Supplementary Information (SI) for Molecular Systems Design & Engineering.
This journal is © The Royal Society of Chemistry 2025

Supporting Information for:

Computational Investigation of the Impact of Metal-Organic Framework Topology on Hydrogen Storage Capacity

Kunhuan Liu[1], Haoyuan Chen[2], Timur Islamoglu[3], Andrew S. Rosen[1], Xijun Wang[1], Omar K. Farha[1,3], Randall Q. Snurr[1]*

[1] Department of Chemical & Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States

[2] Department of Chemistry, Southern Methodist University, Dallas, Texas 75275, United States

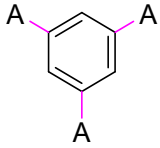
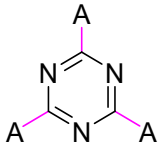
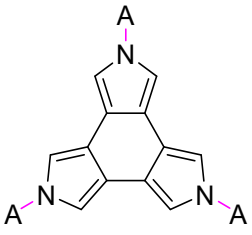
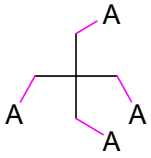
[3] Department of Chemistry, Northwestern University, Evanston, Illinois 60208, United States

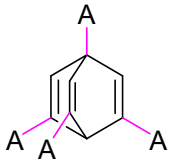
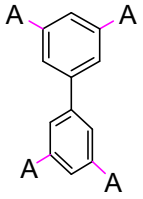
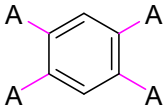
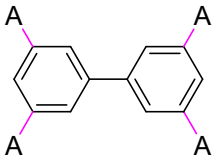
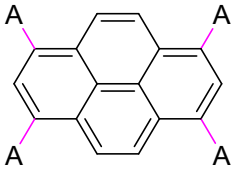
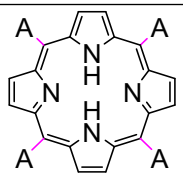
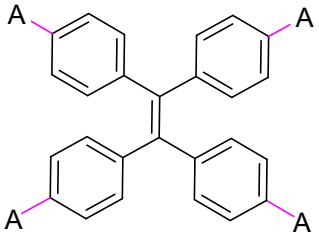
1 Building blocks

The 3d structures of all building blocks and the composition, textual properties and simulation results of the MOFs studied are available on a Zenodo repository.[1]

Partway through our work, we noticed that there was a mistake where edge building block E34B (Table S3) was a variant of E25A with the same chemical structure. The resulting MOF variants have slight structural differences, likely due to differences in the building block 3d coordinates. Therefore, we kept both variants in the data analysis.

Table S1. Organic node (ON) building blocks used in MOF construction. The ON code corresponds to the identifiers used in our dataset. The identifiers for these building blocks as found in the ToBaCCo database[2] are listed in brackets. In the chemical diagram, "A" denotes an "any" atom placeholder, and the purple bond indicates the connection to other building blocks.

ON code [ToBaCCo code]	**Structure**	**SMILES**
ON1 [Tob_ON_2]		[*]C1=CC([*])=CC([*])=C1
ON2 [Tob_ON_1]		[*]C1=NC([*])=NC([*])=N1
ON3 [Tob_ON_5]		[*]N1C=C(C2=CN([*])C=C2C3=CN([*])C=C34)C4=C1
ON4		[*]CC(C[*])(C[*])C[*]

ON5 [Tob_On_7]		[*]C1=CC2([*])C=C([*])C1C([*])=C2
ON6 [Tob_On_8]	 (tetrahedral conformation)	[*]C1=CC([*])=CC(C2=CC([*])=CC([*])=C2)=C1
ON7 [Tob_On_9]		[*]C1=C([*])C=C([*])C([*])=C1
ON8 [Tob_On_10]	 (planar conformation)	[*]C1=CC([*])=CC(C2=CC([*])=CC([*])=C2)=C1
ON9 [Tob_On_11]		[*]C1=C(C=C2)C3=C4C2=C([*])C=C([*])C4=CC=C3C([*])=C1
ON10 [Tob_On_12]		[*]C1=C2C=CC(=N2)C(=C2C=CC(=C(C3=NC(=C(C4NC1=CC=4)[*])C=C3)[*])N2)[*]
ON11 [Tob_On_13]		[*]C1C=CC(=CC=1)C(=C(C1C=CC(=CC=1)[*])C1C=CC(=CC=1)[*])C1C=CC(=CC=1)[*]

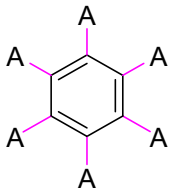
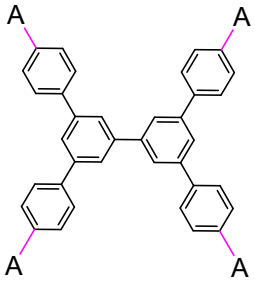
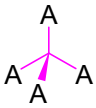
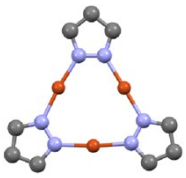
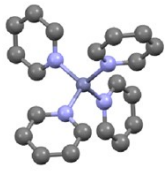
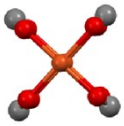
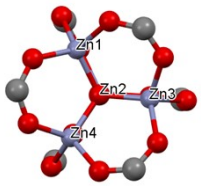
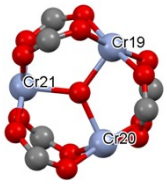
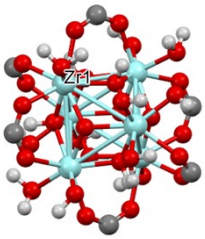
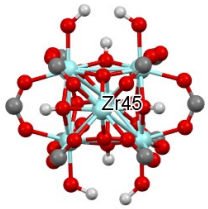
ON12 [Tob_On_14]		[*]C1C([*])=C([*])C(=C(C=1[*])[*])[*]
ON13		[*]C1=CC=C(C2=CC(C3=CC=C([*])C=C3)=CC(C4=CC(C5=CC=C([*])C=C5)=CC(C6=CC=C([*])C=C6)=C4)=C2)C=C1
ON14 [Tob_On_6]		[*]C([*])([*])[*]

Table S2. Metal cluster node (MC) building blocks used in MOF construction. The "MC" code corresponds to the identifiers used in our dataset. The identifiers for these building blocks as found in the ToBaCCo database[2] are listed in brackets.

MC code [ToBaCCo code]	**Structure**	**Description**
MC1 [sym_3_mc_0]		3-c Cu
MC2 [sym_4_mc_1]		4-c Zn
MC3 [sym_5_mc_2]		4-c Cu paddlewheel

MC4 [sym_6_mc_3]		6-c Zn
MC5 [sym_7_mc_4]		6-c Cr
Mc6 [sym_9_mc_5]		6-c Zr
MC7 [sym_8_mc_7]		8-c Co
Mc8 [sym_8_mc_9]		8-c Zr
Mc9 [sym_10_mc_10]		10-c Zr

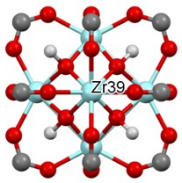
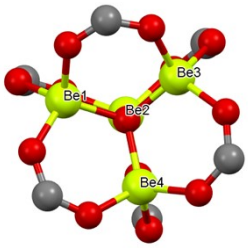
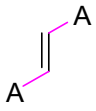
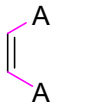
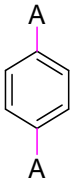
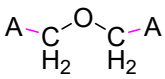
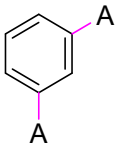
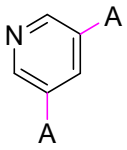
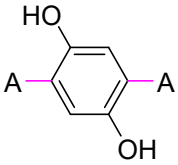
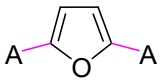
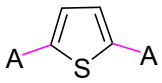
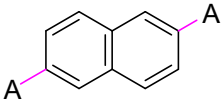
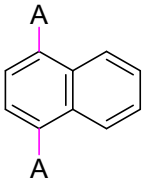
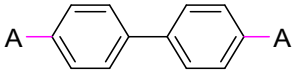
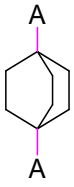
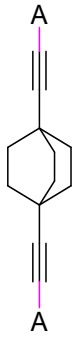
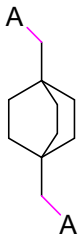
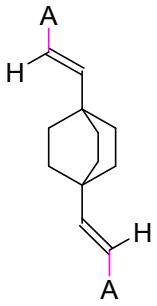
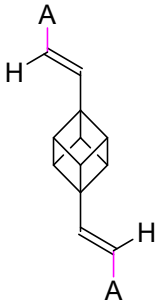
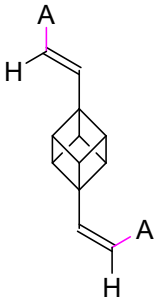
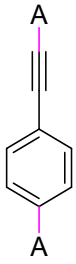
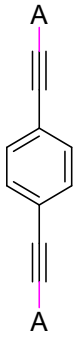
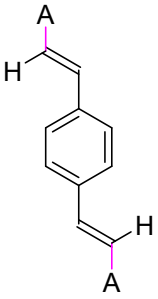
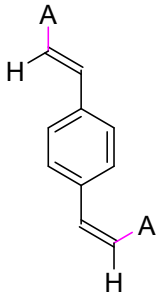
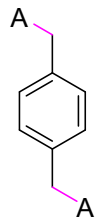
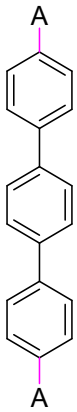
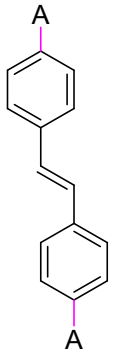
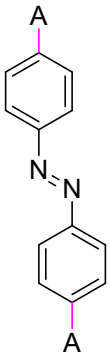
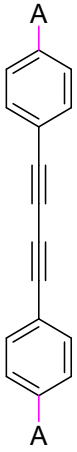
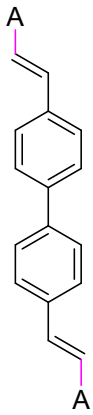
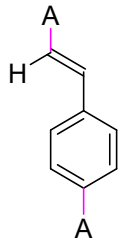
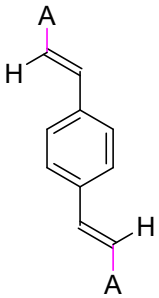
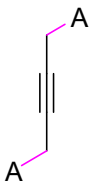
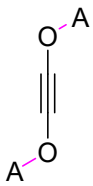
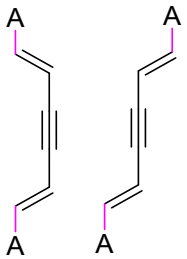
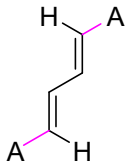
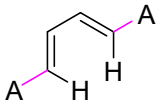
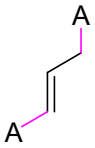
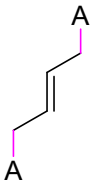
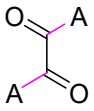
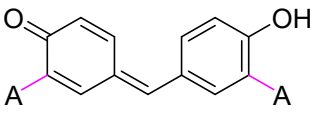
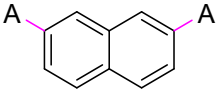
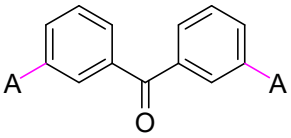
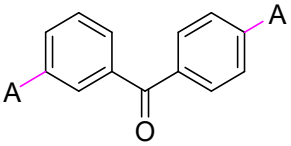
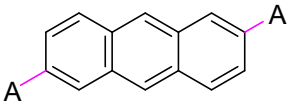
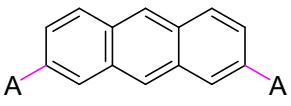
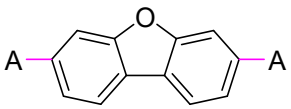
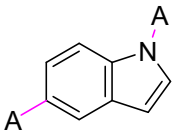
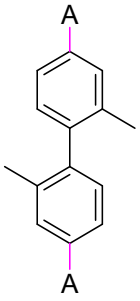
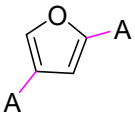
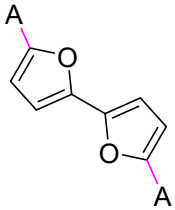
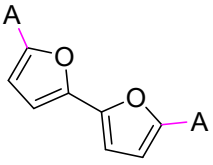
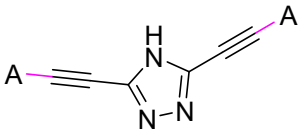
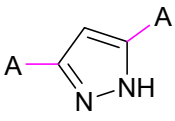
Mc10 [sym_12_mc_11]		12-c Zr
Mc11 [sym_24_mc_13]		24-c Cu
Mc12		6-c Be

Table S3. Edge (E) building blocks used in MOF construction. The shaded rows indicate the identifiers used in this dataset, and the corresponding ToBaCCo database[2] identifiers are listed in brackets, if available. The chemical diagrams are shown under the respective identifiers. In the chemical diagram, the purple bond indicates the connection to other building blocks, and the "A" denotes an atom placeholder replaceable with either metal cluster or organic node connection point (usually a carbon atom).

E0 [NTN]	E1 [CBB_2]	E2	E3
(No real atom)			
E4 [CBB_12]	E5	E6	E7
			
E8	E9	E10 [CBB_5]	E11
			
E12 [CBB_17]	E13 [CBB_8]	E14 [CBB_24]	E15 [CBB_11]
			

E16	E17	E18	E19
			
E20	**E21A (“EE”)**	**E21B (“EZ”)**	**E22**
			
E23	**E24**	**E25A (“EE”)**	**E25B (“EZ”)**
			
E26	**E27**	**E28**	**E29**

			
E30	E31	E32B	E33
			
E34B	E35	E36	E37A and E37B
			
E38	E39A	E39B	E40

			
E41	E42	E43	E44
			
E45	E46	E47 [CBB_28]	E48
			
E49 [CBB_21]	E50	E51	E52
			
E53A	E53B	E54	E55
			

2 Additional discussion of MOF structure generation

We note that 534 structures (out of 105,764) showed a structural collapse (certain atoms are unphysically close to each other) after the geometric optimization. The percentage is much lower than what was observed in our prior work[3], which focused on a tri-nodal net to construct MOFs. We discarded these structures prior to all data analysis and the list of these structures is available in the Zenodo repository.[1]

We also found that the geometry optimization of MOFs with certain organic linkers with phenylethynyl groups sometimes suffered from getting trapped in local energy minima with the UFF force field, and consequently the linkers had a bent shape (Figure S1). This issue happened consistently with gradient-based minimization algorithms (including CG and FIRE), but it could be resolved to some extent by using Newton's method or quasi-Newton algorithms. The bending is observed for the phenylethynyl based molecules shown in Table S4 even in the gas phase (no MOF) with all the gradient-based optimizations in three software packages: Open Babel, Materials Studio and LAMMPS. To reduce the bending in the long linkers in LAMMPS, we used the implemented Hessian-free truncated Newton algorithm as the minimizer for the atom positions for the MOFs with **tpt** topology and the long edge building blocks. Further analyzing two cases, **f5_ssc_46** and **f5_ssc_87**, we note that the angle at the connection points between the linker and the metal node could deviate too far from what is realistic, whereas the bond lengths and angles inside the building blocks are well modeled. To get around this issue, future work that improves the optimization algorithm may focus on 1) improving the building block placement algorithm and 2) performing geometric relaxation and/or molecular dynamics with building block groups constrained to reduce the degrees of freedom during the optimization.

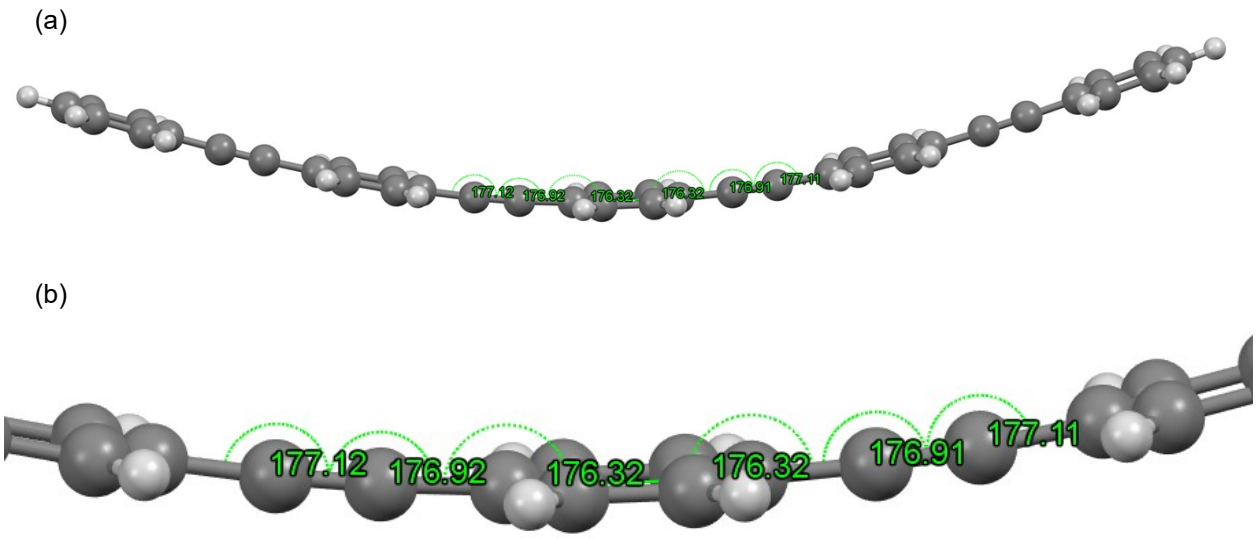
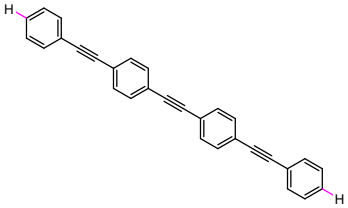
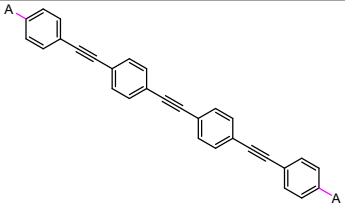
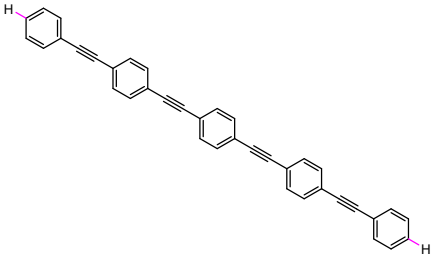
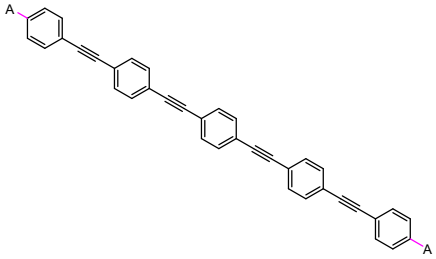
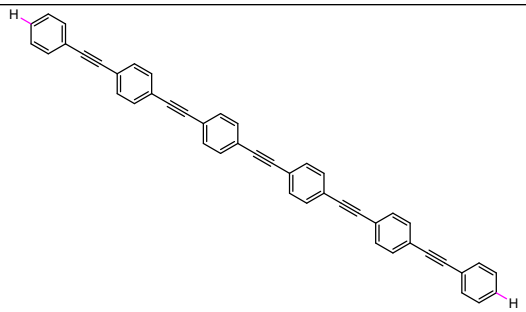
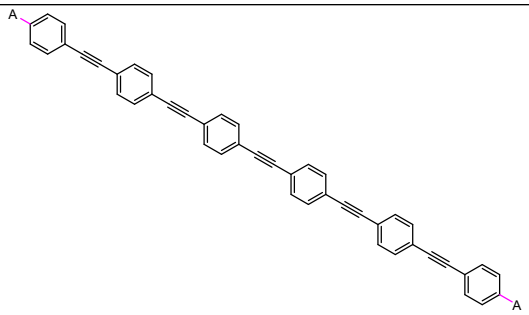


Figure S1. Example molecule that features phenylethynyl groups and shows bent angles (shown in green) between carbons (dark grey) as a result of being trapped in a local energy minimum.

Table S4. The derived molecules based on the long edge building blocks. An edge building block is a molecular fragment with connection points to connect with other building blocks.

Molecules	Edge building blocks
 1,1′-(1,2-ethynediyl)bis[4-(2-phenylethynyl)benzene]	
 1,4-bis[2-[4-(2-phenylethynyl)phenyl]ethynyl]benzene	

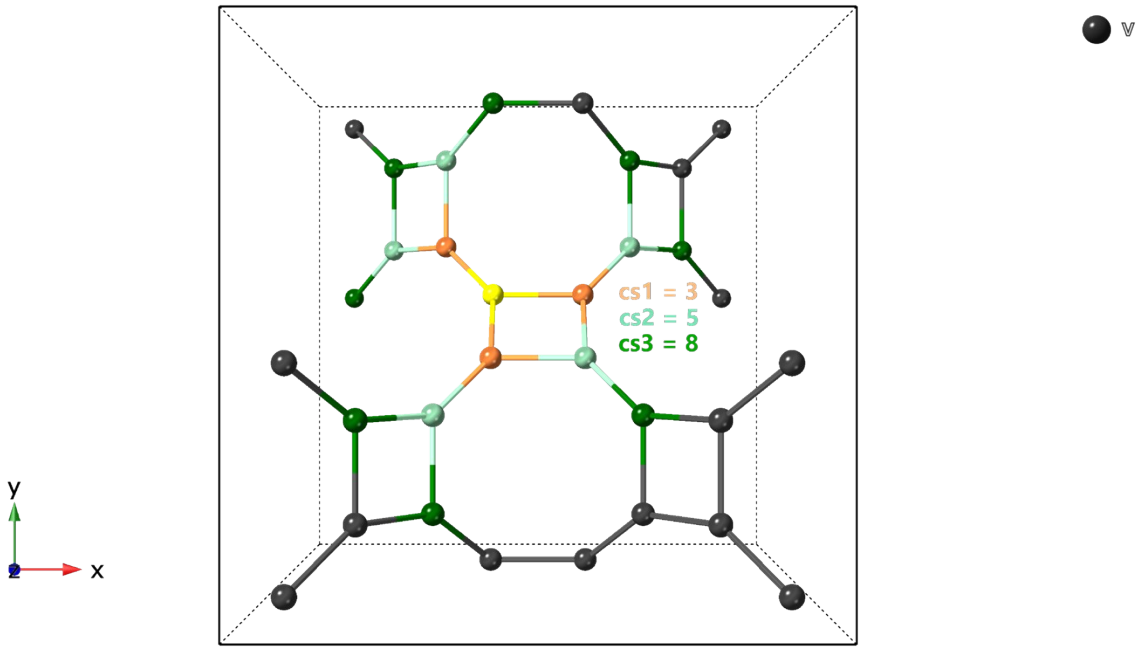
 1,1′-(1,2-ethynediyl)bis[4-[2-[4-(2-phenylethynyl)phenyl]ethynyl]benzene]	

3 Definitions and discussion of topological densities: net density and td10

O'Keeffe proposed the net density as a packing density descriptor that allows comparison of the density of sphere-packing and non-sphere-packing nets.[4] The net density is defined as $\rho = N/V$ where N is the number of circles or spheres with non-overlapping unit diameter in a volume V. In other words, the distance between closest in-contact circles or spheres is normalized to a distance unit of 1, and the unit cell volume is derived accordingly. The net density is closely related to the occupied fractional volume ϕ by a factor of $\pi/4$ for circles or $\pi/6$ for spheres.

Other mathematical descriptors often used for the vertex density of a graph include the coordination sequences and their cumulative sums (known as "topological densities").[5,6] The coordination sequence is a set of numbers (cs_1, cs_2, cs_3,…) that represent the number of vertices included in the neighbor shells and not counted yet (example shown in Figure S2) for any arbitrary vertex. The topological density td10 is defined as the cumulative tenfold sum of the coordination sequences, $td10 = \sum_{i=1}^{i=10} cs_i$. For nets with multiple kinds of vertices that have different coordination sequences, the td10 is computed as the weighted average of the vertices.

(a)



(b)

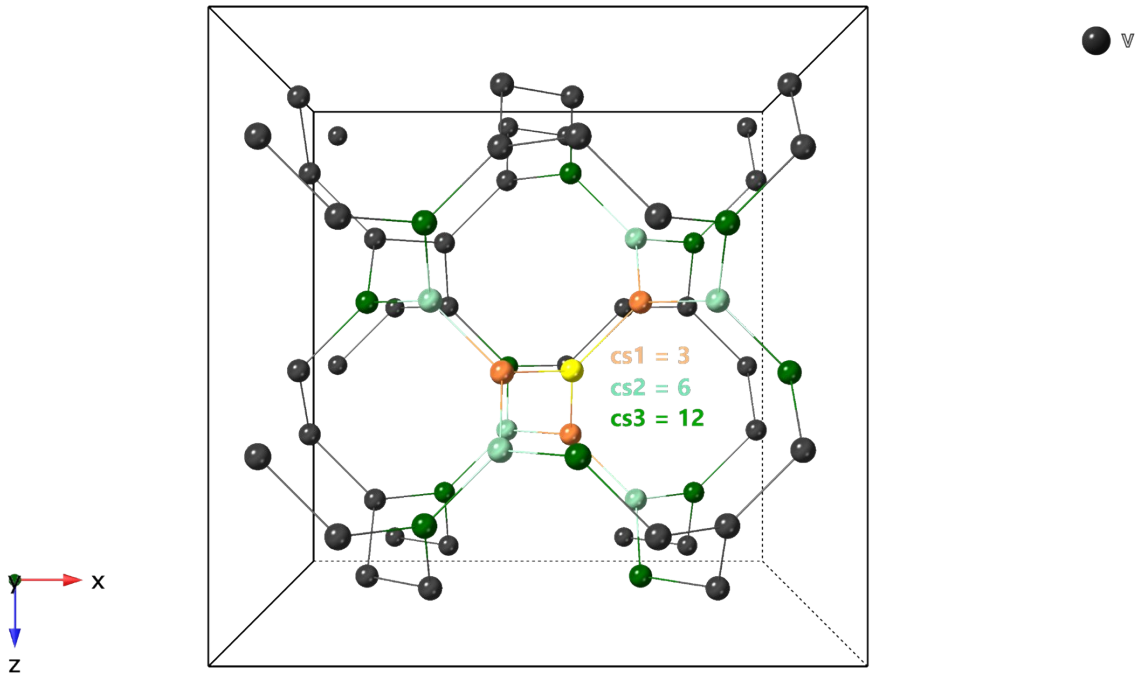


Figure S2. Illustration of coordination sequence (cs) in (a) **lvt-a** and (b) **srs** nets. The first shell around the yellow vertex is highlighted in orange (cs_1); the second shell is highlighted in light green (cs_2); the third shell is highlighted in dark green (cs_3).

4 Comments on the difference between td10 and net density

In the exploratory data analysis, both net density and td10 showed positive correlations with the upper limit of the VDC. While both higher td10 and net density suggest higher interconnectivity between vertices, the spatial cutoff is different. The net density describes the vertex density over a unit volume constructed with non-overlapping unit diameter spheres, whereas the td10 describes the neighboring number density over 10 coordination shells from a reference vertex, without considering the spatial information. This makes td10 an insufficient descriptor for describing the interconnectivity over a specific unit volume. For example, both **crs** and **bcs** nets have a higher td10 than **pcu**, and both of them have 16 vertices in their cubic unit cell, but with different symmetry. After normalizing the unit cell volume, **crs** has a lower dense packing than **pcu**, while **bcs** has a denser packing than **pcu** (Figure S3). The influences are clearly seen by comparing the resulting MOFs.

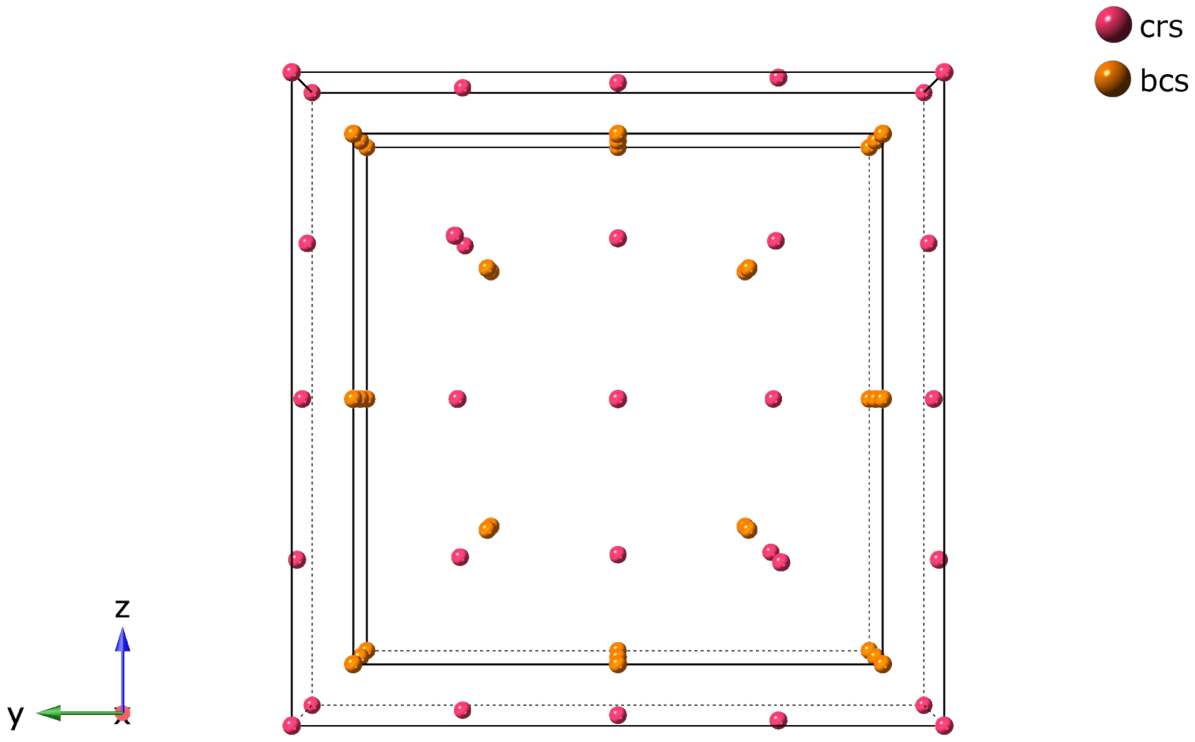


Figure S3. Comparison of unit cell and vertex density of **crs** and **bcs** nets. Red: **crs** net (lower net density), orange: **bcs** net (higher net density). The shortest vertex distance (edge length) in both nets is normalized to the same length, and therefore the unit cell sizes are different.

5 Details about ML model hyperparameters and performance

Table S5. LASSO model performance metrics for different energy histogram parameters tested. Coefficient of determination R^2, mean absolute error (MAE), root mean squared error (RMSE) are shown for both training set and testing set. Spearman's rank correlation coefficient ρ (ρ_s) and Kendall's τ (τ_k) are shown for testing set. The results show a decrease in predictive performance as the bin range is extended to -12 kJ/mol and an improvement when the bin width is refined to 0.5 kJ/mol.

Parameter set		Training set metrics			Testing set metrics				
Bin range [kJ/mol]	Bin width [kJ/mol]	R^2	MAE [g/L]	RMSE [g/L]	R^2	MAE [g/L]	RMSE [g/L]	ρ_s	τ_k
-10 to 0	1	0.926	1.388	1.832	0.916	1.457	1.992	0.940	0.803
-12 to 0	1	0.926	1.395	1.841	0.885	1.508	2.345	0.938	0.802
-10 to 0	0.5	0.952	1.057	1.468	0.939	1.126	1.700	0.961	0.845

Table S6. The coefficients of the best LASSO model parameterized using bin ranges between -10 and 0 kJ/mol and 0.5 kJ/mol bin width.

Bin range [kJ/mol]	Coefficient
(Intercept)	32.505
-Inf to -10	-498.36
-10 to -9.5	-49.452
-9.5 to -9	-3213.1
-9 to -8.5	0
-8.5 to -8	0
-8 to -7.5	0
-7.5 to -7	0
-7 to -6.5	0
-6.5 to -6	183.09
-6 to -5.5	195.35
-5.5 to -5	93.86
-5 to -4.5	141.64
-4.5 to -4	110.43
-4 to -3.5	121.87
-3.5 to -3	90.265

-3 to -2.5	87.052
-2.5 to -2	68.023
-2 to -1.5	73.198
-1.5 to -1	13.873
-1 to -0.5	53.871
-0.5 to 0	0
0 to Inf	-21.693

Table S7. Random forest model performance as a function of training size. Model performance metrics are shown for training set predictions, unbiased out-of-bag predictions, and testing set predictions.

	Training set			Out of bag			Testing set		
Training size	RMSE [g/L]	R^2	MAE [g/L]	RMSE [g/L]	R^2	MAE [g/L]	RMSE [g/L]	R^2	MAE [g/L]
1251	0.609	0.993	0.420	1.488	0.956	1.035	1.492	0.956	1.028
2502	0.549	0.994	0.374	1.350	0.964	0.925	1.407	0.960	0.959
3753	0.534	0.994	0.361	1.302	0.965	0.892	1.340	0.964	0.898
5004	0.516	0.995	0.347	1.266	0.967	0.858	1.286	0.967	0.865
6255	0.497	0.995	0.334	1.219	0.970	0.825	1.273	0.967	0.849
7503	0.495	0.995	0.331	1.216	0.969	0.819	1.244	0.969	0.830
8754	0.488	0.995	0.326	1.199	0.971	0.806	1.246	0.969	0.825
10005	0.481	0.995	0.321	1.183	0.971	0.796	1.231	0.969	0.812
11256	0.478	0.995	0.317	1.177	0.971	0.785	1.221	0.970	0.807
12507	0.475	0.996	0.314	1.165	0.972	0.777	1.212	0.970	0.802
13569	0.473	0.995	0.312	1.165	0.972	0.774	1.206	0.971	0.794

Table S8. RF hyperparameter tuning results. Training size: the number of data points included for training. mtry: number of variables used for each node. RMSE: root mean square error. Rsquared: the coefficient of determination. MAE: mean absolute error. SD: standard deviation (calculated based on the 5-fold cross validation).

Training size	mtry	RMSE	Rsquared	MAE	RMSE SD	Rsquared SD	MAE SD
1251	2	2.0278	0.9203	1.4769	0.0907	0.0295	0.0684
	12	1.5117	0.9503	1.0663	0.1055	0.0202	0.0693
	22	1.5307	0.9486	1.0747	0.1130	0.0199	0.0779
2502	2	1.7635	0.9435	1.2775	0.1037	0.0087	0.0605
	12	1.3818	0.9629	0.9538	0.1089	0.0062	0.0602
	22	1.3924	0.9621	0.9565	0.1249	0.0067	0.0630
3753	2	1.7048	0.9451	1.2376	0.0695	0.0065	0.0460
	12	1.3364	0.9637	0.9144	0.0708	0.0047	0.0371

	22	1.3610	0.9622	0.9273	0.0838	0.0044	0.0473
5004	2	1.6320	0.9485	1.1728	0.0525	0.0079	0.0381
	12	1.2809	0.9658	0.8733	0.0779	0.0063	0.0396
	22	1.3187	0.9635	0.8869	0.0771	0.0065	0.0364
6255	2	1.5889	0.9517	1.1394	0.0705	0.0040	0.0475
	12	1.2395	0.9687	0.8417	0.0437	0.0022	0.0222
	22	1.2597	0.9675	0.8498	0.0439	0.0020	0.0215
7503	2	1.5460	0.9537	1.1135	0.0557	0.0055	0.0324
	12	1.2209	0.9690	0.8312	0.0562	0.0036	0.0244
	22	1.2394	0.9679	0.8373	0.0614	0.0036	0.0227
8754	2	1.5195	0.9555	1.0961	0.0491	0.0047	0.0183
	12	1.2162	0.9695	0.8209	0.0497	0.0045	0.0202
	22	1.2423	0.9680	0.8322	0.0486	0.0048	0.0219
10005	2	1.4998	0.9567	1.0786	0.0287	0.0016	0.0269
	12	1.1997	0.9705	0.8090	0.0396	0.0022	0.0255
	22	1.2289	0.9689	0.8219	0.0458	0.0025	0.0289
11256	2	1.4748	0.9578	1.0623	0.0315	0.0035	0.0297
	12	1.1955	0.9705	0.8039	0.0371	0.0027	0.0200
	22	1.2215	0.9691	0.8120	0.0403	0.0029	0.0181
12507	2	1.4586	0.9592	1.0493	0.0396	0.0021	0.0211
	12	1.1792	0.9716	0.7906	0.0133	0.0007	0.0120
	22	1.2055	0.9702	0.8019	0.0184	0.0011	0.0128
13569	2	1.4459	0.9592	1.0391	0.0237	0.0036	0.0069
	12	1.1785	0.9711	0.7900	0.0285	0.0029	0.0111
	22	1.2068	0.9696	0.8015	0.0309	0.0031	0.0153

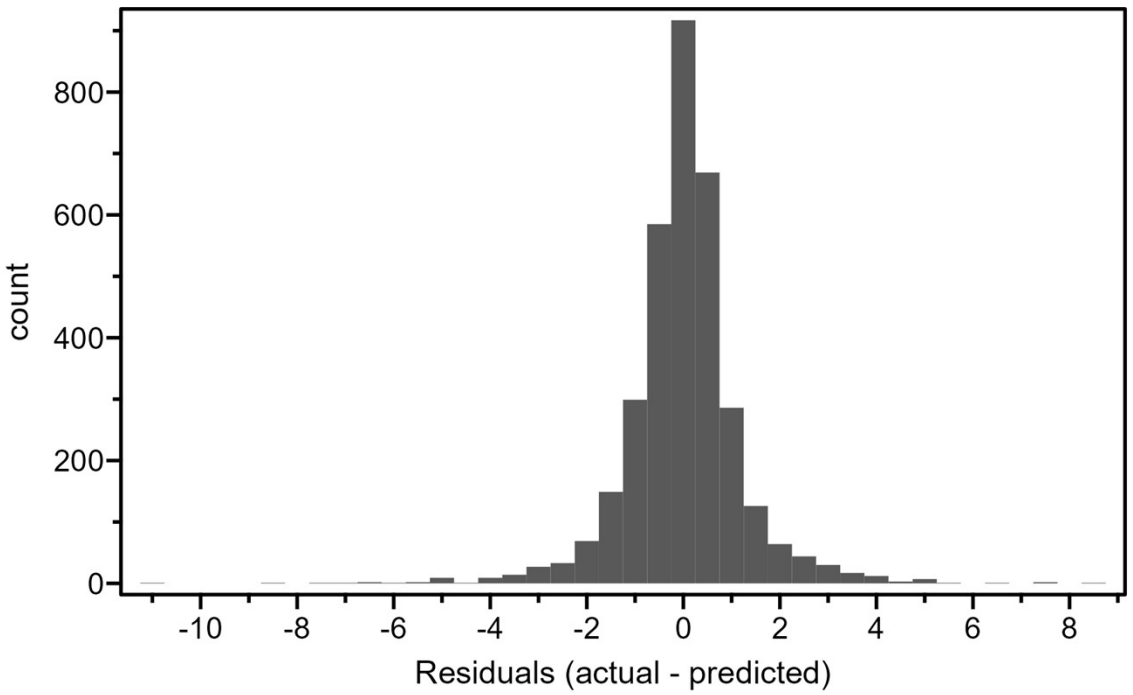


Figure S4. Histogram of prediction residuals of the testing set for the final RF model. The prediction errors for most MOFs in the testing set are within 2 g/L.

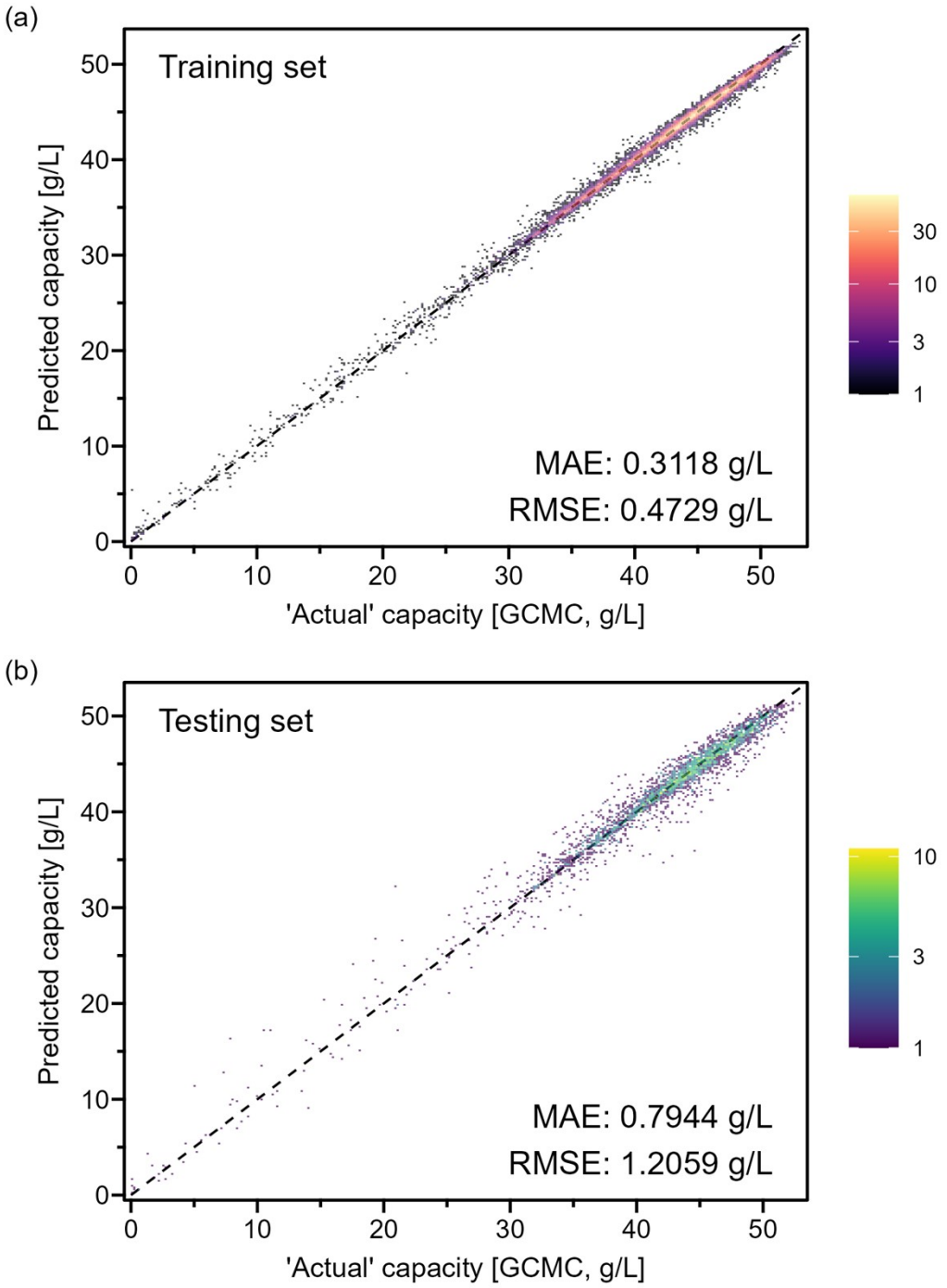


Figure S5. Parity plots of hydrogen deliverable capacity for the full dataset from the random forest (RF) model versus results from GCMC simulation: (a) training set results, (b) testing set results. Predictions are shown for the best RF model after hyperparameter tuning. R^2 is 0.995 for the training set, and 0.971 for the testing set.

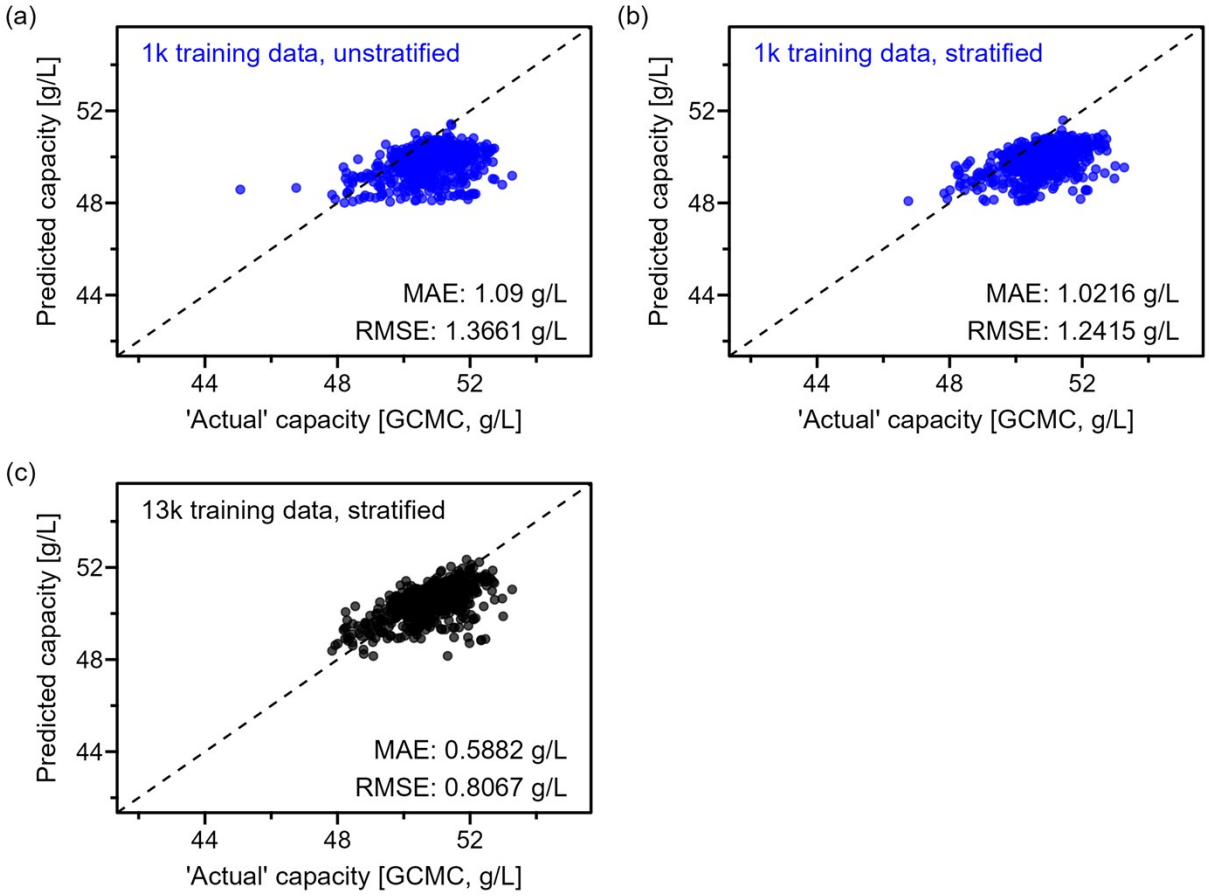


Figure S6. Prediction results for top MOFs (predicted capacity > 48g/L) are improved by sample stratification and increased sample size. The MOFs plotted were selected based on the predictions from the RF model trained with (a) approximately 1,000 data points without stratification, (b) approximately 1,000 data points with stratification, (c) approximately 13,000 data points with stratification. The GCMC simulations were subsequently performed to obtain the capacity of these MOFs.

