Supplementary Information

Efficient Exploration of Transition-Metal Decorated MXene for Carbon Monoxide Sensing Using Integrated Active Learning and Density Functional Theory

Kajjana Boonpalit,[‡,a] Jiramet Kinchagawat,[‡,a] Chanatkran Prommin,[a] Sarana Nutanong,*,[a] and Supawadee Namuangruk*,[b]

[a]*School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong 21210, Thailand*

[b]*National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency (NSTDA), Klong Luang, Pathum Thani 12120, Thailand*

[‡]*These authors contributed equally to this work.*

*E-mail: snutanon@vistec.ac.th (SN), supawadee@nanotec.or.th (SN)

Table S1 CO adsorption on pristine MXene

Surface	E_{ads} (eV)	Surface	E_{ads} (eV)	Surface	E_{ads} (eV)
Ti_2CO_2	-0.14	Zr_2CO_2	-0.13	Hf_2CO_2	-0.14
$Ti_3C_2O_2$	-0.14	$Zr_3C_2O_2$	-0.15	$Hf_3C_2O_2$	-0.15
$Ti_4C_3O_2$	-0.15	$Zr_4C_3O_2$	-0.15	$Hf_4C_3O_2$	-0.15
Ti_2CS_2	-0.13	Zr_2CS_2	-0.13	Hf_2CS_2	-0.13
$Ti_3C_2S_2$	-0.14	$Zr_3C_2S_2$	-0.13	$Hf_3C_2S_2$	-0.13
$Ti_4C_3S_2$	-0.14	$Zr_4C_3S_2$	-0.13	$Hf_4C_3S_2$	-0.16

Table S2 Pre-computed initial dataset for initializing the active learning screening

Initial Dataset					
Surface	Method	E_{ads} (eV)	Surface	Method	E_{ads} (eV)
Fe@Zr_2CO_2	Random	-1.32	Cu@$Ti_3C_2O_2$	Random	-1.75
Fe@Hf_3C2O_2	Random	-1.83	Os@$Ti_3C_2O_2$	Random	-2.12
Co@$Hf_4C_3S_2$	Random	-1.45	Nb@$Ti_3C_2S_2$	Random	-1.65
Ta@Ti_3C2O_2	Random	-1.55	Ta@$Ti_3C_2S_2$	Random	-1.97
Fe@Hf_2CO_2	Random	-1.73	Ru@$Ti_4C_3O_2$	Random	-1.55
Au@Hf_2CO_2	Random	-1.80	Nb@$Ti_4C_3O_2$	Random	-1.54
Ta@Hf_2CS_2	Random	-1.89	Ta@$Ti_4C_3S_2$	Random	-1.98
Y@Hf_2CS_2	Random	-0.67	Ir@$Ti_4C_3S_2$	Random	-1.46
W@$Hf_3C_2O_2$	Random	-1.97	Ag@Zr_2CO_2	Random	-1.12
Rh@$Hf_3C_2O_2$	Random	-1.96	Os@Zr_2CO_2	Random	-2.53
Re@$Hf_3C_2S_2$	Random	-2.12	Nb@Zr_2CS_2	Random	-1.54
Mo@$Hf_3C_2S_2$	Random	-1.81	Au@Zr_2CS_2	Random	-1.63
Pd@Hf_4C3O_2	Random	-1.76	Hf@Zr_3C2O_2	Random	-1.17
Co@Hf_4C3O_2	Random	-2.33	Ni@Zr_3C2O_2	Random	-1.88
V@Hf_4C3S_2	Random	-1.54	Pt@Zr_3C2S_2	Random	-1.24
Fe@Hf_4C3S_2	Random	-1.19	Ti@Zr_3C2S_2	Random	-1.21
Ta@Ti_2CO_2	Random	-1.51	Sc@Zr_4C3O_2	Random	-0.80
Ti@Ti_2CO_2	Random	-1.08	Zr@Zr_4C3O_2	Random	-1.08
Mo@Ti_2CS_2	Random	-1.86	Ni@Zr_4C3S_2	Random	-1.23
Mn@Ti_2CS_2	Random	-1.09	Cr@Zr_4C3S_2	Random	-1.17

Table S3 Predicted and actual adsorption energy with their value difference of query datapoints at each iteration.

Iteration 1					Iteration 2				
Surface	**Method**	**$E_{ads,DFT}$ (eV)**	**$E_{ads,ML}$ (eV)**	**Error (eV)**	**Surface**	**Method**	**$E_{ads,DFT}$ (eV)**	**$E_{ads,ML}$ (eV)**	**Error (eV)**
Ag@Hf$_2$CO$_2$	Explore	-1.07	-1.66	0.59	Ag@Ti$_3$C$_2$S$_2$	Explore	-1.07	-1.63	0.56
Au@Ti$_2$CO$_2$	Explore	-2.80	-1.32	-1.48	Co@Ti$_2$CO$_2$	Explore	-1.68	-1.36	-0.32
Cr@Ti$_2$CO$_2$	Explore	-1.52	-1.36	-0.16	Cu@Zr$_2$CS$_2$	Explore	-1.06	-1.49	0.43
Ir@Zr$_2$CO$_2$	Explore	-2.56	-1.58	-0.98	Ir@Hf$_2$CO$_2$	Explore	-2.76	-1.66	-1.10
Mn@Ti$_2$CO$_2$	Explore	-1.19	-1.39	0.20	Ni@Hf$_2$CS$_2$	Explore	-1.35	-1.52	0.17
Pt@Hf$_2$CO$_2$	Explore	-3.45	-1.65	-1.80	Pd@Hf$_2$CS$_2$	Explore	-1.20	-1.50	0.30
Rh@Zr$_4$C$_3$O$_2$	Explore	-1.96	-1.38	-0.58	Pt@Ti$_2$CS$_2$	Explore	-2.01	-1.65	-0.36
Ta@Hf$_3$C$_2$O$_2$	Explore	-1.74	-1.66	-0.08	Ta@Hf$_4$C$_3$O$_2$	Explore	-1.77	-1.64	-0.13
V@Hf$_2$CO$_2$	Explore	-1.39	-1.66	0.27	Ti@Hf$_2$CO$_2$	Explore	-1.28	-1.69	0.41
Zr@Zr$_2$CO$_2$	Explore	-1.17	-1.56	0.39	W@Zr$_2$CS$_2$	Explore	-2.13	-1.51	-0.62
Explore RMSE = 0.8586 eV			Exploit RMSE = N/A		Explore RMSE = 0.5127 eV			Exploit RMSE = N/A	
Iteration 3					**Iteration 4**				
Surface	**Method**	**$E_{ads,DFT}$ (eV)**	**$E_{ads,ML}$ (eV)**	**Error (eV)**	**Surface**	**Method**	**$E_{ads,DFT}$ (eV)**	**$E_{ads,ML}$ (eV)**	**Error (eV)**
Au@Hf$_2$CS$_2$	Explore	-1.84	-1.31	-0.53	Pt@Hf$_3$C$_2$O$_2$	Explore	-3.37	-2.80	-0.57
Au@Ti$_2$CS$_2$	Explore	-1.82	-1.62	-0.20	Sc@Hf$_2$CO$_2$	Explore	-1.06	-0.45	-0.61
Ir@Ti$_2$CO$_2$	Explore	-2.81	-1.61	-1.20	Sc@Hf$_3$C$_2$O$_2$	Explore	-0.93	-1.06	0.13
Pd@Hf$_2$CO$_2$	Explore	-1.76	-1.78	0.02	Sc@Ti$_2$CO$_2$	Explore	-0.75	-0.32	-0.43
Pd@Zr$_2$CS$_2$	Explore	-1.13	-1.41	0.28	Sc@Zr$_2$CO$_2$	Explore	-0.93	-0.29	-0.64
Pt@Hf$_2$CS$_2$	Explore	-1.49	-1.59	0.10	Sc@Ti$_3$C$_2$O$_2$	Exploit	-0.71	-0.78	0.07
Pt@Hf$_3$C$_2$S$_2$	Explore	-1.34	-1.60	0.26	Sc@Zr$_3$C$_2$S$_2$	Exploit	-0.75	-0.78	0.03
Pt@Ti$_3$C$_2$S$_2$	Explore	-1.97	-2.07	0.10	Y@Zr$_2$CS$_2$	Exploit	-0.65	-0.73	0.08
Pt@Zr$_2$CS$_2$	Explore	-1.38	-1.79	0.41	Y@Zr$_3$C$_2$O$_2$	Exploit	-0.72	-0.67	-0.05
Sc@Ti$_2$CS$_2$	Explore	-0.77	-1.57	0.80	Y@Zr$_4$C$_3$O$_2$	Exploit	-0.70	-0.76	0.06
Explore RMSE = 0.5230 eV			Exploit RMSE = N/A		Explore RMSE = 0.5115 eV			Exploit RMSE = 0.0605 eV	
Iteration 5					**Iteration 6**				
Surface	**Method**	**$E_{ads,DFT}$ (eV)**	**$E_{ads,ML}$ (eV)**	**Error (eV)**	**Surface**	**Method**	**$E_{ads,DFT}$ (eV)**	**$E_{ads,ML}$ (eV)**	**Error (eV)**
Fe@Zr$_4$C$_3$S$_2$	Exploit	-1.15	-0.78	-0.37	Sc@Hf$_4$C$_3$S$_2$	Exploit	-0.76	-0.89	0.13
Sc@Hf$_2$CS$_2$	Exploit	-0.78	-0.78	0.00	Sc@Ti$_4$C$_3$O$_2$	Exploit	-0.71	-0.78	0.07
Sc@Zr$_2$CS$_2$	Exploit	-0.76	-0.67	-0.09	Ti@Zr$_2$CO$_2$	Exploit	-1.22	-0.93	-0.29
Sc@Zr$_3$C$_2$O$_2$	Exploit	-0.80	-0.67	-0.13	Y@Hf$_3$C$_2$O$_2$	Exploit	-0.85	-0.89	0.04
Sc@Zr$_4$C$_3$S$_2$	Exploit	-0.74	-0.67	-0.07	Y@Hf$_3$C$_2$S$_2$	Exploit	-0.65	-0.76	0.11
Y@Hf$_4$C$_3$S$_2$	Exploit	-0.64	-0.80	0.16	Y@Hf$_4$C$_3$O$_2$	Exploit	-0.80	-0.93	0.13
Y@Ti$_2$CO$_2$	Exploit	-0.67	-0.62	-0.05	Y@Ti$_2$CS$_2$	Exploit	-0.67	-0.64	-0.03
Y@Ti$_3$C$_2$O$_2$	Exploit	-0.61	-0.64	0.03	Y@Ti$_3$C$_2$S$_2$	Exploit	-0.66	-0.82	0.16
Y@Ti$_4$C$_3$O$_2$	Exploit	-0.60	-0.80	0.20	Y@Zr$_3$C$_2$S$_2$	Exploit	-0.63	-0.58	-0.05
Y@Zr$_2$CO$_2$	Exploit	-0.86	-0.70	-0.16	Y@Zr$_4$C$_3$S$_2$	Exploit	-0.63	-0.63	0.00
Explore RMSE = N/A			Exploit RMSE = 0.1617 eV		Explore RMSE = N/A			Exploit RMSE = 0.1286 eV	

