npj computational materials

www.nature.com/npjcompumats

REVIEW ARTICLE OPEN

Check for updates

Machine-learning atomic simulation for heterogeneous catalysis

Dongxiao Chen[1], Cheng Shang[1,2] and Zhi-Pan Liu[1,2,3]✉

Heterogeneous catalysis is at the heart of chemistry. New theoretical methods based on machine learning (ML) techniques that emerged in recent years provide a new avenue to disclose the structures and reaction in complex catalytic systems. Here we review briefly the history of atomic simulations in catalysis and then focus on the recent trend shifting toward ML potential calculations. The advanced methods developed by our group are outlined to illustrate how complex structures and reaction networks can be resolved using the ML potential in combination with efficient global optimization methods. The future of atomic simulation in catalysis is outlooked.

npj Computational Materials (2023)9:2 ; https://doi.org/10.1038/s41524-022-00959-5

SHORT HISTORY: FROM DFT TO MACHINE LEARNING

Heterogeneous catalysis is renowned for its great complexity in catalyst structure, and thus revealing how reactions occur on catalyst surfaces is an outstanding challenge. A major difficulty stems from the intimate coupling between catalyst surfaces and molecules during the catalytic conversion[1,2]. The catalyst surface may well reconstruct caused by the molecular adsorption, and molecules can in turn choose the best surface sites to achieve the highest reaction kinetics. Therefore, new techniques that can be operated under reaction conditions and have high spatial-temporal resolution were continuously pursued to characterize the in situ catalyst structure and to reveal the reaction mechanisms. However, to date, most experimental techniques remain frustrating to work properly under even the most common catalytic conditions, such as high-pressure conditions[3,4] and solid-liquid reaction conditions[5–7]. On the other hand, atomic simulations, particularly those based on quantum mechanics (QM) calculations, have attracted much attention in the past two decades: they are not only an indispensable complement to experimental characterization techniques, but also could provide insightful predictions to guide experimental catalyst search[8,9]. It is the purpose of this review to introduce the advance of the latest machine learning atomic simulations in heterogeneous catalysis.

Theoretical calculation emerges as a key player in catalysis research as early as the 1980s. With the advent of density functional theory calculations for periodic systems in the 1990s[10–12], the key obstacle in computing extended solid surfaces was removed, and the first principles atomic simulations soon became a popular and trustful tool in the community. Compared to empirical force field calculations and the finite system DFT calculations, the plane-wave DFT calculations for periodic systems can reach much better accuracy in computing the potential energy surface (PES) of surfaces and the interaction between molecules and surfaces[13–15]. This offers the possibility to systematically compare the adsorption and reaction of different molecules on different surfaces. Because of the relatively low scaling ($O(N\ln N)$) of DFT calculations, the systems under 100 atoms (in periodicity) can now be routinely performed on modern workstation computers. To date, tremendous progress thus has been achieved in catalysis via atomic simulation, where perhaps all known important catalytic systems have been explored by using the single-crystal bulk structures and well-defined low Miller index surfaces as the model[16–22]. Even though, DFT calculations appear still not a game changer for at least two reasons. First, the catalytic systems are generally much more complex than single-crystal surfaces, which involve many more atoms and more complex surface geometries[23–25]. Second, the computation of chemical reactions that need to locate the reaction transition state (TS) is typically one or two orders of magnitude more expensive than that of the adsorption state (initial and final states)[26–31]. Therefore, the predictive power of DFT calculations for catalytic reactions is limited to even smaller systems, not to mention the concerned DFT intrinsic errors in computing the reaction kinetics.

To allow for efficient PES exploration as required, for example, to identify the best catalytic site, many elegant theoretical methods have been proposed. While the real catalyst structure evolution must be in contact with the reaction environment (e.g., gas molecules), most theoretical simulations have to work with fixed numbers of particles in the system (constant number of atoms). For example, molecular dynamics (MD) can produce a structure evolution trajectory in real-time on a finite-temperature free energy surface[32,33], and enhanced MD methods can further speed up the exploration of designed chemical reaction events[34,35]. Unfortunately, due to the speed limit in PES evaluation, the time scale of MD simulations is generally below nano-seconds, which is far not enough for finding the most stable structure at a given composition, a global optimization problem. To this end, a number of global optimization methods were proposed since 1990 to break the time scale of MD simulation, which abandons the strict detailed-balance rules and the finite-temperature free energy effects in kinetics and focuses on finding the global minima (GM) by randomized structure perturbation. These methods include basin hopping[36,37], genetic algorithm (GA)[38–40], particle swarm optimization (PSO)[41,42], and the stochastic surface walking (SSW) method[43–45]. As a complement, variable-atom global optimization methods have also been developed,

[1]Collaborative Innovation Center of Chemistry for Energy Materials (iChEM), Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China. [2]Shanghai Qi Zhi Institution, Shanghai 200030, China. [3]Key Laboratory of Synthetic and Self-Assembly Chemistry for Organic Functional Molecules, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China. ✉email: zpliu@fudan.edu.cn

which can find the most stable structure at different compositions[46]. Even with these advances, these current global optimization methods still cannot replace the grand canonical Monte-Carlo (GCMC) method[47–50] that realizes the dynamic atom exchange under the fixed chemical potential, despite the fact that the GCMC method with the atom-wise operation in structure evolution is generally too expensive to apply to catalytic systems.

Another attractive feature of atomic simulations is the feasibility to locate the lowest energy reaction pathway, which is often hard to conclude from the experiment due to the transient nature of intermediates. The main stream of atomic simulation methods for finding reaction pathways is based on the Markov chain theory and transition state theory[51,52], where a chemical process is considered to be composed by many elementary steps and the rate of each elementary step is controlled by the barrier height between its TS and the initial state (IS). For multiple-step chemical process, the mean-field microkinetics[53,54] or kinetic Monte-Carlo (kMC) simulations[55–57] are required to transfer the knowledge of the microscopic barrier of elementary reactions to macroscopic kinetics of catalysis. While the barrier is typically determined by TS location method as the energy barrier at zero K[26–31], the free energy barrier of an elementary step could be computed by enhanced MD-based approaches once the reaction coordinate is known, such as the umbrella sampling (US) method that integrates out the potential of mean force along a designed reaction coordinate[58–60]. Nevertheless, all of these methods require a priori knowledge or pre-guess on the reaction mechanism to obtain the reaction coordinate for each elementary step. Compared to structure exploration, the finding of reaction pathways is thus not only much more demanding in computation but also needs a priori knowledge as input from experienced users.

In recent years, machine learning (ML) based atomic simulations are evolving rapidly, achieving huge progresses on both the methodology for PES evaluations to the algorithms for structure and reaction pathway sampling[61,62]. In particular, the latest neural network (NN) potential calculations can be more than 10^4 faster than DFT calculations without significant loss of accuracy[63,64]. These ML atomic simulations bypass the heavy QM calculations and utilize ML models to link the atomic coordinates with the total energies by learning the PES data from QM calculations[64]. These new techniques are bound to reshaping profoundly the research in heterogeneous catalysis[61,62,65,66].

As a representative of ML potentials, the high-dimensional neural network (NN) proposed by Behler and Parrinello is one of the most widely utilized ML framework[64,67,68]. They propose that the total energy of a system is the sum of each atomic energy, which allows to learn the total energy, an extensive quantity, by using atom-wise NNs. The input layer of atomic NN is then a set of permutation-, translation- and rotation-invariant numerical functions, the so-called atom-centered symmetry functions, that can sensitively reflect the local chemical environment of atom. To date, there are many other flavors ML models, although most are still based on the same atom-wise NN framework[68]. The structural descriptors and thus the input layer of NN are generally the most different parts among different ML potentials, for example, the convolutional neural network (CNN) and graphic neural network (GNN) are utilized to distinguish the chemical environment of atom[69–72]. In our group, we proposed a set of power-type structural descriptors (PTSDs) as the structural descriptors for fitting the global PES data, where the spherical functions and the four-body terms are introduced[73]. The complex PTSD descriptors allow the ML atomic simulation in complex material systems, such as boron cluster[73], zeolite[66,74,75], phase interface[76], and heterogeneous reactions[77–79].

