



Supplementary Materials for

## Energy-Based Descriptors to Rapidly Predict Hydrogen Storage in Metal-Organic Frameworks

Benjamin J. Bucior, N. Scott Bobbitt, Timur Islamoglu, Subhadip Goswami, Arun Gopalan, Taner Yildirim, Omar K. Farha, Neda Bagheri, Randall Q. Snurr

### Contents

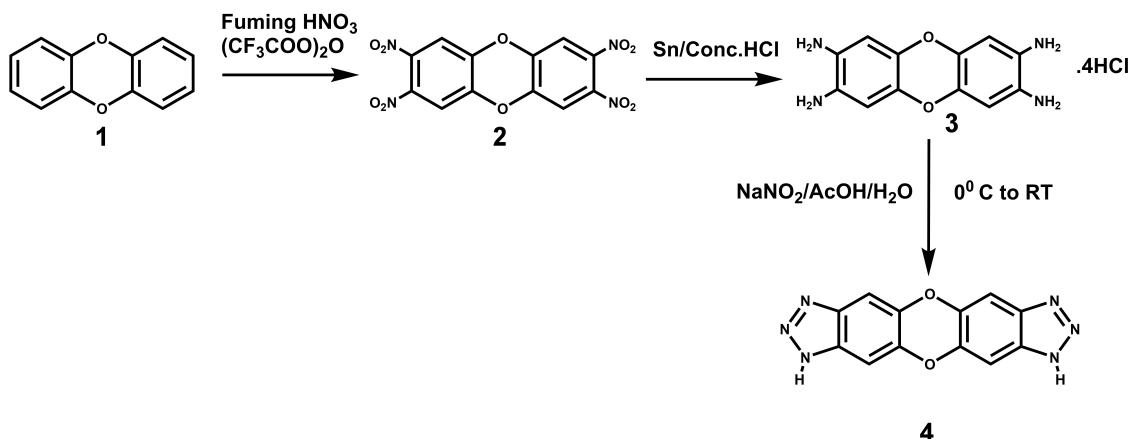
<b>1 Experimental details</b>	<b>2</b>
1.1 Synthesis of MFU-4l . . . . .	2
1.2 Characterization . . . . .	3
1.3 Absolute and excess uptake calculations . . . . .	8
<b>2 Molecular simulation details</b>	<b>10</b>
2.1 Interatomic potential parameters . . . . .	10
2.2 Feynman-Hibbs effects . . . . .	11
2.3 GCMC Error Bars . . . . .	12
2.4 Benchmarking the charge on hydrogen model . . . . .	13
<b>3 Diversity of properties in MOF databases</b>	<b>15</b>
<b>4 Details and tests from the machine learning analysis</b>	<b>18</b>
4.1 Effect of regularization . . . . .	18
4.2 Model coefficients and determination of energy range . . . . .	19
4.3 Detailed model performance . . . . .	22
4.4 Coarse energy grid sampling . . . . .	26
<b>5 Top 25 Candidates from Screening</b>	<b>28</b>
<b>6 Screenshot of Hydrogen in MOF</b>	<b>31</b>
<b>7 MOFs used for model fitting</b>	<b>31</b>
7.1 hMOF database . . . . .	31
7.2 ToBaCCo MOFs . . . . .	32
7.3 Mixed training set . . . . .	35
7.4 CCDC MOFs . . . . .	37

## 1 Experimental details

The starting material for the synthesis of linker **4** was purchased from Combi-blocks and used without any further purification. The synthesis of **2** and **4** was performed according to the reported procedure.<sup>1</sup> Nitration of **2** was achieved by following reported procedure.<sup>2</sup>

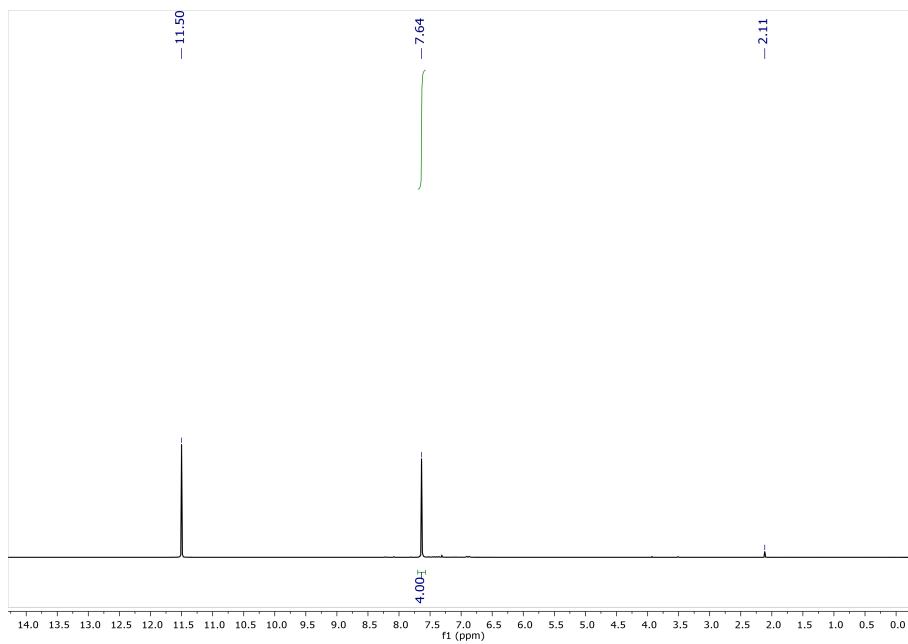
### 1.1 Synthesis of MFU-4I

A published procedure was followed for the synthesis of MFU-4I.<sup>1</sup> H2-BTDD (100 mg, 0.366 mmol) was dissolved in DMF (100 mL) under stirring and heating at 145 °C until clear solution was achieved. Anhydrous ZnCl<sub>2</sub> (1.03 g, 7.55 mmol) was added to the linker solution after it was cooled to around 50 °C and stirred until the zinc chloride was completely dissolved. The resulting solution was stirred under reflux at 145 °C for 18 h and then cooled down to room temperature. The MOF crystals were recovered by centrifuge, washed with DMF (3 times), methanol (3 times) and DCM (3 times) (40 mL each and soaked 30 min each time) and soaked in DCM overnight. The solvent exchanged MOF crystals were then activated under vacuum at 180 °C for 18 h prior to gas adsorption measurements.

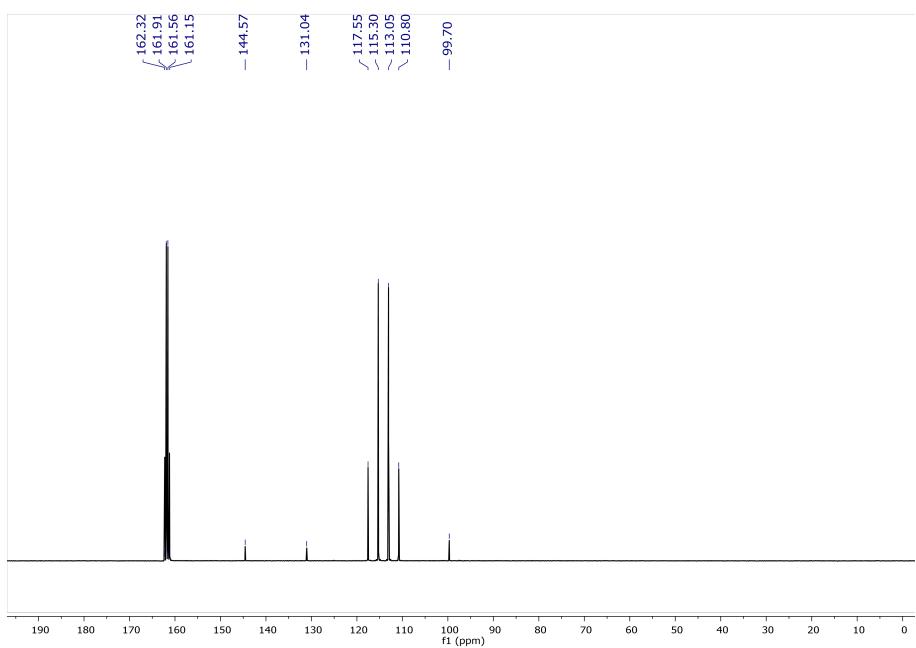


**Figure S1** Synthetic scheme for the linker **4**.

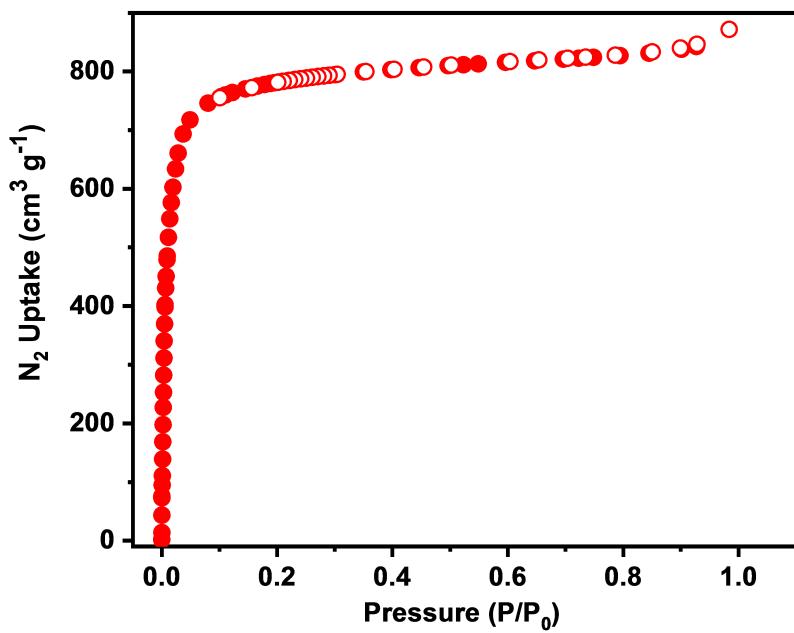
## 1.2 Characterization



**Figure S2**  $^1\text{H}$  (500 MHz) NMR spectra of **4** in  $\text{CF}_3\text{COOD}$ .



**Figure S3**  $^{13}\text{C}$  (125 MHz) NMR spectra of **4** in  $\text{CF}_3\text{COOD}$ .



**Figure S4** Experimental nitrogen isotherms MFU-4I obtained at 77 K. Closed symbols are adsorption and open symbols are desorption.

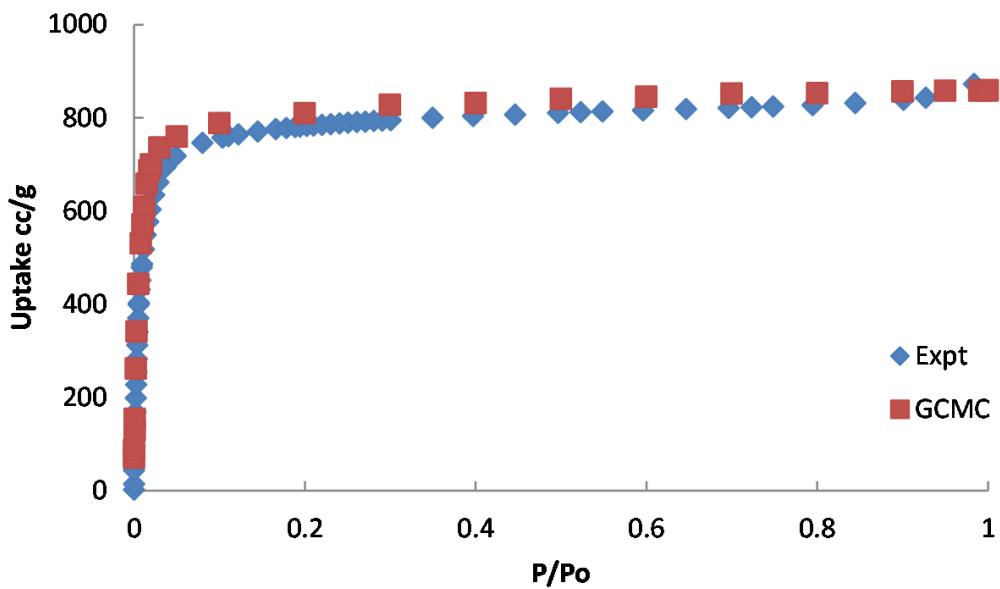


Figure S5 Experimental nitrogen isotherms of MFU-4I obtained at 77 K (blue diamonds) compared with GCMC simulated isotherms (red squares).

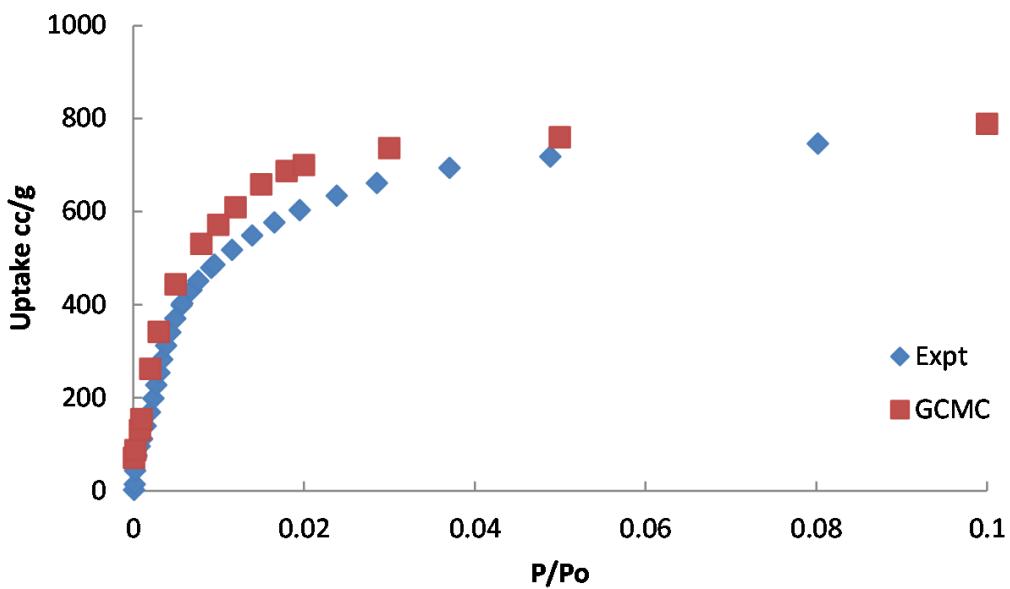


Figure S6 Experimental nitrogen isotherms of MFU-4I obtained at 77 K (blue diamonds) compared with GCMC simulated isotherms (red squares).

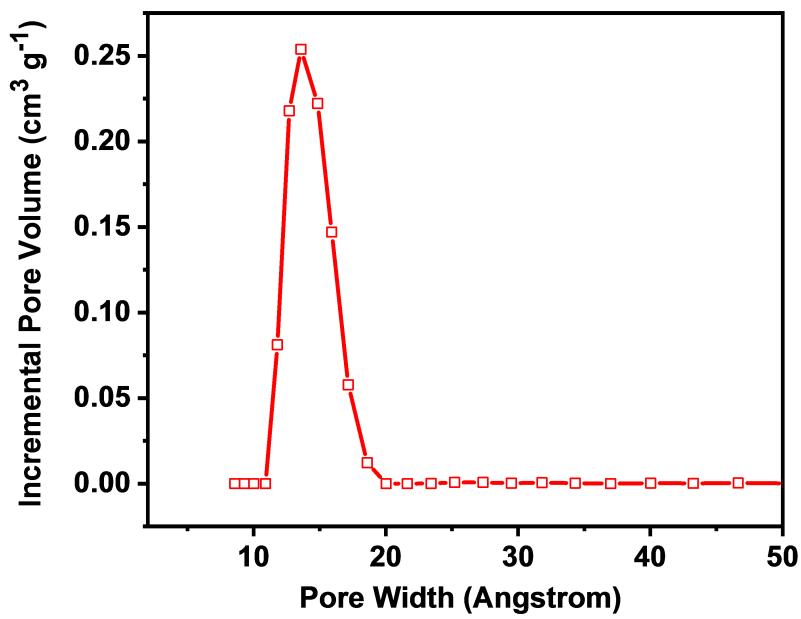


Figure S7 Pore size distribution of MFU-4I obtained from N<sub>2</sub> isotherm at 77 K.

