RESEARCH ARTICLE

WILEY

Data-driven structural descriptor for predicting platinum-based alloys as oxygen reduction electrocatalysts

Xue Zhang[1] | Zhuo Wang[1,2] | Adam Mukhtar Lawan[1] | Jiahong Wang[1,3] | Chang-Yu Hsieh[4] | Chenru Duan[5] | Cheng Heng Pang[2] | Paul. K. Chu[6] | Xue-Feng Yu[1,3] | Haitao Zhao[1,2]

[1]Shenzhen Engineering Center for the Fabrication of Two-Dimensional Atomic Crystals, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, the People's Republic of China

[2]Department of Chemical and Environmental Engineering, The University of Nottingham Ningbo China, Ningbo, the People's Republic of China

[3]Hubei Three Gorges Laboratory, Yichang, Hubei, the People's Republic of China

[4]Innovation Institute for Artificial Intelligence in Medicine, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, the People's Republic of China

[5]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[6]Department of Physics, Department of Materials Science and Engineering, and Department of Biomedical Engineering, City University of Hong Kong, Kowloon, Hong Kong, the People's Republic of China

Correspondence
Haitao Zhao and Xue-Feng Yu, Shenzhen Engineering Center for the Fabrication of Two-Dimensional Atomic Crystals, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, People's Republic of China.
Email: ht.zhao@siat.ac.cn and xf.yu@siat.ac.cn

Abstract

Owing to increasing global demand for carbon neutral and fossil-free energy systems, extensive research is being conducted on efficient and inexpensive electrocatalysts for catalyzing the kinetically sluggish oxygen reduction reaction (ORR) at the cathode of fuel cells. Platinum (Pt)-based alloys are considered promising candidates for replacing expensive Pt catalysts. However, the current screening process of Pt-based alloys is time-consuming and labor-intensive, and the descriptor for predicting the activity of Pt-based catalysts is generally inaccurate. This study proposed a strategy by combining high-throughput first-principles calculations and machine learning to explore the descriptor used for screening Pt-based alloy catalysts with high Pt utilization and low Pt consumption. Among the 77 prescreened candidates, we identified 5 potential candidates for catalyzing ORR with low overpotential. Furthermore, during the second and third rounds of active learning, more Pt-based alloys ORR candidates are identified based on the relationship between structural features of Pt-based alloys and their activity. In addition, we highlighted the role of structural features in Pt-based alloys and found that the difference between the electronegativity of Pt and heteroatom, the valence electrons number of the heteroatom, and the ratio of heteroatoms around Pt are the main factors that affect the activity of ORR. More importantly, the combination of those structural features can be used as

Xue Zhang and Zhuo Wang contributed equally to this study.

structural descriptor for predicting the activity of Pt-based alloys. We believe the findings of this study will provide new insight for predicting ORR activity and contribute to exploring Pt-based electrocatalysts with high Pt utilization and low Pt consumption experimentally.

1 | INTRODUCTION

The oxygen reduction reaction (ORR) in the cathode of proton exchange membrane fuel cells is a challenging reaction because of the slow electrochemical kinetics of ORR reaction. To overcome this obstacle, a stringent and effective catalyst that can enhance the ORR electrochemical kinetics is required.[1] At present, Pt is considered as a promising catalyst; however, its high-cost limits its wide application. One of the major strategies employed to overcome this limitation is nanostructuring or alloying of pure platinum (Pt)-based catalysts with nonprecious metal catalysts.[2,3] Interestingly, Pt-based alloys have demonstrated remarkable stability and show good electrocatalytic performance and hence are considered as substantive alternative to pure Pt catalysts.[4–11] However, owing to the presence of different elements with varying mixing ratios, alloyed-catalyst have numerous chemical features and adsorption sites compared to their pristine form.[12,13] In addition, its ever-changing chemical space makes it difficult to screen functional alloyed-catalysts.[14]

The trial-and-error approach is generally adopted to design ORR catalysts in traditional experiments. However, it is limited by cumbersome synthetic procedures and demands in situ characterization.[15] With the development of first-principles methods and computational resources, theoretical modeling provides new opportunities for rational catalyst design. Creating large databases based on the first-principles results and identifying materials with desired properties from the databases is a simple and powerful approach for designing materials.[16,17] However, this method usually requires a reliable descriptor model that can evaluate and correlate a candidate material's intrinsic properties easily with target properties such as activity and selectivity.[18] Therefore, accurate identification of such descriptors can speed up and enhance the catalyst selection process. Extensive research has been conducted on searching and using descriptors for establishing properties correlations in materials simply; for example, the linear relationship between the reaction free energy and activation energy in heterogeneous catalysis[19] and the linear relationship between the d-band center of a clean surface and the adsorption energy of molecules on that surface.[20,21] Although these descriptors have been studied in detail and widely used because of their simplicity and well-defined physical meaning, they are imprecise; therefore, more and more researchers have focused on overcoming the limitations imposed by corresponding approximations.[22,23]

In addition, structural effects such as the structural diversity of catalysts and the local environment around the adsorption sites have become important factors that affect the performance of catalysts.[24,25] Under realistic conditions, the complex relationship between the catalyst performance and structure hinders the reliable description of the catalytic performance.[26,27] Fortunately, the potentially complex relationship between catalytic activity and structure can be revealed using machine learning (ML) techniques and high-throughput calculations, which can not only achieve accurate and efficient structure optimization,[28] but also provide insights into the catalytic properties of materials as well as predict the catalytic properties of unknown materials.[29–33] Therefore, for the accurate modeling of alloyed catalyst properties, more reliable approaches should be utilized by incorporating data-driven descriptors and chemical descriptors (adsorption energy, coordination numbers, etc.).[34–38]

This study adopts a workflow that utilizes ML and high-throughput calculation to accelerate the discovery of Pt-based alloy catalyst. By combining first-principles calculations and compressed-sensing data-analytics methodology, the Pt-based alloys for ORR applications are prescreened by identifying the descriptors based on the properties of their different compositions and structures. We predict the ratio of heteroatoms around Pt, the difference of the electronegativity between Pt and heteroatom, and the valance electrons of Pt and heteroatoms that can be used to predict the ORR activity in the alloy. The recently developed Sure Independence Screening and Sparsifying Operator (SISSO), a state-of-the-art compressed-sensing-based approach, is used to identify the key descriptive parameters.[39] We identified the ratio of the heteroatoms around Pt, the difference in electronegativity between Pt atoms and heteroatoms, and their

valences as structural descriptors that are capable of predicting the ORR activity in an alloyed catalyst. We believe that these results can provide a useful dataset for experimentalists to further scrutinize the predicted ORR activity as well as for data scientists to construct ML models for ORR performance predictions.

2 | RESULTS AND DISCUSSION

2.1 | Data collection and material prescreening

We propose a data-driven scheme to explore potential Pt-based alloys as highly efficient ORR electrocatalysts. First, we curate datasets of Pt-based alloys from the created materials database (Inorganic Crystal Structure Database [ICSD] and Material Project [MP]). Then, we use high-throughput screening techniques to screen the Pt-based binary alloys that exhibit highly efficient ORR performance. The high-throughput density functional theory (DFT) calculation is used to study the ORR reaction mechanism and to identify the rate-determining step. Based on the redox potential of the rate-determining step, ML is used to obtain descriptors that are capable of identifying and classifying highly efficient catalysts from the datasets by using SISSO, which identifies the best low-dimensional descriptors in an immensity of offered candidates. Finally, we establish the structure–activity relationship and the prediction model, which can be utilized to screen the candidate materials with suitable properties for ORR. Active learning and reverse design are also included for the second and third round screening. Scheme 1 summarizes the process as below. The overall goal is to identify Pt_nM_m alloys that are stable and poses good ORR activity.

