Support Information

Rational Design of Single-Atom Catalysts Stabilized within the Interlayers of Covalent Organic Frameworks for the Oxygen Evolution Reaction

Ran Duan,[a] Lei Sun,[*ac] Xin Wei,[a] Mengqian Xu,[b] Yuwei Pan,[b] Lei Xiao,[a] Yunjie Lang,[b] Wanying Xie,[b] Shaohuang Li,[b] Na Jia,[a] Dong Zhai,[b] Weiqiao Deng,[b] and Li Yang[*b]

[a] College of Chemistry, Chemical Engineering and Resource Utilization, Northeast Forestry University, 26 Hexing Road, Harbin 150040, P. R. China

[b] Institute of Frontier Chemistry, School of Chemistry and Chemical Engineering, Shandong University, Qingdao, 266237, P. R. China

[c] College of Chemistry, Chemical Engineering and Resource Utilization, and Center for Innovative Research in Synthetic Chemistry and Resource Utilization, Northeast Forestry University, 26 Hexing Road, Harbin 150040, P. R. China

[*]Corresponding authors.

E-mail Address: slei@nefu.edu.cn (L. Sun), yangli12@sdu.edu.cn (L. Yang).

Section S1. Density functional theory calculations：

$$\Delta G_1 = G_{*\mathrm{OH}} + G_{\mathrm{H}^+} - G_* - G_{\mathrm{H_2O}} - eU$$
$$\Delta G_2 = G_{*\mathrm{O}} + G_{\mathrm{H}^+} - G_{*\mathrm{OH}} - eU$$
$$\Delta G_3 = G_{*\mathrm{OOH}} + G_{\mathrm{H}^+} - G_{*\mathrm{O}} - G_{\mathrm{H_2O}} - eU$$
$$\Delta G_4 = G_* + G_{\mathrm{H}^+} + G_{\mathrm{O_2}} - G_{*\mathrm{OOH}} - eU$$

Here, the asterisks denote the active sites of the model catalyst. The free energy changes associated with each elementary step are described by equations S1–S4, wherein U represents the applied electrode potential. The adsorption free energies of ΔG_{*OH}, ΔG_{*OOH}, and ΔG_{*O} are referenced to the free energies of stoichiometrically appropriate amounts of H_2O (g) and H_2 (g). Specifically, the free energy of OH^- is referenced to the free energy of H_2, and its relationship is expressed by the following equation:

$$G_{\mathrm{H}^+} = \frac{1}{2}G_{\mathrm{H_2}} - RT\ln 10 * pH$$

The free energies of H_2O (l) and O_2 (g) were calculated based on the thermodynamic data of H_2 (g) and H_2O (g)：

$$G_{\mathrm{H_2O(l)}} = G_{\mathrm{H_2O(g)}} + RT\ln\left(\frac{P}{P_0}\right)$$
$$G_{\mathrm{O_2(g)}} = 2G_{\mathrm{H_2O(g)}} - 2G_{\mathrm{H_2}} + 4.92$$

Here, the free energies of H_2O (g) and H_2 (g) were directly obtained from density functional theory (DFT) calculations. R denotes the ideal gas constant, T was set to 298.15 K, P was taken as 0.035 bar, and P0 corresponds to 1 bar. The onset potential is calculated by:

$$U_{OER}^{onset} = max\{\Delta G_A, \Delta G_B, \Delta G_C, \Delta G_D\}$$

The theory overpotential is calculated by:

$$\eta_{\text{theory}} = max\left\{\Delta G_A, \Delta G_B, \Delta G_C, \Delta G_D\right\} - 1.23$$

Section S2

All ML algorithms were conducted by the open-source code Scikitlearn in the Python3 environment. Given the limited sample size and high feature dimensionality of the dataset in this work, we employed five distinct regression algorithms under a supervised learning framework to construct predictive models. These algorithms encompassed various paradigms, including ensemble learning, kernel methods, probabilistic modeling, and lazy learning, providing multi-faceted insights into the relationship between data features and the target variable. Detailed theoretical formulations, applicability conditions, and algorithmic advantages of each algorithm were provided in the Supplementary Information. Two indicators used to describe prediction errors, the RMSE and R^2 score, were applied herein to evaluate the accuracy of the ML models. Their expressions were as follows:

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_i)^2} \qquad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where $\hat{y}_i$, y_i, and μ_i denote the ground truth, the prediction of the model, and the mean value, respectively. The R^2 score ranged from 0 to 1, and the prediction accuracy of the model was desired when the R^2 score approached the value of 1. The RMSE represented the loss between the prediction and the ground truth. The lower RMSE loss meant a better model performance.

Section S3

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import r2_score, mean_squared_error
import matplotlib.pyplot as plt
import optuna
import shap


# ====== 1. 读取数据并打乱 ======
df = pd.read_excel("potential.xlsx")   # ← 替换为你的路径
df = df.sample(frac=1, random_state=42).reset_index(drop=True)

X = df.drop(columns=["ηTheory"])
y = df["ηTheory"]

# ====== 2. 划分训练/测试集 ======
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.15, random_state=42)

# ====== 3. Optuna 自动调参 ======
def objective(trial):
    params = {
        "n_estimators": trial.suggest_int("n_estimators", 100, 1500),
        "max_depth": trial.suggest_int("max_depth", 1, 4),
        "learning_rate": trial.suggest_float("learning_rate", 0.001, 0.3),
        "subsample": trial.suggest_float("subsample", 0.5, 1.0),
        "min_samples_split": trial.suggest_int("min_samples_split", 2, 20),
        "min_samples_leaf": trial.suggest_int("min_samples_leaf", 2, 20),
        "max_features": trial.suggest_categorical("max_features", ["sqrt", "log2", None]),
        "random_state": 42
    }
    model = GradientBoostingRegressor(**params)
    model.fit(X_train, y_train)
    preds = model.predict(X_test)
    rmse = mean_squared_error(y_test, preds, squared=False)
```

