

## Supplementary information

### Micro-kinetic Modelling of CO Reduction Reaction on Single Atom Catalysts Accelerated by Machine Learning

Qing-Meng Zhang <sup>a</sup>, Zhao-Yu Wang <sup>a</sup>, Hao Zhang <sup>a,\*</sup>, Xiao-Hong Liu <sup>b, c,\*</sup>, Wei Zhang <sup>b</sup>, and Liu-Bin Zhao <sup>a,\*</sup>

<sup>a</sup> Department of Chemistry, School of Chemistry and Chemical Engineering, Southwest University, Chongqing, 400715, China.

<sup>b</sup> Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, 400714, China.

<sup>c</sup> National University of Singapore (Chongqing) Research Institute, Chongqing 401123, China.

## Supplemental Note1

### 1.1 Computation of machine learning

The performance of feature set also directly affects the accuracy of the model. Too many features will also lead to redundancy problems, and the correlation between each feature of a qualified model should be independent. Therefore, Pearson correlation coefficient ( $p$ ) was used to evaluate the performance of the feature set, as shown in the following formula (1).

$$p = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}} \quad (1)$$

where  $X_i$  and  $Y_i$  represent different features. The Pearson correlation coefficient is obtained by dividing the covariance by the standard deviation of the two variables. This is a value between  $-1$  and  $1$ : when the linear relationship between two variables is strong, the value approaches  $1$  or  $-1$ , and if the correlation coefficient is equal to  $0$ , there is no linear relationship between them.

Two scoring criteria, namely coefficient of determination ( $R^2$ ) and root mean square error (RMSE), were used to reflect the prediction accuracy of the model, as shown in the following formula (2) and formula (3)

$$R^2 = 1 - \frac{\sum_i (Y_i - y_i)^2}{\sum_i (Y_i - \bar{Y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_i (Y_i - y_i)^2} \quad (3)$$

Where  $Y_i$  is the data calculated by DFT and  $y_i$  is the result predicted by machine learning.  $\bar{Y}$  is the average value of the DFT data. In general, the  $R^2$  and RMSE of the ideal model are close to  $1$  and  $0$  respectively.

All ML algorithms are implemented using scikit-learn, an open source code in the Python3 environment. In this work, four different supervised ML algorithms were selected, namely Gradient Boosting Regression (GBR), Random Forest Regression (RFR), Adaptive Boost Regression (ABR), and Gaussian Process Regression (GPR).

## 1.2 Computation of reorganization energies

The total reorganization energy consists of three parts, including internal reorganization energy ( $\lambda_{R/P}^i$ ), electron transfer solvent reorganization energy ( $\lambda_{R/P}^{o,ET}$ ) and proton transfer solvent reorganization energy ( $\lambda_{R/P}^{o,PT}$ ).

$$\lambda_{R/P} = \lambda_{R/P}^i + \lambda_{R/P}^{o,ET} + \lambda_{R/P}^{o,PT} \quad (4)$$

The inner reorganization energies defined as electronic energy difference along the general reaction coordinate from the equilibrium configuration of reactant to the equilibrium configuration of product.

$$\lambda_R^i = E_R(P_e) - E_R(R_e) \quad (5)$$

$$\lambda_P^i = E_P(P_e) - E_P(R_e) \quad (6)$$

Where  $E_R(R_e)$  and  $E_P(P_e)$  are electronic energy of the reactant and product in their equilibrium configurations, respectively.  $E_R(P_e)$  denotes single point energy of reactant under the equilibrium configuration of the product and  $E_P(R_e)$  denotes single point energy of product under the equilibrium configuration of the reactant. To keep the conservation of matter, one hydrogen atom is removed from the product in single point energy calculation of  $E_R(P_e)$ , while one hydrogen atom is added to the reactant in single point energy calculation of  $E_P(R_e)$ .

The outer (solvent) reorganization energies are obtained from polarization continuum electrostatic model.<sup>1,2</sup>

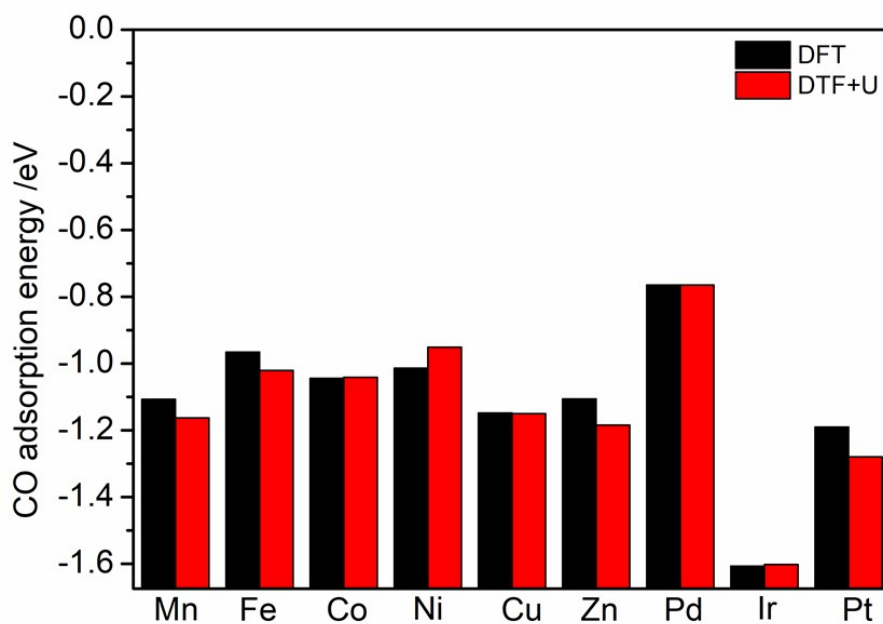
$$\lambda_{R/P}^{o,ET} = \frac{e_0^2}{4\pi\epsilon_0} \left( \frac{1}{\epsilon_{op}} - \frac{1}{\epsilon_S} \right) \frac{1}{2a} \quad (7)$$

$$\lambda_{R/P}^{o,PT} = \frac{1}{4\pi\epsilon_0} \left( \frac{\epsilon_S - 1}{2\epsilon_S + 1} - \frac{\epsilon_{op} - 1}{2\epsilon_{op} + 1} \right) \frac{(\mu_R - \mu_P)^2}{a^3} \quad (8)$$

Where “ $a$ ” is the radius of the sphere, which is determined from the beginning of the calculation of the volume of the optimized geometry of the reactants.  $\epsilon_{op}$  and  $\epsilon_S$  are optical and static dielectric constants.  $\mu_R$  and  $\mu_P$  are the dipole moments of the reactants

and products.

## Supplemental Note 2



**Fig. S1.** CO adsorption energy of  $C_3$  configuration of 9 transition metal by applying DFT+U (Hubbard  $U - J = 3$  eV), compared with DFT method.

