nature catalysis

Article

https://doi.org/10.1038/s41929-025-01398-3

General reactive element-based machine learning potentials for heterogeneous catalysis

Received: 23 August 2024

Accepted: 12 July 2025

Published online: 23 September 2025

Check for updates

Changxi Yang[1,2,5], Chenyu Wu[1,3,5], Wenbo Xie[1,5] ✉, Daiqian Xie[3] & P. Hu[1,4] ✉

Developing truly universal machine learning potentials for heterogeneous catalysis remains challenging. Here we introduce our element-based machine learning potential (EMLP), trained on a unique random exploration via imaginary chemicals optimization (REICO) sampling strategy. REICO samples diverse local atomic environments to build a representative dataset of atomic interactions, making the EMLP inherently general and reactive, capable of accurately predicting elementary reactions without explicit structural or reaction pathway inputs. We demonstrate the generality and reactivity of our approach by building a Ag-Pd-C-H-O EMLP targeting Pd–Ag catalysts interacting with C/H/O-containing species, achieving quantitative agreement with density functional theory even for complex scenarios such as surface reconstruction, coverage effects and solvent environments, cases for which existing foundation models typically fail. Our method paves the way to replace density functional theory calculations for large and intricate systems in heterogeneous catalysis, and offers a general framework that can readily be extended to other catalytic systems, and to broader fields such as materials science.

For decades, quantum chemistry, particularly density functional theory (DFT), has been indispensable in offering microscopic insights into catalytic reactions and profoundly influencing the field of heterogeneous catalysis[1,2]. Despite its successes, it has recently reached a plateau: its application is severely limited by the computational costs associated with the scaling of quantum mechanical calculations. This scaling issue makes the study of large, complex systems or the exploration of dynamic and complex reaction networks impractical[3]. At the same time, empirical methods, such as classical force fields, provide a less costly approach to simulate reactions but at the expense of quantitative accuracy and generality compared to quantum mechanical methods. The emergence of neural-network-based machine learning potentials (MLPs) began to mitigate this compromise, achieving computational efficiency similar to classical force fields while maintaining DFT-level accuracy[4–6]. The rapid innovations in MLP models, including LASP[7,8], DeepMD[9], REANN[10], NequIP[11] and MACE[12], along with training strategies such as active learning[13–15] and delta-learning[16,17], are constantly refining the capability of MLPs and their predicting power. However, the generality of MLPs to perform on-the-fly calculations of arbitrary structures, as DFT can, remains a distant goal in heterogeneous catalysis.

Current MLP sampling approaches in heterogeneous catalysis are predominantly system dependent because they are trained on datasets containing very specific configurations of structures. They usually begin with different phases or facets of the catalyst, vary the orientation and number of reactants/products on the catalyst, and use molecular dynamics (MD) simulations to sample atomic positions

[1]School of Physical Science and Technology, ShanghaiTech University, Shanghai, China. [2]State Key Laboratory of Green Chemical Engineering and Industrial Catalysis, Centre for Computational Chemistry and Research Institute of Industrial Catalysis, School of Chemistry and Molecular Engineering, East China University of Science and Technology, Shanghai, China. [3]Key Laboratory of Mesoscopic Chemistry of MOE, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, China. [4]School of Chemistry and Chemical Engineering, The Queen's University of Belfast, Belfast, UK. [5]These authors contributed equally: Changxi Yang, Chenyu Wu, Wenbo Xie. ✉e-mail: xiewb1@shanghaitech.edu.cn; hupj@shanghaitech.edu.cn

https://doi.org/10.1038/s41929-025-01398-3

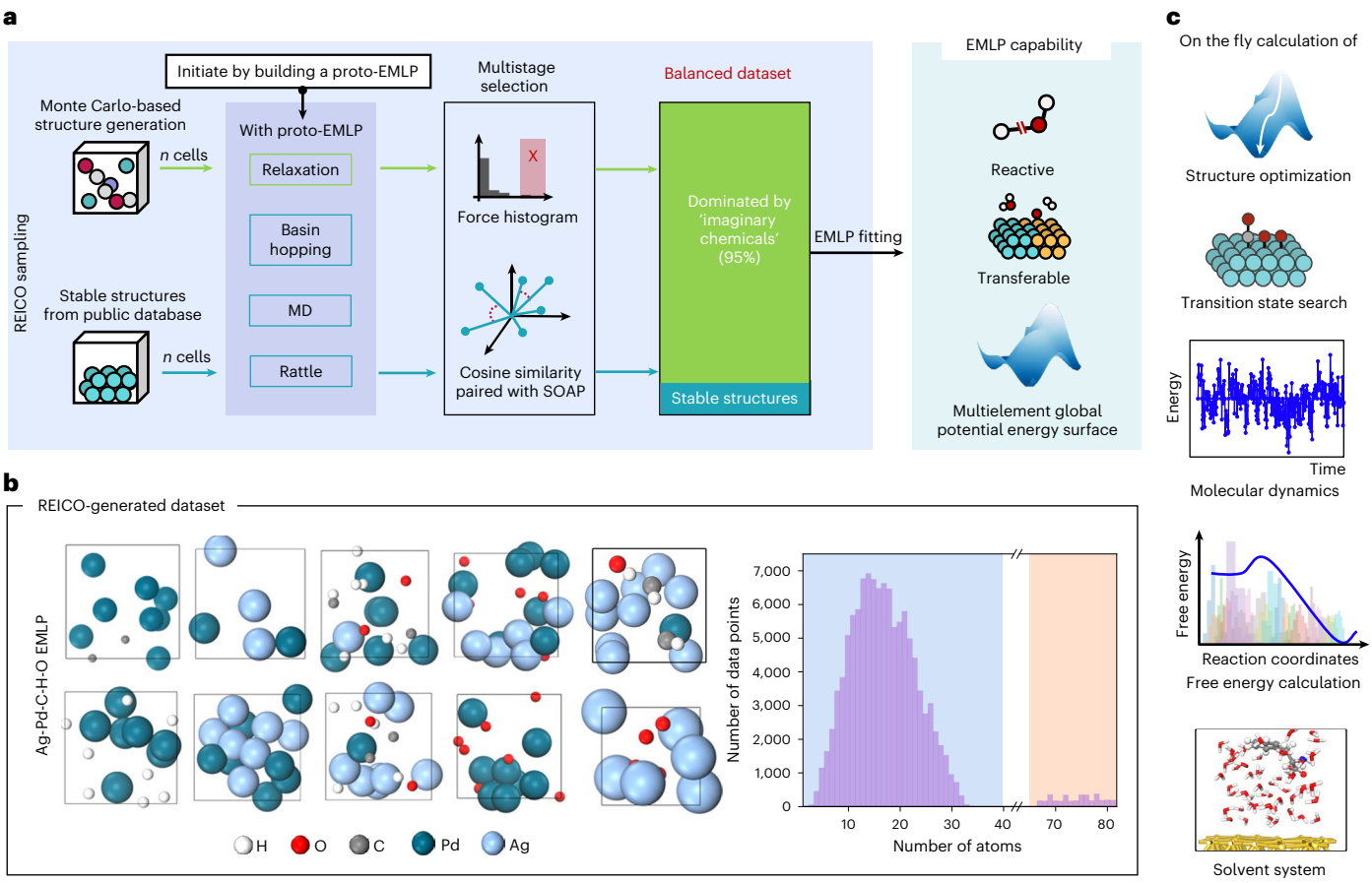


Fig. 1 | Summary of the REICO sampling, EMLP training workflow and applications. a, Workflow for the EMLP. Initially, a proto-EMLP can be trained using a small dataset of random structures and low-precision DFT relaxation. In the REICO-sampling phase, additional random structures are relaxed using the proto-EMLP (alternatively, perform all relaxations with low-precision DFT). The dataset is then supplemented by stable structures with sampling techniques to enrich the local atomic environment in the dataset, as the stable, highly symmetric atomic arrangments are unlikely to be visited by REICO. The dataset is then balanced through a multistage selection scheme, utilizing force histograms and cosine similarity paired with the SOAP descriptor. The final dataset undergoes high-precision DFT calculations before training the EMLP. **b**. Example structures and the system size distribution of the Ag-Pd-C-H-O EMLP training dataset. As can be seen, most calculations were performed on systems containing ~15 atoms. **c**. As the EMLP can perform on-the-fly calculations of arbitrary structures that are not directly targeted during training, it can replace DFT in a series of benchmark calculations representative of heterogeneous catalysis.

and to explore reactive chemical space[18–20]. For example, Liu et al. built a Cu-Zn-C-H-O MLP to explore reaction pathways for methanol synthesis from CO and CO_2 on ZnO/Cu catalysts, revealing the dominance of CO_2 hydrogenation and the complex effects of zinc coverage on catalyst activity[21]. Han et al. utilized a Zn-Cr-O MLP to investigate the chromium-doped ZnO ternary system on the ZnO(1010) surface, exploring the influence of chromium and oxygen vacancies on CO activation for syngas conversion[22,23]. These system-dependent MLPs are good at predicting the targeted reaction systems, often achieving root mean square error (RMSE) of energies <5 meV per atom. However, their application is inherently limited to the targeted systems.

Beginning with well-organized surface structures inevitably results in a multitude of similar local environments. This is particularly true when using the widely adopted slab model, which typically includes layers with very similar atomic arrangements. This leads to the consensus that MLPs can only predict the energies of the structures (or similar structures) for which they have been previously trained[24]. Another issue is that many systems in heterogeneous catalysis are too computationally demanding if sampled this way or have limited known structures, such as catalysts with multicompositions, or reaction systems involving solvent environments[20,25] and coverage effects[26,27]. It is fair to state that the generality of MLPs is arguably the most challenging problem if they are going to replace DFT as the main computational engine.

