A generalized model for estimating adsorption energies of single atoms on doped carbon materials

Maria G. Minotaki[1,2], Julian Geiger[1,2], Andrea Ruiz-Ferrando[1,2], Albert Sabadell-Rendon[1,*], and Núria López[1,*]

[1]*Institute of Chemical Research of Catalonia (ICIQ), The Barcelona Institute of Science and Technology (BIST), Av. Països Catalans 16, 43007 Tarragona, Spain.*

[2]*Universitat Rovira i Virgili, Av. Catalunya, 35, 43002 Tarragona, Spain.*

[*]Corresponding author(s): Núria López, nlopez@iciq.es

March 20, 2024

S1 Supporting Notes

Note S1: Interquartile Range (IQR)

The interquartile range (IQR) is a statistical measure of the dispersion of the data and it is defined as the difference between the 75th and the 25th quartile, IQR = Q3 - Q1. We consider a datapoint as an outlier when its value is below Q1 -1.5 * IQR or above Q3 +1.5 * IQR. We applied IQR for the adsorption energy of the metal-doped and cavity-modified carbon systems. This method identified outliers with large cavity structural rearrangements.

Note S2: Performance Metrics

In assessing the performance of machine learning models, we employ key performance metrics, including the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE). These metrics serve as valuable indicators of the model's predictive accuracy and provide quantitative measures of the variance explained, the average magnitude of errors, and the overall deviation between predicted and actual values.

R^2 is a statistical measure that assesses the proportion of the variance in the dependent

variable explained by the independent variables in a regression model. It is calculated using the formula:

$$\mathrm{R}^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{S1}$$

Here, SS_{res} represents the residual sum of squares: $SS_{res} = \sum_{i=1}^{n}(y_i - y_i')^2$, and SS_{tot} denotes the total sum of squares: $SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y})^2$. In these formulas, n represents the total number of data points in the dataset, y_i corresponds to the true value, y_i' represents the predicted value of the sample i of the target variable, and $\bar{y}$ signifies the mean of the true data. The R^2 value, ranging from 0 to 1, serves as a measure of the fitting of a regression model. A value of zero indicates that the model does not explain any of the variance in the data, while a value of one indicates that the model captures all the observed variance.

In the context of K-fold cross-validation, the RMSE (Root Mean Squared Error) is obtained by taking the square root of the MSE (Mean Squared Error):

$$\mathrm{RMSE} = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i}^{n}(y_i - y_i')^2} \tag{S2}$$

The MSE measures the average of the squares of errors, which represent the differences between the predicted (y_i') and actual values (y_i). By considering both the variance and the bias, the MSE provides a comprehensive assessment of the model's performance. The RMSE value is always non-negative, and a value of zero indicates a perfect fit for the data. A lower RMSE signifies a better performance, as it reflects a smaller average error between the predicted and actual values. In the case of adsorption energy (E_{ads}) values, it is desirable to have RMSE values lower than 0.5 eV, indicating accurate predictions.

The final metric utilized for evaluation was the MAE (Mean Absolute Error), which measures the average absolute difference between the predicted values (y_i') and the actual values (y_i). It represents the average magnitude of the errors, disregarding their direction. The MAE is obtained by calculating the mean of the absolute differences:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i}^{n}|y_i - y_i'| \tag{S3}$$

The MAE directly represents the average error in the same units as the target variable in a regression model. Lower MAE values indicate a better model performance, as it signifies smaller average deviations between the predicted and actual values.

Note S3: Bootstrap Aggregating and Hyperparameter Tuning in Random Forest Regression

During the training process of the model, we employed bootstrap aggregating to enhance its performance. This technique involves repeatedly sampling from the training data with replacement. By creating multiple subsets of the data, each containing some duplicated and some omitted samples, we ensured robustness and mitigated the risk of overfitting. For each iteration, a regression tree was trained on a different subset. To make predictions for unseen samples, the model combined the predictions from all the trees on unseen data ($f_b(x')$) by averaging them for B times according to the number of subsets. This averaging process is represented by the equation:

$$\hat{\mathrm{f}} = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \tag{S4}$$

Bootstrap aggregating helps improve the random forest regression's generalization capabilities and overall performance.

To optimize the performance of the random forest regression model, several crucial hyperparameters were carefully tuned. The number of trees in the forest was set to 128, striking a balance between improved performance and computational efficiency. The maximum depth of each tree was limited to 8, ensuring a controlled level of complexity and mitigating the risk of overfitting. To promote diversity in tree construction, 40% of the available features were randomly selected for each split. By diligently adjusting these hyperparameters and

analyzing their impact on model performance, we successfully achieved improved accuracy and generalization capabilities for our specific problem. Supplementary Figures S4 to S6 provides additional insights into the fine-tuning process.

Note S4

The iterative adjustment process is explained in Figure S4. First, we constructed the data set, a pool of pristine carbon materials with cavities of various sizes and geometries including N, S, and P as different anchoring atoms, and several single metal atoms. Then, we extracted the features and cleaned the DFT data using the interquartile Range (IQR) to remove outliers from the dataset. The generated chemical space is deposited in ioChem-BD. Next, we applied the data preprocessing utilizing the SAC labels of the cleaned data and standardizing the feature space. Afterwards, we trained the different models Gaussian Process Regression (GPR), Support Vector Regression (SVR), Random Forest Regression (RFR), and Bayesian Machine Scientist (BMS). Then, we included as subroutines the 5-fold cross-validation, the hyperparameter tuning, and the Sequential Feature Selection. Finally, we evaluated the model according to the R-squared (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) metrics. RFR and BMS exhibit the most robust performance, and reported the optimal values for the Metal-Leave-One-Group-Out Cross-Validation (M-LOGOCV) for SACs design.

Note S5: Feature Sets

According to Supplementary Table S8, the initial feature space consists of six sets. Set 1 comprises descriptors of the single metal atoms obtained directly from the literature. We included 20 descriptors proposed by O'Connor et al[1], such as Pauling's electronegativity (χ_P), electron affinity (E_{ea}), and Miedema parameters $\eta^{1/3}$ and ϕ. Additionally, we expanded Set 1 by incorporating the d-band center of the metal in the bulk phase (ϵ_d), magnetic moment of the d orbitals (μ_d), and Fermi energy (ϵ_F) of the bulk phase metal.

Set 2 incorporates cavity properties, including the coordination number (n^{cc}), the number of heteroatoms (N, P, S) (N_X), the number of carbon atoms in the anchoring sites of the cavity (N_C), the fraction of carbon atoms (f_C), the fraction of heteroatoms (f_X), the number of heteroatoms in the pyridine ring (N_{pyrid}), the fraction of heteroatoms in the pyridine ring (f_{pyrid}), the number of heteroatoms in the pyrrole ring (N_{pyrr}), and the fraction of heteroatoms in the pyrrole ring (f_{pyrr}). Additionally, we included descriptors obtained after geometry optimization of the pristine cavity structures (without the single metal atom). These descriptors encompass the minimum and maximum strain effect ($\varepsilon_{\min}$ and $\varepsilon_{\max}$), the Fermi energy (ϵ_F^{cc}), and the surface area of the pristine of the carbon cavity (S^{cc}).

To determine the electronic occupancies near the Fermi level for the pristine cavities of the SACs within doped carbon materials (ϵ_F^{cc}), we utilized Gaussian smearing. This approach was chosen to effectively account for the intricate metallic behavior of the graphene monolayer when modified to form SACs over doped carbons. The smearing width was tuned to 0.05 eV, a value suitable for accurately capturing the electronic states in the vicinity of the Fermi level and smoothing out numerical artifacts from highly localized energy distributions.

The anchoring heteroatoms (N, S, P) and carbons outline the region on the geometry-optimized pristine cavities. Thus, we consider the surface area of the cavity of the SACs in doped carbons as an open surface, for structures prior to the introduction of the metal center. The prevalence of trigonal and tetragonal geometries among the studied cavities, and the requirement for a universal descriptor applicable to cavities of complex geometries, suggest the employment of the geometrically versatile formula of the surface area of regular polygon for this characterization, Equation S5:

$$S^{cc} = \frac{1}{2} P\alpha = \frac{1}{4n} P^2 cot(\frac{\pi}{n}) \tag{S5}$$

The surface area S^{cc} of the pristine cavity is defined based on the perimeter P and the apothem α of the polygon. The apothem α indicates the distance from the center of the polygon to the midpoint of a side. Considering the variations stemming from side lengths and

post-relaxation geometries, we utilize the perimeter term, $\mathrm{P} = \sum_{i=1}^{n} l_n$, where l_n the individual side lengths, extracted using the ASE package[2]. The parameter n denotes the number of sides in the polygon. For cavities that include the *tr* prefix, n takes the value 3, while for *sq* cavities, n corresponds to 4.

Set 3 includes statistical descriptors of Pauling's electronegativities assigned to atoms (C, N, P, S) located within the cavity region. These features consist of metrics such as the minimum, mean, maximum, standard deviation, and the sum of electronegativities (χ_P). These descriptors highlight the heterogeneity of the atoms in proximity to the cavity and provide valuable insights into their tendencies to withdraw or donate electrons. By quantifying the relative variations in electronegativity among the heteroatoms within the cavity, these features offer valuable information about the electronic characteristics of the system.

Set 4 contains statistical metrics of the Cordero covalent radii (r_{cov}) for heteroatoms (C, N, P, S) within the cavity region. These radii indicate atom sizes and their bonding tendencies. Similarly to set 3, we used statistical metrics to capture the heterogeneity of anchoring atoms, revealing their participation in bonding interactions within the cavity area.

Set 5 and Set 6 focus on the distance measurements between neighboring atoms in the boundary of the geometry-optimized pristine carbon cavity (d_X). Set 5 captures the distances before the geometry optimization process, while Set 6 represents the distances after the optimization. These distances provide insights into the spatial arrangement of atoms within the cavity boundary. Statistical measurements on these distances, provide further characterization of the atomic arrangements. The distances obtained after geometry optimization are specifically denoted with a dagger symbol ($d_X^{\dagger}$).

Note S6: Sequential Feature Selection

To identify the most informative physical and chemical features that accurately represent the single metal atoms, heteroatoms, and cavities of the SACs, we utilized Sequential Feature Selection (SFS) for the K-fold cross-validation of GPR, SVR, and RFR. It is a feature

selection technique that aims to discover the optimal subset of features from a given feature space.

As a wrapper-based method, Sequential Feature Selection (SFS) goes beyond simple individual feature evaluation. It considers the interaction and synergy between features by iteratively adding or removing them based on a metric criterion. This approach allows for a more comprehensive exploration of feature combinations and their impact on the model's performance.

We specifically employed the forward selection variant of the Sequential Feature Selection (SFS) procedure. This approach begins with an empty set and systematically adds one feature at a time to evaluate its impact on the model's performance. The selection process continues until no further improvement in the Root Mean Square Error (RMSE) score is observed. For each machine learning algorithm used in our study, we derived a set of eight features that proved to be most relevant and informative (Supplementary Table S9. By iteratively evaluating different feature subsets, helps in identifying the most informative features, reducing the dimensionality of the data, and improving the model's performance, interpretability, and efficiency.

Note S7: Miedema Model

Miedema semi-empirical model states that the stability of bimetallic alloys is expressed via their heat of formation: $\Delta H = f(c)$ [-Pe$(\Delta\Phi^*)^2$ + Q$(\Delta\ n_{WS})^2$ -R], and involves the electronic charge (e), two constants (P and Q), an unspecified concentration-dependent function (f(c)), and a correction term (R). The first term relates to the difference in chemical potentials ($\Delta\Phi^*$) between the two metals. At the micro-scale, it is understood in terms of electronegativity, while at the macro-scale relates to the experimentally measurable work function. The second term considers the difference in electron density (Δn_{WS}) at the boundaries of the Wigner-Seitz cell between the atoms of the bimetallic alloy and it is a measure of the electron distribution around the individual metals[3].

Note S8: Details on the symbolic model

The symbolic model obtained with the Bayesian Machine Scientist (BMS) is presented in Equation S6. The values of the fitting constants are listed in Table S11. Each one of $i = \{\text{Covalent}, \text{Ionic}, \text{Cross-term I}, \text{Cross-term II}\}$ the terms in Equation S6 was obtained sequentially, subtracting the results of the previous model from the total E_{ads}, leading to $E'_{ads,i} = E_{ads} - \sum_{i}^{N-1} E_{ads,i}^{Predicted}$, and then fitting $E'_{ads,i}$ in the next step[4]. Thus, the total prediction is the sum of all the $E_{ads,i}^{Predicted}$ terms.

We have started including covalent variables only, two for the metal and three for the cavity: the Miedema parameter of the metal, the surface area of the cavity, the empirical covalent radius of the metal, and the sum and the mean of the Cordero radius of the cavity. Starting by the covalent term is not arbitrary: since the BMS tends to not converge for complex datasets, particularly when the variables space is reduced (less than 5 variables), a first ordering of the system was required before considering the ionic contributions (only three variables). Indeed, if the ionic term is used as the first approach, the BMS returned discontinuous functions.

Then, we estimated the ionic term considering one contribution for the metal, the Martynov-Batsanov electronegativity, and two for the cavity, the Fermi level, and the mean Pauling electronegativity. As explained in the main text, this term is extremely convoluted. However, it can be simplified as the $cos\left(cos\left(\epsilon_F\right)^2\right)$ can be approximated to $0.64\epsilon_F^2 + 4.04\epsilon_F + 6.91$ (MAE=0.006 eV, RMSE=0.007 eV, R^2=0.992). Additionally, this term segregates the different cavities according to their mean Pauling electronegativity, depending if they contain C, N, P, or S, Figure S20.

Next, we estimated the first and second cross-terms (CT1 and CT2 respectively) considering all the eight variables identified by the RFR and two extra variables, the coordination number of the metal and the fraction of heteroatoms. Here we observe an almost linear dependence of the span and the dispersion of CT1 with the Miedema parameter for the same metals, Figure S21. The dependence on the cosine of the surface is interpreted as a measure

of the flexibility of the cavity. Finally, regarding CT2, we observe that this term is almost 0 for metals with small covalent radii and Ag, Figure S22.

Finally, cross-validation was performed by finding the constant values that better fit the training sets and then predicting the values for the testing. This process was applied both for the 5-fold and Metal-Leave-One-Group-Out Cross-Validation (M-LOGOCV). The mean values for the metrics and the associated errors are reported in Tables S12 and S13 respectively. The parity plots between the DFT and predicted energies are illustrated in Figure S25.

$$
\begin{aligned}
E_{ads}^{pred} = & \overbrace{\frac{\left(c_1 + c_2^{-c_2 \bar{r}_{cov}^{cc}} S^{cc}\right)\left(-r_{cov}^{sb} + \frac{\Sigma r_{cov}^{cc}\left(\bar{r}_{cov}^{cc} + S^{cc} e^{-\bar{r}_{cov}^{cc}}\right)(c_1 + 2c_3 + S^{cc})}{c_4 + \Sigma r_{cov}^{cc}}\right)}{\bar{r}_{cov}^{cc}}}^{\text{Covalent}} + \\
& \overbrace{\frac{\chi_{MB}\left(\epsilon_F + \left(\left(c_5^2 \cos^2\left(c_6 + e^{\chi_{MB}}\right)\right)^{c_6}\right)^{\frac{(\bar{\chi}_P)^{-\bar{\chi}_P}(c_5+f)}{c_7}} + \frac{\bar{\chi}_P\left(c_7 + \epsilon_F\left(\chi_{MB} + \sin\left(\frac{c_7 \bar{\chi}_P + c_7}{\epsilon_F}\right)\right)\right)}{c_7}\right)}{\cos\left(\cos\left(\epsilon_F^2\right)\right)}}^{\text{Ionic}} + \\
& \overbrace{\frac{(S^{cc})^2\left(c_8\left(\frac{(c_{10} + \Sigma r_{cov}^{cc})^2}{c_{11}(c_{11} + c_{12})^2}\right)^{-S^{cc}}\left(c_9\left(\bar{\chi}_P\right)^3 S^{cc} + \bar{\chi}_P \eta^{1/3}\right) + c_{13}\right)\cos\left(S^{cc}\right)}{c_{14}\left(-c_{11} + c_{14}\epsilon_F - c_{15} - \Sigma r_{cov}^{cc}\right)}}^{\text{Cross term } I} + \\
& \overbrace{\frac{\left(-c_{16} + r_{cov}^{sb} S^{cc}\left(\left(\eta^{1/3}\right)^{\log\left(r_{cov}^{sb}\right)} + n_c\right)\right)\left(c_{17} + e^{\chi_{MB}}\right)\left(\tan\left(\chi_{MB} c_{18}\right) + \frac{Q_H}{c_{18} S^{cc}}\right)}{c_{19}\left(\chi_{MB}^2 - c_{20}\right)}}^{\text{Cross term } II}
\end{aligned}
\tag{S6}
$$

S2 Supplementary Figures

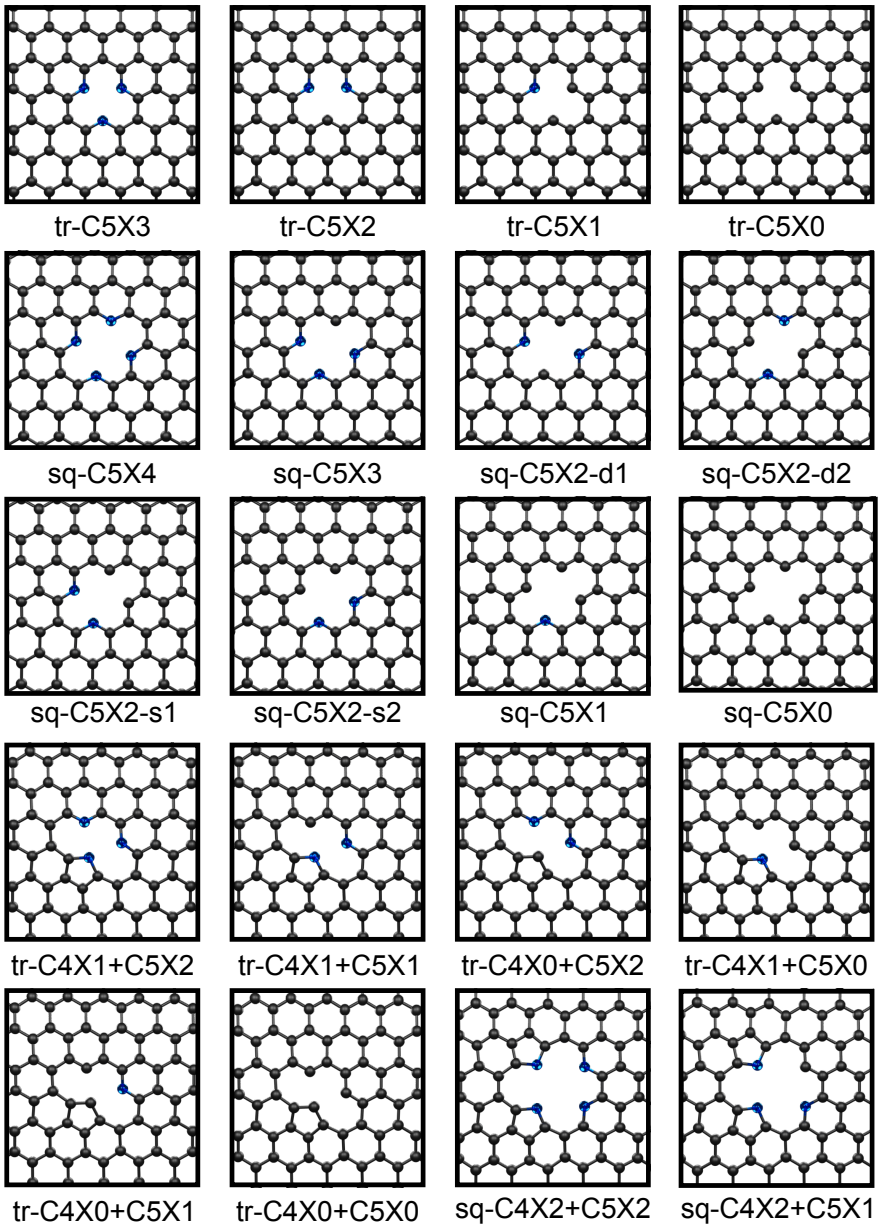

Fig. S1: Schematic representation of the stoichiometric cavity types studied in this work. In these structures, carbon and heteroatom atoms (N, S, P) are denoted with black and blue color, respectively.

