

General reactive element-based machine learning potentials for heterogeneous catalysis

In the format provided by the
authors and unedited

Table of Contents

Supplementary Notes

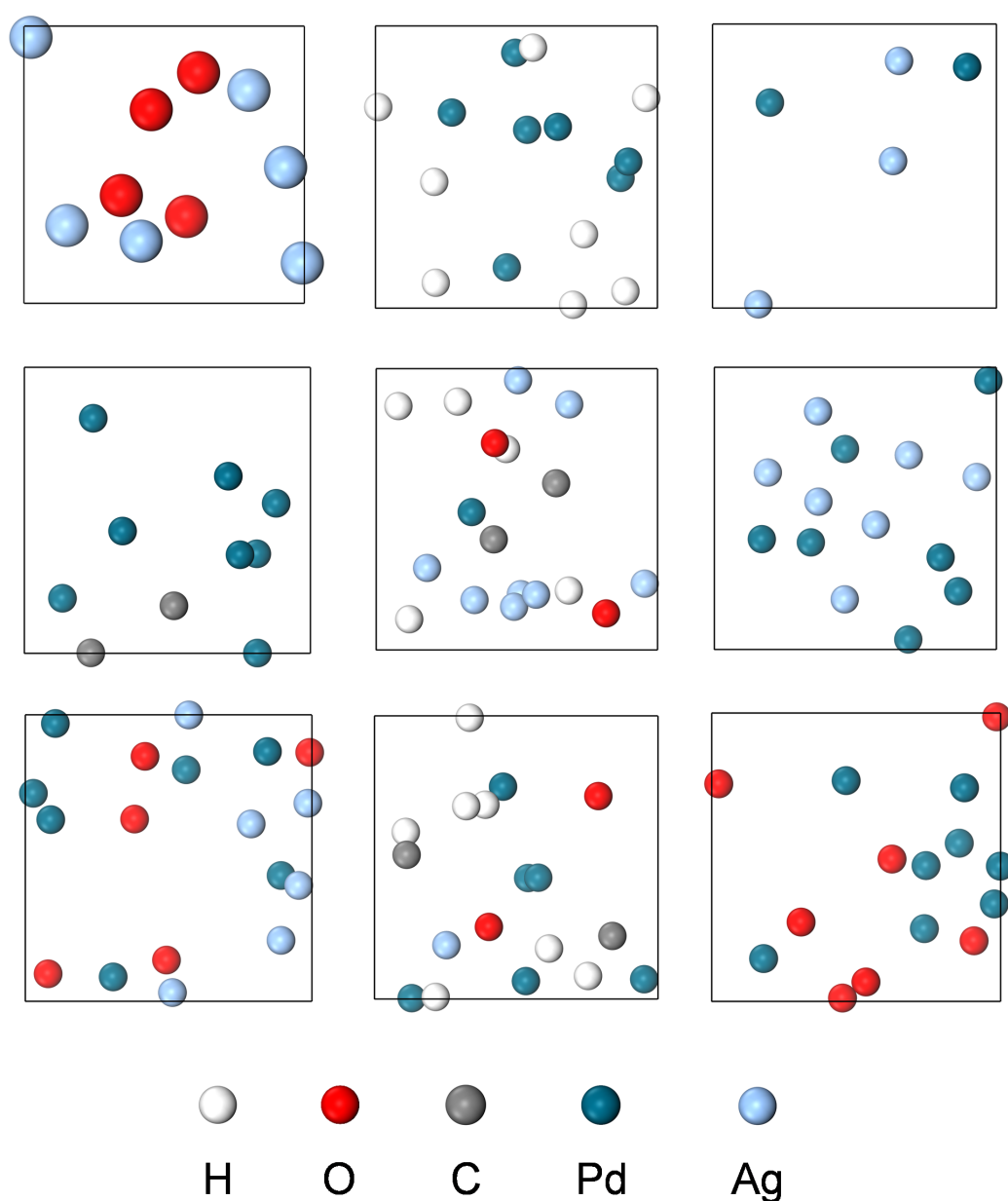
Supplementary Note 1. The Ag-Pd-C-H-O EMLP	2
Supplementary Note 2. The MD-MLP	7
Supplementary Note 3. Data coverage	9
Supplementary Note 4. Generality test results	11
Supplementary Note 5. Energy validation of free cluster MD	12
Supplementary Note 6. Ethylene epoxidation on Ag(100).....	14
Supplementary Note 7. CO oxidation on a range of Pd/Ag-based surfaces.....	16
Supplementary Note 8. Acetylene hydrogenation on Pd(111), H-covered Pd(111), and PdAg(111).....	18
Supplementary Note 9. Fischer–Tropsch (FT) process	20
Supplementary Note 10. Standard test reactions	22
Supplementary Note 11. The dynamics of H ₂ adsorption and dissociation on the Ag(100).	23
Supplementary Note 12. Liquid methanol simulation.....	24
Supplementary Note 13. Lattice constant.....	25
Supplementary Note 14. Global optimization of Ag-AgO	26
Supplementary Note 15. Relaxation of rattled unlearn structures.....	27
Supplementary Note 16. Melting temperature MD	28
Supplementary Note 17. Data convergence.....	29
Supplementary Note 18. Computational cost of training the AgPdCHO EMLP	32
Supplementary Note 19. Spin Test	33

Supplementary References	33
---------------------------------------	-----------

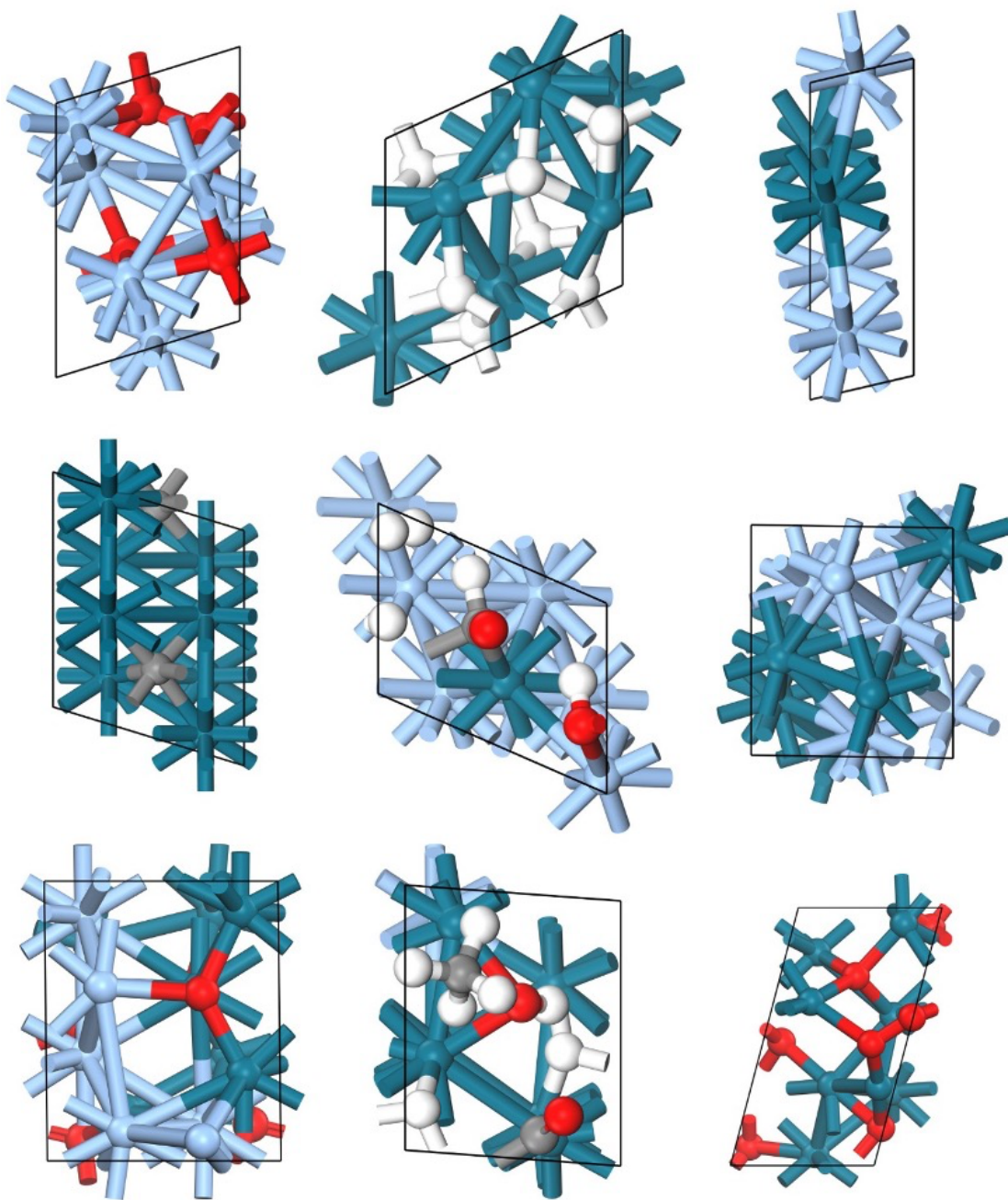
Supplementary Note 1. The Ag-Pd-C-H-O EMLP

(1) Examples of structures generated by Random Exploration via Imaginary Chemicals Optimization (REICO) method

This section shows structures generated by REICO. As shown in **Supplementary Figure 1**, REICO generates structures unbiased by specifying the kind of elements. We then optimized these structures including cell and atom positions as shown in **Supplementary Figure 2**. Since the initial structures are extremely unstable, many unexpected structures are included in the optimized trajectories.