Reaction lies at the heart of heterogeneous catalysis, but developing a reactive MLP is also challenging for two reasons. First, the structural space is enormous, and exhausting such a vast space is a daunting, if not impossible, task. Second, traditional MLP training requires sampling sufficient structures that are similar to the transition state for each elementary step that may be encountered in any reactive system. However, any such transition states are difficult to locate, even using DFT, due to the peculiar nature of transition state structures in the energy landscape. Hence, the initial and final states in a reaction can be predicted with reasonable accuracy, but transition states result in notable errors with traditional MLPs. A recent study by Ceder and co-workers has shown a systematic softening issue in current foundation models due to the near-equilibrium atomic arrangements in their database and the lack of high-energy states[28]. Therefore, reactive MLPs typically employ targeted, model-aware sampling strategies that rely on chemical intuition for the chemical space involved in targeted reactions. Methods such as stochastic surface walking, enhanced sampling or microkinetic model-guided approaches are used to sample rare chemical reaction events[7,21,29,30]. Such strategies thus often require sampling every potential reaction pathway, a task that is both extremely challenging and resource intensive as the training dataset grows. Even so, the resulting MLP would be tuned to only perform well on the reaction network for which it is trained.

https://doi.org/10.1038/s41929-025-01398-3

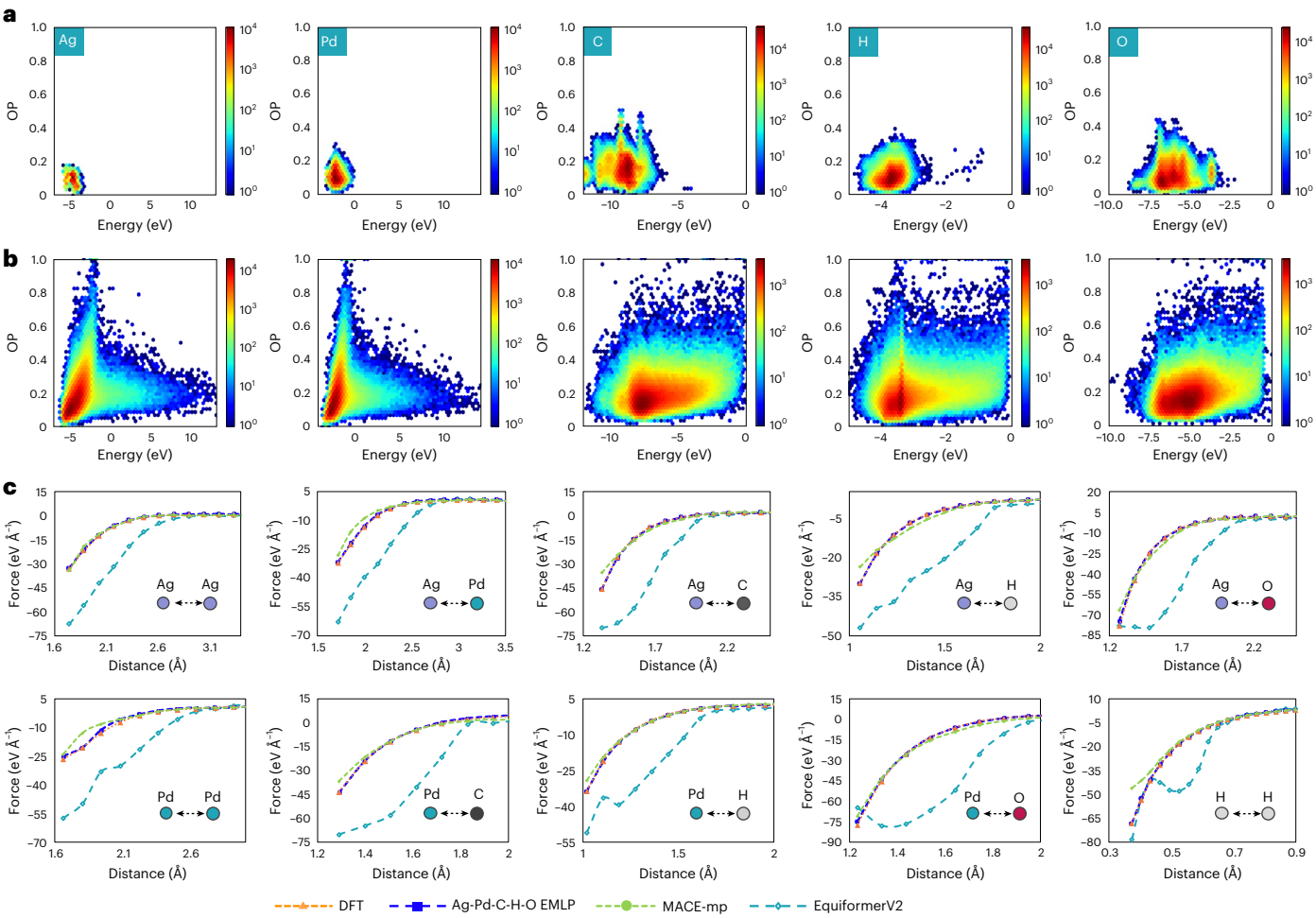


Fig. 2 | Analysis of the RECIO-generated dataset and dimer scan performance. **a**,**b**, Data diversity and coverage analysis of the whole dataset generated by MD (**a**) with structures from a public database compared to REICO (**b**) in this work. The *x* axis shows the per-atom energy, the *y* axis is the Steinhart order parameter (OP) which measures the distribution of local atomic arrangements (arbitrary units), and the colour (*z* axis) is the number of data points in each energy–OP bin ($\log_{10}$ scale/counts). The visualizations highlight that the REICO dataset not only covers most regions sampled by the MD dataset but also expands beyond them substantially in terms of energy and local atomic environments. **c**. Generality tests conducted through rigid dimer scans, which are absent from the training datasets, making them ideal for assessing the generality of the EMLP. The tests involve ten dimer combinations of self- and cross-interactions between the elements silver, palladium, carbon, hydrogen and oxygen. Results from DFT calculations serve as the benchmark, and are compared with predictions from the EMLP, MACE-mp and EquiformerV2. M3GNet failed to perform the dimer test. The EMLP results align closely with the DFT benchmarks, whereas the MACE-mp and EquiformerV2 predictions deviate substantially. The colour code is as in Fig. 1.

In this work, we introduce our element-based machine learning potential (EMLP) approach, which focuses on learning the diverse interactions between elements rather than relying on fixed structural arrangements or predefined reaction coordinates. Central to our approach is random exploration via imaginary chemical optimization (REICO), a sampling procedure that constructs datasets that are detached from structural space and focus solely on atomic interactions. As previous works by Deringer et al.[31,32] and Smith et al.[33] reported, adding a level of randomness can enhance the robustness and generality of the MLP. The REICO method avoids traditional dependencies on a specific system; instead, it uses complete randomness sampling. Leveraging small systems composed of randomly generated structures and their subsequent optimization trajectories, our EMLP can proficiently navigate a vast array of unique local atomic environments. Such a strategy significantly enhances the diversity and representativeness of training datasets, thereby extending the applicability of MLPs across a broader spectrum of chemical systems. The resulting EMLP can tackle arbitrary reactive catalytic processes without needing to sample the reaction pathways. To demonstrate, we built an EMLP that contains five elements: palladium and silver, which are commercially used as metal catalysts, along with carbon, hydrogen and oxygen, which form various reactants in heterogeneous catalysis. We conducted a comprehensive set of generality tests and benchmarks, from data diversity analysis to MD behaviours. We have also applied the EMLP to various reaction systems in heterogeneous catalysis and beyond. Across this broad spectrum of applications, the EMLP can predict results that are consistent with chemical intuition and DFT calculations, without the need for retaining or fine-tuning. In addition, we have benchmarked our potential with a MD-MLP trained using a traditional MD sampling method, and with MACE-mp[34], EquiformerV2[35] and M3GNet[36], three well-recognized foundation models trained on millions of DFT data. Across the wide array of tested systems, the EMLP consistently outperforms other models.

Results

REICO-generated dataset

Building an EMLP requires only knowledge of the involved elements, with no domain expertise necessary. Figure 1 outlines the EMLP training workflow, designed to efficiently map and adapt to various local atomic environments within the chemical space (see also Supplementary Figs. 1–5). To verify that our approach achieves the

https://doi.org/10.1038/s41929-025-01398-3

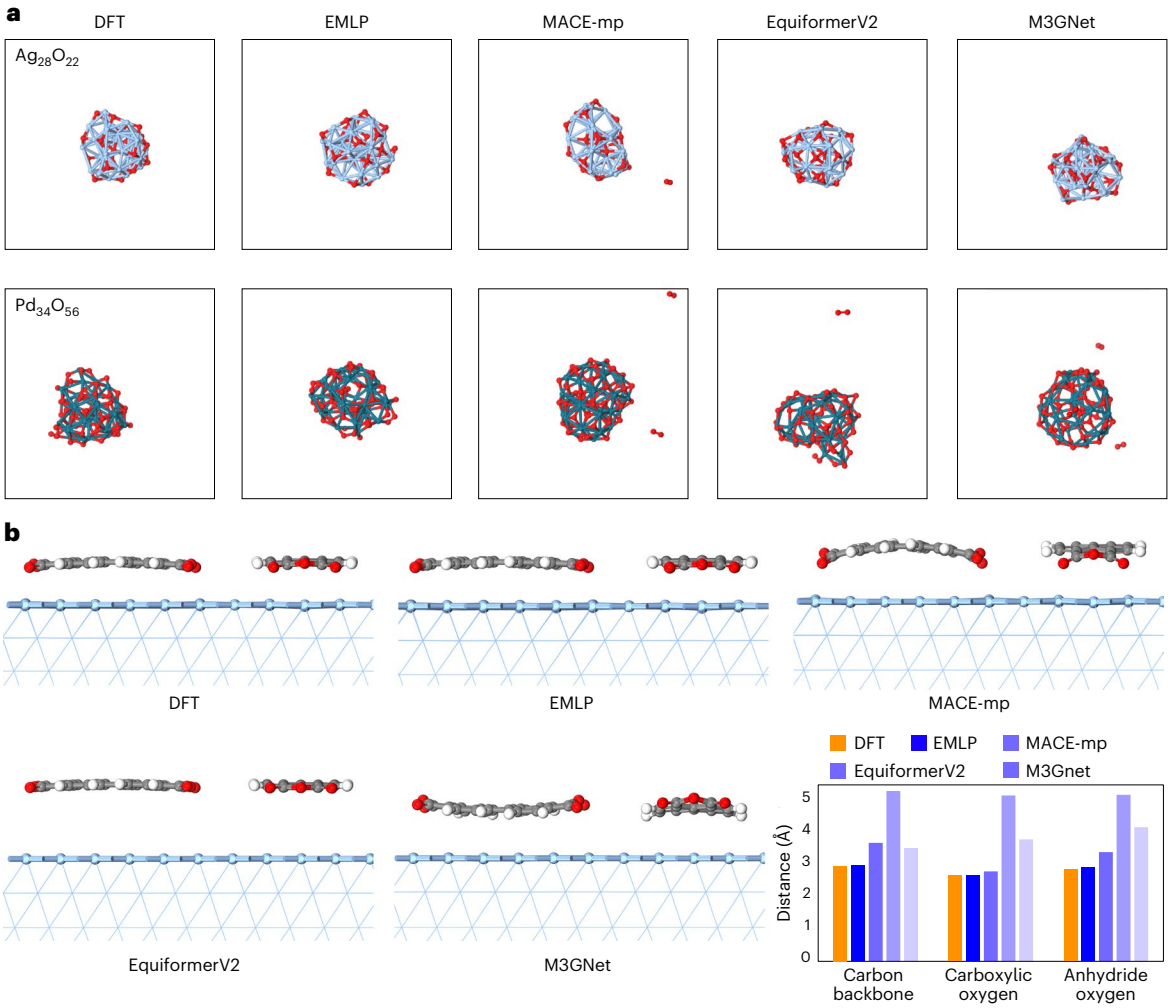


Fig. 3 | Structural predictions from the EMLP and other foundation models in MD and surface adsorption. a, Snapshots from the MD simulation of $Ag_{28}O_{22}$ and a $Pd_{34}O_{56}$ cluster by the EMLP, M3Gnet, EquiformerV2 and MACE-mp with reference to DFT. No O_2 molecules were released during the MD simulation using our EMLP, which is similar to the case of DFT, while some O_2 molecules were produced during the MD simulations from other MLPs. **b**, Structure relaxation of two PTCDA molecules on Ag(111) surface using DFT, EMLP, M3GNet, EquiformerV2 and MACE-mp. The average distance between the carbons on different parts of PTCDA and the Ag(111) surface is shown. The colour code is as in Fig. 1.