To curate the Pt_nM_m alloys datasets, a high-throughput screening approach is adopted to screen elements from the periodic table that should be alloyed with Pt while being as stable as possible. We focus on only the materials that have been reported at the current stage from experiments and theoretical calculations. There are ~160 000 and ~140 000 materials in ICSD and MP, respectively. And ~1500 Pt-based binary alloys were found in the database. As shown in Figures 1 and S1, five criteria were applied to prescreen alloy formation by Pt and the other metal elements in the ICSD and the MP database based on the radius, orbital configuration of the transition metal atom, formation energy, crystal system, and the atom ratio between the transition metal and Pt atom of the compounds. These criteria are considered while training ML models based on the label of suitability for catalyzing ORR. The difference of atomic radius between Pt and the other elements is the first prescreening criterion, which can be used to evaluate the stability of an alloy during the ORR reaction process.[40,41] In

this step, the upper bound for the screening is set to 0.3 Å; a higher difference of atomic radius would make it difficult to keep the morphology of the alloy during the electrochemical tests. The second criterion is the atom orbital of the metal elements, where the 3d-5d metals are chosen. Furthermore, the formation energy of the alloy is the third criterion, because a stable material is needed for high-performance electrochemical catalysts. Considering the possible uncertainties/errors of the formation energy associated with DFT in the database, we slightly loosen the restriction to 50 meV as the threshold value. The fourth criterion is the crystal system, and cube crystal is chosen, which is the same as that of Pt unit cell. By restricting our calculations, the PtM, Pt_3M, and PtM_3 alloys are screened since they are widely studied and experimental evidence suggests that more excellent activity of ORR over the surfaces of these alloys was observed than that over the pure Pt surface.[42–44] It is noteworthy that the Pt alloy containing lanthanide and actinide metals is not excluded in this study.

The trained models are expected to make predictions and classify the unknown materials into two categories. The training dataset comprises 555 potential material candidates with different structural configurations, which is divided into training data and test data with the ratio of 9:1. Because of the small quantity of the training data and to avoid the distribution variation during the validation process, a 10-fold cross-validation scheme is selected to conduct hyperparameter tuning. In the cross-validation stage, the training data are divided equally into 10 groups and each of them group is taken as the validation data to validate and evaluate the model trained by the dataset comprising the remaining 9 groups of data. The performance of each validation is recorded, and the mean value of the 10 performance is considered as the final score of the trained model. We selected the decision tree (DT) and random forest classifiers to implement the classification application. Through the above prescreening criteria, we identified 77 materials as catalyst candidates toward ORR. It should be noted that in the above prescreening procedure, the reference values of the prescreening criteria are tunable parameters, which can be altered to achieve different sizes of screened datasets. Further loosening this criterion may allow us to include more materials, such as the PtBi, which has relatively low stability.[45] Considering the requirement on the formation energy, decreasing the $E_{formaiton}$ criteria threshold from 0.01 to 0.001 eV would filter out only three more 2D materials. For an alloy, the more open the surfaces, the stronger the intermediates bind, and eventually the surface get blocked. Although the utilization efficiency of step sites is much higher compared to the conventional basal plane, the basal plane sites are dominant. Therefore, another implicit assumption is that these dominate the activity of polycrystalline Pt, and the step sites and basal sites are considered. All surfaces are shown in Figure 2.
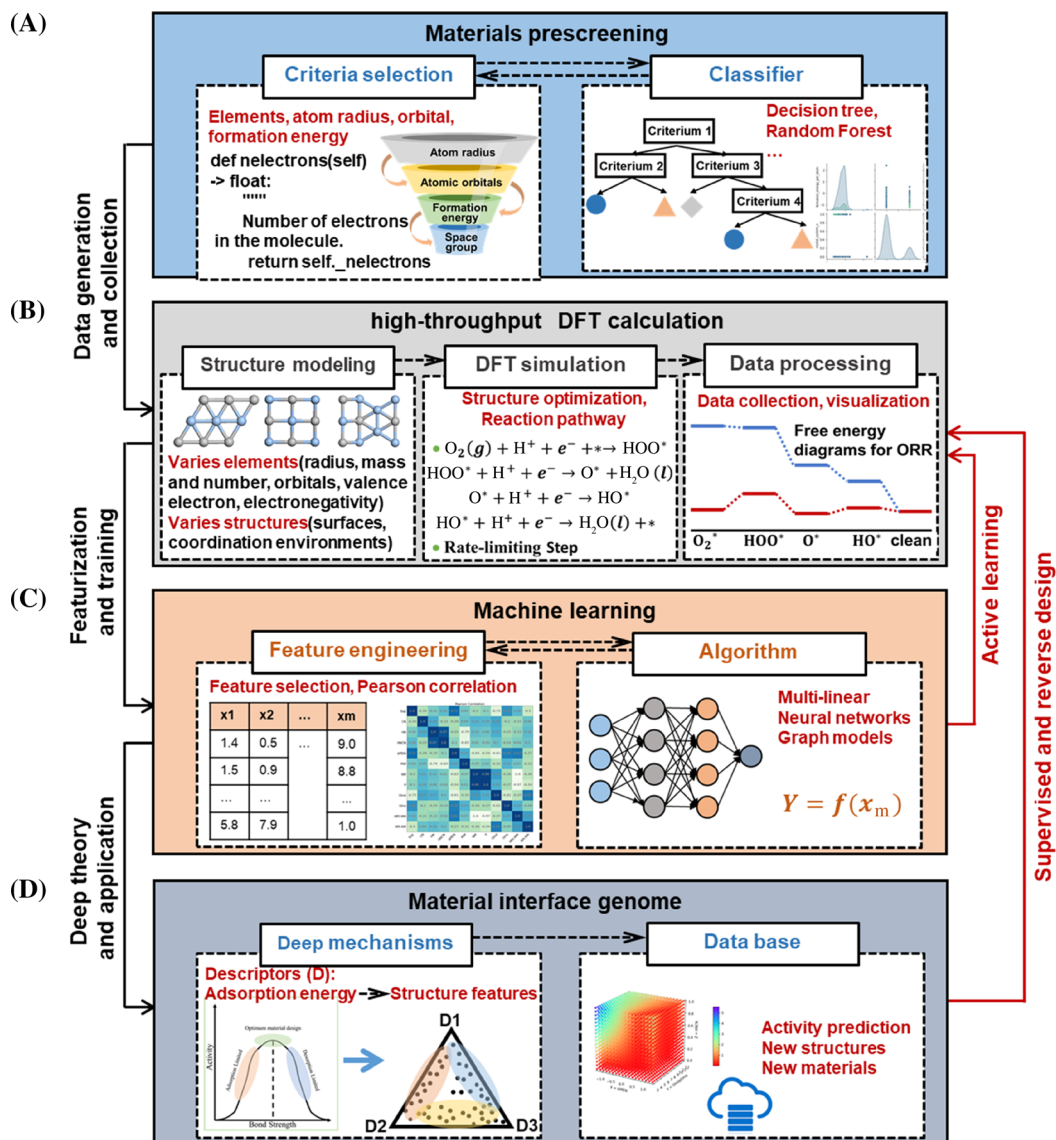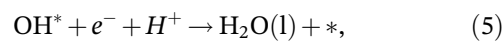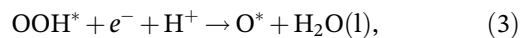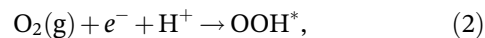
SCHEME 1 A workflow for constructing machine learning (ML) models for predicting Platinum-based alloys as oxygen reduction reaction (ORR) electrocatalysts. Four major steps are involved in this workflow: (A) material prescreening, (B) high-throughput density functional theory (DFT) calculation, (C) machine learning, and (D) material interface genome. Based on these steps, the data are generated and collected, as well as featured and trained to produce the deep theory and further application. Various databases and model packages have enabled a much easier experience of model construction.

