

**Supplementary Materials:**

Materials and Methods

Fig S1-S35

## Supplementary Information

**Title:** Predicting performance limits of methane gas storage in zeolites with artificial neural network

**Authors:** Sangwon Lee<sup>1</sup>, Baekjun Kim<sup>1</sup>, Jihan Kim<sup>1,\*</sup>

**Affiliations:** <sup>1</sup>Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of South Korea

\*Correspondence to: [jihankim@kaist.ac.kr](mailto:jihankim@kaist.ac.kr)

**This PDF file includes:**

**S1: Materials and Methods**

- S1-1: Zeolite structures
- S1-2: Monte Carlo simulation
- S1-3: Architecture of discriminator and generator
- S1-4: Data pre-processing
- S1-5: Effect of free energy matching

**S2: Tables/Results**

- S2-1: All input zeolites
- S2-2: Rest input zeolites
- S2-3: Interpolation between Energy Shapes

## S1 Materials and Methods

### S1-1 Zeolite structures

IZA (21) and PCOD (22, 23) crystal structures for zeolites were used to compute methane adsorption properties. In our work, a reduced set consisting of just the orthorhombic unit cell structures were used.

### S1-2 Monte Carlo simulation

Classical force-field based molecular simulations using the Monte Carlo (MC) algorithm were performed. The Henry coefficient ( $K_H$ ) and the heat of adsorption ( $Q_{st}$ ) of methane were computed using the Widom insertion Monte Carlo moves. For the unitless  $K_H$  values, we just took the average of the Boltzmann weighted sums from random samples. In addition, the void fraction was computed by counting the percentage of the random insertion moves within the zeolite unit cells occupied that were deemed to be low energy (where low energy was defined to be below  $15k_B T$  (24). Details behind energy-based void fraction algorithm can be found elsewhere (S1). Adsorption isotherms were obtained using grand canonical Monte Carlo (GCMC) method. To accelerate the molecular simulations, a high-throughput code using graphics processing unit (GPU) was utilized (24, S2).

To compute the adsorption properties such as Henry coefficients, heats of adsorption, and adsorption isotherms for a large number of zeolites, both the in-house GPU-based code and the CPU code has been used. In this work, we utilized GeForce GTX TITAN Z and GeForce GTX 780. To accelerate computations further, energy grid with a spacing of 0.15 Å was generated first. The interaction energies between gas molecule and framework atoms were calculated using a 12-6 Lennard-Jones (LJ) potential model (eq. 1) with a cut-off distance of 12 Å.

$$U_{LJ}(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (1)$$

where  $U_{LJ}$  is the potential energy,  $\epsilon$  is the well-depth,  $\sigma$  is the equilibrium distance, and  $r$  is the distance between interacting particles. The LJ force-field parameters for methane and zeolites were taken from Garcia-Perez et al. (S3) To compute interaction energies between dissimilar atoms, the Lorentz-Berthelot mixing rules was used.

### S1-3 Architecture of discriminator and generator

The architectures of discriminator and generator of ESGAN are based on DCGAN (S4). The size of all kernels used in 3D transposed convolution and 3D convolution is  $5 \times 5 \times 5$ . All weights are initialized with the zero-centered normal distribution with the standard deviation of 0.02. The Adam optimizer (S5) is used with a learning rate of 0.0001 and  $\beta_1 = 0.5$  for both generator and discriminator. The dropout rate is 0.5 and multivariate normal distribution  $N(\mathbf{0}, \mathbf{I})$  is chosen as a noise distribution  $P_z$  of the size of 1024 where  $\mathbf{I}$  is the identity matrix. The mini-batch size is 32. ESGAN is implemented by Tensorflow (S6) and we utilized GeForce GTX 1080 Ti to train our ANN model. It takes about a week to train a model on a single GPU. . In this work.,

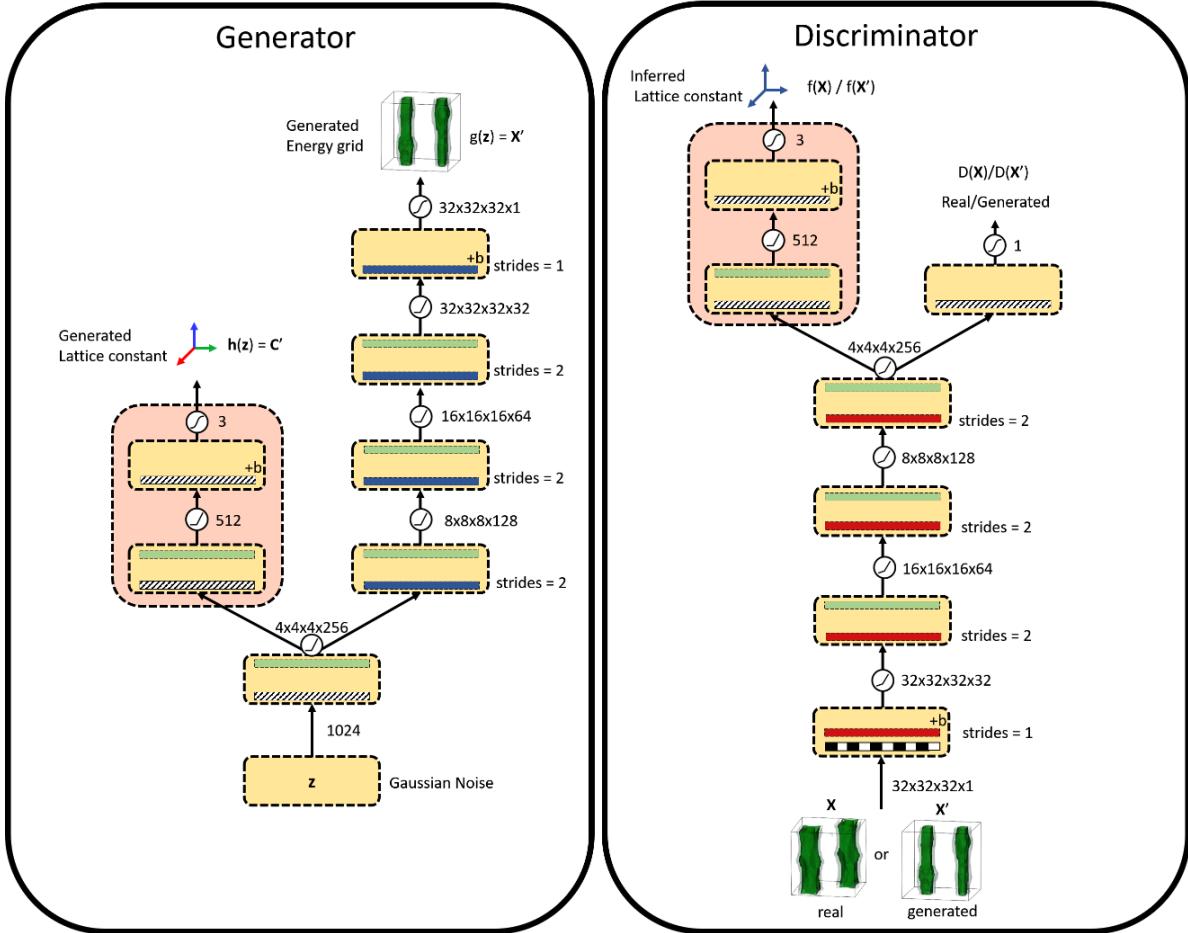
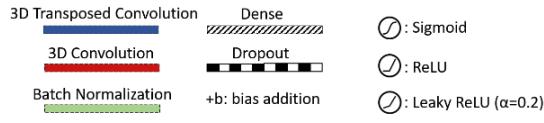


