Dynamics and Kinetic Exploration of Oxygen Reduction Reaction at Fe-N_4/C-water Interface Accelerated by Machine Learning Force Field

Qinghan Yu[1], Pai Li[2*], Xing Ni[1], Youyong Li[1,3], Lu Wang[1,4*]

[1]*Institute of Functional Nano & Soft Materials (FUNSOM), Soochow University, Suzhou, 215123, Jiangsu, China*

[2] State Key Laboratory of Materials for Integrated Circuits, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China

[3]*Macao Institute of Materials Science and Engineering, Macau University of Science and Technology, Taipa 999078, Macau SAR, China*

[4]*Jiangsu Key Laboratory of Advanced Negative Carbon Technologies, Soochow University, Suzhou, 215123, Jiangsu, China*

E-mail: lipai@mail.sim.ac.cn; lwang22@suda.edu.cn

S1 Computational details

S1.1 Density Functional Theory Calculations

The Vienna *ab initio* simulation package (VASP)[1] was employed for spin-polarized Density Functional Theory (DFT) calculations using a pseudopotential plane wave approach. We utilized the generalized gradient approximation (GGA) with the Perdew-Burke-Ernzerhof (PBE) exchange correlation functional[2] for single-point energy calculations, geometry optimization, and ab initio molecular dynamics (AIMD) to accurately describe electronic interactions. A plane-wave basis set with a cutoff energy of 480 eV and Fermi smearing with a width of 0.1 eV was implemented. Brillouin zone sampling was conducted using the Gamma-center method[3] with a k-spacing of 0.38 $Å^{-1}$. The convergence thresholds were set to 10^{-5} eV for the self-consistent field and 10^{-4} eV/Å for geometry optimization. The D3 scheme of Van der Waals correction[4] was employed in all DFT calculations. AIMD simulations were performed with a timestep of 1 fs. Temperature and pressure were controlled using Nose-Hoover thermostats for the canonical (NVT) ensemble and Langevin thermostats for the isothermal-isobaric (NPT) ensemble. Details of temperature, pressure and simulation duration are provided in Table S1. A previous study has demonstrated that hybrid functional/meta-GGA and PBE (D3) exhibit consistent trends for this system[5]. Thus, the functional employed in this study is adequate for describing the system under investigation.

S1.2 Machine Learning Force Field (MLFF)Training

The DeepMD-kit v2.2.1[6, 7] was utilized to train a MLFF. The network architecture

consisted of an embedding network with layers of sizes 25 × 50 × 100, and a fitting network of sizes 240 × 240 × 240. We adopted a radial cutoff of 6 Å with the se_e2_a descriptor. For activate learning iterations, the model was initially trained for 1×10^5 steps. Upon achieving convergence, the production model was trained for 1×10^6 steps. The learning rate was gradually reduced from 1.0×10^{-3} to 3.5×10^{-8}. The loss function's energy and force prefactors were adjusted from 0.02 to 1 and from 1,000 to 1, respectively. We have analyzed the time required for MLFF training and dataset computation, including the initial dataset and 10 rounds of active learning. All training and MD simulations were performed on a single V100 GPU and all DFT calculations were performed on a 24-core CPU node. 10 rounds activate learning required 40 models and 80 MD trajectories, costing 80 hours and 400 hours respectively. All dataset consisted of 6008 single-point energy calculations costing about 900 hours.

S1.3 MLFF-based molecular dynamic and classical molecular dynamic simulations

Molecular dynamic (MD) simulations utilizing the DP Potential were executed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[8] with DeePMD-kit plugin. The classical MD simulations, using the TIP3P[9] force field, were conducted by LAMMPS for pure water system sampling. Periodic boundary conditions, a time step of 0.5 fs and a temperature of 300 K were employed for all simulations. The Nose-Hoover thermostat was applied for both the NVT and NPT ensembles. The T damp parameter was set to 0.05 for both ensembles, and the P damp parameter was set to 5 for the NPT ensemble. An external pressure of 1 bar was applied to the NPT

ensemble.

S1.4 Metadynamics and free energy calculation

All enhanced sampling simulations in this study were performed with a well-tempered metadynamic approach implemented with Plumed package v2.8[10] which was interfaced with LAMMPS. A path collective variable (CV)[11, 12] was employed to explore the free energy variations along a specified reaction coordinate. In practical research, the path CV was defined as follows:

$$s = \frac{\sum_{i=1}^{N} i e^{-\lambda |X - X_i|^2}}{\sum_{i=1}^{N} e^{-\lambda |X - X_i|^2}} \tag{1}$$

$$z = -\frac{1}{\lambda} ln\left[\sum_{i=1}^{N} e^{-\lambda |X - X_i|^2}\right] \tag{2}$$

where s represents progress along a predefined pathway through an ordered sequence of atomic configurations, and *z* quantifies the deviation from the predefined pathway. N is the number of reference structure; X is a measurement value during metadynamics simulations and X_i are preassigned reference structures. The path CV was discretized using seven reference structures with different coordination numbers (CNs). The parameters and equations for calculating CNs were provided in Table S2, and the details of reference structures were provided in Table S3. The parameter λ was set to 0.25. For metadynamics, the initial height of Gaussian bias potential for *s* and *z* was set to 0.05 eV, and the width for s and z was set to 0.05 Å and 1 Å, respectively. The simulations