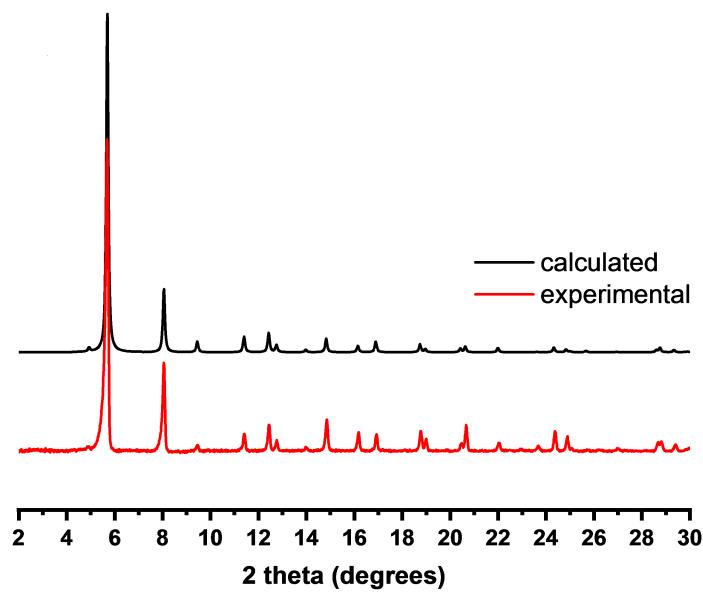
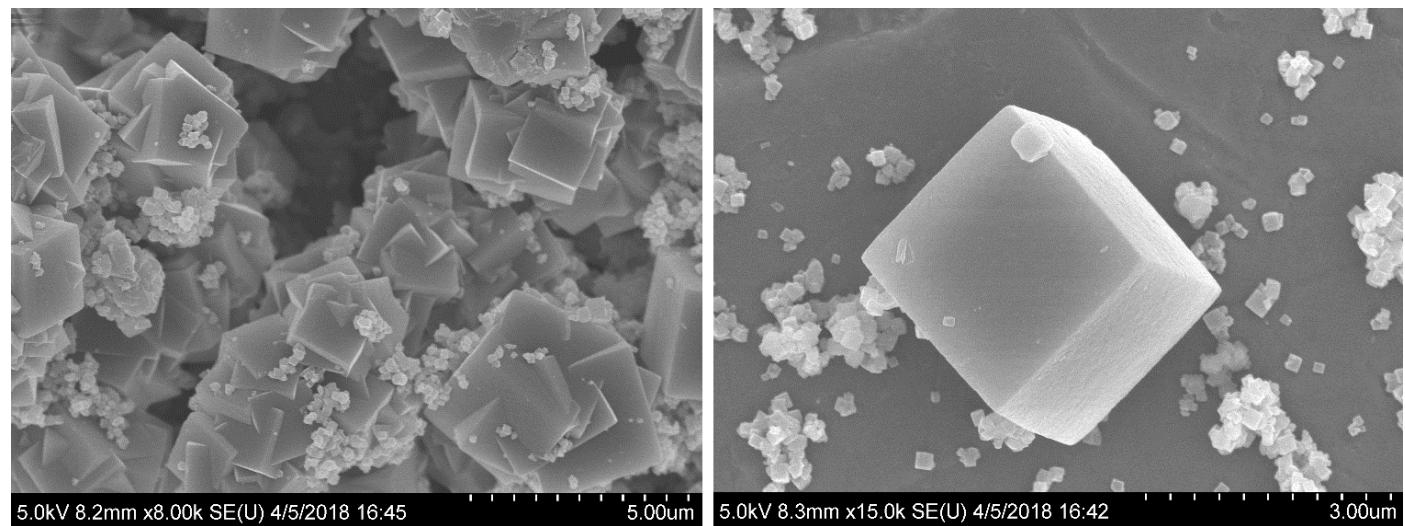


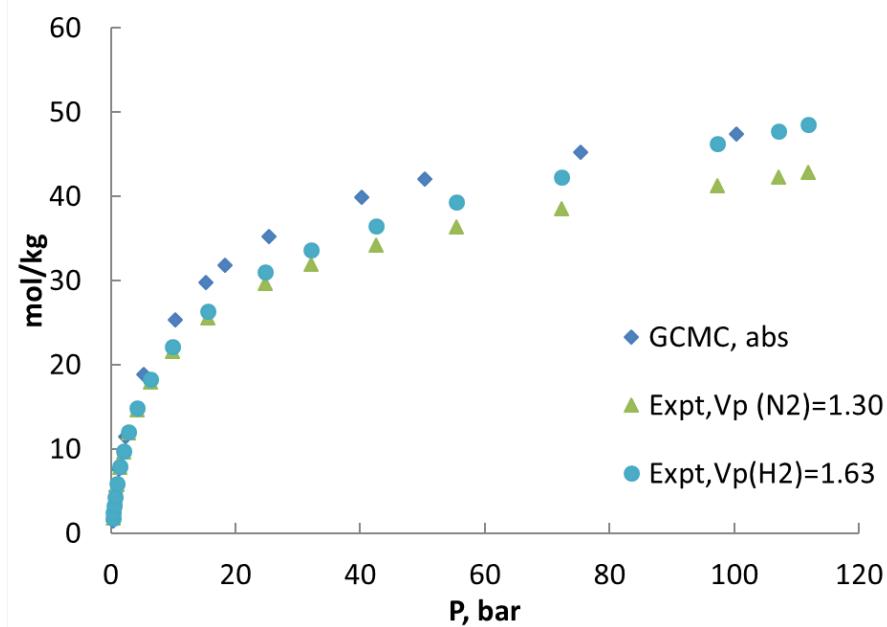
Figure S8 Simulated (top) and experimental (bottom) powder x-ray diffraction patterns of MFU-4I.



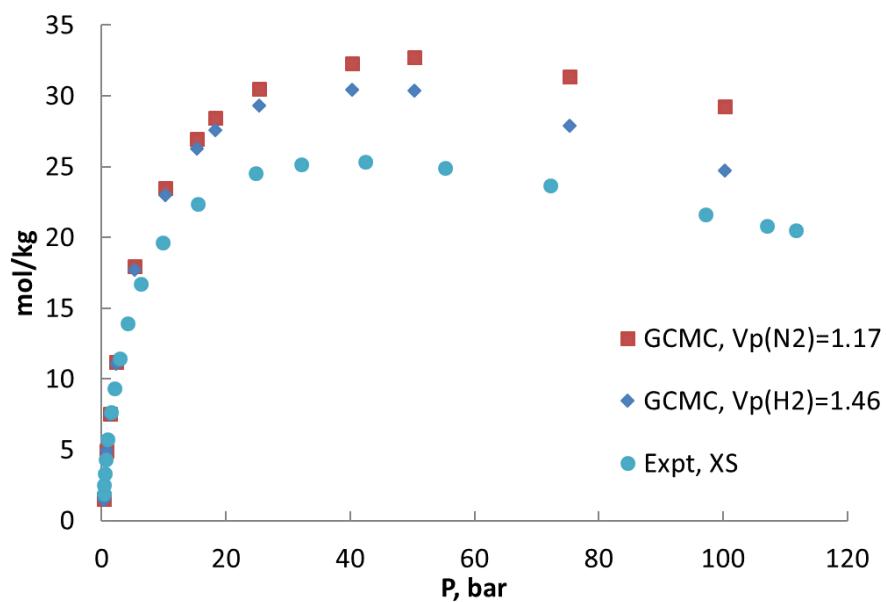
**Figure S9** SEM images of MFU-4I.

### 1.3 Absolute and excess uptake calculations

Experimental measurements of adsorption uptake report excess uptake, while GCMC simulations compute the absolute uptake (that is, the total number of gas molecules in the system). In order to compare these values we convert from absolute to excess or vice versa using  $N_{abs} = N_{xs} + V_{pore} * \rho(T,P)$  where the density is taken from either an equation of state or GCMC simulations of gas in a box at a given temperature and pressure. This calculation is sensitive to the pore volume, which can vary slightly depending on the probe gas used to measure it. Here we have calculated the pore volume with both hydrogen and nitrogen. The data reported in the main text is calculated using the experimental pore volume measured with nitrogen, 1.3 cc/g.



**Figure S10** Absolute H<sub>2</sub> uptake at 77 K calculated with GCMC (blue diamonds) and converted from experimental measurements of excess uptake using pore volumes measured with nitrogen (green triangles) and hydrogen (blue circles).



**Figure S11** Excess H<sub>2</sub> uptake at 77 K measured experimentally (blue circles) and converted from GCMC calculations of absolute uptake using calculated pore volumes from GCMC for nitrogen (red square) and hydrogen (blue diamond).

## 2 Molecular simulation details

### 2.1 Interatomic potential parameters

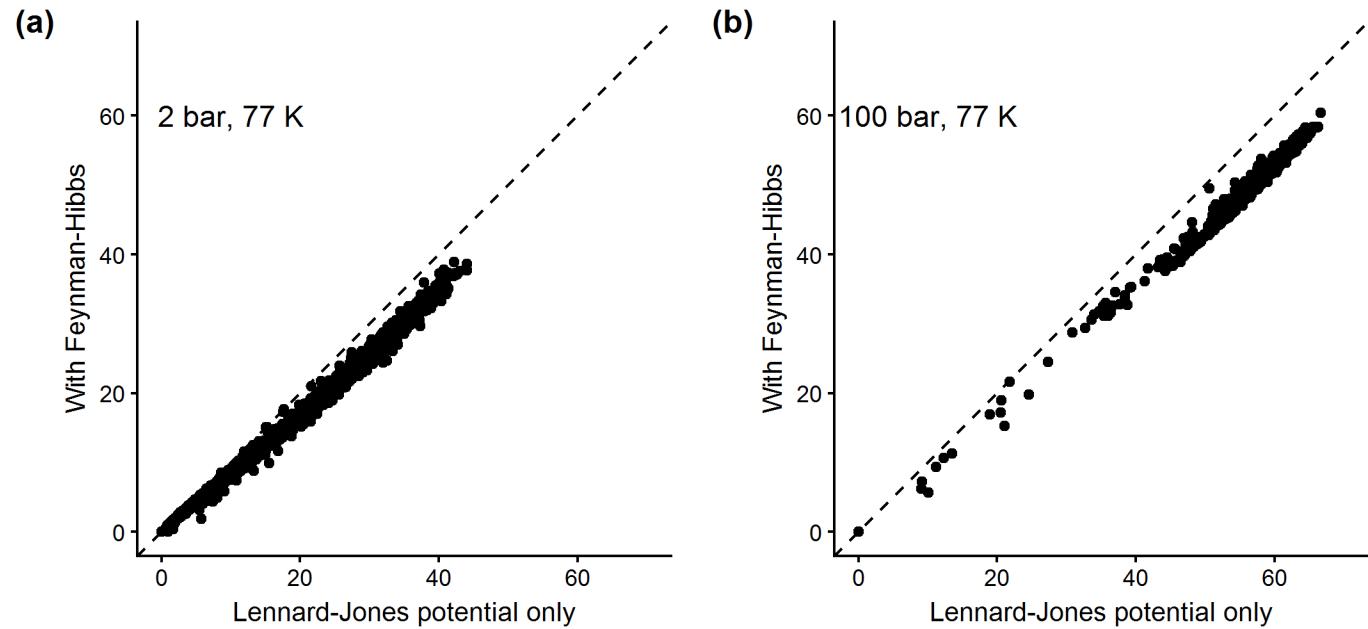
**Table S1** LJ interactions for host framework atoms

Element	$\epsilon/k_B$ [K]	$\sigma$ [\AA]	mass (amu)	Element	$\epsilon/k_B$ [K]	$\sigma$ [\AA]	mass (amu)
Ac	16.600	3.100	227.000	N	34.720	3.260	14.007
Ag	18.110	2.800	107.868	Na	15.090	2.660	22.991
Al	254.090	4.010	26.982	Nb	29.690	2.820	92.906
Am	7.040	3.010	243.000	Nd	5.030	3.180	144.240
Ar	93.080	3.450	39.948	Ne	21.130	2.660	20.180
As	155.470	3.770	74.922	Ni	7.550	2.520	58.693
At	142.890	4.230	210.000	No	5.530	2.890	259.000
Au	19.620	2.930	196.967	Np	9.560	3.050	237.000
B	90.570	3.640	10.811	O	30.190	3.120	15.999
Ba	183.150	3.300	137.327	Os	18.620	2.780	190.230
Be	42.770	2.450	9.012	P	153.460	3.690	30.974
Bi	260.630	3.890	208.980	Pa	11.070	3.050	231.036
Bk	6.540	2.970	247.000	Pb	333.590	3.830	207.200
Br	126.290	3.730	79.904	Pd	24.150	2.580	106.420
C	52.830	3.430	12.011	Pm	4.530	3.160	145.000
Ca	119.750	3.030	40.078	Po	163.520	4.200	210.000
Cd	114.720	2.540	112.411	Pr	5.030	3.210	140.908
Ce	6.540	3.170	140.116	Pt	40.250	2.450	195.078
Cf	6.540	2.950	251.000	Pu	8.050	3.050	244.000
Cl	114.210	3.520	35.453	Ra	203.270	3.280	226.000
Cm	6.540	2.960	247.000	Rb	20.130	3.670	85.468
Co	7.040	2.560	58.933	Re	33.210	2.630	186.207
Cr	7.550	2.690	51.996	Rh	26.670	2.610	102.906
Cs	22.640	4.020	132.905	Rn	124.780	4.250	222.000
Cu	2.520	3.110	63.546	Ru	28.180	2.640	101.070
Dy	3.520	3.050	162.500	S	137.860	3.590	32.066
Er	3.520	3.020	167.260	Sb	225.910	3.940	121.760
Es	6.040	2.940	252.000	Sc	9.560	2.940	44.956
Eu	4.030	3.110	151.964	Se	146.420	3.750	78.960
F	25.160	3.000	18.998	Si	202.270	3.830	28.086
Fe	6.540	2.590	55.845	Sm	4.030	3.140	150.360
Fm	6.040	2.930	257.000	Sn	285.280	3.910	118.710
Fr	25.160	4.370	223.000	Sr	118.240	3.240	87.620
Ga	208.810	3.900	69.723	Ta	40.750	2.820	180.948
Gd	4.530	3.000	157.250	Tb	3.520	3.070	158.925
Ge	190.690	3.810	72.610	Tc	24.150	2.670	98.000
H	22.140	2.570	1.008	Te	200.250	3.980	127.600
He	10.900	2.640	4.003	Th	13.080	3.030	232.038
Hf	36.230	2.800	178.490	Ti	8.550	2.830	47.867
Hg	193.710	2.410	200.590	Tl	342.140	3.870	204.383
Ho	3.520	3.040	164.930	Tm	3.020	3.010	168.934
I	170.570	4.010	126.904	U	11.070	3.020	238.029
In	301.390	3.980	114.818	V	8.050	2.800	50.942
Ir	36.730	2.530	192.217	W	33.710	2.730	183.840
K	17.610	3.400	39.098	Xe	167.040	3.920	131.290
Kr	110.690	3.690	83.800	Y	36.230	2.980	88.906
La	8.550	3.140	138.906	Yb	114.720	2.990	173.040
Li	12.580	2.180	6.941	Zn	62.390	2.460	65.390
Lr	5.530	2.880	266.000	Zr	34.720	2.780	91.224
Lu	20.630	3.240	174.967				
Md	5.530	2.920	258.000				
Mg	55.850	2.690	24.305				
Mn	6.540	2.640	54.938				
Mo	28.180	2.720	95.940				

