Supporting information

Elucidating the impact of oxygen functional groups on the catalytic activity of M-N_4-C catalysts for oxygen reduction reaction: A density functional theory and machine learning approach

Liang Xie[b], Wei Zhou[a, b*], Yuming Huang[b], Zhibin Qu[b], Longhao Li[b], Chaowei Yang[b], Yani Ding[b], Junfeng Li[b], Xiaoxiao Meng[b], Fei Sun[b], Jihui Gao[b*], Guangbo Zhao[b], Yukun Qin[b].

a State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin, Heilongjiang, 150001 P. R. China

b School of Energy Science and Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, 150001 P. R. China

* Corresponding author:

E-mail addresses: hitzhouw@hit.edu.cn (Wei Zhou)

E-mail addresses: gaojh@hit.edu.cn (Jihui Gao)

1. Method for calculating the electronegativity of OGs

In order to investigate the effect of OGs on MN_4, we consider the OGs as a whole and calculate their average electronegativity (χ_{OGs}). The formula for calculating by harmonic average method is as follows:

$$\chi_{OGs} = \frac{N_{OGs}}{\frac{N_C}{\chi_C} + \frac{N_H}{\chi_H} + \frac{N_O}{\chi_O}}$$

Where N_{OGs} represents the total number of atoms in OGs, N_C, N_H and N_O respectively represent the number of atoms C, H and O in OGs, χ_C, χ_H and χ_O respectively represent the Pauli electronegativity of C, H and O.

2. Machine learning model used in this work

The ML models used for training and prediction include RFR, KNN, SVR, NN, XGBR, LINER, GBR, GPR, and LASSO. Here is a brief introduction to them:

(1) RFR (Random Forest Regressor): RFR is an integrated learning model based on decision trees, which improves prediction accuracy by constructing multiple decision trees and taking their average values. Applicable to regression problems. (Type: integrated learning model)

(2) KNN (K-Nearest Neighbors): Description: KNN performs classification or regression based on the nearest neighbor relationship of the sample. For a given sample, predictions are made by looking at its nearest k neighbors, either by majority voting or by average. (Type: Supervised learning model)

(3) SVR (Support Vector Regressor):SVR is an application of support vector machines that focuses on regression problems. By finding the support vector in the data, an optimal hyperplane is constructed to minimize the prediction error. (Type: Support vector machine model)

(4) NN (Neural Network): NN is a model that simulates the structure of human brain neural network. Composed of input layers, hidden layers, and output layers, it learns weights to achieve complex non-linear relationships and is suitable for a variety of tasks, including classification and regression. (Type: Deep learning model)

(5) XGBR (XGBoost Regressor): XGBR is a gradient lifting algorithm that improves performance by integrating multiple decision trees. It is excellent at handling structured data and regression problems. (Type: integrated learning model)

(6) LINER (Linear Regression): LINER is a simple but powerful linear model that performs regression by fitting linear relationships in the data. Suitable for problems where linear relationships are obvious. (Type: Linear regression model)

(7) GBR (Gradient Boosting Regressor): GBR is a gradient lifting algorithm that improves prediction performance by iteratively training a weak model and correcting the errors of the previous model. Applicable to regression problems. (Type: integrated learning model)

(8) GPR (Gaussian Process Regressor): GPR is based on Bayesian inference, which treats predictions as Gaussian distributions over the underlying functions. It is suitable for small sample regression problems and provides uncertainty estimation. (Type: Gaussian process model)

(9) LASSO (Least Absolute Shrinkage and Selection Operator): LASSO is a linear regression model that achieves feature selection by adding L1 regularization to the coefficients. It is suitable for regression problems in high dimensional data sets. (Type: Linear regression model)

3.Supplementary Figures

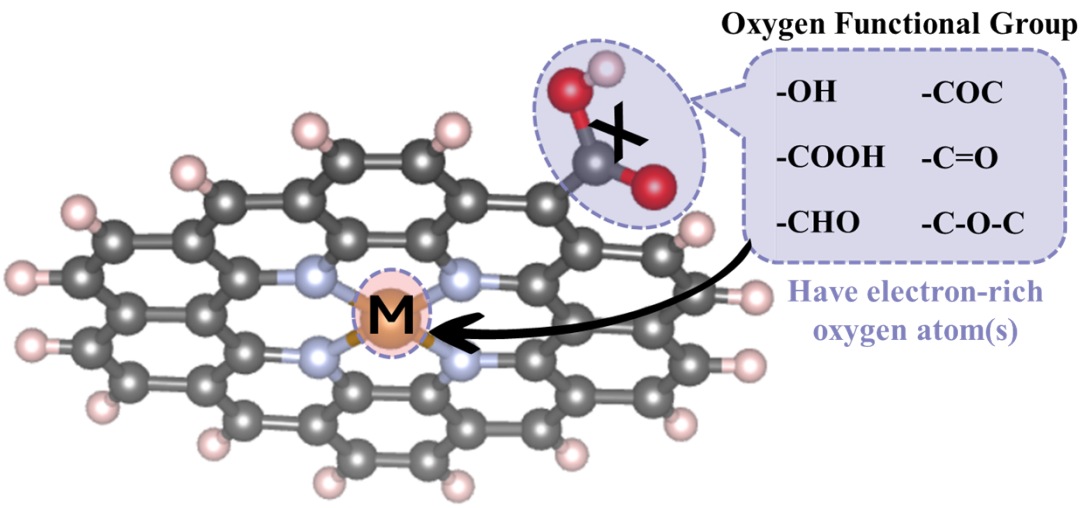

Figure S1. In OGs@MN_4 catalyst, oxygen functional groups (OH, COOH, CHO, COC, C-O-C, C=O, etc.) have electron-rich oxygen atom(s).

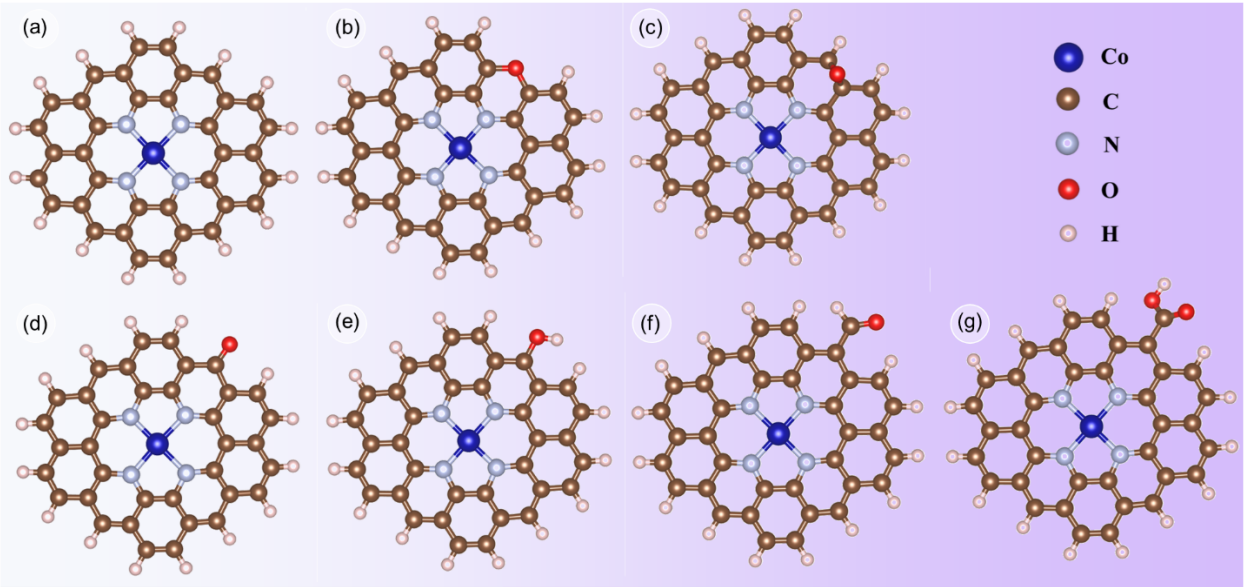

Figure S2. Optimized structures of single doped OGs@CoN_4. (a) CoN_4, (b) COC@CoN_4, (c) C-O-C@CoN_4, (d) C=O@CoN_4, (e) OH@CoN_4, (f) CHO@CoN_4 and (g) COOH@CoN_4.