In this review, we will highlight the recent progress on ML-based methodologies for solving complex PES problems in heterogeneous catalysis, both on the structure determination and reaction pathway finding. We finally discuss the future directions in theoretical catalysis, focusing on large-scale atomic simulations.

ATOMISTIC SIMULATIONS FOR EXPLORING STRUCTURES

Since catalysts are dynamic, reconstructing and interacting constantly with coming molecules under catalytic conditions, atomic simulations to explore unknown catalyst structures ideally need to be held under the grand canonical (GC) ensemble, which means particles in a system can exchange with the environment as driven by the chemical potential[80,81]. Although the grand canonical Monte-Carlo (GCMC) simulations were developed for such a purpose[47–50], it is difficult to combine GCMC with QM calculations due to the slow self-consistent-field cycles in solving Schrodinger equations. Additionally, the MC method as utilized in GCMC evolves structure via random perturbation, which is often not efficient enough to find stable configurations. Here we will introduce the progress of ML-powered global optimization and grand canonical structure search for catalyst structures, which are benefited significantly by the high speed and high accuracy of ML potentials.

Machine learning global optimization

The global optimization method was introduced to search for unknown structures from 1990s[82]. Compared to the typical MD and MC methods, the global optimization methods can better surmount high barriers between minima and avoid trapping at high-entropy regions of PES. Due to the dramatic structure change between configurations and the high computational costs, the global optimization methods have long been limited to a few toy models and small clusters[83].

Unlike most global optimization methods, the stochastic surface walking (SSW) method developed in our group utilizes small structure perturbations as driven by bias potentials to move a structure from one minimum to another[43–45]. The transition region between minima can be properly visited by SSW which allows the reaction pathway searching during the global optimization. The early applications of SSW include both the finding of the global minimum of clusters[43], and the phase transition pathways between bulk crystal materials[45]. The SSW method provides a powerful and convenient solution for PES exploration and thus becomes an ideal tool for PES data generation. Indeed, in 2017 a global-to-global scheme as introduced by our group to combine SSW with NN endows the global optimization ability to ML potentials[84]. The so-called global NN potential (G-NN) have good transferability and can be utilized to explore unknown structures with arbitrary compositions starting from random structures. The SSW-NN methods have been applied to a wide range of PES problems, in particularly those related to complex surface and interface structures in heterogeneous catalysis[85].

Figure 1a gives an example of how SSW-NN determines the famous silver surface oxide structure, $Ag_{12}O_6$ phase on p(4×4) periodicity of Ag(111)[86–88]. First, a Ag-O global NN potential is generated by SSW-NN method, which is iteratively trained by learning a wide range of different Ag-O structures, including the bulk and surfaces for metal Ag and AgO_x oxides. In searching for the surface oxide structure, the initial structure ('IS' in the figure) can randomly generated, for example, from a 1 ns of high-temperature (2000 K) NN-based MD simulations that contains O_2 species, surface oxygen, subsurface oxygen, and silver vacancies. Next, the SSW-NN is conducted to locate the global minimum (GM) of this structure. A 2000-step SSW-NN trajectory at a given $Ag_{12}O_6$ composition is shown in Fig. 1a, where the energy drops rapidly in the early stage, for example, the structure in step-372 is only 0.2 eV higher than the GM. The SSW trajectory also reveals the stable structure patterns of surface oxide, namely (i) the subsurface oxygen is not stable (str **1**

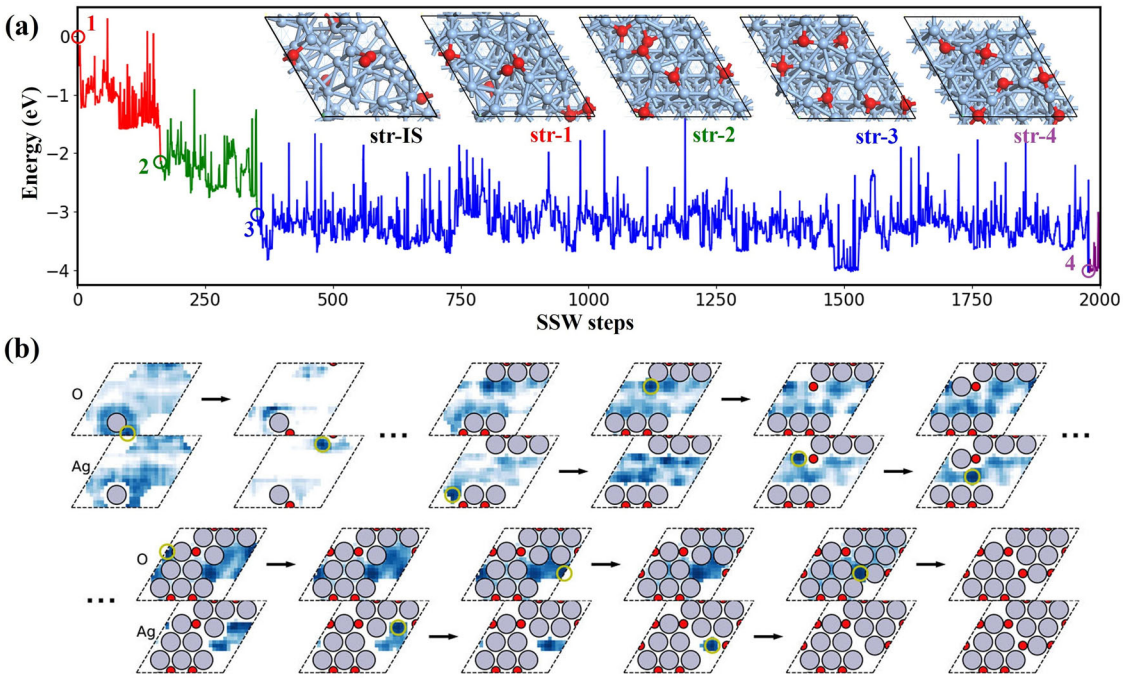

Fig. 1 The identification of silver surface oxide. a The 2000-step SSW-NN trajectory for identifying the structure of silver surface oxide ($Ag_{12}O_6$ phase on p(4 × 4) supercell of Ag(111)). The initial structure (IS) and four representative structures under different stage of optimization during SSW-NN are shown. **b** The building of $Ag_{12}O_6$ optimal surface phase with the given input of p(4 × 4) supercell and $Ag_{12}O_6$ composition by reinforcement learning based atomistic structure learning algorithm, in which the depth of blue in the figure represents the Q-values (the preference for the location of the next atom) predicted by the CNN, reproduced with permission from ref. [89]. Copyright American Institute of Physics, 2019.

→ **2**, red part), (ii) molecular O_2 species, if present, will dissociate on the surface (str **2** → **3**, green part), and (iii) the optimal Ag-O arrangement (str **3** → **4**, blue part). Finally, the GM structure, identical to that reported previously[87,88], is identified after 1979 SSW steps, which is verified by continuing SSW search for another 10,000 steps and no more stable structures is found.

For the same Ag-O surface oxide system, Hammer group has developed the atomistic structure learning algorithm (ASLA) to identify the global minimum, as shown in Fig. 1b[89,90]. The method builds 2D structures using an atom-by-atom strategy based on a reinforcement learning framework. In this method, CNN is used to acquire the knowledge of a surface structure and output the Q-values used for locating the next atom in ε-greedy policy. DFT calculations are performed to evaluate the stability of the predicted structures. It takes more than 3990 episodes (one trial for building $Ag_{12}O_6$ phase is one episode) to resolve the $Ag_{12}O_6$ global minimum phase. Apparently, if DFT can be replaced by ML potentials, the searching speed should be dramatically dropped. This demonstrates, on the other hand, that ML techniques can be combined versatility with PES evaluation methods to solve complex PES problems.

ASOP algorithm

The fast global optimization ability of SSW-NN opens the door for searching structures with variable compositions in a GC ensemble. We have developed the automated search for optimal surface phases (ASOP)[91], which can explore the composition space under a predefined chemical potential condition. As shown in Fig. 2a, the ASOP simulation is based on a multi-grid framework to scan all likely compositions. It takes only simple inputs, including the bulk crystal structure, the surface Miller index, and the chemical potentials of exchanging particles, and outputs a phase diagram, including a list of stable phases.