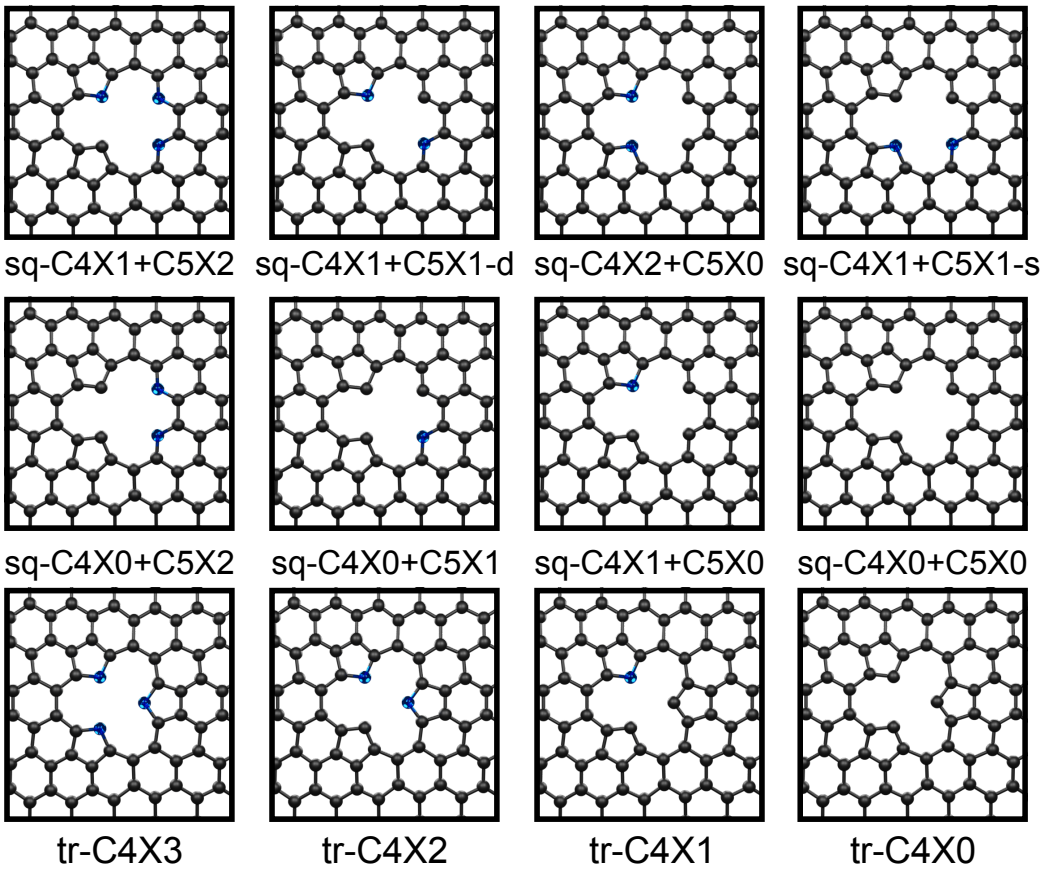

Fig. S2: Schematic representation of the stoichiometric cavity types studied in this work. In these structures, carbon and heteroatom atoms (N, S, P) are denoted with black and blue color, respectively.

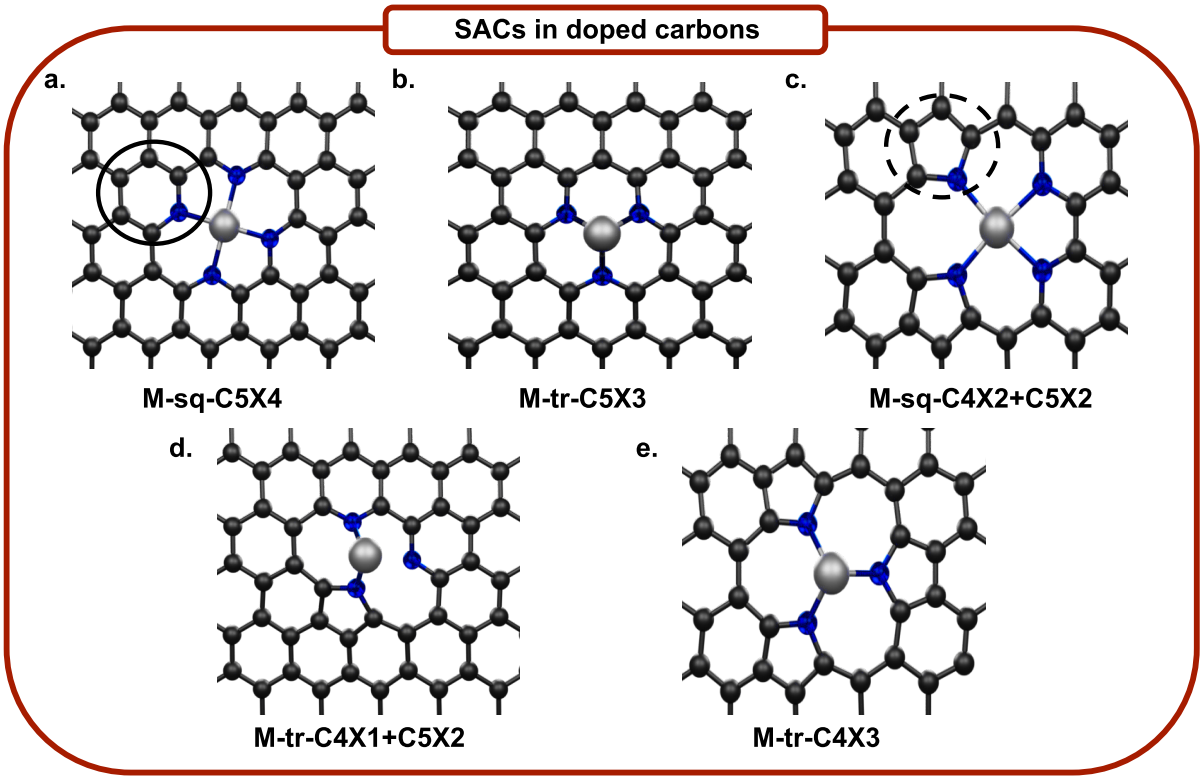

Fig. S3: Key structural configurations of Single Atom Catalysts (SACs) within doped carbon frameworks: a. M-sq-C5X4, b. M-tr-C5X3, c. M-sq-C4X2+C5X2, d. M-tr-C4X1+C5X2, and e. M-tr-C4X3 . In these representations, carbon atoms are depicted in black, anchoring sites (N, S, P, and C) in blue, and the metal center (M) in silver. The pyridinic and pyrrolic rings are distinguished using solid and dashed lines, respectively. The combination of geometric and stoichiometric characteristics with the transition metal centers derives a chemical space of 1153 configurations.

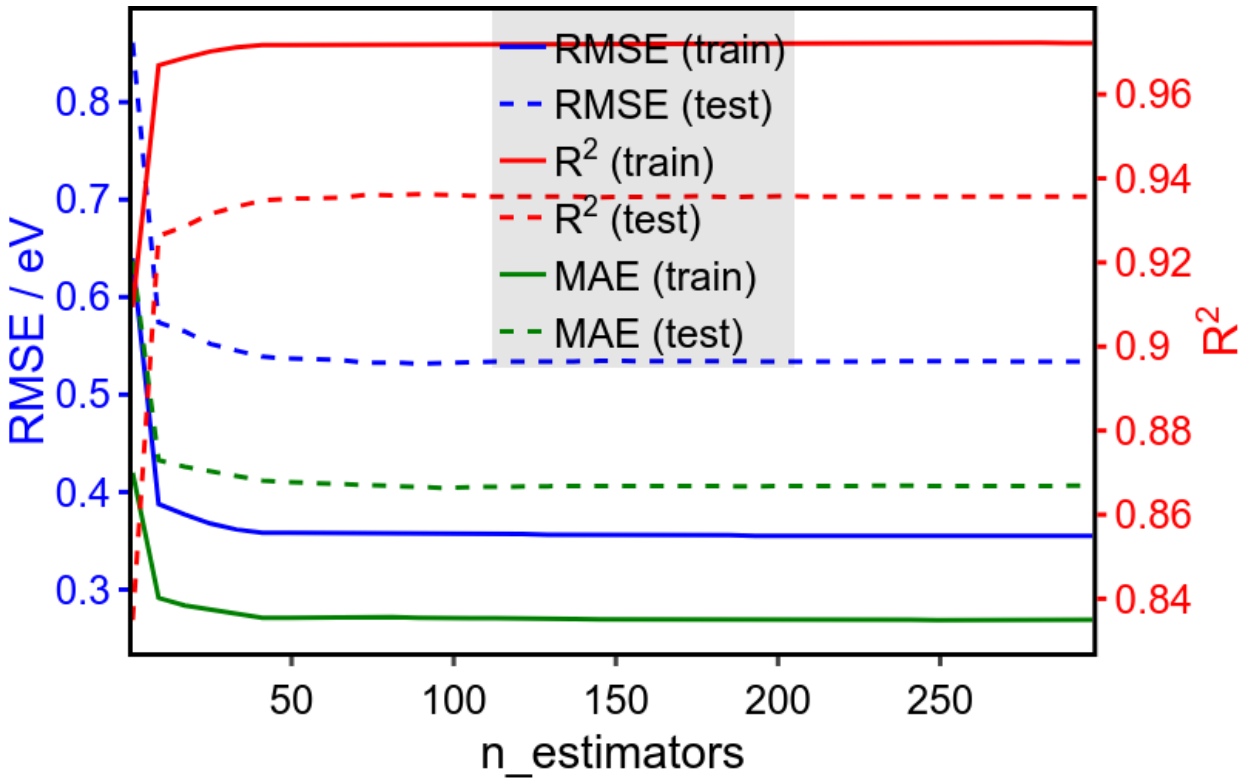

Fig. S4: Hyperparameter Tuning for Random Forest Regression (RFR) model on the number of trees parameters for SACs in carbon cavities.

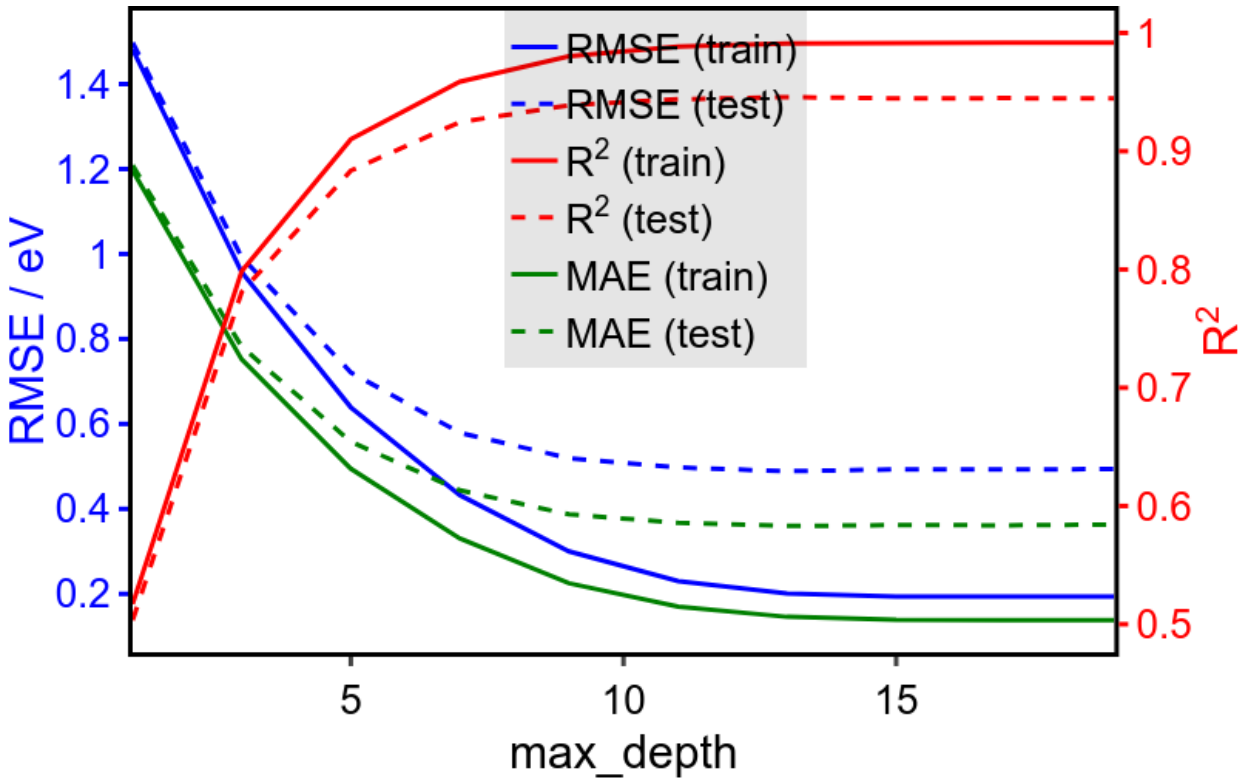

Fig. S5: Hyperparameter Tuning for Random Forest Regression (RFR) model on the depth of trees for SACs in carbon cavities.

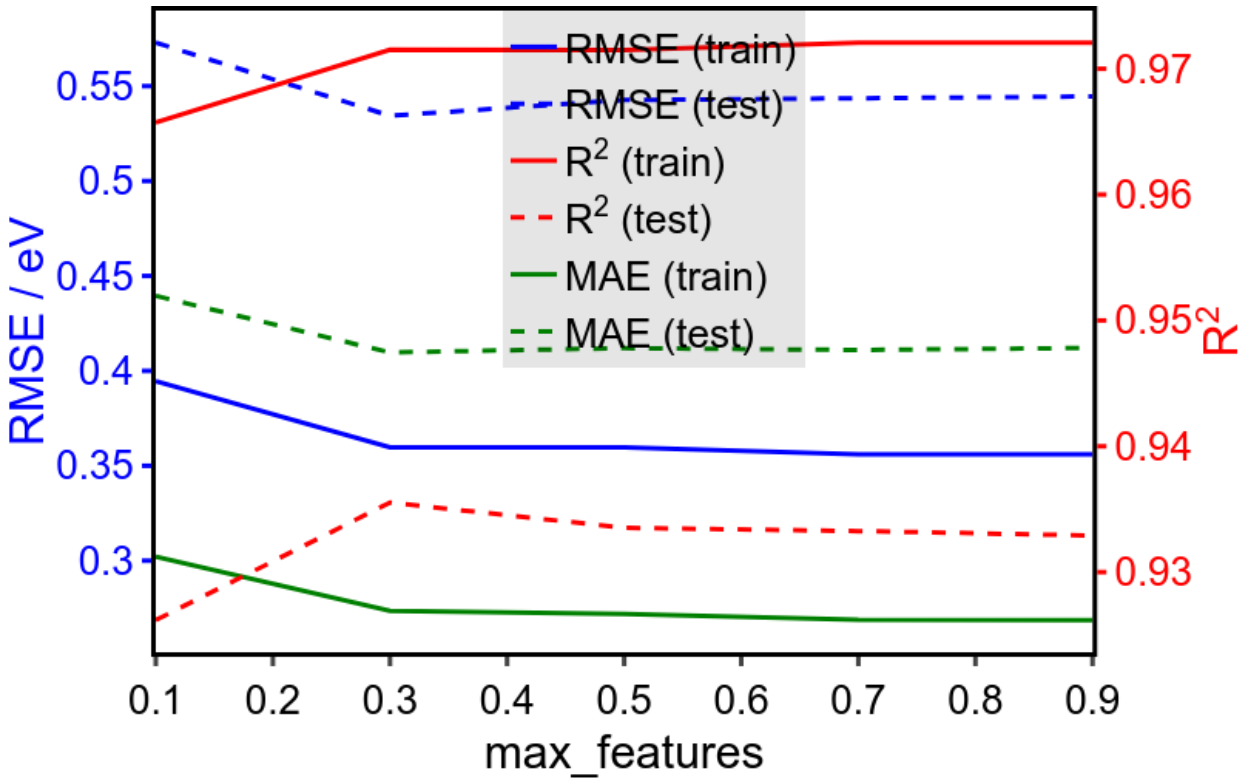

Fig. S6: Hyperparameter Tuning for Random Forest Regression (RFR) model on the feature subset size for SACs in carbon cavities.

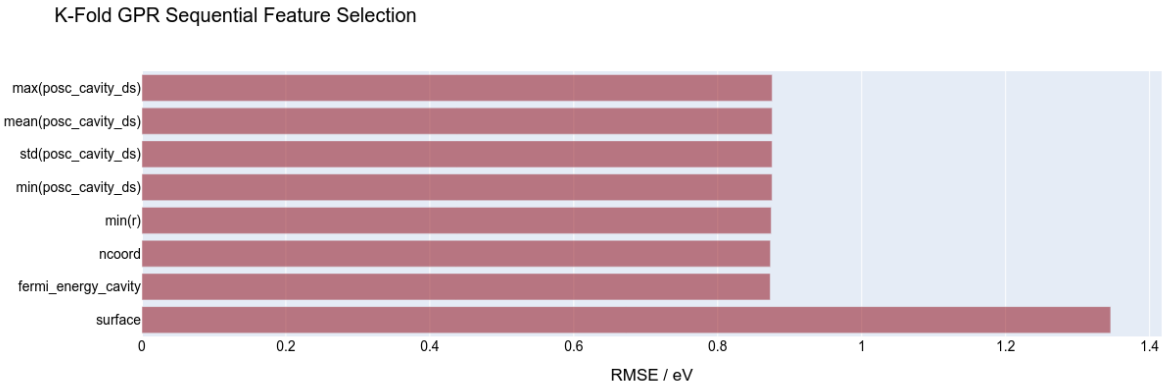

Fig. S7: The courses of RMSE values for GPR models during forward sequential feature selection (SFS), evaluated via five K-fold cross-validations. The R^2, RMSE, and MAE values and the selected primary features are listed in Supplementary Table S4.).

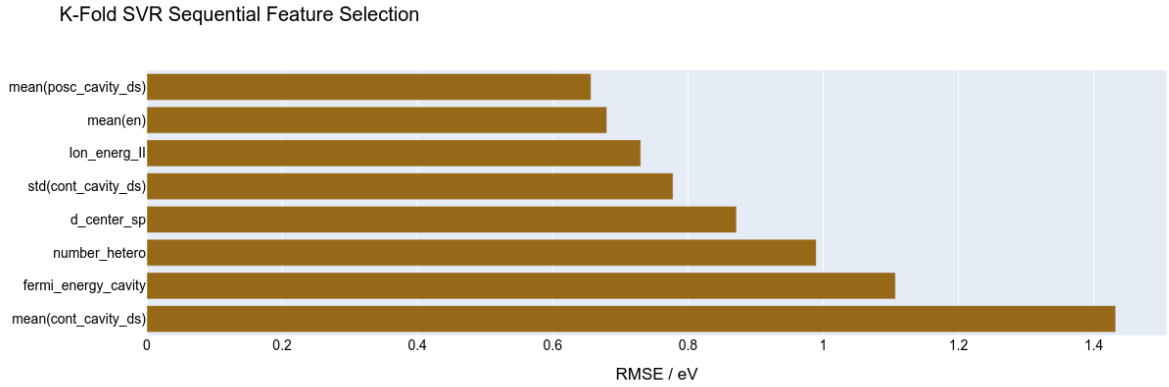

Fig. S8: The courses of RMSE values for SVR models during forward sequential feature selection (SFS), evaluated via five K-fold cross-validations. The R^2, RMSE, and MAE values and the selected primary features are listed in Supplementary Table S4.