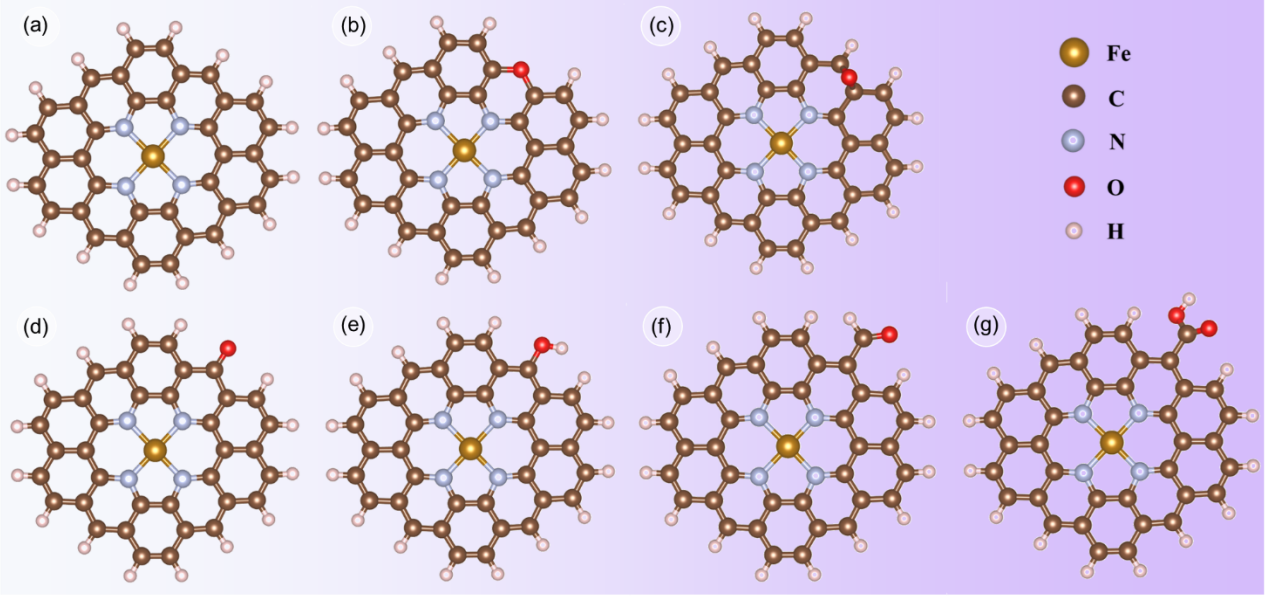

Figure S3. Optimized structures of single doped OGs@FeN_4. (a) FeN_4, (b) COC@FeN_4, (c) C-O-C@FeN_4, (d) C=O@FeN_4, (e) OH@FeN_4, (f) CHO@FeN_4 and (g) COOH@FeN_4.

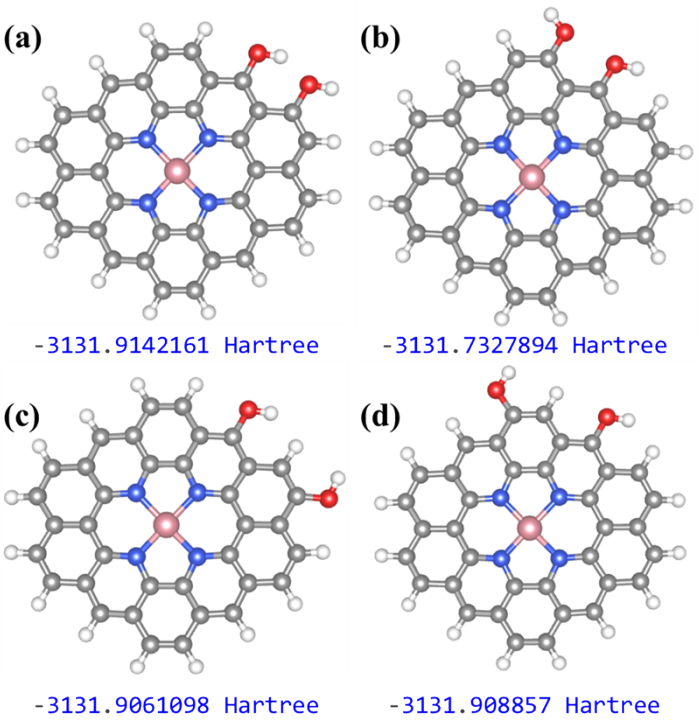

Figure S4. Four configurations of 2OH@CoN_4 are considered in this work.

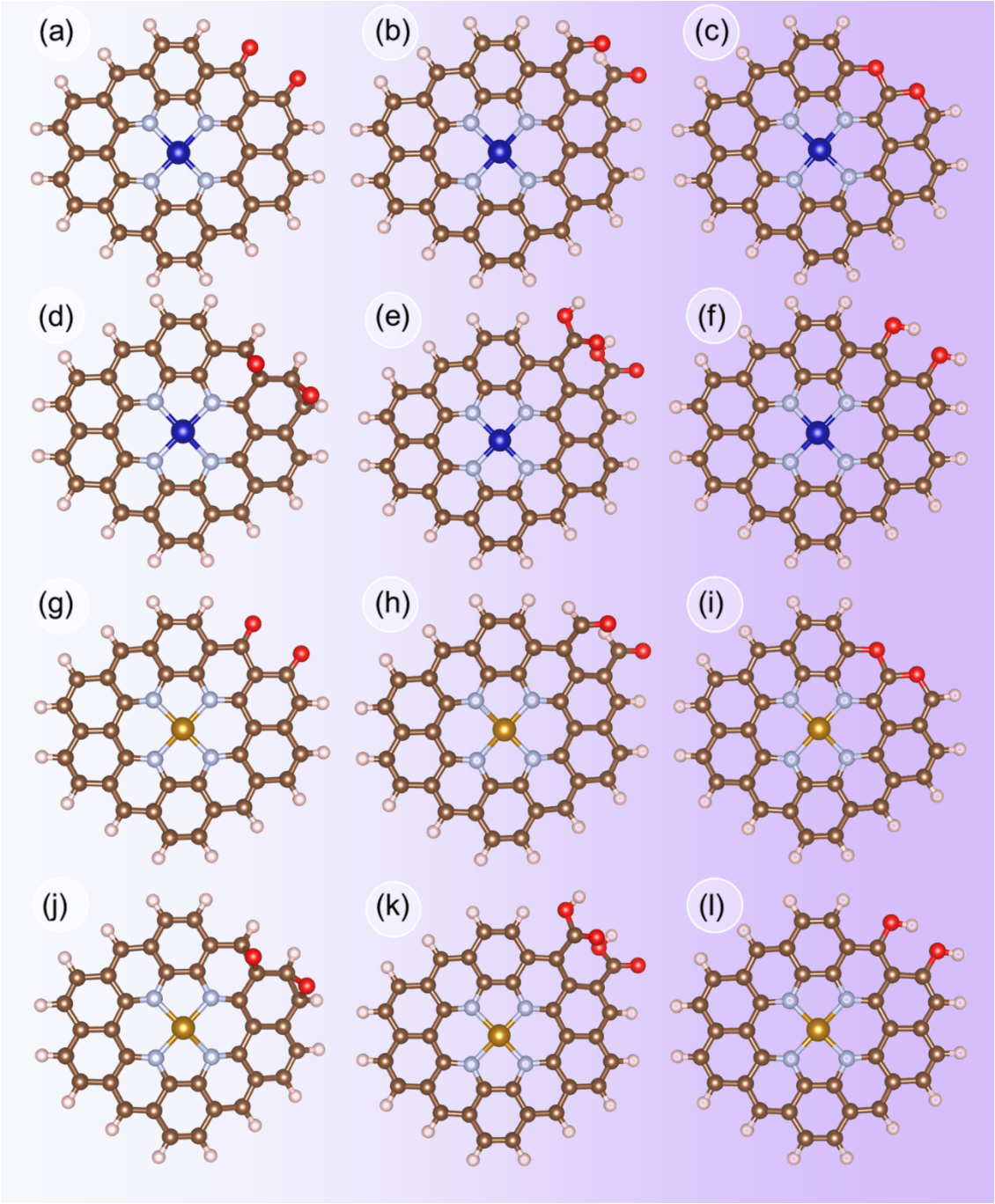

Figure S5. Optimized structures of double doped OGs@MN_4. (a) 2C=O@CoN_4, (b) 2CHO@CoN_4, (c) 2COC@CoN_4, (d) 2C-O-C@CoN_4, (e) 2COOH@CoN_4, (f) 2OH@CoN_4, (g) 2C=O@FeN_4, (h) 2CHO@FeN_4, (i) 2COC@FeN_4, (j) 2C-O-C@FeN_4, (k) 2COOH@FeN_4, (l) 2OH@FeN_4,