Fig. S1. Illustration of the ESGAN architecture. (a) Generator, (b) Discriminator.

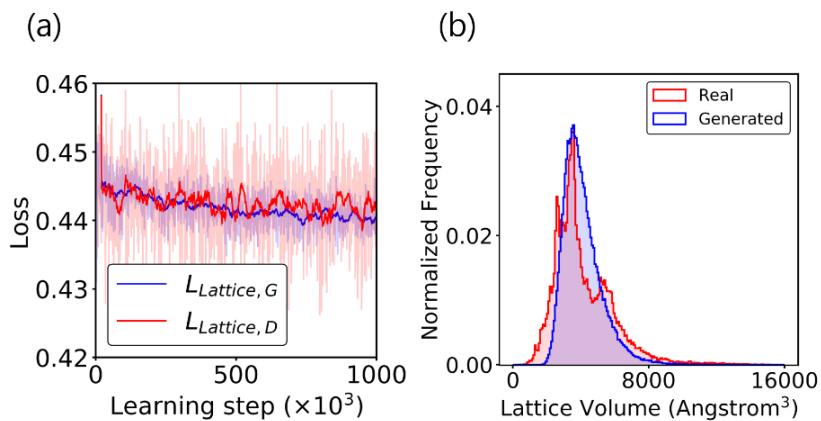


Fig. S2. (a)  $L_{Lattice, G}$  and  $L_{Lattice, D}$  over training steps, (b) distribution of unit cell volume of real and generated energy shape.

Over training steps, the orders of  $L_{Lattice, G}$  and  $L_{Lattice, D}$  are similar to each other. And the average of the absolute length difference between the lattice constant and inferred lattice constant is also similar to each other. This means that the relationship of the generated energy

grid and lattice constant is similar to that of the real energy grid and lattice constant. Also, the unit cell distribution of generated energy shapes is in good agreement with that of real energy shapes. Furthermore, visually, we do not observe any “squeezed” energy shapes, so the energy grid and lattice constant are well matched.

The size of  $\mathbf{C}$  is three because we treat the orthorhombic crystals (no angle variables) and the size of  $\mathbf{X}$  is  $32 \times 32 \times 32$ . For a typical molecular simulation, this is a very small energy size and we used this reduced size for memory purposes in our ANN model. Therefore, we validate the choice of  $32 \times 32 \times 32$  grid size given that this is in fractional dimensions and the actual grid size in Cartesian dimensions vary from one cell to another. We tested the effect of grid size on the IZA set by looking at the computed Henry coefficient and the void fraction. The results of the test simulations are shown in Fig. S3.

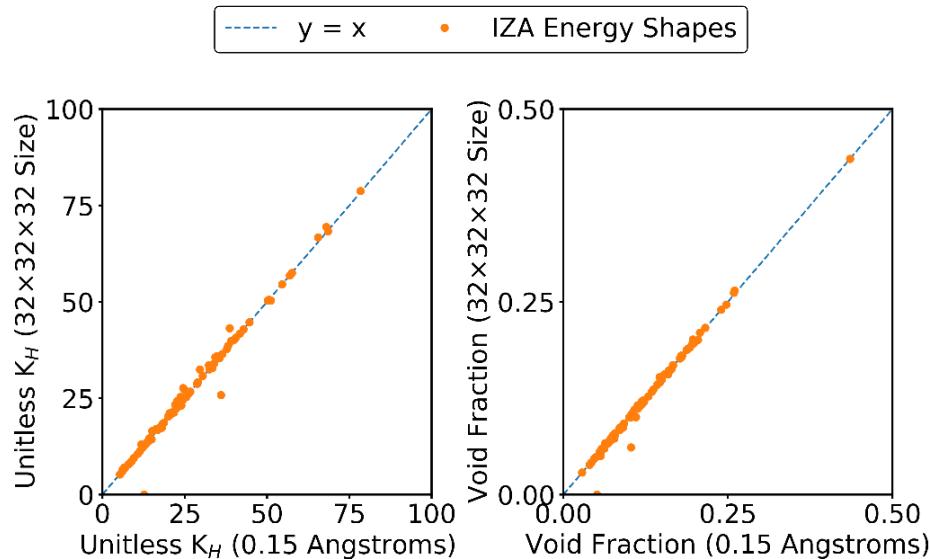


Fig. S3. Adsorption property comparison between different grid sizes. (a) Henry coefficient (b) void fraction. The property values of x-axis are obtained from the energy grid with 0.15 Angstroms on Cartesian coordinate and the values of y-axis are obtained from the energy grid with size 1/32 on fractional coordinate. The results indicate that regardless of the choice in the grid size, the properties are more or less the same and thus justifying our usage of the uniform grid size in the fractional space.