2.2 | First-principal calculations and feature engineering

The overall ORR can be described as the following equation:

$$O_2(g) + 4e^- + 4H^+ \rightarrow 2H_2O(l). \tag{1}$$
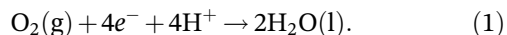
The four-electron reaction pathway (2–5) of O_2 reduction in acidic media is considered (pH = 0)[46]:

$$O_2(g) + e^- + H^+ \rightarrow OOH^*, \tag{2}$$

$$OOH^* + e^- + H^+ \rightarrow O^* + H_2O(l), \tag{3}$$

$$O^* + e^- + H^+ \rightarrow OH^*, \tag{4}$$

$$OH^* + e^- + H^+ \rightarrow H_2O(l) + *, \tag{5}$$

where the asterisk denotes an adsorbed site on the surface, (g) and (l) refer to the gas and liquid phases, respectively.
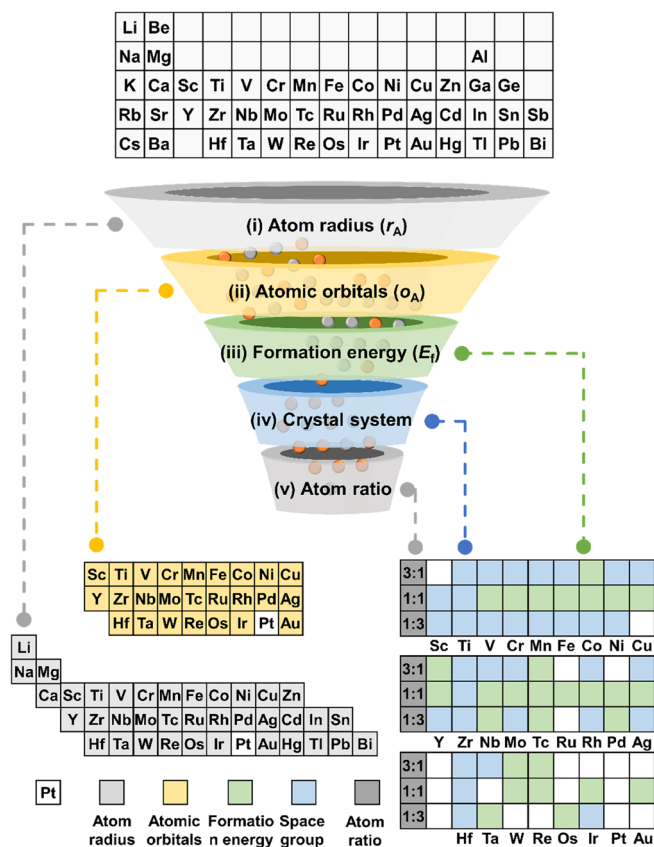
FIGURE 1 The approach for materials prescreening. From a large number of possible compounds with the criteria of (i) the radius and (ii) the orbital configuration of the transition metal atom, (iii) formation energy, (iv) crystal system, and (v) the atom ratio between the transition metal and the Pt atom, the materials considered to be calculated by density functional theory (DFT) are generated. Using machine learning techniques, we classify materials based on the decision tree.

The ORR mechanism on more than 62 Pt-based alloys is calculated; this large dataset comprised the adsorption energy of the intermediates and the free energy of the element steps at the surfaces of the Pt alloyed with 62 transition metal. The intermediates are placed at different sites and the energy for the most favorable site is included in the dataset. Complete information on adsorption sites and the corresponding reaction free energy is provided in Table S1. The changes of the free energy calculated by DFT during the ORR show that two reaction steps are sluggish that involve a positive change in free energy: the third electron and proton transfer for forming the adsorbed OH (ΔG_3) and the last transfer for removing OH from the surface to form water (ΔG_4). To screen the Pt-based alloy more easily, we use the values of overpotential (η) transferred from the ΔG for these steps and equilibrium potential as a measure of the activity. The overpotential for ORR is calculated by the equation:

$$\eta = \frac{\max\{\Delta G_1, \Delta G_2, \Delta G_3, \Delta G_4\}}{e} + 1.23, \qquad (6)$$

