

# Supporting Information for “First Principles Calculation of Water pK<sub>a</sub> Using the Newly Developed SCAN Functional”

*Ruiyu Wang<sup>1,2</sup>, Vincenzo Carnevale<sup>3,4\*</sup>, Michael L. Klein<sup>1,2,3</sup> and Eric Borguet<sup>1,2\*</sup>*

<sup>1</sup>Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, United States

<sup>2</sup>Center for Complex Materials from First Principles (CCM), Temple University, 1925 North  
12th Street, Philadelphia, Pennsylvania 19122, United States

<sup>3</sup>Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania  
19122, United States

<sup>4</sup>Department of Biology, Temple University, Philadelphia, Pennsylvania 19122, United States

## **Simulation Details**

### **Constrained Molecular Dynamics Simulation Using OH Distance**

First principles Car-Parrinello molecular dynamics (CPMD)<sup>1</sup> simulations based on the SCAN functional<sup>2</sup> were carried out using Quantum Espresso 6.2.1.<sup>3-4</sup> The simulation system contained one OH<sup>-</sup>-H<sub>3</sub>O<sup>+</sup> pair and 53 water molecules in a cubic box of dimensions 11.81 × 11.81 × 11.81 Å<sup>3</sup> with periodic boundary conditions, corresponding to a concentration of about 1 M. A simulation box containing 55 water molecules is large enough that finite size effects are small: others have shown that further increasing the box size does not significantly change the pK<sub>a</sub> of solutes.<sup>5</sup> We used the norm conserving pseudopotentials HSCV<sup>6</sup> for H and O atoms,

with a cutoff of 90 Ry<sup>7</sup>. The fictitious electronic mass was set to 100 atomic units. Nuclei were treated classically and the equations of motion were integrated numerically using a time step of 2 atomic units, about 0.048 fs. Simulations were carried out in the NVT ensemble at 300 K using a Nose-Hoover thermostat with a frequency of 400 cm<sup>-1</sup>.<sup>8-10</sup> The reference simulations were carried out using the CP2K<sup>11-12</sup>/QUICKSTEP<sup>13</sup> package based on Bohn-Oppenheimer MD (BOMD) and the Becke-Lee-Yang-Parr (BLYP)<sup>14-15</sup> exchange and correlation functional with DFT-D3<sup>16</sup> correction is used, the BLYP-D3 has been proved to reproduce the O-O radial distribution function of bulk water.<sup>17</sup> CP2K uses both a Gaussian basis set and auxiliary plane wave basis set.<sup>18</sup> We used a triple- $\zeta$  doubly polarized (TZV2P) basis set and a plane wave cutoff of 400 Ry. BLYP-D3 simulations were carried out using norm conserving GTH pseudopotentials.<sup>19-20</sup> The time-step for BOMD is 0.5 fs and the time constant for the Nose-Hoover thermostat is 100 fs. The reference system had the same number of water molecules and box dimensions as the SCAN simulations. Constrained MD simulations<sup>21</sup> were carried out by constraining the distance of one OH bond to values ranging from 0.8 Å to 1.7 Å in increments of 0.05 Å, for a total of 18 windows. All initial configurations were generated using the SPC/E water model<sup>22</sup> and the GROMACS<sup>23</sup> molecular dynamics package. After a short equilibration with SPC/E, the system underwent geometry optimization (using SCAN or BLYP-D3) before the production runs. For each window, an NVT simulation was performed to collect a trajectory of about 7 ps, of which the last 5 ps were used for analysis. The error on the free energy is estimated by:

$$\delta F(r) = \sum_{i=r_e}^r \sigma(i) \times \Delta r \quad (S1)$$

where  $r_e$  is the last window,  $r_{OH}=1.65 \text{ \AA}$ ;  $\Delta r = 0.05 \text{ \AA}$  is the increment between consecutive windows and  $\sigma(i)$  is the standard deviation of the constraint force, in the  $i^{\text{th}}$  window, estimated by sampling from the time-series of the constraint force at intervals of 1 fs.

