Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

Accelerating Screening and Assembly of Promising MOFs with Open Cu Sites for Isobutene and Isobutane Separation with a Data-Driven Approach

Xi Sun[1], Wang-qiang Lin[1], Kun Jiang[2], Heng Liang[1] and Guang-hui Chen[1,*]

[1]Department of Chemistry and Key Laboratory for Preparation and Application of Ordered Structural Materials of Guangdong Province, Shantou University, Guangdong 515063, P. R. China. *E-mail: ghchen@stu.edu.cn

[2]Department of Natural Science, Shantou Polytechnic, Shantou 515041, Guangdong, China.

Table S1 Part of fitted Lennard-Jones and Morse potential parameters of modified force field (MFF) used for CBMC and MD simulations.

Lennard-Jones parameters

Atom Type	ε (J)	σ (Å)	Atom Type	ε (J)	σ (Å)	Atom Type	ε (J)	σ (Å)
Li_	3.00683	2.451	O_OH	85.5469	2.96	C_benz	30.7	3.6
N_	34.722	3.66	C_CH_3I	51.22	4.12	H_benz	25.45	2.36
F_	36.4834	3.0932	I_CH_3I	324.06	2.2	CH_4_sp3	36.27	4.2372
B_	47.8058	3.58141	H_CH_3I	10.01	3.75	CH_3_sp3	77.92	3.69
P_	161.03	3.69723	C_ch2o	52.838	2.96	CH_2_sp3	23.8621	4.0677
S_	173.107	3.59032	O_ch2o	105.676	2.846	CH_sp3	31.38	4.15
Cl_	142.562	3.51932	H_ch2o	7.64893	3.05	CH_2_sp2	88.72	4.09
Br_	186.191	3.51905	O_co2	79	2.8	C_sp3	11.438	3.8983
I_	40.77	4.5	C_co2	27	3.31	CH_sp2	50.4	3.95
H_	7.64893	2.84642	N_n2	36	3.12	C_sp2	46.75	4.32
Cu_	2.5161	3.11369	O_coo1	85.5469	3			
Si_	155.998	3.80414	C_coo1	52.8378	3.75			

Morse potential parameters

Atom pair	D_0(K)	R_0(Å)	α
Cu-CH_3_sp3	275.2189	3.4893	11.4184
Cu-CH_sp3	351.3712	2.6056	10.5356
Cu-CH_2_sp2	517.7893	3.0893	10.4184
Cu-C_sp2	825.3661	2.9515	9.2101

Table S2 Evaluation of seven ML algorithms for the isobutene uptake.

property	model	MAE	RMSE	R^2
$N_{isobutene}$	Ridge	1.4721	2.3106	0.7322
	LASSO	1.4721	2.3106	0.7322
	Elastic Net	1.4721	2.3106	0.7322
	SVM	1.9505	3.0620	0.5297
	ByR	1.4733	2.3091	0.7325
	ANN	0.9939	1.8374	0.8306
	XGBoost	0.5330	0.8054	0.9675

Table S3 Details of ANN and XGBoost models.

ML algorithms	Function	Parameters
ANN	neural_network. MLPClassifier	solver=adam activation= relu alpha=0.0001 hidden_layer_sizes=100 learning_rate_init =0.01
XGB	xgb.XGB Regressor	min_child_weight=0.5 max_depth=10 eta=0.1 gamma=0.2 subsample=0.8 colsample_bytree=0.9 reg_lambda=0.7 reg_alpha=0.8

Table S4 The physical parameters as well as XGBoost predicted and CBMC calculated uptakes, APS and selectivities on isobutene/isobutane (1:1) mixture of top 100 G-MOFs, 298 K and 100 kPa.

species	XGBoost predicted isobutene loading (mmol/g)	XGBoost predicted APS	CBMC calculated isobutene loading (mmol/g)	CBMC calculated APS	CBMC calculated selectivity	species	XGBoost predicted isobutene loading (mmol/g)	XGBoost predicted APS	CBMC calculated isobutene loading (mmol/g)	CBMC calculated APS	CBMC calculated selectivity
Cu_2O_8-mof177-TDPAT_No730	21.51	108.76	21.92	111.35	5.08	Cu_2O_8-pcn6_B-TDPAT_No474	20.79	108.16	20.53	97.39	4.75
Cu_2O_8-mof177-TDPAT_No772	20.98	108.05	21.07	108.65	5.16	Cu_2O_8-pcn6_B-TDPAT_No797	23.93	108.51	21.21	97.37	4.59
Cu_2O_8-mof177-TDPAT_No728	22.22	109.35	22.00	107.42	4.88	Cu_2O_8-mof177-mofHTB_No759	21.39	108.42	20.55	97.30	4.75
Cu_2O_8-BTC_A-mof177_No22	21.27	108.14	21.21	105.07	4.96	Cu_2O_8-pcn6_B-TDPAT_No395	20.79	108.16	20.49	97.30	4.75
Cu_2O_8-mof177-TDPAT_No35	22.44	109.04	22.15	104.74	4.73	Cu_2O_8-pcn6_B-TDPAT_No150	20.79	108.16	20.54	97.29	4.74
Cu_2O_8-mof177-TDPAT_No327	22.44	109.04	21.92	104.59	4.77	Cu_2O_8-pcn6_B-TDPAT_No718	23.93	108.53	21.26	97.09	4.57
Cu_2O_8-mof177-TDPAT_No328	21.13	109.34	20.63	104.50	5.07	Cu_2O_8-mof177-mofHTB_No379	21.04	107.99	20.22	96.86	4.80
Cu_2O_8-pcn6_B-TDPAT_No309	20.78	107.85	21.15	104.02	4.92	Cu_2O_8-pcn6_B-TDPAT_No556	23.93	108.51	21.25	96.60	4.55
Cu_2O_8-mof177-TDPAT_No323	21.02	109.54	20.97	104.00	4.96	Cu_2O_8-pcn6_B-TDPAT_No231	20.79	108.16	20.39	95.56	4.69
Cu_2O_8-mof177-TDPAT_No774	21.13	108.01	20.73	103.77	5.02	Cu_2O_8-BTC_B-mof200_No29	20.81	107.95	19.55	95.48	4.92
Cu_2O_8-mofHTB-TDPAT_No329	22.14	108.02	21.42	103.73	4.85	Cu_2O_8-pcn6_B-TDPAT_No473	23.96	108.77	20.99	95.41	4.55
Cu_2O_8-pcn6_B-TDPAT_No715	20.79	108.45	20.60	103.57	5.03	Cu_2O_8-pcn6_B-TDPAT_No71	20.56	108.09	20.11	95.36	4.74
Cu2O8-mof177-TDPAT_No36	21.09	109.32	20.76	103.10	4.97	Cu_2O_8-mof177-TDPAT_No326	18.11	109.38	17.86	95.33	5.36
Cu_2O_8-mof177-mofHTB_No335	17.48	108.11	18.54	102.43	5.66	Cu_2O_8-mof177-TDPAT_No715	21.20	108.49	19.70	95.27	4.84
Cu_2O_8-mof177-TDPAT_No165	22.44	109.04	21.94	102.23	4.66	Cu_2O_8-pcn6_B-TDPAT_No716	23.96	108.82	20.99	95.23	4.54
Cu_2O_8-pcn6_B-TDPAT_No714	22.32	108.18	21.71	102.13	4.71	Cu_2O_8-pcn6_B-TDPAT_No798	20.79	108.16	20.29	95.18	4.69
Cu_2O_8-mof177-TDPAT_No768	22.28	108.14	21.61	101.70	4.71	Cu_2O_8-pcn6_B-TDPAT_No152	20.79	108.16	20.36	95.18	4.68

