Supplementary Information for p-d Orbital Coupling in Silicon-Based Dual-Atom Catalysts for Enhanced CO_2 Reduction: Insight into Electron Regulation of Active Center and Coordination Atoms

Meijie Wang,[†] Yaowei Xiang,[†] Yuxing Lin,[†] Yang Sun,[†] Zi-zhong Zhu,[†] Shunqing Wu,[†] and Xinrui Cao[*,†,‡]

†*Department of Physics, Xiamen University, Xiamen 361005, China*

‡*Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Xiamen University, Xiamen 361005, China*

E-mail: xinruicao@xmu.edu.cn

Computational Details

The number of p_z electrons (N_{p_z}) is estimated by the following equation:

$$N_{p_z} = \int_{-\infty}^{0} D(E)\mathrm{d}E \tag{1}$$

Considering that the transition metal d orbital is localized, whereas the Si p orbital is delocalized, we have set different intervals. The d band center (ε_d) is estimated by the following equation:

$$\varepsilon_d = \frac{\int_{-6.4}^{+5.4} E \cdot D(E)\mathrm{d}E}{\int_{-6.4}^{+5.4} D(E)\mathrm{d}E} \tag{2}$$

The p_z band center (ε_{p_z}) is estimated by the following equation:

$$\varepsilon_{p_z} = \frac{\int_{-\infty}^{+\infty} E \cdot D(E)\mathrm{d}E}{\int_{-\infty}^{+\infty} D(E)\mathrm{d}E} \tag{3}$$

Different structures have different energy intervals, so we normalize ε_{p_z} to obtain ε'_{p_z}:

$$\varepsilon'_{p_z} = \frac{1}{\int_{-\infty}^{+\infty} E\mathrm{d}E} \cdot \varepsilon_{p_z} = \frac{\int_{-\infty}^{+\infty} E \cdot D(E)\mathrm{d}E}{\int_{-\infty}^{+\infty} D(E)\mathrm{d}E \cdot E\mathrm{d}E} \tag{4}$$

where D(E) is the density of state (DOS) of the band at a given energy E. The Fermi level (E_F) is calibrated to $E_F = 0$.

Machine Learning

First, the subset consisting of 85 $SiTMN_5An$ structures with 24 different transition metals across 12 coordination environments was employed to train for predicting ΔG_{*CO_2} and ΔG_{*OH}. The distribution of ΔG_{*CO_2} and ΔG_{*OH} of $SiTMN_5An$ DACs in the DFT calculated database of this work is plotted in Figure S1(a)(b), respectively. The ΔG_{*CO_2} ranges from -1.94 to 0.24 eV, while ΔG_{*OH} ranges from -2.00 to -0.65 eV. The ΔG_{*CO_2} and ΔG_{*OH} values are distributed in a wide range, which is beneficial for the machine learning model to learn

the relationship between the structure and adsorption free energy.

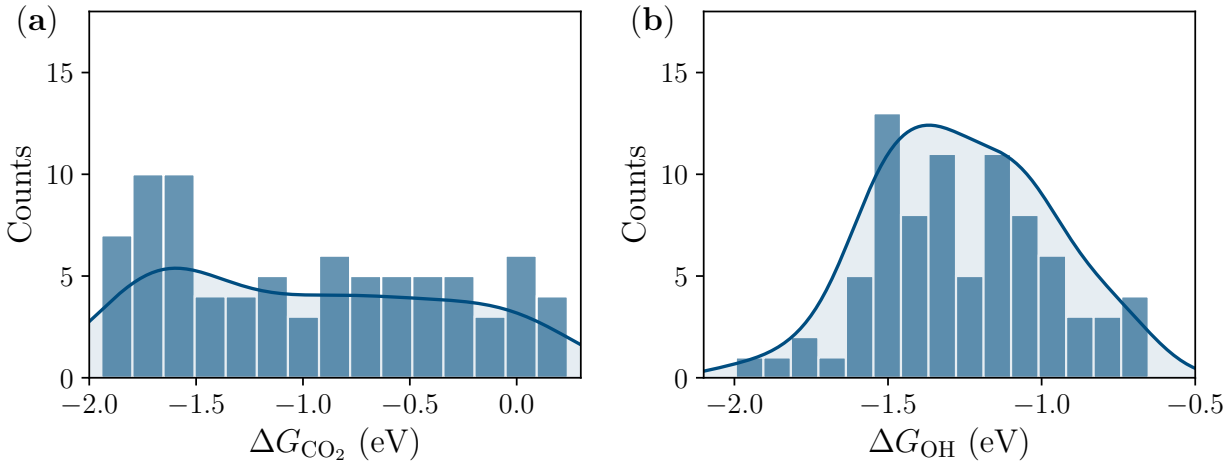

Figure S1: Histogram of (a) ΔG_{*CO_2} and (b) ΔG_{*OH} in the DFT calculated database of $SiTMN_5An$ DACs, respectively.

Our primary target was the adsorption free energy (ΔG) on $SiTMN_5An$ DACs. To predict ΔG, we incorporated 22 input features relating to the structure and various atomic properties that may influence adsorption free energy. These features included 14 attributes pertaining to transition metals and 8 derived from one-hot encoding of coordination environments, as detailed in Table S1.

Table S1: 22 Initial Features with their Description

Features	**Description**	**Features**	**Description**
N	Atomic number	C	Atomic cycle
G	Atomic group	M	Atomic weight
R	Atomic radii	α	Polarizability
E_A	Electron affinity	χ	Electronegativity
E_l	Lonization energy	N_l	Number of lone pair electrons
N_s	Number of s electrons	N_p	Number of p electrons
N_d	Number of d electrons	N_V	Number of valence electrons
D_B	Doping with B atom	D_C	Doping with C atom
P_1	Doping of position 1	P_2	Doping of position 2
P_3	Doping of position 3	P_4	Doping of position 4
P_5	Doping of position 4	P_6	Doping of position 6

To refine our feature set, we first analyzed the Pearson correlation coefficients among

the 14 non-coordination features, as shown in Figure S2. We identified and removed five redundant features (N, C, G, α and N_p) that exhibited correlations exceeding 0.8, reducing the feature count to 17. The Recursive Feature Elimination[1] (RFE) method, supported by the Gradient Boosting Regression (GBR) algorithm, was then employed to further minimize the feature set. Root mean squared error (RMSE) and determination coefficient (R^2), averaged from 10 training runs, guided the iterative removal of features, a process depicted in Figure S3.

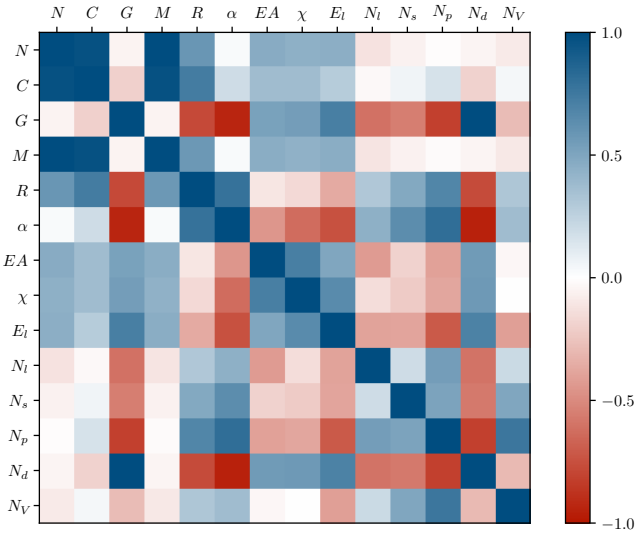

Figure S2: The Pearson correlation coefficients of the 14 features about transition metal atoms.

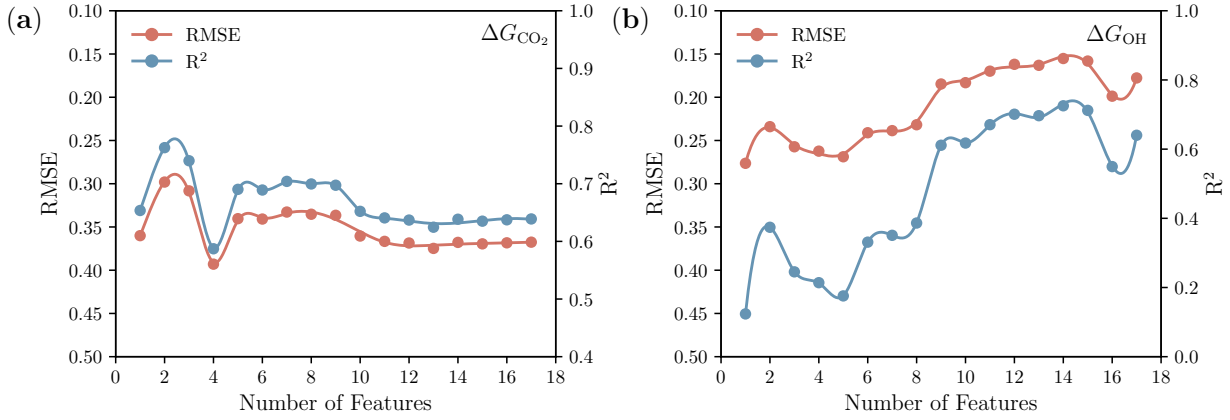

Figure S3: Extraction of important features by RFE for (a) $\Delta G_{*\mathrm{CO_2}}$ and (b) $\Delta G_{*\mathrm{OH}}$.

Combining the results of RFE and considering the interpretability and accuracy of the

model, ultimately, the features N_d, R, and eight coordination-related features were selected for the final model, as listed in Table S2. These features did not require additional density functional theory (DFT) calculations, thereby facilitating the exploration of a broader chemical space.

Table S2: 10 Features Used in the Final Model in Predicting $\Delta G_{*\mathrm{CO_2}}$ and $\Delta G_{*\mathrm{OH}}$

Features	**Description**	**Features**	**Description**
N_d	Number of d electrons	R	Atomic radii
D_B	Doping with B atom	D_C	Doping with C atom
P_1	Doping of position 1	P_2	Doping of position 2
P_3	Doping of position 3	P_4	Doping of position 4
P_5	Doping of position 4	P_6	Doping of position 6

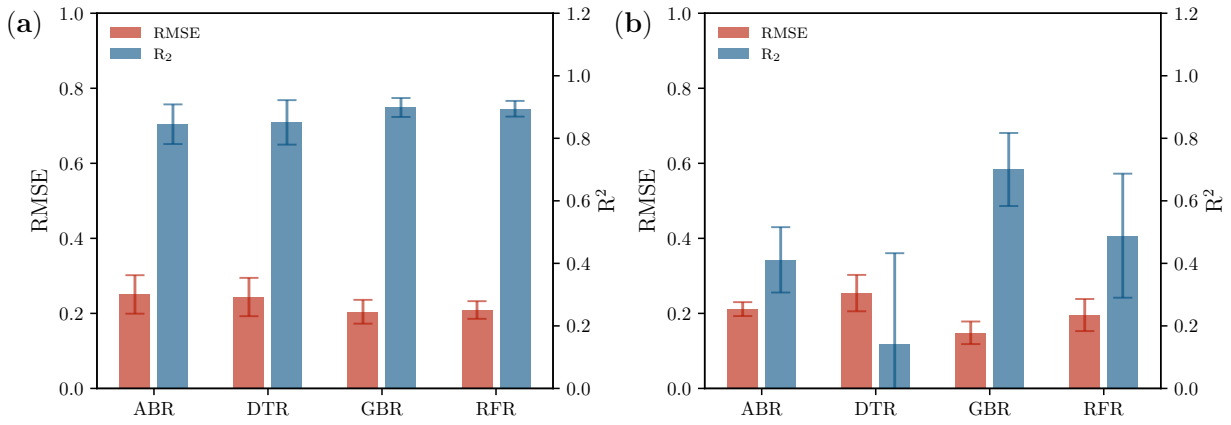

Figure S4: Performance of the models constructed through different algorithms: ABR, DTR, GBR and RFR in predicting (a) $\Delta G_{*\mathrm{CO_2}}$ and (b) $\Delta G_{*\mathrm{OH}}$.