#### S1-4 Data pre-processing

Our ANN dataset consists of tuple data with lattice constant and energy grid information. The values of interaction energy between methane and zeolites are saved in the Kelvin unit. The interaction energy value goes to infinite when the methane molecule overlaps with the zeolite framework. To prevent infinite values in the energy grid, we cut off the energy value higher than 5000 K, which is approximately the energy value of the inaccessible region at the moderate temperature. The energy range is set to  $[-3500, 5000]$  in Kelvin unit because 99.92% of materials are within that range. The cell lengths are also normalized within the range  $[0, 100]$  in Angstrom unit, which is sufficiently large for all the zeolites except one.

To make the value of inaccessible region go to zero, we applied additional operation given as

$$x \leftarrow 1 - x, \quad (8)$$

to each energy value after the min-max normalization. So afterward, the value of 5000 K becomes zero and the value of -3500 K becomes one.

Because the crystals have no bias for translation and rotation in space, we applied two kinds of data augmentations to each energy shape: translational augmentation and rotational augmentation. Image descriptions for these augmentations are shown in Fig. S4

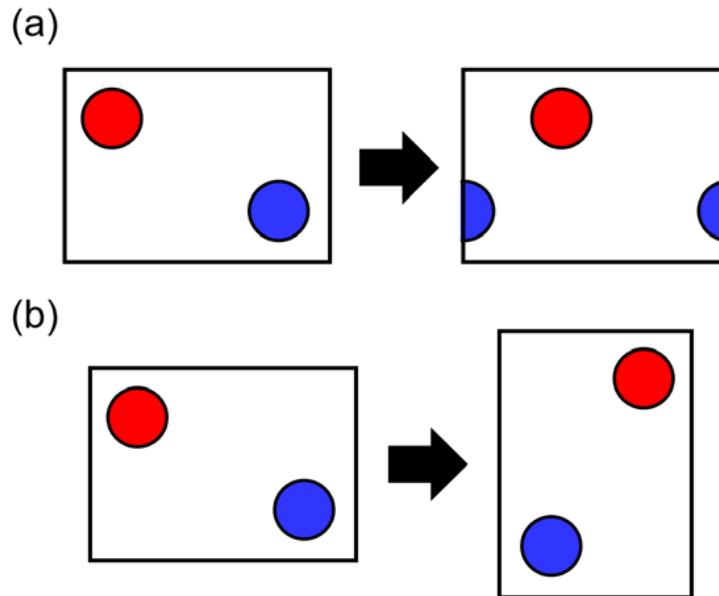


Fig. S4. Image description of two types of energy grid augmentation. (a) translational augmentation, (b) rotational augmentation. Because of the translational and rotational invariance of crystals, all images in the figure are identical structures.

Because the size of the grid is 32 in all directions, the total number of augmentation by translation for each energy grid is  $32^3$ . Also, the total number of augmentation by rotation is three because the only three permutations of the axis can be obtained by rotation.

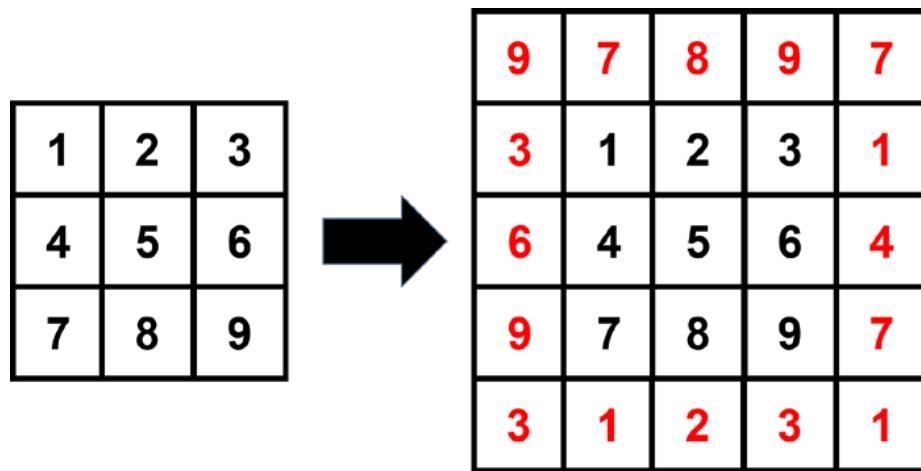


Fig. S5. Image description of the periodic padding. The illustrated example is two-dimensional, but we applied a three-dimensional version of periodic padding to our neural network.

### S1-5 Effect of free energy matching

To show the effect of free energy matching, we used the energy range  $[-3500, 5000]$  with free energy matching up to 600k steps and obtained the pre-trained parameters. From there, we changed the energy range to  $[-5000, 5000]$  and compared the results with and without free energy matching. Fig. S6 shows the results for free energy matching.