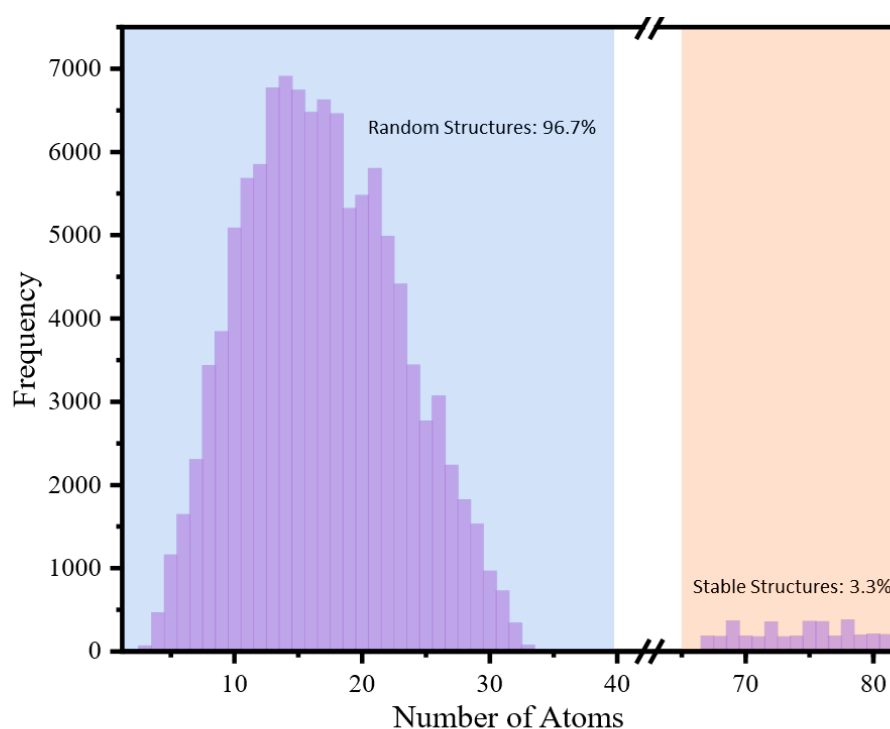
Supplementary Figure 1. Example structures. The structure generated by REICO.



Supplementary Figure 2. Example structures. The final optimized structures generated by REICO.

(2) Ag-Pd-C-H-O EMLP training set composition

Most of the datasets used in this work comprise structures generated by REICO, with the number of atoms less than 35. This range was selected to ensure the efficiency of density functional theory (DFT) calculations. The stabilized structures consist of four layers of 4×4 PdAg slabs adsorbed by various carbon-based small molecules. Structural diversity was enhanced through the random replacement of metal atoms within the slabs and the application of the Basin Hopping strategy. As shown in **Supplementary Figure 3**, the blue and yellow regions represent the random structures and the stabilized structures, respectively. The random structures constitute 96.7% of the total.

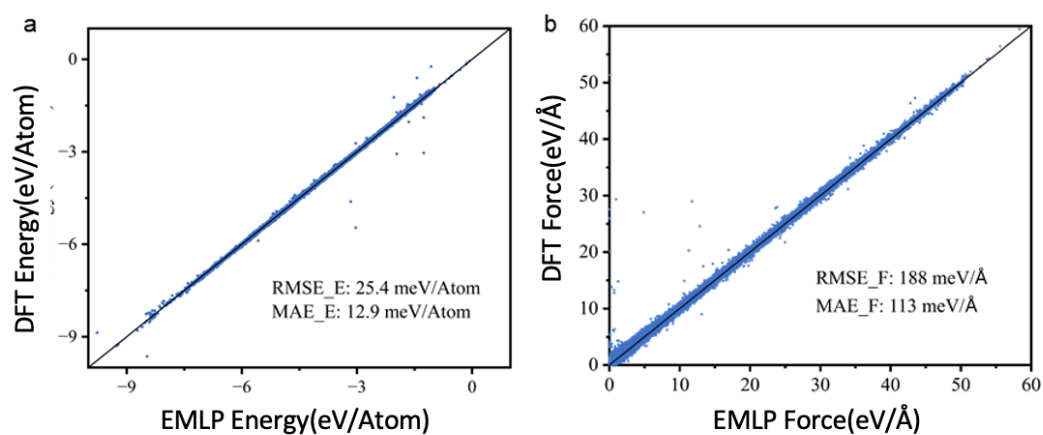


Supplementary Figure 3. Structure size distribution in the AgPdCHO EMLP training set. Histogram of the number of atoms in the Ag-Pd-C-H-O EMLP training dataset. The y-axis shows the frequency (counts), and the x-axis gives the number of atoms per structure.

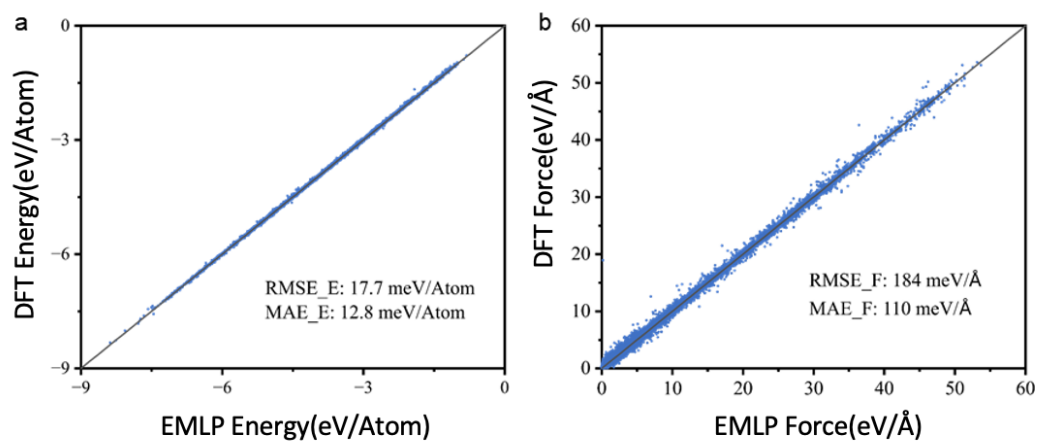
(3) Settings and hyperparameters of the Ag-Pd-C-H-O EMLP model

In this study, NequIP¹ was utilized to fit the EMLP, which comprises a total of 452,280 weights. The cutoff radius is set as 5 Å. In the equivariant feature representations, the maximum rotation order employed in this study is 2, with a total of 32 features. This configuration ensures that the diverse structures present in the dataset are adequately described. During training, the Adam optimizer was employed with an initial learning rate of 0.005 and a batch size of 128 to minimize the error of energy and force on the training set. The validation loss was used as the metric for learning rate decay. When the learning rate fell below 10^{-5} , the Neural Network Potential (NNP) was deemed to have reached convergence, and the training was terminated.

(4) The performances of Ag-Pd-C-H-O EMLP model



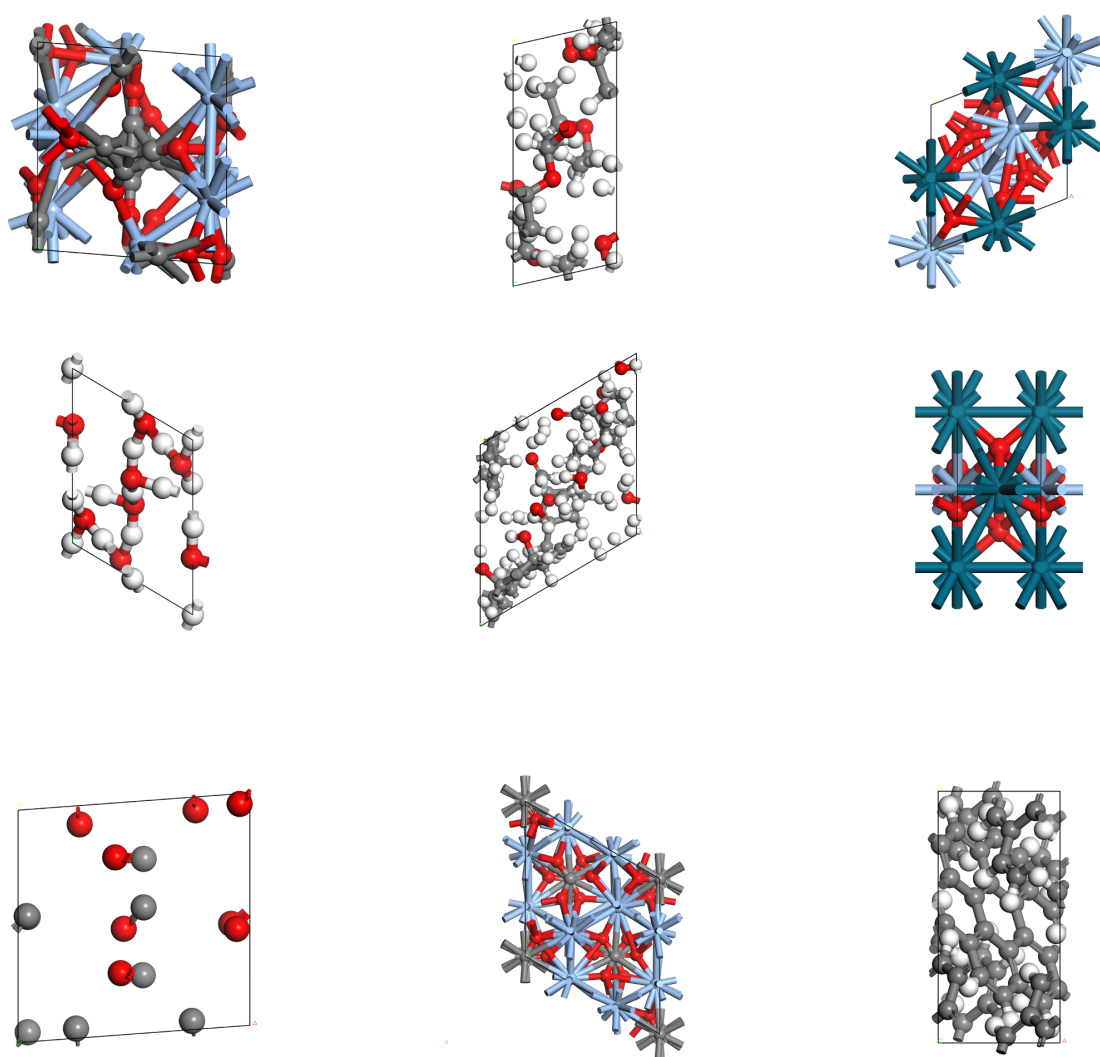
Supplementary Figure 4. Energy and force correlation plots for the training set.



Supplementary Figure 5. Energy and force correlation plots for the validation set.

Supplementary Note 2. The MD-MLP

The dataset of MD-MLP contains 50,599 structures and the initial structures are selected from the M3GNet database² that contained combinations of the Ag, Pd, C, H, and O elements. Each structure underwent MD calculations at 800 K and then one structure was selected from every 40 MD steps to form the dataset. Examples of the initial structures are shown in **Supplementary Figure 6**, and the training set composition is listed in **Supplementary Table 1**. The MD-MLP is trained with the same hyperparameters as EMLP. Its performance on the validation set has a RMSE of 0.248 meV/Atom and 53.9 meV/Å for energy and force, respectively.