where ΔG_1, ΔG_2, ΔG_3, and ΔG_4 denote the reaction free energies in Equations (2) to (5), respectively. According to thermodynamics, the smaller the overpotential, the higher the corresponding activity of ORR is; therefore, the performance of candidates is better for catalyzing ORR. The implicit assumption in this analysis is that the kinetic relationship is closely related to thermodynamics and can be simplified by thermodynamics. Because there will be an activation free energy in four elemental steps at the equilibrium potential of 1.23 eV, which is at least equal to the largest of the reaction free energies and the corresponding step is therefore likely the rate-limiting step.[16,46–48] Forming a Pt alloy is one way to modify the electronic structure of the Pt surface to tune the stability of these critical intermediates. The stabilities of OH intermediates ($E_{ads}(OH)$), in turn, scale roughly with the stability of adsorbed O ($E_{ads}(O)$), as shown in Figure S2. Therefore, this parameter is particularly useful for characterizing both ΔG_1 and ΔG_2. The ORR activity trends on different metal surfaces are summarized in Figure 3A. Plotting measured activities (overpotential of the rate-determining step) for a series of different catalysts as a function of the calculated OH adsorption energy results in a simple "volcano" relationship (Figure 3A). If ΔE_{OH} becomes increasingly positive, adsorbed H_2O is destabilized and can desorb from the surface more easily. However, if ΔE_{OH} keeps increases in the positive range, it becomes easier to break the Pt—OH bonds, which makes the OH* formation difficult. This appears to be a reasonable conjecture, given that more open surfaces tend to bind intermediates considerably stronger and become blocked. As shown in Figure 3A, a surface that binds OH with the adsorption energy of 1.1 eV exhibits an optimal ORR activity. The decreasing of overpotential indicates the increasing ORR activity and the activity is closely related to the behavior of OH adsorption. The weaker OH adsorption on the surface results in the lower ORR activity. The rate-determining step of the left half branch is H_2O formation. In this branch, the ORR activity becomes better as the adsorption strength of OH decreases. When the OH adsorption is too strong, the H_2O formed after OH hydrogenation is difficult to desorb from the catalyst surface; therefore, weaker OH adsorption strength is beneficial for ORR. The rate-determining step of the right half branch is OH formation. In this branch, the ORR activity becomes worse as the OH adsorption strength decreases. When the OH adsorption is too weak, it is difficult to form OH from O hydrogenation. With the calculated overpotential for the catalyst candidates, it is desirable to leverage such data to examine whether the OH adsorption is simply correlated
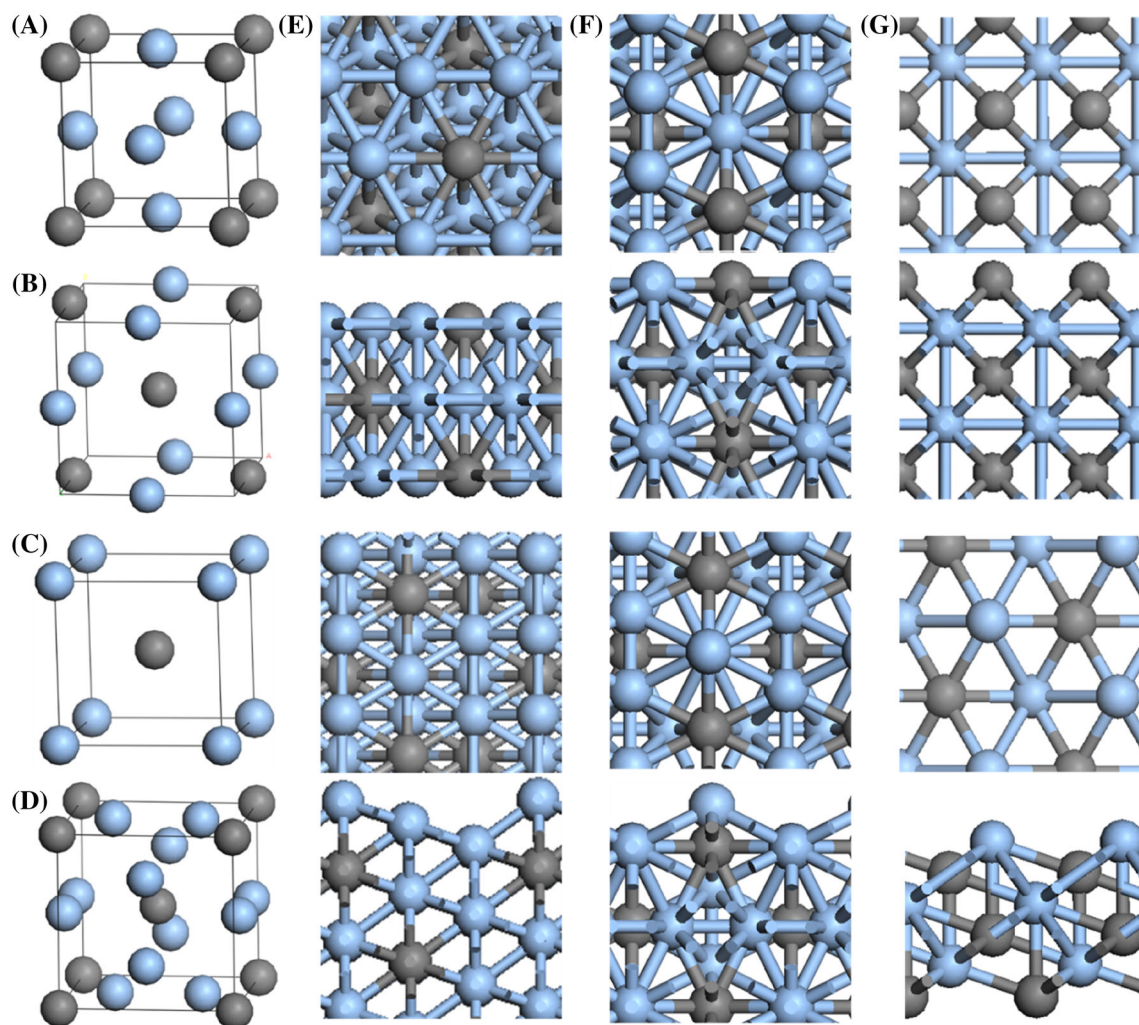
FIGURE 2 The unite cell of the platinum (Pt)-based alloy and surface structures used in this study. (A–D) The crystal structures of the Pt-based alloy prescreened from the database. (E) The (111), (211), (100) surfaces of crystal structure (A). (F) The (111), (211), (100) surfaces of crystal structure (B). (G) The (111), (211), (100) surfaces of crystal structure (C).

with a certain intrinsic property of a given material. Such simple correlations are usually established on a series of adsorption systems bearing similar atomic structures, leading to the predominance of the electronic effect. Although all computational results are perfect on the monocrystalline surfaces in experiments, a mixture of single crystal, vacuum-annealed polycrystalline and Ar-sputtered polycrystalline surfaces is always used. The resulting structural differences introduce deviations from our single-crystal models, thereby indicating modest changes in the ORR activity. However, despite on the polycrystalline of the Pt-based alloy, the site corresponding to the most stable configuration can always be model on the single crystal surface and the differences do not substantially alter the trends described above.[49,50] Although computationally based electrocatalyst discovery is the principal aim of this approach, more generally and, perhaps, more importantly, we probe our present understanding of the ORR. In the

field of catalysis, there is no stronger evidence for accuracy of a theoretical framework than the ability to use that framework to identify new active materials.

Furthermore, we carried out a correlation study on all the factors that affected the O_2 reduction activity on Pt-based alloys, including the coordination number of Pt, the number of heteroatoms around Pt, the electronegativity difference between Pt and heteroatoms, the difference between Pt and heteroatoms, the ratio of Pt and heteroatoms, the atomic number and period of the heteroatoms, the number of valence electrons (d and s electrons) of the heteroatoms, the difference in the relative mass and the difference of the atomic radius between Pt and the heteroatoms. All the abbreviations of the features are summarized in Table 1, which indicate various important or typical factors of a catalyst to influence the activity of ORR. It should be noted that the structure features are based on the final optimized structures by DFT
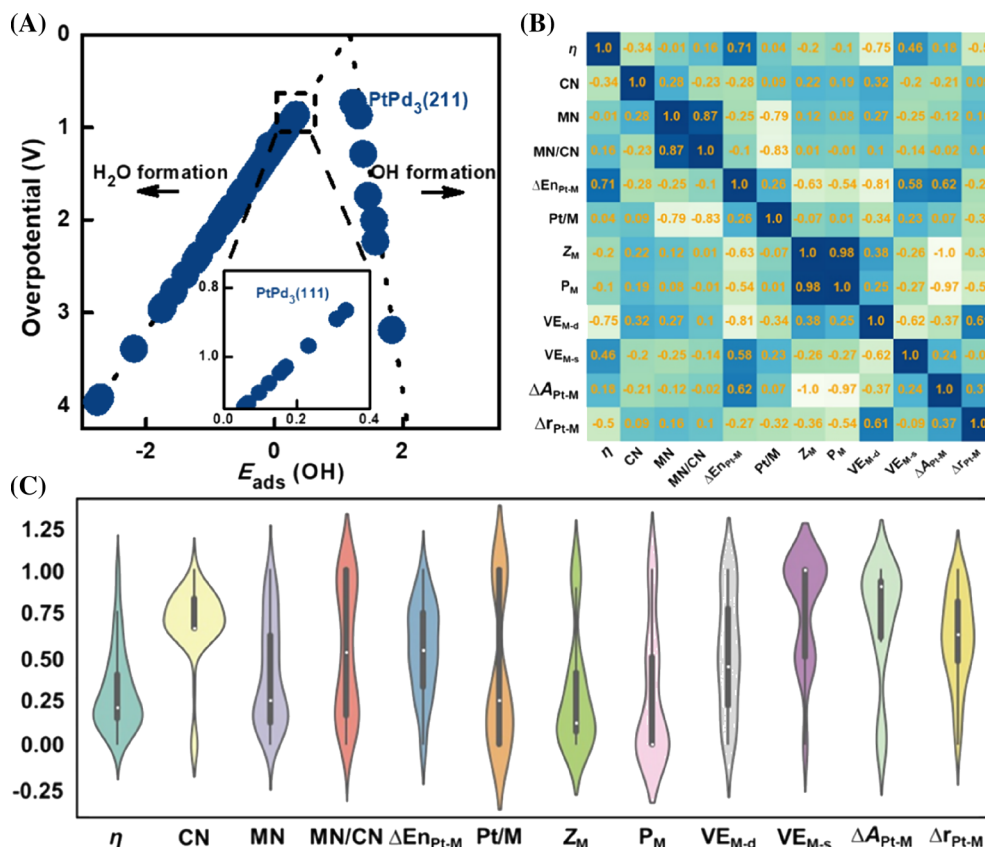
FIGURE 3 The activity of oxygen reduction reaction (ORR) and potential features. (A) The ORR activity trends on different metal surfaces, (B) the in-pair Pearson correlation coefficients of the selected potential features and the overpotential, and (C) a violin plot of the distribution of the variables after min–max normalization. Every colorful area is the density plot of corresponding variables. The black box in each density plot represents the range from 25% to 75% percentiles where the white point marks the mean value, and the whiskers denote 95% and 5%.