In terms of machine learning models, we evaluated four regression algorithms from the scikit-learn[1] library: Adaptive Boosting Regressor (ABR), Decision Tree Regressor (DTR), Gradient Boosted Regression (GBR) and Random Forest Regressor (RFR). These models were assessed based on their RMSE and R^2, with results showcased in Figure S4. A fivefold cross-validation on the training dataset was conducted to fine-tune the hyperparameters. Although most models performed adequately for predicting $\Delta G_{*\mathrm{CO_2}}$, only the GBR model demonstrated robust performance for $\Delta G_{*\mathrm{OH}}$, making it the preferred choice for subsequent ML modeling.

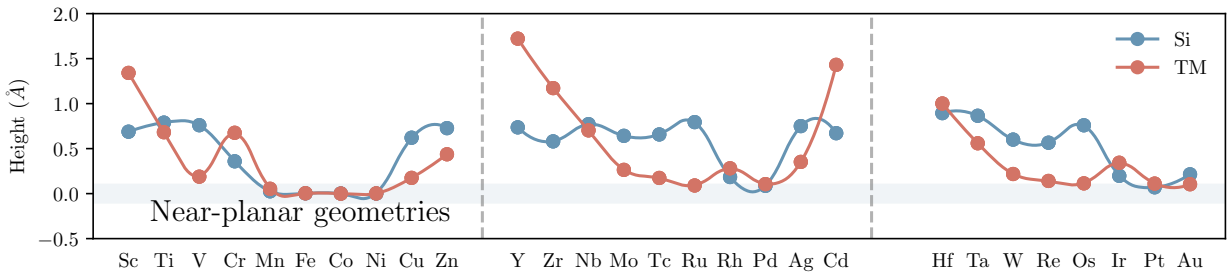

Figure S5: The height of Si and TM atoms away from the graphene plane in different $SiTMN_6$ DACs.

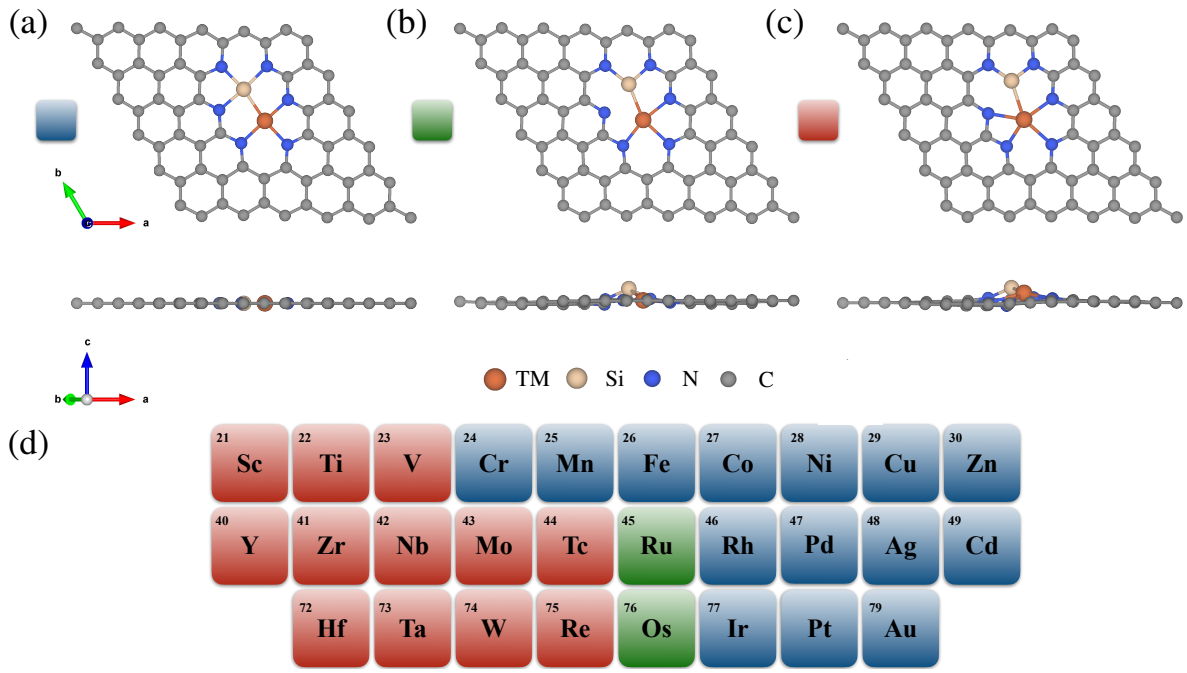

Figure S6: Three possible $SiTMN_6$ DACs structures. And distribution heatmap of different structures.

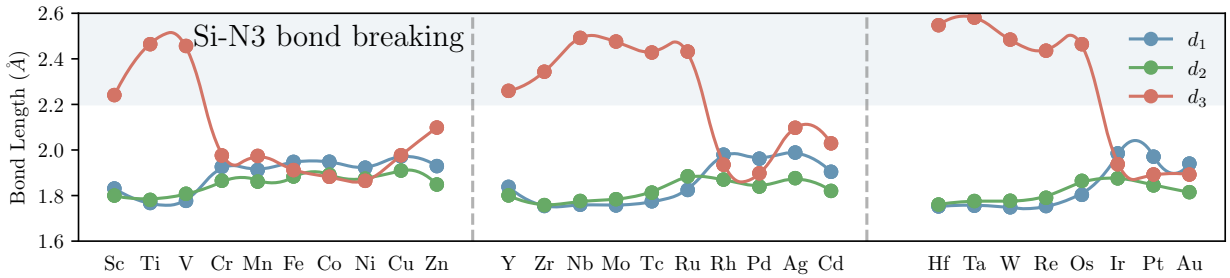

Figure S7: The bond length between Si atoms and N atoms on different $SiTMN_6$ DACs.

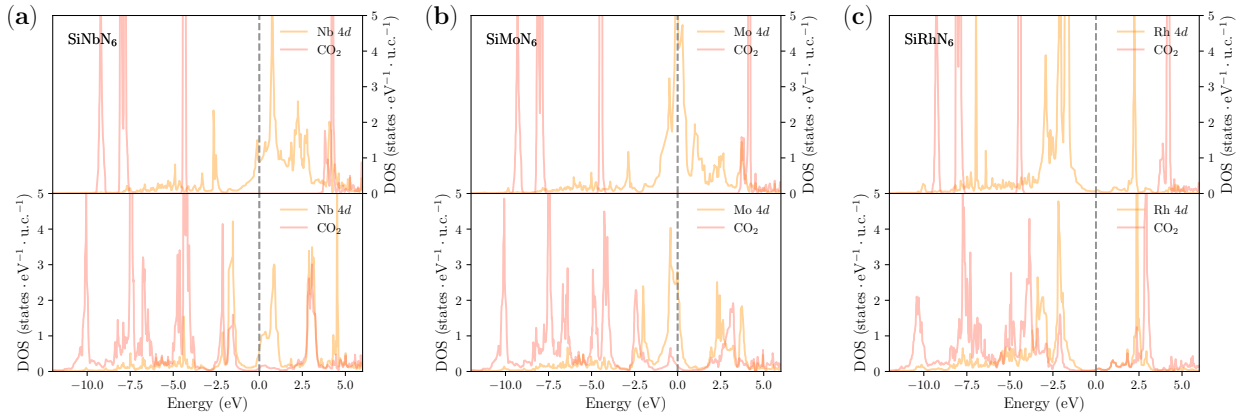

Figure S8: Before and after adsorption, the partial density of states (PDOS) for the TM d and CO_2 orbitals of $SiNbN_6$, $SiMoN_6$ and $SiRhN_6$ DACs, respectively. The top image represents the state before adsorption, while the bottom image shows the state after adsorption.

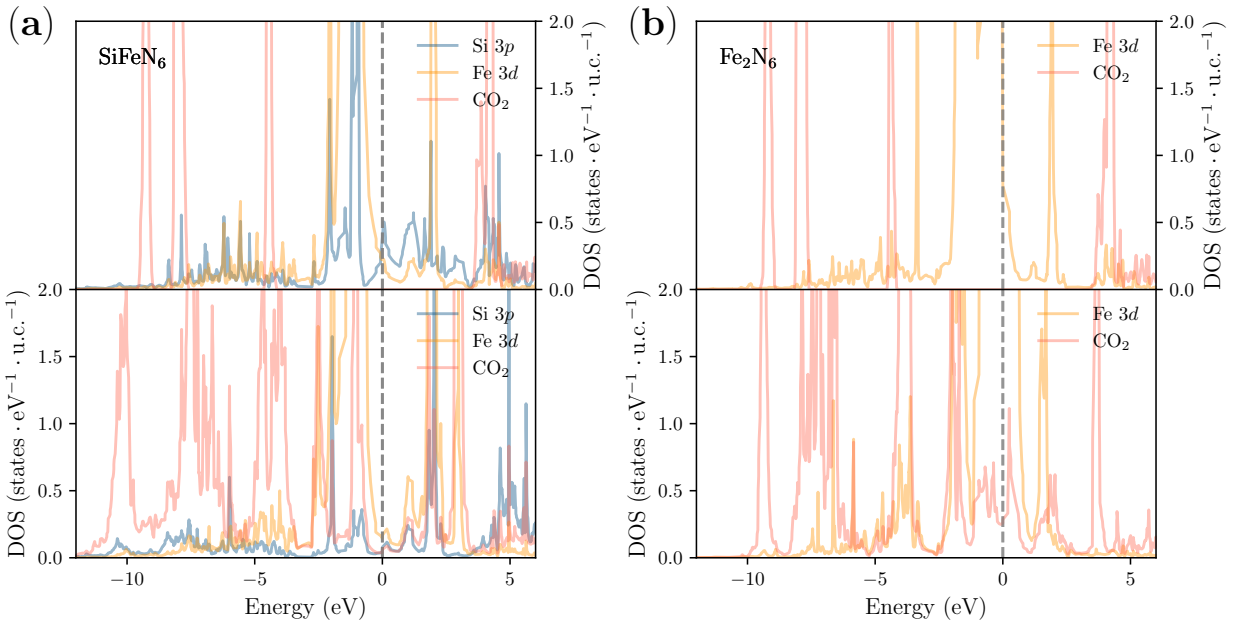

Figure S9: Before and after adsorption, the PDOS for the Si $3p$, Fe $3d$ and CO_2 orbitals of $SiFeN_6$ and Fe_2N_6 DACs, respectively. The top image represents the state before adsorption, while the bottom image shows the state after adsorption.