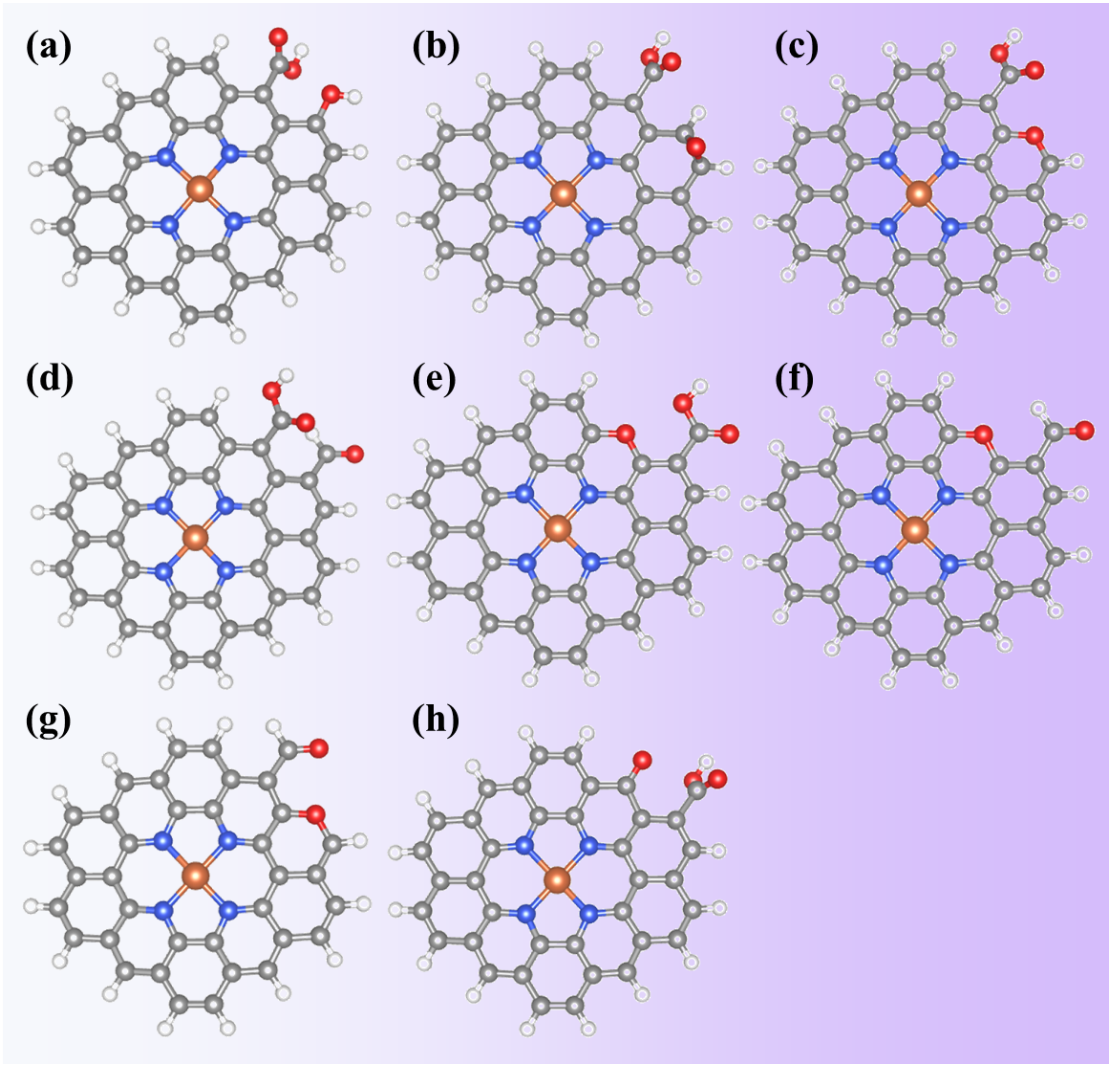

Figure S6. Optimized structures of double doped OGs@MN_4. (a) COOH+OH@MN_4, (b) COOH+C-O-C@MN_4, (c) COOH+COC@MN_4, (d) COOH+CHO@MN_4, (e) COC+COOH@MN_4, (f) COC+CHO@MN_4, (g) CHO+COC@MN_4, (h) C=O+COOH@MN_4.

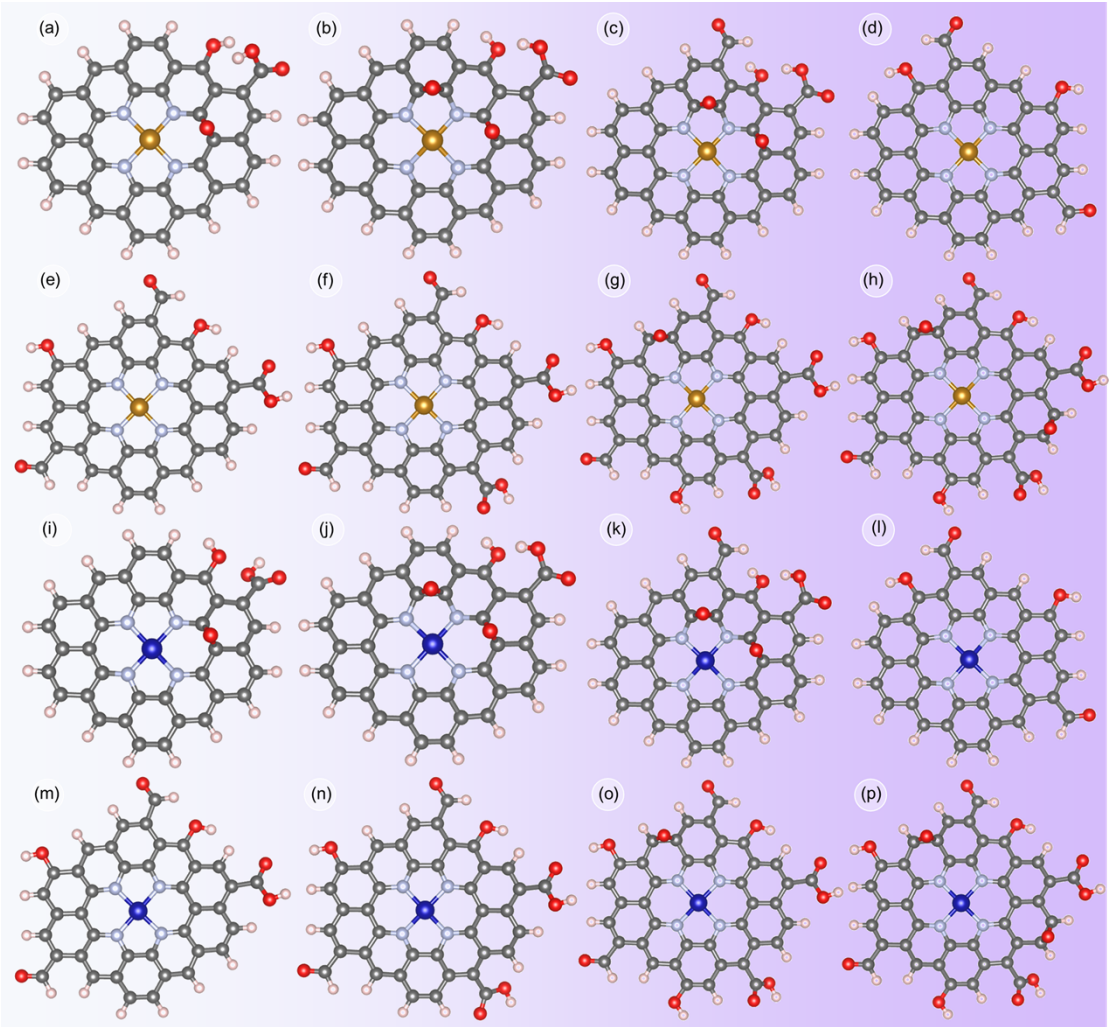

Figure S7. Optimized structures of multiple oxygen doped OGs@MN_4. (a) OH+COOH+C-O-C@CoN_4, (b) OH+COOH+2C-O-C@CoN_4, (c) OH+COOH+2C-O-C+CHO@CoN_4, (d) 2OH+2CHO@CoN_4, (e) 2OH+2CHO+COOH@CoN_4, (f) 2OH+2CHO+2COOH@CoN_4, (g) 2OH+2CHO+2COOH+C-O-C@CoN_4, (h) 2OH+2CHO+2COOH+2C-O-C@CoN_4, (i) OH+COOH +C-O-C@FeN_4, (j) OH+COOH+2C-O-C@FeN_4, (k) OH+COOH+2C-O-C+CHO@FeN_4, (l) 2OH+2CHO@FeN_4, (m) 2OH+2CHO+COOH@FeN_4, (n) 2OH+2CHO+2COOH@FeN_4, (o) 2OH+2CHO+2COOH+C-O-C@FeN_4, (p) 2OH+2CHO+2COOH+2C-O-C@FeN_4,