were conducted at 300 K with a bias factor of 75 and a deposition rate of 25 fs. All parameters about CNs and metadynamics were set based on the previous work[13, 14] with further testing. Upon the convergence of metadynamics, the free energy was calculated based on the Hills file. For two-dimensional free energy profiles, a string method[15, 16] was employed to determine the minimum energy path. The initial configuration of the global free energy surface calculation (Figure 5) includes catalyst, 29 water molecules with one O_2 molecule and 1 H_3O^+. The initial configuration of free energy surface calculation of PCET 1&2 (Figure 6) includes the same thing as the global free energy surface calculation. The initial configuration of free energy surface calculation of PCET 3&4 (Figure S7) includes catalyst, 30 water molecules, 1 OH adsorbed on Fe site and 1 H_3O^+. The MD simulation duration of all free energy surface calculation are 1 ns.

S1.5 MLFF validation and trajectory analysis

A comparison between energies/forces calculated by DFT and those predicted by the MLFF was conducted for all structures in the training set. The same comparison was conducted on structures from a 5 ns MD trajectory with metadynamics, used as a validation set. Various error metrics, including mean absolute error (MAE), root mean square error (RMSE) and the coefficient of determination (R^2)[17], were used to evaluate the performance of the model trained. Radial distribution functions (RDF) were calculated by OVITO[18] python interface based on a 5 ps MD trajectory of a system containing 60 water molecules. The MD simulation duration corresponding to the reactant, product and MetaD (Figure 3a) is 1 ns. The initial configuration of the reactant

and MetaD includes catalyst, 30 water molecules and one oxygen adsorbed on Fe site. The initial configuration of the product includes catalyst, 31 water molecules in the solution and one water molecule adsorbed on Fe site. The structural features of the training set, generated by the se_e2_a descriptor, were visualized using the t-distributed stochastic neighbor embedding (t-SNE) method[19]. For the entire training set and the training set containing only water molecule systems, principal component analysis (PCA) was used for initialization embedding (inside t-SNE). For all Fe-N_4/C-water systems, the random method was used for initialization embedding. The adsorption energy of O_2 and H_2O was defined as:

$$E_{ads} = E_{all} - E_{catal} - E_{mol} \tag{3}$$

where E_{ads} represents adsorption energy, E_{all} is the energy when molecule was adsorbed onto the site, E_{catal} is the energy of Fe-N_4/C, E_{mol} is the energy of a single O_2 or H_2O molecule in a vacuum. For DFT energy calculations, VASP was employed with the same parameter as in section 2.1. The MLFF energy predictions were performed via the ASE[20] interface of DeepMD-kit, and structural optimization was performed using the BFGS algorithm with a force threshold value of 10^{-4} eV/Å. For hydrogen bond lifetime, water density, and water configuration space calculations, MDAnalysis[21] was employed on a 1ns MD trajectory within an NPT ensemble.

S2 Supporting Tables

Table S1. Training set composition.

Category	chemical	Notes	Numbers	Stage

	symbol			
Water	H_4O_2	A cluster composed of two water molecules; MLFF-MD at 300 K, 1 bar, 500 ps, NPT, with Metadynamic.	27	activate learning
		2 water molecules selected from 80 water molecules, ρ=1 g/cm^3; MLFF-MD at 300 K, 1 ns; NVT.	30	
	$H_{12}O_6$	6 water molecules selected from 80 water molecules, ρ=1 g/cm^3; MLFF-MD at 300 K, 1 ns; NVT.	45	
	$H_{16}O_8$	8 water molecules, ρ=0.1 g/cm^3; classical MD at 300 K, 10 ns; NVT.	45	initial
	$H_{24}O_{12}$	12 water molecules, ρ=0.15 g/cm^3; classical MD at 300 K, 10 ns; NVT.	45	initial
	$H_{32}O_{16}$	16 water molecules with ice phase as initial structure; AIMD accelerated by one-the-fly MLFF at 0 K-1000 K, 5 ps, NVT.	368	
	$H_{36}O_{18}$	18 water molecules, ρ=0.24 g/cm^3; classical MD at 300 K, 10 ns; NVT.	45	
	$H_{64}O_{32}$	32 water molecules with ice phase as initial structure; AIMD accelerated by one-the-fly MLFF at 100 K-600 K, 1 bar, 2 ps, NPT.	204	
	$H_{128}O_{64}$	64 water molecules, ρ=0.84 g/cm^3; classical MD at 300 K, 10 ns; NVT.	45	
Water-O_2	$H_{76}O_{40}$	38 water and one O_2 molecules; AIMD accelerated by one-the-fly MLFF at 100 K-600 K, 1 bar, 5 ps, NPT.	572	
Graphene	C_{32}	4*4 graphene; AIMD accelerated by one-the-fly MLFF at 100 K-600 K, 20 ps, NVT.	79	