Supplementary Figure 6. Example structures. The example structures selected from M3GNet database² that contained combinations of the Ag, Pd, C, H, and O elements.

Supplementary Table 1. The dataset of MD-MLP. The structures with same atom composition are combined in the table.

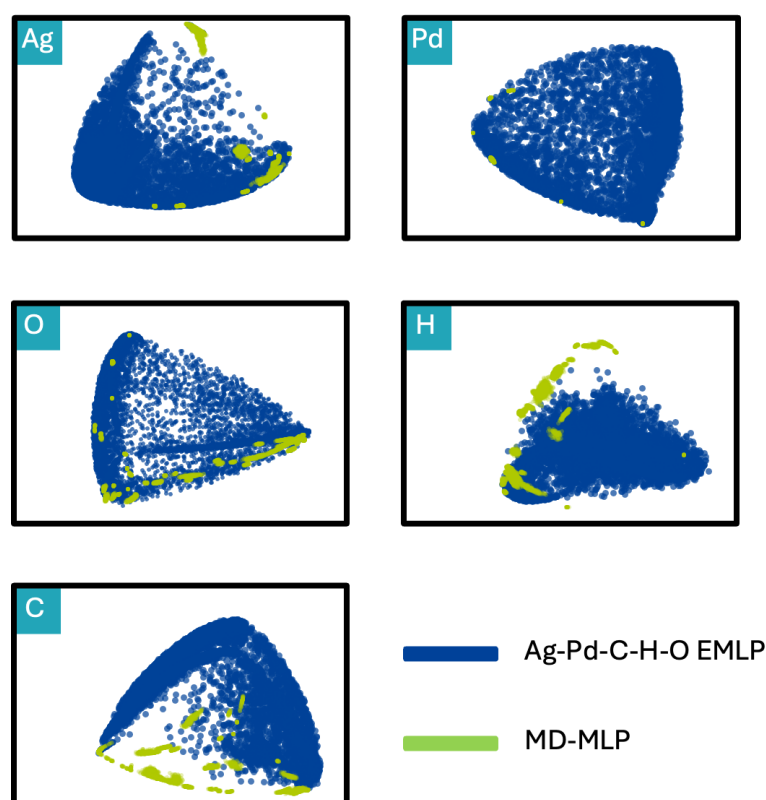
composition	no. of data	composition	no. of data	composition	no. of data
C ₄ O ₈	2004	O ₂ Pd ₄	501	C ₄₀ H ₈₈ O ₁₂	1002
H ₂₈ O ₁₄	501	C ₁₀ H ₈ O ₄	501	C ₆ H ₂	501
C ₃₈ H ₆₈	1002	C ₄ H ₈	501	Ag ₆ O ₂	501
H ₁₁ O ₈	501	C ₄ Ag ₄ O ₈	501	H ₁₆ O ₈	500
C ₂₈ H ₁₆ O ₄	501	Ag ₈ O ₄	501	C ₁₆ H ₈ O ₁₆	501
C ₁₂ H ₃₂ O ₈	501	Ag ₄ O ₁₂ Pd ₄	501	C ₁₂ H ₈ O ₁₆	501
C ₈	2505	C ₈ O ₈	501	C ₆ Ag ₁₂ O ₁₈	1002
C ₁₄₀	501	C ₆ H ₁₂ O ₆	501	Ag ₂ O ₄ Pd ₂	501
C ₄ H ₂ O ₄	501	C ₁₂₀	501	C ₈ H ₄ O ₈	501
H ₂ Pd ₆	501	Ag ₈ O ₁₂	501	H ₄ O ₂	501
H ₈ O ₈	1002	H ₂₄ Ag ₈ O ₁₆	501	Ag ₄ O ₂	501
Ag ₄ O ₆	1002	C ₁₂ H ₂	501	C ₁₆ Ag ₈ O ₁₆	1002
C ₆₀	1502	C ₂ Ag ₄ O ₆	501	H ₆₄ Ag ₈ O ₃₆	501
C ₄ O ₄	1002	H ₃₂ Ag ₄ O ₂₀	501	C ₂ H ₄ O ₆	501
H ₈ O ₄	2505	C ₁₂	1002	H ₁₄ O ₈	501
H ₅₆ Ag ₄ O ₃₂	501	Ag ₆ O ₃	501	H ₃₆ Ag ₄ O ₂₀	501
Ag ₆ O ₈	501	C ₄₈ O ₃₆	501	C ₄₈ H ₂₄	501
O ₂₄	501	C ₈ H ₈	501	C ₁₆ H ₃₂ O ₈	501
C ₂ Ag ₂ O ₆	501	C ₂₄ O ₁₆	501	C ₂₀ H ₄₀ O ₂₀	501
H ₂₄ O ₁₂	3006	Ag ₈ O ₈	501	O ₄₈ Pd ₄₂	501
C ₂ O ₆	501	C ₄ H ₂	501	C ₂ O ₄	501
C ₈ O ₁₆	501	O ₄ Pd ₂	501	C ₈ O ₁₆	501
C ₄ H ₈ O ₈	1503	H ₂ Ag ₂ O ₄	501		

Supplementary Note 3. Data coverage

After the datasets finally established, two-dimensional visualizations of the local atomic environments present in REICO and MD benchmark dataset (**Supplementary Figure 7**) are generated by kernel principal component analysis (kernel PCA³) implemented in scikit-learn package to achieve non-linear dimensionality reduction through the use of kernels⁴. The SOAP kernel⁵:

$$K^{\text{SOAP}}(\mathbf{x}, \mathbf{y}) = \left(\frac{\mathbf{x} \cdot \mathbf{y}}{\sqrt{(\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y})}} \right)^{\zeta} \quad (1)$$

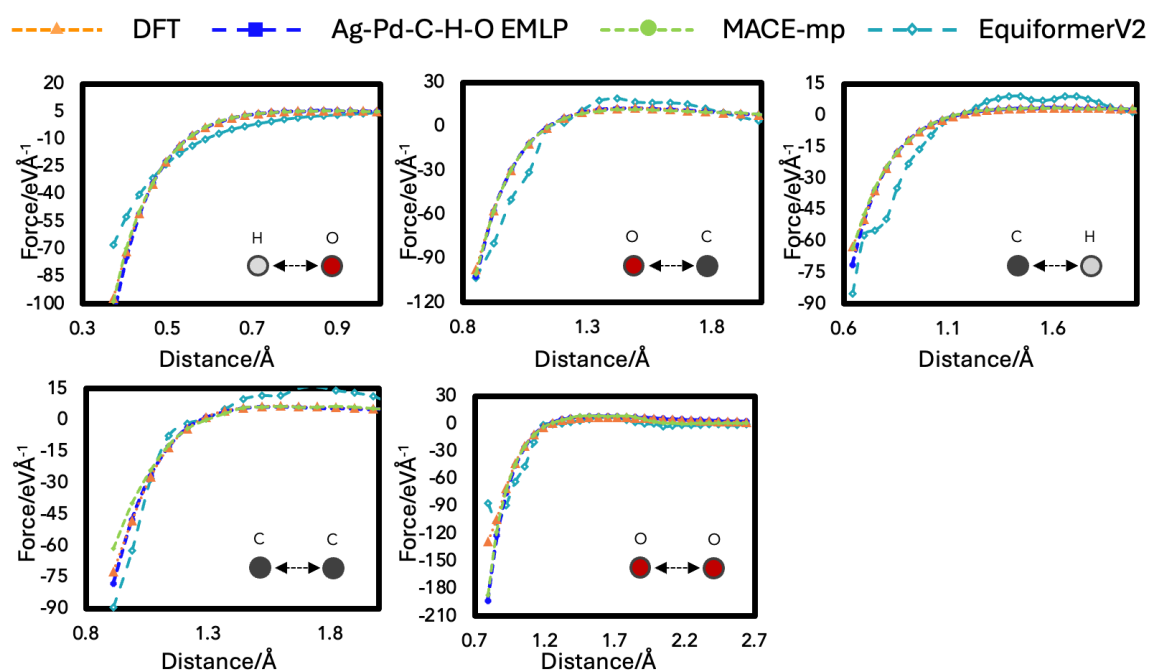
is used to measure the distance between atomic environments $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are partial power spectrum vectors of SOAP and ζ is any positive integer (here, ζ is 1). For visual clarity and avoiding out of memory, only a random subset (10,000 local atomic environments) of each element is sampled from REICO and MD benchmark dataset respectively. Then SOAP descriptors of these atomic environments are generated by the implementation from DScibe package⁶, which serve as the input of kernel PCA.



Supplementary Figure 7. Data coverage analysis. The dataset generated by REICO(blue) in this work compared to a MD generated dataset(green) with structures from public database. Two-dimensional PCA visualizations using the SOAP kernel illustrate the diversity and coverage of local atomic environments within datasets generated by REICO (blue) and MD

(green). Each figure includes 10,000 randomly selected local atomic environments featuring each element of Ag, Pd, C, H, and O. The visualizations highlight that the REICO dataset not only covers most regions sampled by the MD dataset but also fills gaps between these regions and expands beyond them substantially.

Supplementary Note 4. Generality test results

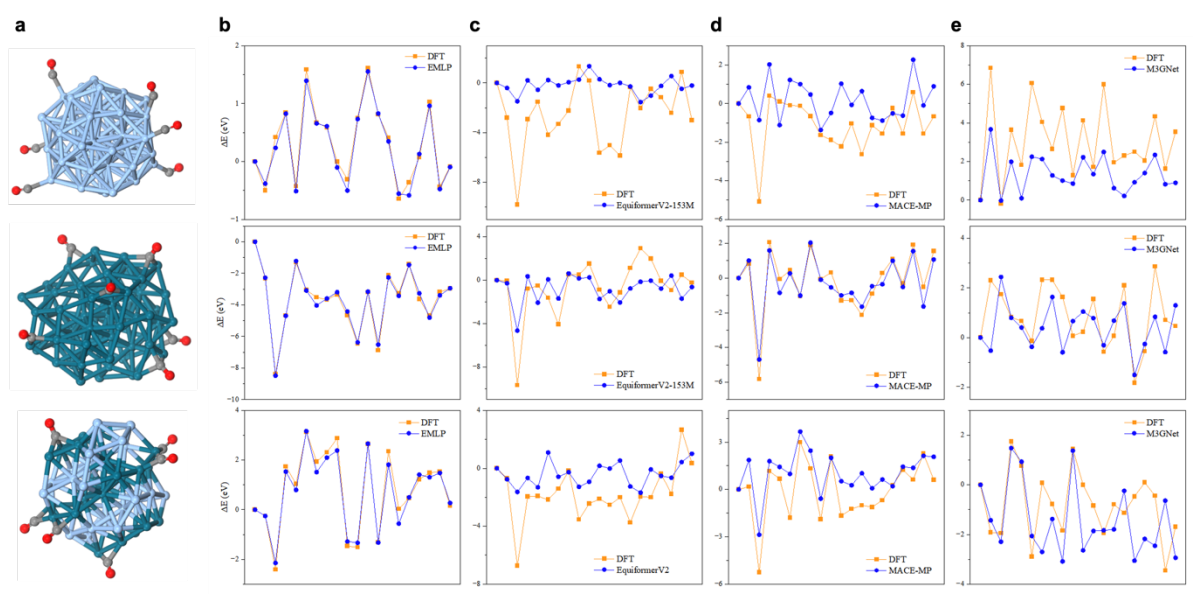


Supplementary Figure 8. The rest of generality tests conducted through rigid dimer scans.