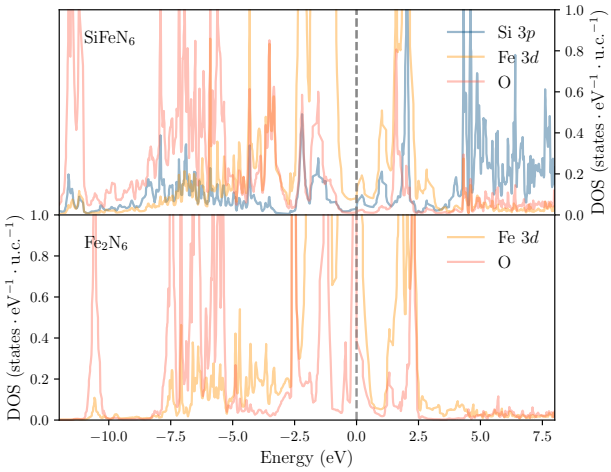

Figure S10: The PDOS for the Si 3p, Fe 3d and O in *CHO on $SiFeN_6$ DACs. And the PDOS for the Fe 3d and O in *CHO on Fe_2N_6 DACs.

Table S3: $\Delta G_{\max}$ of 27 $SiTMN_6$ DACs, Product and Potential-determining Step.

Group	**TM**	$\boldsymbol{\Delta G_{\max}}$	**Product**	**Potential-determining Step**
IIIB	Sc	0.33	CH_4	$^*OH \rightarrow ^* H_2O$
	Y	0.26	CH_4	$^*OH \rightarrow ^* H_2O$
IVB	Ti	0.41	CH_4	$^*OH \rightarrow ^* H_2O$
	Zr	0.43	CH_4	$^*OH \rightarrow ^* H_2O$
	Hf	0.44	CH_4	$^*OH \rightarrow ^* H_2O$
VB	V	0.49	CH_4	$^*OH \rightarrow ^* H_2O$
	Nb	0.46	CH_4	$^*OH \rightarrow ^* H_2O$
	Ta	0.55	CH_4	$^*OH \rightarrow ^* H_2O$
VIB	Cr	0.53	CH_4	$^*OH \rightarrow ^* H_2O$
	Mo	0.50	CH_4	$^*OH \rightarrow ^* H_2O$
	W	0.50	CH_4	$^*OH \rightarrow ^*H_2O$
VIIB	Mn	0.33	CH_4	$^*OH \rightarrow ^* H_2O$
	Tc	0.59	CH_4	$^*CO +^* OH \rightarrow ^* CO +^* H_2O$
	Re	0.58	CH_4	$^*CO +^* OH \rightarrow ^* CO +^* H_2O$
VIII	Fe	0.28	CH_4	$^*OH \rightarrow ^*H_2O$
	Ru	0.53	CH_4	$^*OH \rightarrow ^* H_2O$
	Os	0.52	CH_4	$^*OH \rightarrow ^* H_2O$
	Co	0.17	CH_3OH	$^*OH \rightarrow ^* H_2O$
	Rh	0.13	CH_3OH	$^*OH \rightarrow ^* H_2O$
	Ir	0.08	CH_4	$^*OH \rightarrow ^* H_2O$
	Ni	0.26	CH_4	$^*OH \rightarrow ^* H_2O$
	Pd	0.22	CH_4	$^*OH \rightarrow ^* H_2O$
	Pt	0.18	CH_4	$^*OCH_2OH \rightarrow ^* OCH_2OH_2$
IB	Cu	0.54	CH_4	$^*OH \rightarrow ^* H_2O$
	Ag	0.52	CH_4	$^*OH \rightarrow ^* H_2O$
IIB	Zn	0.23	CH_4	$^*OH \rightarrow ^* H_2O$
	Cd	0.32	CH_4	$^*OH \rightarrow ^* H_2O$

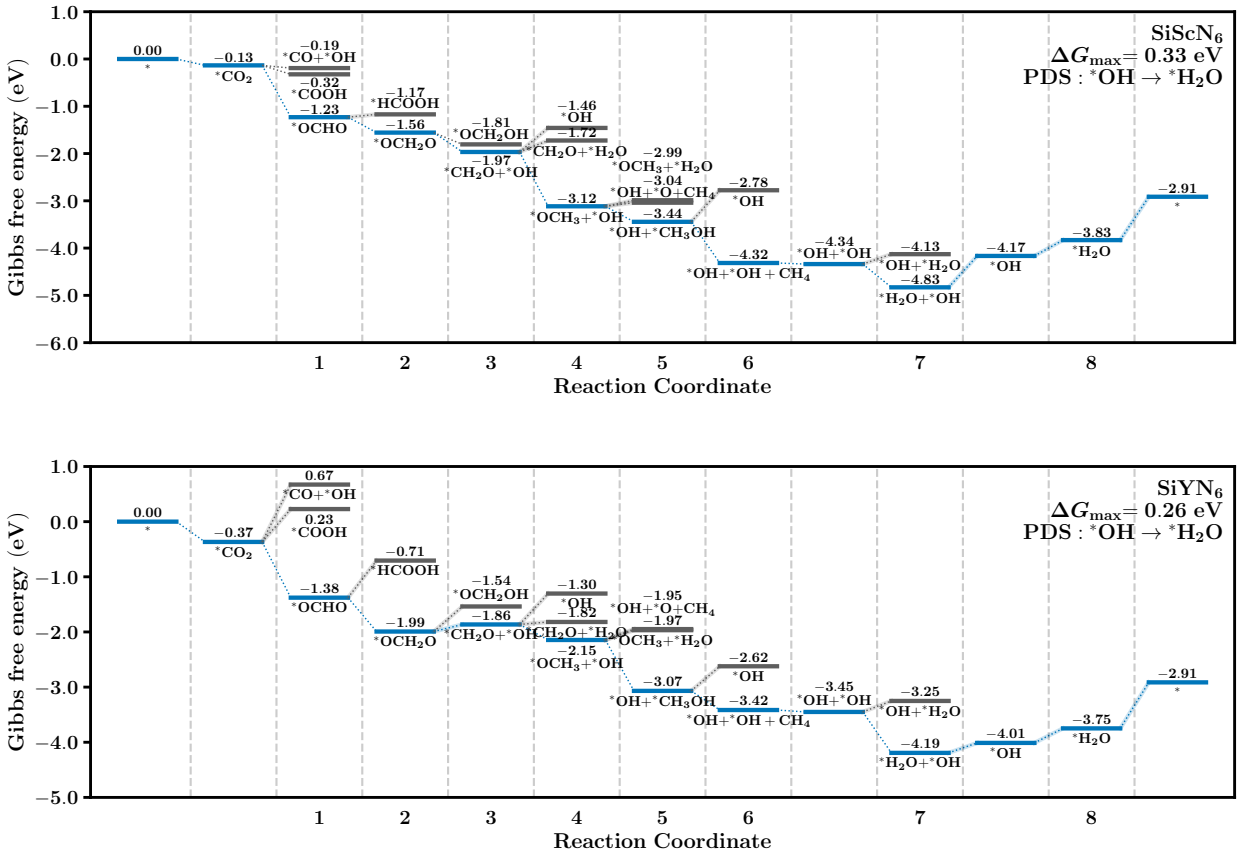

Figure S11: The relative Gibbs free energies in CO_2RR on $SiScN_6$ and $SiYN_6$ DACs under the applied potential of 0 V.

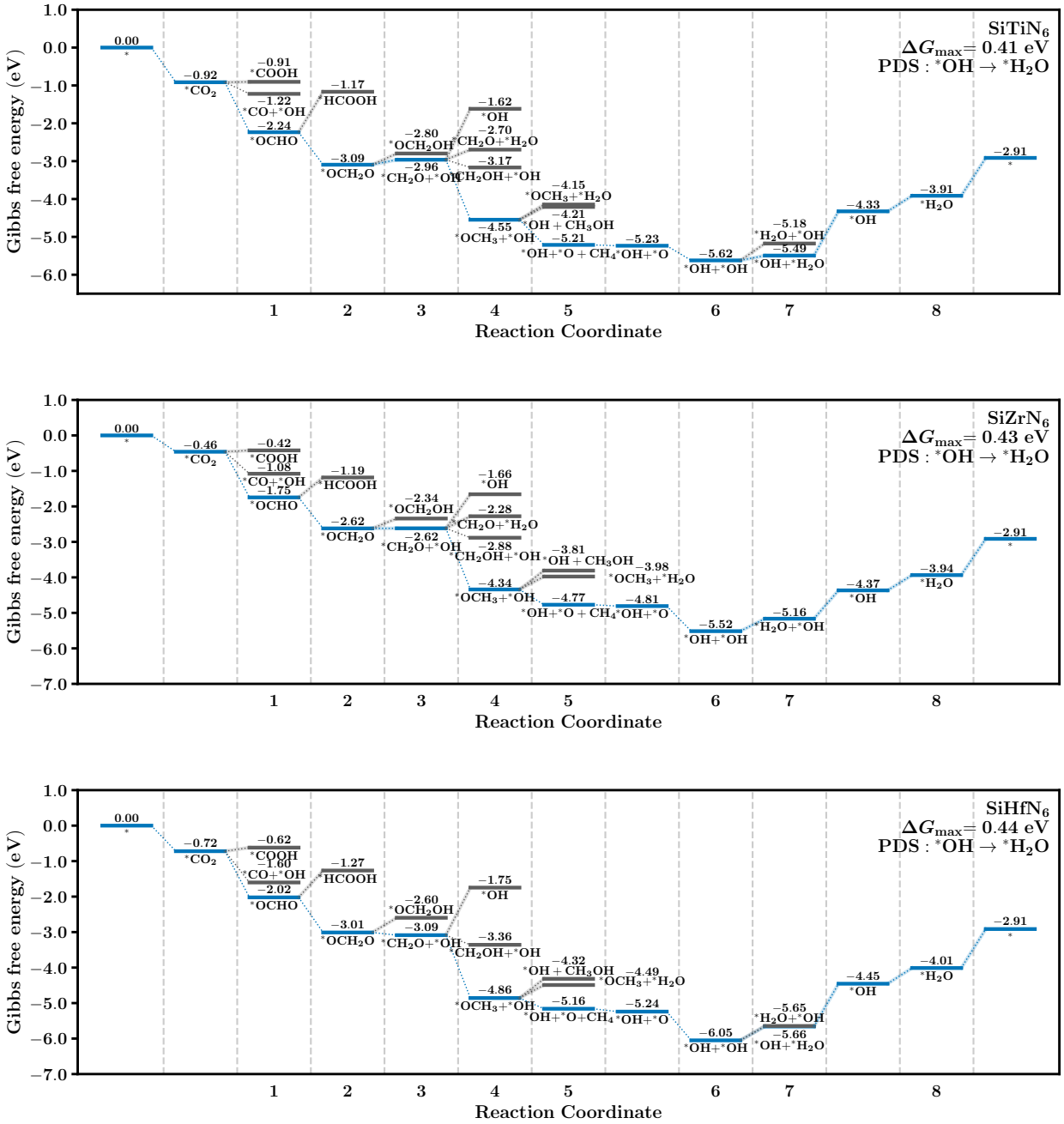

Figure S12: The relative Gibbs free energies in CO_2RR on $SiTiN_6$, $SiZrN_6$ and $SiHfN_6$ DACs under the applied potential of 0 V.