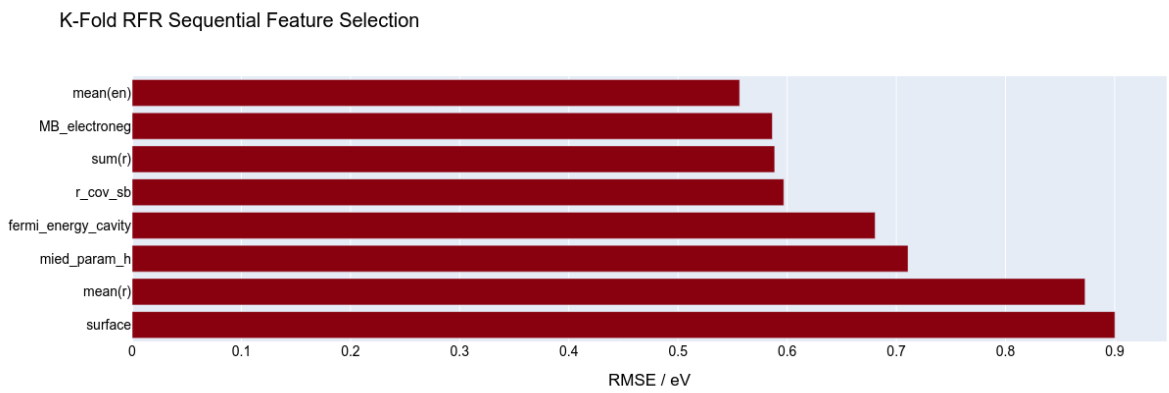

Fig. S9: The courses of RMSE values for RFR models during forward sequential feature selection (SFS), evaluated via five K-fold cross-validations. The R^2, RMSE, and MAE values and the selected primary features are listed in Supplementary Table S4.

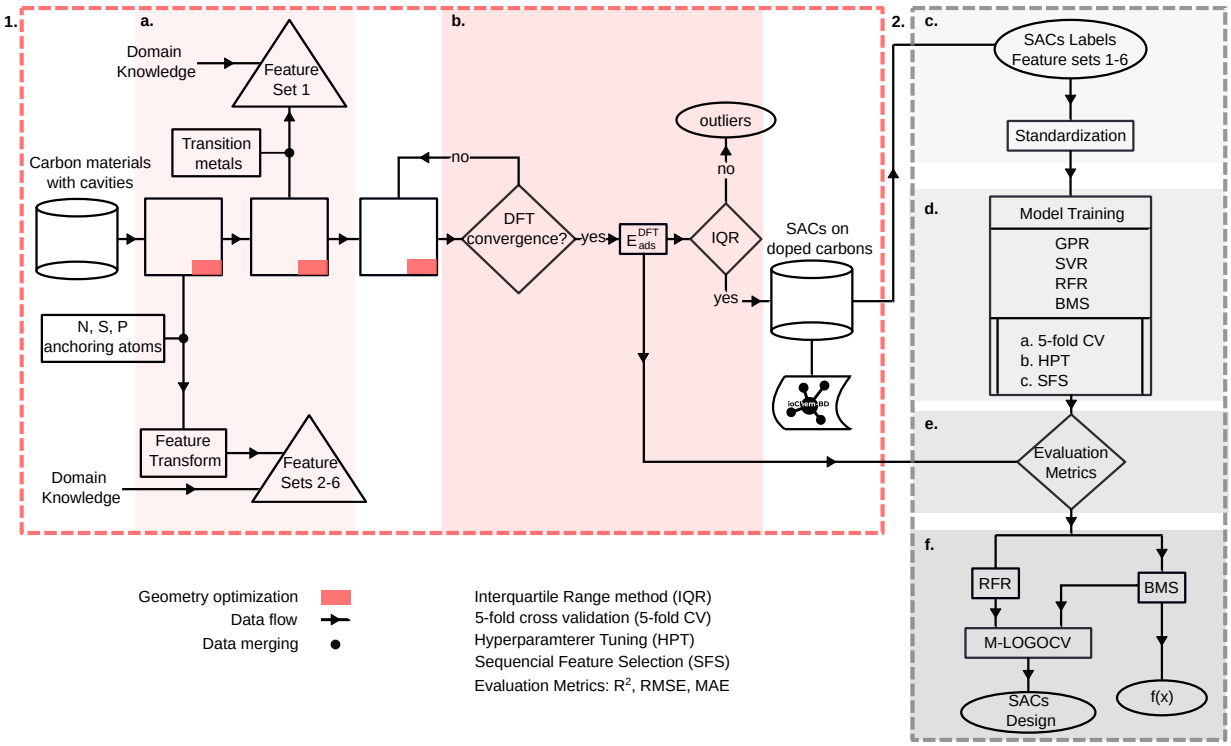

Fig. S10: Flow Chart for SACs on Doped Carbons: The process is divided into two main frameworks: the Theoretical Modeling Framework and the Machine Learning Procedure, enclosed in red and grey frames, respectively. a. Dataset Construction and Feature Extraction b. Data Cleaning Procedure. c. Data Preprocessing. d. Model Training. e. Model Evaluation. f. M-LOGOCV.

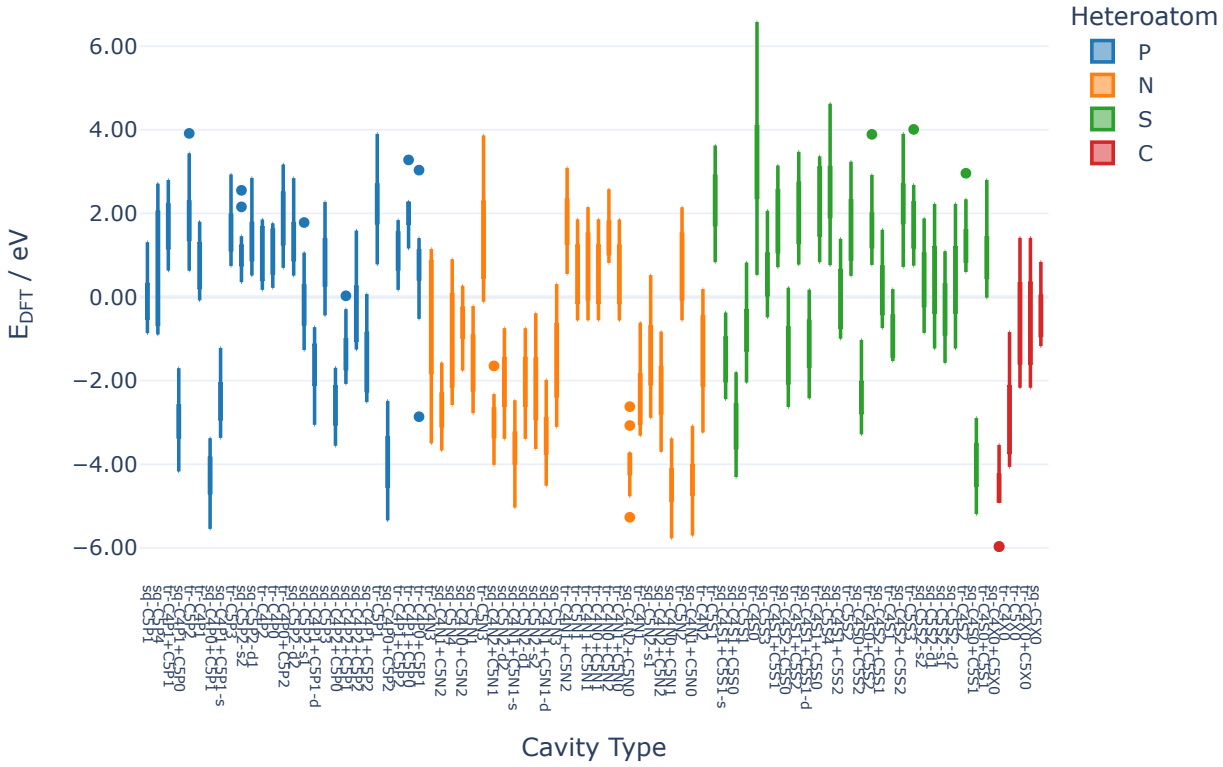

Fig. S11: Box plots illustrating the simulated adsorption energies (E_{DFT}^{ads}) among SACs in carbon cavities, grouped by the stoichiometric cavity type.

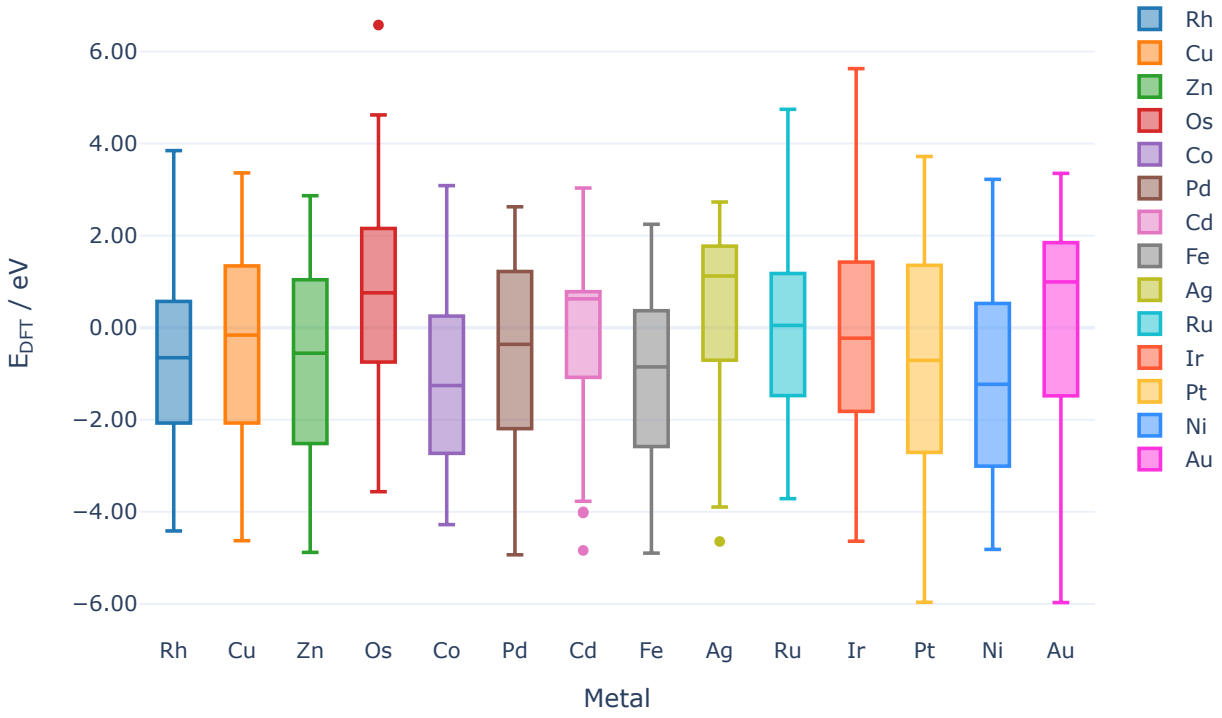

Fig. S12: Box plots illustrating the simulated adsorption energies (E^{ads}_{DFT}) among SACs in carbon cavities, grouped by the type of single metal atom SACs.

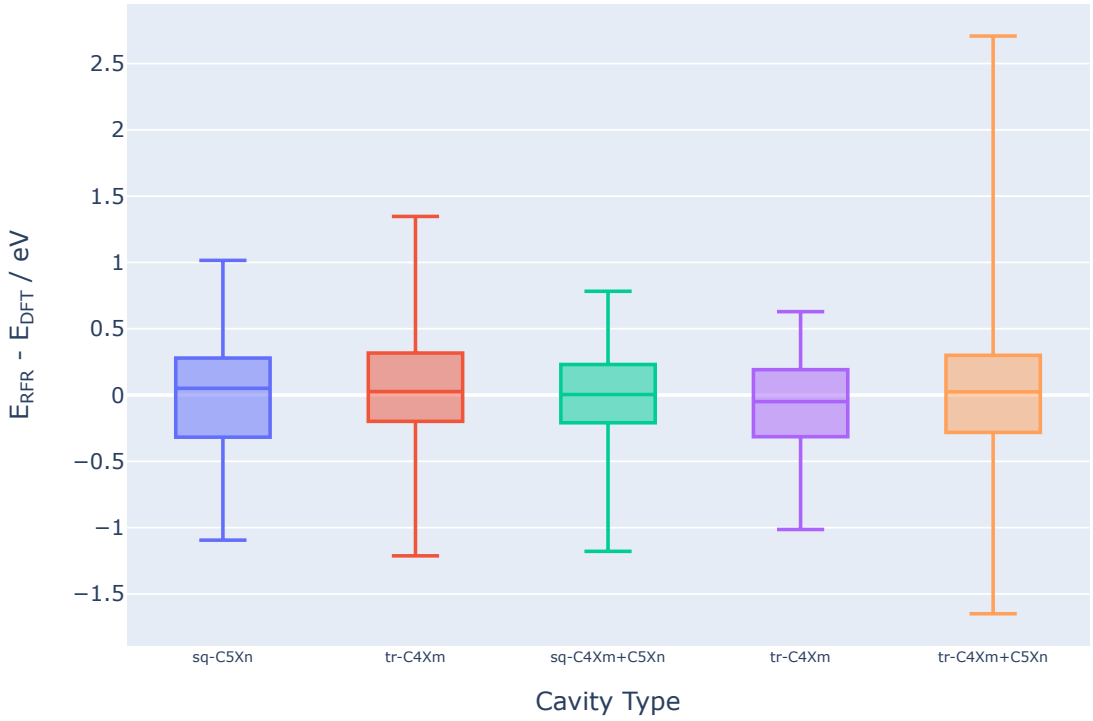

Fig. S13: The error distribution between the predicted energy of the RFR model (E_{RFR}) and the DFT calculated energy (E_{DFT}) grouped by the cavity type on the test set of the 5 K-fold cross-validation model.

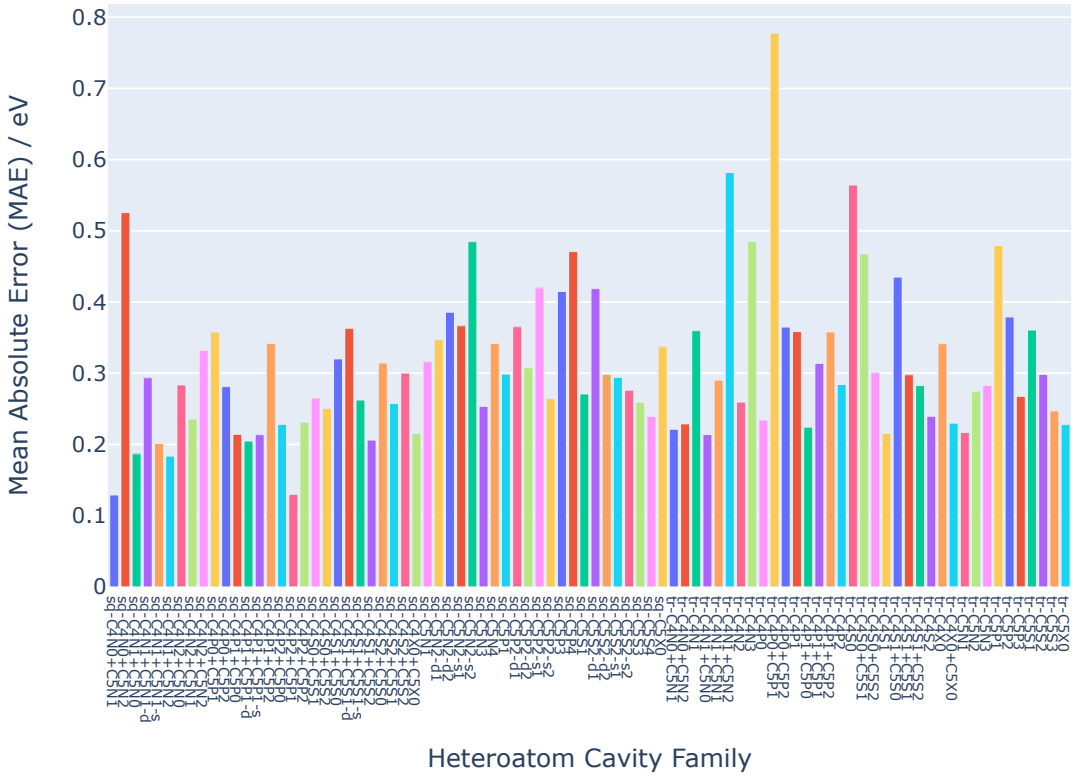

Fig. S14: Mean absolute error grouped by heteroatom cavity family in the test set of the 5-fold cross-validation for the highest performance RFR model.

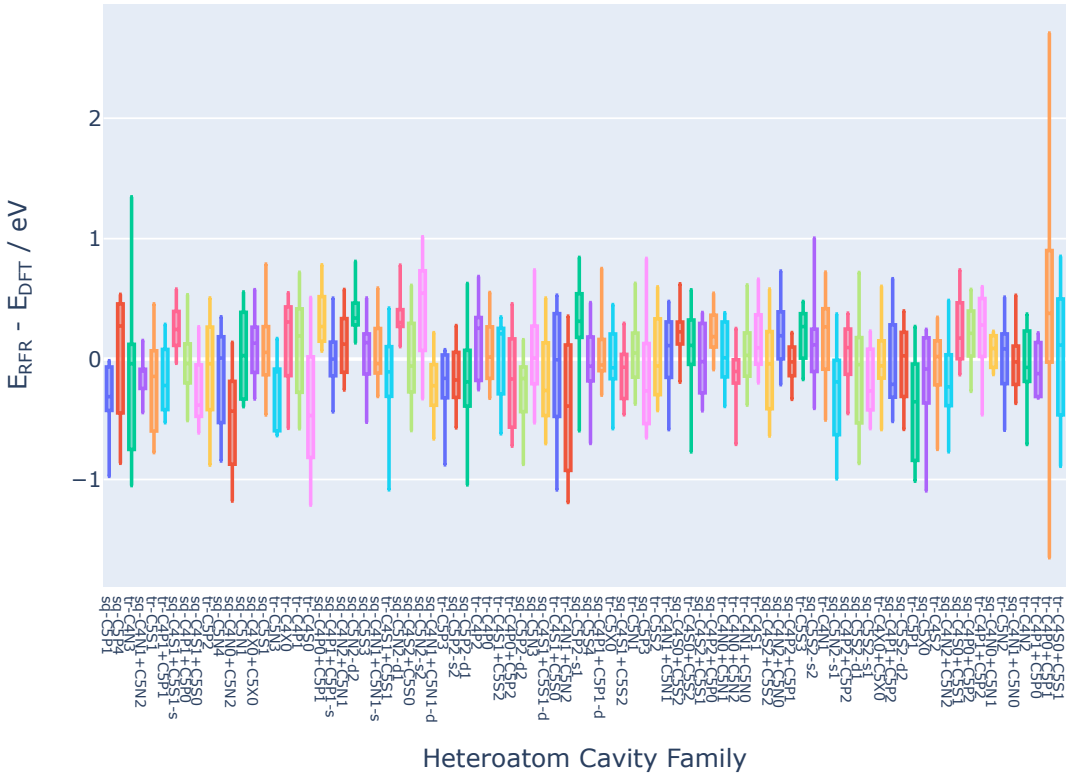

Fig. S15: The error distribution between the predicted energy of the RFR model (E_{RFR}) and the DFT calculated energy (E_{DFT}) grouped by cavity and heteroatom type on the test set of the 5 K-fold cross-validation model.