Catalyst	$C_{26}N_4Fe$	Fe single atom catalyst with 4 N coordination; AIMD accelerated by one-the-fly MLFF at 100 K-600 K, 5 ps, NVT.	101	
Graphene-water	$C_{32}H_{64}O_{32}$	A 4*4 graphene with 32 water molecules; AIMD accelerated by one-the-fly MLFF at 100 K-600 K, 2 ps, NVT.	234	
Catalyst-water	$C_{26}N_4FeO_{16}H_{32}$	Catalyst adsorbed with OH, 14 water molecules, 1 H_3O^+; AIMD accelerated by one-the-fly MLFF at 100 K-600 K, 5 ps, NVT.	412	
	$C_{26}N_4FeO_{32}H_{61}$	Catalyst, 29 water molecules with one O_2 molecule, 1 H_3O^+; AIMD accelerated by one-the-fly MLFF at 100 K and 250K, 1 bar, 5 ps, NPT.	299	
		Catalyst adsorbed with O_2, 29 water molecules, 1 H_3O^+; MLFF-MD at 300 K, 1 bar, 28 ns, NPT, with Metadynamic.	730	activate learning
	$C_{26}N_4FeO_{32}H_{62}$	Catalyst adsorbed with OOH, 29 water molecules with one O_2 molecule, 1 H_3O^+; AIMD accelerated by one-the-fly MLFF at 100 K-600 K and 250K, 1 bar, 5 ps, NPT.	144	initial
	$C_{26}N_4FeO_{32}H_{64}$	Catalyst, 32 water molecules; AIMD at 300K, 1 bar, 0.5 ps, NPT.	450	
		Catalyst, 32 water molecules; AIMD accelerated by one-the-fly MLFF at 300 K, 1 bar, 2 ps, NPT.	40	
		Catalyst, 32 water molecules; AIMD accelerated by one-the-fly MLFF at 100 K-600 K, 5 ps, NVT.		
		Catalyst adsorbed with OH, 30 water molecules, 1 H_3O^+; AIMD at 300 K, 1 bar, 10 fs, NPT.	14	

		Catalyst adsorbed with OH, 30 water molecules, 1 H_3O^+; AIMD accelerated by one-the-fly MLFF at 100 K-600 K, 5 ps, NVT.	279	
		Catalyst adsorbed with OH, 30 water molecules, 1 H_3O^+; MLFF-MD at 300 K, 1 bar, 20 ns, NPT, with Metadynamic.	1753	activate learning
		Catalyst, 32 water molecules; MLFF-MD at 300 K, 1 bar, 3 ns, NPT, with Metadynamic.	47	
Total			6008	

Note: AIMD accelerated by one-the-fly MLFF is a new function of VASP v6.3.0 using Bayesian linear regression[22]. Activate learning means activate learning dataset, initial means initial dataset.

Table S2. Parameters and equations for calculating coordination numbers.

CV	Definition	Parameters
CN O^2-H	$CN\ O^2-H=\sum_{i\in H_w}\frac{1-(\frac{r_{O^2,\ i}}{r_0})^m}{1-(\frac{r_{O^2,\ i}}{r_0})^n}$	$r_{O^2,\ i}$: distance between atom O far away from Fe in *O_2 and i-*th* H atom in water; r_0=1.5 Å; m=8; n=16
CN O^1-H	$CN\ O^1-H=\sum_{i\in H_w}\frac{1-(\frac{r_{O^1,\ i}}{r_0})^m}{1-(\frac{r_{O^1,\ i}}{r_0})^n}$	$r_{O^1,\ i}$: distance between atom O near Fe in *O_2 and i-*th* H atom in water; r_0=1.5 Å; m=8; n=16
CN O^1-O^2	$CN\ O^1-O^2=\frac{1-(\frac{r_{O^1,\ O^2}}{r_0})^m}{1-(\frac{r_{O^1,\ O^2}}{r_0})^n}$	$r_{O^1,\ O^2}$: distance between O atoms in *O_2; r_0=1.8 Å; m=6; n=12
CN Fe-O^1	$CN\ Fe-O^1=\frac{1-(\frac{r_{Fe,\ O^1}}{r_0})^m}{1-(\frac{r_{Fe,\ O^1}}{r_0})^n}$	$r_{Fe,\ O^1}$: distance between Fe atom and O atom near Fe in *O_2; r_0=2.2 Å; m=6; n=12

Table S3. Coordination numbers of different reference structures in ORR path.

CV	*O_2	*OOH	*O--OH	*O--H_2O	*O	*OH	*H_2O
CN O^2-H	0.0	1.0	1.3	2.0	2.0	2.0	2.0
CN O^1-H	0.0	0.0	0.0	0.0	1.0	2.0	2.0
CN O^1-O^2	0.9	0.8	0.3	0.3	0.0	0.0	0.0
CN Fe-O^1	0.9	0.9	0.9	0.9	0.9	0.9	0.0