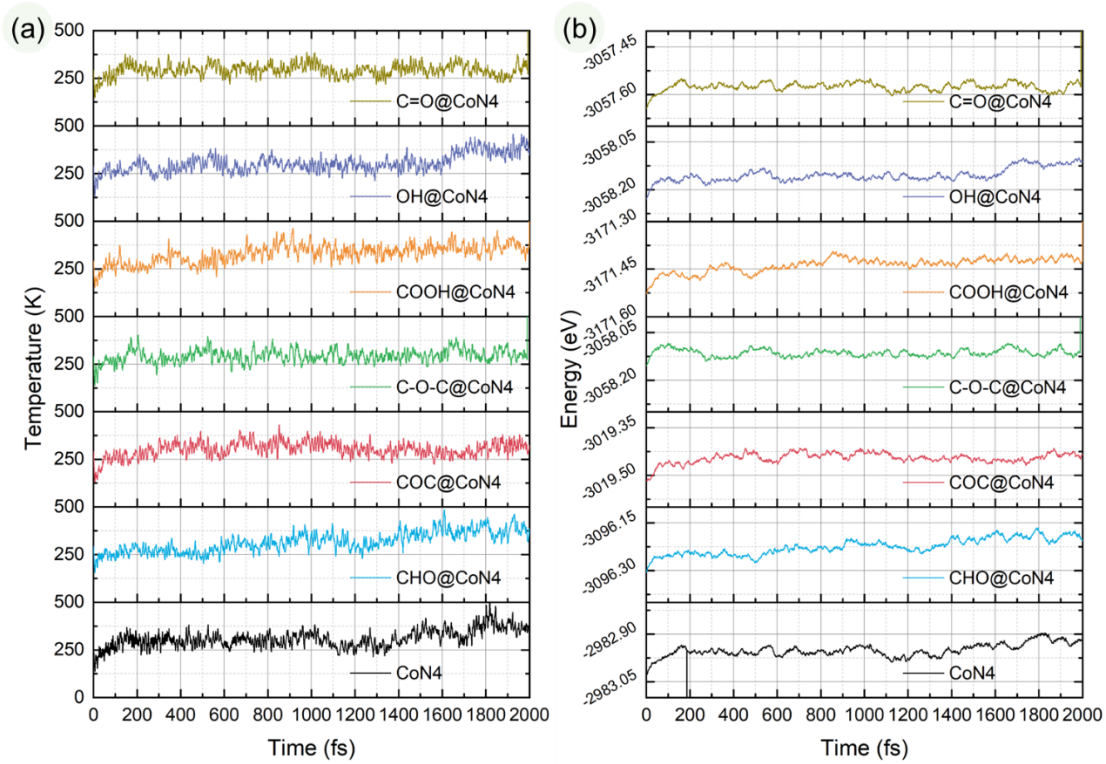

Figure S8. The variations of (a) temperature and (b) energy versus the AIMD simulation time for 2 ps of OGs@CoN_4 under 298.15 K.

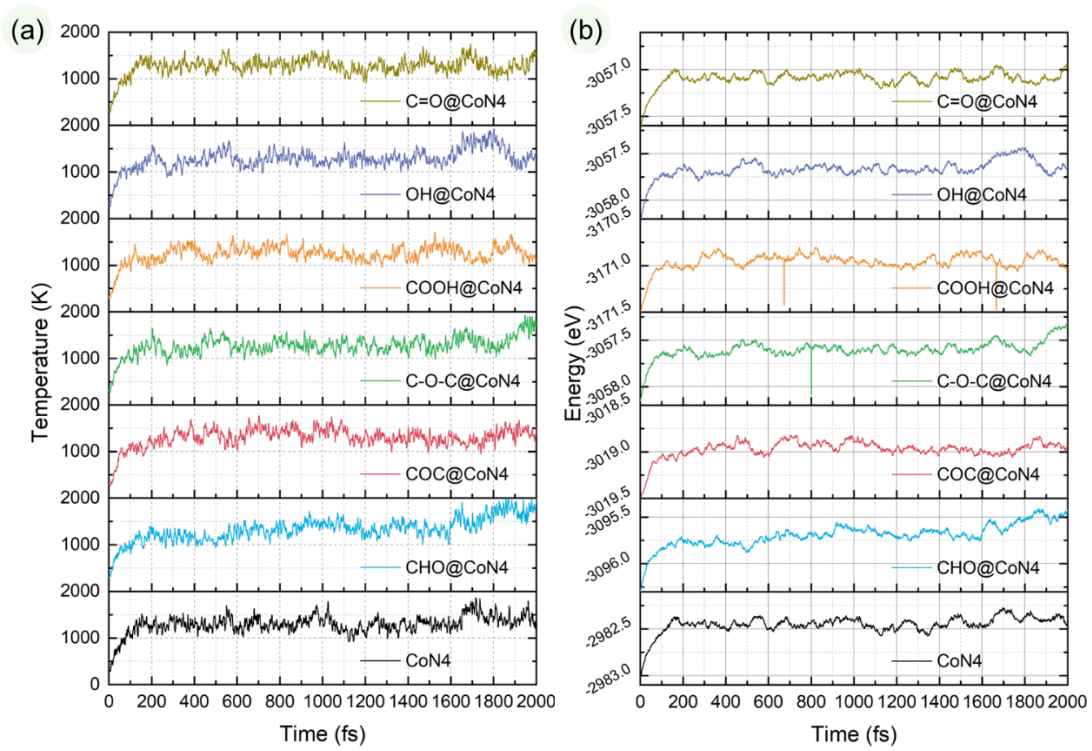

Figure S9. The variations of (a) temperature and (b) energy versus the AIMD simulation time for 2 ps of OGs@CoN_4 under 1298.15 K.

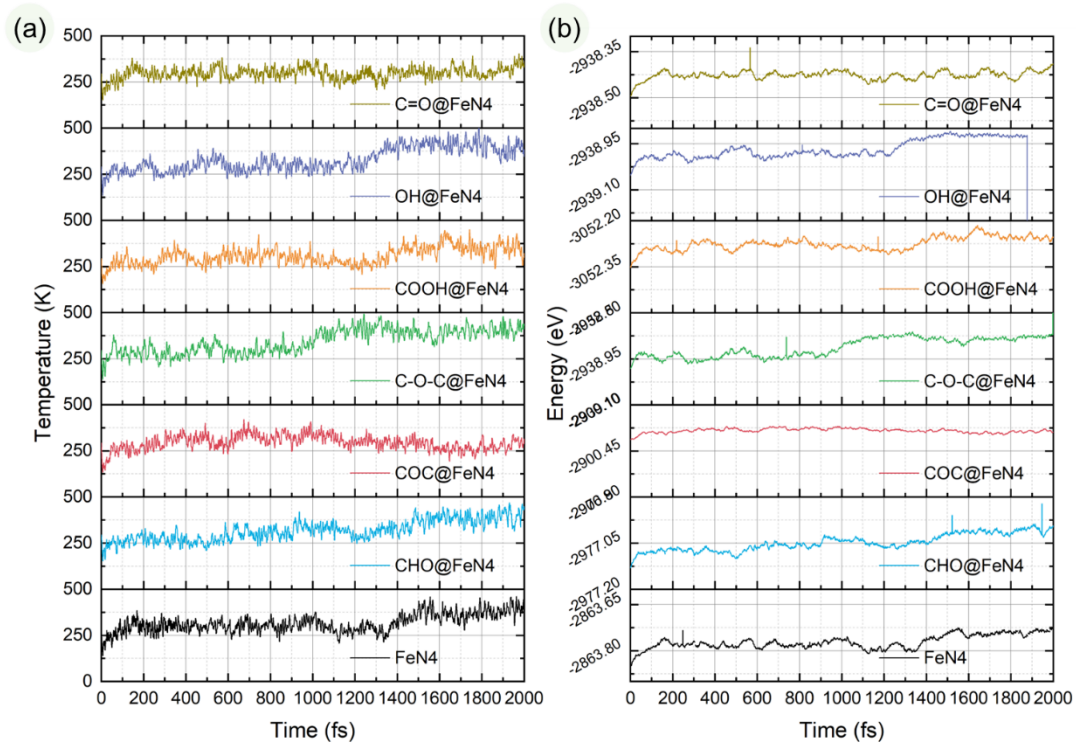

Figure S10. The variations of (a) temperature and (b) energy versus the AIMD simulation time for 2 ps of OGs@FeN_4 under 298.15 K.