```
    return rmse

study = optuna.create_study(direction="minimize")
study.optimize(objective, n_trials=5000)
best_params = study.best_params
print("✅ 最佳参数：", best_params)
# ====== 4. 最佳模型训练与预测 ======
model = GradientBoostingRegressor(**best_params)
model.fit(X_train, y_train)
y_train_pred = model.predict(X_train)
y_test_pred = model.predict(X_test)
r2_train = r2_score(y_train, y_train_pred)
r2_test = r2_score(y_test, y_test_pred)
rmse_train = mean_squared_error(y_train, y_train_pred, squared=False)
rmse_test = mean_squared_error(y_test, y_test_pred, squared=False)
print(f"训练集 R²: {r2_train:.3f}, RMSE: {rmse_train:.3f}")
print(f"测试集 R²: {r2_test:.3f}, RMSE: {rmse_test:.3f}")
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
plt.scatter(y_train, y_train_pred, color='dodgerblue', alpha=0.7)
plt.plot([y_train.min(), y_train.max()], [y_train.min(), y_train.max()], 'r--')
plt.xlabel("True ηTheory")
plt.ylabel("Predicted")
plt.title(f"Train: R²={r2_train:.3f}, RMSE={rmse_train:.3f}")
plt.subplot(1, 2, 2)
plt.scatter(y_test, y_test_pred, color='seagreen', alpha=0.7)
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'r--')
plt.xlabel("True ηTheory")
plt.ylabel("Predicted")
plt.title(f"Test: R²={r2_test:.3f}, RMSE={rmse_test:.3f}")
plt.tight_layout()
plt.savefig("gbr_eta_scatter.png", dpi=300)
plt.show()

# ====== 5. SHAP 分析 ======
explainer = shap.Explainer(model)
shap_values = explainer(X)
plt.figure()
shap.plots.bar(shap_values, max_display=12)
plt.title("SHAP Feature Importance ")
plt.savefig("shap_bar_theory.png", dpi=1500, bbox_inches='tight')
plt.show()
plt.figure()
shap.plots.beeswarm(shap_values, max_display=12)
```