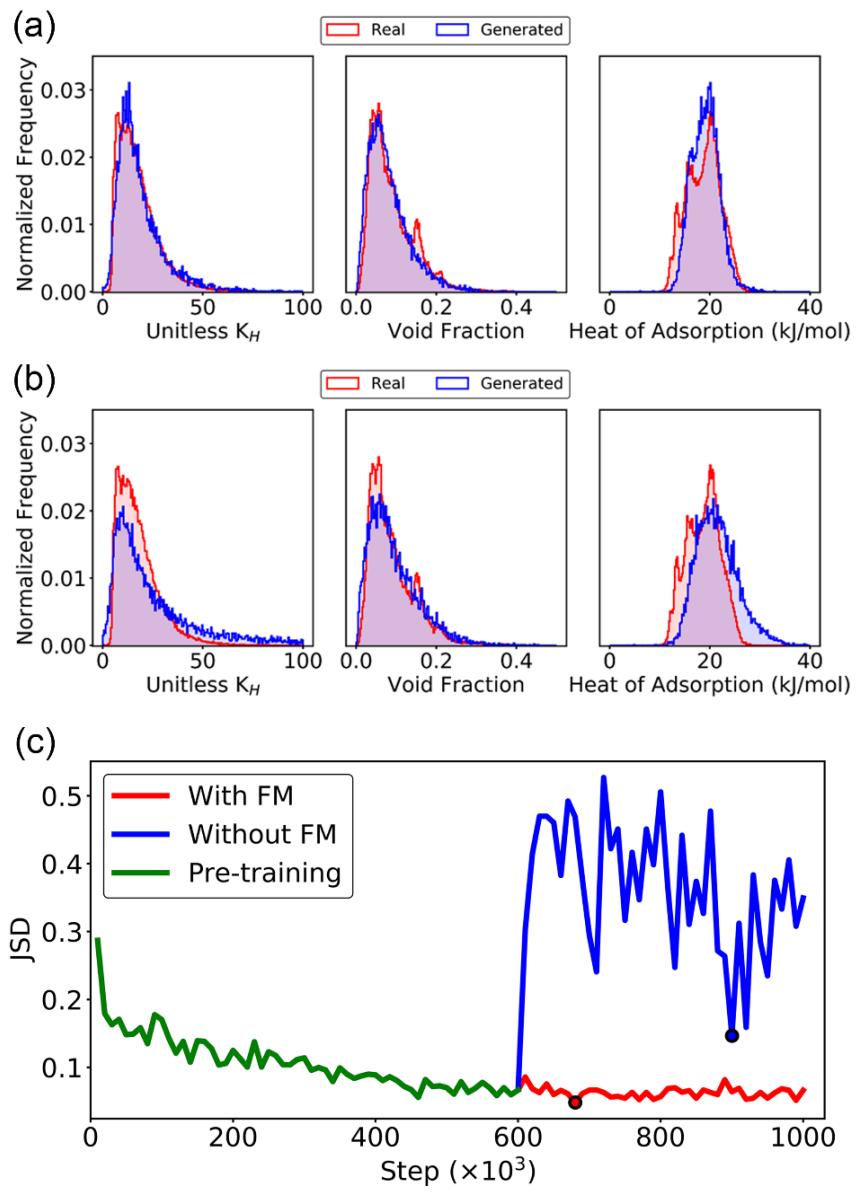


Fig. S6. (a) Property distributions of the model of minimum JSD with free energy matching. (b) Property distributions of the model of minimum JSD without free energy matching. (c) JSD over training step with and without free energy matching. The points on the JSD curve indicate the minimum.

We found that JSD converges slowly and is highly oscillating without free energy matching and the distribution is less likely to match with that of free energy matching and highly sensitive to the range of energy normalization.

