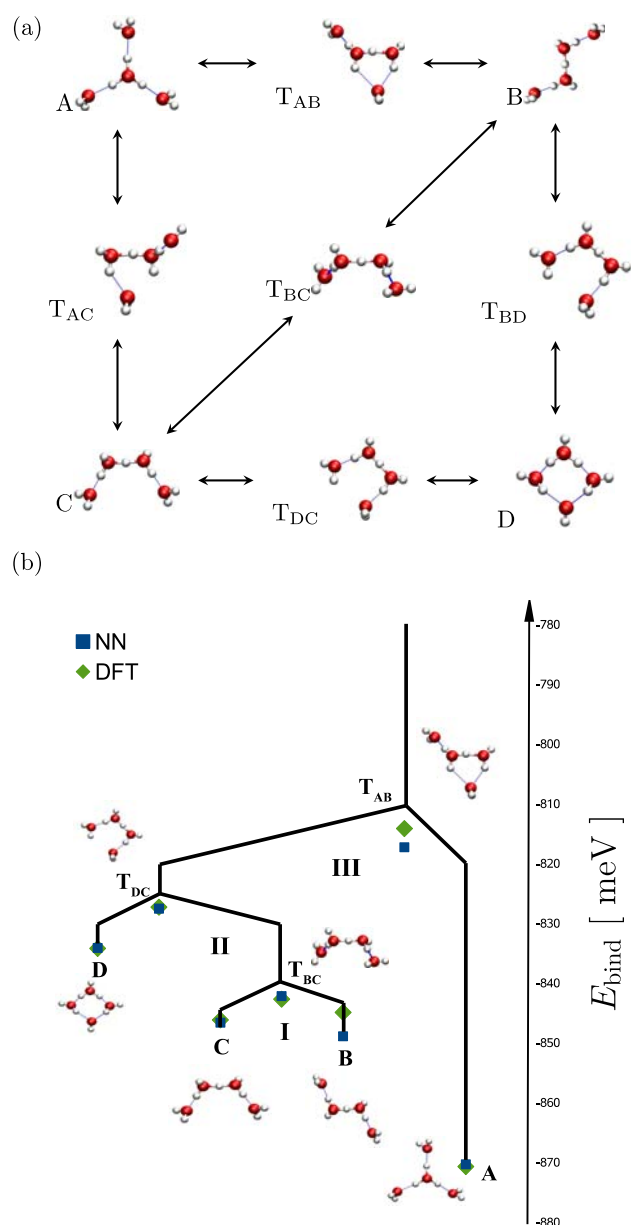


Fig. 11 Pathways connecting the minima A, B, C, and D as well as the intermediate transition states of the protonated water tetramer are shown in (a). The disconnectivity graph is shown in (b). Disconnectivity graph of the potential-energy surface of the protonated water tetramer showing basins I, II, and III. The minima (A, B, C, and D) have been determined using the neural network (NN) potential and density-functional theory (DFT)-based geometry optimizations, and the DFT energies for the transition states (T_{AB}, T_{BC}, T_{DC}) have been obtained by recalculating the NN-derived transition state structures.

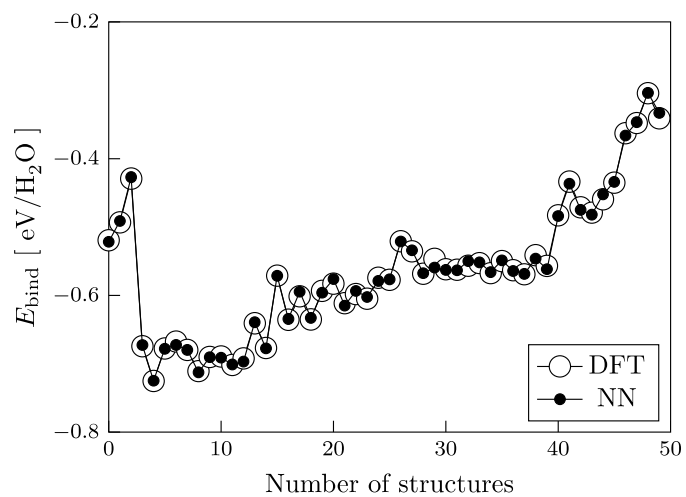


Fig. 12 Comparison of the density-functional theory (DFT) and neural network (NN) energies for a number of non-equilibrium structures of the protonated water heptamer obtained in replica exchange molecular dynamics simulations.