6 Case studies of MOFs and nets

Table S9. Structural properties of the two example structures in Figure 7 with "vacant" topologies.

Structure id	LCD [Å]	PLD [Å]	VSA [m^2/cc]	GSA [m^2/g]	PV [cm^3/g]
fl_srsa_6_1x1x1	72.98	70.18	195.4	6370	32.00
fl_diaf_5_1x1x1	69.33	62.85	227.0	7280	31.49

7 Topologies of the best performing MOFs

Table S10. Nets that result in MOFs with simulated VDC > 52.0 g/L.

Net		**Source**
bcs	6-c	Synthesized MOF / coordination polymer
ctn	3,4-c	
pyr	3,6-c	
tsx	3,6-c	
rob	6-c	Natural / synthesized minerals
ibd	4,6-c	Theoretical MOF topologies
ssc	4,4-c	
czz	3,6-c	No references
esg	3,6-c	
icd	4,4-c	

8 Effect of different optimization algorithms on top performing MOFs

To understand how different optimization methods may affect the MOF porosity and subsequently the predicted storage performance, we studied two example MOFs from our top performing candidates: **f1_pyr_142** and **f1_tsx_103** by comparing their textural properties and simulated hydrogen isotherms. For **f1_pyr_142**, we compared the different versions of the structure from 1) not optimized (straight from the structural assembly), 2) optimized with LAMMPS using the CG/FIRE algorithm (default algorithm used for all structures), 3) optimized with LAMMPS using only CG algorithm and 4) optimized with extended tight binding (xTB) method. For **f1_tsx_103**, we compared the structure versions optimized with 1) LAMMPS, 2) xTB and 3) DFT.

Table S11 presents the textural properties of different versions of the two MOFs, with relative differences compared to the default MOF structure (LAMMPS with CG/FIRE). We found that compared to the default MOF structure, the unoptimized structure of **f1_pyr_142** has a much larger pore size, with ~40% larger PLD and LCD, and 50% larger PV. In addition, we also observed ~10% differences in LCD, surface area, and PV between other optimized structures and the default **f1_pyr_142**. For **f1_tsx_103**, the changes in the textural properties across all three different optimized structures are less than 5%.

Table S11. Textural properties of different crystal structure versions of **f1_pyr_142** and **f1_tsx_103**. The percentages in brackets denote the relative difference compared to the default structure optimized in LAMMPS (in bold).

Cif name	Structure version	LCD [Å]	PLD [Å]	Crystal density [g/cc]	Volume [Å3]	VSA [m2/cc]	GSA [m2/g]	VF	PV [cc/g]
f1_pyr_142	not optimized	13.36 [+38%]	10.46 [+41%]	0.36 [-29%]	25678 [+40%]	1965 [-26%]	5493 [+3%]	0.82 [+10%]	2.28 [+54%]
	LAMMPS (CG/FIRE)	**9.67**	**7.42**	**0.50**	**18343**	**2669**	**5331**	**0.74**	**1.48**
	LAMMPS (CG only)	11.47 [+19%]	7.42 [-0%]	0.46 [-8%]	20026 [+9%]	2460 [-8%]	5364 [+1%]	0.76 [+3%]	1.66 [+12%]
	xTB	9.24	7.00	0.54	17093	2614	4865	0.72	1.35

		[-4%]	[-6%]	[+7%]	[-7%]	[-2%]	[-9%]	[-2%]	[-9%]
f1_tsx_103	**LAMMPS**	**11.45**	**7.85**	**0.43**	**12214**	**2285**	**5328**	**0.76**	**1.76**
	xTB	11.78 [+3%]	7.78 [-1%]	0.43 [-0%]	12264 [+0.4%]	2302 [+1%]	5389 [+1%]	0.76 [-0%]	1.77 [+0.6%]
	DFT	11.73 [+2%]	7.80 [-1%]	0.42 [-3%]	12602 [+3%]	2284 [-0%]	5495 [+3%]	0.76 [+1%]	1.83 [+4%]

We observe slight differences in the amount of H_2 adsorbed under low and medium pressure comparing the unoptimized **f1_pyr_142** and the optimized counterparts (Figure S7), but the differences in volumetric uptakes at 100 bar between optimized structures are small. As a result of changes in crystal density, the gravimetric uptakes at 100 bar between optimized structures are significantly different. The differences in PV are correlated with the differences in predicted gravimetric hydrogen uptakes in the medium to high pressure region under both 77 K and 160 K conditions. This observation is in agreement with previous studies.[3,7] For **f1_tsx_103**, the simulated H_2 uptakes are in good agreement across all three versions of the structures (Figure S8).

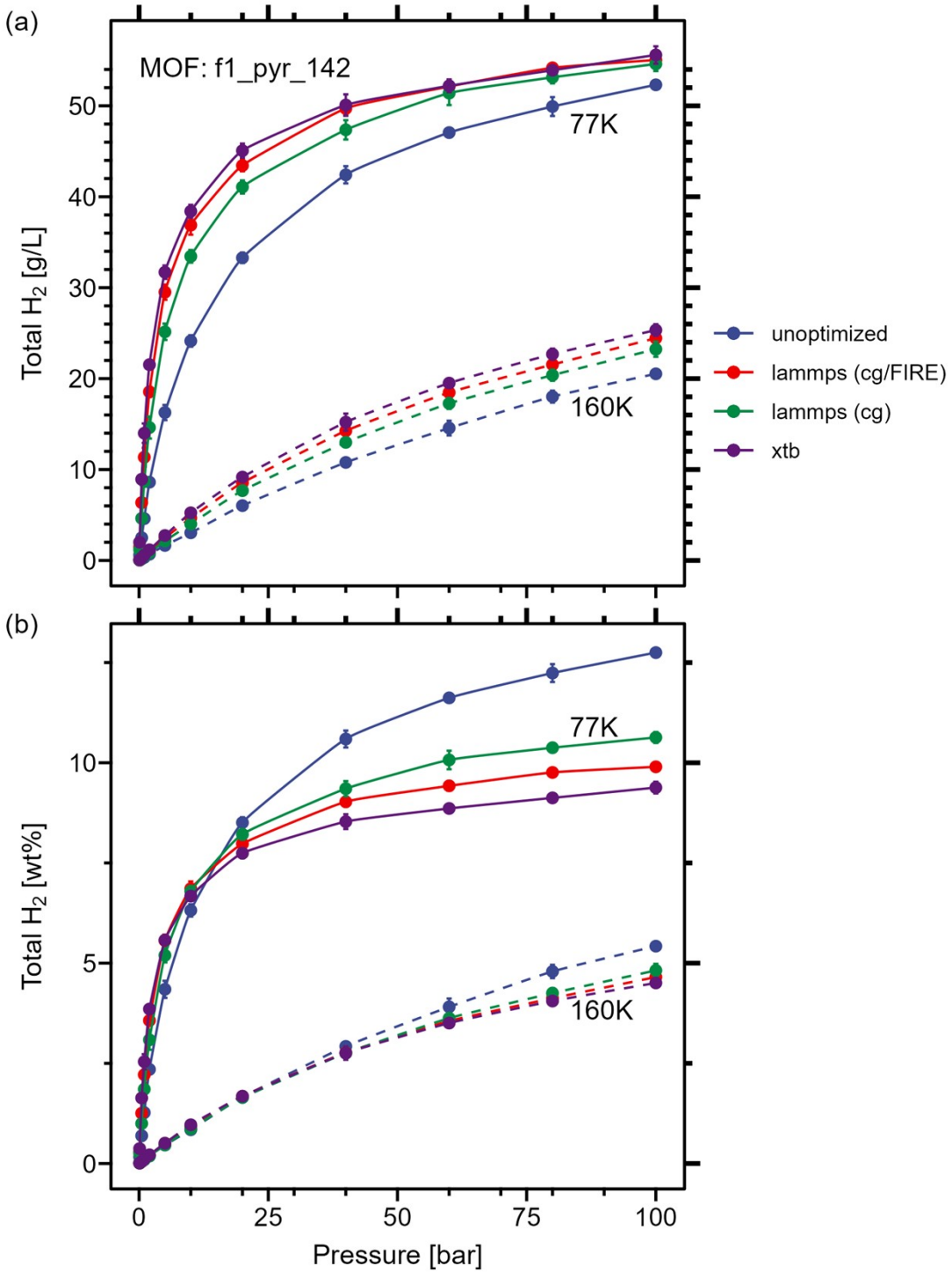


Figure S7. Simulated H_2 isotherms of different versions of **f1_pyr_142** at 77 K and 160 K, with the absolute uptake in (a) volumetric units (g/L) and (b) gravimetric units (wt%). Blue: the unoptimized structure, red: optimized using LAMMPS with CG/FIRE algorithm (default), green: optimized using LAMMPS with CG algorithm, purple: optimized using xTB method.

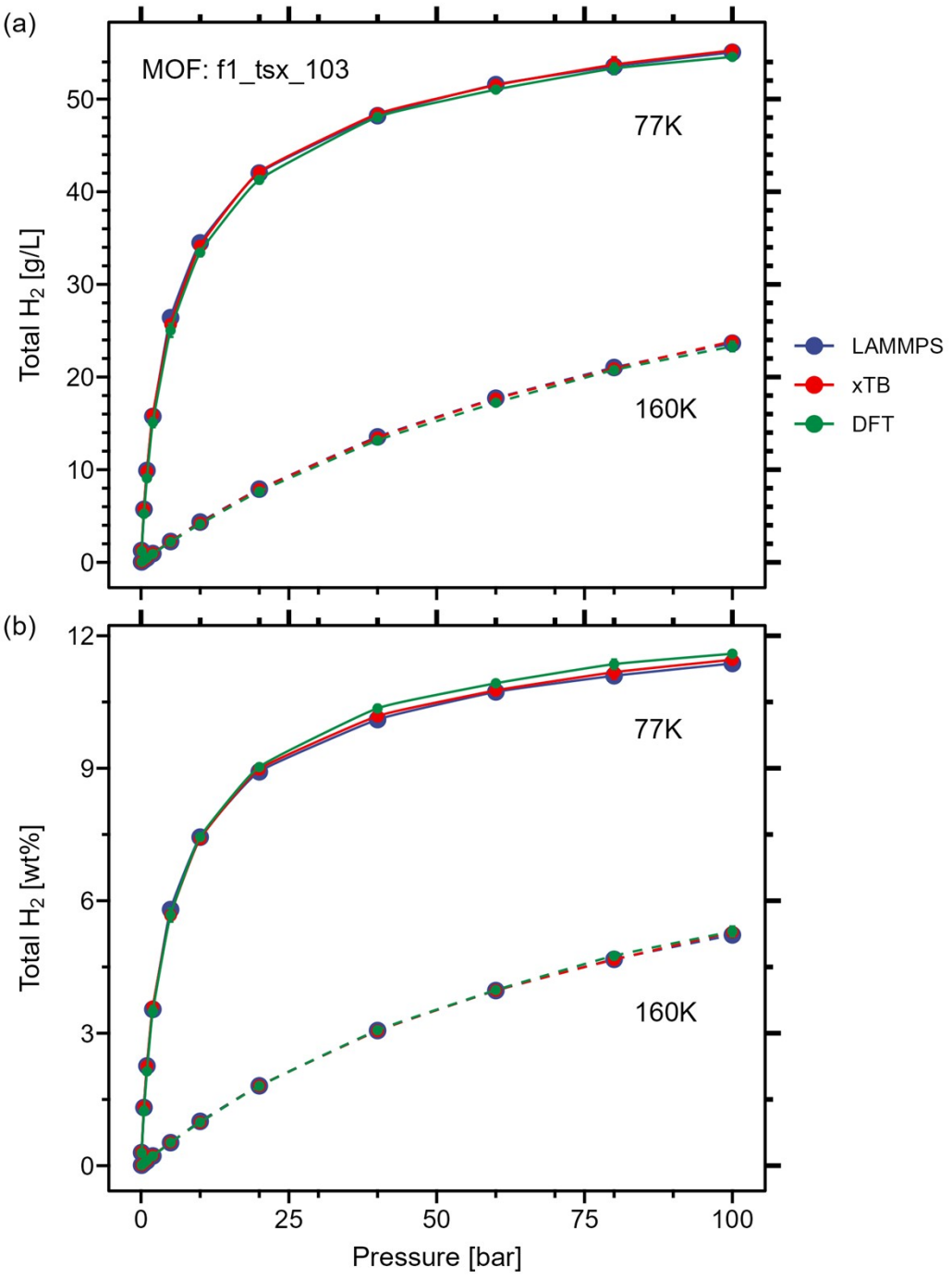


Figure S8. Simulated H_2 isotherms of different versions of **f1_tsx_103** at 77 K and 160 K, with the absolute uptake in (a) volumetric units (g/L) and (b) gravimetric units (wt%). Blue: optimized using LAMMPS with CG/FIRE algorithm (default), red: optimized using xTB method, green: optimized with DFT method.

We also compared the deliverable capacity of these structures. To our surprise, the VDC values of different versions of **f1_pyr_142** differ only minimally (Table S12); the difference between the deliverable capacity of the unoptimized structure and the optimized structure is less than 3%. Nevertheless, optimization methods lead to differences in crystal density and PV, which lead to significant difference in GDC. When overlapping the optimized structures as shown in Figure S9, we indeed observe a smaller unit cell and slightly shorter bonds overall for the xTB-optimized structure (Figure S10) compared to the default structure (Figure S9a). On the other hand, there is some linker rotation in the structure optimized with LAMMPS-CG compared to the default structure (Figure S9b), suggesting that the geometric optimization may have ended up in another local minima.

Table S12. Hydrogen absolute uptake at 77 K and 160 K and deliverable capacity in different units for all different crystal structures of **f1_pyr_142** and **f1_tsx_103**. The relative difference compared to the default structure (in bold) is shown in the brackets.

Cif name	Structure version	Uptake at 77 K/100 bar [g/L]	Uptake at 160 K/5 bar [g/L]	Deliverable [g/L]	Deliverable [mg/g]	Deliverable [wt%]
f1_pyr_142	not optimized	52.76	1.66	51.10 [-3%]	142.75 [+36%]	12.49 [+31%]
	LAMMPS (CG/FIRE)	**55.20**	**2.52**	**52.68**	**105.15**	**9.51**
	LAMMPS (CG only)	54.45	2.15	52.31 [-1%]	113.96 [+8%]	10.23 [+8%]
	xTB	55.25	2.74	52.51 [-0.2%]	97.79 [-7%]	8.91 [-6%]
f1_tsx_103	**LAMMPS**	**54.97**	**2.25**	**52.73**	**122.90**	**10.95**
	xTB	55.16	2.26	52.89 [+0.3%]	123.88 [+1%]	11.02 [+1%]
	DFT	54.55	2.19	52.36 [-1%]	125.85 [+2%]	11.18 [+2%]

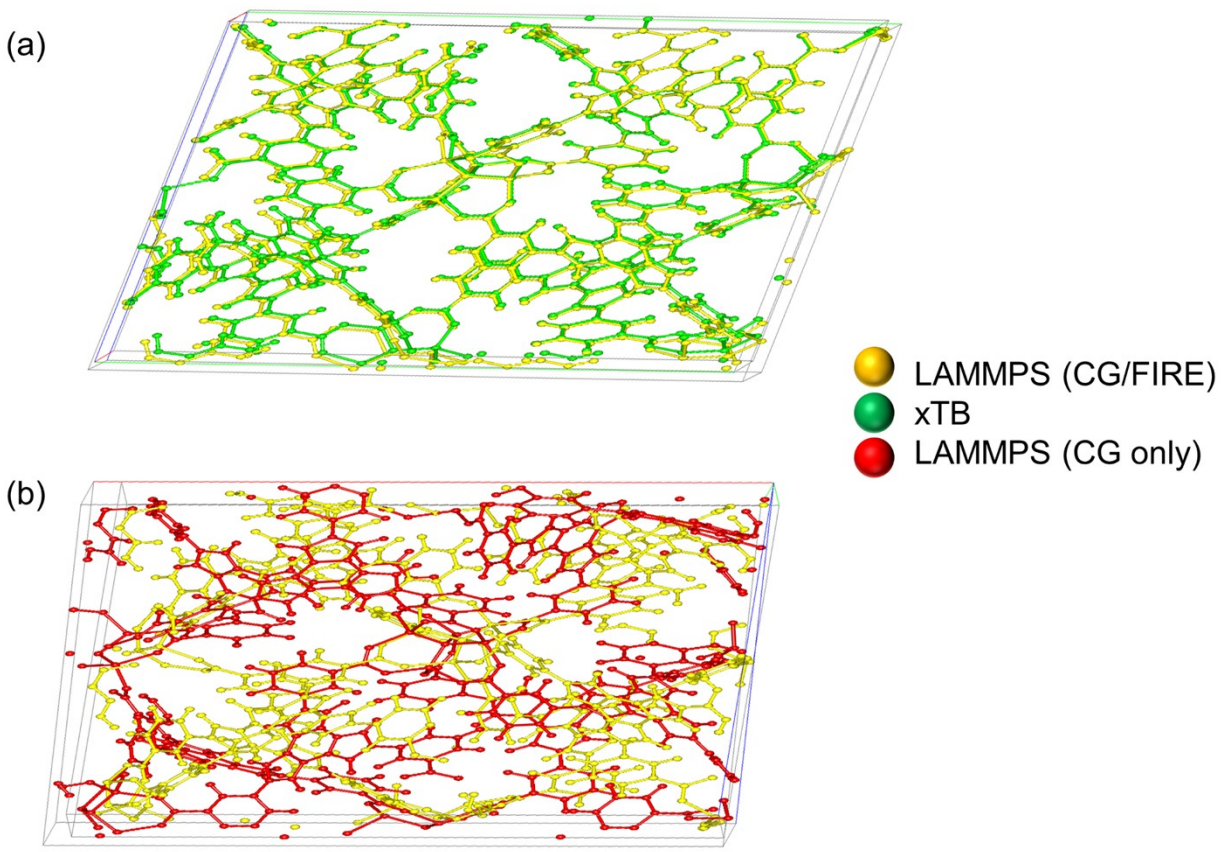


Figure S9. Overlapping crystal structures of **f1_pyr_142** optimized with (a) LAMMPS (CG/FIRE) algorithm versus xTB method and (b) LAMMPS(CG/FIRE) algorithm versus LAMMPS (CG only) algorithm. The visualization was generated in the CrystalCMP program.[8]

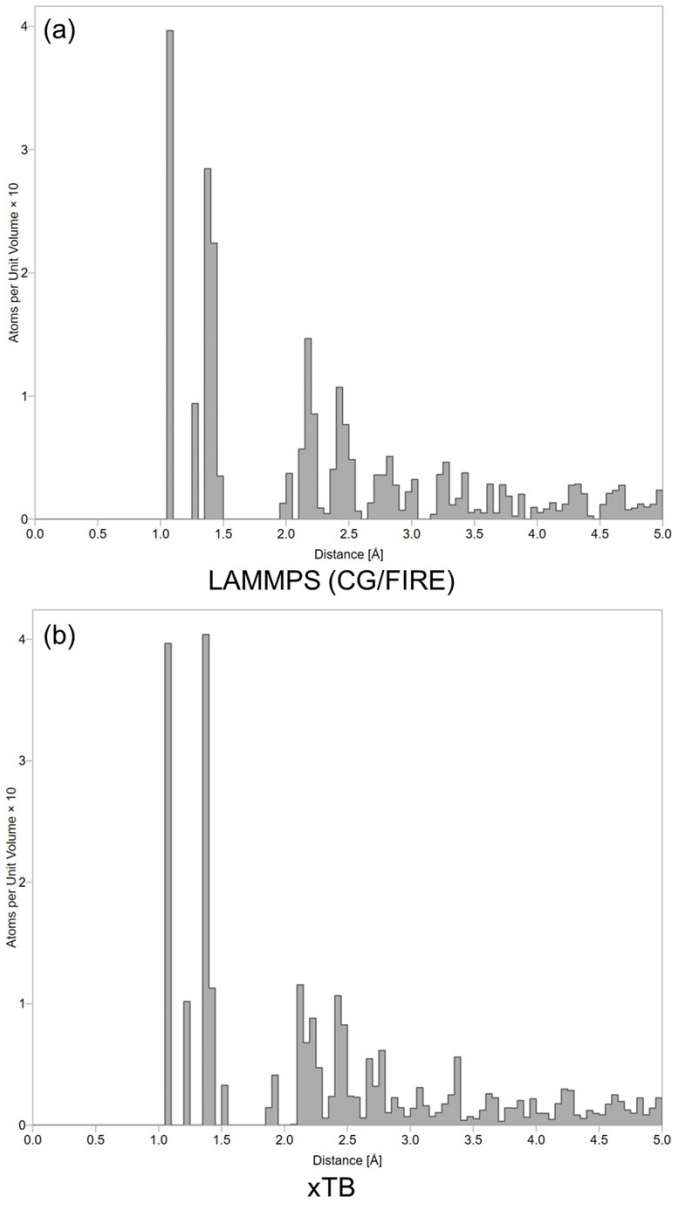


Figure S10. Pair distribution function g(r) between all atoms in the crystal structures optimized with (a) LAMMPS (CG/FIRE) and (b) xTB. The most notable change in pairwise distance is around 2.0 Å.

It is known that structural optimization is important to make structures realistic. However, our case studies suggest that different optimization methods can lead to up to 10% difference in the textural properties, where VF is the least sensitive textural property to the structural changes. The pore shape may become different due to linker rotations, in agreement with other work. Although differences in volumetric uptakes are observed in simulated isotherms particularly in the low and medium pressure region, the VDC is not very sensitive to the structural differences (Table S12). In contrast, gravimetric

uptakes and GDC can be impacted by the significant changes in PV and crystal density due to changes in force field and relaxation algorithms.

9 Investigation of very strong adsorption sites

We noticed that the energy grids for certain structures in the dataset exhibit energetic interactions stronger than -9.5 kJ/mol, and some even exceeding -10 kJ/mol. These values are higher than the typical binding energies associated with hydrogen physisorption, in the range of around 4-6 kJ/mol for MOFs[7] and graphene.[9] Among the 105,764 structures analyzed, 1032 structures contain sites with binding energies between -10 and -9.5 kJ/mol and 675 structures contain energy sites stronger than -10 kJ/mol. We analyzed two such MOFs, **f3_sxh_9** (Figure S11) and **f3_unw_9** (Figure S12) in detail. They are constructed with the same ditopic organic linker but different metal clusters: a 6-coordinated trinuclear Cr cluster and a 4-coordinated copper paddlewheel, respectively, as detailed in Table S13.

Notably, both structures contain strong adsorption sites encircled by phenyl groups from the side of the edge building blocks, forming an interesting bowl-shaped geometry. In the case of **f3_unw_9**, the strongest binding site (-9.81 kJ/mol) is surrounded by 3 equidistant phenyl rings (Figure S14) where the closest carbon atoms are more than 3.5 Å away. We also note that the heat of adsorption differs from the magnitude of the strongest energy sites, since the heat of adsorption, q, is related to the average energy of adsorption and at finite temperature molecules sample many sites beyond the strongest binding sites.

In both MOFs mentioned, the number of strong binding sites is quite low. In **f3_sxh_9**, only 5 locations out of 64,000 grid points exhibit binding energies between [-10,9.5) kJ/mol; in **f3_unw_9**, merely 1 grid in the unit cell was found in this geometric configuration, indicating that the bowl-shape geometry is not consistently maintained given the rotational degree of freedom of the linkers. Consequently, the hydrogen storage performances for both structures are modest, with predicted capacities of 38.5 g/L and 43 g/L, respectively. Their heats of adsorption at zero loading at 77 K are 8.6 kJ/mol and 7.9 kJ/mol, respectively.

For additional context, we also plotted a distribution overview of properties for the structures with strong binding sites (Figure S15). The edge building block E13, as used in **f3_sxh_9** and **f3_unw_9**,

yields the highest number of MOFs with strong binding sites compared to other edge building blocks (Figure S16). This observation may guide future research to develop and design edge side groups that form specific binding pockets, potentially enhancing gas storage capabilities.

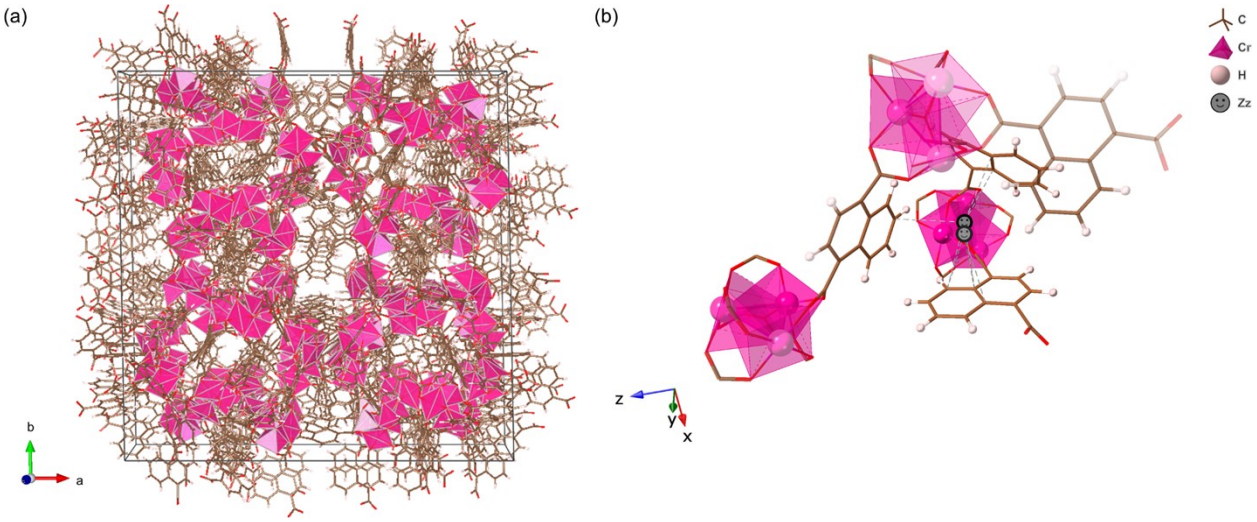


Figure S11. The structure f3_sxh_9 shown with (a) the crystal structure and (b) the strong binding sites (smiley faces) and their local environment in perspective view.

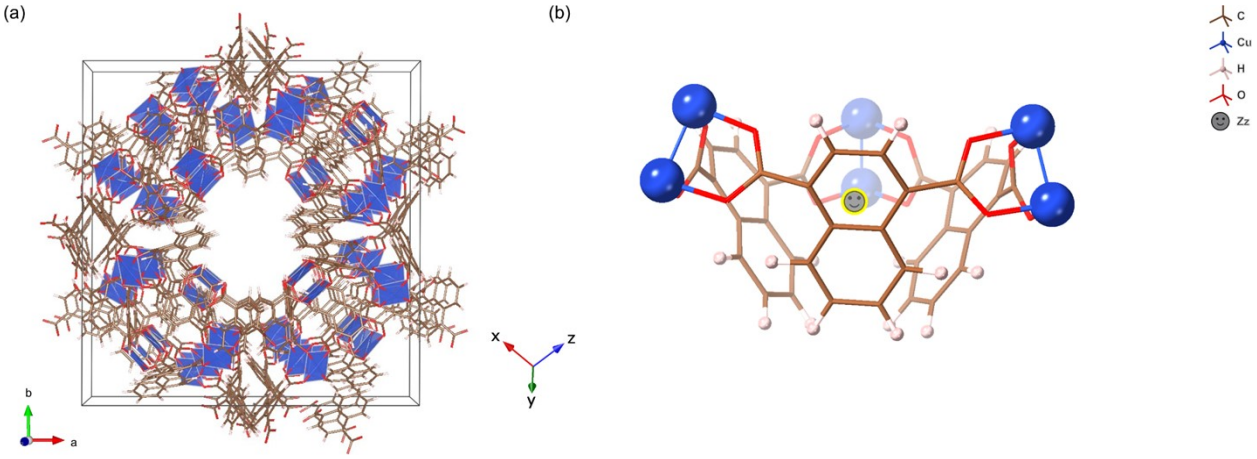


Figure S12. The structure f3_unw_9 shown with (a) the crystal structure and (b) the strong binding sites (smiley faces) and their local environment in perspective view.

Table S13. Composition of the two MOFs investigated for strong binding sites. The column “V1 coordination” denotes the coordination number of the vertex.

Cif name	Topology	Topology type	V1 coordination	Vertices	Edges
f3_sxh_9	sxh	single vertex	6	v1-mc5	E13
f3_unw_9	unw	single vertex	4	v1-mc3	E13

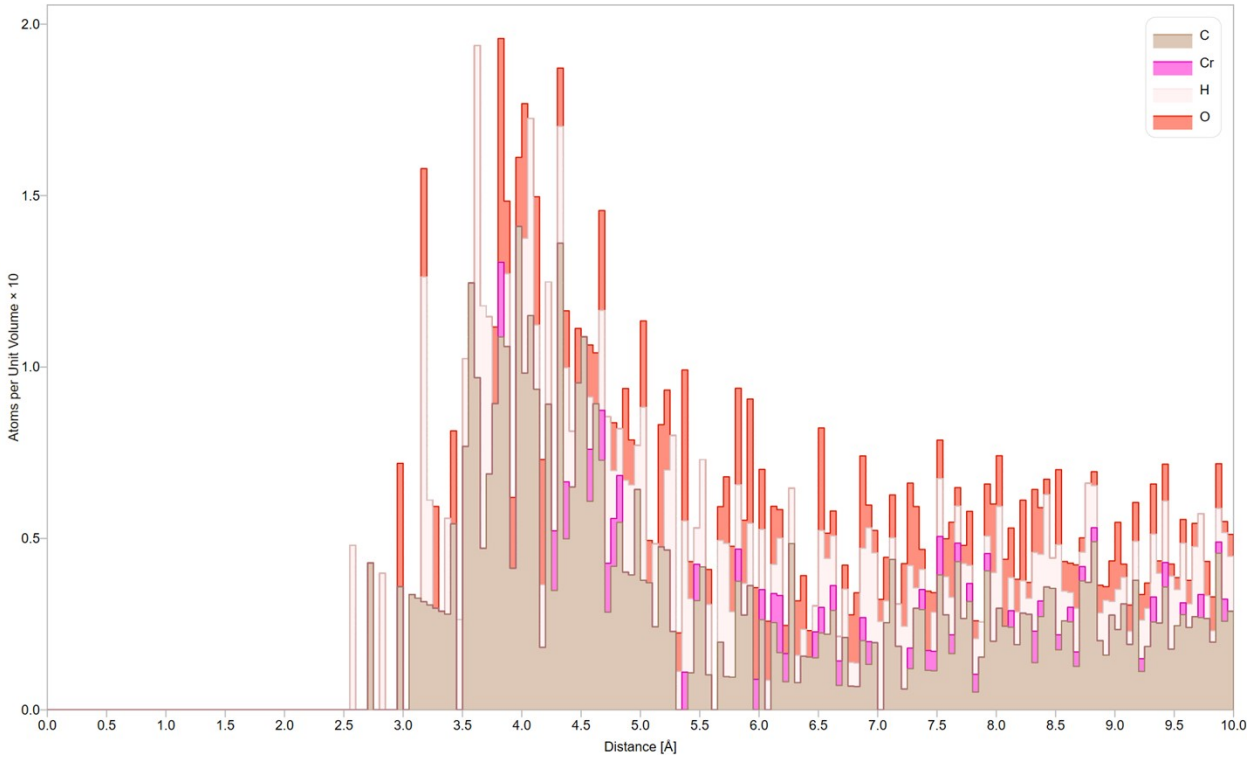


Figure S13. Pair distribution function g(r) for the energetic site (smiley face) to the neighboring atoms in **f3_sxh_9**.

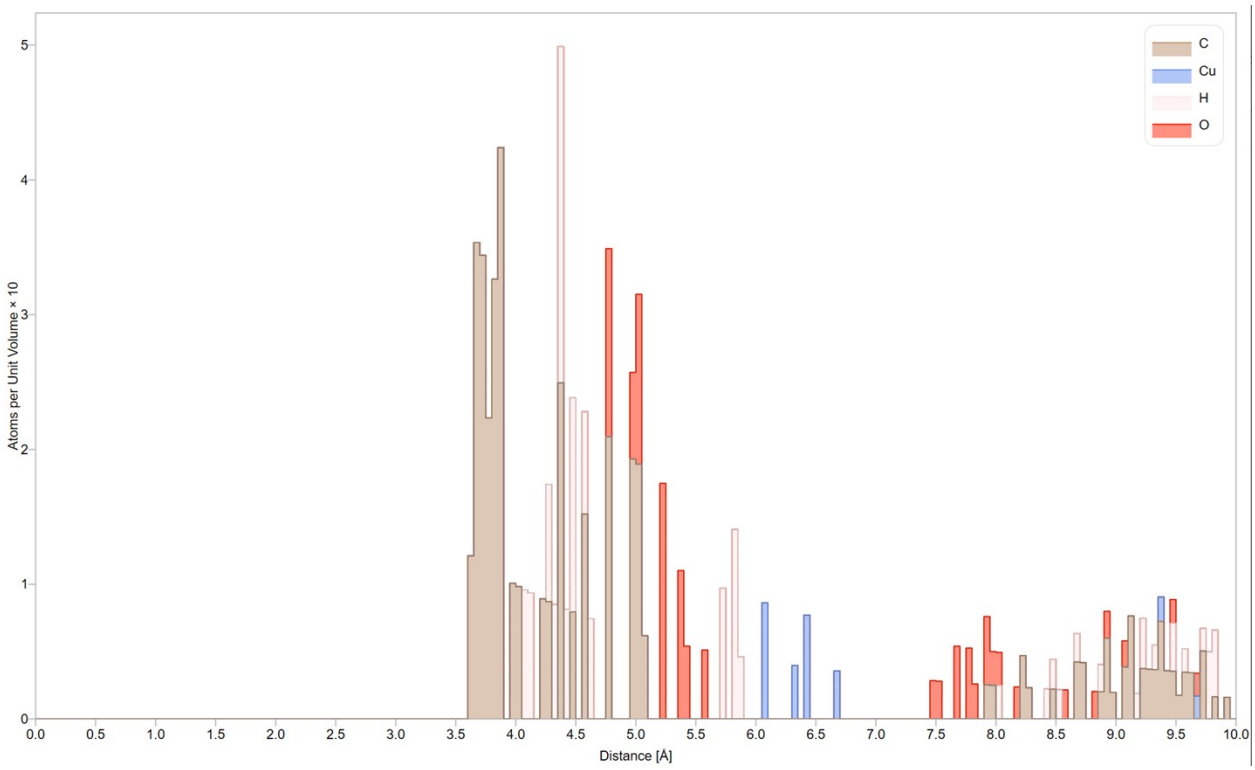


Figure S14. Pair distribution function g(r) between the energetic site (smiley face) to the neighboring atoms in **f3_unw_9**.

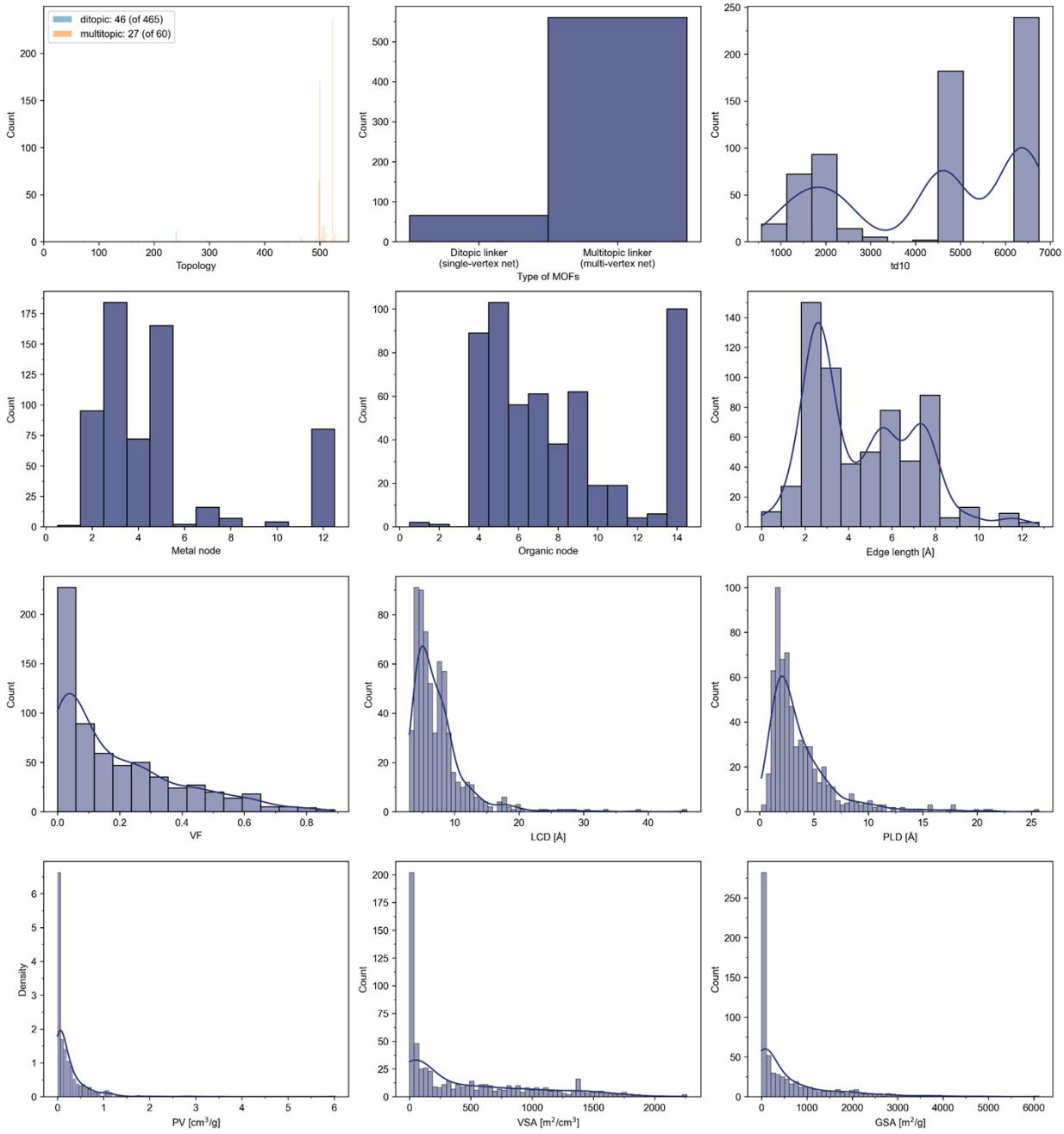


Figure S15. The distribution overview of the MOF subset having strong binding sites with lower than -10 kJ/mol binding energy with hydrogen.

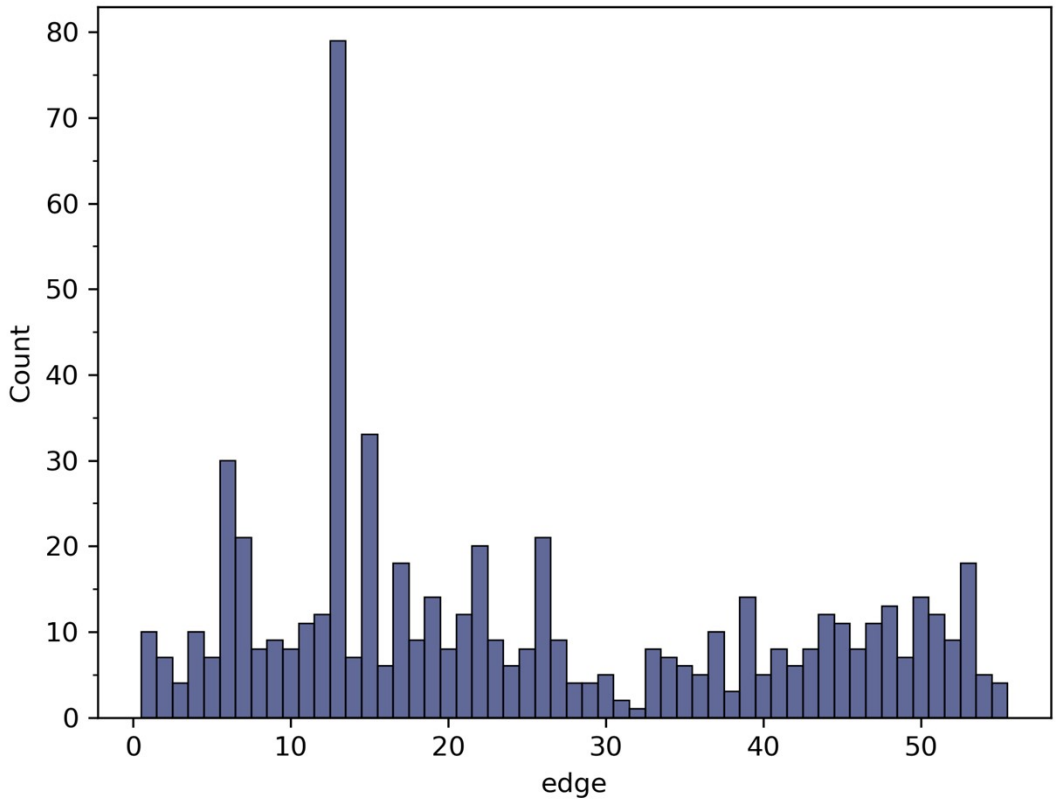


Figure S16. Histogram of edge composition of the MOF subset having strong binding sites with lower than -10 kJ/mol binding energy with hydrogen. The x axis shows the edge id.

10 Unsupervised learning of underlying structure–property relationships.

To better recognize the underlying patterns in high dimensional data, we used dimension reduction techniques for data visualization. One of the modern dimension reduction techniques is Uniform Manifold Approximation and Projection (UMAP).[10] It works by identifying a lower dimensional structure that approximates the manifold structure of the data points. Compared to other techniques such as t-SNE, the major advantage and why it has been increasingly popular is its ability to effectively preserve both local and global structures. It is also faster and more scalable.

We investigated the relationship between the underlying nets and hydrogen deliverable capacity using UMAP. Instead of using net names as labels, we employed a set of numerical net descriptors available

on the RCSR database[11] and listed in Table S14 to describe the nets. For the full dataset of MOF structures, we conducted UMAP reduction using only the net descriptors, and then with both net descriptors and porosity descriptors. Different normalization and scaling procedures were applied to the descriptors as appropriate to their mathematical nature (Table S14).

Table S14. Normalization and scaling of the net descriptors and the porosity descriptors for UMAP analysis. Normalization is not conducted on net descriptors where the numerical distance has meaning. Max absolute scaling is conducted for those which have physical meaning at zero.

	Normalization	**Scaling**
Net descriptors:		
Net type	/	NA (binary)
Coordination of V1	/	Min-max
Coordination of V2	/	Min-max
Net density	Power transform (Yeo-Johnson)	Max absolute
td10	Power transform (Yeo-Johnson)	Max absolute
Genus	/	Min-max
D-size	/	Min-max
Average vertex order	/	Max absolute
Smallest ring size	/	Min-max
Porosity descriptors:		
Pore volume	Power transform (Yeo-Johnson)	Max absolute
Density	Power transform (Yeo-Johnson)	Max absolute
VSA	Power transform (Yeo-Johnson)	Max absolute
GSA	Power transform (Yeo-Johnson)	Max absolute

We first explored how *nets* were related based on the net descriptors using UMAP. The net data points were colored with statistical measures computed from the MOF data samples of each net. The UMAP analysis on the 529 nets, colored according to the average predicted VDC (Figure S17) and maximum predicted VDC (Figure S18), showed a distinguishable cluster of nets that are mathematically similar

(Table S15), and this subset consists mostly of theoretically derived **6-c** nets. Additionally, we noticed clusters of nets with darker color (lower upper VDC) and a qualitative transition from the brighter color to darker color on the left island in the visualization.

Table S15. Details of a cluster of nets identified with UMAP which have similar net descriptors.

net	type	v1C	v2C	Net density	td10	Genus	D-size	Average vertex order	Smallest ring size
ana-e	di	6	0	0.70	1825	97	48	1	3
pcu-m	di	6	0	0.68	1394	97	40	1	3
sxl	di	6	0	0.71	1317	97	52	1	3
sxm	di	6	0	0.99	2177	97	44	1	3
sxn	di	6	0	0.64	1189	97	36	1	3
sxo	di	6	0	1.08	1553	97	68	1	3
sxp	di	6	0	1.15	2657	97	44	1	3
sxq	di	6	0	0.95	2269	97	0	1	3

We then conducted UMAP reduction on the MOF structures with only net descriptors as features. MOFs formed clusters of varying sizes (Figure S19). Interestingly, we also see big clusters with black dots (nonporous MOFs) and yellow dots (best performing MOFs), which corroborates our earlier observation that the crowded nets yield nonporous MOFs but also the best performing MOFs. Coloring in the net type or index, the clusters of MOF data points corresponded to the single- and multi-vertex nets (Figure S21).

When both topological descriptors and porosity descriptors (PV, crystal density, VSA and GSA) are used as features for UMAP reduction, we observed a different relationship between the best performing MOFs and the nonporous MOFs. As shown in Figure S20, there are two major clusters of nonporous MOFs; one (upper left) cluster crosses with the best performing MOFs, whereas the other cluster is not adjacent to any best performing MOFs in the lower left region.