Iteration 7				
Surface	**Method**	**$E_{ads,DFT}$ (eV)**	**$E_{ads,ML}$ (eV)**	**Error (eV)**
Ag@Ti$_2$CS$_2$	Exploit	-1.06	-1.04	-0.02
Ag@Zr$_3$C$_2$S$_2$	Exploit	-1.04	-1.10	0.06
Ag@Zr$_4$C$_3$S$_2$	Exploit	-1.05	-1.05	0.00
Fe@Ti$_3$C$_2$S$_2$	Exploit	-1.43	-1.10	-0.33
Sc@Hf$_3$C$_2$S$_2$	Exploit	-0.76	-0.84	0.08
Sc@Hf$_4$C$_3$O$_2$	Exploit	-0.89	-0.97	0.08
Sc@Ti$_3$C$_2$S$_2$	Exploit	-0.77	-0.80	0.03
Sc@Ti$_4$C$_3$S$_2$	Exploit	-0.78	-1.00	0.22
Y@Hf$_2$CO$_2$	Exploit	-1.04	-1.03	-0.01
Y@Ti$_4$C$_3$S$_2$	Exploit	-0.67	-0.81	0.14
Explore RMSE = N/A		Exploit RMSE = 0.1395 eV		

Table S4 Summary of candidates having adsorption energy between -0.6 and -0.8 eV. The distance between the anchored-TM atom and the C atom of CO (d(TM-C)), CO adsorption energy (E_{ads}), recovery time (τ) calculated at 300 K, formation energy (E_f), and charge transfer (Δq) are listed.

No	Surface	Iteration	Method	E_{ads} (eV)	τ (s)	E_f (eV)	Δq (e-)
1	Y@Hf$_4$C$_3$O$_2$	6	Exploit	-0.80	27.24	-1.89	-0.29
2	Y@Zr$_3$C$_2$O$_2$	4	Exploit	-0.72	1.24	-1.73	-0.25
3	Y@Ti$_2$CO$_2$	5	Exploit	-0.67	0.18	-2.53	-0.22
4	Y@Zr$_4$C$_3$O$_2$	4	Exploit	-0.70	0.57	-2.82	-0.22
5	Sc@Zr$_3$C$_2$O$_2$	5	Exploit	-0.80	27.24	-1.49	-0.18
6	Sc@Zr$_4$C$_3$O$_2$	0	Random	-0.80	27.24	-2.55	-0.18
7	Y@Ti$_3$C$_2$O$_2$	5	Exploit	-0.61	0.02	-4.63	-0.15
8	Sc@Ti$_2$CO$_2$	4	Explore	-0.75	3.94	-2.29	-0.15
9	Y@Ti$_4$C$_3$O$_2$	5	Exploit	-0.60	0.01	-3.40	-0.14
10	Y@Ti$_3$C$_2$S$_2$	6	Exploit	-0.66	0.12	-3.02	-0.12
11	Sc@Ti$_3$C$_2$O$_2$	4	Exploit	-0.71	0.84	-4.38	-0.12
12	Y@Ti$_2$CS$_2$	6	Exploit	-0.67	0.18	-2.83	-0.12
13	Sc@Ti$_2$CS$_2$	3	Explore	-0.77	8.54	-2.61	-0.10
14	Sc@Ti$_3$C$_2$S$_2$	7	Exploit	-0.77	8.54	-2.81	-0.09
15	Y@Ti$_4$C$_3$S$_2$	7	Exploit	-0.67	0.18	-2.98	-0.09
16	Y@Hf$_2$CS$_2$	0	Random	-0.67	0.18	-3.12	-0.09
17	Sc@Ti$_4$C$_3$O$_2$	6	Exploit	-0.71	0.84	-3.11	-0.07
18	Sc@Ti$_4$C$_3$S$_2$	7	Exploit	-0.78	12.57	-2.75	-0.07
19	Y@Hf$_3$C$_2$S$_2$	6	Exploit	-0.65	0.08	-3.26	-0.07
20	Y@Hf$_4$C$_3$S$_2$	5	Exploit	-0.64	0.06	-3.61	-0.07
21	Sc@Hf$_2$CS$_2$	5	Exploit	-0.78	12.57	-2.83	-0.07
22	Sc@Hf$_3$C$_2$S$_2$	7	Exploit	-0.76	5.80	-2.97	-0.05
23	Y@Zr$_4$C$_3$S$_2$	6	Exploit	-0.63	0.04	-3.78	-0.05
24	Y@Zr$_2$CS$_2$	4	Exploit	-0.65	0.08	-3.75	-0.05
25	Sc@Hf$_4$C$_3$S$_2$	6	Exploit	-0.76	5.80	-3.33	-0.05
26	Sc@Zr$_2$CS$_2$	5	Exploit	-0.76	5.80	-3.43	-0.04
27	Y@Zr$_3$C$_2$S$_2$	6	Exploit	-0.63	0.04	-3.73	-0.03
28	Sc@Zr$_4$C$_3$S$_2$	5	Exploit	-0.74	2.68	-3.48	-0.03
29	Sc@Zr$_3$C$_2$S$_2$	4	Exploit	-0.75	3.94	-3.41	-0.02

Table S5. The example of the CO adsorption energy at different kinetic energy cutoff. The structure is sampled from different type of decorated transition metal and MXene.

Surface	CO adsorption energy (eV)			
	400 eV	450 eV	500 eV	550 eV
Sc@Zr$_3$C$_2$O$_2$	-0.84	-0.80	-0.80	-0.81
Y@Zr$_3$C$_2$O$_2$	-0.74	-0.72	-0.73	-0.72
Ag@Ti$_2$CS$_2$	-1.08	-1.06	-1.05	-1.06
Pd@Hf$_4$C$_3$O$_2$	-1.79	-1.76	-1.75	-1.76

Table S6. The hyperparameter, total number of fitting parameter, and the average RMSE of test set from five different seeds. The hyperparameter optimization is done on initial dataset. The training set, validation set, and test set ratio are 0.8:0.1:0.1 in hyperparameter optimization.