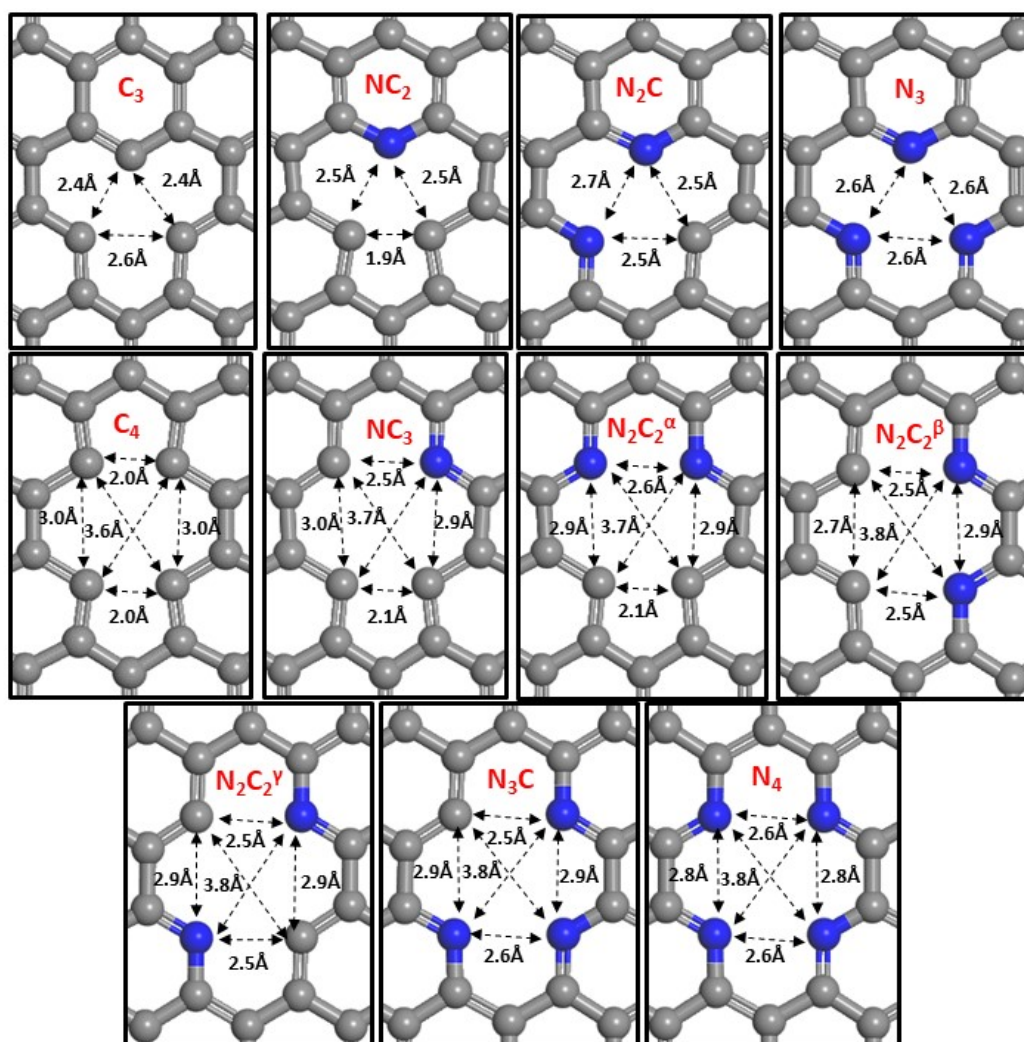
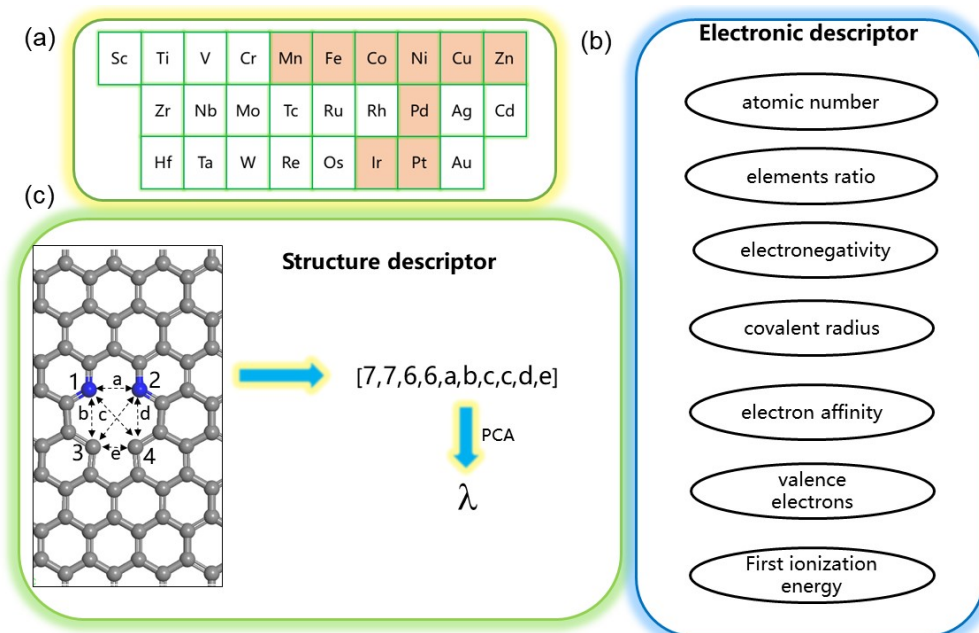
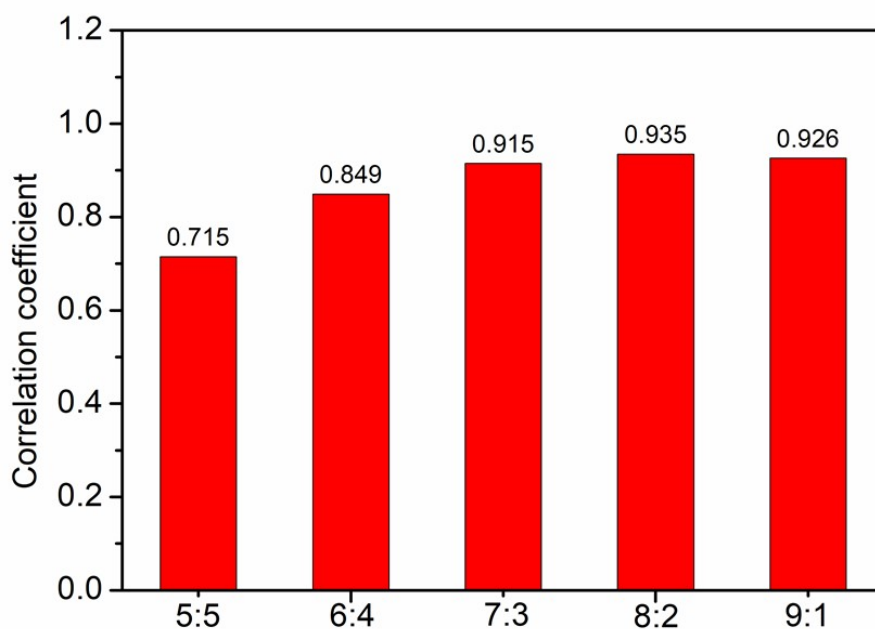


Fig. S2. The distances between coordination atoms in the 11 configurations



**Fig. S3.** (a) The figure shows all the transition metals involved, where the brown metal represents the data used in the DFT calculations for training the model. (b) Represents all selected electronic descriptors. (c) The structure descriptor of the design



**Fig. S4** Correlation coefficient ( $R^2$ ) in machine learning models with different training set and test set ratios from 5:5 to 9:1.

**Table S1.** Initial thermodynamic dataset using DFT calculations

Structure	$E_s$	$E_{ad}(\text{CO})$	$\Delta G(\text{COH})$	$\Delta G(\text{CHO})$	$\Delta G(^*\text{H})$
Mn-C <sub>3</sub>	5.74	-1.16	1.82	0.55	0.47
Mn-C <sub>4</sub>	6.69	-1.67	1.17	0.94	0.26
Mn-N <sub>2</sub> C	9.58	-3.87	1.32	0.56	-2.27
Mn-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	7.27	-2.47	0.92	0.63	-0.45
Mn-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	5.85	-2.14	0.80	0.54	-0.16
Mn-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	5.42	-2.01	1.07	0.34	0.13
Mn-N <sub>3</sub>	8.49	-2.00	1.53	0.78	-0.04
Mn-N <sub>3</sub> C	6.58	-2.56	0.97	0.76	-0.30
Mn-N <sub>4</sub>	7.40	-2.54	2.48	0.96	-0.41
Mn-NC <sub>2</sub>	7.90	-2.10	1.30	0.42	-0.59
Mn-NC <sub>3</sub>	5.92	-1.91	1.31	0.50	0.15
Fe-C <sub>3</sub>	3.53	-1.02	1.82	0.50	0.39
Fe-C <sub>4</sub>	4.94	-1.74	1.38	0.75	0.10
Fe-N <sub>2</sub> C	5.70	-1.88	1.90	0.61	0.00
Fe-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	4.29	-1.68	1.26	0.73	0.29
Fe-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	3.99	-2.46	1.12	0.67	-0.69
Fe-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	3.45	-2.03	1.02	0.65	-0.65
Fe-N <sub>3</sub>	6.85	-1.86	1.19	0.56	0.08
Fe-N <sub>3</sub> C	4.39	-2.50	1.26	\	-0.55
Fe-N <sub>4</sub>	4.97	-2.31	1.90	0.80	-0.40
Fe-NC <sub>2</sub>	5.27	-1.15	1.32	0.10	-0.05
Fe-NC <sub>3</sub>	4.25	-1.90	0.97	0.42	-0.35
Co-C <sub>3</sub>	0.59	-1.04	2.04	0.30	0.14
Co-C <sub>4</sub>	1.70	-1.54	1.34	0.52	0.13
Co-N <sub>2</sub> C	2.76	-1.20	1.19	0.66	0.12
Co-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	1.13	-1.72	1.41	0.62	-0.06
Co-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	-0.04	-1.79	1.48	0.66	-0.05
Co-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	-0.19	-1.77	1.47	0.59	0.13
Co-N <sub>3</sub>	3.73	-1.99	1.44	0.68	0.01
Co-N <sub>3</sub> C	0.29	-1.56	1.97	0.48	0.08
Co-N <sub>4</sub>	1.04	-0.84	1.82	-0.11	0.20
Co-NC <sub>2</sub>	2.30	-1.23	1.35	0.12	-0.06
Co-NC <sub>3</sub>	0.61	-1.71	1.31	0.59	-0.31
Ni-C <sub>3</sub>	0.82	-0.95	1.84	0.60	0.48
Ni-C <sub>4</sub>	0.94	-1.03	1.38	0.23	-0.70
Ni-N <sub>2</sub> C	2.71	-1.83	1.56	1.10	-0.16
Ni-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	-0.12	-0.68	2.08	0.53	0.22
Ni-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	-1.18	-0.68	1.93	0.30	-0.07
Ni-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	-1.29	-0.83	2.19	0.46	-0.19
Ni-N <sub>3</sub>	3.40	-2.03	1.16	0.72	-0.03
Ni-N <sub>3</sub> C	-0.70	-0.19	1.86	0.06	0.08
Ni-N <sub>4</sub>	0.46	-0.23	2.53	0.77	1.51