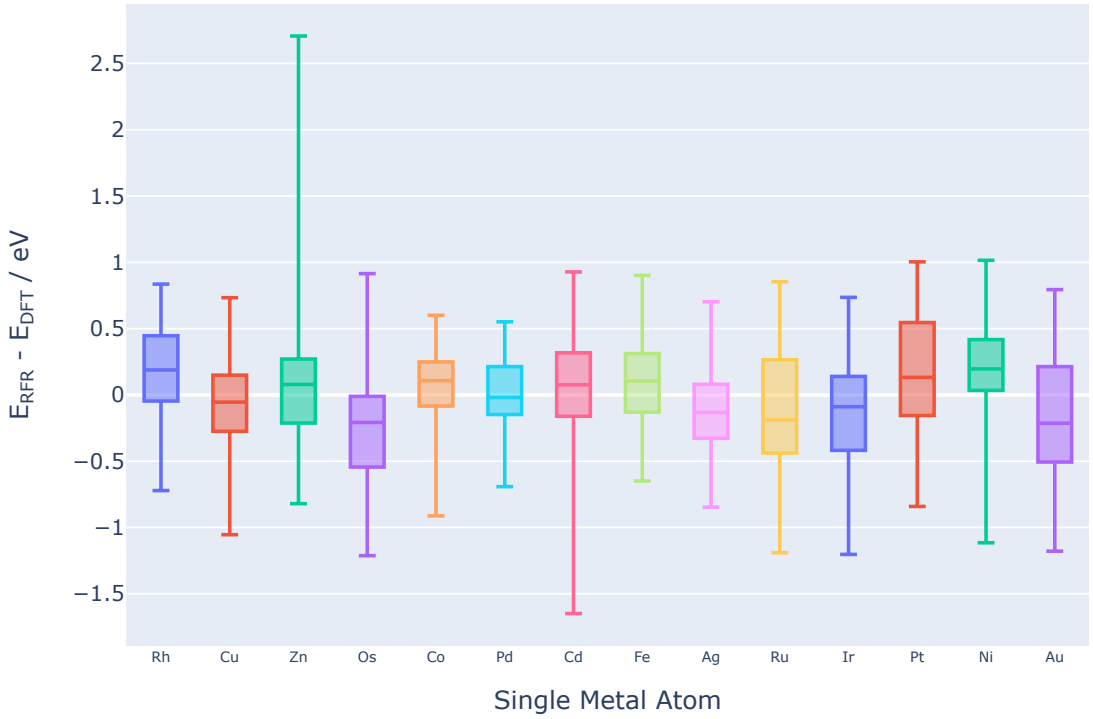

Fig. S16: The error distribution between the predicted energy of the RFR model (E_{RFR}) and the DFT calculated energy (E_{DFT}) grouped by the single metal atom on the test set of the 5 K-fold cross-validation model.

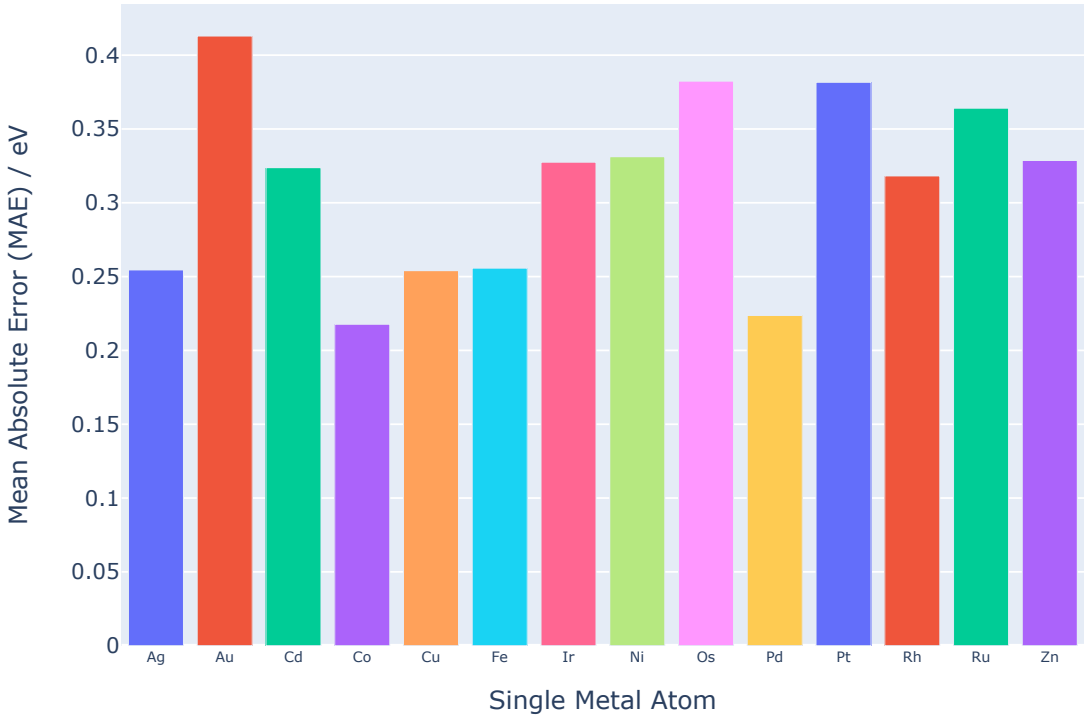

Fig. S17: Mean absolute error grouped by single metal atom in the test set of the 5-fold cross-validation for the highest performance RFR model.

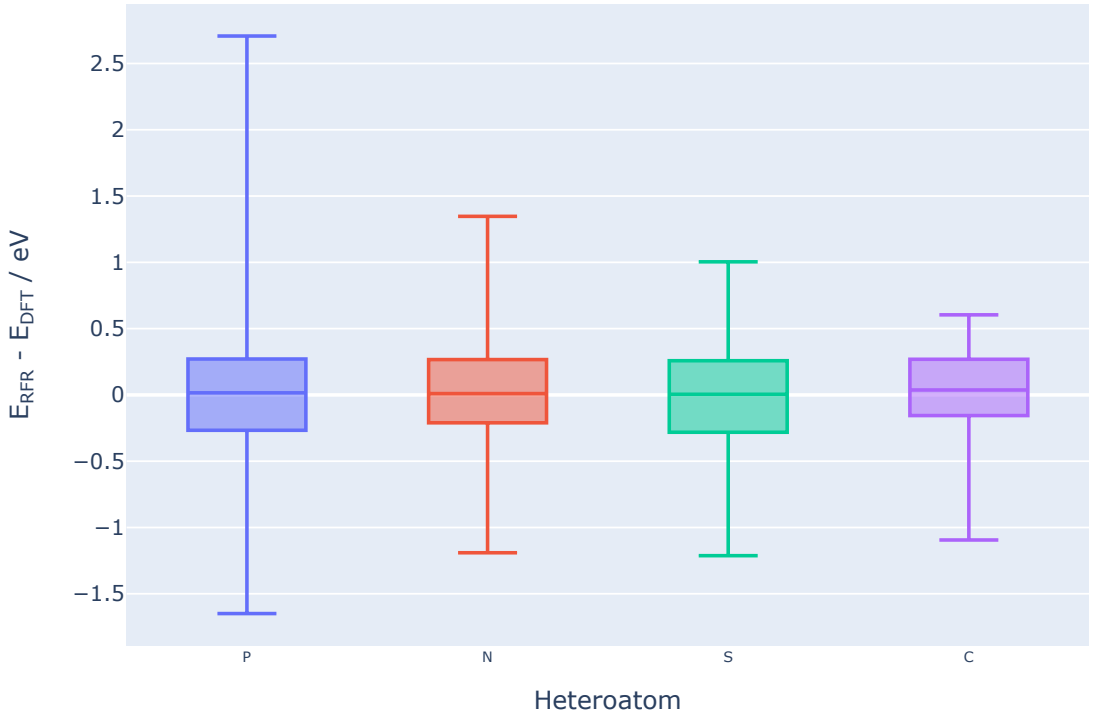

Fig. S18: The error distribution between the predicted energy of the RFR model (E_{RFR}) and the DFT calculated energy (E_{DFT}) grouped by heteroatom in the anchoring site on the test set of the 5 K-fold cross-validation model.

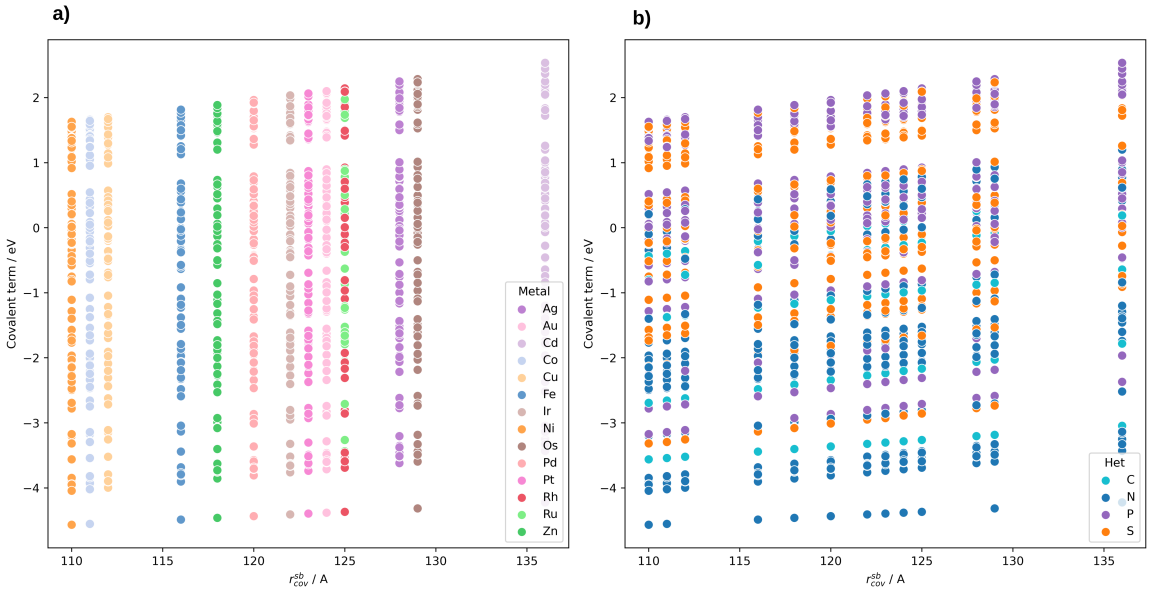

Fig. S19: Dependence of the Covalent term in Equation S6 with the metal covalent radius r^{sb}_{cov}

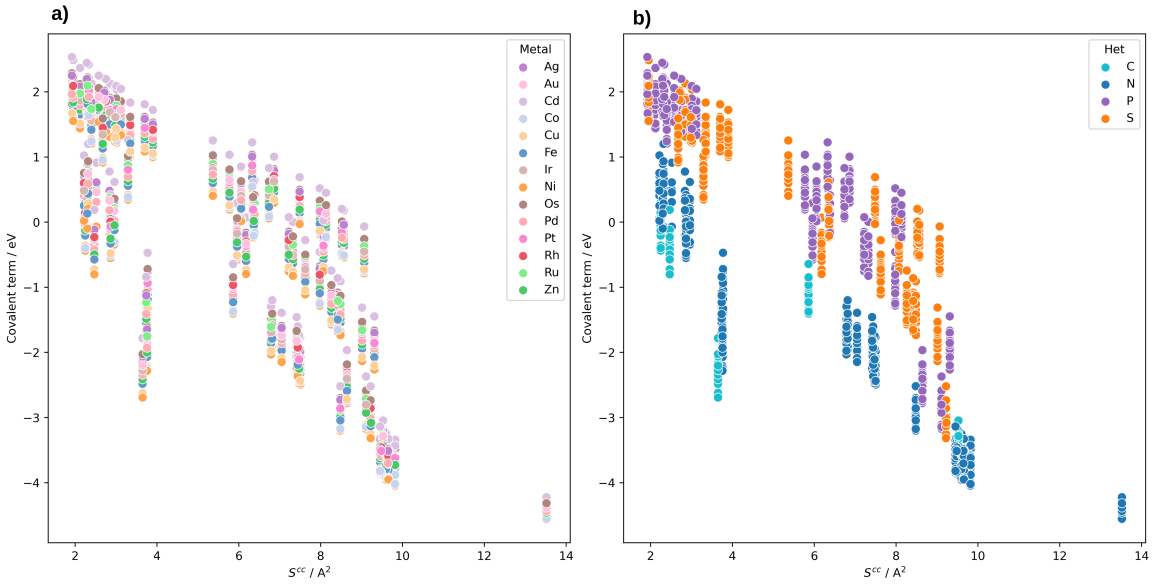

Fig. S20: Dependence of the Covalent term in Equation S6 with the surface of the cavity S^{cc}

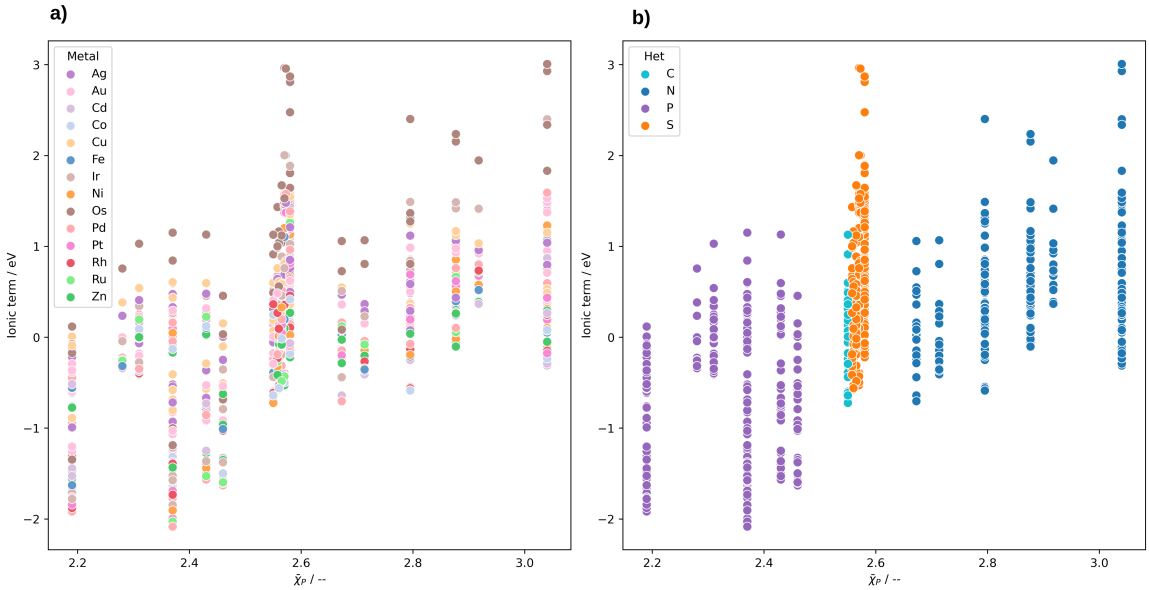

Fig. S21: Dependence of the Ionic term in Equation S6 with the mean Pauling electronegativity $\bar{\chi}_P$

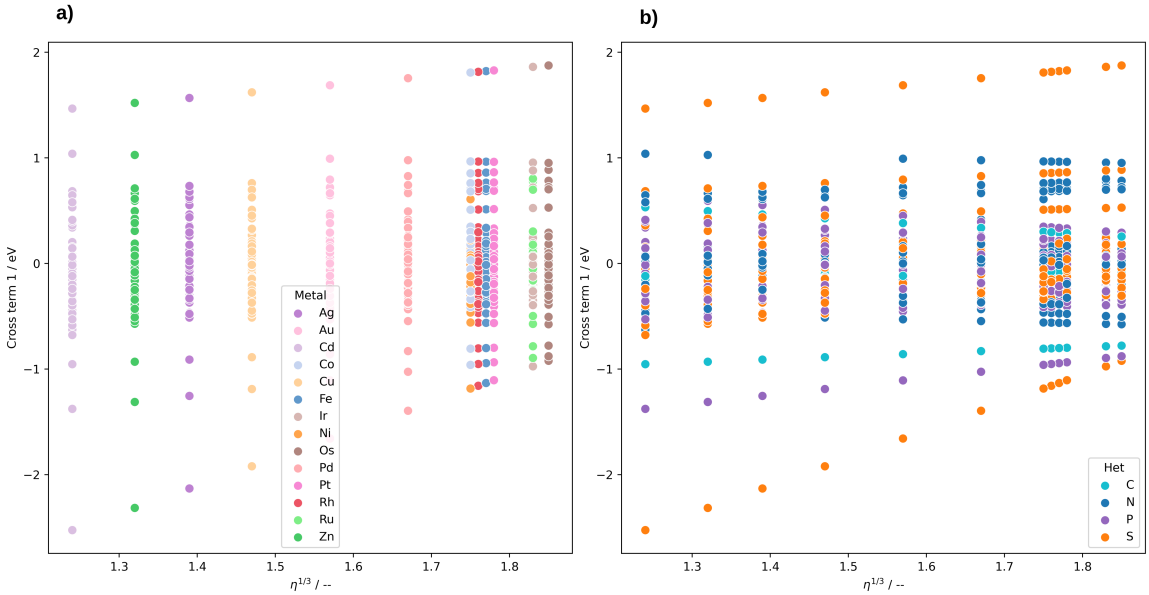

Fig. S22: Dependence of the First Cross-term in Equation S6 with Miedema parameter $\eta^{1/3}$

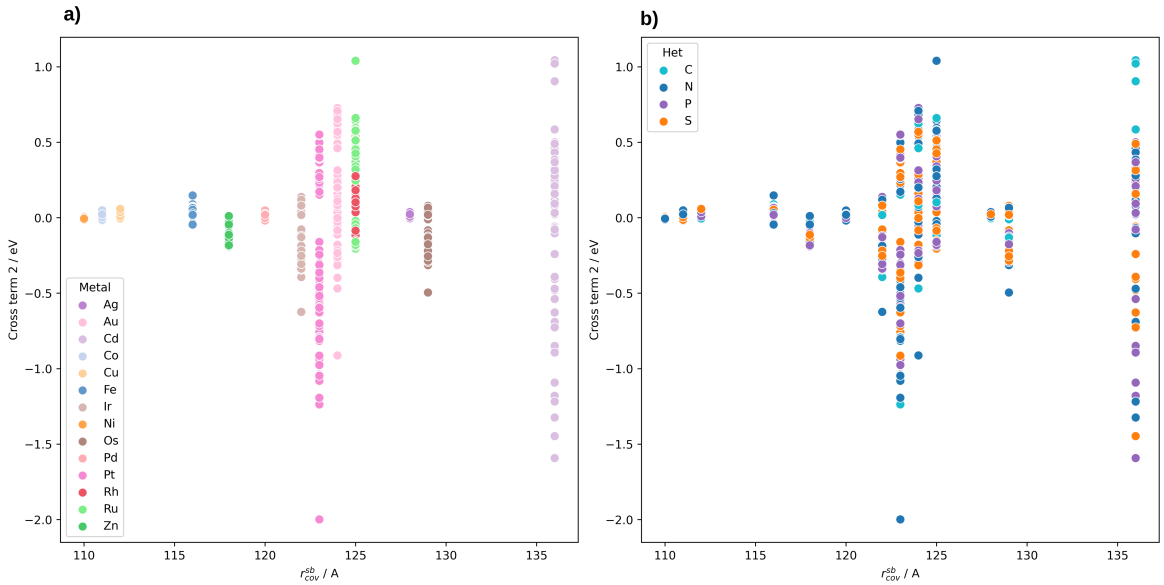

Fig. S23: Dependence of the Second Cross-term in Equation S6 with the metal covalent radius r^{sb}_{cov} .