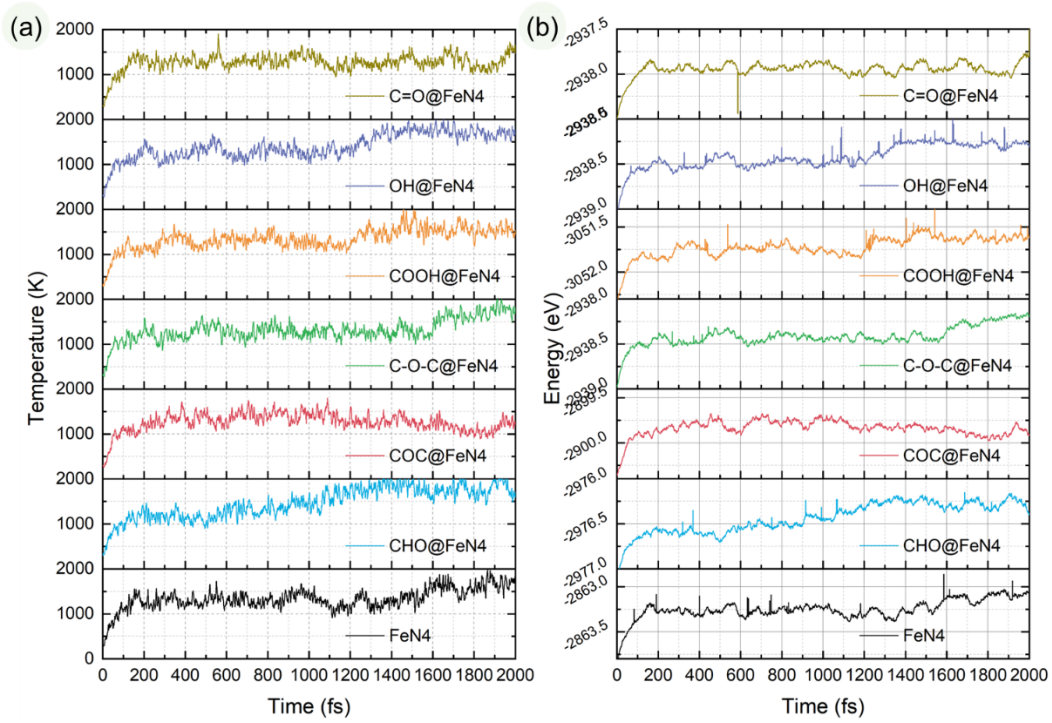

Figure S11. The variations of (a) temperature and (b) energy versus the AIMD simulation time for 2 ps of OGs@FeN_4 under 1298.15 K.

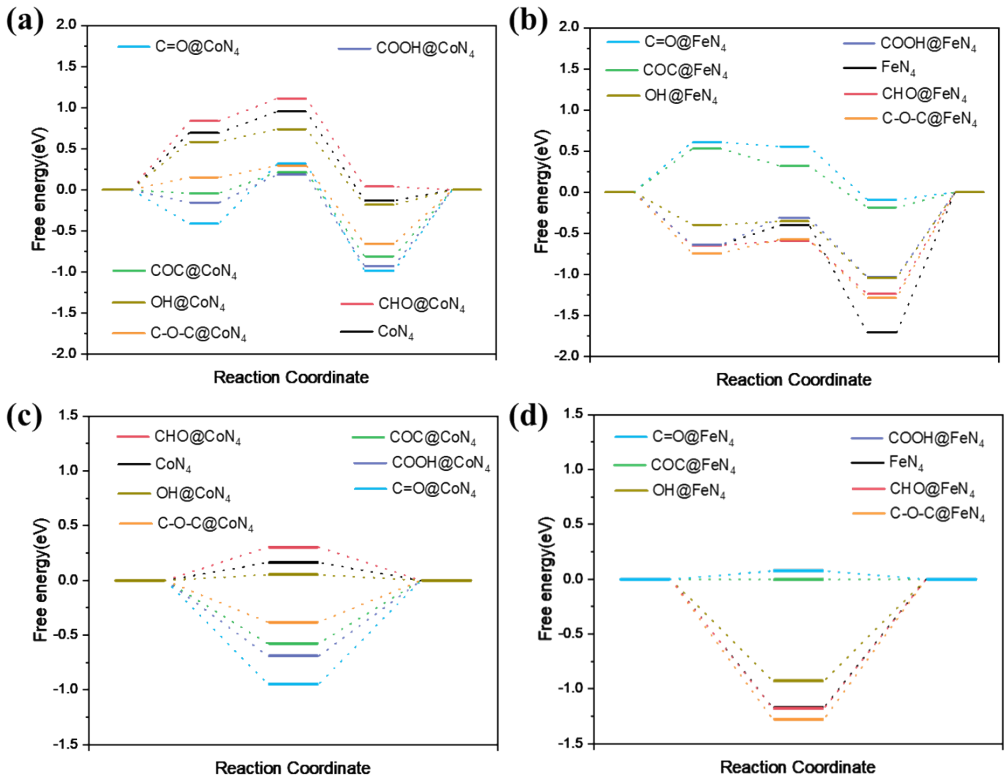

Figure S12. The free energy diagram of 4eORR and 2eORR pathway on OGs@CoN_4 and OGs@FeN_4 in implicit water solvent.

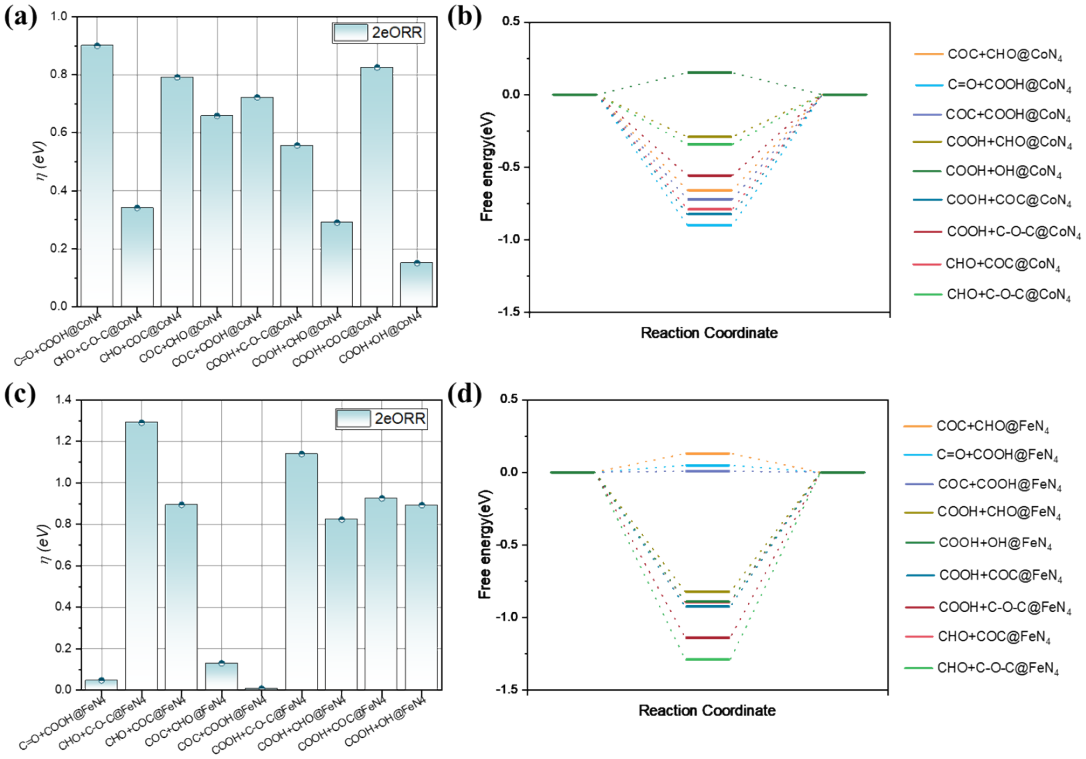

Figure S13. The overpotential and free energy diagram of 2eORR pathway on double doped OGs@CoN_4 and OGs@FeN_4 in implicit water solvent.