In one ASOP simulation, the whole composition space is discretized into a series of grids from coarse to fine, in which each grid represents a unique surface periodicity ordered by the surface area from small to large. Then, several cycles of SSW-NN structure exploration and Monte-Carlo selection are conducted for the compositions in each grid. The simulation starts from the coarsest grid and progressively explores the larger grids, where the knowledge of the stability of compositions is inherited from grid to grid. The SSW-NN sweeps over each composition with only a few steps (e.g., smaller than 400) since the energetically favorable compositions will be biased and visited again in the subsequent cycles. The Monte-Carlo scheme is utilized to select the focus zone containing energetically favorable compositions.

Taking Ag(100) oxidation under typical ethene epoxidation conditions (500 K, 1 bar of O_2) as the example, the ASOP algorithm can identify the top stable silver surface oxides within 90 hours on 80 CPU cores. Figure 2b shows the PES map and energy spectrum of surface oxides, in which the optimal zones locate at the Ag and O coverages being 0.6~1.0 and 0.5~0.9 ML, respectively (see the blue zone in the figure). Figure 2c plots the geometries of phase-1, the most stable structure from ASOP, which differs from an experimentally observed surface oxide at low O_2 pressure, phase-2[92,93]. The phase-1 possesses a Ag_7O_5 stoichiometry with O coverage of 0.625 ML, which contains the planar coordinated [AgO_4] motif with high O-density. By contrast, the phase-2 has a lower O coverage (0.50 ML) and is less sable (0.025 J/m^2, see energy spectrum of Fig. 2b), which features a missing-row reconstruction induced by atomic O adsorption. These results indicate that Ag(100) surface under typical ethene epoxidation conditions are in fact at a deeper oxidized state comparing to the surface science experiment. The Ag catalyst surface is highly dynamic under different O_2 pressures.

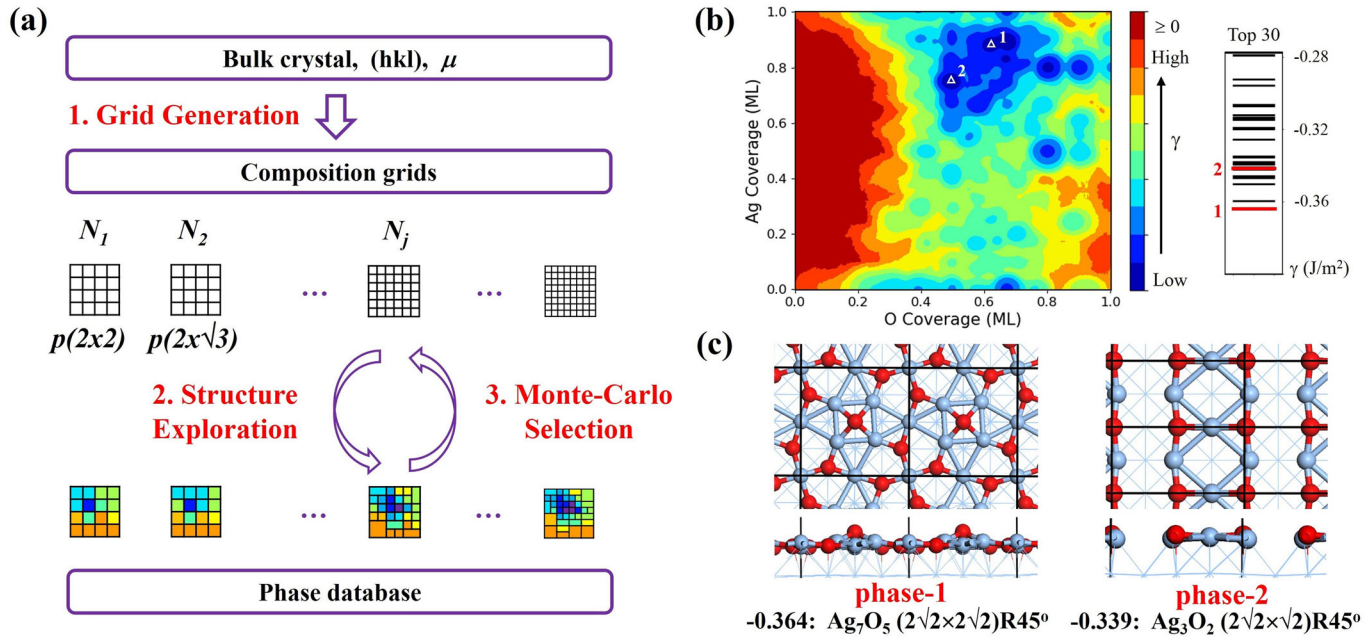

Fig. 2 The ASOP algorithm and its application. a Flowchart of the ASOP algorithm that contains three steps. **b** The PES contour map and energy spectrum for silver surface oxides on Ag(100) from ASOP simulation. **c** The geometries of two phases, as labeled in both PES map and energy spectrum. Reproduced with permission from ref. [91]. Copyright American Institute of Physics, 2022.

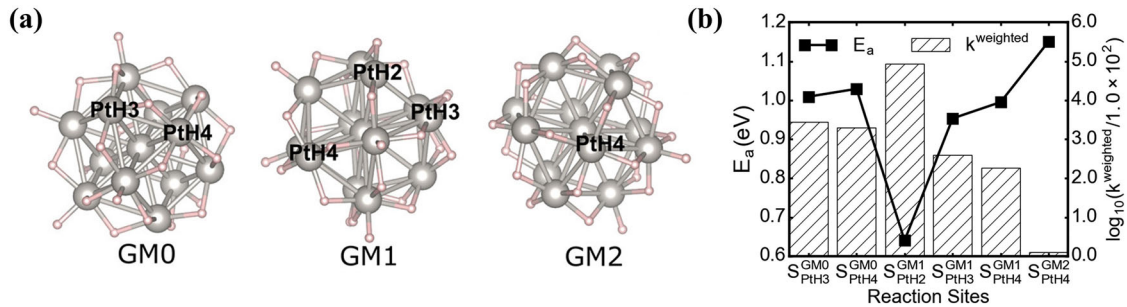

Fig. 3 Pt clusters for methane activation. a The geometries of top three stable $Pt_{13}H_{26}$ clusters (denoted as GM0, GM1, and GM2) as explored by the HDNN potential together with modified genetic algorithm, in which the note 'PtH_x' represents the coordination number (x) of Pt. **b** The activation energies (E_a) and relative contributions to the reaction rate of methane activation under 673 K. Reproduced with permission from ref. [97]. Copyright American Chemical Society, 2018.

REACTION ACTIVITY PREDICTION

The determination of the catalyst structures is generally not enough for predicting reactivity in heterogeneous catalysis. The molecules can adsorb, diffuse and react at a vast number of local sites, which may only have trivial differences in geometry but can provide very different reactivity. Traditionally, the reaction pathway at each site needs to be determined by locating all likely transition states (TS) in order to predict correctly the overall activity[51,52]. The TS search can be accelerated by the NN potential, but the pre-guess of reaction coordinate remains to be a key obstacle in unknown reactions. The Brønsted-Evans-Polanyi (BEP) relationship[94–96], that correlates the reaction energy (adsorption energy difference between the initial and the final states) with the reaction barrier is often utilized to speed up the reactivity prediction, although the accuracy of BEP is generally poor and may only be used as a coarse screening tool.

Recently, Sun and Sautet demonstrated further the complexity of reaction pathways on nanocatalysts using a concept of catalyst fluxionality[97–99]. They utilized the HDNN potential together with a modified genetic algorithm to explore extensively the low energy metastable ensemble (LEME) of $Pt_{13}H_x$ clusters and then evaluates the hydrogen evolution reaction (HER) and methane activation activities on these clusters[97]. Fig. 3a shows the top three stable $Pt_{13}H_{26}$ clusters in its LEME, in which the Pt atoms exhibit various coordination numbers, such as four-coordinated Pt (PtH_4, see GM0, GM1, and GM2), three-coordinated Pt (PtH_3, see GM0 and GM1), and two-coordinated Pt (PtH_2, see only GM1). The kinetic evaluation of all these clusters in Fig. 3b indicate the PtH_2 site is the most active site with the lowest activation energy, which locates on energetically less favored GM1 compared to GM0. Notably, even with the consideration of presence probability, the weight reaction rate of GM1 is still 30 times higher than GM0, indicating the fluxionality of active site from the most stable structure to metastable isomer, which can hardly be noticed without the machine learning accelerated large-scale structure sampling.