Ni-NC <sub>2</sub>	2.25	-1.22	1.56	0.27	0.01
Ni-NC <sub>3</sub>	-0.40	-0.97	1.53	\	-0.72
Cu-C <sub>3</sub>	3.38	-1.15	1.67	0.56	0.22
Cu-C <sub>4</sub>	1.68	-0.24	\	\	0.24
Cu-N <sub>2</sub> C	3.87	-1.59	1.72	0.92	0.21
Cu-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	1.27	-0.03	2.23	0.73	1.47
Cu-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	0.11	-0.03	\	0.81	1.64
Cu-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	0.11	-0.04	\	0.47	1.28
Cu-N <sub>3</sub>	4.22	-2.29	2.67	1.48	0.28
Cu-N <sub>3</sub> C	1.08	-0.04	\	-0.35	-0.03
Cu-N <sub>4</sub>	2.42	-0.05	\	0.67	1.51
Cu-NC <sub>2</sub>	4.48	-1.81	1.97	0.80	-0.14
Cu-NC <sub>3</sub>	0.41	-0.08	2.02	0.63	0.05
Zn-C <sub>3</sub>	0.93	-1.18	2.17	0.55	-0.23
Zn-C <sub>4</sub>	-0.81	-0.11	\	0.94	1.55
Zn-N <sub>2</sub> C	0.21	-0.84	2.60	0.59	0.17
Zn-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	-0.58	-0.03	2.28	0.73	1.17
Zn-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	-1.70	-0.04	2.38	0.74	1.27
Zn-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	-2.23	-0.06	\	0.75	1.44
Zn-N <sub>3</sub>	0.79	-1.00	2.31	-0.28	-0.92
Zn-N <sub>3</sub> C	-1.39	-0.24	\	0.58	0.85
Zn-N <sub>4</sub>	-0.98	-0.20	2.28	0.46	0.73
Zn-NC <sub>2</sub>	1.14	-0.97	2.41	0.66	-0.03
Zn-NC <sub>3</sub>	-1.36	-0.05	\	0.36	1.04
Pd-C <sub>3</sub>	-0.20	-0.76	2.14	0.54	0.51
Pd-C <sub>4</sub>	0.36	-0.84	1.45	0.15	-0.82
Pd-N <sub>2</sub> C	1.83	-1.66	1.83	1.46	0.05
Pd-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	-0.45	-0.45	1.95	0.47	0.84
Pd-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	-1.57	-0.42	1.78	0.17	0.60
Pd-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	-1.63	-0.53	2.05	0.38	-0.40
Pd-N <sub>3</sub>	3.32	-2.26	0.96	0.45	-0.39
Pd-N <sub>3</sub> C	-1.04	-0.03	0.76	0.22	0.10
Pd-N <sub>4</sub>	-0.08	-0.04	1.44	1.00	1.88
Pd-NC <sub>2</sub>	1.65	-1.26	1.67	0.80	-0.31
Pd-NC <sub>3</sub>	-0.77	-0.70	1.40	-0.01	-0.97
Pt-C <sub>3</sub>	1.56	-1.28	1.82	0.57	-0.05
Pt-C <sub>4</sub>	1.29	-0.92	1.15	0.20	-0.24
Pt-N <sub>2</sub> C	4.50	-2.70	1.50	1.69	-0.94
Pt-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	0.96	-0.62	1.86	0.36	0.59
Pt-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	-0.20	-0.58	1.60	0.11	0.37
Pt-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	-0.20	-0.76	1.86	0.23	-0.11
Pt-N <sub>3</sub>	6.35	-3.48	0.62	0.72	-1.32
Pt-N <sub>3</sub> C	0.53	-0.03	1.61	-0.23	0.24
Pt-N <sub>4</sub>	1.58	-0.04	2.72	0.77	1.53

Pt-NC <sub>2</sub>	3.73	-1.83	1.28	0.04	-0.99
Pt-NC <sub>3</sub>	0.39	-0.95	1.28	0.13	0.14
Ir-C <sub>3</sub>	2.79	-1.60	1.63	0.27	-0.64
Ir-C <sub>4</sub>	3.29	-1.83	1.22	0.59	-0.41
Ir-N <sub>2</sub> C	6.23	-2.76	1.12	0.54	-1.18
Ir-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	3.86	-2.28	1.02	0.33	-0.80
Ir-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	2.53	-2.30	1.20	0.64	-0.57
Ir-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	2.14	-1.98	1.01	0.60	-0.37
Ir-N <sub>3</sub>	8.22	-2.91	0.65	0.40	-1.34
Ir-N <sub>3</sub> C	3.20	-2.07	1.67	0.59	-0.40
Ir-N <sub>4</sub>	4.18	-1.27	1.57	-0.21	-0.33
Ir-NC <sub>2</sub>	4.92	-2.40	1.97	0.75	-0.77
Ir-NC <sub>3</sub>	2.52	-2.02	1.18	0.66	-0.25

---

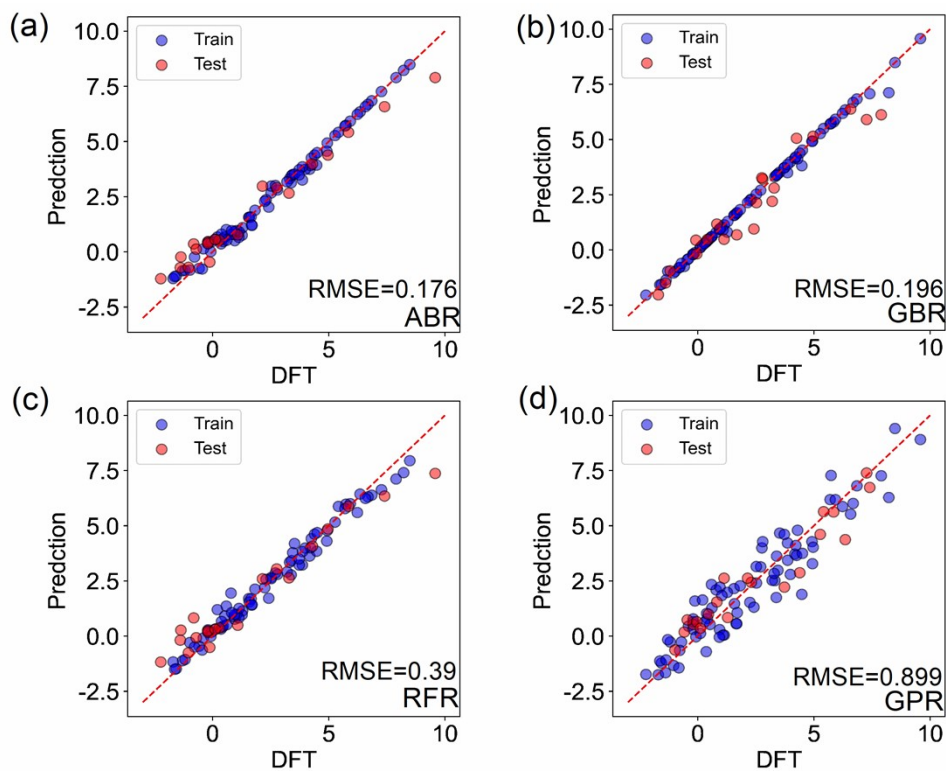
**Table S2.** Initial kinetic dataset using DFT calculations

Structure	$\lambda_R(\text{CHO})$	$\lambda_P(\text{CHO})$	$\lambda_R(\text{COH})$	$\lambda_P(\text{COH})$
Mn-C <sub>3</sub>	2.39	2.29	2.12	1.90
Mn-C <sub>4</sub>	2.92	1.68	2.11	1.94
Mn-N <sub>2</sub> C	2.37	1.81	2.23	2.08
Mn-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	3.42	2.29	1.93	1.82
Mn-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	3.27	2.08	1.82	2.15
Mn-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	2.72	2.13	2.01	1.89
Mn-N <sub>3</sub>	2.50	1.81	2.12	1.95
Mn-N <sub>3</sub> C	3.65	2.40	2.06	1.89
Mn-N <sub>4</sub>	3.55	1.65	3.07	\
Mn-NC <sub>2</sub>	2.25	1.87	1.78	2.56
Mn-NC <sub>3</sub>	3.38	2.37	2.37	2.12
Fe-C <sub>3</sub>	1.87	2.05	1.98	2.49
Fe-C <sub>4</sub>	3.05	2.01	2.10	1.87
Fe-N <sub>2</sub> C	2.84	2.07	2.67	1.92
Fe-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	3.04	2.86	2.23	1.93
Fe-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	3.02	2.36	1.84	1.87
Fe-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	3.04	1.80	1.89	2.06
Fe-N <sub>3</sub>	2.45	8.18	1.83	\
Fe-N <sub>3</sub> C	1.50	\	2.05	1.82
Fe-N <sub>4</sub>	2.46	2.25	2.05	1.79
Fe-NC <sub>2</sub>	1.69	1.84	2.16	2.34
Fe-NC <sub>3</sub>	2.67	2.01	1.88	1.94
Co-C <sub>3</sub>	0.55	2.33	2.21	1.66
Co-C <sub>4</sub>	1.51	2.32	2.33	2.05
Co-N <sub>2</sub> C	0.95	0.86	1.62	2.37
Co-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	1.01	2.27	2.38	2.02
Co-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	1.49	2.95	2.44	1.84
Co-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	0.87	2.27	2.09	1.74
Co-N <sub>3</sub>	1.41	1.39	2.16	2.45
Co-N <sub>3</sub> C	0.96	2.79	2.27	1.85
Co-N <sub>4</sub>	0.50	3.27	2.31	1.82
Co-NC <sub>2</sub>	0.35	2.52	1.91	2.30
Co-NC <sub>3</sub>	1.65	3.09	2.26	1.96
Ni-C <sub>3</sub>	2.59	2.15	2.29	\
Ni-C <sub>4</sub>	2.54	2.26	2.34	1.90
Ni-N <sub>2</sub> C	3.09	1.48	2.38	2.31
Ni-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	1.84	2.85	2.31	2.00
Ni-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	2.13	4.27	2.15	2.02
Ni-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	1.81	3.05	2.27	\
Ni-N <sub>3</sub>	2.74	2.13	2.17	2.10
Ni-N <sub>3</sub> C	1.66	3.38	2.03	1.99
Ni-N <sub>4</sub>	1.75	1.99	2.78	2.67