### Constrained Molecular Dynamics Simulation Using Coordination Number

The setup is similar to that used for constraining the OH distance, but the reaction coordinate (or collective variable) is replaced by coordination number and we only carry this calculation using the SCAN functional. Here, a soft, smoothed definition of coordination number is used based on the distance between two atoms:

$$n_H = \sum_{i=1}^{N_H} S(|\mathbf{r}_{H_i} - \mathbf{r}_{O^*}|) \quad (S2)$$

The sigmoid function  $S$  is defined using Fermi form:

$$S(r) = \frac{1}{\exp[\kappa(r - r_c)] + 1} \quad (S3)$$

in which  $r_c = 1.384 \text{ \AA}$  and  $\kappa = 10 \text{ \AA}^{-1}$ . When the coordination number is used in constrained MD, the generalized force is calculated using both Lagrange multiplier  $\lambda$  and Cartesian coordinates dependent factors:

$$\xi' = n_H \text{ and } n_H \in \{1.0, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.85, 1.9, 1.95, 2.0\} \quad (S4)$$

$$f_{\xi'} = \frac{\langle Z^{-1/2}[\lambda - kTG] \rangle_{\xi'}}{\langle Z^{-1/2} \rangle_{\xi'}} \quad (S5)$$

$$Z = \sum_i^N \frac{1}{m_i} \left( \frac{\partial \xi}{\partial \mathbf{r}_i} \right)^2 \quad (S6)$$

$$G = \frac{1}{Z^2} \sum_{i,j}^N \frac{1}{m_i m_j} \frac{\partial \xi}{\partial \mathbf{r}_i} \frac{\partial^2 \xi}{\partial \mathbf{r}_i \partial \mathbf{r}_j} \frac{\partial \xi}{\partial \mathbf{r}_j} \quad (S7)$$

The derivation of these equations can be found in the seminal paper by Sprik.<sup>24</sup>

For the purpose of comparing the results of the two simulations, we kept all the other settings unchanged, including the constrained MD as enhanced sampling technique. We choose smoothed coordination numbers from 1.0 to 2.0, with 0.1 as interval and a total of 11 windows in the first stage, and later we added additional windows at 1.15, 1.85 and 1.95. To calculate the correct free energy difference from potential of mean force (PMF), one needs to identify the thermodynamic states of interest with the corresponding values of the reaction coordinate. For a metastable state, the PMF should show a local minimum; consequently, the derivative of the PMF should be 0 at those points. Due to the definition of smoothed coordination number and the limited number of windows, we find that the smoothed coordinate does not show the values of 2 and 1 for H<sub>2</sub>O and OH<sup>-</sup>, respectively. For the H<sub>2</sub>O state,  $f_{\xi'=1.9}$  is positive while  $f_{\xi'=2.0}$  is negative and by reading Figure S1, we choose  $\xi' = 1.96$  as state of H<sub>2</sub>O. For OH<sup>-</sup> state,  $f_{\xi'}$  never reaches 0, but both  $f_{\xi'=1.1}$  and  $f_{\xi'=1.15}$  are quite small. As a result, the average free energy between  $\xi' = 1.1$  and  $\xi' = 1.15$  is 31.775 kT, corresponding to pK<sub>a</sub> 13.8. When compared to the pK<sub>a</sub> predicted using BLYP and a similar protocol, SCAN achieves a better performance, i.e. it 1 pK<sub>a</sub> unit closer to the target value.