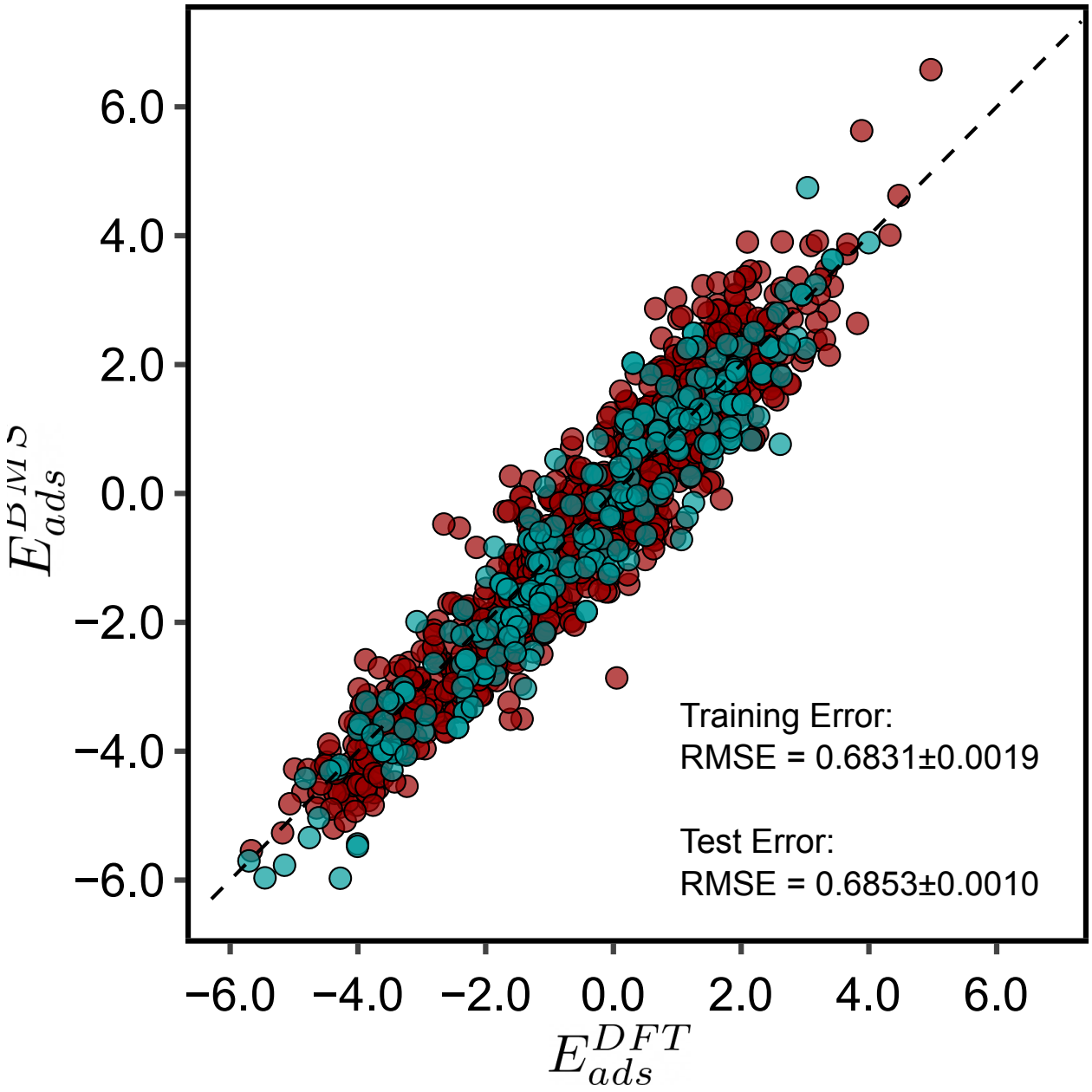

Fig. S24: BMS 5fold CV

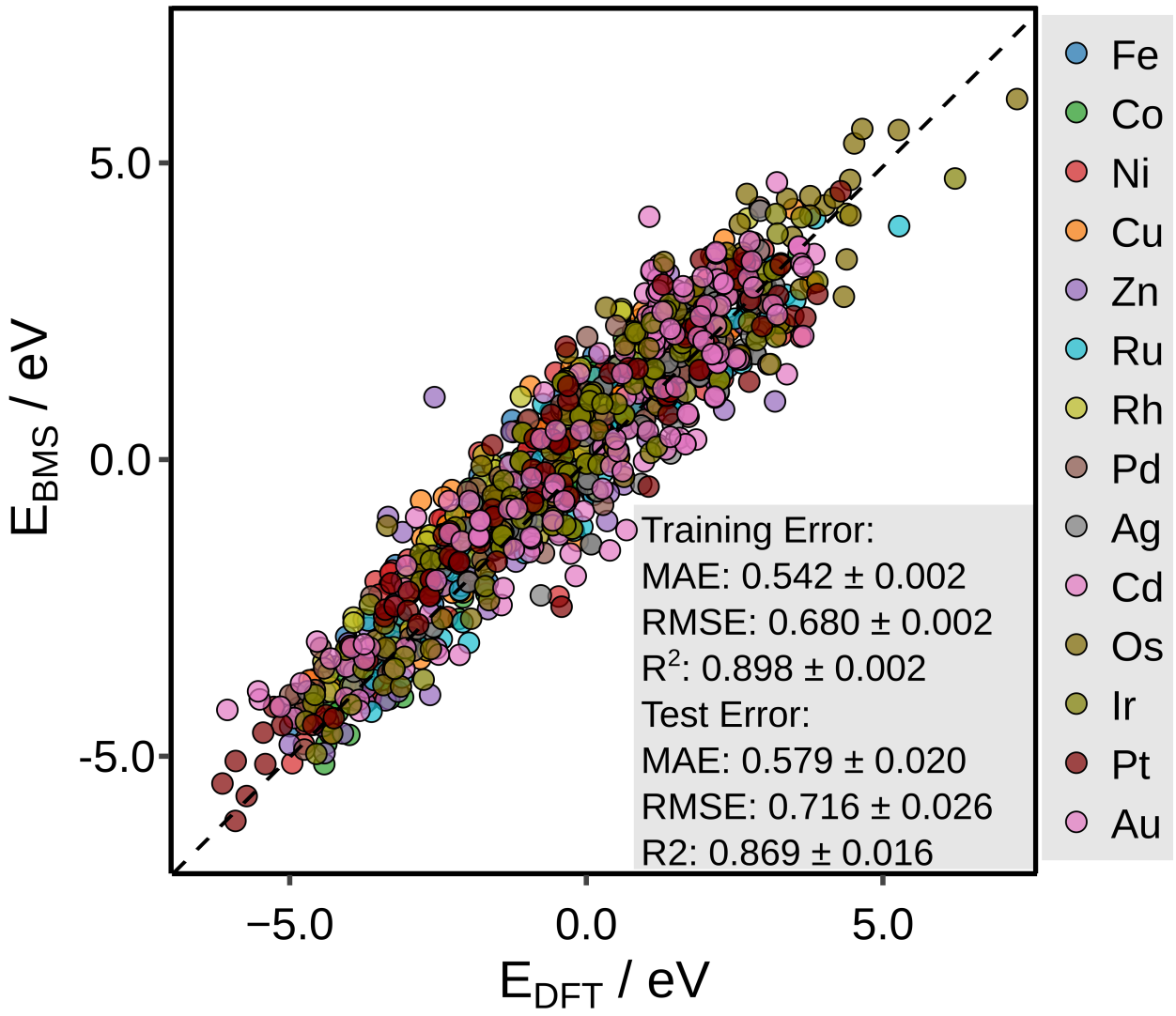

Fig. S25: BMS Metal Leave-One-Group-Out (M-LOGOCV) cross-validation.

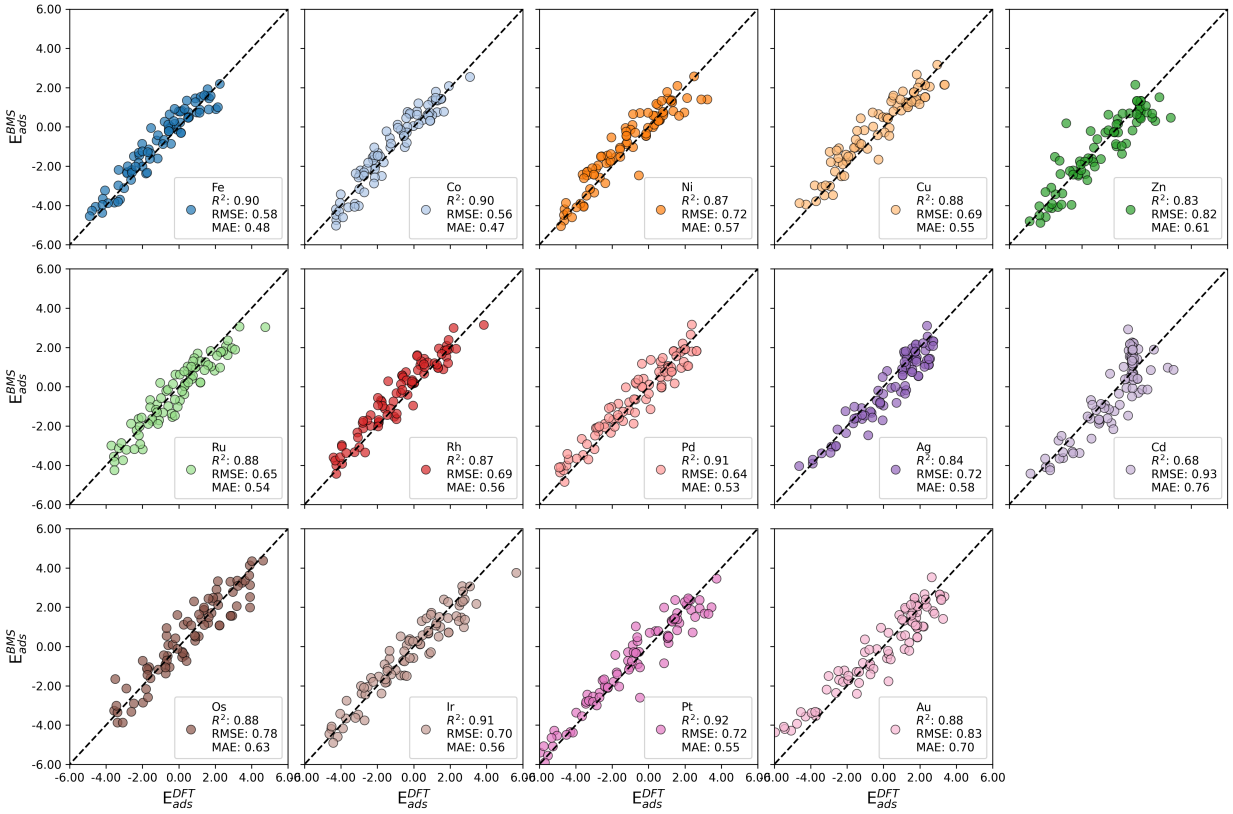

Fig. S26: BMS Metal Leave-One-Group-Out (M-LOGOCV) cross-validation by metal.

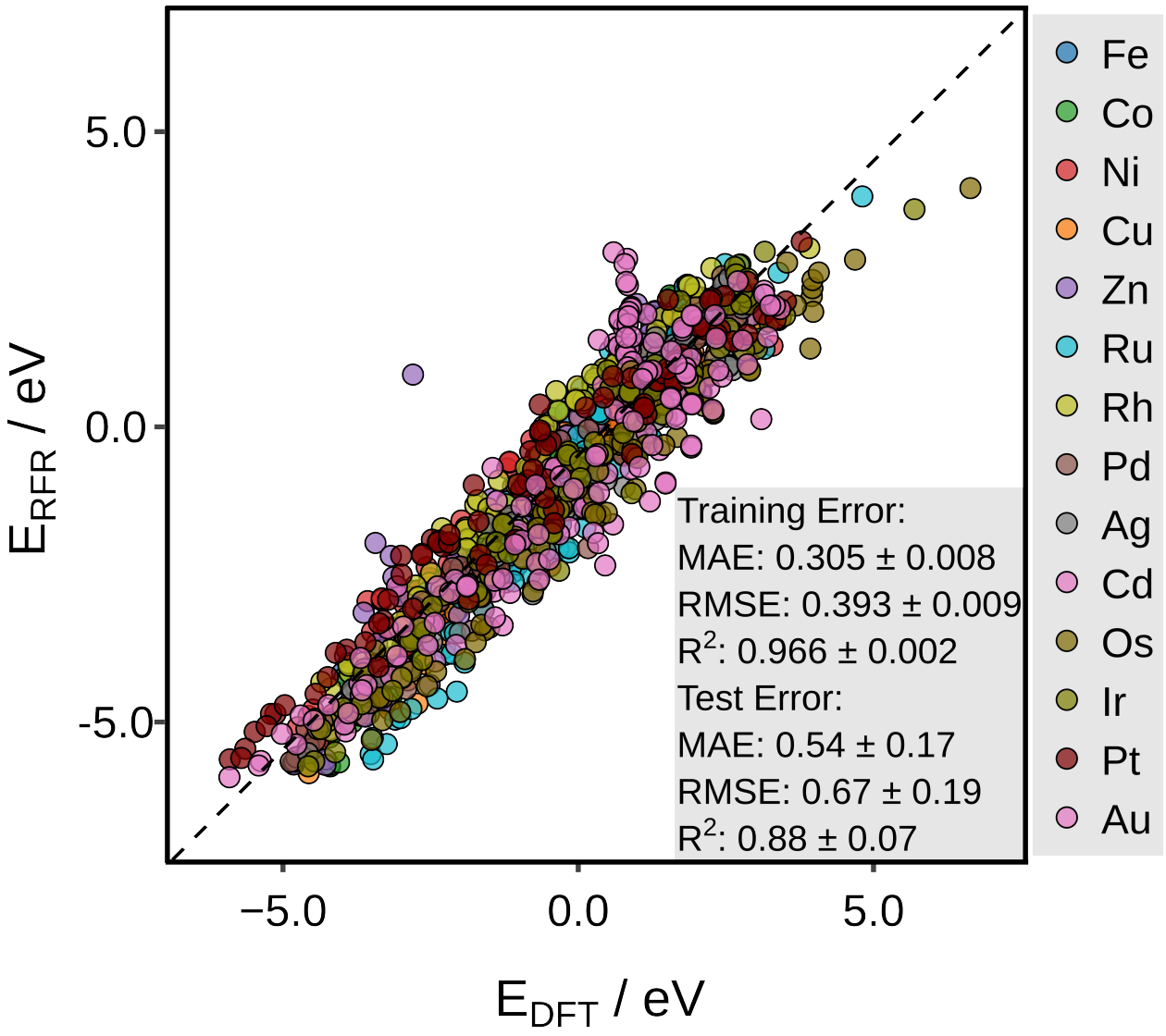

Fig. S27: RFR M-LOGOCV

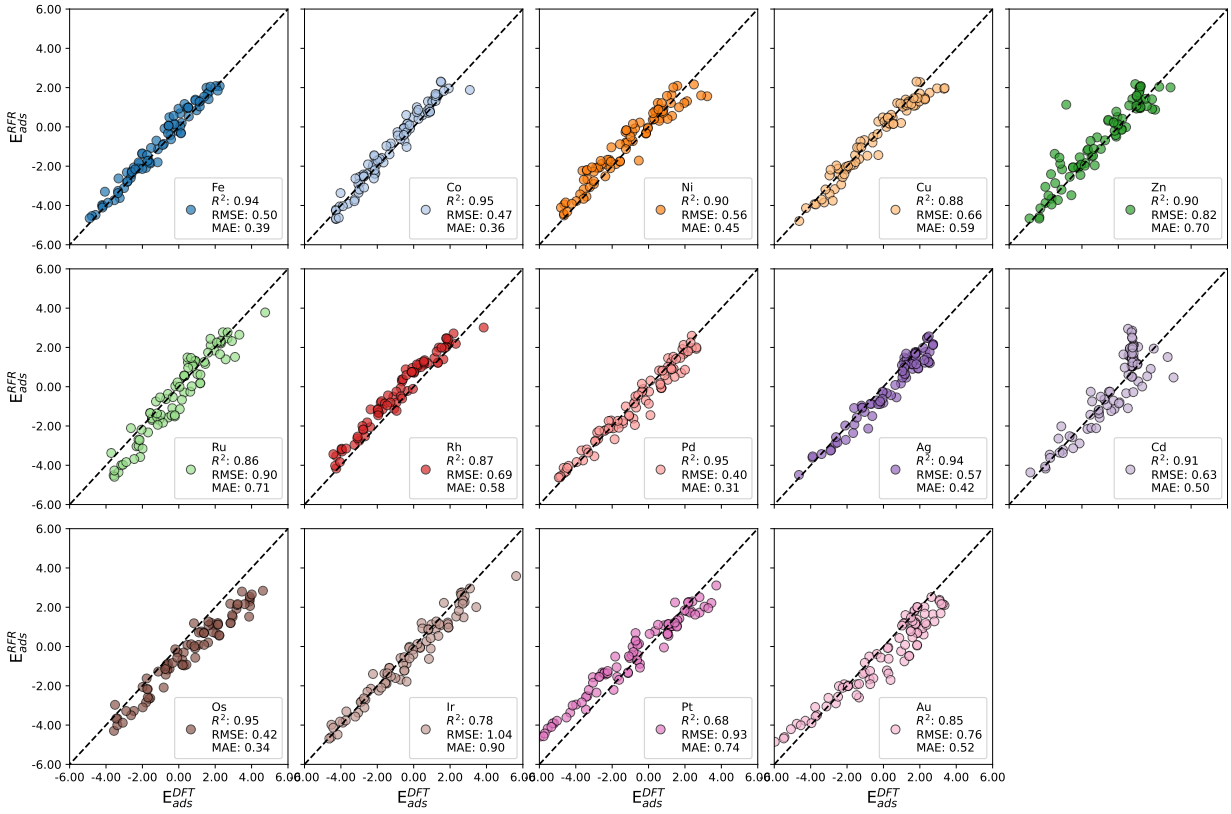

Fig. S28: RFR M-LOGOCV by metal.

