Supplementary information: Online Test-time Adaptation for Better Generalization of Interatomic Potentials to Out-of-distribution Data

List of Tables

List of Figures

Hyperparameters

We show the detailed hyperparameters for different datasets in Table S1, Table S2, and Table S3. We employ the learning rate adjustment strategy as [1]. Initially, we warm up the learning rate, followed by utilizing StepLR learning rate strategies during training. Specifically, the learning rate is reduced by a decay ratio at every fixed epoch, which we refer to as the step size. For all experiments, some hyperparameters are set to fixed values without optimization, while others are optimized through grid search.

Table S1 Search space for hyperparameters on MD17.

Hyperparameters	Search space
Number of interaction blocks (SchNet)	6
Number of interaction blocks (PaiNN)	3,4
Cutoff distance	5.0, 6.0, 10.0
Batch size	1, 2, 4, 16, 32
Initial learning rate	1e-3, 5e-4, 1e-4
Learning rate strategy	StepLR
Learning rate decay ratio	0.4, 0.5, 0.6
Learning rate step size	50, 100, 200
Learning rate warmup factor	0.7
Max noisy intensity T	50
Mask ratio	0.15
Scaling factor γ	1, 2, 5
Max # of Epochs	500, 1000, 2000

Table S2 Search space for hyperparameters on ISO17.

Hyperparameters	Search space
Number of interaction blocks (SchNet)	5,6
Number of interaction blocks (PaiNN)	3,4
Cutoff distance	5.0, 6.0
Batch size	32, 64, 128
Initial learning rate	1e-3, 1e-4, 5e-5
Minimum learning rate	1e-7
Learning rate cosine decay ratio	0.0, 0.1, 0.2
Learning rate step size	50, 100, 200
Learning rate warmup factor	0.7
Max noisy intensity T	50
Mask ratio	0.15
Scaling factor γ	1, 2, 5
Max # of Epochs	50

Table S3 Search space for hyperparameters on Electrolyte/Liquid water.

Hyperparameters	Search space
Number of interaction blocks (SchNet)	5, 6
Number of interaction blocks (PaiNN)	3, 4, 5
Batch size	1, 2, 4, 8
Initial learning rate	1e-3, 5e-4, 1e-3
Learning rate strategy	StepLR
Learning rate decay ratio	0.4, 0.5, 0.6
Learning rate step size	50, 100, 200
Learning rate warmup factor	0.7
Max noisy intensity T	50
Mask ratio	0.15
Scaling factor γ	1, 2, 5
Max # of Epochs	500, 1000, 2000

Additional Results

Table S4 Performance of TAIP trained on varying amounts of data from ISO17 dataset. We randomly sampled 1%, 5%, and 30% of the training data from ISO17 dataset to evaluate the performance of TAIP on the same test set. Energy and force MAEs are reported in units of kcal/mol and kcal/mol/Å, respectively. Results of TAIP-based models are shown in bold.

		SchNet	SchNet-TAIP	PaiNN	PaiNN-TAIP
1%	Energy	97.10	**39.51**	88.74	**22.31**
	Force	4.52	**3.93**	3.39	**2.73**
5%	Energy	23.83	**13.22**	18.34	**11.34**
	Force	3.33	**2.89**	2.46	**1.68**
30%	Energy	11.18	**4.99**	3.93	**1.32**
	Force	2.60	**2.07**	1.29	**0.87**
100%	Energy	2.40	**1.13**	0.92	**0.66**
	Force	2.18	**1.78**	0.89	**0.68**

Table S5 Ablation study on liquid water dataset. Energy and force MAEs are reported in unit of eV and eV/Å, respectively. T1 indicates the noise intensity prediction task, T2 indicates the atom feature recover task, T3 indicates the pseudo force recover task, and TAIP indicates all the 3 tasks are involved.