```
plt.title("SHAP Beeswarm Plot ")
plt.savefig("theory_shap_beeswarm.png", dpi=1500, bbox_inches='tight')
plt.show()
```

Section S4. Computational Details for Periodic VASP Calculations

Periodic density functional theory (DFT) calculations were performed using the Vienna Ab initio Simulation Package (VASP). The projector-augmented wave (PAW) method was employed to describe the ion–electron interactions. The plane-wave energy cutoff was set to 400 eV. Geometry optimizations were carried out using a Γ-centered 1 × 1 × 1 k-point mesh due to the large supercell size. The electronic self-consistency criterion was set to 1×10^{-5} eV (EDIFF = 1E-5), and the structures were relaxed until the residual forces were below 0.05 eV $Å^{-1}$ (EDIFFG = −0.05 eV $Å^{-1}$). A maximum of 1000 ionic steps (NSW = 1000) and 120 electronic iterations (NELM = 120) were allowed. Spin-polarized calculations were performed (ISPIN = 2). Initial magnetic moments were assigned to the transition metal atoms. The DFT-D3 method with Becke–Johnson damping (IVDW = 12) was applied to account for van der Waals interactions. The PREC was set to Normal, and LREAL was set to Auto.

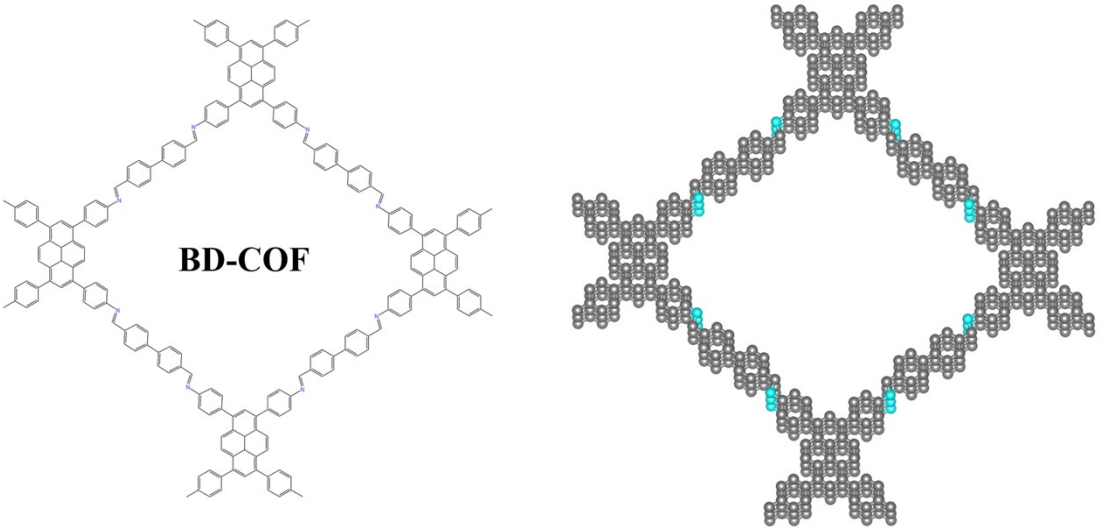

Fig. S1 Structure of BD-COF and AA stacking configuration. Blue represents N atoms.

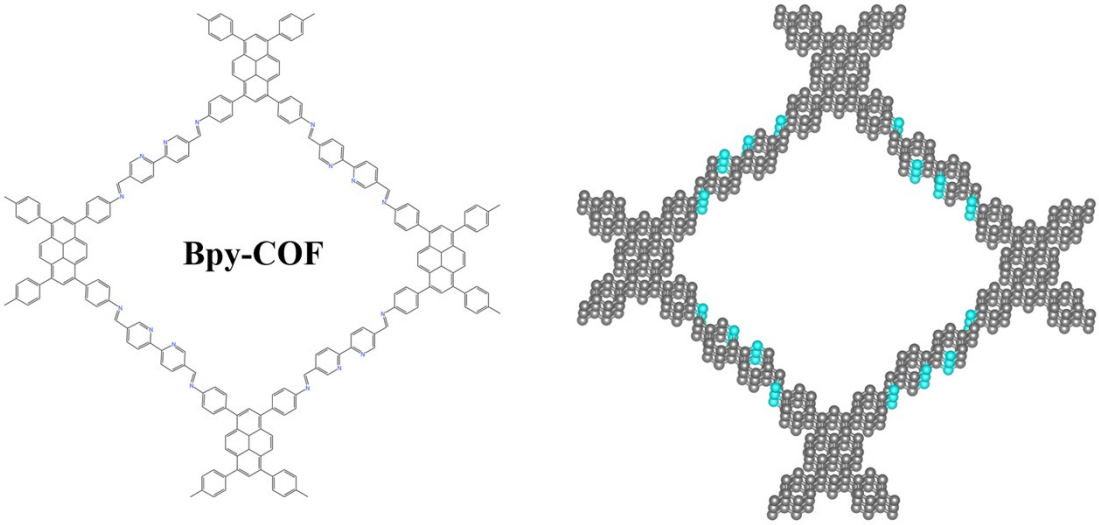

Fig. S2 Structure of Bpy-COF and AA stacking configuration.

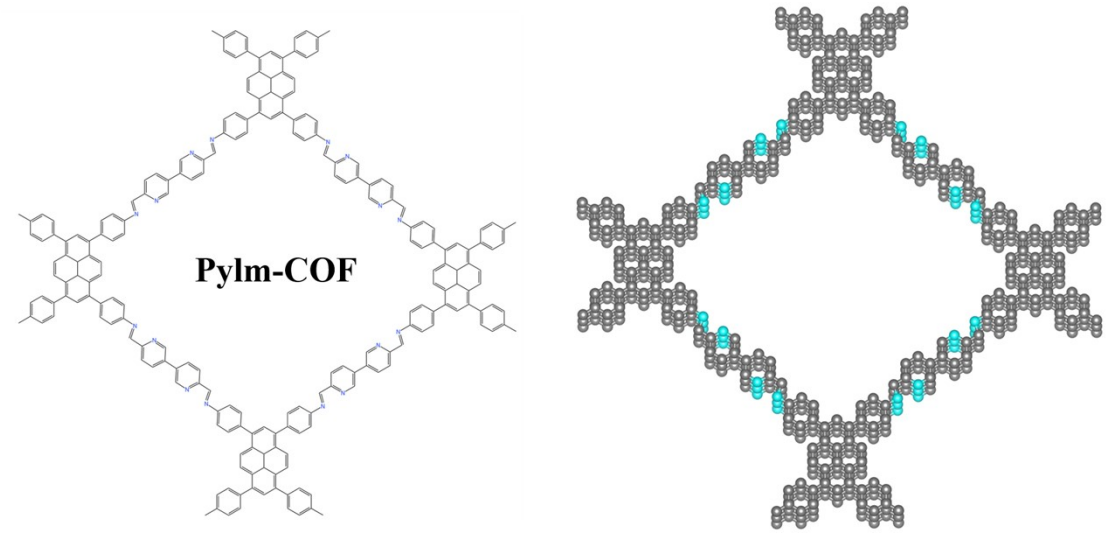

Fig. S3 Structure of Pylm-COF and AA stacking configuration.

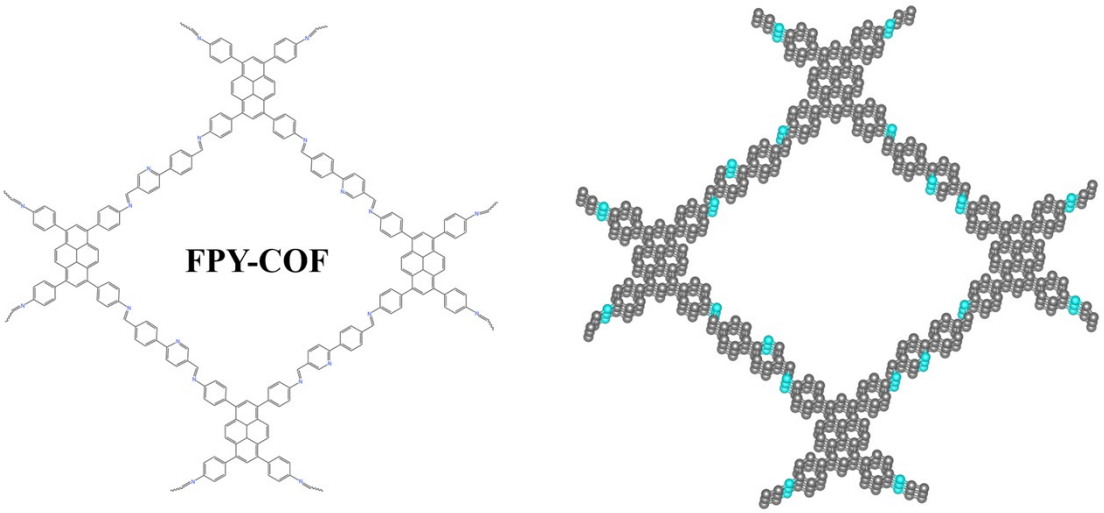

Fig. S4 Structure of FPY-COF and AA stacking configuration.

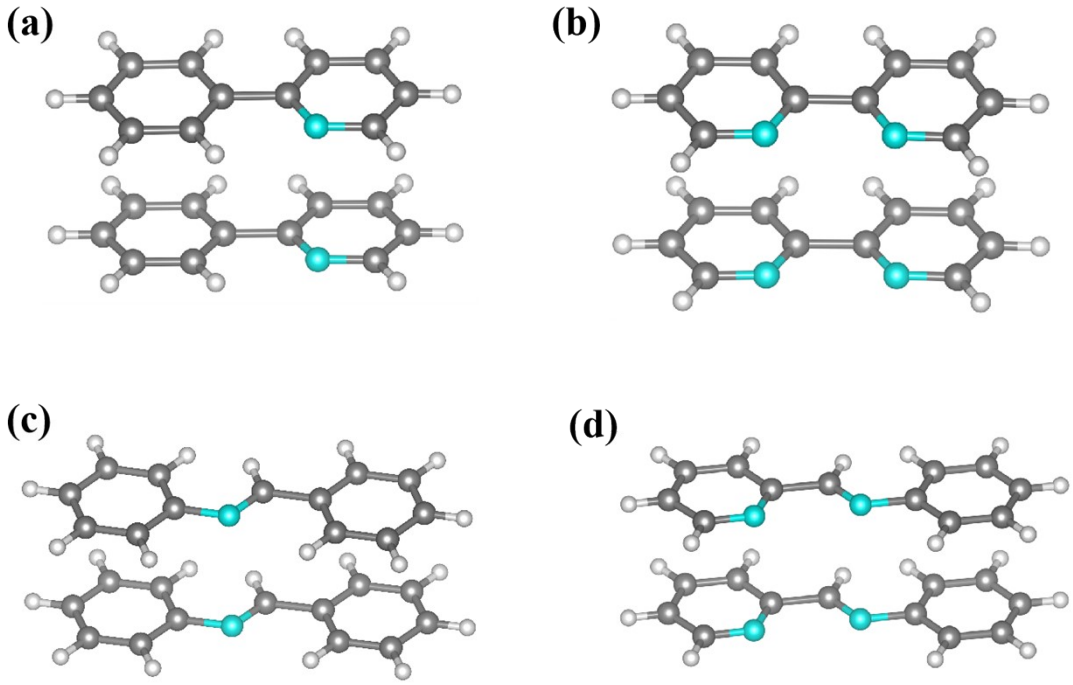