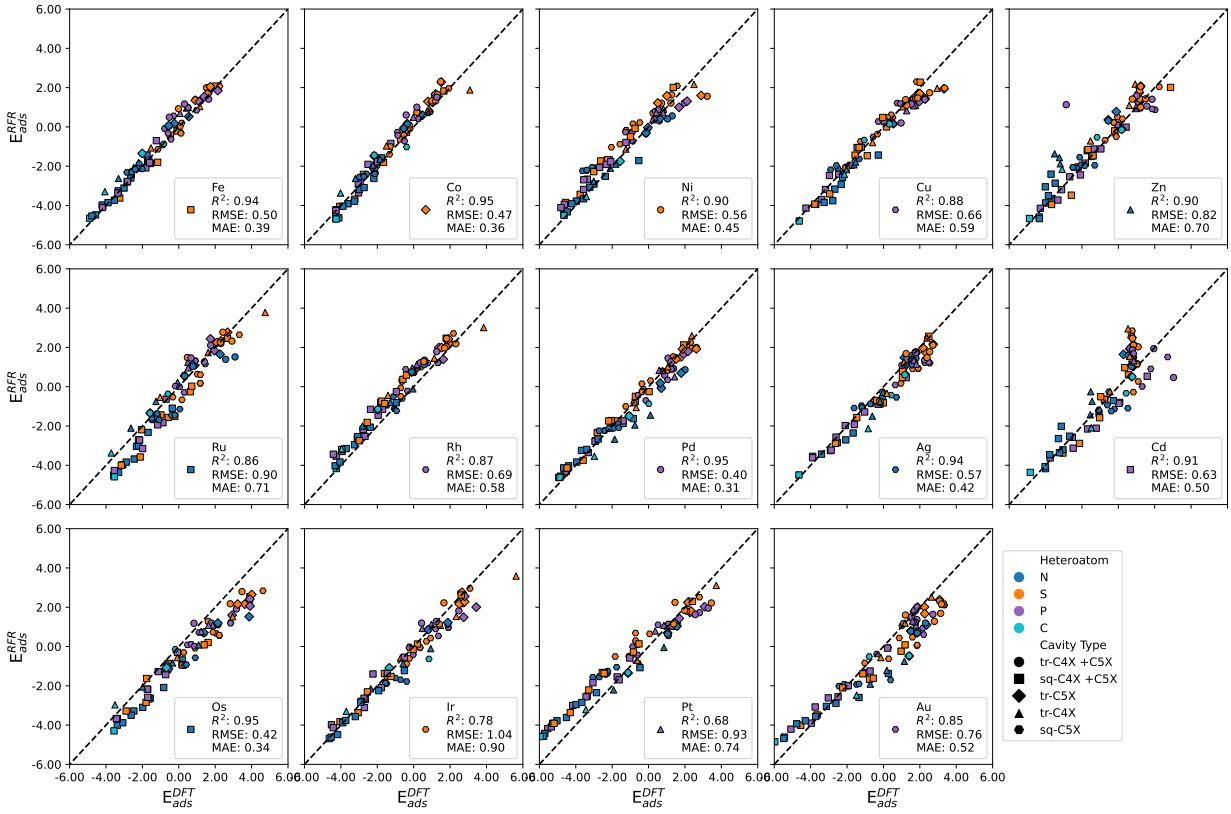

Fig. S29: RFR M-LOGOCV illustrated by cavity and heteroatom

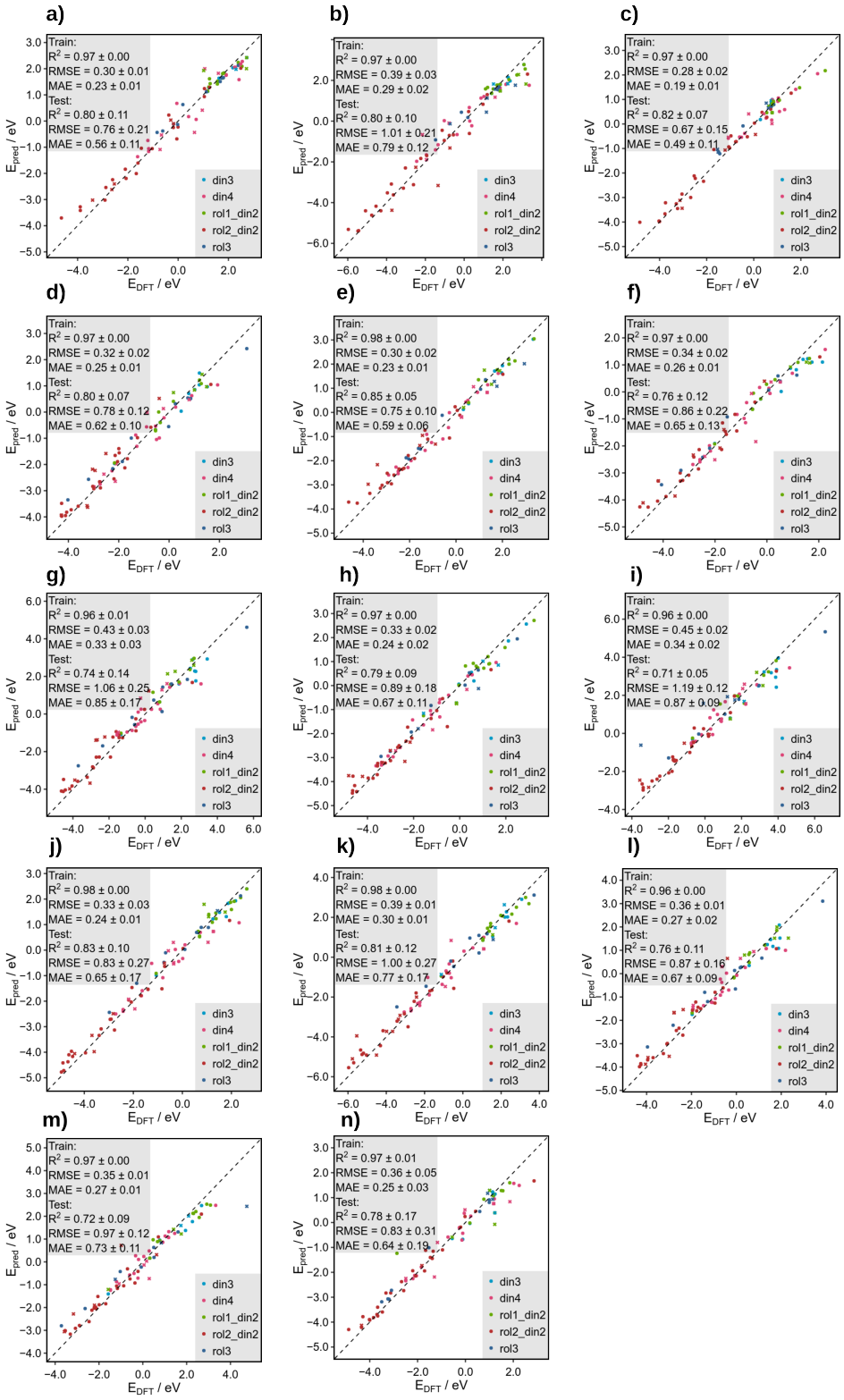

Fig. S30: Performance of the RFR by metal: **a)** Ag, **b)** Au, **c)** Cd, **d)** Co, **e)** Cu, **f)** Fe, **g)** Ir, **h)** Ni, **i)** Os, **j)** Pd, **k)** Pt, **l)** Rh, **m)** Ru, and **n)** Zn.

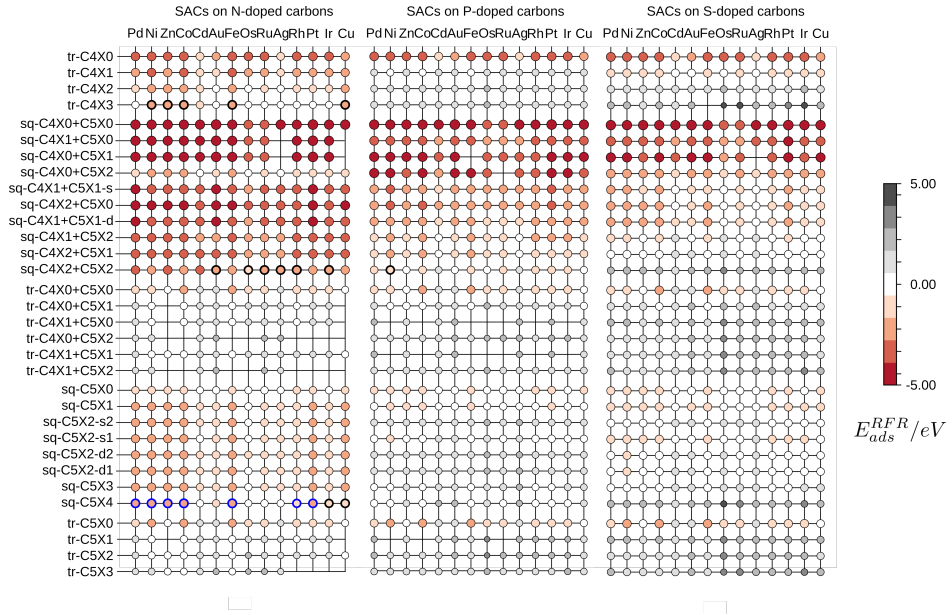

Fig. S31: High-throughput screening of SACs on N, P, and S doped carbons. Each circle represents a unique structure, with color and size indicating the adsorption energy E_{ads}^{RFR} based on the colormap. More stable structures are represented in a red spectrum, while less stable structures are denoted by grey areas. Circles highlighted in black indicate structures predicted by RFR as stable within the energy spectrum of -1.00 to -2.50 eV. Additionally, blue highlighted circles represent structures that were not only predicted but also have been experimentally manufactured and theoretically investigated for validation of our model. The areas with no circles are for the structures unable to converge during the optimization procedure.

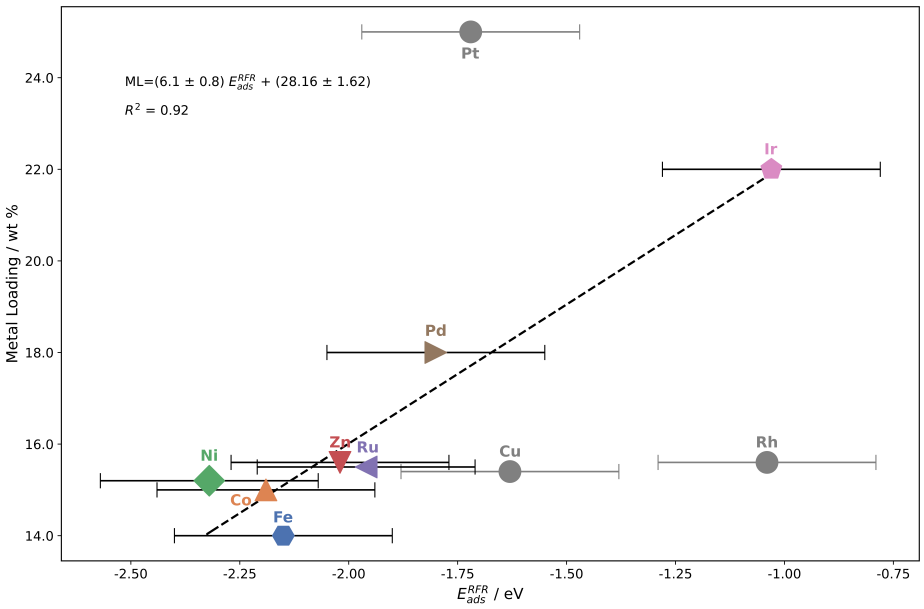

Fig. S32: Metal loading in % in weight that Carbon Nitride (CN) can disperse as single atoms[5] as a function of the adsorption energies of the metals estimated with our models.

S3 Supplementary Tables

Table S1: Electron configurations and corresponding calculated cohesive energies (E_{coh}^{DFT}) for single metal atoms.

Metal	Configuration	E_{coh}^{DFT} / eV
Fe	$3d^6\ 4s^2$	8.62
Ru	$4d^7\ 5s^1$	9.81
Os	$4f^{14}\ 5d^6\ 6s^2$	11.84
Co	$3d^7\ 4s^2$	7.38
Rh	$4d^8\ 5s^1$	7.85
Ir	$4f^{14}\ 5d^7\ 6s^2$	9.56
Ni	$3d^8\ 4s^2$	5.87
Pd	$4d^{10}$	5.79
Pt	$4f^{14}\ 5d^9\ 6s^1$	6.87
Cu	$3d^{10}\ 4s^1$	4.23
Ag	$4d^{10}\ 5s^1$	3.21
Au	$4f^{14}\ 5d^{10}\ 6s^1$	3.88
Zn	$3d^{10}\ 4s^2$	1.62
Cd	$4d^{10}\ 5s^2$	1.23

Table S2: Summary statistics of the DFT simulated adsorption energies (E_{ads}^{DFT} / eV) on SACs catalysts in doped carbons.

Summary Statistics							
count	mean	std	min	25%	50%	75%	max
1153	-0.460	2.134	-5.971	-2.099	-0.249	1.186	6.576

Table S3: Summary statistics of the DFT simulated adsorption energies (E_{ads}^{DFT} / eV) on SACs in doped carbons grouped by the metal center.

Summary Statistics								
	count	mean	std	min	25%	50%	75%	max
Ag	83	0.440	1.825	-4.644	-0.691	1.123	1.773	2.731
Au	86	0.014	2.429	-5.971	-1.478	0.995	1.846	3.354
Cd	86	-0.207	1.643	-4.838	-1.074	0.628	0.783	3.035
Co	81	-1.178	1.816	-4.278	-2.727	-1.256	0.250	3.087
Cu	78	-0.362	2.025	-4.627	-2.068	-0.160	1.342	3.364
Fe	80	-1.095	1.870	-4.897	-2.581	-0.851	0.341	2.248
Ir	81	-0.274	2.289	-4.638	-1.783	-0.227	1.414	5.629
Ni	83	-1.217	2.049	-4.816	-3.005	-1.229	0.521	3.224
Os	82	0.628	2.246	-3.563	-0.745	0.758	2.154	6.576
Pd	86	-0.651	2.162	-4.933	-2.192	-0.361	1.185	2.628
Pt	83	-0.826	2.585	-5.965	-2.708	-0.709	1.333	3.721
Rh	81	-0.811	1.929	-4.415	-1.971	-0.651	0.571	3.848
Ru	82	-0.091	1.894	-3.714	-1.475	0.051	1.178	4.745
Zn	81	-0.863	1.996	-4.881	-2.495	-0.554	1.011	2.868

Table S4: Summary statistics of the DFT simulated adsorption energies (E_{ads}^{DFT} / eV) on SACs in doped carbons grouped by the heteroatom.

Summary Statistics								
	count	mean	std	min	25%	50%	75%	max
C	70	-1.897	1.950	-5.971	-3.556	-1.527	-0.646	1.415
N	348	-1.514	1.904	-5.769	-2.981	-1.749	-0.105	3.865
P	359	-0.082	1.994	-5.543	-1.448	0.514	1.361	3.914
S	376	0.421	1.951	-5.192	-0.776	0.716	1.853	6.576

Table S5: Summary statistics of the DFT simulated adsorption energies (E_{ads}^{DFT} / eV) on the SACs in doped carbons grouped by cavity type and heteroatom.

Summary Statistics								
	count	mean	std	min	25%	50%	75%	max
sq-C4N0+C5N1	12	-4.504	0.695	-5.769	-4.813	-4.440	-4.208	-3.382
sq-C4N0+C5N2	14	-0.644	0.680	-1.758	-0.923	-0.491	-0.275	0.270
sq-C4N1+C5N0	12	-4.344	0.757	-5.702	-4.642	-4.261	-4.044	-3.081
sq-C4N1+C5N1-d	14	-3.288	0.648	-4.511	-3.689	-3.256	-2.943	-1.982
sq-C4N1+C5N1-s	14	-3.683	0.644	-5.033	-3.949	-3.688	-3.295	-2.471
sq-C4N1+C5N2	14	-2.762	0.581	-3.668	-3.059	-2.796	-2.395	-1.571
sq-C4N2+C5N0	14	-3.995	0.660	-5.267	-4.216	-3.917	-3.801	-2.622
sq-C4N2+C5N1	14	-3.060	0.639	-4.013	-3.322	-3.115	-2.734	-1.647
sq-C4N2+C5N2	14	-2.268	0.750	-3.699	-2.736	-2.311	-1.747	-0.830
sq-C4P0+C5P1	13	-4.318	0.608	-5.543	-4.628	-4.277	-3.895	-3.378
sq-C4P0+C5P2	13	-3.950	0.801	-5.340	-4.519	-4.193	-3.389	-2.484
sq-C4P1+C5P0	14	-2.978	0.667	-4.170	-3.335	-3.078	-2.624	-1.703
sq-C4P1+C5P1-d	14	-1.787	0.639	-3.055	-2.050	-1.814	-1.315	-0.720
sq-C4P1+C5P1-s	14	-2.477	0.646	-3.368	-2.884	-2.612	-2.116	-1.218
sq-C4P1+C5P2	14	-1.364	0.807	-2.511	-2.082	-1.310	-0.893	0.072
sq-C4P2+C5P0	14	-2.685	0.572	-3.557	-3.013	-2.807	-2.279	-1.693
sq-C4P2+C5P1	14	-1.278	0.615	-2.076	-1.695	-1.424	-1.067	0.029
sq-C4P2+C5P2	14	-0.299	0.918	-1.256	-1.010	-0.607	0.186	1.588
sq-C4S0+C5S1	13	-3.968	0.663	-5.192	-4.476	-4.049	-3.670	-2.894
sq-C4S0+C5S2	14	-2.313	0.615	-3.282	-2.723	-2.389	-2.051	-1.025
sq-C4S1+C5S0	14	-3.005	0.704	-4.305	-3.509	-3.024	-2.595	-1.801
sq-C4S1+C5S1-d	14	-1.138	0.754	-2.420	-1.638	-1.206	-0.630	0.174
sq-C4S1+C5S1-s	14	-1.495	0.621	-2.442	-1.950	-1.634	-1.023	-0.372
sq-C4S1+C5S2	14	0.014	0.786	-0.996	-0.689	-0.116	0.611	1.391
sq-C4S2+C5S0	14	-1.350	0.892	-2.630	-1.977	-1.513	-0.749	0.222
sq-C4S2+C5S1	14	0.210	0.696	-0.742	-0.316	0.006	0.621	1.608
sq-C4S2+C5S2	14	2.266	0.806	0.719	1.820	2.225	2.658	3.903
sq-C4X0+C5X0	14	-4.686	0.701	-5.971	-4.878	-4.644	-4.356	-3.538
sq-C5N1	14	-1.584	0.823	-2.777	-2.198	-1.741	-1.003	-0.216
sq-C5N2-d1	14	-2.052	0.761	-3.392	-2.560	-2.182	-1.495	-0.746
sq-C5N2-d2	14	-2.052	0.761	-3.392	-2.559	-2.182	-1.495	-0.745
sq-C5N2-s1	14	-1.276	1.031	-2.887	-2.055	-1.388	-0.720	0.525
sq-C5N2-s2	14	-2.128	0.996	-3.632	-2.891	-2.321	-1.518	-0.387
sq-C5N3	14	-1.556	0.980	-3.106	-2.315	-1.700	-0.748	0.310
sq-C5N4	14	-0.999	1.176	-2.581	-2.030	-1.225	-0.088	0.904
sq-C5P1	14	-0.063	0.624	-0.864	-0.482	-0.245	0.216	1.311
sq-C5P2-d1	14	1.346	0.690	0.517	0.902	1.222	1.652	2.847
sq-C5P2-d2	14	1.344	0.690	0.514	0.900	1.223	1.652	2.843

Table S6: Summary statistics of the DFT simulated adsorption energies (E_{ads}^{DFT} / eV) on SACs in doped carbons grouped by cavity type and heteroatom.

Summary Statistics								
	count	mean	std	min	25%	50%	75%	max
sq-C5P2-s1	14	-0.110	0.834	-1.266	-0.627	-0.267	0.210	1.782
sq-C5P2-s2	14	1.110	0.602	0.356	0.782	0.953	1.211	2.552
sq-C5P3	14	0.942	0.842	-0.438	0.297	1.070	1.363	2.273
sq-C5P4	14	0.421	1.327	-0.895	-0.642	-0.210	1.705	2.712
sq-C5S1	14	-0.786	0.795	-2.043	-1.260	-0.877	-0.343	0.823
sq-C5S2-d1	14	0.417	1.065	-1.229	-0.330	0.178	1.144	2.226
sq-C5S2-d2	14	0.419	1.065	-1.229	-0.328	0.180	1.146	2.227
sq-C5S2-s1	14	-0.255	0.854	-1.571	-0.858	-0.276	0.259	1.092
sq-C5S2-s2	14	0.452	0.824	-0.858	-0.131	0.454	0.987	1.877
sq-C5S3	14	0.550	0.729	-0.483	0.071	0.304	0.998	2.062
sq-C5S4	14	2.479	0.931	0.763	1.994	2.281	3.019	4.623
sq-C5X0	14	-0.431	0.644	-1.174	-0.892	-0.616	-0.039	0.839
tr-C4N0+C5N1	12	0.595	0.807	-0.553	-0.113	0.743	1.131	1.859
tr-C4N0+C5N2	6	1.559	0.628	0.817	1.126	1.566	1.779	2.575
tr-C4N1	14	-2.229	0.871	-3.314	-2.963	-2.194	-1.878	-0.613
tr-C4N1+C5N0	12	0.590	0.807	-0.558	-0.118	0.739	1.127	1.853
tr-C4N1+C5N1	14	0.742	0.844	-0.554	0.109	0.686	1.431	2.146
tr-C4N1+C5N2	6	1.801	0.875	0.552	1.361	1.781	2.233	3.088
tr-C4N2	14	-1.271	0.986	-3.239	-1.979	-1.269	-0.503	0.190
tr-C4N3	14	-0.602	1.454	-3.499	-1.734	-0.055	0.659	1.149
tr-C4P0	14	1.036	0.519	0.221	0.620	0.900	1.565	1.763
tr-C4P0+C5P1	14	0.597	1.302	-2.863	0.473	0.859	1.080	3.035
tr-C4P0+C5P2	14	1.861	0.683	0.700	1.368	1.799	2.353	3.169
tr-C4P1	14	0.781	0.650	-0.078	0.265	0.682	1.262	1.801
tr-C4P1+C5P0	7	2.054	0.646	1.162	1.799	1.947	2.196	3.280
tr-C4P1+C5P1	7	1.737	0.732	0.634	1.356	1.772	2.121	2.799
tr-C4P1+C5P2	14	1.068	0.556	0.174	0.689	1.232	1.508	1.839
tr-C4P2	14	0.968	0.576	0.173	0.508	0.840	1.526	1.854
tr-C4S0	13	3.224	1.705	0.533	2.390	2.948	3.848	6.576
tr-C4S0+C5S1	14	1.103	0.816	-0.014	0.493	1.151	1.405	2.801
tr-C4S0+C5S2	14	1.684	0.901	0.770	1.209	1.237	1.939	3.891
tr-C4S1	14	-0.812	0.575	-1.528	-1.370	-0.703	-0.492	0.188
tr-C4S1+C5S0	14	2.218	0.847	0.834	1.498	2.315	2.994	3.364
tr-C4S1+C5S1	14	1.867	0.769	0.714	1.192	1.816	2.477	3.145
tr-C4S1+C5S2	14	2.196	0.876	0.777	1.396	2.400	2.685	3.472
tr-C4S2	14	1.293	0.656	0.599	0.908	1.129	1.484	2.960
tr-C4X0	14	-2.989	1.065	-4.061	-3.711	-3.430	-2.357	-0.837
tr-C4X0+C5X0	14	-0.689	1.232	-2.168	-1.574	-1.099	0.291	1.415
tr-C5N1	12	0.593	0.807	-0.556	-0.114	0.742	1.131	1.857

Table S7: Summary statistics of the DFT simulated adsorption energies (E_{ads}^{DFT} / eV) on SACs in doped carbons grouped by cavity type and heteroatom.