Batch Size	Learning Rate	# Convolution Layer	Number of fitting parameter	Test set RMSE (eV)
4	0.00003	2	58433	0.3668
4	0.00003	4	102465	0.3575
4	0.00003	6	146497	0.3862
4	0.0003	2	58433	0.3294
4	0.0003	4	102465	0.3225
4	0.0003	6	146497	0.3273
4	0.003	2	58433	0.3377
4	0.003	4	102465	0.3529
4	0.003	6	146497	0.3852
8	0.00003	2	58433	0.3627
8	0.00003	4	102465	0.3696
8	0.00003	6	146497	0.3556
8	0.0003	2	58433	0.3617
8	0.0003	4	102465	0.3212
8	0.0003	6	146497	0.3637
8	0.003	2	58433	0.3832
8	0.003	4	102465	0.3329
8	0.003	6	146497	0.3494
16	0.00003	2	58433	0.3433
16	0.00003	4	102465	0.3744
16	0.00003	6	146497	0.3073
16	0.0003	2	58433	0.3641
16	0.0003	4	102465	0.3634
16	**0.0003**	**6**	**146497**	**0.3062**
16	0.003	2	58433	0.3655
16	0.003	4	102465	0.3567
16	0.003	6	146497	0.4034

Fig. S1 Actual adsorption energy of labeled datapoints and predicted adsorption energy of unlabeled datapoints at iteration 1. The yellow shaded area represents the target adsorption energy range of -0.6 to -0.8 eV.