intended diversity and generality, it is crucial to quantitatively assess the generated datasets. We first analysed the diversity and completeness of our dataset compared to traditional methods. We created a benchmark MLP dataset using conventional MD techniques, selecting initial structures from the M3GNet database[36] that contained different combinations of the elements silver, palladium, carbon, hydrogen and oxygen. Figure 2a,b and Supplementary Fig. 7 provide visualizations of these high-dimensional datasets in terms of atomic energy and local atomic environments. It shows a comparison between the datasets generated by our REICO (112,732 structures) and traditional MD (104,104 structures). The REICO dataset, not only encompasses most of the scope of the MD dataset but also has a coverage significantly greater than the chemical space covered by MD sampling. The atomic interactions within the MD-sampled structures are similar, especial for metal elements where the periodic and symmetric nature of stable metal structures from public databases prevails. This can ultimately lead to overfitting, as the model becomes overly specialized in recognizing these familiar patterns and fails to generalize well to new, unseen systems.

Generality tests

The EMLP is designed for broad applicability across diverse chemical systems, extending beyond the specific configurations on which it was trained. To demonstrate its range, we conducted a series of benchmarks.

First, we started with basic atomic interactions. As shown in Fig. 2c, we evaluated the performance of the EMLP by comparing the interactions between all dimer combinations of silver, palladium, carbon, hydrogen and oxygen (see also Supplementary Figure 8 for the rest). These dimers, not specifically present in our training dataset, serve to test the EMLP across a range of bond lengths, showcasing the EMLP's capability to accurately replicate DFT results. Our EMLP predictions align closely with DFT calculations, effectively capturing the repulsion at shorter interatomic distances and accurately depicting the decreasing force as the atoms move apart—something that other models fail to achieve. These characteristics are crucial to allow the EMLP to perform robust and accurate MD simulations without experiencing unphysical holes (see below). In comparison, M3GNet fails to predict the forces in these dimer systems; EquiformerV2's predictions deviate substantially from DFT; and although MACE-mp performs better than the other two and approaches EMLP-like performance, it still shows an average error of around 10% relative to DFT.

Next, we wanted to assess whether the predicted forces yield physically meaningful dynamics. We performed MD simulations on free clusters of $Ag_{28}O_{22}$ and $Pd_{34}O_{56}$. MD simulations were performed

https://doi.org/10.1038/s41929-025-01398-3

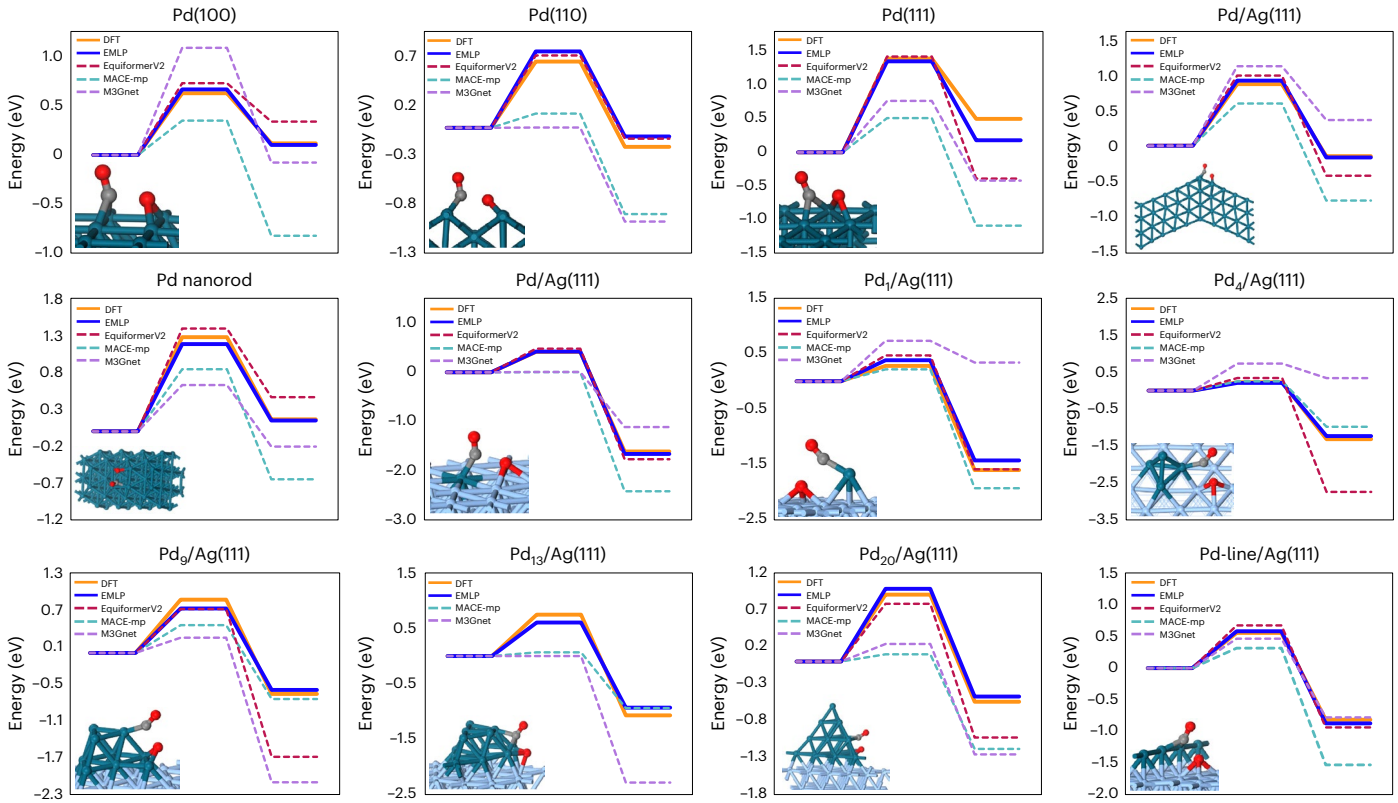


Fig. 4 | Reaction predictions of CO oxidation on various palladium and Pd/Ag surfaces. Each panel compares DFT with model predictions from EMLP, EquiformerV2, MACE-mp and M3GNet. The vertical axis shows the reaction energy referenced to the initial adsorbate state. The transition state structures obtained from EMLP are presented. The colour code is as in Fig. 1.

at 700 K and ran for 500 ps (15 ps for DFT). The snapshots are shown in Fig. 3a. Notably, only EMLP's trajectories align well with the DFT reference, while the other models show unphysical holes (oxygen atoms form O_2 and leave the cluster), causing deformations. Furthermore, to investigate the EMLP's ability to capture subtle energetic changes in finite systems, a critical requirement for studying catalyst behaviour under reaction conditions, we performed MD simulations on three distinct cluster models: a pure silver cluster, a pure palladium cluster and a bimetallic Pd/Ag cluster. As shown in Supplementary Fig. 9, the 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 show notable deviations for one or more of the tested cluster systems.

Finally, large and complex adsorbates. There are many reaction systems that are just too hard to sample using system-dependent sampling methods, especially for large reactants such as perylenetetracarboxylic dianhydride (PTCDA) which have more than 100 atoms. Previous studies, including experimental work by Hauschild et al.[37] and DFT analysis by Ruiz et al.[38], established consistent configurations of PTCDA adsorbed on the Ag(111) surface. However, building a MLP that provides accurate results for such a system often involves complex and computationally intensive sampling. To demonstrate the robust learning capabilities of our EMLP, which contains mostly training data with 5–30 atoms, we addressed the challenge of the PTCDA system containing 364 atoms. When relaxing PTCDA on Ag(111), shown in Fig. 3b, EMLP-produced geometries align closely with DFT results, accurately capturing both the shape of the adsorbed molecule and its vertical distance to the surface. In contrast, EquiformerV2 yields a nearly flat PTCDA structure but places it at an incorrect distance from the surface. M3GNet and MACE-mp generate structures with noticeable distortions (for example, bending or arching of PTCDA), suggesting that these models struggle with complex, flexible adsorbates.

Reaction systems in heterogeneous catalysis

In this section, we focus on three particularly challenging aspects of reactions in heterogeneous catalysis: complex catalyst surfaces, surface coverage and extended reaction networks. To demonstrate our EMLP's advantages for describing reactivity, we selected several classic reaction systems, including CO oxidation, acetylene hydrogenation, the Fischer–Tropsch process and ethylene epoxidation (Supplementary Fig. 10). We also evaluated M3GNet, EquiformerV2 and MACE-mp, paying special attention to transition state searches.

CO oxidation, one of the most used model systems in the study of heterogeneous catalysis, has greatly contributed to advances in fundamental catalytic research. Recent studies have highlighted the CO-induced reaction-driven metal–metal bond breaking in metal catalytic surfaces even under relatively mild conditions[39,40]. These findings emphasize the need for theoretical predictions of potential surface reconstructions under reactive conditions, which would enhance our understanding of active sites on metal catalysts and guide future experimental efforts. To test whether our EMLP can recognize different surface facets, we built three kinds of surfaces to test the reactivity of the models, namely, simple surfaces (Pd(100), Pd(110), Pd(111)), special surfaces (twinned boundary palladium and nanorod structure palladium), and various palladium-on-Ag(111) surfaces (Pd/Ag(111) single-atom catalyst, palladium adatom on Ag(111), palladium clusters on Ag(111) and palladium-line/Ag(111)). As shown in Fig. 4, the EMLP consistently delivers the most accurate reaction barriers and enthalpy changes. In contrast, MACE-mp and M3GNet often show significant deviations, especially near transition states. EquiformerV2 can occasionally match EMLP's accuracy on symmetric surfaces and bimetallic surfaces for barrier but struggles with enthalpy predictions.