Summary Statistics								
	count	mean	std	min	25%	50%	75%	max
tr-C5N2	14	0.741	0.844	-0.554	0.108	0.685	1.430	2.145
tr-C5N3	10	1.325	1.231	-0.113	0.498	0.881	2.146	3.865
tr-C5P1	11	2.168	0.865	0.783	1.809	2.141	2.549	3.901
tr-C5P2	14	1.912	0.888	0.633	1.383	1.756	2.233	3.914
tr-C5P3	14	1.644	0.618	0.744	1.165	1.627	1.885	2.930
tr-C5S1	14	2.347	0.815	0.835	1.788	2.469	2.867	3.622
tr-C5S2	14	1.691	0.796	0.510	0.992	1.596	2.294	3.235
tr-C5S3	14	1.957	0.839	0.750	1.296	1.958	2.231	4.009
tr-C5X0	14	-0.691	1.228	-2.167	-1.573	-1.097	0.284	1.415

Table S8: Feature space of the primary features

Feature Set	Symbol	Description	Unit
Set 1	m_a	Atomic mass	-
	r_w	Van der Waals radius	Å
	r_{cov}^{sb}	Empirical covalent radius (single bond)	Å
	r_{cov}^{db}	Empirical covalent radius (double bond)	Å
	$\mathcal{P}$	Polarizability of the metal atom	$Å^2\ s^4\ g^{-1}$
	r_{ion}	Ionic radius	Å
	χ_P	Pauling's electronegativity	-
	χ_{MB}	Martynov-Batsanov electronegativity	-
	E_{ea}	Electron affinity	eV
	r_{cov}	Covalent radius	Å
	Z	Atomic number	-
	IE_{n-1}	Ionization energy (n-1)	eV
	IE_n	Ionization energy n	eV
	r_s	Zunger orbital radius	a.u.
	r_p	Cohen orbital radius	a.u.
	r_w	Waber orbital radius	Å
	$\eta^{1/3}$	Miedema parameter $\eta^{1/3}$	$(\text{electrons A}^3)^{(1/3)}$
	ϕ	Miedema parameter ϕ	Volt
	HOMO	HOMO of the isolated metal atom from DFT calculations	eV
	LUMO	LUMO of the isolated metal atom from DFT calculations	eV
	ϵ_d	d-band center of the metal in the bulk phase	eV
	μ_d	Magnetic moment of d orbitals of the metal in the bulk phase	MeV c^{-1}
	ϵ_F	Fermi energy of the metal in the bulk phase	eV
Set 2	η^{cc}	Coordination number	-
	N_X	Number of heteroatoms (N, S, P)	-
	N_C	Number of Carbon atoms in the anchoring sites of the cavity	-
	f_C	Fraction of Carbon atoms in the anchoring sites of the cavity	-
	f_X	Fraction of heteroatoms (N, S, P)	-
	N_{pyrid}	Number of heteroatoms (N, S, P) in pyridine ring	-
	f_{pyrid}	Fraction of heteroatoms (N, S, P) in pyridine ring	-
	N_{pyrr}	Number of heteroatoms (N, S, P) in pyrrole ring	-
	f_{pyrr}	Fraction of heteroatoms (N, S, P) in pyrrole ring	-

Feature Set	Symbol	Description	Unit
	$\varepsilon_{\min}$	Minimum strain effect of the pristine cavity-oriented carbon	Å
	$\varepsilon_{\max}$	Maximum strain effect of the pristine cavity-oriented carbon	Å
	S^{cc}	Surface area of the pristine cavity-oriented carbon	Å
	ϵ_F^{cc}	Fermi energy of the pristine cavity-oriented carbon	eV
Set 3	$\chi_{P,\min}^{cc}$	The minimum of Pauling's electronegativity of the anchoring atoms in the cavity area	-
	$\overline{\chi_P^{cc}}$	The mean of Pauling's electronegativity of the anchoring atoms in the cavity area	-
	$\chi_{P,\max}^{cc}$	The maximum of Pauling's electronegativity of the anchoring atoms in the cavity area	-
	$\sigma_{\chi_P^{cc}}$	The standard deviation of Pauling's electronegativity of the anchoring atoms in the cavity area	-
	$\Sigma\chi_P^{cc}$	The sum of Pauling's electronegativity of the anchoring atoms in the cavity area	-
Set 4	$r_{cov,\min}^{cc}$	Minimum of Cordero covalent radius of the anchoring atoms in the cavity area	Å
	$\overline{r_{cov}^{cc}}$	Mean of Cordero covalent radius of the anchoring atoms in the cavity area	Å
	$r_{cov,\max}^{cc}$	Maximum of Cordero covalent radius of the anchoring atoms in the cavity area	Å
	$\sigma_{r_{cov}^{cc}}$	Standard deviation of Cordero covalent radius of the anchoring atoms in the cavity area	Å
	Σr_{cov}^{cc}	Sum of Cordero covalent radius of the anchoring atoms in the cavity area	Å
Set 5	$d_{X,\min}$	Minimum distance between heteroatoms of cavity area	Å
	$d_{X,\max}$	Maximum distance between heteroatoms of cavity area	Å
	$\overline{d}_X$	Mean distance between heteroatoms of cavity area	Å
	σ_{d_X}	Standard deviation of distance between heteroatoms of cavity area	Å
Set 6	$d_{X,\min}^{\dagger}$	Minimum distance between heteroatoms of cavity area	Å
	$d_{X,\max}^{\dagger}$	Maximum distance between heteroatoms of cavity area	Å
	$\overline{d}_{X^{\dagger}}$	Mean distance between heteroatoms of cavity area	Å
	$\sigma_{d_{X^{\dagger}}}$	Standard deviation between heteroatoms of distance of cavity area	Å

Table S9: Performance of the Gaussian Process Regression (GPR), Support Vector Regression (SVR), and Random Forest Regression (RFR) models during forward Sequential Feature Selection, evaluated via K-Fold Cross-Validation up to the inclusion of eight primary features for SACs in carbon cavities, according to Supplementary Note S5.

SACs in doped carbons				
Model	**Feature**	**R^2**	**RMSE / eV**	**MAE / eV**
GPR	S^{cc}	0.61 ± 0.01	1.35 ± 0.07	1.03 ± 0.04
	ϵ_F^{cc}	0.85 ± 0.01	0.87 ± 0.03	0.70 ± 0.03
	n^{cc}	0.85 ± 0.01	0.87 ± 0.03	0.70 ± 0.03
	$r_{cov,\mathrm{min}}^{cc}$	0.85 ± 0.01	0.87 ± 0.04	0.70 ± 0.03
	$d_{X,\mathrm{min}}$	0.85 ± 0.01	0.87 ± 0.04	0.70 ± 0.03
	σ_{d_X}	0.85 ± 0.01	0.87 ± 0.04	0.70 ± 0.03
	$\overline{d_X}$	0.85 ± 0.01	0.88 ± 0.04	0.70 ± 0.03
	$d_{X,\mathrm{max}}$	0.85 ± 0.01	0.88 ± 0.04	0.70 ± 0.03
SVR	$\overline{d_X^\dagger}$	0.55 ± 0.01	1.43 ± 0.07	1.06 ± 0.04
	ϵ_F^{cc}	0.74 ± 0.01	1.11 ± 0.02	0.87 ± 0.02
	N_X	0.79 ± 0.02	0.99 ± 0.02	0.78 ± 0.02
	ϵ_d	0.84 ± 0.01	0.87 ± 0.03	0.68 ± 0.02
	$\sigma_{d_X^\dagger}$	0.88 ± 0.01	0.78 ± 0.03	0.60 ± 0.02
	IE_n	0.70 ± 0.01	0.73 ± 0.01	0.56 ± 0.01
	$\overline{\chi_P^{cc}}$	0.91 ± 0.01	0.68 ± 0.01	0.52 ± 0.01
	$\overline{d_X}$	0.92 ± 0.01	0.66 ± 0.02	0.49 ± 0.01
RFR	S^{cc}	0.84 ± 0.01	0.90 ± 0.04	0.72 ± 0.03
	$\overline{r_{cov}^{cc}}$	0.85 ± 0.01	0.87 ± 0.04	0.67 ± 0.03
	$\eta^{1/3}$	0.93 ± 0.03	0.71 ± 0.03	0.55 ± 0.03
	ϵ_F^{cc}	0.93 ± 0.02	0.68 ± 0.02	0.53 ± 0.03
	r_{cov}^{sb}	0.95 ± 0.01	0.60 ± 0.03	0.46 ± 0.02
	Σr_{cov}^{cc}	0.96 ± 0.01	0.59 ± 0.03	0.46 ± 0.02
	χ_{MB}	0.95 ± 0.02	0.59 ± 0.02	0.45 ± 0.02
	$\overline{\chi_P}$	0.96 ± 0.01	0.56 ± 0.03	0.43 ± 0.02

Table S10: Training performance of RFR model with five-fold cross-validation on different feature subsets for SACs in doped carbons, according to the Supplementary Note S4.

SACs in doped carbons				
Model	**Feature space**	**R^2**	**RMSE / eV**	**MAE / eV**
GPR	Set 1	0.07 ± 0.01	2.06 ± 0.02	1.74 ± 0.02
	Set 2	0.87 ± 0.01	0.78 ± 0.01	0.62 ± 0.01
	Set 1-2	1.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01
	Set 1-3	1.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01
	Set 1, 2, 4	1.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01
	Set 1-4	1.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01
	Set 1- 5	1.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01
	Set 1-4, 6	1.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01
	Set 1-6	1.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01
SVR	Set 1	0.04 ± 0.01	2.09 ± 0.03	1.71 ± 0.03
	Set 2	0.83 ± 0.01	0.88 ± 0.01	0.67 ± 0.01
	Set 1-2	0.88 ± 0.01	0.73 ± 0.01	0.50 ± 0.01
	Set 1-3	0.91 ± 0.01	0.64 ± 0.01	0.41 ± 0.01
	Set 1, 2, 4	0.90 ± 0.01	0.67 ± 0.01	0.44 ± 0.01
	Set 1-4	0.91 ± 0.01	0.64 ± 0.01	0.41 ± 0.01
	Set 1-5	0.93 ± 0.01	0.55 ± 0.01	0.34 ± 0.01
	Set 1-4, 6	0.94 ± 0.01	0.52 ± 0.01	0.32 ± 0.01
	Set 1-6	0.95 ± 0.01	0.49 ± 0.01	0.30 ± 0.01
RFR	Set 1	0.07 ± 0.01	2.06 ± 0.02	1.74 ± 0.02
	Set 2	0.87 ± 0.00	0.78 ± 0.01	0.62 ± 0.01
	Set 1-2	0.98 ± 0.01	0.33 ± 0.01	0.26 ± 0.01
	Set 1-3	0.98 ± 0.01	0.32 ± 0.01	0.24 ± 0.01
	Set 1, 2, 4	0.98 ± 0.01	0.32 ± 0.01	0.24 ± 0.01
	Set 1-4	0.98 ± 0.01	0.31 ± 0.01	0.23 ± 0.01
	Set 1-5	0.98 ± 0.01	0.32 ± 0.01	0.24 ± 0.01
	Set 1-4, 6	0.98 ± 0.01	0.33 ± 0.01	0.25 ± 0.01
	Set 1-6	0.98 ± 0.01	0.33 ± 0.01	0.25 ± 0.01

Table S11: Constant values for Equation S6.

Constant	Value
c_1	-0.03
c_2	5.60
c_3	5.74
c_4	-1.57
c_5	0.99
c_6	6.04e3
c_7	-50.06
c_8	4.09e-2
c_9	0.13
c_{10}	-57.84
c_{11}	81.42
c_{12}	-75.74
c_{13}	-0.33
c_{14}	3.88
c_{15}	-93.98
c_{16}	1.02e4
c_{17}	-4.71
c_{18}	-8.96
c_{19}	8.08e4
c_{20}	1.96

Table S12: Metrics for the 5-fold cross-validation for Equation S6.

	MAE / eV	RMSE / eV	R^2
Train	0.545 ± 0.002	0.683 ± 0.002	0.897 ± 0.002
Test	0.541 ± 0.011	0.685 ± 0.010	0.897 ± 0.004

Table S13: Metrics for the M-LOGOCV for Equation S6. MAE and RMSE are in eV.

Metal	Cu	Ru	Os	Pd	Fe	Co	Ni	Ir	Zn	Rh	Au	Pt	Ag	Cd
MAE	0.55	0.54	0.63	0.53	0.48	0.47	0.57	0.56	0.61	0.56	0.70	0.55	0.58	0.76
RMSE	0.69	0.65	0.78	0.64	0.58	0.56	0.72	0.70	0.82	0.69	0.83	0.72	0.72	0.93
R^2	0.88	0.88	0.88	0.91	0.90	0.90	0.87	0.91	0.83	0.87	0.88	0.92	0.84	0.68

Table S14: Metrics for the M-LOGOCV for RFR. MAE and RMSE are in eV.

Metal	Cu	Ru	Os	Pd	Fe	Co	Ni	Ir	Zn	Rh	Au	Pt	Ag	Cd
MAE	0.59	0.71	0.34	0.31	0.39	0.36	0.45	0.90	0.70	0.58	0.52	0.74	0.42	0.50
RMSE	0.66	0.90	0.42	0.40	0.50	0.47	0.56	1.04	0.82	0.69	0.76	0.93	0.57	0.63
R^2	0.88	0.86	0.95	0.95	0.94	0.95	0.90	0.78	0.90	0.87	0.85	0.68	0.94	0.91

Table S15: Metrics for the M-LOGOCV

		MAE / eV	RMSE / eV	R^2
for RFR.	Train	0.300 ± 0.010	0.390 ± 0.010	0.970± 0.010
	Test	0.54 ± 0.17	0.670± 0.190	0.88 ± 0.07

Table S16: Predicted adsorption energy (E_{ads}^{RFR}) for Fe/Cu, Ni/Pd and Ru/Pt, sorted by cavity type and in descending values for thermodynamic stability. The Mean Absolute Error (MAE) and Mean Square Error (MSE) are included in comparison to the DFT simulated energy (E_{ads}^{DFT}).

Cavity Type	$E_{ads,Fe}^{RFR}$ / eV	$E_{ads,Cu}^{RFR}$ / eV	MAE_{Fe} / eV	MSE_{Fe} / eV	MAE_{Cu} / eV	MSE_{Cu} / eV
sq-C4X0+C5X0	-4.650	-4.792	0.247	0.061	0.164	0.027
sq-C4P0+C5P2	-4.105	-3.697	0.088	0.008	0.307	0.094
sq-C4S0+C5S1	-4.023	-3.932	0.159	0.025	0.146	0.021
sq-C4N2+C5N0	-3.980	-3.946	0.233	0.054	0.231	0.053
sq-C4N1+C5N1-s	-3.859	-3.846	0.031	0.001	0.586	0.344
sq-C4N1+C5N1-d	-3.733	-3.759	0.192	0.037	0.973	0.948
sq-C4S1+C5S0	-3.620	-3.412	0.375	0.141	0.458	0.210
sq-C4P1+C5P0	-3.368	-3.187	0.026	0.001	0.001	0.000
tr-C4X0	-3.301	-2.410	0.760	0.578	0.260	0.068
sq-C4N2+C5N1	-3.258	-3.168	0.075	0.006	0.494	0.244
sq-C4N1+C5N2	-3.114	-2.931	0.103	0.011	0.609	0.371
sq-C4P2+C5P0	-2.803	-2.475	0.035	0.001	0.491	0.241
sq-C4P1+C5P1-s	-2.777	-2.416	0.040	0.002	0.041	0.002
tr-C4N1	-2.624	-2.176	0.690	0.476	0.142	0.020
sq-C4N2+C5N2	-2.600	-2.559	0.037	0.001	0.315	0.100
sq-C4S0+C5S2	-2.435	-2.123	0.175	0.031	0.063	0.004
sq-C5N2-s2	-2.317	-2.180	0.537	0.288	0.723	0.523
sq-C5N2-d1	-2.305	-2.152	0.277	0.077	0.340	0.116
sq-C5N2-d2	-2.263	-2.124	0.318	0.101	0.368	0.136
sq-C5N3	-2.159	-1.991	0.234	0.055	0.052	0.003
sq-C5N2-s1	-2.140	-2.040	0.133	0.018	0.030	0.001
sq-C5N1	-2.112	-1.966	0.101	0.010	0.621	0.385
sq-C4P1+C5P1-d	-2.088	-1.953	0.278	0.077	0.167	0.028
tr-C4N2	-2.067	-1.930	0.405	0.164	0.311	0.097
sq-C5N4	-1.877	-1.634	0.299	0.089	0.303	0.092
sq-C4S2+C5S0	-1.863	-1.438	0.300	0.090	0.122	0.015
sq-C4P1+C5P2	-1.822	-1.038	0.221	0.049	0.181	0.033
tr-C4N3	-1.808	-1.793	0.398	0.159	0.260	0.067
sq-C4S1+C5S1-d	-1.801	-1.468	0.639	0.408	0.582	0.338
sq-C4S1+C5S1-s	-1.730	-1.059	0.019	0.000	0.279	0.078
sq-C4P2+C5P1	-1.703	-1.048	0.047	0.002	0.399	0.159
sq-C4N0+C5N2	-1.435	-1.435	0.310	0.096	1.153	1.330
tr-C4X0+C5X0	-1.373	0.151	0.640	0.410	0.176	0.031
tr-C5X0	-1.348	0.178	0.664	0.441	0.138	0.019
sq-C5S1	-1.139	-0.739	0.010	0.000	0.524	0.274
tr-C4S1	-1.072	-0.798	0.457	0.208	0.201	0.041
sq-C5S2-s1	-0.920	-0.365	0.041	0.002	0.040	0.002
sq-C5X0	-0.865	-0.670	0.042	0.002	0.384	0.148
sq-C4P2+C5P2	-0.700	0.278	0.531	0.282	0.224	0.050
sq-C5P1	-0.630	-0.509	0.202	0.041	0.056	0.003
sq-C4S1+C5S2	-0.422	-0.001	0.041	0.002	0.167	0.028
sq-C5P2-s1	-0.329	0.274	0.264	0.070	0.560	0.313
sq-C5S2-d1	-0.324	0.447	0.442	0.195	0.210	0.044
sq-C5S2-d2	-0.303	0.473	0.420	0.177	0.231	0.053
sq-C4S2+C5S1	-0.207	0.479	0.112	0.013	0.518	0.268
sq-C5P4	-0.085	1.213	0.697	0.486	0.459	0.211
sq-C5S2-s2	0.026	0.477	0.053	0.003	0.034	0.001

Table S17: Table displaying the stability of SACs in doped carbons within the energy window of E_{ads}^{RFR} from -1.00 to -2.50 eV, as predicted by the Random Forest Regression (RFR) model.