## S2 Tables/Results

### S2-1 All input zeolites

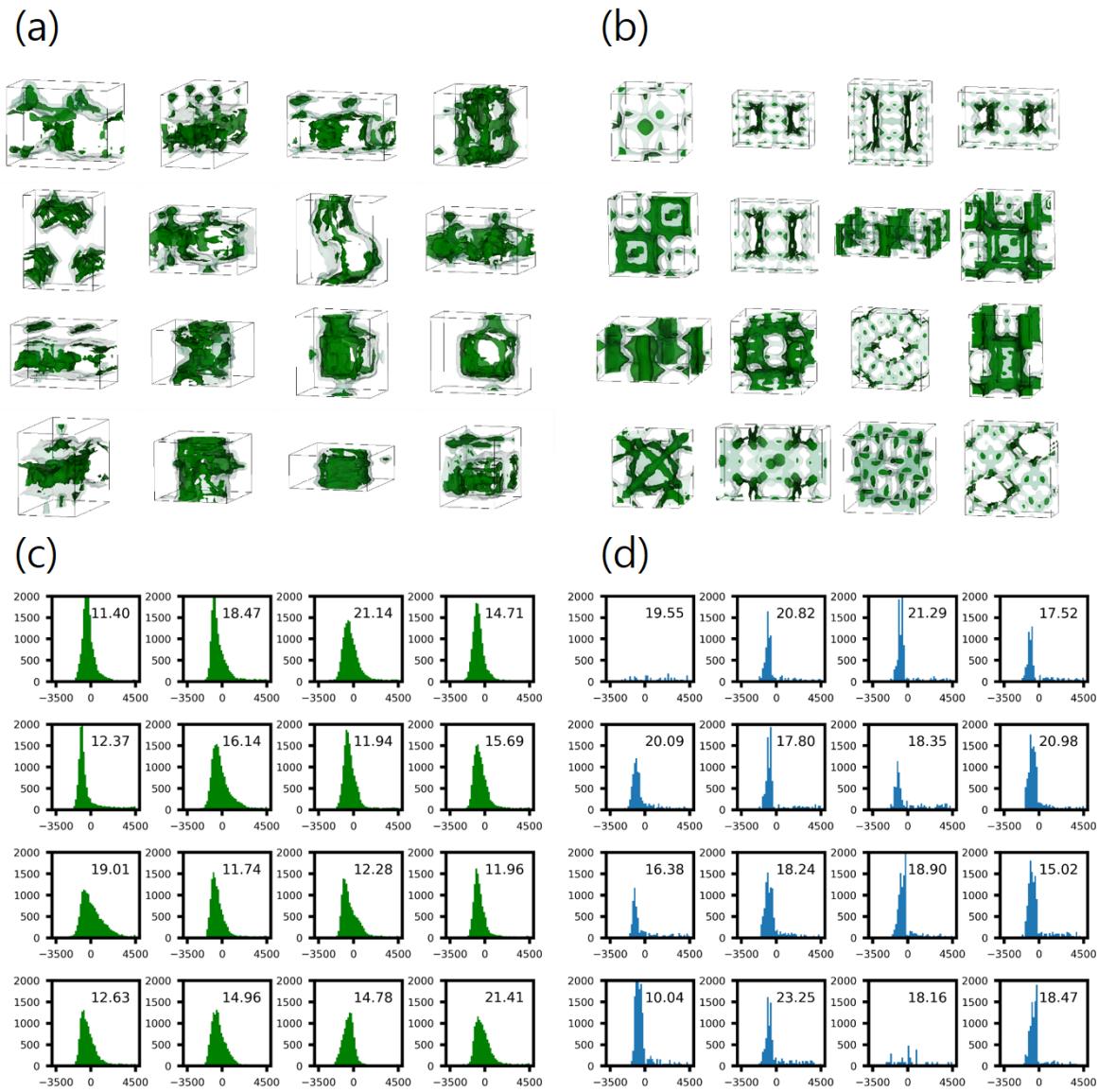


Fig. S7. Top performing energy shapes and the histogram of the value of energy grid. (a) generated energy shapes, (b) zeolite energy shapes, (c) energy histograms of each generated energy grid (the order of histogram is same as (a)), (d) energy histogram of each energy grid of zeolites (the order of histogram is same as (b)). The inset value in each subplot in (c) and (d) is the heat of adsorption in kJ/mol unit.

Table. S1. Top 5 Properties values of real and generated energy shapes (All zeolites)

Rank	Unitless $K_H$		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	400.02	0.55	0.53	32.22	29.29
2	264.63	331.85	0.51	0.53	31.89	28.57
3	258.14	211.15	0.48	0.52	31.47	28.47
4	250.54	202.51	0.47	0.50	31.21	28.37
5	245.07	186.55	0.47	0.48	30.99	28.01
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	156.74	159.41	232.76	224.82	143.90	133.56
2	144.32	155.42	231.68	220.80	115.02	130.82
3	143.49	149.55	207.14	217.17	113.50	121.07
4	142.93	142.49	204.16	215.27	112.10	120.96
5	141.16	141.79	202.39	206.11	107.31	120.79

S2-2 Rest input zeolites

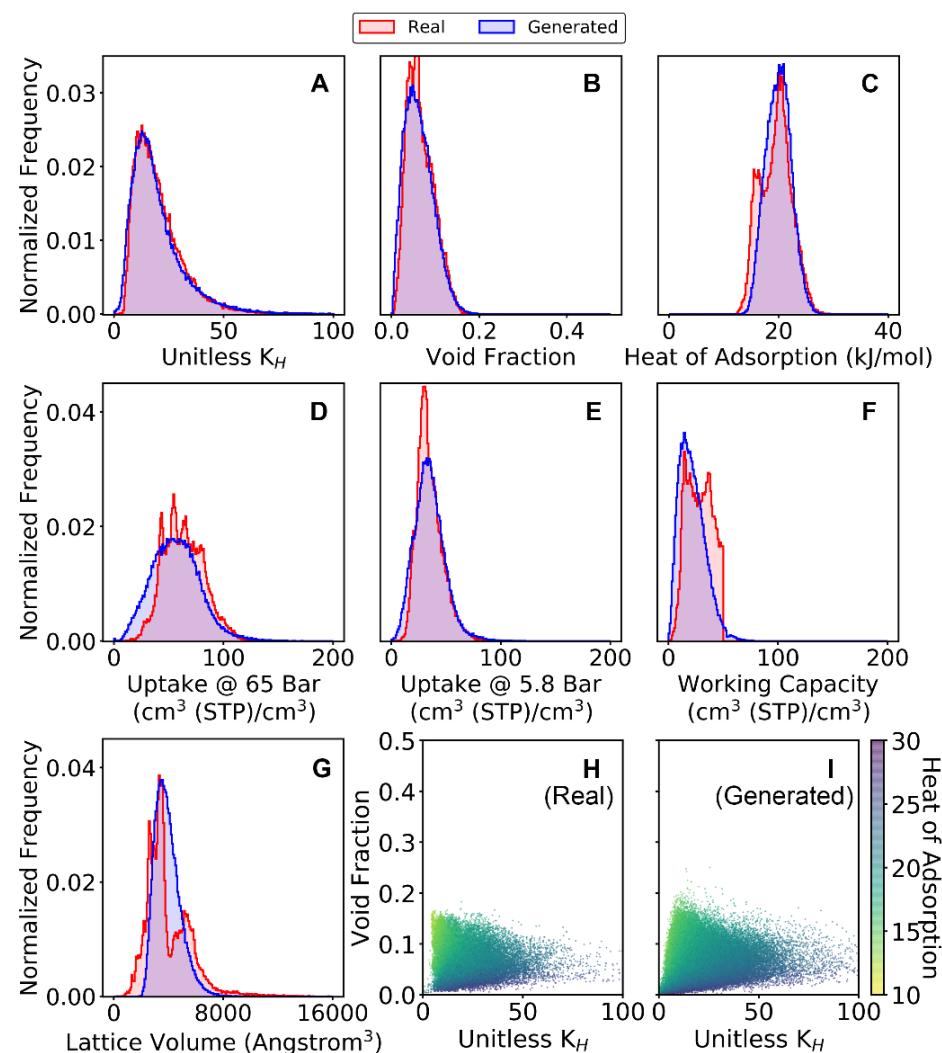


Fig. S8. Property distributions of dataset 1 (WC under  $50 \text{ cm}^3/\text{cm}^3$ )

