Supplementary Material to the article “Proton Transfer in Water”

ReaxFF system energy. The key concept behind the ReaxFF potential, which underlies the dynamic creation and breaking of bonds, as well as the calculation of energy and forces between atoms, is bond order. Bond order can be understood as a chemical bond order, but if it were real number. The bond order for each pair of atoms, i and j, is calculated based on the distance between them. After that, an adjustment is made to obtain an adjusted bond order, which is then used in the calculation system's energy. The formulas for calculating bond order are provided in the original article on the potential [1] and are not included here for brevity. The calculated bond order is used to determine whether there is a chemical bond between two atoms. If the value of the bond order between atoms i and j, denoted as $\mathrm{BO}_{i,j}$, is greater than a certain threshold value, it is considered that a bond exists between them. If the value is below the threshold, it means that there is no bond between the two atoms. The threshold value can be adjusted for different pairs of atoms, depending on their types. The presence or absence of a connection between atoms is not required for calculating energy and forces. However, it is essential to study the system of atoms at each point in time, identify molecules within it, and track their movements and changes.

The energy of a system of atoms, modeled using the ReaxFF potential, can be generally written as:

$$E_{\text{system}} = E_{\text{bond}} + E_{\text{over}} + E_{\text{under}} + E_{\text{angular}} + E_{\text{tors}} + \\ + E_{\text{specific}} + E_{\text{vdWaals}} + E_{\text{Coulomb}}, \tag{S1}$$

where each of the terms is the sum of all atoms or pairs/triples/fours of atoms and denotes:

E_{bond} – bond energy. In fact, the total energy of each bond in each molecule.

E_{over} and E_{under} – the energy that regulates the coordination number of each atom. This energy introduces a penalty into the system if some atom has an excessive (over-coordination) or insufficient (under-coordination) number of bonds compared to the optimal number. Accordingly, E_{over} grows rapidly when the coordination number of an atom exceeds the optimal value and quickly becomes zero when the coordination number is lower than the optimum. E_{under}, on the other hand, grows rapidly with an insufficient number of bonds and is practically zero for optimal or excessive numbers.

E_{angular} – valence angles energy. In fact, the total energy of each bond angle in each molecule. The bond angle is defined as the angle between two bonds in the molecule.

E_{tors} – torsion angles energy. In fact, the total energy for each torsional angle between the four atoms in each molecule. This contribution to energy is not significant when modeling water, although it still has some relevance when modeling the H_3O^+ ion and the complex ions formed from H_3O^+ and OH^- ions, such as the Zundel cation $H_5O_2^+$. In addition to the bond order, the energy of a torsional angle also depends on the value of two valence angles, which are multiplied by the sine of these angles. The energy for a torsional angle is zero if the two valence angles in the triplet of atoms that form it are equal to 0 or π.

E_{specific} – this contribution to the energy includes all specific contributions from the system under consideration. In the general formula of ReaxFF potential this term is kept in its original form and is not further broken down. In the original paper on ReaxFF [1] the potential was tuned for hydrocarbons, so the specific energy included two terms: penalties for the stability of double-bonded hydrocarbons and the energy of conjugated systems. The specific contributions that are relevant for water modeling are treated separately.

E_{vdWaals} – Van der Waals interaction energy between all pairs of atoms in a system. This energy contribution does not depend on the bond order between atoms, but only on the distance between them. To calculate the Van der Waals energy, the expression with a near-zero screening factor is used. This means that the standard formula for Van der Waals energy changes smoothly near zero distance, so that at $r = 0$, the energy is constant. This is done to prevent a sharp increase in the energy of interaction between atoms that become close to each other during a molecular dynamic calculation. This means that a huge repulsive force will not appear, and two atoms that are nearby will not fly apart at a high speed, which could lead to the non-physical avalanche-like destruction of the system due to subsequent numerical errors associated with the finite integration step. Even if we exclude the possibility of such destruction, the accidental bringing of atoms close together followed by high-speed scattering is still unphysical. The finite integration step is used, which means that there is a non-zero probability for two atoms to be close together, and this probability increases with an increase in the integration step. This is the problem of probable non-physical behavior in molecular dynamics, which is solved by using screening.

E_{Coulomb} – Coulomb interaction energy between all pairs of atoms in a system. This energy contribution does not depend on the bond order between atoms, but only on the distance between them. The similar expression with a near-zero screening factor is used.

Specific energy contributions that are relevant for the particular system under consideration need to be addressed separately. For a system of water molecules, the specific energy E_{spec} can be expressed as follows [2, 3]:

$$E_{\text{specific}} = E_{\text{lone pairs}} + E_{\text{Hydrogen bonds}}, \tag{S2}$$

where each of the terms is the sum of all atoms or pairs of atoms and denotes:

$E_{\text{lone pairs}}$ – the energy of atom's valence electrons pairs (lone pairs) that do not participate in the bonds formation. In water, oxygen has two lone pairs and hydrogen has none. This energy depends on the difference between the ideal and actual number of lone pairs. The ideal number of pairs remains constant, while the actual number is calculated during the simulation process.