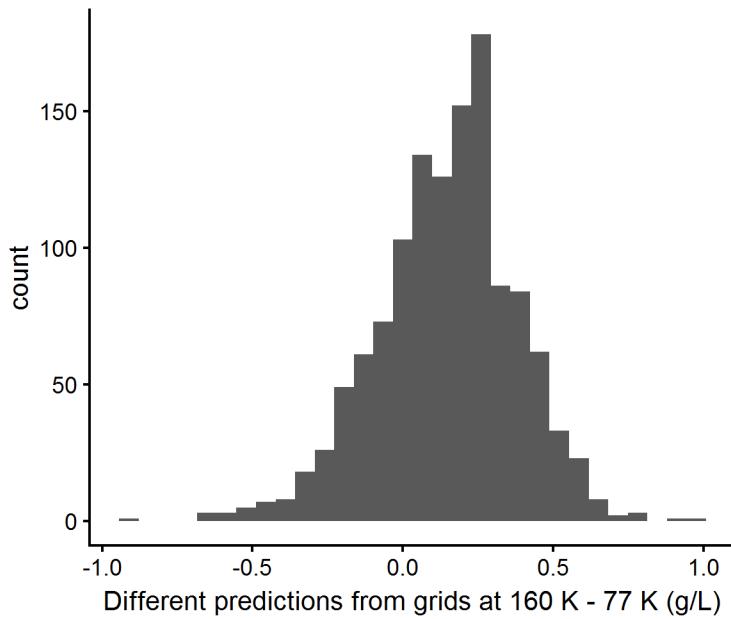
**Table S2** Parameters for guest adsorbate molecules

Pseudo-atom	$\varepsilon/k_B$ [K]	$\sigma$ [ $\text{\AA}$ ]	mass (amu)	$q$
H <sub>com</sub>	36.700	2.958	2.016	-0.936
H <sub>h2</sub>	0.000	0.000	-	0.468
N <sub>com</sub>	0.000	0.000	-	0.8096
N <sub>n2</sub>	36.000	3.310	14.007	-0.4048
CH <sub>4</sub>	148.000	3.730	16.042	-

## 2.2 Feynman-Hibbs effects



**Figure S12** Effect of Feynman-Hibbs corrections on simulated hydrogen uptake in some ToBaCCo MOFs.



**Figure S13** Quantifying the differences between energy grids at 160 K and 77 K, which result from the temperature dependence of the Feynman-Hibbs correction. 1,500 hMOFs were selected for testing. We calculated grids at both temperatures, processed them with the same trained regression model (without refitting), and plotted the difference in predicted deliverable capacity. Overall, the model prediction is generally within 1 g/L at the two temperatures when used interchangeably without retraining the model. Based on these results, we chose to solely use the 77 K grids for training models and making predictions, i.e. the deliverable capacity between 100 bar, 77 K and 5 bar, 160 K.

### 2.3 GCMC Error Bars

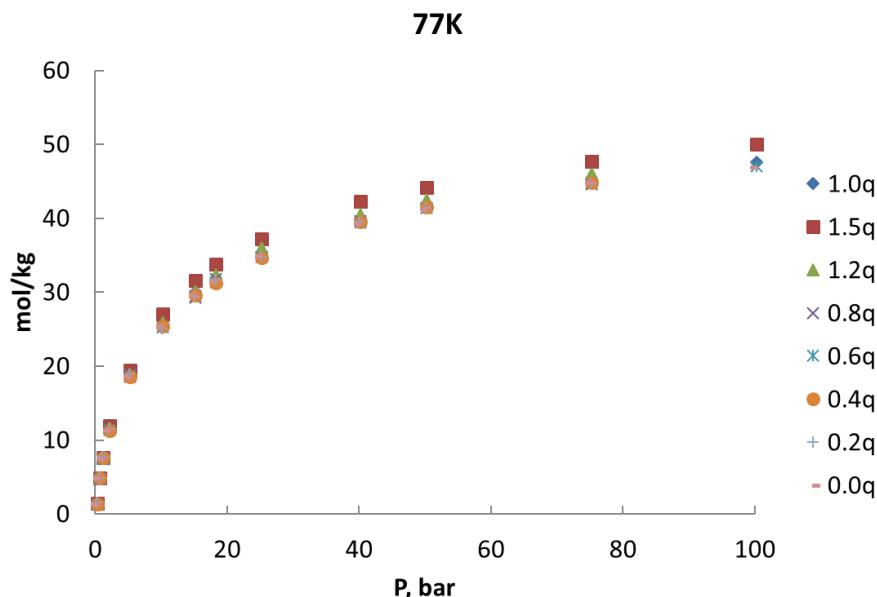
**Table S3** Error statistics for the 1000 MOFs from CCDC for which we performed GCMC simulations.

	mg/g	g/L
Average	0.48	0.51
Median	0.34	0.40
Minimum	0.00	0.00
Maximum	7.03	5.04
Standard deviation	0.64	0.48

## 2.4 Benchmarking the charge on hydrogen model

In this work we used the Darkrim-Levesque model of hydrogen<sup>3</sup> as described in the main text. This hydrogen model uses a single Lennard-Jones site at the center of mass and three charged Coulomb sites. In order to ascertain the effect these charges have on the hydrogen uptake, we ran simulations with various values of the charges while leaving the Lennard-Jones parameters unchanged. The values used for the simulations are shown in Table S4.

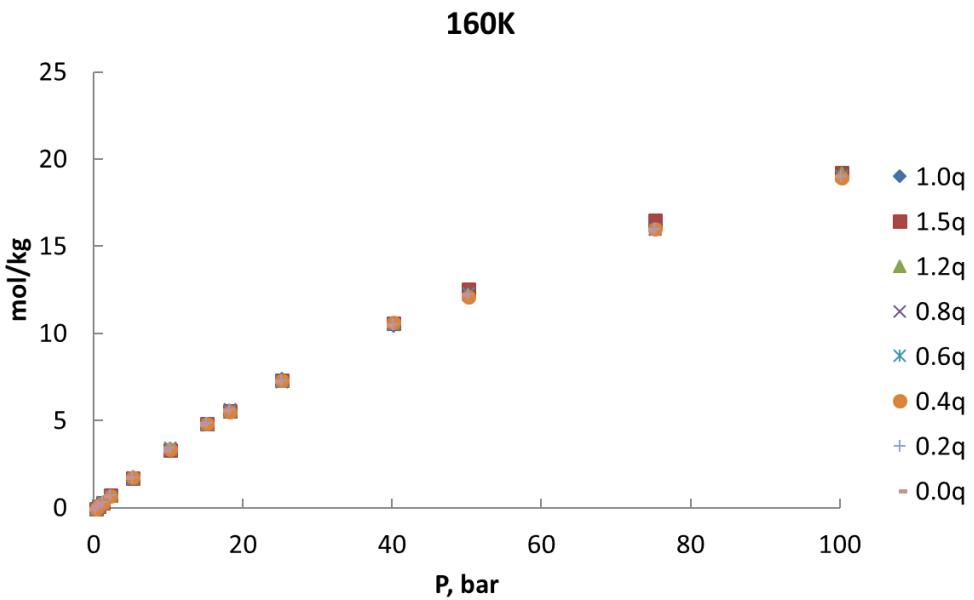
In this work, there are no charges on the MOF frameworks, so the charge on the hydrogen molecule contributes only to the adsorbate-adsorbate interaction. Ultimately, the strength of the charge on the hydrogen model contributes only a little to the total uptake, as shown in Figure S14. Removing the charge entirely reduces the uptake only 0.7 mol/kg or 1.5%. Increasing the strength of the charge by a factor of 1.5 increases uptake by 2.5 mol/kg or about 5%. The hydrogen-MOF interactions dominate over the adsorbate-adsorbate interaction, and the Coulombic contribution is only about five percent of the total adsorbate-adsorbate energy at 77 K (Figure S16). At the higher temperature of 160 K, varying the charge on hydrogen make essentially no difference in the total uptake, as shown in Figure S15.



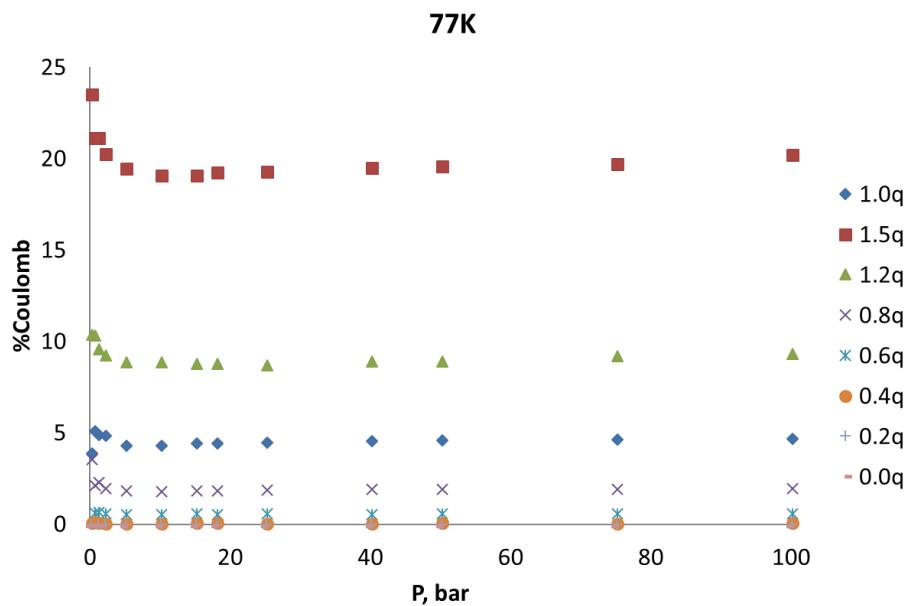
**Figure S14** Calculated hydrogen uptake in mol/kg at 77 K using different charge models for hydrogen molecule as detailed in Table S4. The blue diamonds (1.0q) correspond to the standard Darkrim-Levesque hydrogen model used in this work.

**Table S4** Parameters used for hydrogen simulations. Data set 0 is the standard Darkrim-Levesque model. Units of lengths are in angstroms, and energy is in Kelvin. Units of charges are fundamental charge, e.

Data Set	Sigma, Å (H-H <sub>2</sub> )	Eps, K (H-H <sub>2</sub> )	Sigma, Å (H_com)	Eps, K (H_com)	Scale factor	q (H <sub>2</sub> )	q (com)
0	0	0	2.958	36.7	1.0	0.4680	-0.9360
1	0	0	2.958	36.7	1.5	0.7020	-1.4040
2	0	0	2.958	36.7	1.2	0.5616	-1.1232
3	0	0	2.958	36.7	0.8	0.3744	-0.7488
4	0	0	2.958	36.7	0.6	0.2808	-0.5616
5	0	0	2.958	36.7	0.4	0.1872	-0.3744
6	0	0	2.958	36.7	0.2	0.0936	-0.1872
7	0	0	2.958	36.7	0.0	0.0000	0.0000

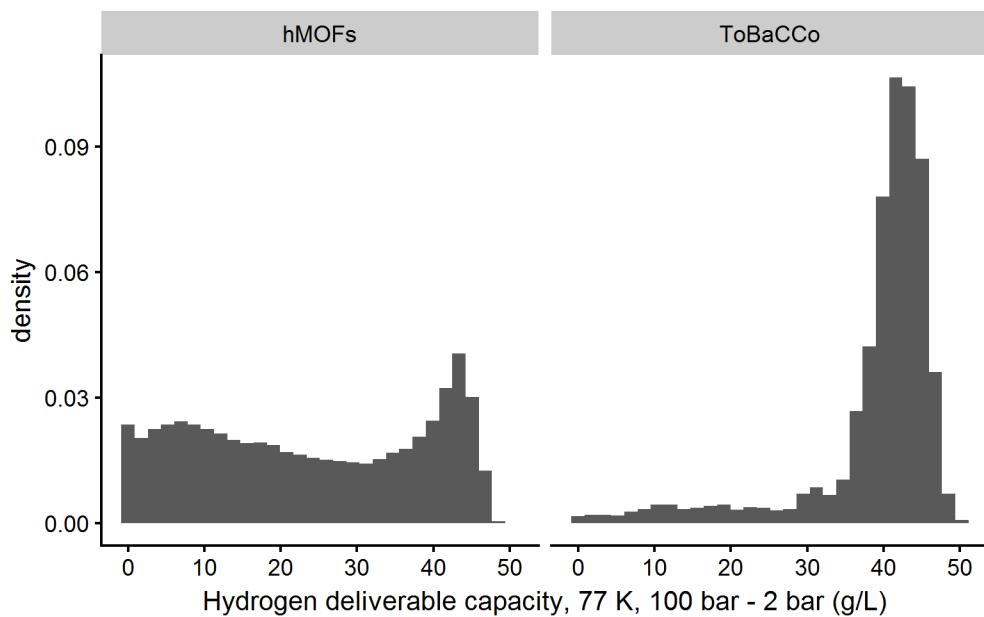


**Figure S15** Calculated hydrogen uptake in mol/kg at 160 K using different charge models for hydrogen molecule as detailed in Table S4. The blue diamonds (1.0q) correspond to the standard Darkrim-Levesque hydrogen model used in this work.

