

Accelerating the atomic catalyst discovery by theoretical calculations-machine learning strategy

Mingzi Sun, Dr. Alan William Dougherty, Prof. Bolong Huang, Prof. Yuliang Li, Prof. Chun-Hua Yan*

M. Sun, Dr. A. W. Dougherty and Prof. B. Huang
Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China

E-mail: bhuang@polyu.edu.hk

Prof. Y. Li
Key Laboratory of Organic Solids, Institute of Chemistry, Chinese Academy of Sciences, Beijing, 100190, PR China

Prof. C.-H. Yan
Key Laboratory of Nonferrous Metal Chemistry and Resources Utilization of Gansu Province, State Key Laboratory of Applied Organic Chemistry and College of Chemistry and Chemical Engineering, Research Center of Biomedical Nanotechnology, Lanzhou University, Lanzhou, 730000, PR China

Keywords: Graphdiyne, Atomic Catalyst, Density functional theory, Machine learning, Hydrogen evolution reaction

The atomic catalysts (AC) are emerging as a highly attractive research topic, especially in sustainable energy fields. Lack of a full picture of the hydrogen evolution reaction (HER) impedes the future development of potential electrocatalysts. In this work, the systematic investigation of HER process of graphdiyne (GDY) based AC has been presented regarding the adsorption energies, adsorption trend, electronic structures, reaction pathway, and active sites. This comprehensive work innovatively reveals GDY based AC for HER covering all the transition metals (TM) and lanthanide (Ln) metals, supplying the screening of potential catalysts. The density functional theory (DFT) calculations have carefully explored the HER performance beyond the comparison of sole H adsorption. Therefore, the screened catalysts candidates not only match with experimental results but also provide significant references for novel catalysts. Moreover, we also innovatively utilize the machine learning (ML) technique bag-tree approach based on the fuzzy model for data separation and converse prediction of the

HER performances, which indicates a similar result of the theoretical calculations. From two independent theoretical perspectives (DFT and ML), this work proposes pivotal guidelines for experimental catalyst design and synthesis. The proposed advanced research strategy shows great potential as a general approach in other energy-related areas.

1. Introduction

To actualize the goal of the sustainable energy system as the solution for the present energy crisis, searching for the economic electrocatalyst for the efficient production of hydrogen and oxygen by water-splitting is becoming the focus of the scientific community.^[1] In particular, the hydrogen evolution reaction (HER) process is of pivotal significance for fuel generation in many energy conversion systems.^[2] Presently, Pt-based electrocatalysts are still the dominant selections for the HER, which demands the more potential candidates due to the high cost and limited resources. As an alternative for Pt, the high electroactivity and highly stable materials with low costs have attracted tremendous attention, especially on the transition metal-based catalysts.^[3] Meanwhile, the non-metal electrocatalysts have also made great progress in recent years.^[4] However, compared to the performance of Pt-based electrocatalysts, those electrocatalysts still cannot satisfy the present requirements for efficient catalysts for practical applications. To overcome such challenge, atomic catalyst (AC) with the maximum explosion of active sites based on the minimum usage of electroactive metals have become the optimal solutions for the future sustainable energy solutions.^[5] In recent years, many noble metal based AC (i.e Pt, Au, Ru) have been reported with great potential in different electrocatalysis.^[6] Due to the high surface energy, the aggregation of AC is the main challenge during the synthesis and electrocatalytic process, which significantly lowers the active sites, electron transfer efficiency as well as the durability of the electrocatalysts. Thus, searching for stable anchoring AC with the appropriate support is a pivotal strategy to develop future electrocatalysts.

To achieve stable anchoring AC, various supports have been considered. For example, metal oxides have been reported by many groups in stabilizing the single transition metals.^[5a, 6a, 6d, 7] Zhang and his groups have actualized both Ir and Pt AC on the FeO_x, which showed high electroactivity for water gas shift reactions and CO oxidations, respectively. The single-atom transition metals show strong interactions with the FeO_x to facilitate the generation of surface defects, leading to remarkable performance.^[8] Li *et al.* have investigated the CeO₂ support for Au, which unravels the critical role of single Au atom in gold nanocatalysis and supplied rational design methodology.^[9] On the other hand, the non-metal 2D materials are also becoming the ideal support for AC due to the following merits including specific anchoring sites to stabilize the single atoms, the sufficient and flexible surface area to allow the high loading of catalytic sites, and the superior electron transfer capability to facilitate electrocatalysis. In particular, 2D materials such as graphene, C₃N₄, BN and carbon matrix also demonstrate superior electrocatalysis performances.^[6b, 10]

Recently, the graphdyine (GDY) has emerged as the novel electrocatalysts owing to their unique structural and electronic features. GDY possesses abundant benzene rings with *sp*²-hybridized carbon atoms and the diacetylenic groups with *sp*-hybridized carbon atoms as the connection chains, which exhibits excellent chemical/mechanical stability^[11]. Such a structure demonstrates a porous 2D structure with sufficient space for the anchoring of the AC. Compared with traditional 2D carbon materials, the unique electronic structure in GDY endows the materials with great potential in the electrocatalyst. Moreover, GDY also delivers an ideal electronic structure, which displays a high electronic conductivity and charge carrier mobility, leading to their remarkable performance in both electrocatalysis and photocatalysis.^[12] Especially, they have been reported as the optimal support to re^[13]alize the zero valence AC^[14]. The potential of GDY in the HER electrocatalysts is still waiting to be

unrevealed since only a few works have reported the synthesis of GDY based atomic catalysts (GDY-AC) ^[14a, 15]. Until now, most of the research of GDY is investigated case by case, which lacks the whole picture of the appropriate selection of anchoring metals. This challenge still suffers from limited experimental works and theoretical guidance to efficiently screen out the efficient electrocatalyst. In conventional routine calculations, the adsorption energy of the single H atom has usually been seen as the most pivotal parameters to evaluate the HER performance. However, such simple calculations cannot completely reveal the HER mechanism accurately since the multi-parameters in different reaction processes are equally crucial. The machine-learning (ML) technique is emerging as a promising solution to efficiently tackle the problem in the complex, multidimensional, and dynamic process in catalysis. ^[16] Recently, the ML technique has supplied key information in the design, synthesis, characterization and different applications of catalytic materials. Beyond discovering novel materials, ML technique also enables a deeper understanding of the underlying correlation between the material structure and their catalytic properties, which is beneficial for the establishment of the design principle and modification strategy to further improve the electrocatalytic efficiency.

In this work, we introduce the theoretical calculations with machine learning techniques to map out the potential GDY-AC electrocatalyst to achieve the efficient HER, which aims to reveal a comprehensive mapping regarding the HER mechanism and adsorption energies. We firstly comprehensively calculate the adsorption energies of H-species during HER on GDY-M (transition metal and lanthanide metals) for the first time regarding different active sites. Then we evaluate and screen out the potential GDY-AC electrocatalysts to reveal their electronic structures. Following the holistic mapping of all the GDT-AC systems, we introduce the machine learning technique to demonstrate the prediction of adsorption energies by consideration of multi parameters. This work supplies an innovative theoretical

understanding and guidance of GDY-based AC for HER through two independent approaches of theoretical investigations and machine learning techniques, which facilitates searching for more potential electrocatalysts to achieve efficient HER.

2. Results and Discussion

2.1. GDY-Transition metal Atomic Catalyst

Based on the first-principle DFT calculations to study the HER mechanism, we have considered the key steps including Volmer, Heyrovsky, and Tafel reactions, which are all essential in affecting the HER performances rather than one simple parameter. Initially, the overall reaction energies for all the transition metals by comparing the initial adsorption of H and final desorption of H₂ are displayed, which demonstrate the reaction trend from a direct perspective. For 3d transition metals, we notice that most metals show higher energies of desorption than the initial adsorptions, however, the energy difference is relatively small, except for Mn and Zn. These results indicate that Mn and Zn cannot be a good potential candidate for GDY-AC due to the low energetic preference of the HER process. Notably, Ni is the only metal that shows spontaneous reaction trend in 3d elements, indicating the overall good HER performance, which is also consistent with our previous work (**Figure 2a**).^[14a] In comparison, the 4d elements show a much larger energy difference between the initial adsorption and final desorption process, especially in from Y to Rh. Similar with Ni, Pd and Cd show the only energetically favorable HER performance. More importantly, the desorption energy of H₂ on GDY-Pd is highly close to the neutral line ($E = 0$ eV), supporting the efficient generation (**Figure 2b**). For the 5d metal-based GDY-AC, we notice that most metals are not the appropriate selections due to the overall endothermic HER, which requires the large applied potential to overcome. However, Pt shows a distinct behavior with other 5d metals, which shows a high energetic favorable reaction trend with a neglectable energy cost

for the final desorption of formed H_2 . Although Hg shows an exothermic reaction trend, the hinder for the initial H adsorption leads to the low efficiency of HER (**Figure 2c**).