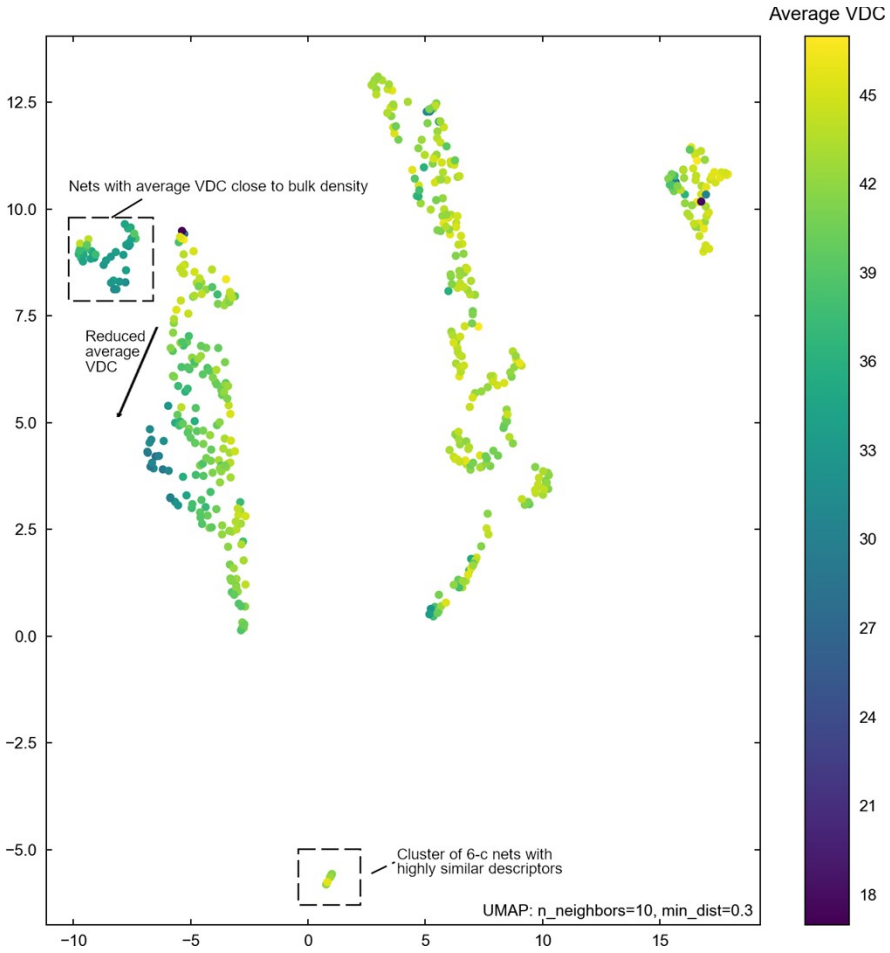


Figure S17. UMAP representation of the net data points, colored by the average VDC of each net.

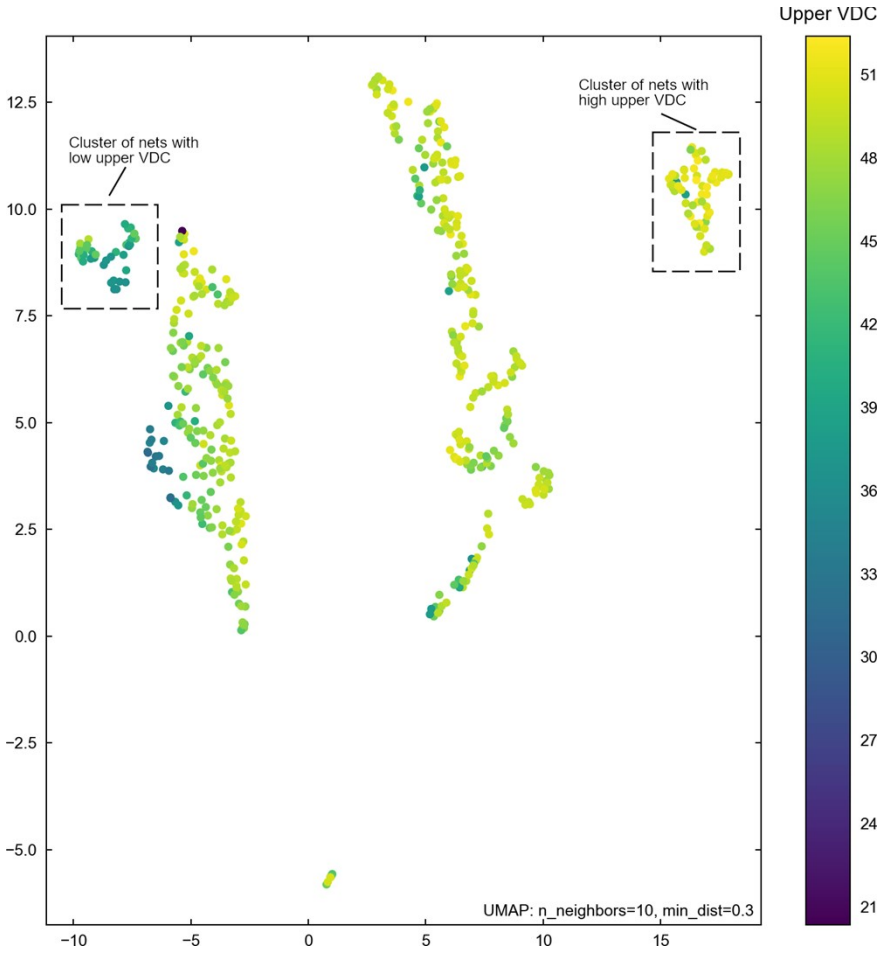


Figure S18. UMAP representation of the net data points, colored by the upper VDC of each net.

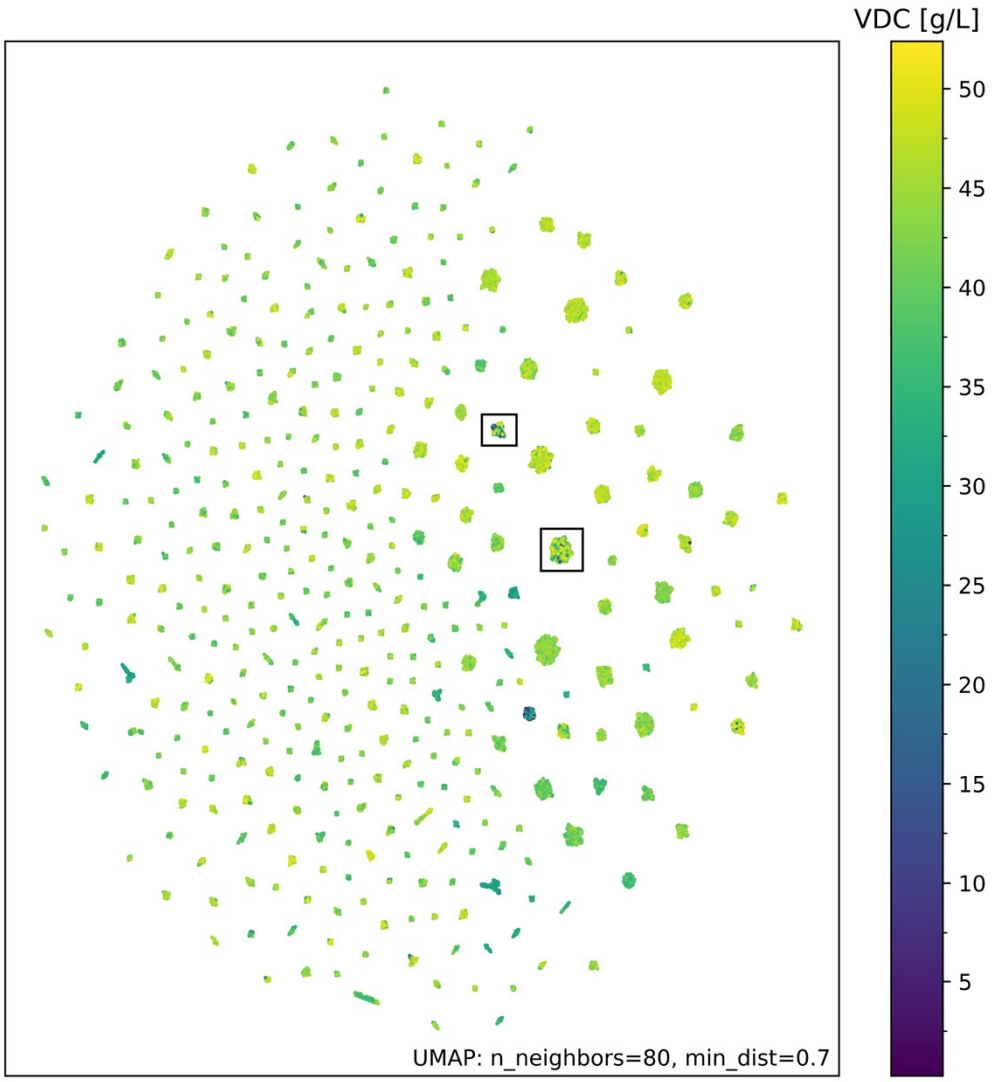


Figure S19. UMAP representation of the MOF data points based only on the topological descriptors.

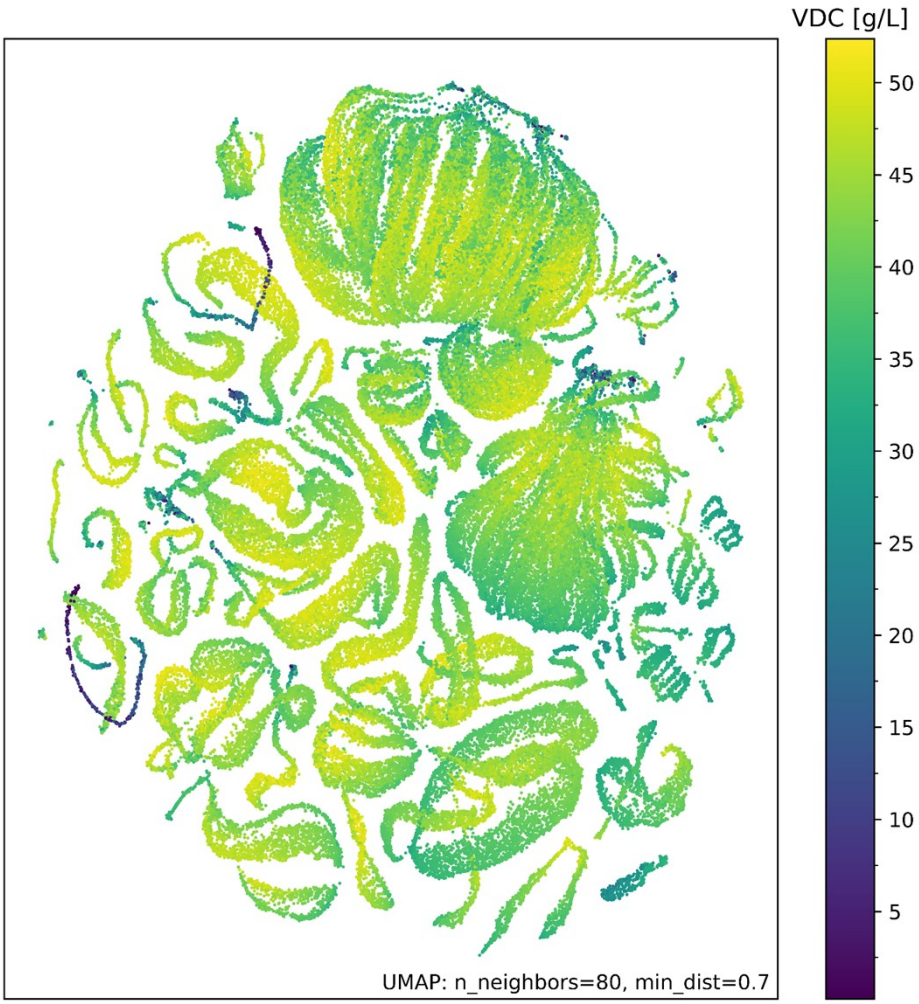


Figure S20. UMAP representation of the MOF data points based on topological descriptors and pore descriptors (crystal density ρ, PV, VSA and GSA). Dark purple denotes MOFs with low VDC, and bright yellow denotes MOFs with high VDC.

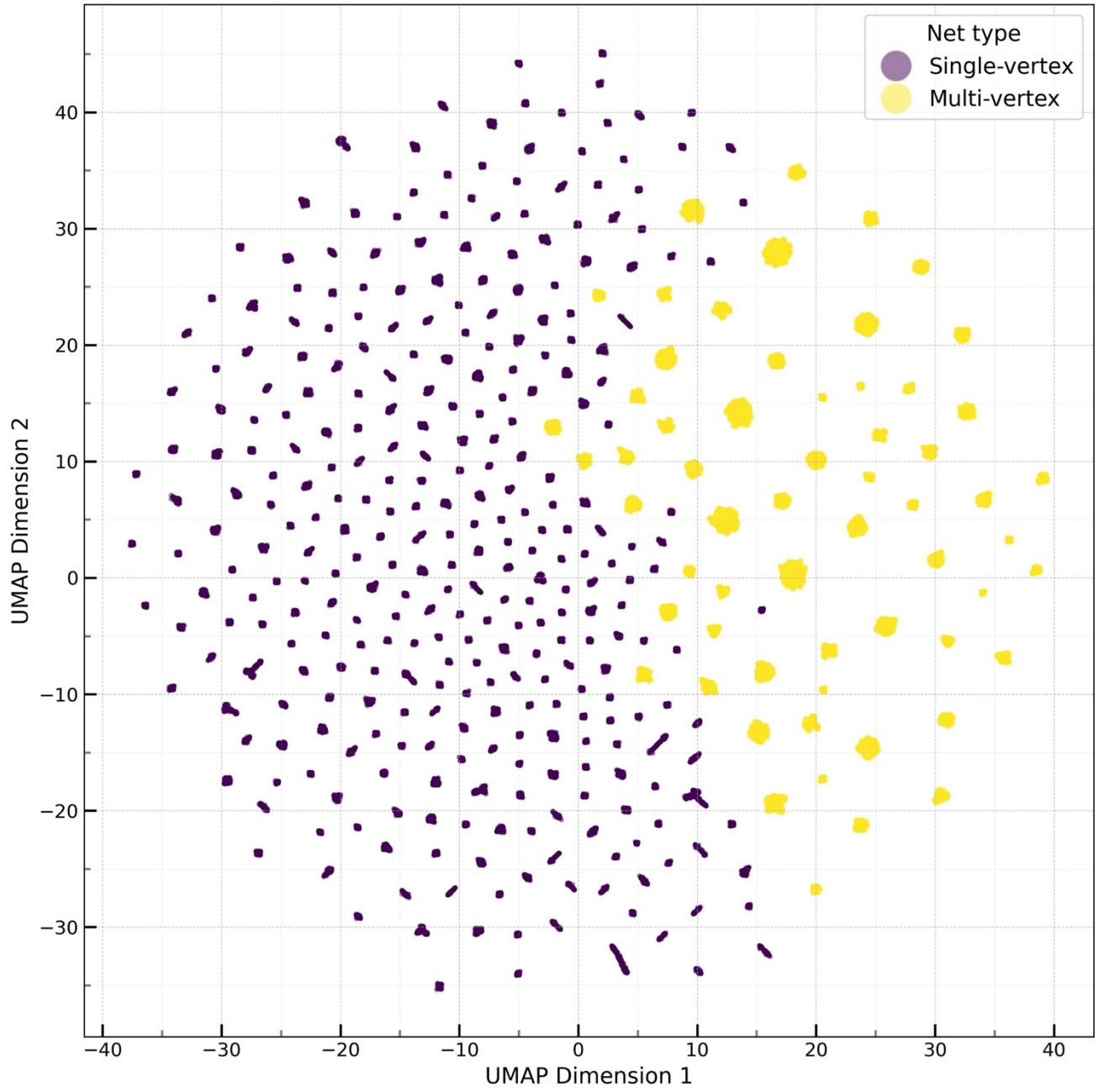


Figure S21. UMAP representation of the MOF data points based only on topological descriptors, colored by topology types. Purple: MOFs with single-vertex nets. Yellow: MOFs with multi-vertex nets. Data points are set to be half-transparent.

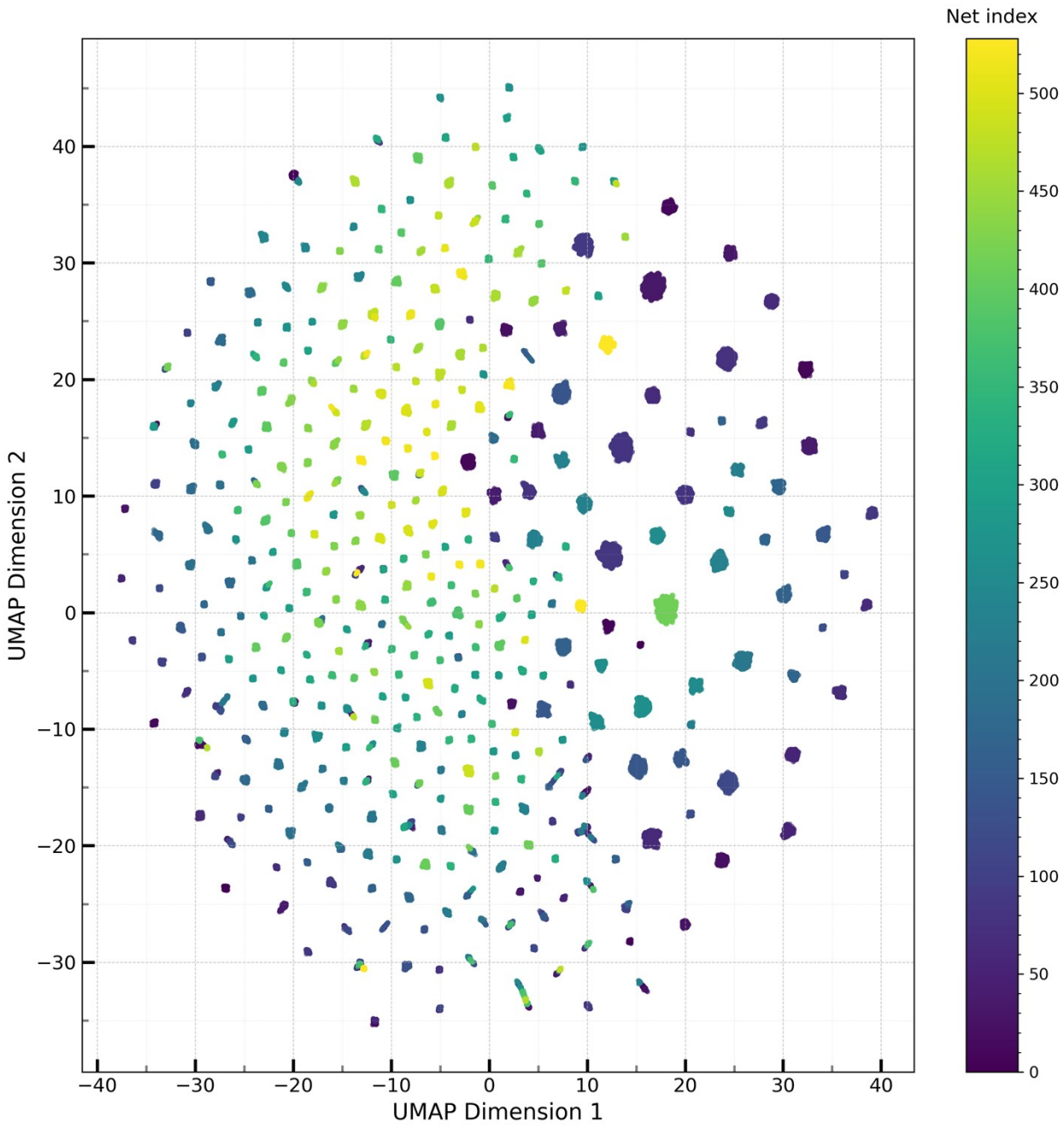


Figure S22. UMAP representation of the MOF data points based only on topological descriptors, colored by the topology index.

11 Additional data analysis

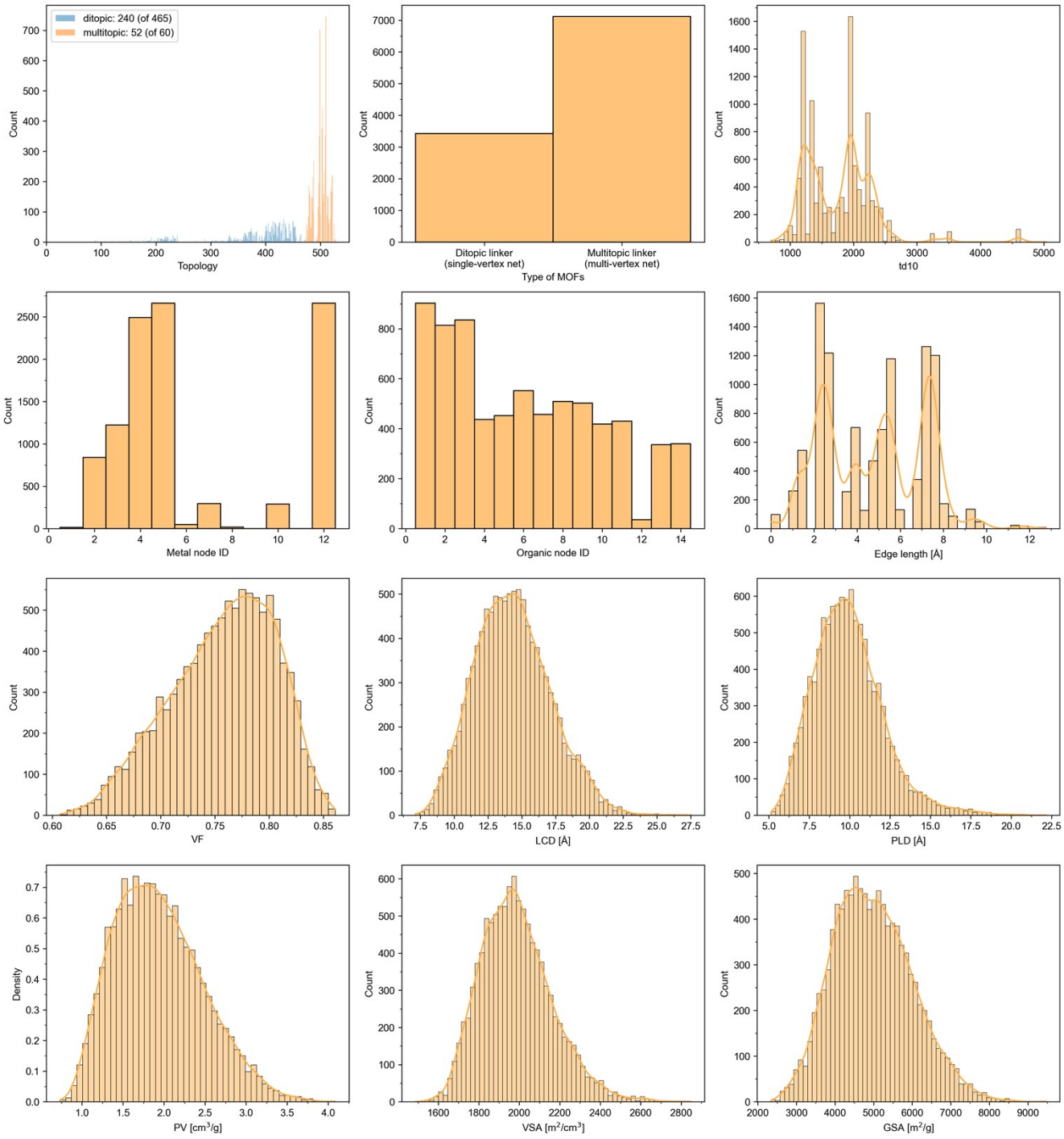


Figure S23. Overview of the top 10% of MOFs with the highest VDC. Top six panels: distribution of topologies and building blocks in the dataset. The first panel shows the structure count of each single-vertex and multi-vertex nets. The bottom six panels show the distribution of void fraction (VF), largest

cavity diameter (LCD), pore limiting diameter (PLD), pore volume (PV), volumetric surface area (VSA) and gravimetric surface area (GSA).

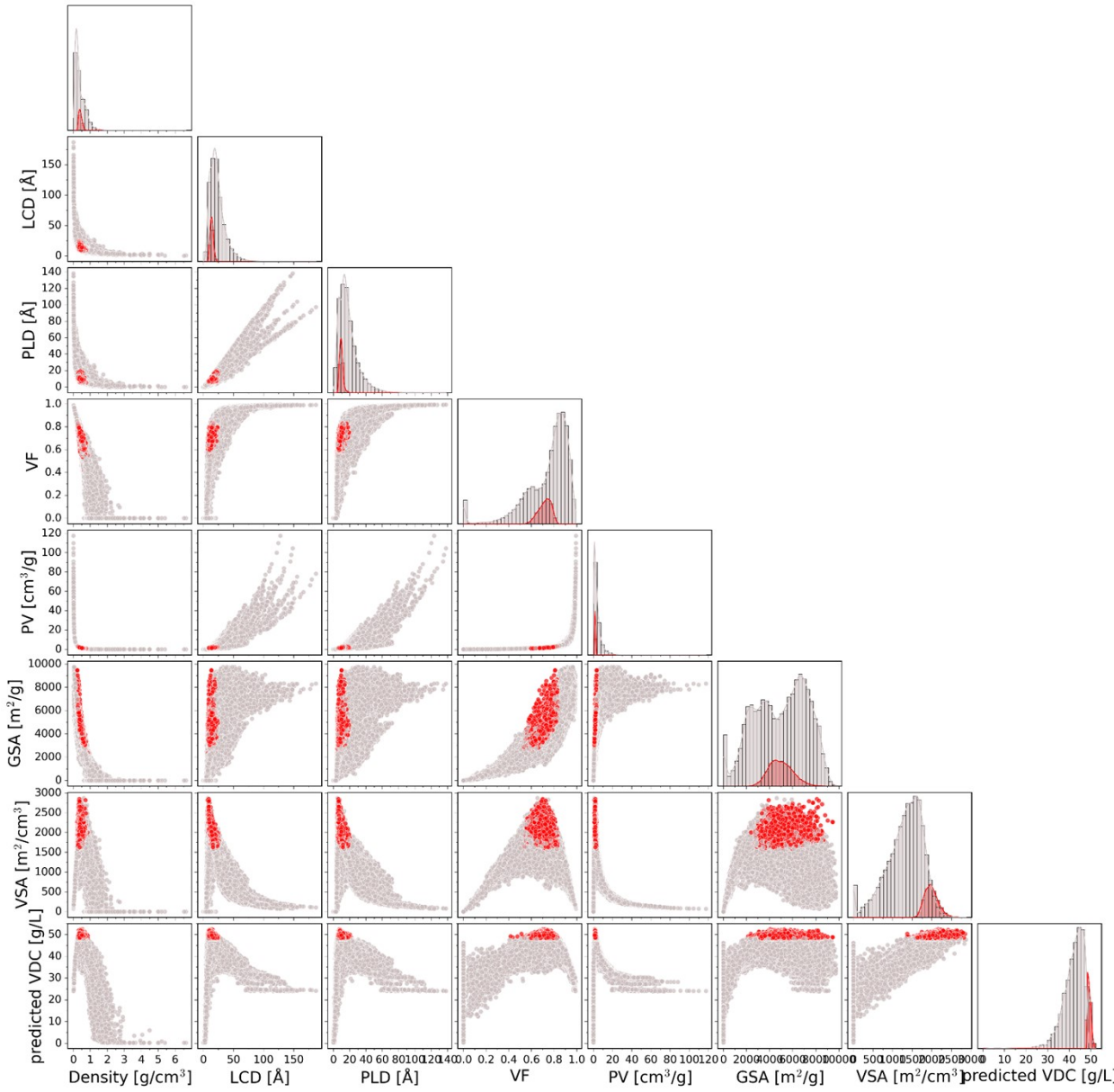


Figure S24. Textural property correlations of overall dataset (grey) and top 10% of the MOFs (red). Diagonal plots are frequency histogram of the respective property and the other plots are scatterplots.

Additional range-based plots

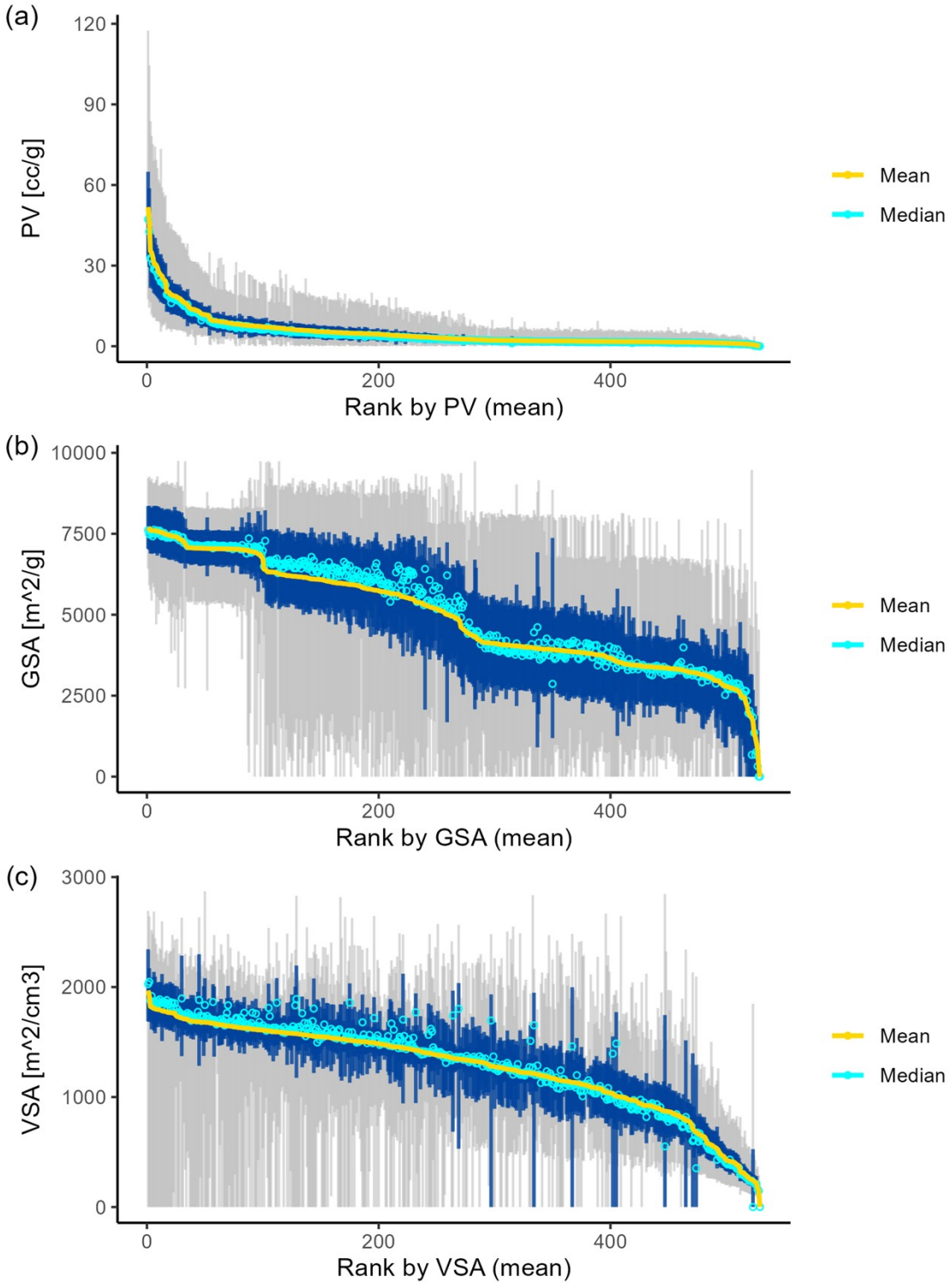


Figure S25. Distribution of textural properties of MOFs grouped by the underlying nets, for (a) pore volume (PV), (b) gravimetric surface area (GSA), (c) volumetric surface area (VSA). Zero denotes the

net with the highest mean value. The grey line represents the range of the property, dark blue represents the 25/75 quartile, light blue bubbles denotes the median, and yellow line denotes the average.

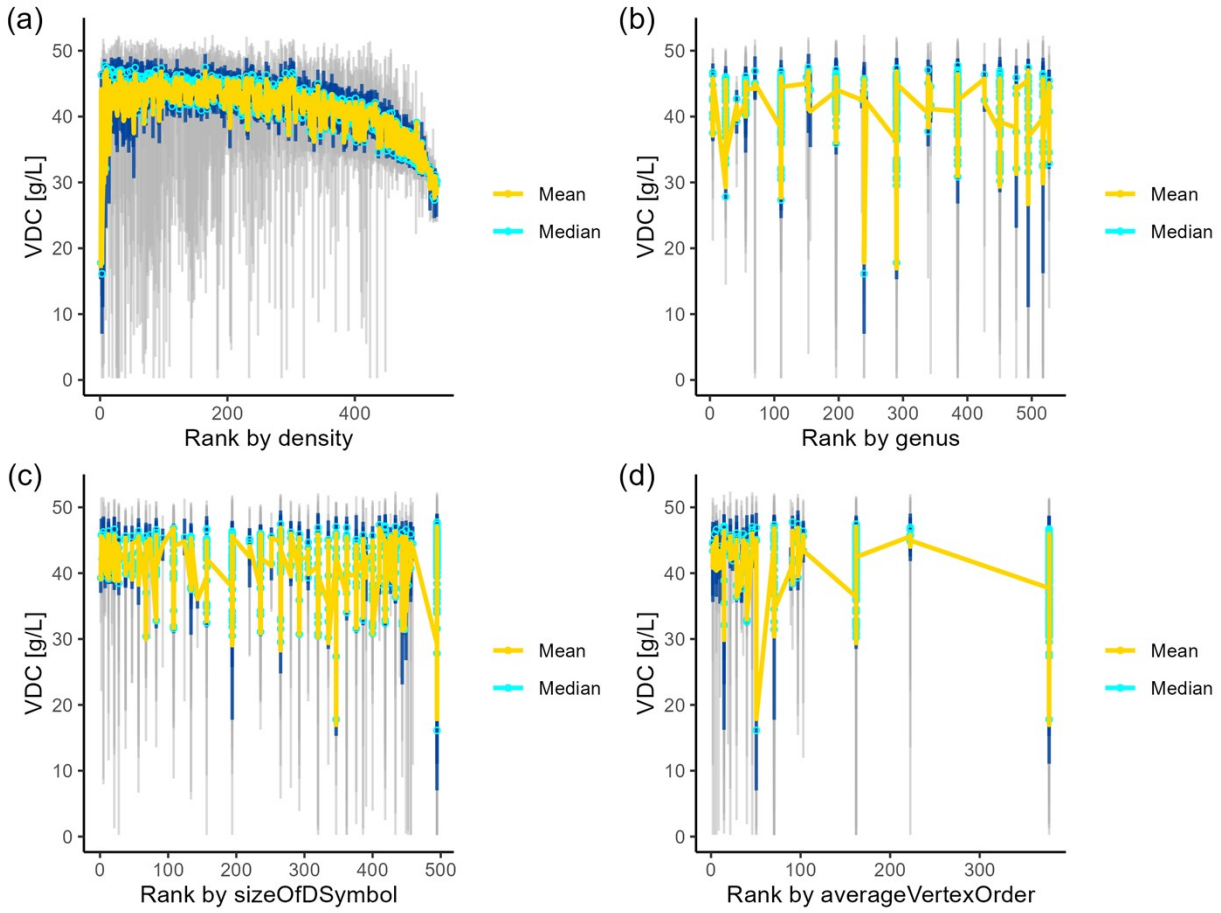


Figure S26. VDC as function of topological descriptors of MOFs grouped by the underlying nets for (a) net density, (b) genus, (c) D-symbol size, and (d) average vertex order. Zero indicates the net with the highest mean value. Grey represents the range of VDC, dark blue represents the 25/75 quartile, light blue denotes the median, and yellow denotes the average for each topology.

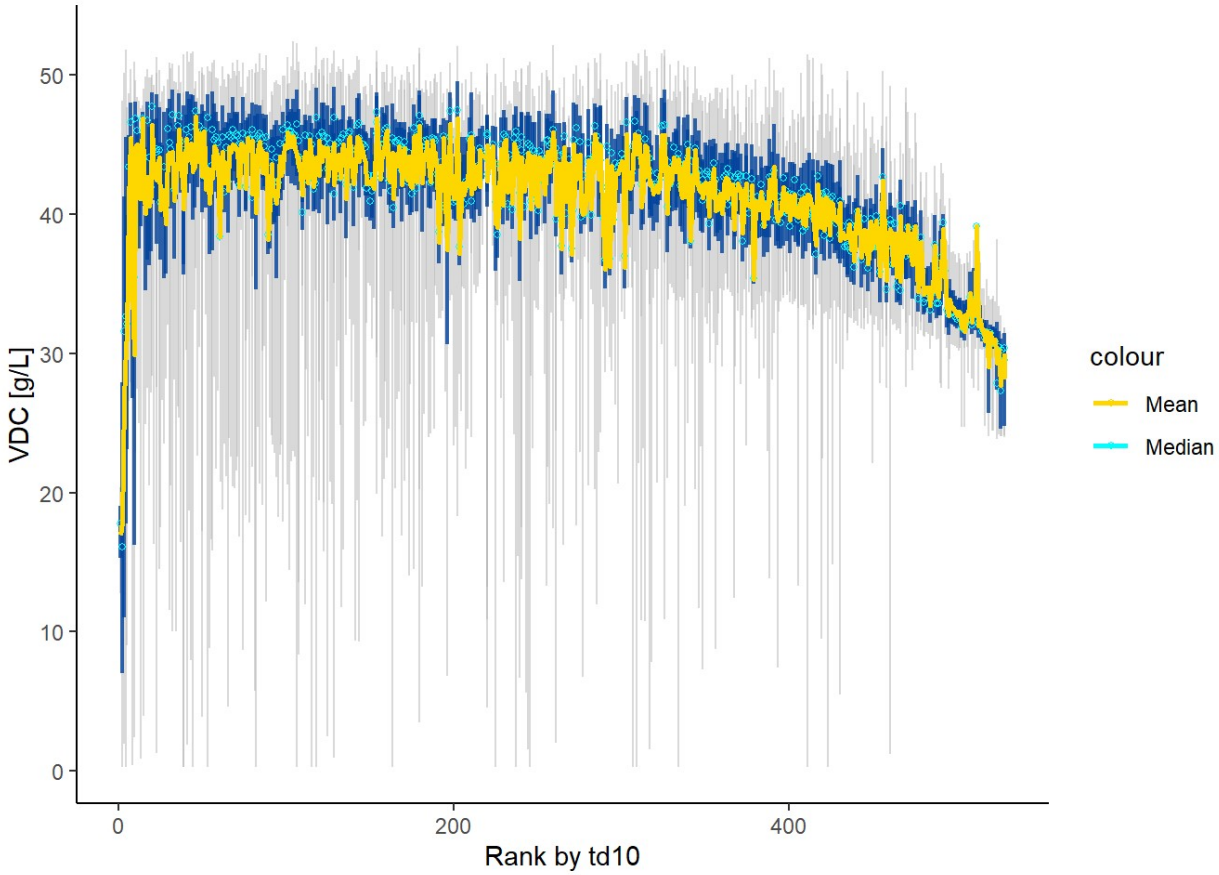


Figure S27. VDC as function of topological descriptor, td10.

The p-values for the correlation heatmap

Table S16. P-values corresponding to the correlation heatmap in Figure 8. This table details the p-values for each Spearman correlation coefficient, with color intensity indicating the statistical significance. White: no statistical significance. Light yellow: $p<0.05$; medium yellow: $p<0.01$; deep yellow: $p<0.001$.

	Net type	Coordination of V1	Coordination of V2	Net density	td10	Genus	D-size	Average vertex order	Smallest ring size
Lowest predicted capacity [g/L]	5.18×10^{-6}	2.24×10^{-2}	4.24×10^{-6}	8.98×10^{-17}	2.85×10^{-12}	3.26×10^{-4}	4.68×10^{-2}	1.69×10^{-4}	5.87×10^{-1}
Average predicted capacity [g/L]	1.89×10^{-6}	1.40×10^{-22}	7.16×10^{-7}	2.45×10^{-68}	1.18×10^{-58}	2.86×10^{-2}	2.22×10^{-1}	1.08×10^{-2}	3.94×10^{-5}
Highest predicted capacity [g/L]	1.53×10^{-15}	5.98×10^{-16}	1.87×10^{-15}	1.48×10^{-70}	1.28×10^{-55}	1.72×10^{-1}	8.37×10^{-1}	3.57×10^{-6}	4.32×10^{-2}
Lowest VSA [m^2/cm^3]	1.14×10^{-6}	7.78×10^{-2}	1.01×10^{-6}	1.69×10^{-8}	1.85×10^{-4}	3.99×10^{-7}	3.98×10^{-4}	1.73×10^{-5}	5.46×10^{-4}
Average VSA [m^2/cm^3]	1.64×10^{-1}	2.60×10^{-48}	1.20×10^{-1}	6.26×10^{-119}	9.02×10^{-110}	3.66×10^{-3}	1.21×10^{-1}	9.54×10^{-2}	6.70×10^{-14}
Highest VSA [m^2/cm^3]	2.82×10^{-8}	1.59×10^{-5}	3.63×10^{-8}	1.29×10^{-47}	9.98×10^{-34}	7.07×10^{-4}	3.56×10^{-1}	1.82×10^{-4}	4.56×10^{-1}
Lowest PV [cm^3/g]	2.28×10^{-6}	3.21×10^{-26}	1.28×10^{-6}	1.34×10^{-76}	7.37×10^{-69}	7.61×10^{-1}	6.99×10^{-1}	9.63×10^{-4}	1.01×10^{-7}
Average PV [cm^3/g]	7.95×10^{-1}	8.56×10^{-116}	9.45×10^{-1}	1.29×10^{-185}	1.56×10^{-231}	1.55×10^{-8}	1.59×10^{-1}	1.33×10^{-1}	7.49×10^{-37}
Highest PV [cm^3/g]	7.59×10^{-2}	1.36×10^{-119}	1.15×10^{-1}	3.89×10^{-174}	3.41×10^{-212}	8.92×10^{-10}	1.23×10^{-1}	2.70×10^{-1}	6.94×10^{-38}

Additional figures of MOF data points segregated by net density

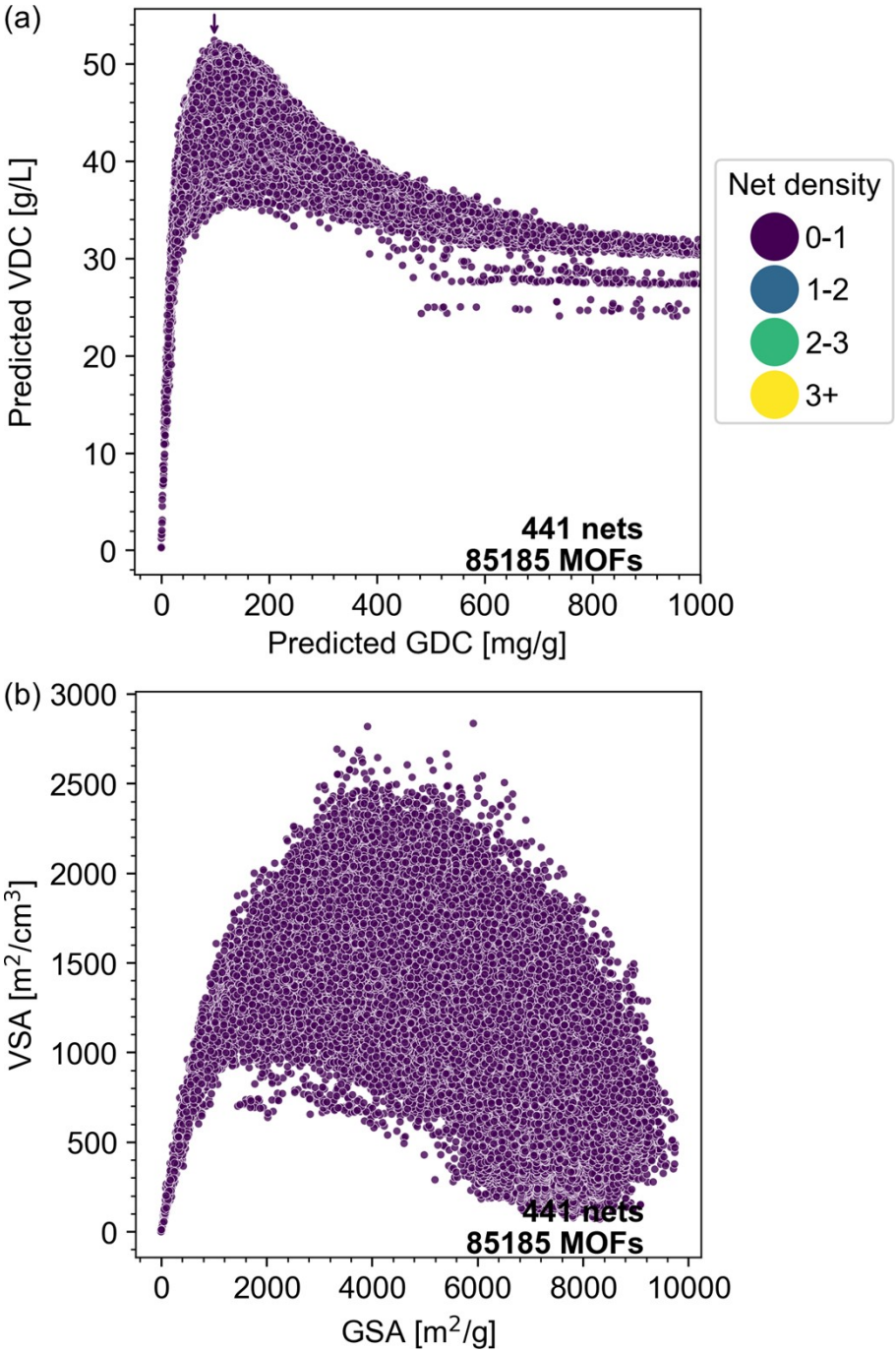


Figure S28. Analysis of MOF data points in Figure 9 segregated by net density with only the MOFs having net density between 0–1 plotted.