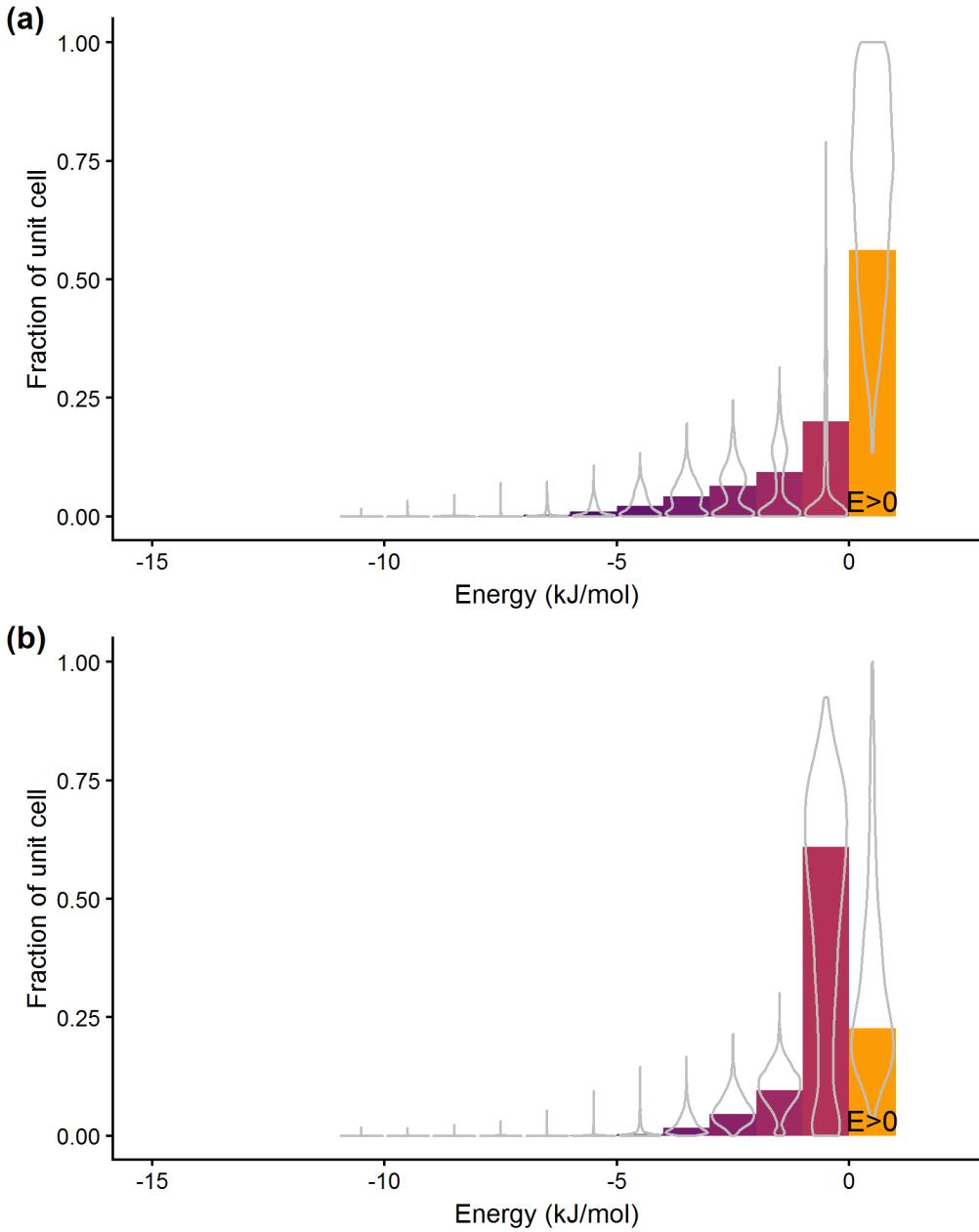


**Figure S16** The percentage of the total adsorbate-adsorbate energy attributed to Coulomb interactions as a function of pressure. The blue diamonds (1.0q) correspond to the standard Darkrim-Levesque hydrogen model used in this work.

### 3 Diversity of properties in MOF databases

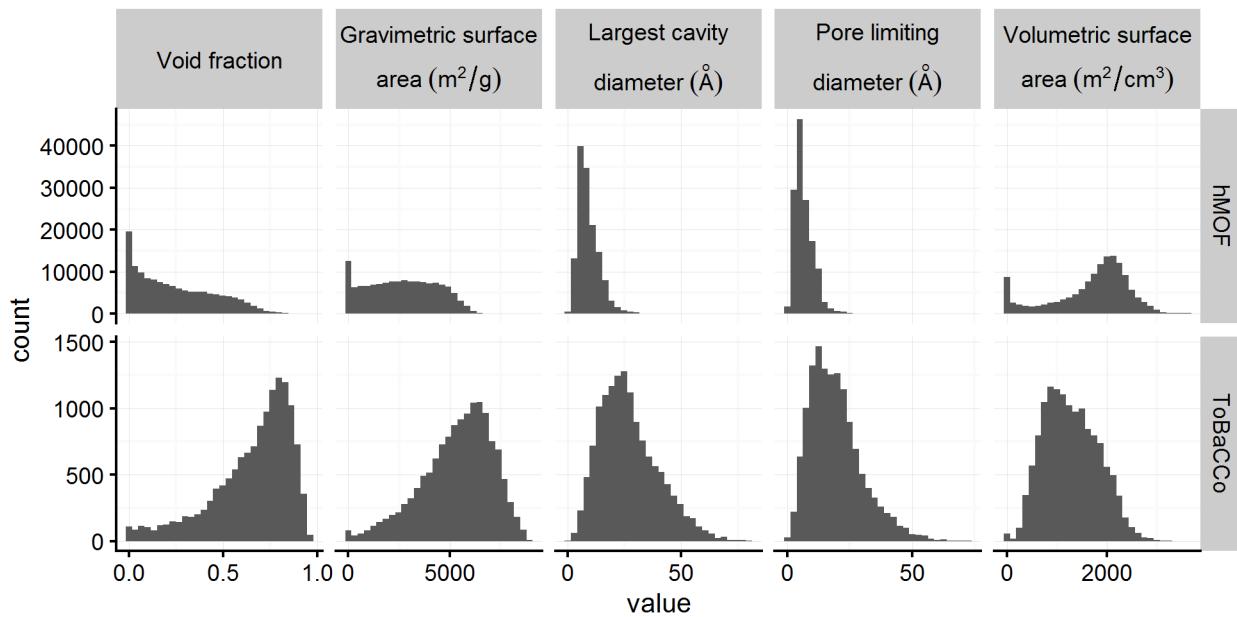


**Figure S17** The hMOFs and ToBaCCo MOFs have different distributions of the response variable  $y$ . Deliverable capacity is calculated from GCMC simulations at 77 K between 100 bar and 2 bar. In this histogram, density refers to a probability density.

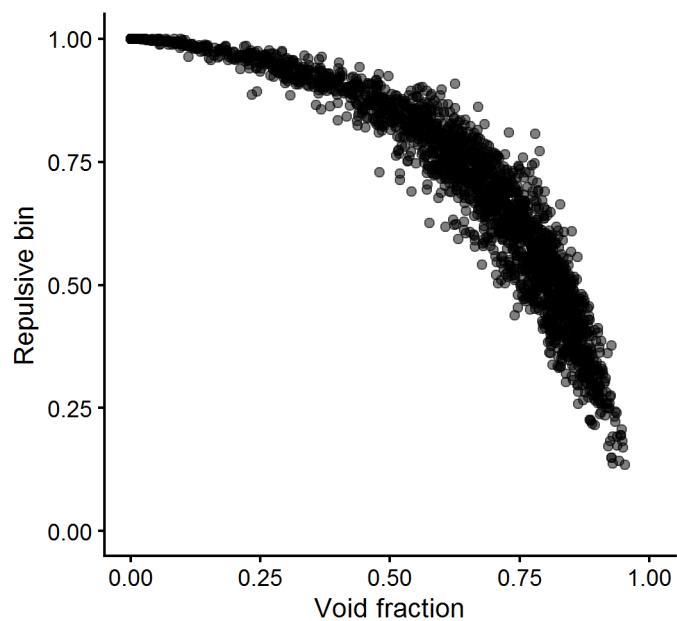


**Figure S18** Variability of energy histogram bins ( $x$  variables in regression) for the (a) hMOF and (b) ToBaCCo databases. Background bars show the mean value of a bin. The superimposed violin plot gives the distribution of bin values for a MOF database. A wider portion of the curve indicates that many MOFs have the histogram bin at a certain value (the curves are like a rotated probability density plot for the possible values of each  $x_i$ ). The maximum width of each "violin" is scaled to be the same as the width of a bar.

For example, we can see that the hMOF database has many MOFs with a large repulsive fraction (low void fraction), shown by the width of the "violin" of the orange bar at high values. By contrast, in the ToBaCCo MOF database the most common repulsive value is approximately 0.25 (corresponding with a void fraction  $\approx 0.75$ ), although the distribution of values also has a long tail including fully repulsive MOFs (no attractive region for hydrogen adsorption). In general, the ToBaCCo MOFs have a greater fraction of the unit cell in the weakly attractive region, which explains the higher average deliverable capacity from the previous figure.



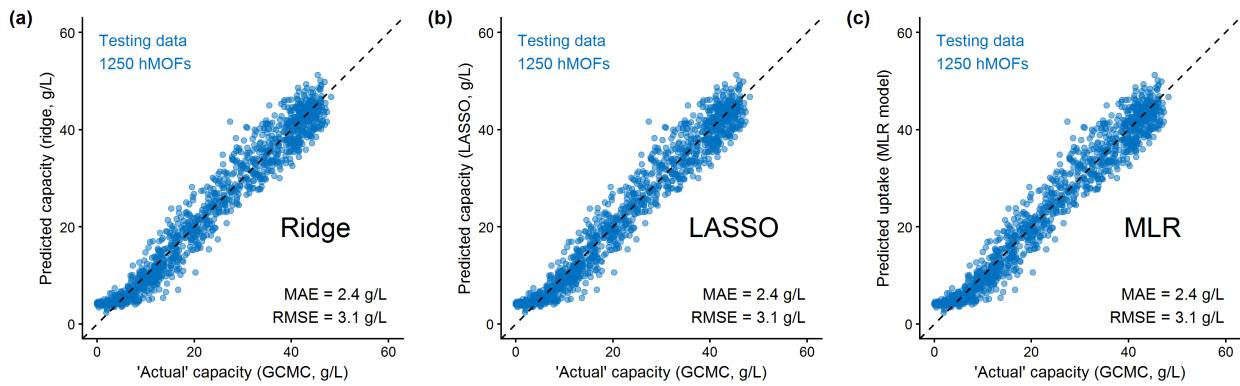
**Figure S19** The hMOFs and ToBaCCo MOFs have different distributions of textural properties. Since the largest PLD in the hMOF database is 24.75  $\text{\AA}$ , it does not successfully train a model to consider large-pore effects found in the ToBaCCo database.



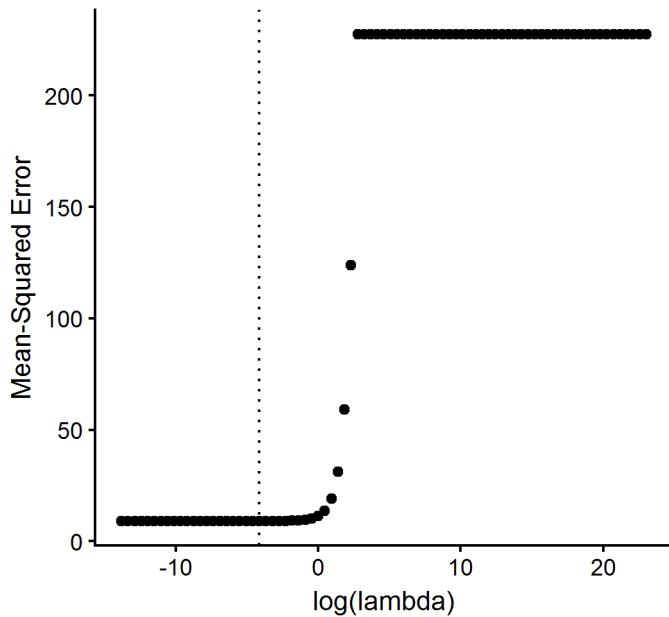
**Figure S20** Inverse relationship between void fraction and the repulsive bin for the hMOFs. There is not an exact relationship because void fraction often uses a different geometric or energetic cutoff than 0 kJ/mol  $\text{H}_2$ , but the plot highlights the general trend that as the top repulsive bin increases, there is less attractive void space remaining.

## 4 Details and tests from the machine learning analysis

### 4.1 Effect of regularization



**Figure S21** Ridge regression, LASSO, and MLR yield consistent model performance



**Figure S22** Determination of the regularization parameter  $\lambda$  for LASSO regression. Mean-squared error is calculated using 10-fold cross-validation (using the `cv.glmnet` function), and we selected the value of  $\lambda$  that minimizes this error (dotted line). We fit models for 100 values of  $\lambda$  spaced between  $1e-2$  and  $1e10$ .

## 4.2 Model coefficients and determination of energy range

**Table S5** Determination of the range of bins considered for the energy histograms and regression models, based on range of energies within the hMOF materials (training set). Counts are the number of MOFs containing a nonzero value for the bin (i.e. at least one grid point with the specified energy) within the 1,000 MOFs randomly selected for training (or in the case of the CCDC database, validation). We selected the lower bound of explicitly calculated bins, -10 kJ/mol, by considering the number of hMOF samples used in training. For example, if the lower cutoff were set to -12 kJ/mol instead of -10 kJ/mol, one of the hMOF  $\beta$  coefficients would be trained with a sample size of only  $n = 7$  representative MOFs.

Bin	CCDC Validation	hMOF	ToBaCCo
-15 to -14 kJ/mol	1	0	1
-14 to -13 kJ/mol	0	1	1
-13 to -12 kJ/mol	3	2	4
-12 to -11 kJ/mol	9	7	8
-11 to -10 kJ/mol	62	28	20
-10 to -9 kJ/mol	159	79	40
-9 to -8 kJ/mol	298	187	97
-8 to -7 kJ/mol	426	365	216
-7 to -6 kJ/mol	512	597	419
-6 to -5 kJ/mol	580	849	741
-5 to -4 kJ/mol	614	954	945
-4 to -3 kJ/mol	644	971	995
-3 to -2 kJ/mol	641	969	996
-2 to -1 kJ/mol	663	962	997
-1 to 0 kJ/mol	668	966	996
0 to Inf kJ/mol	1000	1000	1000

**Table S6** Regression coefficients for the trained models using ridge regression, LASSO, and MLR, and the number of MOFs with nonzero bin values used for training

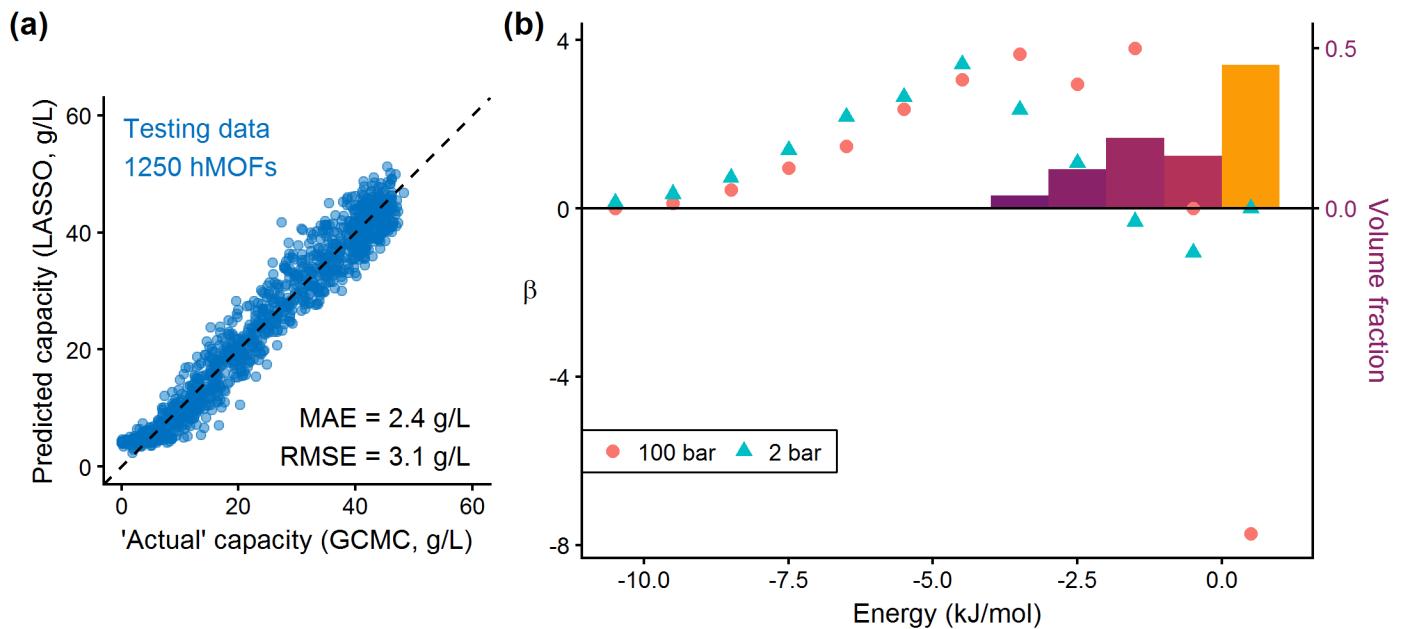
Bin	Ridge	LASSO	MLR	Number of nonzero samples
-Inf to -10 kJ/mol	-163	-157	-123	28
-10 to -9 kJ/mol	-79	-80	-40	79
-9 to -8 kJ/mol	-67	-69	-28	187
-8 to -7 kJ/mol	-43	-45	-4	365
-7 to -6 kJ/mol	-51	-54	-13	597
-6 to -5 kJ/mol	2	0	41	849
-5 to -4 kJ/mol	4	0	42	954
-4 to -3 kJ/mol	59	56	98	971
-3 to -2 kJ/mol	53	49	91	969
-2 to -1 kJ/mol	80	77	118	962
-1 to 0 kJ/mol	24	21	63	966
0 to Inf kJ/mol	-20	-23	18	1000
(Intercept)	24	27	-14	981

**Table S7** Determination of the range of bins considered for the methane energy histograms and regression models, based on range of energies within the hMOF materials. Table is the CH<sub>4</sub> equivalent to Table S5.