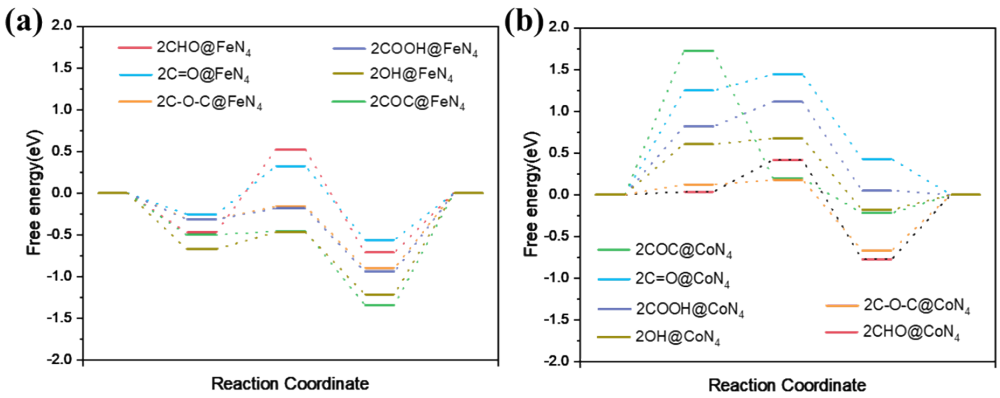

Figure S14. The free energy diagram of 2eORR pathway on double doped OGs@CoN_4 and OGs@FeN_4 in implicit water solvent.

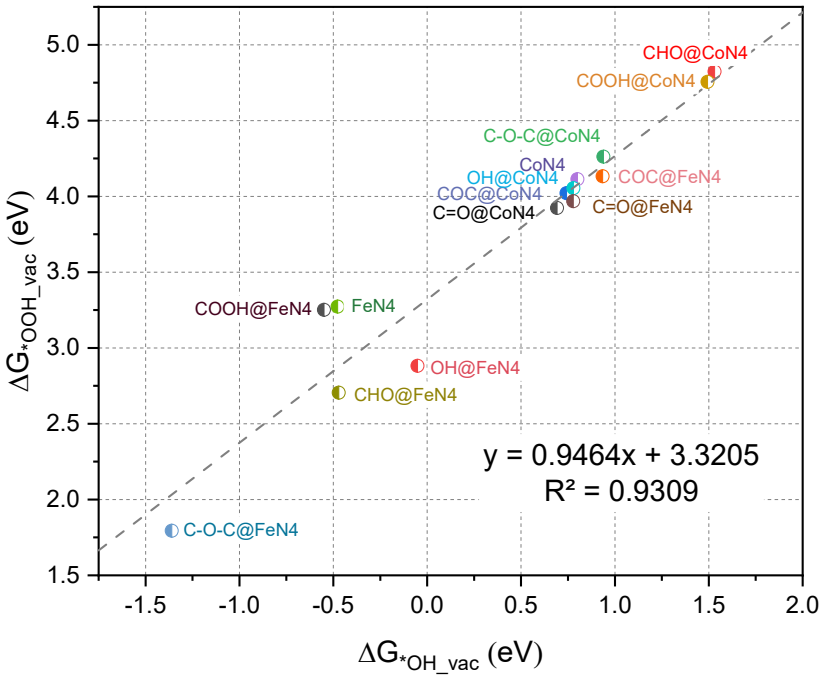

Figure S15. Linear correlation between ΔG_{*OOH_vac} and ΔG_{*OH_vac}.

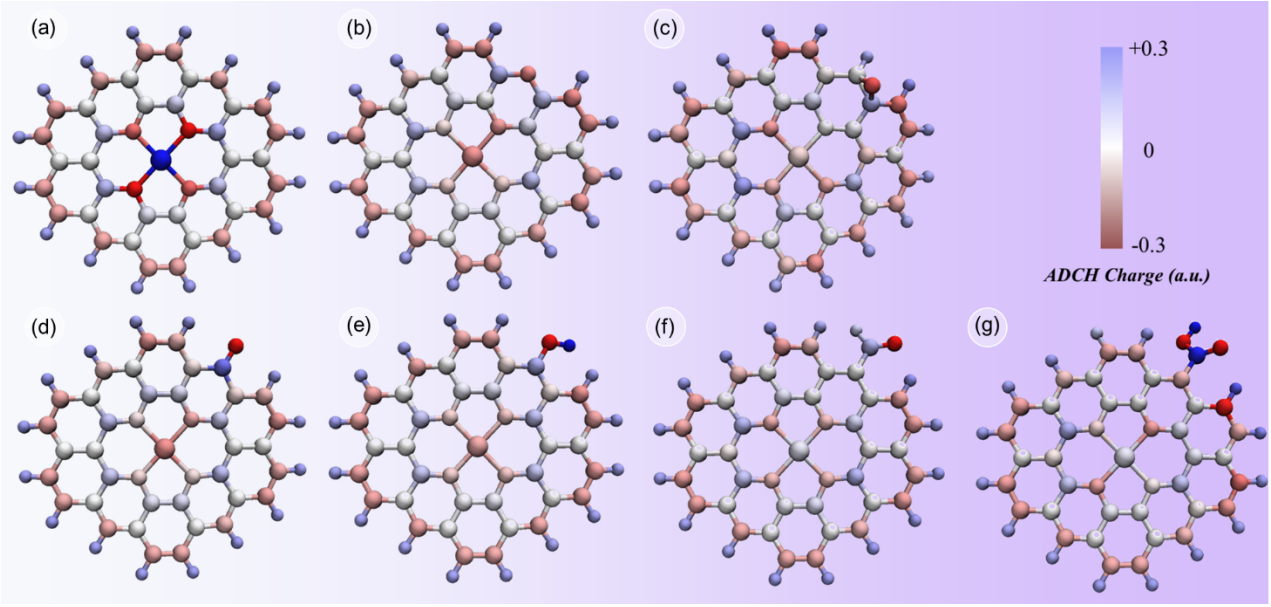

Figure S16. Charge distribution of single doped OGs@CoN_4. (a) CoN_4, (b) COC@CoN_4, (c) C-O-C@CoN_4, (d) C=O@CoN_4, (e) OH@CoN_4, (f) CHO@CoN_4 and (g) COOH@CoN_4.

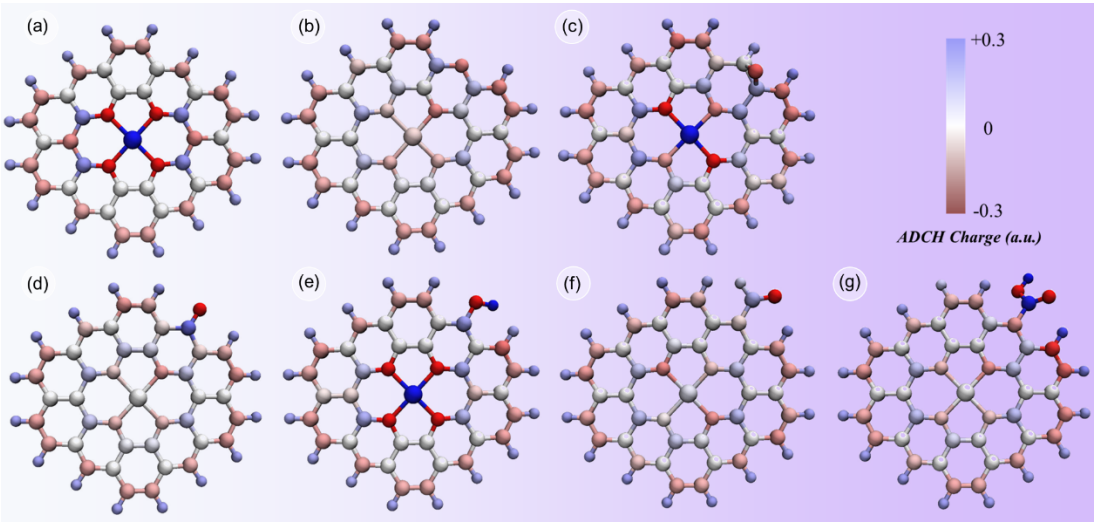

Figure S17. Charge distribution of single doped OGs@FeN_4. (a) FeN_4, (b) COC@FeN_4, (c) C-O-C@FeN_4, (d) C=O@FeN_4, (e) OH@FeN_4, (f) CHO@FeN_4 and (g) COOH@FeN_4.