Cu_2O_8-mof177-TDPAT_No202	21.06	107.92	20.71	101.49	4.90	Cu_2O_8-mof177-TDPAT_No404	22.47	109.35	20.60	95.03	4.61
Cu_2O_8-mof177-TDPAT_No813	22.44	109.16	21.63	101.42	4.69	Cu_2O_8-pcn6_B-TDPAT_No147	22.28	108.27	20.41	94.89	4.65
Cu_2O_8-mof177-TDPAT_No489	22.51	109.14	21.67	101.32	4.68	Cu_2O_8-mof177-TDPAT_No566	21.56	108.38	20.47	94.70	4.63
Cu_2O_8-mof177-TDPAT_No647	21.02	109.76	20.58	101.22	4.92	Cu_2O_8-mof177-TDPAT_No607	20.94	107.92	19.97	94.59	4.74
Cu_2O_8-mof177-TDPAT_No652	21.06	109.34	20.42	101.22	4.96	Cu_2O_8-pcn6_B-TDPAT_No228	22.37	108.52	20.28	94.54	4.66
Cu_2O_8-mof177-TDPAT_No161	21.56	108.61	21.22	101.22	4.77	Cu_2O_8-pcn6_B-TDPAT_No311	23.93	108.51	20.95	94.48	4.51
Cu_2O_8-pcn6_B-TDPAT_No312	20.79	108.16	20.69	100.96	4.88	Cu_2O_8-mofHTB-TDPAT_No314	21.26	108.06	19.57	94.47	4.83
Cu_2O_8-mof177-TDPAT_No84	22.44	109.04	21.43	100.89	4.71	Cu_2O_8-mof177-TDPAT_No406	20.92	108.72	19.98	94.41	4.73
Cu_2O_8-pcn6_B-TDPAT_No800	20.79	108.16	20.59	100.56	4.89	Cu_2O_8-pcn6_B-TDPAT_No472	20.86	108.40	19.40	94.15	4.86
Cu_2O_8-mofHTB-TDPAT_No854	21.34	108.02	20.67	100.47	4.86	Cu_2O_8-mof177-TDPAT_No485	22.30	115.05	20.25	94.08	4.65
Cu_2O_8-mof177-TDPAT_No770	22.28	108.22	21.57	100.30	4.65	Cu_2O_8-mofHTB-TDPAT_No699	22.41	107.92	20.09	94.05	4.68
Cu_2O_8-pcn6_B-TDPAT_No717	20.79	108.16	20.73	100.24	4.84	Cu_2O_8-mof177-TDPAT_No526	21.77	115.33	19.70	93.81	4.76
Cu_2O_8-pcn6_B-TDPAT_No719	20.79	108.16	20.60	100.13	4.86	Cu_2O_8-pcn6_B-TDPAT_No799	23.93	108.53	20.89	93.68	4.49
Cu_2O_8-pcn6_B-TDPAT_No393	20.86	108.16	20.69	100.07	4.84	Cu_2O_8-pcn6_B-TDPAT_No313	23.93	108.53	21.01	93.31	4.44
Cu_2O_8-mof177-TDPAT_No163	21.74	109.61	20.68	100.03	4.84	Cu_2O_8-mof177-mofHTB_No641	17.96	108.07	17.85	93.30	5.30
Cu_2O_8-mof177-mofHTB_No377	21.54	108.63	20.82	100.01	4.81	Cu_2O_8-pcn6_B-TDPAT_No637	23.93	108.53	20.75	93.30	4.50
Cu_2O_8-mof177-TDPAT_No651	22.44	109.04	21.54	99.80	4.63	Cu_2O_8-mof177-TDPAT_No369	23.92	108.47	20.45	93.26	4.56
Cu_2O_8-pcn6_B-TDPAT_No638	20.64	108.16	20.50	99.42	4.85	Cu_2O_8-mof177-mofHTB_No756	22.22	107.94	20.32	93.18	4.59
Cu_2O_8-mof177-TDPAT_No205	21.15	108.06	20.18	99.41	4.93	Cu_2O_8-pcn6_B-TDPAT_No391	20.79	108.45	19.60	93.05	4.75
Cu_2O_8-pcn6_B-TDPAT_No555	20.79	108.16	20.64	99.15	4.81	Cu_2O_8-mof177-TDPAT_No650	18.11	109.38	17.66	92.96	5.28
Cu_2O_8-pcn6_B-TDPAT_No557	20.79	108.16	20.60	99.05	4.81	Cu_2O_8-mof177-TDPAT_No246	17.82	108.78	18.03	92.89	5.24
Cu_2O_8-pcn6_B-TDPAT_No314	20.64	108.16	20.56	99.03	4.82	Cu_2O_8-pcn6_B-TDPAT_No151	23.96	108.53	20.73	92.67	4.47
Cu_2O_8-mof177-TDPAT_No490	21.11	110.67	20.40	98.94	4.85	Cu_2O_8-mof177-TDPAT_No767	23.94	108.81	20.67	92.31	4.47
Cu_2O_8-mof177-TDPAT_No166	21.00	109.58	20.38	98.69	4.84	Cu_2O_8-pcn6_B-TDPAT_No635	23.93	108.51	20.78	92.02	4.43
Cu_2O_8-mof177-TDPAT_No242	20.88	108.72	20.64	98.36	4.77	Cu_2O_8-pcn6_B-TDPAT_No394	23.93	108.82	20.53	91.79	4.47

Cu_2O_8-mof177-TDPAT_No486	20.54	111.20	19.59	98.36	5.02	Cu_2O_8-mofHTB-TDPAT_No328	22.12	108.02	19.79	91.78	4.64
Cu_2O_8-mofHTB-TDPAT_No653	22.07	108.18	20.75	98.30	4.74	Cu_2O_8-mof177-TDPAT_No568	21.67	109.59	19.78	91.74	4.64
Cu_2O_8-pcn6_B-TDPAT_No636	20.79	108.16	20.40	98.20	4.82	Cu_2O_8-pcn6_B-TDPAT_No232	23.93	108.51	20.73	91.71	4.42
Cu_2O_8-pcn6_B-TDPAT_No22	20.86	108.16	20.59	98.12	4.77	Cu_2O_8-mof177-mofHTB_No333	22.22	108.00	19.95	91.68	4.60
Cu_2O_8-pcn6_B-TDPAT_No233	20.79	108.16	20.48	97.72	4.77	Cu_2O_8-pcn6_B-TDPAT_No230	23.93	108.82	20.64	91.63	4.44
Cu_2O_8-pcn6_B-TDPAT_No69	20.79	108.16	20.40	97.71	4.79	Cu_2O_8-pcn6_B-TDPAT_No293	21.12	108.28	18.98	91.26	4.81
Cu_2O_8-mofHTB-pcn6_B_No308	22.12	107.97	20.59	97.60	4.74	Cu_2O_8-pcn6_B-TDPAT_No68	23.93	108.82	20.56	90.86	4.42
Cu_2O_8-pcn6_B-TDPAT_No476	20.79	108.16	20.44	97.54	4.77	Cu_2O_8-pcn6_B-TDPAT_No21	23.96	108.51	20.49	90.65	4.42