SACs in doped carbons	E_{ads}^{RFR} / eV
Ni-sq-C4P2+C5P2	-1.02
Ir-sq-C5N4	-1.03
Rh-sq-C5N4	-1.04
Os-sq-C4N2+C5N2	-1.36
Cu-sq-C5N4	-1.63
Pt-sq-C5N4	-1.72
Cu-tr-C4N3	-1.79
Pd-sq-C5N4	-1.80
Ni-tr-C4N3	-1.83
Co-tr-C4N3	-1.85
Fe-tr-C4N3	-1.86
Ru-sq-C4N2+C5N2	-1.96
Au-sq-C4N2+C5N2	-1.96
Zn-sq-C5N4	-2.02
Ag-sq-C4N2+C5N2	-2.08
Fe-sq-C5N4	-2.15
Co-sq-C5N4	-2.19
Ni-sq-C5N4	-2.32
Ir-sq-C4N2+C5N2	-2.34
Zn-tr-C4N3	-2.36
Rh-sq-C4N2+C5N2	-2.42

Table S18: Continuation of Table S16.

Cavity Type	$E^{RFR}_{ads,Fe}$ / eV	$E^{RFR}_{ads,Cu}$ / eV	MAE_{Fe} / eV	MSE_{Fe} / eV	MAE_{Cu} / eV	MSE_{Cu} / eV
tr-C4N1+C5N1	0.123	0.144	0.363	0.131	0.400	0.160
tr-C5N2	0.168	0.167	0.409	0.167	0.375	0.141
sq-C5S3	0.198	0.615	0.038	0.001	0.190	0.036
tr-C4P0+C5P1	0.490	0.200	1.010	1.019	0.769	0.591
sq-C5P3	0.565	1.281	0.847	0.717	0.083	0.007
tr-C4P1	0.690	0.824	0.620	0.384	0.448	0.201
sq-C5P2-s2	0.848	1.186	0.132	0.017	0.260	0.068
tr-C4S0+C5S1	0.925	1.098	0.939	0.881	0.180	0.032
tr-C4P2	0.962	1.074	0.535	0.286	0.580	0.337
sq-C5P2-d1	0.982	1.193	0.465	0.216	0.149	0.022
sq-C5P2-d2	0.984	1.202	0.471	0.222	0.141	0.020
tr-C4P0	0.994	1.098	0.480	0.230	0.665	0.442
tr-C4S2	1.040	1.023	0.113	0.013	0.035	0.001
tr-C4P1+C5P2	1.166	1.161	0.854	0.729	0.368	0.136
tr-C5P3	1.281	1.294	0.149	0.022	0.677	0.458
tr-C4S0+C5S2	1.343	1.390	0.145	0.021	0.170	0.029
tr-C5S2	1.351	1.461	0.442	0.195	0.204	0.042
tr-C4S1+C5S1	1.388	1.470	0.478	0.228	0.414	0.172
tr-C4P0+C5P2	1.412	1.447	0.228	0.052	0.329	0.108
tr-C5P2	1.539	1.453	0.145	0.021	0.819	0.671
tr-C5S3	1.683	1.700	0.114	0.013	0.303	0.092
tr-C4S1+C5S2	1.766	1.737	0.112	0.013	0.802	0.643
tr-C5P1	1.848	1.424	0.293	0.086	0.867	0.753
tr-C4S1+C5S0	1.993	1.975	0.471	0.222	1.388	1.928
tr-C5S1	2.047	1.941	0.310	0.096	1.384	1.916
sq-C5S4	2.077	2.302	0.171	0.029	0.484	0.234
sq-C4S2+C5S2	2.079	2.278	0.040	0.002	0.260	0.067
Cavity Type	$E^{DFT}_{ads,Ni}$ / eV	$E^{DFT}_{ads,Pd}$ / eV	MAE_{Ni} / eV	MSE_{Ni} / eV	MAE_{Pd} / eV	MSE_{Pd} / eV
sq-C4X0+C5X0	-4.505	-4.618	0.139	0.019	0.253	0.064
sq-C4N0+C5N1	-4.439	-4.622	0.215	0.046	0.312	0.097
sq-C4N1+C5N0	-4.327	-4.555	0.142	0.020	0.283	0.080
sq-C4P0+C5P1	-4.110	-4.388	0.706	0.499	0.241	0.058
sq-C4P0+C5P2	-3.943	-4.311	0.687	0.472	0.212	0.045
sq-C4N2+C5N0	-3.914	-4.133	0.303	0.092	0.534	0.285
sq-C4S0+C5S1	-3.847	-4.136	0.709	0.503	0.340	0.116
sq-C4N1+C5N1-s	-3.776	-4.030	0.193	0.037	0.365	0.133
sq-C4N1+C5N1-d	-3.663	-3.835	0.087	0.008	0.115	0.013
tr-C4X0	-3.549	-3.539	0.134	0.018	0.561	0.315
sq-C4S1+C5S0	-3.432	-3.752	0.327	0.107	0.155	0.024
sq-C4P1+C5P0	-3.184	-3.354	0.383	0.147	0.040	0.002
sq-C4N2+C5N1	-3.070	-3.243	0.159	0.025	0.444	0.197
sq-C4N1+C5N2	-2.911	-3.193	0.091	0.008	0.202	0.041
tr-C4N1	-2.763	-2.675	0.248	0.061	0.812	0.659
sq-C4P2+C5P0	-2.692	-2.783	0.865	0.749	0.160	0.026
sq-C4P1+C5P1-s	-2.662	-2.751	0.700	0.491	0.156	0.024
sq-C4N2+C5N2	-2.517	-2.660	0.139	0.019	0.109	0.012
sq-C4S0+C5S2	-2.279	-2.329	0.731	0.534	0.493	0.243
sq-C5N2-s2	-2.235	-2.241	1.397	1.952	0.473	0.224
sq-C5N2-d1	-2.222	-2.244	1.170	1.370	0.173	0.030
sq-C5N2-d2	-2.194	-2.229	1.198	1.436	0.188	0.035

Table S19: Continuation of Table S16.

Cavity Type	$E^{RFR}_{ads,Ni}$ / eV	$E^{RFR}_{ads,Pd}$ / eV	MAE_{Ni} / eV	MSE_{Ni} / eV	MAE_{Pd} / eV	MSE_{Pd} / eV
tr-C4N2	-2.103	-1.950	0.004	0.000	1.199	1.439
sq-C5N2-s1	-2.080	-2.116	0.807	0.651	0.178	0.032
sq-C5N1	-2.077	-2.089	0.700	0.490	0.452	0.204
sq-C4P1+C5P1-d	-2.063	-2.102	0.625	0.391	0.020	0.000
sq-C5N3	-2.060	-2.127	1.047	1.095	0.066	0.004
sq-C4P1+C5P2	-1.910	-1.872	0.417	0.174	0.470	0.221
tr-C4N3	-1.895	-1.447	0.095	0.009	1.549	2.399
sq-C5N4	-1.890	-1.890	0.691	0.477	0.221	0.049
sq-C4P2+C5P1	-1.773	-1.739	0.303	0.092	0.338	0.114
tr-C4X0+C5X0	-1.764	-1.460	0.194	0.038	0.391	0.153
tr-C5X0	-1.748	-1.498	0.180	0.032	0.432	0.186
sq-C4S2+C5S0	-1.732	-1.743	0.899	0.807	0.449	0.202
sq-C4N0+C5N2	-1.715	-1.622	1.179	1.390	0.788	0.620
sq-C4S1+C5S1-d	-1.677	-1.739	0.478	0.228	0.047	0.002
sq-C4S1+C5S1-s	-1.676	-1.722	0.582	0.339	0.333	0.111
tr-C4S1	-1.129	-1.061	0.117	0.014	0.378	0.143
sq-C5S1	-1.038	-1.002	1.005	1.011	0.249	0.062
sq-C5S2-s1	-0.880	-0.842	0.691	0.477	0.045	0.002
sq-C5X0	-0.797	-0.867	0.377	0.142	0.881	0.776
sq-C4P2+C5P2	-0.792	-0.767	0.452	0.204	0.036	0.001
sq-C5P1	-0.724	-0.701	0.140	0.020	0.552	0.305
sq-C4S1+C5S2	-0.501	-0.492	0.495	0.245	0.230	0.053
tr-C4N1+C5N0	-0.341	-0.064	0.196	0.039	0.765	0.585
tr-C4N0+C5N1	-0.322	-0.058	0.185	0.034	0.762	0.580
tr-C5N1	-0.307	-0.074	0.167	0.028	0.777	0.604
sq-C5P2-s1	-0.264	-0.436	1.002	1.003	0.137	0.019
sq-C5P4	-0.255	-0.119	0.630	0.397	0.168	0.028
sq-C5S2-d1	-0.175	-0.263	1.054	1.112	0.093	0.009
sq-C5S2-d2	-0.164	-0.228	1.065	1.134	0.125	0.016
sq-C4S2+C5S1	-0.076	-0.250	0.666	0.443	0.279	0.078
tr-C4N1+C5N1	-0.072	0.118	0.037	0.001	0.526	0.277
tr-C5N2	-0.037	0.186	0.001	0.000	0.457	0.209
sq-C5S2-s2	0.157	0.151	1.015	1.030	0.518	0.268
sq-C5S3	0.220	0.312	0.704	0.495	0.205	0.042
tr-C4P0+C5P1	0.308	0.583	0.280	0.078	0.302	0.091
tr-C4P1	0.368	0.505	0.136	0.018	0.928	0.861
sq-C5P3	0.399	0.348	0.090	0.008	0.701	0.491
tr-C4N0+C5N2	0.405	0.925	0.639	0.408	0.450	0.202
tr-C5N3	0.431	0.697	0.184	0.034	1.095	1.199
tr-C4N1+C5N2	0.561	0.862	0.734	0.539	1.140	1.299
tr-C4S0+C5S1	0.700	0.845	0.451	0.204	0.231	0.054
tr-C4P0	0.722	0.828	0.081	0.007	0.248	0.062
tr-C4P2	0.726	0.822	0.309	0.095	0.052	0.003
sq-C5P2-s2	0.798	0.799	0.442	0.196	0.066	0.004
tr-C4P1+C5P2	0.807	0.955	0.022	0.000	0.266	0.071
sq-C5P2-d1	0.881	0.892	0.343	0.118	0.371	0.137
sq-C5P2-d2	0.891	0.885	0.358	0.128	0.381	0.145
tr-C4S2	0.901	1.016	0.085	0.007	0.044	0.002
tr-C5P3	1.000	1.255	0.675	0.455	0.007	0.000
tr-C4S1+C5S1	1.210	1.468	0.496	0.246	0.279	0.078
tr-C4S0+C5S2	1.215	1.523	0.012	0.000	0.159	0.025

Table S20: Continuation of Table S16.

Cavity Type	$E^{RFR}_{ads,Ni}$ / eV	$E^{RFR}_{ads,Pd}$ / eV	MAE_{Ni} / eV	MSE_{Ni} / eV	MAE_{Pd} / eV	MSE_{Pd} / eV
tr-C5S2	1.218	1.414	0.708	0.502	0.100	0.010
tr-C4P0+C5P2	1.218	1.455	0.746	0.557	0.384	0.147
tr-C5P2	1.309	1.660	0.804	0.646	0.266	0.071
tr-C4S1+C5S0	1.560	1.998	1.664	2.768	0.628	0.394
tr-C5S3	1.579	1.917	0.570	0.325	0.053	0.003
tr-C5S1	1.601	1.928	1.288	1.658	0.700	0.489
tr-C4S1+C5S2	1.630	1.961	0.365	0.133	0.417	0.174
sq-C4S2+C5S2	2.008	2.129	0.651	0.424	0.203	0.041
sq-C5S4	2.090	2.255	0.512	0.262	0.058	0.003
tr-C4S0	2.162	2.593	0.333	0.111	0.218	0.048
Cavity Type	$E^{RFR}_{ads,Ru}$ / eV	$E^{RFR}_{ads,Pt}$ / eV	MAE_{Ru} / eV	MSE_{Ru} / eV	MAE_{Pt} / eV	MSE_{Pt} / eV
sq-C4X0+C5X0	-4.588	-4.586	1.050	1.103	1.379	1.903
sq-C4N0+C5N1	-4.508	-4.566	0.928	0.862	1.203	1.448
sq-C4N1+C5N0	-4.359	-4.428	1.051	1.106	1.274	1.623
sq-C4P0+C5P1	-4.268	-4.178	0.724	0.524	1.366	1.865
sq-C4S0+C5S1	-3.995	-3.911	0.832	0.692	1.282	1.643
sq-C4N2+C5N0	-3.973	-3.910	0.900	0.810	1.357	1.841
sq-C4N1+C5N1-s	-3.838	-3.781	0.947	0.897	1.252	1.568
sq-C4N1+C5N1-d	-3.682	-3.613	1.231	1.517	0.899	0.808
sq-C4S1+C5S0	-3.579	-3.364	1.459	2.128	0.941	0.886
tr-C4X0	-3.375	-3.212	0.339	0.115	0.235	0.055
sq-C4P1+C5P0	-3.149	-3.004	1.159	1.344	1.167	1.362
sq-C4N2+C5N1	-3.033	-2.956	0.710	0.504	1.027	1.055
sq-C4N1+C5N2	-3.015	-2.848	0.810	0.656	0.821	0.674
sq-C4P2+C5P0	-2.713	-2.564	0.560	0.313	0.872	0.760
sq-C4P1+C5P1-s	-2.696	-2.527	0.478	0.229	0.842	0.708
sq-C4N2+C5N2	-2.326	-2.348	0.636	0.405	0.515	0.265
sq-C4S0+C5S2	-2.190	-2.213	0.142	0.020	1.069	1.144
tr-C4N1	-2.108	-2.209	0.523	0.274	0.287	0.082
sq-C4P1+C5P1-d	-1.941	-1.843	0.782	0.612	1.213	1.471
sq-C4P1+C5P2	-1.834	-1.696	0.963	0.927	0.078	0.006
sq-C5N2-d1	-1.694	-1.550	0.220	0.048	1.159	1.342
sq-C5N2-s2	-1.691	-1.560	0.607	0.368	1.502	2.257
sq-C5N2-d2	-1.617	-1.532	0.142	0.020	1.176	1.384
sq-C4S1+C5S1-d	-1.565	-1.336	0.981	0.963	1.084	1.176
sq-C4P2+C5P1	-1.553	-1.548	0.360	0.129	0.176	0.031
sq-C4S2+C5S0	-1.512	-1.322	1.294	1.675	1.221	1.491
sq-C5N3	-1.458	-1.421	1.056	1.116	0.935	0.874
sq-C5N2-s1	-1.449	-1.363	1.228	1.508	0.891	0.794
sq-C4S1+C5S1-s	-1.435	-1.361	0.450	0.202	1.081	1.170
sq-C5N1	-1.363	-1.248	0.164	0.027	0.998	0.995
tr-C4X0+C5X0	-1.344	-1.325	0.232	0.054	0.195	0.038
tr-C5X0	-1.342	-1.353	0.233	0.054	0.225	0.051
sq-C5N4	-1.142	-1.053	1.197	1.433	0.700	0.490
sq-C4N0+C5N2	-1.094	-1.078	0.736	0.541	0.603	0.364
tr-C4N2	-0.752	-0.878	0.504	0.254	0.402	0.161
sq-C5S2-s1	-0.665	-0.499	0.838	0.701	0.583	0.340
sq-C4P2+C5P2	-0.631	-0.585	0.067	0.005	0.445	0.198
tr-C4S1	-0.543	-0.697	0.487	0.237	0.068	0.005
sq-C5S1	-0.530	-0.510	0.202	0.041	1.324	1.752

Table S21: Continuation of Table S16.

Cavity Type	$E^{RFR}_{ads,Ru}$ / eV	$E^{RFR}_{ads,Pt}$ / eV	MAE_{Ru} / eV	MSE_{Ru} / eV	MAE_{Pt} / eV	MSE_{Pt} / eV
sq-C5X0	-0.367	-0.441	0.231	0.054	0.474	0.224
sq-C5P1	-0.291	-0.426	0.587	0.345	0.182	0.033
sq-C4S1+C5S2	-0.257	-0.280	0.886	0.785	0.551	0.303
sq-C5P2-s1	-0.006	-0.010	0.053	0.003	0.867	0.751
sq-C4S2+C5S1	0.024	0.135	0.686	0.471	0.661	0.437
sq-C5P4	0.026	0.083	0.220	0.048	0.696	0.484
sq-C5S2-d1	0.175	0.267	1.002	1.005	0.977	0.954
sq-C5S2-d2	0.179	0.300	1.000	0.999	1.008	1.016
tr-C4N3	0.222	-0.049	0.292	0.085	0.894	0.799
tr-C4N0+C5N1	0.544	0.638	0.222	0.049	0.410	0.168
tr-C4N1+C5N0	0.545	0.640	0.229	0.053	0.403	0.162
tr-C5N1	0.551	0.692	0.233	0.054	0.356	0.127
sq-C5P3	0.576	0.543	0.284	0.081	0.549	0.301
sq-C5S2-s2	0.614	0.686	0.573	0.328	1.402	1.964
sq-C5S3	0.642	0.643	0.379	0.143	0.585	0.342
tr-C4P1	0.961	0.791	0.448	0.201	0.424	0.180
tr-C4N1+C5N1	1.035	1.013	0.257	0.066	0.547	0.299
tr-C5N2	1.107	1.049	0.332	0.110	0.512	0.262
sq-C5P2-s2	1.158	1.107	0.297	0.088	0.586	0.343
tr-C4P0	1.243	1.042	0.742	0.551	0.224	0.050
tr-C4P0+C5P1	1.265	1.034	0.141	0.020	0.253	0.064
tr-C4P2	1.281	1.081	0.441	0.195	0.240	0.058
sq-C5P2-d1	1.314	1.138	0.404	0.164	0.044	0.002
sq-C5P2-d2	1.319	1.147	0.411	0.169	0.061	0.004
tr-C4P1+C5P2	1.464	1.218	0.793	0.628	0.225	0.050
tr-C4S0+C5S1	1.487	1.149	1.016	1.031	0.230	0.053
tr-C4S2	1.736	1.341	0.126	0.016	0.242	0.059
tr-C5P3	1.839	1.429	0.112	0.012	0.262	0.069
tr-C4P0+C5P2	2.105	1.623	0.352	0.124	0.937	0.879
tr-C4S1+C5S0	2.203	2.245	0.110	0.012	0.791	0.626
tr-C5S2	2.248	1.801	0.156	0.024	0.533	0.284
tr-C4S1+C5S1	2.258	1.803	0.031	0.001	0.278	0.077
tr-C4S0+C5S2	2.323	1.844	0.597	0.357	0.141	0.020
tr-C5S1	2.328	2.267	0.017	0.000	0.069	0.005
tr-C5P2	2.440	1.733	0.702	0.492	0.547	0.300
sq-C4S2+C5S2	2.465	2.300	0.209	0.044	0.111	0.012
sq-C5S4	2.645	2.512	0.681	0.464	0.285	0.081
tr-C5S3	2.769	2.234	0.087	0.008	0.067	0.004
tr-C4S1+C5S2	2.774	2.218	0.351	0.123	1.235	1.524
tr-C4S0	3.779	3.109	0.966	0.934	0.611	0.374

References

(1) O'Connor, N. J.; Jonayat, A. S.; Janik, M. J. et al. Interaction trends between single metal atoms and oxide supports identified with density functional theory and statistical learning. *Nat. Catal.* **2018**, *1*, 531–539.

(2) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J. et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(3) Miedema, A. A simple model for alloys. *Philips tech. Rev.* **1973**, *33*, 6.

(4) Geiger, J.; Sabadell-Rendón, A.; Daelman, N. et al. Data-driven models for ground and excited states for Single Atoms on Ceria. *npj Comput. Mater.* **2022**, *8*, 171.

(5) Hai, X.; Xi, S.; Mitchell, S. et al. Scalable two-step annealing method for preparing ultra-high-density single-atom catalyst libraries. *Nat. Nanotechnol.* **2022**, *17*, 174–181.