Fig. S5 Four interlayer COF configurations: (a) Ben-Pyr (2N), (b) Pyr (4N), (c) Ben-N (2N) and (d) Pyr-N (4N).

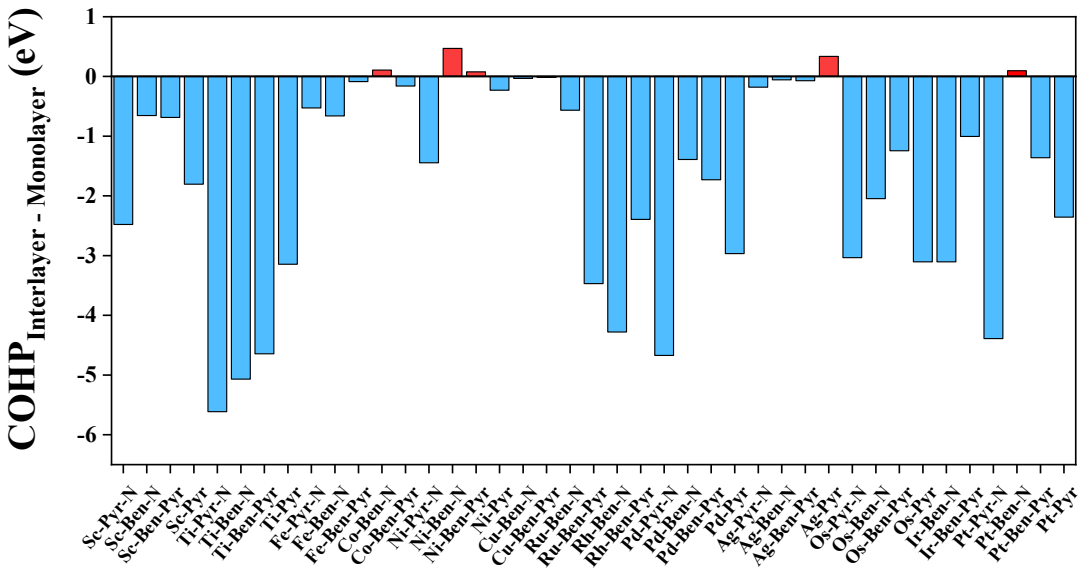

Fig. S6 Under conditions favoring enhanced interlayer stability, shows the difference between COHP Interlayer and Monolayer (COHP $_{\text{Interlayer}}$ - COHP $_{\text{Monolayer}}$).

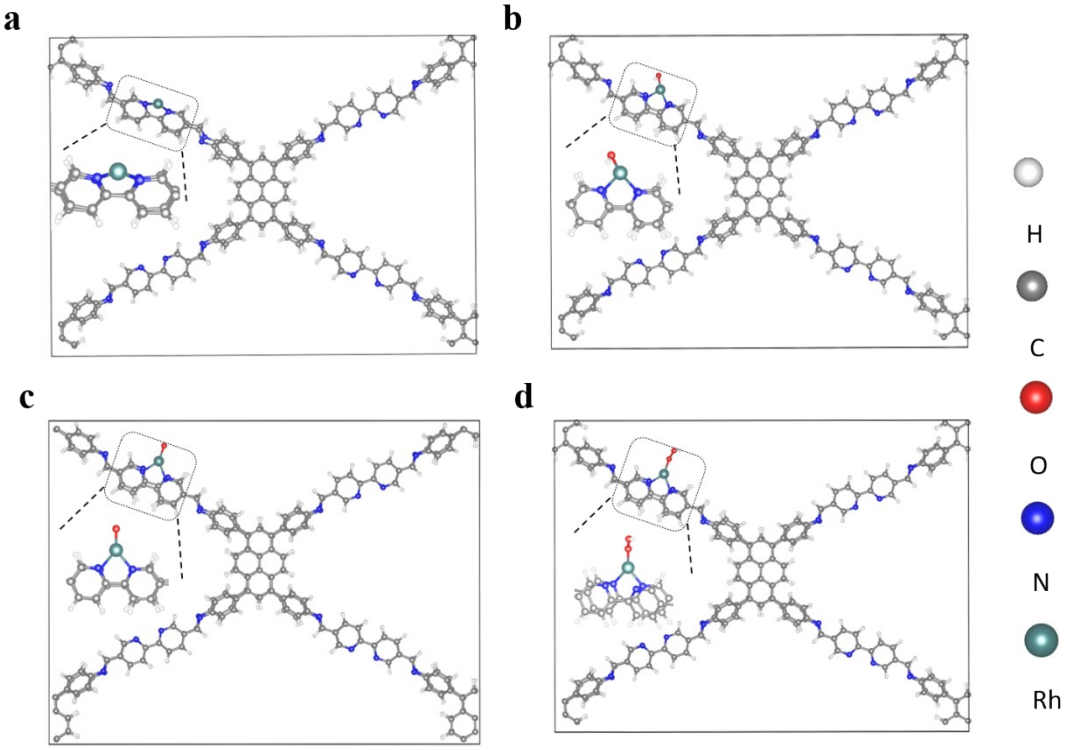

Fig. S7 Adsorption states of Rh-Pyr/Interlayer (4N) and three key intermediates (a) clean; (b) *OH; (c) *O; (d) *OOH.

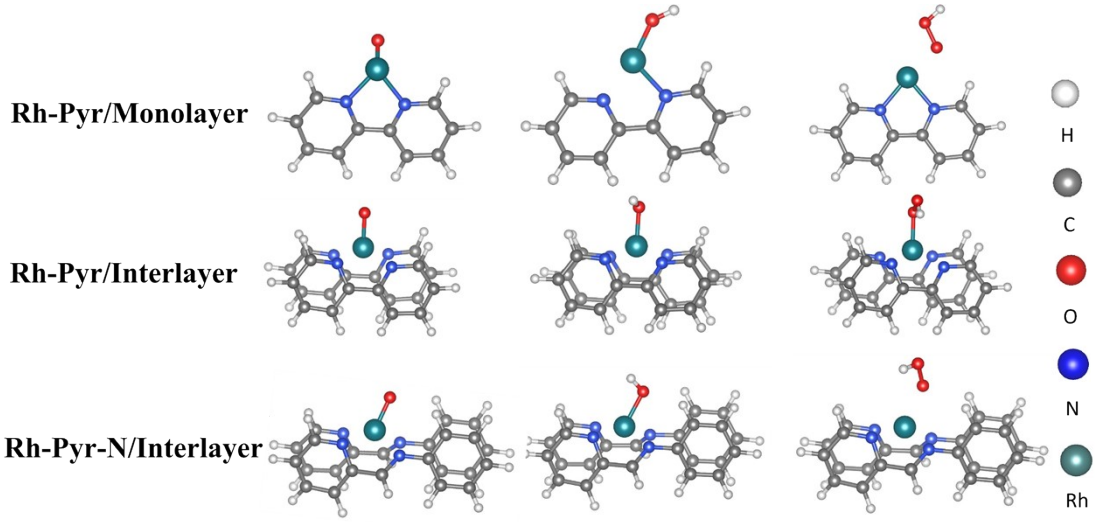

Fig. S8 Adsorption configurations of key intermediates on Rh-Pyr/Monolayer (2N), Rh-Pyr/Interlayer (4N), and Rh-Pyr-N/Interlayer (4N) catalysts.

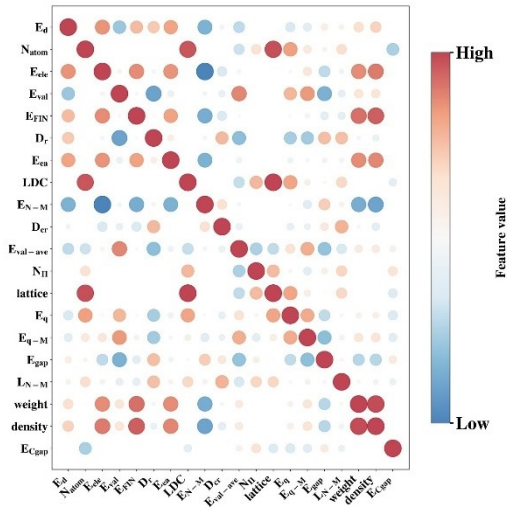

Fig. S9 Pearson correlation coefficients of 20 descriptors. Color represents correlation, and the darker the color, the higher the correlation.

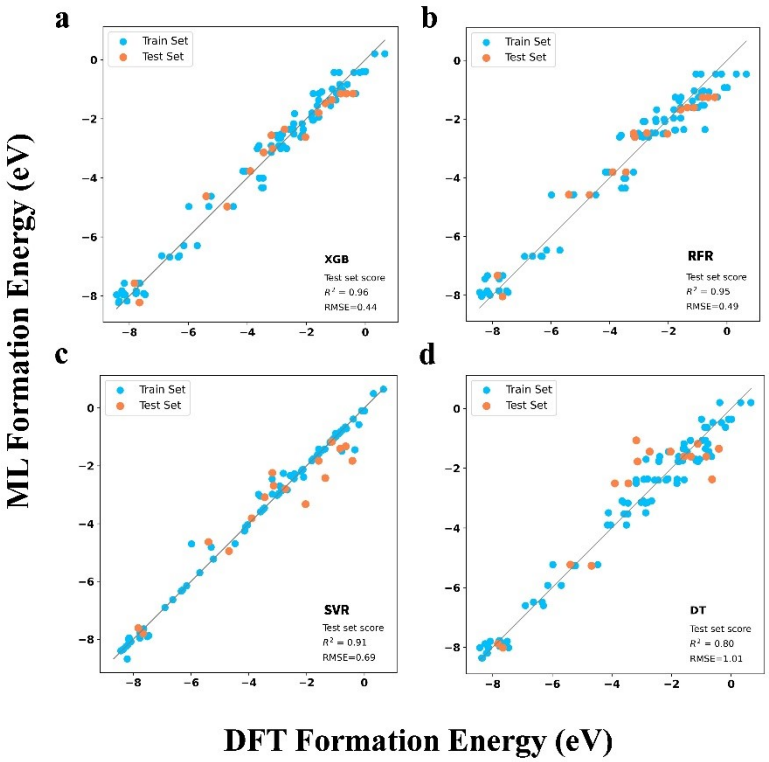

Fig. S10 Four machine learning models predict formation energy (a) XGB; (b) RFR; (c) SVR and (d) DT.

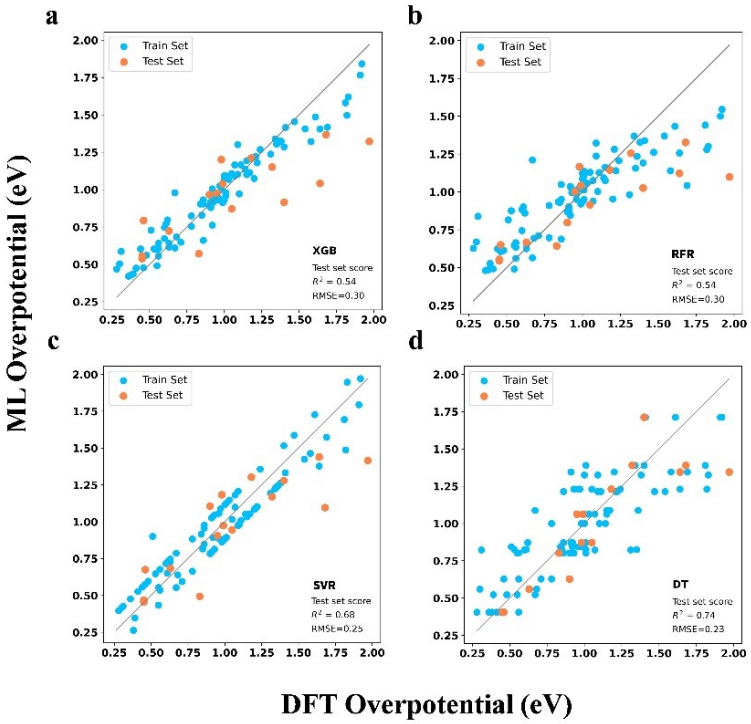

Fig. S11 Four machine learning models predict overpotential (a) XGB; (b) RFR; (c) SVR and (d) DT.

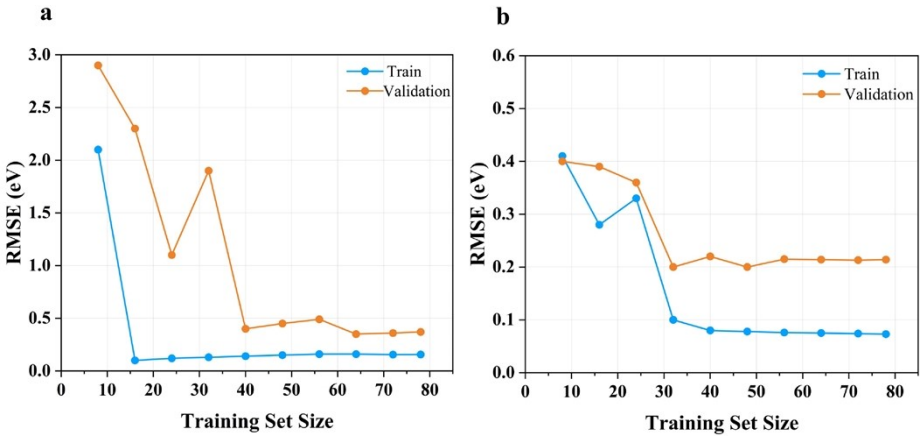

Fig. S12 Learning curves of the GBR model illustrating training and validation RMSE as a function of training set size: (a) formation energy prediction; (b) theoretical overpotential prediction.

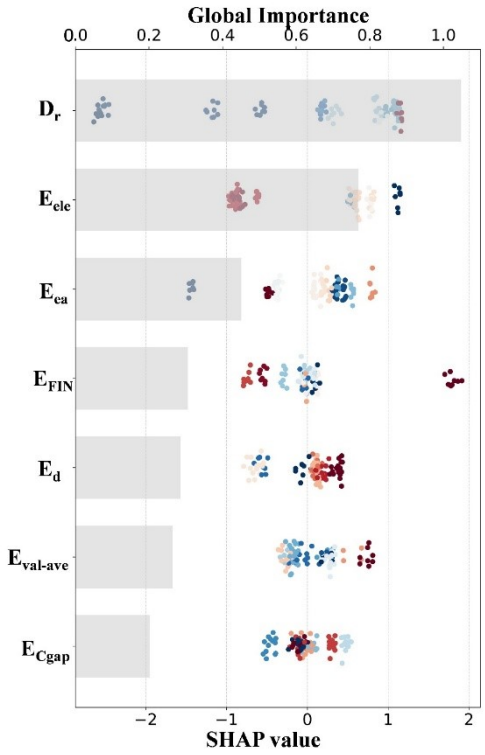

Fig. S13 SHAP global feature importance analysis for formation energy prediction.

Table S1. The LDC and stacking modes of four different COFs.

Name	LDC/ **Å**	Stack
Ben-Pyr	4.33	AA
Pyr	4.33	AA
Ben-N	3.92	AA
