Using Machine Learning to Greatly Accelerate Path Integral Ab Initio Molecular Dynamics

*Chenghan Li and Gregory A. Voth**

Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, 60637, USA

KEYWORDS. molecular dynamics, path integral, nuclear quantum effects, machine learning, excess proton, solvation

ABSTRACT *Ab initio* molecular dynamics (AIMD) has become one of the most popular and robust approaches for modeling complicated chemical, liquid, and material systems. However, the formidable computational cost often limits its widespread application in simulations of the largest-scale systems. The situation becomes even more severe in cases where the hydrogen nuclei may be better described as quantized particles using a path integral representation. Here, we present a computational approach that combines machine learning with recent advances in path integral contraction schemes, and we achieve a two orders-of-magnitude acceleration over direct path integral AIMD simulation while at the same time maintaining its accuracy.

In molecular systems, due to the large masses of the nuclei, electrons are often considered to move infinitely fast compared to the nuclear degrees of freedom, and as a result, the nuclei are thought to evolve on a potential energy surface (PES) created by the electronic ground-state according to the Born-Oppenheimer approximation. If the nuclei are heavy enough to be treated as classical particles, their motion will be well-described by Newtonian dynamics on the electronic ground-state PES. In the *ab initio* molecular dynamics (AIMD) approach[1-3] the electronic structure is solved "on the fly" as the nuclei move, so there is no need to first perform a sometimes challenging and time-consuming parameterization of the system's interatomic and intermolecular interactions, provided the underlying electronic structure in the AIMD algorithm is accurate enough.

However, in some scenarios, nuclear quantum effects (NQEs) can be important for an accurate description of the system of interest, which is frequently associated with light atoms (such as protons) and/or at low temperatures. For example, NQEs are found to be important in liquid water,[4-5] proton solvation and transfers,[6-8] and various enzymatic processes.[9] Feynman's imaginary time path integral approach[10] has been adopted widely to compute the quantum statistics and dynamics by mapping the quantum sampling into the sampling problem of a system consisting of fictitious particles or "beads" (replicas) of the corresponding classical nuclear system.[5, 11-12] For an *N*-particle system governed by the Hamiltonian

$$\hat{H} = \hat{T} + \hat{V} = \frac{1}{2}\hat{\boldsymbol{p}}^{\mathrm{T}}\mathbf{M}^{-1}\hat{\boldsymbol{p}} + \hat{V}(\boldsymbol{r}) \qquad 1$$

the thermal density matrix in the real-space representation can be expressed by the path integral formalism

$$\rho_{\boldsymbol{r}\boldsymbol{r}'} = \frac{1}{Z}\langle \boldsymbol{r}|e^{-\beta\widehat{H}}|\boldsymbol{r}'\rangle$$

$$\propto \lim_{P\to\infty} \int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_{P-1} \exp\left\{-\beta \sum_{j=1}^{P} \frac{P}{2\beta^2\hbar^2}\left(\boldsymbol{r}_j - \boldsymbol{r}_{j-1}\right)^{\mathrm{T}}\mathbf{M}\left(\boldsymbol{r}_j - \boldsymbol{r}_{j-1}\right)\right. \quad (2)$$

$$\left. + \frac{V(\boldsymbol{r}_j) + V(\boldsymbol{r}_{j-1})}{2}\right\}\Bigg|_{\boldsymbol{r}_0=\boldsymbol{r}}^{\boldsymbol{r}_P=\boldsymbol{r}'}$$

where $\mathbf{M}$ is a $3N \times 3N$ diagonal matrix of the particle masses, and $\boldsymbol{r}$ is the $3N$-dimensional vector representing the atomic coordinates. The energy and force calculations must be performed for all such replicas ($\{\boldsymbol{r}_j\}$) of the system, which typically increases the computational cost by at least an order of magnitude depending on the number of beads (P) required to converge the path integral. Considering the already formidable computational cost in the underlying electronic structure calculation with AIMD, modeling NQEs on an *ab initio* PES has not been typically viable for many realistic systems until recent developments involving some computational approximations.[12] Aspects of this body of work include the ring-polymer contraction (RPC),[13] the ring-polymer interpolation,[14] and the combined path integral and generalized Langevin equation (PI+GLE) approach.[15]

In RPC, the system PES is divided into two terms, a dominant term and a slowly varying term. This partitioning in AIMD can be accomplished by employing a lower level quantum approach as the reference potential, and its deviation from the original high-level PES is regarded as a perturbation term, such that

$$V_{\mathrm{high}}(\boldsymbol{r}) = V_{\mathrm{low}}(\boldsymbol{r}) + V_{\mathrm{diff}}(\boldsymbol{r}) \quad (3)$$

The reference potential V_{low} is evaluated for all the P beads $\{\boldsymbol{r}_j\}$ while the deviation term V_{diff} is only evaluated on a "contracted" ring polymer with $P' < P$. In the special case when $P' = 1$, i.e., the centroid contraction, V_{diff} is only evaluated on the centroid:

$$\bar{V}_{\mathrm{diff}} = V_{\mathrm{high}}(\boldsymbol{r}_c) - V_{\mathrm{low}}(\boldsymbol{r}_c) \qquad 4$$

where $\boldsymbol{r}_c$ is the imaginary time path centroid defined as $\boldsymbol{r}_c = \frac{1}{P}\sum_{j=1}^{P} \boldsymbol{r}_j$, the zero-frequency normal mode of the ring polymer.[10] Then, the potential at each path integral bead is simply approximated by

$$V_{\mathrm{high}}(\boldsymbol{r}_j) \approx V_{\mathrm{low}}(\boldsymbol{r}_j) + \bar{V}_{\mathrm{diff}} \qquad 5$$

We note that the path centroid variables of the particles are the most classical-like variables in a quantized imaginary time (equilibrium) system, as has been elaborated in considerable detail previously.[16] The particle centroids are therefore arguably the most natural variables to choose for a path integral contraction scheme. If the computational cost of V_{low} is negligible, then RPC speeds up path integral simulations by a factor of P/P' because the exact potential is only evaluated on the contracted ring polymer. The partitioning of the potential (eq 3) can also be useful for accelerating classical AIMD via a multiple time scale (MTS) integration scheme,[17] where V_{high} is only evaluated every M steps (outer timestep) but V_{low} is evaluated every single step (inner timestep). Similar to RPC, if V_{low} can be evaluated much more efficiently, the MTS will accelerate the AIMD by a factor of M. In practice, the MTS and RPC can be both employed to accelerate the quantum simulation as formulated in refs.[18-20]

One commonly used low-level electronic structure method in MTS/RPC is the self-consistent-charge density-functional tight-binding (SCC-DFTB) method.[21-22] However, the SCC-DFTB approach is largely semi-empirical and known to have limitations, e.g., in describing the solvation structure and dynamics of liquid water and hydrated excess protons.[23-25] In principle, the accuracy of the MTS/RPC approximation is subject to the accuracy of the underlying reference potential, and the efficiency gain is also strongly influenced by the computational efficiency of the low-level method. Hence, it is of considerable value to find an accurate enough reference potential which can also be evaluated efficiently numerically.