S3 Supporting Figures

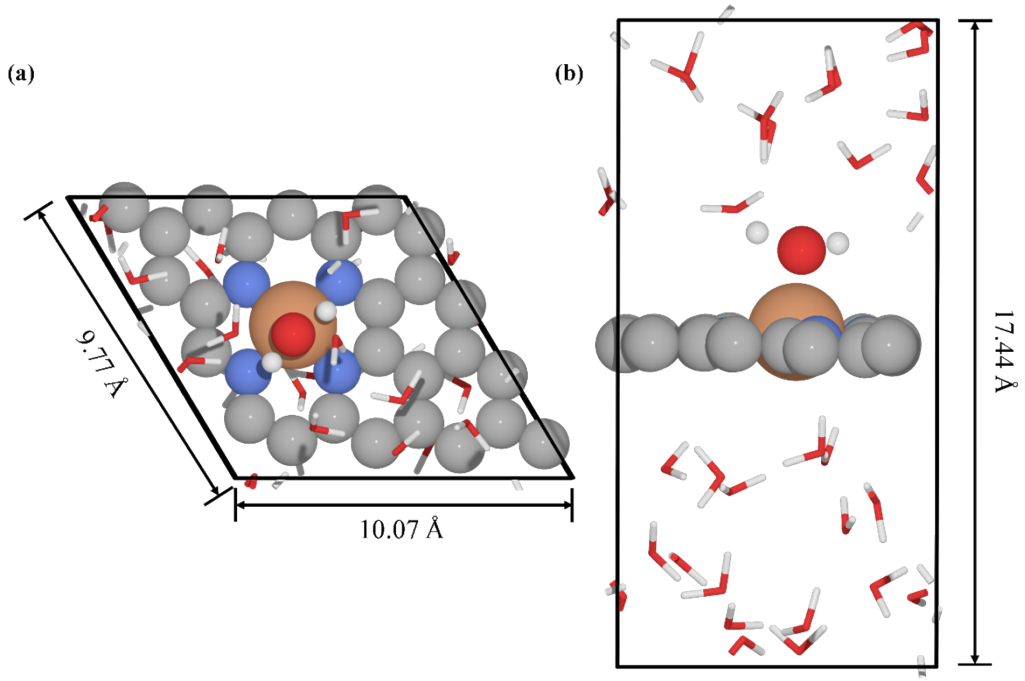

Figure S1. Dimensions and relative positions of configuration. (a) Top view of the configuration. Fe-N_4/C catalyst modeled by a layer of graphene with Fe-N_4 embedded in the center. (b) Side view of the configuration. The catalyst is located in the middle of the solution.

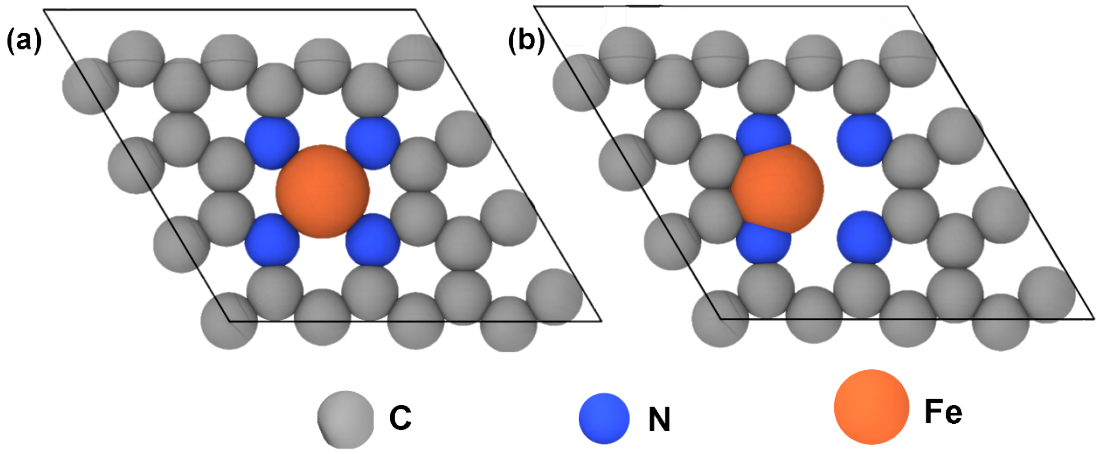

Figure S2. Configuration of Fe with different coordinates. (a) Fe atom with 4 N coordinated. (b) Configuration of Fe with 2 N coordinated. This configuration (2 N coordinated) originated from early active learning and helped MLFF determine the correct potential energy surface boundary.