The tests involve five dimer combinations of self and cross interactions between the elements Ag, Pd, C, H, and O. Results from DFT calculations serve as the benchmark, and are compared with predictions from the EMLP (blue dashes), MACE-mp, and EquiformerV2.

Supplementary Note 5. Energy vailadaiton of free cluster MD

To investigate EMLP's ability to capture subtle energetic changes in finite systems—a critical requirement for studying catalyst behavior under reaction conditions—we performed MD simulations on three distinct cluster models: a pure Ag cluster, a pure Pd cluster, and a bimetallic Pd/Ag cluster. In each of these clusters, we adsorbed six CO molecules to introduce additional complexity and test how well EMLP and other models can handle adsorbate–cluster interactions, shown in **Supplementary Figure 9a**.



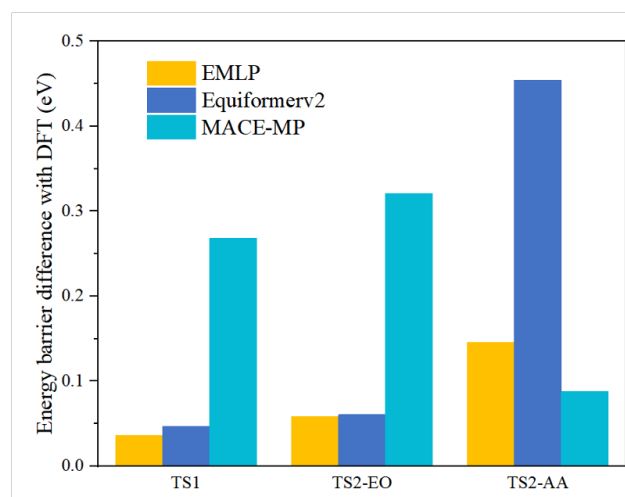
Supplementary Figure 9. Energy prediction errors on random MD snapshots. (a) Structures of Pd, Ag, and Pd/Ag clusters with COs used in the MD. (b–e) Absolute energy deviations ΔE (eV) relative to DFT for (b) EMLP, (c) EquiformerV2, (d) MACE-mp, and (e) M3GNet.

During each MD run, we randomly extracted 20 snapshots to capture a representative range of configurations. For every snapshot, we recalculated the single-point energy with DFT and compared it against the energy predicted by EMLP. We then referenced all energies to the initial structure of each simulation for monitoring energy trends, a standard way to evaluate the thermodynamic accuracy of MLPs.⁷ As shown in **Supplementary Figure 9b**, EMLP consistently reproduces the DFT energy trends across the snapshots, demonstrating its ability to handle subtle structural rearrangements and energetic changes. In contrast, other models (MACE-mp, EquiformerV2, and M3GNet) show significant deviations for one or more of the tested cluster systems, shown is **Supplementary Figure 9c-e**. EquiformerV2 and M3GNet fail

to capture the correct energy variation trend for the Ag cluster, leading to large quantitative and even qualitative errors. MACE-mp is considered better overall but still exhibits instances where the predicted energy is opposite in sign to DFT.

Supplementary Note 6. Ethylene epoxidation on Ag(100)

Achieving reactivity in MLPs without targeted, model-aware sampling strategies presents a significant challenge in ensuring their generality. Ethylene epoxidation on Ag catalysts is the sole industrial method for producing ethene oxide (EO), with acetaldehyde (AA) being a common side product that typically undergoes full oxidation to CO₂ and water. Previous studies established that the selectivity for EO on silver surfaces is approximately 50%^{8,9}. This balance between two competing reaction pathways offers a robust test case for evaluating the accuracy and reactivity of our EMLP. We utilized the EMLP to model both reaction pathways of ethene epoxidation on the Ag(100) surface and validated the EMLP results with DFT calculations. As illustrated in **Supplementary Figure 10**, the EMLP effectively reproduced DFT results, identifying the rate-determining step in the reaction, pinpointing it at the transition state TS2 where the oxametallacycle intermediate transforms into either EO or AA. The comparison revealed an average discrepancy in energy barriers of merely 0.08 eV between the EMLP and DFT results. Moving to other models, all models can locate the TSs for the competing pathways. EquiformerV2 achieve an average energy deviation of less than 0.1 eV from DFT, whereas MACE-MP and M3GNet show larger errors (0.15–0.2 eV). Moreover, in terms of consistency (as measured by maximum deviation), EMLP remains far more reliable (0.13 eV) compared to 0.3–0.4 eV for the others.



Supplementary Figure 10. Model performances on Ethylene epoxidation on Ag(100). Reaction barrier difference of ethylene epoxidation on Ag catalysts between model predictions made by EMLP(yellow), EquiformerV2(blue), MACE-mp(teal), and M3Gnet compared with DFT.

Supplementary Table 2. The DFT vibrational frequency calculations of transition states found by different models. The transition state structures with very small or no imaginary frequency are highlighted in red.

	EMLP	MACE-mp	EquiformerV2
TS1	f/i=338.21	f/i=92.11	f/i=85.69
TS2-EO	f/i=359.69	f/i=84.59	f/i=299.69
TS2-AA	f/i=845.86	f/i=1289.79	f/i=660.46

Supplementary Table 3. The structure similarity check between transition states found by DFT and different models. The predicated structure that is closest to the DFT result is highlighted in bold.

	EMLP	MACE-mp	EquiformerV2
TS1	98.67%	91.96%	69.81%
TS2-EO	98.46%	85.42%	86.26%
TS2-AA	92.09%	95.55%	96.51%

From **Supplementary Table 2-3**, we can see that EMLP not only achieves significantly lower MAE but also maintains high structural fidelity, with over 92% similarity to DFT-calculated TSs and 100% confirmation of single imaginary frequencies. In contrast, MACE-MP and EquiformerV2 exhibit higher MAEs, substantial energy deviations, and often fail to accurately identify or validate TS structures.

Supplementary Note 7. CO oxidation on a range of Pd/Ag-based surfaces

We rigorously checked the validity of TS structures of the reactions in **Figure 4**(main text) predicted by each model, listed in **Supplementary Table 4**. After DFT frequency calculations, the EMLP-derived TS structures consistently featured a single imaginary frequency (f/i), confirming that they are true first-order saddle points. In contrast, some TS candidates identified by MACE-MP and EquiformerV2 correspond to incorrect stationary points, evidenced by small imaginary frequencies ($\sim 100 \text{ cm}^{-1}$) or no imaginary frequency. For example, transition state structures for CO oxidation on Pd-TB and Ag-Pd(S) found by using MACE-mp were proven to be incorrect. EquiformerV2 also failed to find the correct transition state structures on Ag-Pd₄ and Ag-Pd₁₃. Also among the transition state structures found by MACE-mp and EquiformerV2 there appears to be many small imaginary frequency (~ 100). This might indicate that the potential energy surface is nearly flat along the reaction coordinate near the transition state, which is not the case as seen in **Figure 4**. Thus, it might mean that the structure is not a true transition state but perhaps a higher-order saddle point or even a local minimum.

Supplementary Table 4. The DFT vibrational frequency calculations of transition states found by different models. The transition state structures with very small or no imaginary frequency are highlighted in red.

	EMLP	MACE-mp	EquiformerV2
Pd(100)	$f/i=339.97$	$f/i=102.37$	$f/i=287.87$
Pd(110)	$f/i=448.41$	$f/i=105.18$	$f/i=383.63$
Pd(111)	$f/i=466.29$	$f/i=31.47$	$f/i=89.83$
Pd-TB	$f/i=422.84$	$f=55.44$	$f/i=508.34$
Pd-Nanorod	$f/i=373.60$	$f/i=250.07$	$f/i=403.81$
Ag(111)-Pd(SAC)	$f/i=355.34$	$f=32.32$	$f/i=226.10$
Ag(111)-Pd ₁	$f/i=263.31$	$f/i=133.36$	$f/i=307.78$
Ag(111)-Pd ₄	$f/i=316.23$	$f/i=141.20$	$f=47.10$
Ag(111)-Pd ₉	$f/i=359.60$	$f/i=143.94$	$f/i=126.02$
Ag(111)-Pd ₁₃	$f/i=356.09$	$f/i=253.98$	-
Ag(111)-Pd ₂₀	$f/i=419.12$	$f/i=376.26$	$f/i=314.54$
Ag(111)-Pd(line)	$f/i=328.50$	$f/i=134.90$	$f/i=317.32$

In addition, structural similarity analyses reveal that EMLP-derived TS structures maintain at least a 97% similarity to the DFT references. However, MACE-MP and EquiformerV2 occasionally deviate by up to 10%, despite occasionally achieving seemingly reasonable energy predictions. For instance, on Pd(100) and Pd(110) surfaces, the TS structures found by EquiformerV2 show about a 10% difference in similarity compared to the DFT-obtained TS, yet the predicted energies remain reasonably accurate, suggesting that error cancellation may be occurring. Conversely, MACE-mp achieves a 99.57% structural similarity to the DFT TS on Pd(111), but the energy still deviates by nearly 1 eV.

Supplementary Table 5. The structure similarity check between transition states found by DFT and different models. The predicated structure that is closest to the DFT result is highlighted in bold.