Simulating surface coverage effects, a critical factor in surface reaction kinetics, using DFT approaches is notably time consuming[26,27]. The iterative method, currently considered state-of-the-art for

https://doi.org/10.1038/s41929-025-01398-3

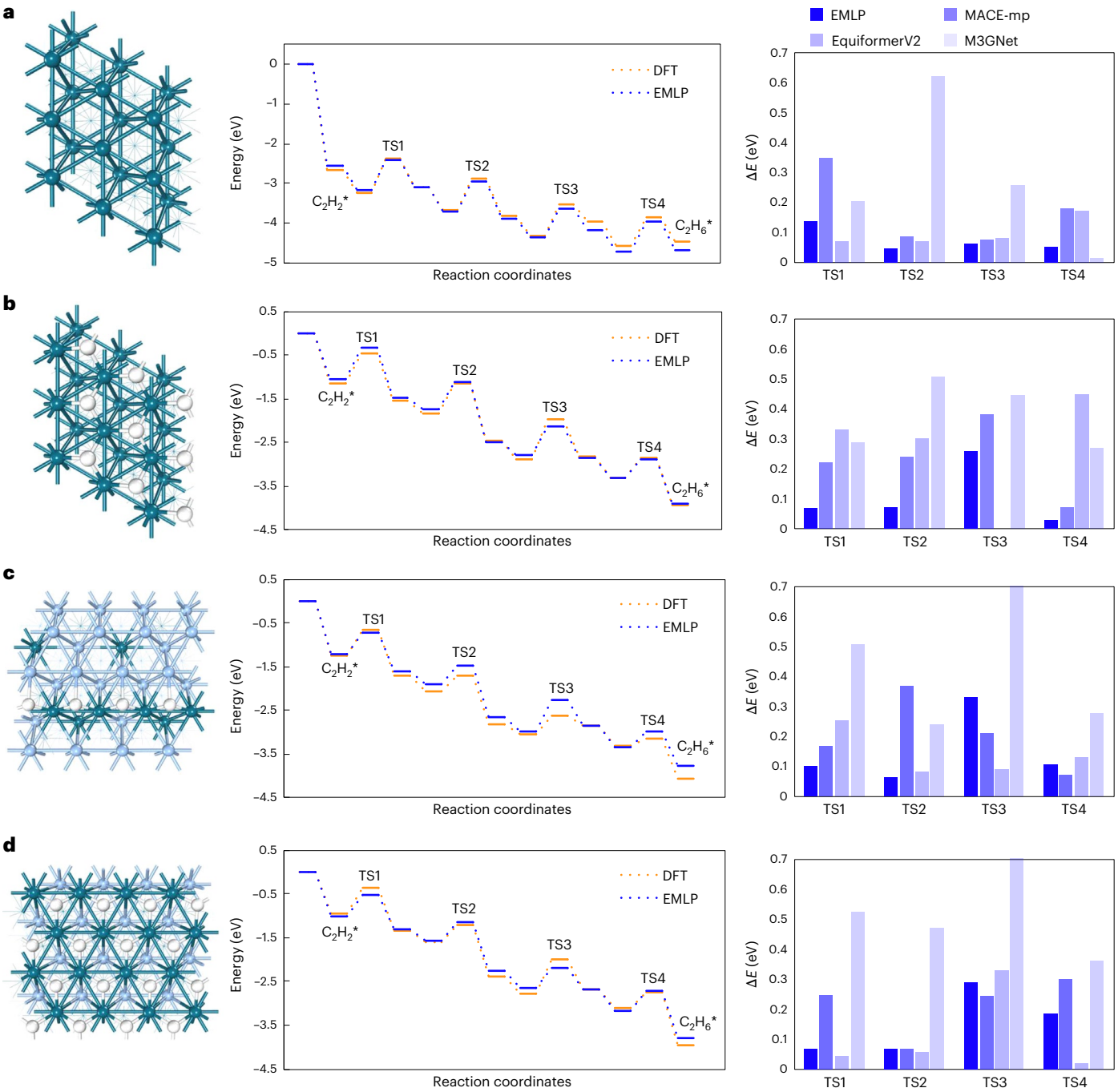


Fig. 5 | Reaction predictions of C_2H_2 hydrogenation (the stepwise addition of hydrogen atoms to C_2H_2 to form C_2H_6) on various palladium-based surfaces with coverage effect. a–d, The surface model, the energy profile comparison between DFT and EMLP, and the reaction barrier difference of each transition state from different model predictions (EMLP, EquiformerV2, MACE-mp, and M3GNet). The surfaces are: Pd(111) (**a**), Pd(111) covered by 1 ML hydrogen (**b**), $PdAg_3$(111) covered by 0.25 ML hydrogen (**c**) and PdAg(111) covered by 1 ML hydrogen (**d**). For **a** the reaction starts with the adsorption of C_2H_2 and H, whereas for **b**–**d** the reaction starts with adsorption of C_2H_2 and a H from the surface coverage. The colour code is as in Fig. 1.

calculating reaction kinetics[26,41,42], demands accurate knowledge of surface coverage effects, necessitating detailed configurations of all surface species interactions. To validate the reliability of our EMLP for such a system, we calculated the reaction pathways for acetylene hydrogenation on both clean Pd(111) and hydrogen-covered Pd(111) surfaces. As shown in Fig. 5, the energy profiles calculated by the EMLP for both coverage-independent and coverage-dependent reaction systems align closely with those calculated by DFT. Additionally, we tested the EMLP on metal alloys, particularly PdAg alloys, which are also highly regarded for acetylene hydrogenation from a previous work by Li et al.[43]. They found that the Pd/Pd_1Ag_1(111) surface with a hydrogen coverage of 1 ML (ML, monolayer) and the Pd_1Ag_3/Pd_1Ag_3(111) surface with a hydrogen coverage of 0.25 ML exhibited the lowest reaction energy barriers. Using the EMLP, we reproduced their DFT results for these surfaces, while the other models struggle to match this performance. EquiformerV2 and MACE-mp are better than M3GNet in terms of energy predictions. This demonstrates that EMLP is not only capable of effectively capturing complex interspecies interactions on catalyst surfaces but also provides a robust platform for predicting reaction kinetics across various catalytic systems.

https://doi.org/10.1038/s41929-025-01398-3

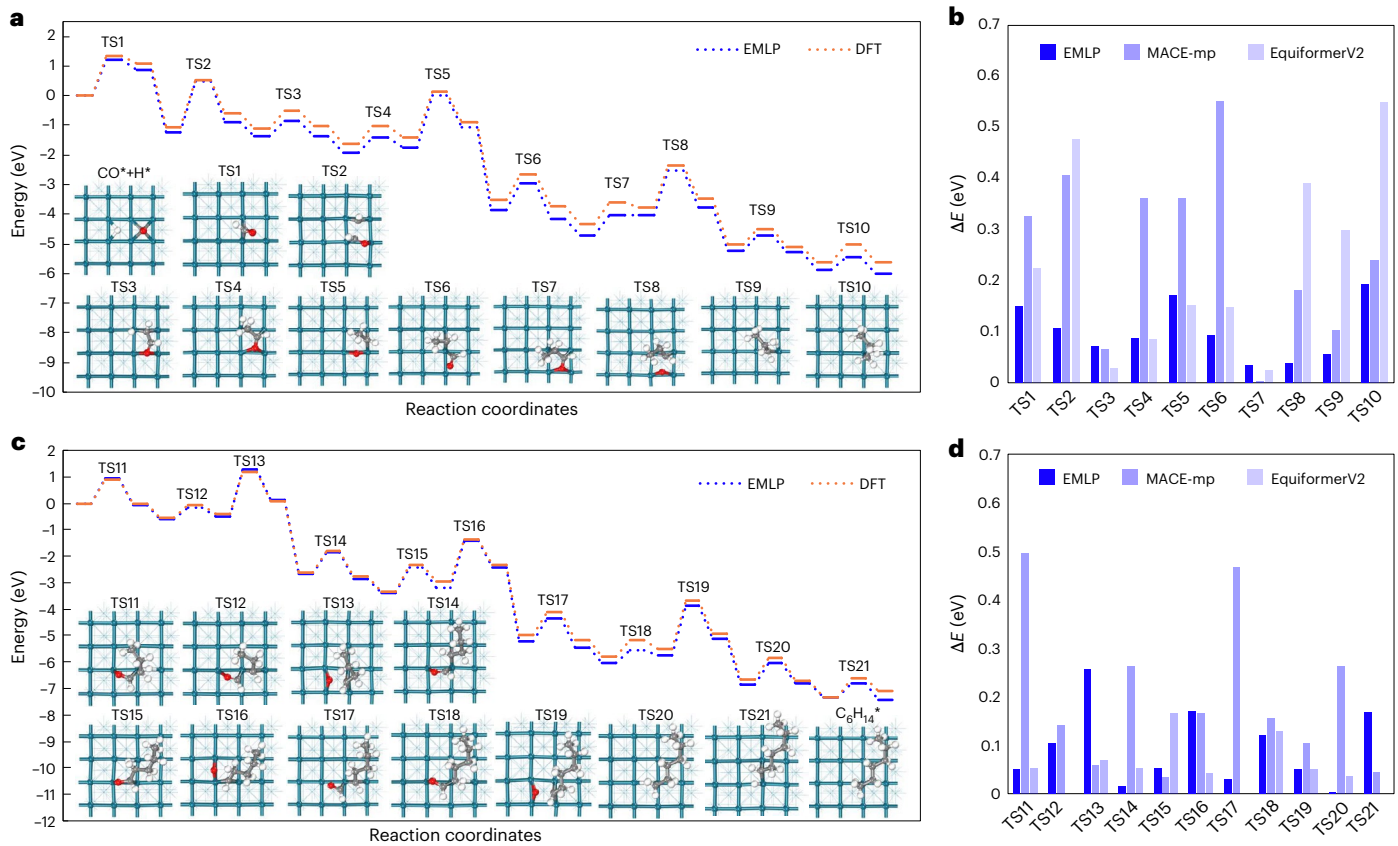


Fig. 6 | Reaction predictions of an extended reaction pathway for the Fischer–Tropsch process on Pd(100). a,**c**, The reaction energy profiles for the Fischer–Tropsch process on Pd(100) with a selected reaction pathway that contains multiple elementary steps starting with CO and ending with the formation of C_6H_{14} (see Supplementary Table 8 for a list of the elementary steps). The energy profile predicted by EMLP is compared with reference to DFT. The transition states are labelled along the reaction coordinates, with corresponding structures found by the EMLP shown below. **b**,**d**, The absolute energy deviation (ΔE) of EMLP, MACE-mp and EquiformerV2 relative to DFT for each transition state. The colour code is as in Fig. 1.