Fig. S2 mMDS of embedding after training of surrogate model at iteration 1. Circles denote unlabeled data, squares denote labeled data, and crosses denote queried data by K-Center
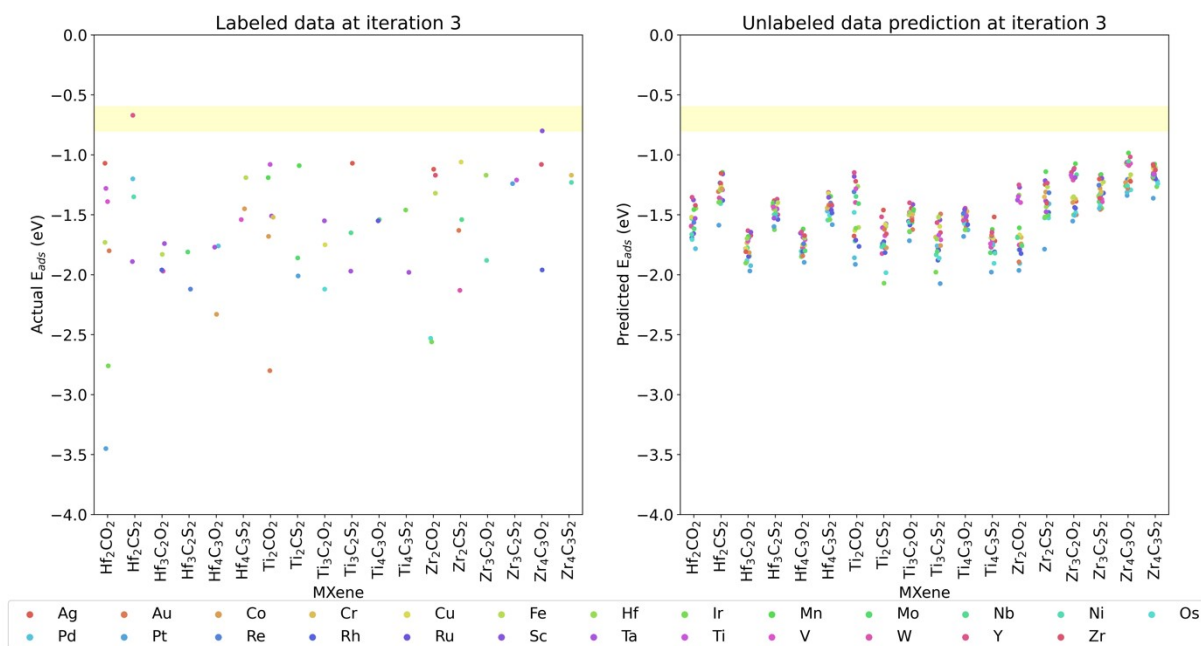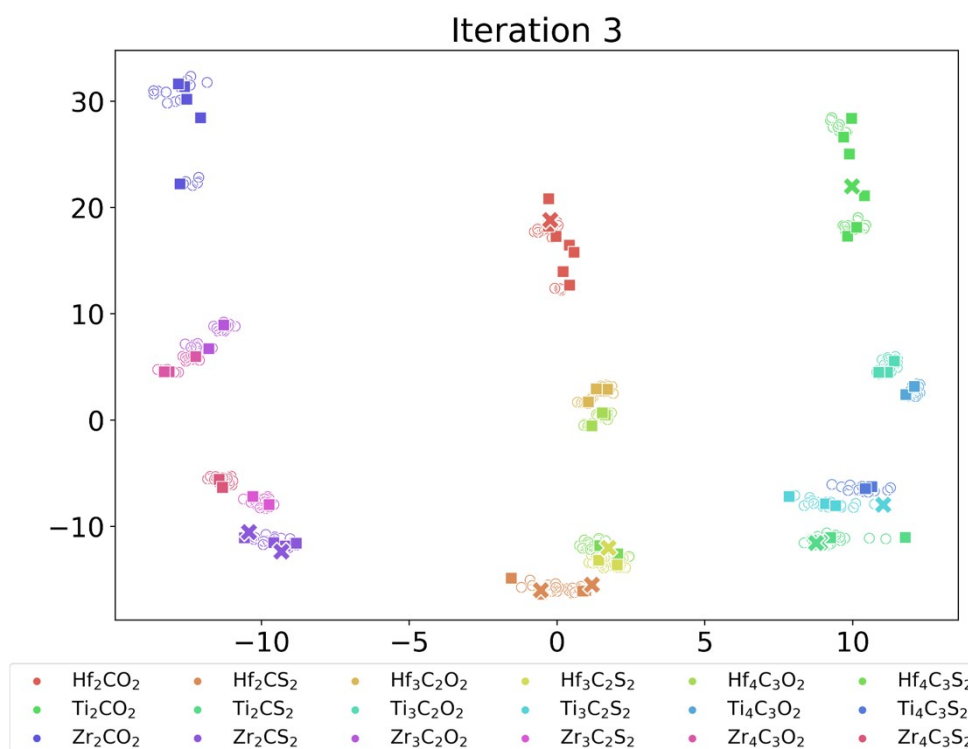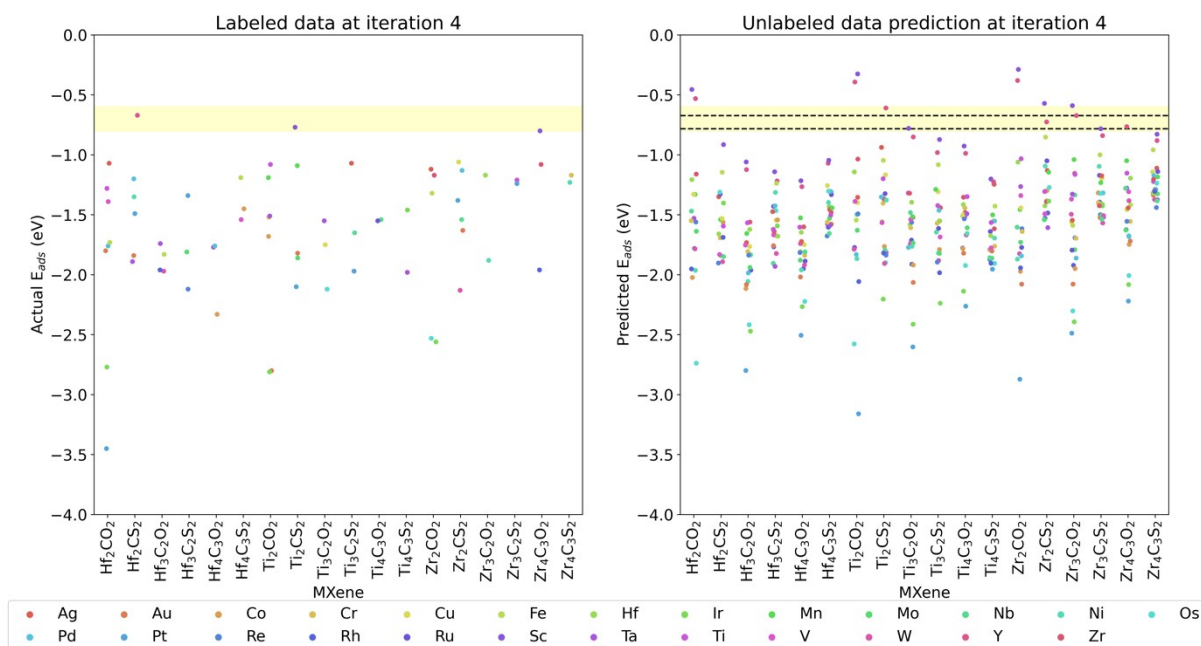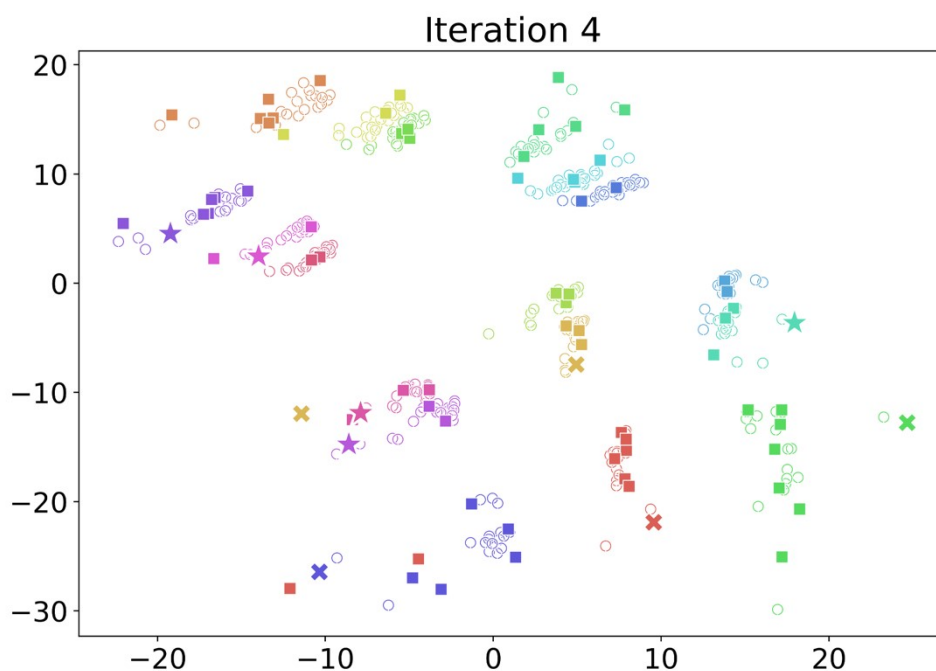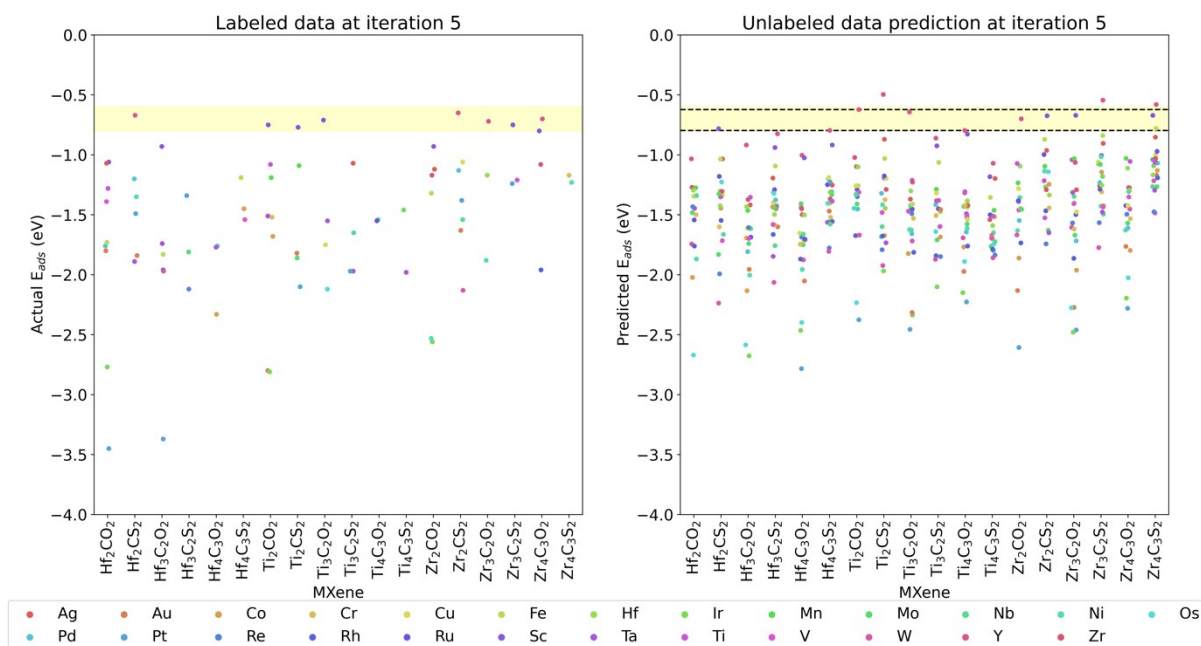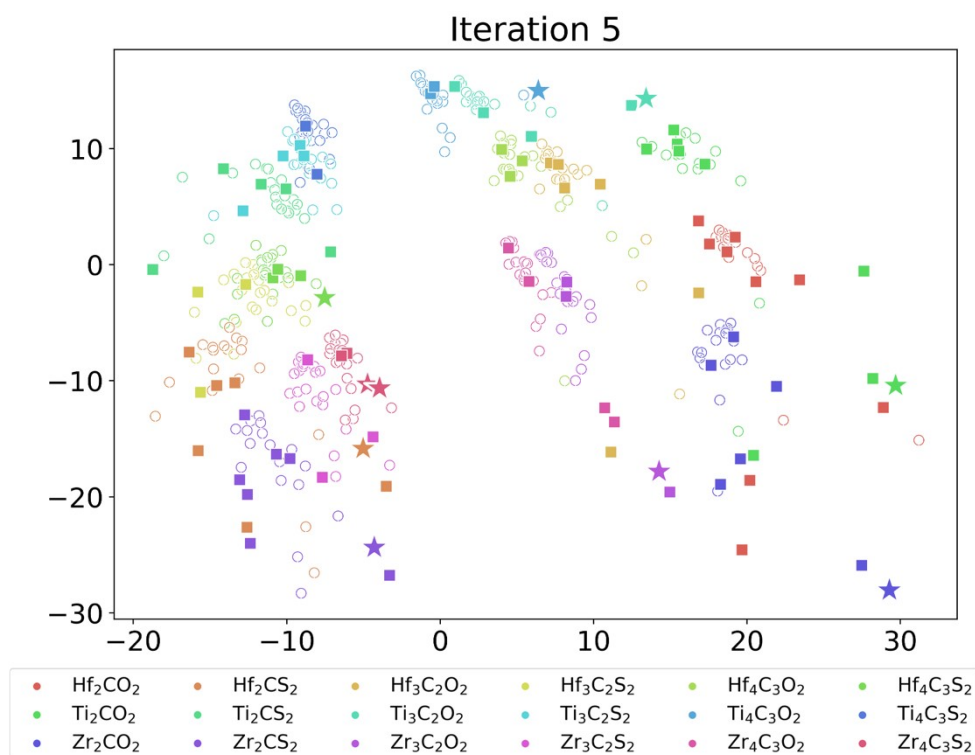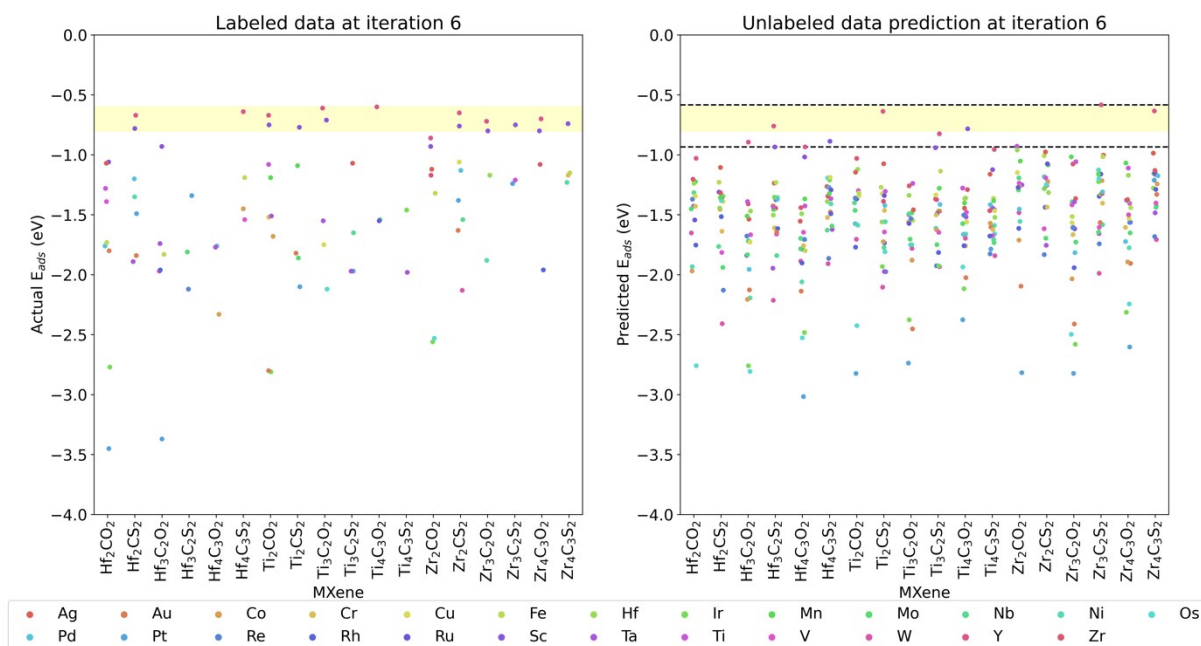
Fig. S3 Actual adsorption energy of labeled datapoints and predicted adsorption energy of unlabeled datapoints at iteration 2. The yellow shaded area represents the target adsorption energy range of -0.6 to -0.8 eV.