	EMLP	MACE-mp	EquiformerV2
Pd(100)	98.71%	91.37%	91.50%
Pd(110)	99.65%	96.86%	90.35%
Pd(111)	97.25%	99.57%	58.06%
Pd-TB	99.20%	92.60%	88.86%
Pd-Nanorod	98.85%	89.92%	91.78%
Ag(111)-Pd(SAC)	99.81%	95.17%	82.78%
Ag(111)-Pd₁	99.17%	90.00%	90.64%
Ag(111)-Pd₄	97.13%	81.55%	74.08%
Ag(111)-Pd₉	99.95%	72.78%	97.41%
Ag(111)-Pd₁₃	98.49%	59.85%	-
Ag(111)-Pd₂₀	99.41%	63.85%	85.36%
Ag(111)-Pd(line)	99.20%	94.96%	92.18%

Supplementary Note 8. Acetylene hydrogenation on Pd(111), H-covered Pd(111), and PdAg(111)

The accuracy of TS predictions in of the reactions in **Figure 5**(main text) are further checked through the frequency analysis and structural comparisons summarized in **Supplementary Table 6** and **Supplementary Table 7**. In **Supplementary Table 6**, the DFT-based frequency calculations confirm that EMLP's predicted TS structures all exhibit a single imaginary frequency, confirming their status as true first-order saddle points. In contrast, some of the TS structures identified by EquiformerV2 are incorrect, despite appearing energetically reasonable. This suggests that EquiformerV2's seemingly good performance may be coincidental due to its inconsistent performance rather than truly reliable. This discrepancy again confirms that good energy agreement alone does not guarantee a physically meaningful TS structure. As for MACE-mp, aside from one TS having a slightly smaller imaginary frequency, all other identified TS structures are confirmed to be correct by the frequency calculations. Moreover, **Supplementary Table 7**, which assesses structural similarity between the DFT-derived TS structures and those found by each model, reveals that EMLP's TS structures consistently align with DFT by over 98% which is clearly unrivaled in identifying accurate TS structures. MACE-mp and EquiformerV2 show fluctuating similarity scores, reflecting their weaker performance in accurately capturing right configurations.

Supplementary Table 6. The DFT vibrational frequency calculations of transition states found by different models. The transition state structures with very small or no imaginary frequency are highlighted in red.

		EMLP	MACE-mp	EquiformerV2
Pd(111)	TS1	$f/i=812.85$	$f/i=703.40$	$f/i=991.30$
	TS2	$f/i=766.63$	$f/i=688.29$	$f/i=780.67$
	TS3	$f/i=383.85$	$f/i=96.43$	$f=63.45$
	TS4	$f/i=907.80$	$f/i=892.52$	$f/i=874.24$
Pd(111)-Hcov	TS1	$f/i=972.62$	$f/i=937.65$	$f/i=963.60$
	TS2	$f/i=889.59$	$f/i=938.80$	$f/i=930.44$
	TS3	$f/i=735.67$	$f/i=902.83$	$f=94.39$
	TS4	$f/i=763.66$	$f/i=962.46$	$f=94.39$
PdAg(111)-Hcov	TS1	$f/i=628.32$	$f/i=672.56$	$f/i=816.99$

	TS2	<i>f</i> /i=626.20	<i>f</i> /i=791.91	<i>f</i> /i=689.28
	TS3	<i>f</i> /i=701.20	<i>f</i> /i=927.04	-
	TS4	<i>f</i> /i=830.65	<i>f</i> /i=955.64	<i>f</i> /i=754.87
PdAg₃(111)-Hcov	TS1	<i>f</i> /i=730.04	<i>f</i> /i=813.96	<i>f</i> /i=920.83
	TS2	<i>f</i> /i=703.44	<i>f</i> /i=745.67	<i>f</i> /i=641.81
	TS3	<i>f</i> /i=1060.81	<i>f</i> /i=981.39	<i>f</i> /i=1094.72
	TS4	<i>f</i> /i=863.50	<i>f</i> /i=843.13	<i>f</i> /i=760.04

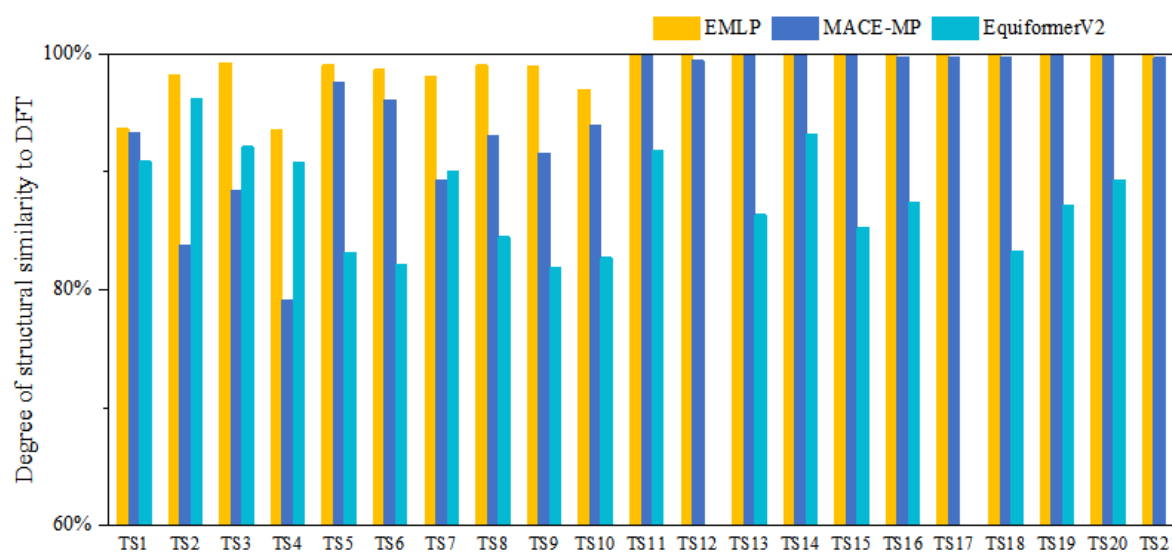
Supplementary Table 7. The structure similarity check between transition states found by DFT and different models. The predicated structure that is closest to the DFT result is highlighted in bold.

		EMLP	MACE-mp	EquiformerV2
Pd(111)	TS1	99.69%	95.55%	93.39%
	TS2	99.92%	96.50%	93.08%
	TS3	99.65%	94.97%	70.22%
	TS4	99.08%	83.67%	90.64%
Pd(111)-Hcov	TS1	99.69%	99.77%	95.14%
	TS2	99.97%	99.81%	95.51%
	TS3	99.53%	93.20%	-
	TS4	99.51%	98.82%	90.65%
PdAg(111)-Hcov	TS1	99.73%	99.87%	97.89%
	TS2	99.73%	92.41%	96.18%
	TS3	99.96%	97.09%	92.23%
	TS4	99.00%	99.23%	93.81%
PdAg₃(111)-Hcov	TS1	98.57%	93.91%	94.28%
	TS2	97.93%	94.58%	94.83%
	TS3	98.19%	95.93%	96.72%
	TS4	98.42%	97.71%	94.19%

Supplementary Note 9. Fischer–Tropsch (FT) process

Supplementary Table 8. Elementary steps for FT reactions on Pd(100).

$\text{CO} * + \text{H} * \rightleftharpoons \text{HCO} * \text{ (R1)}$	$\text{C}_2\text{H}_5\text{CHCHO} * + \text{H} * \rightleftharpoons \text{C}_2\text{H}_5\text{CH}_2\text{CHO} * \text{ (R12)}$
$\text{HCO} * + \text{CH} * \rightleftharpoons \text{CHCHO} * \text{ (R2)}$	$\text{C}_2\text{H}_5\text{CH}_2\text{CHO} * \rightleftharpoons \text{C}_2\text{H}_5\text{CH}_2\text{CH} * + \text{O} * \text{ (R13)}$
$\text{CHCHO} * + \text{H} * \rightleftharpoons \text{CH}_2\text{CHO} * \text{ (R3)}$	$\text{C}_3\text{H}_7\text{CH} * + \text{HCO} * \rightleftharpoons \text{C}_3\text{H}_7\text{CHCHO} * \text{ (R14)}$
$\text{CH}_2\text{CHO} * + \text{H} * \rightleftharpoons \text{CH}_3\text{CHO} * \text{ (R4)}$	$\text{C}_3\text{H}_7\text{CHCHO} * + \text{H} * \rightleftharpoons \text{C}_3\text{H}_7\text{CH}_2\text{CHO} * \text{ (R15)}$
$\text{CH}_3\text{CHO} * \rightleftharpoons \text{CH}_3\text{CH} * + \text{O} * \text{ (R5)}$	$\text{C}_4\text{H}_9\text{CHO} * \rightleftharpoons \text{C}_4\text{H}_9\text{CH} * + \text{O} * \text{ (R16)}$
$\text{CH}_3\text{CH} * + \text{HCO} * \rightleftharpoons \text{CH}_3\text{CHCHO} * \text{ (R6)}$	$\text{C}_4\text{H}_9\text{CH} * + \text{HCO} * \rightleftharpoons \text{C}_4\text{H}_9\text{CHCHO} * \text{ (R17)}$
$\text{CH}_3\text{CHCHO} * + \text{H} * \rightleftharpoons \text{CH}_3\text{CH}_2\text{CHO} * \text{ (R7)}$	$\text{C}_4\text{H}_9\text{CHCHO} * + \text{H} * \rightleftharpoons \text{C}_4\text{H}_9\text{CH}_2\text{CHO} * \text{ (R18)}$
$\text{CH}_3\text{CHCHO} * \rightleftharpoons \text{CH}_3\text{CH}_2\text{CH} * + \text{O} * \text{ (R8)}$	$\text{C}_4\text{H}_9\text{CH}_2\text{CHO} * \rightleftharpoons \text{C}_4\text{H}_9\text{CH}_2\text{CH} * + \text{O} * \text{ (R19)}$
$\text{CH}_3\text{CH}_2\text{CH} * + \text{H} * \rightleftharpoons \text{CH}_3\text{CH}_2\text{CH}_2 * \text{ (R9)}$	$\text{C}_4\text{H}_9\text{CH}_2\text{CH} * + \text{H} * \rightleftharpoons \text{C}_4\text{H}_9\text{CH}_2\text{CH}_2 * \text{ (R20)}$
$\text{CH}_3\text{CH}_2\text{CH}_2 * + \text{H} * \rightleftharpoons \text{CH}_3\text{CH}_2\text{CH}_3 * \text{ (R10)}$	$\text{C}_5\text{H}_{11}\text{CH}_2 * + \text{H} * \rightleftharpoons \text{C}_6\text{H}_{14} * \text{ (R21)}$
$\text{C}_2\text{H}_5\text{CH} * + \text{HCO} * \rightleftharpoons \text{C}_2\text{H}_5\text{CHCHO} * \text{ (R11)}$	



Supplementary Figure 11. Transition state structural similarity check. Degree of structural similarity to DFT from for F-T process on Pd(100) between model predictions made by EMLP(yellow), Mace-mp (blue), and EquiformerV2 (teal).