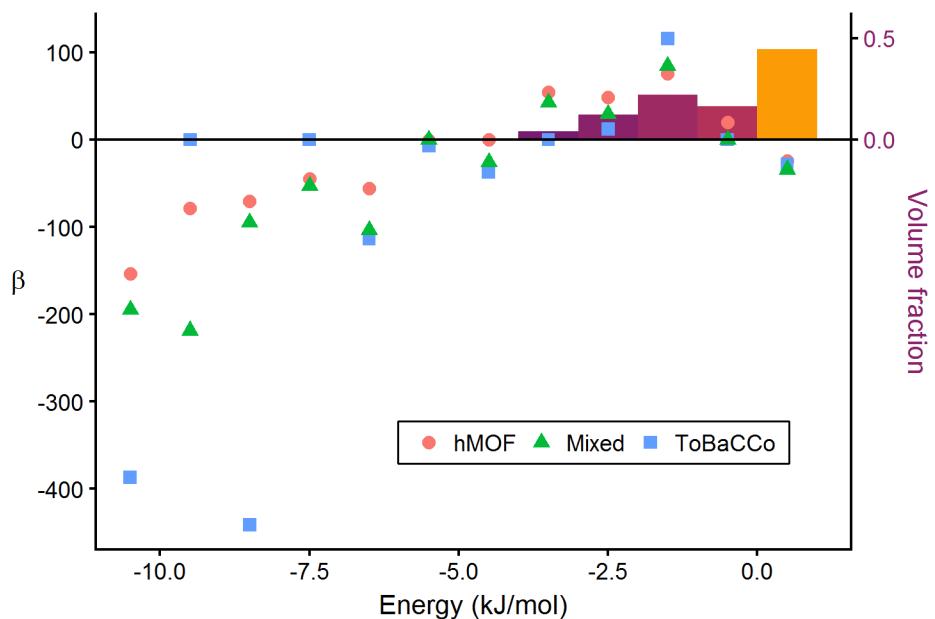
Bin	nsample
-36 to -34 kJ/mol	1
-34 to -32 kJ/mol	3
-32 to -30 kJ/mol	4
-30 to -28 kJ/mol	8
-28 to -26 kJ/mol	37
-26 to -24 kJ/mol	73
-24 to -22 kJ/mol	157
-22 to -20 kJ/mol	272
-20 to -18 kJ/mol	426
-18 to -16 kJ/mol	583
-16 to -14 kJ/mol	770
-14 to -12 kJ/mol	895
-12 to -10 kJ/mol	954
-10 to -8 kJ/mol	966
-8 to -6 kJ/mol	973
-6 to -4 kJ/mol	965
-4 to -2 kJ/mol	970
-2 to 0 kJ/mol	971
0 to Inf kJ/mol	1000

**Table S8** Regression coefficients for the trained methane model using LASSO, and the number of MOFs with nonzero bin values used for training (CH<sub>4</sub> equivalent to Table S6).

Bin	beta	Number of nonzero samples
-Inf to -26 kJ/mol	-1937	39
-26 to -24 kJ/mol	-471	73
-24 to -22 kJ/mol	-931	157
-22 to -20 kJ/mol	-418	272
-20 to -18 kJ/mol	-101	426
-18 to -16 kJ/mol	0	583
-16 to -14 kJ/mol	68	770
-14 to -12 kJ/mol	322	895
-12 to -10 kJ/mol	419	954
-10 to -8 kJ/mol	452	966
-8 to -6 kJ/mol	330	973
-6 to -4 kJ/mol	241	965
-4 to -2 kJ/mol	253	970
-2 to 0 kJ/mol	0	971
0 to Inf kJ/mol	-25	1000
(Intercept)	66	1000

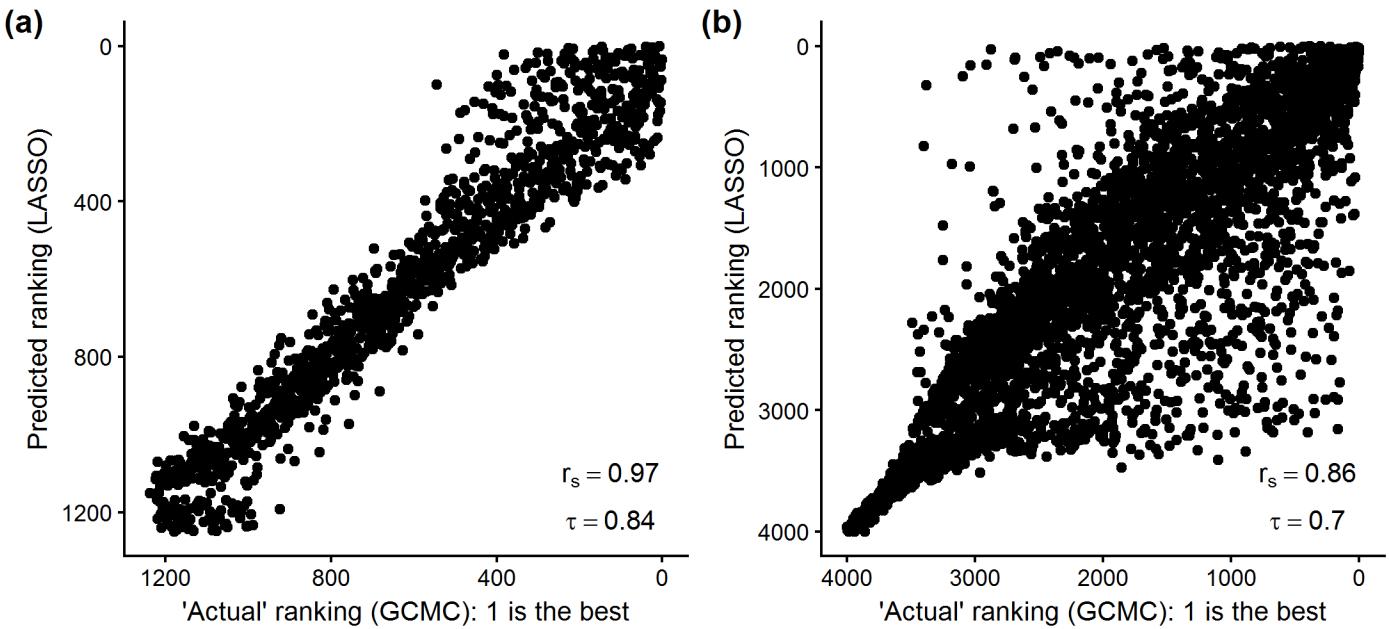


**Figure S23** Generating a hydrogen model on the hMOF database using the z-scores of the histogram bins instead of the direct values of the bins. (a) Parity plot for testing data and (b)  $\beta$  coefficient plot, whose effect size (magnitude of  $\beta$ ) tends toward zero at large values of energy.



**Figure S24** Comparison between LASSO regression  $\beta$  coefficients trained on the hMOFs, ToBaCCo, and a mixed training set. The background energy histogram is the average over the hMOF database.

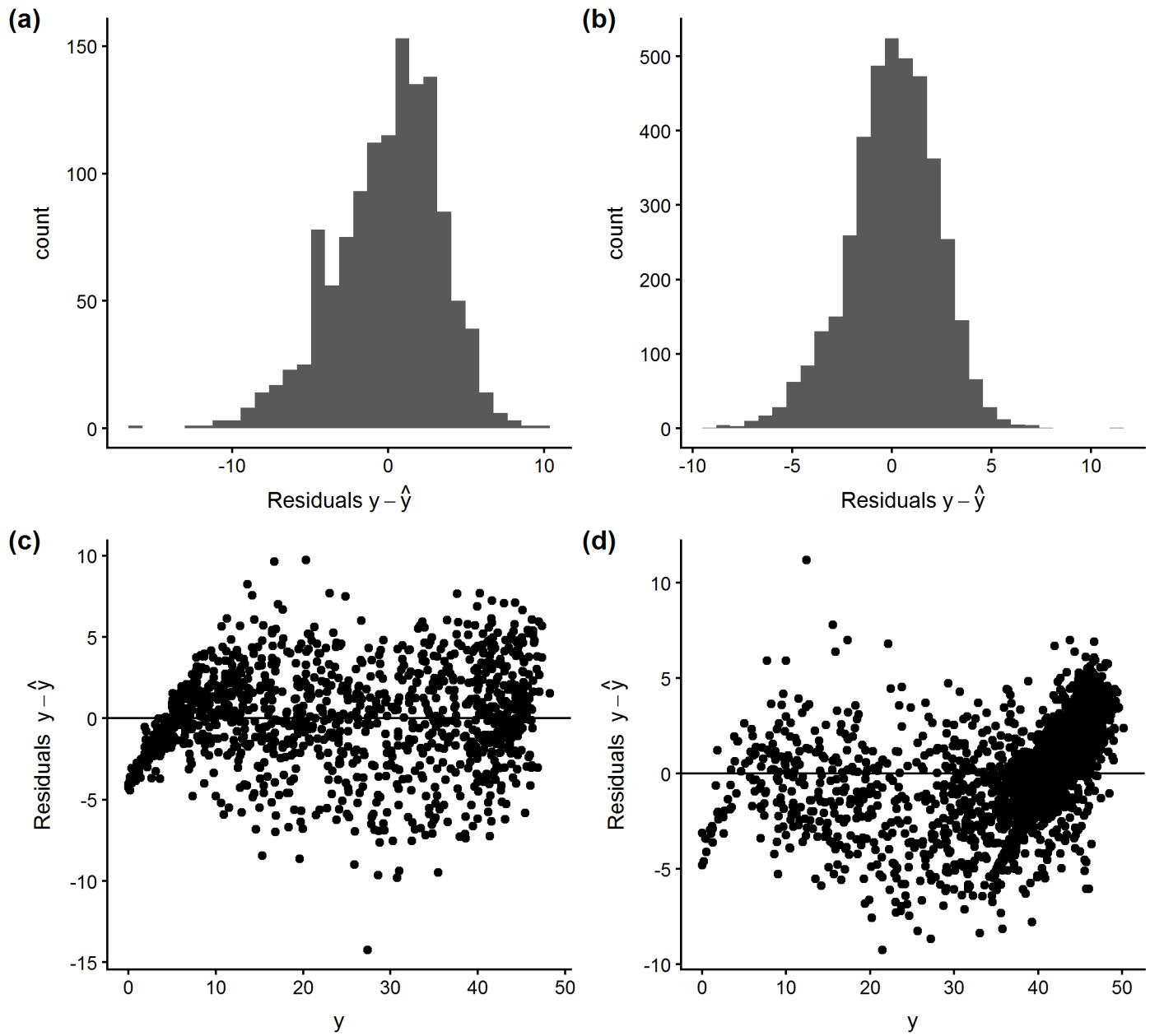
#### 4.3 Detailed model performance



**Figure S25** Performance of the trained “mixed” model to rank candidates from testing data in the (a) hMOFs and (b) ToBaCCo MOFs. The Spearman rank correlation coefficient  $r_s$  and the Kendall rank correlation coefficient  $\tau$  confirm that the mixed model can accurately rank order the databases according to their hydrogen deliverable capacity. The ToBaCCo database is more difficult to rank because of a narrower distribution of deliverable capacities (85% of the ToBaCCo MOFs have a deliverable capacity in the range from 35 g/L to 50 g/L. See also Figure S17).

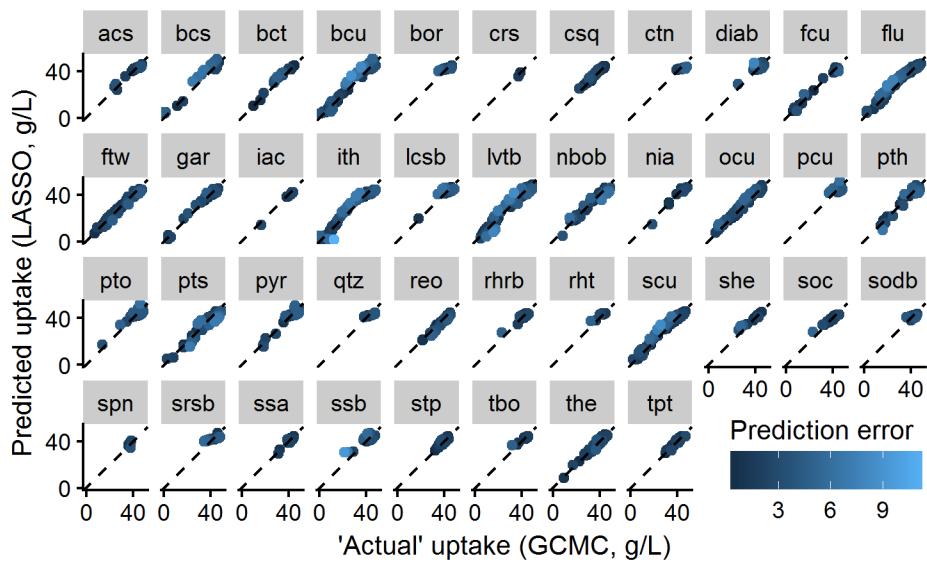
**Table S9** Using different amounts of training data to fit the sparse regression model, with statistics from a separate set of testing data.

Number of MOFs	RMSE	MAE	$r_{spearman}$	$r_{kendall}$
30	5.1	3.7	0.93	0.78
50	3.5	2.8	0.97	0.83
75	3.3	2.6	0.97	0.84
100	3.4	2.6	0.97	0.84
150	3.2	2.6	0.97	0.86
200	3.2	2.6	0.97	0.85
300	3.1	2.5	0.97	0.85
400	3.1	2.4	0.97	0.86
500	3.1	2.5	0.97	0.86
1000	3.1	2.4	0.97	0.86

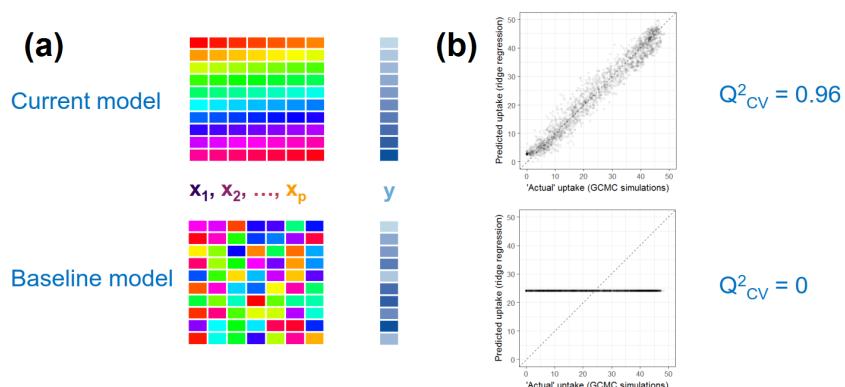


**Figure S26** Diagnostic plots for residuals from using the mixed regression model to predict capacity in the hMOF and ToBaCCo databases.