Then, we move to the chemisorption of H as an important indicator for the preference of H coverage (**Figure 2d-2f**). For most of the transition metals, the chemisorption of H is energetically favorable, which demonstrates a durable H adsorption capability as a pivotal foundation for Heyrovsky reactions in HER. For 3d metals, we notice that the only Cu shows a negative trend for further H adsorption. It is worth noticed that Ni not only shows both spontaneous and mild adsorption trend for H, which prevents the GDY-Ni from the overbinding effect of H. The strong binding trend for H on GDY-Zn induces the overbinding, which leads to the large energy cost for final desorption of H_2 , supporting our previous findings (**Figure 2d**). The similar overbinding effects are discovered in Mo and Ru. Both Y and Ag shows the energy increase for chemisorption of H. Meanwhile, GDY-Pd shows a slightly lower adsorption trend than GDY-Cd (**Figure 2e**). Interestingly, although GDY-Au shows an extremely strong H adsorption, the chemisorption trend is still unfavorable, indicating a low efficiency of HER. The Pt and Hg show the most facile chemisorption trend of H, which further confirms their potential for efficient HER (**Figure 2f**). To reveal a more direct trend of the overall reaction of HER, we sketch out the energy difference between the initial adsorption and the final desorption of GDY-AC. For 3d metals, Zn locates at the valley point, indicating the most the energy unfavorable of HER. To achieve efficient HER without overbinding, the energy difference between the initial adsorption and desorption should remain in a small range of variation. The ultra-positive or ultra-negative variation leads to either the poor efficiency for H_2 formation or the overbinding effect, which both significantly affect HER performance. The large positive variation of the energy difference represents the final formation of H_2 requires a large energy cost, which lowers the formation efficiency. On the contrary, the ultra-negative variation indicates that the initial adsorption of H^* is too

strong, which results in the overbinding effect with low conversion efficiency to HER. From 3d-5d, the valley points are gradually decreasing. For 3d, Ni, Fe, Co are the most possible candidate for HER. Sc and Cr show slightly larger energy variation (**Figure 2g**). For 4d metals, Cd and Pd shows the smallest energy difference, supporting our conclusion at the chemisorption. In contrast, Ru locates at the middle valley point, demonstrating a relatively weak reduction capability to achieve the HER, which is highly consistent with the understanding of typical Ru catalysts as the efficient and widely used water-splitting catalysts (**Figure 2h**). Au exhibits the lowest point in 5d metals and the energy variation further increases. Only Pt shows a high potential for the HER with a nearly zero-energy variation. Compared to 3d, most 4d and 5d metals show the energy variation larger than 0.4 eV, indicating a slightly lower HER efficiency (**Figure 2i**).

2.2. GDY-Lanthanide metal atomic catalyst

Beyond the transition metals, the lanthanide elements are also investigated since rare earth has been applied as superior catalysts.^[7c, 17] Notably, most element shows the moderate energy variation between the initial adsorption and final desorption. Interestingly, the most widely used electrocatalyst Ce shows the poor thermodynamic trend for HER, which is ascribed to their strong oxidation capability rather than transfer electrons from H. In contrast, several other elements display great potential for promoting the HER including Sm, Eu, Gd and Ho, in which Sm, Eu, and Ho shows the overall spontaneous reaction in generating H₂ (**Figure 3a**). Moreover, a more direct reaction trend of the HER reaction based on GDY-Ln electrocatalysts. Pm shows the abnormal energy variation due to the extremely high energy cost for the initial H adsorption. Besides Pm, La, and Ce, all the lanthanide ions show the energy variation in a limited range, in which Sm, Gd, and Eu locate closest to the natural line. Pr, Tm, and Ho are also potential candidates with slightly larger energy (**Figure 3b**). To further screen the potential electrocatalyst, the energy pathway of the HER process regarding

the transition state energies are illustrated to reveal a practical thermodynamic energetic trend of the Ln-GDY for HER. We have considered the chemisorption energies for the H* to examine the formation efficiency of H₂. Although some atomic catalysts show the similar H* adsorption, the different chemisorption energies result in the distinct HER performances.^[18] Notably, both Sm and Eu only require small energy costs within 0.2 eV for the adsorption of the second H atom. Meanwhile, the desorption of H₂ only requires subtle energy cost, which endows them great potential in facilitating the HER. Pm shows a dramatic energy drop for the consecutive adsorption of H atoms, which laid a formidable energetic impedance for the desorption of H₂. For Ce and La, they all meet an evident uphill thermodynamic trend for the formation of H₂, which lowers HER efficiency (**Figure 3c**). As the atomic radius becomes smaller, the energy variation range of consecutive H adsorption becomes larger. Tm shows facile adsorption of the second H atom on the surface with 0.2 eV energy cost for the formation of H₂. Although Dy and Gd show the overall spontaneous reaction trend of formation of H₂, however, the obvious energetic unfavorable for the reaction [2H* → *H₂] excludes them from the HER electrocatalyst candidate (**Figure 3d**). Therefore, based on the careful investigation of the reaction energy change, we have screened out GDY-Eu and GDY-Sm as two validate electrocatalysts to achieve the superior HER performances, which supplies valuable references and guidance for the experimental synthesis.

2.3. Electronic structures

After the comparison from the energetic view, we also investigate the electronic structures based on the projected partial density of states (PDOSs) to further unravel the reaction trend. The Fermi level is the surface where all the electrons below the Fermi level are localized. With closer position to the E_F, the electrons meet smaller barriers to be depleted to delocalization and further transfer to the adsorbates within the HER process. Therefore, the closer position to the E_F correlates to the higher electroactivity in promoting the electron

transfer for electrocatalysis. For the 3d metals, the d orbitals show a very diverse trend. From Sc to V, the dominant peak of 3d locates above the Fermi level (E_F). Compare to Mn, Cr shows a much closer position to the E_F , which is similar to the positions of Ni, Co, and Fe, indicating a good electron transfer capability as well. For Cu and Zn with d10 electronic configurations, both 3d orbitals are electron-rich, locating at a deep position from E_F , which significantly limits the electron transfer pathway (**Figure 4a**). For 4d elements, the gradual shift of the 4d dominant peaks is noticed. From Mo to Rh, the evident gap between t_{2g} and e_g demonstrate the barrier for electron transfer in HER. Pd shows the high electroactivity due to the dominant 4d peak close to E_F . Ag and Cd show similar electronic structures with Cu and Zn, demonstrating the low possibility of Cd to be the electrocatalyst (**Figure 4b**). From the PDOSs, Os, Ir, and Pt both show the electroactivity due to the close positions of d orbitals to E_F . Distinct from Zn and Cd, the dominant 5d peak of Hg shift towards $E_V-4.5$ eV ($E_V=0$), which is similar to the Au-5d (**Figure 4c**). Then, we further demonstrate the detailed PDOSs of the screened electrocatalyst. Compared to Ni, the Co-3d peak is even closer to E_V , illustrating the even stronger electron transfer capability for HER. The Co-3d peak locates at $E_V-0.5$ eV while Ni-3d peak locates $E_V-1.0$ eV (**Figure 4d-4e**). GDY-Pd also exhibits a similar d orbital position with high electroactivity, which supports our previous work.^[15b] Moreover, GDY-Pt demonstrates a slightly lower 5d position at $E_V-1.5$ eV (**Figure 4f-4g**). These trends further confirm that the electroactivity of GDY-AC follows the order as 3d > 4d > 5d, which is in a good agreement with the energetic trend. In addition, the electronic structures of screened rare earth candidates are also reflected. Both Eu-4f and Sm-4f orbitals are highly electroactive for the electron depletion based reduction. These electronic structures are closely consistent with the energetic trends of GDY-Eu and GDY-Sm, which also confirms the screening results of catalysts from the energetic mapping (**Figure 4h-4i**).