Pyr-N	3.60	AA

Table S2. Overpotential、the formation energy and gibbs free energy of all 104 groups of catalysts.

Layer	Element	COF	ΔG_1	ΔG_2	ΔG_3	ΔG_4	η	E_f
Interlayer	Sc	Pyr-N	0.8	2.59	0.95	0.58	1.36	-0.81
Interlayer	Sc	Ben-N	0.93	2.24	1.26	0.48	1.01	-0.89
Interlayer	Sc	Ben-Pyr	0.59	2.84	0.53	0.96	1.61	-0.38
Interlayer	Sc	Pyr	1.47	2.61	0.32	0.52	1.38	-1.12
Interlayer	Ti	Pyr-N	0.16	1.19	2.4	1.17	1.17	-8.35
Interlayer	Ti	Ben-N	0.27	1.34	2.09	1.22	0.86	-8.18
Interlayer	Ti	Ben-Pyr	0.12	0.51	3.04	1.24	1.81	-7.78
Interlayer	Ti	Pyr	0.21	1.68	1.9	1.13	0.67	-8.08
Interlayer	Fe	Pyr-N	0.08	2.38	0.6	1.87	1.15	-2.14
Interlayer	Fe	Ben-N	-0.25	2.45	0.54	2.18	1.22	-2.8
Interlayer	Fe	Ben-Pyr	-0.4	2.77	1.77	0.78	1.54	-2.43
Interlayer	Fe	Pyr	0.7	2	-0.34	2.55	1.32	-2.43
Interlayer	Co	Pyr-N	1.64	1.73	2.22	-0.67	0.99	-0.78
Interlayer	Co	Ben-N	1.38	2.01	0.94	0.59	0.78	-1.58
Interlayer	Co	Ben-Pyr	1.17	1.92	2.37	-0.55	1.14	-1
Interlayer	Co	Pyr	2.28	1.93	1.48	-0.77	1.05	-0.86
Interlayer	Ni	Pyr-N	1.15	2.3	0.81	0.66	1.07	-3.02
Interlayer	Ni	Ben-N	2.09	1.96	0.93	-0.06	0.86	-3.61
Interlayer	Ni	Ben-Pyr	0.99	2.21	1.08	0.65	0.98	-3.19
Interlayer	Ni	Pyr	1.81	2.87	-0.46	0.7	1.64	-3.19
Interlayer	Cu	Pyr-N	0.94	2.26	0.94	0.77	1.03	-1.35
Interlayer	Cu	Ben-N	1.45	2.38	0.72	0.37	1.15	-2.4
Interlayer	Cu	Ben-Pyr	0.99	2.23	0.97	0.73	1	-1.61
Interlayer	Cu	Pyr	1.79	2.41	-1.05	1.77	1.18	-1.36
Interlayer	Ru	Pyr-N	1.5	0.91	2.16	0.35	0.93	-8.37
Interlayer	Ru	Ben-N	1.15	0.89	1.53	1.35	0.3	-8.21
Interlayer	Ru	Ben-Pyr	0.84	0.6	2.13	1.35	0.9	-7.46
Interlayer	Ru	Pyr	1	1.11	2.24	0.57	1.01	-7.66
Interlayer	Rh	Pyr-N	1.45	1.33	1.69	0.45	0.46	-3.51
Interlayer	Rh	Ben-N	1.25	1.29	2.01	0.36	0.78	-4.15