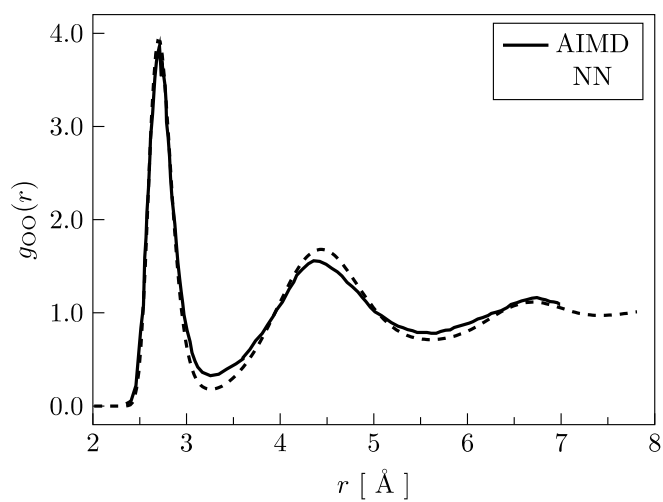


Fig. 13 Oxygen-oxygen radial distribution function g_{OO} of liquid water at 300 K from simulations with the neural network potential (NN) compared to results from ab initio simulations (AIMD, Ref. 88).

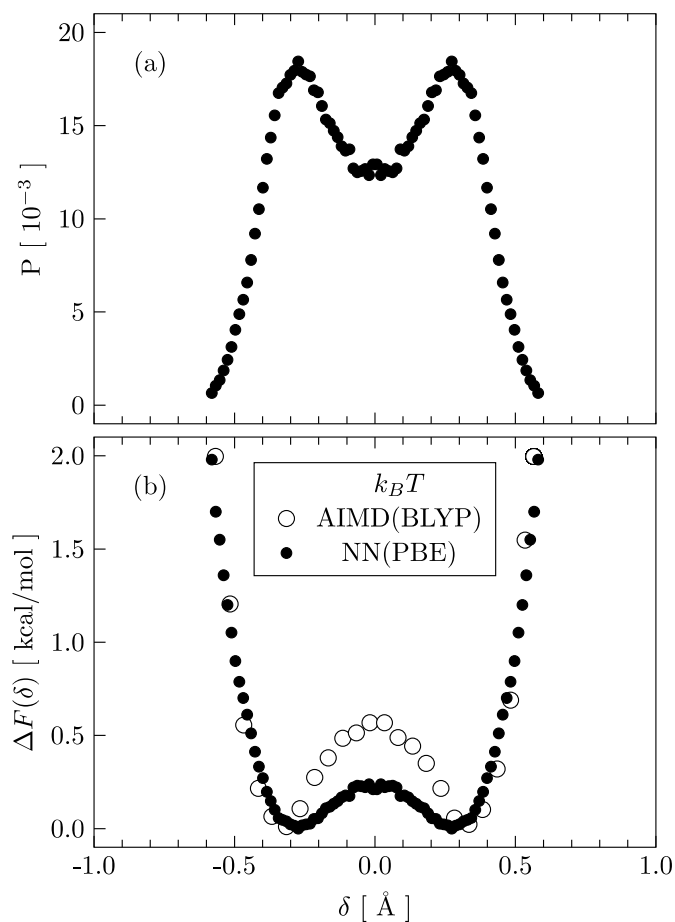


Fig. 14 Distribution function P of the proton transfer coordinate δ of an excess proton in liquid water at 300 K (a) and the corresponding free-energy profile $\Delta F(\delta)$ (b) obtained from neural network (NN) simulations based on the PBE functional. For comparison data from ab initio simulations (AIMD, Ref. 89) using the BLYP functional is shown. The red dashed line indicates the thermal energy $k_B T$ at 300 K.

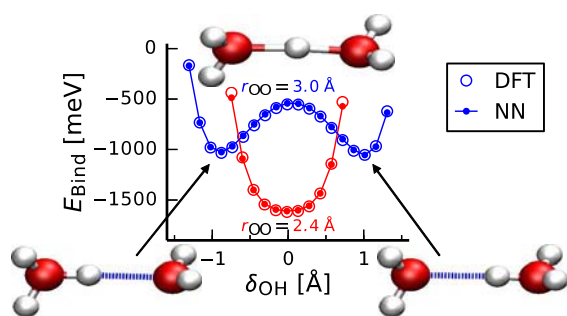


Fig. 15 TOC picture