The Fischer–Tropsch process—one of the most important and complex reaction systems in heterogeneous catalysis—involves the elementary steps of chain initiation, chain growth and chain termination. The reaction systems include a large number of possible elementary steps and reaction intermediates. Here, we have selected a reaction pathway from C1 to C6. Although this process normally makes use of iron-based catalysts, for benchmarking purposes we are using Pd(100) as the catalyst in this case. The list of the elementary steps considered is shown in Supplementary Table 8. In Fig. 6a,c, we compare the entire energy profile predicted by our EMLP with the energy profile computed using DFT. The close agreement between the EMLP and DFT curves across numerous steps and complex intermediates highlights the generality and accuracy of the EMLP in capturing the energetics of intricate catalytic reaction networks. In terms of the success rate of finding transition states, from Fig. 6b,d, EMLP and MACE-mp were able to successfully find the transition states in each step of the reaction pathway, while EquiformerV2 failed to find the transition states in three reactions, R7, R12 and R21.

We have summarized the performance of our EMLP alongside two of the stronger foundation models, MACE-mp and EquiformerV2, across a range of heterogeneous catalysis reactions. In addition to evaluating energy predictions, we rigorously assessed the validity of the transition state (TS) structures predicted by each model through frequency calculations and structural similarity analysis. The average values for each system are listed in Table 1; details are shown in Supplementary Fig. 11 and Supplementary Tables 4–7 and 9. The EMLP achieves ~0.1 eV mean absolute error (MAE) for both E_a and ΔE across all the systems while maintaining a tight worst-case deviation (maximum deviation (MaxDev) ≤ 0.34 eV). In contrast, MACE-mp exhibits MAEs up to 0.59 eV and large outliers (MaxDev > 1.5 eV), and EquiformerV2 shows intermediate performance (MAE = 0.07–0.36 eV, MaxDev > 1.4 eV). Equally important, the EMLP achieves 100% transition-state localization success and ≥ 96%geometric fidelity in every case, whereas MACE-mp and EquiformerV2 struggle. Taking the Fischer–Tropsch example, from Supplementary Fig. 11, we see that the EMLP is the best in both error and transition-state structure similarity, and more importantly, it maintains a high degree of consistency in such a complex catalytic reaction network. Namely, there is a certain correspondence between the transition-state energy error and the structure error. In contrast, both MACE-mp and EquiformerV2 are unable to maintain consistence: it appears that either the structures are very similar to the DFT structures, but the energies differ by >0.5 eV, such as TS6, TS11 and TS17 for MACE-mp, or the energies are accurate but the structures differ by >20%, such as TS13, TS18 and TS19. That is, even if some of the energies are correctly calculated, this is based on coincidental error cancellation rather than the generality of the model itself, which is also consistent with our previous conclusions. The performance of MACE-mp can be attributed to the fact that catalytic activity is often surface and interface dependent, which is simply not captured in bulk datasets, thus imposing a hard ceiling on predictive performance. It is also interesting to see that EquiformerV2, which is trained on a dataset focused on surfaces with adsorbed species, failed most test cases. It might also suggest that slab models are indeed not ideal training data, as it contains too many similar atomic arrangements, limiting the model's ability to generalize.

Table 1 | Overall comparison of EMLP against MACE-mp and EquiformerV2

EMLP							
Reaction system	**No. of reactions**	**MAE E_a (eV)**	**MAE ΔE (eV)**	**MaxDev E_a (eV)**	**MaxDev ΔE (eV)**	**TS frequency**	**TS similarity**
CO+O (simple)	3	**0.06**	**0.15**	0.11	**0.32**	**100.00%**	**98.53%**
CO+O (special)	2	**0.07**	**0.02**	**0.09**	**0.02**	**100.00%**	**99.02%**
CO+O (bimetallic)	7	**0.07**	**0.09**	**0.14**	**0.17**	**100.00%**	**99.02%**
C_2H_4+H (Pd)	4	**0.07**	**0.10**	**0.14**	**0.17**	**100.00%**	**99.59%**
C_2H_4+H (Pd_H_{cov})	4	**0.11**	**0.08**	**0.26**	**0.15**	**100.00%**	**99.68%**
C_2H_4+H (PdAg$_3$_H_{cov})	4	0.15	**0.13**	**0.29**	**0.34**	**100.00%**	**98.28%**
C_2H_4+H (PdAg_H_{cov})	4	0.15	0.15	0.33	**0.25**	**100.00%**	**99.60%**
Fischer–Tropsch	21	**0.10**	**0.10**	**0.26**	**0.30**	**100.00%**	**98.82%**
MACE-mp							
CO+O (simple)	3	0.59	1.08	0.89	1.59	0.00%	95.93%
CO+O (special)	2	0.35	0.72	0.43	0.81	50.00%	91.26%
CO+O (bimetallic)	7	0.38	0.43	0.81	0.81	85.71%	79.73%
C_2H_4+H (Pd)	4	0.17	0.21	0.35	0.54	75.00%	92.67%
C_2H_4+H (Pd_H_{cov})	4	0.23	0.19	0.38	0.38	100.00%	97.90%
C_2H_4+H (PdAg$_3$_H_{cov})	4	**0.11**	0.24	0.33	0.53	100.00%	95.53%
C_2H_4+H (PdAg_H_{cov})	4	0.20	0.30	0.37	0.40	100.00%	97.15%
Fischer–Tropsch	21	0.23	0.12	0.55	0.34	90.48%	95.42%
EquiformerV2							
CO+O (simple)	3	0.07	0.40	**0.10**	0.89	66.67%	79.97%
CO+O (special)	2	0.12	0.46	0.12	0.62	100.00%	90.32%
CO+O (bimetallic)	7	0.22	0.61	0.75	1.43	71.43%	86.72%
C_2H_4+H(Pd)	4	0.10	0.46	0.17	0.76	75.00%	86.83%
C_2H_4+H (Pd_H_{cov})	4	0.36	0.96	0.45	1.36	50.00%	93.77%
C_2H_4+H (PdAg$_3$_H_{cov})	4	**0.11**	0.24	0.33	0.53	75.00%	95.10%
C_2H_4+H (PdAg_H_{cov})	4	**0.14**	**0.14**	**0.25**	0.33	100.00%	95.03%
Fischer–Tropsch	21	0.16	0.21	0.55	0.80	80.95%	87.66%

For each reaction system, we list the number of reactions evaluated; the MAE and MaxDev in predicted reaction barriers (E_a) and reaction enthalpies (ΔE); the transition state search success rate (one negative frequency at TS); and the average TS geometry similarity to DFT. Best performances are highlighted in bold. H_{cov}, surface covered by hydrogen ML.

Beyond heterogeneous catalysis

To fully demonstrate the capability and generality of our EMLP, we extended its application beyond heterogeneous catalysis; a broader range of systems from organic chemistry to surface dynamics were tested. We were pleasantly surprised by how well the EMLP performed across all test cases, showing its versatility and potential across diverse domains.

We have adapted the standard test reactions given by Baker and Chan[44], which cover a range of different reaction types and have long been used as a benchmark for transition-state searching. Nine out the 25 reactions were chosen as our EMLP only covers carbon, hydrogen and oxygen, as shown in Fig. 7a. For simpler or smaller molecular systems (for example, HCCH ↔ CCH_2, H_2CO ↔ H_2 + CO), E_a and ΔH values from the EMLP differ from DFT calculations by only around 0.1–0.2 eV. For more complex reactions, such as the cyclization of butadiene + ethylene → cyclohexene, the predicted activation barrier (0.67 eV) from the EMLP slightly overshoots the DFT value (0.46 eV) by about 0.2 eV. Nevertheless, this remains a modest deviation. The ΔE predictions match even more closely, indicating that the EMLP maintains good fidelity for moderately complicated transformations. In the case of CH_3CH_3 → CH_2CH_2 + H_2, which exhibits higher activation barriers, DFT gives rise to an E_a of 4.70 eV, while the EMLP predicts 4.23 eV, yielding an absolute difference of ~0.5 eV. However, this is still on the order of a 10% relative error, which can be considered reasonable for high-energy transition states. For reactions sensitive to conformational shifts, such as *trans*-butadiene ↔ *cis*-butadiene isomerization or vinyl alcohol ↔ acetaldehyde, the EMLP also provides E_a and ΔH values that closely align with DFT. This implies that the EMLP can effectively capture nuanced changes in molecular geometry and their corresponding energy variations.

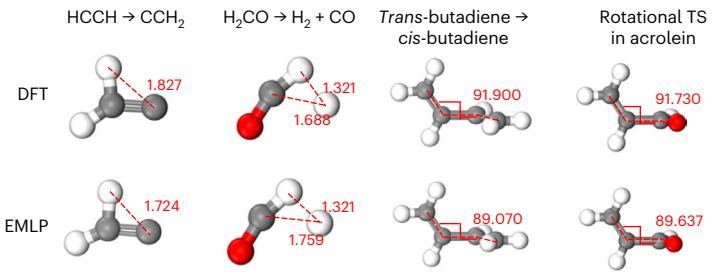
Predicting the free energy accurately with MLPs, especially in complex reaction environments such as water, presents significant challenges. Luo et al. showed that their system-specific MLP can be used for enhanced MD simulations, namely, umbrella sampling, to compute free-energy profiles for the reaction of CO with atomic oxygen on the Pt(111) surface at both solid–gas and solid–liquid interfaces[25]. We applied the EMLP to replicate and extend these findings to the Pd(111) surface, by performing enhanced MD simulations to obtain free-energy profiles for CO oxidation on Pd(111) under similar environmental conditions, including challenging aqueous environments. Figure 7b shows the structures and the free-energy profile versus the reaction coordinates for CO* + O* in the presence of water molecules, while Fig. 7c illustrates the CO* + O* reaction in the absence of water molecules at 300 K from umbrella sampling simulations. The free-energy barrier computed by the EMLP closely matches the energy barrier calculated by DFT, and the overall shape of the free-energy curve is almost identical, indicating that our general-purpose EMLP was able to capture the relevant chemistry as well as the system-specific MLP.

