



Title	Deep Neural Network Potential for Modeling Formic Acid Dissociation in Aqueous Solution
Author(s)	Fadilla, Rizka Nur; Morikawa, Yoshitada
Citation	サイバーメディアHPCジャーナル. 2026, 16, p. 92-95
Version Type	VoR
URL	https://doi.org/10.18910/106244
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Deep Neural Network Potential for Modeling Formic Acid Dissociation in Aqueous Solution

Rizka Nur Fadilla and Yoshitada Morikawa

大阪大学 大学院工学研究科

1. Introduction

Acid dissociation in aqueous solutions is a fundamental process that influences numerous chemical phenomena. [1-3] Despite its apparent simplicity, acid dissociation involves both the transfer of protons from the acid to adjacent water molecules and reorganization of the solvent to stabilize the resulting anion and hydronium ion. Accurately capturing these processes continues to pose a substantial challenge for computational modeling.

Ab initio molecular dynamics (AIMD) provides a powerful approach for modeling acid dissociation. This method captures changes during chemical reactions, enabling the analysis of proton transfer mechanisms, solvation dynamics, and intermediate states, which cannot be accurately represented by classical molecular dynamics or empirical force fields. Such atomic-level insights are crucial for understanding acid-base behavior in chemical and biological environments and improving predictions of reaction kinetics, thermodynamics, and mechanisms. However, AIMD's computational demand for AIMD arises from solving the electronic Schrödinger equation at every simulation step, which significantly limits the accessible trajectory length. This constraint restricts extensive sampling for the free-energy calculations required for the acid

dissociation constant calculations.

In recent years, machine learning interatomic potentials (MLIPs) have demonstrated remarkable capability in reproducing the potential energy surfaces predicted by ab-initio calculations while reducing the computational cost. [4,5] Among these approaches, the Deep Potential (DP) framework [6] has emerged as a powerful method for enabling long-timescale molecular dynamics simulations with near ab initio accuracy. Consequently, Deep Potential Molecular Dynamics (DPMD) has become an effective approach for investigating acid dissociation using an explicit solvation model.

Formic acid, among weak acids, serves as a suitable benchmark system due to its simple molecular structure, well-characterized dissociation mechanism, and extensive availability of both experimental and theoretical data. Its moderate acidity, with an experimental pK_a of approximately 3.75 at room temperature, provides a good test for assessing the capability of DP to accurately describe proton transfer and the solvent-mediated stabilization of dissociated ions.

2. Computational Method

2.1 Deep Potential (DP)

We utilized the DP framework to develop MLIP. Within this framework, the total energy of

a system is decomposed into atomic energies, each of which is calculated based on the local environment within a pre-defined local cutoff radius. This approach transforms the local atomic environment into a descriptor that maintains translational, rotational, and permutational symmetry. This descriptor is then fed into a deep neural network to predict the atomic energy and the corresponding atomic forces. The neural network was trained on atomic configurations, along with their respective energies and forces, using DFT data. Following training, the DP model was employed to perform atomistic simulations. The DP framework was implemented in DeePMD-kit [7].

2.2 Density-functional Theory (DFT)

Density Functional Theory (DFT) is a quantum mechanical approach employed to investigate the electronic structures of molecules and materials. In this study, DFT was utilized to obtain the energy and atomic forces that served as training data for constructing the DP model. The VASP software [8,9] was used to calculate the energy and atomic forces, using r^2 SCAN and r^2 SCAN0 as the exchange-correlation functionals along with the PAW pseudopotential. The training data included two systems: pure bulk water and formic acid in an aqueous solution. Fig. 1 and 2 depict the typical training structures used in this study.

2.3 Umbrella Sampling

Due to the infrequent nature of acid dissociation events, conventional molecular dynamics simulations are insufficient for adequate sampling. Therefore, umbrella sampling [10] was utilized to enhance

sampling along the reaction coordinate. This reaction coordinate is defined by the coordination number. The reaction coordinate was divided into fifty windows, and the simulations were constrained with harmonic biasing potential. The biased probability distribution obtained were then integrated using umbrella integration [10] to reconstruct the dissociation free energy profile hence pK_a can be calculated. The simulations were performed using LAMMPS [11] patched with Plumed [12].