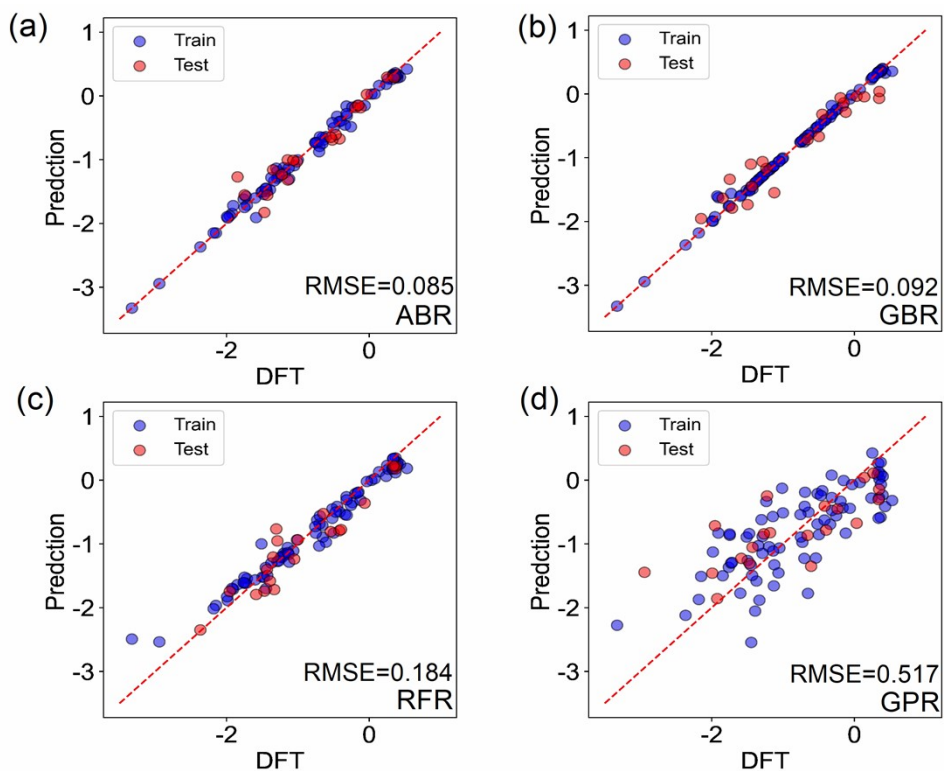
Ni-NC <sub>2</sub>	1.91	2.43	1.98	2.02
Ni-NC <sub>3</sub>	\	\	2.36	1.98
Cu-C <sub>3</sub>	\	\	2.18	2.05
Cu-C <sub>4</sub>	\	\	\	\
Cu-N <sub>2</sub> C	2.58	1.57	2.16	1.83
Cu-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	1.76	\	2.36	\
Cu-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	1.87	1.94	\	\
Cu-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	1.80	2.39	\	\
Cu-N <sub>3</sub>	2.37	1.95	1.91	1.70
Cu-N <sub>3</sub> C	2.32	3.36	\	\
Cu-N <sub>4</sub>	1.59	2.40	\	\
Cu-NC <sub>2</sub>	2.29	2.13	2.31	2.12
Cu-NC <sub>3</sub>	\	\	2.29	2.10
Zn-C <sub>3</sub>	2.48	2.53	2.15	1.80
Zn-C <sub>4</sub>	1.85	2.48	\	1.36
Zn-N <sub>2</sub> C	2.16	2.68	2.43	2.13
Zn-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	\	\	2.47	2.33
Zn-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	\	\	2.60	2.20
Zn-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	1.64	1.95	\	\
Zn-N <sub>3</sub>	1.82	3.14	2.28	2.33
Zn-N <sub>3</sub> C	1.90	2.64	\	\
Zn-N <sub>4</sub>	1.84	2.79	2.14	2.35
Zn-NC <sub>2</sub>	2.51	2.70	2.52	\
Zn-NC <sub>3</sub>	1.67	\	\	\
Pd-C <sub>3</sub>	1.99	2.51	2.66	2.35
Pd-C <sub>4</sub>	2.30	3.79	2.53	2.45
Pd-N <sub>2</sub> C	3.35	1.74	2.55	2.45
Pd-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	1.74	3.05	2.26	2.23
Pd-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	2.00	6.67	2.27	2.15
Pd-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	1.73	3.03	2.18	1.89
Pd-N <sub>3</sub>	2.61	1.64	2.07	1.87
Pd-N <sub>3</sub> C	1.65	2.88	\	\
Pd-N <sub>4</sub>	1.81	2.09	\	\
Pd-NC <sub>2</sub>	1.87	2.42	2.40	2.39
Pd-NC <sub>3</sub>	2.29	3.51	2.17	1.79
Pt-C <sub>3</sub>	2.20	2.69	2.75	2.28
Pt-C <sub>4</sub>	2.45	2.91	2.28	2.17
Pt-N <sub>2</sub> C	4.01	1.66	2.62	2.41
Pt-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	1.85	3.42	2.41	2.24
Pt-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	2.04	2.45	2.40	2.30
Pt-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	1.84	3.60	2.58	2.13
Pt-N <sub>3</sub>	3.36	1.56	2.02	1.85
Pt-N <sub>3</sub> C	1.62	2.82	2.08	\
Pt-N <sub>4</sub>	1.90	2.35	2.36	2.08

Pt-NC <sub>2</sub>	2.47	3.80	2.33	2.33
Pt-NC <sub>3</sub>	2.44	3.90	2.23	1.77
Ir-C <sub>3</sub>	2.13	2.82	2.08	1.92
Ir-C <sub>4</sub>	2.74	2.15	2.19	1.99
Ir-N <sub>2</sub> C	3.19	2.01	2.27	2.46
Ir-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	2.99	3.47	2.34	2.15
Ir-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	3.26	3.04	2.41	2.00
Ir-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	2.79	2.36	2.14	2.06
Ir-N <sub>3</sub>	2.84	2.08	1.78	2.60
Ir-N <sub>3</sub> C	2.56	3.30	2.24	1.94
Ir-N <sub>4</sub>	2.05	3.97	2.60	2.42
Ir-NC <sub>2</sub>	2.70	3.19	2.91	2.14
Ir-NC <sub>3</sub>	3.32	3.00	2.48	2.22