The ultimate goal of reaction activity prediction in the heterogeneous reaction is to discover reaction channels in an automated way, which should take into account the intimate coupling between catalyst structure and molecular reactions. Since the complexity of the reaction network grows exponentially when many elementary reactions and likely reaction sites are present, an intelligent on-the-fly pathway explorer is essential in order to identify the kinetically relevant reactions by the algorithm itself. For this purpose, new methods have developed by

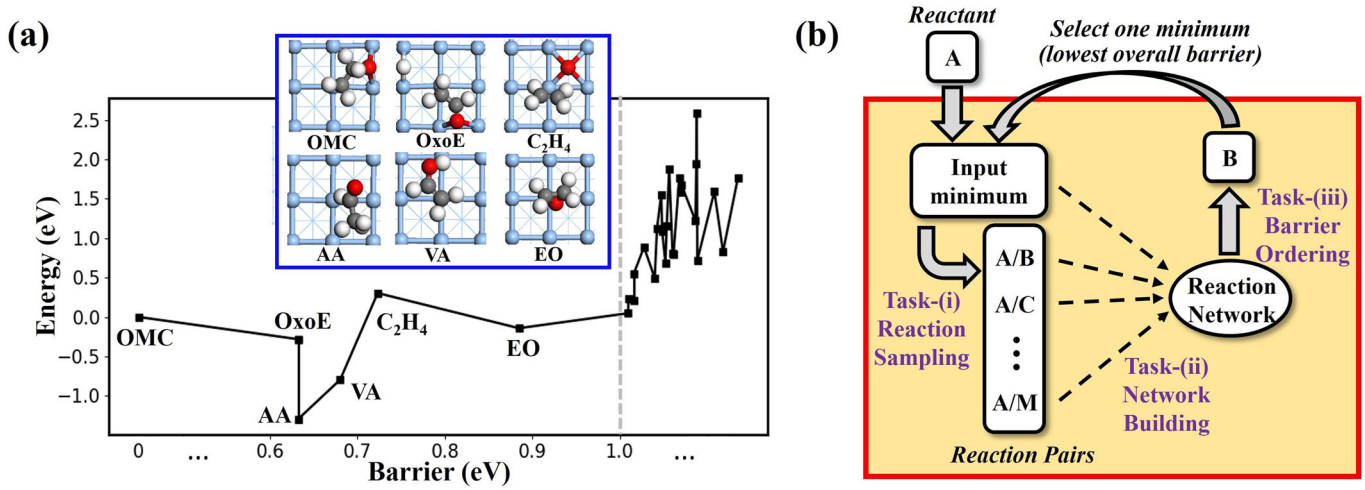

Fig. 4 The SSW-RS method and its application. a The barrier-energy map of the sampled products of oxometallacycle (OMC) intermediate on p(4 × 4) supercell of Ag(100) after 10000-step SSW-RS. Abbreviations: OxoE, 2-oxoethyl; AA, acetaldehyde; VA, vinyl alcohol; EO, ethene oxide. **b** The diagram of the general SSW-RS based pathway sampling algorithm, reproduced with permission from ref. [79]. Copyright American Chemical Society, 2021.

exploiting the SSW global optimization for PES search and the G-NN potential for PES evaluation.

SSW-RS method

The SSW reaction sampling (SSW-RS) method modifies the SSW global optimization for pathway collection during the SSW PES search. Specifically, in the SSW-RS simulation, the structure characteristics, such as the bond matrix and chirality of the initial structure, are remembered and used to judge whether a new structure from SSW is identical to the initial structure. If it is false, SSW will output a reaction pair for the subsequent pathway search and the current structure is back to the initial structure to continue the SSW search. Finally, the TS location for all reaction pairs is performed by the double-ended surface walking (DESW)[31], which can identify iteratively all TSs along the pathway connecting a reaction pair. The SSW-RS method can be combined with DFT and G-NN calculations for studying different type of reactions, from gas phase reactions[100–102] to solid phase transitions[103–105].

Figure 4(a) illustrates an SSW-RS example for finding the reaction pathways for the oxometallacycle (OMC) intermediate on Ag(100), which is known to be a critical intermediate in ethene epoxidation[106–108]. It was generally believed that OMC can further oxidize to AA and EO with similar barriers on metal Ag sites and thus the EO selectivity on Ag catalyst is about 50 %[106,109,110].

By using SSW-RS, a total 864 qualified reaction pairs are collected from 10000 SSW-RS steps, and after DESW pathway search, 29 distinct products are identified. In these products, only five of them are kinetically favored with a barrier smaller than 1 eV (see Fig. 4a), i.e., OxoE (2-oxoethyl), AA (acetaldehyde), VA (vinyl alcohol), ethene, EO (ethene oxide). The SSW-RS identifies a new intermediate OxoE and much lower formation barrier of AA than EO, in which the route from OMC to AA should be indirect via OMC → OxoE → AA[79]. This finding rules out the metal Ag sites as the active site for EO production, and the active site for EO production must be Ag surface oxides.

For multiple-step reactions, SSW-RS can be iteratively performed and the whole reaction network can be resolved finally, as shown in Fig. 4b[79]. Starting from a given reactant A, the first SSW-RS is performed to sample the possible reaction pairs connecting to the reactant A and locate all the TSs, then the reaction network is updated based on the sampled new reaction channels. Next, according to the order of the calculated barriers, the kinetically favorable intermediates are chosen to start new SSW-RS tasks. By performing SSW-RS to scan all likely intermediates, the lowest reaction pathways from reactant A to any products can in principle be obtained in an automated manner. This simple algorithm meets difficulties when the reaction network is too complex with too many elementary reactions. The AI-Cat and MMLPS algorithms are thus developed to cope with the challenges relating to complex reaction networks.

AI-Cat algorithm

Kang et al. proposed an end-to-end artificial intelligence framework for the activity prediction of heterogeneous catalytic systems (AI-Cat)[111]. The AI-Cat method is a general framework for predicting the kinetics of heterogeneous catalytic processes where the reaction network is too complex to resolve by using SSW-RS. The method first needs a general reaction database containing information on common elementary reactions, including the structures and energetics of initial state (IS), final state (FS), and their TS. The database can be obtained by the SSW-RS simulations introduced above. The structures in the reaction database are then encoded by Surface-sensitive atom-centered Extended-Connectivity FingerPrint (s-ECFP). To learn the reaction information, two neural networks, namely the reaction pattern (R-Pat) unit and the kinetics information (K-Info) unit, are built for predicting the most likely reaction patterns (RPs) (i.e., the coding of IS/FS pair) from a given IS and predicting the kinetics data (reaction barrier and energy) of this IS/FS pair, respectively. The reaction pathways between any given reactant and product are determined using the Monte-Carlo tree search method, which inquires the R-Pat unit and the K-Info unit in making decisions.

Figure 5a shows a MC tree search step, namely the expansion step, from a father node to child nodes. The input reactant, the father node, firstly go through R-Pat unit to generate all the possible RPs. Next, the top RPs are selected and go through K-Info unit to predict the associated reaction energies and barriers. Finally, the kinetically favorable products, the child nodes are determined.