Fig. S4 mMDS of embedding after training of surrogate model at iteration 2. Circles denote unlabeled data, squares denote labeled data, and crosses denote queried data by K-Center

Fig. S5 Actual adsorption energy of labeled datapoints and predicted adsorption energy of unlabeled datapoints at iteration 3. The yellow shaded area represents the target adsorption energy range of -0.6 to -0.8 eV.
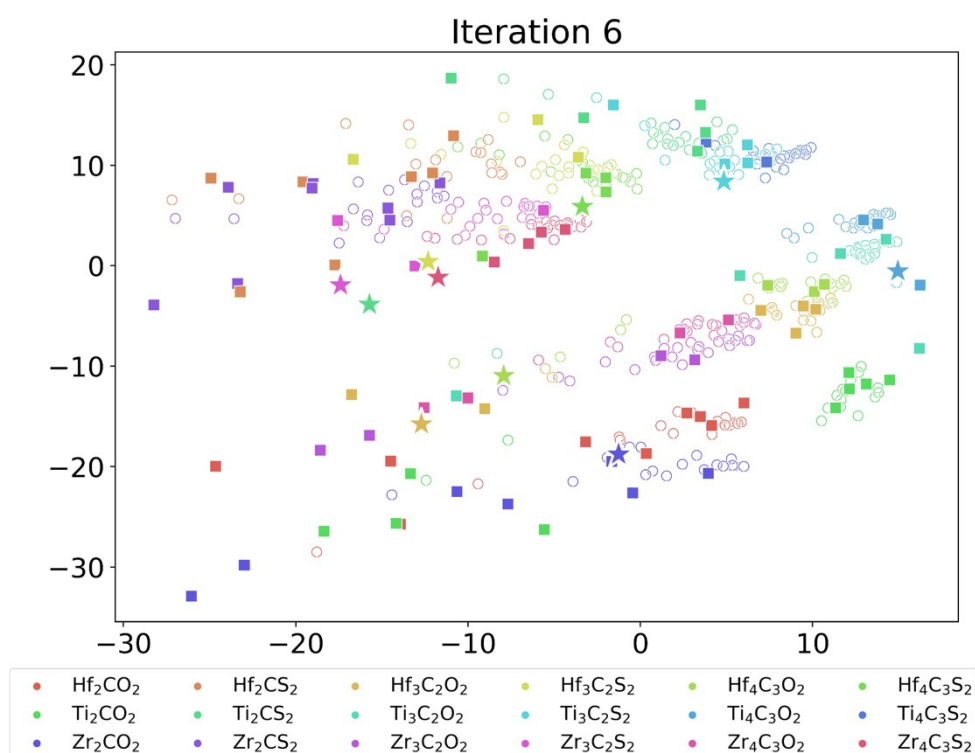

Fig. S6 mMDS of embedding after training of surrogate model at iteration 3. Circles denote unlabeled data, squares denote labeled data, and crosses denote queried data by K-Center
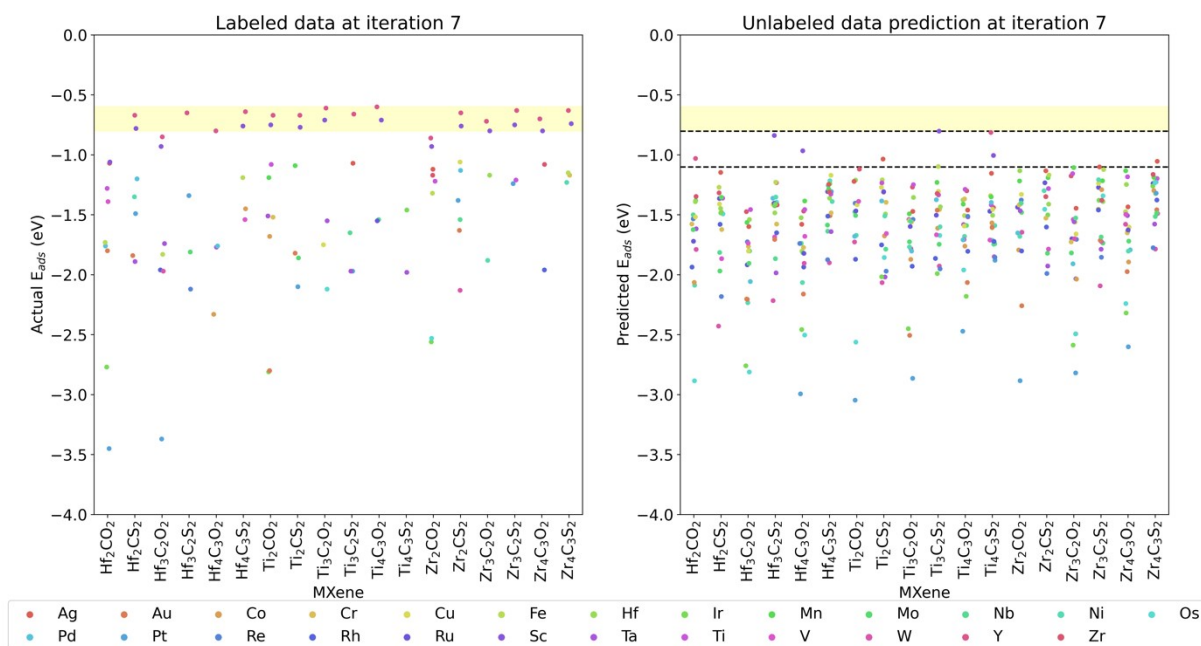
Fig. S7 Actual adsorption energy of labeled datapoints and predicted adsorption energy of unlabeled datapoints at iteration 4. The yellow shaded area represents the target adsorption energy range of -0.6 to -0.8 eV. The dash lines denote the range of queried data by Greedy.

Fig. S8 mMDS of embedding after training of surrogate model at iteration 4. Circles denote unlabeled data, squares denote labeled data, crosses denote queried data by K-Center, and stars denote queried data by Greedy.