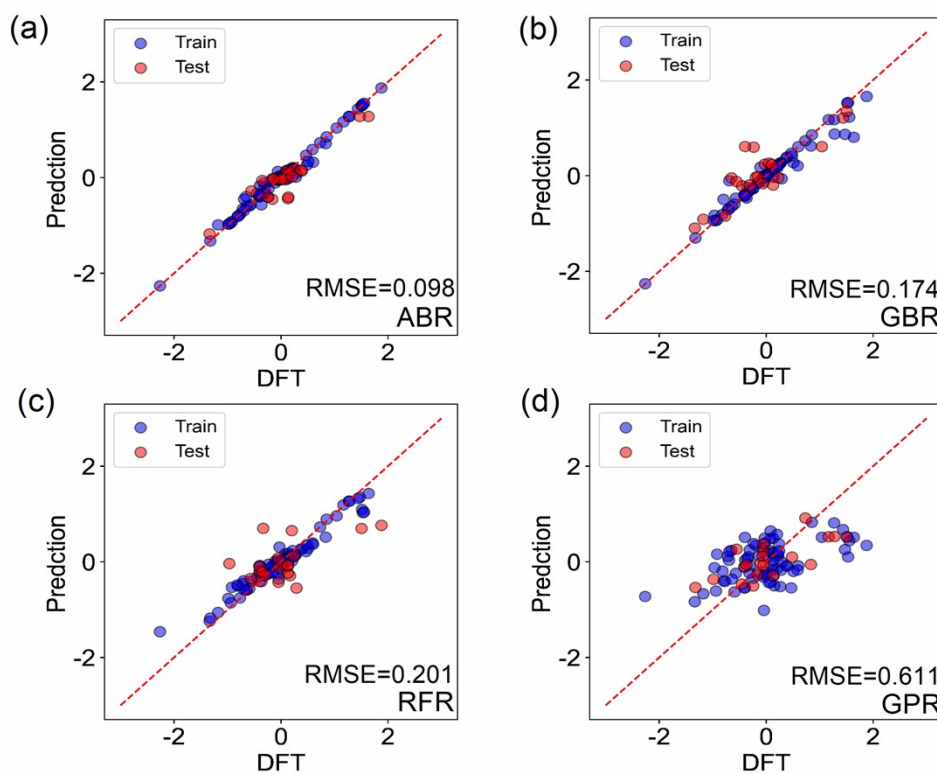
---



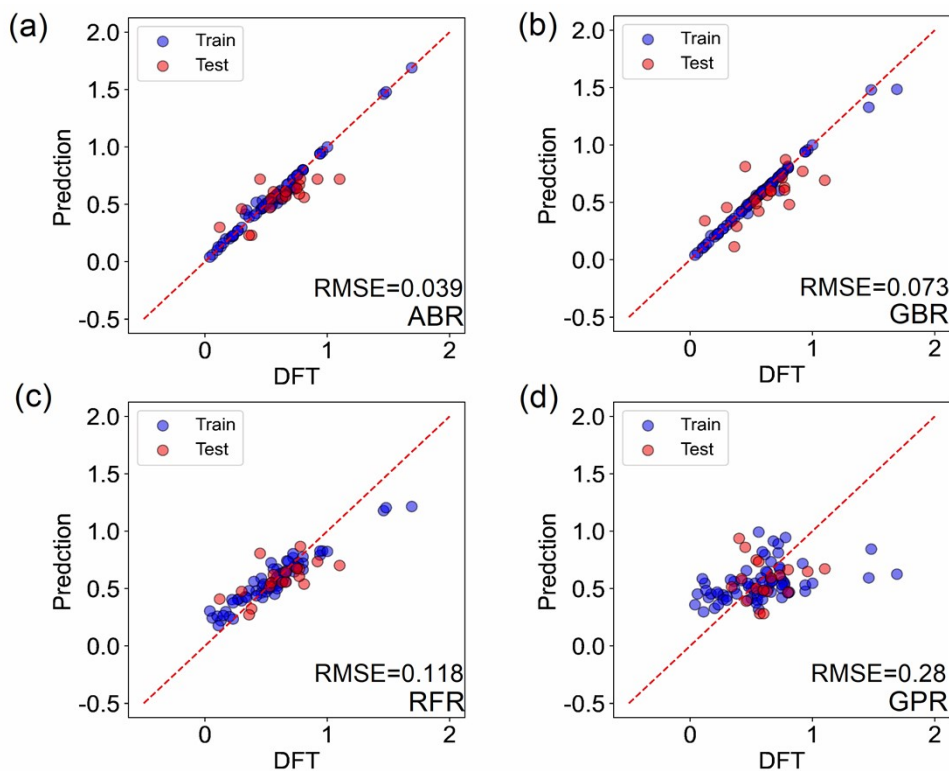
**Fig. S5.** Training and prediction of  $E_s$  by four machine learning algorithms



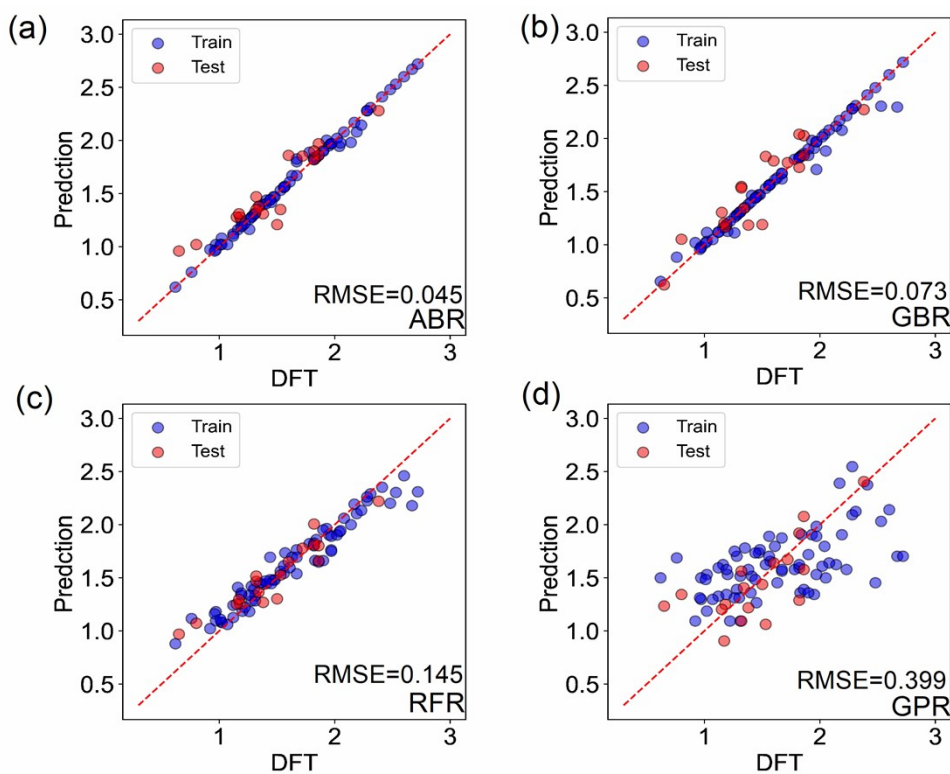
**Fig. S6.** Training and prediction of  $\Delta G^*_{CO}$  by four machine learning algorithms



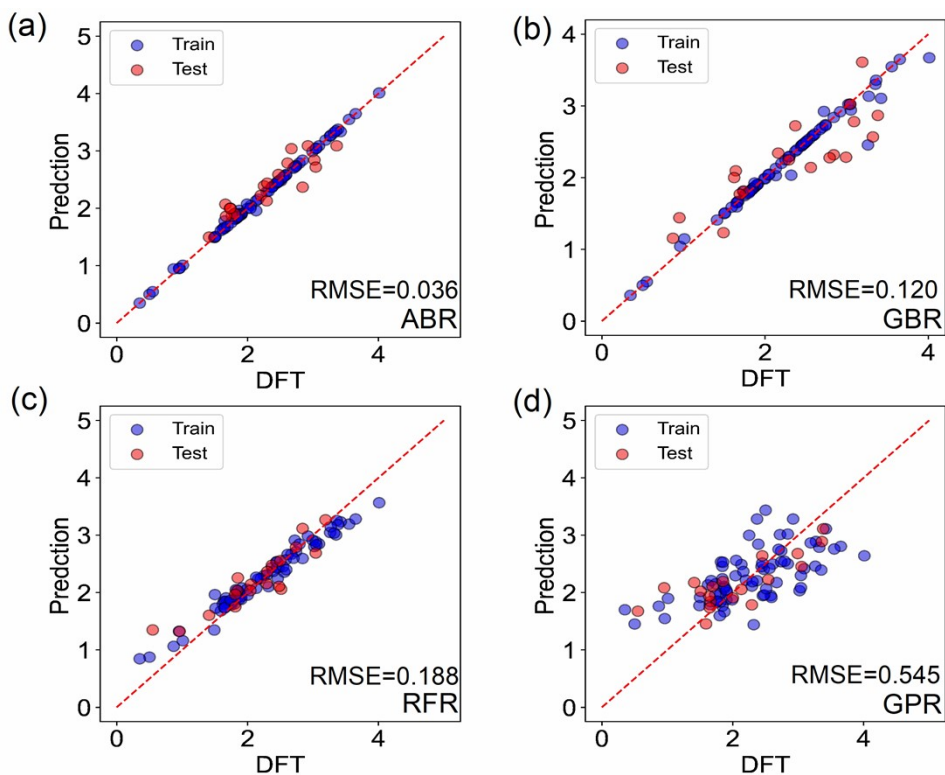
**Fig. S7.** Training and prediction of  $E_{ad}(*H)$  by four machine learning algorithms