Table S5 The calculated physical parameters as well as their uptakes, working capacities, selectivities, regenerabilities and APS on isobutene/isobutane (1:1) mixture of assembled MOFs, 298 K and 100 kPa.

species	VOL(Å^3)	ρ(g/cm^3)	GCD(Å)	PLD(Å)	LCD(Å)	Sa(Å^2)	φ
Cu_2O_8-irmof14_A-core-1_No1	4667.57	0.60	8.53	7.08	8.53	1527.37	0.74
Cu_2O_8-irmof12_A-core-1_No1	4680.06	0.59	8.66	6.12	8.65	1534.55	0.73
Cu_2O_8-irmof10_A-core-1_No1	4817.63	0.54	8.66	7.11	8.66	1534.58	0.77

species	Isobutene uptake (mmol/g)	Isobutane uptake (mmol/g)	Isobutene working capacity (mmol/g)	*S_{ads}	R%	APS	Reference
Cu_2O_8-irmof14_A-core-1_No1	7.09	1.42	6.89	4.99	97.27	34.43	this work
Cu_2O_8-irmof12_A-core-1_No1	5.97	1.30	5.87	4.58	98.38	26.86	this work
Cu_2O_8-irmof10_A-core-1_No1	5.56	1.43	5.54	3.89	99.65	21.56	this work

*Selectivities (S_{ads}) were calculated by IAST method.

Table S6 Simulated diffusion selectivities of top five MOFs on isobutene/isobutane adsorption separation.

species	$D_{isobutane}$ (10^{-8} m^2/s)	$D_{isobutene}$ (10^{-8} m^2/s)	S_{diff}
AVAKEP	1.81	1.38	0.76
XAHPON	1.18	1.13	0.96
HUNCIE	1.30	0.90	0.69
Cu_2O_8-mof177-TDPAT_No730	1.40	0.91	0.65
Cu_2O_8-BTC_B-core-4_No1	1.35	0.93	0.69

Diffusion selectivity S_{diff}: $D_{isobutene}/D_{isobutane}$

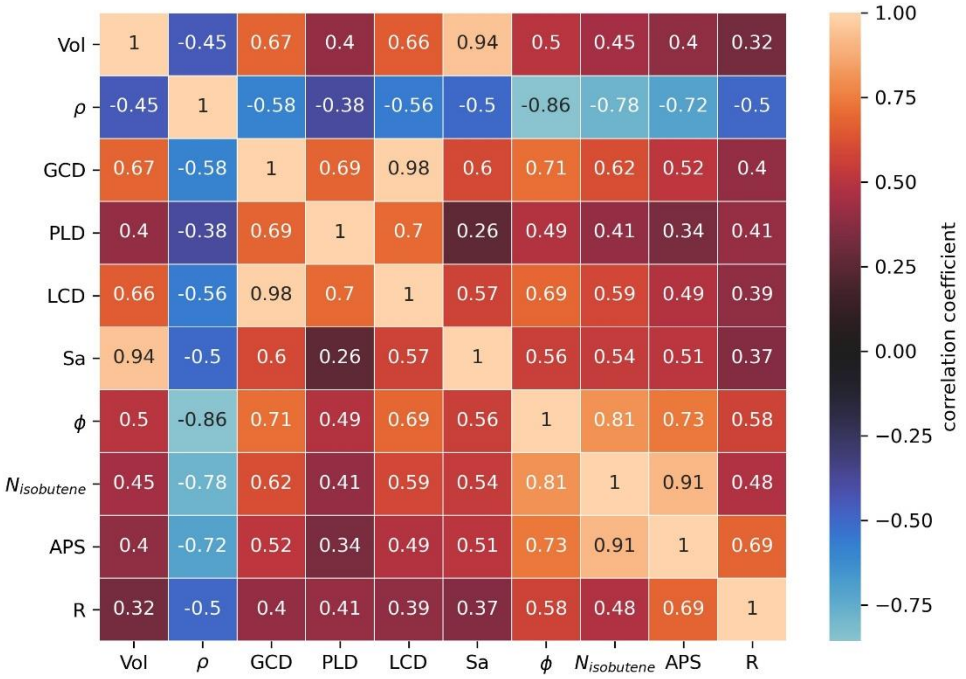


Figure S1. Correlation diagram of material features and adsorption properties of isobutene/isobutane based on the CoRE MOF database. Note that the color bars represent the size of the Pearson correlation coefficients.

descriptors / species	node1	node2	.....	linker1	linker2	.....
MOF1	1	0	.....	1	0	.....
MOF2	0	1	.....	0	1	.....
.....	.....	.....	.....	.....	.....	.....

Figure S2. One-hot encoding is a process of converting categorical variables into a form that is easy to use for ML algorithms in order to prepare it for an algorithm and get a better prediction. Here, with one-hot, we convert each categorical value into a new categorical column and assign a binary value of 1 or 0, corresponding to the presence and absence of the linker or node, respectively, and then use it as a feature of model training.