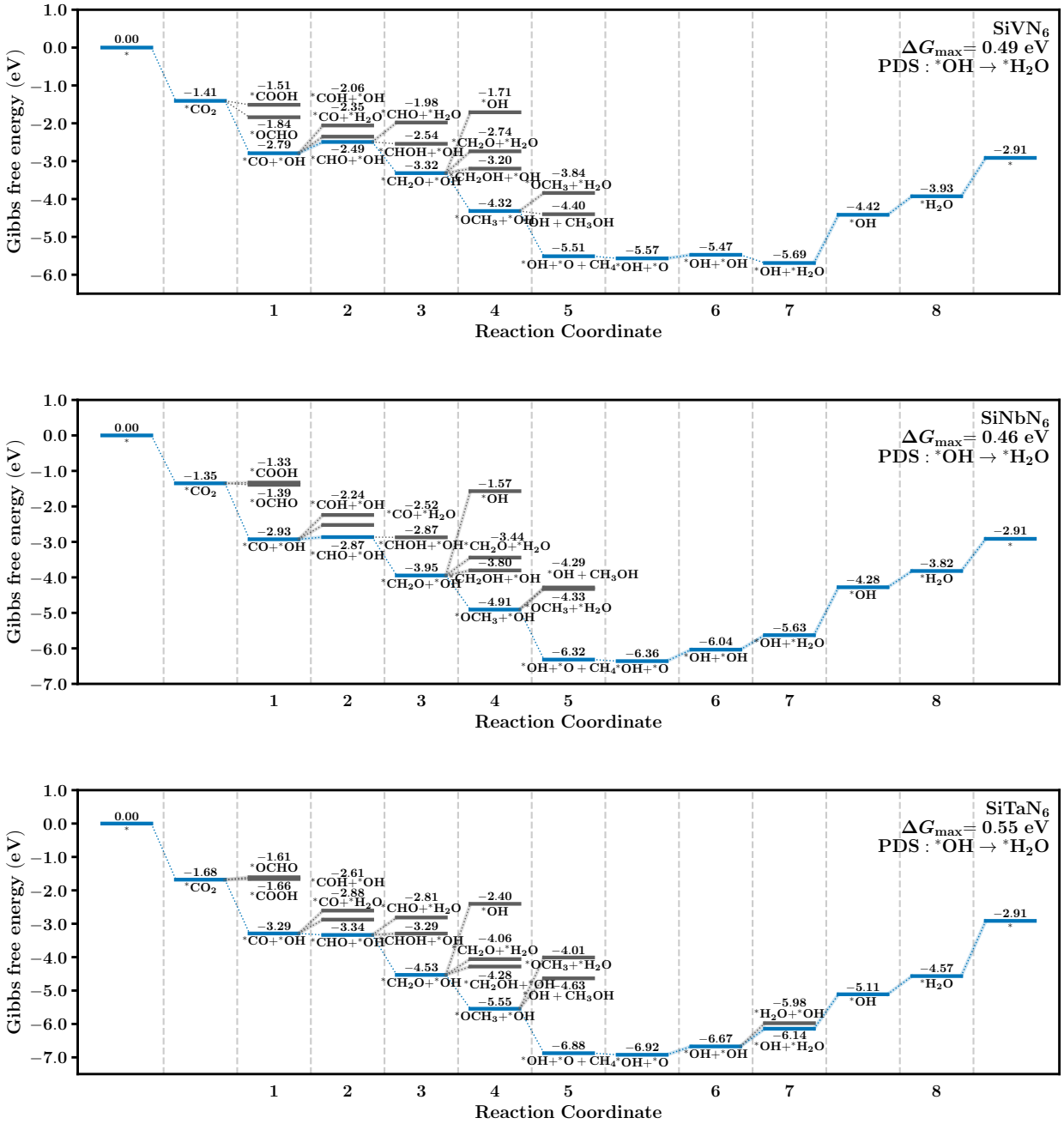

Figure S13: The relative Gibbs free energies in CO_2RR on $SiVN_6$, $SiNbN_6$ and $SiTaN_6$ DACs under the applied potential of 0 V.

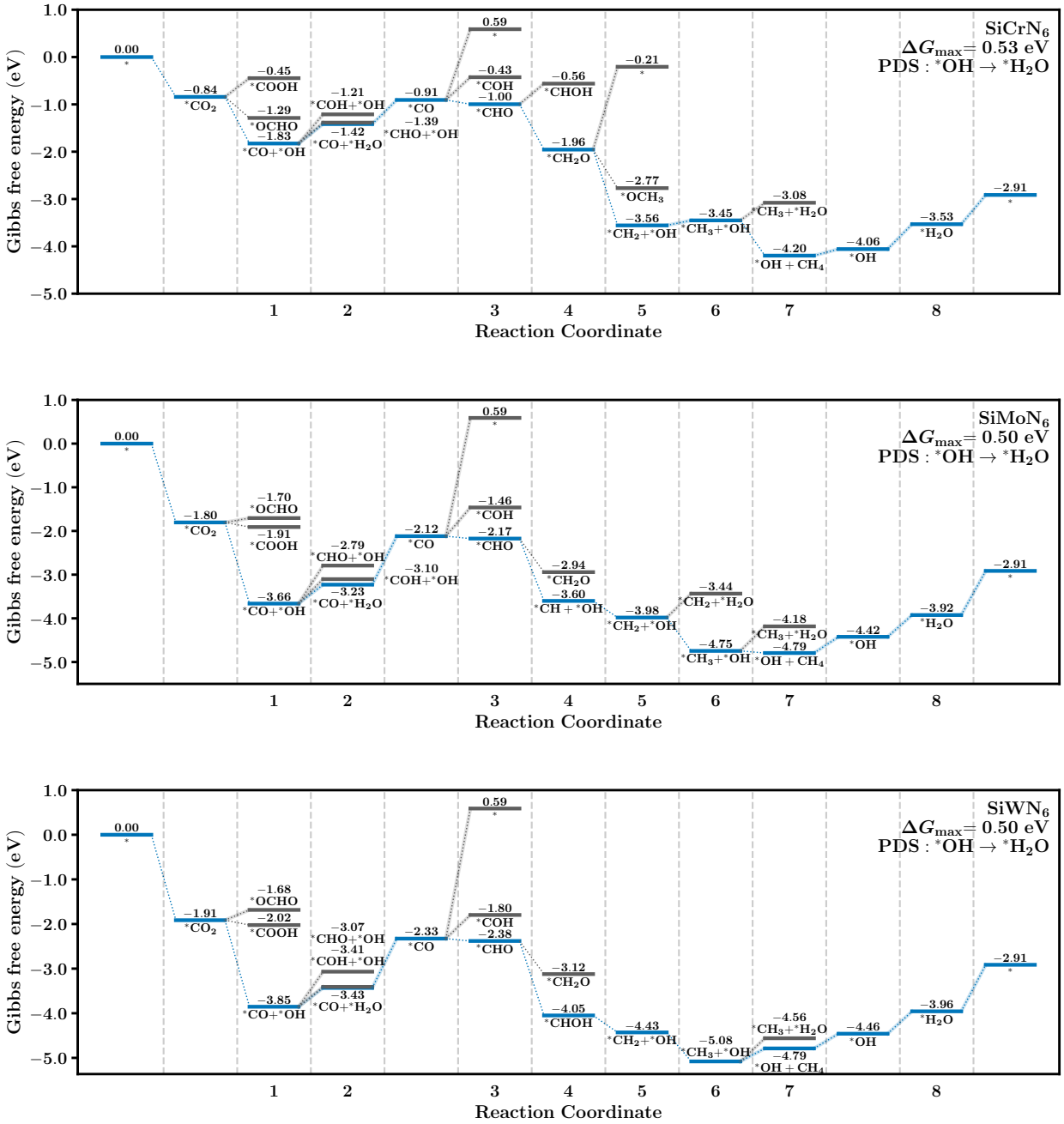

Figure S14: The relative Gibbs free energies in CO_2RR on $SiCrN_6$, $SiMoN_6$ and $SiWN_6$ DACs under the applied potential of 0 V.

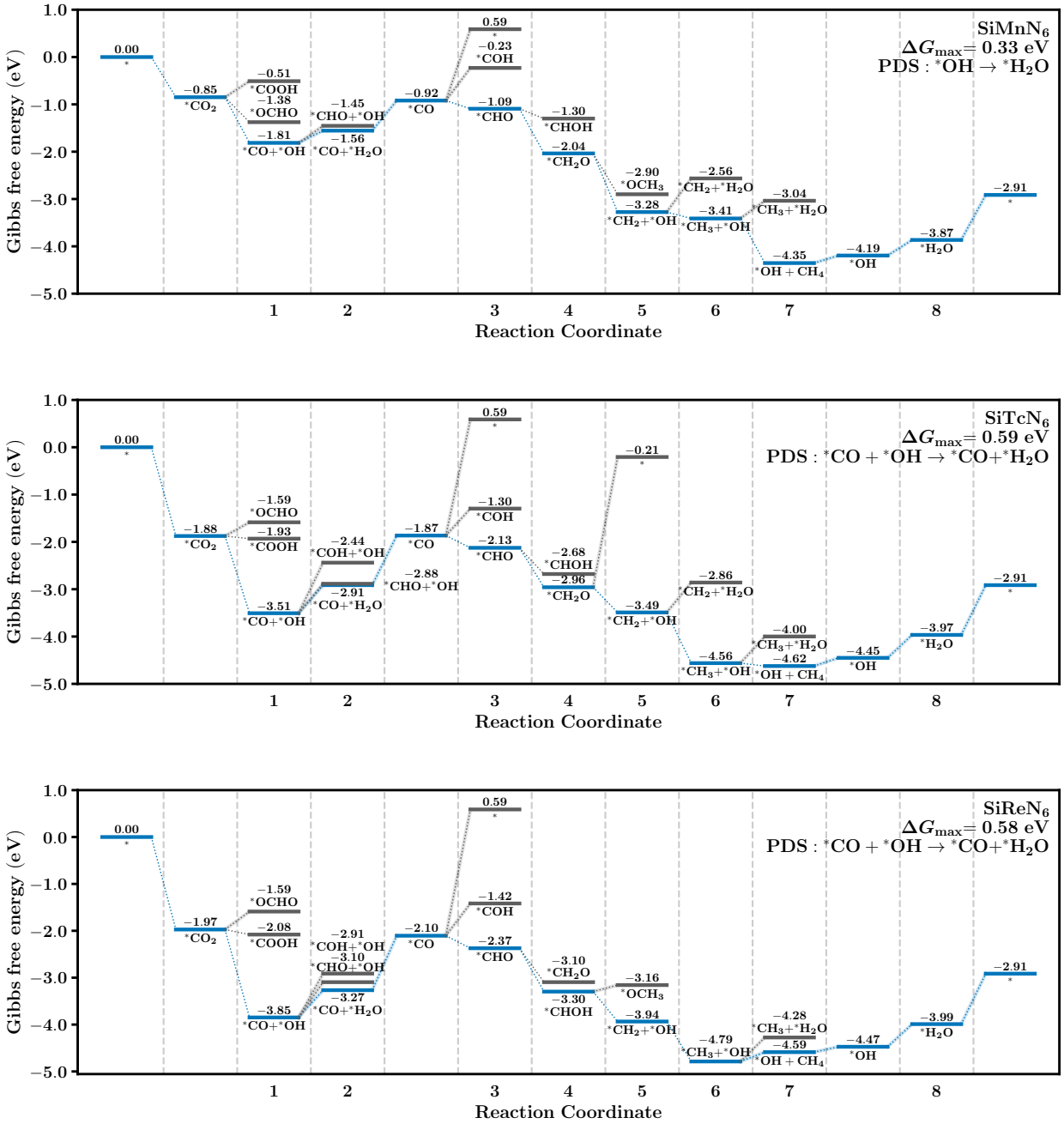

Figure S15: The relative Gibbs free energies in CO_2RR on $SiMnN_6$, $SiTcN_6$ and $SiReN_6$ DACs under the applied potential of 0 V.