2.4. Active sites and reaction energy mapping

The determination of active sites in GDY is of great importance as well. To unravel the most electroactive sites for the HER, we have summarized the initial H adsorption energies on different sites of GDY-AC as shown in **Figure 5a**. Based on the previous works,^[14a, 15b] active sites for H mainly concentrate on the C-chain (denoted as C1 and C2) and the single metal atom (denoted as M1). Notably, we can see that C2 are the most stable sites for the initial adsorption, which is well-matched with our previous reports. For the newly screened electrocatalyst, Co and Pt also prefer the C2 sites. On the contrary, the atomic metal position is not an ideal position for stable H adsorption (**Figure 5a**). On the other hand, the lowest adsorption energies of formed H₂ are also revealed. Obviously, the single metal atom displays a strong binding trend with the formed H₂, which is possible to induce the overbinding effect. C1 shows the most unstable adsorption of H₂, which lays a good foundation for the release of formed H₂ in HER. The active site investigation further confirms that C2 on the GDY chain is the most electroactive site, which displays both stable adsorptions of H and the stable desorption of H₂ (**Figure 5b**). We also summarize all the chemisorption energies of elements along the periodic table order from III group to II group. Overall, the chemisorption shows limited regulation with the number of d electrons. However, from the 3d-5d, we indeed notice the increase of chemisorption, which demonstrates the H coverage preference is decreasing. These behaviors further verify our previous conclusion on the change of electroactivity for HER in the periodic table (**Figure 5c**). In the end, we also carefully map out and compare the energy pathway of HER process in all the GDY-TM systems. For 3d metals, both consecutive H adsorption and final desorption of H₂ are thermodynamically spontaneous while the formation of H₂ is not energetically favorable, especially for Sc to Mn. The formation cost of H₂ follows the order of Fe > Co > Ni. In particular, GDY-Ni shows very subtle energy change for the HER process, supporting the highly efficient HER observed in experiments (**Figure 5d**). The energy cost of [2H* → *H₂] is also the main potential determining criteria for 4d elements. The energetically favorable candidates are Pd, Ag, and Cd (**Figure 5e**).

Furthermore, the spontaneous reaction trend of $[2\text{H}^* \rightarrow \text{H}_2]$ only occurs in Pt and Hg. The left 5d elements all require high energy costs to the formation of H_2 , representing the anticipated poor HER performance (**Figure 5f**). The further summary of the energetic reaction trend not only confirms our previous screening on the electrocatalysts but also identifies the active sites for GDY-AC.

2.5. Machine learning (ML)

In previous work,^[19] we have presented the potential approach Fuzzy C-Means (FCM) model for the data separation by ML. Based on such a concept, we introduce a more in-depth bag-tree algorithm in ML techniques to develop the potential strategy for identifying the GDY catalysts and determining factors. The input data for ML includes our theoretical calculation results, the electronegativity, electron affinity, ionization potential, d/f electrons, active sites, and mass numbers. With all these parameters, we compare the predicted values of adsorption energies of H, 2H, H_2 with our calculated values. The predicted values have shown a relatively high accuracy since ML usually requires a much larger database to achieve highly accurate results. However, the chemisorption energies of H show abnormal patterns of the predicted values, which indicates the limitation of ML in the prediction of H coverage based on the present algorithm (**Figure 6a-6d**). In our previous mapping work, we have proposed the concept of the redox energy barrier to quantify the electron transfer capability of GDY-AC.^[19] Thus, we further supply the redox energy barrier of all the elements as a new parameter to improve the prediction by ML. For the H adsorption, the prediction shows similar results. In comparison, the prediction accuracy of 2H^* adsorption energies has been improved with closer results. The prediction accuracy for adsorption of H_2 shows limited changes as well. These results indicate that the adsorption of 2H^* tightly correlates to the redox energy barrier, which is attributed to the energy transfer in 2H^* instead of the non-

electron exchange physical adsorption. This also explains the pivotal role of $[2\text{H}^* \rightarrow ^*\text{H}_2]$ reaction in HER (**Figure 6e-6h**).

Finally, we compared the energetic trend between theoretical and ML techniques to validate the potential of ML in the future screening of electrocatalysts. Due to the highest ML prediction accuracy, we compared the energies of 2H^* between theoretical calculation results and ML predicted values. Overall, the ML technique reproduces the adsorption energies in relatively high accuracy. The overall adsorption energies trend has been well preserved while the detailed order between some elements still requires further improvements in the ML algorithm. However, the highly similar trend already demonstrates the capability of ML in facilitating the prediction of electroactivity of novel catalysts. Both DFT and ML present a similar trend in the electroactivities, which validate the screening catalyst candidates for future experimental synthesis (**Figure 7a-7b**). In this work, the innovative investigations of both theoretical calculations and ML mapping technique deepens our understanding of the atomic catalysts, which open new opportunities in future design and synthesis of efficient AC for the experimental scientists as well.

3. Conclusion

In this work, by taking the GDY-M based systems as the example, we have revealed the HER performance based on both DFT calculations and machine learning strategy. Beyond the simple comparison of only H adsorption energies, we investigate the HER performance from a comprehensive perspective including adsorption energies of all intermediates, electronic structures, and active sites. We have successfully predicted the potential GDY based atomic catalyst for HER including rare-earth based material for the first time. Meanwhile, it is also the first time, we apply the bagged-tree method as the machine learning algorithm to train the system to predict the adsorption energies for different AC systems, which is developed based

on our previously proposed fuzzy model concept. With a limited database of GDY based AC, we have obtained a similar trend of the electroactivity between theoretical calculations and machine learning, which not only reveals a full picture of HER process but also opens the opportunities of machine-learning driven research strategy for broad fields in future energy-related applications.

4. Methods

Theoretical Calculations: The density functional theory (DFT) has been applied for all the calculations to investigate the electronic and energetic of GDY based atomic catalysts within the CASTEP codes ^[20]. For the functionals, the generalized gradient approximation (GGA) with the Perdew-Burke-Ernzerhof (PBE) is chosen to reveal the exchange-correlation energy of electron interactions ^[21]. The plane-wave cutoff energy has been set to 380 eV within the ultrasoft pseudopotential scheme for all the geometry optimizations ^[22]. For the k-point mesh to achieve the energy minimization, the Broyden-Fletcher-Goldfarb-Shannon (BFGS) algorithm is applied ^[23]. The convergence criteria for the geometry optimizations have been set as that the Hellmann-Feynman forces will be less than 0.001 eV/Å meanwhile the total energy should not exceed 5×10^{-5} eV per atom. The vacuum space in the z-direction was set as 15 Å to prevent the interaction between periodic unit cells.

The adsorption energy was calculated according to the equation below:

$$E_{\text{ads}} = E_{\text{total}} - E_{\text{slab}} - E_{\text{adsorbate}} \quad (1)$$

where the E_{ads} is the adsorption energy, E_{total} is the total energy of the system with adsorbate, E_{slab} is the energy of the optimized GDY and $E_{\text{adsorbate}}$ is the energy of the isolated adsorbate in the gas phase. In this framework, the free energy of the electron-proton pair ($\text{H}^+ + \text{e}^-$) can be referenced to the chemical potential of gaseous H_2 at equilibrium (0 V vs standard hydrogen electrode). The change of free energy can be calculated as follows,

$$\Delta G = \Delta E + \Delta E_{\text{ZPE}} - T\Delta S \quad (2)$$

where ΔE is the electronic energy difference directly obtained by DFT calculations. ΔE_{ZPE} is the change in the zero-point energy (ZPE), T is the temperature and ΔS is the entropy change. Considering the substrate vibration, the zero point energy of the GDY is around 1.5 eV to 3.0 eV for different atomic catalyst systems. Thus, with the consideration of thermal perturbation effect under different temperatures, some of the adsorption energies obtain a positive value.