	PaiNN	PaiNN-T1	PaiNN-T2	PaiNN-T1+T2	PaiNN-T2+T3	PaiNN-TAIP
Energy	1.193	0.574	0.871	0.563	0.549	**0.423**
Force	0.083	0.065	0.078	0.062	0.060	**0.055**

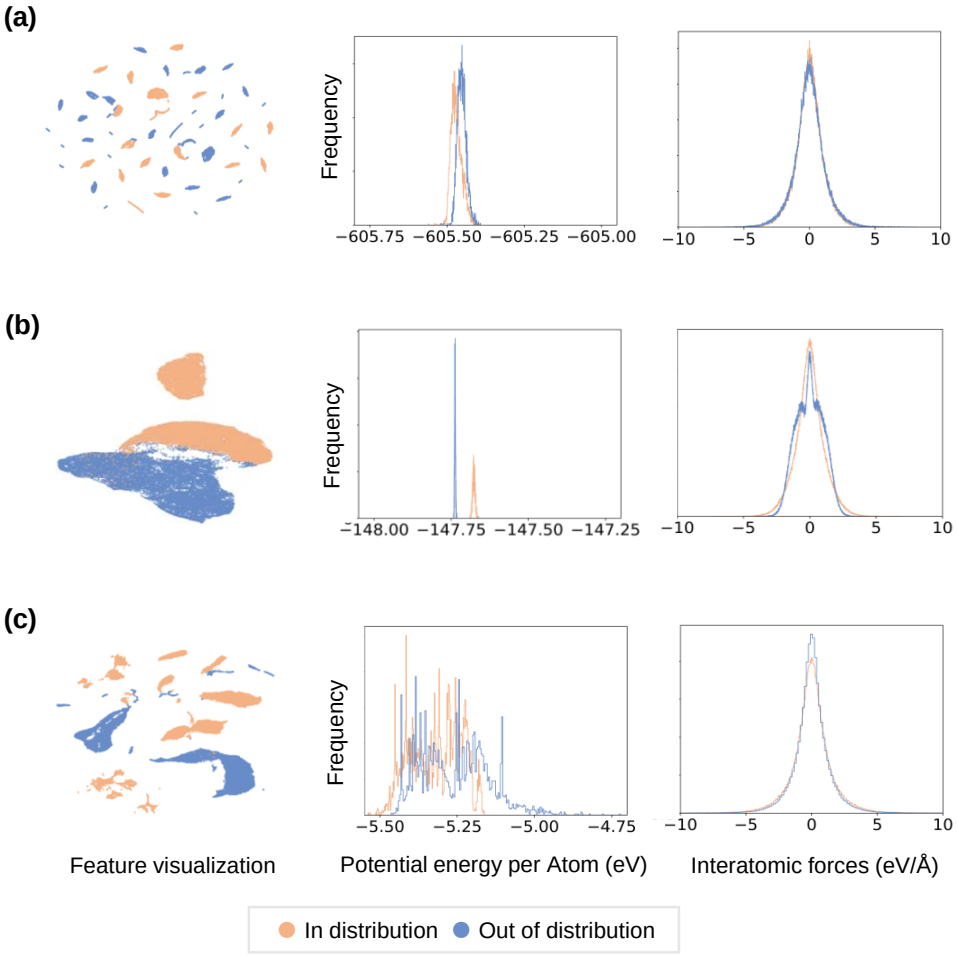

Fig. S1 Distribution shift characterization. (a) ISO17. (b) Water. (c) Electrolyte solutions. Left: feature visualization. The representation of atom features are generated by UMAP dimension reduction method on SOAP descriptors. Middle: distribution of potential energy. Right: distribution of force.