**Fig. S8.** Training and prediction of  $\Delta G_{CHO}$  by four machine learning algorithms

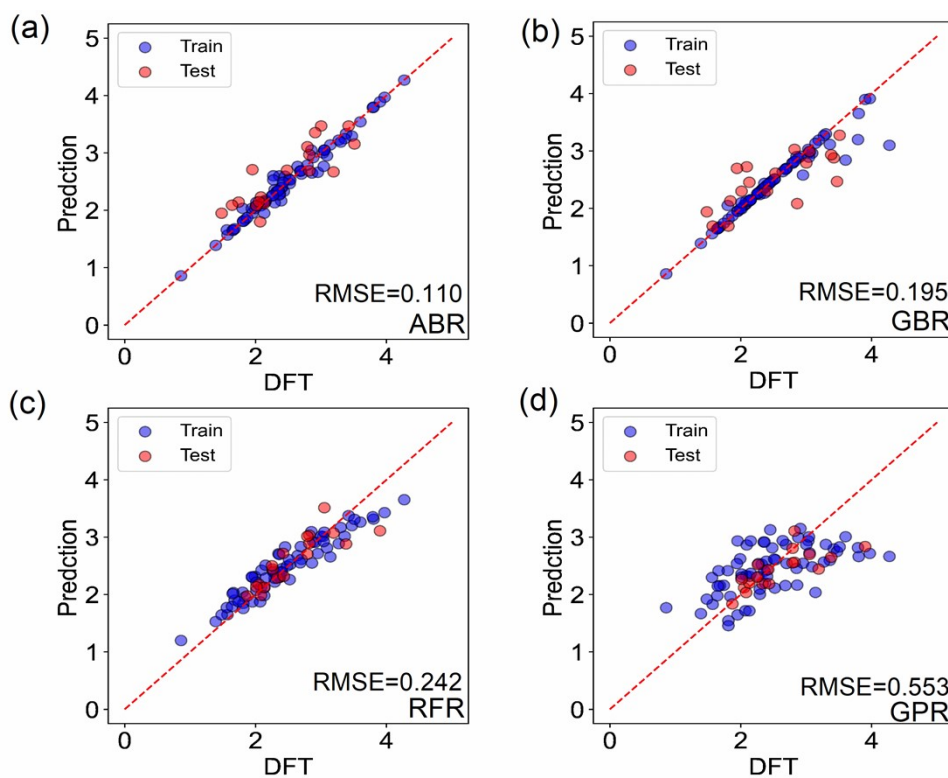


**Fig. S9.** Training and prediction of  $\Delta G_{\text{COH}}$  by four machine learning algorithms

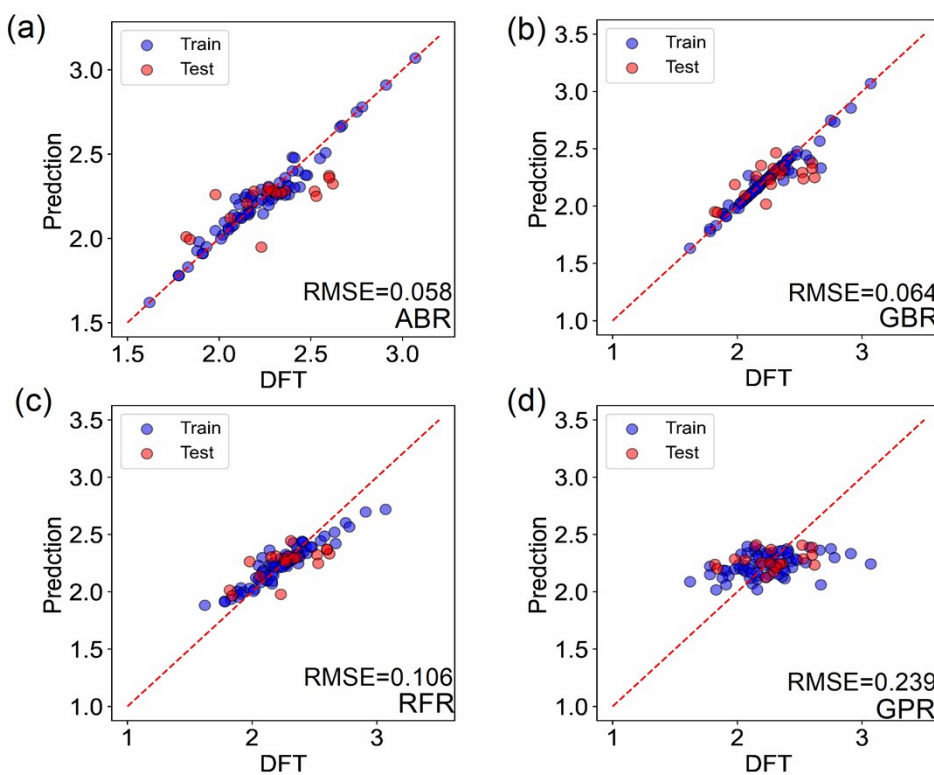


**Fig. S10.** Training and prediction of  $\lambda_{\text{R}}(\text{CHO})$  by four machine learning algorithms

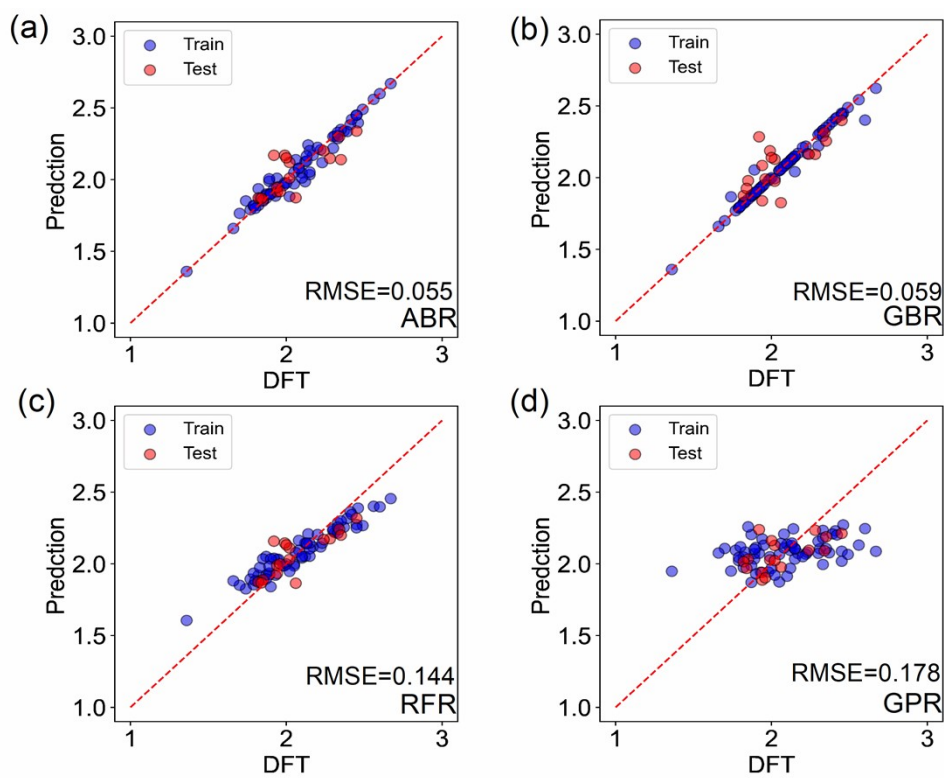




**Fig. S11.** Training and prediction of  $\lambda_p(\text{CHO})$  by four machine learning algorithms



**Fig. S12.** Training and prediction of  $\lambda_R(\text{COH})$  by four machine learning algorithms

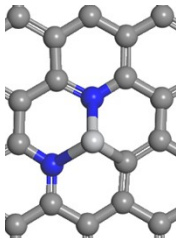
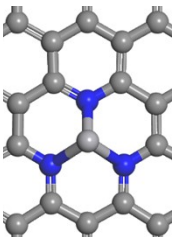
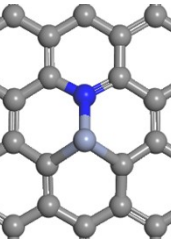
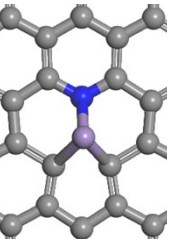
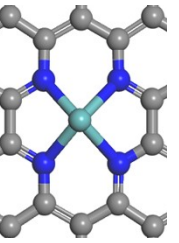
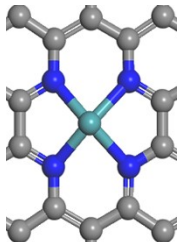
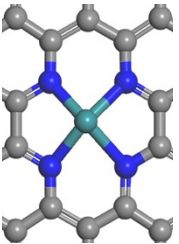
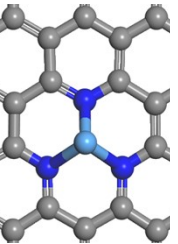
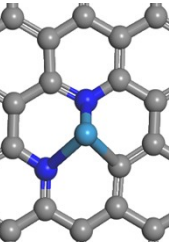
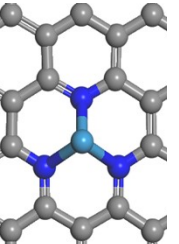
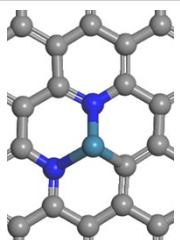
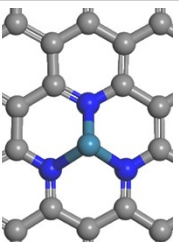
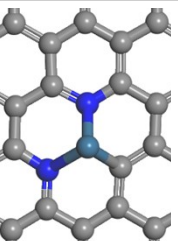
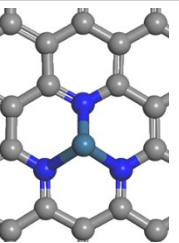
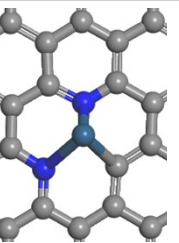
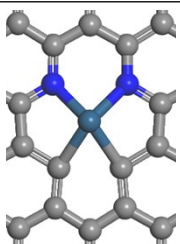
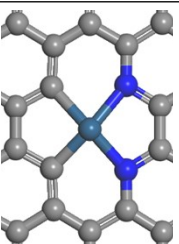
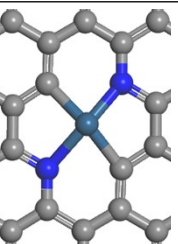
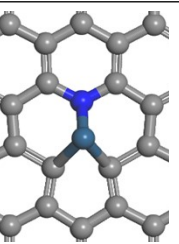
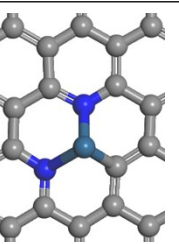
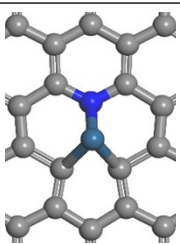
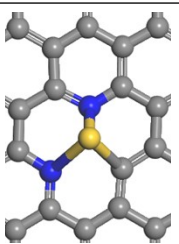
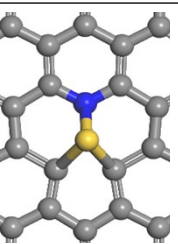
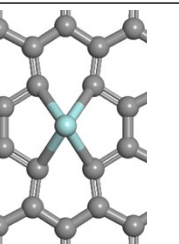
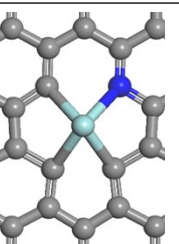