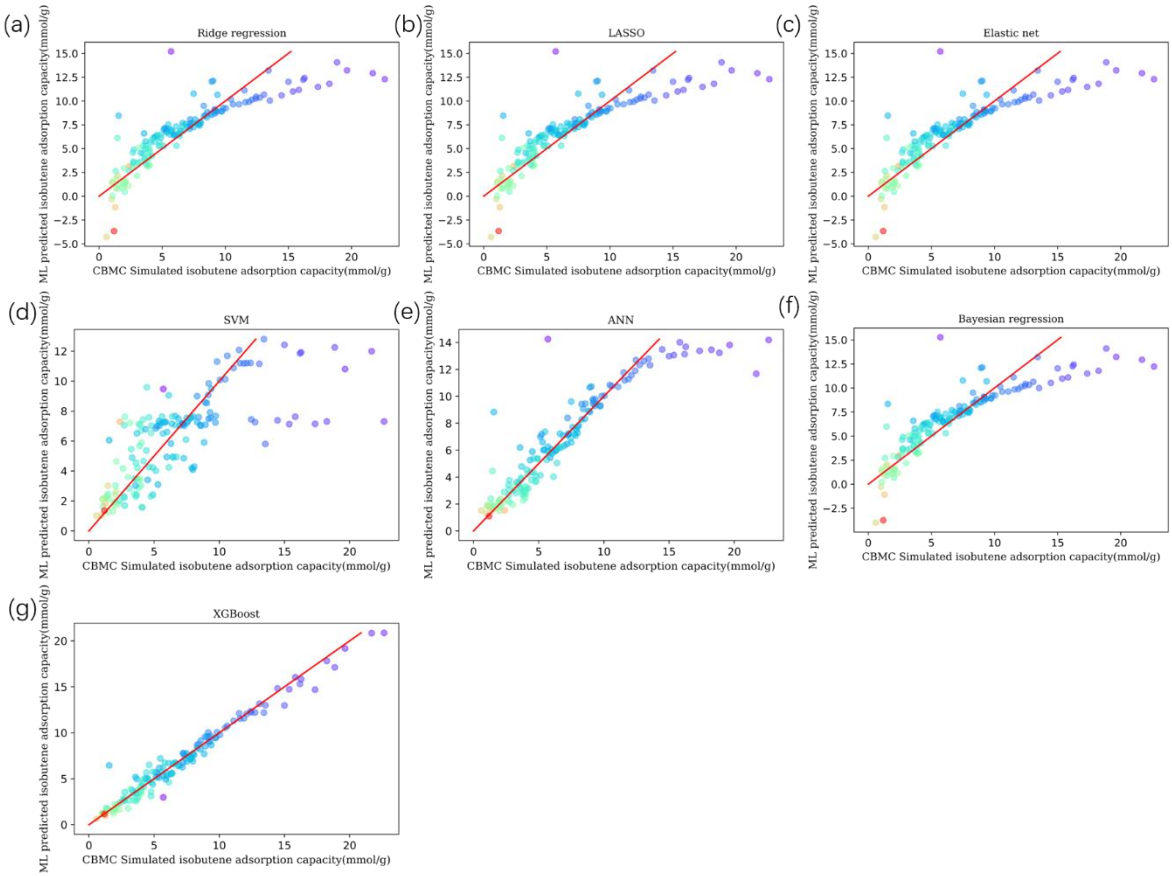
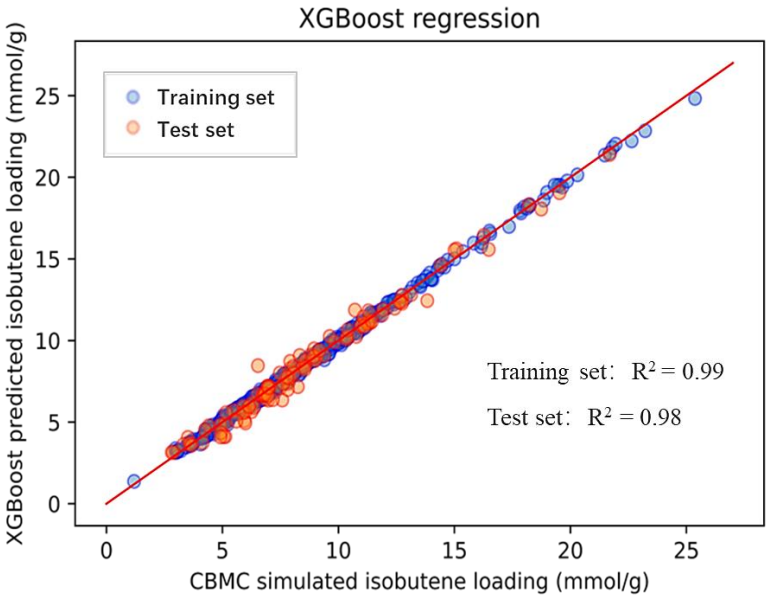
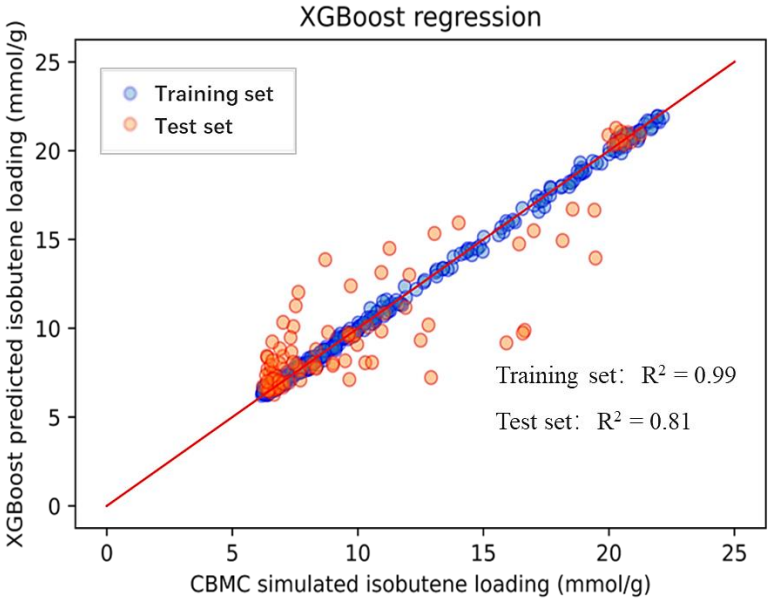


Figure S3. The parity plots representing the predicted and simulated adsorption capacity of MOFs data for the (a) ridge regression (b) LASSO (c) Elastic Net (d) SVM (e) ANN (f) Bayesian regression and (g) XGBoost models.



(a)



(b)

Figure S4. Comparison of isobutene adsorption capacities between the predicted values of XGBoost model and CBMC simulation for (a) CoRE MOF and (b) G-MOF databases, where red and blue dots represent structures in training and test sets, respectively.