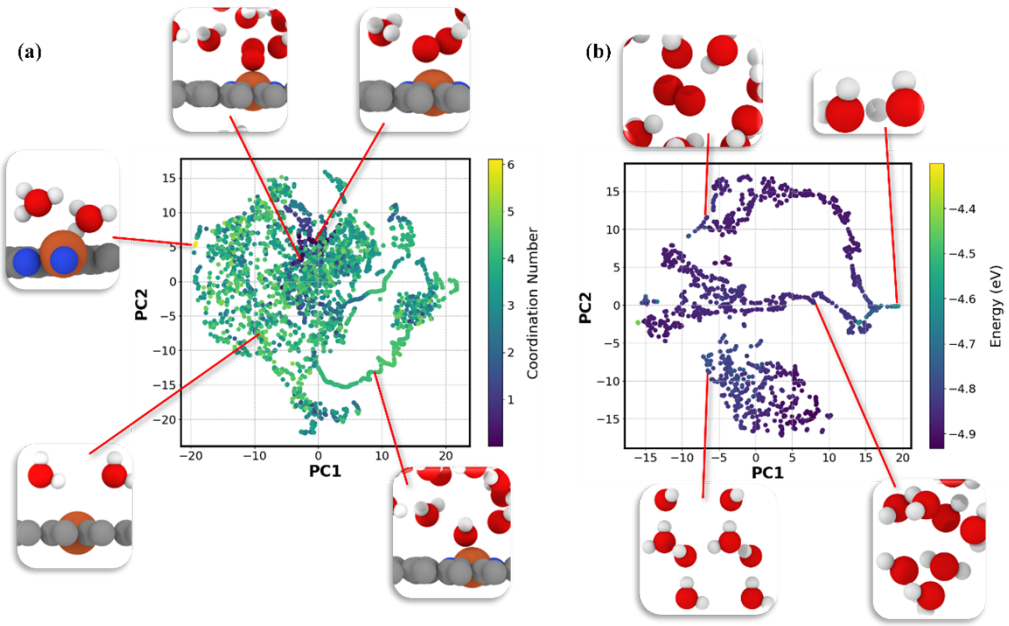

Figure S3. Visualization of the t-distributed stochastic neighbor embedding (t-SNE) results for various structures. (a) t-SNE results of all catalyst-water configurations, colored by the coordination number between adsorbed O_2 and H. (b) t-SNE results of all water, ice and water-O_2 configurations, colored by the average atomic energy.

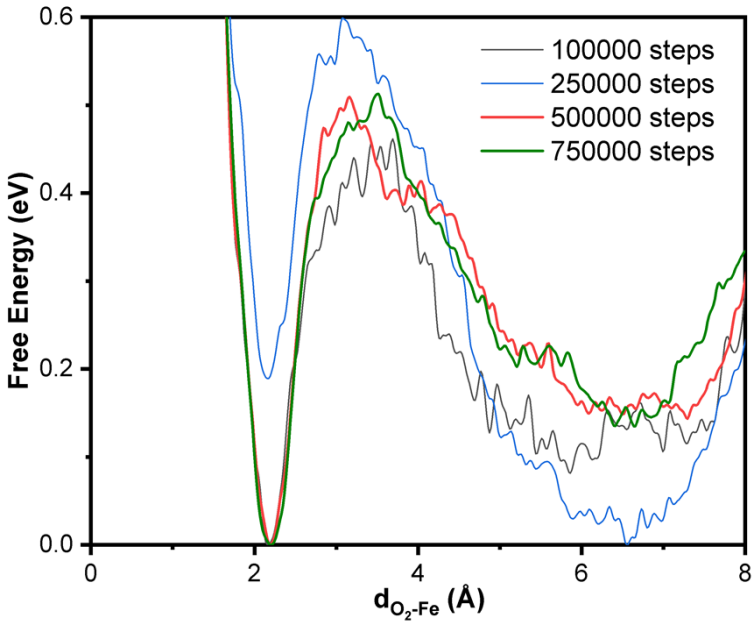

Figure S4. Comparison of free energy at different steps for O_2 adsorption. As the simulation steps is above 500000, the free energy no longer changes significantly.

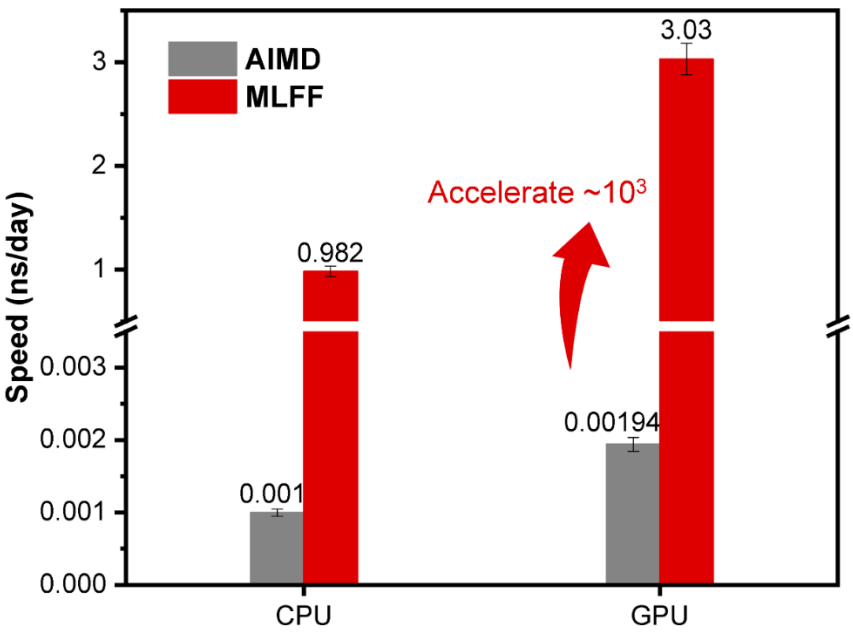

Figure S5. Comparison of computational speeds between AIMD and MLFF-MD. There were 124 atoms in the evaluation system, and MD simulation was executed without metadynamics.