calculation. The electronegativity difference and valence electron number showed the greatest correlation with the ORR activity (Figure 3B). For the 12 features, we showed a violin plot of the distribution of the variables after min–max normalization in Figure 3C and the in-detail calculation is provided in the Supporting Information. The violin plot synergistically combines the box plot and the density trace (or smoothed histogram) into a single display that reveals structure found within the data, which can provide us a quick overview of the combination of the box plot and density trace. The corresponding box plot is shown in Figure S3. According to the density plot, we observed that the material with higher CN (CN = 8 and 9) and lower MN (MN = 3 and 4) tend to be more promising as candidates. Furthermore, materials with 1:3 and 3:1 ratios of Pt atom over transition metal atom are dominant in the screening stage, which is in agreement with the distribution of the MN/CN value. Variables such as ΔEn_{Pt-M}, VE_{M-d}, and Δr_{Pt-M} show a relatively even distribution. The VE_{M-s} and relative atomic mass of most materials are 2 and approximately 140, respectively.

2.3 | ML discovery of descriptors and establishment of structure–activity relationship

To obtain the best low-dimensional descriptors using SISSO, the descriptors with lower Root Mean Square Error (RMSE) are chosen. Here, the descriptor can be one feature or the combination of various features, which is used in a model of the relationship between the activity and the feature(s) of a catalyst. As shown in Figure 4A, the three-dimensional (3D) models were found to be more accurate. Therefore, we chose 3D descriptors. From the 3D model's descriptors, we obtained the following relationships as the formula:

For the rate-determining step of H_2O formation,

$$0.237 * \frac{\Delta EN_{Pt-M}^2 * (VE_{M-d} + VE_{M-s})}{MN/CN} + 0.645, \quad (7)$$

For the rate-determining step of OH formation,

$$8.995 * \frac{(MN/CN) * \Delta EN_{Pt-M}}{VE_{M-d} + VE_{M-s}} - 0.0003724 * \exp(P_M) * \frac{MN}{CN}$$
$$* (\Delta r_{Pt-M}) + 1.1765.$$

$$(8)$$

In this work, we used the hold-out validation method to verify the model because the training dataset for the SISSO model is relatively small and imbalanced, and the

TABLE 1 A set of the 11 least-correlated primary features used for the descriptor construction

Features	Description
CN	Coordination number of Pt
MN	Number of heteroatom around Pt
MN/CN	Ratio of heteroatom around Pt
ΔEN_{Pt-M}	Difference of the electronegativity between Pt and heteroatom
Pt/M	Atomic ratio of Pt and heteroatom in the alloy
Z_M	Atomic number of the heteroatom
P_M	Period of a heteroatom in the periodic table
VE_{M-d}	Number of valence electrons in the d orbital
VE_{M-s}	Number of valence electrons in the s orbital
ΔA_{Pt-M}	Difference of relative atomic mass between Pt and heteroatom
Δr_{Pt-M}	Difference of atomic radius between Pt and heteroatom

use of cross-validation might introduce extra bias if the random 10-fold splitting is uneven, although we try to ensure the data of every type of material is included in the train set, where the trained SISSO model has an RMSE on the validation set far below 0.001 eV.[51] The activity predicted by this model was found to be consistent with the model calculated by DFT, irrespective of whether for H_2O formation or OH formation was the rate-determining step (Figure 4B,C).

Based on the above model, we added 11 structures as the second active learning. The overpotential and OH adsorption energy are presented in Table S2. The new data followed the general trends well (Figure 5A), which indicated that our model exhibited true predictive ability on describing the trends of ORR activity on Pt-based alloys. Furthermore, Pt_3Co (211) showed more optimized performance with the OH formation as the rate-determining step. Then, the third active learning was carried out, and the four structures were calculated. The overpotential and OH adsorption energies are shown in Table S3. It was found that $Pt_3Ni(111)$ was upshifted to the top along the left branch with H_2O formation as the rate-determining step. As the volcano plot shows, the $Pt_3Co(211)$ showed the most optimized performance among the 77 structures. In Figure 5B, we included more detailed calculations on the $Pt_3Co(211)$ surface. The free energy changes of the elemental steps were found to be negative at 0 V versus Standard Hydrogen Electrode (SHE). At the equilibrium potential of 1.23 V, the OOH formation, OH formation and H_2O formation were endothermic and the OH formation exhibited the highest free energy change. Therefore, the energy level diagram of ORR on the $Pt_3Co(211)$ surface is the rate-determining
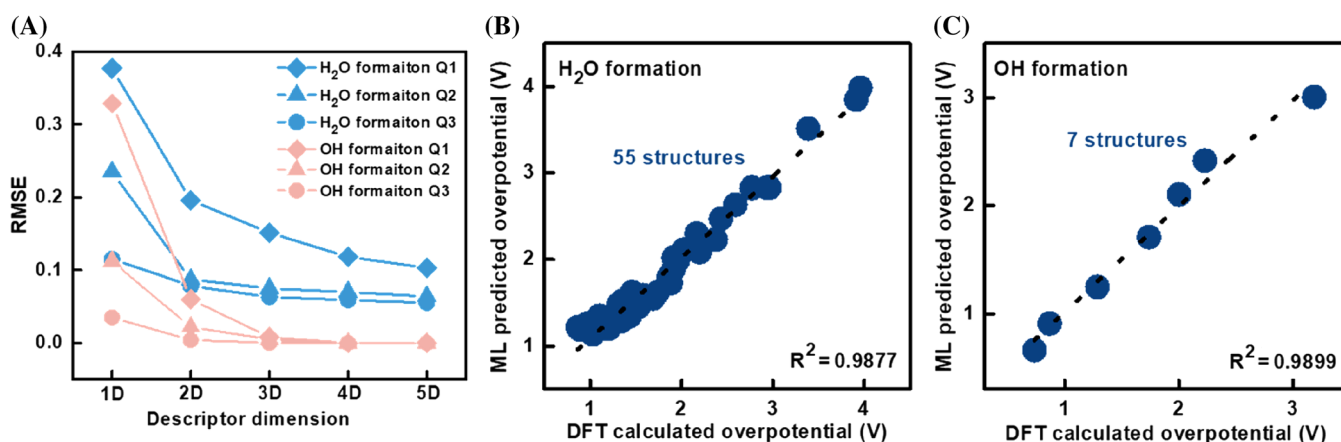