Machine Learning Algorithms: We introduce a possible algorithm in machine-learning to classify the complex database and screen out potential candidates. The Fuzzy C-Means (FCM) algorithm has a strong ability to classify a data-set into groups based on data-clustering, whilst maintaining any inherent uncertainty between these groups. The FCM algorithm is an unsupervised learning concept. The optimization objective function of the standard FCM algorithm $J = \sum_{i=1}^N \sum_{j=1}^M u_{i,j}^p \|x_i - c_j\|^2$ has an important uncertainty term $u_{i,j}^p$ which is known as the membership function and i, j represent the vector.

$$J_{SAFCM} = \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^{\Omega} u_{k|i,j}^p \|\eta_{i,j} c_k\|^2 + \lambda \sum_{i=1}^N \sum_{j=1}^M (\eta_{i,j} - \Gamma(\eta_{i,j}))^2 \quad (3)$$

Additionally, one must note that the FCM algorithm works much like other clustering algorithms, such that, it forms classes by looking at a particular distance metric between the data points and the class-cluster mean. The class cluster Mean in $J = \sum_{i=1}^N \sum_{j=1}^M u_{i,j}^p \|x_i - c_j\|^2$ is denoted by c , and the distance between the data-point and the mean is often found using distance metrics such as the Euclidean distance. Eq. (3) shows a modified FCM algorithm which considers this spatial relationship between the data-points, $\eta_{i,j}$ is a local data term that biases certain data-points likelihood to belonging to a cluster based on the amount of other similar-valued data, which exist locally to it. An additional term Γ , which is a filter response

around a local area of any data-point x , is also added to the equation for Lagrangian optimization of $\eta_{i,j}$.

To further enhance results, if a data-set has some clusters which are not easily linearly separable, we can further develop a machine-learning algorithm to utilize Kernel functions. Essentially, the kernel function is a mapping function, which can take any n-dimensional data-set and map it to a higher N-dimensional one to allow for easier separation. In the case of the FCM, by combining the ability of Kernel function with its ability to model uncertainties within data-sets, we can significantly increase the likelihood of correctly classifying a data-set, which has unusually distributed and overlapping features. One such example of a Kernel-based version of Eq. (1) is shown in Eq. (2), where K is the radial basis function Kernel [24].

$$J_{KAFCM} = \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^{\Omega} u_{k|i,j}^p (1 - K(x_i, \eta_{i,j} c_k)) + \lambda \sum_{i=1}^N \sum_{j=1}^M (1 - K(\eta_{i,j}, \eta_{i,j} D_{i \in \phi})) \quad (4)$$

With the abovementioned data separation model, we further apply the more detailed machine learning algorithm to realize the pathway in this work. Due to the small size of the available dataset, and the need for the results to be interpretable, we had initially chosen to use commercially available machine learning toolboxes within MATLAB to train a model for prediction. The reason behind a desire for interpretability is for gaining further insight into the dataset for future algorithm development, i.e. which features are heavily influencing prediction and where should we focus future data acquisition.

The bagged-tree method [25] was empirically found to give the best results without obvious over-fitting for this dataset. For example, it would be possible to obtain "near-perfect" results

for a small amount of input data using an algorithm such as mixed Gaussian models. However, this was practically observed to have an obvious overfitting problem on training data leading to poor validation results. One advantage that the decision tree method is that all the actions are fully under control and well-understood, which supplies significant knowledge to guide future models.

As an ensemble learning method, bagging is advantageous for small size datasets due to its smoothing operation and combination of weak learners to form an overall strong learner. Bagging uses multiple individual decision trees which would be the so-called "weak learner", and when a new classification needs to be performed, it is done so on each of the individual trees. The final classification is decided to be the one that gained the most votes from each of the individual trees.^[26]

All of the models trained in this work were done so using 20-fold cross-validation, which was found empirically to give the best results overall. As can be seen from the results, the predicted versus real values track relatively well with few outliers, representing the validity of such a model for prediction.

Acknowledgements

The authors gratefully acknowledge the support of the Natural Science Foundation of China (Grant No.: NSFC 21771156), and the Early Career Scheme (ECS) fund (Grant No.: PolyU 253026/16P) from the Research Grant Council (RGC) in Hong Kong.

Received: ((will be filled in by the editorial staff))

Revised: ((will be filled in by the editorial staff))

Published online: ((will be filled in by the editorial staff))

References

- [1] a) I. Roger, M. A. Shipman, M. D. Symes, *Nature Reviews Chemistry* **2017**, *1*; b) C. Wei, R. R. Rao, J. Peng, B. Huang, I. E. L. Stephens, M. Risch, Z. J. Xu, Y. Shao-

- Horn, *Adv. Mater.* **2019**, *31*, e1806296; c) J. Zhang, Q. Zhang, X. Feng, *Adv. Mater.* **2019**, *31*, 1808167; d) C. Hu, L. Zhang, J. Gong, *Energ Environ Sci* **2019**, *12*, 2620-2645.
- [2] a) Y. Zhu, W. Zhou, Y. Zhong, Y. Bu, X. Chen, Q. Zhong, M. Liu, Z. Shao, *Advanced Energy Materials* **2017**, *7*, 1602122; b) A. Kondori, M. Esmailirad, A. Baskin, B. Song, J. Wei, W. Chen, C. U. Segre, R. Shahbazian - Yassar, D. Prendergast, M. Asadi, *Advanced Energy Materials* **2019**, *9*, 1900516; c) H. Sun, Z. Yan, F. Liu, W. Xu, F. Cheng, J. Chen, *Adv. Mater.* **2019**, e1806326.
- [3] a) M. Zeng, Y. Li, *Journal of Materials Chemistry A* **2015**, *3*, 14942-14962; b) Z. Chen, X. Duan, W. Wei, S. Wang, B.-J. Ni, *Journal of Materials Chemistry A* **2019**, *7*, 14971-15005.
- [4] a) Y. Zheng, Y. Jiao, Y. Zhu, L. H. Li, Y. Han, Y. Chen, A. Du, M. Jaroniec, S. Z. Qiao, *Nat Commun* **2014**, *5*, 3783; b) S. Yao, X. Zhang, A. Chen, Z. Zhang, M. Jiao, Z. Zhou, *Journal of Materials Chemistry A* **2019**, *7*, 19290-19296.
- [5] a) A. Wang, J. Li, T. Zhang, *Nature Reviews Chemistry* **2018**, *2*, 65-81; b) A. Beniya, S. Higashi, *Nature Catalysis* **2019**, *2*, 590-602; c) B. Qiao, A. Wang, X. Yang, L. F. Allard, Z. Jiang, Y. Cui, J. Liu, J. Li, T. Zhang, *Nature Chemistry* **2011**, *3*, 634-641; d) W. H. Lai, Z. Miao, Y. X. Wang, J. Z. Wang, S. L. Chou, *Advanced Energy Materials* **2019**, *9*, 1900722; e) S. Ding, M. J. Hülsey, J. Pérez-Ramírez, N. Yan, *Joule* **2019**, *3*, 2897-2929.
- [6] a) X. Li, W. Bi, L. Zhang, S. Tao, W. Chu, Q. Zhang, Y. Luo, C. Wu, Y. Xie, *Adv. Mater.* **2016**, *28*, 2427-2431; b) H. Yan, H. Cheng, H. Yi, Y. Lin, T. Yao, C. Wang, J. Li, S. Wei, J. Lu, *J. Am. Chem. Soc.* **2015**, *137*, 10484-10487; c) L. Cao, Q. Luo, J. Chen, L. Wang, Y. Lin, H. Wang, X. Liu, X. Shen, W. Zhang, W. Liu, Z. Qi, Z. Jiang, J. Yang, T. Yao, *Nat Commun* **2019**, *10*, 4849; d) R. Lang, W. Xi, J. C. Liu, Y. T. Cui, T. Li, A. F. Lee, F. Chen, Y. Chen, L. Li, L. Li, J. Lin, S. Miao, X. Liu, A. Q. Wang, X. Wang, J. Luo, B. Qiao, J. Li, T. Zhang, *Nat Commun* **2019**, *10*, 234; e) L. Zhang, L. Han, H. Liu, X. Liu, J. Luo, *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 13694-13698.
- [7] a) S. F. Hackett, R. M. Brydson, M. H. Gass, I. Harvey, A. D. Newman, K. Wilson, A. F. Lee, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 8593-8596; b) F. Li, Y. Li, X. C. Zeng, Z. Chen, *ACS Catalysis* **2014**, *5*, 544-552; c) Y. G. Wang, D. Mei, V. A. Glezakou, J. Li, R. Rousseau, *Nat Commun* **2015**, *6*, 6511; d) J.-C. Liu, Y.-G. Wang, J. Li, *J. Am. Chem. Soc.* **2017**, *139*, 6190-6199; e) Y. Tang, Y.-G. Wang, J. Li, *The Journal of Physical Chemistry C* **2017**, *121*, 11281-11289; f) K. Ding, A. Gulec, A. M. Johnson, N. M. Schweitzer, G. D. Stucky, L. D. Marks, P. C. Stair, *Science* **2015**, *350*, 189-192.
- [8] a) J. Lin, A. Wang, B. Qiao, X. Liu, X. Yang, X. Wang, J. Liang, J. Li, J. Liu, T. Zhang, *J. Am. Chem. Soc.* **2013**, *135*, 15314-15317; b) B. Qiao, A. Wang, X. Yang, L. F. Allard, Z. Jiang, Y. Cui, J. Liu, J. Li, T. Zhang, *Nat Chem* **2011**, *3*, 634-641; c) H. Wei, X. Liu, A. Wang, L. Zhang, B. Qiao, X. Yang, Y. Huang, S. Miao, J. Liu, T. Zhang, *Nat Commun* **2014**, *5*, 5634.
- [9] J. C. Liu, Y. G. Wang, J. Li, *J. Am. Chem. Soc.* **2017**, *139*, 6190-6199.
- [10] a) C. Gao, S. Chen, Y. Wang, J. Wang, X. Zheng, J. Zhu, L. Song, W. Zhang, Y. Xiong, *Adv. Mater.* **2018**, *30*, e1704624; b) M. Fan, J. D. Jimenez, S. N. Shirodkar, J. Wu, S. Chen, L. Song, M. M. Royko, J. Zhang, H. Guo, J. Cui, K. Zuo, W. Wang, C. Zhang, F. Yuan, R. Vajtai, J. Qian, J. Yang, B. I. Yakobson, J. M. Tour, J. Lauterbach, D. Sun, P. M. Ajayan, *ACS Catalysis* **2019**, *9*, 10077-10086; c) X. Liu, Y. Jiao, Y. Zheng, M. Jaroniec, S. Z. Qiao, *J. Am. Chem. Soc.* **2019**, *141*, 9664-9672; d) H. Zhang, P. An, W. Zhou, B. Y. Guan, P. Zhang, J. Dong, X. W. D. Lou, *Sci Adv* **2018**, *4*, eaao6657; e) H. Zhang, W. Zhou, T. Chen, B. Y. Guan, Z. Li, X. W. Lou, *Energ Environ Sci* **2018**, *11*, 1980-1984.