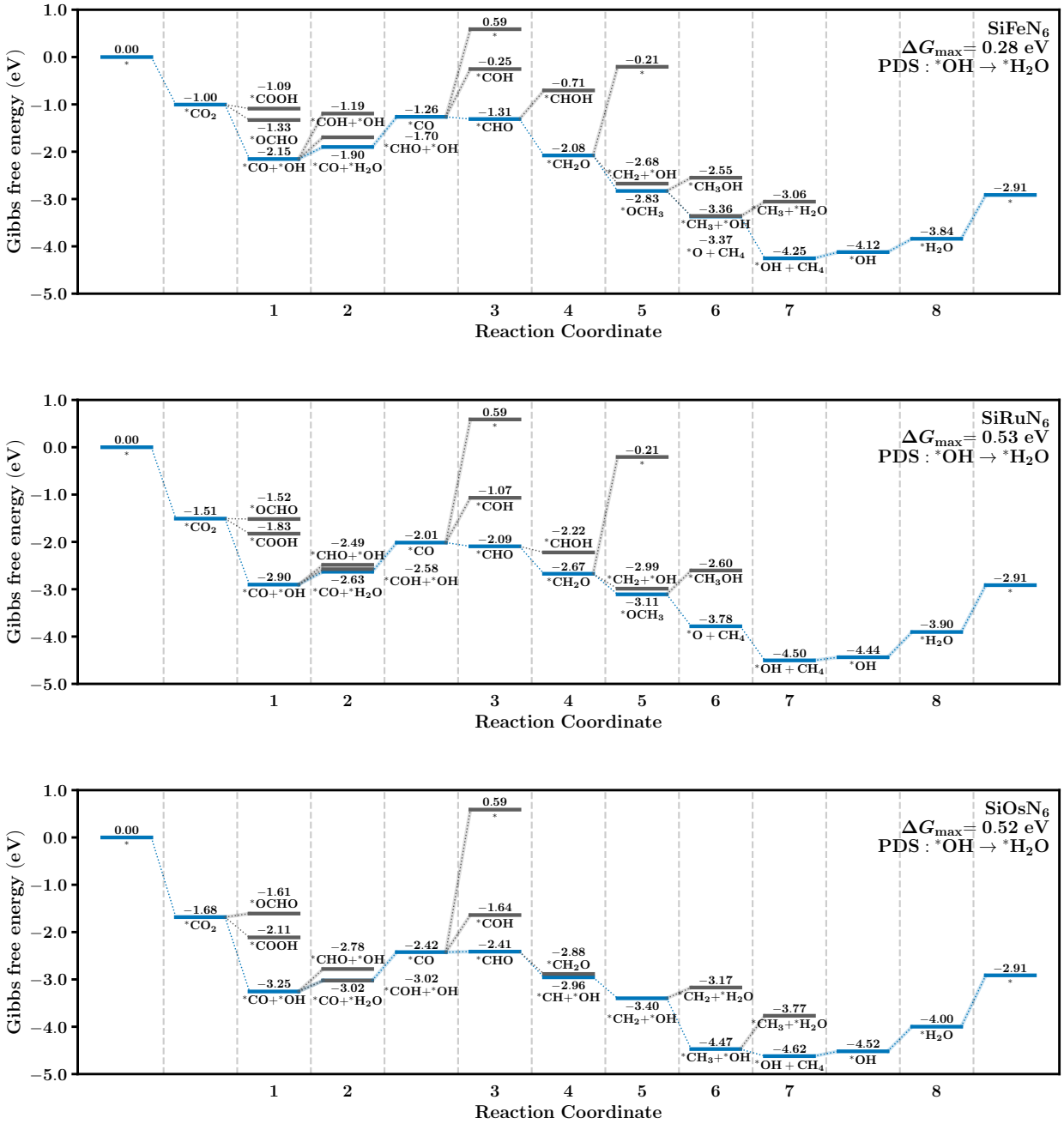

Figure S16: The relative Gibbs free energies in CO_2RR on $SiFeN_6$, $SiRuN_6$ and $SiOsN_6$ DACs under the applied potential of 0 V.

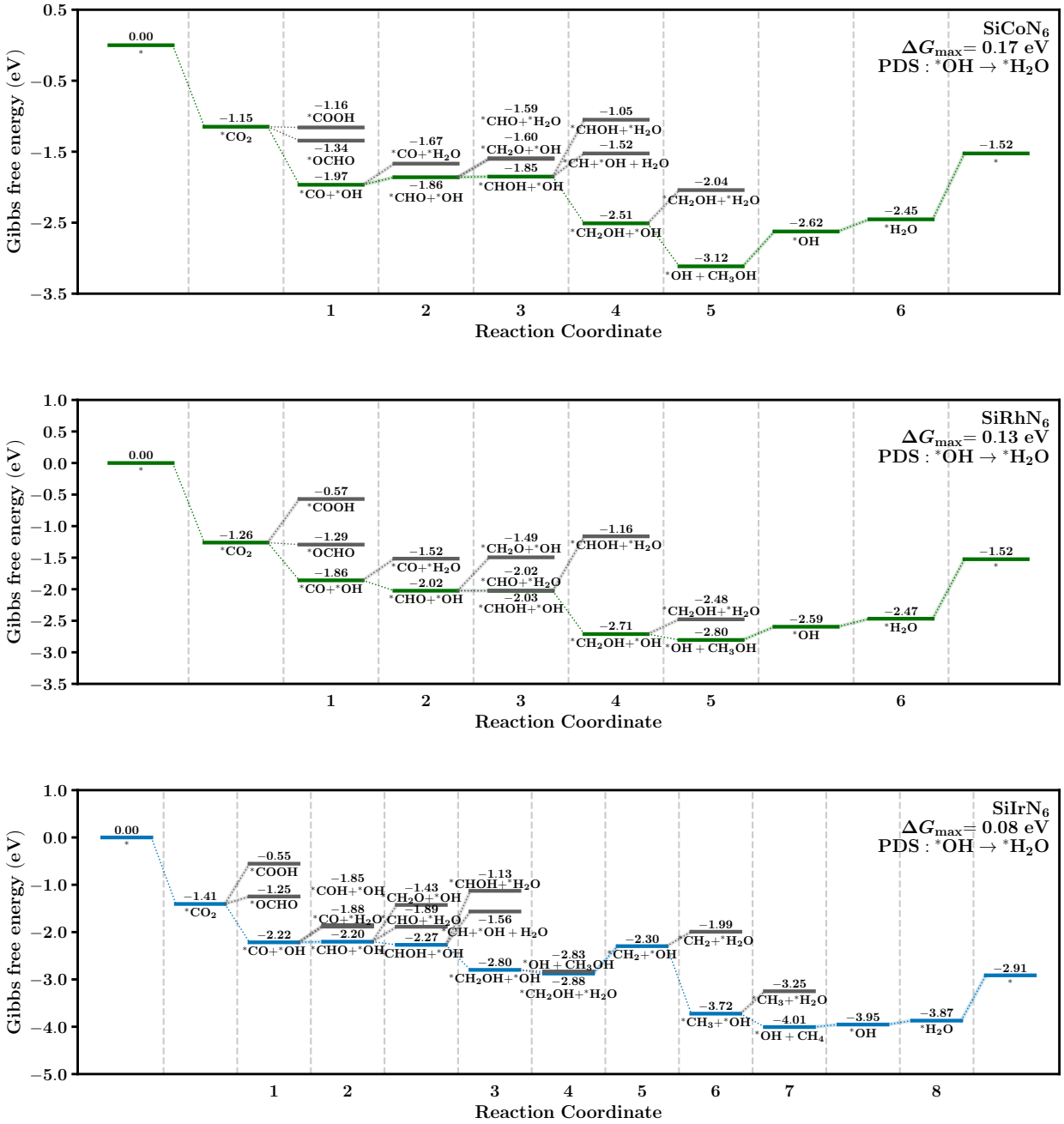

Figure S17: The relative Gibbs free energies in CO_2RR on $SiCoN_6$, $SiRhN_6$ and $SiIrN_6$ DACs under the applied potential of 0 V.