$E_{\text{Hydrogen bonds}}$ – hydrogen bonds energy. Hydrogen bonds exist between the hydrogen atom of one water molecule or ion and the oxygen atom of another water molecule or ion. The ReaxFF potential model considers this energy as the sum of the atoms triples energies of the form $O_1H_1O_2$, where O_1 и H_1 – is an oxygen and hydrogen of one molecule, and O_2 – is an oxygen of another molecule. The energy depends on the bond order between the oxygen and hydrogen atoms in the first molecule, the distance between the hydrogen in the first molecule and the oxygen in the second molecule, as well as the sine of the angle $O_1H_1O_2$. If this angle is zero, the hydrogen bond energy is 0 or π.

The exact formulas for all the terms mentioned above are not provided due to the complexity of their mathematical expressions.

Algorithms for tracking molecules/ions. The issue of tracking water molecules and ions H_3O^+ and OH^- during modeling requires special attention. Due to the dynamic formation of bonds, molecules at each step of integration may consist of different atoms. This is because at some step, there may have been an exchange of atoms or a proton jump, which can change the composition of a molecule or ion. After this change, even if the chemical formula remains the same, the molecule or ion can no longer be considered the same and should be monitored separately. These features of the ReaxFF potential impose additional requirements on postprocessing atomic trajectories.

The LAMMPS [4] package generates two files at each integration step: a file containing the coordinates of each atom and its type (oxygen or hydrogen), and a file that lists the existing bonds between atoms in the "atom 1 – atom 2" format. Using these files, it is possible to reconstruct all the molecules that were present in the system at each integration step, similar to the task of reconstructing the connectivity components of a graph from a list of edges. The descriptions of our algorithms for tracking molecules and ions in a system and calculating diffusion coefficients are provided below.

An algorithm for calculating the self-diffusion coefficient of water molecules. The initial moment of time is chosen, and the first output of the LAMMPS calculation is made. At this point, all the water molecules that are currently in the system are stored. In the next step, all the molecules are saved too, and so on. Next, the molecules that existed in the system at the previous step and the current step are compared. If a molecule existed in the same form (containing exactly the same atoms with the same identifiers) at both steps, then its new position in the trajectory is recorded. If at the current step, there is a molecule that was not present at the previous step, it is considered to be a new molecule. For this new molecule, this current time step is chosen as the starting point in time, and its trajectory is preserved from there, meaning that any deviation in space is measured relative to its position at this current time. If at the current step, there is no molecule that was present at the previous step in the integration process, where the coordinates were generated, then we can assume that this molecule has broken down or transformed into the new one. Therefore, its trajectory is considered completed and it will not continue, even if, after some time, the molecule may reassemble in the same composition again, with exactly the same atoms. At the point of reappearance, this molecule is considered to be a new one.

After the entire calculation process has been completed, all active molecules at the time of the last coordinate file will move to the "completed" state. The trajectories of all these molecules are then collected for further processing. The calculation algorithm is specifically designed for water molecules, so only the trajectories of those with the formula H_2O are remembered. All other formations, including ions, are not tracked.

The trajectories can have different lengths, as different molecules may exist in the system unchanged at different times. Trajectories that are too short are excluded. Sufficiently long trajectories are truncated to the same length so that the time dependence of the squared displacements for all trajectories can be averaged. If the calculations are long enough, several segments of the same length can be selected from long trajectories, but with the segments separated in time by a distance that ensures the independence of the molecules' positions in different segments. After averaging, the time-dependence of the mean squared displacement of molecules is obtained, and the coefficient of self-diffusion of water molecules is calculated from the slope of the line.

An algorithm for calculating the diffusion coefficient of ions in water. The tracking algorithm for ions is slightly different from that used for tracking water molecules, although it follows similar principles.

For the sake of simplicity in calculations and to model an infinitely small concentration of ions in water, we have considered only one type of ion, either H_3O^+ or OH^-, in all calculations in this work. Therefore, it is necessary to track the trajectory of this particular ion. Due to the fact that ion transport in water occurs not only through the movement of the ion as a whole, but also through the transfer of the ion from one molecule to another, the principle of following only a molecule with the same composition is not applicable in this case. However, due to the fact that there is only one ion, it is sufficient to monitor only the formation that is not H_2O.

More specifically, at each step, the cation of the form $H_3O\left(H_2O\right)_n$ or $HO\left(H_2O\right)_n$, was searched for in the system, when considering, respectively, the ions H_3O^+ or OH^-. Here n is the number of water molecules bonded to the ion. At each step, the center of mass of the cation is remembered. Due to the relatively small size of the system under consideration, the long-lived ions H_3O^+ and OH^- did not form from two water molecules. Therefore, the ion present in the system could not be confused with others, and its trajectory could be accurately tracked.

After processing all the coordinate files generated during the LAMMPS calculation, a long ion trajectory was obtained.

After processing all the coordinate files output during the LAMMPS calculation, one long ion trajectory was obtained. The number of trajectory points equals the number of output files containing the coordinates of atoms. Further, several segments of the same length and with independent positions of molecules were selected from this trajectory. Based on these, the average time dependence of the squared ion displacements was calculated. And further, the diffusion coefficient of the corresponding ions, H_3O^+ or OH^-, was determined.