FIGURE 4 Validation of the model by machine learning. (A) Training and validation RMSE from descriptor dimension for all tested Sure Independence Screening and Sparsifying Operator (SISSO) parameters, (B) the calculated activity based on our density functional theory (DFT) model as well as a dashed line indicating predicted activity for the left branch with the rate-determining step of H_2O formation, and (C) the calculated activity based on our DFT model as well as a dashed line indicating predicted activity for the left branch with the rate-determining step of OH formation.
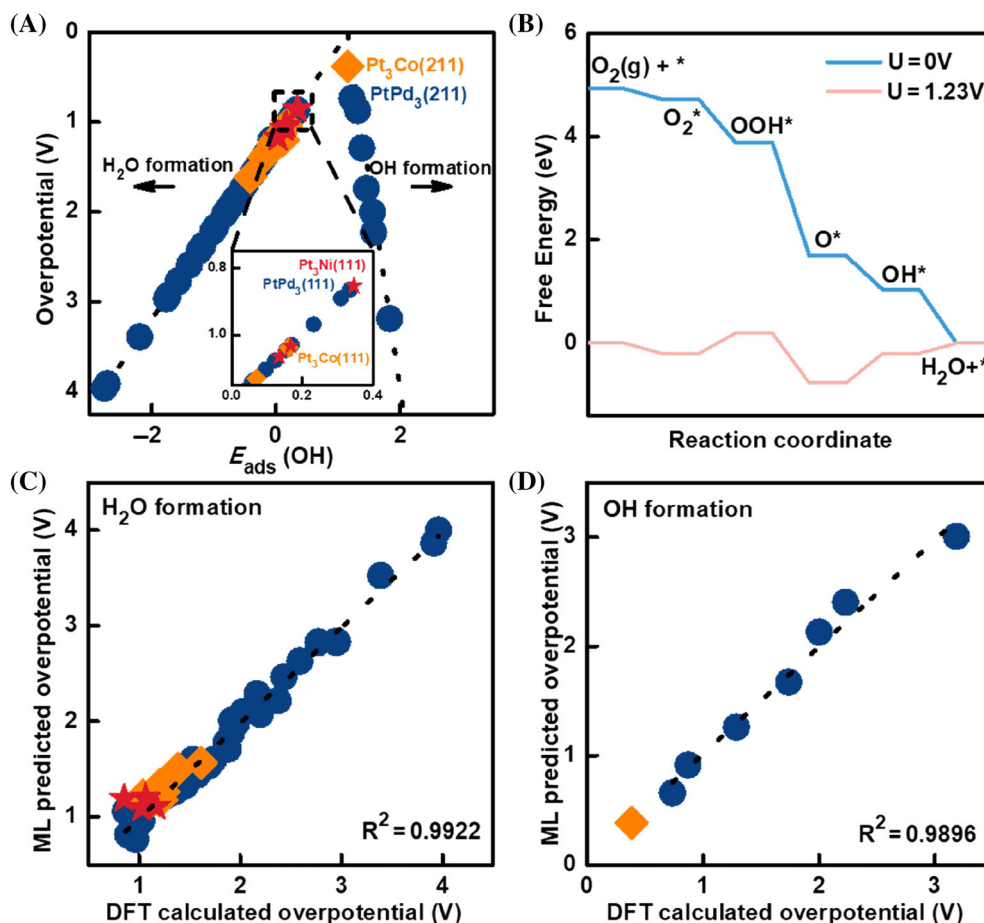
FIGURE 5 The active learning for platinum (Pt)-based alloy. (A) Active learning for another 11 (second round, orange squares) and 4 structures (third round, red stars); (B) the energy diagram of oxygen reduction reaction (ORR) on Pt$_3$Co (211) at 0 and 1.23 V versus SHE; (C) the calculated activity based on our DFT model as well as a dashed line indicating predicted activity for the left branch with the rate-determining step of H$_2$O formation; and (D) The calculated activity based on our DFT model as well as a dashed line indicating the predicted activity for the left branch with the rate-determining step of OH formation.

step. The activity predicted by this model is still consistent with the model calculated by DFT, irrespective of whether H$_2$O formation or OH formation was the rate-determining step (Figures 5C,D). Here, we used $|\eta| \leq 1$ eV as the final screening criterion and identified five potential catalyst candidates: Pt$_3$Co(211) (0.38 V), PtPd$_3$(211) (0.74 V), Pt$_3$Ni (111) (0.85 V), PtPd$_3$(111) (0.86 eV), and PtAu(111) (0.870 eV), on which OH adsorption is nearly thermoneutral for ORR at low overpotential. It is worth noting that Pt$_3$Co has been experimentally proven with an efficient ORR activity.

It was observed that based on the model, the difference between the electronegativity of Pt and heteroatom, the valence electrons number of the heteroatom, and the ratio of heteroatoms around Pt have the most obvious effects on the ORR performance. This is because the difference in the electronegativity between Pt atoms and heteroatoms and the number of valence electrons of heteroatoms can distinguish the types of alloys, the

coordination number of Pt atoms can reflect the different surfaces, and the ratio of heteroatoms around Pt atoms can reflect the doping ratio of the heteroatoms. The three factors in the model are scaled as x, y, and z, respectively. The 3D plot and cross-sectional view in the middle of each axis are shown in Figure 6, where the blank part is ascribed to the negative overpotential. As for the strong OH adsorption, when the electronegativity of Pt is lower than that of the heteroatom, the more the number of valence electrons of heteroatom is, and the lower the ratio of the number of heteroatoms around the Pt is, the higher the activity is. As for the weak OH adsorption, its activity is mainly determined by the electronegativity difference between Pt and heteroatoms, which is within 0.5; therefore, the activity is higher. It is worth noting that the activity can be predicted from the structural information for a given structure, which is very convenient and direct for new materials prediction without the need for tedious electronic structure calculations.
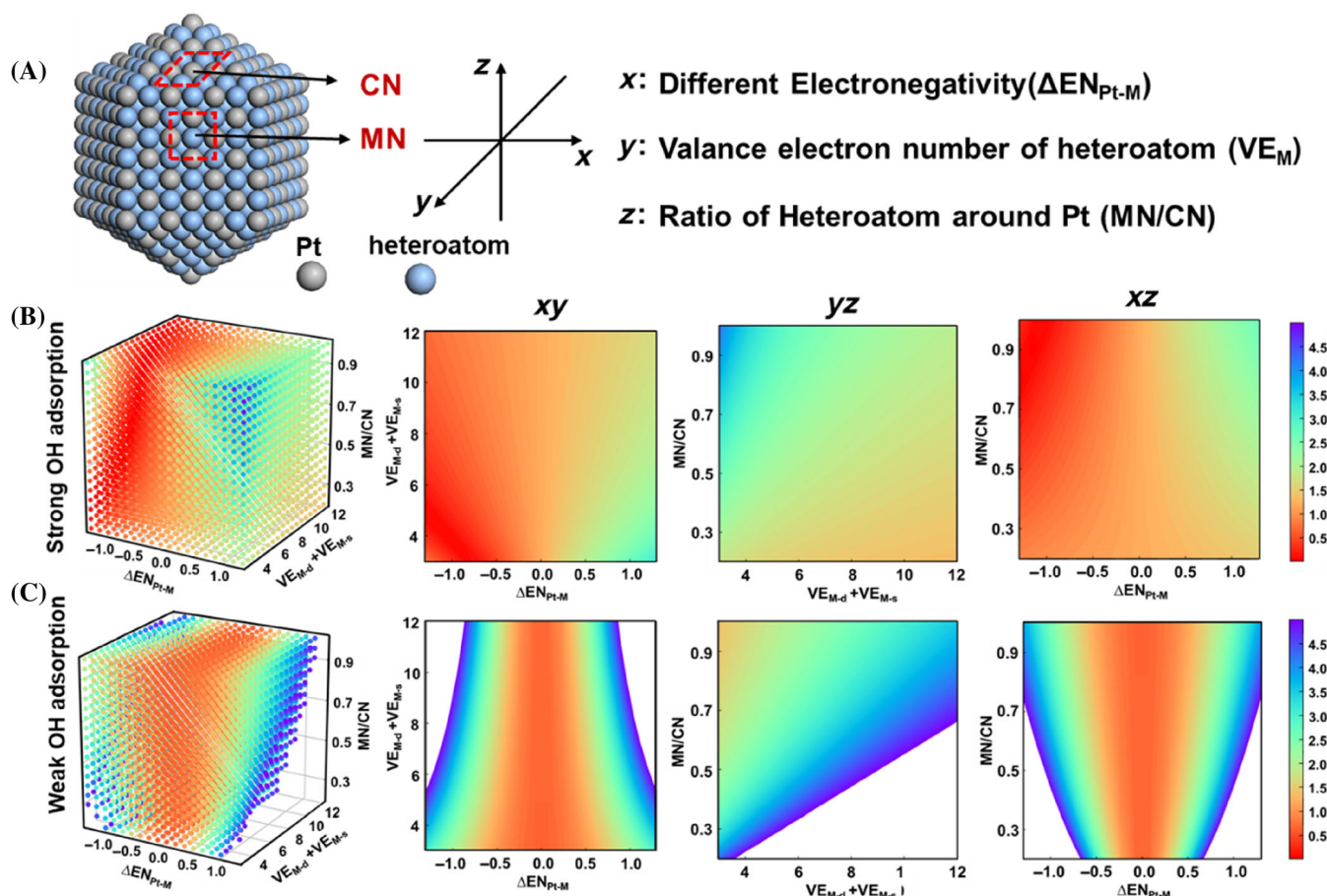
FIGURE 6 The representation of descriptors and the relationship of the structure–activity. (A) Schematic of descriptors on Pt-based alloys. The three-dimensional plot of oxygen reduction reaction (ORR) activity with the three descriptors scaled in x, y, z and the cross-sectional view at the middle of each axis, where (B) the rate-determining step is H_2O formation and (C) the rate-determining step is OH formation.