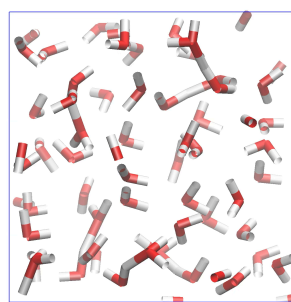


Fig. 1: Bulk water. The red and white colors denote oxygen and hydrogen atoms, respectively.

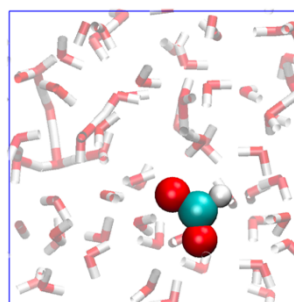


Fig. 2: Formic acid in aqueous solution. The red, white, and green colors denote oxygen, hydrogen, and carbon atoms, respectively

3. Validation of Deep Potential

3.1 Energy and Force

Fig. 3 displays parity plots that compare the energy and atomic forces determined by DFT and machine learning (ML) across 420

test data points. These test data points were not included in the training process. For DP@r²SCAN, the root mean square errors (RMSE) for energy and atomic forces are 1.15 meV/atom and 55.35 meV/Å, respectively, which indicates a good accuracy for the MLIP. MLIP@r²SCAN0 demonstrated a comparable level of accuracy. This accuracy level was also consistent with our previous works [13, 14].

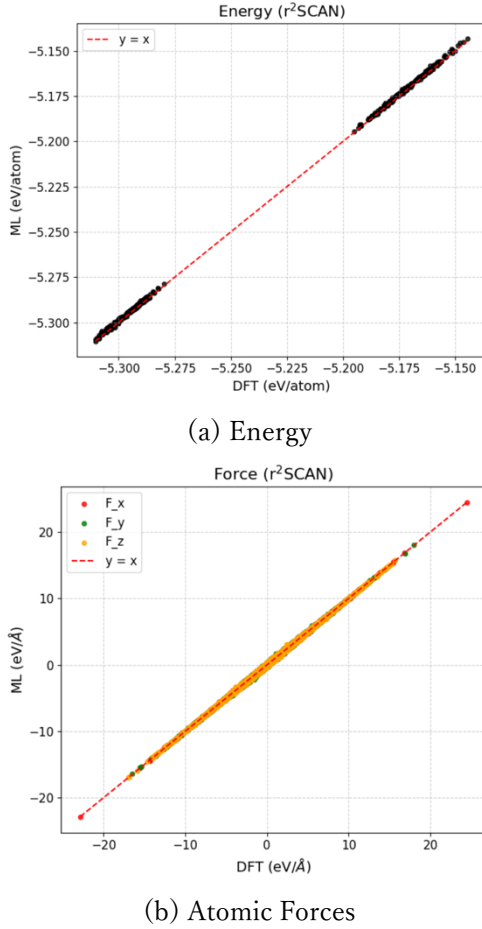


Fig. 3: Parity plot for DP@r²SCAN.

3.2 Bulk Water Properties

Fig. 4 presents the oxygen-oxygen radial distribution function (OO-RDF) of water predicted by DP@r²SCAN and DP@r²SCAN0 at 300 K and 350 K. DP predicts overstructured bulk water at 300 K for both r²SCAN and r²SCAN0, with r²SCAN0 predicting a less overstructured configuration. This overstructuring arises from the omission of

nuclear quantum effects (NQE) from simulations. As noted in previous studies, an approach to partially simulate the NQE is to increase the temperature. Our results demonstrate that an additional 50 K can approximate the experimental OO-RDF at 300 K, which is in agreement with previous studies. We also performed NPT simulations to predict equilibrium bulk water density, resulting in r²SCAN and r²SCAN0 prediction of 0.96 gr/cm³, close to the experimental bulk water density of 0.997 gr/cm³.

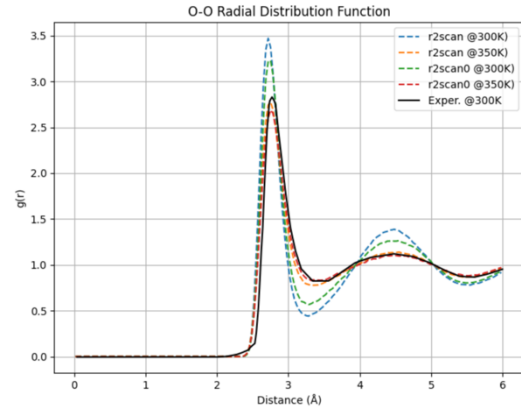


Fig. 4: Oxygen-oxygen radial distribution function of water.