Supplementary Table 9. The DFT vibrational frequency calculations of transition states found by different models. The transition state structures with very small or no imaginary frequency are highlighted in red.

	EMLP	MACE-mp	EquiformerV2
F-T	TS1	$f/i=262.36$	$f/i=30.10$
	TS2	$f/i=475.11$	$f/i=480.04$
	TS3	$f/i=735.84$	$f/i=698.48$
	TS4	$f/i=853.43$	$f/i=682.98$
	TS5	$f/i=358.18$	$f/i=149.60$
	TS6	$f/i=473.32$	$f/i=384.69$
	TS7	$f/i=909.76$	$f/i=879.55$
	TS8	$f/i=411.27$	$f/i=11.89$
	TS9	$f/i=874.97$	$f/i=509.38$
	TS10	$f/i=991.08$	$f/i=171.25$
	TS11	$f/i=469.41$	$f/i=433.24$
	TS12	$f/i=818.36$	-
	TS13	$f/i=441.00$	$f/i=336.53$
	TS14	$f/i=428.04$	$f/i=430.01$
	TS15	$f/i=1056.65$	$f/i=357.79$
	TS16	$f/i=346.66$	$f/i=121.39$
	TS17	$f/i=484.91$	-
	TS18	$f/i=712.24$	$f=13.80$
	TS19	$f/i=408.41$	$f/i=68.71$
	TS20	$f/i=898.29$	$f/i=720.21$
	TS21	$f/i=835.00$	-

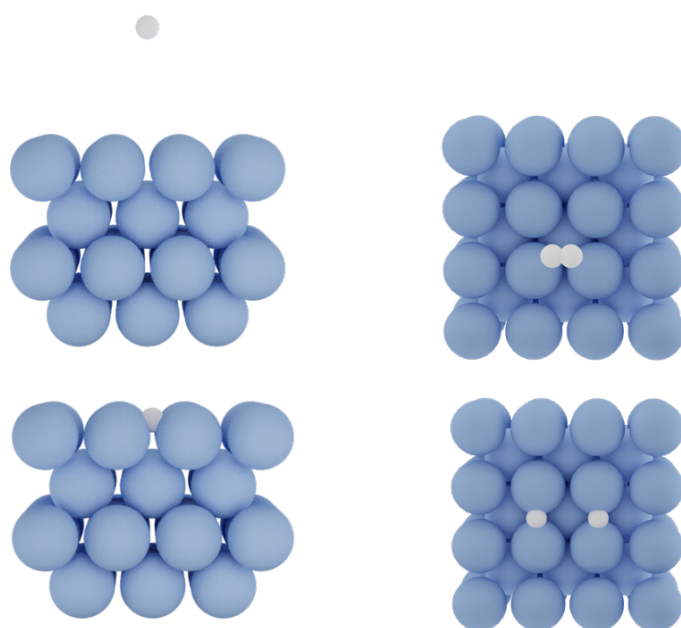
Supplementary Note 10. Standard test reactions

We have also conducted the same tests using MACE-mp and EquiformerV2, shown in **Supplementary Table 10**. EquiformerV2 has failed to calculate 3 out of 9 reaction and have an average error more than 1 eV. Mace-mp can calculate all 9 reactions but also suffer from a very high error compared to DFT results.

Supplementary Table 10. Model performance comparison. Comparison of reaction barrier (E_a , eV) and reaction enthalpies (ΔE , eV) obtained from DFT and Mace-mp and EquiformerV2 for a range of reactions.

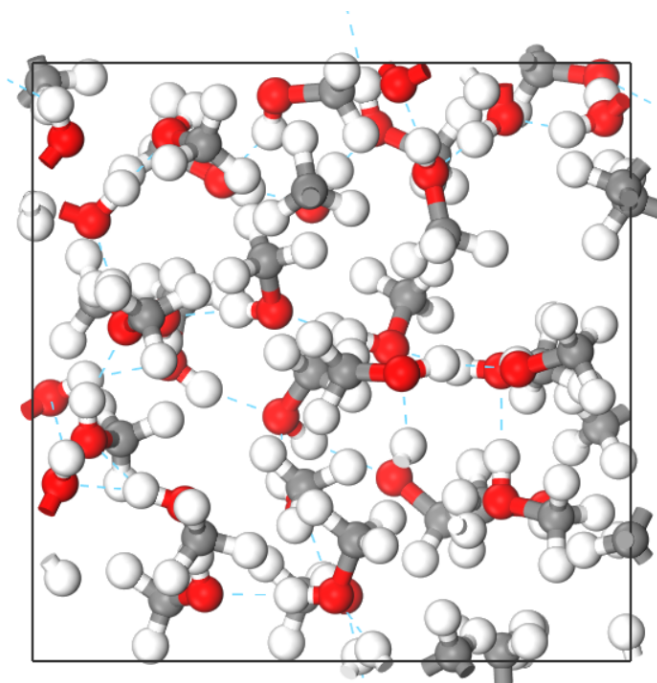
Reactions	DFT		MACE-mp		EquiformerV2	
	E_a /eV	ΔE /eV	E_a /eV	ΔE /eV	E_a /eV	ΔE /eV
HCCH $\leftrightarrow$ CCH ₂	1.98	-1.97	1.92	-1.40	0.73	-0.34
H ₂ CO $\leftrightarrow$ H ₂ + CO	2.83	0.58	1.61	0.96	-	-
bicyclo[1,1,0] $\leftrightarrow$ butane TS1	2.66	-0.84	1.87	-1.16	-	-
butadiene + ethylene $\leftrightarrow$ cyclohexene	0.46	1.86	0.78	1.94	0.03	1.03
trans-butadiene $\leftrightarrow$ cis-butadiene	0.32	-0.17	0.22	0.05	0.27	0.06
CH ₃ CH ₃ $\leftrightarrow$ CH ₂ CH ₂ + H ₂	4.70	-1.65	3.47	-1.97	-	-
vinyl alcohol $\leftrightarrow$ acetaldehyde	2.25	0.37	1.89	0.18	0.78	1.03
claisen rearrangement	1.65	-0.74	1.81	-0.75	0.67	0.94
rotational TS in acrolein	0.38	-0.10	0.18	0.03	0.48	1.77
Mean absolute error (MAE)			0.49	0.25	0.71	1.15

Supplementary Note 11. The dynamics of H₂ adsorption and dissociation on the Ag(100).



Supplementary Figure 12. Selected structures from the dynamics simulations of H₂ adsorption and dissociation on the Ag(100).

Supplementary Note 12. Liquid methanol simulation



Supplementary Figure 13. Structure used for the liquid methanol RDF simulation.

Supplementary Note 13. Lattice constant

The lattice constant is one of the fundamental parameters for describing the crystal structure. Accurate description of the lattice constant is crucial as it significantly reflects the performance of the EMLP. In this work, various crystal systems of Pd, Ag, Pd_xAg_y, Pd_xO_y, and Ag_xO_y were selected from the Materials Project¹⁰. Their lattice constants were calculated using the trained EMLP and compared with the results obtained from DFT. As shown in **Supplementary Table 11**, our EMLP exhibits an average error of only 0.63% compared to DFT calculations. This accuracy demonstrates the capability of the EMLP to effectively handle the complex task of identifying unknown crystal structures.

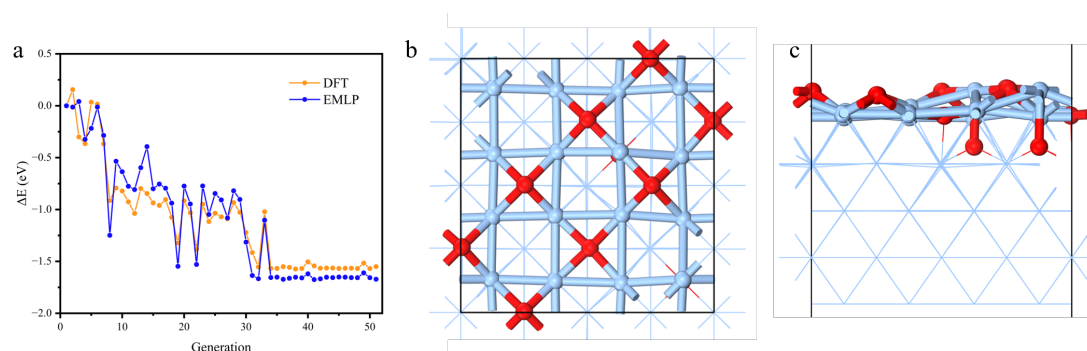
Supplementary Table 11. Lattice constants calculated by EMLP vs. DFT.

Bulk	DFT (Å)			EMLP (Å)			Δ
Ag ₂ O-Cubic (mp-353)	4.75	4.75	4.75	4.74	4.74	4.74	0.21%
Ag ₃ O ₄ -Monoclinic (mp-1605)	3.57	9.26	5.67	3.66	9.29	5.71	1.18%
Ag ₃ Pd-Tetragonal (mp-985296)	4.04	4.04	7.97	4.03	4.03	8.04	0.46%
Ag-cubic (mp-124)	4.07	4.07	4.07	4.07	4.07	4.07	0.00%
Ag-Hex (mp-8566)	2.87	2.87	9.47	2.88	2.88	9.38	0.55%
AgPd-Trigonal (mp-1229012)	2.8	2.8	13.8	2.8	2.8	13.9	0.24%
Ag-Trigonal (mp-989737)	2.86	2.86	21.38	2.88	2.88	21.11	0.89%
Pd ₂ O-cubic (mp-10729)	4.56	4.56	4.56	4.47	4.47	4.47	1.97%
Pd-cubic (mp-2)	3.89	3.89	3.89	3.88	3.88	3.88	0.26%
Pd-Hexagonal (mp-1186427)	2.74	2.74	9.06	2.75	2.75	8.98	0.54%
Average	-	-	-	-	-	-	0.63%

Supplementary Note 14. Global optimization of Ag-AgO

Due to the structural complexity of catalysts in heterogeneous catalysis, catalysts are likely to undergo reconstruction upon molecular adsorption, which can significantly impact their catalytic activity. Therefore, finding a reasonable catalyst model is crucial for understanding catalytic reactions. From a thermodynamic perspective, catalysts tend to adopt the most stable structure. Global optimization methods can help identify this most stable configuration. However, global searches inevitably require finding numerous local minima, and using DFT for global optimization is resource intensive. Traditional force fields and biased machine learning potentials based on MD are limited by their accuracy and often fail to identify the reasonable global minimum.