Interlayer	Rh	Ben-Pyr	1.69	1.43	1.45	0.36	0.46	-3.9
Interlayer	Rh	Pyr	1.51	1.46	1.29	0.66	0.28	-3.46
Interlayer	Pd	Pyr-N	1.78	1.55	1.37	0.22	0.55	-2.95
Interlayer	Pd	Ben-N	1.46	2.09	1.11	0.26	0.86	-3.66
Interlayer	Pd	Ben-Pyr	1.79	1.67	1.21	0.26	0.56	-3.14
Interlayer	Pd	Pyr	1.76	2.06	0.9	0.2	0.83	-2.88
Interlayer	Ag	Pyr-N	1.84	2.23	0.42	0.42	1	-0.84
Interlayer	Ag	Ben-N	1.77	0.41	2.28	0.46	1.05	-1.77
Interlayer	Ag	Ben-Pyr	1.82	0.37	2.24	0.49	1.01	-1.57
Interlayer	Ag	Pyr	1.87	2.16	0.49	0.4	0.93	-0.97

Interlayer	Os	Pyr-N	0.11	0.44	2.64	1.73	1.41	-7.77
Interlayer	Os	Ben-N	0.37	0.29	2.91	1.36	1.68	-8.17
Interlayer	Os	Ben-Pyr	0.24	-0.02	3.15	1.55	1.92	-7.83
Interlayer	Os	Pyr	0.71	0.14	3.14	0.94	1.91	-8.25
Interlayer	Ir	Pyr-N	1.19	1.43	1.62	0.68	0.39	-2.19
Interlayer	Ir	Ben-N	0.86	1.94	1.11	1.01	0.71	-2.44
Interlayer	Ir	Ben-Pyr	1.36	0.93	1.67	0.95	0.44	-2.1
Interlayer	Ir	Pyr	1.67	1.29	1.71	0.24	0.48	-2.03
Interlayer	Pt	Pyr-N	1.78	1.36	1.58	0.21	0.55	-5.4
Interlayer	Pt	Ben-N	1.54	1.57	1.59	0.23	0.36	-5.98
Interlayer	Pt	Ben-Pyr	1.9	1.08	1.72	0.22	0.67	-5.31
Interlayer	Pt	Pyr	1.64	1.64	1.42	0.22	0.41	-5.23
Monolayer	Sc	Pyr-N	0.56	2.63	0.95	0.78	1.4	-0.19
Monolayer	Sc	Ben-N	1.07	1.01	2.58	0.26	1.35	0.32
Monolayer	Sc	Ben-Pyr	0.62	2.63	1.29	0.38	1.4	0.66
Monolayer	Sc	Pyr	1.62	1.76	1.75	-0.21	0.53	-1.06
Monolayer	Ti	Pyr-N	0.44	1.38	2.23	0.88	1	-7.77
Monolayer	Ti	Ben-N	0.49	1.19	3.06	0.18	1.83	-6.3
Monolayer	Ti	Ben-Pyr	0.58	1.23	2.32	0.78	1.09	-6.9
Monolayer	Ti	Pyr	0.41	1.84	1.63	1.04	0.61	-7.51
Monolayer	Fe	Pyr-N	0.2	2.08	1.06	1.59	0.85	-2.12
Monolayer	Fe	Ben-N	-0.05	2.34	0.94	1.69	1.11	-1.18