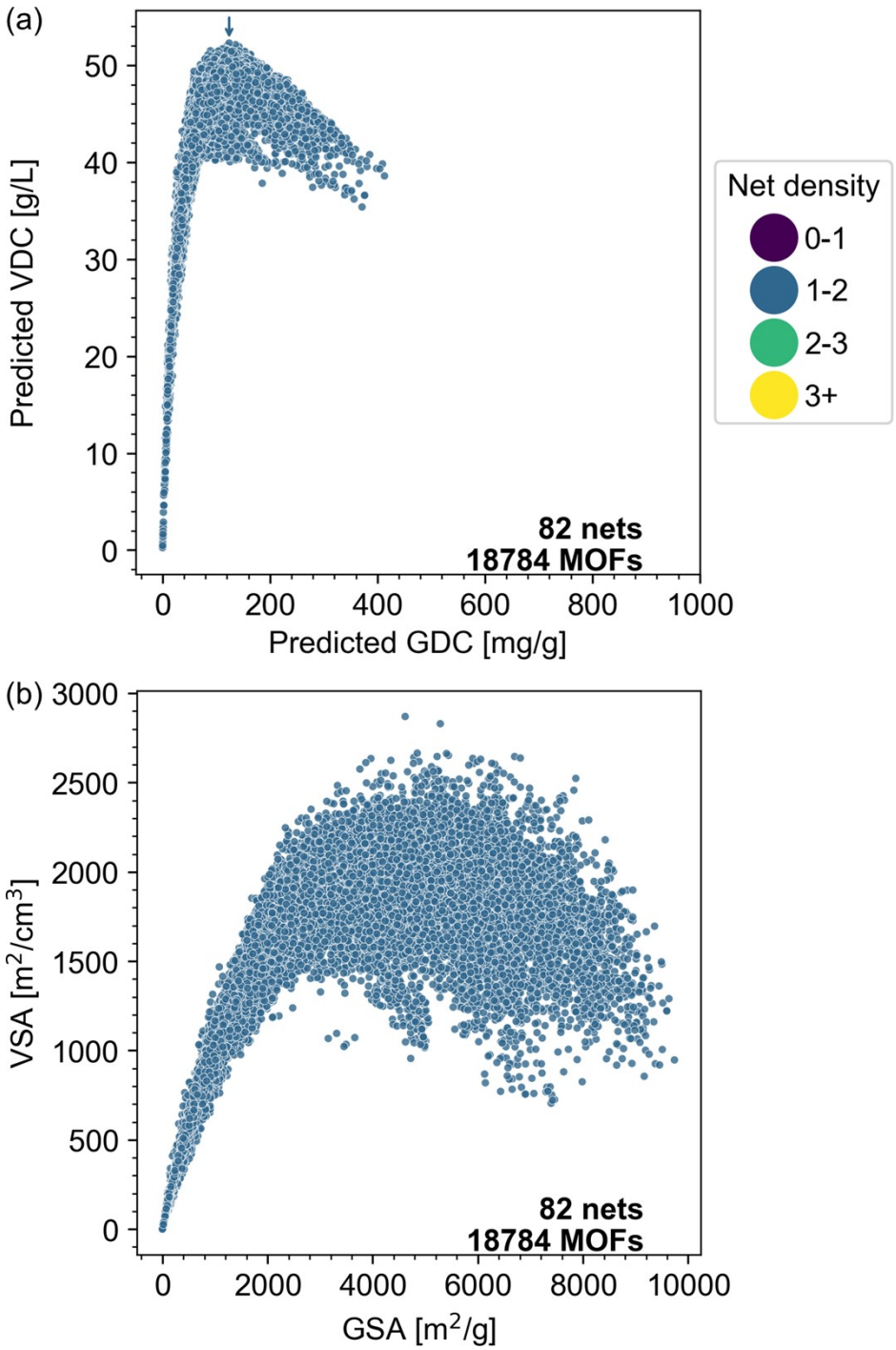


Figure S29. Analysis of MOF data points in Figure 9 segregated by net density with only the MOFs having net density between 1–2 plotted.

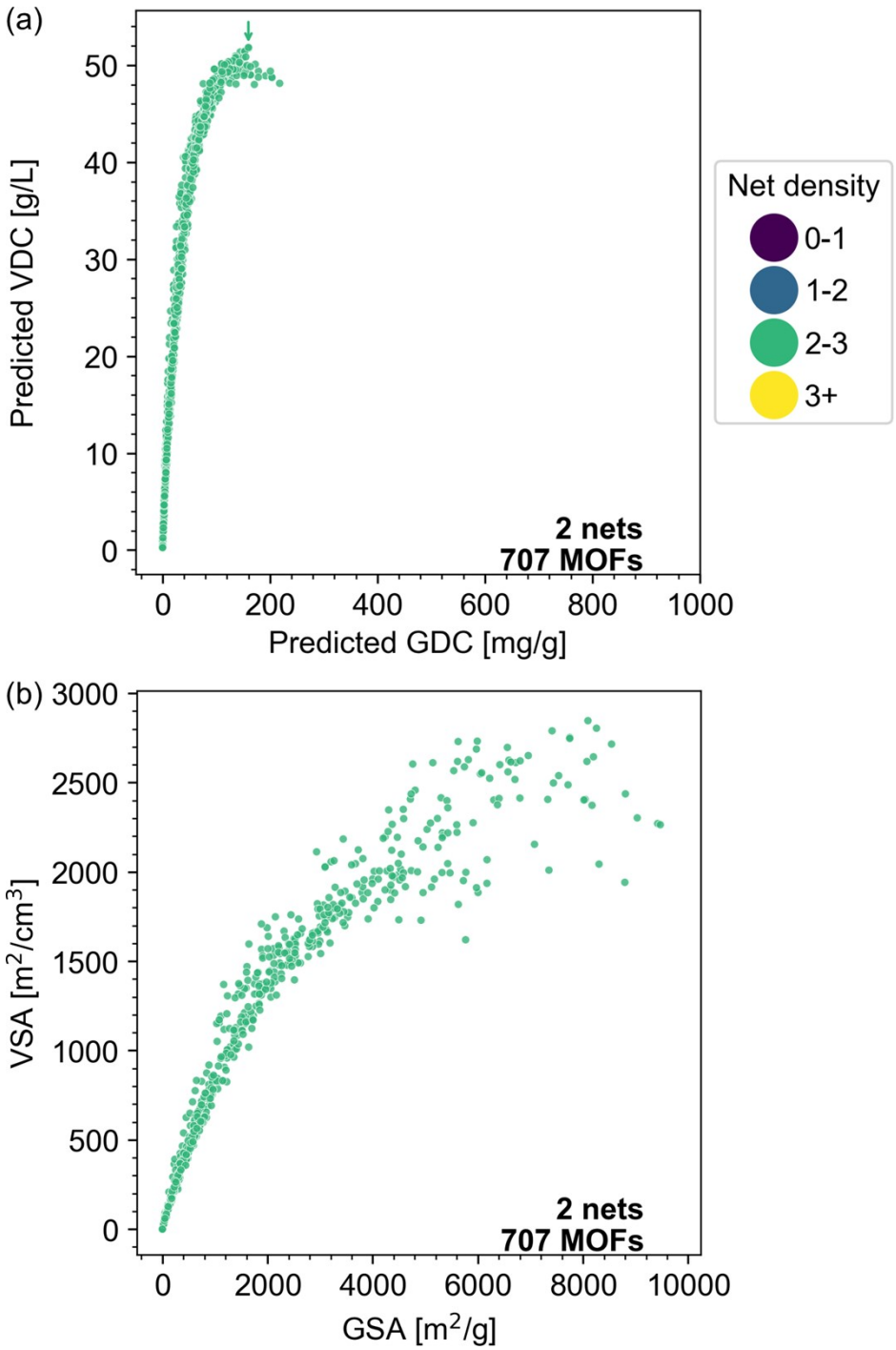


Figure S30. Analysis of MOF data points in Figure 9 segregated by net density with only the MOFs having net density between 2–3 plotted.

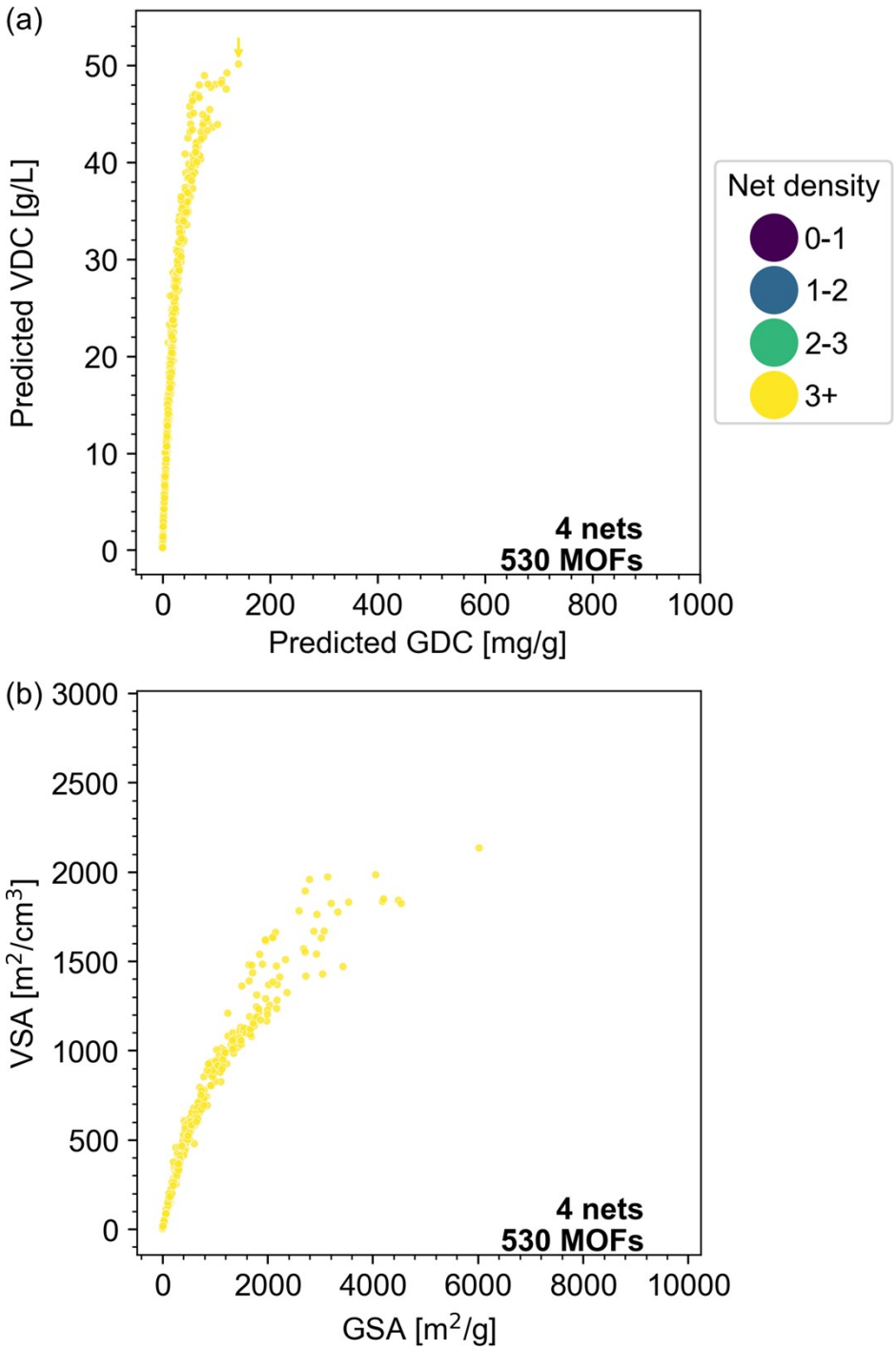


Figure S31. Analysis of MOF data points in Figure 9 segregated by net density with only the MOFs having net density between 3–4 plotted.

Analysis based on the surface area landscape of the dataset

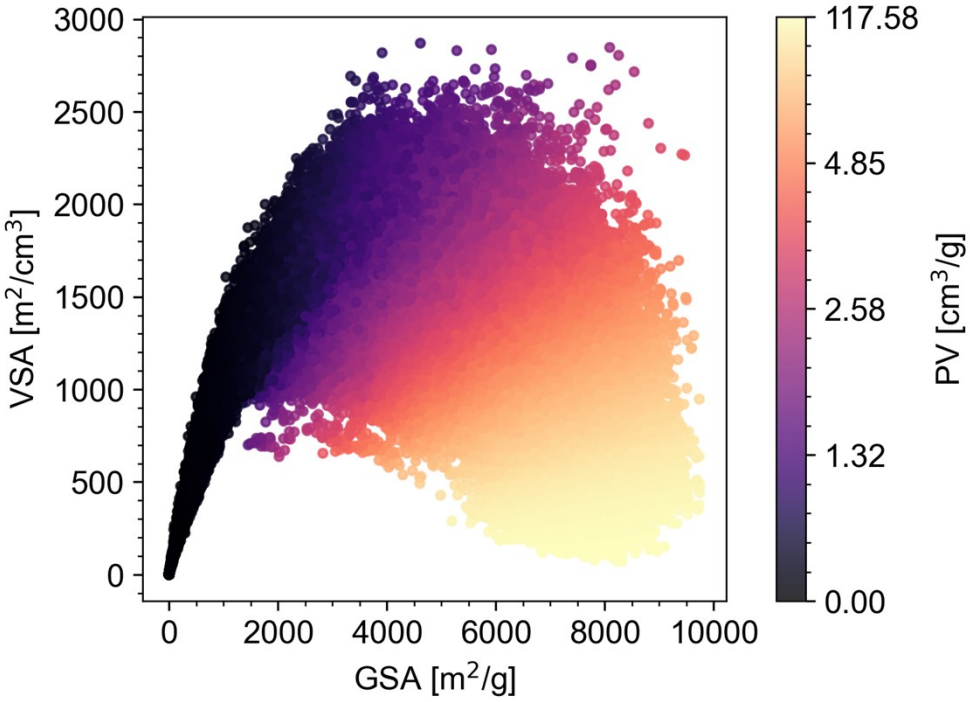


Figure S32. Surface areas of 105,206 MOFs colored by the pore volume. The pore volume of MOFs with high VSA increases as GSA increases.

Effect of metal clusters on deliverable capacity

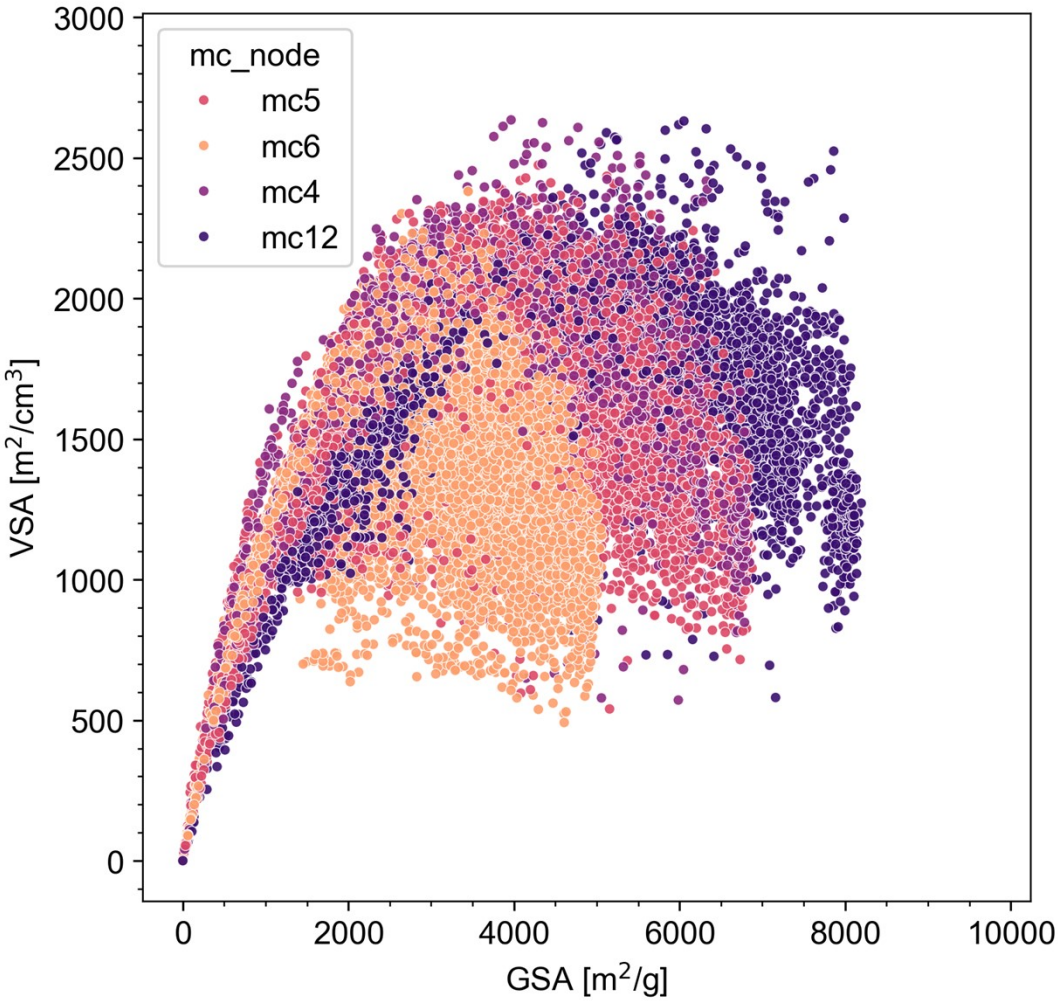


Figure S33. Surface areas of the MOFs with 6-c nets, colored by the metal clusters. Different metal clusters result in multiple volcano patterns.

Identifying the MOF topology using MOFid

Table S17. Mismatch between MOFid topology and the actual topology in the dataset. The folder column denotes subdirectories used to store the MOFs in the Zenodo repository.[1] The MOFid count indicates the number of structures for which a MOFid was successfully generated, and the mismatch count denotes the number of structures where the topology from MOFid did not match the topology used for construction.

Folder	MOFid count	Mismatch count	Error rate
f1	19514	983	5.04%
f2	3946	251	6.36%
f3	1853	97	5.23%
f4	20477	119	0.58%
f5	18110	170	0.94%
f6	22308	332	1.49%
f7	19556	329	1.68%
(Total)	105764	2281	2.16%

12 Topology related illustrations

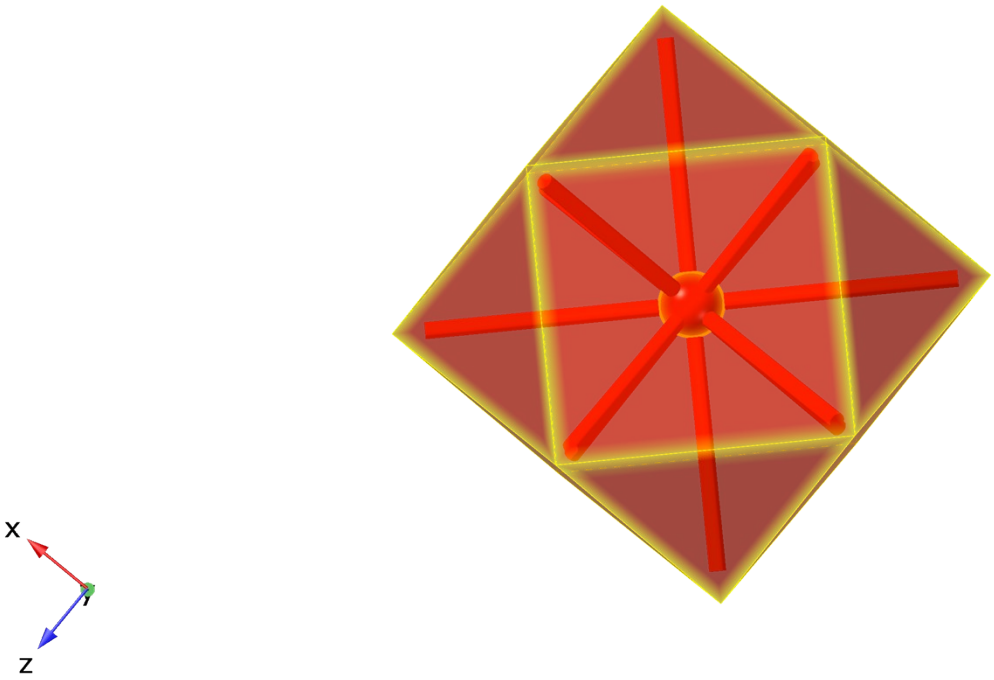


Figure S34. The cuboctahedron shape of the vertex of **fcu**.

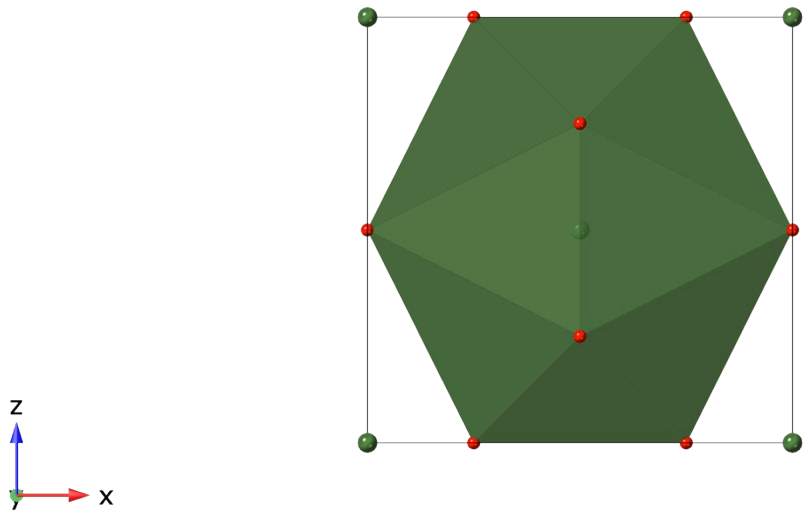


Figure S35. The icosahedron shape of the 12-coordinated vertex of **ith**. Green: 12-coordinated vertex. Red: 4-coordinated vertex.

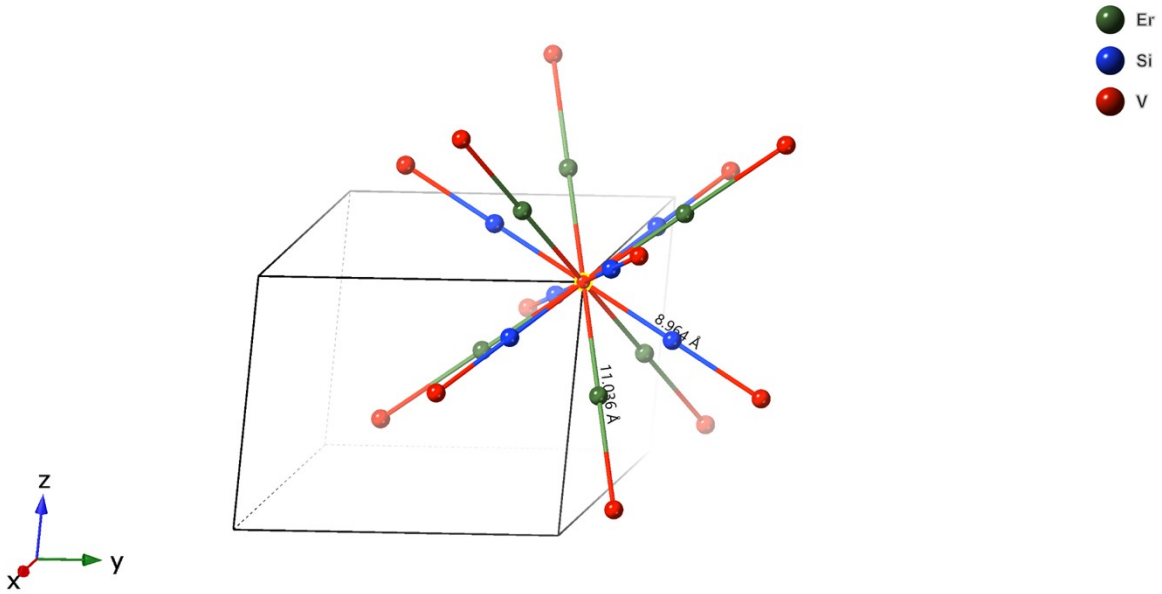


Figure S36. Illustration of **ild** net highlighting the connectivity of one reference vertex. Vertices are shown in red. There are two types of the neighboring vertices, which are 8.964 Å apart (shorter edge, whose center is shown as blue sphere) and 11.036 Å apart (longer edge, whose center is shown as green sphere).

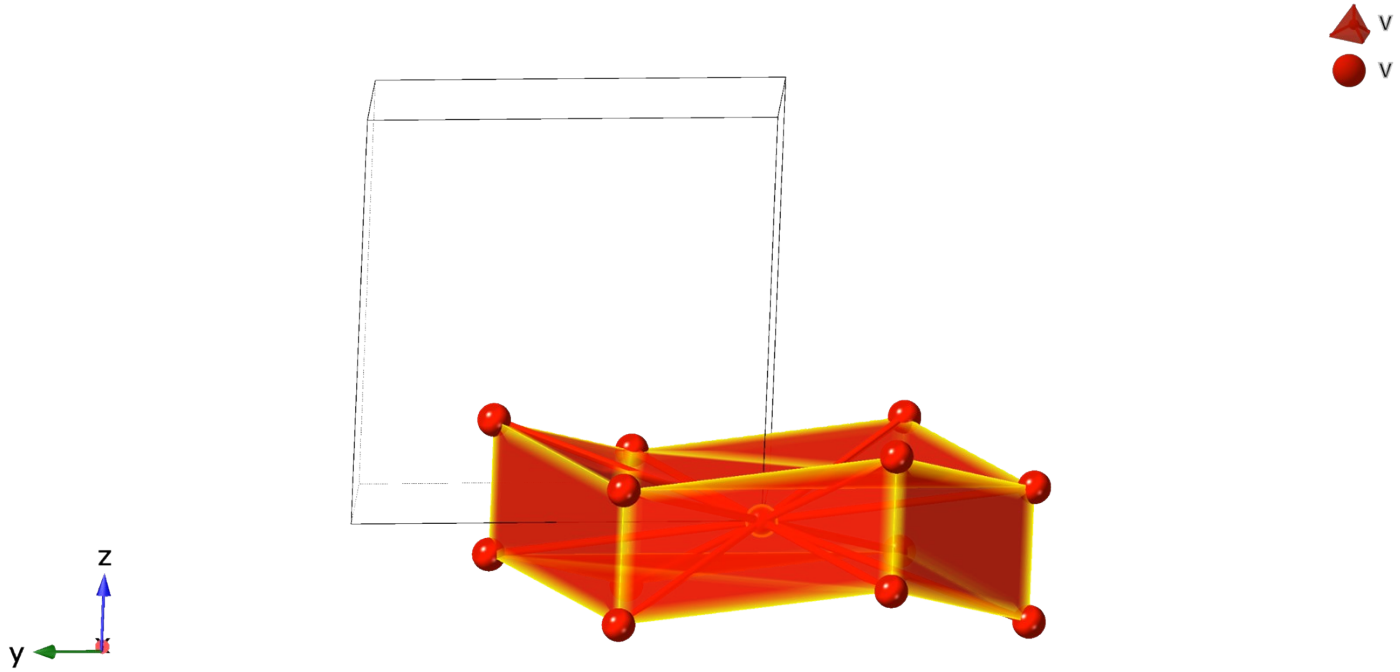


Figure S37. The polyhedral shape of the 12-coordinated vertex of **ild**.

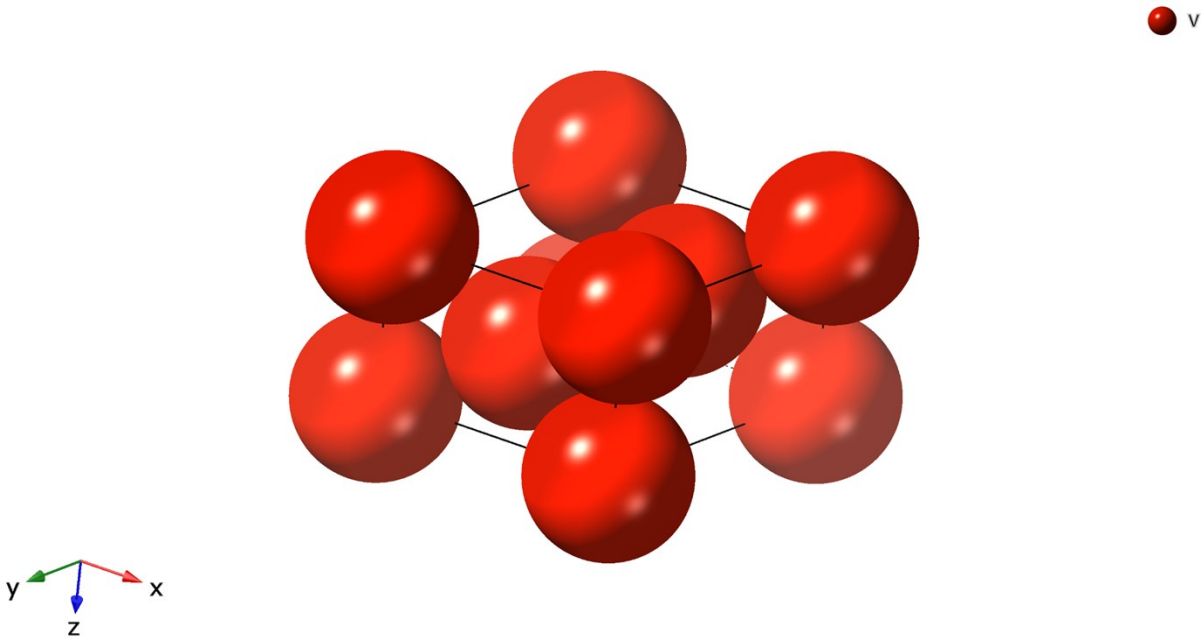


Figure S38. Illustration of **ild** net with all vertices shown as spheres. Vertices are shown in red. When the central left vertex and the top left vertex touch each other, the central left vertex and the bottom left vertex overlap as a result of different edge lengths.

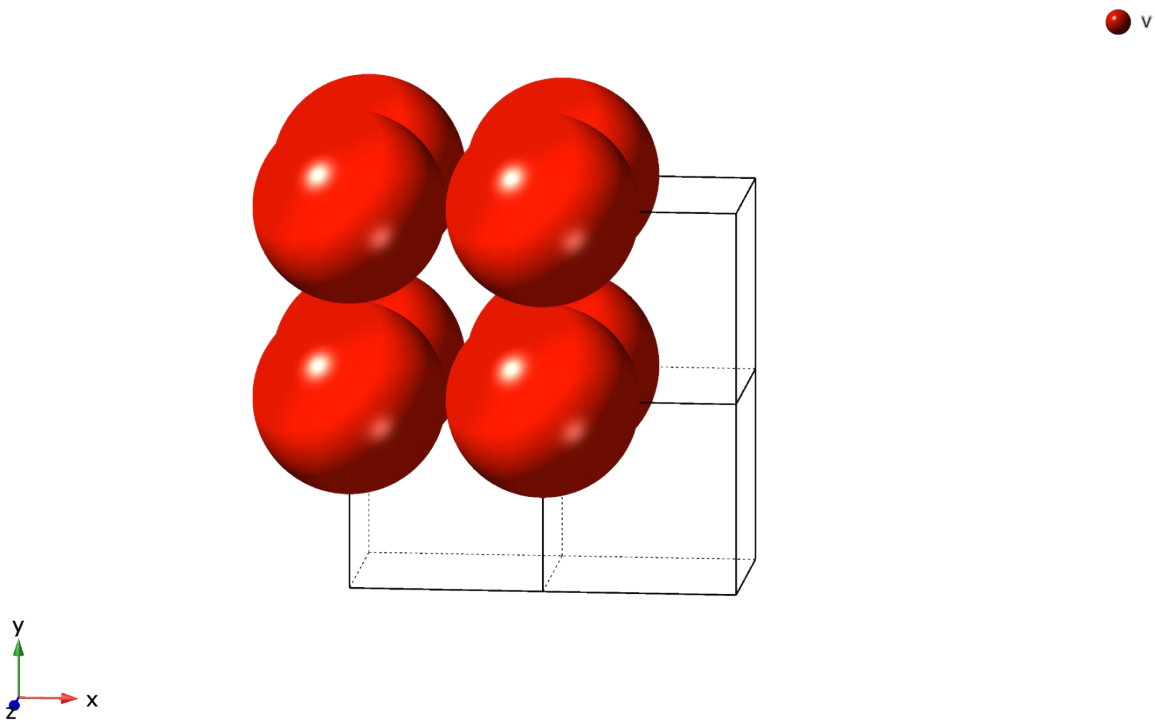


Figure S39. Illustration of **pcu** net with all vertices shown as spheres. Vertices are shown in red. It is the densest 6-c sphere-packing net possible.[12]

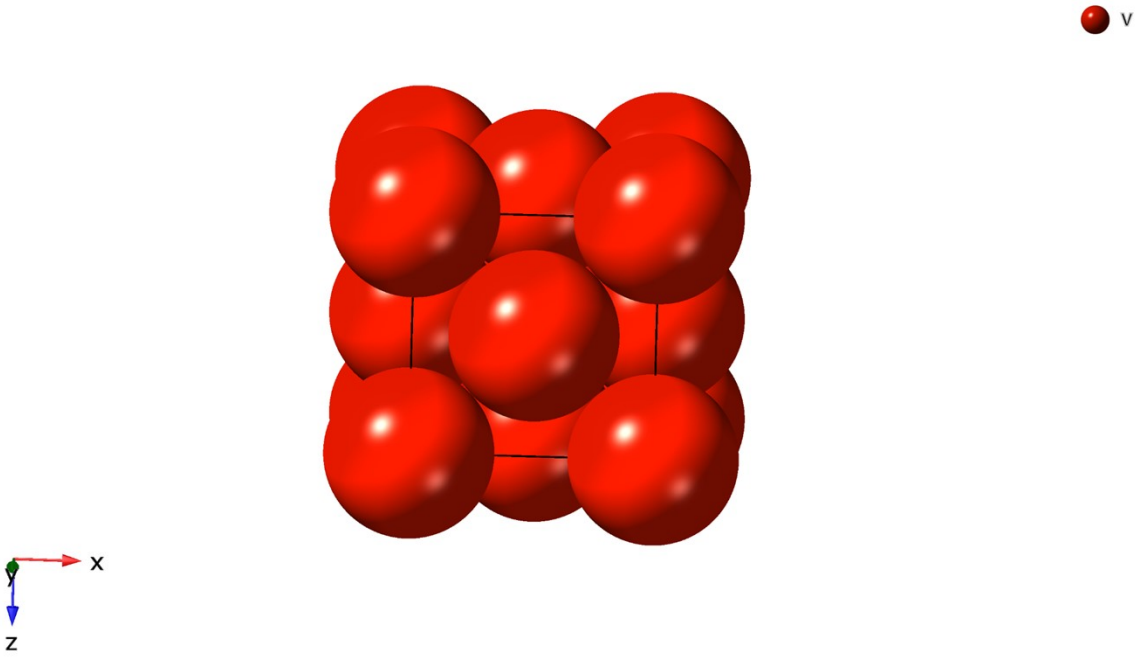


Figure S40. Illustration of **fcu** net with all vertices shown as spheres. Vertices are shown in red. It is the densest 12-c sphere-packing net possible.[12]

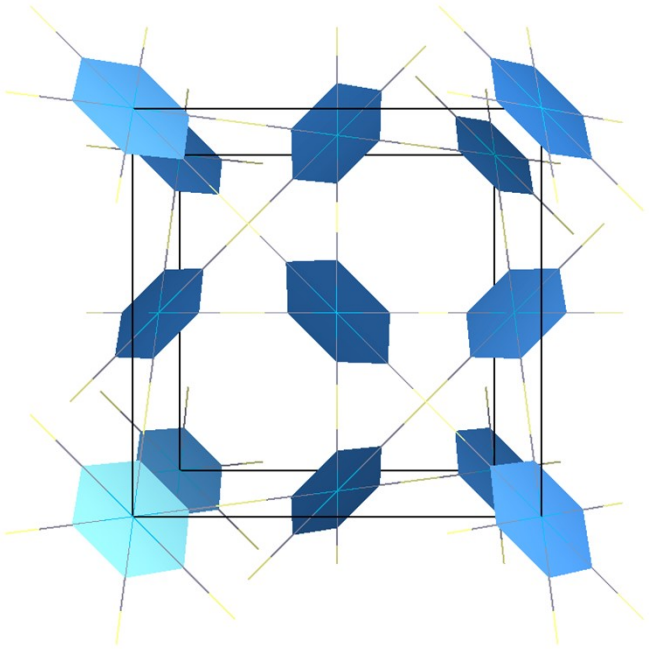


Figure S41. Illustration of **hxg** net with all vertices (blue) shown as polyhedra based on the coordination. Edge centers are shown in yellow.

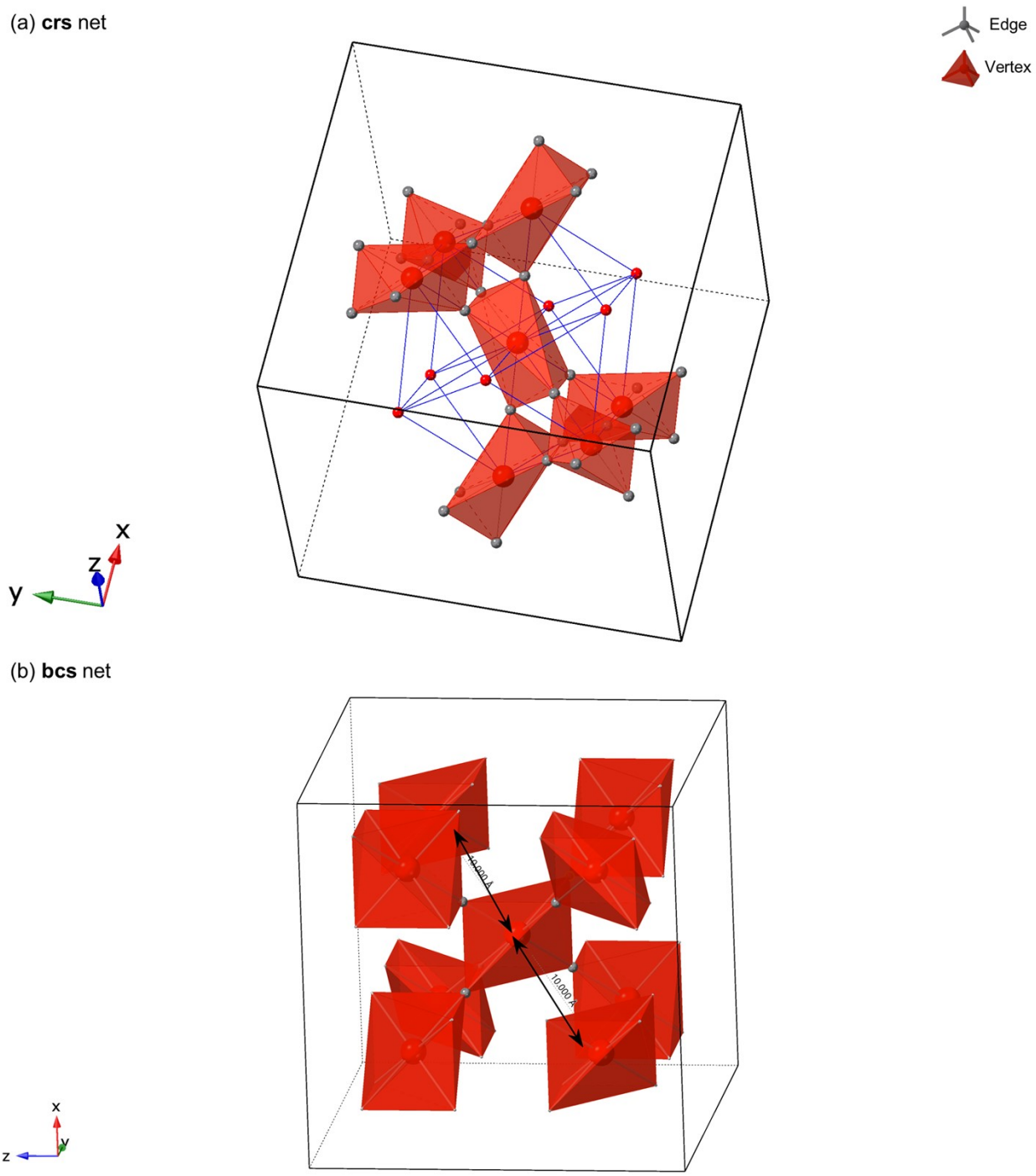


Figure S42. Illustration of non-sphere packing nature of the (a) **crs** net and (b) **bcs** net. (a) Connected vertices are shown as red polyhedra where connection points are shown as grey dots. Red spheres without polyhedra represent the neighboring six vertices unconnected to the central vertex. They are at the same distance from the central vertex as the six connected vertices. (b) The neighboring two

unconnected vertices are highlighted, with annotations denoting their distance of 10 Å , which is the same unit distance between the connected vertices.

13 Figures to validate the correlation of net density and deliverable capacities and surface areas

The following figures accompany the discussion of controlling the variance in net coordination, organic node, and metal cluster. Additional plots colored based on td10 are available in the Zenodo repository.[1] We also recognized that higher coordinated metal clusters (namely, 12-connected Zr-cluster) had fewer compatible nets (Table S18), and, vice versa, nets based on higher coordinated vertex had fewer compatible metal nodes (Table S19). For example, among the 4,12-c nets, the **ith** net is experimentally observed with icosahedron shaped Zr cluster (Figure S35) in MOF-812[13], a metal cluster that is not considered in the dataset.