By using SSW-RS to collect the reaction data for common C1-C4 organic compounds on Cu(111), Cu(100), and Cu(211), and training R-Pat unit and K-Info unit based on the database, the AI-Cat method is able to explore the glycerol hydrolysis pathways on Cu(111), which is significant in biomass-derived polyol utilization[112,113]. Fig. 5b shows the Gibbs free energy profile for the AI-Cat predicted lowest-energy pathways from glycerol to 1,2-propanediol (1,2-PDO, blue solid line) and 1,3-propanediol (1,3-

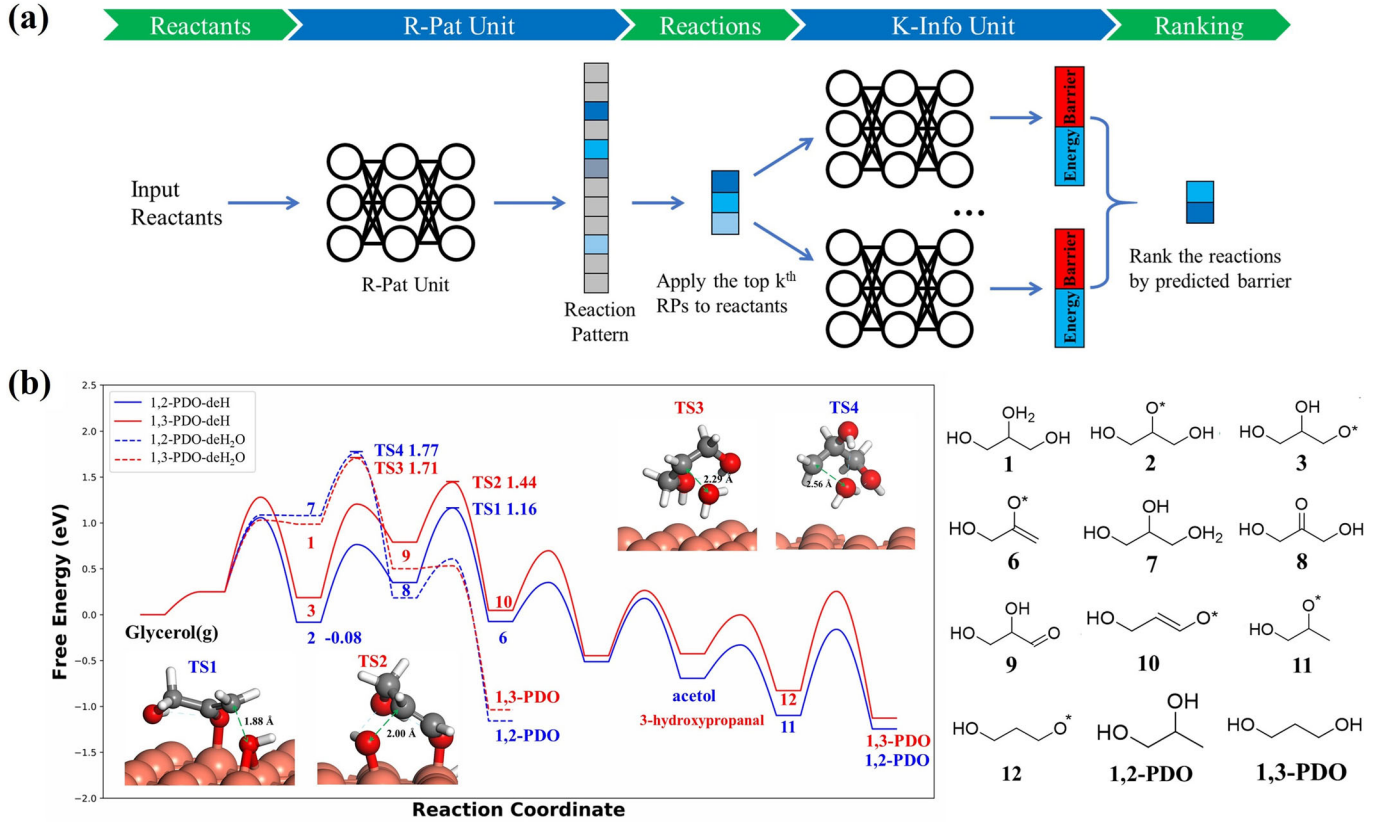

Fig. 5 The AI-Cat method and its application. a The expansion from father nodes (reactants) to child nodes (products) in AI-Cat method. **b** The Gibbs free energy profile for four low energy pathways of glycerol hydrogenolysis to 1,2-PDO and 1,3-PDO under typical experimental conditions (473 K, total pressure of 1 atm with glycerol: H_2: H_2O = 1: 140: 12). Reproduced with permission from ref. [111]. Copyright Royal Society of Chemistry, 2022.

PDO, red solid line). In the rate-determining step (TS1 and TS2), the overall barrier of 1,2-PDO path is 0.22 eV lower than 1,3-PDO path (TS2 − (TS1 + state 2)), indicating the hydroxyl on the terminal C favors dissociation more than central C. Therefore, the 1,2-PDO product after TS1 is more selective than 1,3-PDO after TS2, thus explaining the long-standing high selectivity puzzle of 1,2-PDO on Cu surfaces[114]. It is worth noting that the popular glyceraldehyde mechanism in literature that involves a dehydration process (TS3 and TS4)[115,116], is neither kinetically favored nor able to explain the high 1,2-PDO selectivity (see TS1, TS2 < TS3, TS4 and TS4 > TS3 in the figure).

MMLPS algorithm

Shi et al. proposed a microkinetics-guided machine learning pathway search method (MMLPS) method to speed up the buildup of the reaction database by SSW-RS[117]. The MMLPS aims to fast build a reaction database for a target reaction and identify the kinetically favorable pathway. Compared to AI-Cat method which is more general and can predict unknown reactions, the MMLPS is the tool to get the accurate kinetics for a target reaction. Specifically, the MMLPS simulation divides a reaction network into several parts with different molecules and surface coverages, then each SSW-RS branch samples independently different parts of the reaction PES as guided by a fast microkinetics solver. A reaction dataset is established by merging reactions from all branches, and the microkinetics simulation is performed to identify the lowest barrier reaction pathway.

Figure 6(a) illustrates the complete 2D reaction map of CO and CO_2 hydrogenation on Cu and Zn-alloyed Cu surface, as plotted from 14958 reaction pairs sampled by MMLPS. These reaction pairs are collected from three batches, i.e., $CO_2 + H_2$, HCOOH + H_2, and HCHO + H_2, which depicts the whole reaction channel from CO/CO_2 to methanol. On all surfaces, CO_2 hydrogenates via the formate pathway (CO_2 − HCOO* − HCOOH* − H_2COOH* − HCHO* − CH_3O* − CH_3OH* − CH_3OH), and CO hydrogenates pathway via the formyl pathway (CO − CO* − CHO* − HCHO* − CH_3O* − CH_3OH* − CH_3OH), as shown by the free energy profile in Fig. 6b, c. On Cu(211), only 1.40 eV is required for CO_2 hydrogenation, which is 0.05 eV lower than that of CO, indicating that CO_2 is the main carbon source in methanol products. The microkinetics based on MMLPS data reveals that Zn alloying has no obvious kinetics promotion to CO_2/CO hydrogenation reaction and a high coverage of Zn would even poison the catalyst (yellow and blue lines in the figure).

CONCLUSIONS

This review outlines recent advances in ML potential-based atomic simulations for heterogeneous catalysis. The high accuracy and high speed of ML potentials accelerate greatly the PES exploration and thus allow the development of new algorithms to solve long-standing challenges in heterogeneous catalysis.

In particular, we illustrate a few key advances in realizing ML potential atomic simulations, including the atomic-based ML model to fit the total energy that is assumed to be a sum of single atom energies, the structure descriptors to distinguish the chemical environment of atoms, and the global-to-global scheme to generate global PES dataset and train ML potential. As the representative, the SSW-NN method and the methods developed upon SSW-NN, such as ASOP, AI-Cat, and MMLPS, can now tackle complex catalysis problems, such as the surface phase diagram under the grand canonical ensemble, the end-to-end reaction