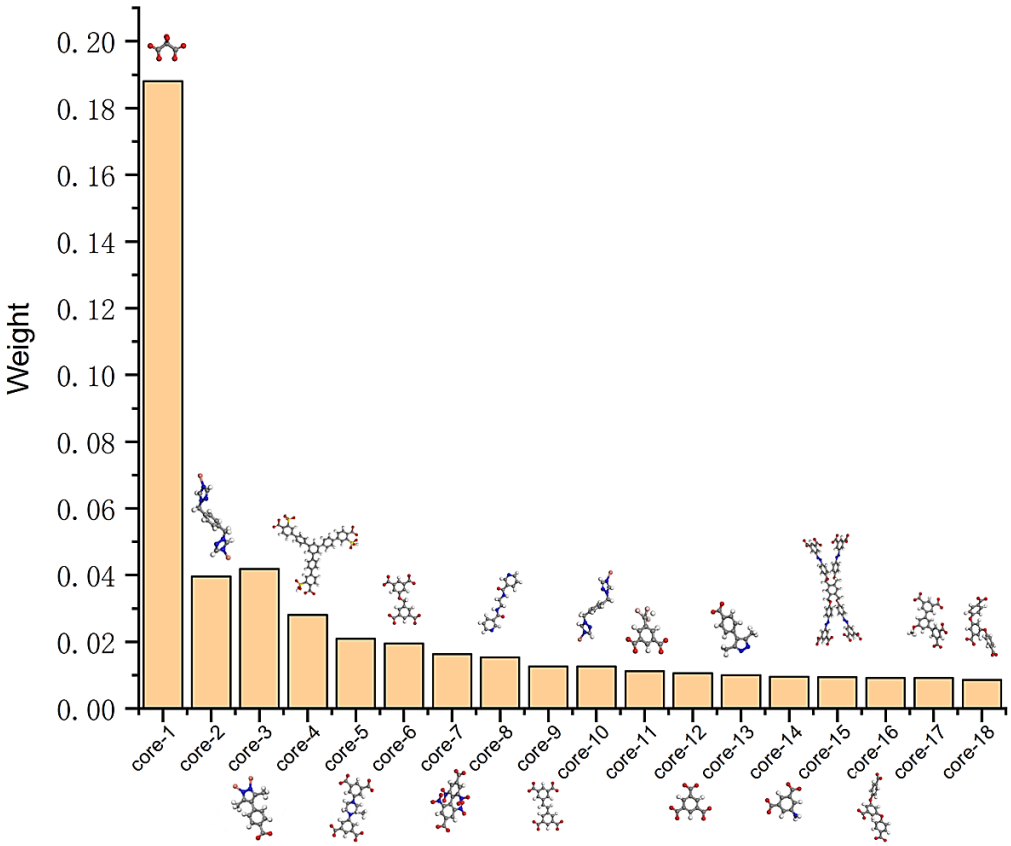


Figure S5. The predicted top genes of frameworks in CoRE MOF with large influence weights.

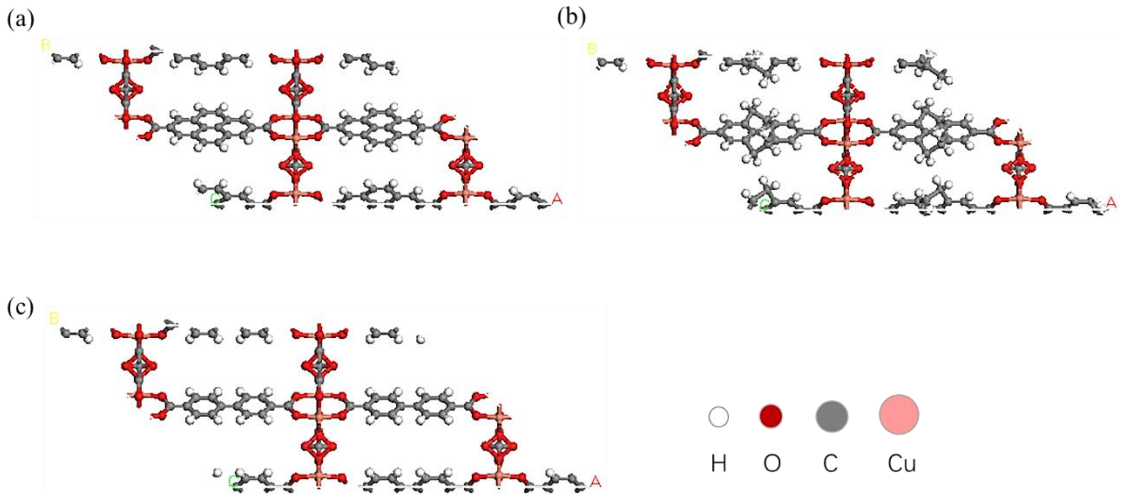


Figure S6. The optimized crystal structures of assembled (a) Cu_2O_8-irmof14_A-core-1_No1, (b) Cu_2O_8-irmof12_A-core-1_No1 and (c) Cu_2O_8-irmof10_A-core-1_No1, where gray, white, red and pink dots represent carbon, hydrogen, oxygen and copper atom, respectively.

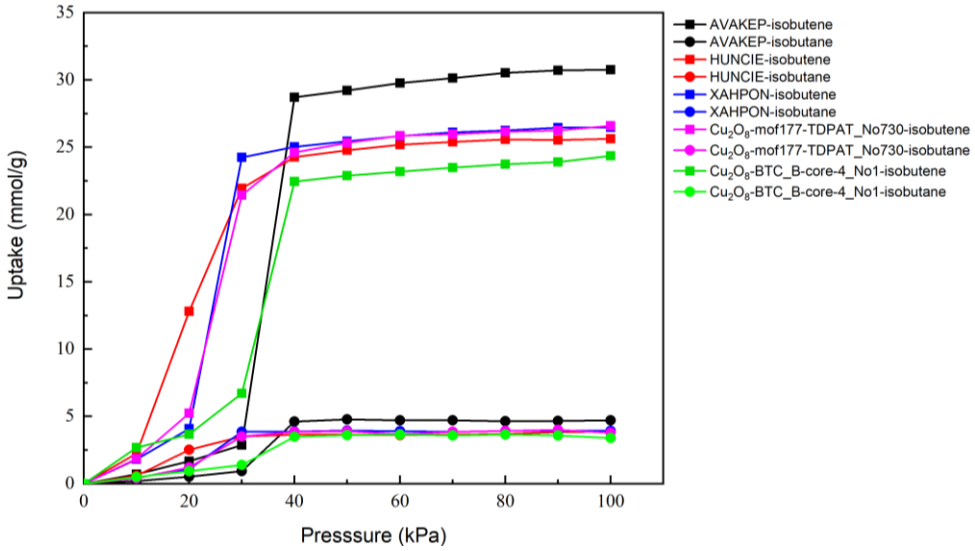
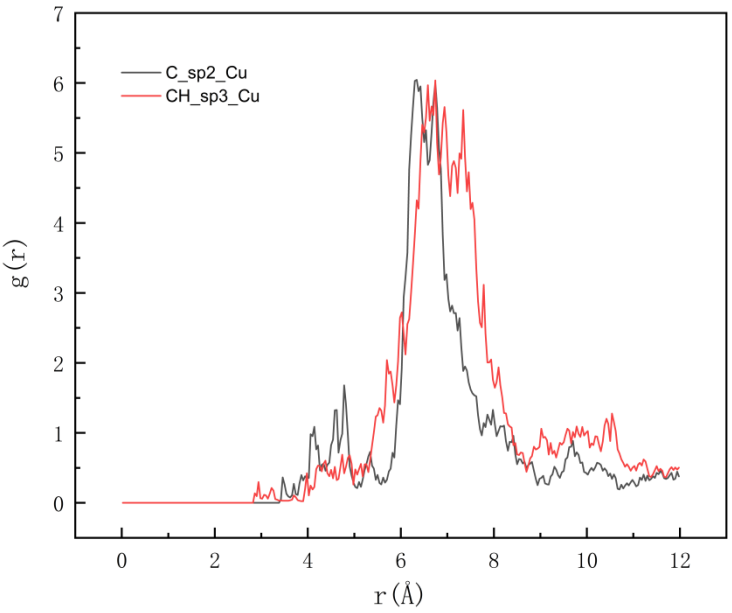
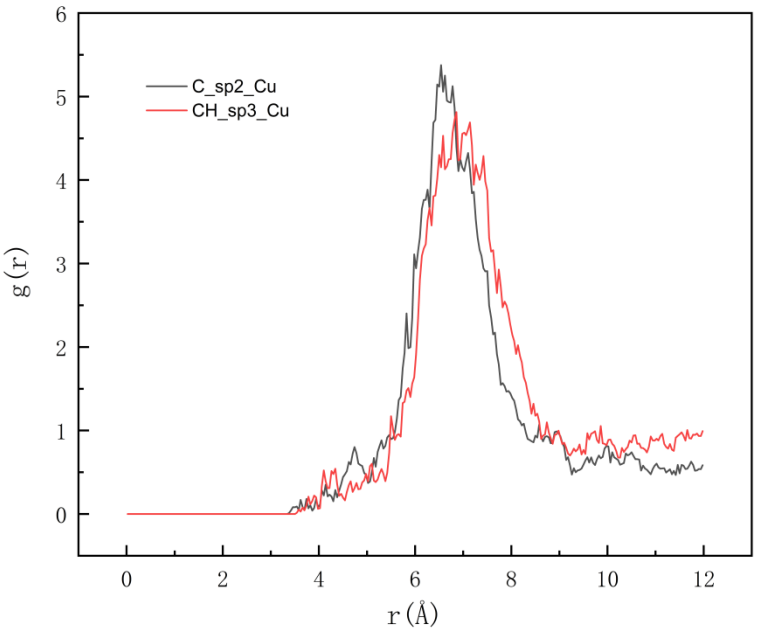