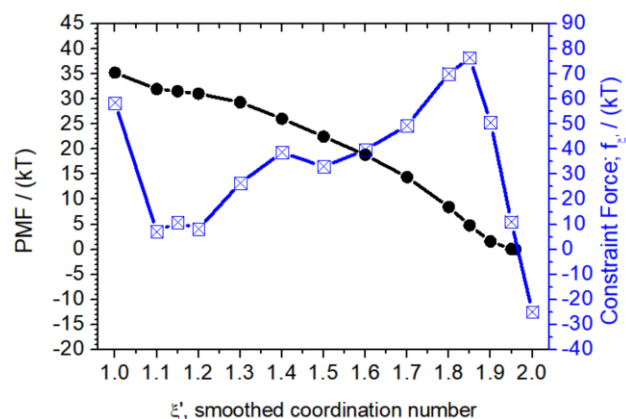


Figure S1. PMF and constraint force from constrained MD using smoothed coordination number as the reaction coordinate. The PMF (black) shows the free energy change from the

equilibrium state of H<sub>2</sub>O ( $\xi' = 1.96$ ) and OH<sup>-</sup> ( $\xi'$  between 1.1 and 1.15). Both the PMF and the constraint force have units of energy, kT (since the coordination number and thus the displacement is dimensionless).

### **Further comparison of reaction coordinates, OH distance and smoothed coordination number**

The selection of reaction coordinates is important to correctly describe the process of interest. For water or other acid dissociation, both OH distance and smoothed coordination number are widely used reaction coordinates.<sup>26-28</sup> However, none of them is perfect. One common issue for OH distance is that when OH distance is long, close to the dissociated state, weak acid anion may accept a proton from neighboring waters, i.e. the OH<sup>-</sup> diffuses away yet the collective variable does not change. These recombination events become extremely frequent at large separation distance of the OH<sup>-</sup>-H<sub>3</sub>O<sup>+</sup> ion pair, therefore estimating the long-range part of the PMF is impossible. As a result, Sprik proposed that smoothed coordination number is a better choice of collective variable,<sup>25</sup> because by constraining the coordination number near the deprotonated state, such recombination cannot happen. However,  $g_{O^*H}(r)$  at different windows shows that even using coordination number as the reaction coordinate, sampling of large OH<sup>-</sup>-H<sub>3</sub>O<sup>+</sup> separation distance is still very poor (Figure S2). In particular, in the windows  $\xi' = 1.1$  and  $\xi' = 1.15$ , the distance between the nearest hydrogen atom H\* and O\* is between 1.6 to 1.7 Å (Figure S2), indicating that these simulations do not improve the sampling compared to simulations in which the OH distance is constrained at 1.65 Å. Further decreasing  $\xi'$  to 1.0 not only pushes the H\* atom away, but also destroys the hydration shell of OH<sup>-</sup> because it reduces the coordination number of O\*; what is more, the covalent bond of OH<sup>-</sup> is longer, indicating the

OH<sup>-</sup> is losing its proton as well (Figure S2), further supporting the notion that the equilibrium state of OH<sup>-</sup> lies between  $\xi' = 1.1$  and 1.15.

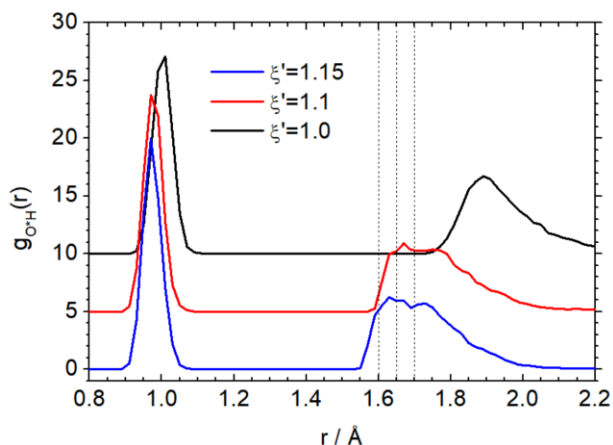


Figure S2. Radial distribution function between O\* and all hydrogen atoms for windows  $\xi' = 1.0$ ,  $\xi' = 1.1$  and  $\xi' = 1.15$ . Three vertical dashed lines highlight the values 1.6 Å, 1.65 Å and 1.7 Å.

## REFERENCES

- (1) Car, R.; Parrinello, M., Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471-2474.
- (2) Sun, J.; Ruzsinszky, A.; Perdew, J. P., Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.
- (3) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I., QUANTUM ESPRESSO: a Modular and Open-source Software Project for Quantum Simulations of Materials. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.