The algorithm is designed to work with only one ion in the system, and will not function for cases with two or more ions. This is not inconvenient, as exactly one ion was modeled in the system. Implementing an algorithm that can handle multiple ions simultaneously becomes more challenging, as it requires developing a logic for handling multiple ions interaction. As long as the ions do not approach each other, they can be tracked using the methods described above. However, the task becomes more complex when these ions approach each other and combine or swap atoms. In practice, when simulating several ions in a system, we have noticed that these ions, along with several water molecules, often combine into one larger ion. This combination can then be further broken down into another composition. This requires a more sophisticated approach to handling these situations and accurately measuring the movement of each ion.

All the described algorithms have been implemented in the Python programming language.

T, K	P, atm	ρ, g/cm^3
273.16	0.00604	0.99979
300	1	0.997
400	9.8692327 (= 1 MPa)	0.937871
500	100	0.83815
550	100	0.76201

Table S1. The values of temperature, pressure, and density of the system used in liquid water modeling.

Modeling system parameters. The values of temperature, pressure, and density of the liquid water systems simulated in this work are presented in Table S1.

Diffusion coefficients for OH^- under normal conditions. Diffusion coefficients for ion OH^- in liquid water under normal conditions are shown in Fig. S1. Similarly to H_3O^+, it is not possible to detect dimensional effects with the available accuracy.

Pressure dependence of diffusion coefficients for OH^-. Fig. S2 shows the pressure dependencies of OH^- ion diffusion coefficients for four isotherms. Similar to the H_3O^+ ion, no significant pressure dependence is observed, suggesting that the obtained temperature dependence of the ion diffusion coefficient remains valid across a wide range of pressures.

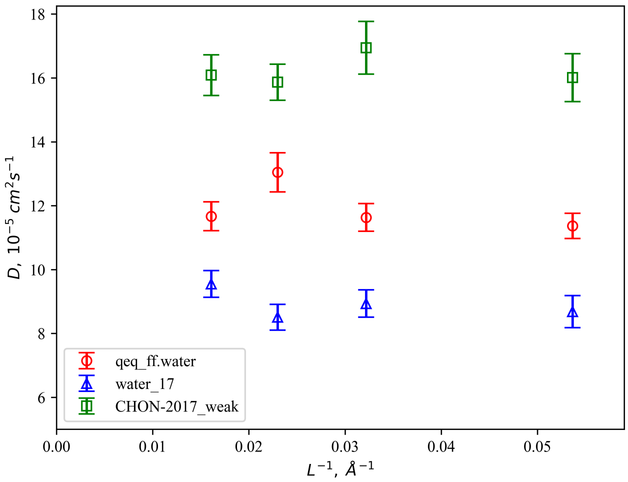

Fig. S1. OH^- diffusion coefficients in water for the three sets of ReaxFF potential coefficient and four system sizes depending on the inverse system's linear size. The error crosses indicate a value of one standard deviation. Normal conditions: $P = 1$ atm, $T = 300$ K, $\rho = 0.997$ g/cm^3. 1 ion in the system; the number of water molecules: 215, 999, 2743 and 7999, respectively, from the smallest system to the largest.

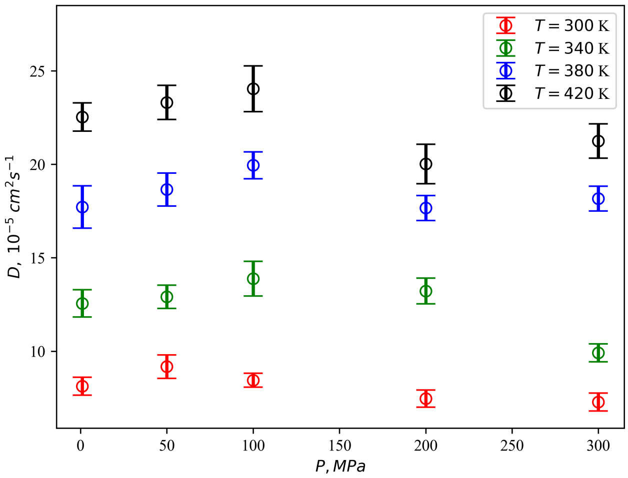

Fig. S2. Diffusion coefficients for OH^- ions as a function of system pressure for four isotherms. Error crosses indicate one standard deviation. The system contains 999 water molecules and 1 corresponding ion.

1. A. C. T. van Duin, S. Dasgupta, F. Lorant, W. A. Goddard III, J. Phys. Chem. A., **105**, 9396 (2001).
2. A. C. T. van Duin, C. Zou, K. Joshi, V. S. Bryantsev, W. A. Goddard, Comp. Catal., **14**, 223 (2013).
3. W. Zhang, A. C. T. van Duin, J. Phys. Chem. B., **121**, 6021 (2017).
4. S. Plimpton, J. Comp. Phys., **117** (1), 1 (1995).