Fig. S7 Actual adsorption energy of labeled datapoints and predicted adsorption energy of unlabeled datapoints at iteration 5. The yellow shaded area represents the target adsorption energy range of -0.6 to -0.8 eV. The dash lines denote the range of queried data by Greedy.
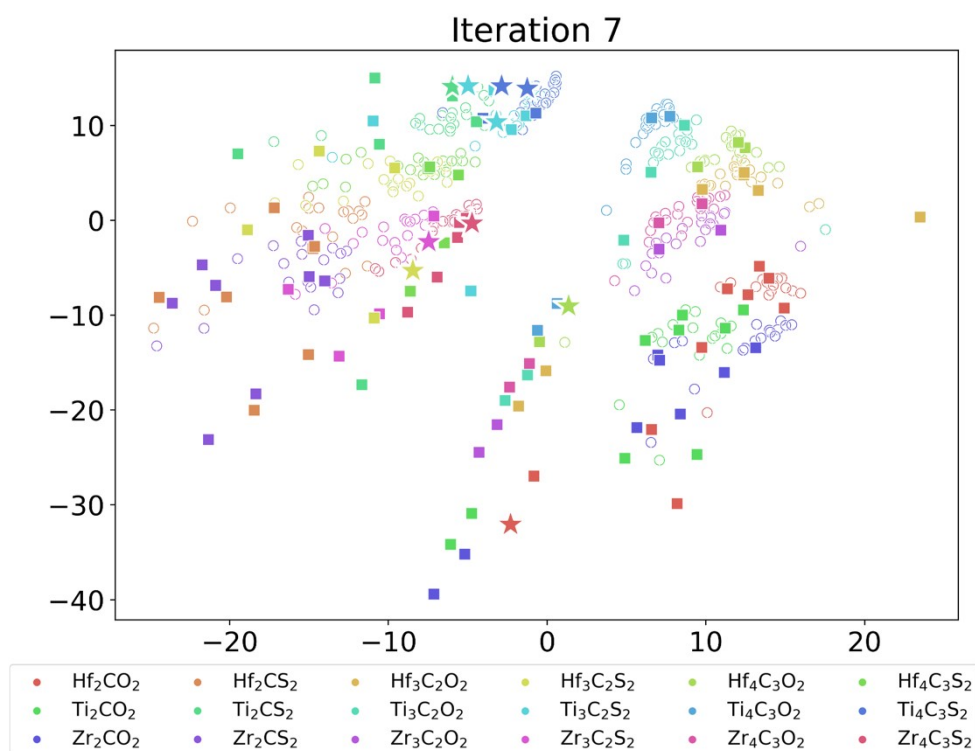

Fig. S8 mMDS of embedding after training of surrogate model at iteration 5. Circles denote unlabeled data, squares denote labeled data, and stars denote queried data by Greedy.
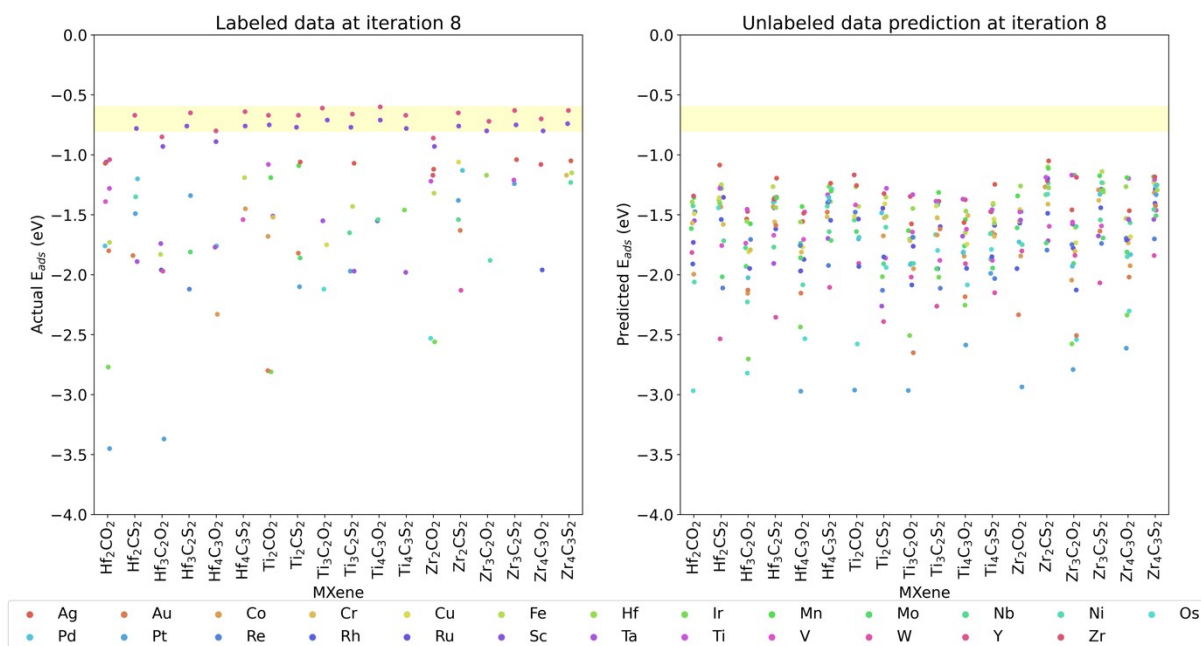
Fig. S9 Actual adsorption energy of labeled datapoints and predicted adsorption energy of unlabeled datapoints at iteration 6. The yellow shaded area represents the target adsorption energy range of -0.6 to -0.8 eV. The dash lines denote the range of queried data by Greedy.
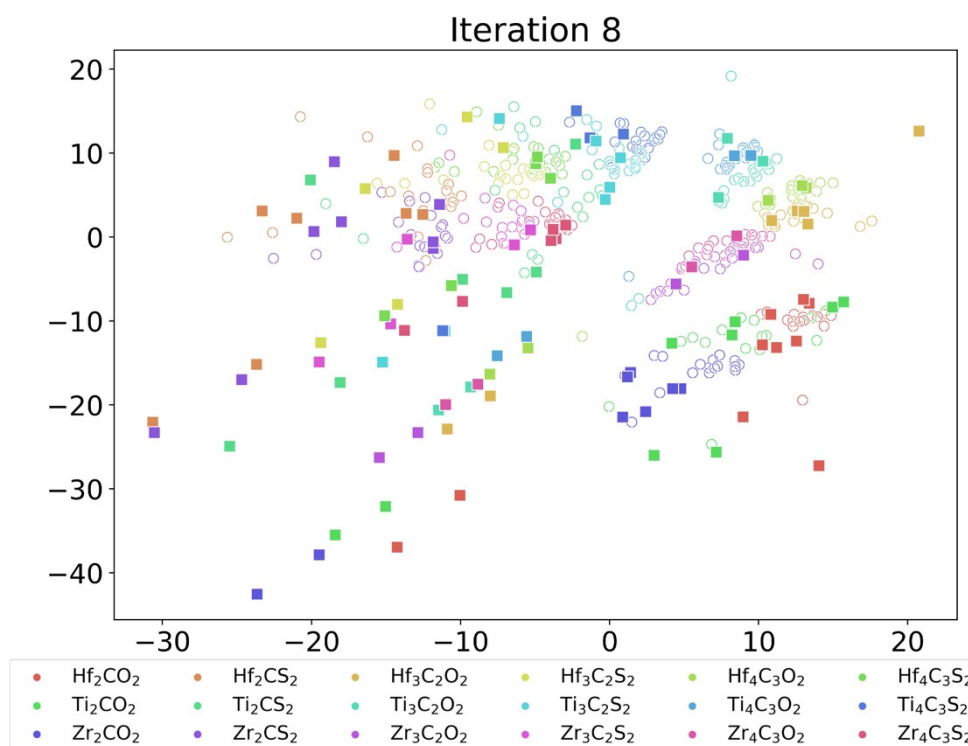