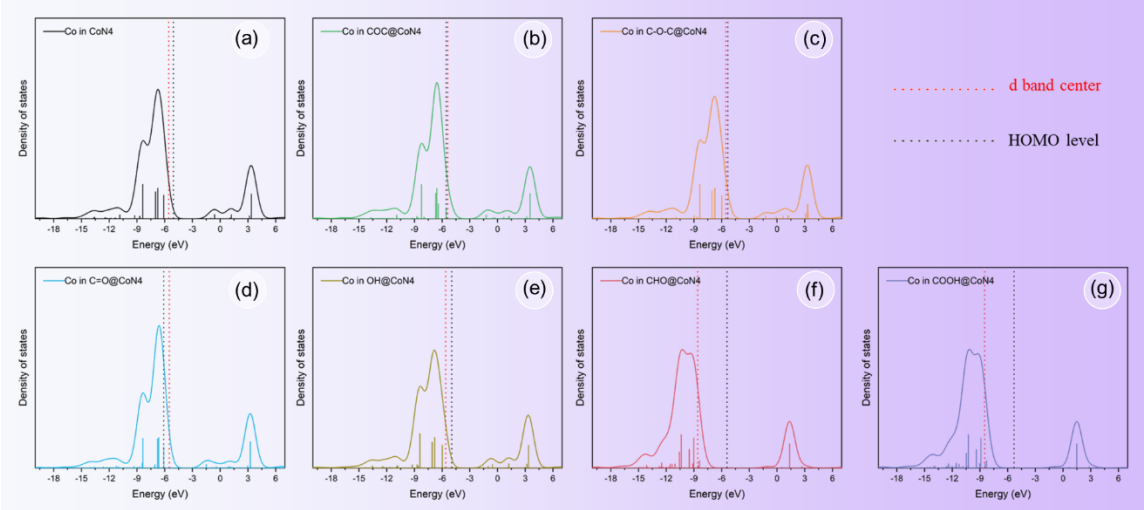

Figure S18. D band center of single doped OGs@CoN_4. (a) CoN_4, (b) COC@CoN_4, (c) C-O-C@CoN_4, (d) C=O@CoN_4, (e) OH@CoN_4, (f) CHO@CoN_4 and (g) COOH@CoN_4.

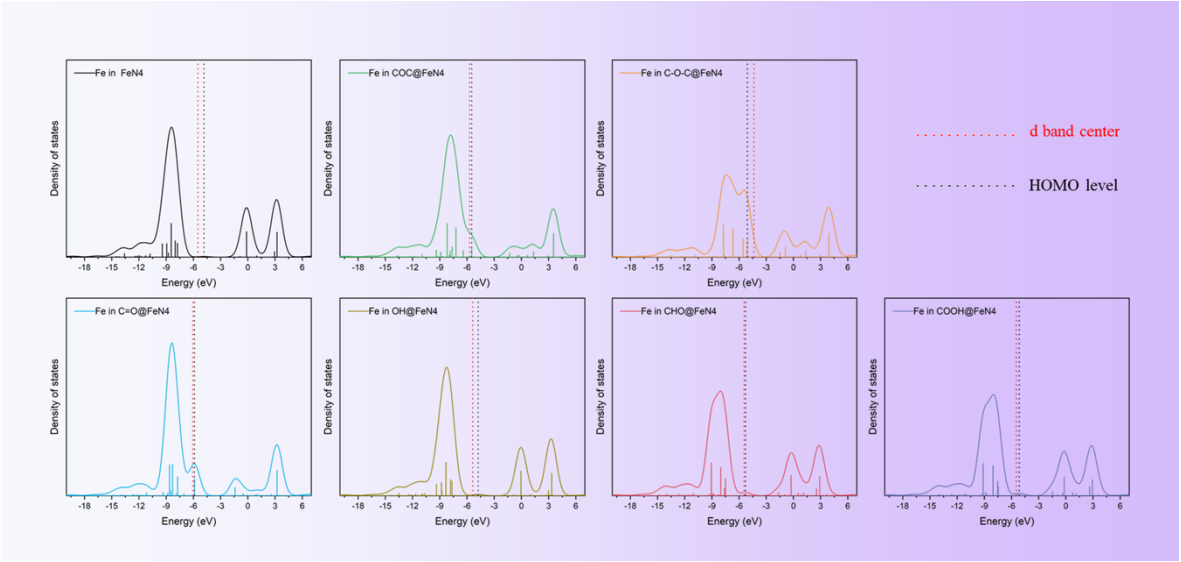

Figure S19. D band center of single doped OGs@FeN_4. (a) FeN_4, (b) COC@FeN_4, (c) C-O-C@FeN_4, (d) C=O@FeN_4, (e) OH@FeN_4, (f) CHO@FeN_4 and (g) COOH@FeN_4.

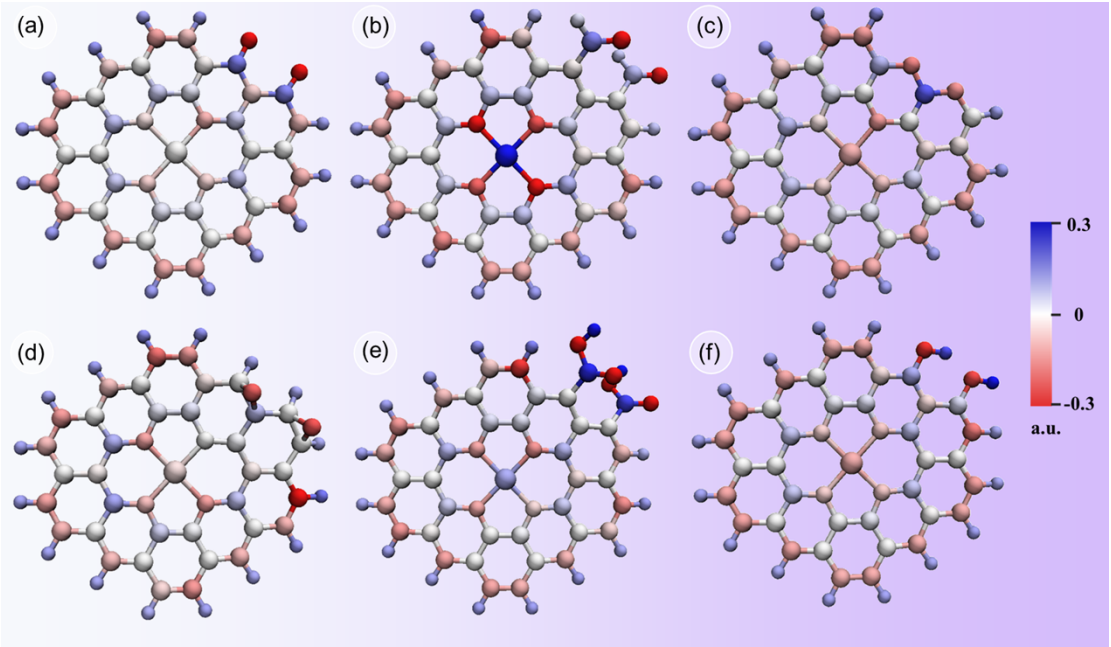

Figure S20. Charge distribution of double doped OGs@CoN_4. (a) 2C=O@ CoN_4, (b) 2CHO@ CoN_4, (c) 2COC@ CoN_4, (d) 2C-O-C@ CoN_4, (e) 2COOH@ CoN_4, and (f) 2OH@ CoN_4.