Monolayer	Fe	Ben-Pyr	0.41	2.24	0.37	1.9	1.01	-1.63
Monolayer	Fe	Pyr	0.89	1.83	0.65	1.55	0.6	-2.86
Monolayer	Co	Pyr-N	2.15	1.72	1.38	-0.33	0.92	-1.12
Monolayer	Co	Ben-N	1.71	1.7	1.74	-0.22	0.51	-0.09
Monolayer	Co	Ben-Pyr	1.65	1.75	2.333	-0.8	1.1	0.01
Monolayer	Co	Pyr	2.2	1.83	1.66	-0.78	0.97	-1.19
Monolayer	Ni	Pyr-N	1.71	2.07	0.55	0.58	0.84	-2.79
Monolayer	Ni	Ben-N	1.29	2.18	1.5	-0.05	0.95	-1.58
Monolayer	Ni	Ben-Pyr	1.61	2.15	0.34	0.83	0.92	-1.82
Monolayer	Ni	Pyr	2.03	3.05	-0.89	0.74	1.82	-2.93
Monolayer	Cu	Pyr-N	1.93	2.22	0.95	-0.18	0.99	-1.58
Monolayer	Cu	Ben-N	1.7	2.19	0.34	0.69	0.96	-0.62
Monolayer	Cu	Ben-Pyr	2.01	2.21	0.19	0.51	0.98	-0.86
Monolayer	Cu	Pyr	2.47	2.2	-0.02	0.27	1.24	-1.81
Monolayer	Ru	Pyr-N	1.83	1.82	1.69	-0.42	0.6	-8.43
Monolayer	Ru	Ben-N	1.52	0.66	2.81	-0.07	1.58	-6.34
Monolayer	Ru	Ben-Pyr	1.36	1	1.54	1.02	0.31	-6.63
Monolayer	Ru	Pyr	2.43	1.59	2.05	-1.14	1.2	-8.14
Monolayer	Rh	Pyr-N	1.76	1.79	1	0.38	0.56	-4.04
Monolayer	Rh	Ben-N	2.3	1.08	1.55	-0.01	1.07	-3.19
Monolayer	Rh	Ben-Pyr	2.54	1.16	1.91	-0.69	1.31	-3.45
Monolayer	Rh	Pyr	2.13	1.66	1.3	-0.17	0.9	-4.12

Monolayer	Pd	Pyr-N	1.6	1.86	0.67	0.8	0.63	-2.67
Monolayer	Pd	Ben-N	1.5	1.86	0.72	0.84	0.63	-1.76
Monolayer	Pd	Ben-Pyr	1.38	1.86	0.78	0.9	0.63	-1.78
Monolayer	Pd	Pyr	1.61	1.91	0.61	0.8	0.68	-2.87
Monolayer	Ag	Pyr-N	1.7	2.14	0.6	0.48	0.91	-0.66
Monolayer	Ag	Ben-N	1.94	2.14	0.36	0.48	0.91	-0.33
Monolayer	Ag	Ben-Pyr	1.85	0.9	1.65	0.51	0.62	-0.41
Monolayer	Ag	Pyr	1.58	2.19	0.63	0.52	0.96	-0.64
Monolayer	Os	Pyr-N	0.89	0.92	2.44	0.67	1.21	-7.75
Monolayer	Os	Ben-N	1.31	0.19	2.32	1.1	1.09	-5.7
Monolayer	Os	Ben-Pyr	0.69	0.71	2.57	0.95	1.34	-6.15
Monolayer	Os	Pyr	0.91	0.69	2.7	0.62	1.47	-7.64