- [11] a) Y. Zhao, J. Wan, H. Yao, L. Zhang, K. Lin, L. Wang, N. Yang, D. Liu, L. Song, J. Zhu, L. Gu, L. Liu, H. Zhao, Y. Li, D. Wang, *Nat Chem* **2018**, *10*, 924-931; b) Y. Zhao, J. Wan, H. Yao, L. Zhang, K. Lin, L. Wang, N. Yang, D. Liu, L. Song, J. Zhu, L. Gu, L. Liu, H. Zhao, Y. Li, D. Wang, *Nature Chemistry* **2018**, *10*, 924-931.
- [12] a) S. Guo, Y. Jiang, F. Wu, P. Yu, H. Liu, Y. Li, L. Mao, *ACS Appl Mater Interfaces* **2019**, *11*, 2684-2691; b) X. Gao, H. Liu, D. Wang, J. Zhang, *Chem. Soc. Rev.* **2019**, *48*, 908-936; c) R. Sakamoto, R. Shiotsuki, K. Wada, N. Fukui, H. Maeda, J. Komeda, R. Sekine, K. Harano, H. Nishihara, *Journal of Materials Chemistry A* **2018**, *6*, 22189-22194; d) Z. Zuo, D. Wang, J. Zhang, F. Lu, Y. Li, *Adv. Mater.* **2019**, *31*, e1803762; e) J. Li, X. Gao, X. Jiang, X.-B. Li, Z. Liu, J. Zhang, C.-H. Tung, L.-Z. Wu, *ACS Catalysis* **2017**, *7*, 5209-5213.
- [13] L. Hui, Y. Xue, H. Yu, Y. Liu, Y. Fang, C. Xing, B. Huang, Y. Li, *J. Am. Chem. Soc.* **2019**, *141*, 10677-10683.
- [14] a) Y. Xue, B. Huang, Y. Yi, Y. Guo, Z. Zuo, Y. Li, Z. Jia, H. Liu, Y. Li, *Nat Commun* **2018**, *9*, 1460; b) H. Yu, Y. Xue, B. Huang, L. Hui, C. Zhang, Y. Fang, Y. Liu, Y. Zhao, Y. Li, H. Liu, Y. Li, *iScience* **2019**, *11*, 31-41.
- [15] a) X.-P. Yin, H.-J. Wang, S.-F. Tang, X.-L. Lu, M. Shu, R. Si, T.-B. Lu, *Angew. Chem. Int. Ed.* **2018**, *57*, 9382-9386; b) H. Yu, Y. Xue, B. Huang, L. Hui, C. Zhang, Y. Fang, Y. Liu, Y. Zhao, Y. Li, H. Liu, *iScience* **2019**, *11*, 31-41.
- [16] a) T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K.-I. Shimizu, *ACS Catalysis* **2019**; b) G. H. Gu, J. Noh, I. Kim, Y. Jung, *Journal of Materials Chemistry A* **2019**, *7*, 17096-17117.
- [17] a) K. Tomishige, Y. Furusawa, Y. Ikeda, M. Asadullah, K. Fujimoto, *Catal. Lett.* **2001**, *76*, 71-74; b) P. Mazierski, A. Mikolajczyk, B. Bajorowicz, A. Malankowska, A. Zaleska-Medynska, J. Nadolna, *Applied Catalysis B: Environmental* **2018**, *233*, 301-317.
- [18] a) V. Pallassana, M. Neurock, L. B. Hansen, B. Hammer, J. K. Nørskov, *Physical Review B* **1999**, *60*, 6146-6154; b) G. Papoian, J. K. Nørskov, R. Hoffmann, *J. Am. Chem. Soc.* **2000**, *122*, 4129-4144.
- [19] M. Sun, T. Wu, Y. Xue, A. W. Dougherty, B. Huang, Y. Li, C.-H. Yan, *Nano Energy* **2019**, *62*, 754-763.
- [20] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, M. C. Payne, *Zeitschrift Fur Kristallographie* **2005**, *220*, 567-570.
- [21] a) J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865-3868; b) P. J. Hasnip, C. J. Pickard, *Comput. Phys. Commun.* **2006**, *174*, 24-29; c) J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais, *Physical Review B* **1992**, *46*, 6671-6687.
- [22] D. Vanderbilt, *Physical Review B* **1990**, *41*, 7892-7895.
- [23] a) J. D. Head, M. C. Zerner, *Chem. Phys. Lett.* **1985**, *122*, 264-270; b) M. I. J. Probert, M. C. Payne, *Physical Review B* **2003**, *67*.
- [24] A. W. Dougherty, J. You, **2017**, 198-205.
- [25] a) L. Breiman, *Machine Learning* **2001**, *45*, 5-32; b) A. Liaw, M. Wiener, *R news* **2002**, *2*, 18-22.
- [26] S. Athey, J. Tibshirani, S. Wager, *The Annals of Statistics* **2019**, *47*, 1148-1178.