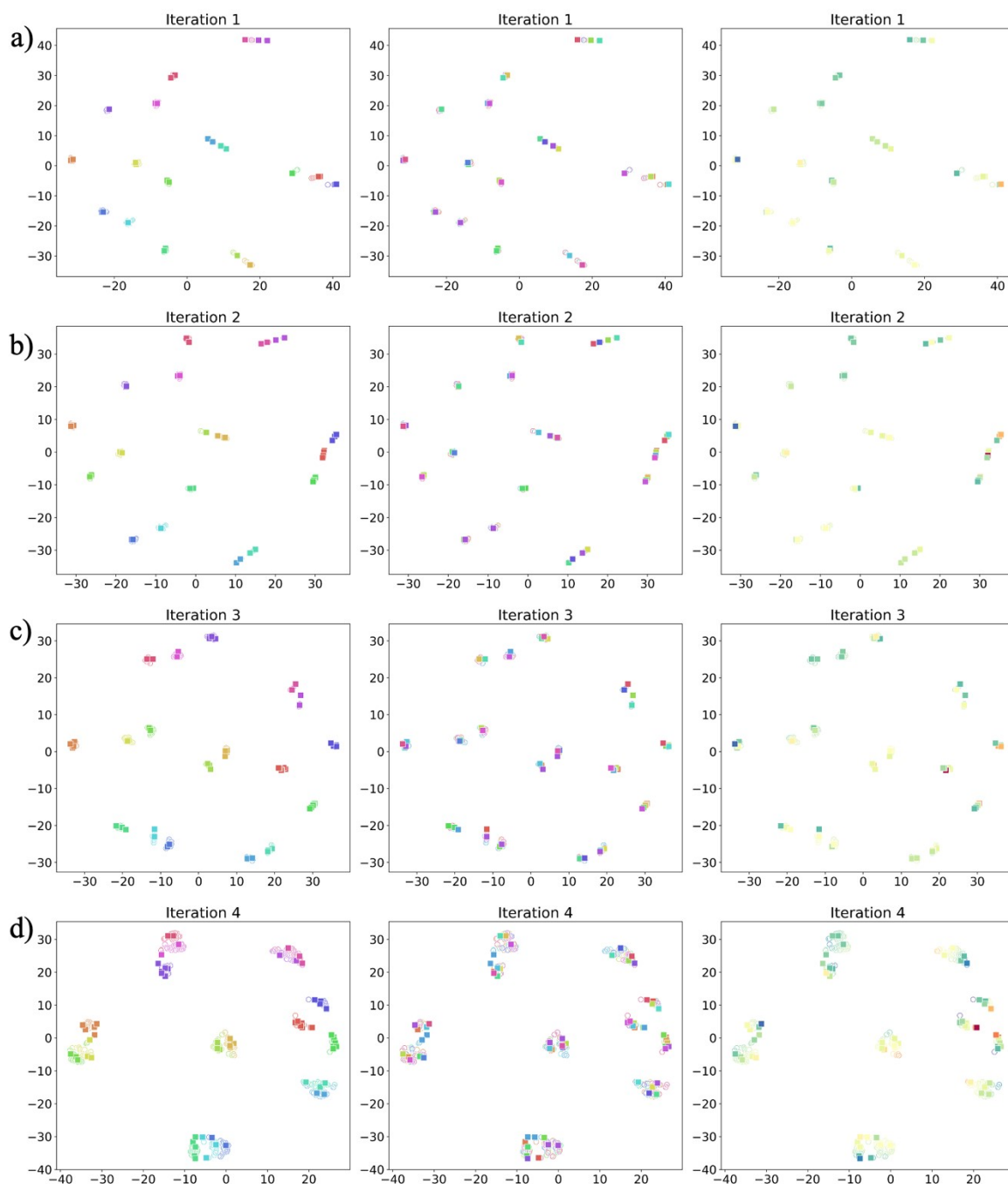**Fig. S10** mMDS of embedding after training of surrogate model at iteration 6. Circles denote unlabeled data, squares denote labeled data, and stars denote queried data by Greedy.

Fig. S11 Actual adsorption energy of labeled datapoints and predicted adsorption energy of unlabeled datapoints at iteration 7. The yellow shaded area represents the target adsorption energy range of -0.6 to -0.8 eV. The dash lines denote the range of queried data by Greedy.

Fig. S12 mMDS of embedding after training of surrogate model at iteration 7. Circles denote unlabeled data, squares denote labeled data, and stars denote queried data by Greedy.

Fig. S13 Actual adsorption energy of labeled datapoints and predicted adsorption energy of unlabeled datapoints at iteration 8. The yellow shaded area represents the target adsorption energy range of -0.6 to -0.8 eV.

Fig. S14 mMDS of embedding after training of surrogate model at iteration 8. Circles denote unlabeled data, and squares denote labeled data.

Fig. S15 T TSNE of the embedding at iteration 1-4, **a-d** respectively. Colors are based on (Left) different surfaces. (Middle) different anchored metal. These two columns use the same color as the legends of **Fig. S1** and **S2** respectively. (Right) Colors represent the actual E_{ads} for labeled data and the predicted E_{ads} for unlabeled data, with lower E_{ads} color being red and higher E_{ads} being blue. Circles denote unlabeled data, and squares denote labeled data.
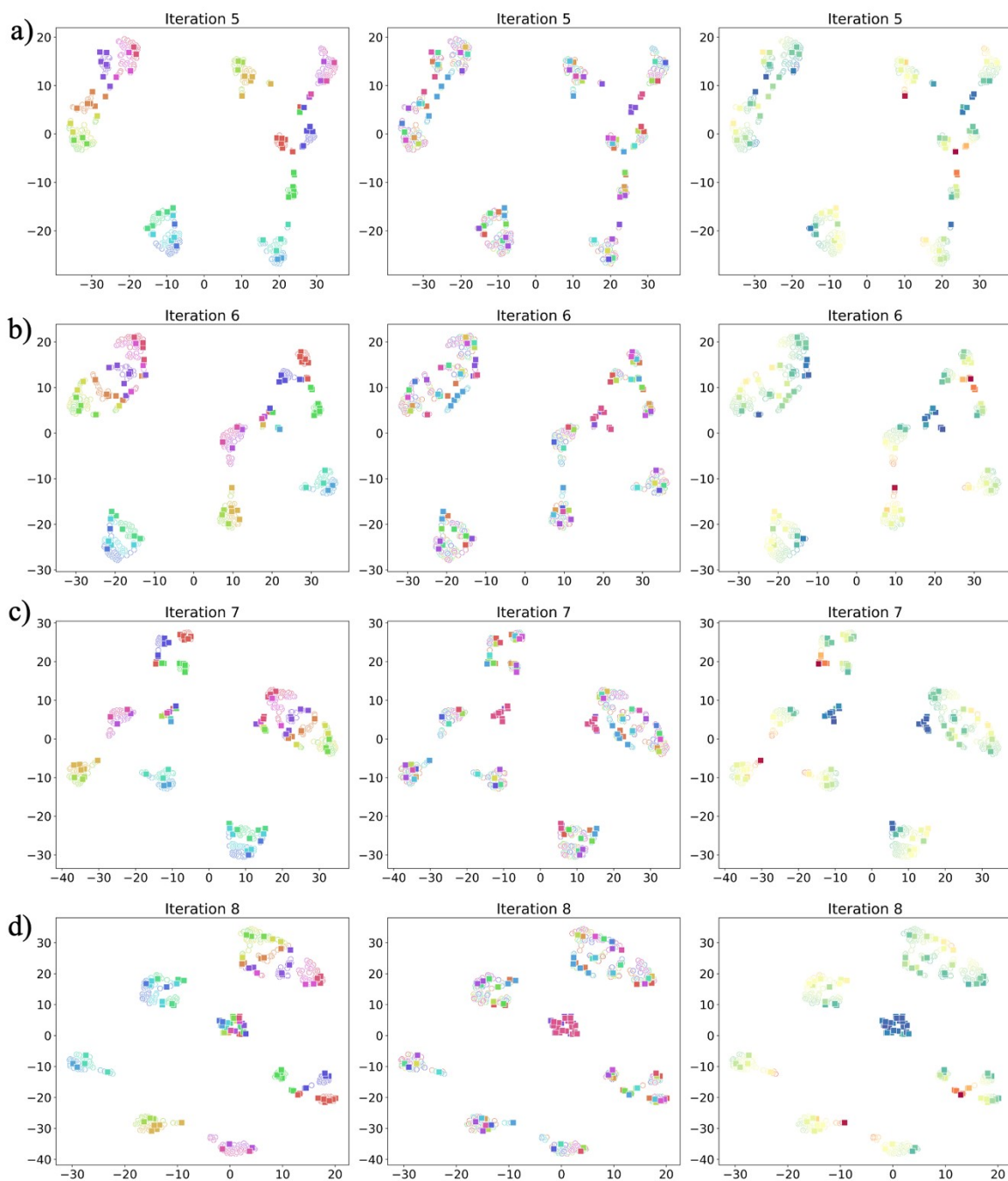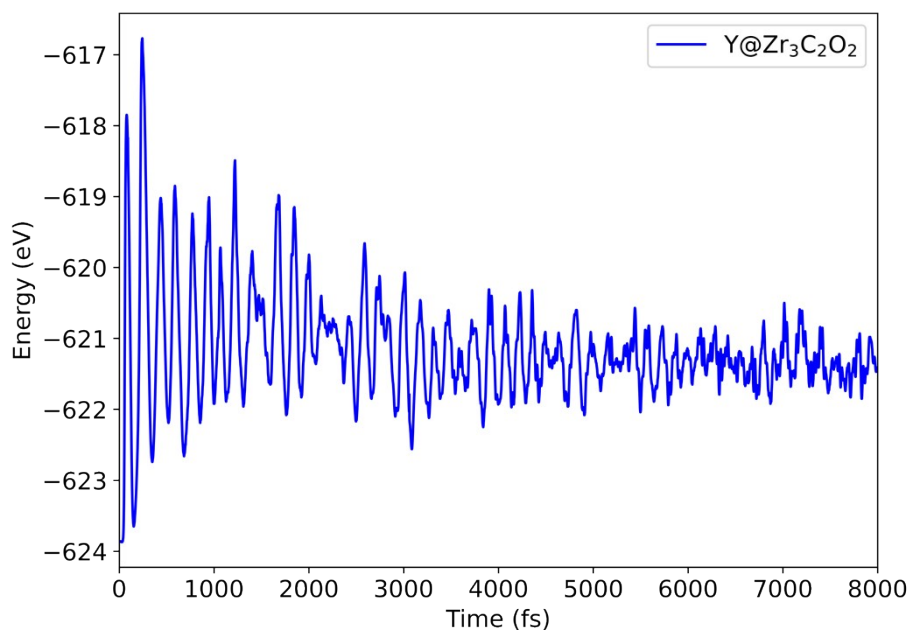
Fig. S16 TSNE of the embedding at iteration 5-8, **a-d** respectively. Colors are based on (Left) different surfaces. (Middle) different anchored metal. These two columns use the same color as the legends of **Fig. S1** and **S2** respectively. (Right) Colors represent the actual E_{ads} for labeled data and the predicted E_{ads} for unlabeled data, with lower E_{ads} color being red and higher E_{ads} being blue. Circles denote unlabeled data, and squares denote labeled data.