Table S18. The count … columns denote the m… and their coordination … combinations, with th… …s. The … nodes

Coordination	MC Node	3-c	4-c		6-c				8-c		10-c	12-c	24-c
		mc1	mc2	mc3	mc4	mc5	mc6	mc12	mc7	mc8	mc9	mc10	mc11
Coordination	ON Node												
3-c	on1	339 [3]	122 [2]	122 [2]	907 [15]	973 [16]	774 [13]	907 [15]	61 [1]	61 [1]	0 [0]	61 [1]	0 [0]
	on2	344 [3]	122 [2]	122 [2]	908 [15]	973 [16]	776 [13]	908 [15]	61 [1]	61 [1]	0 [0]	61 [1]	0 [0]
	on3	359 [3]	122 [2]	122 [2]	912 [15]	971 [16]	771 [13]	912 [15]	61 [1]	61 [1]	0 [0]	61 [1]	0 [0]
4-c	on4	122 [2]	230 [2]	546 [7]	397 [7]	292 [5]	300 [5]	397 [7]	121 [2]	122 [2]	0 [0]	60 [1]	0 [0]
	on5	122 [2]	230 [2]	543 [7]	398 [7]	294 [5]	300 [5]	398 [7]	121 [2]	122 [2]	0 [0]	60 [1]	0 [0]
	on6	183 [3]	595 [8]	1004 [9]	444 [8]	407 [7]	272 [5]	443 [8]	180 [3]	181 [3]	0 [0]	101 [2]	0 [0]
	on7	183 [3]	596 [8]	909 [9]	365 [6]	362 [6]	293 [5]	365 [6]	240 [4]	234 [4]	0 [0]	118 [2]	0 [0]
	on8	182 [3]	540 [8]	851 [9]	366 [6]	365 [6]	292 [5]	366 [6]	237 [4]	218 [4]	0 [0]	104 [2]	0 [0]
	on9	176 [3]	542 [8]	842 [9]	366 [6]	366 [6]	287 [5]	366 [6]	240 [4]	220 [4]	0 [0]	101 [2]	0 [0]
	on10	122 [2]	548 [7]	973 [8]	427 [7]	365 [6]	366 [6]	427 [7]	241 [4]	186 [4]	0 [0]	180 [3]	0 [0]
	on11	183 [3]	608 [8]	1020 [9]	488 [8]	425 [7]	425 [7]	488 [8]	235 [4]	184 [4]	0 [0]	180 [3]	0 [0]
	on13	244 [4]	548 [7]	925 [9]	530 [10]	468 [8]	463 [8]	530 [10]	270 [5]	255 [5]	0 [0]	176 [3]	0 [0]
	on14	121 [2]	167 [2]	482 [7]	371 [7]	282 [5]	240 [4]	372 [7]	119 [2]	120 [2]	0 [0]	59 [1]	0 [0]
6-c	on12	234 [4]	122 [2]	121 [2]	59 [1]	60 [1]	0 [0]	59 [1]	61 [1]	60 [1]	0 [0]	0 [0]	0 [0]
No organic node		3589 [60]	10182 [169]	8676 [152]	5954 [112]	11774 [202]	9830 [171]	5949 [112]	466 [9]	393 [8]	0 [0]	112 [2]	0 [0]

Table S19. The count of MOFs and corresponding nets grouped by the net coordination. The columns and rows denote the vertex coordination of the nets. The row NA denotes single-vertex nets where the vertex coordination is denoted in the column. Each entry denotes a group of nets of the same vertex coordination, and shows the count of MOFs, with the number of included topologies (nets) indicated in brackets.

vertex coordination	NA	3-c	4-c	6-c	8-c	12-c
NA	0 [0]	3406 [57]	18249 [185]	33325 [216]	859 [9]	112 [2]
3-c	--	1225 [3]	2370 [4]	10926 [19]	366 [1]	183 [1]
4-c	--	--	13308 [10]	15411 [11]	3846 [5]	1139 [3]
6-c	--	--	--	360 [2]	121 [1]	0 [0]
8-c	--	--	--	--	0 [0]	0 [0]
12-c	--	--	--	--	--	0 [0]

Subgroups divided by the coordination of nets

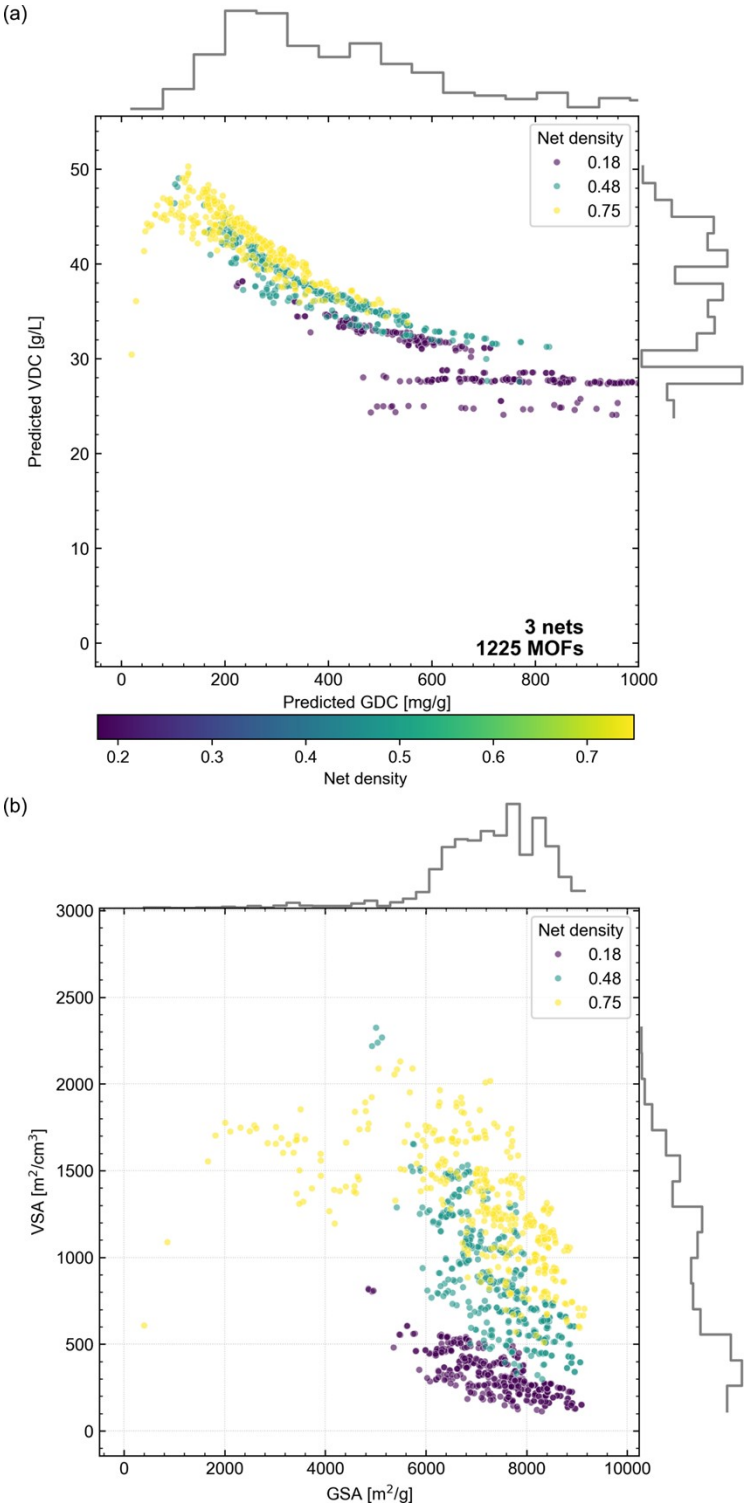


Figure S43. Bivariate plots of the MOFs with the 3,3-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

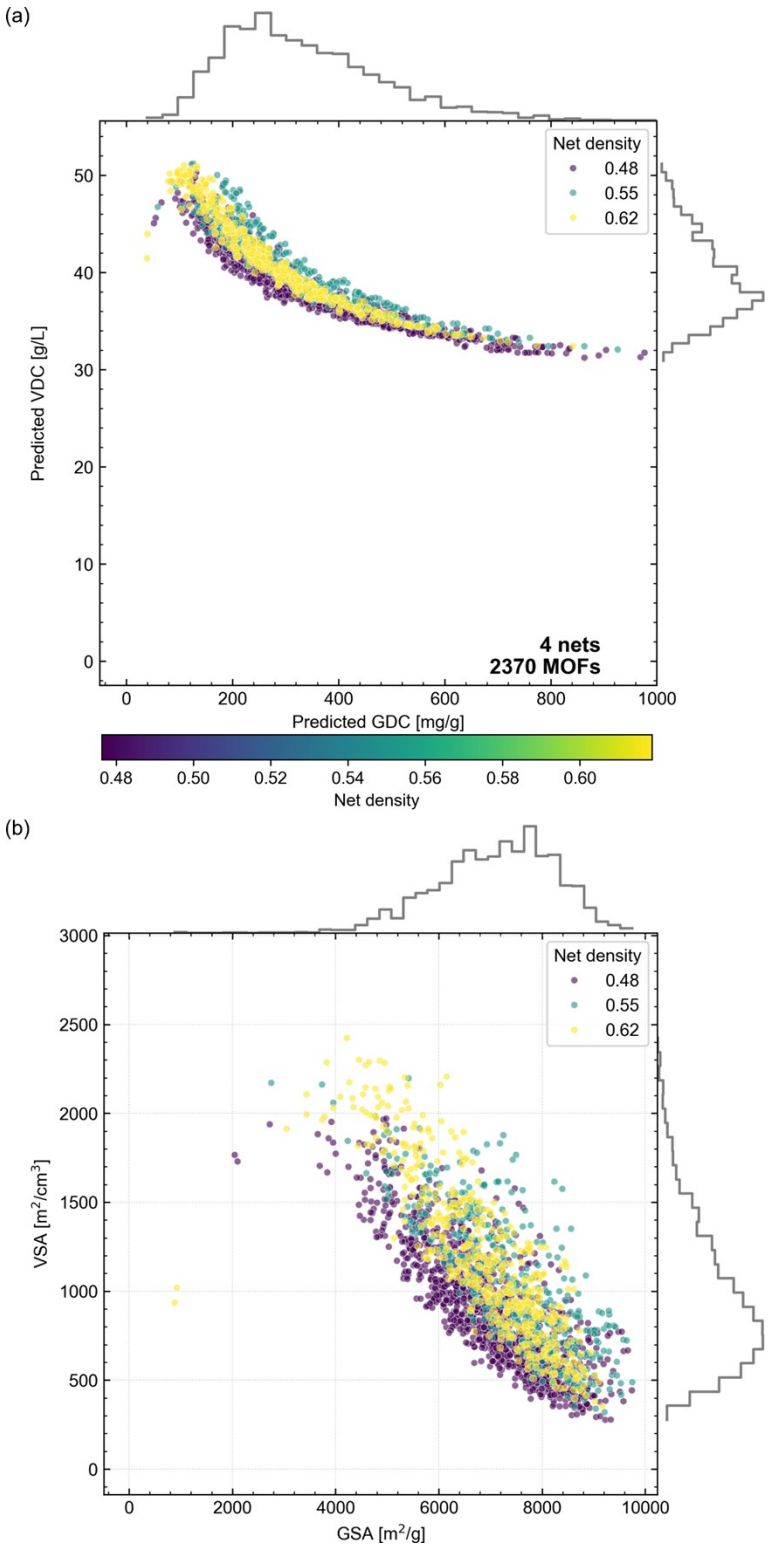


Figure S44. Bivariate plots of the MOFs with the 3,4-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

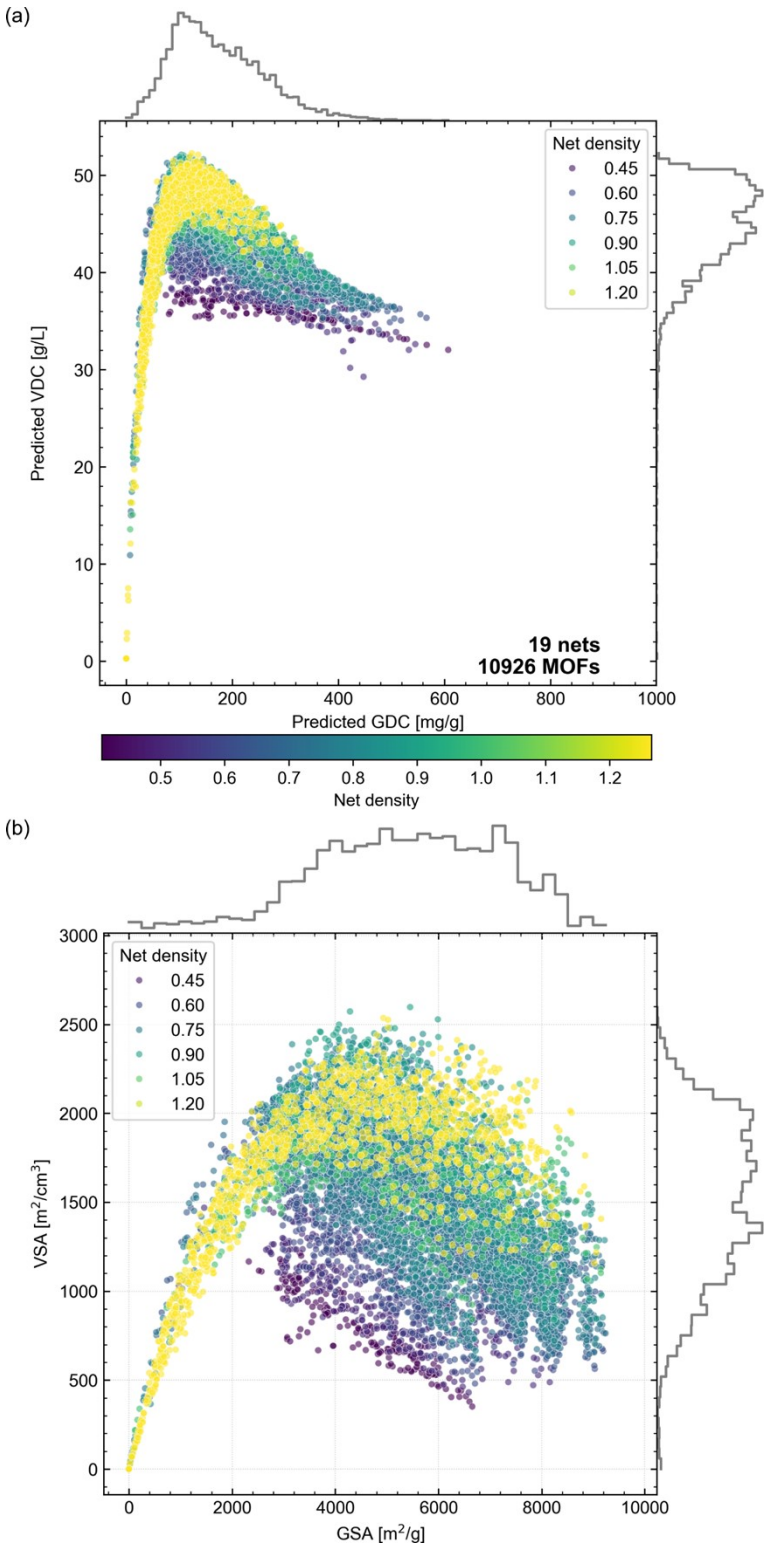


Figure S45. Bivariate plots of the MOFs with the 3,6-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

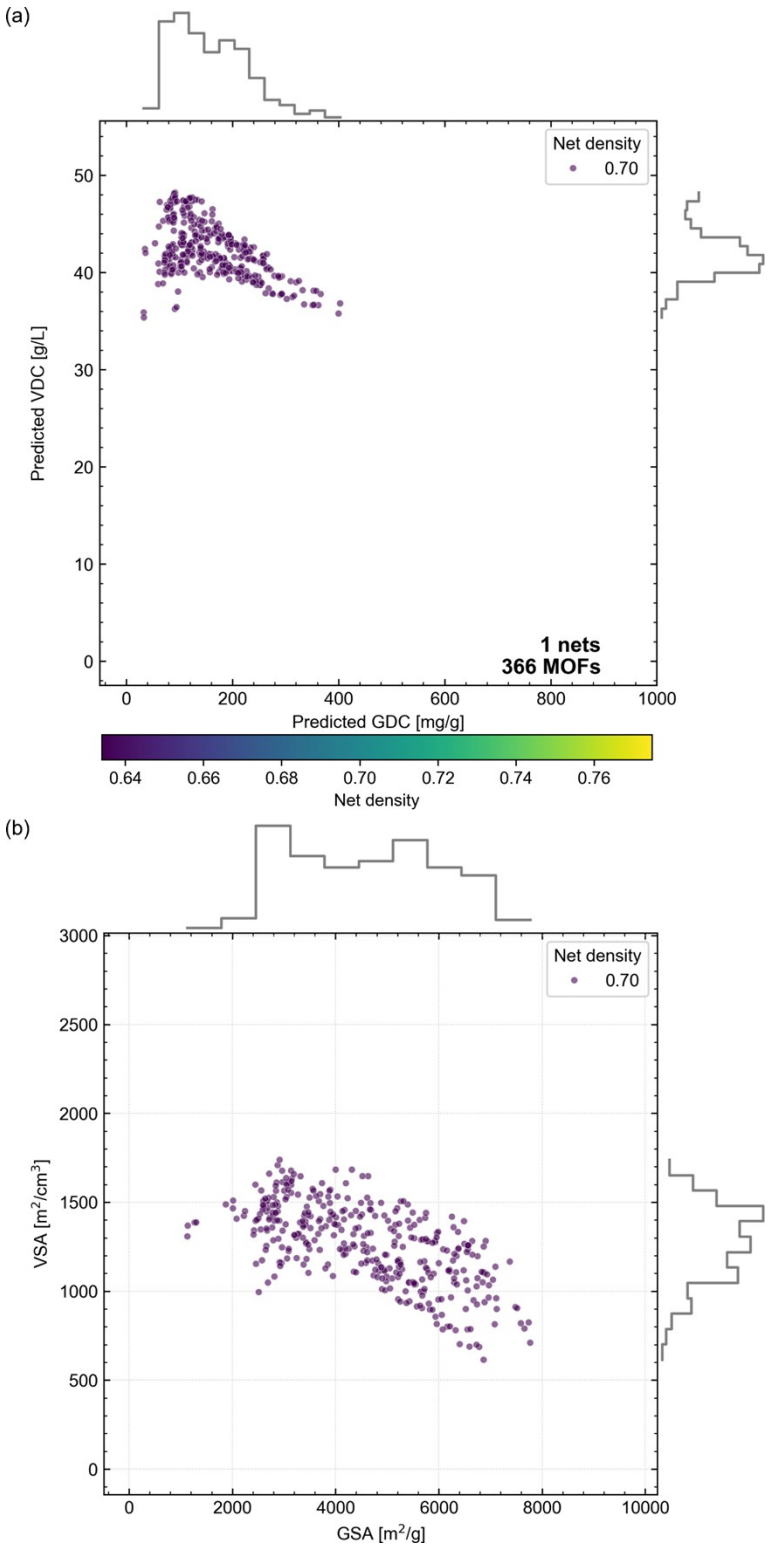


Figure S46. Bivariate plots of the MOFs with the 3,8-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

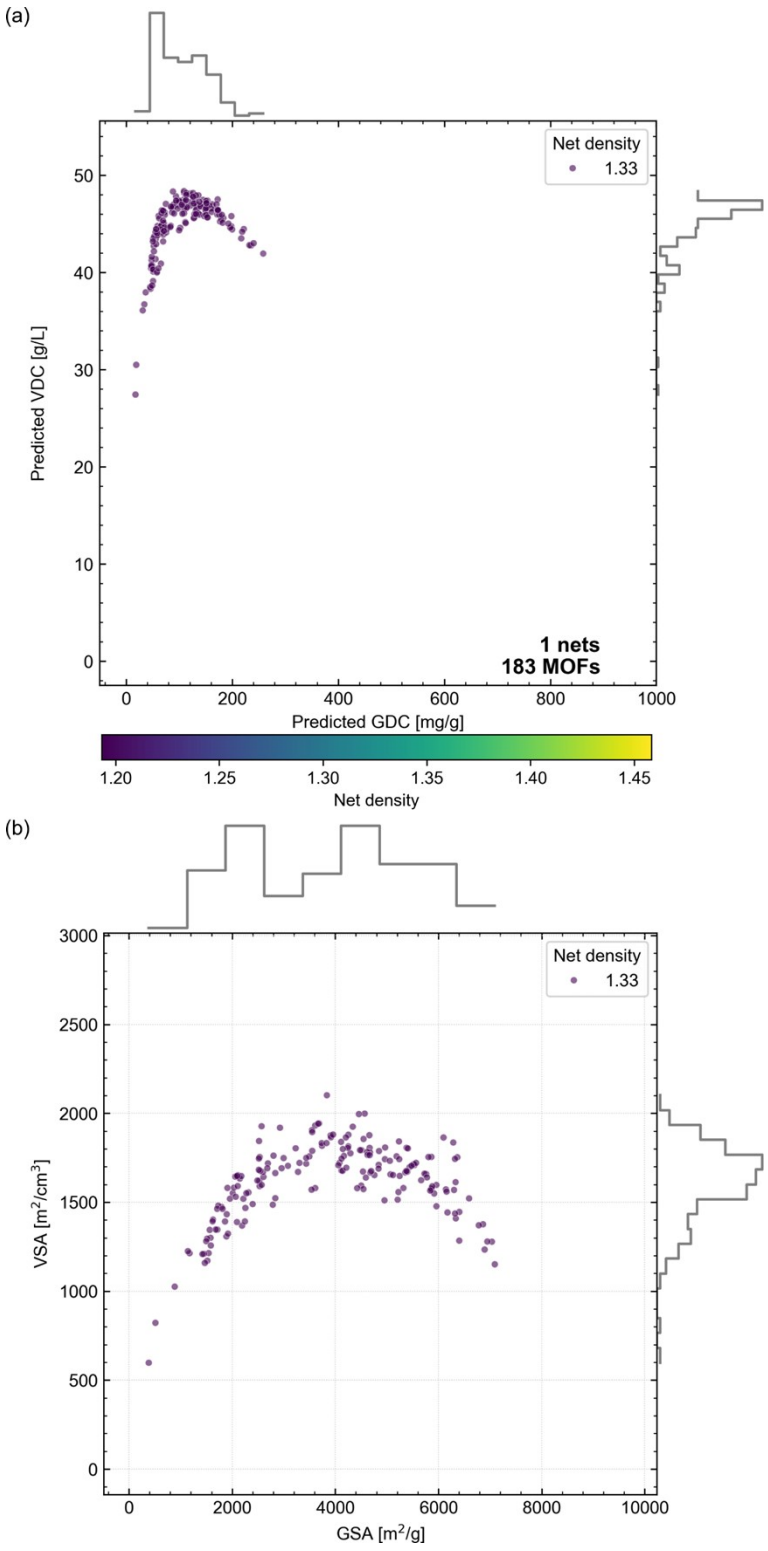


Figure S47. Bivariate plots of the MOFs with the 3,12-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

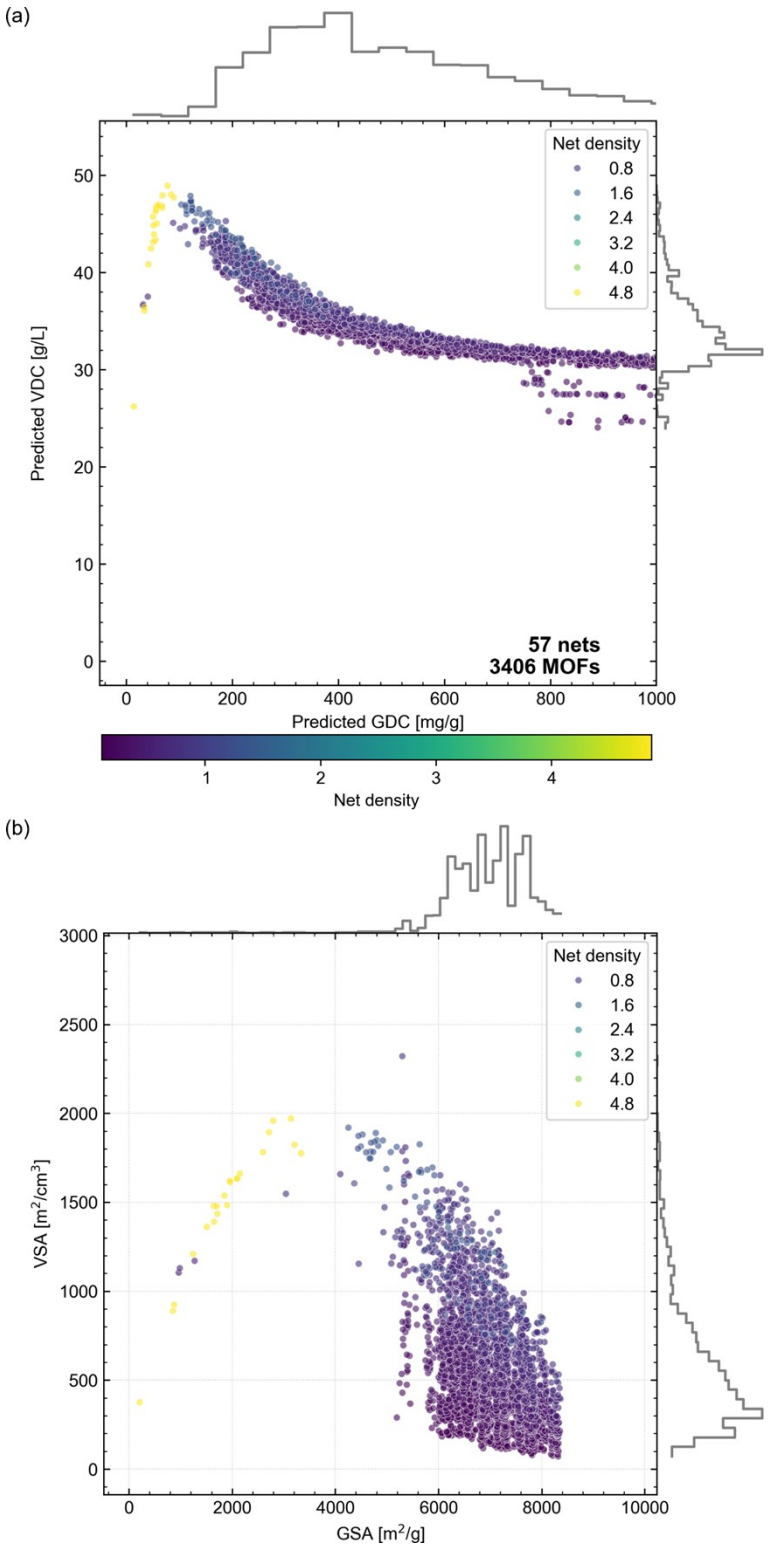


Figure S48. Bivariate plots of the MOFs with the 3-c net (uninodal net), colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

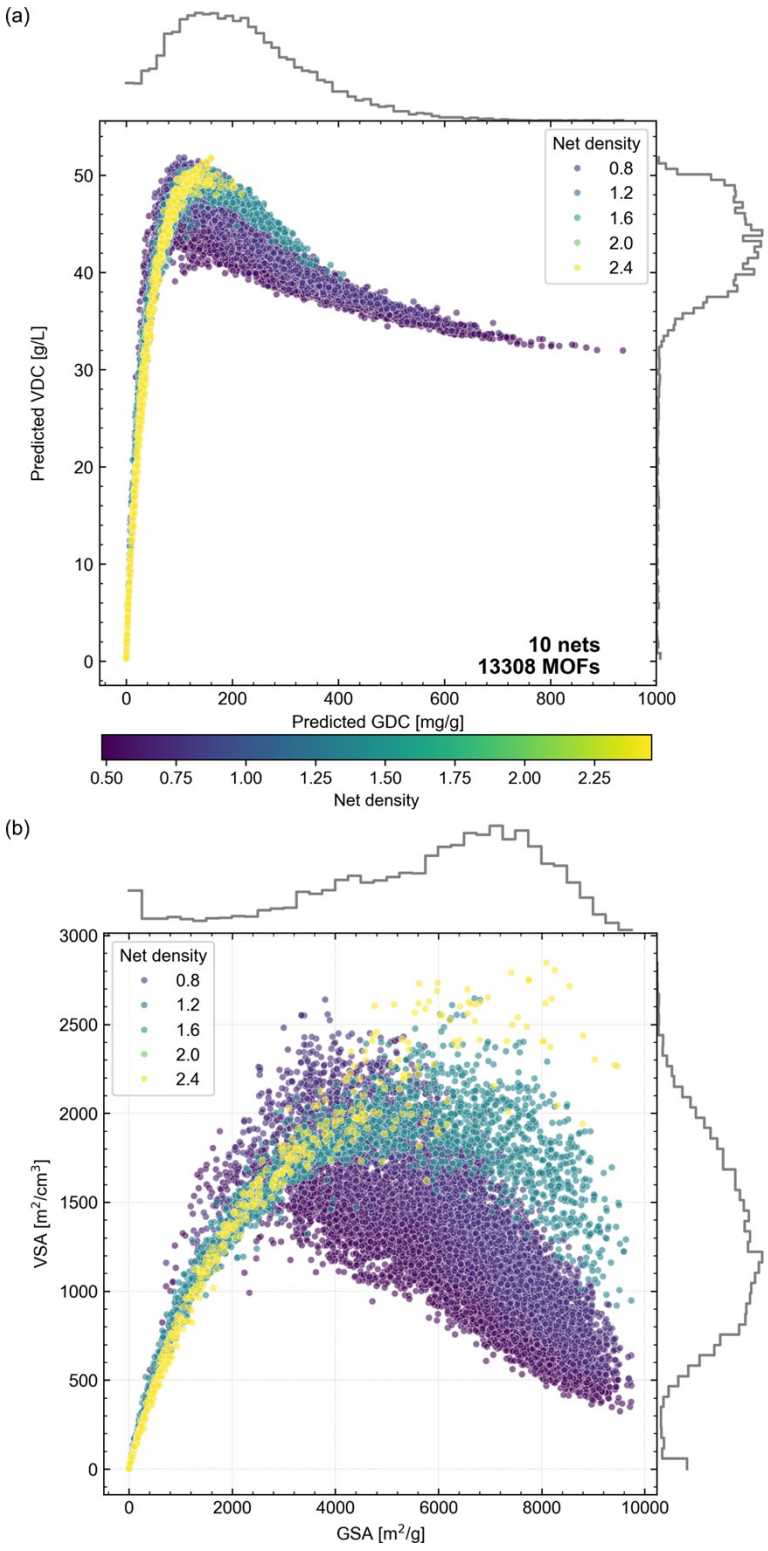


Figure S49. Bivariate plots of the MOFs with the 4,4-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

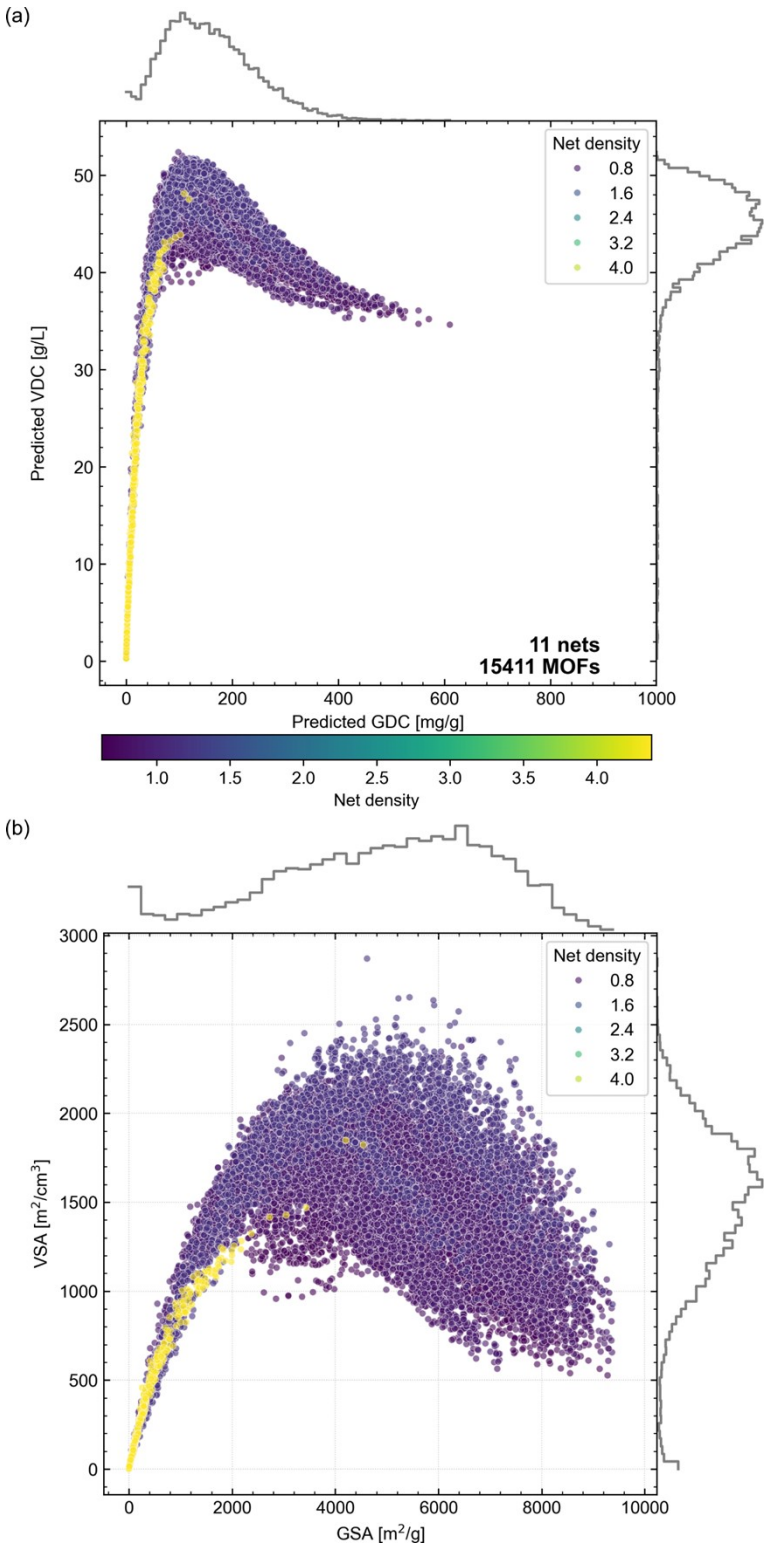


Figure S50. Bivariate plots of the MOFs with the 4,6-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

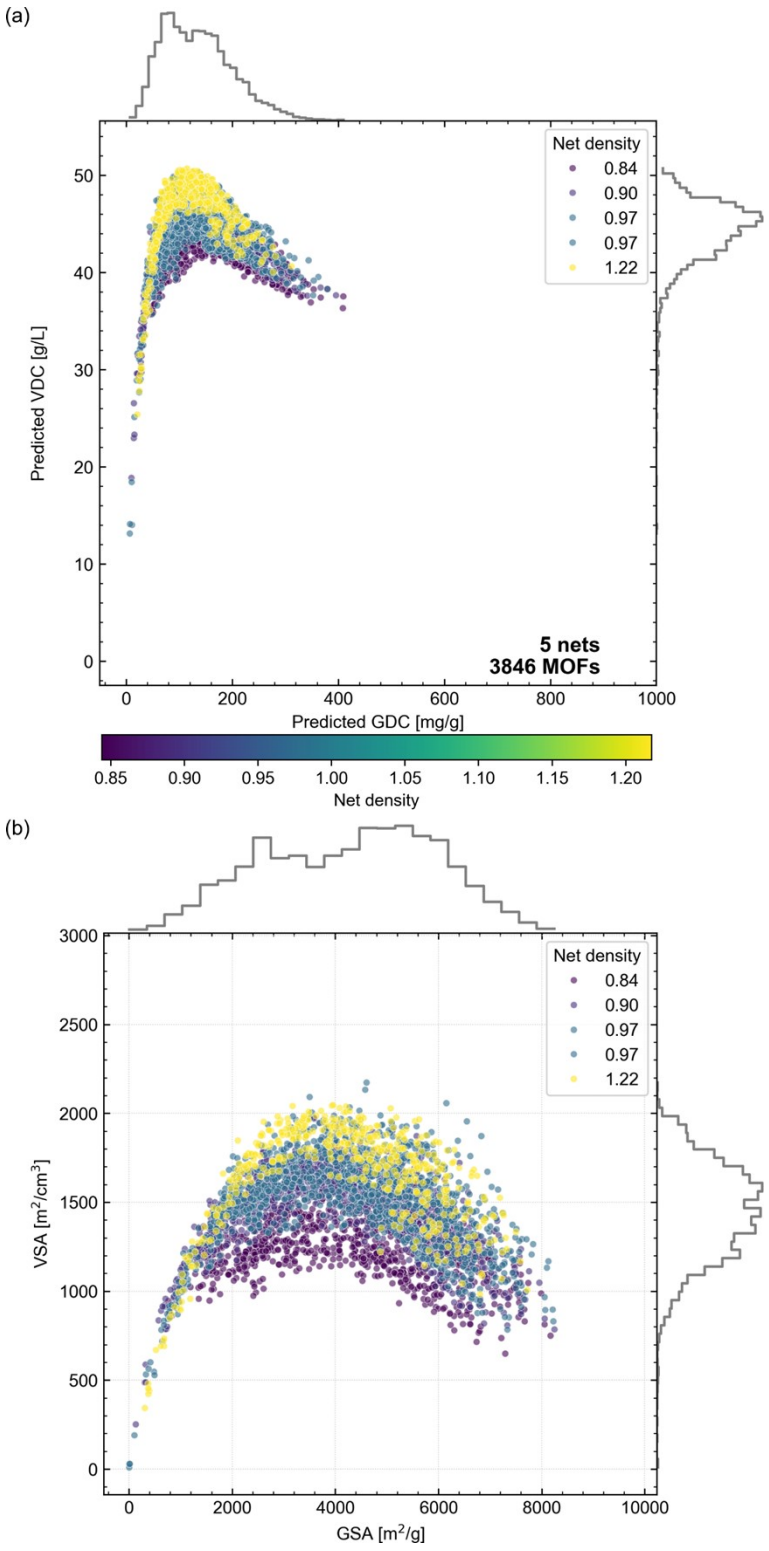


Figure S51. Bivariate plots of the MOFs with the 4,8-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

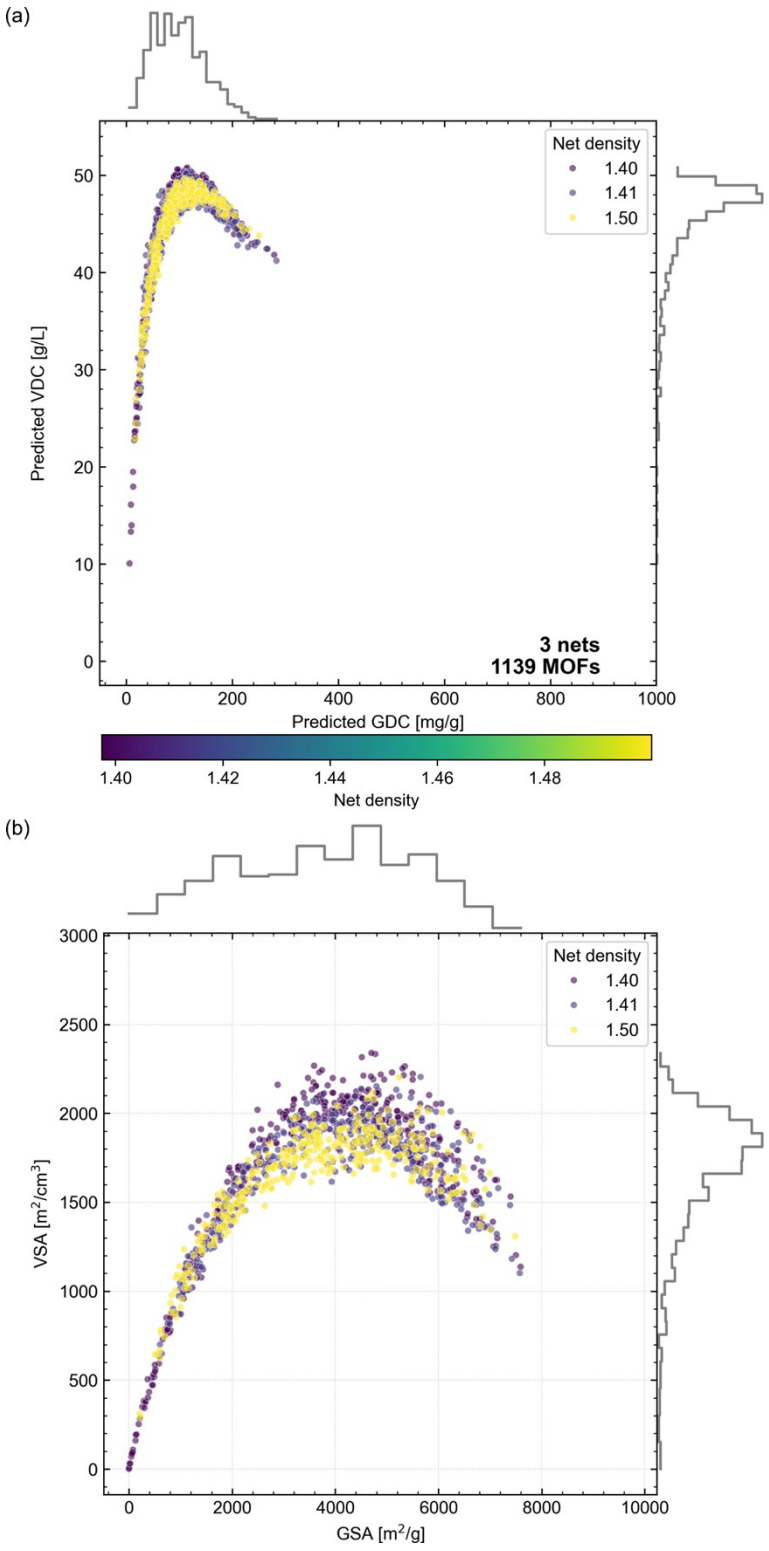


Figure S52. Bivariate plots of the MOFs with the 4,12-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

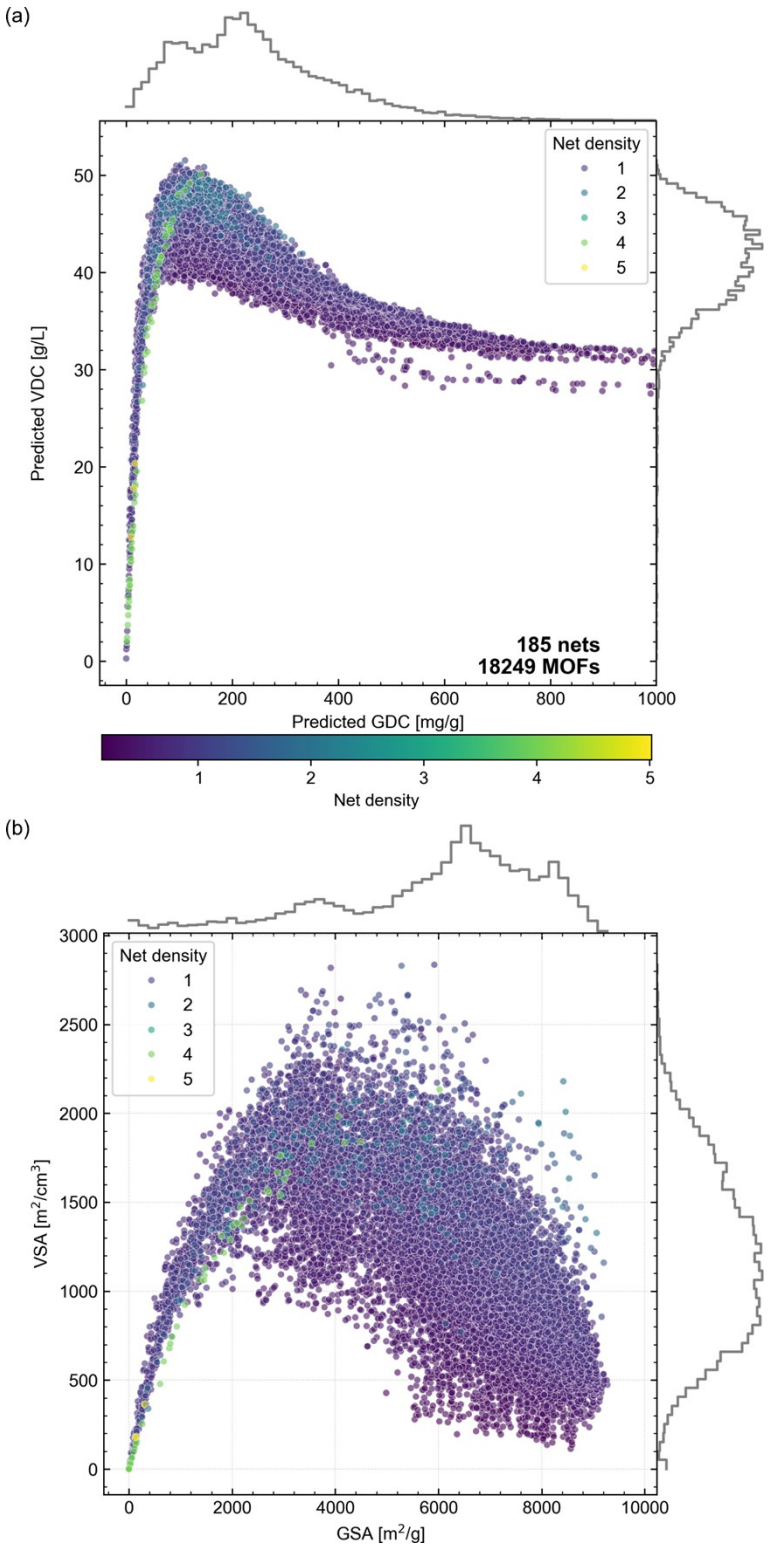


Figure S53. Bivariate plots of the MOFs with the 4-c net (uninodal), colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

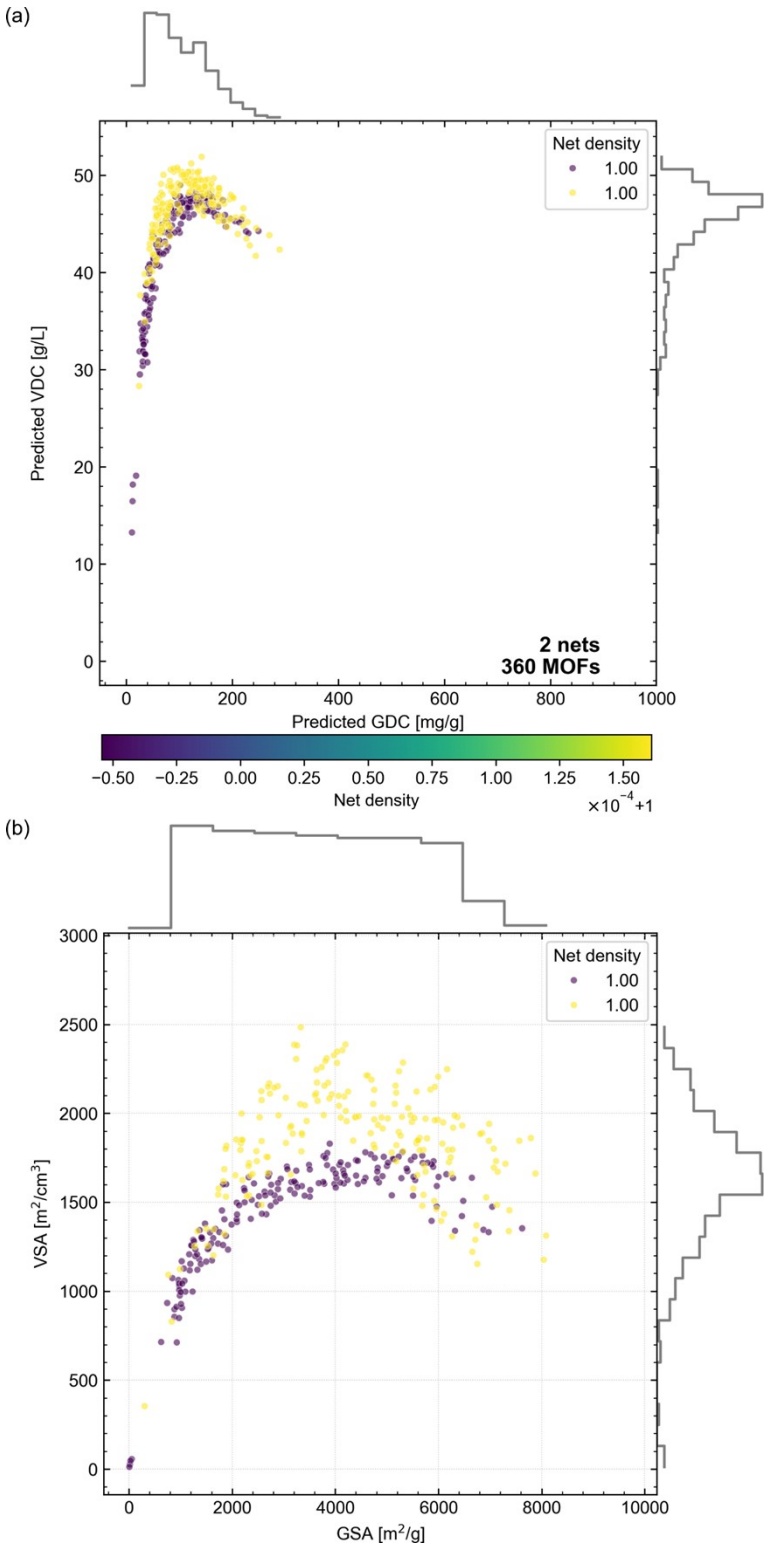


Figure S54. Bivariate plots of the MOFs with the 6,6-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

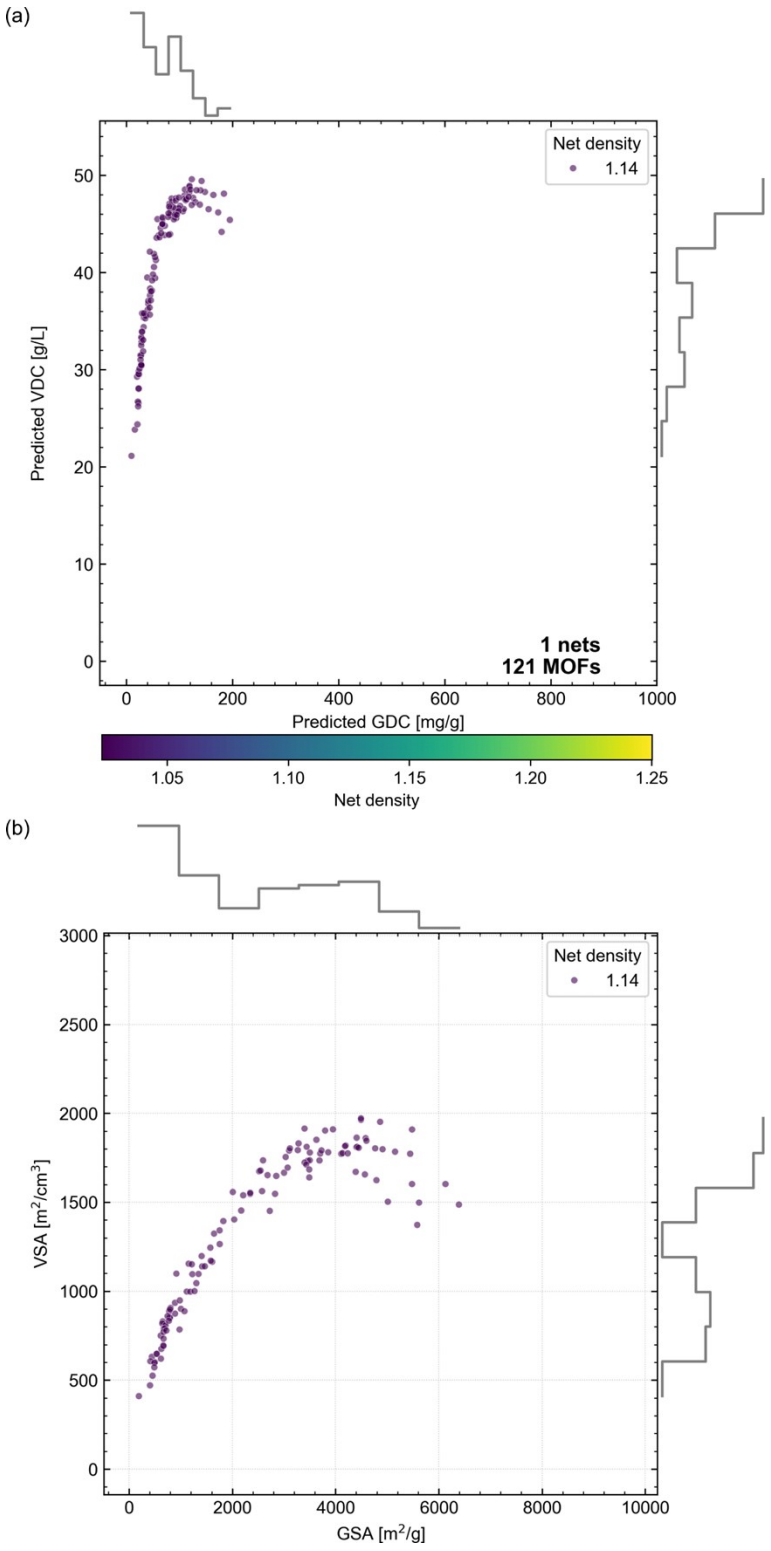


Figure S55. Bivariate plots of the MOFs with the 6,8-c net, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