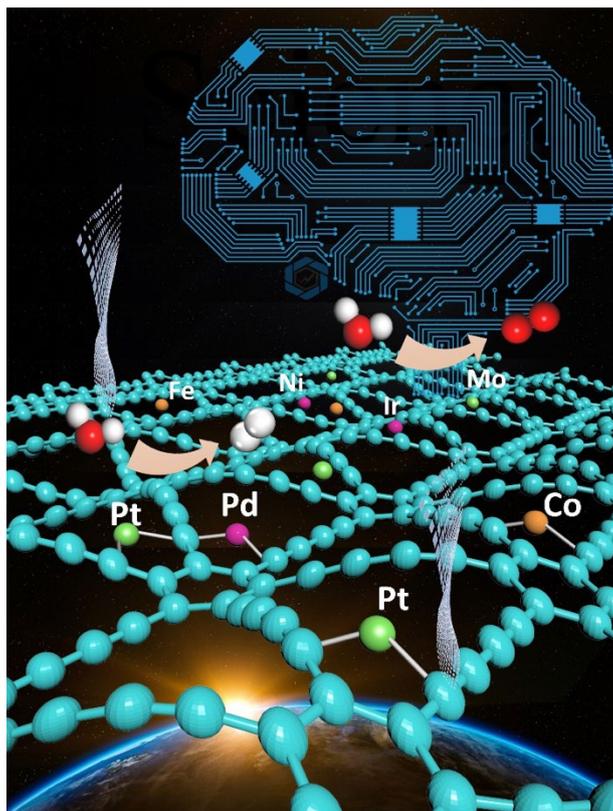


Figure 1. Schematic Figure for the introduction of machine-learning for screening GDY based AC for HER.

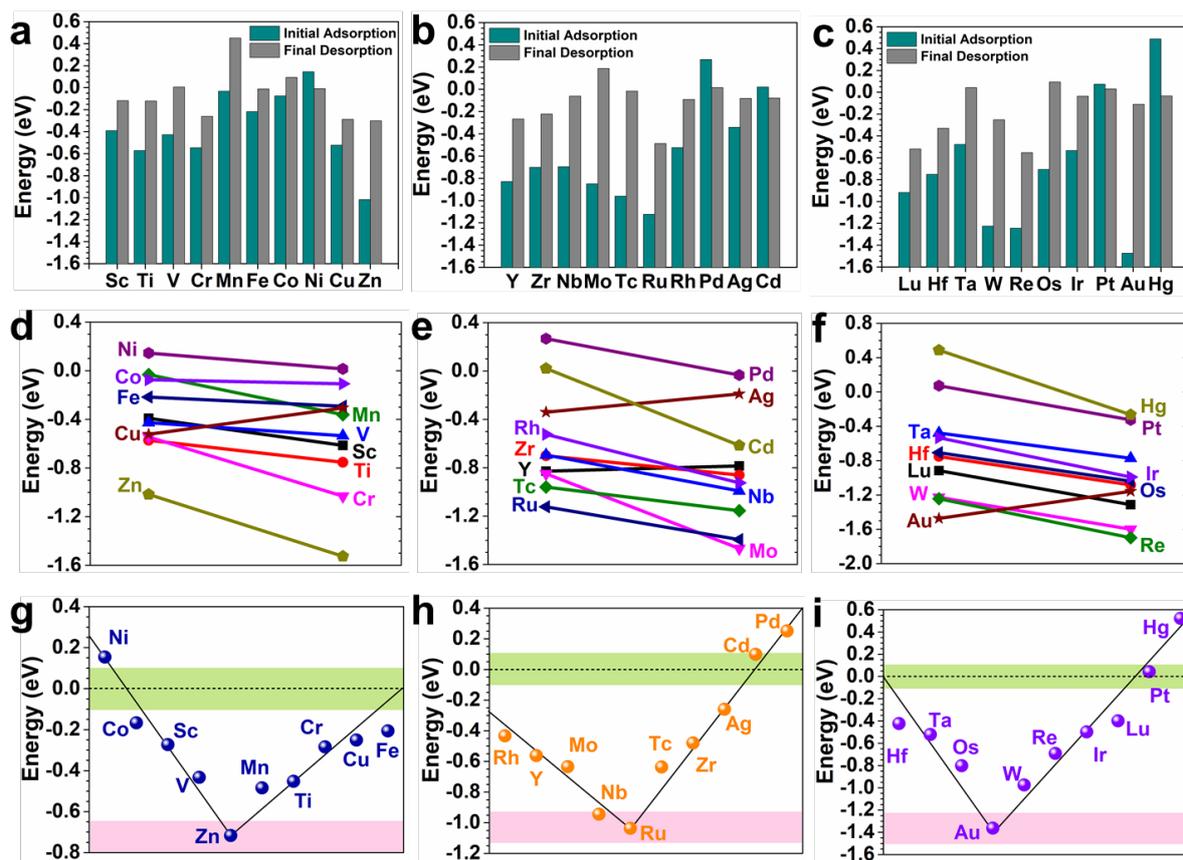


Figure 2. DFT calculations for energies diagram of GDY-transition-metal based AC. The energy of initial adsorption and final desorption of (a) 3d, (b) 4d and (c) 5d transition metals based GDY AC. Without considering the energy changes within the HER process, the direct comparison of the final and initial reactions demonstrates the reaction preference. The chemisorption trend of H on (d) 3d, (e) 4d and (f) 5d transition metals based GDY AC. The volcano plot of the HER reaction energy of (g) 3d, (h) 4d and (i) 5d transition metals.

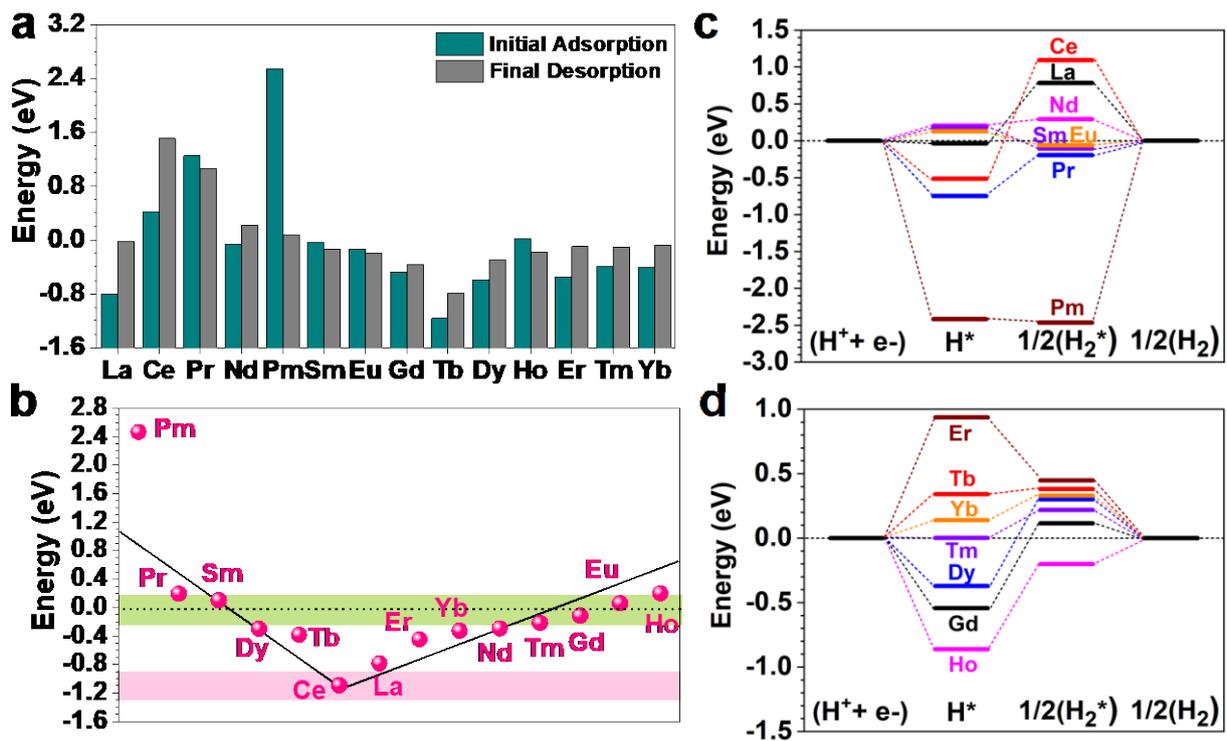


Figure 3. DFT calculations for energies diagram of GDY-lanthanide-metal based AC. (a) The energy of initial adsorption and final desorption of HER on the lanthanide-based GDY AC. Without considering the energy changes within the HER process, the direct comparison of the final and initial reactions demonstrates the reaction preference. **(b)** The volcano plot of the reaction energy of HER. **(c, d)** The free energy diagram of the HER process.