The recent advances in machine learning[26-27] (ML) have opened up new opportunities to enable efficient simulations of complicated systems by approximating the many-body quantum PES by a flexible *ansatz*, usually an artificial neural network (ANN). However, it is usually impossible to learn perfectly the *ab initio* PES by the network training, due to the many-body nature of molecular systems and the finite expressive capabilities of an ANN. The situation becomes more challenging for a heterogeneous system because a data-driven approach is heavily affected by the amount of data available, and it remains an open question as to how to efficiently generate a high-quality training set that adequately samples the interactions associated with each species and components in the system, especially for dilute ones.

It is thus desirable to have a "monitor" that could keep verifying whether the ML potential deviates from the exact *ab initio* potential during the production run and if it does, correcting it in some way. Although this may be achieved by employing an online learning approach that utilizes the newly sampled data to refine the underlying ML model on the fly, there is no guarantee that this combined learning-sampling procedure will eventually converge to a determined solution due to the stochastic nature of the network training and MD sampling. Here, we instead propose to use

the MTS/RPC framework to achieve such a "monitor" task by employing the ML model as the reference potential in MTS/RPC. The ML potential in MTS/RPC is not used to directly propagate the system; instead it is constantly "corrected" by regular evaluations of the *ab initio* PES, alleviating the concern about the learned potential accuracy. In addition, as we will discuss below, a ML potential can indeed be an excellent choice of the reference potential in MTS/RPC to improve its computational efficiency and accuracy.

We note that the assumption made in RPC is that the V_{diff} part of the exact potential V_{high} is a slowly varying function with respect to atomic positions, and the approximation becomes exact if V_{diff} has no position-dependency at all. Hence, the gradient of V_{diff} with respect to atomic coordinates is ideally close to zero ($\boldsymbol{\nabla} V_{\mathrm{diff}} \to \mathbf{0}$). By noting eq 3 and that the negative potential gradient defines the force, an ideal reference potential V_{low} would minimize the force residual

$$\chi^2_{\mathrm{RPC}} = \langle |\boldsymbol{\nabla} V_{\mathrm{diff}}|^2 \rangle = \left\langle \left| \boldsymbol{F}_{high} - \boldsymbol{F}_{low} \right|^2 \right\rangle \qquad 6$$

where brackets represent some sort of ensemble average, which is usually performed by the average over the training set. In MTS, the "slowly varying" means the potential V_{diff} does not change significantly in real-time, measured by the time derivative of the potential

$$\chi^2_{\mathrm{MTS}} = \left\langle \left| \frac{\mathrm{d}V_{\mathrm{diff}}}{\mathrm{d}t} \right|^2 \right\rangle = \langle |\boldsymbol{\nabla} V_{\mathrm{diff}} \cdot \dot{\boldsymbol{r}}|^2 \rangle \qquad 7$$

which is also modulated by the V_{diff} gradient, and can thus be minimized if χ^2_{RPC} is minimized. Equations 6 and 7 suggest that a ML potential trained using the force residual as the loss function can thus serve as an optimal low-level potential in MTS/RPC approach.

Herein, we benchmarked the ML-MTS/RPC approach for a heterogeneous system which is consisted of one excess proton and 128 water molecules described by the Becke-Lee-Yang-Parr exchange-correlation functional[28-29] with D3 dispersion correction[30] and the experiment directed simulation (EDS) correction parametrized for aqueous systems with an excess proton.[31-32] The details of the simulations are given in the Supporting Information (SI), and the high-level, low-level, and difference force are plotted in Figure S1.

In Figure 1, we show the O-O, O-H and H-H radial distribution functions (RDFs) of the simulated system with both classical and quantum nuclei. Since the excess proton (H^+) is dilute in the system, these RDFs basically represent the pure water structures. Because the ANN training set contains a large number of samples for water-water interactions, it is expected that the data-driven ML approach would be able to learn the water potential with high fidelity. It is indeed what we observed that the ML potential itself already represents a good description of the water-water interactions, as seen by the good agreement of the RDFs between the ML ones and the reference EDS corrected BLYP-D3 (EDS-BLYP-D3) AIMD ones (red vs. blue in Figure 1A,C,E). Since the ML potential itself is sufficiently accurate for the water structure, it is understandable that a classical AIMD simulation using the ML potential as the MTS reference reproduces the EDS-BLYP-D3 water RDFs (green curves vs. blue curves in Figure 1). Interestingly, although the SCC-DFTB approach alone was found to underestimate the second-shell water structures, using it in MTS does present a problem here in the water RDFs (purple vs. blue in Figure 1). This is attributed to the evaluations of the *ab initio* energy and force every several time steps, which serve as a "monitor" that corrects on-the-fly the underlying low-level force.