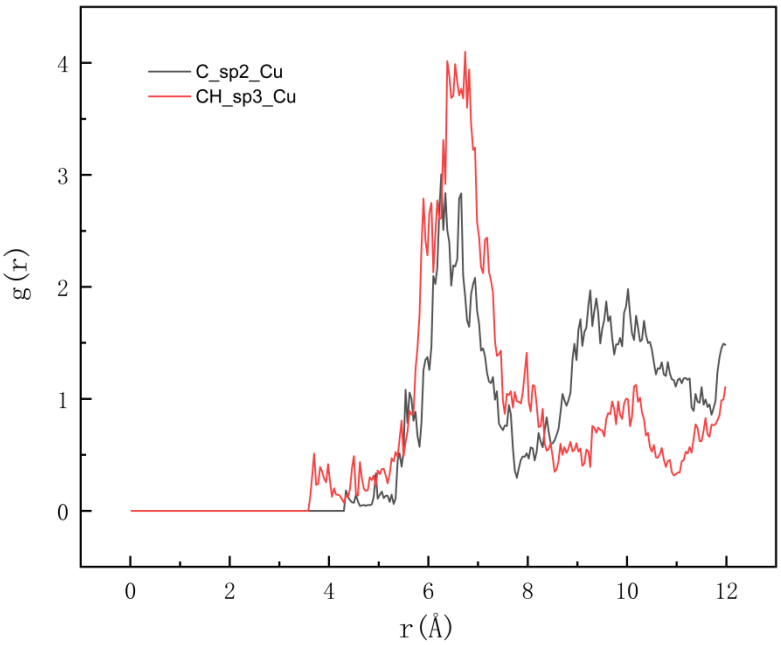
Figure S7. The simulated adsorption isotherms of isobutene and isobutane adsorbed on the five promising MOFs at 273 K.


C_sp2_Cu
CH_sp3_Cu
g(r)
r (Å)
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12

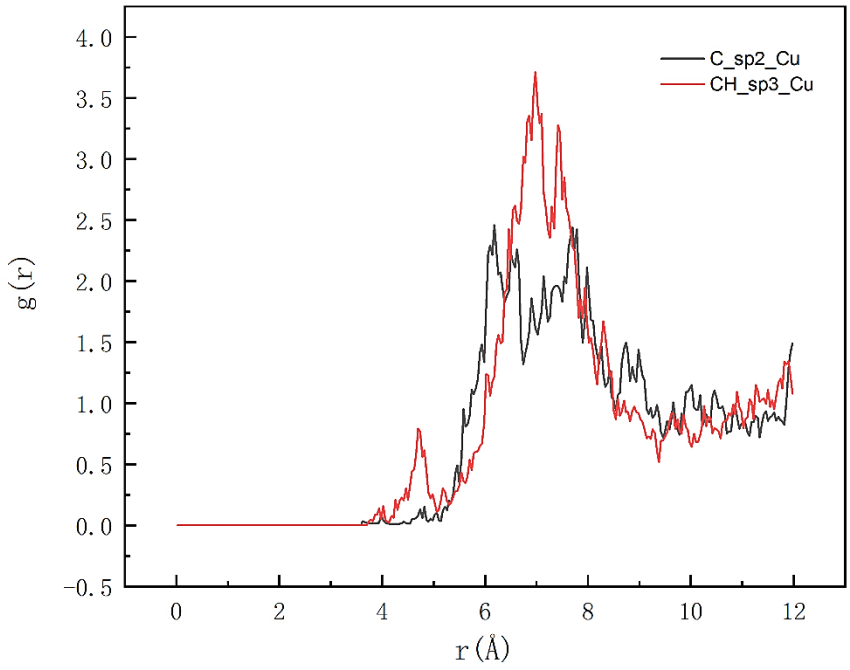
(a)


C_sp2_Cu
CH_sp3_Cu
g(r)
r (Å)
0
1
2
3
4
5
6
0
2
4
6
8
10
12

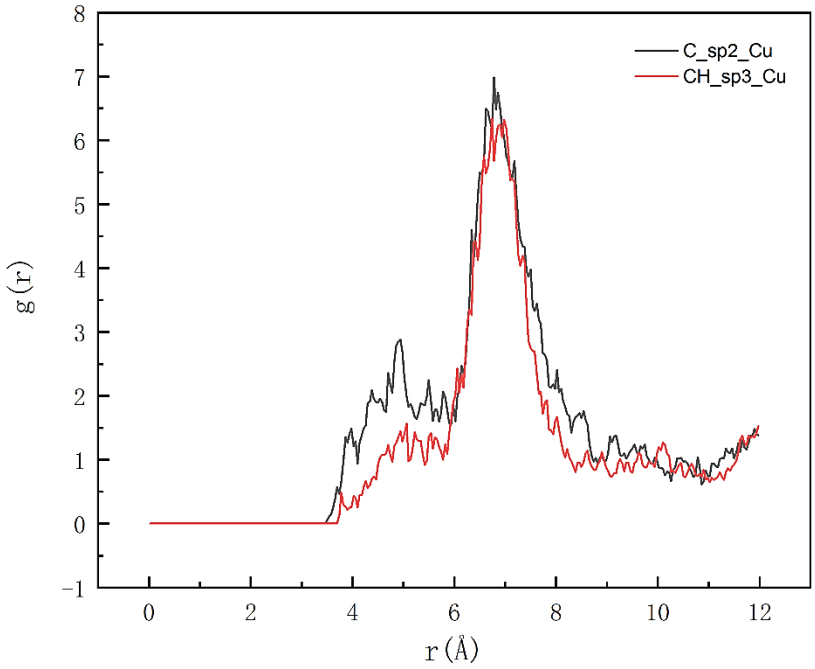
(b)


C_sp2_Cu
CH_sp3_Cu
g(r)
r(Å)
0
1
2
3
4
0
2
4
6
8
10
12

(c)


C_sp2_Cu
CH_sp3_Cu
g(r)
r(Å)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
0
2
4
6
8
10
12

(d)



(e)

Figure S8. The RDF plots for isobutene and isobutane on (a) AVAKEP (b)XAHPON (c)HUNCIE (d) Cu_2O_8-mof177-TDPAT_No730 and (e) Cu_2O_8-BTC_B-mof200 at 1 bar and 298 K.