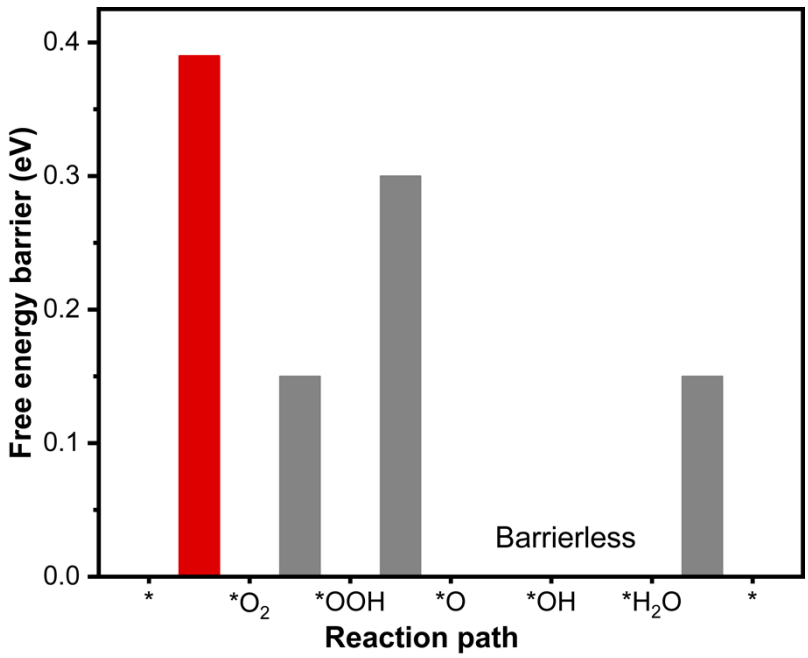

Figure S6. Free energy barriers along the reaction pathway. The highest energy barrier is the oxygen adsorption process, which is the rate determining step. There is no barrier no of $*O + H^+ + e^- \rightarrow *OH$ and $*OH + H^+ + e^- \rightarrow * + H_2O$ process.

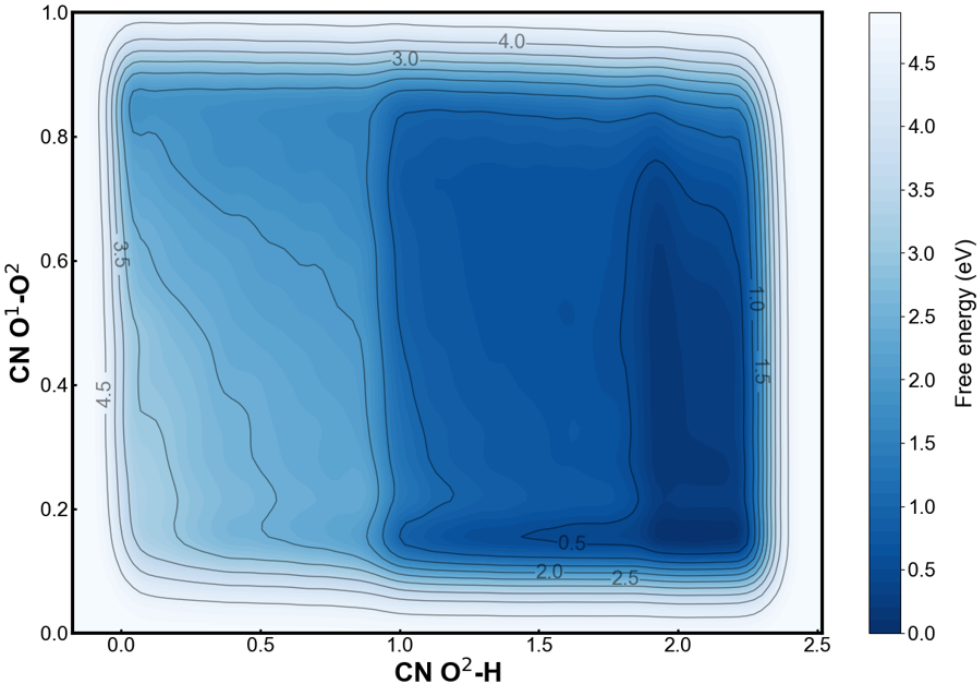

Figure S7. Free energy profile for *O + $2H^+$ + $2e^-$ → * + H_2O. The free energy was calculated by MLFF-MD with Metadynamics. The CN O^1-H and CN Fe-O^1 were employed as collective variables, where CN O^1-H meant coordination number between the O atom near Fe in *O_2 (O^2) and all H atoms, CN Fe-O^1 meant coordination number between the O atom near Fe(O^1) and Fe atom.