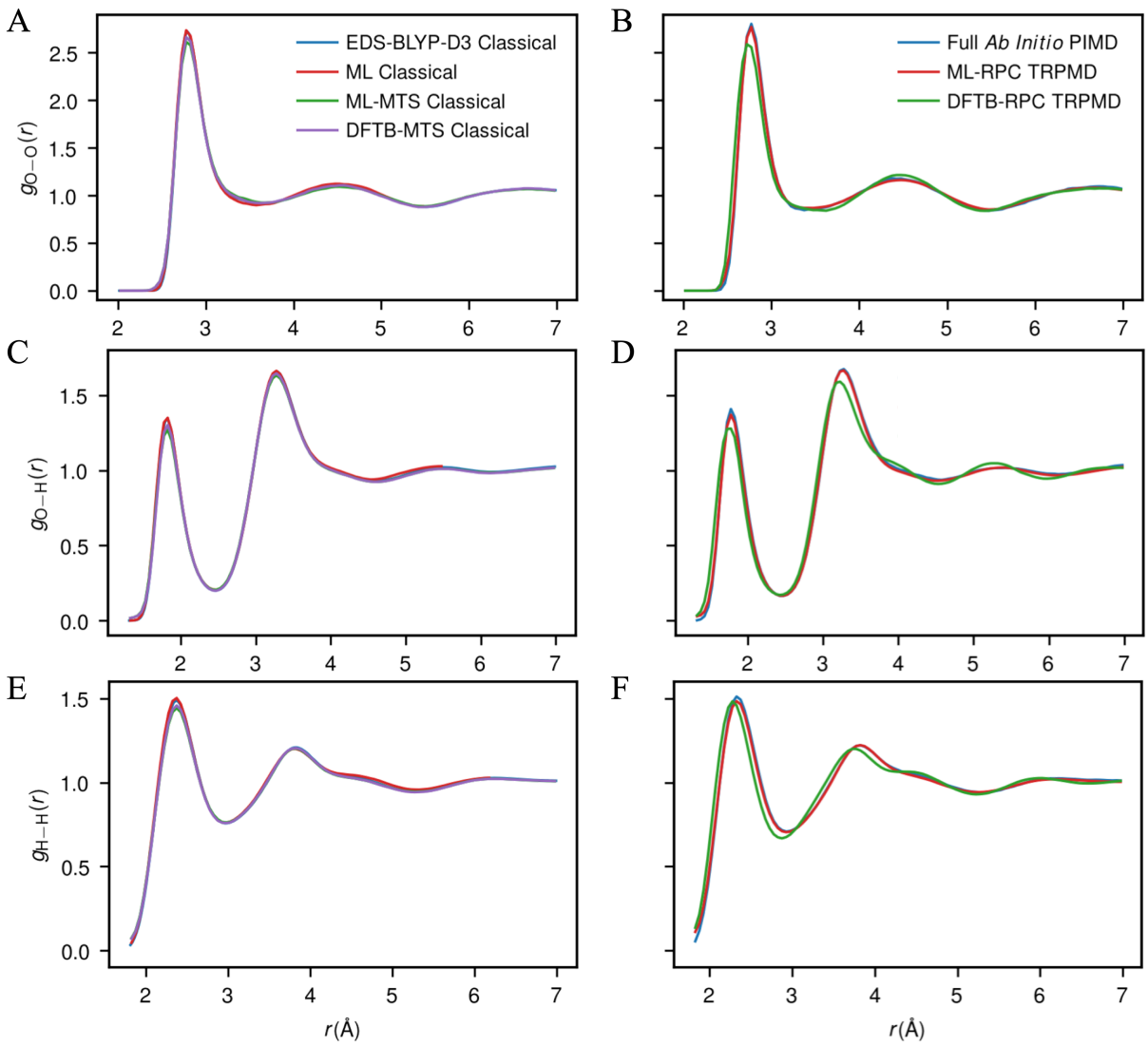


Figure 1. Water solvation structures in terms of the O-O RDF (A & B), O-H RDF (C & D), and H-H RDF (E & F) in classical nuclear AIMD simulations (A, C & E) and quantum simulations (B, D & F). In A, C and E, the classical AIMD using EDS-BLYP-D3 is shown in blue, the simulation directly driven by the ML potential is shown in red, the AIMD using the ML potential as MTS reference is in green, and the AIMD using the SCC-DFTB as MTS reference is in green. In B, D and F, the full *ab initio* PIMD is shown in blue, the TRPMD using ML as RPC reference is in red, and SCC-DFTB as RPC reference in green. The quantum RDFs are reported here as the path centroid results, while the "real" RDFs computed from the averages over the path integral beads are provided in the SI. The statistical errors in the RDFs are all within 0.02.

In **Figure 1**B, D, F, we show the RDFs in quantum simulations. The full *ab initio* PIMD which evaluates the energy and force explicitly on every path integral bead is used as the reference. Employing the ML-RPC approach in the thermostatted ring polymer molecular dynamics (TRPMD)[33] again reproduces the RDFs of the exact quantum results (red vs. blue in **Figure 1**B,D,F). However, while SCC-DFTB performs well as a reference potential in classical simulations, it causes some minor but noticeable variations from the reference quantum results when used in RPC TRPMD. This shows that, while the MTS/RPC approach allows for some tolerance in the underlying reference potential, the accuracy of such an approximation is still dependent on the quality of the reference potential as one might expect. The RDFs computed from the average over the path integral beads (Figure S2) are fully consistent with the discussion here.

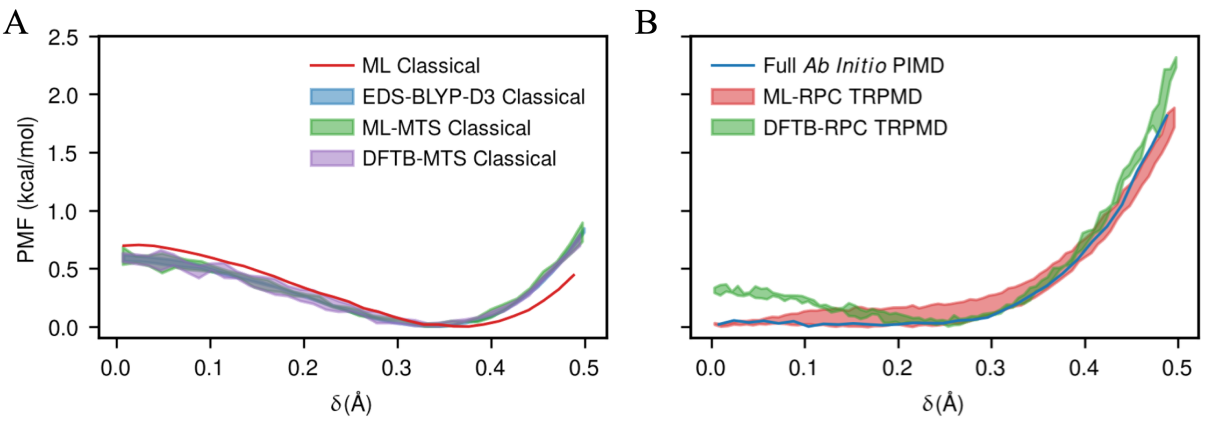


Figure 2. Hydrated excess proton solvation structure measured by the potential of mean force of the proton sharing coordinate (δ) (eq 8). By construction, the lower value in δ represents configurations where the excess proton is more evenly shared. Statistical errors estimated from standard deviations of independent runs are shown as intervals around the curves. (A) The classical EDS-BLYP-D3 AIMD is in blue, the classical simulation driven by the ML potential is in red, the AIMD that uses the ML potential in MTS is in green, and the AIMD using SCC-DFTB in MTS is in purple. (B) The full *ab initio* PIMD reference is shown in blue, the TRPMD using ML potential as RPC reference is in red, the TRPMD using SCC-DFTB as the reference is in green. The quantum PMFs are computed from the centroid trajectories, while the PMFs computed from the averages over the beads are provided in the SI.