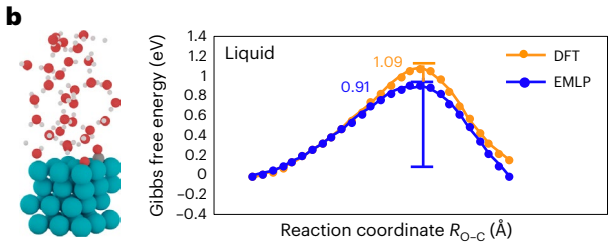
We also examined H_2 adsorption and dissociation dynamics on the Ag(100) surface. To map the potential energy surface (PES) of this

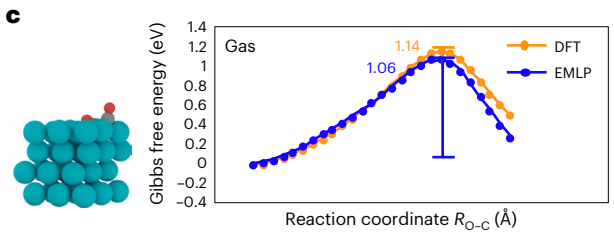
https://doi.org/10.1038/s41929-025-01398-3

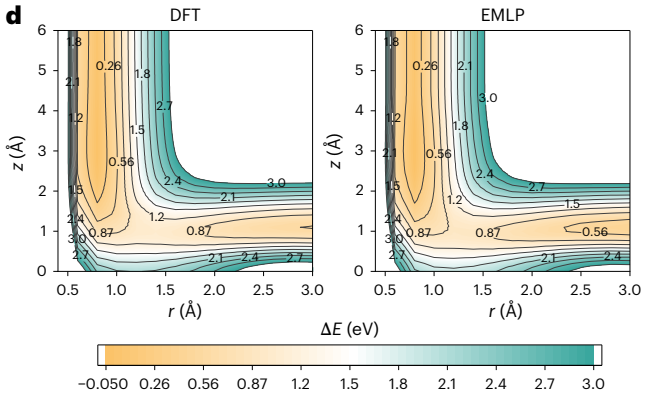
a

Reactions	No. of atoms	DFT		EMLP	
		E_a (eV)	ΔE (eV)	E_a (eV)	ΔE (eV)
HCCH → CCH_2	4	1.98	−1.97	2.09	−2.06
H_2CO → H_2 + CO	4	2.83	0.58	3.05	0.52
Bicyclo[1,1,0] → butane TS1	10	2.66	−0.84	2.75	−1.00
Butadiene + ethylene → cyclohexene	14	0.46	1.86	0.67	1.76
Trans-butadiene → *cis*-butadiene	10	0.32	−0.17	0.37	−0.12
CH_3CH_3 → CH_2CH_2 + H_2	8	4.70	−1.65	4.23	−1.84
Vinyl alcohol → acetaldehyde	7	2.25	0.37	2.02	0.59
Claisen rearrangement	14	1.65	−0.74	1.67	−0.58
Rotational TS in acrolein	8	0.38	−0.10	0.38	−0.06
MAE				0.16	0.12









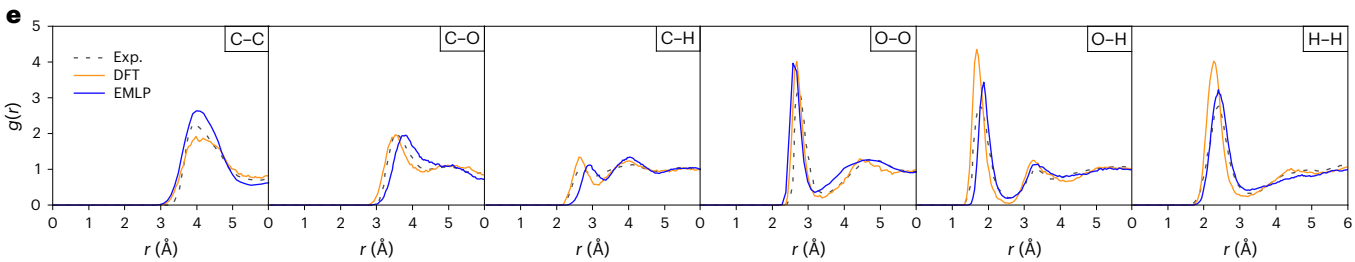


Fig. 7 | Additional benchmarks across solid, liquid and gas phases. **a**, Comparison of activation energies (E_a) and reaction enthalpies (ΔE) obtained from DFT and the EMLP for a range of reactions picked from standard test reactions given by Baker and Chan[44], along with some transition-state structural comparisons (distances given in Å). **b**,**c**, Comparison between the free-energy profiles calculated using MD with umbrella sampling from DFT and the EMLP for the CO oxidation reaction on a Pd(111) surface in the presence of explicit water molecules (**b**) and in the absence of water molecules (**c**). **d**, Comparison of two-dimensional contour plots of the DFT and EMLP PESs of scattering dynamics of H_2 on Ag(100) as a function of z and r (Å), with the molecular centre and orientation fixed at the hollow site. **e**, RDFs for liquid methanol, comparing DFT results with those from MD simulations using the EMLP. Each panel corresponds to a specific atom pair: O–O, O–H, H–H, C–C, C–H and C–O. For hydrogen-containing pairs, only hydroxyl hydrogens (from the –OH group) are included, and intramolecular pairs are excluded. The colour code is as in Fig. 1.

system and enable a direct comparison with our EMLP predictions, we conducted 854 single-point DFT and EMLP calculations at various combinations of vertical distance (z) and lateral spacing (r), while keeping the molecular centre and orientation fixed at the hollow site. Supplementary Note 11 shows structures from the dynamics simulations of H_2 adsorption and dissociation on Ag(100). The resulting two-dimensional contour plots (Fig. 7d) reveal that the EMLP-derived PES closely mirrors the DFT reference. The smooth progression of energy contours indicates that the EMLP correctly captures the interplay between vertical adsorption height and lateral site variation. The overall similarity in contour shapes, spacing and energy gradients demonstrates that the EMLP provides a near-DFT-level representation of the potential energy surface.

Moreover, we have also tested our EMLP's performance for liquid-phase simulations (Supplementary Fig. 13). Figure 7e presents six radial distribution function (RDF) plots, O–O, O–H, H–H, C–C, C–H and C–O, illustrating how closely the EMLP matches the DFT reference. Overall, the EMLP successfully replicates the peak positions and approximate intensities, demonstrating its ability to capture the key structural features of potential energy surface of methanol. Some discrepancies in the C–C or C–H RDFs may result from limited sampling of these interactions because the EMLP is primarily designed for heterogeneous catalysis. The RDF results of liquid methanol confirm that the EMLP-based MD simulation successfully reproduces the well-defined hydrogen-bonding network (indicated by the O–O and O–H peaks), the arrangement of methyl groups captured by the C–C peaks and the intermediate ordering reflected in the H–H, C–H and C–O correlations.

We also carried out further validation to demonstrate the EMLP's broad applicability, including lattice constant calculations (Supplementary Table 11), global optimization of Ag–AgO

(Supplementary Fig. 14), relaxation of rattled surface structures (Supplementary Fig. 15) and high-temperature MD (Supplementary Fig. 16).

Discussion

The consistent performance of our EMLP, which offers low average and worst-case errors, accurate transition-state finding, and high fidelity across reaction systems in heterogeneous catalysis and beyond, demonstrates its readiness for diverse applications in chemistry. This capability arises primarily from a fundamental shift in how we conceive and approach the sampling space. Building an MLP for heterogeneous catalysis is no longer just a regression problem, but fundamentally a sampling problem. Previously, discussions of sampling space in heterogeneous catalysis emphasized structural space, categorized intuitively into chemically meaningful groups such as bulk phases, surfaces or transition states. However, such categorizations carry little intrinsic significance to machine learning models. The sampling space we perceive and what the machine learning models actually see and learn are fundamentally different. There is a subtle and important reframing: the MLP does not need to understand an entire structure; it only needs to understand the physics of local atomic environments. From the model's perspective, structures are fragmented into local atomic environments, each defined by its cut-off radius. While techniques such as message-passing and attention can incorporate the information from the surroundings, the essential part still lies within the cut-off.

In this view, constructing the training set becomes an entirely different task, detached from conventional notions of meaningful structures. Based on this idea, we developed REICO, specifically designed to proficiently navigate and generate diverse and representative local atomic environments. Structures generated through optimization trajectories of imaginary chemicals encompass a rich array of bond formation and breaking, varied coordination patterns and intricate elemental interactions. This targeted diversity provides a robust foundation enabling the model, like DFT, to truly describe interatomic interactions, thereby accurately calculating arbitrary structures and capturing reactive chemistry in heterogeneous catalysis and beyond. In addition, we have briefly tested the data convergence of our method by making 25,000- and 47,000-structure versions of the Ag-Pd-C-H-O EMLP (Supplementary Figs. 17 and 18). The results demonstrate that our use of ~10,000 structures/element in the training dataset seems to be sufficient. In addition, the modest CPU and GPU requirements of our EMLP (Supplementary Fig. 19) highlight the method's efficiency and broad accessibility.

Our results show that our EMLP achieves a balance of generality and reactivity within the targeted elemental realm with only a marginal accuracy trade-off. Consequently, extending the elemental coverage of EMLP efficiently and improving its accuracy effectively pose important open challenges. We are interested in exploring whether scaling laws hold true when applied to datasets constructed via REICO. Incorporating more intricate chemical information, such as spin states, magnetism, long-range interactions and external fields, also represents a critical frontier. Encouragingly, REICO's flexibility in labelling accuracy, afforded by its use of small-scale structures, enables convenient relabelling at higher theoretical levels, enhancing our capacity to systematically explore and refine the broader potential of EMLPs. However, the highly random and disordered nature of these structures can pose convergence difficulties for high-level electronic-structure methods.

Overall, we argue that the generalizability of MLPs hinges on explicitly encoding diverse atomic interaction regimes. This principle may not only broaden MLP applicability to any system addressable by first-principles methods but also position EMLPs as a viable, efficient replacement for DFT in large-scale simulations. Our EMLP can directly simulate chemical reactions, including the precise, consistent identification of transition states across various systems in heterogeneous catalysis, a capability absent in other MLP methods. Our EMLP opens new avenues for studying reaction mechanisms and kinetics at scale. In addition, our framework can predict properties in solid, liquid and gas phases without system-specific sampling, overcoming the traditional limitation of MLPs to narrow, predefined systems. We anticipate that these advancements could contribute to the training of next-generation foundational MLPs and significantly reduce reliance on DFT calculations not only in heterogeneous catalysis but also in general chemical research and materials science in the future.

Methods

EMLP training workflow

Initialization. Our approach starts with a random structure generator. To thoroughly sample across chemical space, we generate structures with random unit cells, random numbers of atoms and random ratios between target elements. We restrict the atom distance based on covalent radii[45] to ensure no two atoms are overly proximate. For the Ag-Pd-C-H-O EMLP, we first generated 400 random configurations via REICO and relaxed them using low-precision DFT, providing sufficient data at minimal cost. A multistage selection scheme was used to filter the initial dataset to isolate the most representative 50,000 from the relaxation trajectories, which were then subjected to high-precision single-point DFT calculations to create a training dataset for the proto-EMLP. NequIP[11] was then used to train the proto-EMLP. This initialization step is optional but recommended, and one may instead perform the entire REICO sampling with low-precision DFT, as discussed in the computational cost section in Supplementary Note 18.