3 | CONCLUSIONS

In summary, high-throughput first-principles calculations is performed to screen high-performance catalysts toward ORR and the structure–activity is obtained. Based on the relationship, we identified more excellent Pt-based alloys via the second and third rounds of active learning. Among the 77 prescreened candidates, five candidates exhibited the thermodynamic capability of ORR with the lowest overpotential, which indicated their potential applications for catalyzing ORR. Furthermore, the difference between the electronegativity of Pt and heteroatom, the valence electrons number of the heteroatom, and the ratio of heteroatoms around Pt were found to have the most obvious effect on the ORR performance based on the model. We believe the results of this study will provide a useful dataset for experimentalists to further scrutinize the predicted ORR activity or for data scientists to construct ML models for ORR performance predictions. In addition, this study may contribute to the exploration of catalysts for other electrocatalytic processes, such as water electrolysis.

4 | DFT CALCULATION

All DFT calculations were performed using the Vienna ab initio simulation program, along with projected augmented wave[52] pseudopotentials and the Perdew–Burke–Ernzerhof functional.[53] According to the different crystal systems screened by the critera, the (111), (100), and (211) surfaces were cleavaged to simulate the planar and step sites. Neighboring slabs were separated by a vacuum of 15 Å to avoid spurious self-interactions. The surface irreducible Brillouin zone was sampled on the k-point mesh generated by the Monkhorst−Pack scheme. An energy cutoff of 400 eV was employed for the plane-wave basis set. The convergence threshold for electronic steps in geometry optimization was 1×10^{-5} eV. Geometries were deemed converged when the forces on each atom

were below 0.02 eV Å^{-1}. A frequency analysis was carried out on the stable states to confirm that these represent genuine minima. All the electronic energies were corrected for zero-point energy contributions.

AUTHOR CONTRIBUTIONS

Xue Zhang conceived this project and determined the framework of the article. Xue Zhang and Zhuo Wang carried out the DFT calculation, data collection and analysis, and contributed toward writing this manuscript. Adam Mukhtar Lawan, Jiahong Wang, Chang-Yu Hsieh, Chenru Duan, Cheng Heng Pang, and P. K. Chu revised the manuscript. Haitao Zhao and Xue-Feng Yu supervised this project. All authors discussed and commented on the manuscript. Authors Xue Zhang and Zhuo Wang contributed equally to this work.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Xue Zhang 🄯 https://orcid.org/0000-0003-1913-719X
Haitao Zhao 🄯 https://orcid.org/0000-0002-2448-8448

REFERENCES

1. Wang X, Li Z, Qu Y, et al. Review of metal catalysts for oxygen reduction reaction: from nanoscale engineering to atomic design. *Chem*. 2019;5(6):1486-1511.
2. Jaouen F, Proietti E, Lefèvre M, et al. Recent advances in non-precious metal catalysis for oxygen-reduction reaction in polymer electrolyte fuel cells. *Energ Environ Sci*. 2011;4(1):114-130.
3. Beermann V, Gocyla M, Kühl S, et al. Tuning the electrocatalytic oxygen reduction reaction activity and stability of shape-controlled Pt-Ni nanoparticles by thermal annealing—elucidating the surface atomic structural and compositional changes. *J Am Chem Soc*. 2017;139(46):16536-16547.
4. Chung DY, Jun SW, Yoon G, et al. Highly durable and active PtFe nanocatalyst for electrochemical oxygen reduction reaction. *J Am Chem Soc*. 2015;137(49):15478-15485.
5. Fan C, Wen P, Li G, et al. Facile synthesis of Pt_5La nanoalloys as the enhanced electrocatalysts for oxygen reduction reaction and methanol oxidation reaction. *J Alloys Compd*. 2022;894: 161892.
6. Gong L, Liu J, Li Y, et al. An ultralow-loading platinum alloy efficient ORR electrocatalyst based on the surface-contracted hollow structure. *Chem Eng J*. 2022;428:131569.
7. Hwang SJ, Yoo SJ, Jang S, Lim T-H, Hong SA, Kim S-K. Ternary Pt-Fe-Co alloy electrocatalysts prepared by electrodeposition: elucidating the roles of Fe and Co in the oxygen reduction reaction. *J Phys Chem C*. 2011;115(5):2483-2488.
8. Liu X, Hao S, Zheng G, et al. Ultrasmall Pt_2Sr alloy nanoparticles as efficient bifunctional electrocatalysts for oxygen reduction and hydrogen evolution in acidic media. *J Energy Chem*. 2022;64:315-322.
9. Stamenkovic VR, Fowler B, Mun BS, et al. Improved oxygen reduction activity on $Pt_3Ni(111)$ via increased surface site availability. *Science*. 2007;315(5811):493-497.
10. Vej-Hansen UG, Escudero-Escribano M, Velázquez-Palenzuela A, et al. New platinum alloy catalysts for oxygen electroreduction based on alkaline earth metals. *Electrocatalysis*. 2017;8(6):594-604.
11. Wu D, Shen X, Pan Y, Yao L, Peng Z. Platinum alloy catalysts for oxygen reduction reaction: advances, challenges and perspectives. *ChemNanoMat*. 2020;6(1):32-41.
12. Li X, Li B, Yang Z, Chen Z, Gao W, Jiang Q. A transferable machine-learning scheme from pure metals to alloys for predicting adsorption energies. *J Mater Chem A*. 2022;10(2):872-880.
13. Peng J, Tao P, Song C, Shang W, Deng T, Wu J. Structural evolution of Pt-based oxygen reduction reaction electrocatalysts. *Chin J Catal*. 2022;43(1):47-58.
14. Goldsmith BR, Esterhuizen J, Liu J-X, Bartel CJ, Sutton C. Machine learning for heterogeneous catalyst design and discovery. *AIChE J*. 2018;64(7):2311-2323.
15. Wang Y, Wang C, Li M, Yu Y, Zhang B. Nitrate electroreduction: mechanism insight, in situ characterization, performance evaluation, and challenges. *Chem Soc Rev*. 2021; 50(12):6720-6733.
16. Chen A, Zhang X, Zhou Z. Machine learning: accelerating materials development for energy storage and conversion. *InfoMat*. 2020;2(3):553-576.
17. Wei J, Chu X, Sun X-Y, et al. Machine learning in materials science. *InfoMat*. 2019;1(3):338-358.
18. Wu P, He T, Zhu H, et al. Next-generation machine vision systems incorporating two-dimensional materials: progress and perspectives. *InfoMat*. 2022;4(1):e12275.
19. van Santen RA, Neurock M, Shetty SG. Reactivity theory of transition-metal surfaces: a Brønsted–Evans–Polanyi linear activation energy-free-energy analysis. *Chem Rev*. 2010;110(4): 2005-2048.
20. Nørskov JK, Bligaard T, Rossmeisl J, Christensen CH. Towards the computational design of solid catalysts. *Nat Chem*. 2009; 1(1):37-46.
21. Hammer B, Norskov JK. Why gold is the noblest of all the metals. *Nature*. 1995;376(6537):238-240.
22. Zhang X, Li Y, Guo P, et al. Theory on optimizing the activity of electrocatalytic proton coupled electron transfer reactions. *J Catal*. 2019;376:17-24.
23. Ding R, Chen Y, Chen P, et al. Machine learning-guided discovery of underlying decisive factors and new mechanisms for