In **Figure 2**, we show the potential of mean force (PMF) as a function of proton sharing coordinate δ for the excess proton between two water molecules. The δ coordinate is defined as the difference in O-H distances of a shared proton,

$$\delta = r_{\mathrm{OH}^*} - r_{\mathrm{O}^*\mathrm{H}^*} \quad (8)$$

where O^* represents the most hydronium-like oxygen, which is defined as the oxygen with the closest three hydrogen atoms, the O represents the hydrogen bond acceptor oxygen of H^*, and H^* is one of the three bonded protons in the hydronium that maximizes the δ value. The PMF of δ thus provides a measure of the excess proton solvation structure. In classical simulations (**Figure 2**A), we found the direct use of ML (red curve) will drive the MD results into some subtle deviations from what the model was supposed to learn (blue curve). As was discussed earlier, this is not surprising because there is only one excess proton in the system (vs. 128 waters) and thus the sampling of the hydrated H^+ interactions is much less optimal than the sampling of the many water interactions. As the result, it is challenging for the ML to "learn" perfectly those excess proton interactions given the limited information encoded in the ANN training data. We anticipate this issue to be more pronounced for even more heterogeneous systems with multiple elements and components, which further prevents the application of a ML potential directly to those systems. However, if the ML potential is used only as the reference potential in MTS, the correct proton sharing free energy can be reproduced (**Figure 2**A, green). And, despite the fact that the SCC-DFTB used in this work may predict some abnormal proton solvation structures, such as the second shell water encroaching into the hydrated proton first solvation shell,[25] the use of SCC-DFTB in MTS is still able to predict the correct δ PMF (**Figure 2**A purple vs. blue). This again highlights the tolerance of the MTS approach to the underlying reference potential, and is an

encouraging result for the application of an imperfect ML potential in the RPC approach for quantum path integral AIMD (PI-AIMD) simulations.

In **Figure 2**B, we show the excess proton sharing PMF in quantum simulations. In contrast to the water RDFs (**Figure 1**), the NQEs play a non-negligible role and almost flatten the proton transfer barrier along δ. Although the ML potential shows some deficiency in describing the excess proton solvation structures (**Figure 2**A red), the ML-RPC approach is able to reproduce the δ PMF of the full *ab initio* PIMD result within statistical error. By contrast, the SCC-DFTB-RPC approach lacks sufficient accuracy for capturing the barrierless feature in the small-δ region, but predicts a barrier height between the classical value and the quantum value. We note that this difference is not as pronounced in the δ PMFs computed from averaging over the path integral beads (Figure S3 of the SI), which is consistent with the fact that the DFTB-RPC approach is able to predict the proton sharing structure in the gas-phase Zundel ($H_5O_2^+$) cation as reported in ref[13]. However, the deviation shown in the centroid distribution (Figure 2B) does reveal a certain discrepancy of the DFTB-RPC ensemble with the full *ab initio* PIMD ensemble. It is worth noting that the path centroid distribution provides the most "classical-like" underlying behavior of a quantum system at equilibrium, so the quantum delocalization seen in the full bead averaged result seems to mask these core differences between the DFTB-RPC and ML-RPC outcomes. These contrasting results in the present work for a condensed phase hydrated proton imply that utilizing SCC-DFTB in RPC quantum PIMD simulations may require careful benchmarking when applied to new systems and/or for simulations of the same system using a new underlying electronic density functional in the AIMD.

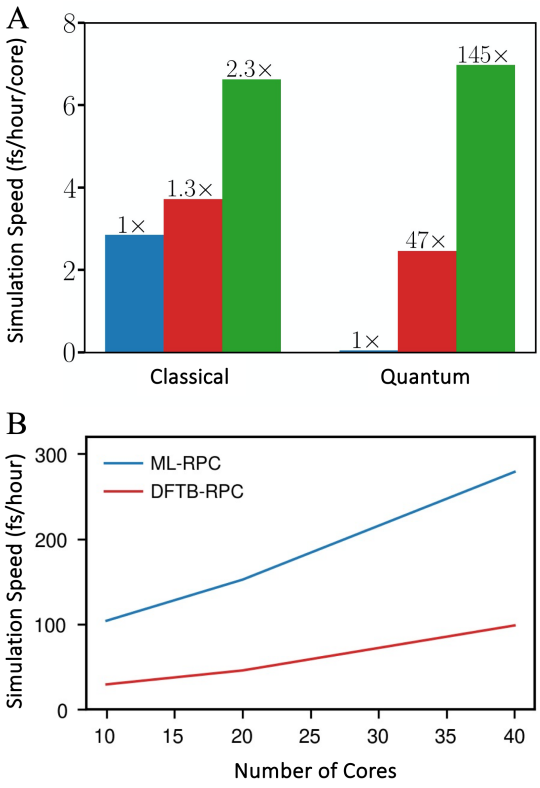


Figure 3. Simulation efficiency benchmarks. (A) Simulation speed defined as simulation time obtained (fs) divided by wall time (hour) and number of cores in classical and quantum simulations. The blue bars are plain AIMD runs without using MTS or RPC. The red bars use SCC-DFTB and the green bars use the ML potential to accelerate simulations via MTS or RPC. The acceleration compared to the plain runs are denoted above the bar tops. A simulation using 600 cores is used to calculate the speed for the plain quantum simulation for a faster overall simulation speed, while the 40-core data is used for all other simulation protocols. Note that the simulation speeds plotted are scaled by the number of cores used. (B) Simulation speed defined as simulation time (fs) divided by wall time (hour) as a function of core number used.

Due to the reduced number of energy evaluations on the *ab initio* PES, the MTS and RPC approaches can accelerate classical and quantum AIMD simulations given that the reference ML potential is computationally efficient. In **Figure 3**A, we show the simulation speed per CPU core in classical and quantum simulations. In both DFTB- and ML-MTS, we used an outer timestep of 2 fs and an inner timestep of 0.25 fs whereas the timestep in plain classical AIMD was 0.5 fs. If the computational cost for the reference potential is negligible, then the MTS approach should

have a 4-fold speedup. However, due to the computational overhead of SCC-DFTB and the ML potential, the acceleration is 1.3-fold for DFTB-MTS and 2.3-fold for ML-MTS. We note that the ANNs used in the ML potentials can be more efficiently evaluated on specialized hardware, such as a graphics processing unit (GPU) and tensor processing unit (TPU). Hence, we expect the 4-fold speedup is eventually achievable for the ML-MTS approach if those specialized computation architectures are employed. For quantum simulations, the acceleration becomes very large where the DFTB-RPC approach accelerates the simulation by a factor of 47 while the ML-RPC approach reduces the computational cost by 2 orders of magnitude (while even benchmarked in a pure CPU environment).