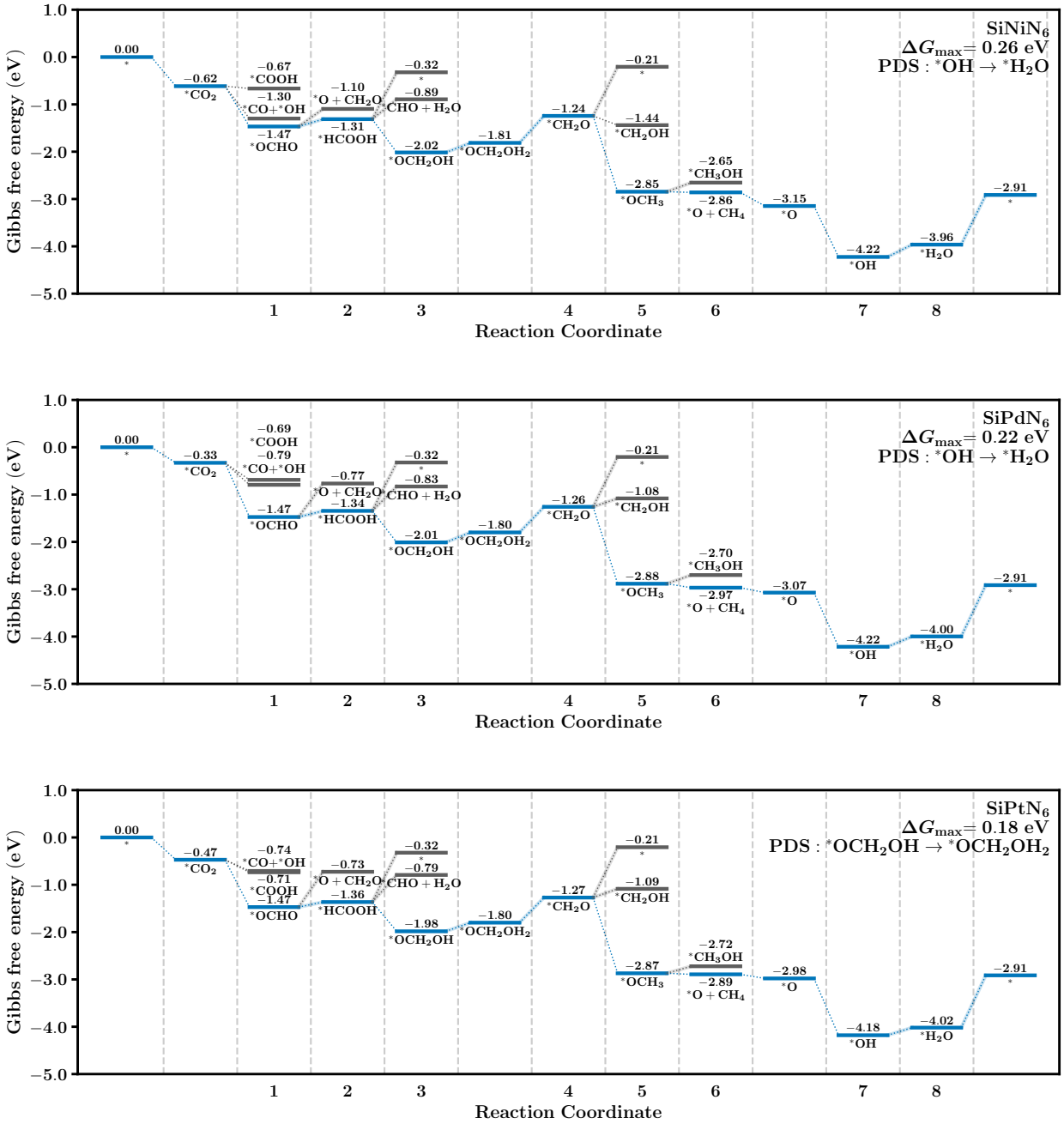

Figure S18: The relative Gibbs free energies in CO_2RR on $SiNiN_6$, $SiPdN_6$ and $SiPtN_6$ DACs under the applied potential of 0 V.

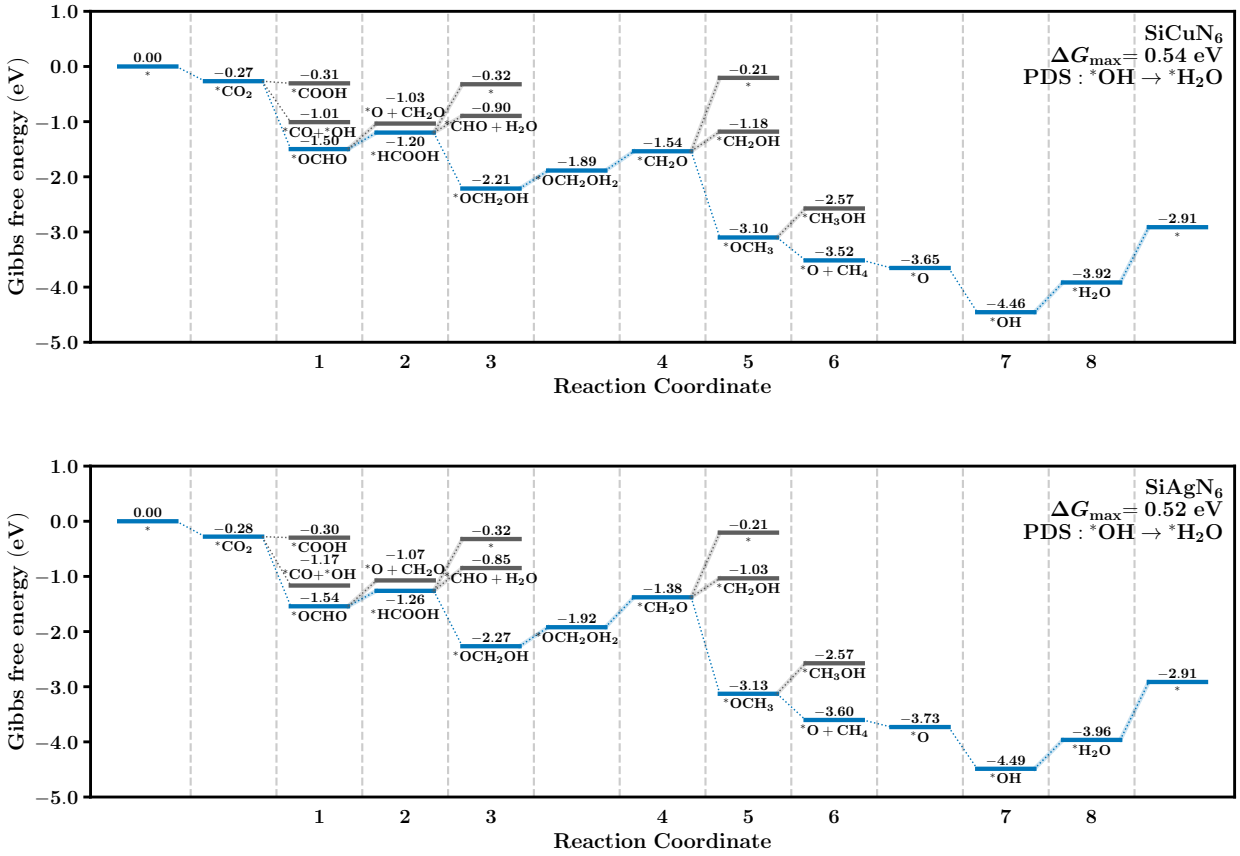

Figure S19: The relative Gibbs free energies in CO_2RR on $SiCuN_6$ and $SiAgN_6$ DACs under the applied potential of 0 V.

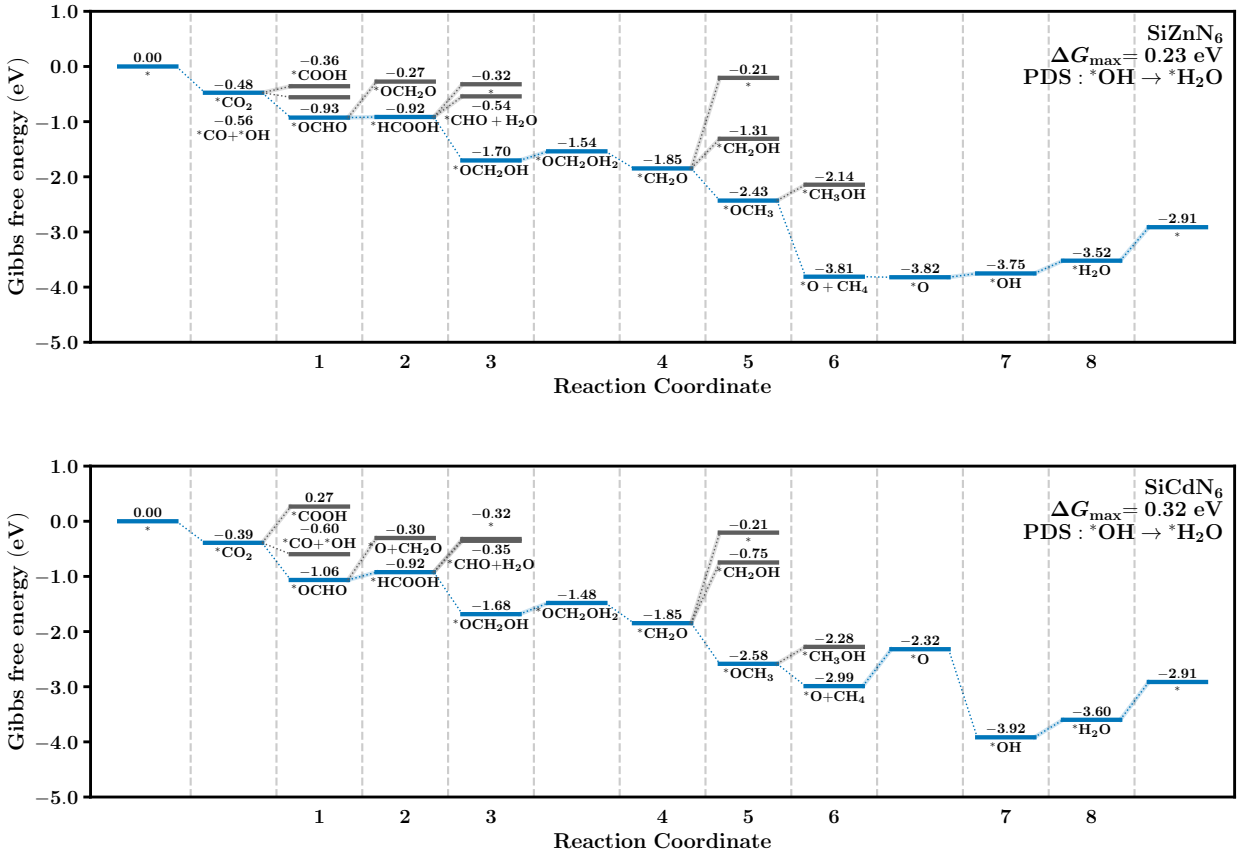

Figure S20: The relative Gibbs free energies in CO_2RR on $SiZnN_6$ and $SiCdN_6$ DACs under the applied potential of 0 V.

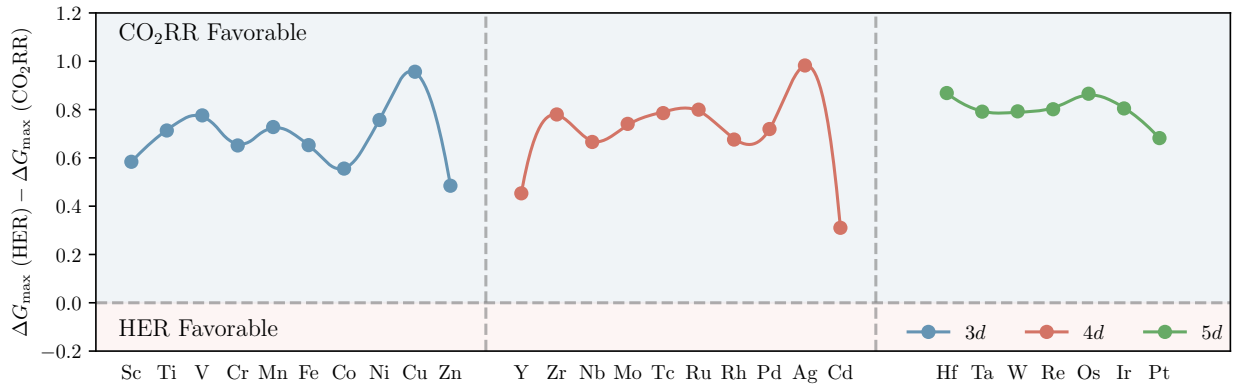

Figure S21: The ΔG_{max} difference between HER with CO_2RR on different $SiTMN_6$ DACs.

Table S4: Computed dissolution potential (U_{diss}) for selected structures.