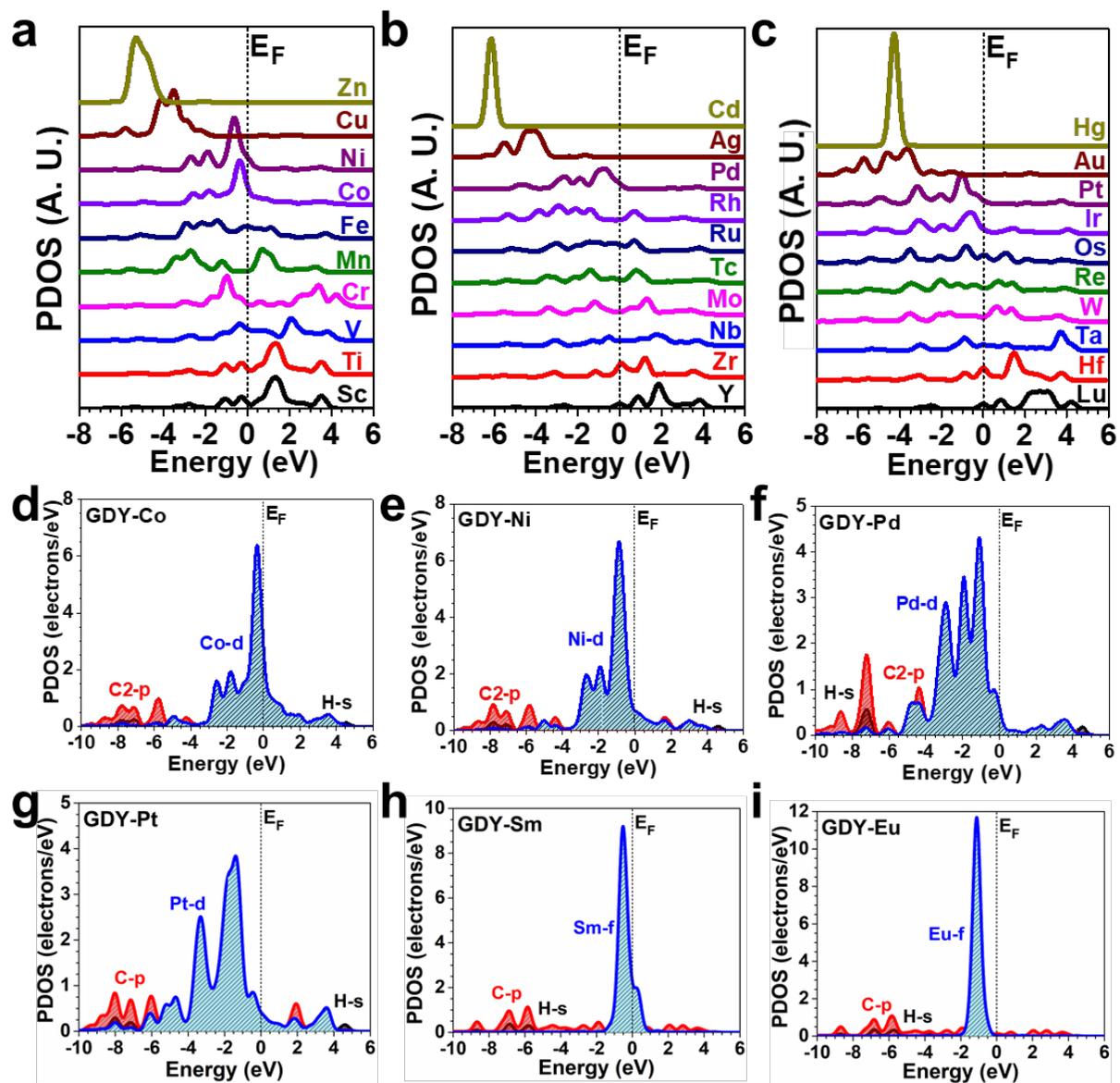


Figure 4. Electronic structures for GDY-metal based AC. The PDOSs of GDY-transition-metal based AC of (a) 3d, (b) 4d and (c) 5d. (d-i) The PDOSs of screened AC for HER. (d) PDOS of GDY-Co. (e) PDOS of GDY-Ni. (f) PDOS of GDY-Pd. (g) PDOS of GDY-Pt. (h) PDOS of GDY-Sm. (i) PDOS of GDY-Eu.

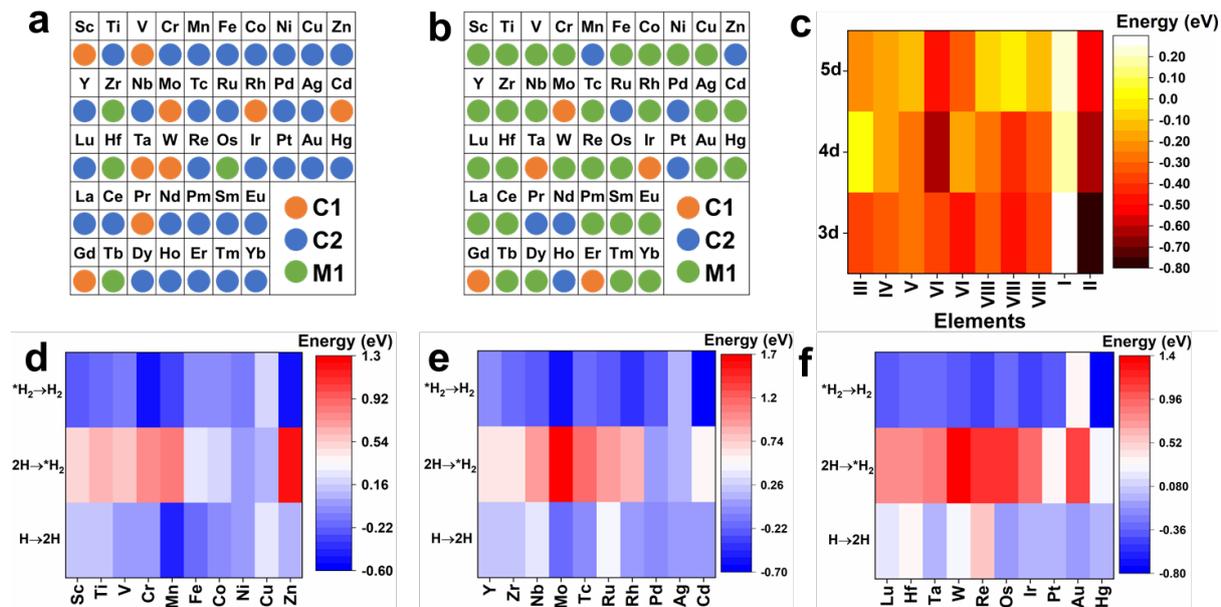


Figure 5. DFT mapping of active sites and reaction energies for GDY-metal based AC. (a) Mapping of the preferable initial adsorption sites for H* in HER. (b) Mapping of the final desorption sites for H₂ in HER. (c) Mapping of the chemisorption energies for GDY-transition-metal based AC. Mapping of the reaction energies of HER for (d) 3d, (e) 4d and (f) 5d GDY-transition-metal based AC.