One must also always achieve some degree of certainty that the ML approach is able to learn "perfectly" an inherently many-body *ab initio* PES. This concern may be more prevalent in a heterogeneous system where training a fully reliable ML model is challenging. In this work, using a hydrated excess proton in water case, we demonstrated it is can be the case where a ML potential has difficulty in describing the solute at a low concentration. However, we also discovered that even with a less-than-ideal ML potential, the RPC technique may bridge the "accuracy gap" between the trained ML model and the exact *ab initio* PES, and accurately reproduce the full quantum results. By taking advantage of the low computational cost of evaluating an ANN, the ML-RPC is also two orders of magnitude faster than a plain full *ab initio* PI- AIMD simulation. This holds true even after accounting for the computational work required to generate the training set. If the quantum TRPMD simulations in this paper were carried out with pure *ab initio* path integral, the amount of energy and gradient evaluations would be more than 100 times higher than those producing the training data. More importantly, the training is a "one-time deal" and, once completed, both classical and quantum AIMD can be run cheaply for the same system or even for

similar systems considering the tolerance of the MTS and RPC approaches. The accuracy and the efficiency of our approach thus demonstrates its promise and potential for the future *ab initio* modeling of complex systems with NQEs taken into account.

In the present work, the ML potential was trained by targeting both the *ab initio* energies and forces (see Supporting Information) to improve the training convergence. We note that the resulting ML model, combined with RPC, is almost able to reproduce everything from the exact PIMD simulation except some small deviations in the high free-energy region of the bead-averaged δ PMF (see Figure S3). This behavior could possibly be improved by involving force matching only in the training process to best fulfill the slow-varying assumption made in the MTS and RPC frameworks (eqs 6 and 7). In addition, the quantum fluctuations vary for different nuclei due to their mass difference (the thermal de Broglie wavelength of an atom is on the order of $1/\sqrt{m}$), and thus the current approach could be further improved by introducing masses as weights in the force residual, such as

$$\chi^2 = \left\langle \left(\boldsymbol{F}_{\text{high}} - \boldsymbol{F}_{\text{low}}\right)^{\mathrm{T}} \mathbf{M}^{-1} \left(\boldsymbol{F}_{\text{high}} - \boldsymbol{F}_{\text{low}}\right) \right\rangle \qquad 9$$

to focus more on the forces related to light atoms. We note that the velocity fluctuation of a classical particle is also inversely proportional to the mass, $\sqrt{\langle \dot{\boldsymbol{r}}^2 \rangle} \propto 1/\sqrt{m}$, so then the mass-weighted force residual of eq 9 may also be helpful for deriving a more accurate reference potential for MTS according to eq 7.

To summarize, machine learning with an ANN has been developed in this work and employed for path integral AIMD simulations within the MTS and RPC schemes. For the examples studied here, the accuracy of the approach is excellent and the speed-up over direct PI-AIMD simulation is dramatic (two orders-of-magnitude). These results suggest that nuclear quantized AIMD

simulations may become rather routine for many systems, especially as further advances in machine learning of *ab initio* force fields are achieved.

ASSOCIATED CONTENT

Supporting Information

The following files are available free of charge: Simulation details of both classical and quantum EDS-BLYP-D3 AIMD; training details of the ANN used in this work; figures showing the quantum RDFs and δ PMFs computed via averaging the results over path integral beads.

AUTHOR INFORMATION

Corresponding Author

*gavoth@uchicago.edu

Author Contributions

The manuscript was conceived and written through contributions of both authors. The authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Prof. Tom Markland for making his group's DFTB-RPC code available to us. This research was supported in part by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award Numbers DE-SC0018648. The research was also supported in part by the National Institute of General Medical Sciences (NIGMS) of the U.S. National Institutes of Health (NIH) through grant R01 GM053148. Computational resources were provided by the Research Computing Center (RCC) at the University of Chicago. We thank Dr. Yining Han for helpful discussions.

REFERENCES

1. Tuckerman, M. E., Ab initio molecular dynamics: basic concepts, current trends and novel applications. *J. Phys.: Condens. Matter* **2002,** *14*, R1297.

2. Carloni, P.; U. Rothlisberger; M. Parrinello, The role and perspective of ab initio molecular dynamics in the study of biological systems. *Acc. Chem. Res.* **2002,** *35*, 455-464.

3. Marx, D.; J. Hutter, Ab initio molecular dynamics: Theory and implementation. *Modern methods and algorithms of quantum chemistry* **2000,** *1*, 141.

4. Marsalek, O.; T. E. Markland, Quantum dynamics and spectroscopy of ab initio liquid water: The interplay of nuclear and electronic quantum effects. *J. Phys. Chem. Lett.* **2017,** *8*, 1545-1551.

5. Ceriotti, M.; W. Fang; P. G. Kusalik; R. H. McKenzie; A. Michaelides; M. A. Morales; T. E. Markland, Nuclear quantum effects in water and aqueous systems: Experiment, theory, and current challenges. *Chem. Rev.* **2016,** *116*, 7529-7550.

6. Tuckerman, M. E.; D. Marx; M. L. Klein; M. Parrinello, On the quantum nature of the shared proton in hydrogen bonds. *Science* **1997,** *275*, 817-820.

7. Cassone, G., Nuclear quantum effects largely influence molecular dissociation and proton transfer in liquid water under an electric field. *J. Phys. Chem. Lett.* **2020,** *11*, 8983-8988.

8. Napoli, J. A.; O. Marsalek; T. E. Markland, Decoding the spectroscopic features and time scales of aqueous proton defects. *J. Chem. Phys.* **2018,** *148*, 222833.

9. Hammes-Schiffer, S., Quantum–classical simulation methods for hydrogen transfer in enzymes: a case study of dihydrofolate reductase. *Curr. Opin. Struct. Biol.* **2004,** *14*, 192-201.