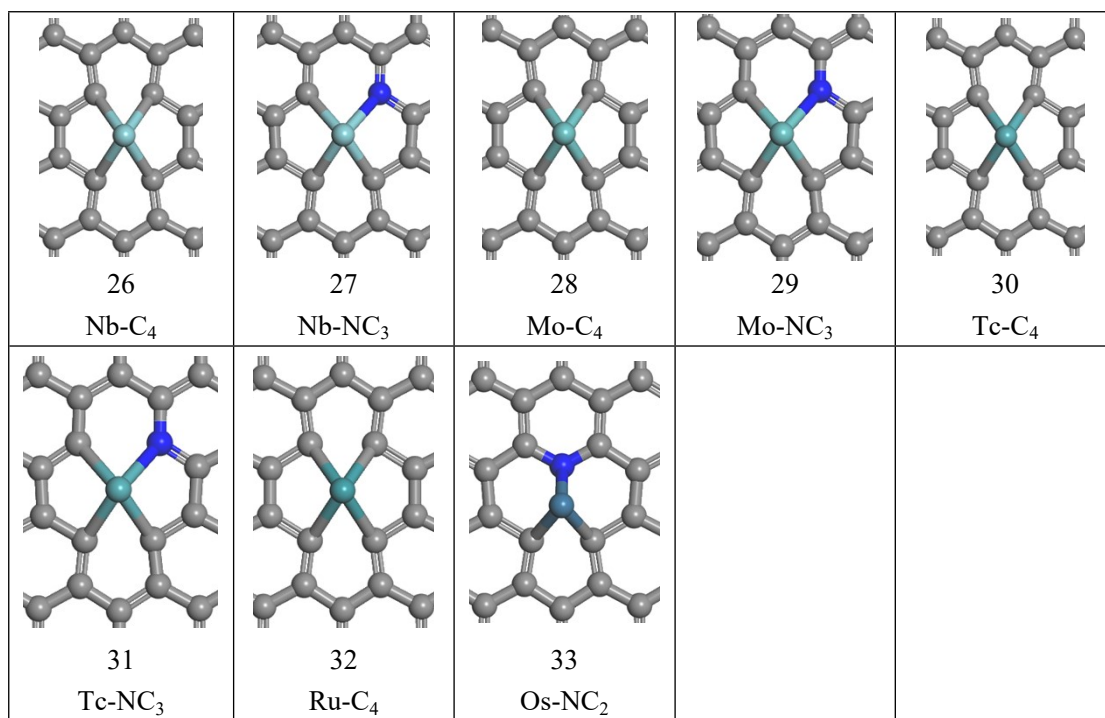
**Fig. S13.** Training and prediction of  $\lambda_p(\text{COH})$  by four machine learning algorithms

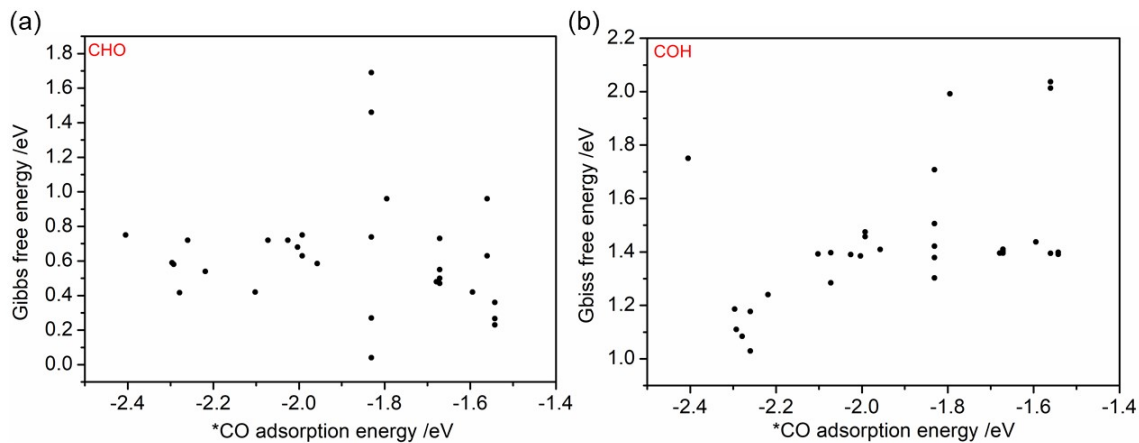
**Table S3.** The hyperparameters of four model algorithms.

<b>Algorithm</b>	<b>Hyperparameter</b>
Adaptive Boost Regression	base_estimator = DecisionTreeRegressor(max_depth=7), n_estimators=5000
Gradient Boosting Regression	learning_rate=0.02, loss='absolute_error', max_depth=4, min_samples_split=3, n_estimators=5000
Random Forest Regression	n_estimators = 1000, random_state = 28
Gaussian Process Regression	kernel = DotProduct(), random_state=1

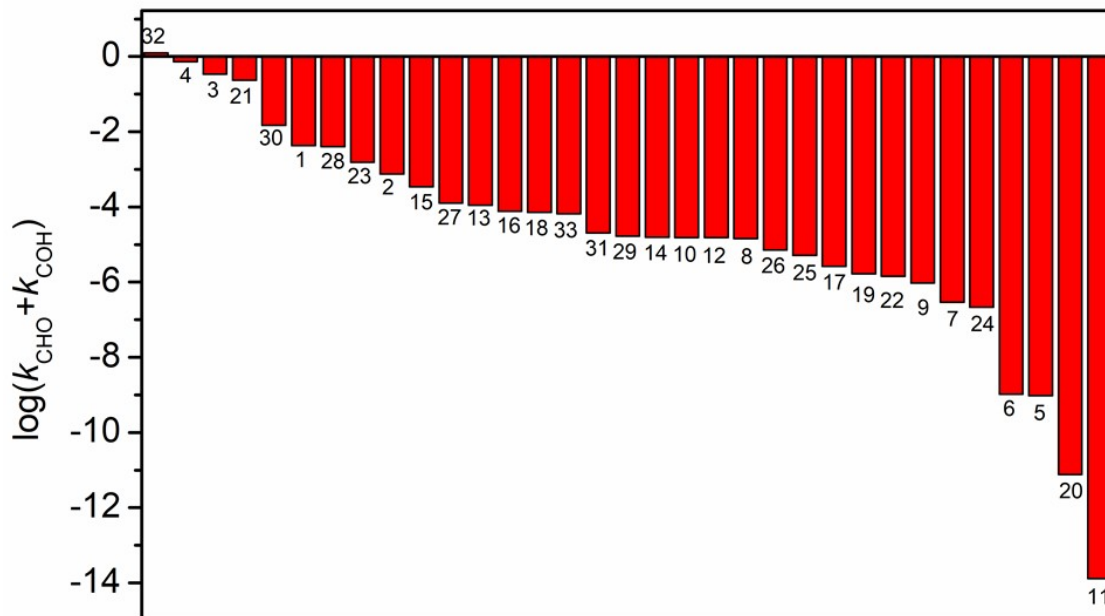
**Table S4.** After two rounds of screening, the 34 catalyst configurations were numbered

 1 Ti-N <sub>2</sub> C	 2 V-N <sub>3</sub>	 3 Cr-NC <sub>2</sub>	 4 Mn-NC <sub>2</sub>	 5 Mo-N <sub>4</sub>
 6 Tc-N <sub>4</sub>	 7 Ru-N <sub>4</sub>	 8 Ta-N <sub>3</sub>	 9 W-N <sub>2</sub> C	 10 W-N <sub>3</sub>
 11 Re-N <sub>2</sub> C	 12 Re-N <sub>3</sub>	 13 Os-N <sub>2</sub> C	 14 Os-N <sub>3</sub>	 15 Ir-N <sub>2</sub> C
 16 Ir-N <sub>2</sub> C <sub>2</sub> <sup>α</sup>	 17 Ir-N <sub>2</sub> C <sub>2</sub> <sup>β</sup>	 18 Ir-N <sub>2</sub> C <sub>2</sub> <sup>γ</sup>	 19 Ir-NC <sub>2</sub>	 20 Pt-N <sub>2</sub> C
 21 Pt-NC <sub>2</sub>	 22 Au-N <sub>2</sub> C	 23 Au-NC <sub>2</sub>	 24 Zr-C <sub>4</sub>	 25 Zr-NC <sub>3</sub>





**Fig. S14.** The relation of the remained 33 catalysts between (a)  $E_{\text{ad}}(\text{CO})$  and  $\Delta G_{\text{CHO}}$  (b)  $E_{\text{ad}}(\text{CO})$  and  $\Delta G_{\text{COH}}$ .