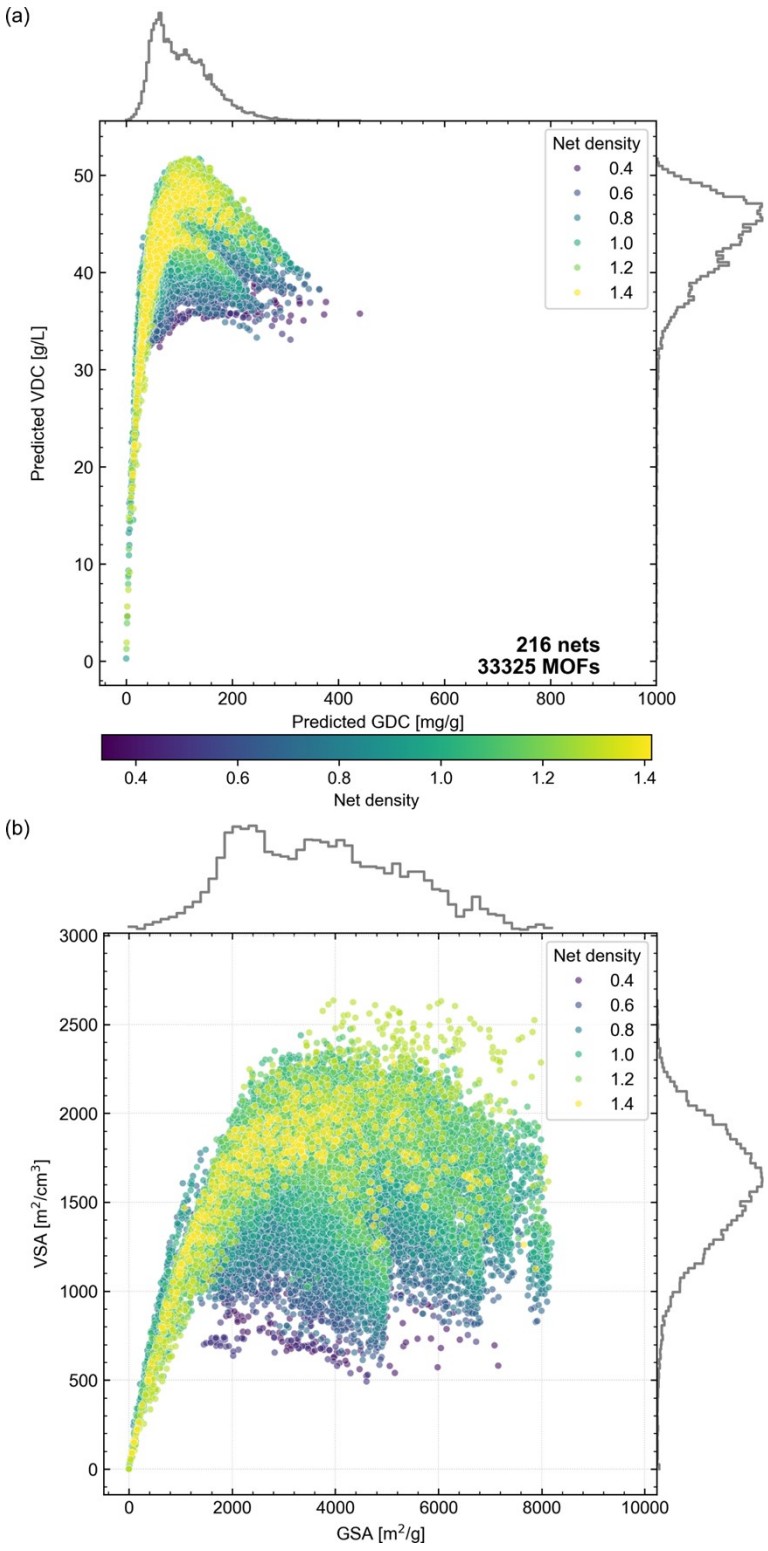


Figure S56. Bivariate plots of the MOFs with the 6-c net (uninodal), colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

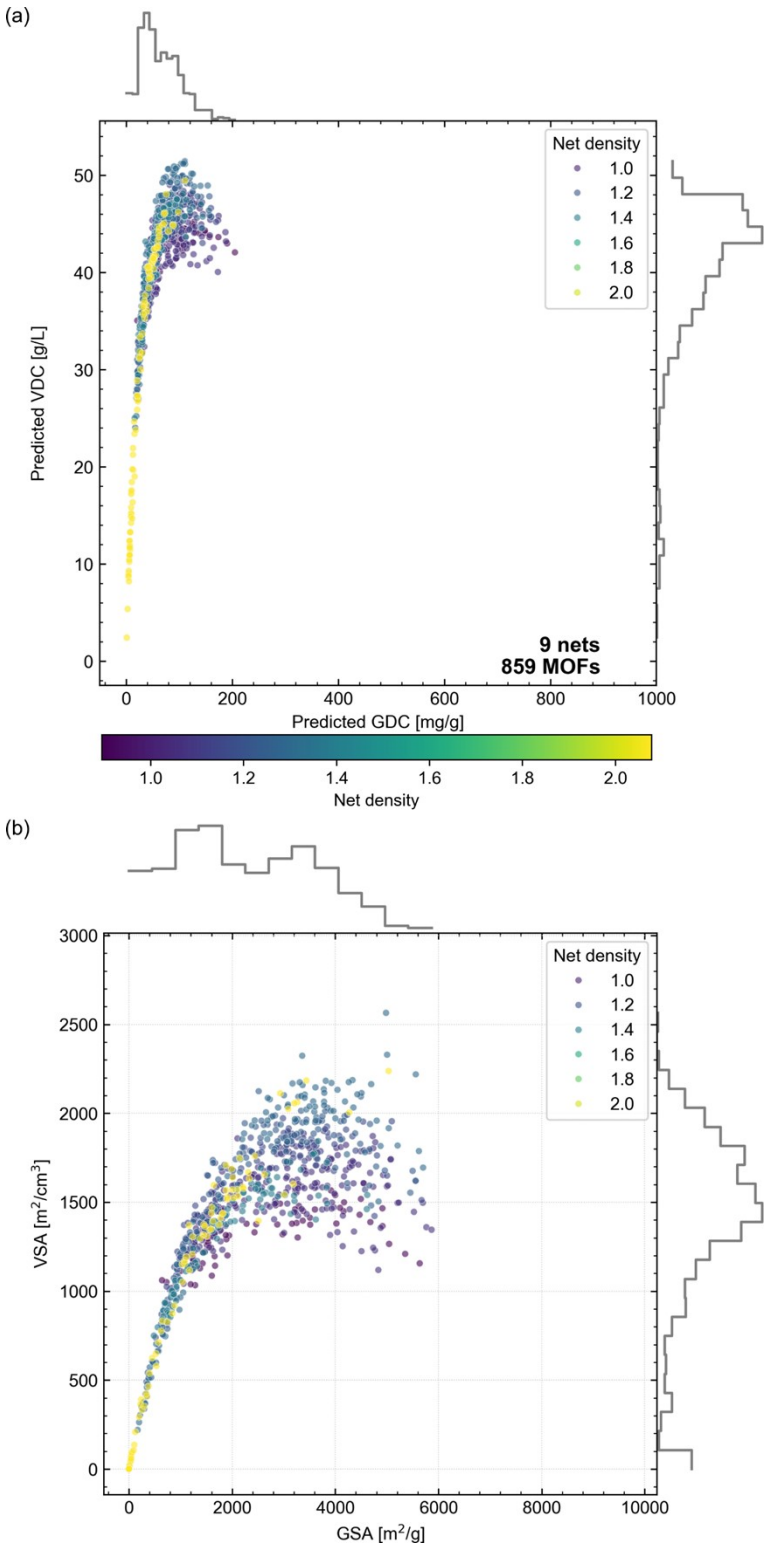


Figure S57. Bivariate plots of the MOFs with the 8-c net (uninodal), colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

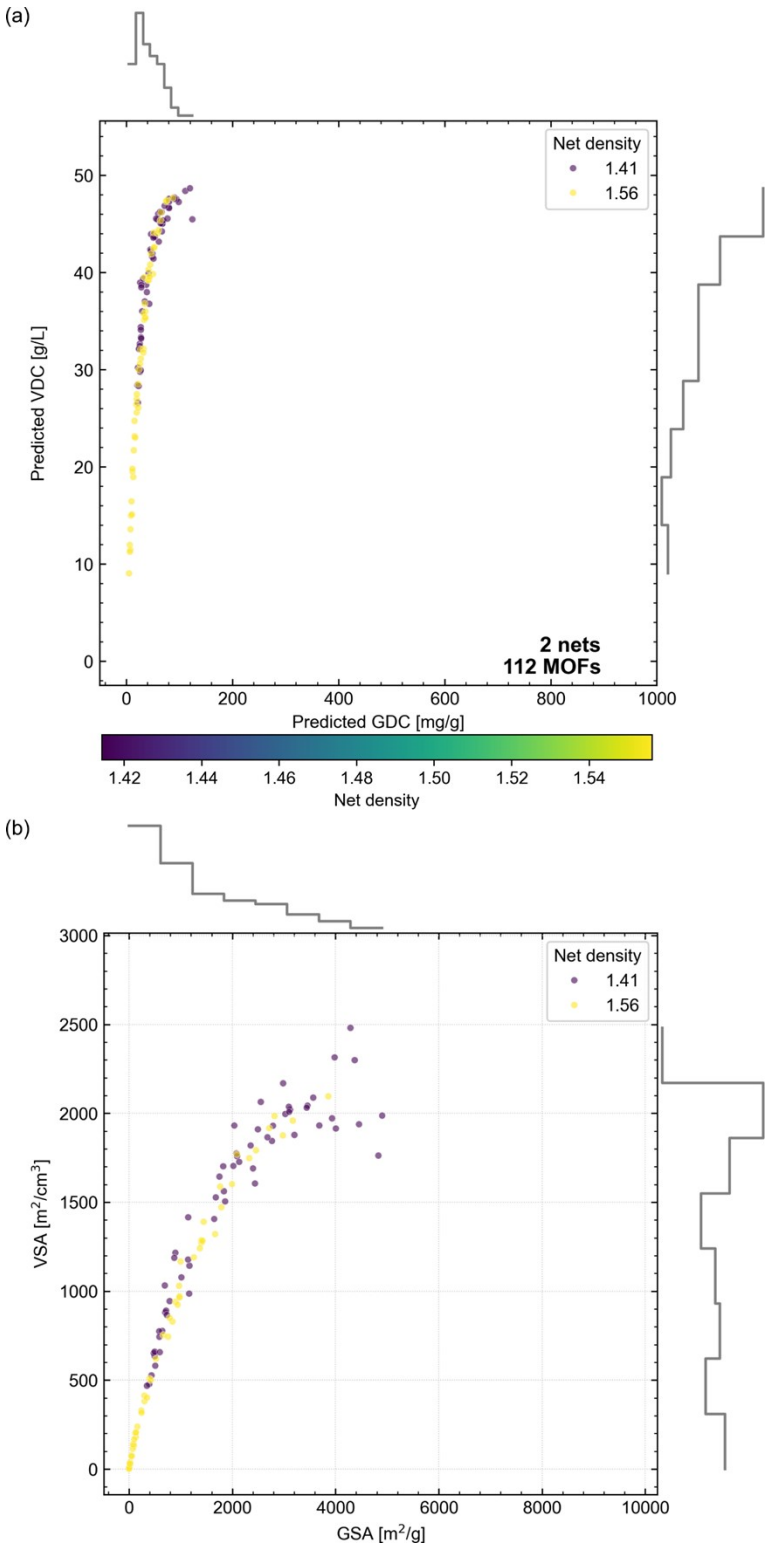


Figure S58. Bivariate plots of the MOFs with the 12-c net (uninodal), colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative net density values to facilitate interpretation.

Subgroups divided by the metal building blocks

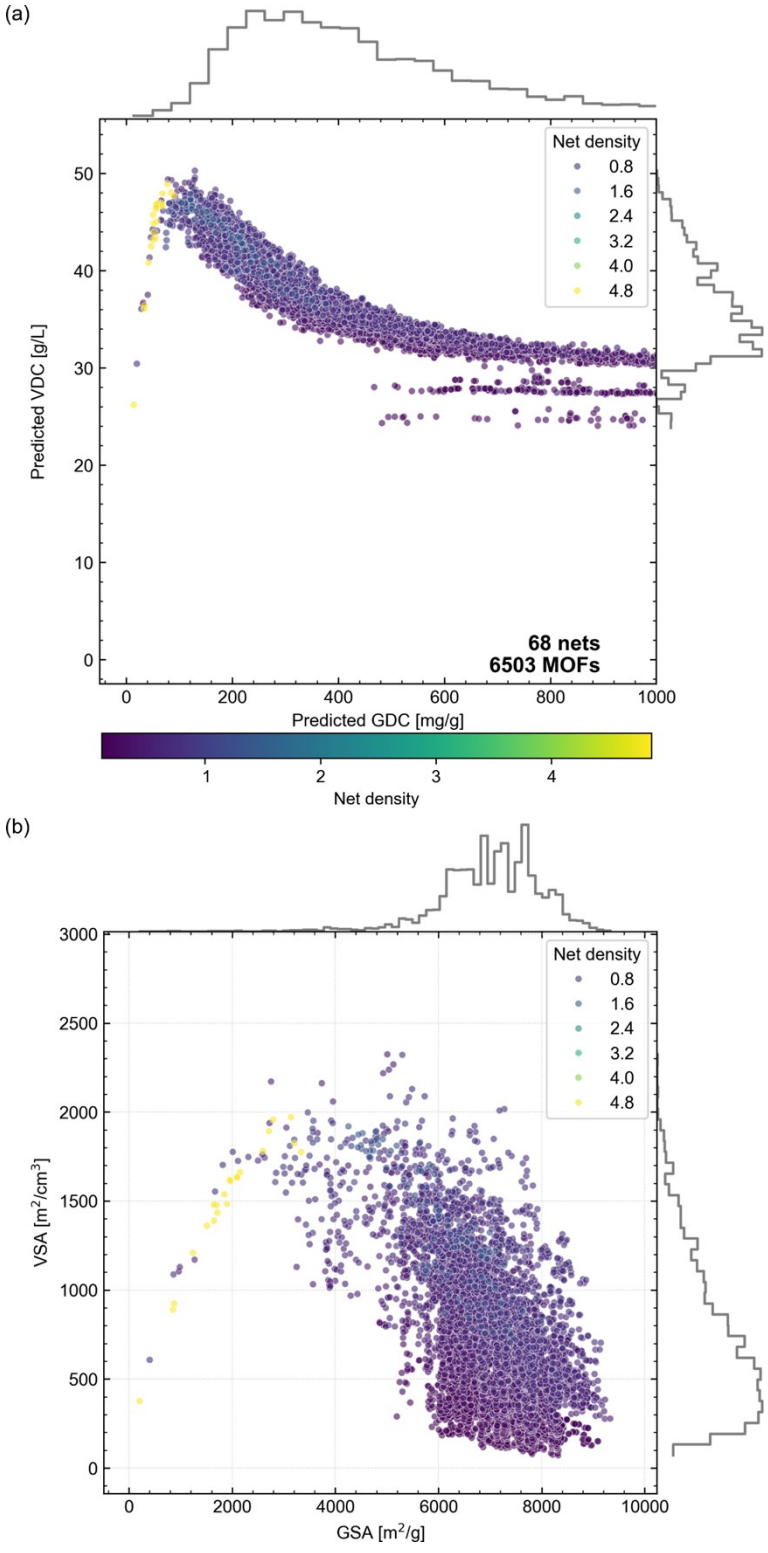


Figure S59. Bivariate plots of the MOFs with the 3-c mc1 cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

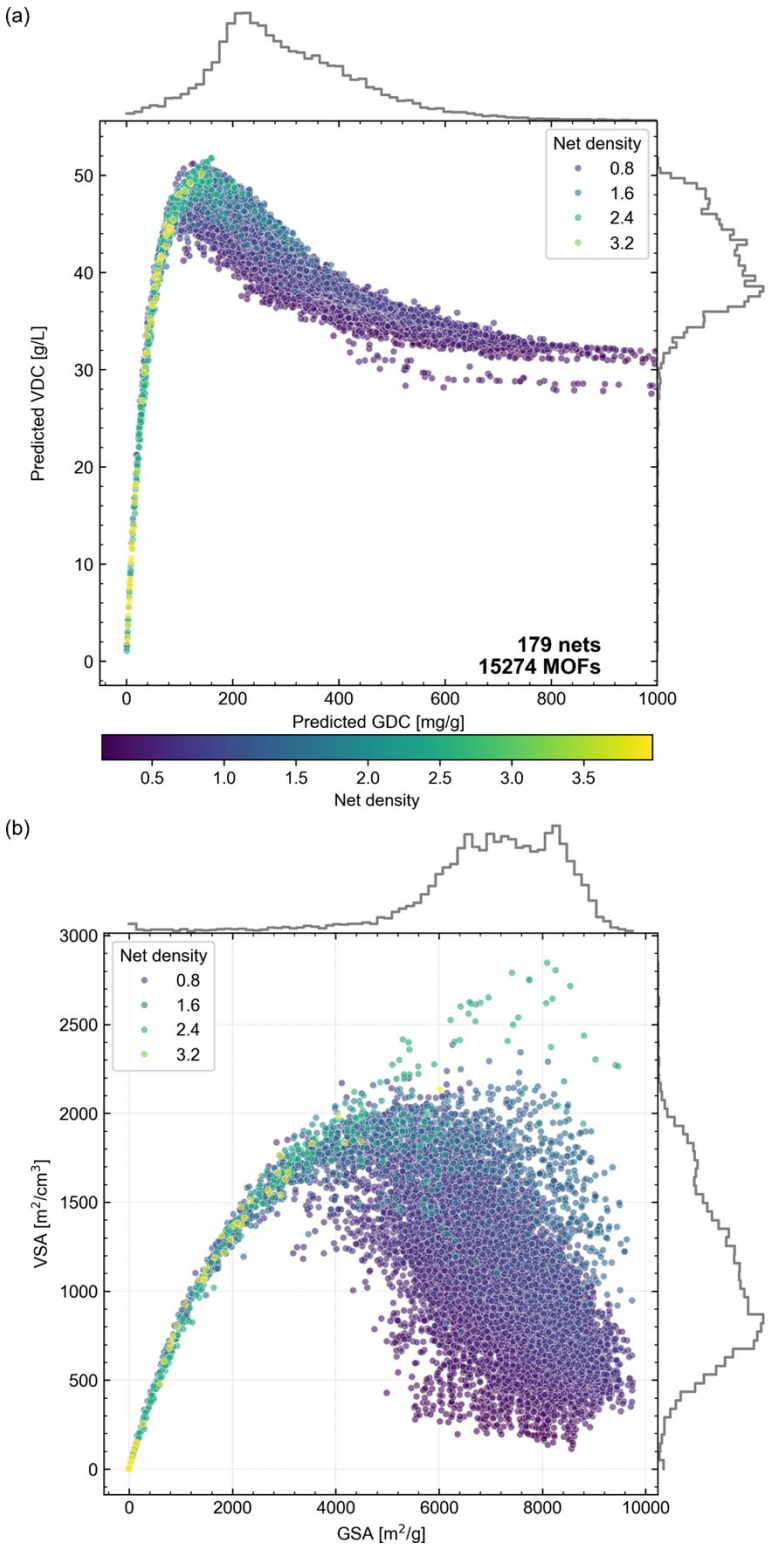


Figure S60. Bivariate plots of the MOFs with the 4-c mc2 cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

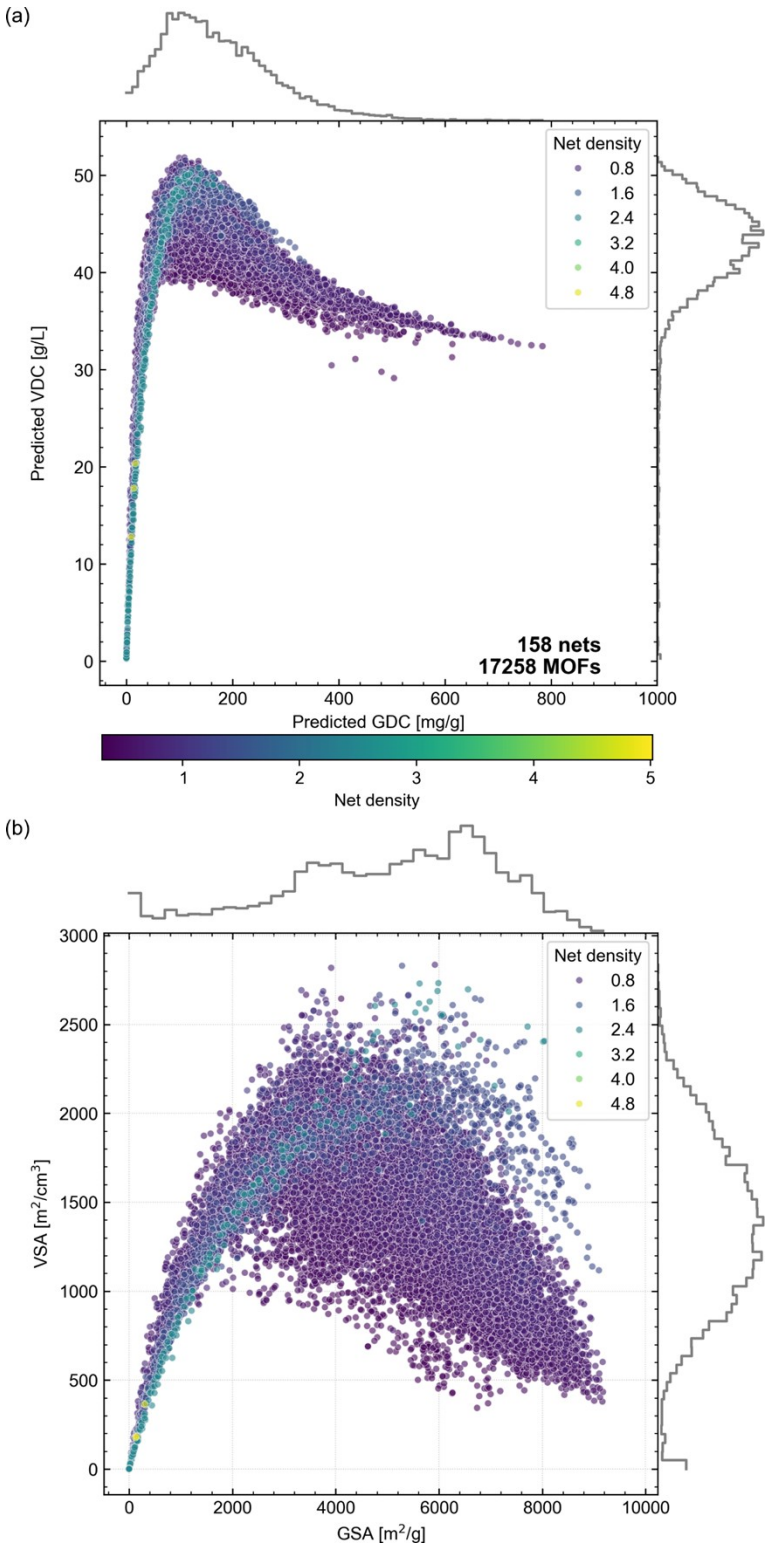


Figure S61. Bivariate plots of the MOFs with the 4-c mc3 cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

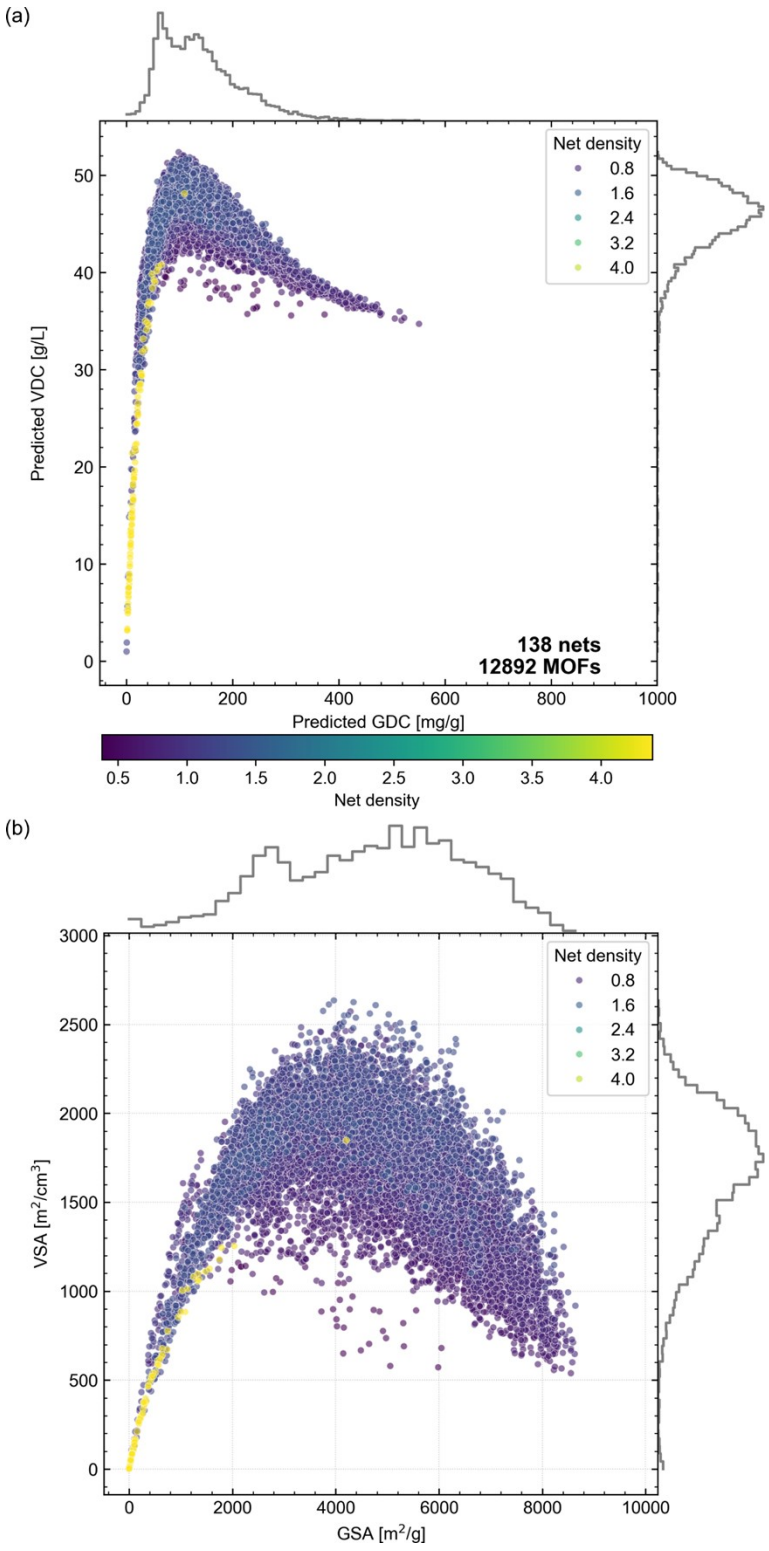


Figure S62. Bivariate plots of the MOFs with the 6-c mc4 cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

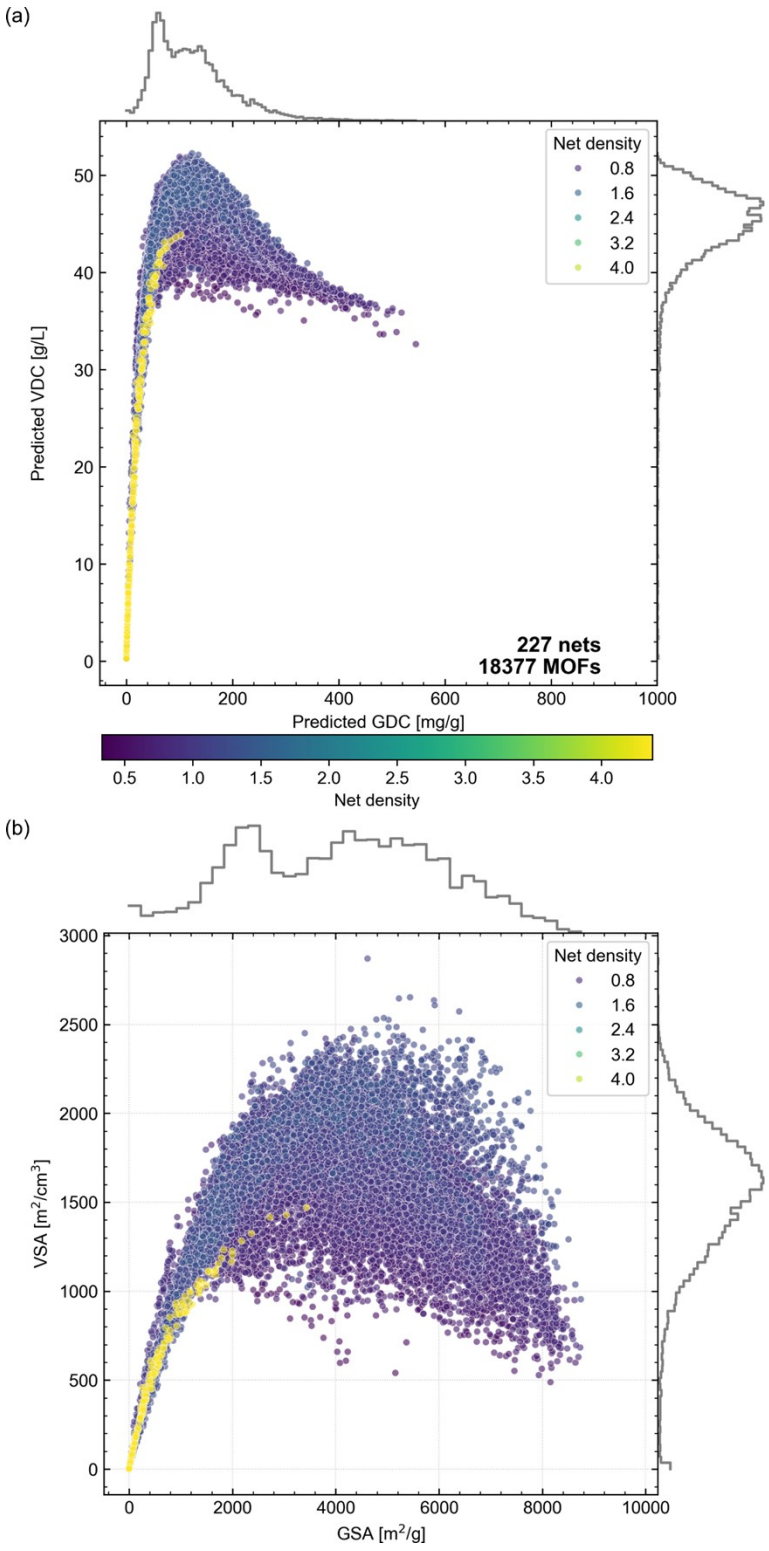


Figure S63. Bivariate plots of the MOFs with the 6-c mc5 cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

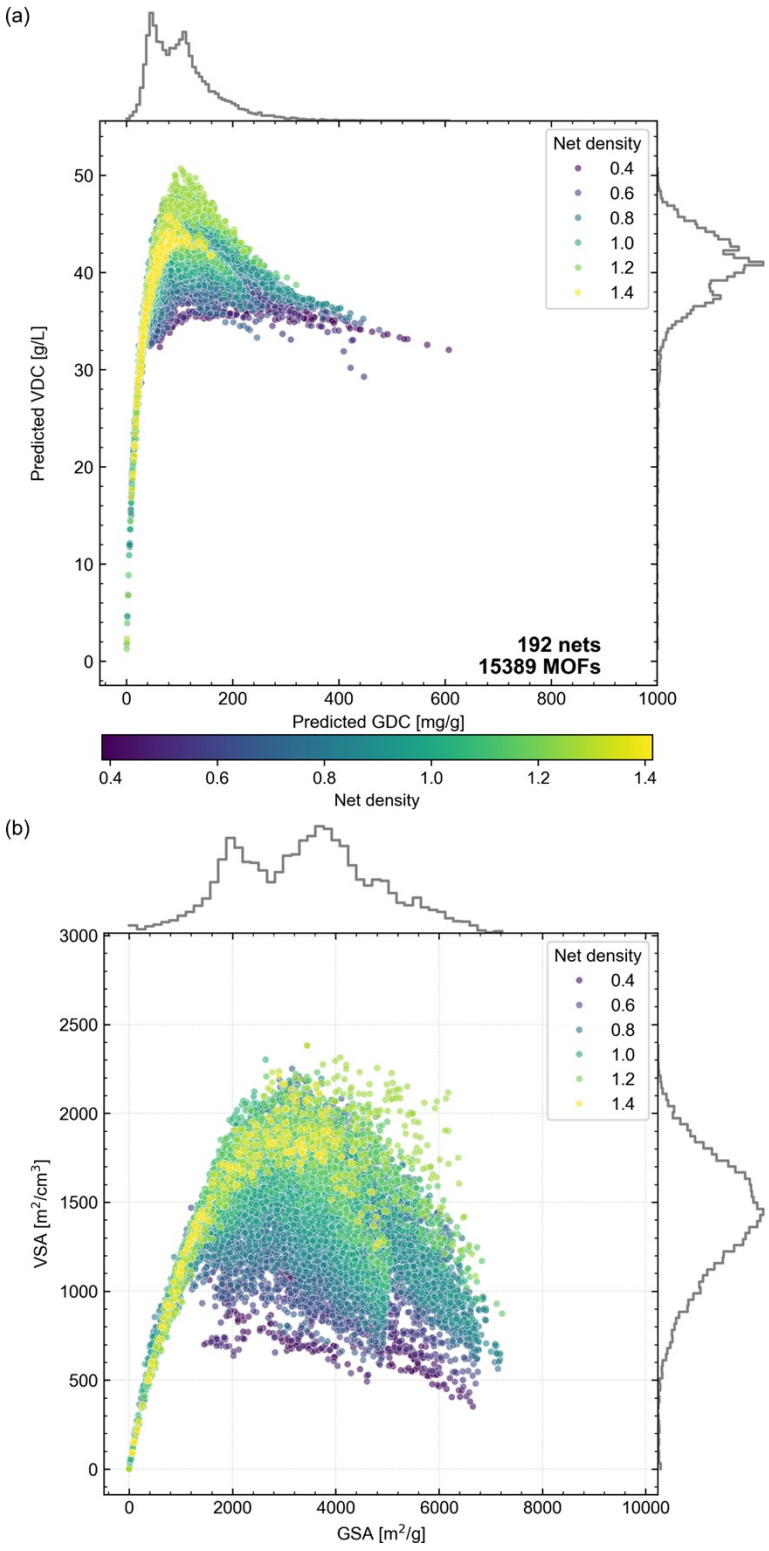


Figure S64. Bivariate plots of the MOFs with the 6-c **mc6** cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

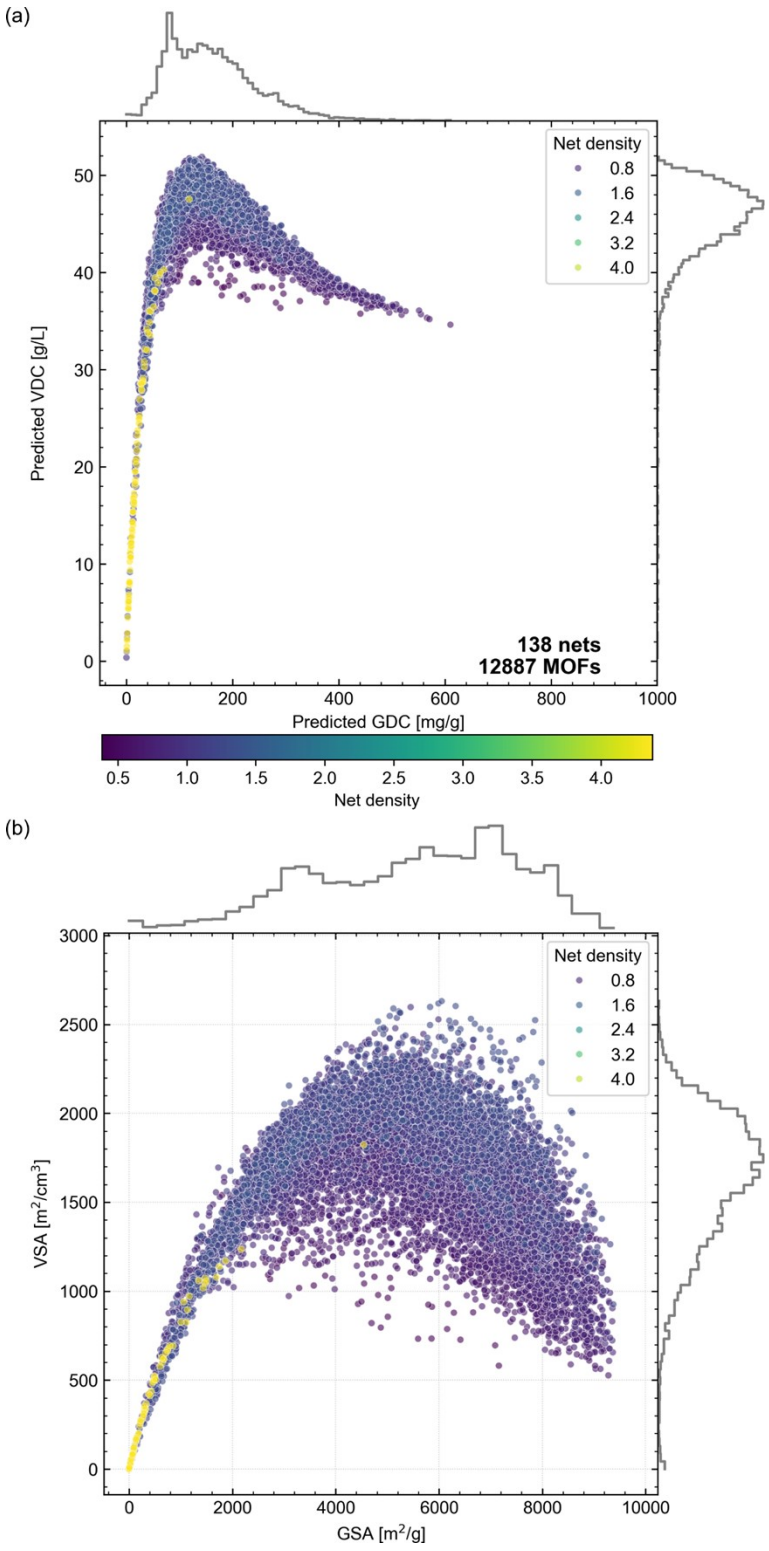


Figure S65. Bivariate plots of the MOFs with the 6-c **mc12** cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

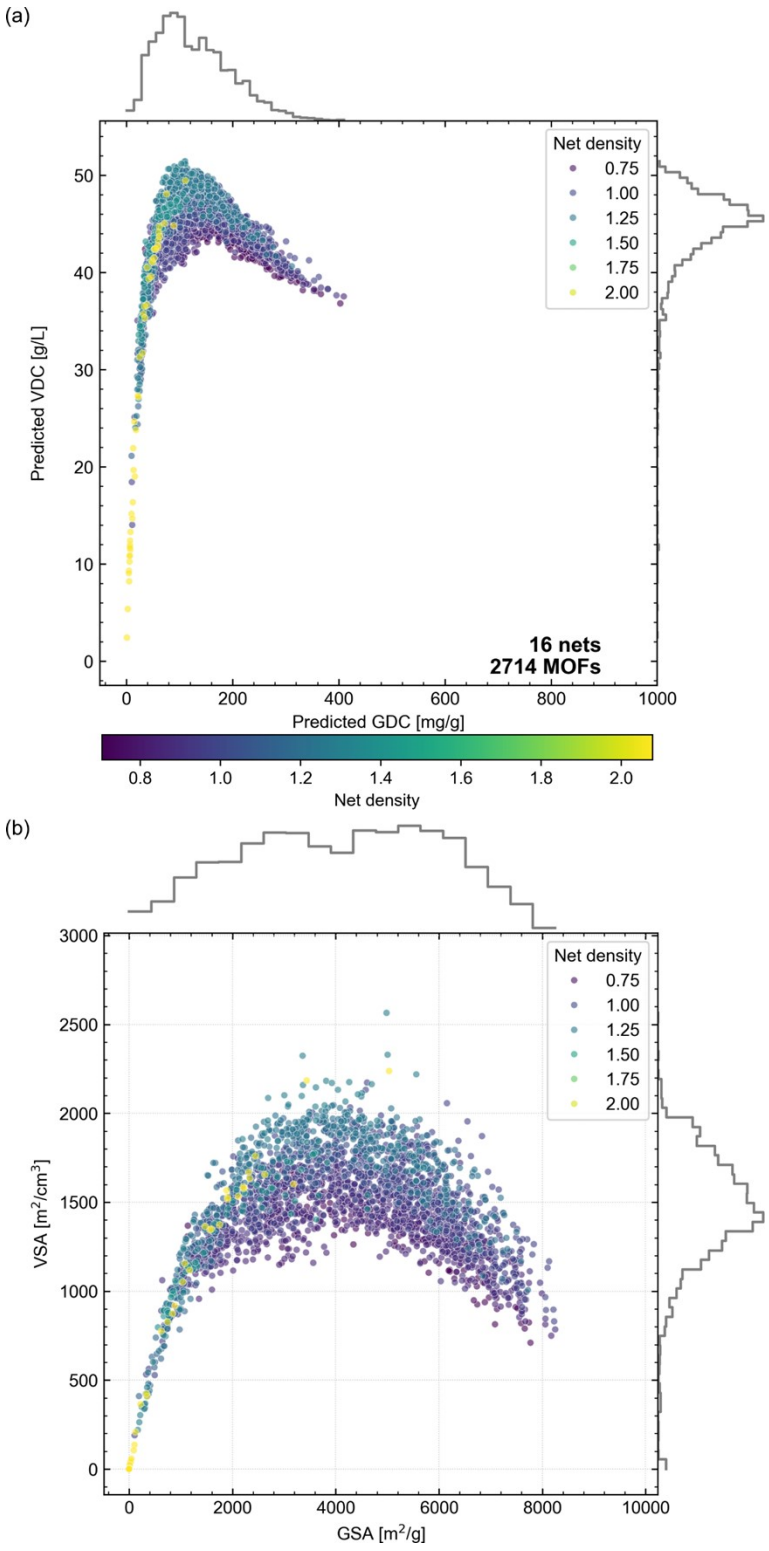


Figure S66. Bivariate plots of the MOFs with the 8-c mc7 cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

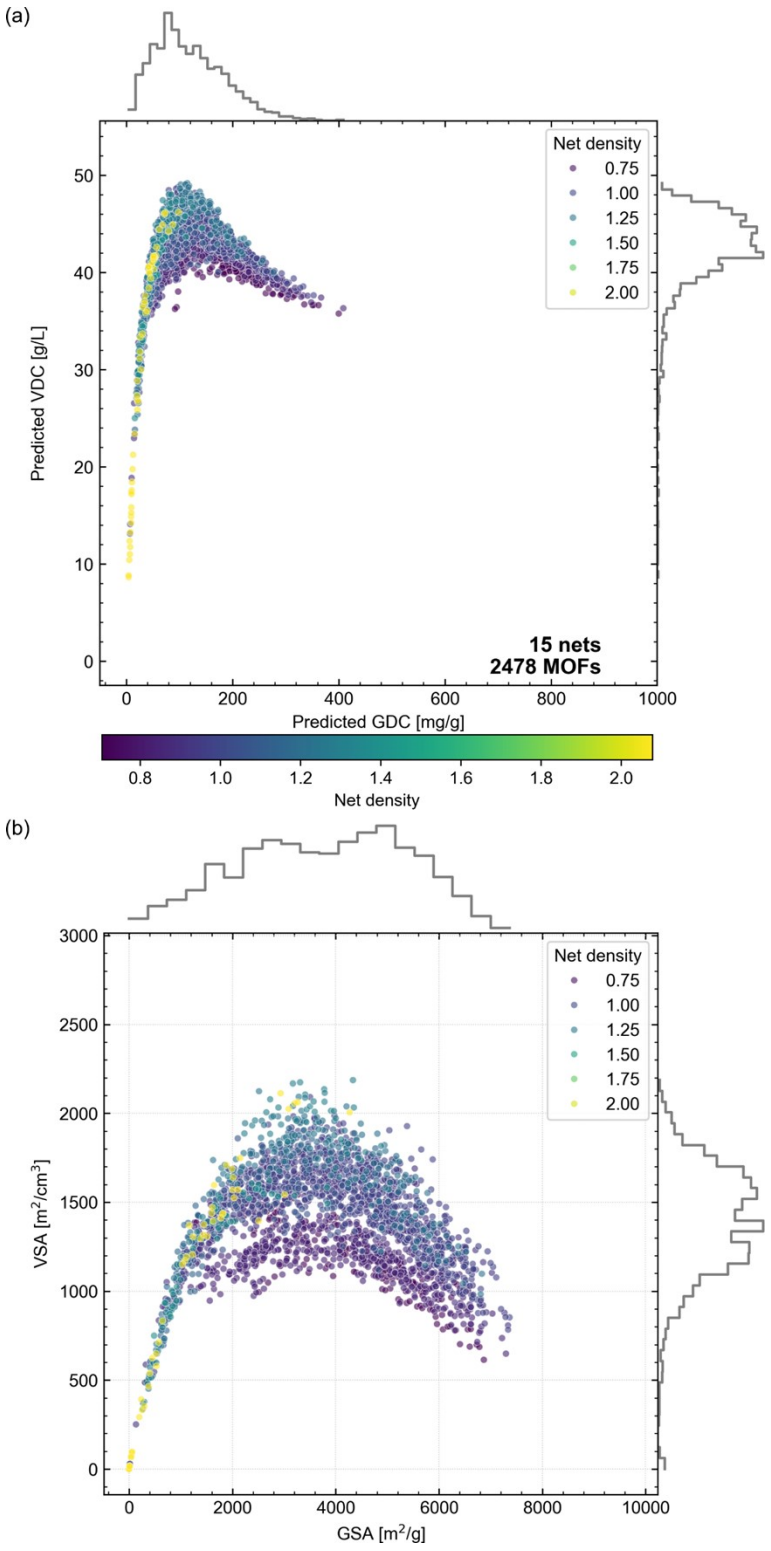


Figure S67. Bivariate plots of the MOFs with the 8-c mc8 cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