10. Feynman, R. P.; A. R. Hibbs, *Quantum mechanics and path integrals*. McGraw-Hill: New York,, 1965; p xiv, 365 p.

11. Marx, D.; M. Parrinello, Ab initio path integral molecular dynamics: Basic ideas. *J. Chem. Phys.* **1996,** *104*, 4077-4082.

12. Markland, T. E.; M. Ceriotti, Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.* **2018,** *2*, 1-14.

13. Markland, T. E.; D. E. Manolopoulos, An efficient ring polymer contraction scheme for imaginary time path integral simulations. *J. Chem. Phys.* **2008,** *129*, 024105.

14. Buxton, S. J.; S. Habershon, Accelerated path-integral simulations using ring-polymer interpolation. *J. Chem. Phys.* **2017,** *147*, 224107.

15. Ceriotti, M.; D. E. Manolopoulos, Efficient first-principles calculation of the quantum kinetic energy and momentum distribution of nuclei. *Phys. Rev. Lett.* **2012,** *109*, 100604.

16. Cao, J.; G. A. Voth, The formulation of quantum statistical mechanics based on the Feynman path centroid density. I. Equilibrium properties. *J. Chem. Phys.* **1994,** *100*, 5093-5105.

17. Steele, R. P., Communication: Multiple-timestep ab initio molecular dynamics with electron correlation. *J. Chem. Phys.* **2013,** *139*, 011102.

18. Kapil, V.; J. VandeVondele; M. Ceriotti, Accurate molecular dynamics and nuclear quantum effects at low cost by multiple steps in real and imaginary time: Using density functional theory to accelerate wavefunction methods. *J. Chem. Phys.* **2016,** *144*, 054111.

19. Marsalek, O.; T. E. Markland, Ab initio molecular dynamics with nuclear quantum effects at classical cost: Ring polymer contraction for density functional theory. *J. Chem. Phys.* **2016,** *144*, 054112.

20. Cheng, X.; J. D. Herr; R. P. Steele, Accelerating ab initio path integral simulations via imaginary multiple-timestepping. *J. Chem. Theory Comput.* **2016,** *12*, 1627-1638.

21. Elstner, M.; D. Porezag; G. Jungnickel; J. Elsner; M. Haugk; T. Frauenheim; S. Suhai; G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998,** *58*, 7260.

22. Gaus, M.; Q. Cui; M. Elstner, DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* **2011,** *7*, 931-948.

23. Goyal, P.; M. Elstner; Q. Cui, Application of the SCC-DFTB method to neutral and protonated water clusters and bulk water. *J. Phys. Chem. B* **2011,** *115*, 6790-6805.

24. Maupin, C. M.; B. Aradi; G. A. Voth, The self-consistent charge density functional tight binding method applied to liquid water and the hydrated excess proton: benchmark simulations. *J. Phys. Chem. B* **2010,** *114*, 6922-6931.

25. Liang, R.; J. M. Swanson; G. A. Voth, Benchmark study of the SCC-DFTB approach for a biomolecular proton channel. *J. Chem. Theory Comput.* **2014,** *10*, 451-462.

26. Zhang, L. F.; J. Q. Han; H. Wang; R. Car; E. Weinan, Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018,** *120*,

27. Zhang, L.; J. Han; H. Wang; W. Saidi; R. Car; W. E In *End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems*, Advances in Neural Information Processing Systems, Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; Garnett, R., Eds. Curran Associates, Inc.: 2018.

28. Becke, A. D., Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988,** *38*, 3098-3100.

29. Lee, C. T.; W. T. Yang; R. G. Parr, Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B* **1988,** *37*, 785-789.

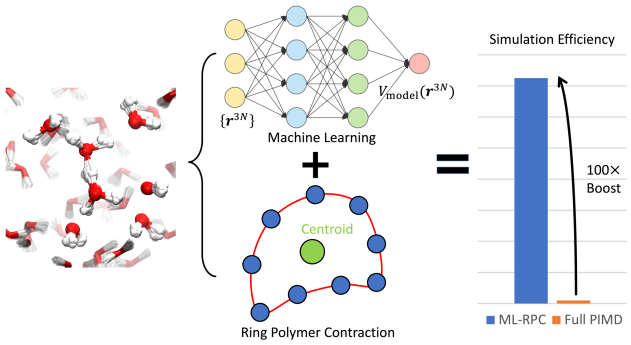
30. Grimme, S.; J. Antony; S. Ehrlich; H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010,** *132*, 154104.

31. Calio, P. B.; G. M. Hocky; G. A. Voth, Minimal Experimental Bias on the Hydrogen Bond Greatly Improves Ab Initio Molecular Dynamics Simulations of Water. *J. Chem. Theory Comput.* **2020,** *16*, 5675-5684.

32. Li, C.; G. A. Voth, Using Constrained Density Functional Theory to Track Proton Transfers and to Sample Their Associated Free Energy Surface. *J. Chem. Theory Comput.* **2021,** *17*, 5759–5765.

33. Rossi, M.; M. Ceriotti; D. E. Manolopoulos, How to remove the spurious resonances from ring polymer molecular dynamics. *J. Chem. Phys.* **2014,** *140*, 234116.

TOC Figure


$\{\boldsymbol{r}^{3N}\}$
$V_{\mathrm{model}}(\boldsymbol{r}^{3N})$
Machine Learning
+
Centroid
Ring Polymer Contraction
=
Simulation Efficiency
100×
Boost
ML-RPC
Full PIMD

Supporting Information for

Using Machine Learning to Greatly Accelerate Path Integral *Ab Initio* Molecular Dynamics

*Chenghan Li and Gregory A. Voth**

Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, 60637, USA

*Corresponding author: gavoth@uchicago.edu

Simulation Details

EDS-BLYP-D3 simulations were performed with 1 H^+ solvated in 128 water molecules in a cubic simulation box of side length = 15.64 Å, and the net charge was neutralized by a uniform background. The Goedecker-Teter-Hutter (GTH) pseudopotentials[1] were used to describe the core electrons, a TZV2P basis set was used to expand the Kohn-Sham orbitals, and a plane wave basis set with a cutoff of 400 Ry was used to expand the electron density. The wave function at each step was optimized by the orbital transformation (OT) method,[2] using a self-consistent field (SCF) convergence criterion of 1×10^{-7} a.u.