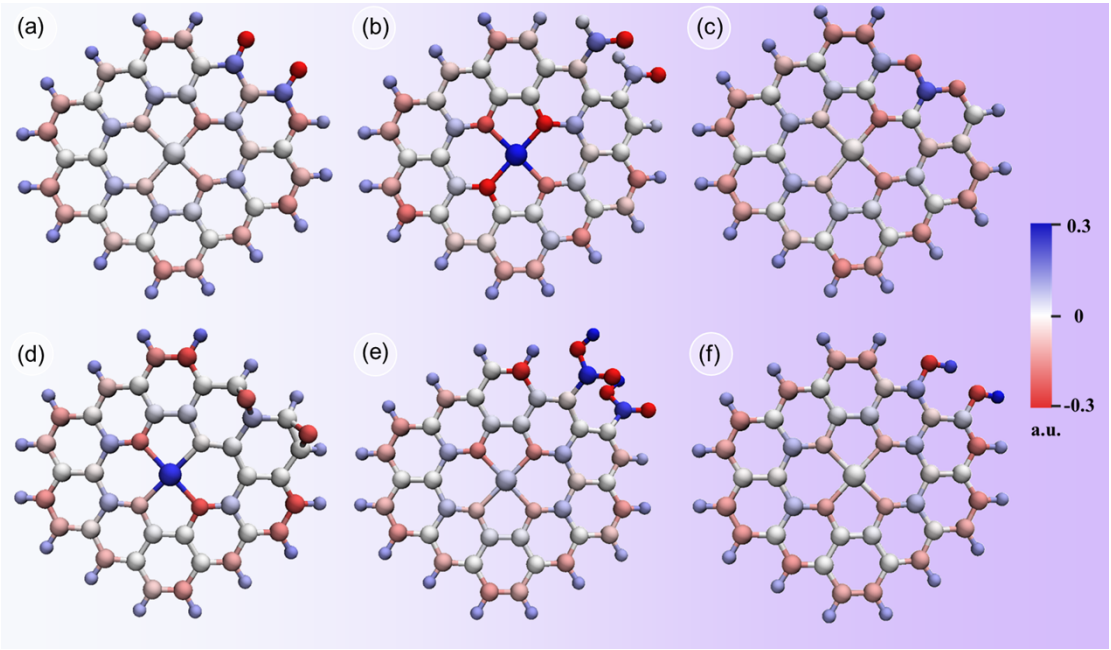

Figure S21. Charge distribution of double doped OGs@FeN_4. (a) 2C=O@ FeN_4, (b) 2CHO@ FeN_4, (c) 2COC@ FeN_4, (d) 2C-O-C@ FeN_4, (e) 2COOH@ FeN_4, and (f) 2OH@ FeN_4.

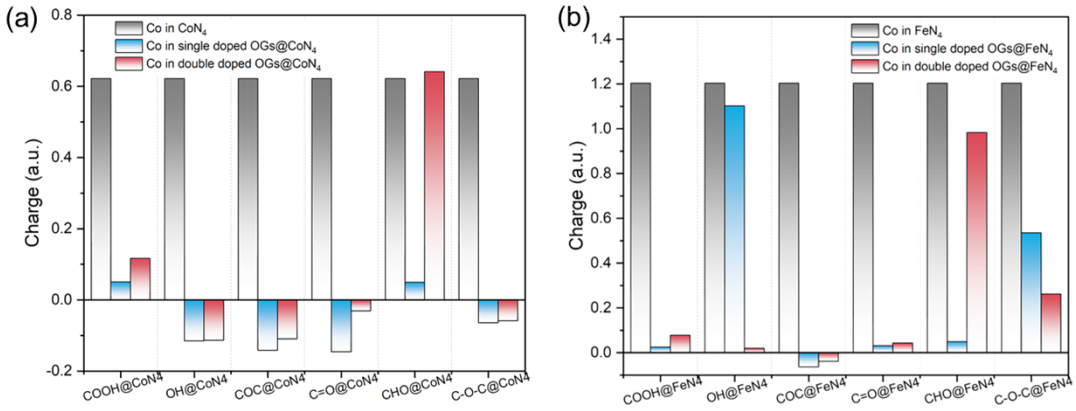

Figure S22. Charge of (a) Co in CoN_4, single doped OGs@CoN_4, double doped OGs@CoN_4, and (b) Fe in FeN_4, single doped OGs@FeN_4, double doped OGs@FeN_4.

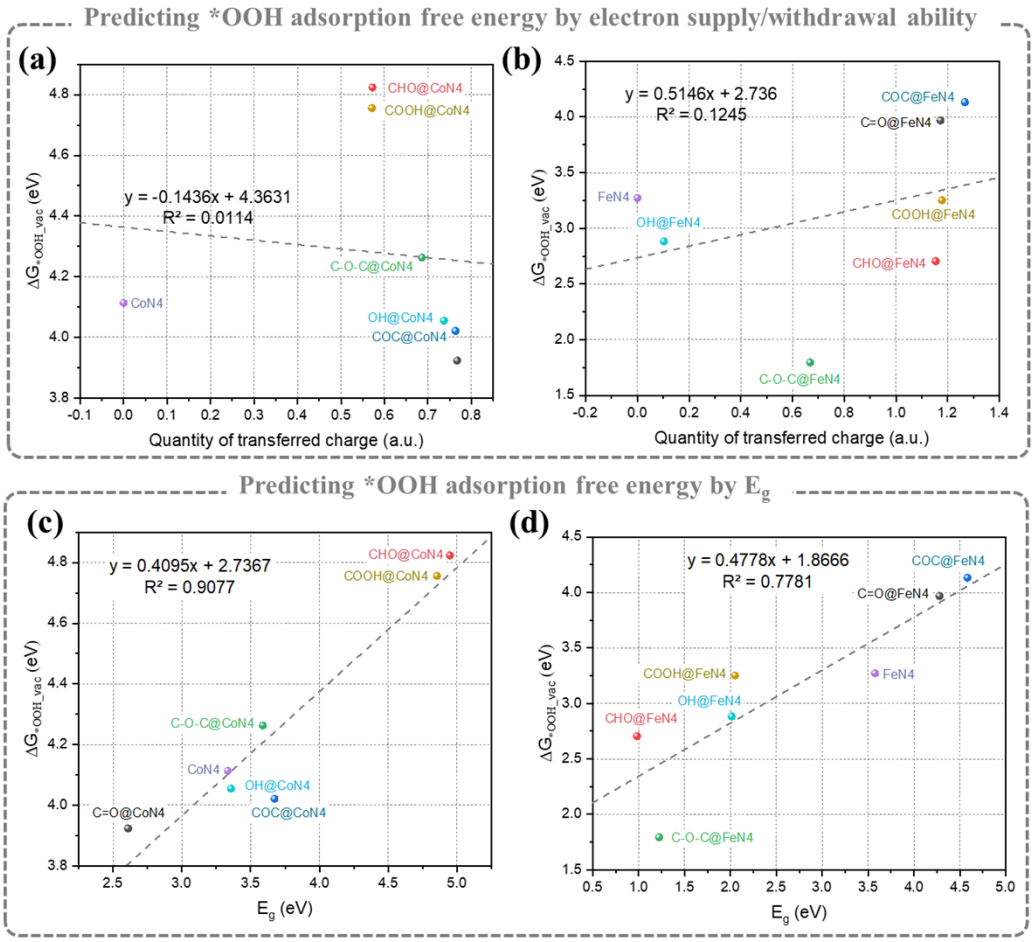

Figure S23. The conventional charge based method and the E_g based method proposed in this study were used to predict *OOH adsorption free energy.

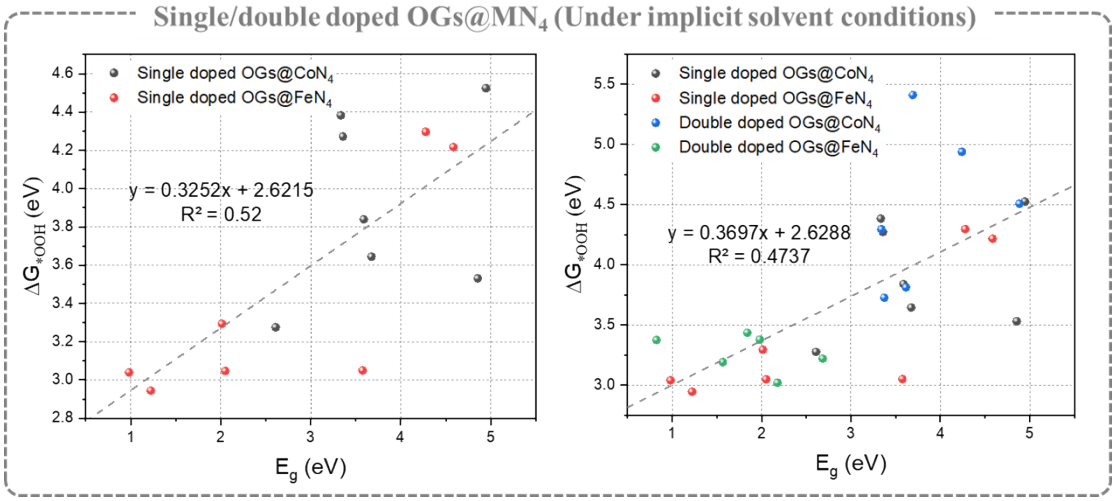

Figure S24. The correlation between the calculated OOH adsorption free energy and the fundamental gap under implicit solvent conditions.