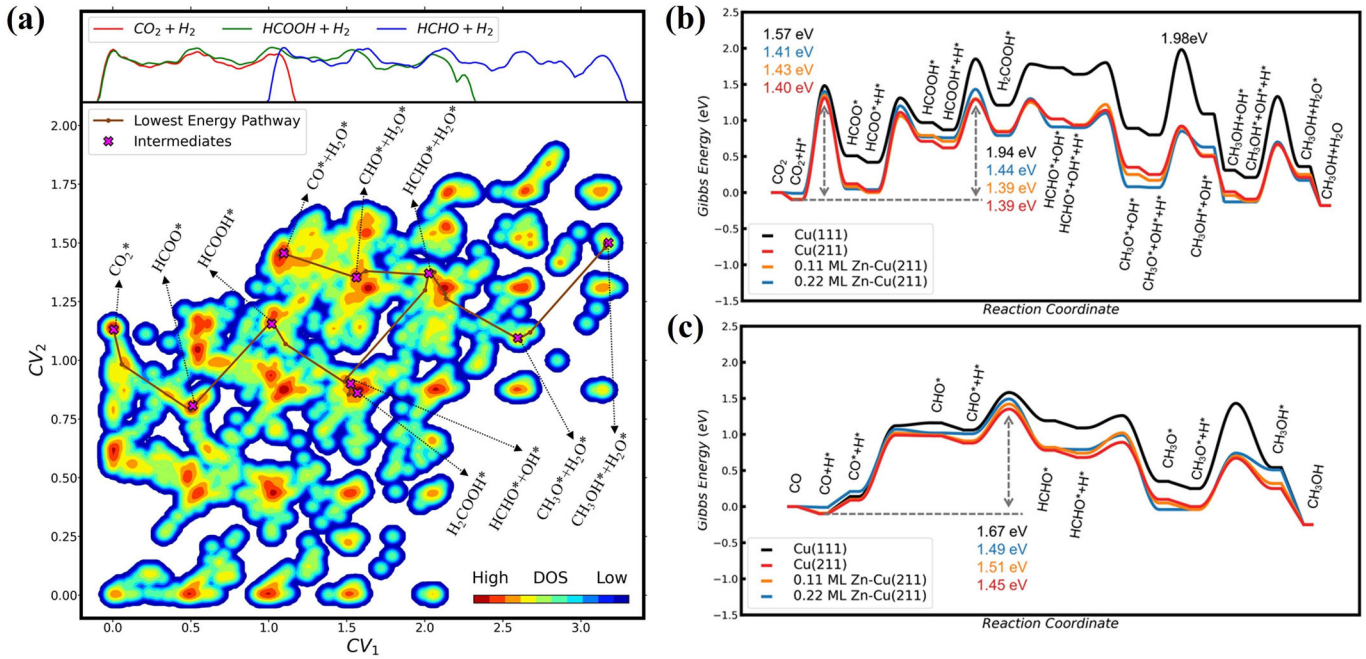

Fig. 6 The MMLPS method and its application. a Contour plot for 14958 reaction pair (IS, TS, FS) obtained by MMLPS on Cu(211). **b** CO_2 and **c** CO hydrogenation Gibbs energy profile on Cu and Zn alloyed Cu surface. Reproduced with permission from ref. [117]. Copyright American Chemical Society, 2022.

prediction based on the known reaction database, and the automatic determination of the reaction kinetics in a complex reaction network.

Nevertheless, the current algorithms, even with ML potential calculations, generally only consider the weak coupling between catalyst structure and reaction by either focusing on the thermodynamics in structure evolution or resolving the reaction kinetics on well-defined surfaces. This is apparently due to the too-large degrees of freedom if both surface flexibility and the reaction varieties are treated at the same time. In some catalytic systems, this approximation may lead to misleading or even wrong conclusions, for example, the reactions on nanoclusters as shown in reactions on supported Pt_{13} cluster. Ag-catalyzed ethene epoxidation is another example: despite the metal Ag is now ruled out as the active site for epoxidation, the true active sites of Ag surface oxides remain largely unclear, which is strongly influenced by the reaction atmosphere (ethene and O_2 pressures).

One possible and feasible solution to treat strong-coupling systems could be an automatic focus on transient structures during PES search, including the meta-stable phases and the TS structures of molecular reactions. This is likely achieved by the on-the-fly constrained PES searching procedure[98,100], which identifies the critical reaction coordinate and the lowest energy reaction channel by constraining PES search at the target reaction coordinate (e.g., important chemical bonds). During such exploration of reaction pathways, the surface phases should be allowed to change by considering the grand canonical reaction conditions. Given the recent progresses, one can be sure that the methodology advance and better applications of ML techniques in heterogeneous catalysis are coming in the future.

Received: 29 October 2022; Accepted: 22 December 2022;
Published online: 07 January 2023

REFERENCES

1. Schlögl, R. Heterogeneous catalysis. *Angew. Chem., Int. Ed.* **54**, 3465–3520 (2015).
2. Vogt, C. & Weckhuysen, B. M. The concept of active site in heterogeneous catalysis. *Nat. Rev. Chem.* **6**, 89–111 (2022).
3. Liu, H. Ammonia synthesis catalyst 100 years: Practice, enlightenment, and challenge. *Chin. J. Catal.* **35**, 1619–1640 (2014).
4. Santos, R. Gdos & Alencar, A. C. Biomass-derived syngas production via gasification process and its catalytic conversion into fuels by Fischer Tropsch synthesis: A review. *Int. J. Hydrog. Energ.* **45**, 18114–18132 (2020).
5. Fang, Y.-H. & Liu, Z.-P. Electrochemical reactions at the electrode/solution interface: Theory and applications to water electrolysis and oxygen reduction. *Sci. China-Chem.* **53**, 543–552 (2010).
6. Deng, W. et al. Catalytic amino acid production from biomass-derived intermediates. *Proc. Natl Acad. Sci.* **115**, 5093–5098 (2018).
7. Lee, K., Jing, Y., Wang, Y. & Yan, N. A unified view on catalytic conversion of biomass and waste plastics. *Nat. Rev. Chem.* **6**, 635–652 (2022).
8. Chen, B. W. J., Xu, L. & Mavrikakis, M. Computational methods in heterogeneous catalysis. *Chem. Rev.* **121**, 1007–1048 (2021).
9. Greeley, J. Theoretical heterogeneous catalysis: scaling relationships and computational catalyst design. *Annu. Rev. Chem. Biomol.* **7**, 605–635 (2016).
10. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
11. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
12. 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).
13. Perdew, J. P., Ernzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **105**, 9982–9985 (1996).
14. Ernzerhof, M. & Scuseria, G. E. Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *J. Chem. Phys.* **110**, 5029–5036 (1999).
15. Bajaj, A. & Kulik, H. J. Eliminating delocalization error to improve heterogeneous catalysis predictions with molecular DFT + U. *J. Chem. Theory Comput.* **18**, 1142–1155 (2022).
16. Gong, X.-Q., Liu, Z.-P., Raval, R. & Hu, P. A systematic study of CO Oxidation on metals and metal oxides: density functional theory calculations. *J. Am. Chem. Soc.* **126**, 8–9 (2004).
17. Huo, C.-F., Li, Y.-W., Wang, J. & Jiao, H. Insight into CH4 formation in iron-catalyzed Fischer−Tropsch synthesis. *J. Am. Chem. Soc.* **131**, 14713–14721 (2009).
18. Cheng, J. et al. Density functional theory study of iron and Cobalt Carbides for Fischer−Tropsch synthesis. *J. Phys. Chem. C.* **114**, 1085–1093 (2010).
19. Broos, R. J. P., Zijlstra, B., Filot, I. A. W. & Hensen, E. J. M. Quantum-chemical DFT study of direct and H- and C-assisted CO dissociation on the χ-Fe5C2 Hägg Carbide. *J. Phys. Chem. C.* **122**, 9929–9938 (2018).