Structure	U_{diss}
$SiCoN_6$	0.09
$SiRhN_6$	0.75
$SiIrN_6$	1.08
$SiCoN_5B$	0.43
$SiCoN_5C$	0.35
$SiRhN_5B$	1.14
$SiRhN_5C$	1.05
$SiIrN_5B$	1.34
$SiIrN_5C$	1.30

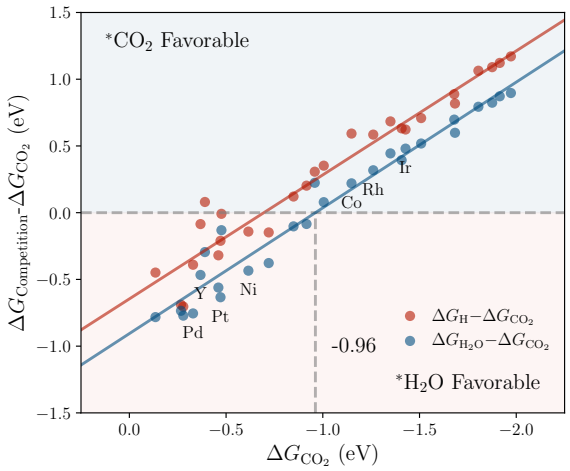

Figure S22: Comparison of the adsorption free energies of *H, *H_2O, and *CO_2.

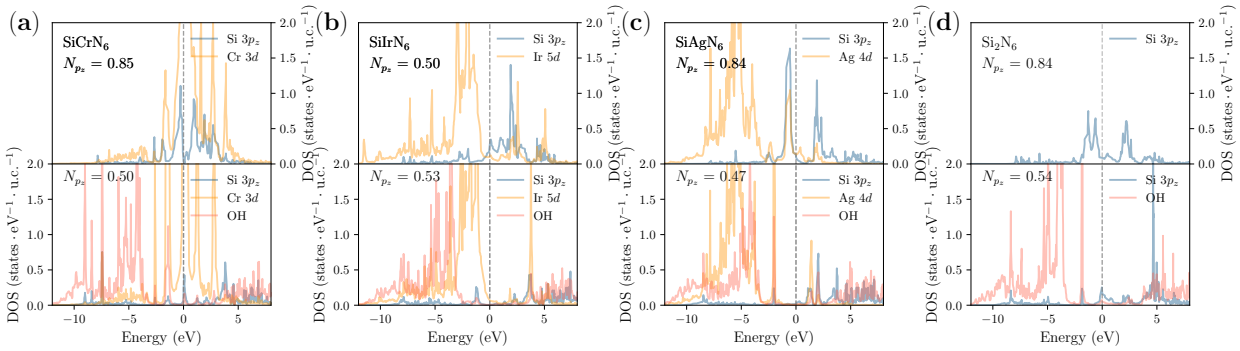

Figure S23: Before and after adsorption, the PDOS for the TM d, Si $3p_z$ and CO_2 orbitals of $SiCrN_6$, $SiIrN_6$, $SiAgN_6$ and Si_2N_6 DACs, respectively. The top image represents the state before adsorption, while the bottom image shows the state after adsorption.

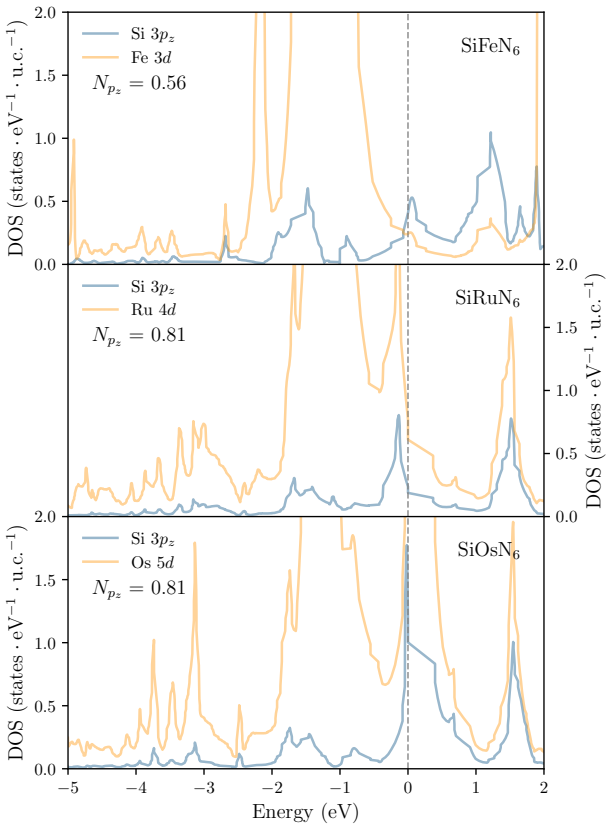

Figure S24: The PDOS for the TM d and Si $3p_z$ on SiFeN$_6$, SiRuN$_6$ and SiOsN$_6$ DACs, respectively.

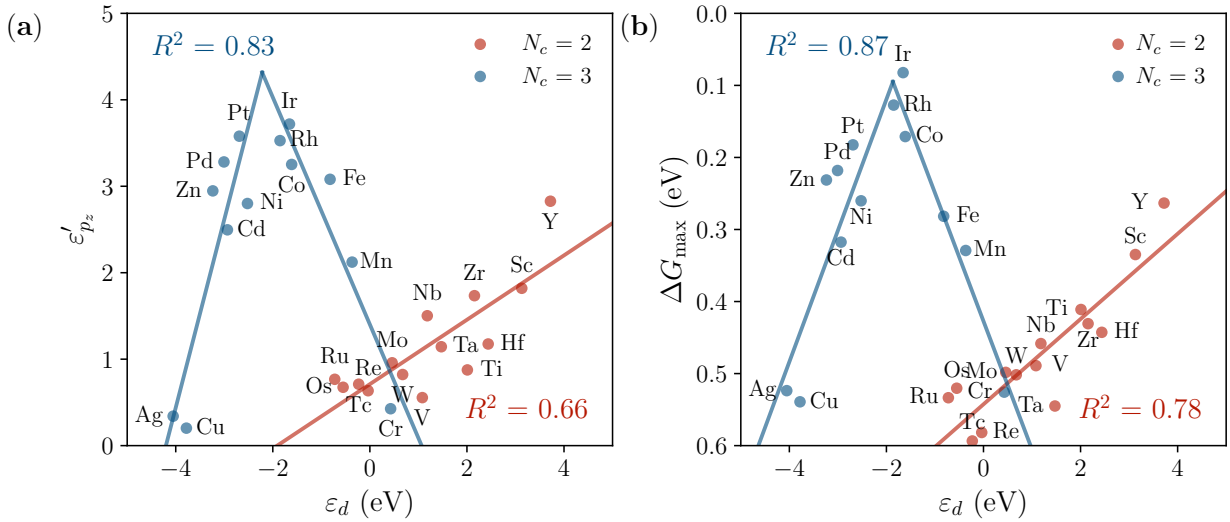

Figure S25: (a) Correlations between ε_d and ε'_{p_z} for different SiTMN$_6$ DACs; (b) Correlations between ε_d and $\Delta G_{\max}$ for different SiTMN$_6$ DACs.

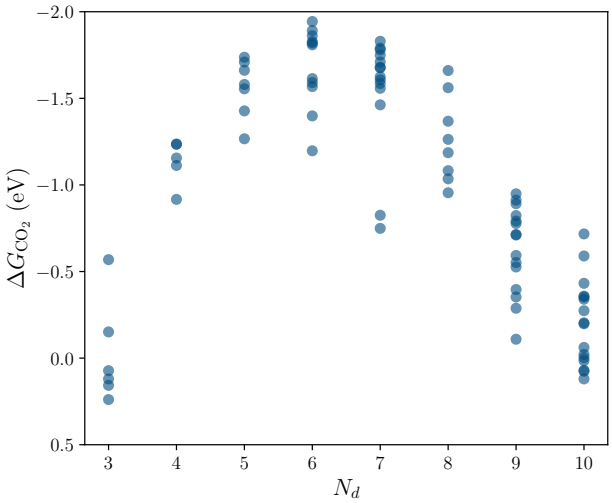

Figure S26: Distribution of ΔG_{*CO_2} in the training data set as functions of feature of N_d.

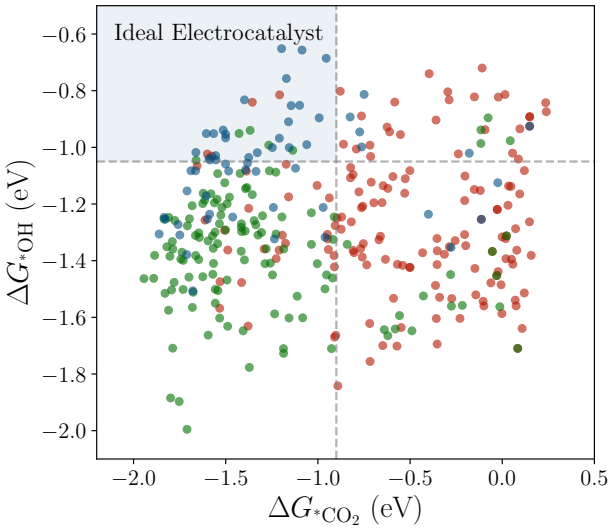

Figure S27: ΔG_{*CO_2} and ΔG_{*OH} for different $SiTMN_5An$.