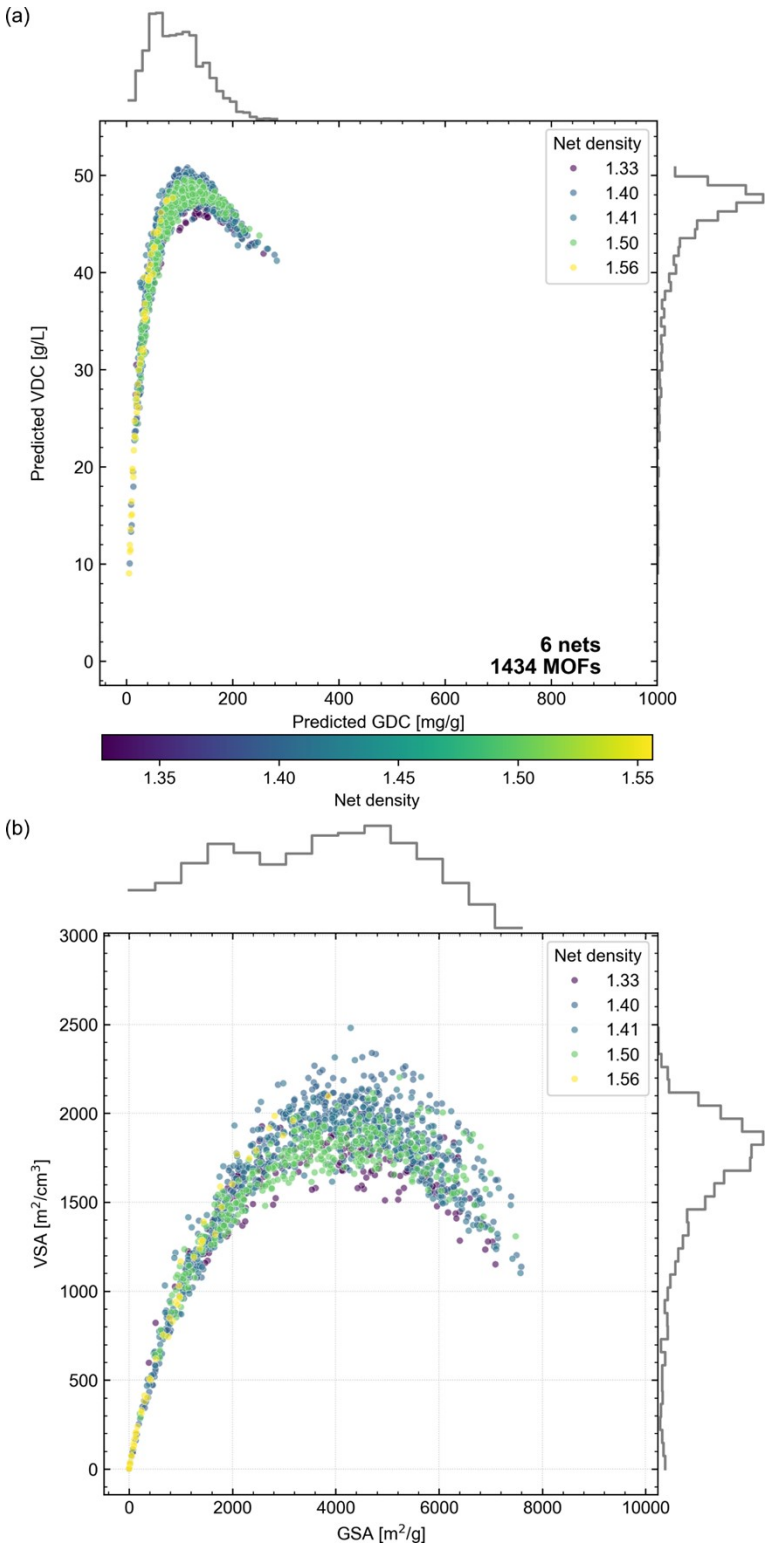


Figure S68. Bivariate plots of the MOFs with the 12-c mc10 cluster, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

Subgroups divided by the organic building blocks

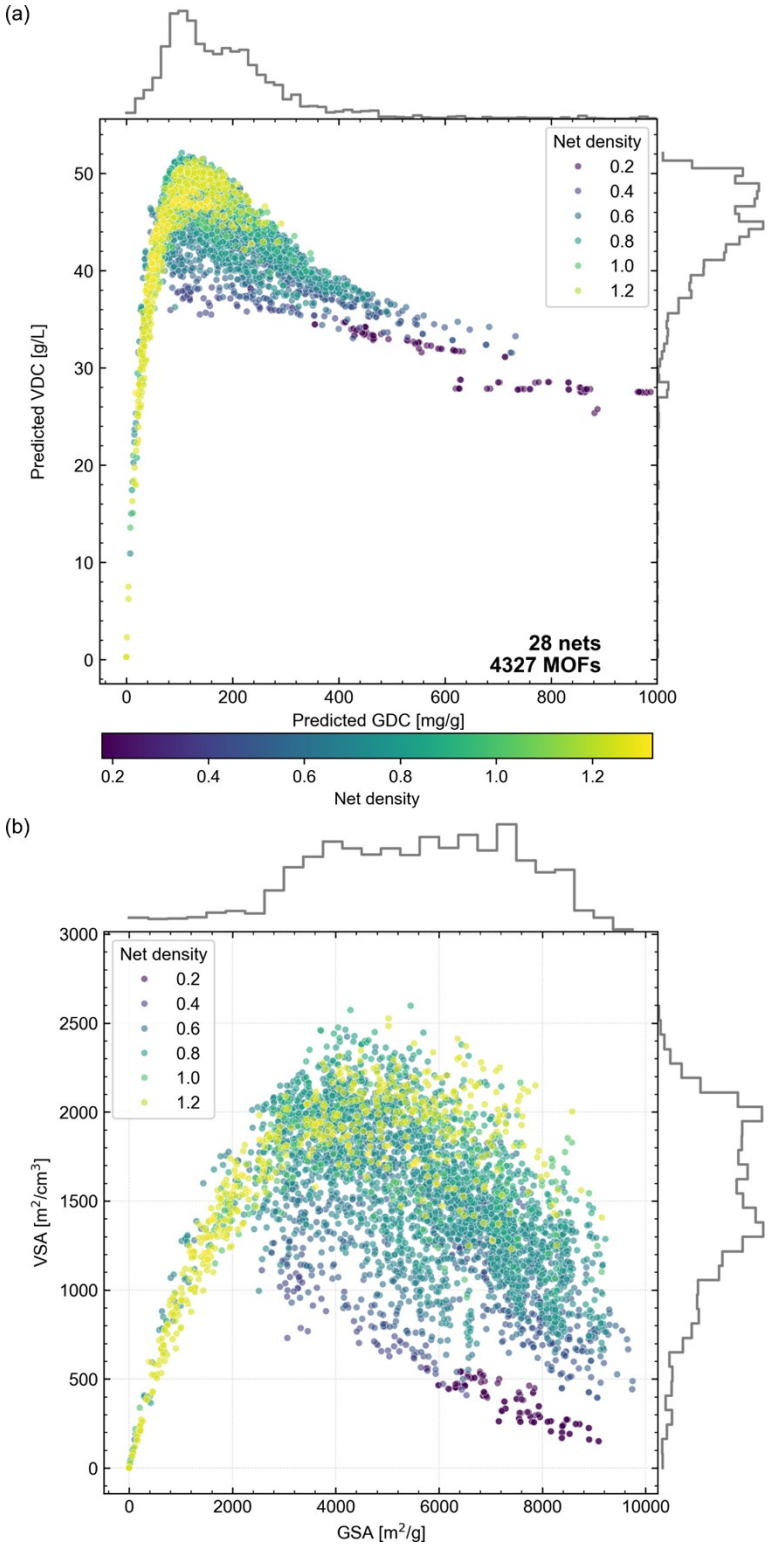


Figure S69. Bivariate plots of the MOFs with the 3-c on1 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

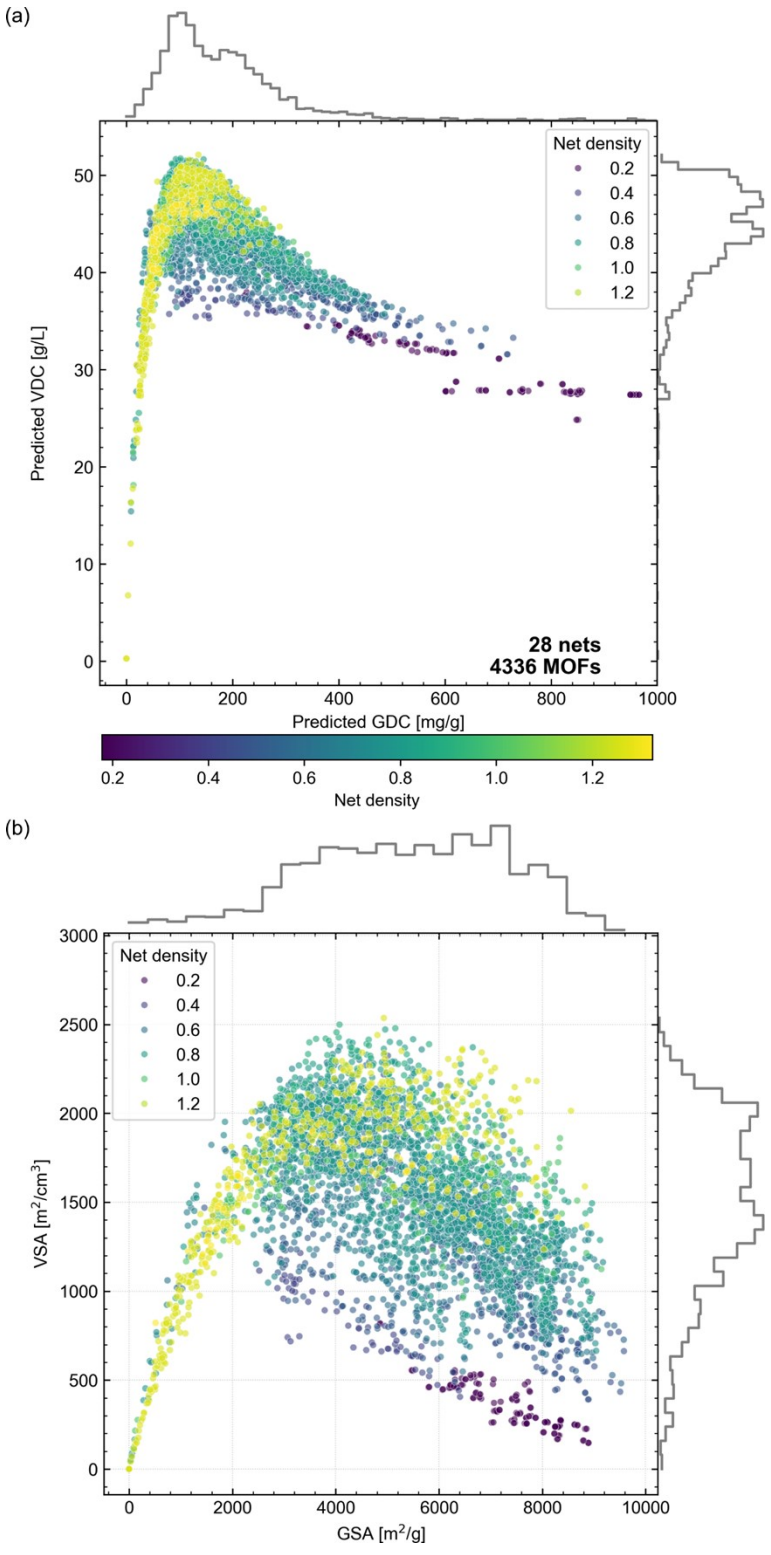


Figure S70. Bivariate plots of the MOFs with the 3-c on2 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

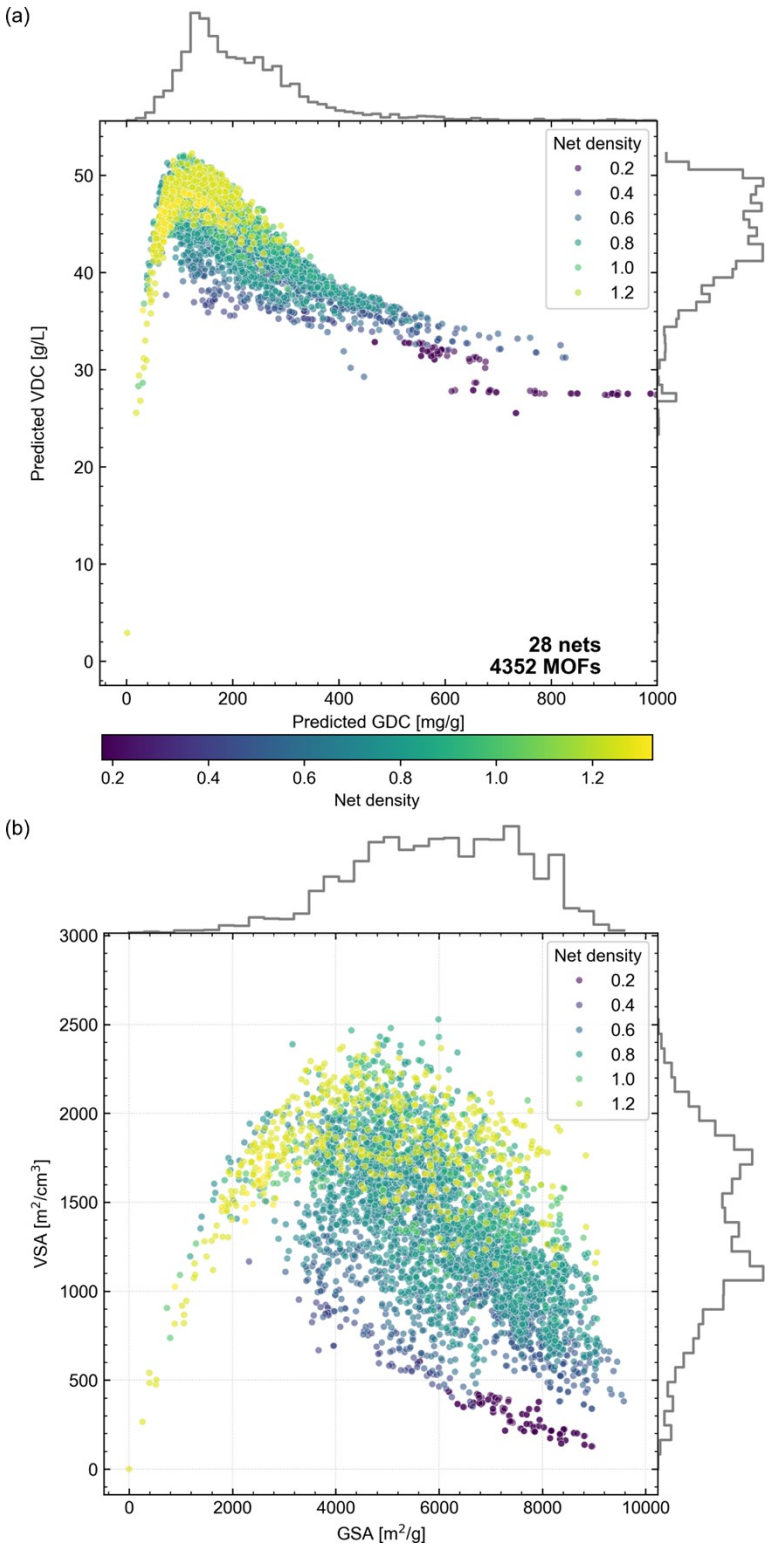


Figure S71. Bivariate plots of the MOFs with the 3-c on3 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

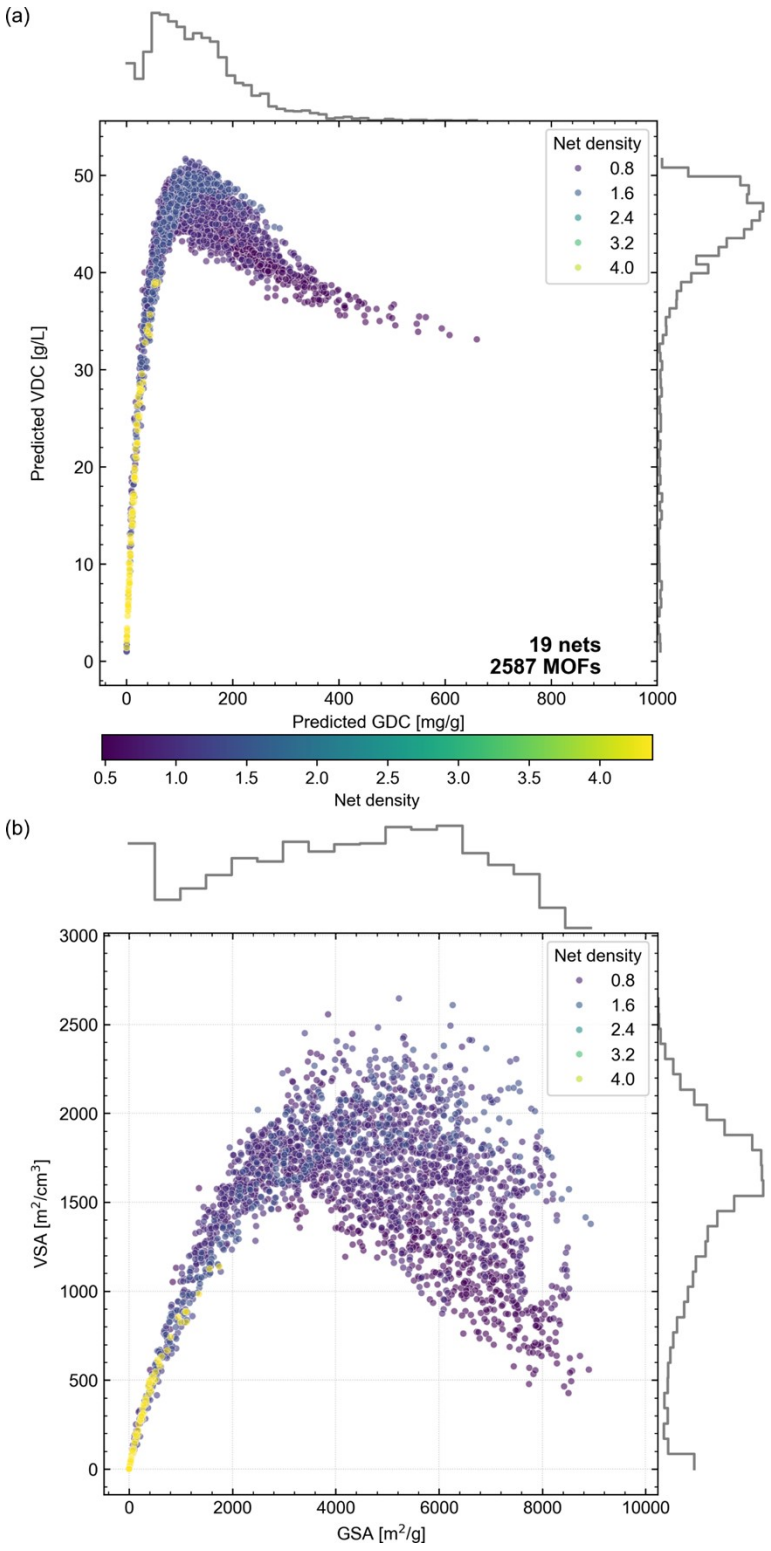


Figure S72. Bivariate plots of the MOFs with the 4-c on4 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

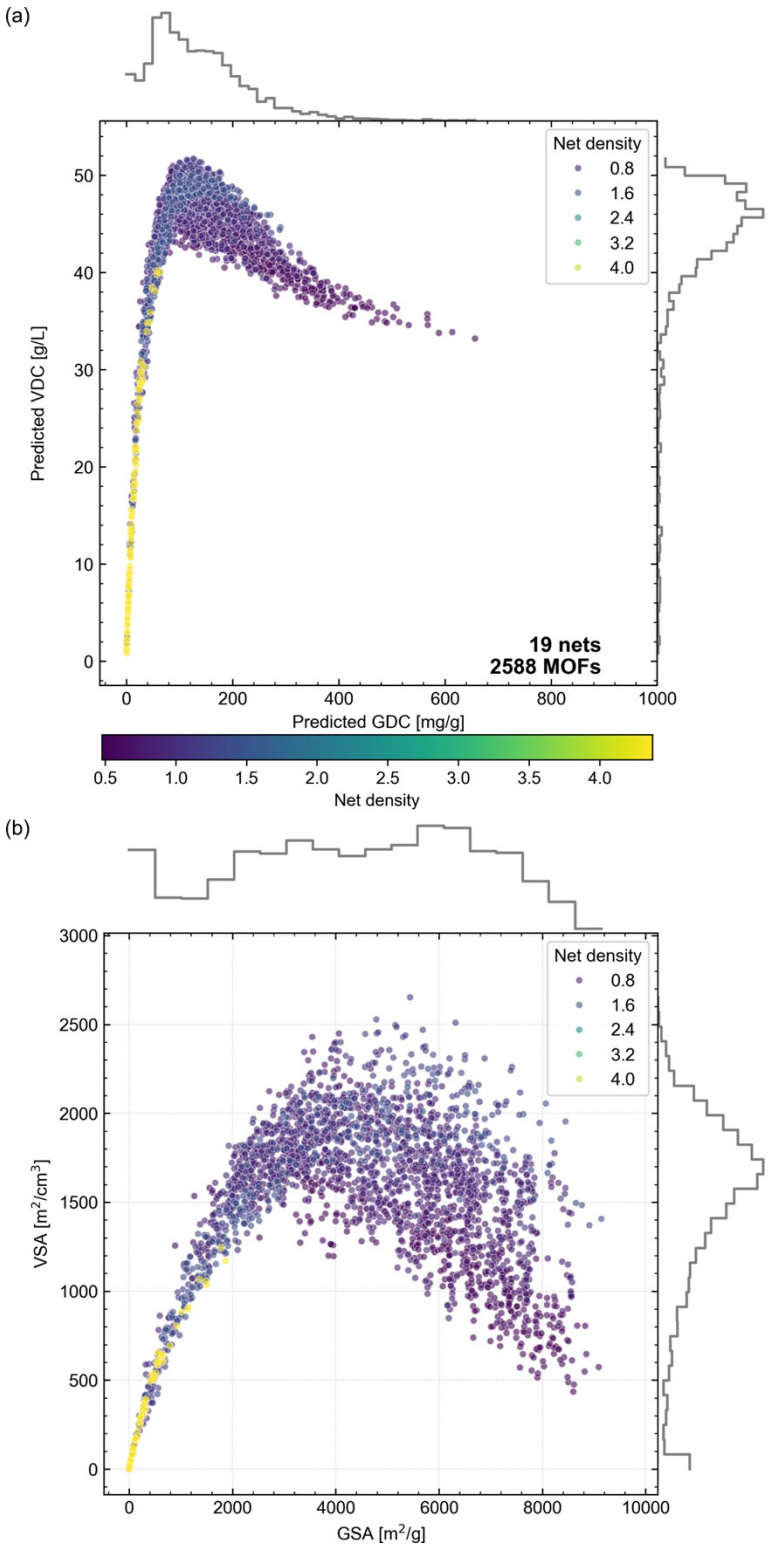


Figure S73. Bivariate plots of the MOFs with the 4-c on5 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

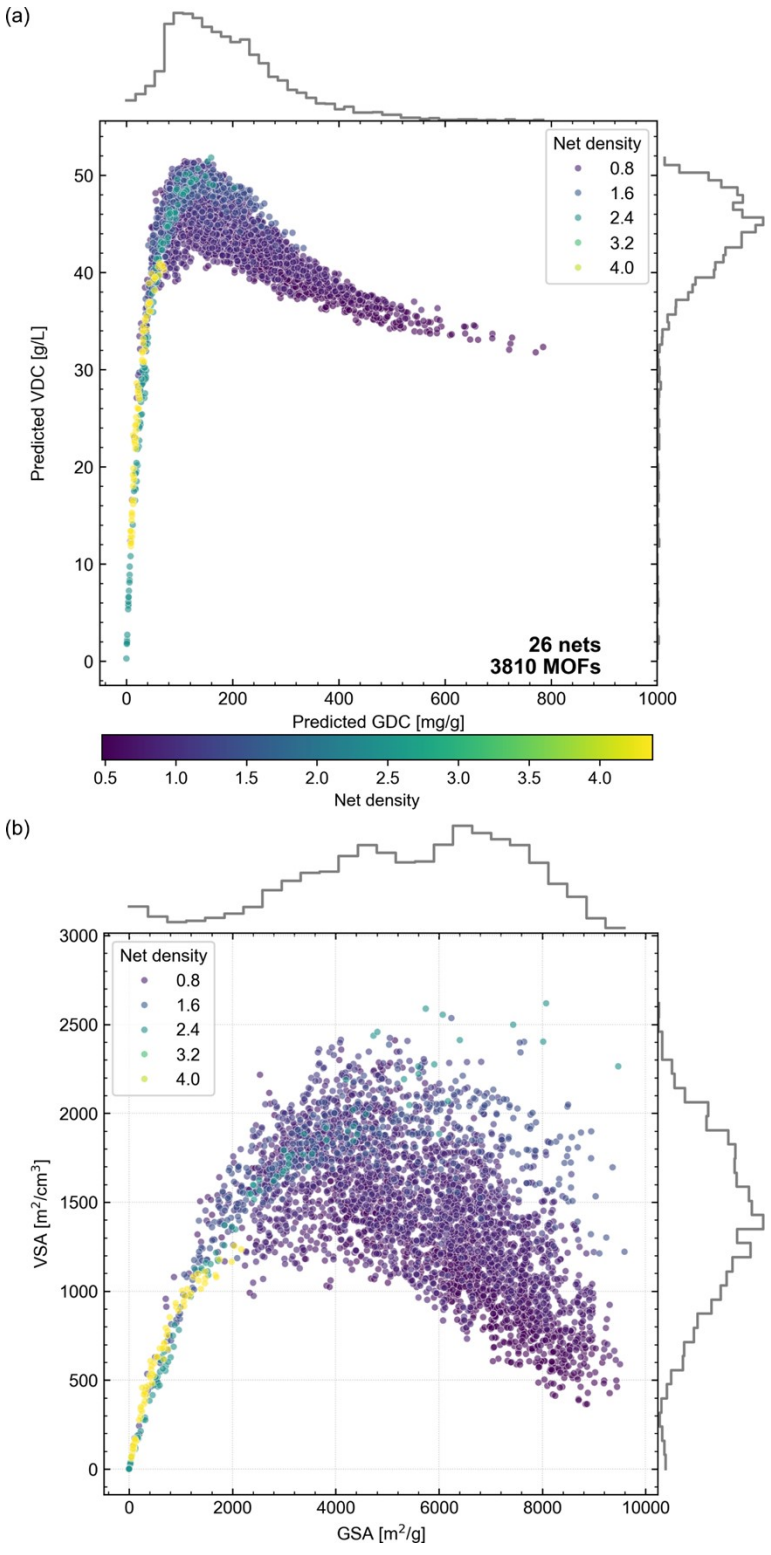


Figure S74. Bivariate plots of the MOFs with the 4-c on6 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

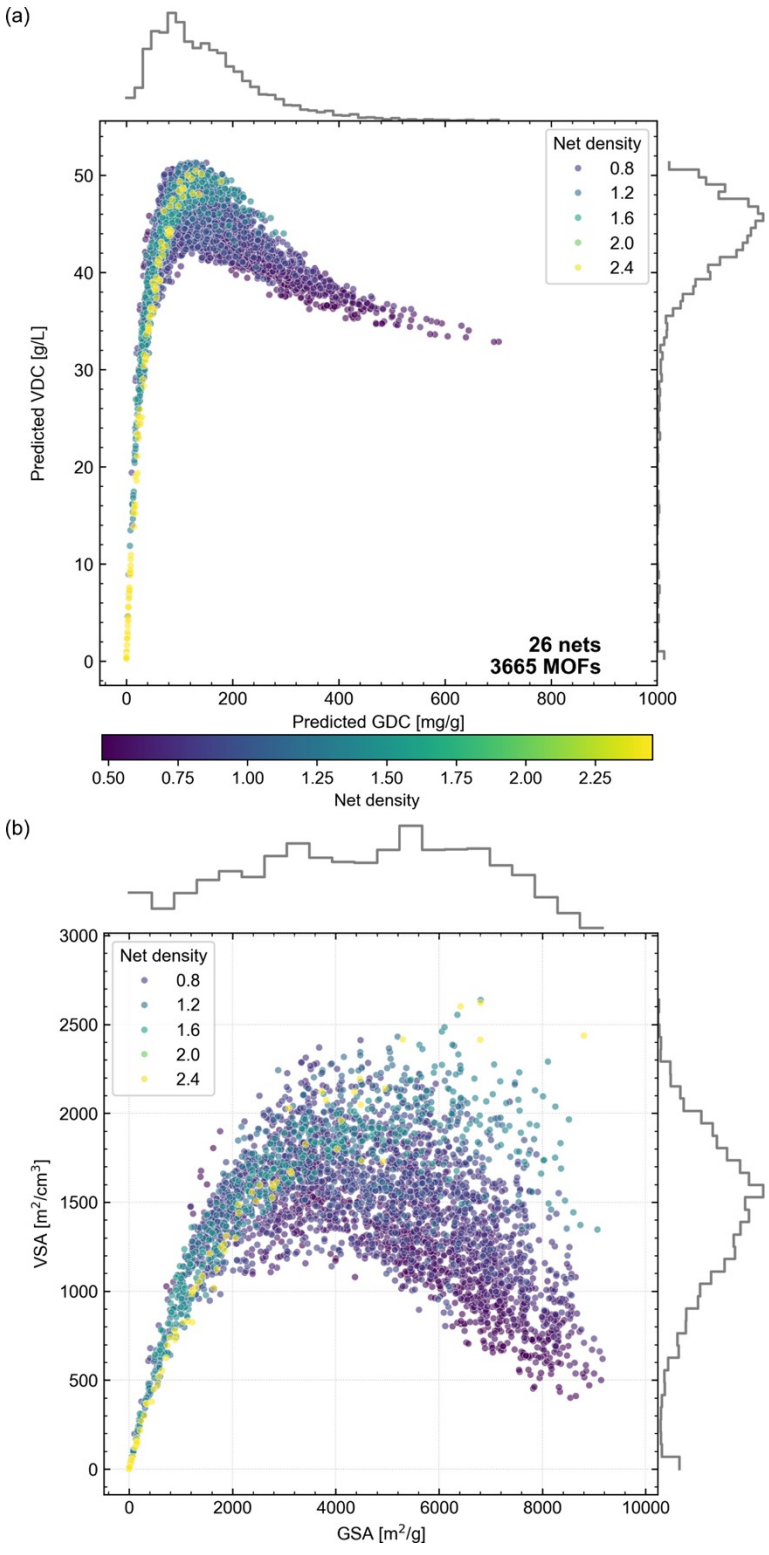


Figure S75. Bivariate plots of the MOFs with the 4-c on7 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

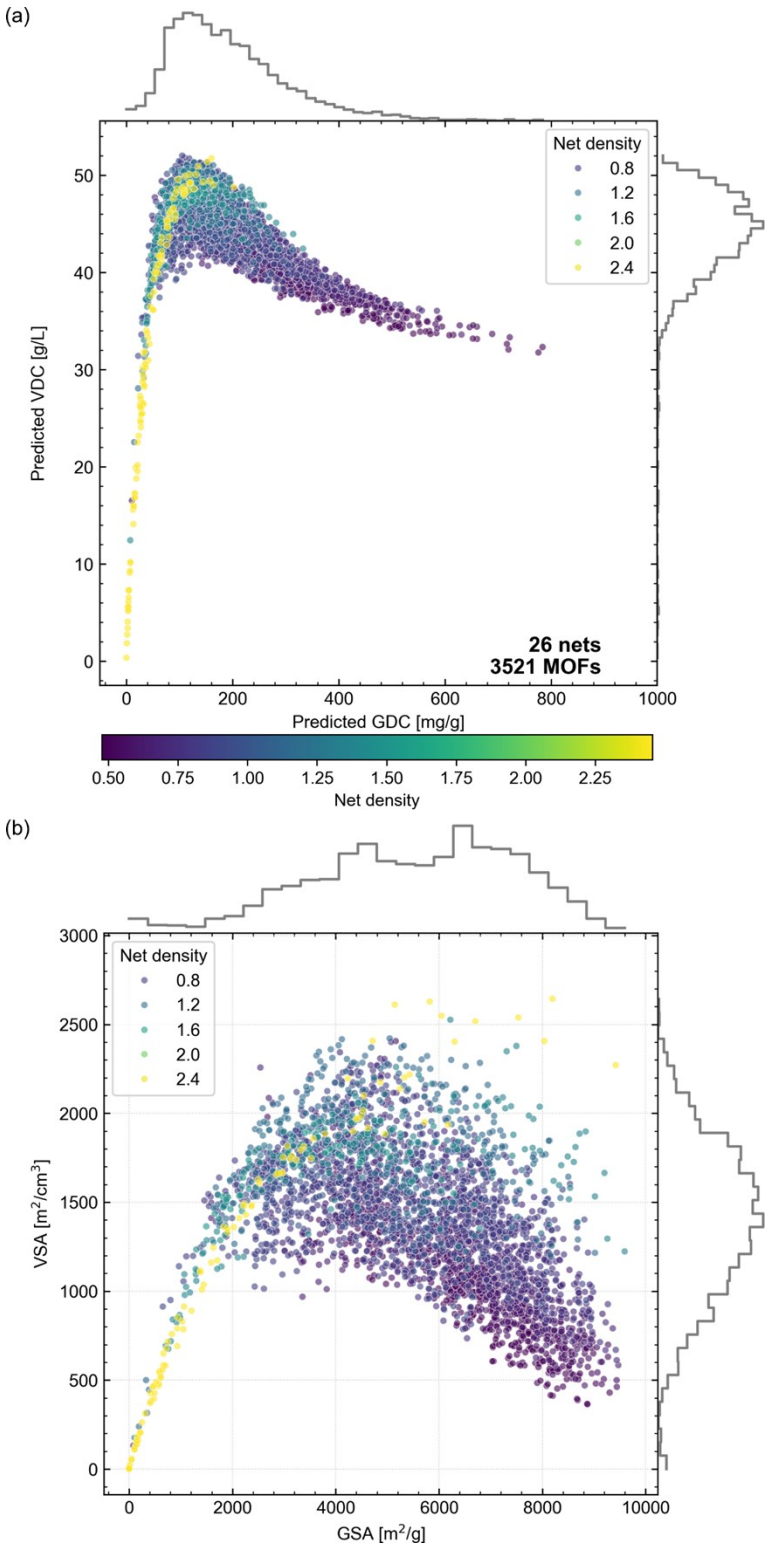


Figure S76. Bivariate plots of the MOFs with the 4-c on8 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

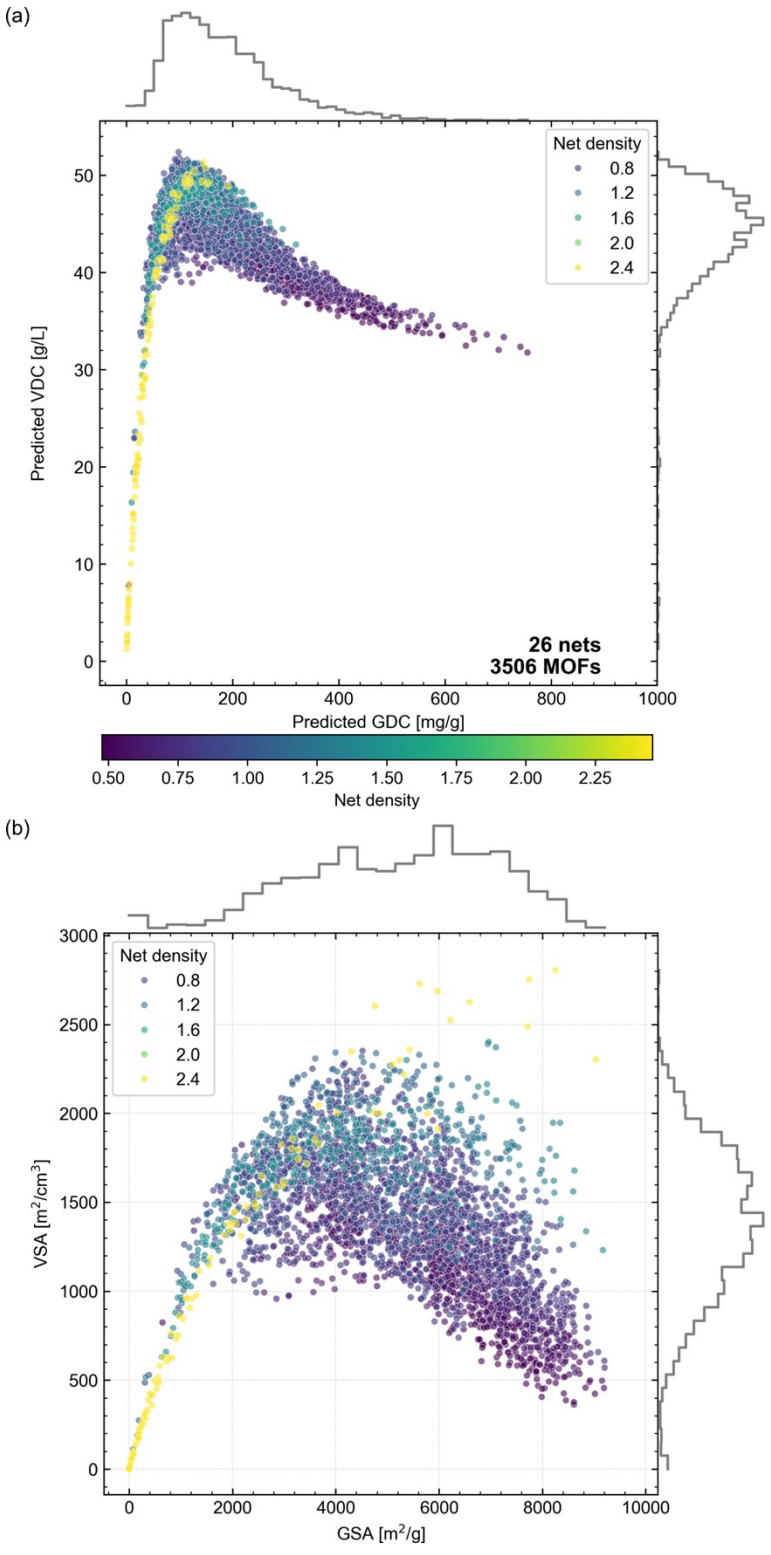


Figure S77. Bivariate plots of the MOFs with the 4-c on9 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

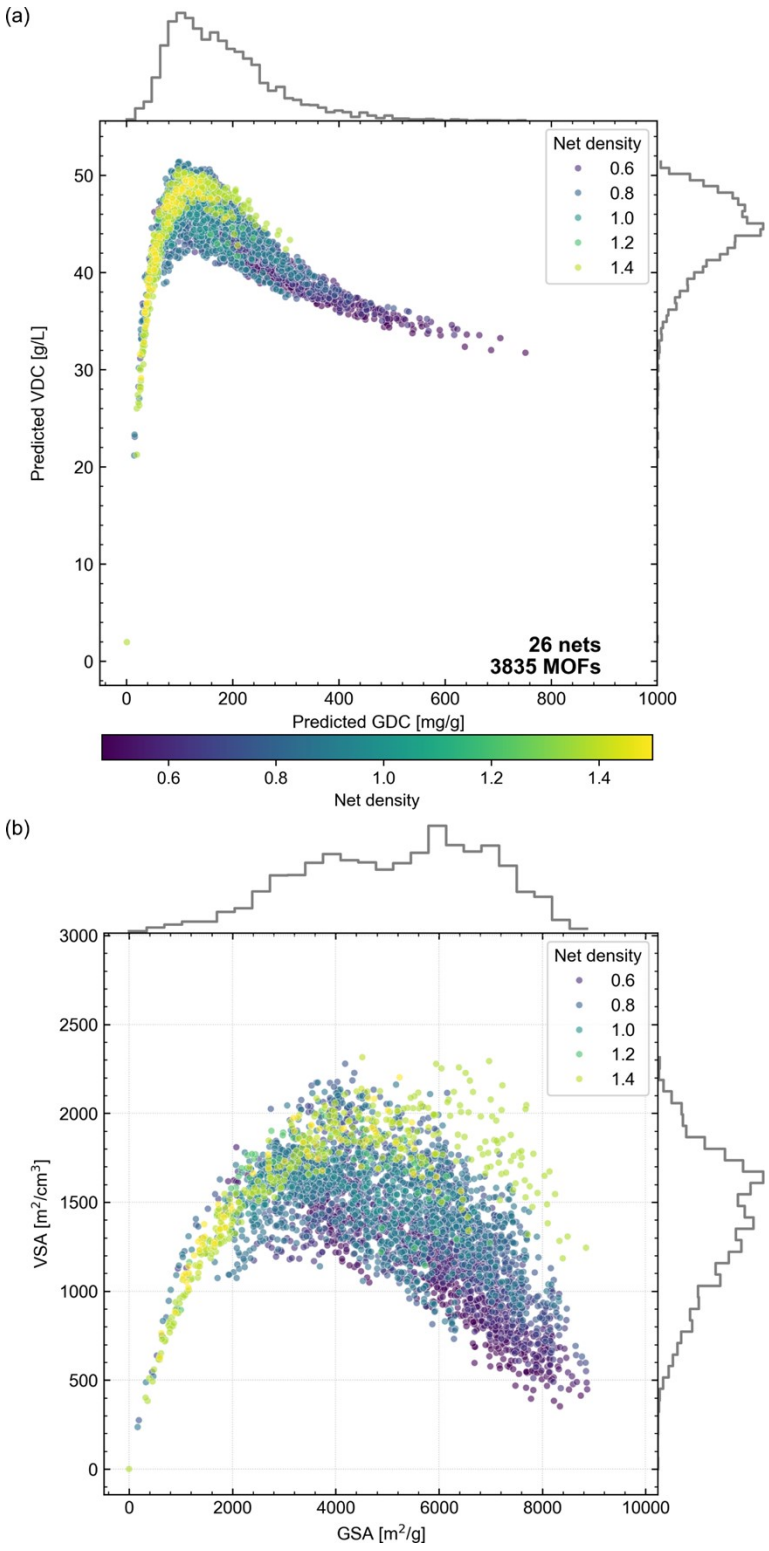


Figure S78. Bivariate plots of the MOFs with the 4-c on10 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

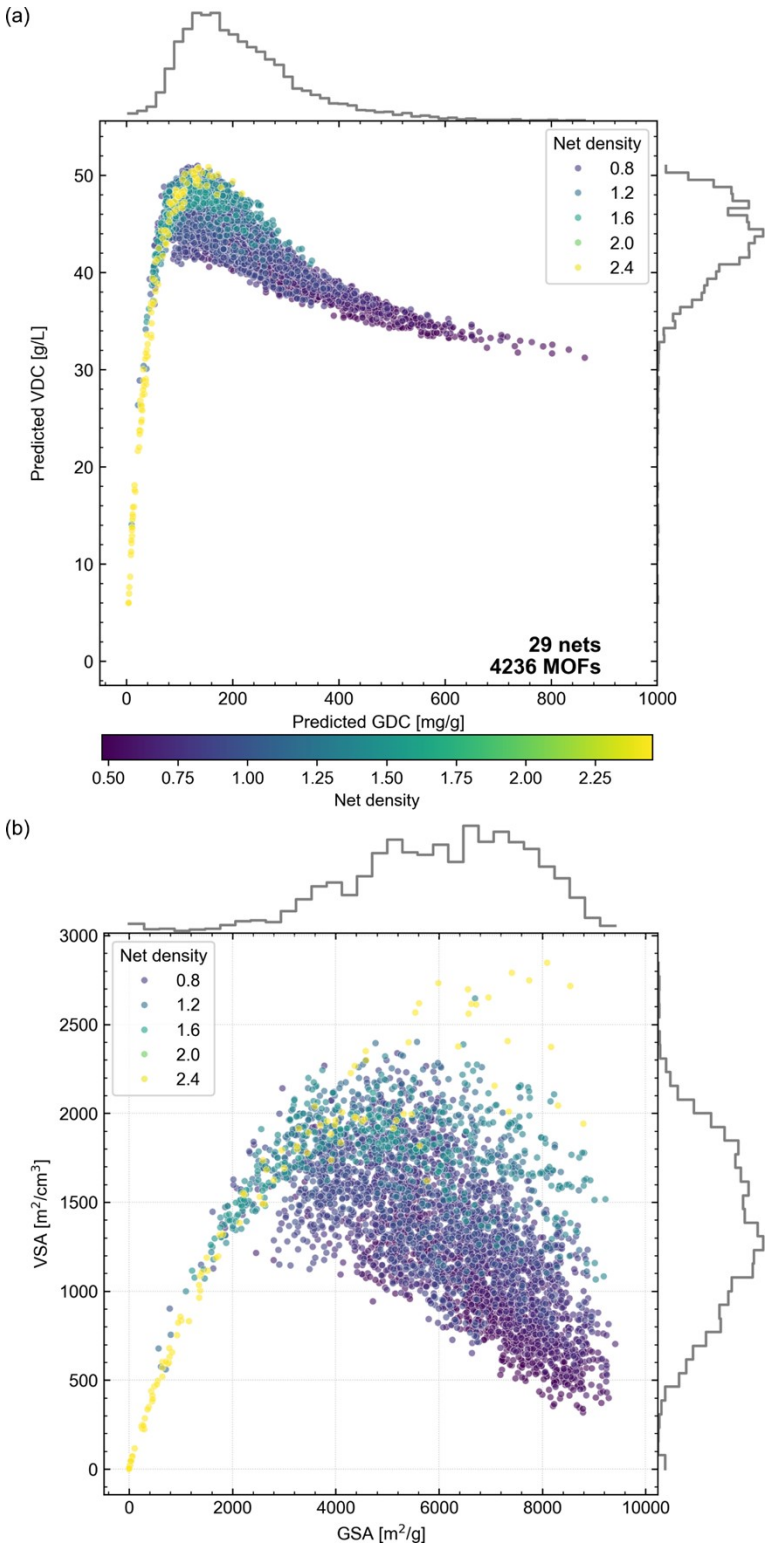


Figure S79. Bivariate plots of the MOFs with the 4-c on11 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