20. Montoya, J. H., Tsai, C., Vojvodic, A. & Nørskov, J. K. The challenge of electrochemical ammonia synthesis: a new perspective on the role of nitrogen scaling relations. *ChemSusChem* **8**, 2180–2186 (2015).
21. An, Q., Mcdonald, M., Fortunelli, A. & Goddard, W. A. I. Si-doped Fe catalyst for ammonia synthesis at dramatically decreased pressures and temperatures. *J. Am. Chem. Soc.* **142**, 8223–8232 (2020).
22. Feaster, J. T. et al. Understanding selectivity for the electrochemical reduction of carbon dioxide to formic acid and carbon monoxide on metal electrodes. *ACS Catal.* **7**, 4822–4827 (2017).
23. Martin, N. M. et al. High-coverage oxygen-induced surface structures on Ag(111). *J. Phys. Chem. C.* **118**, 15324–15331 (2014).
24. Peng, M. et al. Fully Exposed Cluster Catalyst (FECC): Toward rich surface sites and full atom utilization efficiency. *ACS Cent. Sci.* **7**, 262–273 (2021).
25. Jin, R., Li, G., Sharma, S., Li, Y. & Du, X. Toward active-site tailoring in heterogeneous catalysis by atomically precise metal nanoclusters with crystallographic structures. *Chem. Rev.* **121**, 567–648 (2021).
26. Shang, C. & Liu, Z.-P. Constrained Broyden minimization combined with the dimer method for locating transition state of complex reactions. *J. Chem. Theory Comput.* **6**, 1136–1144 (2010).
27. Peng, C., Ayala, P. Y., Schlegel, H. B. & Frisch, M. J. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.* **17**, 49–56 (1996).
28. Zimmerman, P. M. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.* **138**, 184102 (2013).
29. Henkelman, G. & Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **113**, 9978–9985 (2000).
30. 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).
31. Zhang, X.-J., Shang, C. & Liu, Z.-P. Double-ended surface walking method for pathway building and transition state location of complex reactions. *J. Chem. Theory Comput.* **9**, 5745–5753 (2013).
32. Hu, S. & Li, W.-X. Sabatier principle of metal-support interaction for design of ultrastable metal nanocatalysts. *Science* **374**, 1360–1365 (2021).
33. Kim, H. Y., Lee, H. M. & Henkelman, G. CO oxidation mechanism on CeO_2-supported Au nanoparticles. *J. Am. Chem. Soc.* **134**, 1560–1570 (2012).
34. Bernardi, R. C., Melo, M. C. R. & Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *BBA-GEN Subj.* **1850**, 872–877 (2015).
35. Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).
36. Wales, D. J. & Doye, J. P. K. Global optimization by Basin-Hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997).
37. Calvo, F., Schebarchov, D. & Wales, D. J. Grand and semigrand canonical Basin-Hopping. *J. Chem. Theory Comput.* **12**, 902–909 (2016).
38. Sierka, M. et al. Oxygen adsorption on Mo(112) surface studied by ab initio genetic algorithm and experiment. *J. Chem. Phys.* **126**, 234710 (2007).
39. Vilhelmsen, L. B. & Hammer, B. A genetic algorithm for first principles global structure optimization of supported nano structures. *J. Chem. Phys.* **141**, 044711 (2014).
40. Wang, Q., Oganov, A. R., Zhu, Q. & Zhou, X.-F. New reconstructions of the (110) surface of Rutile TiO_2 predicted by an evolutionary method. *Phys. Rev. Lett.* **113**, 266101 (2014).
41. Bunting, R. J., Cheng, X., Thompson, J. & Hu, P. Amorphous surface PdOX and its activity toward methane combustion. *ACS Catal.* **9**, 10317–10323 (2019).
42. Lu, S., Wang, Y., Liu, H., Miao, M. & Ma, Y. Self-assembled ultrathin nanotubes on diamond (100) surface. *Nat. Commun.* **5**, 3666 (2014).
43. Shang, C. & Liu, Z.-P. Stochastic surface walking method for structure prediction and pathway searching. *J. Chem. Theory Comput.* **9**, 1838–1845 (2013).
44. Zhang, X.-J., Shang, C. & Liu, Z.-P. From atoms to fullerene: stochastic surface walking solution for automated structure prediction of complex material. *J. Chem. Theory Comput.* **9**, 3252–3260 (2013).
45. Shang, C., Zhang, X.-J. & Liu, Z.-P. Stochastic surface walking method for crystal structure and phase transition pathway prediction. *Phys. Chem. Chem. Phys.* **16**, 17845–17856 (2014).
46. Liu, X., Niu, H. & Oganov, A. R. COPEX: co-evolutionary crystal structure prediction algorithm for complex systems. *npj Comput. Mater.* **7**, 199 (2021).
47. Tylianakis, E. & Froudakis, G. E. Grand canonical Monte Carlo method for gas adsorption and separation. *J. Comput. Theor. Nanosci.* **6**, 335–348 (2009).
48. Wexler, R. B., Qiu, T. & Rappe, A. M. Automatic prediction of surface phase diagrams using ab Initio grand canonical Monte Carlo. *J. Phys. Chem. C.* **123**, 2321–2328 (2019).
49. Fantauzzi, D. et al. Growth of Stable surface oxides on Pt(111) at near-ambient pressures. *Angew. Chem., Int. Ed.* **56**, 2594–2598 (2017).
50. Senftle, T. P., Meyer, R. J., Janik, M. J. & van Duin, A. C. T. Development of a ReaxFF potential for Pd/O and application to palladium oxide formation. *J. Chem. Phys.* **139**, 044109 (2013).
51. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107–115 (1935).
52. Truhlar, D. G. & Garrett, B. C. Variational transition state theory. *Annu. Rev. Phys. Chem.* **35**, 159–189 (1984).
53. Motagamwala, A. H. & Dumesic, J. A. Microkinetic modeling: a tool for rational catalyst design. *Chem. Rev.* **121**, 1049–1076 (2021).
54. Bossche, M. Vden & Grönbeck, H. Methane oxidation over PdO(101) revealed by first-principles kinetic modeling. *J. Am. Chem. Soc.* **137**, 12035–12044 (2015).
55. Bortz, A. B., Kalos, M. H. & Lebowitz, J. L. A new algorithm for Monte Carlo simulation of Ising spin systems. *J. Comput. Phys.* **17**, 10–18 (1975).
56. Stamatakis, M. & Vlachos, D. G. Unraveling the complexity of catalytic reactions via kinetic Monte Carlo simulation: current status and frontiers. *ACS Catal.* **2**, 2648–2663 (2012).
57. Kattel, S., Ramírez, P. J., Chen, J. G., Rodriguez, J. A. & Liu, P. Active sites for CO_2 hydrogenation to methanol on Cu/ZnO catalysts. *Science* **355**, 1296–1299 (2017).
58. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
59. 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).
60. Kästner, J. Umbrella sampling. *WIREs Comput. Mol. Sci.* **1**, 932–942 (2011).
61. Xu, J., Cao, X.-M. & Hu, P. Perspective on computational reaction prediction using machine learning methods in heterogeneous catalysis. *Phys. Chem. Chem. Phys.* **23**, 11155–11179 (2021).
62. Schlexer Lamoureux, P. et al. Machine learning for computational heterogeneous catalysis. *ChemCatChem* **11**, 3581–3601 (2019).
63. Kang, P.-L., Shang, C. & Liu, Z.-P. Large-scale atomic simulation via machine learning potentials constructed by global potential energy surface exploration. *Acc. Chem. Res.* **53**, 2119–2129 (2020).
64. Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem., Int. Ed.* **56**, 12828–12840 (2017).
65. Ma, S. & Liu, Z.-P. Machine learning for atomic simulation and activity prediction in heterogeneous catalysis: current status and future. *ACS Catal.* **10**, 13213–13226 (2020).
66. Ma, S. & Liu, Z.-P. Machine learning potential era of zeolite simulation. *Chem. Sci.* **13**, 5055–5068 (2022).
67. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
68. Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).
69. Zhang, L., Han, J., Wang, H., Car, R. & Weinan, E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
70. Han, J., Zhang, L., Car, R. & Weinan, E. Deep potential: a general representation of a many-body potential energy surface. *Commun. Comput. Phys.* **23**, 629–639 (2018).
71. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).
72. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
73. Huang, S.-D., Shang, C., Kang, P.-L. & Liu, Z.-P. Atomic structure of boron resolved using machine learning and global sampling. *Chem. Sci.* **9**, 8644–8655 (2018).
74. Ma, S., Shang, C., Wang, C.-M. & Liu, Z.-P. Thermodynamic rules for zeolite formation from machine learning based global optimization. *Chem. Sci.* **11**, 10113–10118 (2020).
75. Ma, S. & Liu, Z.-P. Zeolite-confined subnanometric PtSn mimicking mortise-and-tenon joinery for catalytic propane dehydrogenation. *Nat. Commun.* **13**, 2716 (2022).
76. Li, Y.-F. & Liu, Z.-P. Smallest Stable Si/SiO_2 Interface that Suppresses Quantum Tunneling from Machine-Learning-Based Global Search. *Phys. Rev. Lett.* **128**, 226102 (2022).
77. Liu, Q.-Y., Shang, C. & Liu, Z.-P. In situ active site for CO activation in Fe-catalyzed Fischer-Tropsch synthesis from machine learning. *J. Am. Chem. Soc.* **143**, 11109–11120 (2021).
78. 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).