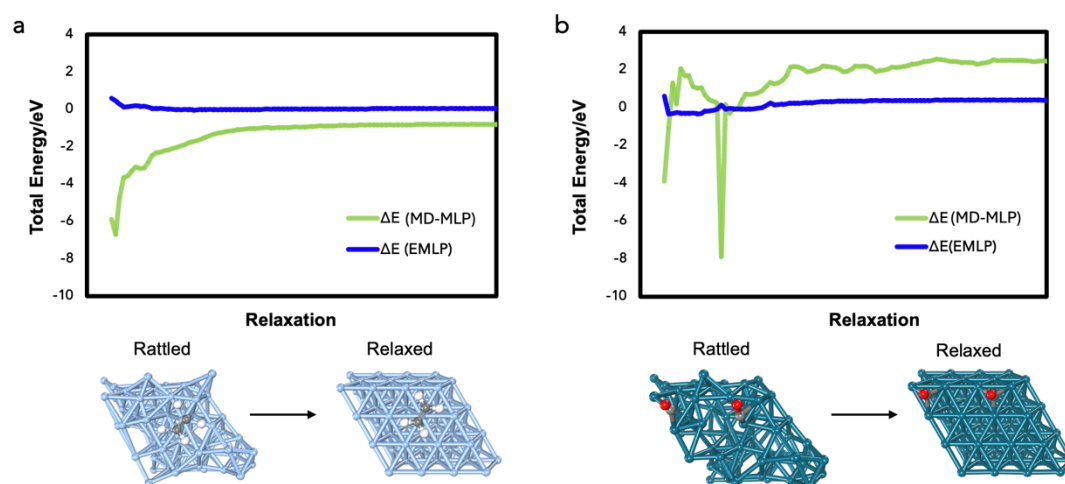
In this work, we utilized the genetic algorithm module within GDPx(<https://gdp.readthedocs.io/en/latest/>) to perform global optimization of AgO/Ag surfaces using EMLP. Simultaneously, we employed DFT to perform single-point energy calculations on the most stable structures identified in each generation, thereby validating the accuracy of the EMLP. As shown in **Supplementary Figure 14**, during EMLP global optimization, the energy of the most stable structure exhibits a decreasing trend with increasing generation. Furthermore, results from DFT calculations corroborate the validity of the EMLP. Additionally, the most stable structures identified by EMLP exhibit a certain degree of symmetry, consistent with our chemical understanding.



Supplementary Figure 14. Global optimization of surface structures using EMLP and GA: (a) Ag₁₆O₁₂/Ag(111) surface, where Ag(111) represents a 4 layers 4×4 surface. In the optimization process, the bottom 2 layers of atoms were fixed. (b), (c) show the most stable Ag₁₆O₁₂/Ag(111) structure of generation 50.

Supplementary Note 15. Relaxation of rattled unlearn structures

We performed generality and stability tests on two structures that are not in the training set: (a) C_2H_4 on $\text{Ag}(111)$; and (b) 2CO on $\text{Pd}(111)$. Each system was initially rattled into an unstable state and then allowed to relax, with the process illustrated in **Supplementary Figure 15**. The EMLP ability to track the relaxation trajectories and energy profiles closely mirrored those derived from DFT, highlighting its force accuracy on each atom and energy shifts during relaxation. In contrast, the MD-MLP showed poor performance, with significant inconsistencies in ΔE , highlighting its limitations in dealing with unlearn structures. The consistency of EMLP results with DFT across these tests reaffirms the accuracy of our method and its potential for general applications in simulating complex catalytic processes.

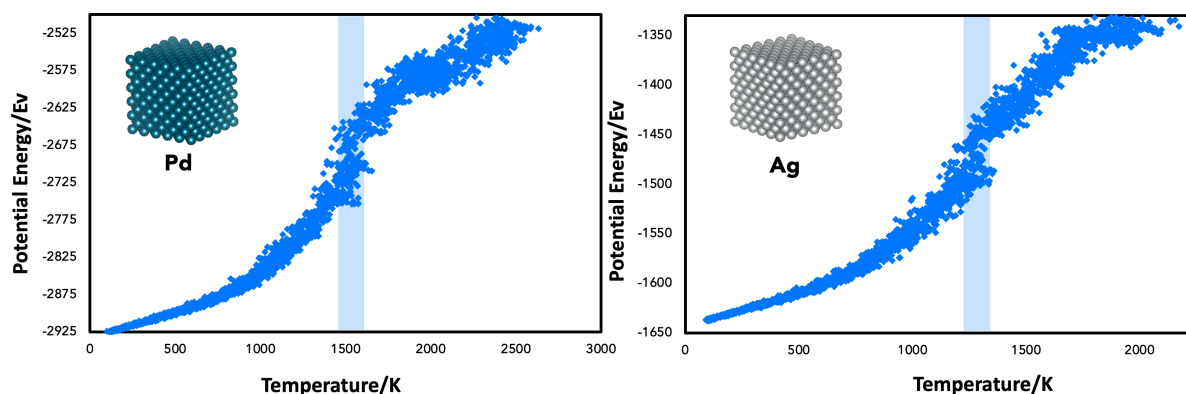


Supplementary Figure 15. Generality test through the relaxation of rattled structures.

The figure illustrates the energy relaxation trajectories for two structures that are not in the training set: (a) C_2H_4 on $\text{Ag}(111)$; and (b) 2CO on $\text{Pd}(111)$. In each case, the system was initially rattled into an unstable state and then allowed to relax. The energy differences (ΔE) between the results from EMLP/MD-MLP and DFT are plotted to assess the performance of each method. The results demonstrate that while the EMLP closely aligns with the DFT results throughout the relaxation process, indicating its robustness and accuracy, the MD-MLP shows significant variations in ΔE , suggesting its inability to accurately handle the structural relaxations.

Supplementary Note 16. Melting temperature MD

The melting temperature is primarily a bulk property, and as such we have calculated the melting temperature for Ag and Pd in its bulk form using the single-phase method. It's worth noting that this discrepancy falls within a reasonable error range inherent to the single-phase method itself (such as, the size of the system and the absence of surface sites that typically nucleate a new phase).^{11–13}



Supplementary Figure 16. Melting MD simulations. Potential energy of Pd and Ag bulks as a function of temperature using the EMLP model.

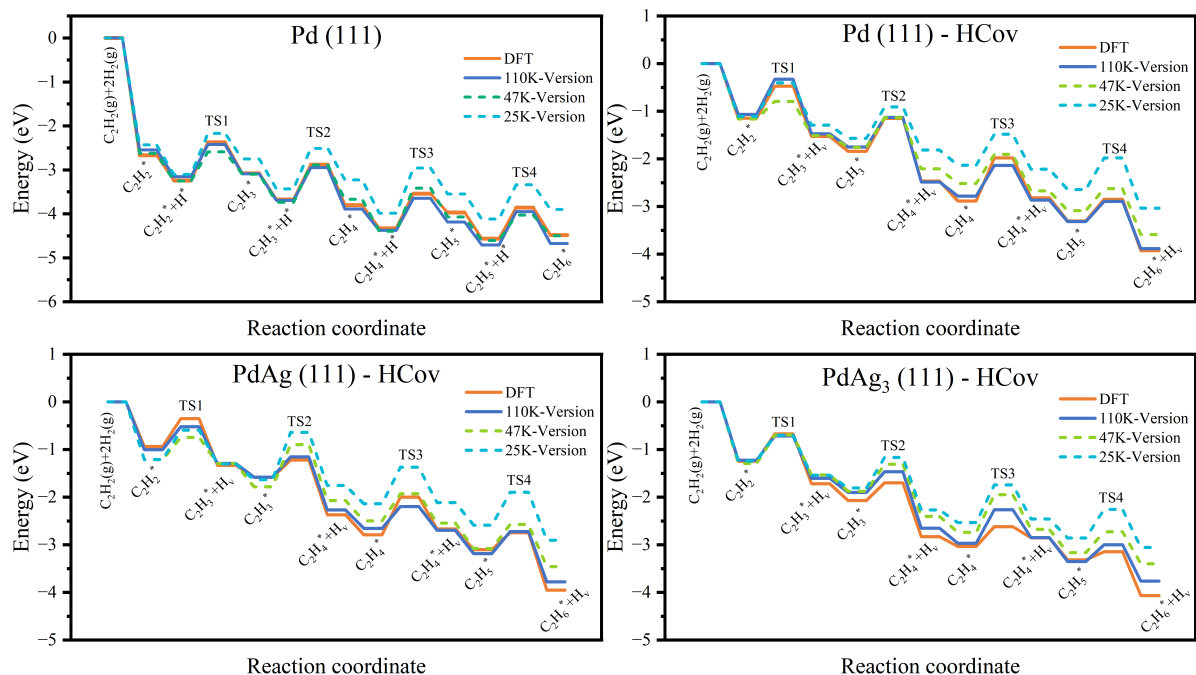
Our results indicate a melting point of bulk Pd is around 1550-1650 K with the reference of 1828 K, while the melting point of bulk Ag is around 1175-1275K with the reference of 1235K.¹⁴ These values are slightly different from the experimental values (~ 10%) demonstrates the capability of our EMLP to describe the bulk properties of the system with the need to simple such systems.