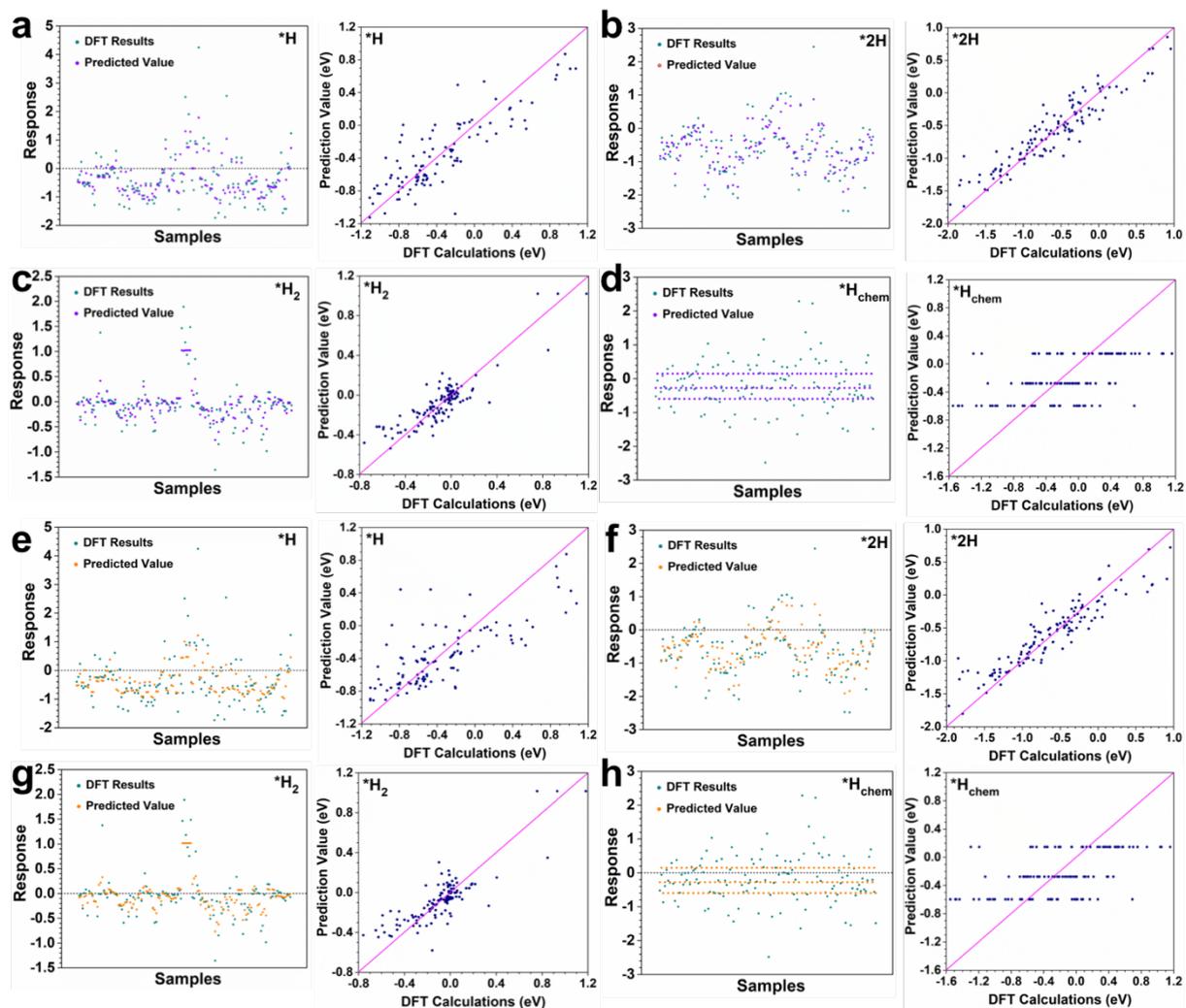


Figure 6. Introduction of machine learning for transition metal AC on GDY. (a) – (d) The machine learning predicted values and comparison with theoretical calculations by considering the fundamental chemical properties of each metal. (e)–(g) The improved machine learning predicted values and comparison with theoretical calculations by further considering the redox barrier of electron transfer.

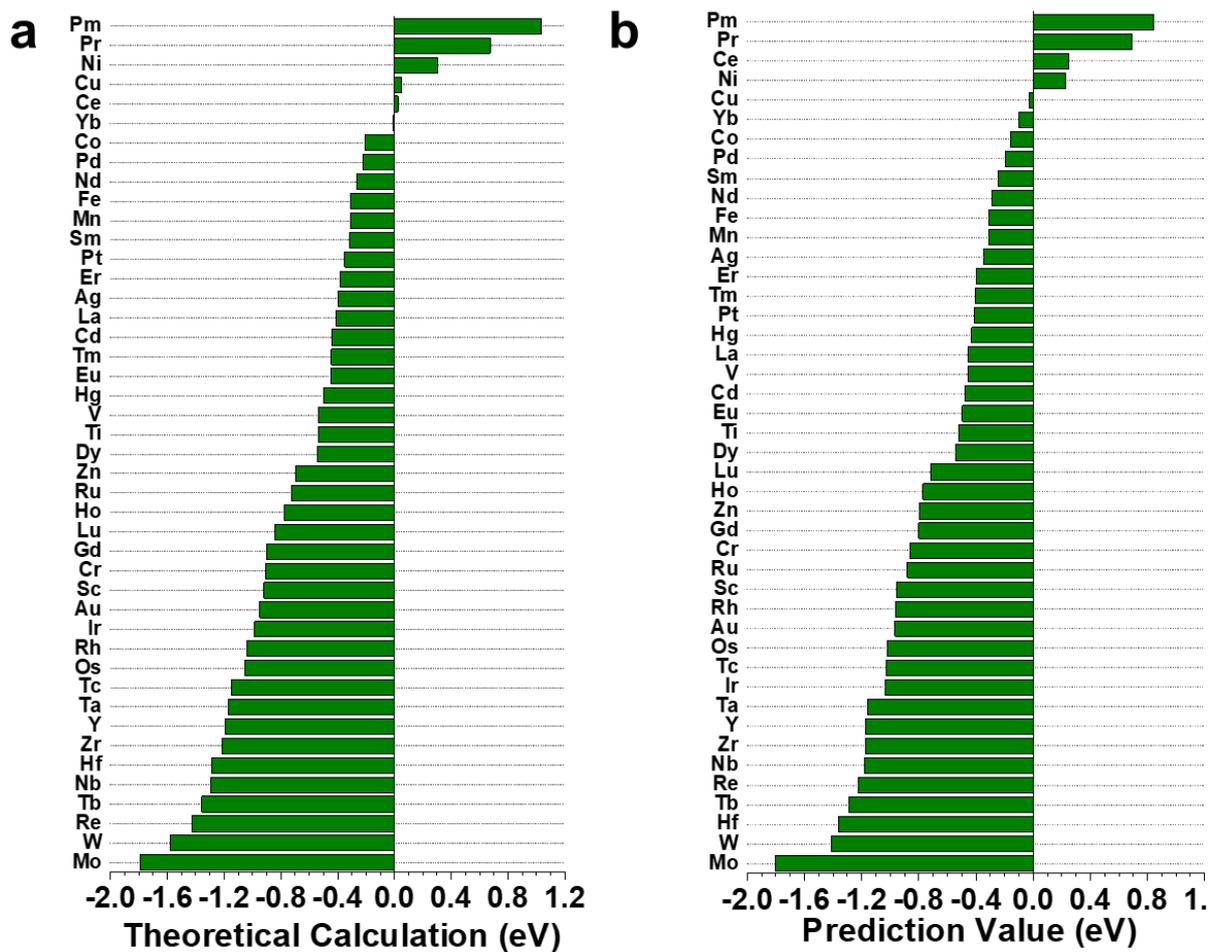


Figure 7. Comparison of (a) theoretical calculation and (b) Machine learning prediction values of 2H adsorption energies.

With the assistance of theoretical calculations and machine learning, an advanced mapping approach for GDY based atomic catalysts is proposed. This strategy has predicted several novel electrocatalysts for achieving efficient HER for the first time, which supplies directional guidelines for the synthesis and modulation of potential catalysts through an effective theoretical approach.

Keyword Electrocatalysts

Mingzi Sun, Dr. Alan William Dougherty, Prof. Bolong Huang, Prof. Yuliang Li, Prof. Chun-Hua Yan*

Accelerating the atomic catalyst discovery by theoretical calculations-machine learning strategy