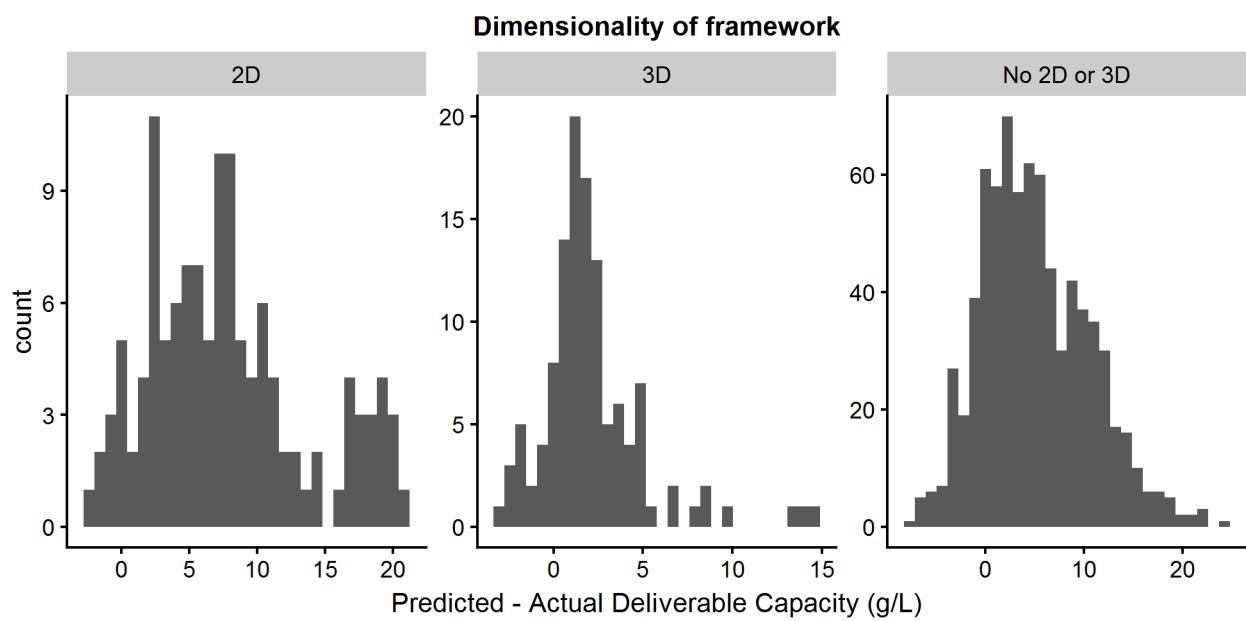
Histograms for the (a) hMOF and (b) ToBaCCo databases confirm that the residuals approximately follow a normal distribution. The residual plots for the (c) hMOF and (d) ToBaCCo databases do not show much systematic difference in error at different values of the response variable.



**Figure S27** Assessing hydrogen uptake for different topologies in the ToBaCCo MOFs. The accuracy of the mixed model is consistent across topologies. The actual amount of gas uptake, however, shows topological dependence, which corroborates results by Colón, Gómez-Gualdrón, and coworkers.<sup>4,5</sup>

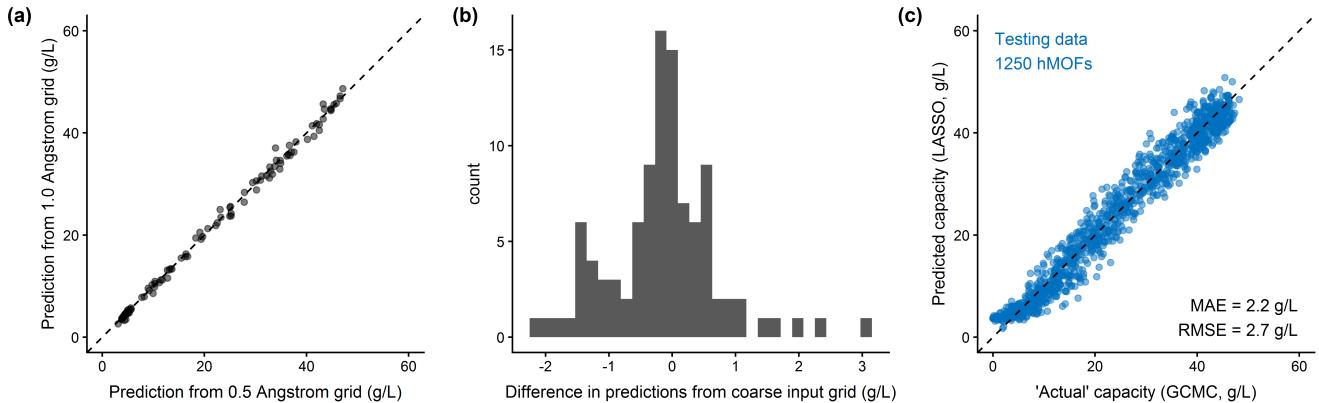


**Figure S28** Determining a baseline for  $Q^2$  using randomized data. (a) cartoon depicting randomization of the original  $x$  and  $y$  data. Colors are a visual guide to illustrate the shuffling of data points and are not indicative of their values. (b) Parity plots for the original and randomized data. The results indicate that the model is not learning spurious correlations, since the model has no predictive power when the variables are disassociated.



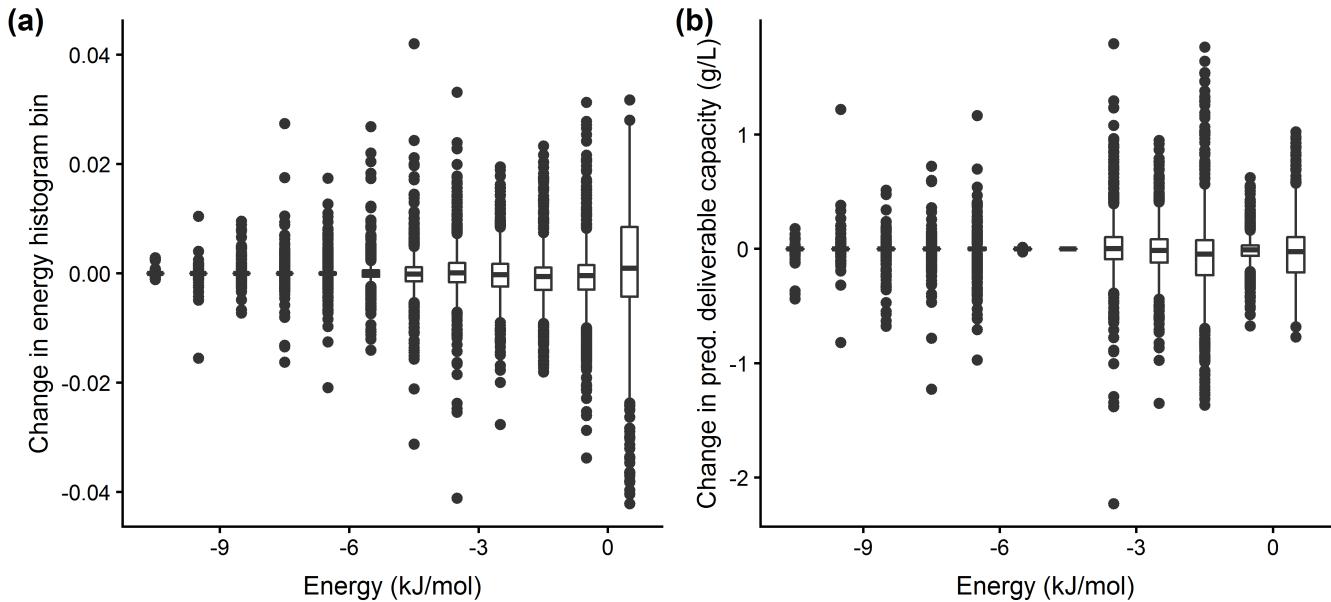
**Figure S29** Residuals for the top 500 CCDC MOFs predicted by the regression model compared against GCMC simulations. Dimensionality of the framework was calculated using Zeo++.

#### 4.4 Coarse energy grid sampling



**Figure S30** Testing the effect of using a finer grid spacing than 1.0 Angstrom for calculating the energy histogram and regression predictions.

(a) Parity plot between model predictions using the 1.0 Angstrom grid spacing and a finer 0.5 Angstrom grid. The calculations are sufficiently converged, especially for ranking purposes, based on the proximity of the points to the parity line. (b) Histogram showing the difference between the two coordinates plotted in part a. The error from using a somewhat coarse grid of 1.0 Angstrom spacing is not substantial given the other errors embedded in the regression approach, which also neglect spatial dependence, explicit guest-guest interactions, etc. The differences reported in this histogram are much smaller than the overall model residuals, shown in Figure S26. (c) Training and testing a model using the 0.5 Angstrom grid spacing. The finer resolution is more accurate than the 1.0 Angstrom model used in the paper (compare against Figure 4 in the main text). However, increasing the resolution from 1.0 Angstrom to 0.5 Angstrom increases the computational requirements by eight-fold, because there is a factor of two in each of the three Cartesian directions. The 1.0 Angstrom grid spacing is sufficiently accurate for screening at a lower computational cost.



**Figure S31** Differences in the energy histogram results by bin, defined as the results from a 1.0 Angstrom grid spacing minus the results with a finer 0.5 Angstrom grid, calculated on 1,000 hMOFs. The boxplots for each bin show the distribution of changes (often called the “five number summary”) between these two settings. The lower edge, middle line, and top edge of the boxes give the first quartile ( $Q_1$ ), median, and third quartile ( $Q_3$ ), respectively. Thus, the box contains the middle 50% of the data. The interquartile range (IQR) is calculated as  $Q_3 - Q_1$ . Explicitly plotted points are outliers, defined as points more than  $1.5 * IQR$  away from the box. Lines, or “whiskers”, connect the boxes to all non-outlier points. (a) Using a finer grid spacing does not have a large effect on the value of the energy histogram bins, which are fed into the model. The x-axis depicts the location of the bin and the y-axis is the change in the value of the feature, which was the proportion of the MOF unit cell with the specified host-guest interactions. Based on the figure, the coarse energy grids are sufficient to obtain accuracy with 2% of the unit cell (a raw value of 0.02) for most MOFs. (b) The discrepancies between the 1 Angstrom energy grid and a finer, “true” distribution will ultimately propagate into model predictions. We calculated these values by multiplying the results from Subfigure a with the LASSO coefficients in Table S6. This subfigure shows that the using a coarse energy sampling often has a negligible effect on the model prediction, as seen by the size of the boxes and whiskers. Except for some outliers, nearly all of the bins individually have an effect of less than 1 g/L, which supports the evidence from Figure S30 that while the coarse energy grid may have some effect on model predictions for a MOF, the effect is small enough to still allow rapid screening.

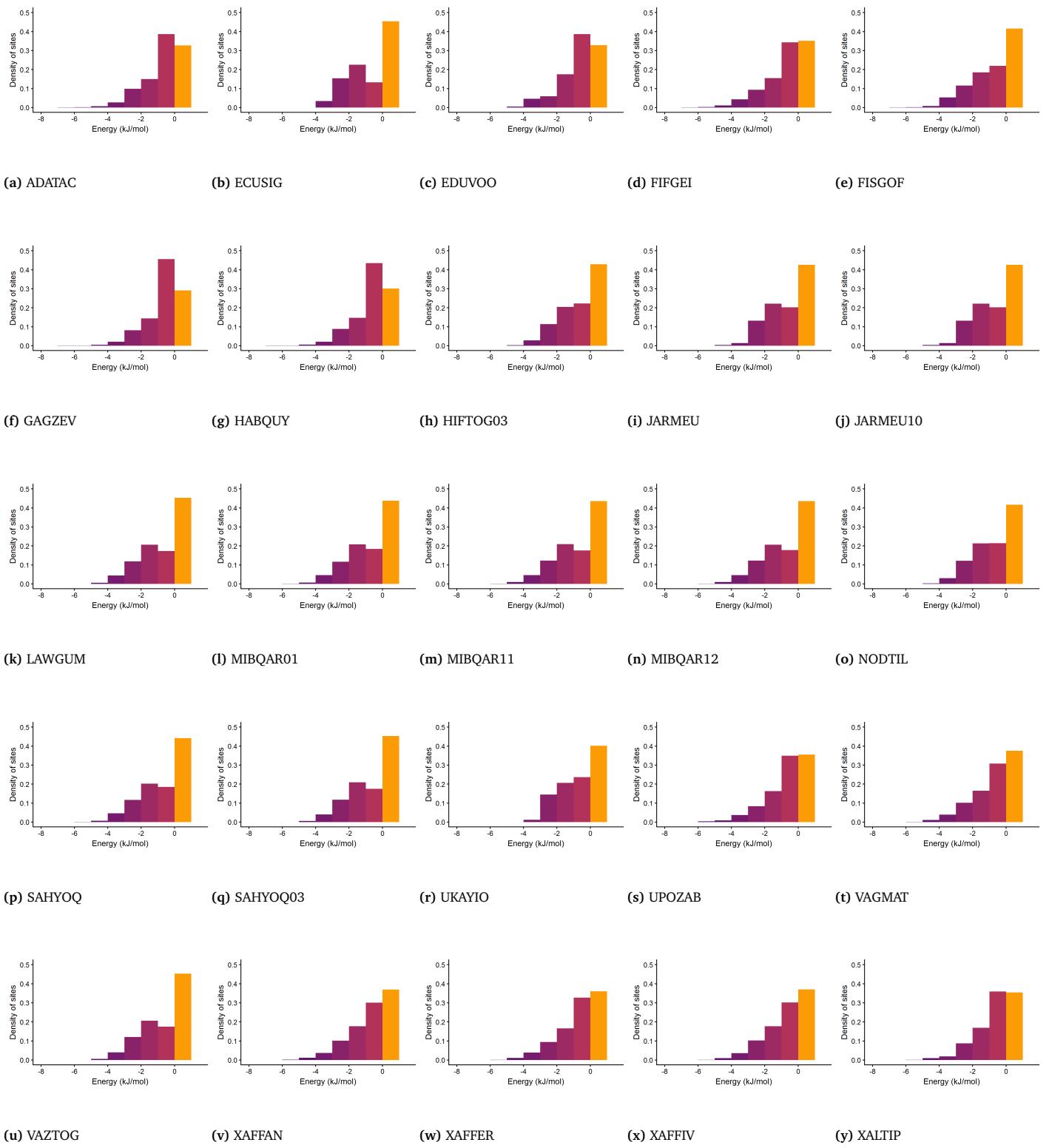
## 5 Top 25 Candidates from Screening

**Table S10** The uptake and deliverable capacity data for the top 25 candidates identified in this screening. Ranking is based on deliverable capacity between 100 bar, 77 K and 2 bar, 77 K from GCMC