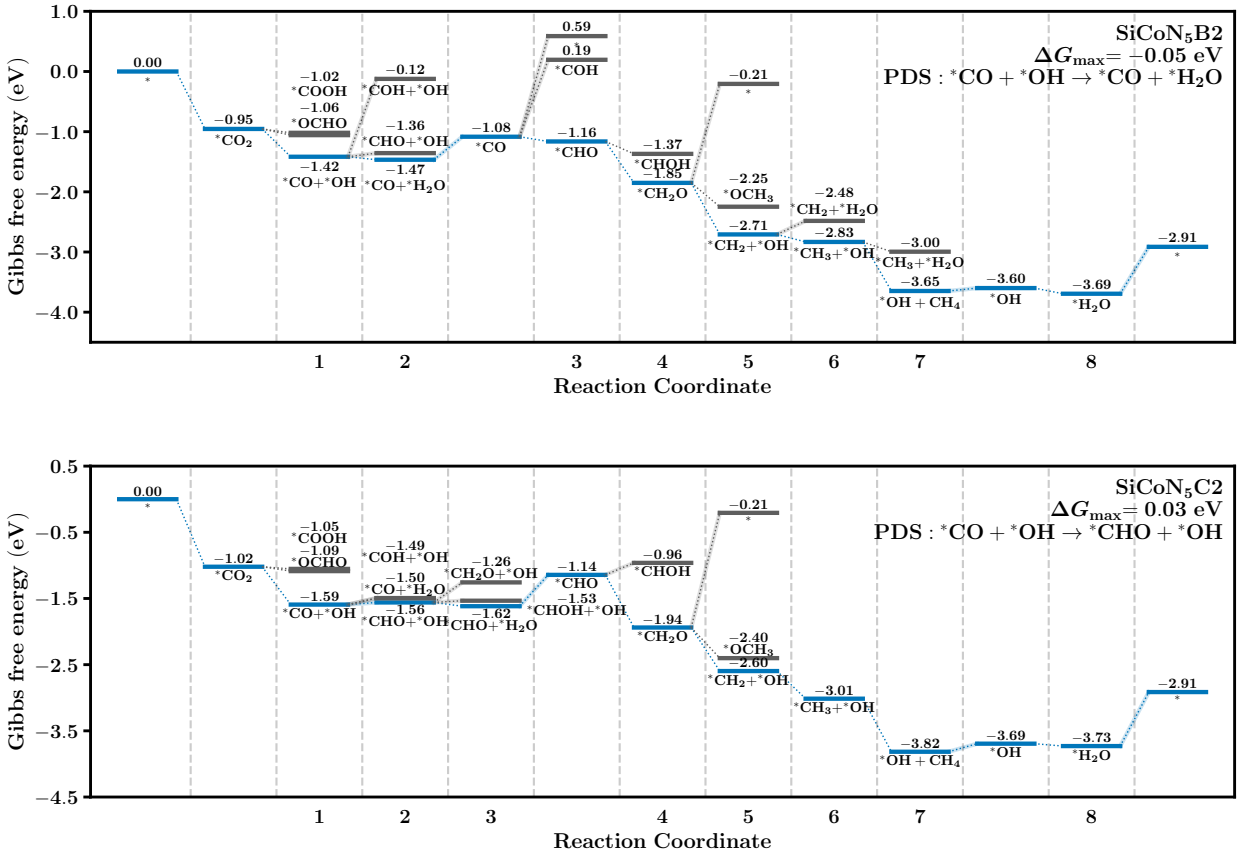

Figure S28: The relative Gibbs free energies in CO_2RR on $SiCoN_5$B2 and $SiCoN_5$C2 DACs under the applied potential of 0 V.

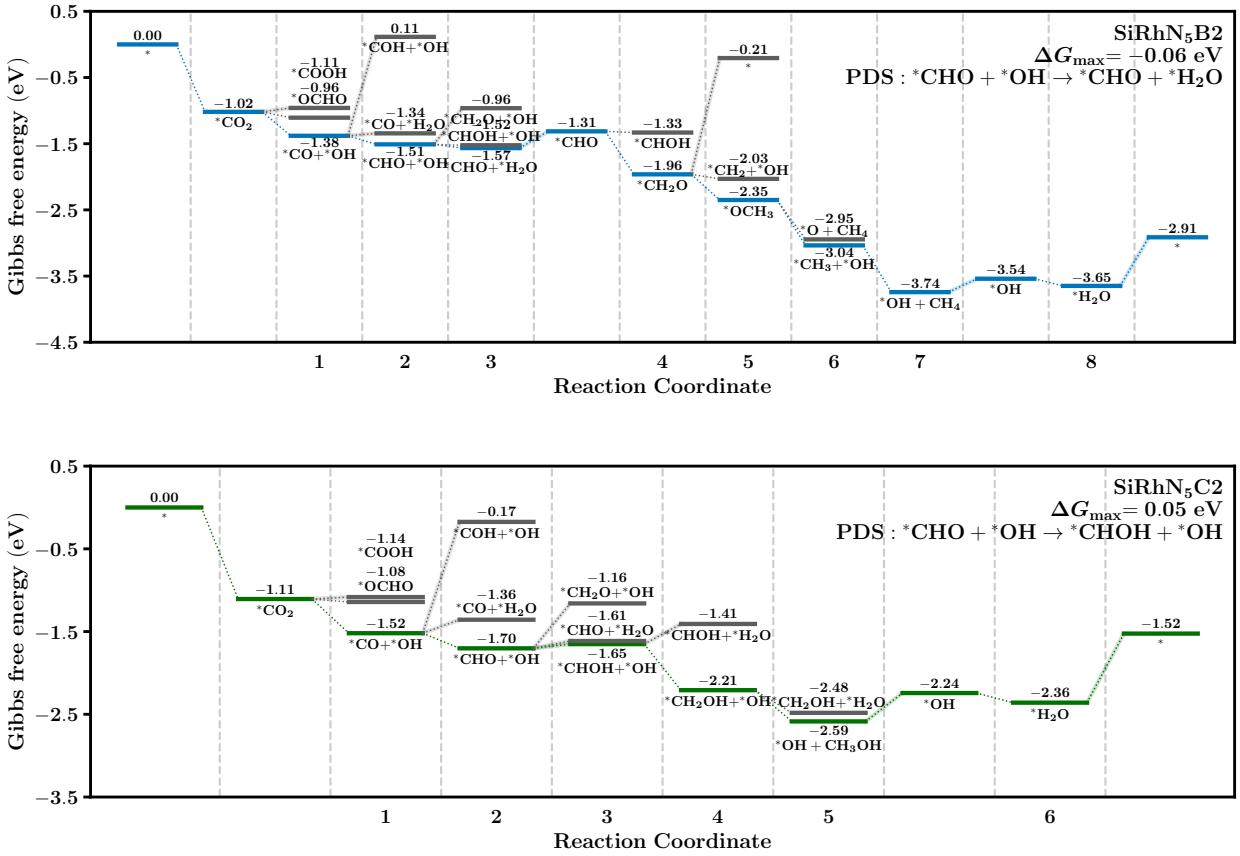

Figure S29: The relative Gibbs free energies in CO_2RR on $SiRhN_5$B2 and $SiRhN_5$C2 DACs under the applied potential of 0 V.

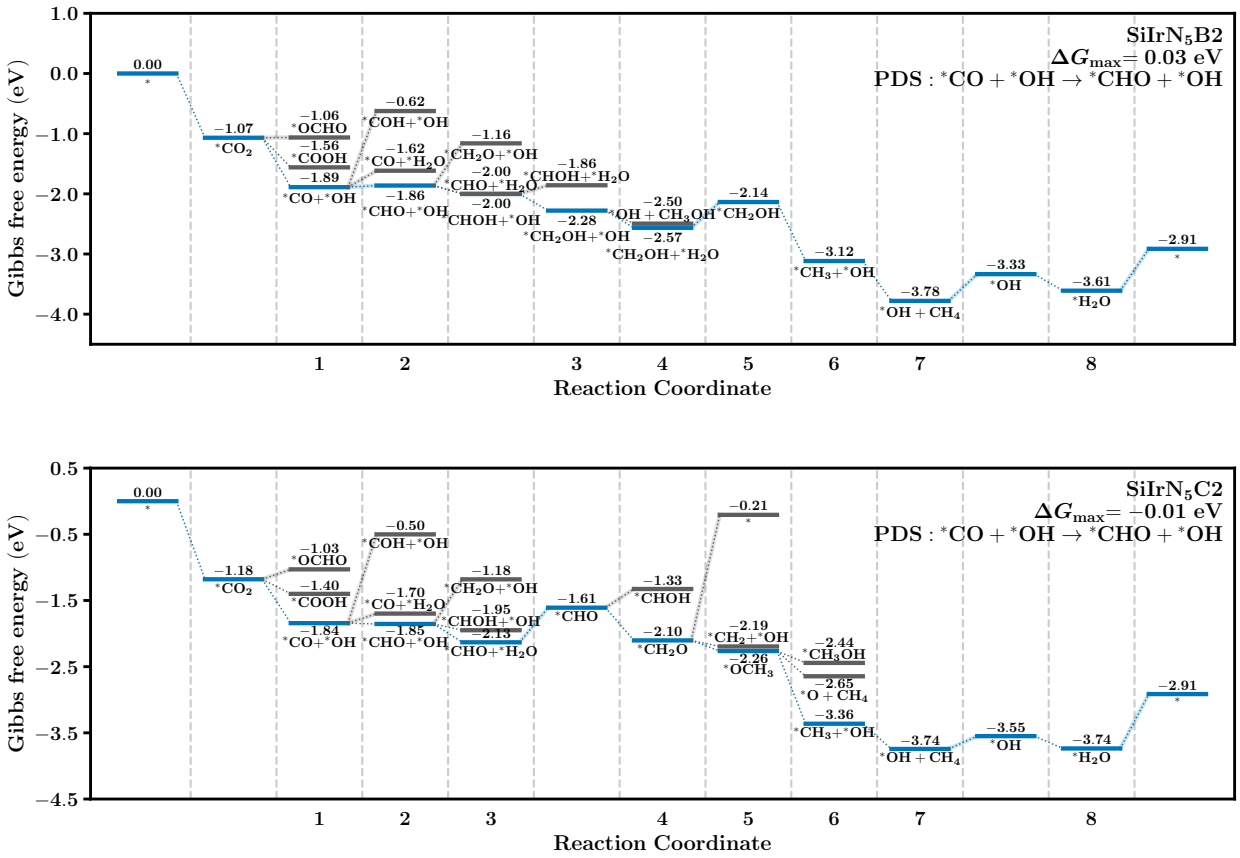

Figure S30: The relative Gibbs free energies in CO_2RR on SiIrN$_5$B2 and SiIrN$_5$C2 DACs under the applied potential of 0 V.

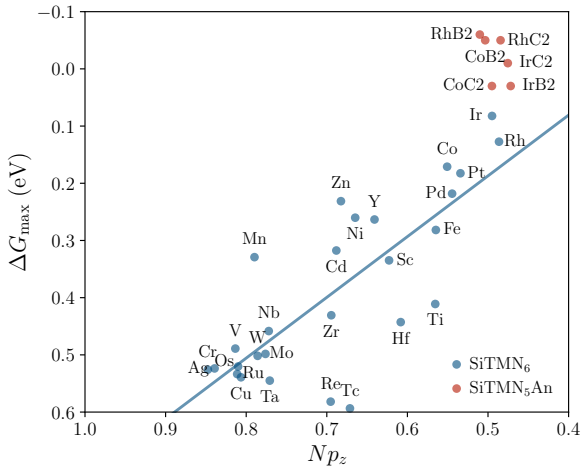

Figure S31: Correlations between N_{p_z} and $\Delta G_{\max}$ of SiTMN$_6$ DACs and screened SiTMN$_5$An DACs.

Table S5: Screening Results of Dopant Candiates Identified after Considering the ΔG_{*CO_2} and ΔG_{*OH}.

Number	**Structure**	$\boldsymbol{\Delta G_{OH \rightarrow H_2O}}$	**Number**	**Structure**	$\boldsymbol{\Delta G_{OH \rightarrow H_2O}}$
01	IrB2	-0.27	18	VB2	0.11
02	IrC2	-0.19	19	TcB2	0.11
03	RhC2	-0.12	20	MnC2	0.12
04	RhB2	-0.11	21	RuB2	0.13
05	CoB2	-0.09	22	CrB2	0.13
06	CoC2	-0.04	23	NbB2	0.16
07	IrC4	0.00	24	MoC2	0.17
08	IrC1	0.01	25	WC2	0.21
09	MoC1	0.02	26	IrB1	0.21
10	RhC6	0.03	27	RuC2	0.22
11	IrC6	0.06	28	VC2	0.22
12	MoB2	0.06	29	CrC2	0.24
13	MnB2	0.07	30	MoC3	0.26
14	ReB2	0.07	31	CoC5	0.26
15	OsB2	0.08	32	IrC5	0.26
16	RhC1	0.09	33	CrC1	0.32
17	WB2	0.10			

References

(1) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.