- (4) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M., Advanced Capabilities for Materials Modelling with Quantum ESPRESSO. *J. Phys.: Condens. Matter* **2017**, *29*, 465901.
- (5) Tummanapelli, A. K.; Vasudevan, S., Ab Initio Molecular Dynamics Simulations of Amino Acids in Aqueous Solutions: Estimating pKa Values from Metadynamics Sampling. *J. Phys. Chem. B* **2015**, *119*, 12249-12255.
- (6) Vanderbilt, D., Optimally Smooth Norm-conserving Pseudopotentials. *Phys. Rev. B* **1985**, *32*, 8412-8415.
- (7) DelloStritto, M. J.; Piontek, S. M.; Klein, M. L.; Borguet, E., Effect of Functional and Electron Correlation on the Structure and Spectroscopy of the Al<sub>2</sub>O<sub>3</sub>(001)–H<sub>2</sub>O Interface. *J. Phys. Chem. Lett.* **2019**, *10*, 2031-2036.
- (8) Hoover, W. G., Canonical Dynamics: Equilibrium Phase-space Distributions. *Phys. Rev. A* **1985**, *31*, 1695-1697.
- (9) Nosé, S., A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511-519.
- (10) Martyna, G. J.; Klein, M. L.; Tuckerman, M., Nosé–Hoover Chains: The Canonical Ensemble via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97*, 2635-2643.
- (11) The CP2K developers group, <http://cp2k.berlios.de/>. **2004**.
- (12) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J., CP2K: Atomistic Simulations of Condensed Matter Systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 15-25.

- (13) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J., Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167*, 103-128.
- (14) Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785-789.
- (15) Becke, A. D., Density-functional Exchange-energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098-3100.
- (16) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., 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.
- (17) Bankura, A.; Karmakar, A.; Carnevale, V.; Chandra, A.; Klein, M. L., Structure, Dynamics, and Spectral Diffusion of Water from First-Principles Molecular Dynamics. *J. Phys. Chem. C* **2014**, *118*, 29401-29411.
- (18) Lippert, G.; Hutter, J.; Parrinello, M., A hybrid Gaussian and Plane Wave Density Functional Scheme. *Mol. Phys.* **1997**, *92*, 477-488.
- (19) Hartwigsen, C.; Goedecker, S.; Hutter, J., Relativistic Separable Dual-space Gaussian Pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641-3662.
- (20) Goedecker, S.; Teter, M.; Hutter, J., Separable Dual-space Gaussian Pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703-1710.
- (21) Sprik, M.; Ciccotti, G., Free Energy from Constrained Molecular Dynamics. *J. Chem. Phys.* **1998**, *109*, 7737-7744.



- (22) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269-6271.
- (23) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High Performance Molecular Simulations through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*, 19-25.
- (24) Sprik, M., Coordination Numbers as Reaction Coordinates in Constrained Molecular Dynamics. *Faraday Discuss.* **1998**, *110*, 437-445.
- (25) Sprik, M., Computation of the pK of Liquid Water Using Coordination Constraints. *Chem. Phys.* **2000**, *258*, 139-150.
- (26) Davies, J. E.; Doltsinis, N. L.; Kirby, A. J.; Roussev, C. D.; Sprik, M., Estimating pKa Values for Pentaoxyphosphoranes. *J. Am. Chem. Soc.* **2002**, *124*, 6594-6599.
- (27) Doltsinis, N. L.; Sprik, M., Theoretical pKa Estimates for Solvated P(OH)<sub>5</sub> from Coordination Constrained Car-Parrinello Molecular Dynamics. *PCCP* **2003**, *5*, 2612-2618.
- (28) Sandmann, N.; Bachmann, J.; Hepp, A.; Doltsinis, N. L.; Müller, J., Copper(II)-mediated Base Pairing Involving the Artificial Nucleobase 3H-imidazo[4,5-f]quinolin-5-ol. *Dalton Trans.* **2019**, *48*, 10505-10515.