4. Formic Acid Dissociation

Structures were sampled from umbrella sampling trajectories in the initial state (non-dissociated, IS), transition state (TS), and final state (dissociated, FS). For each ensemble, the O*-H RDF was calculated, as depicted in Fig. 5. The progressive shift of the first peak position towards larger values confirms the dissociation process. The coordination number, which is indicative of the hydration structure, can be determined by integrating the RDF, shown in the right axis of Fig. 5. To enhance clarity, the hydration structures of IS and FS are visualized in Fig. 6. Upon verifying the reliability of the sampled trajectory, we

will proceed to the free energy calculations and ensure their convergence, which is our ongoing work.

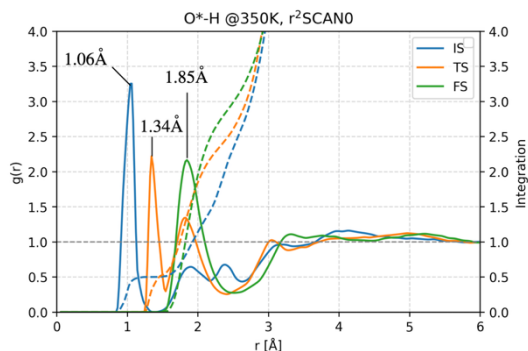


Fig. 5: O*-H radial distribution function of the initial states (IS), transition states (TS), to final states (FS).

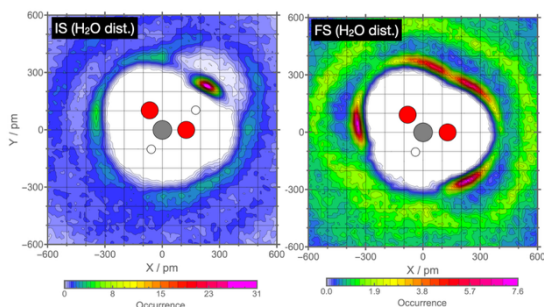


Fig. 6: Hydration structure of formic acid (left) and formate (right) in aqueous solution.

5. Conclusion

We developed an MLIP using a DP framework to model formic acid dissociation with explicit solvation. The accuracy of the DP was validated against the DFT reference data and experimental bulk water properties, demonstrating reasonable agreement and indicating an accurate DP model. Subsequently, DP was integrated with umbrella sampling simulations to enhance the observation of dissociation, which was confirmed to progress reasonably. We are currently calculating the dissociation free energy. Following the completion of the

formic acid case, we aimed to benchmark the applicability of this framework to more complex cases, such as molecules with multiple dissociation sites, and to investigate in detail the contributions of steric effects present in complex molecules.

Bibliography

- (1) D. T. Manallack, et al., *Chem. Soc. Rev.*, **42**, 485-496 (2013)
- (2) L. Mamy, et al., *Crit. Rev. Environ. Sci. Technol.*, **45**, 1277-1377 (2015)
- (3) P. Patel, et al., *Trends Pharmacol. Sci.*, **42**, 448-460 (2021)
- (4) J. Behler, et al., *Phys. Rev. Lett.*, **98**, 146401 (2017)
- (5) K. T. Schütt, et al., *Nat. Commun.*, **8**, 1-8 (2017)
- (6) L. Zhang, et al., *Phys. Rev. Lett.*, **120**, 143001 (2018)
- (7) H. Wang, et al., *Comput. Phys. Commun.*, **228**, 178-184 (2018)
- (8) G. Kresse, et al., *Phys. Rev. B.*, **59**, 1758 (1999)
- (9) G. Kresse, et al., *Phys. Rev. B.*, **54**, 11169 (1996)
- (10) J. Kästner, et al., *J. Chem. Phys.*, **123**, 144104 (2005)
- (11) A. P. Thompson, et al., *Comput. Phys. Commun.*, **271**, 108171 (2022)
- (12) G. A. Tribello, et al., *Comput. Phys. Commun.*, **185**, 604-612 (2014)
- (13) R. N. Fadilla, et al. *Cybermedia HPC Journal*, **14**, 89-92 (2024)
- (14) R. N. Fadilla, et al. *Cybermedia HPC Journal*, **15**, 67-70 (2025)