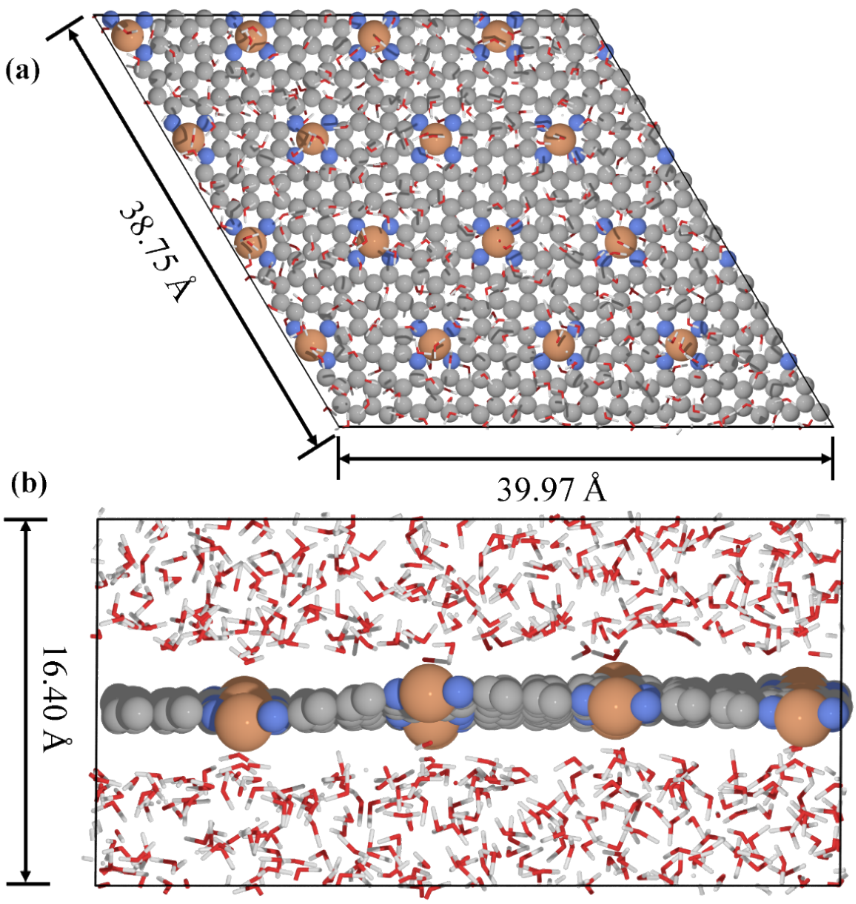

Figure S8. Dimensions and relative positions of 4*4*1 extended model. (a) Top view of the configuration. Fe-N_4/C catalyst modeled by a layer of graphene with Fe-N_4 embedded in the center. (b) Side view of the configuration. The catalyst is located in the middle of the solution. There are 510 water molecules, totally 1984 atoms.