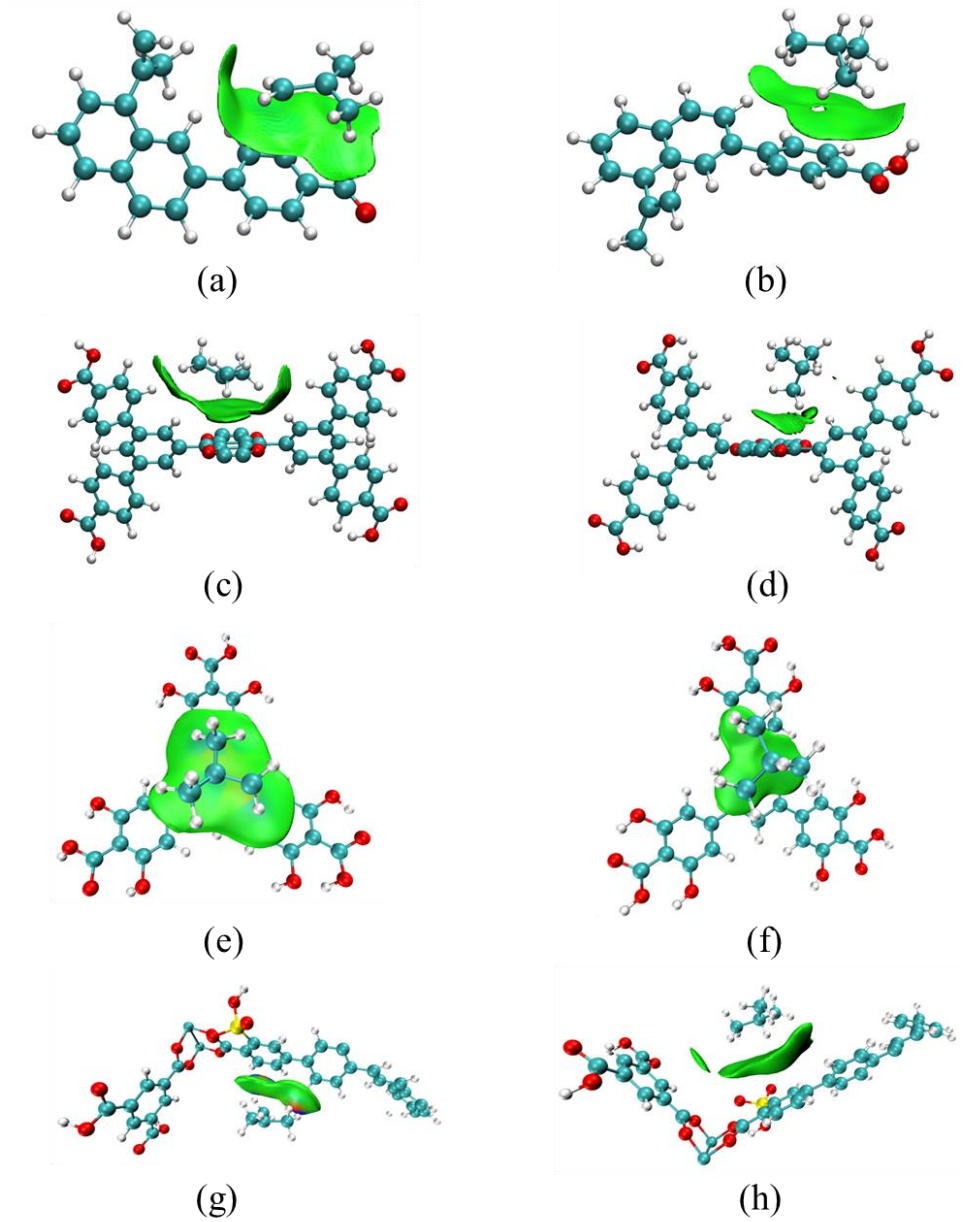


Figure S9. The IGMH analyses of the weak interactions, including (a) XAHPON-isobutene, (b) XAHPON-isobutane, (c) HUNCIE-isobutene, (d) HUNCIE-isobutane, (e) Cu_2O_8-mof177-TDPAT_No730-isobutene, (f) Cu_2O_8-mof177-TDPAT_No730-isobutane, (g) Cu_2O_8-BTC_B-mof200-isobutene, and (h) Cu_2O_8-BTC_B-mof200-isobutane, respectively. Note that all the isosurfaces are set as 0.005 a.u., blue represents attractive interaction, green for van der Waals interaction, and red for repulsive interaction.