For the classical AIMD simulations with EDS-BLYP-D3, the system was equilibrated in the canonical (constant NVT) ensemble for 250 ps at 298 K. In all simulations, a Nose-Hoover chain thermostat with a characteristic frequency of 3000 cm^{-1} was coupled to all degrees of freedom, and a timestep of 0.5 fs was used to integrate the Newton's equations of motion. All simulations were carried out with the CP2K software package.[3-4] The EDS correction (described in more detail below) was added via a modified version of PLUMED2[5] coupled to CP2K, with the EDS parameters taken from ref [6]. The MST classical and RPC quantum simulations were initiated from the constant NVT equilibrated configurations. For DFTB-MTS/RPC, the SCC-DFTB method with D3 dispersion was used as the low-level potential (reference potential), and the parameters

provided by CP2K for H and O atoms were used. Three independent replicas of classical DFTB-MTS simulations were performed for 40-60 ps in the microcanonical (constant NVE) ensemble with an inner timestep of 0.25 fs and an outer timestep of 2 fs. Three independent replicas of DFTB-RPC TRPMD simulations with P = 30 path integral beads were performed for 65-70 ps with centroid contraction and the same MTS setup as the classical simulation. For ML-MTS/RPC, a deep learning potential (DP) trained on the EDS-BLYP-D3 data was used as the reference potential (see below for training details). Three independent replicas of classical ML-MTS simulations were performed for ~50 ps in the constant NVE ensemble with an inner timestep of 0.25 fs and an outer timestep of 2 fs. Six independent replicas of ML-RPC TRPMD simulations with P = 30 beads were performed for ~200 ps with centroid contraction and the same MTS setup as the classical simulation. The full *ab initio* PIMD using EDS-BLYP-D3 was carried out for ~33 ps using a Langevin thermostat with relaxation time of 100 fs. A timestep 0.5 fs was used to integrate the dynamics and the full *ab initio* calculations were carried out on the total 30 path integral beads. All the MTS/RPC and PIMD was performed by i-Pi.[7]

The PMFs were computed using it relation to the histogram, as follows

$$\mathrm{PMF}(\delta) = -k_{\mathrm{B}}T\log\big(P(\delta)\big), \tag{S1}$$

where δ is the proton-sharing coordinated defined in eq 8 in the main text, k_{B} is the Boltzmann constant, T is the simulation temperature, and $P(\delta)$ is the histogram of δ made from MD trajectories. In Figure 2B of the main text, the histogram was computed on the centroid trajectory, while in Figure S2, the histograms were first computed on each path integral bead, and then were averaged to be used in eq S1.

The EDS correction was originally developed for AIMD simulation of water.[6] The EDS potential takes the form of

$$V(\boldsymbol{r}_i) = \sum_{k=0}^{M} \frac{\alpha_k}{\hat{f}_k} \sum_{j\neq i}^{n_{\mathrm{O}}} r_{ij}^{k}\big[1 - u\big(r_{ij} - r_0\big)\big], \tag{S2}$$

which is an additive energy to the DFT PES. When simulating water, the EDS potential (eq S2) is only turned off between O-H pairs within 1.2 Å. The motivation is to only bias the overly strong hydrogen bonds often found in DFT water for certain electronic density functional but keep bonded O-H atoms unperturbed. In order to apply the EDS potential to hydrated protons in water, a

quantitative description of covalency is thus needed to smoothly switch on and off the bias during bond breaking and forming in the Grotthuss proton shuttling mechanism. An approximation to the charge transfer behavior of constrained density functional theory (CDFT) calculations developed in ref[8] was used to compute the CDFT configurational interaction ground-state wave function in the diabatic state representation. Each diabatic state was defined by assigning the excess proton to a different water molecule, resulting into different bonding topologies. The EDS correction was computed in each diabatic state $|i\rangle$ excluding bonded O-H pairs, and the final EDS force was a weighted average of the state-dependent EDS bias force, given by

$$\boldsymbol{F}^{\mathrm{EDS}} = \sum_{i} \boldsymbol{c}_i^2 \boldsymbol{F}_i^{\mathrm{EDS}}. \tag{S3}$$

Artificial Neural Network Training Details

The training set for EDS-BLYP-D3 contained 132000 configurations sampled from the constant NVT equilibration at 298 K. The training was conducted using the DeepMD-kit tool[9] with the smooth edition of deep potential molecular dynamics (DPMD)[10]. The smooth edition of DP using both angular and radial information was employed to featurize the atomic coordinates. The angular descriptors used a cutoff of 5 Å with the smoothing starting from 3.5 Å and the maximum number of neighbor searching was 20 for oxygen atoms and 40 for hydrogen atoms. The embedding network had three hidden layers with 10×20×40 nodes with the axis matrix size of 4. The radial descriptors used a cutoff of 6 Å with the smoothing starting from 1 Å. The maximum number of neighbor searching was 46 for oxygen atoms and 92 for hydrogen atoms. The embedding network had three hidden layers with 5×10×20 nodes. The fitting network had a three-layer architecture with 240×240×240 nodes. The loss function employed in the training process combined both energy residual and force residual

$$loss(t) = w_f(t)|\Delta F|^2 + w_e(t)|\Delta E|^2. \tag{S4}$$

The weights for force error and energy error were changing as a function of training step t via the decaying learning rate

$$w(t) = w_{\mathrm{init}} \frac{lr(t)}{lr(0)} + w_{\mathrm{final}} \left(1 - \frac{lr(t)}{lr(0)}\right), \tag{S5}$$

where $w_{\text{init}} = 10000$ and $w_{\text{final}} = 1$ for force error, and $w_{\text{init}} = 0.02$ and $w_{\text{final}} = 1$ for energy error, and the learning rate $lr(t) = 0.005 \times (0.9473)^{t/5000}$. A mini-batch size of 4 was used and a total of 395500 steps was conducted in the training.

The trained ML potential was benchmarked in a system consist of 1 H^+ and 64 water molecules in a cubic box with side length = 12.447 Å. A NVT run of 400 ps was conducted at 298 K in LAMMPS[11] coupled to the DeepMD-kit.

Supporting Figures

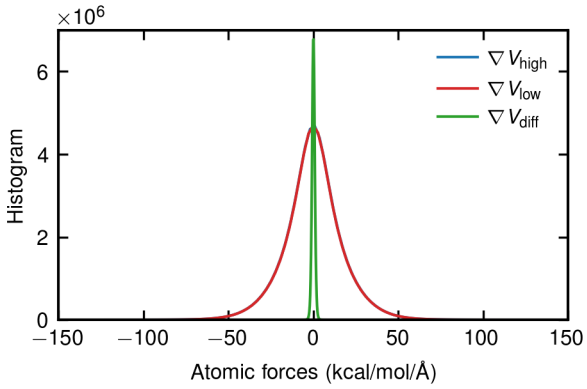


Figure S1. The distribution of atomic forces on the training set. Note that blue and red curves overlap with each other.