MOF ID	Uptake, g/L 2 bar, 77 K	Uptake, g/L 100 bar, 77 K	Uptake, g/L 5 bar, 160 K	Del. Cap g/L 77 K, 77 K	Model Prediction Del Cap, g/L, 77 K, 77 K	Del. Cap 77 K, 160 K
UKAYIO	7.3	57.0	1.5	49.7	47.4	55.5
ECUSIG	10.0	57.9	1.7	47.9	48.4	56.2
VAGMAT	10.6	58.0	1.8	47.4	44.4	56.2
XAFFIV	10.1	57.0	1.7	47.0	45.6	55.3
XAFFAN	10.6	57.5	1.8	46.9	45.6	55.7
HABQUY	7.7	54.5	1.6	46.8	44.4	52.9
JARMEU10	7.9	54.7	1.5	46.7	47.5	53.1
JARMEU	8.0	54.7	1.6	46.7	47.5	53.1
XAFFER	10.5	57.0	1.8	46.6	44.9	55.3
GAGZEV	7.3	53.8	1.6	46.5	44.2	52.2
HIFTOG03	9.1	55.3	1.6	46.3	46.0	53.7
XALTIP	10.3	56.5	1.8	46.2	44.3	54.7
UPOZAB	11.0	57.2	1.9	46.2	44.3	55.3
LAWGUM	11.9	57.9	1.8	46.0	46.1	56.0
VAZTOG	11.5	57.4	1.8	45.9	46.0	55.6
ADATAC	9.3	55.2	1.7	45.9	44.2	53.5
FIFGEI	11.6	57.5	2.0	45.9	44.3	55.6
FISGOF	12.7	58.6	2.0	45.9	45.7	56.6
MIBQAR01	11.8	57.6	1.9	45.8	46.8	55.7
SAHYOQ	12.3	58.1	1.8	45.8	46.2	56.3
MIBQAR11	12.9	58.7	1.9	45.8	47.0	56.8
MIBQAR12	12.8	58.6	2.0	45.8	46.8	56.7
SAHYOQ03	11.8	57.5	1.9	45.7	46.2	55.6
EDUVOO	10.7	56.4	1.8	45.7	46.1	54.6
NODTIL	10.5	56.2	1.8	45.7	47.5	54.5

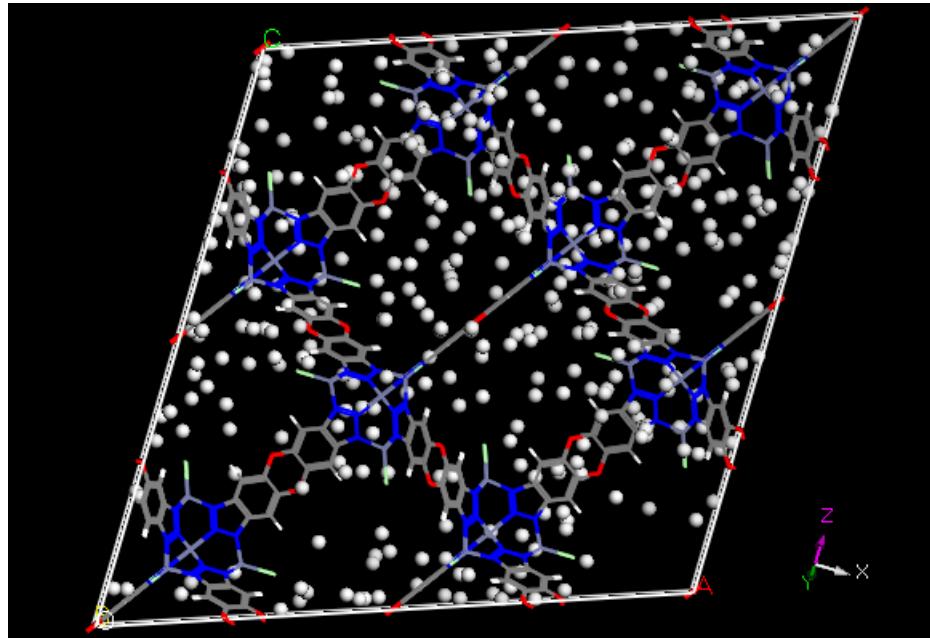
**Table S11** Textural properties for top 25 candidates based on deliverable capacity. LCD stands for Largest Cavity Diameter and is the largest sphere that can be included in the pore. PLD stands for Pore Limiting Diameter and is the diameter of the largest sphere that freely move throughout the MOF. The void fraction is computed using Widom insertions of helium. The surface area is computed by rolling a nitrogen molecule along the surface of the MOF.

MOF ID	LCD Ang	PLD Ang	Void Fraction He	Surf Area m <sup>2</sup> /g	Surf. Area m <sup>2</sup> /cm <sup>3</sup>
UKAYIO	10.7	9.8	0.87	3417	2131
ECUSIG	10.0	8.2	0.85	3484	2320
VAGMAT	14.9	13.3	0.89	5287	1928
XAFFIV	14.2	13.2	0.88	5390	1932
XAFFAN	14.9	13.2	0.88	5284	1930
HABQUY	25.7	12.1	0.91	5761	1667
JARMEU10	11.1	8.3	0.83	6092	2254
JARMEU	11.1	8.3	0.83	6092	2254
XAFFER	14.2	13.3	0.89	5283	1901
GAGZEV	28.6	11.5	0.91	5803	1620
HIFTOG03	15.0	7.9	0.81	3754	2163
XALTIP	18.7	10.4	0.87	3667	1790
UPOZAB	18.7	10.4	0.88	3599	1787
LAWGUM	15.1	7.9	0.82	3837	2213
VAZTOG	15.1	7.9	0.81	3777	2182
ADATAC	26.3	10.3	0.90	5186	1749
FIFGEI	16.2	14.7	0.91	4328	1760
FISGOF	15.8	8.6	0.88	4680	2032
MIBQAR01	15.1	8.0	0.82	3835	2216
SAHYOQ	15.1	8.0	0.81	3768	2183
MIBQAR11	15.0	8.0	0.83	3818	2224
MIBQAR12	15.0	8.0	0.82	3814	2224
SAHYOQ03	15.1	8.0	0.82	3827	2210
EDUVOO	20.9	10.6	0.90	4893	1802
NODTIL	14.8	8.0	0.83	3635	2156



**Figure S32** Energy histograms for top 25 candidates found in this screening.

## 6 Screenshot of Hydrogen in MOF



**Figure S33** Visualized snapshot of hydrogen molecules inside MFU-4I at 77 K and 5 bar.

## 7 MOFs used for model fitting

MOFs for training and validation data were selected from their parent databases using a simple random sample. The results and conclusions of this study were independent of the random seed established for selecting the random samples. Nevertheless, this section includes the lists of MOFs used for training the models reported in this work.

### 7.1 hMOF database

The hypothetical MOF database (denoted “hMOFs” in this work, or WLLFHS database in another ref.<sup>6</sup>) contains 137,953 structures generated using a top-down construction algorithm.<sup>7</sup> We randomly selected 1,000 MOFs for training, which are labeled here by their identification number.

1000106, 1000325, 1000533, 1000844, 1000965, 1001250, 1001374, 1001561, 1001809, 1001873, 1001966, 1002243, 1002355, 1002390, 1002849, 1002896, 1002948, 1002986, 1003516, 1003556, 1003695, 1003754, 1003781, 1004258, 1004379, 1004630, 1004665, 10077, 10291, 10352, 10370, 10532, 10628, 10695, 10741, 10836, 10889, 10916, 10940, 10992, 11039, 11062, 1117, 11425, 11610, 11792, 11822, 12345, 12456, 12548, 12743, 12996, 13115, 13185, 13306, 13423, 1368, 14010, 14148, 14190, 1423, 14297, 14383, 14438, 14479, 14516, 15120, 15185, 15209, 15359, 15529, 15610, 15753, 15837, 15896, 16001, 16033, 16059, 16090, 162, 16267, 16329, 16482, 16595, 16776, 16844, 16974, 17148, 17276, 17498, 17524, 1775, 1783, 17891, 18036, 18076, 18437, 18507, 1881, 19315, 19342, 19589, 1987, 19893, 1997, 2000046, 2000124, 2000183, 2000439, 2000574, 2000724, 2000862, 20061, 20228, 20338, 20504, 20529, 20649, 20695, 20803, 21006, 21233, 21258, 21281, 21404, 21497, 21521, 21737, 21854, 2187, 22011, 22420, 22666, 23019, 23310, 23390, 23467, 23508, 23616, 23761, 23789, 23880, 23901, 23952, 24031, 24367, 24427, 24491, 24556, 24601, 24760, 24805, 2484, 24916, 24952, 24992, 25122, 25209, 25319, 25507, 25894, 25935, 25960, 26021, 26084, 2618, 26289, 26387, 26477, 26499, 26576, 2662, 26768, 26972, 27032, 27228, 27330, 27479, 27602, 27755, 27772, 27885, 28240, 28334, 28457, 28496, 28683, 28874, 29046, 29552, 29782, 29819, 29845, 30000, 3000204, 3000295, 3000483, 3000549, 3000673, 3000717, 3000736, 3000769, 3000833, 3000987, 3001052, 3001292, 3001799, 30049, 30307, 30330, 30441, 30706, 30798, 30823, 30853, 30901, 31212, 31320, 31418, 31477, 31557, 31759, 31848, 31946, 31982, 32113, 3230, 32367, 32390, 32461, 32519, 32604, 32693, 32957, 3303, 33064, 33156, 33548, 33974, 34026, 34045, 34135, 34410, 34477, 34642, 3481, 34859, 34921, 35011, 35047, 35065, 35442, 35477, 35521, 35750, 35815, 35895, 35942, 36045, 36238, 36399, 36630, 36949, 37293, 37676, 3809, 3881, 3920, 4000056, 4000151, 4000208, 4087, 4123, 4267, 4371, 4494, 4672, 4718, 4741, 5000366, 5000407, 5000488, 5000677, 5000755, 5000961, 5001077, 5001111, 5001207, 5001382, 5001405, 5001490, 5001595, 5001730, 5001821, 5001928, 5002023, 5002127, 5002165, 5002202, 5002306, 5002444, 5002558, 5002679, 5002752, 5003105, 5003254, 5003303, 5003408, 5003416, 5003673, 5003891, 5003910, 5004308, 5004620, 5004665, 5004713, 5004903, 5004983, 5005061, 5005132, 5005347, 5005394, 5005539, 5005690, 5005791, 5005836, 5005878, 5005938, 5006058, 5006122, 5006178, 5006327, 5006648, 5006711, 5006770, 5006800, 5006901, 5006971, 5007237, 5007287, 5007508, 5007828, 5007898, 5008041, 5008264, 5008330, 5008374, 5008401, 5008440, 5008747, 5008835, 5008881, 5009522, 5009647, 5009718, 5009744, 5009819, 5010433, 5010533, 5010639, 5010984, 5011044, 5011124, 5011243, 5011463, 5011529, 5011734, 5011789, 5012103, 5012151, 5012219, 5012709, 5012762, 5012807, 5012898, 5013014, 5013119, 5013171, 5013275, 5013325, 5013404, 5013535, 5013619, 5013875, 5014010, 5014414, 5014621, 5014770, 5014792, 5015053, 5015359, 5015410, 5015436, 5015624, 5015655, 5015714, 5015900, 5016096, 5016158, 5016203, 5016243, 5016459, 5016487, 5016542, 5016773, 5016807, 5016974, 5017014, 5017147, 5017306, 5017492, 5017778, 5017898, 5017966, 5018125, 5018203, 5018280, 5018370, 5018766, 5019020, 5019241, 5019292, 5019429, 5019472, 5019510, 5019531, 5019785, 5019848, 5019989, 5020116, 5020328, 5020745, 5020796, 5020843, 5020936, 5021100, 5021208, 5021463, 5021530, 5021766, 5021890, 5021925, 5022029, 5022081, 5022131, 5022221, 5022342, 5022450, 5022605, 5022707, 5022797, 5022883, 5023065, 5023093, 5023153, 5023194, 5023303, 5023374, 5023410, 5023474, 5023520, 5023581, 5023636, 5023712, 5023783, 5023874, 5024243, 5024393, 5024488, 5024888, 5025058, 5025121, 5025175, 5025249, 5025521, 5025617, 5025747, 5025994, 5026056, 5026109, 5026346, 5026677, 5027231, 5027377, 5027560, 5027594, 5028040, 5028480, 5028822, 5029160, 5029206, 5029283, 5029386, 5029615, 5029815, 5029954, 5030109, 5030198, 5030508, 5030616, 5030759, 5031227, 5031277, 5031371, 5031430, 5031591, 5031626, 5031707, 5031828, 5031946, 5031964, 5032009, 5032105, 5032271, 5032361, 5032479, 5032514, 5032566, 5032649, 5032703, 5032932, 5033017, 5033049, 5033153, 5033197, 5033368, 5033435, 5033502, 5033531, 5033743, 5033855, 5034117, 5034232, 5034360, 5034493,











## 7.4 CCDC MOFs

These refcodes are the 1,000 experimental structures from the CCDC MOF subset<sup>8</sup> sampled for the validation study.