79. Chen, D., Kang, P.-L. & Liu, Z.-P. Active site of catalytic ethene epoxidation: machine-learning global pathway sampling rules out the metal sites. *ACS Catal.* **11**, 8317–8326 (2021).
80. Rogal, J., Reuter, K. & Scheffler, M. First-principles statistical mechanics study of the stability of a subnanometer thin surface oxide in reactive environments: CO oxidation at Pd(100). *Phys. Rev. Lett.* **98**, 046101 (2007).
81. Li, W. X., Stampfl, C. & Scheffler, M. Why is a noble metal catalytically active? The role of the O-Ag interaction in the function of silver as an oxidation catalyst. *Phys. Rev. Lett.* **90**, 256102 (2003).
82. Hartke, B. Global optimization. *WIREs Comput. Mol. Sci.* **1**, 879–887 (2011).
83. Wales, D. J. & Scheraga, H. A. Global optimization of clusters, crystals, and biomolecules. *Science* **285**, 1368–1372 (1999).
84. 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).
85. Huang, S.-D., Shang, C., Kang, P.-L., Zhang, X.-J. & Liu, Z.-P. LASP: Fast global potential energy surface exploration. *WIREs Comput. Mol. Sci.* **9**, e1415 (2019).
86. Michaelides, A., Reuter, K. & Scheffler, M. When seeing is not believing: Oxygen on Ag(111), a simple adsorption system? *J. Vac. Sci. Technol. A* **23**, 1487–1497 (2005).
87. Schmid, M. et al. Structure of Ag(111)-p(4x4)-O: No silver oxide. *Phys. Rev. Lett.* **96**, 146102 (2006).
88. Schnadt, J. et al. Revisiting the structure of the p(4x4) surface oxide on Ag(111). *Phys. Rev. Lett.* **96**, 146101 (2006).
89. Jorgensen, M. S. et al. Atomistic structure learning. *J. Chem. Phys.* **151**, 054111 (2019).
90. Mortensen, H. L., Meldgaard, S. A., Bisbo, M. K., Christiansen, M.-P. & Hammer, B. Atomistic structure learning algorithm with surrogate energy model relaxation. *Phys. Rev. B* **102**, 075427 (2020).
91. Chen, D., Shang, C. & Liu, Z.-P. Automated search for optimal surface phases (ASOPs) in grand canonical ensemble powered by machine learning. *J. Chem. Phys.* **156**, 094104 (2022).
92. Rocca, M. et al. Phase transition of dissociatively adsorbed oxygen on Ag(001). *Phys. Rev. B* **61**, 213–227 (2000).
93. Costina, I. et al. Combined STM, LEED and DFT study of Ag(100) exposed to oxygen near atmospheric pressures. *Surf. Sci.* **600**, 617–624 (2006).
94. Bronsted, J. N. Acid and basic catalysis. *Chem. Rev.* **5**, 231–338 (1928).
95. Evans, M. G. & Polanyi, M. Inertia and driving force of chemical reactions. *Trans. Faraday Soc.* **34**, 11–24 (1938).
96. Michaelides, A. et al. Identification of general linear relationships between activation energies and enthalpy changes for dissociation reactions at surfaces. *J. Am. Chem. Soc.* **125**, 3704–3705 (2003).
97. Sun, G. & Sautet, P. Metastable structures in cluster catalysis from first-principles: structural ensemble in reaction conditions and metastability triggered reactivity. *J. Am. Chem. Soc.* **140**, 2812–2820 (2018).
98. Sun, G., Fuller, J. T., Alexandrova, A. N. & Sautet, P. Global activity search uncovers reaction induced concomitant catalyst restructuring for alkane dissociation on model Pt catalysts. *ACS Catal.* **11**, 1877–1885 (2021).
99. Sun, G. & Sautet, P. Active site fluxional restructuring as a new paradigm in triggering reaction activity for nanocluster catalysis. *Acc. Chem. Res.* **54**, 3841–3849 (2021).
100. Zhang, X.-J., Shang, C. & Liu, Z.-P. Stochastic surface walking reaction sampling for resolving heterogeneous catalytic reaction network: A revisit to the mechanism of water-gas shift reaction on Cu. *J. Chem. Phys.* **147**, 152706 (2017).
101. Kang, P.-L., Shang, C. & Liu, Z.-P. Glucose to 5-Hydroxymethylfurfural: Origin of site-selectivity resolved by machine learning based reaction sampling. *J. Am. Chem. Soc.* **141**, 20525–20536 (2019).
102. Kang, P.-L. & Liu, Z.-P. Reaction prediction via atomistic simulation: from quantum mechanics to machine learning. *iScience* **24**, 102013 (2021).
103. Zhu, S.-C., Xie, S.-H. & Liu, Z.-P. Nature of Rutile nuclei in Anatase-to-Rutile phase transition. *J. Am. Chem. Soc.* **137**, 11532–11539 (2015).
104. Xie, Y.-P., Zhang, X.-J. & Liu, Z.-P. Graphite to diamond: origin for kinetics selectivity. *J. Am. Chem. Soc.* **139**, 2545–2548 (2017).
105. Li, Y.-F., Zhu, S.-C. & Liu, Z.-P. Reaction Network of layer-to-tunnel transition of MnO_2. *J. Am. Chem. Soc.* **138**, 5371–5379 (2016).
106. Linic, S. & Barteau, M. A. Formation of a stable surface oxametallacycle that produces ethylene oxide. *J. Am. Chem. Soc.* **124**, 310–317 (2002).
107. Linic, S., Piao, H., Adib, K. & Barteau, M. A. Ethylene epoxidation on Ag: Identification of the crucial surface intermediate by experimental and theoretical investigation of its electronic structure. *Angew. Chem., Int. Ed.* **43**, 2918–2921 (2004).
108. Pu, T., Tian, H., Ford, M. E., Rangarajan, S. & Wachs, I. E. Overview of selective oxidation of ethylene to ethylene oxide by Ag catalysts. *ACS Catal.* **9**, 10727–10750 (2019).
109. Christopher, P. & Linic, S. Engineering selectivity in heterogeneous catalysis: Ag nanowires as selective ethylene epoxidation catalysts. *J. Am. Chem. Soc.* **130**, 11264–11265 (2008).
110. Hus, 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).
111. Kang, P.-L., Shi, Y.-F., Shang, C. & Liu, Z.-P. Artificial intelligence pathway search to resolve catalytic glycerol hydrogenolysis selectivity. *Chem. Sci.* **13**, 8148–8160 (2022).
112. Ruppert, A. M., Weinberg, K. & Palkovits, R. Hydrogenolysis goes bio: from carbohydrates and sugar alcohols to platform chemicals. *Angew. Chem., Int. Ed.* **51**, 2564–2601 (2012).
113. Corma, A., Iborra, S. & Velty, A. Chemical routes for the transformation of biomass into chemicals. *Chem. Rev.* **107**, 2411–2502 (2007).
114. Alonso, D. M., Wettstein, S. G. & Dumesic, J. A. Bimetallic catalysts for upgrading of biomass to fuels and chemicals. *Chem. Soc. Rev.* **41**, 8075–8098 (2012).
115. Zhang, X. et al. Platinum–copper single atom alloy catalysts with high performance towards glycerol hydrogenolysis. *Nat. Commun.* **10**, 5812 (2019).
116. Wang, S., Zhang, Y. & Liu, H. Selective hydrogenolysis of glycerol to propylene glycol on Cu–ZnO composite catalysts: structural requirements and reaction mechanism. *Chem. Asian J.* **5**, 1100–1111 (2010).
117. 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).

ACKNOWLEDGEMENTS

This work received financial support from the National Key Research and Development Program of China (2018YFA0208600), the National Science Foundation of China (12188101, 22033003, 91945301, 91745201, 92145302, 22122301 and 92061112), Fundamental Research Funds for the Central Universities (20720220011) and the Tencent Foundation for XPLORER PRIZE.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception, structuring, and writing of this review.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Zhi-Pan Liu.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.