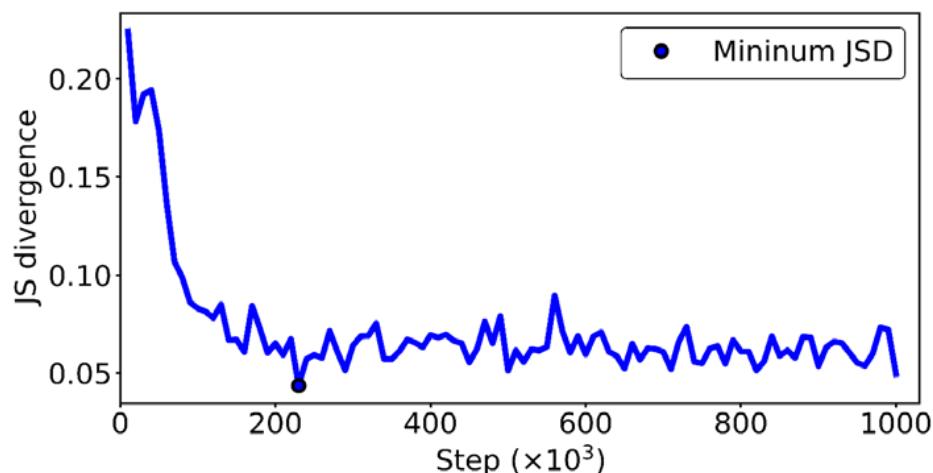


Fig. S9. JSD convergence of dataset 1 (WC under  $50 \text{ cm}^3/\text{cm}^3$ )

Table. S2. Top 5 Properties values of real and generated energy shapes (dataset 1, WC under 50 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	262.12	0.18	0.25	32.22	29.49
2	264.63	189.66	0.17	0.25	31.89	28.63
3	250.54	188.19	0.17	0.24	31.47	28.54
4	245.07	186.12	0.17	0.23	31.21	28.46
5	226.46	186.06	0.17	0.22	30.98	27.97
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	50.00	88.88	161.34	193.70	115.02	157.59
2	50.00	84.94	159.40	192.46	113.50	148.49
3	50.00	83.27	145.07	184.28	104.54	133.28
4	49.99	80.86	143.38	182.35	98.39	129.06
5	49.99	79.85	142.40	181.91	95.68	128.61

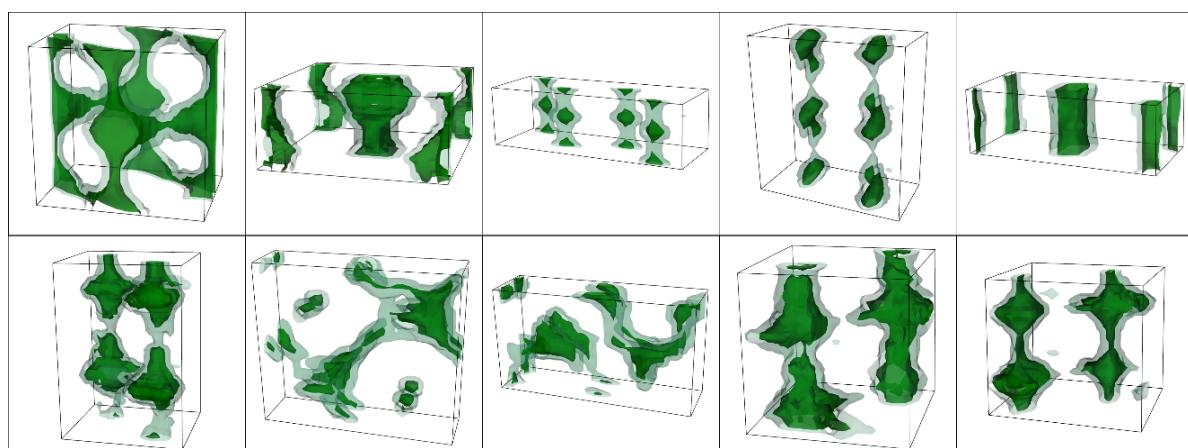


Fig. S10. Visualization of top performing energy shape (dataset 1, WC under 50 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