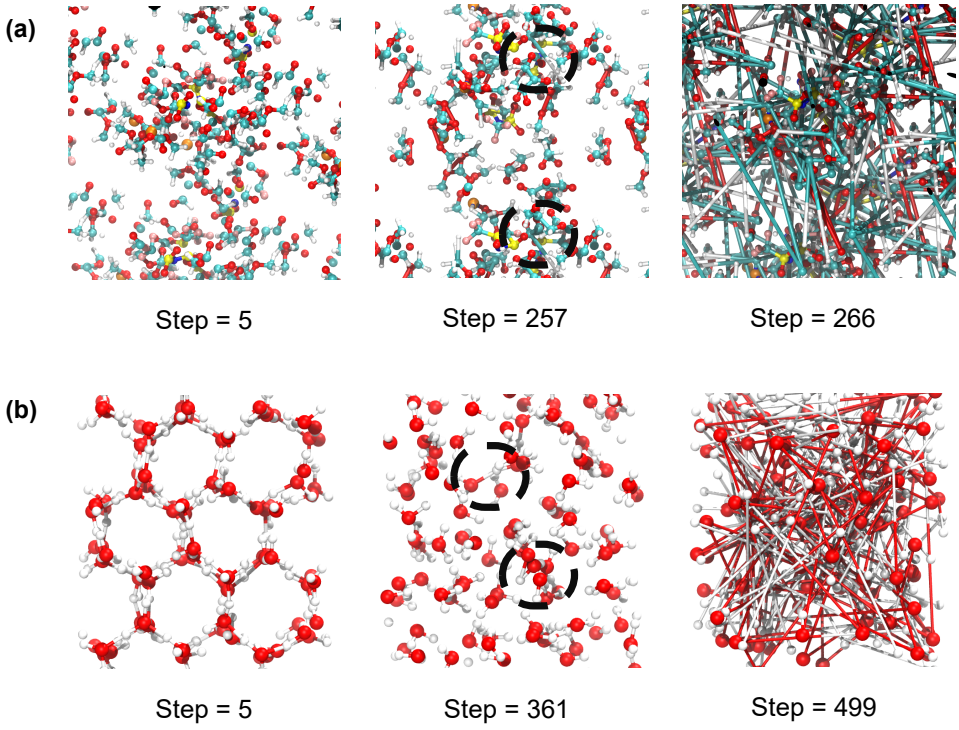

Fig. S2 Trajectories of electrolyte solution and Water. The Snapshots of the MD trajectories of (a) electrolyte solution (1M $[Na^+][Tf_2N^-]$ in EC + DMC) and (b) ice using a schematic baseline PaiNN model trained on a small dataset at 300K. Occurrences of unrealistic molecular configurations (circled by black dashed lines) will cause a collapse during MD simulations using MLIPs.

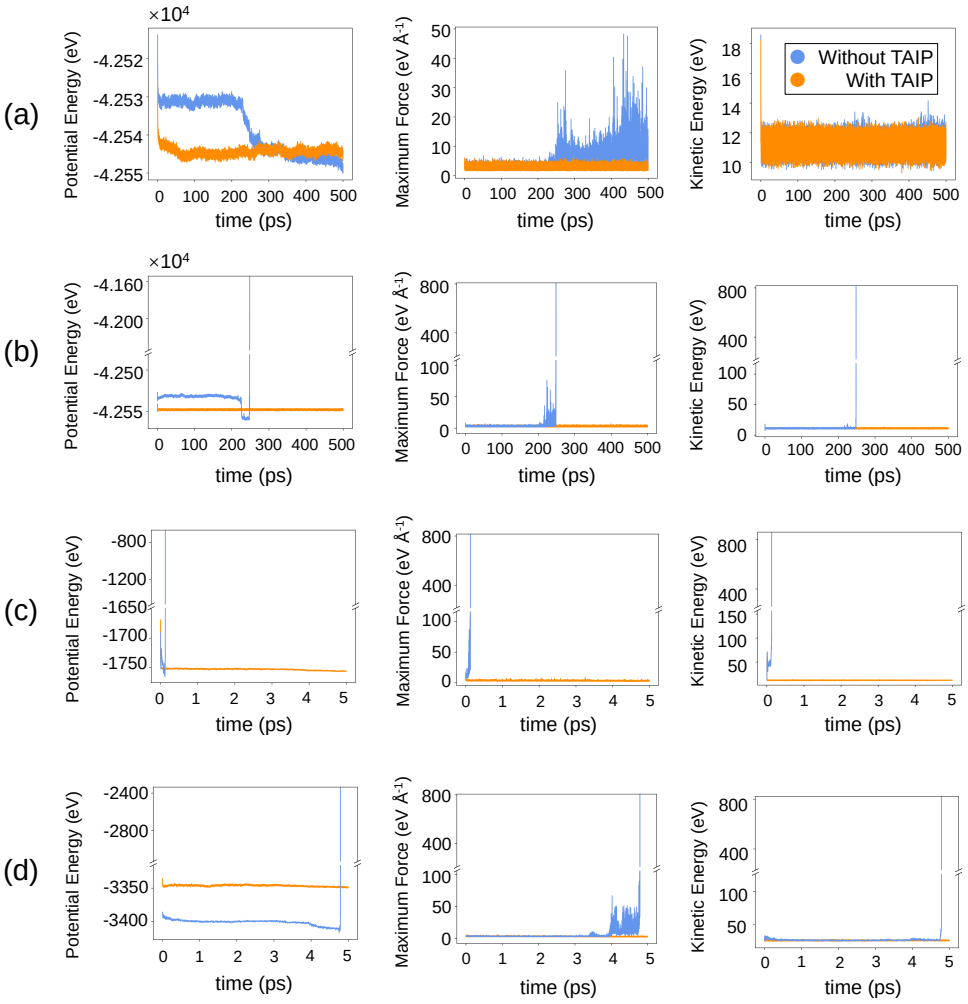

Fig. S3 Results of MD simulations using PaiNN and PaiNN-TAIP. (a) Liquid water. (b) ice. (c) 1M electrolyte solutions. (d) 4M electrolyte solutions. The MD simulations with baseline PaiNNs (blue) occur to breakdown, whereas those with TAIP (yellow) are stable throughout the simulation time.