Supplementary Note 17. Data convergence

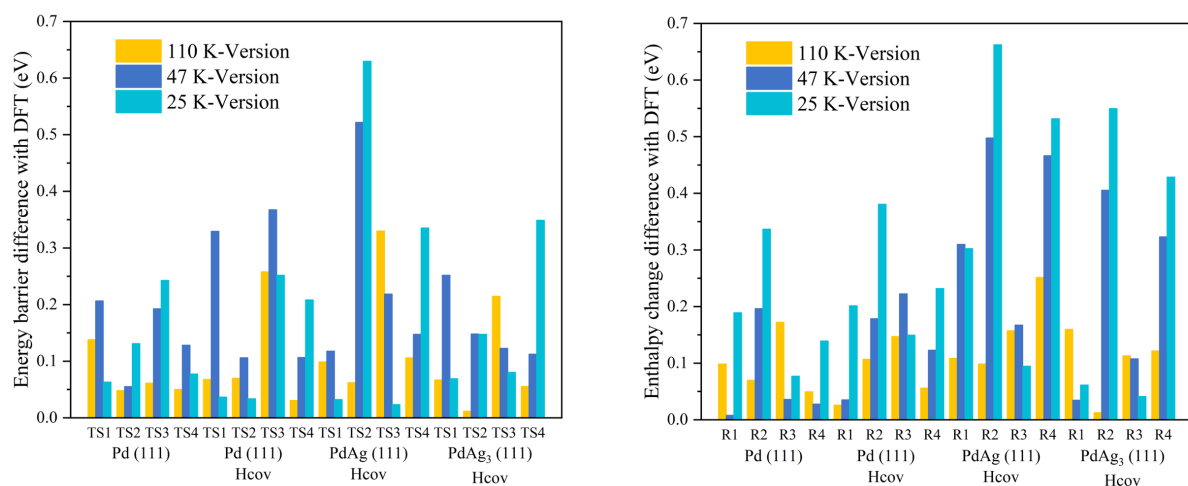
we have conducted additional experiments to clarify how much data is needed to achieve stable and general EMLP performance to provide valuable insights into the scaling of our approach. Based on our experience training EMLPs for various elemental combinations, we have observed a rough heuristic: For achieving a general and reactive EMLP with reasonable accuracy (close to DFT), about *10,000 structures per element* included in the model seems to be sufficient. To illustrate this, we trained three versions of the AgPdCHO EMLP with different dataset sizes:

- **25k structures** (~5,000 structures/element)
- **47k structures** (~9,400 structures/element)
- **110k structures** (~22,000 structures/element), which is the one reported in the manuscript.

Using acetylene hydrogenation on Pd(111), H-covered Pd(111), and PdAg(111) as benchmarking systems, **Supplementary Figure 17** shows that the 47k-version of the EMLP performs almost on par with the 110k-version, with only minor deviations in reaction energetics and barriers compared to DFT. In contrast, the 25k-version, which provides approximately half the “per-element” data density as the 47k-version, shows more noticeable discrepancies. This indicates that increasing the dataset size from 25k to around 47k structures can bring the EMLP’s accuracy much closer to that of the 110k-version.



Supplementary Figure 17. Impact of training set size on reaction predictions. C_2H_2 hydrogenation reaction profiles on the surface of Pd and PdAg alloys. (a) The Pd(111) surface; (b) Pd(111) covered by 1 ML hydrogen; (c) PdAg₃ (111) surface covered by 0.25 ML hydrogen; and (d) PdAg (111) covered by 1 ML hydrogen. The blue and orange lines are the results of 110k version of EMLP(which we used in this paper) and DFT calculations, respectively. The green and teal dash line are the results of the 47k and 25k versions.



Supplementary Figure 18. Impact of training set size on reaction predictions. Reaction barrier and reaction enthalpy change difference of the four C_2H_2 hydrogenation reaction

between model predictions made by EMLP fitted on three different training dataset sizes compared with DFT.

Supplementary Table 12. Impact of training set size on reaction predictions. The MAE (mean absolute error) and MD (maximum deviation) analysis of each model's performance against DFT. Best model results are highlighted in bold

	E_a/eV			ΔE/eV		
Training Data	110k	47k	25k	110k	47k	25k
MAE(eV)	0.10	0.20	0.17	0.11	0.20	0.27
MD(eV)	0.33	0.52	0.63	0.25	0.50	0.66

Supplementary Note 18. Computational cost of training the AgPdCHO EMLP

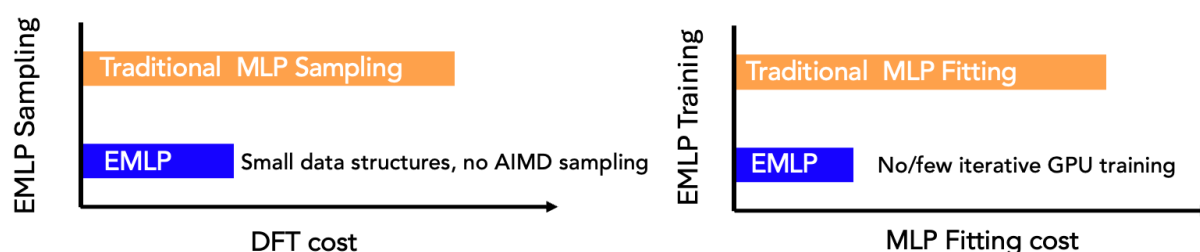
The training framework requires no domain knowledge. The two-step training procedure is straightforward, with DFT used solely for single-point calculations and low-precision structure relaxations, enhancing cost-effectiveness. In principle, our sampling method can be easily integrated with any state-of-the-art atomistic ML models. For example, the AgPdCHO EMLP training dataset is selected from a pool of around 5 million initial structures (relaxation of REICO generated structures which are ~ 30 atoms). The total building time including:

1. **Low-precision relaxation:** If the relaxations were all done using DFT with reasonable settings (gamma, 450 cut-off, $5.0e^{-6}$ eV/Å energy criterion). Generate 1 structure (1 relaxation step) only requires ~ 1.5 s on a 16-CPU node, and it takes about 1 week of computing in total (assuming 10 parallel jobs, 160 CPU). This cost is several orders of magnitude lower than typical AIMD-based sampling. Moreover, one can further accelerate this stage by using a proto-EMLP (i.e. described in the initiation part) or any available classical force field, reserving DFT only for final labelling. Since at this point, we only need to generate diverse local environments rather than final, high-precision geometries.

2. **The high precision labeling:** DFT single-point calculation is used to label the selected training set, each took about ~ 120 s on a 16-CPU node (in total ~ 12 days assuming 10 parallel jobs for all 110k structures).

3. **Model Training:** As we did not apply iterative training, the AgPdCHO EMLP took about 9 days to train on a single GPU (Nvidia 4090D or equivalent).

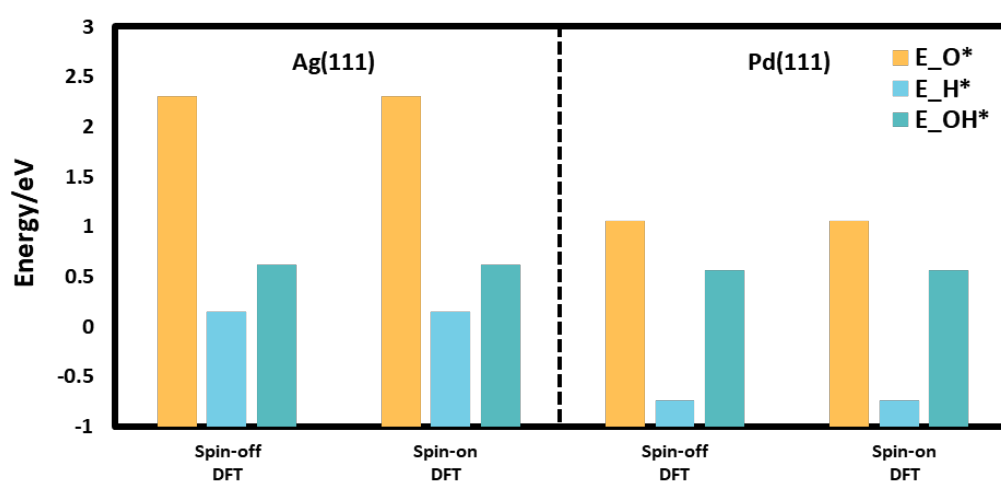
In short, it comes with a very modest computational cost in terms of CPU and GPU, making EMLP generation highly efficient and accessible, as shown in **Supplementary Figure 19**.



Supplementary Figure 19. Computational cost. Comparative computational cost of traditional MLP workflows versus EMLP.

Supplementary Note 19. Spin Test

The primary focus is on PdAg-related catalysts, which typically do not require spin/ magnetism considerations. We have benchmarked the binding energies of O*, H*, and OH* on Ag(111) and Pd(111) surfaces using spin-polarized and non-spin-polarized DFT calculations. As illustrated in **Supplementary Figure 20**, the results indicate exact same values between spin-polarized and non-spin-polarized DFT energies for these systems, thus, all the reference DFT calculations performed in this work do not account for spin/magnetism.



Supplementary Figure 20. DFT spin test. Comparison of adsorption energies for O*, H*, and OH* on Ag(111) and Pd(111) surfaces, computed using non-spin-polarized DFT, and spin-polarized DFT.

Supplementary References

1. Batzner, S. *et al.* E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
2. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
3. Schölkopf, B., Smola, A. & Müller, K.-R. Kernel principal component analysis. in *Artificial Neural Networks — ICANN'97* (eds. Gerstner, W., Germond, A., Hasler, M. & Nicoud, J.-D.) 583–588 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1997).
4. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
5. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
6. Himanen, L. *et al.* DScript: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020).
7. Luo, L.-H., Huang, S.-D., Shang, C. & Liu, Z.-P. Resolving Activation Entropy of CO Oxidation under the Solid–Gas and Solid–Liquid Conditions from Machine Learning Simulation. *ACS Catal.* **12**, 6265–6275 (2022).
8. Huš, M. & Hellman, A. Ethylene Epoxidation on Ag(100), Ag(110), and Ag(111): A Joint Ab Initio and Kinetic Monte Carlo Study and Comparison with Experiments. *ACS Catal.* **9**, 1183–1196 (2019).
9. Linic, S. & Barteau, M. A. On the Mechanism of Cs Promotion in Ethylene Epoxidation on Ag. *J. Am. Chem. Soc.* **126**, 8086–8087 (2004).
10. Alberi, K. *et al.* The 2019 materials by design roadmap. *J. Phys. Appl. Phys.* **52**, 013001 (2019).

11. Wang, S. *et al.* New Ab Initio Based Pair Potential for Accurate Simulation of Phase Transitions in ZnO. *J. Phys. Chem. C* **118**, 11050–11061 (2014).
12. Strachan, A., Çağın, T. & Goddard, W. A. Phase diagram of MgO from density-functional theory and molecular-dynamics simulations. *Phys. Rev. B* **60**, 15084–15093 (1999).
13. Rang, M. & Kresse, G. First-principles study of the melting temperature of MgO. *Phys. Rev. B* **99**, 184103 (2019).
14. CRC handbook of chemistry and physics.