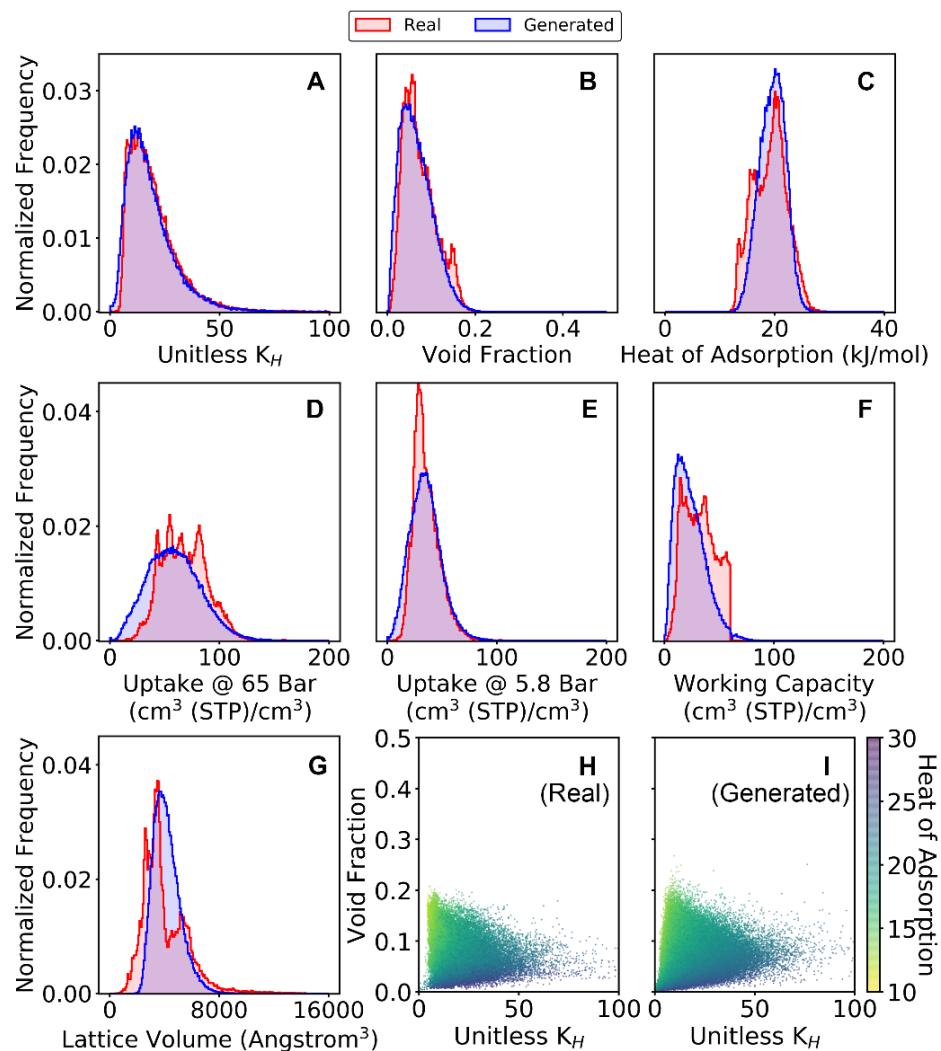


Fig. S11. Property distributions of dataset 2 (WC under  $60 \text{ cm}^3/\text{cm}^3$ )

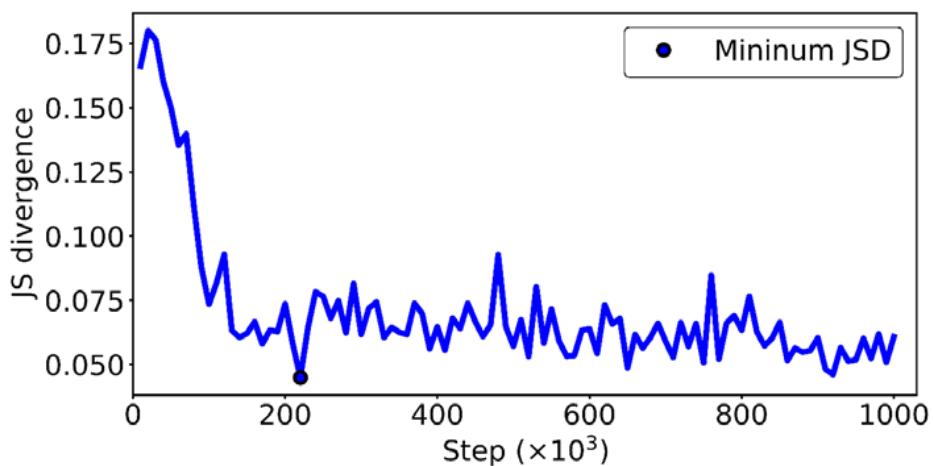


Fig. S12. JSD convergence of dataset 2 (WC under  $60 \text{ cm}^3/\text{cm}^3$ )

Table. S3. Top 5 Properties values of real and generated energy shapes (dataset 2, WC under 60 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	166.38	0.33	0.27	32.22	27.49
2	264.63	149.93	0.29	0.25	31.89	27.15
3	250.54	147.02	0.24	0.25	31.47	27.14
4	245.07	136.89	0.21	0.25	31.21	27.09
5	226.46	133.50	0.21	0.24	30.98	26.99
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	60.00	101.72	166.86	185.88	115.02	126.76
2	60.00	101.30	166.32	180.41	113.50	125.92
3	60.00	99.84	163.47	176.95	107.31	123.21
4	60.00	98.72	163.38	176.52	107.11	110.29
5	60.00	93.71	162.65	171.48	107.10	108.11

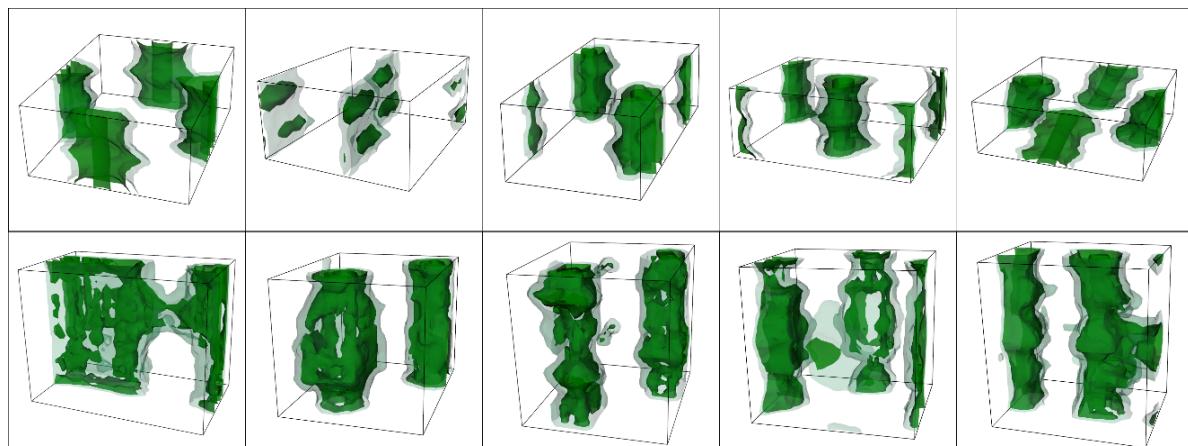


Fig. S13. Visualization of top performing energy shape (dataset 2, WC under 60 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