REICO sampling. Once the proto-EMLP has been trained, it can be used as an engine to sample the vast local atomic environments across chemical space in great efficiency. For the Ag-Pd-C-H-O EMLP, we generated an additional 20,000 random structures, using the proto-EMLP for structure relaxation. To diversify local atomic environments within our dataset, we integrate structures from public databases that represent stable atomic arrangements, which are typically elusive in randomly generated datasets. We have also performed structure search methods, such as basin hopping, MD and rattle to sample stable and metastable local environments. The outcome is a comprehensive dataset comprising 5 million random and 1 million sensible structures, from which 116,516 data points (96.7% from random structures and 3.3% from sensible structures) are selected for the final model training, after labelling with DFT.

Multistage structure selection. For data selection, a two-stage selection is proposed here for building our Ag-Pd-C-H-O EMLP. The first stage is to eliminate any extremely unstable structures by examining the force per atom histogram and setting a threshold (>50 eV A^{-1}). These unstable structures occur because the minimization of random structures using proto-EMLP sometimes brings atoms too close to each other, resulting in unfittable forces. By doing so, we see a great improvement in r.m.s.e. To distinguish between structures, candidate structures are quantified by the Smooth Overlap of Atomic Positions (SOAP)[46] descriptor and paired with cosine similarity, so that we can compare entire unit cells. Due to the high level of randomness in our dataset, we can set the similarity to 0.005 to select the most representative structures, which is not possible in a MD-created stable system-dependent dataset. We were able in this way to select 116,000 data points from a database of 6 million structures. In addition, other potential selection methods, including descriptors such as ACSF[47] and SNAP[48], and algorithms such as CUR[49], can also be used here. Therefore, we believe it is possible to add multiple stages of selection to reduce the size of the dataset and improve the model performance.

Random structure generation

The generation of random structures involved the use of random numbers generated by the NumPy[50] module in Python. Throughout the

process of searching for the different atomic environments, parameters such as lattice constants, types and quantities of elements, and atomic positions, were adjustable through the generation of random numbers. Because the primary objective of this study is to train a global potential, we opted for random generation of lattice constants, element quantities and atomic positions.

DFT calculation

In this study, all DFT computations were conducted using the Vienna Ab initio Simulation Package (VASP)[51], employing the Perdew–Burke–Ernzerhof function within the framework of the generalized gradient approximation[52]. The core–valence electron interaction was described using the projector augmented wave method[53]. We used a Monkhorst–Pack *k*-point grid with a density of 25 *k*-points per reciprocal Å, sampled at multiples of the reciprocal lattice vectors (1/25 $Å^{-1}$). For the single-point calculation, the energy cut-off is set to 450 eV. The convergence criterion for atomic forces was set to 0.05 eV $Å^{-1}$ to ensure sufficient convergence during the optimization of atomic positions, while the energy criterion for convergence was set to 5×10^{-6} eV. Additionally, the DFT-D3 method was used to further correct intermolecular forces[54]. For low-precision DFT relaxation, we set the energy cut-off to 350 eV, and *k*-point sampling utilizes only the gamma point to balance the accuracy and speed of the computation. All DFT calculations are non-spin-polarized because our primary focus is PdAg-related catalysts, which do not require spin/magnetism considerations as shown in Supplementary Note 18.

Machine learning model

This study used the NequIP software package[11] for fitting our EMLP and MLP. NequIP utilizes the E(3)-equivariant graph neural network model, demonstrating excellent fitting and generalization capabilities in MD simulations. The final dataset for the Ag-Pd-C-H-O five-element EMLP used in this study consists of 116,000 configurations calculated using highly accurate DFT.

For the development of the MD-MLP, candidate structures were selectively retrieved from the MPF2021.2.8 dataset[36], incorporating configurations that contain all or a subset of the elements silver, palladium, carbon, hydrogen and oxygen. Our initial selection yielded more than 300 structures. Noting that MPF2021.2.8 includes pairs of structures that are identical except for slight differences in orientation, we reduced redundancy by halving the dataset to 101 unique structures. Each of these structures was then subjected to a high-temperature MD simulation using LAMMPS[55], executed for 20,000 steps with a time step of 0.5 fs. To ensure comprehensive sampling of the MD trajectories, one structure was saved every 20 steps, resulting in a total of 104,104 structures. Half of the dataset were trained using the NequIP software package[12] with the same hyperparameters as the Ag-Pd-C-H-O EMLP to obtain the MD-MLP

Data diversity and coverage

The Steinhart order parameter (OP)[56] and energy of each atom in a REICO-generated dataset (112,732) and a MD-generated dataset (104,104) are calculated and plotted in Fig. 2a to characterize and compare the distribution of local atomic arrangements and environments. The Steinhart OP is calculated by equations (1) and (2):

$$\mathrm{OP}_l = \left(\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |q_{lm}|^2 \right)^{1/2} \tag{1}$$

where

$$q_{lm} = \frac{1}{N} \sum_{j=1}^{N} Y_{lm}(\mathbf{r}_{ij}) \tag{2}$$

in which i is the central atom, $\mathbf{r}_{ij}$ is the vector connecting atom i and its neighbour j, N is the number of neighbours of atom i and Y_{lm} is the spherical harmonic function of degree l and order m. Here, the Python package pyscal[57] is used to identify the neighbours of an atom and calculate the Steinhart OP, where the cut-off radius is 5 Å and l is 4. The energy of each atom is evaluated by our Ag-Pd-C-H-O EMLP, and the plotting reveals the influence of the surrounding environment on the energy of the central atom.

Generality test

We tested the generality of our EMLP by comparing the forces predicted by DFT, EMLP and other MLPs by a rigid scan of the dimer systems. All possible dimer combinations of silver, palladium, carbon, hydrogen and oxygen are calculated by both DFT and MLP models and an atom distance range from 0.6 to 2.0 times the equilibrium bond length is scanned. Here, the sum of the covalent radii of two atoms is considered to be the equilibrium bond length of the dimer, and the covalent radii dataset established in the Atomic Simulation Environment (ASE) package[58] is used in the estimation[45]. A pretrained EquiformerV2[35] checkpoint with 153 million parameters and a validation force MAE of 15.0 meV $Å^{-1}$, which was trained on OC20 S2EF-All+MD splits, is also included in the comparison as another benchmark model.

Free cluster MD

The $Ag_{28}O_{22}$ and $Pd_{34}O_{56}$ cluster models were placed in a 30 Å × 30 Å × 30 Å lattice. Initially, structural optimization was performed using DFT until the maximum atomic force was reduced to <0.05 eV $Å^{-1}$. Subsequently, MD simulations were conducted with a time step of 1 fs at a temperature of 700 K. DFT calculations were performed in VASP using the Nosé–Hoover thermostat under the NVT ensemble for a total simulation time of 15 ps. All MLP calculations (EMLP, M3GNet, EquiformerV2 and MACE-mp) were carried out using ASE, utilizing the Langevin thermostat under the NVT ensemble, with a total simulation time of 500 ps.

Surface models

This study utilized multiple surface models to test the reliability and generality of the neural network potentials. The details of the surface models are as follows.

For calculating the Ag(111)/PTCDA interface, we constructed a four-layer Ag(100) surface, with two layers of silver atoms fixed to simulate the bulk atomic environment, and used a $(6 \times 6\sqrt{3})$ supercell to ensure that two PTCDA molecules could be adsorbed on the surface.

For calculating CO oxidation, this work used a total of five different surfaces. For Pd(100) and Pd(111), four-layer (3 × 3) unicell surfaces were constructed, with the lower two layers of palladium atoms fixed. For Pd(110), a six-layer (3 × 3) unicell surface was used, with the lower two layers of palladium atoms also fixed. Finally, for the model with palladium atoms on the Ag(111) surface, a four-layer $(4 \times 2\sqrt{3})$ supercell was constructed for the Ag(111) surface. The palladium atom models included a Pd_4 cluster and a palladium single atom, with the lower two layers of silver atoms fixed in all models to simulate the bulk environment.

For calculating the reaction pathway for ethylene oxidation, a four-layer (4 × 4) Ag(111) surface was used, and the lower two layers of silver atoms were fixed to simulate the bulk environment.

For calculating the reaction pathway for acetylene hydrogenation, the Pd(111) surface used is consistent with that used in the CO oxidation calculations. The bulk structure of $PdAg_3$ used was derived from the stable structure identified in the work of Li et al.[43], and the hydrogen-covered surface structure used is also consistent with their findings.

In the study of the palladium–water interface, we constructed a four-layer Pd(111) surface using a (3 × 3) unicell and placed 30 water molecules on the surface to ensure the reasonableness of the solution simulation. In the umbrella sampling, the carbon atom in carbon

monoxide and the oxygen atom on the surface were used as collective variables. Each window had a width of approximately 0.05 Å, and a total of 39 windows were created to reconstruct the free-energy profile of carbon monoxide oxidation.

Transition-state search

This study used constrained transition-state searches[59] and the climbing image nudged elastic band[60] method to locate transition states. The climbing image nudged elastic band method was implemented using ASE[58]. All transition states converged to forces <0.05 eV Å^{-1}. DFT is used to check a single imaginary frequency to verify the transition state for all model searched transition states.

Transition-state structure similarity analysis

In this study, the SOAP[46] descriptor is used to analyse the chemical environments of carbon, hydrogen and oxygen atoms within the transition-state structures. The features extracted from these environments are averaged to construct a descriptor for the entire transition-state structure. A cut-off radius of 4.5 Å is chosen to ensure that atoms involved in surface relaxation are adequately captured while minimizing the influence of bulk atoms. For the calculation of structural similarity, this strategy is applied to generate the transition-state descriptors and the cosine similarity between these descriptors is used to quantify the structural similarity.

Umbrella sampling

In the umbrella sampling, the carbon atom in carbon monoxide and the oxygen atom on the surface were used as collective variables (CV). The harmonic potential is chosen to be the bias potential that constrains the CV, shown in equation (3):

$$V = \frac{1}{2}\kappa \times (\xi - \xi_0)^2 \tag{3}$$

The parameters of the potential are the force constant κ and the minimum or the potential ξ_0. Each window had a width of approximately 0.05 Å, and a total of 39 windows were created to reconstruct the free-energy profile of carbon monoxide oxidation using a spring force of 15 eV Å^{-1}. All simulation processes are implemented using LAMMPS and VASP.

Radial distribution functions

In this study, we selected trajectories from the methanol MD simulations over a time window of 5–10 ps, extracting one structure every 5 fs for the calculation of the RDF. All RDF calculations were performed using the VMD package, and the final RDF was obtained by averaging all individual RDF results.

Melting point MD

In this study, the melting points were simulated using the single-phase method based on the NPT ensemble. Silver and palladium models with a 5 × 5 × 5 supercell, each containing 500 atoms, were used. For the calculation of the melting point of silver, a heating rate of 6.3×10^{12} K s^{-1} was applied, with temperatures ranging from 100 K to 2,000 K. For the palladium melting point calculation, the heating rate was set to 8×10^{12} K s^{-1}, with temperatures ranging from 100 K to 2,500 K.