**Fig. S15.** Activity ranking of 33 catalyst candidates.

**Table S5.** Thermodynamic data on the remaining catalysts after two rounds of screening.

SACs	$E_s$	$\Delta G^*_{CO}$	$\Delta G^*_{H}$	$\Delta G_{CHO}$	$\Delta G_{COH}$
Ti-N <sub>2</sub> C	6.85	-1.67	-0.58	0.63	1.46
V-N <sub>3</sub>	7.90	-1.60	-0.59	0.68	1.38
Cr-NC <sub>2</sub>	7.90	-1.40	-0.59	0.42	1.44
Mn-NC <sub>2</sub>	7.93	-1.60	-0.57	0.42	1.39
Mo-N <sub>4</sub>	6.58	-1.75	1.88	0.96	2.01
Tc-N <sub>4</sub>	6.58	-1.75	1.51	0.96	1.99
Ru-N <sub>4</sub>	5.42	-1.75	1.51	0.96	2.04
Ta-N <sub>3</sub>	4.50	-1.47	-0.57	0.72	1.39
W-N <sub>2</sub> C	4.48	-1.84	-0.55	0.75	1.47
W-N <sub>3</sub>	6.23	-1.81	-0.64	0.72	1.18
Re-N <sub>2</sub> C	4.50	-1.30	-0.64	1.46	1.51
Re-N <sub>3</sub>	6.23	-1.78	-0.92	0.72	1.40
Os-N <sub>2</sub> C	4.92	-2.02	-0.94	0.59	1.41
Os-N <sub>3</sub>	8.22	-2.18	-0.98	0.72	1.03
Ir-N <sub>2</sub> C	6.23	-2.16	-0.97	0.54	1.24
Ir-N <sub>2</sub> C <sub>2</sub> <sup>a</sup>	3.86	-1.58	-0.80	0.42	1.08
Ir-N <sub>2</sub> C <sub>2</sub> <sup>b</sup>	2.53	-1.56	-0.57	0.59	1.19
Ir-N <sub>2</sub> C <sub>2</sub> <sup>c</sup>	2.53	-1.56	-0.57	0.58	1.11
Ir-NC <sub>2</sub>	4.92	-1.27	-0.73	0.75	1.75
Pt-N <sub>2</sub> C	4.52	-2.15	-0.91	1.69	1.30
Pt-NC <sub>2</sub>	3.74	-1.27	-0.97	0.05	1.42
Au-N <sub>2</sub> C	4.50	-1.89	-0.94	0.72	1.28
Au-NC <sub>2</sub>	3.47	-1.28	-0.80	0.27	1.38
Zr-C <sub>4</sub>	6.69	-1.12	-0.54	0.73	1.40
Zr-NC <sub>3</sub>	5.92	-1.14	-0.61	0.55	1.41
Nb-C <sub>4</sub>	6.69	-1.12	-0.59	0.63	1.39
Nb-NC <sub>3</sub>	5.92	-1.13	-0.70	0.47	1.41
Mo-C <sub>4</sub>	6.58	-1.12	-0.70	0.27	1.40
Mo-NC <sub>3</sub>	5.92	-1.13	-0.72	0.50	1.40
Tc-C <sub>4</sub>	6.69	-1.12	-0.70	0.36	1.39
Tc-NC <sub>3</sub>	5.92	-1.14	-0.70	0.48	1.40
Ru-C <sub>4</sub>	4.94	-1.09	-0.70	0.23	1.40
Os-NC <sub>2</sub>	4.24	-1.06	-0.60	0.74	1.71

**Table S6.** Kinetic data on the remaining catalysts after two rounds of screening.

SACs	$\lambda_R(\text{CHO})$	$\lambda_P(\text{CHO})$	$\lambda_R(\text{COH})$	$\lambda_P(\text{COH})$
Ti-N <sub>2</sub> C	2.37	1.82	2.23	2.08
V-N <sub>3</sub>	2.46	1.81	2.12	2.10
Cr-NC <sub>2</sub>	2.39	2.01	1.91	2.51
Mn-NC <sub>2</sub>	2.39	1.84	1.78	2.56
Mo-N <sub>4</sub>	3.55	2.11	2.78	2.42
Tc-N <sub>4</sub>	3.55	2.09	2.78	2.42
Ru-N <sub>4</sub>	2.32	2.09	2.78	2.67
Ta-N <sub>3</sub>	2.84	2.10	2.21	2.18
W-N <sub>2</sub> C	3.32	2.07	2.35	2.34
W-N <sub>3</sub>	2.84	2.08	2.07	1.87
Re-N <sub>2</sub> C	3.08	2.09	2.40	2.33
Re-N <sub>3</sub>	2.84	2.08	2.08	1.89
Os-N <sub>2</sub> C	3.19	2.08	2.35	2.34
Os-N <sub>3</sub>	2.84	2.08	2.02	2.33
Ir-N <sub>2</sub> C	3.19	2.08	2.30	2.40
Ir-N <sub>2</sub> C <sub>2</sub> <sup><math>\alpha</math></sup>	2.99	3.30	2.27	2.21
Ir-N <sub>2</sub> C <sub>2</sub> <sup><math>\beta</math></sup>	3.26	2.80	2.30	2.15
Ir-N <sub>2</sub> C <sub>2</sub> <sup><math>\gamma</math></sup>	2.82	2.61	2.25	2.14
Ir-NC <sub>2</sub>	2.70	2.68	2.91	2.25
Pt-N <sub>2</sub> C	4.05	1.62	2.32	2.38
Pt-NC <sub>2</sub>	2.59	3.83	2.31	2.30
Au-N <sub>2</sub> C	3.32	2.15	2.31	2.27
Au-NC <sub>2</sub>	2.70	3.80	2.33	2.30
Zr-C <sub>4</sub>	3.05	2.85	2.26	2.42
Zr-NC <sub>3</sub>	3.27	2.85	2.31	2.25
Nb-C <sub>4</sub>	3.05	2.58	2.28	2.39
Nb-NC <sub>3</sub>	3.27	2.56	2.31	2.39
Mo-C <sub>4</sub>	3.05	3.05	2.28	2.45
Mo-NC <sub>3</sub>	3.27	2.85	2.28	1.90
Tc-C <sub>4</sub>	3.05	2.32	2.26	2.45
Tc-NC <sub>3</sub>	3.32	2.85	2.28	1.99
Ru-C <sub>4</sub>	2.72	2.32	2.26	1.95
Os-NC <sub>2</sub>	2.17	2.63	2.41	2.31



**Table S7.** Comparison of ML prediction and DFT calculation on Mn-NC<sub>2</sub>

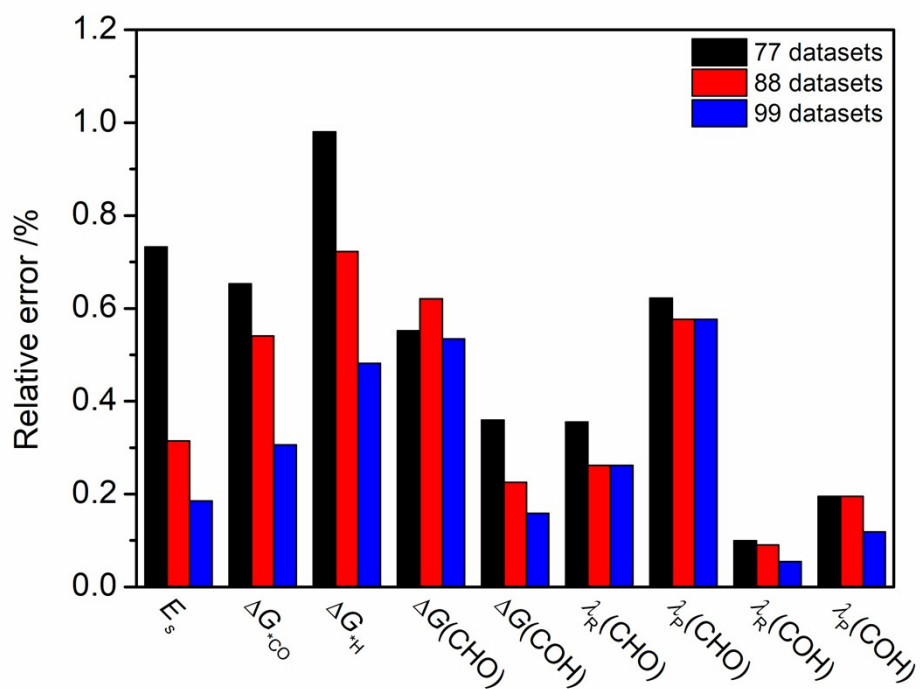
Properties	ML	DFT
$E_s$	7.93	7.90
$\Delta G_{*CO}$	-1.60	-1.59
$\Delta G(*H)$	-0.57	-0.59
$\Delta G(CHO)$	0.42	0.42
$\Delta G(COH)$	1.39	1.30
$\lambda_R(CHO)$	2.39	2.25
$\lambda_P(CHO)$	1.84	1.87
$\lambda_R(COH)$	1.78	1.78
$\lambda_P(COH)$	2.56	2.56

**Table S8.** Comparison of ML prediction and DFT calculation on Pt-NC<sub>2</sub>

Properties	ML	DFT
$E_s$	3.74	3.73
$\Delta G_{*CO}$	-1.27	-1.26
$\Delta G(*H)$	-0.97	-0.99
$\Delta G(CHO)$	0.05	0.04
$\Delta G(COH)$	1.42	1.28
$\lambda_R(CHO)$	2.59	2.47
$\lambda_P(CHO)$	3.83	3.80
$\lambda_R(COH)$	2.31	2.33
$\lambda_P(COH)$	2.30	2.33

**Table S9.** Comparison of ML prediction and DFT calculation on Au-NC<sub>2</sub>

Properties	ML	DFT
$E_s$	3.47	4.26
$\Delta G_{*CO}$	-1.28	-0.98
$\Delta G(*H)$	-0.8	-0.54
$\Delta G(CHO)$	0.27	0.58
$\Delta G(COH)$	1.38	1.64
$\lambda_R(CHO)$	2.7	2.14
$\lambda_P(CHO)$	3.8	2.41
$\lambda_R(COH)$	2.33	2.21
$\lambda_P(COH)$	2.3	2.61



**Fig. S16** Model (ABR) accuracy test of 9 properties on Au-NC<sub>2</sub> by 77, 88 and 99 datasets.

## References

1. C. Costentin, D. H. Evans, M. Robert, J.-M. Savéant and P. S. Singh, *J. Am. Chem. Soc.*, 2005, **127**, 12490-12491.
2. C. Costentin, M. Robert and J.-M. Savéant, *J. Am. Chem. Soc.*, 2006, **128**, 4552-4553.