DOPTEI, YIFFUQ, JUHNOR, SUZVEO, VIWNEV, ALOLES, NISYAR, XIHXUJ, VOLKIR, OHAMER, KUFQEI, XOJYUR, RARKUR01, VIYPUP, VOGPUE, SEQGUR, PEVQEO, POLCIE, DEJGU, DIXWAJ, NEHCAG, ETEHAN, TADCAC, TATQOV, WIHPPIO, XUKUV, YUCDUW, BETFUD, EJAZOF, FARKIT, REFMEW, ICADEX, ZUMJOI, OGIPUR, TOWNOK, CECVOW, ASIVAZ, FEQRQA, MIGWUW, TIPRIT01, RAVNAE, QAQKOB, YOYWEOQ1, TEMPIM, VOXGUM, TINBAU, ZOZHIIH, ACASUU, NODRAB, QITZUP, GIRDUI, SIDWUZ, BTFUD, EJAZOF, FARKIT, REFMEW, ICADEX, ZUMJOI, OGIPUR, TOWNOK, CECVOW, ASIVAZ, FEQRQA, MIGWUW, TIPRIT01, RAVNAE, QAQKOB, YOYWEOQ1, TEMPIM, VOXGUM, TINBAU, ZOZHIIH, ACASUU, NODRAB, QITZUP, GIRDUI, SIDWUZ, VAYCJL, VIWLAQ, QIKKOK, FURCIG, FISKAT, VUNCAJ, HODWUS, YIMGOS, POSZED01, POCWAH, OGENUL, DUPFEZ01, RUGYOJ, WIDWOX, YARJEI, AFUKAP, BPYCUC, CIHMIR, JOFLAS, YUXVIX, ACAJOF, GITBOC, EMOSUV, ZOBSSUG, QICTEC, UJAQUA, POJTOY, AXUDUS, VAWYEZ, LULZAV, WITPUM, LETXEP, EKUGOG01, MAVCOD, WUDNUG, BIYLIG, AFUQID, PIGHIX, HATUVU03, PEVKOR, WONXAZ, KISFAU, XEFCUH01, XACHIS, VOKTIA, XADTED, OLAQQI, TIGVUC, QUSDOY, FABTUZ, RUFRUH, NIZQAO, ZOZXUJ, DAYGOB, OMUHEK, GIKBOS, OHESOL, UZEQEWA, PUWVIN, RELTIM, FODHAJ, DIWCIAO, TOYBEQ, SIHQEF, EYUGEL, IPEBIO01, GIJXUT, LUKWAR, KETGOH, LUNDOP, LIWDEB04, QOZXUZ, TODRIP, GINROL, ABIGEZ06, XAPQOW, YIQNIW, IPAYUT, FAVZEJ, QURUV, MIMKIE, ETEKEU, MAFVOG, CIDWIX, VANGOH, PASRIM, TUMFUE, SIYQEX01, EJONIB, BEMSAO02, HIPGOD, DENVIC, UHUTAT, YATWOG, NEGWEB, XAXSOG, IGEWUM04, IDANOR, DOWHAA, CAPXIC, XECBSA, USUZEQ, VETTOE, GUSTYI, ZOBTOB, SEGDEP, MIHKUK, KIVFEC, NULTAQ, ULUWOO, HERPEA, LOCCIP, DARPAO, KAVMOK, FIPWUY, CEZTUX, VAXJIN02, GOYLOX, MUHNAF, ZAYROI, PUFWIX, DIZGUQ, YINPUI, WOCBUL, DOSWOY, UXUPEJ, QIZNAQ, NETHIE, DEPTAU, EMARIT01, SAMVOT, QUTHET, PUGJEI, TAQFUN01, CUXBUT, FIWBW01, NIIFTOM, SEQCEW, AMEQUF, FIYNOR, XUQBIV, ROBZAL, WUZNEK, QITROA, RIPKOS, PLJDW, EZOBIF04, DAMGUV, HIBLUA, GUSNUOE01, BARQAN, NECWOJ, WEQRAM, KALPET, SIYWUS, AFIDUP, WOQCUB02, VEDFER, DOQGUM, TUFTAR, CIDRUX, RIYXIG, IDATOV, SUSWE01, PAVLUU, OCAHAB, XICLAY, AVEMOE, KOWTIB, FUHPUT, QUGKAE, YIQWOM, UFACEK, XUMKUN, DEFHUR, IHOSUV, AVAWEA, JEHWQJ, MAZBUL, XOKHUB, QCJAB, GENFEM, AGUTOM, YIZCER, SAMCOZ, QUMDUY, HOSXES, CODKUB01, RATVAL, WEDXUA, FEJXIC, WOQRUR, CUZKAK, HIVMA, WUXVUH, NUWVIM, CANYIA02, OGUHII, RIDQAX, NOTDIL, OGAGUZ, VINHIL, TEHYUC, ACLJUT, ROBSAD, KIYZID03, HORHEC, OFIXOR, SODCOE, CEJCAX, ESUNEM, HAYKEZ, FAZGOE, ILAGOS, VOQYAD, PODGEW02, AWEDOW, LORTET, QAQBOA, CIFFU, LERNEB, XILHUW01, VATCAV, GIFWAV, DEHMEJ, IFANUZ, XADWUW, ALOKER, ETESAY, XUMNEA, BOBJAE, YANBAR03, GANPOC, MOCKEV, CODJOW, REGSIH, BAKWER, TEKDUI, JUNSIW, NAKLIW, YOVXEO, OCAGOO, KADHUU, MAWXOZ, ROHCUN, LIRPUZ, VIXKOE, AYUZUP, ECOKAJ, LENREC, QQGQMA01, NAVDIZ, XUZLEK, ENEZAZ, QUKKEN, QAGZUV, QIFHIX, BOHPAQ, SESXEU, UDUXAT, FATKIW, FUBFEP, YUMVAF, PAKTEB, VUSCAN, HIZTOA, FIJBUX, PEPWIR, KEPPOL, IPAGIQ, COTROU, HUPVAR, CORSOR01, ABUCIK, SAGZUW, RICCEM01, HOVFE, OKIRAC, ESAVUQ, AYAMAO, ATTITAY, FAPTIB, ENUKUJ, GUBYEI, ECUWEG, PUDCUN, ELIWUS, EYUDAE, TONVAV, EPELER, QAYCEZ, VIKMOS, BAXKOC, XAWZIG, CEZDAN01, DOCHIN, WICVOU, XEBHOC, EGUGIY, GITWIR, HOYMUD02, IHAVI, NAPGUH, NESHEA, EFOZIK, EPELIV, NENMAV01, RODQEG, HOWQUI, PURDEG, AHONUG, UDITIM, ROZSVU, DUVTOE01, DELFUV, GINTUT, LUYKEV, DAXGOA, KITNIM03, PIJHOH03, VATXIY, OHAMOB, TEQKAC, BUPFEY, SUHRA, YOVBUJ, ALIWAT, DIZQAG, HIZPAJ, KAYZOB, XOLSEY, VUFFAE, XIPCEP, HESSEE, FAZPUT, XEXGOY, QODHUO, WUBZAU, FEPLOD, TUTHOG, RAXCEA, UZIDIR, TIPKEK, ENULUV02, VAJNAW, EMOTOQ, KALGP, GYGINP02, QUXLUR, PAZSOZ, SICDIT, ENOBUE, SAMWIO, DURKIL, BEXDUE, TANNIH, QALTIG, WILPUD, EYUMAN, UWOTAC, CEHWUJ, SOLRAO, BOHMAN, ANOHER, XOMBAE, IMIDZB07, VEPKEI, XIPHIN, PUCRAH, LAGPOA, FEVTEF, LUDPUX, KIPHIC, XULNOJ, SUJVEZ, WIRBUW, QUJGAE, FODLEQ03, VAQDUM01, NUXZIQ, PIWZUR, ODIREZ, XOGXAU, XUBCAZ, QOSJEP, NIQWOB, HISWAJ, BUCWOM, LUXQEC, VASKOR, BUHJOD, GEWWOW, GUMDEZ, QQQBDJ04, ARENIV, ETEUJU, JIMCEO, ZUQFAU, VAZZIG, XIMJAG, MONQUC, VUNNIC, FAVFUF, HEQLEV, VEJPIL, BIQVII, ENSCCO10, NOQJAG, SARFOH, SEFKZ, PIHKOI, NEPDAP, SOJFEC, WAZCIL, GUJUD, NIDBEI, IGEWUM01, PARFOF, SUJWAX, CUBXH, MIRZOE, FAYVOR, JOMWAJ01, RUVMUR, DUPVEQ, ELADAW01, SOYFOD, IBIMEM, NAHGOS, LIKCOY, OQIKUV, CIXDUI, IRIHIA, ZILSES, RIZJAM, QARMAY, UCOCOE, GIKJER, ISAYAD, TUWWIT, LIMKIC, RARYUG, MUKTIX, MOCHOC, HOMMIG, BOFFEJ, NIRGAY, QAQTEJ, GERYOX, KEYSIS, VILKUX, TABPIW, CUSUCC04, YOFSAO, ESOYUH, PEWDIG, AKACIY, MUTMOD, OWAXIU, VAWTAO, BEWSAY, KIGMOE, WIMMAH, MICMER, FUHLAX, NAPHCU02, HEDZUN, POQTAQ, WUCDED, ZODPUF, SOGSUE11, FEJHOS, TIPVEU, SIRJAG, WEFBUG, GOBAJ, FIHYAX, AGMTHY, TIVXED, NEGTOK, SOCPOQ, LIVGEE, MIXPIU, IBUZEM, CEWHAO, OPEQIK, LASMUP, UYQZAK, WINKEL, PUKTIZ, ROVWEG, WOHHEH, FOWJIL, PODKUF, LOMVOA, HIQSUW, VUHNIX, AJCUS, LORJEJ, GAWKUK, YUCWID, TIDEZ, GAHBEX, QAVFEA, RAJNOG, XIWMOG, XUMREE, BEZSIK, MUNKEN, UFOCUP, OGUHII, IAPDEI, NAHVEX, TOVWAE, ZEZSAA, GEJRAP, CATPAL03, RAQLE01, UJJUJF07, FECVEO, YUJQOL, BERQIA, OGLJK, BAZBU02, LIZJIN, KACZIZ, OMZIAG, OFEMAN, LONTAJ, EVUPER, IGADAX, HIZVES, ISUYUR, VANCUI01, REXFIL, MAGVOG27, MEVLOO, PEWFAA, ETEQEAA, AMEWUK, RAFQEV, HUWKIT, JECGII, KOHWIO, ABIGEZ, MURTEZ, WUHNOE, OGOWAJ, RUWXXO, KAYMII, PAHVIE, ISITUZ01, PAGVIC, QIXWOL, GUQVIX, LAZJH, YEYYAC, ROMMOK, FIJFUA, MINJOK, MOPTES, TUCMOU, GAXXAG, XAWZOM, QAMFU03, LIQLH07, FARMOC, GEVNIG, QESLIK, YIHZA02, INIZEK, FEVNOI, NEZBOK, FIGTUM, WITJUG, BUGJAO, XEHGEY, NOJSUC, QIHQLJ, TINFAZ, WUYXOD, AQODOA, QEGGUE, XIRHEM, QUQCIP, KAJGEJ, DAHMEF01, AZOHON, XAGYEL, QULJOX, SOKZOJ, VOXTEJ, AJATEL, NILKUQ, HAWWIM, UGUJUC, RAQVIQ, VEPNIN, BUPGAV, COZGED, HUSHAE, AMOYUX, IMATEV, TARQIM, HIWBOF, IWOXOH01, XOTMEY, PIZREX, BUNHAY, YASMOW, YOYKON, VIHZOD, YORJUL, SADLIT, HOQGAX, ZOZNHE, BIYNN, IGESOE, XOVPII, FAXFAL, JEXJUS, GAPQAQ, ODIBAF, XAPDAU, TIYMAQ, XETTOF, VIYZOU, AYIVIO, INAXIE, QAFNAO, TUMQID, DJJCOP01, OTESUC, NUNVOI01, EXAVAB, COLMUM, SOWYIM, JADVUG01, AWODOG, PURNAT, TUJIN, ZEWPAU, FASSOJ, INOBWI, IGUTUB01, HIZHUT, PEVYUM, LOFFAO, IPAXUT, GUCDUF, TEXFAF, XUDSIY, LOXGEM, POLPUC, PESGEA, YOJFUA, BUHNUN03, TIDROP01, KALPCO, GOPCOF, YIHHED, TEZKAL, MOFYIR, HAPHUD, HIZVAO, QIWPET, YIXGAP, KEYSUE, WEHJIE, OBYEYA, BAWKIV, LESMOL01, IRULAI, MIBQAR10, NIHFIU, XEHGUO, VOJRUI, GELQAO, METMIH, DEYMEA, UZUNAF, PUDSUD, QOTROI, VEJPOR, JIMQEC, TAKPIG, EBIBOH, QOGDM, NIYNAK, REPLUQ, QAKHAL, PUPJEQ, QUBQIP, DOWROX, DIYDIA, TUVGAS, YOBXAQ, YABZAD, ZUHRIF, YIBRUY, HOXFOQ, PITBOK, MASPOM, ZUBTUN, LEYFOM, RAXJUW, WEHSUZ, LUQBF, MALREW, VEHKOK, HOPDOH, BURJUU, LOWBAC, CEJQUF, HOBSIB, JAWGIX, XIFBUL, GOBKUD, YOVVIQ, JOKJEZ, ODOFIW, DASTEY02, TIGGUN, NOQZAW, DOWCOJ, MUDJUR, XOJYEB01, EKAWOD, CAHNK, ECUTEB, OJANIW, BINRUN, VICPH, EXEMIE, FUNDIB, IDUKOJ, OLOSAJ, QUPBU, XATNIP01, QIXJUE, EBONOA, BOFKAJ, FARMIV, SOVQAX, MOSPOB, UMAWAG, EGIXAU, MUGGF, MIYNIR, BIPJOB, TUFWUN, CETPIB02, QIYFEK, JOSQEN, LATJEW, UGUKIR, OKISAD, GIYHON, AMBZPB10, SAWTOZ, NUCROT, RBTNCQ12, QAVBOG, RAGCEI, RITDUV, QARSIM03, AWESUR, FEFJEG, ZEXCIQ, FEWTOS, YORNOK, BISCUD, REWRER, KABNIL, WITNIY, ENUQIO, NUCBIY, MOQCUR, COXSEO, LOZVON, RUQWAD, YUKFEQ, LICDOX01, WOBNUX, SISFAD01, POLDEA, SOFROV, DEXZIR, TIBHIW, EKUPEG, GEVXAI, MOYPUM, ODATEO, BERFOV, NEYKOF, VSNFER, HIPQIG, IXISUD, DINDUZ, NIMZIT01, SUMPOH, UHIXEP, KUTJOA, YOHCP, EKERAQ, MABRIS, TATRAI, VEYVIG, BAGQAC02, PADSII, VOBSEM, CLJHIO02, UDICER, KETKOJ, WONDIO, SOCGB, XIHQOV, WULJOE, MEKLUJ, XAKQOR, TAFKAN, KOMGEZ, FOPQX, TOWXEK, LAHTOE, AFOQOC01, OVUJEV, KICLEP, XASFEC01, POKFOL, CUCROC, MOHHUN, KUQXOK, FOYVIEW, KELJIW, XOKHQI, YIGGEB, LINPEE, VACPOV01, NAZHIH, CIPOD, BIDZQA, TORJER, CERFUB, PEHNIB, ICEPIR, CIDLUY, PULWAV, VOSTII, IVUXUS, AWECOV, WOCXET, MUFGIE, POKVUI, FOBZEB, OLIPOO, LOTPEP, SOCWUD, DIRYEK, PUSDOX, HOQDEX, ICAWFU, MULMOX, MIXTAP, WUXVER, REPPUY, RIPNOV, QIRGED, ZIDVEP, GIWPEJ, RESYIZ, SETQUF, YOMZUX, YECQED, LEITUT, TITHUB, PEJRUT, QAHCJE, REWBIF, EJINOB, PIBFAJ, BIGWIY, PURXUX, MOHLOL01, TAKZEL, ULSIR, GUDYUA, WIZKIB, HOYREU, NEVLAD, HOWROC, JIVRUC, VUGWUR, PIAZZIA, VIBPED, YILKAH, BADFOD, SEPYEI, VOKQU, RAVIU, TOVYEK, FIKTUQ, RIRJEJ, OLIRIK, IPAYIH, JIZKIO, QIYRIA, CEJTHU, GEXCOD, COHDUA, DOSSAG, LEPCEP, LASBOY, JOFQOK, FIQYOV, IBIYAU, ROMTET, NITHOP, XANXOZ, PUDTEO, NUGROX, XUFGOV, VOFGII, SAVLEH, ZUJFOB, ZOMPAU, LUFCKU, CEJXOG, XOVFTUY, AWIPUH, AHXCUD11, EMEFAD, CUMMEE, FUMDOG, PEQSAG, WIYZOU, OFANEQ, VIVXOP, TANNAK, KOZSID, ROGHEA, HAVZOV, WANSUA, PRDSBA10, XOLGEM, RUYFAT, WOSZEL, ITEBEP

## References

- D. Denysenko, M. Grzywa, M. Tonigold, B. Streppel, I. Krkljus, M. Hirscher, E. Mugnaioli, U. Kolb, J. Hanss and D. Volkmer, *Chem. Eur. J.*, 2011, **17**, 1837–1848.
- K. Chichak, U. Jacquemard and N. R. Breda, *Eur. J. Solid State Inorg. Chem.*, 2002, **2002**, 357–368.
- F. Darkrim and D. Levesque, *J. Chem. Phys.*, 1998, **109**, 4981–4984.
- D. A. Gómez-Gualdrón, Y. J. Colón, X. Zhang, T. C. Wang, Y.-S. Chen, J. T. Hupp, T. Yildirim, O. K. Farha, J. Zhang and R. Q. Snurr, *Energy Env. Sci.*, 2016, **9**, 3279–3289.
- Y. J. Colón, D. A. Gómez-Gualdrón and R. Q. Snurr, *Cryst. Growth Des.*, 2017, **17**, 5801–5810.

- 6 Y. G. Chung, D. A. Gómez-Gualdrón, P. Li, K. T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J. F. Stoddart, F. You, J. T. Hupp *et al.*, *Sci. Adv.*, 2016, **2**, e1600909.
- 7 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nat Chem*, 2012, **4**, 83–89.
- 8 P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, *Chem. Mater.*, 2017, **29**, 2618–2625.