Fig. S17 Energy fluctuation of CO adsorbed on $Y@Zr_3C_2O_2$ from AIMD simulation at the temperature of 300 K for 8 ps with a time step of 2 fs. The total simulation consists of 2 ps for the equilibration period and 6 ps for the production period.
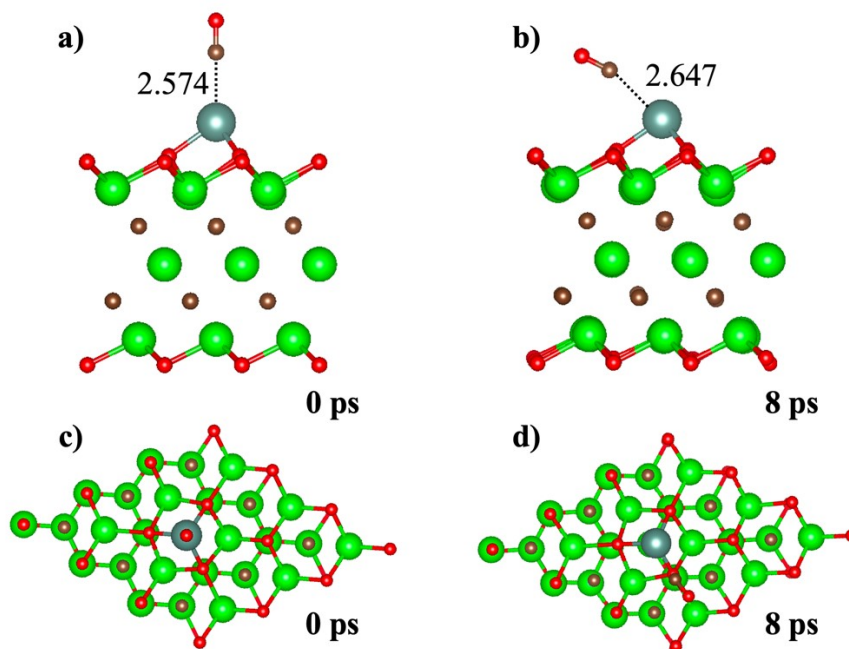

Fig. S18 Snapshot of initail and final structure from AIMD simulation of CO adsorbed on $Y@Zr_3C_2O_2$. (a-b) side views at 0 and 8 ps (c-d) top views at 0 and 8 ps. The distance between Y and C atoms are represented in angstrom.

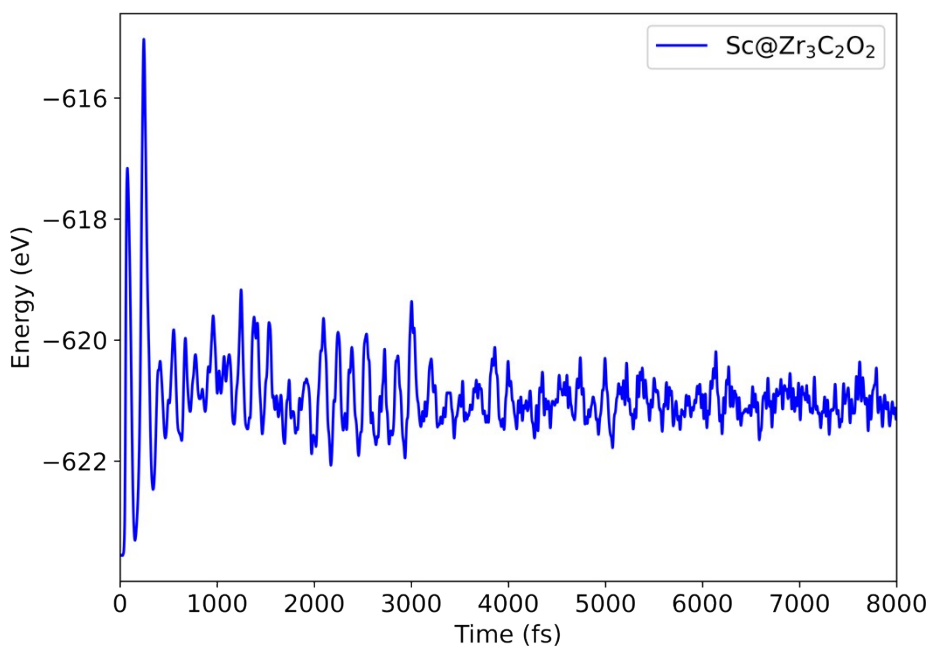**Fig. S19** Energy fluctuation of CO adsorbed on Sc@Zr$_3$C$_2$O$_2$ from AIMD simulation at the temperature of 300 K for 8 ps with a time step of 2 fs. The total simulation consists of 2 ps for the equilibration period and 6 ps for the production period.
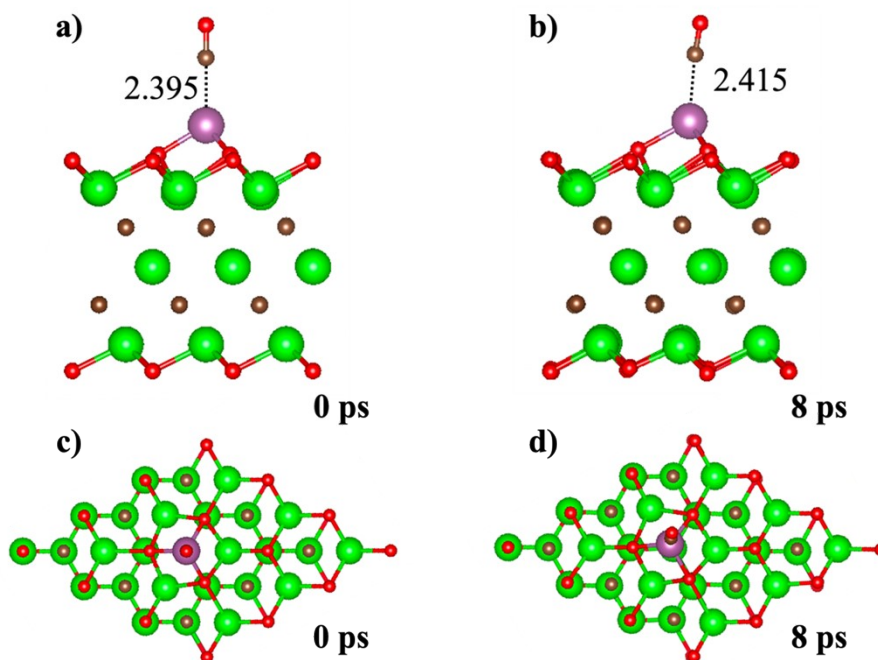

Fig. S20 Snapshot of initail and final structure from AIMD simulation of CO adsorbed on Sc@Zr$_3$C$_2$O$_2$. (a-b) side views at 0 and 8 ps (c-d) top views at 0 and 8 ps. The distance between Sc and C atoms are represented in angstrom.