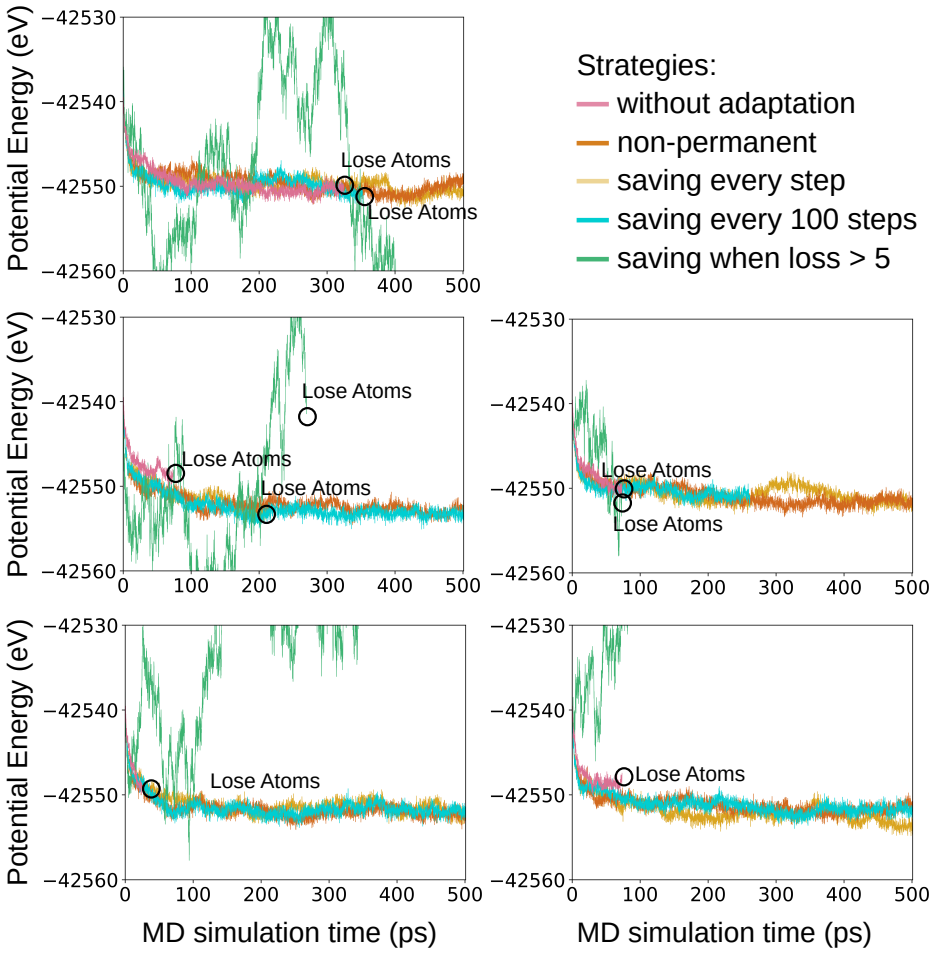

Fig. S4 MD simulations with different test-time adaptation strategies. We implement different adaptation strategies for the MD simulations, including simulate without any adaptation, with non-saving adaptation, with saved adaptation every simulation step, with saved adaptation every 100 simulation steps, and with saved adaptation when the loss of the simulation step given by self-supervised learning tasks is larger than a threshold (here, we define it to be 5, which is approximately the average loss in the liquid water test dataset). The experiments are conducted on liquid water using SchNet-TAIP model at the temperature of 300K with five randomly selected initial configurations. We can observe that the simulations with the simplest non-saving adaptation strategy are quite stable.

References

[1] Cui, T., Tang, C., Su, M., Zhang, S., Li, Y., Bai, L., Dong, Y., Gong, X., Ouyang, W.: Geometry-enhanced pretraining on interatomic potentials. Nature Machine Intelligence, 1–9 (2024)