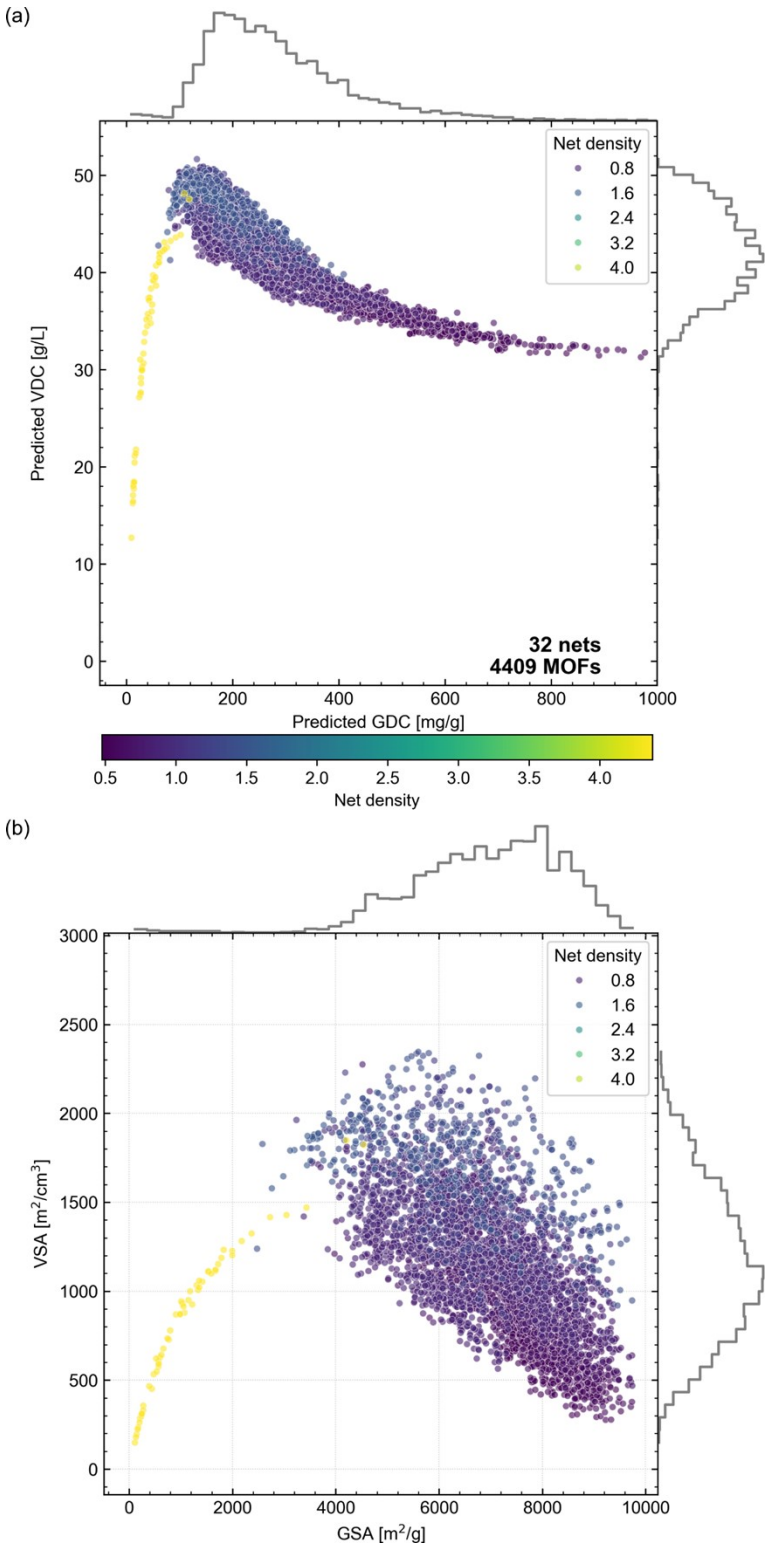


Figure S80. Bivariate plots of the MOFs with the 4-c on13 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

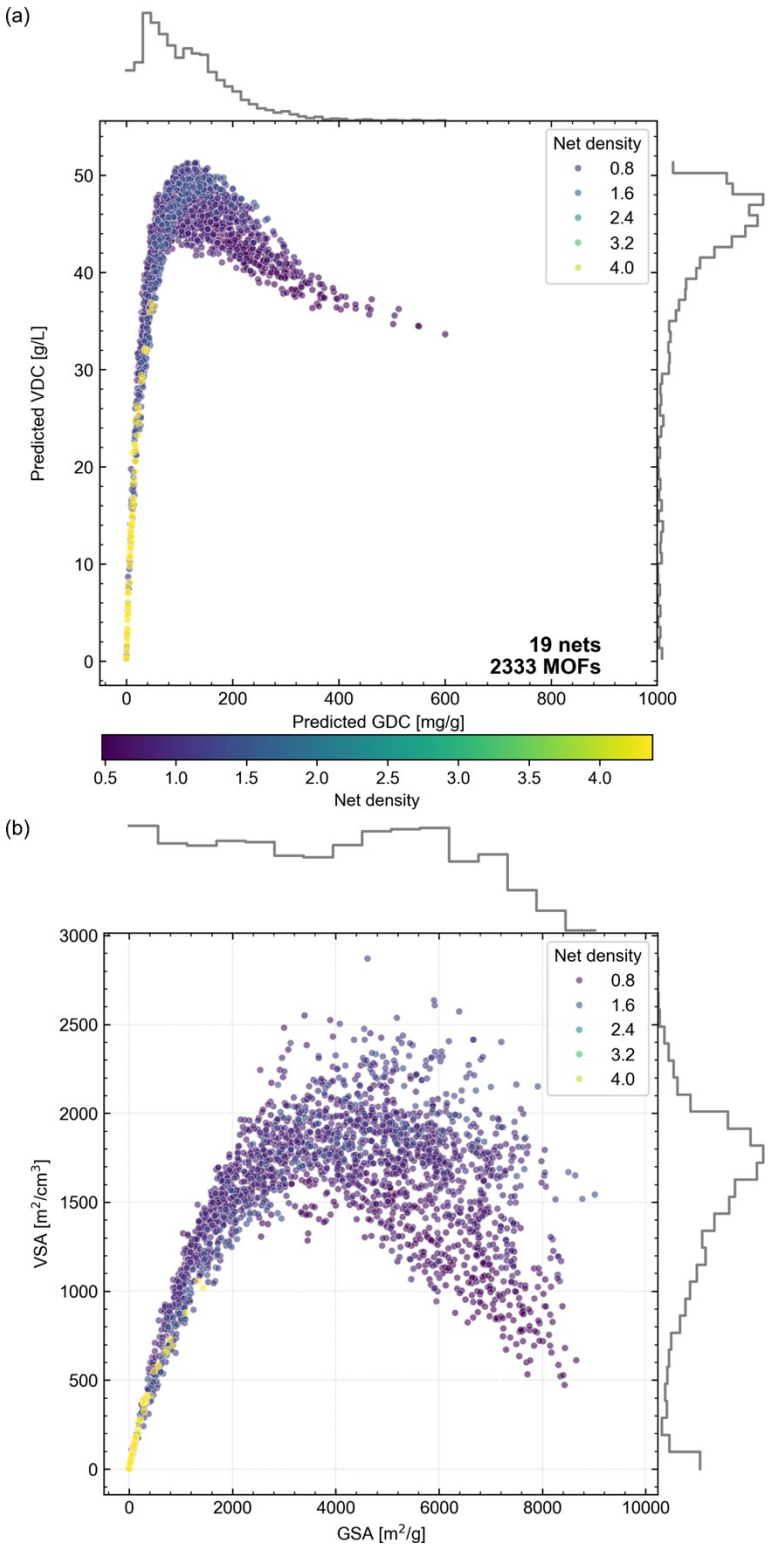


Figure S81. Bivariate plots of the MOFs with the 4-c on14 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

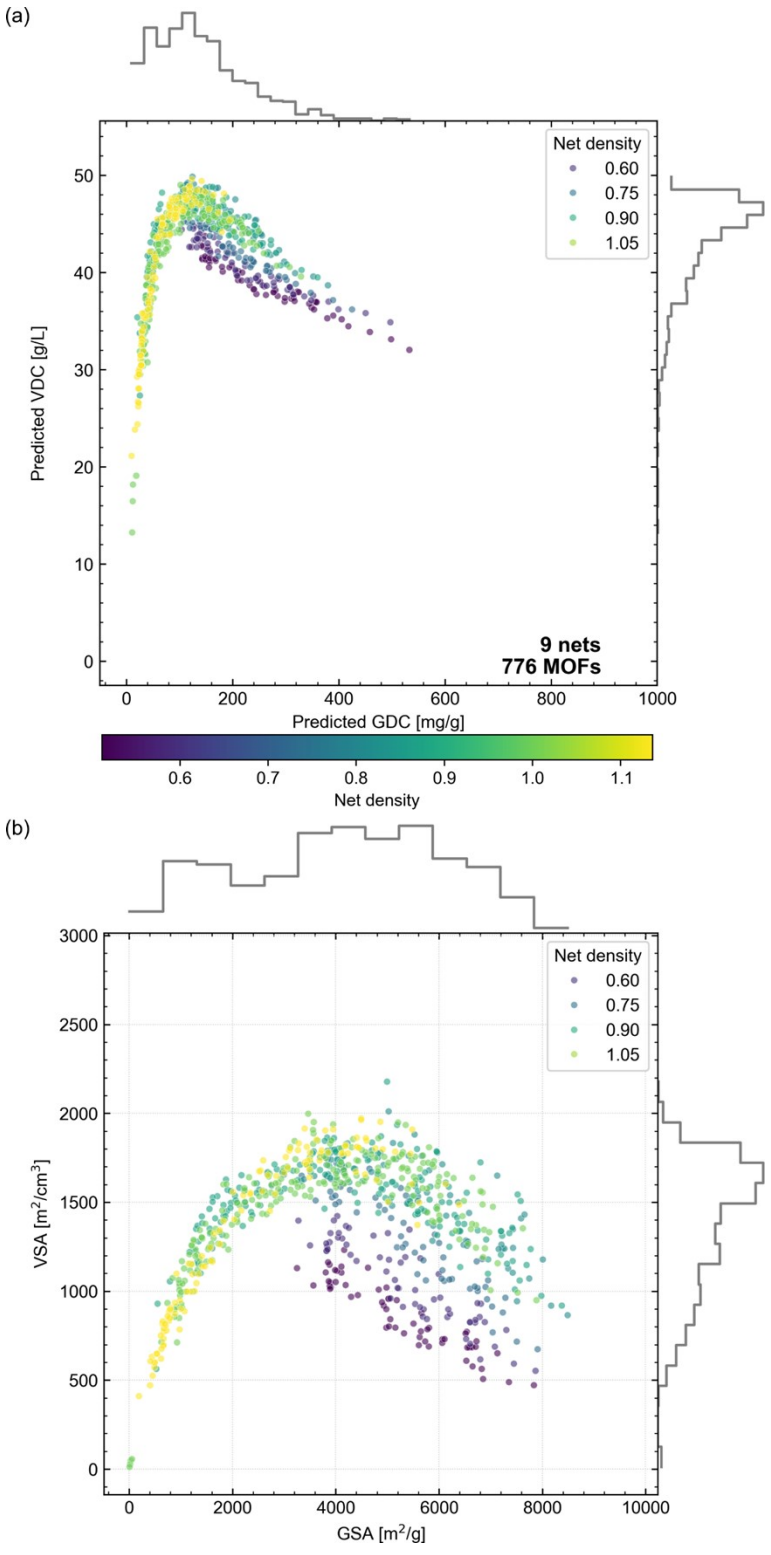


Figure S82. Bivariate plots of the MOFs with the 6-c on12 node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

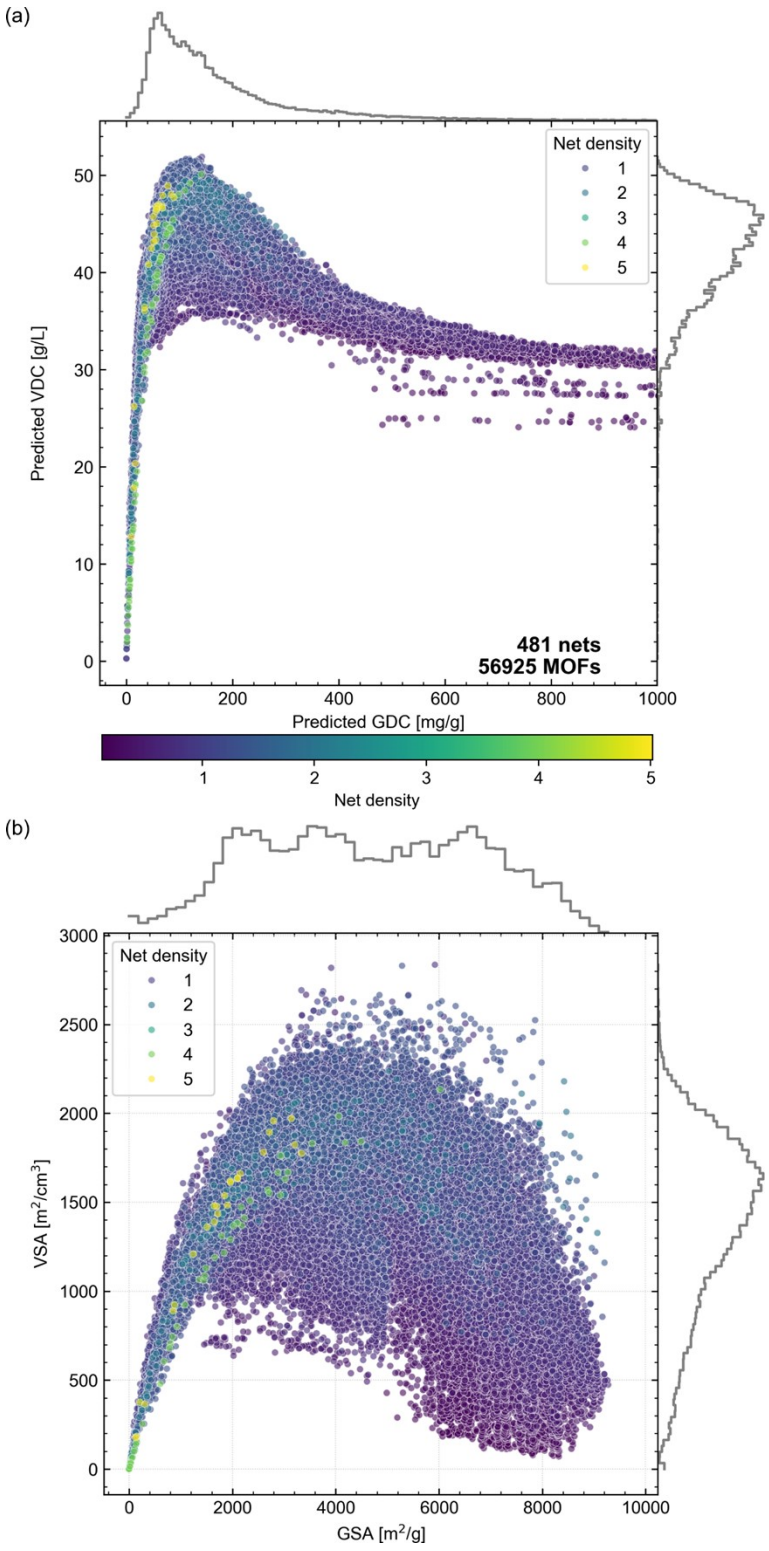


Figure S83. Bivariate plots of the MOFs with no organic node, colored by the net density: (a) deliverable capacities and (b) surface areas. Legends show representative values.

14 Analysis of distributions of MOFs with respect to organic nodes

We observe slight variations in the deliverable capacity landscape between MOFs with different organic node coordination (Figure S84). While the MOFs with the 6-c organic node apparently have a lower peak performance than the other subgroups, there is only one 6-coordinated organic node (a 6-substituted phenyl group), and its chemical space may be underexplored.

When comparing the MOFs with different organic nodes that have identical coordination numbers (Figure S86 – S89), shifts in the distribution are observed. For example, the distribution of the MOFs incorporating the **on13** node demonstrates a higher GDC mode (peak of the distribution) but also a lower VDC mode than those with other 4-connected organic nodes (Figure S87). However, regarding the peak VDC performance, the best performing MOF constructed with **on13** still approaches the Pareto front with a predicted VDC close to 52 g/L.

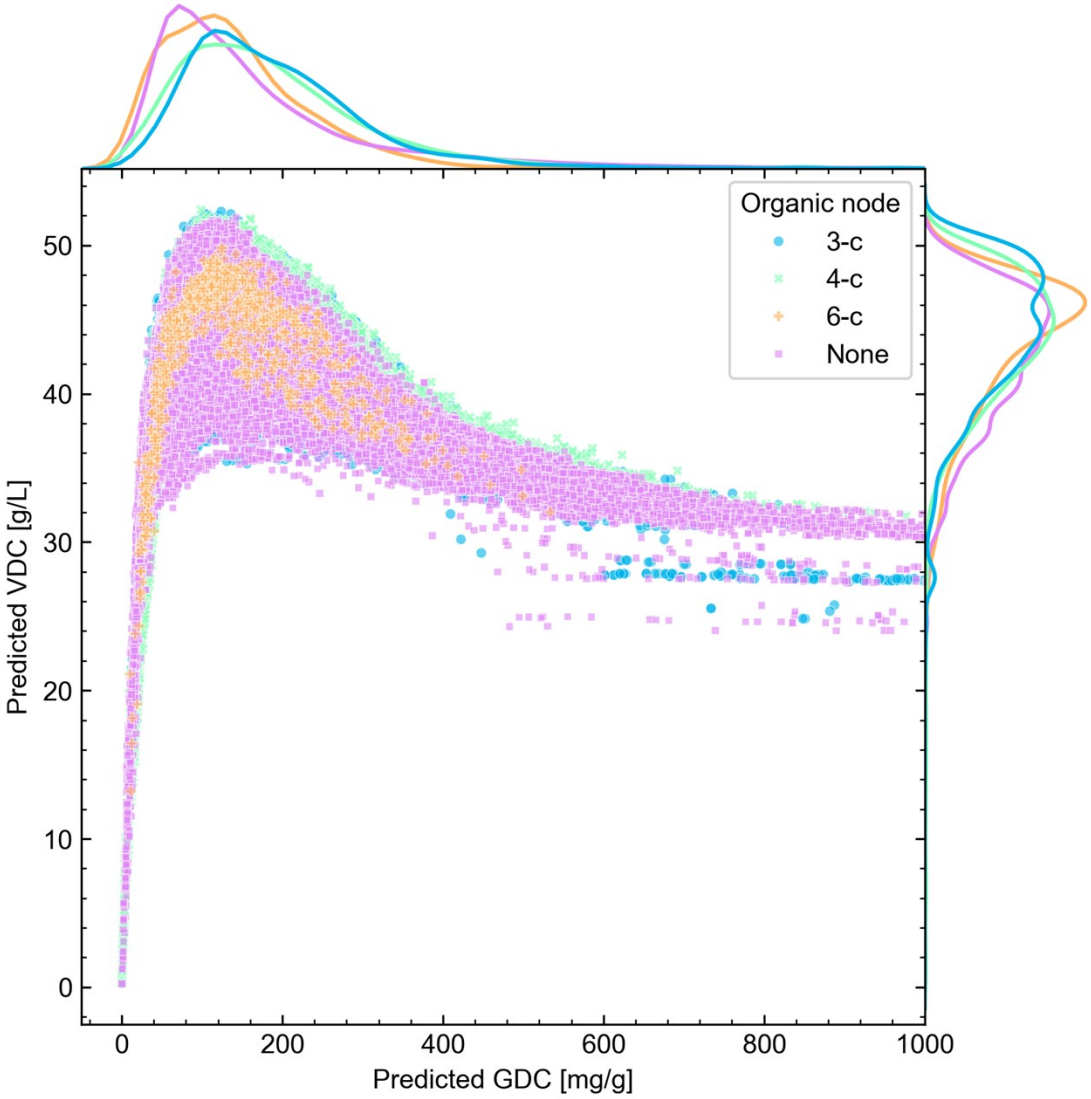


Figure S84. VDC versus GDC for the 105,206 MOFs, colored by the organic node coordination. The figure is truncated at GDC of 1000 mg/g for readability.

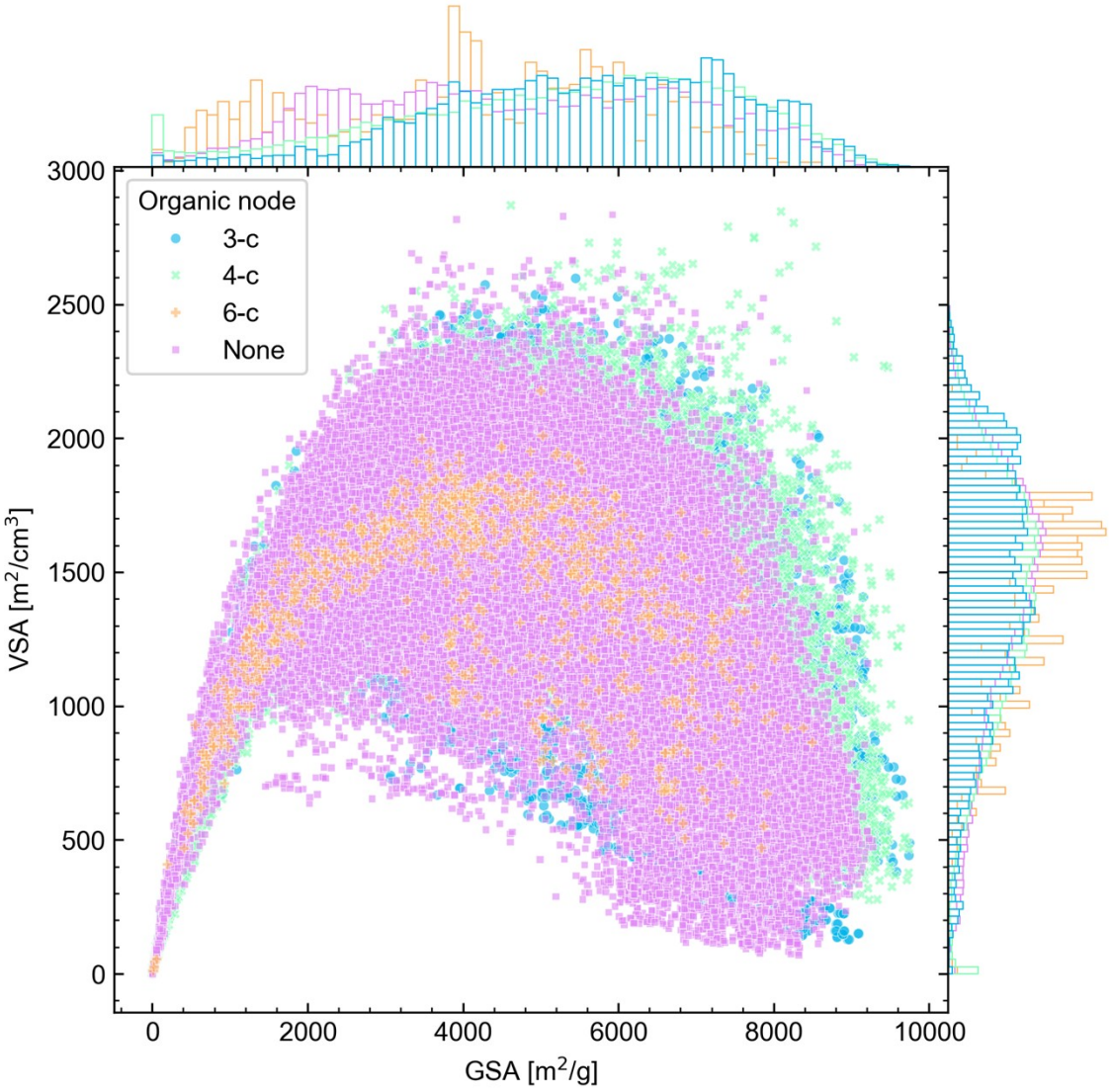


Figure S85. VSA versus GSA for the 105,206 MOFs plotted in VSA against GSA, colored by the organic node coordination.

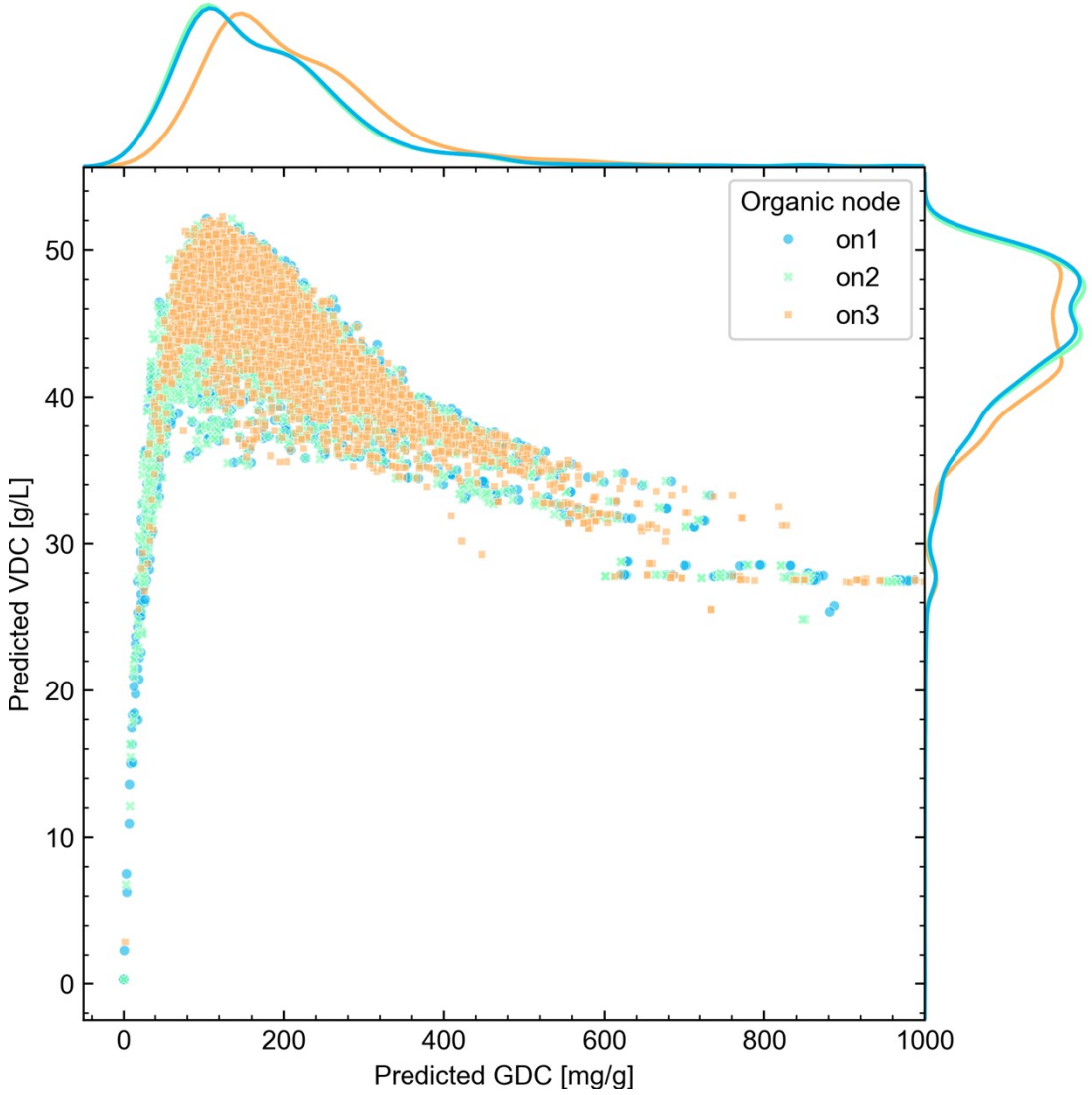


Figure S86. Deliverable capacity landscape of the MOFs with 3-c organic nodes, colored by the organic node coordination. The figure is truncated at GDC of 1000 mg/g for readability.

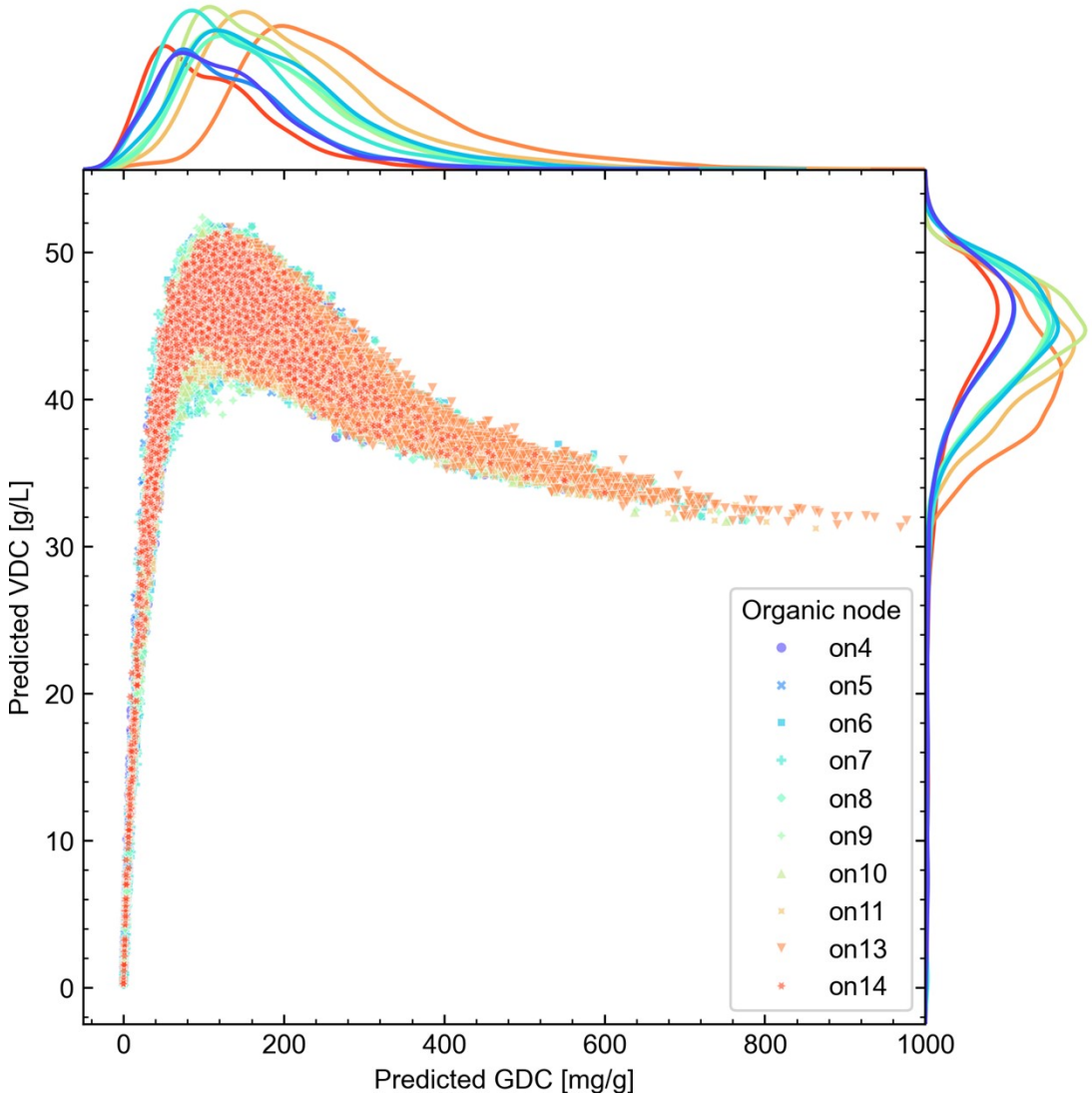


Figure S87. Deliverable capacity landscape of the MOFs with 4-c organic nodes, colored by the organic node coordination. The figure is truncated at GDC of 1000 mg/g for readability.

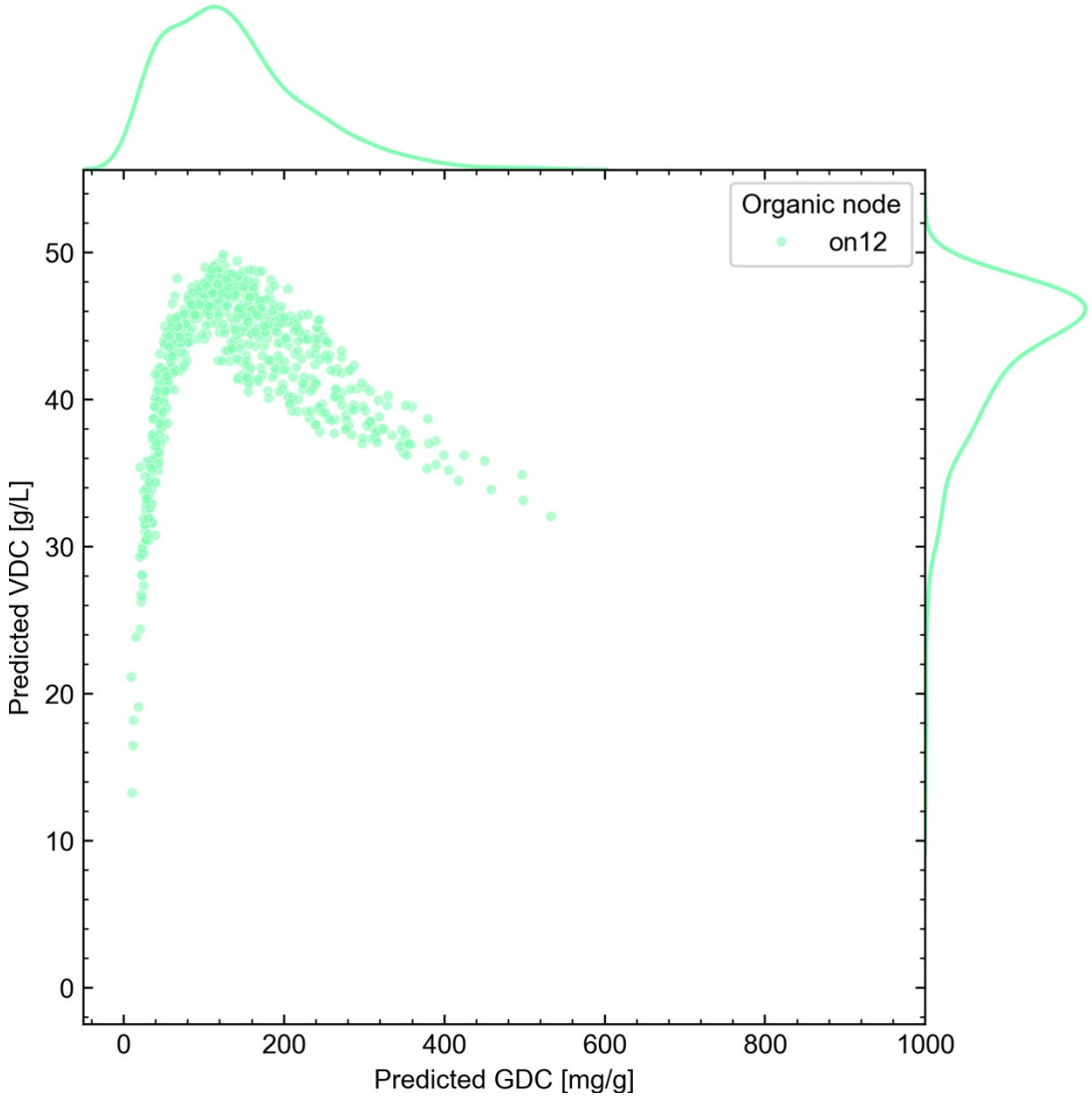


Figure S88. Deliverable capacity landscape of the MOFs with 6-c organic nodes, colored by the organic node coordination. The figure is truncated at GDC of 1000 mg/g for readability.

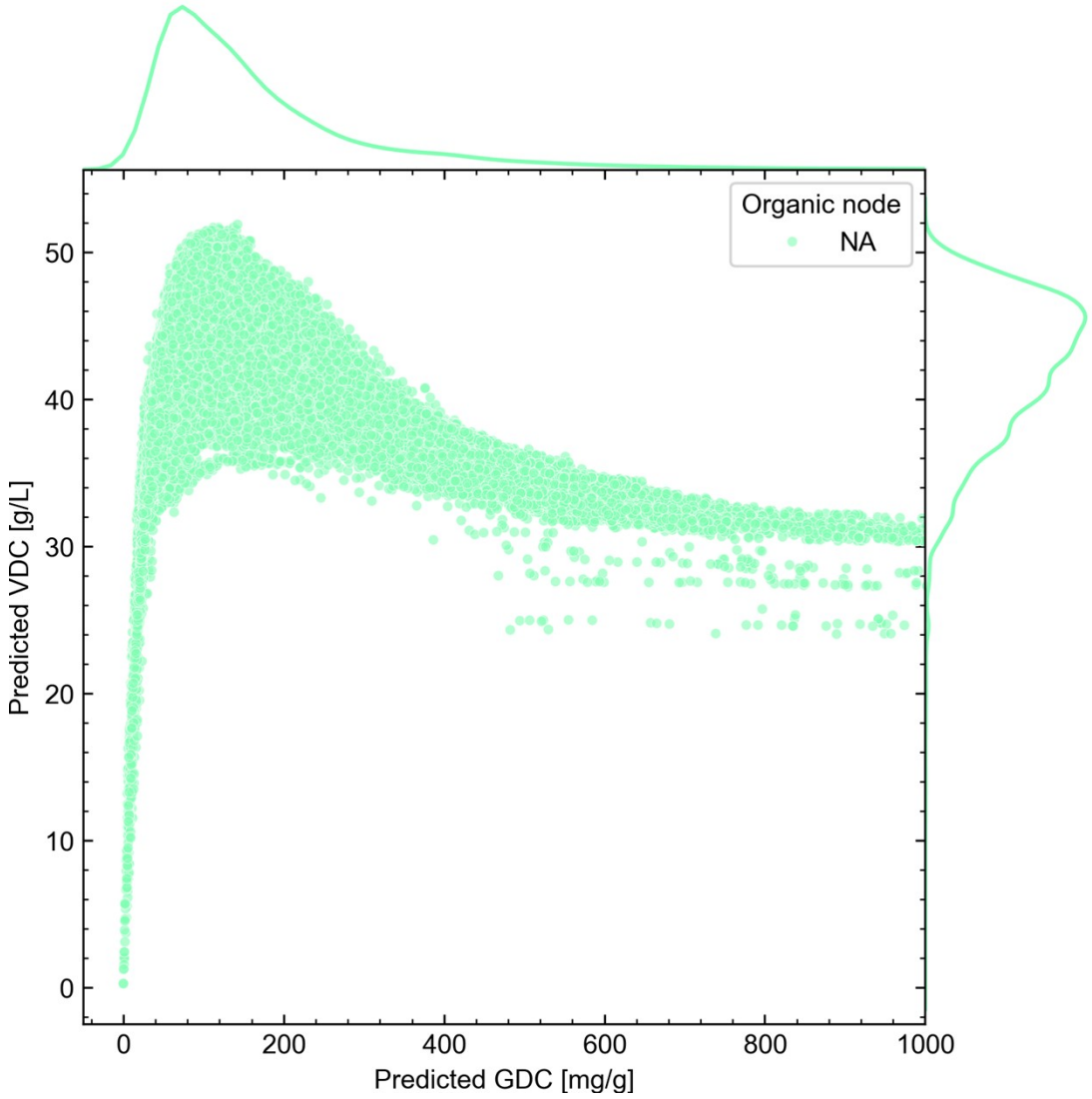


Figure S89. Deliverable capacitylandscape of the MOFs with no organic nodes, colored by the organic node coordination. The figure is truncated at GDC of 1000 mg/g for readability.

15 Exploration of tpt-MOFs

We observe that MOFs with a very dense net tend to have suboptimal pore volume, which prevents the MOFs from reaching the volcano peak of volumetric versus gravimetric surface areas or deliverable capacities (for example, Figure S50). For instance, the 3,6-c **tpt** net is a mathematically discovered net[14] with the highest net density in the dataset. To determine the location of its volcano peak in the surface area space, we explored 420 new **tpt**-MOFs using extra long edge building blocks of approximately 15–25 Å. However, many of the structures showed collapsed pores after geometry optimization and therefore reduced VSA and pore volume (Figure S90). Due to intrinsic limitations of the optimization algorithm (see Section 2), the results remain inclusive.

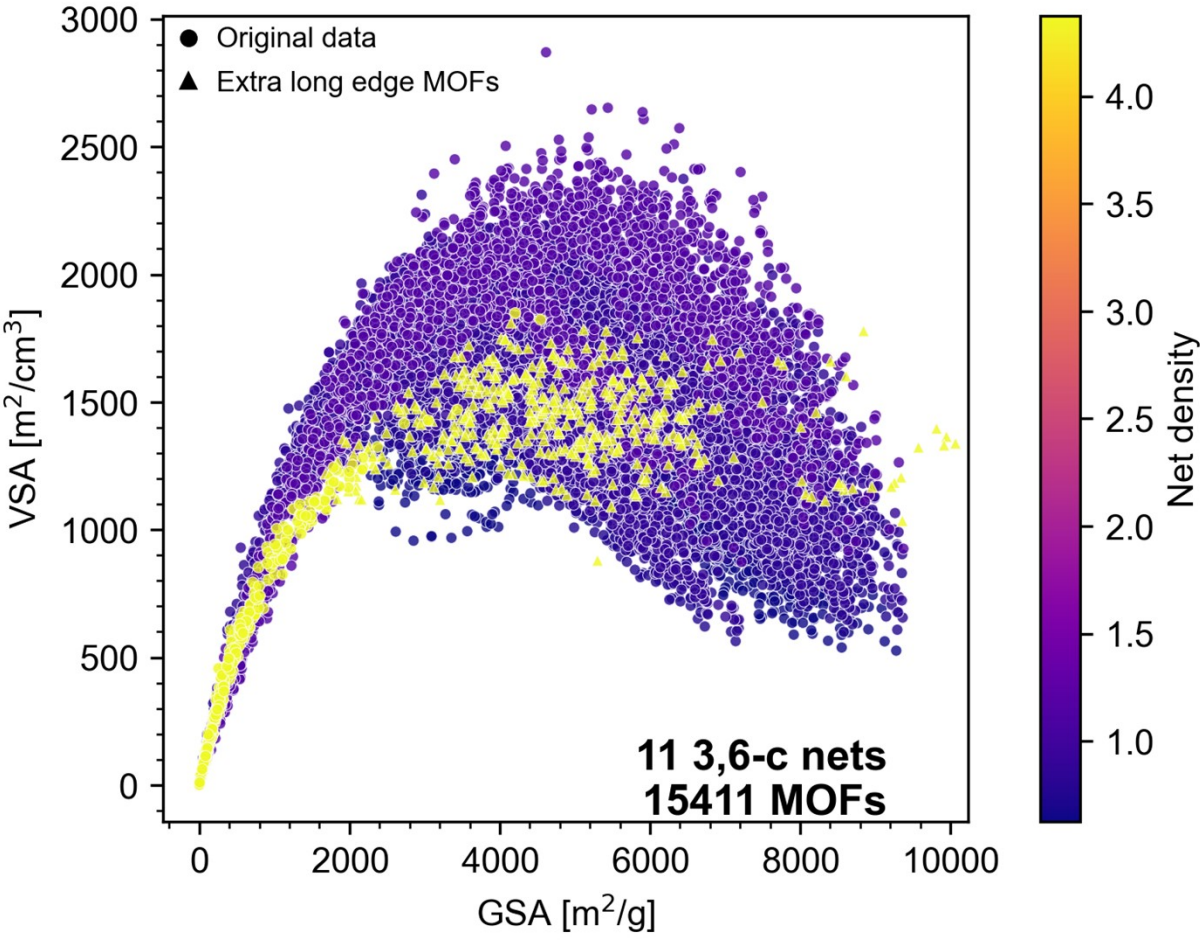


Figure S90. Exploration of structural space for **tpt** net by generating MOFs with extra long edge building blocks. The surface area distributions for all the MOFs with 3,6-c nets are shown colored by the

net density. The original MOFs are indicated by circles, and the new MOFs explored are indicated by triangles.

References

(1) Liu, K.; Chen, H.; Islamoglu, T.; Rosen, A. S.; Wang, X.; Farha, O. K.; Snurr, R. Q. Supplementary Data for "Computational Investigation of the Impact of Metal-Organic Framework Topology on Hydrogen Storage Capacity." https://doi.org/10.5281/zenodo.14997970.

(2) Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. *Crystal Growth & Design* **2017**, *17* (11), 5801–5810. https://doi.org/10.1021/acs.cgd.7b00848.

(3) Liu, K.; Chen, Z.; Islamoglu, T.; Lee, S.-J.; Chen, H.; Yildirim, T.; Farha, O. K.; Snurr, R. Q. Exploring the Chemical Space of Metal–Organic Frameworks with Rht Topology for High Capacity Hydrogen Storage. *J. Phys. Chem. C* **2024**, *128* (18), 7435–7446. https://doi.org/10.1021/acs.jpcc.4c00638.

(4) O'Keeffe, M. Some Properties of Three-Periodic Sphere Packings. In *Science of Crystal Structures: Highlights in Crystallography*; Hargittai, I., Hargittai, B., Eds.; Springer International Publishing: Cham, 2015; pp 155–163. https://doi.org/10.1007/978-3-319-19827-9_17.

(5) Delgado-Friedrichs, O.; Foster, M. D.; O'Keeffe, M.; Proserpio, D. M.; Treacy, M. M. J.; Yaghi, O. M. What Do We Know about Three-Periodic Nets? *Journal of Solid State Chemistry* **2005**, *178* (8), 2533–2554. https://doi.org/10.1016/j.jssc.2005.06.037.

(6) Power, S. C.; Baburin, I. A.; Proserpio, D. M. Isotopy Classes for 3-Periodic Net Embeddings. *Acta Cryst A* **2020**, *76* (3), 275–301. https://doi.org/10.1107/S2053273320000625.

(7) Frost, H.; Düren, T.; Snurr, R. Q. Effects of Surface Area, Free Volume, and Heat of Adsorption on Hydrogen Uptake in Metal−Organic Frameworks. *J. Phys. Chem. B* **2006**, *110* (19), 9565–9570. https://doi.org/10.1021/jp060433+.

(8) Rohlíček, J.; Skořepová, E.; Babor, M.; Čejka, J. CrystalCMP: An Easy-to-Use Tool for Fast Comparison of Molecular Packing. *J Appl Cryst* **2016**, *49* (6), 2172–2183. https://doi.org/10.1107/S1600576716016058.

(9) Costanzo, F.; Silvestrelli, P. L.; Ancilotto, F. Physisorption, Diffusion, and Chemisorption Pathways of H2 Molecule on Graphene and on (2,2) Carbon Nanotube by First Principles Calculations. *J. Chem. Theory Comput.* **2012**, *8* (4), 1288–1294. https://doi.org/10.1021/ct300143a.

(10) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv 2018. https://doi.org/10.48550/ARXIV.1802.03426.

(11) O'Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets. *Acc. Chem. Res.* **2008**, *41* (12), 1782–1789. https://doi.org/10.1021/ar800124u.

(12) Slack, G. The Most-Dense and Least-Dense Packings of Circles and Spheres. *Z. Kristallogr.* **1983**, *165*, 1.

(13) Furukawa, H.; Gándara, F.; Zhang, Y.-B.; Jiang, J.; Queen, W. L.; Hudson, M. R.; Yaghi, O. M. Water Adsorption in Porous Metal–Organic Frameworks and Related Materials. *J. Am. Chem. Soc.* **2014**, *136* (11), 4369–4381. https://doi.org/10.1021/ja500330a.

(14) O'Keeffe, M. Nets, Tiles, and Metal-Organic Frameworks. *APL Materials* **2014**, *2* (12), 124106. https://doi.org/10.1063/1.4901292.