Monolayer	Ir	Pyr-N	2.14	0.89	1.78	0.12	0.91	-2.57
Monolayer	Ir	Ben-N	1.91	-0.13	3.2	-0.06	1.97	-0.75
Monolayer	Ir	Ben-Pyr	1.22	1.47	2.87	-0.64	1.64	-1.47
Monolayer	Ir	Pyr	2.92	0.45	1.94	-0.39	1.69	-2.74
Monolayer	Pt	Pyr-N	1.11	1.61	1.02	1.18	0.38	-4.47
Monolayer	Pt	Ben-N	0.79	1.68	1.25	1.19	0.45	-3.47
Monolayer	Pt	Ben-Pyr	0.7	1.68	1.65	0.89	0.45	-3.59
Monolayer	Pt	Pyr	0.93	1.79	0.96	1.24	0.56	-4.69

Table S3. Comparative analysis of cluster and periodic model calculations for Rh–Pyr/Interlayer

Model	ΔG_A	ΔG_B	ΔG_C	ΔG_D
Cluster	1.51	1.46	1.29	0.66
Periodic	1.47	1.43	1.30	0.93

Table S4. The names and meanings of 20 descriptors.

Descriptors	Detailed explanation of the descriptors
E_d	The number of d orbitals in the TM
N_{atom}	The total number of atoms in the structure
E_{ele}	The electronegativity of TM
E_{val}	The valence electron number of TM
E_{FIN}/ eV	The first ionization energy of TM
D_r/ Å	The ionic radius of TM
E_{ea}/ eV	The electronic affinity energy of TM
LDC/ Å	Interlayer distance of COF
E_{N-M}	The difference between the electronegativity of N atoms and that of TM atoms
D_{cr}/ Å	TM covalent radius
$E_{val-ave}$	Ratio of metal charge to total N and TM numbers in the system
N_{Π}	The number of Π bonds in the structure
lattice	The number of layers of COF
E_q	Ratio of TM valence state to the total number of N atoms in the optimized system
E_{q-M}	Valence state of TM in optimized system
E_{gap} /eV	Bandgap of optimized system
L_{N-M}/ Å	Average bond length between N and TM in optimized system
weight	The relative mass of metal atoms
density/g.cm^{-3}	Density of TM
E_{Cgap}/eV	Bandgap of the COF