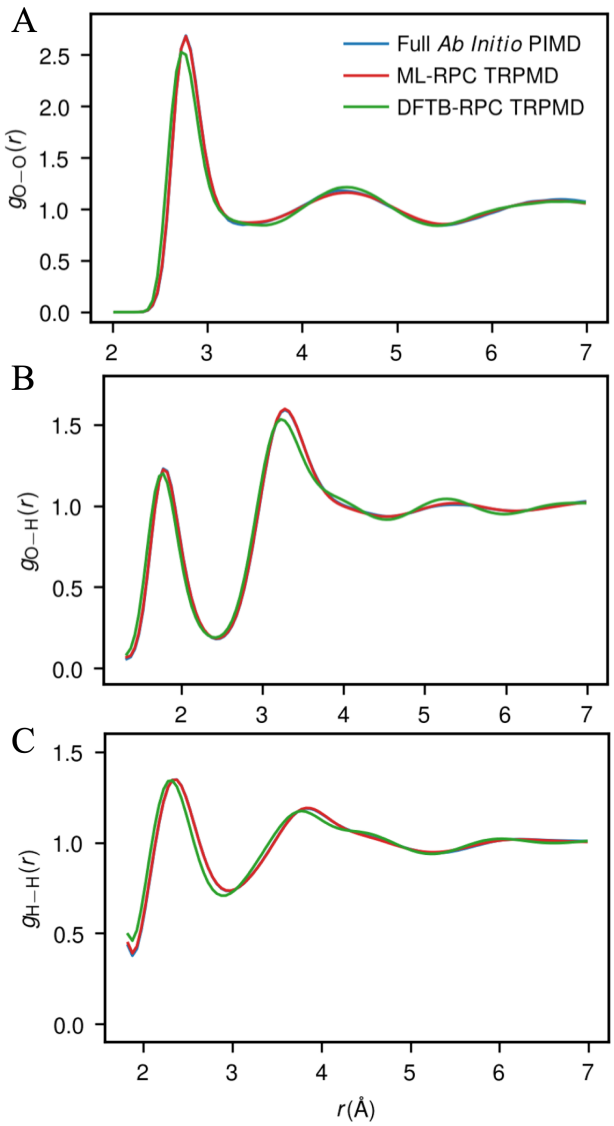


Figure S2. Water solvation structures in terms of the O-O RDF (A), O-H RDF (B), and H-H RDF (C) in quantum simulations. The full *ab initio* PIMD is shown in blue, the TRPMD using ML as RPC reference is in red, and SCC-DFTB as RPC reference in green. All the RDFs were computed independently for each path integral bead, and then averaged over all the beads. Note that the blue and red curves overlap with each other.

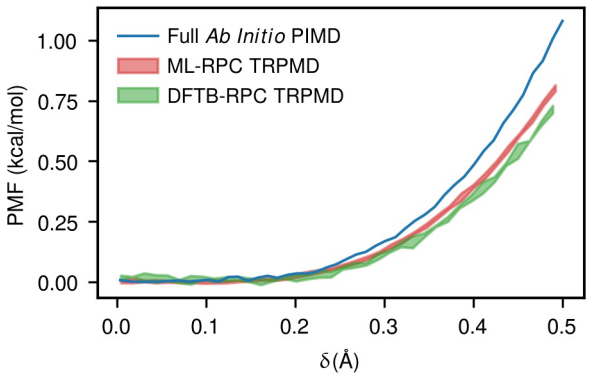


Figure S3. Hydrated excess proton solvation structure measured by the potential of mean force of the proton sharing coordinate (δ) (eq 8). Statistical errors estimated from standard deviations of independent runs are shown as intervals around the curves. The full *ab initio* PIMD reference is shown in blue, the TRPMD using ML potential as RPC reference is in red, the TRPMD using SCC-DFTB as the reference is in green.

Supplemental References

1. Hartwigsen, C.; S. Gœdecker; J. Hutter, Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998,** *58*, 3641.
2. VandeVondele, J.; J. Hutter, An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003,** *118*, 4365-4369.
3. Kuhne, T. D.; M. Iannuzzi; M. Del Ben; V. V. Rybkin; P. Seewald; F. Stein; T. Laino; R. Z. Khaliullin; O. Schutt; F. Schiffmann; D. Golze; J. Wilhelm; S. Chulkov; M. H. Bani-Hashemian; V. Weber; U. Borstnik; M. Taillefumier; A. S. Jakobovits; A. Lazzaro; H. Pabst; T. Muller; R. Schade; M. Guidon; S. Andermatt; N. Holmberg; G. K. Schenter; A. Hehn; A. Bussy; F. Belleflamme; G. Tabacchi; A. Gloss; M. Lass; I. Bethune; C. J. Mundy; C. Plessl; M. Watkins; J. VandeVondele; M. Krack; J. Hutter, CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020,** *152*, 194103.
4. VandeVondele, J.; M. Krack; F. Mohamed; M. Parrinello; T. Chassaing; J. Hutter, QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005,** *167*, 103-128.
5. Tribello, G. A.; M. Bonomi; D. Branduardi; C. Camilloni; G. Bussi, PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014,** *185*, 604-613.
6. Calio, P. B.; G. M. Hocky; G. A. Voth, Minimal Experimental Bias on the Hydrogen Bond Greatly Improves Ab Initio Molecular Dynamics Simulations of Water. *J. Chem. Theory Comput.* **2020,** *16*, 5675-5684.

7. Kapil, V.; M. Rossi; O. Marsalek; R. Petraglia; Y. Litman; T. Spura; B. Cheng; A. Cuzzocrea; R. H. Meißner; D. M. Wilkins, i-PI 2.0: A universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **2019,** *236*, 214-223.
8. Li, C.; G. A. Voth, Using Constrained Density Functional Theory to Track Proton Transfers and to Sample Their Associated Free Energy Surface. *J. Chem. Theory Comput.* **2021,** *17*, 5759–5765.
9. Wang, H.; L. Zhang; J. Han; E. Weinan, DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018,** *228*, 178-184.
10. Zhang, L.; J. Han; H. Wang; W. Saidi; R. Car; W. E In *End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems*, Advances in Neural Information Processing Systems, Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; Garnett, R., Eds. Curran Associates, Inc.: 2018.
11. Plimpton, S., Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J. Comput. Phys.* **1995,** *117*, 1-19.