the design of nonprecious metal electrocatalysts. *ACS Catal.* 2021;11(15):9798-9808.

24. Yang F, Gao L, Lai W, Huang H. Recent advance on structural design of high-performance Pt-based nanocatalysts for oxygen reduction reaction. *Adv Sensor Energy Mater.* 2022;2(1):100022.

25. Kim M, Firestein KL, Fernando JFS, et al. Strategic design of Fe and N co-doped hierarchically porous carbon as superior ORR catalyst: from the perspective of nanoarchitectonics. *Chem Sci.* 2022;13(36):10836-10845.

26. Xu M, Xu M, Miao X. Deep machine learning unravels the structural origin of mid-gap states in chalcogenide glass for high-density memory integration. *InfoMat.* 2022;4(6):e12315.

27. Wu N, Wan S, Su S, Huang H, Dou G, Sun L. Electrode materials for brain-machine interface: a review. *InfoMat.* 2021;3(11):1174-1194.

28. Chen L, Tian Y, Hu X, et al. A universal machine learning framework for electrocatalyst innovation: a case study of discovering alloys for hydrogen evolution reaction. *Adv Funct Mater.* 2022;32(47):2208418.

29. Li Z, Wang S, Chin WS, Achenie LE, Xin H. High-throughput screening of bimetallic catalysts enabled by machine learning. *J Mater Chem A.* 2017;5(46):24131-24138.

30. Noh J, Back S, Kim J, Jung Y. Active learning with non-ab initio input features toward efficient CO_2 reduction catalysts. *Chem Sci.* 2018;9(23):5152-5159.

31. Andersen M, Medford AJ, Nørskov JK, Reuter K. Scaling-relation-based analysis of bifunctional catalysis: the case for homogeneous bimetallic alloys. *ACS Catal.* 2017;7(6):3960-3967.

32. Chen BWJ, Xu L, Mavrikakis M. Computational methods in heterogeneous catalysis. *Chem Rev.* 2021;121(2):1007-1048.

33. Zhong M, Tran K, Min Y, et al. Accelerated discovery of CO_2 electrocatalysts using active machine learning. *Nature.* 2020;581(7807):178-183.

34. Andersen M, Reuter K. Adsorption enthalpies for catalysis modeling through machine-learned descriptors. *Acc Chem Res.* 2021;54(12):2741-2749.

35. Jinnouchi R, Asahi R. Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning algorithm. *J Phys Chem Lett.* 2017;8(17):4279-4283.

36. Zhou J, Shi Q, Ullah S, et al. Phosphorus-based composites as anode materials for advanced alkali metal ion batteries. *Adv Funct Mater.* 2020;30(49):2004648.

37. Weng B, Song Z, Zhu R, et al. Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts. *Nat Commun.* 2020;11(1):3513.

38. Andersen M, Levchenko SV, Scheffler M, Reuter K. Beyond scaling relations for the description of catalytic materials. *ACS Catal.* 2019;9(4):2752-2759.

39. Ouyang R, Ahmetcik E, Carbogno C, Scheffler M, Ghiringhelli LM. Simultaneous learning of several materials properties from incomplete databases with multi-task sisso. *J Phys Mater.* 2019;2(2):024002.

40. Deshmukh AA, Kuthe SA, Palikundwar UA. Understanding the effect of compositions on electronegativity, atomic radius

and thermal stability of Mg-Ni-Y amorphous alloy. *AIP Conf Proc.* 2018;1953:090016.

41. Guo X, Mao D, Lu G, Wang S, Wu G. The influence of La doping on the catalytic behavior of Cu/ZrO_2 for methanol synthesis from CO_2 hydrogenation. *J Mol Catal A Chem.* 2011;345(1-2):60-68.

42. Li J, Alsudairi A, Ma ZF, Mukerjee S, Jia Q. Asymmetric volcano trend in oxygen reduction activity of Pt and non-Pt catalysts: In situ identification of the site-blocking effect. *J Am Chem Soc.* 2017;139(4):1384-1387.

43. Bing Y, Liu H, Zhang L, Ghosh D, Zhang J. Nanostructured Pt-alloy electrocatalysts for pem fuel cell oxygen reduction reaction. *Chem Soc Rev.* 2010;39(6):2184-2202.

44. Greeley J, Stephens IEL, Bondarenko AS, et al. Alloys of platinum and early transition metals as oxygen reduction electrocatalysts. *Nat Chem.* 2009;1(7):552-556.

45. Zhang Y, Zhou YJ, Lin JP, Chen GL, Liaw PK. Solid-solution phase formation rules for multi-component alloys. *Adv Eng Mater.* 2008;10(6):534-538.

46. Nørskov JK, Rossmeisl J, Logadottir A, et al. Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J Phys Chem B.* 2004;108(46):17886-17892.

47. Parada GA, Goldsmith ZK, Kolmar S, et al. Concerted proton-electron transfer reactions in the Marcus inverted region. *Science.* 2019;364(6439):471-475.

48. Mayer JM. Simple Marcus-theory-type model for hydrogen-atom transfer/proton-coupled electron transfer. *J Phys Chem Lett.* 2011;2(12):1481-1489.

49. Stamenkovic VR, Mun BS, Arenz M, et al. Trends in electrocatalysis on extended and nanoscale Pt-bimetallic alloy surfaces. *Nat Mater.* 2007;6(3):241-247.

50. Tian N, Zhou Z-Y, Sun S-G, Ding Y, Wang ZL. Synthesis of tetrahexahedral platinum nanocrystals with high-index facets and high electro-oxidation activity. *Science.* 2007;316(5825):732-735.

51. Sun X, MacManus-Driscoll JL, Wang H. Spontaneous ordering of oxide-oxide epitaxial vertically aligned nanocomposite thin films. *Annu Rev Mater Res.* 2020;50(1):229-253.

52. Blöchl PE. Projector augmented-wave method. *Phys Rev B.* 1994;50(24):17953-17979.

53. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. *Phys Rev Lett.* 1996;77(18):3865-3868.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.