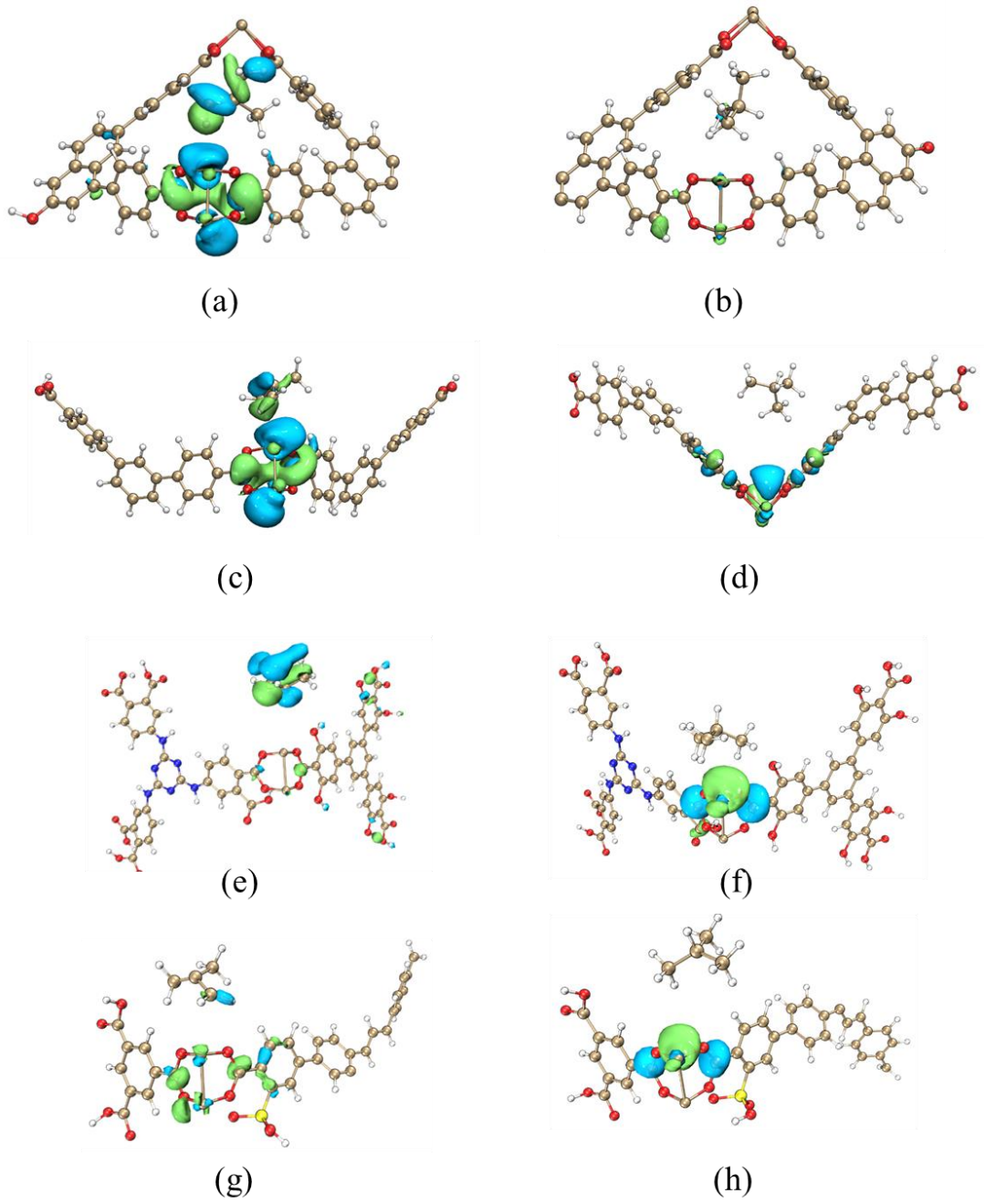


Figure S10. The complex natural orbitals (NO) calculated at PBE level for (a) XAHPON -isobutene, (b) XAHPON -isobutane, (c) HUNCIE -isobutene, (d) HUNCIE -isobutane, (e) Cu_2O_8-mof177-TDPAT_No730-isobutene, (f) Cu_2O_8-mof177-TDPAT_No730-isobutane, (g) Cu_2O_8-BTC_B-mof200-isobutene, and (h) Cu_2O_8-BTC_B-mof200-isobutane complex. The orbital isovalue is set to 0.05, with green and blue parts corresponding to positive and negative regions of orbital phase, respectively.

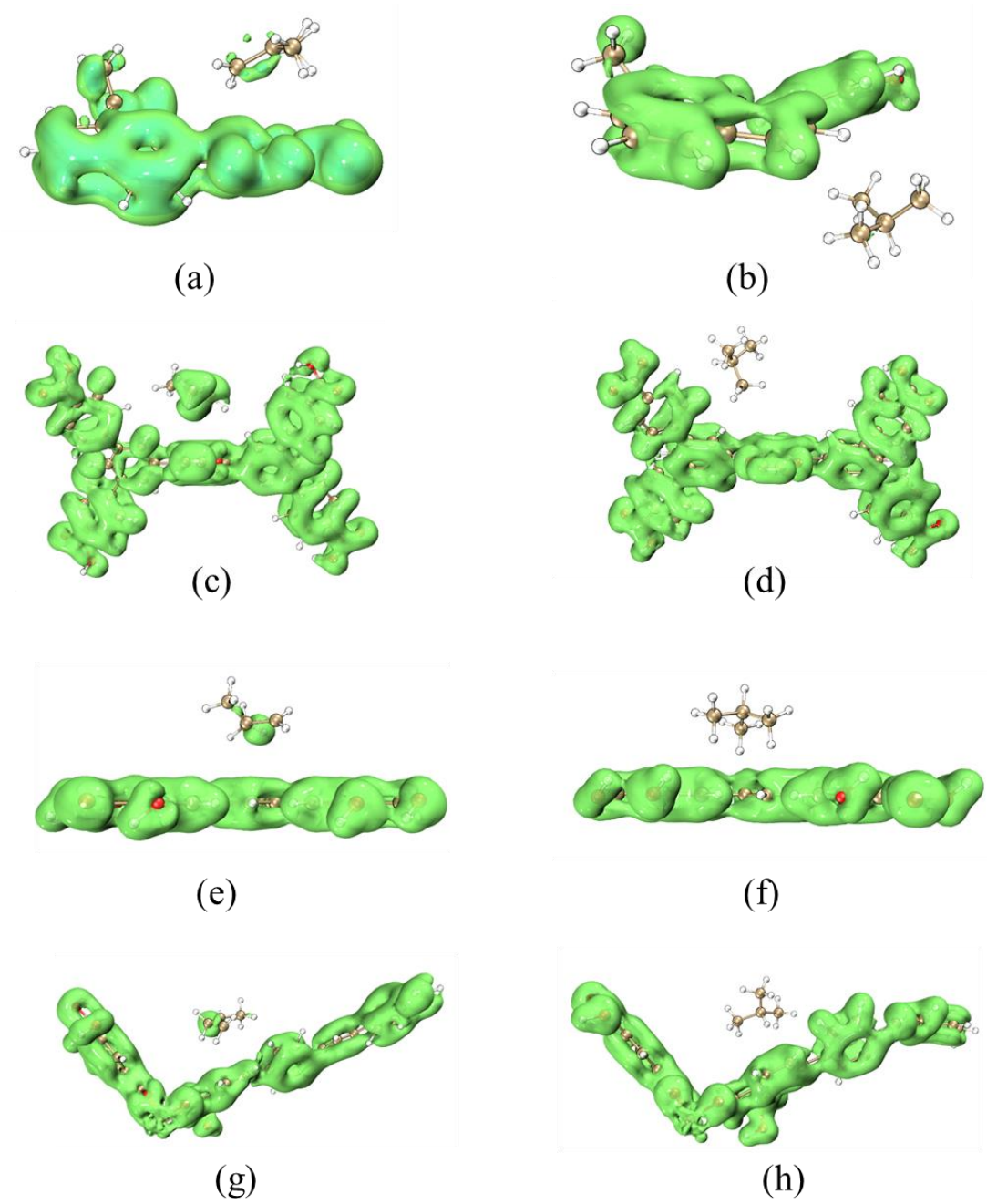


Figure S11. The LOL-π isosurfaces of (a) XAHPON -isobutene, (b) XAHPON -isobutane, (c) HUNCIE -isobutene, (d) HUNCIE -isobutane, (e) Cu_2O_8-mof177-TDPAT_No730-isobutene, (f) Cu_2O_8-mof177-TDPAT_No730-isobutane, (g) Cu_2O_8-BTC_B-mof200-isobutene, and (h) Cu_2O_8-BTC_B-mof200-isobutane, respectively, with isovalue of 0.45.