Data availability

The Ag-Pd-C-H-O model and its training dataset are publicly available under the Ag-Pd-C-H-O EMLP from GitHub via https://github.com/HuGroup-shanghaiTech/REICO. Details are provided in the corresponding section in Methods. All the electronic structure calculations are available from Fisgshare via https://figshare.com/articles/dataset/electronic_structure_calculations/29484686 (ref. 61). Source data are provided with this paper.

Code availability

The REICO code is available from GitHub via https://github.com/HuGroup-shanghaiTech/REICO.

References

1. Liu, Z.-P. & Hu, P. General rules for predicting where a catalytic reaction should occur on metal surfaces: a density functional theory study of C–H and C–O bond breaking/making on flat, stepped, and kinked metal surfaces. *J. Am. Chem. Soc.* **125**, 1958–1967 (2003).
2. Alavi, A., Hu, P., Deutsch, T., Silvestrelli, P. L. & Hutter, J. CO oxidation on Pt(111): an ab initio density functional theory study. *Phys. Rev. Lett.* **80**, 3650–3653 (1998).
3. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331–348 (2019).
4. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
5. Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Ed.* **56**, 12828–12840 (2017).
6. Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
7. Huang, S.-D., Shang, C., Zhang, X.-J. & Liu, Z.-P. Material discovery by combining stochastic surface walking global optimization with a neural network. *Chem. Sci.* **8**, 6327–6337 (2017).
8. Ma, S., Shang, C. & Liu, Z.-P. Heterogeneous catalysis from structure to activity via SSW-NN method. *J. Chem. Phys.* **151**, 050901 (2019).
9. Wang, H., Zhang, L., Han, J. & E, W. DeePMD-kit: a deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **228**, 178–184 (2018).
10. Zhang, Y., Xia, J. & Jiang, B. Physically motivated recursively embedded atom neural networks: incorporating local completeness and nonlocality. *Phys. Rev. Lett.* **127**, 156002 (2021).
11. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
12. Batatia, I. et al. The design space of E(3)-equivariant atom-centred interatomic potentials. *Nat. Mach. Intell.* **7**, 56–67 (2025).
13. Young, T. A., Johnston-Wood, T., Deringer, V. L. & Duarte, F. A transferable active-learning strategy for reactive molecular force fields. *Chem. Sci.* **12**, 10944–10955 (2021).
14. Smith, J. S. et al. Automated discovery of a robust interatomic potential for aluminum. *Nat. Commun.* **12**, 1257 (2021).
15. Xu, J., Xie, W., Han, Y. & Hu, P. Atomistic insights into the oxidation of flat and stepped platinum surfaces using large-scale machine learning potential-based grand-canonical Monte Carlo. *ACS Catal.* **12**, 14812–14824 (2022).
16. Liu, Y. & Guo, H. A Gaussian process based Δ-machine learning approach to reactive potential energy surfaces. *J. Phys. Chem. A* **127**, 8765–8772 (2023).
17. Nandi, A., Qu, C., Houston, P. L., Conte, R. & Bowman, J. M. Δ -machine learning for potential energy surfaces: a PIP approach to bring a DFT-based PES to CCSD(T) level of theory. *J. Chem. Phys.* **154**, 051102 (2021).
18. Morrow, J. D., Gardner, J. L. A. & Deringer, V. L. How to validate machine-learned interatomic potentials. *J. Chem. Phys.* **158**, 121501 (2023).

19. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more: sampling chemical space with active learning. *J. Chem. Phys.* **148**, 241733 (2018).
20. Schran, C. et al. Machine learning potentials for complex aqueous systems made simple. *Proc. Natl Acad. Sci. USA* **118**, e2110077118 (2021).
21. Shi, Y.-F., Kang, P.-L., Shang, C. & Liu, Z.-P. Methanol synthesis from CO_2/CO mixture on Cu–Zn catalysts from microkinetics-guided machine learning pathway search. *J. Am. Chem. Soc.* **144**, 13401–13414 (2022).
22. Han, Y., Xu, J., Xie, W., Wang, Z. & Hu, P. Comprehensive study of oxygen vacancies on the catalytic performance of ZnO for CO/H_2 activation using machine learning-accelerated first-principles simulations. *ACS Catal.* **13**, 5104–5113 (2023).
23. Han, Y., Xu, J., Xie, W., Wang, Z. & Hu, P. Unravelling the impact of metal dopants and oxygen vacancies on syngas conversion over oxides: a machine learning-accelerated study of CO activation on Cr-doped ZnO surfaces. *ACS Catal.* **13**, 15074–15086 (2023).
24. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **31**, 1902765 (2019).
25. 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).
26. Xie, W., Xu, J., Ding, Y. & Hu, P. Quantitative studies of the key aspects in selective acetylene hydrogenation on Pd(111) by microkinetic modeling with coverage effects and molecular dynamics. *ACS Catal.* **11**, 4094–4106 (2021).
27. Xie, W., Xu, J., Chen, J., Wang, H. & Hu, P. Achieving theory–experiment parity for activity and selectivity in heterogeneous catalysis using microkinetic modeling. *Acc. Chem. Res.* **55**, 1237–1248 (2022).
28. Deng, B. et al. Systematic softening in universal machine learning interatomic potentials. *NPJ Comput. Mater.* **11**, 9 (2025).
29. Zeng, J., Cao, L., Xu, M., Zhu, T. & Zhang, J. Z. H. Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation. *Nat. Commun.* **11**, 5713 (2020).
30. Schaaf, L. L., Fako, E., De, S., Schäfer, A. & Csányi, G. Accurate energy barriers for catalytic reaction pathways: an automatic training protocol for machine learning force fields. *NPJ Comput. Mater.* **9**, 180 (2023).
31. Bernstein, N., Csányi, G. & Deringer, V. L. De novo exploration and self-guided learning of potential-energy surfaces. *NPJ Comput. Mater.* **5**, 99 (2019).
32. Morrow, J. D. & Deringer, V. L. Indirect learning and physically guided validation of interatomic potential models. *J. Chem. Phys.* **157**, 104105 (2022).
33. Zhang, S. et al. Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential. *Nat. Chem.* **16**, 727–734 (2024).
34. Chiang, Y. & Benner, P. mace-universal (Revision e5ebd9b). *Hugging Face* https://doi.org/10.57967/hf/1202 (2023).
35. Liao, Y.-L., Wood, B. M., Das, A. & Smidt, T. EquiformerV2: improved equivariant transformer for scaling to higher-degree representations. In *Twelfth International Conference on Learning Representations* (2024).
36. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
37. Hauschild, A. et al. Normal-incidence X-ray standing-wave determination of the adsorption geometry of PTCDA on Ag(111): comparison of the ordered room-temperature and disordered low-temperature phases. *Phys. Rev. B* **81**, 125432 (2010).
38. Ruiz, V. G., Liu, W., Zojer, E., Scheffler, M. & Tkatchenko, A. Density-functional theory with screened van der Waals Interactions for the modeling of hybrid inorganic–organic systems. *Phys. Rev. Lett.* **108**, 146103 (2012).
39. Eren, B. et al. Activation of Cu(111) surface by decomposition into nanoclusters driven by CO adsorption. *Science* **351**, 475–478 (2016).
40. Xu, L. & Mavrikakis, M. Adsorbate-induced adatom formation on lithium, iron, cobalt, ruthenium, and rhenium surfaces. *JACS Au* **3**, 2216–2225 (2023).
41. Xie, W. & Hu, P. Influence of surface defects on activity and selectivity: a quantitative study of structure sensitivity of Pd catalysts for acetylene hydrogenation. *Catal. Sci. Technol.* **11**, 5212–5222 (2021).
42. Wu, J., Chen, D., Chen, J. & Wang, H. Structural and composition evolution of palladium catalyst for CO oxidation under steady-state reaction conditions. *J. Phys. Chem. C* **127**, 6262–6270 (2023).
43. Li, X.-T., Chen, L., Shang, C. & Liu, Z.-P. In situ surface structures of PdAg catalyst and their influence on acetylene semihydrogenation revealed by machine learning and experiment. *J. Am. Chem. Soc.* **143**, 6281–6292 (2021).
44. Baker, J. & Chan, F. The location of transition states: a comparison of Cartesian, Z-matrix, and natural internal coordinates. *J. Comput. Chem.* **17**, 888–904 (1996).
45. Cordero, B. et al. Covalent radii revisited. *Dalton Trans.* https://doi.org/10.1039/B801115J (2008).
46. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
47. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
48. Thompson, A. P., Swiler, L. P., Trott, C. R., Foiles, S. M. & Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).
49. Mahoney, M. W. & Drineas, P. CUR matrix decompositions for improved data analysis. *Proc. Natl Acad. Sci. USA* **106**, 697–702 (2009).
50. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
51. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
52. Perdew, J. P. et al. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **100**, 136406 (2008).
53. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
54. 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.* **132**, 154104 (2010).
55. Thompson, A. P. et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).
56. Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Phys. Rev. B* **28**, 784–805 (1983).
57. Menon, S., Leines, G. & Rogal, J. pyscal: a Python module for structural analysis of atomic environments. *J. Open Source Softw.* **4**, 1824 (2019).
58. Larsen, A. H. et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).

59. Wang, H.-F. & Liu, Z.-P. Comprehensive mechanism and structure-sensitivity of ethanol oxidation on platinum: new transition-state searching method for resolving the complex reaction network. *J. Am. Chem. Soc.* **130**, 10996–11004(2008).
60. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901–9904 (2000).
61. Xie, W. Electronic structure calculations. *Figshare* https://doi.org/10.6084/m9.figshare.29484686.v1 (2025).

Acknowledgements

The authors thank B. Jiang, G. Zhang and J. Xu for fruitful discussions. This work was supported by the NKRDPC (2021YFA1500700) and the National Natural Science Foundation of China NSFC (22433004 and 22403064) and ShanghaiTech University. We are also grateful for the computing time provided by the HPC Platform of ShanghaiTech University and the High-Performance Computing Center (HPCC) at Nanjing University.

Author contributions

P.H. and W.X. conceived the project and guided the research the project. The REICO code was written by C.Y. The Ag-Pd-C-H-O EMLP was trained by C.Y. and C.W. performed the generality tests and trained the MD-MLP. C.Y. performed the validation case studies. C.Y., C.W., W.X., D.X. and H.P discussed the results. W.X. wrote the original draft of the manuscript and designed Figs. 1–3 and 5–7. C.Y. designed Fig. 4. D.X. and H.P. provided the computational resources. All authors edited and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41929-025-01398-3.

Correspondence and requests for materials should be addressed to Wenbo Xie or P. Hu.

Peer review information *Nature Catalysis* thanks Brett Savoie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.