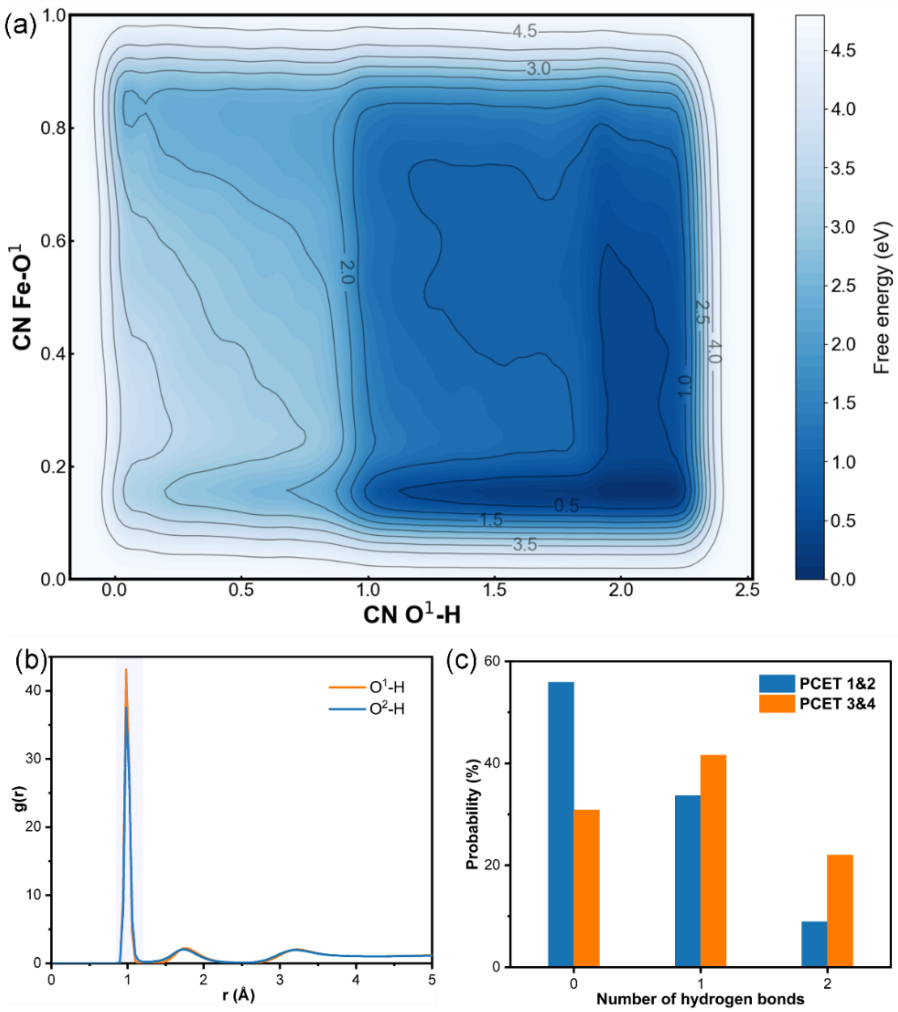

Figure S9. Free energy and interface microenvironment for different PCET processes based on extended model. (a) Free energy profile for $*O + 2H^+ + 2e^- \rightarrow * + H_2O$. The CN O^1-H and CN Fe-O^1 were employed as collective variables, where CN O^1-H meant coordination number between the O atom near Fe in $*O_2$ (O^2) and all H atoms, CN Fe-O^1 meant coordination number between the O atom near Fe(O^1) and Fe atom. (b) Radial distribution function between O^2-H and O^1-H, where O^2 represents the O atom far away from Fe in $*O_2$ and O^1 represents the O atom near Fe. The range of the first peak is marked with a purple band. (c) Probability distribution of the number of hydrogen bonds.

Reference:

1. J. Hafner, Ab-initio simulations of materials using VASP: Density-functional theory and beyond, *J. Comput. Chem.*, 2008, **29**, 2044-2078.
2. J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1996, **77**, 3865.
3. F. Fu, X. Wang, L. Zhang, Y. Yang, J. Chen, B. Xu, C. Ouyang, S. Xu, F. Z. Dai and W. E, Unraveling the Atomic-scale Mechanism of Phase Transformations and Structural Evolutions during (de) Lithiation in Si Anodes, *Adv. Funct. Mater.*, 2023, **33**, 2303936.
4. S. Grimme, J. Antony, S. Ehrlich and 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.
5. J.-W. Chen, Z. Zhang, H.-M. Yan, G.-J. Xia, H. Cao and Y.-G. Wang, Pseudo-adsorption and long-range redox coupling during oxygen reduction reaction on single atom electrocatalyst, *Nat. Commun.*, 2022, **13**, 1734.
6. J. Zeng, D. Zhang, D. Lu, P. Mo, Z. Li, Y. Chen, M. Rynik, L. a. Huang, Z. Li and S. Shi, DeePMD-kit v2: A software package for deep potential models, *J. Chem. Phys.*, 2023, **159**, 054801.
7. H. Wang, L. Zhang, J. Han and E. Weinan, DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics, *Comput. Phys. Commun.*, 2018, **228**, 178-184.
8. A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. In't Veld, A. Kohlmeyer, S. G. Moore and T. D. Nguyen, LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, *Comput. Phys. Commun.*, 2022, **271**, 108171.
9. E. Neria, S. Fischer and M. Karplus, Simulation of activation free energies in molecular systems, *J. Chem. Phys.*, 1996, **105**, 1902-1921.
10. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, PLUMED 2: New feathers for an old bird, *Comput. Phys. Commun.*, 2014, **185**, 604-613.
11. D. Branduardi, F. L. Gervasio and M. Parrinello, From A to B in free energy space, *J. Chem. Phys.*, 2007, **126**, 054103.
12. F. Pietrucci and A. M. Saitta, Formamide reaction network in gas phase and solution via a unified theoretical approach: Toward a reconciliation of different prebiotic scenarios, *Proc. Natl. Acad. Sci.*, 2015, **112**, 15030-15035.
13. X. Yang, A. Bhowmik, T. Vegge and H. A. Hansen, Neural network potentials for accelerated metadynamics of oxygen reduction kinetics at Au–water interfaces, *Chem. Sci.*, 2023, **14**, 3913-3922.
14. M. Yang, L. Bonati, D. Polino and M. Parrinello, Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water, *Catal. Today*, 2022, **387**, 143-149.
15. E. Weinan, W. Ren and E. Vanden-Eijnden, String method for the study of rare

events, *Phys. Rev. B*, 2002, **66**, 052301.

16. W. Ren and E. Vanden-Eijnden, Simplified and improved string method for computing the minimum energy paths in barrier-crossing events, *J. Chem. Phys.*, 2007, **126**, 164103.
17. S. Chang, H. Wang, Y. Ji and Y. Li, Influence Factors of CO Adsorption on C_2N-Supported Dual-Atom Catalysts Unveiled by Machine Learning and Twofold Feature Engineering, *Phys. Chem. Chem. Phys.*, 2024, 9350-9355.
18. A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool, *Modell. Simul. Mater. Sci. Eng.*, 2009, **18**, 015012.
19. G. Claeskens, C. Croux and J. Van Kerckhoven, An information criterion for variable selection in support vector machines, *J. Mach. Learn. Res.*, 2008, **9**, 541-558.
20. A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer and C. Hargus, The atomic simulation environment—a Python library for working with atoms, *J. Phys.: Condens. Matter*, 2017, **29**, 273002.
21. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, MDAnalysis: a toolkit for the analysis of molecular dynamics simulations, *J. Comput. Chem.*, 2011, **32**, 2319-2327.
22. R. Jinnouchi, F. Karsai and G. Kresse, On-the-fly machine learning force field generation: Application to melting points, *Phys. Rev. B*, 2019, **100**, 014105.