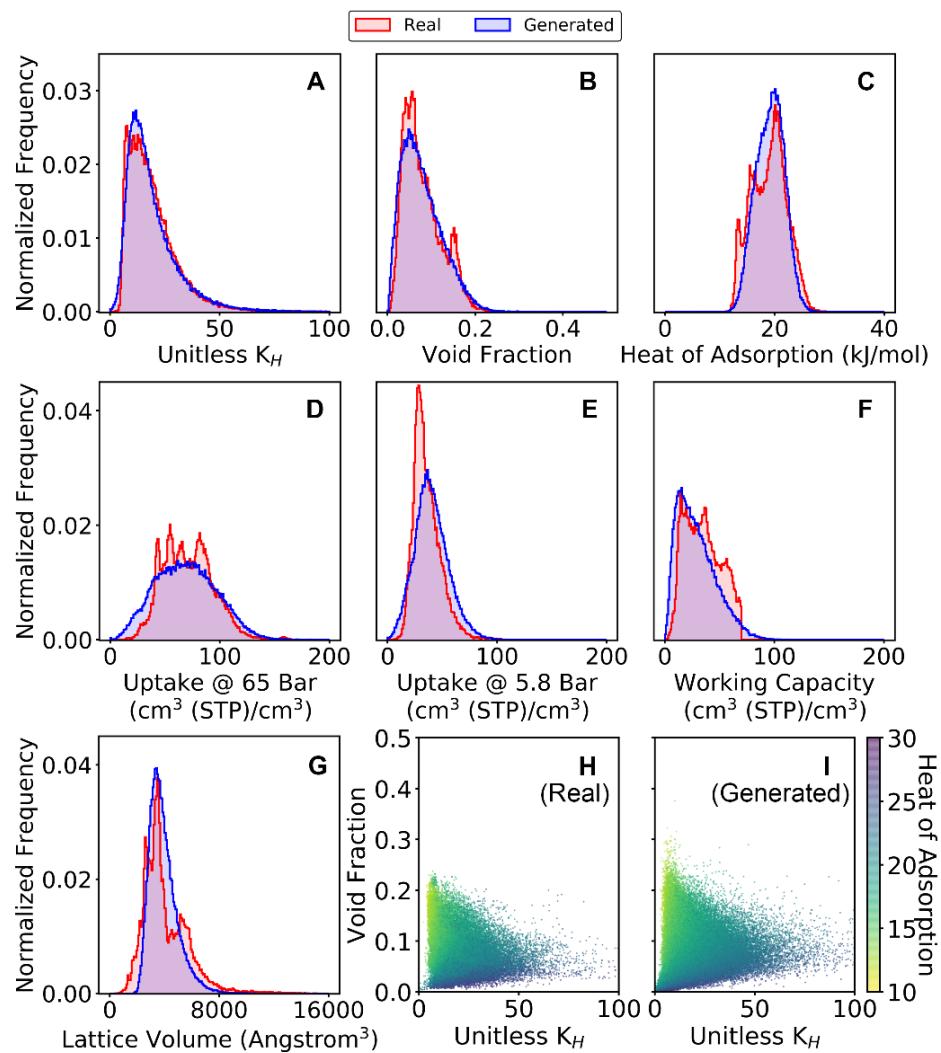


Fig. S14. Property distributions of dataset 3 (WC under 70 cm<sup>3</sup>/cm<sup>3</sup>)

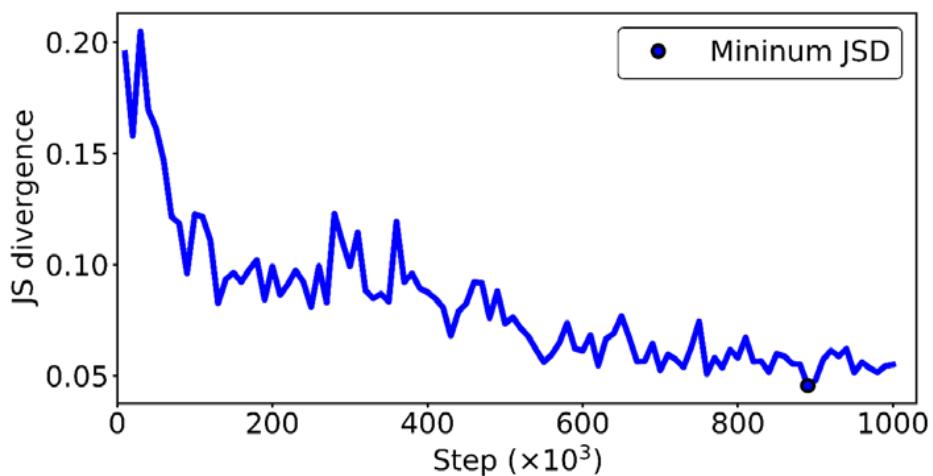


Fig. S15. JSD convergence of dataset 3 (WC under 70 cm<sup>3</sup>/cm<sup>3</sup>)

Table. S4. Top 5 Properties values of real and generated energy shapes (dataset 3, WC under 70 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	266.37	0.33	0.38	32.22	28.44
2	264.63	245.61	0.29	0.32	31.89	27.94
3	258.14	199.70	0.24	0.32	31.47	27.89
4	250.54	191.48	0.24	0.31	31.21	27.74
5	245.07	183.09	0.23	0.30	30.98	27.70
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	70.00	117.53	167.92	208.21	115.02	142.26
2	70.00	117.25	167.01	207.38	113.50	129.65
3	70.00	115.10	166.86	202.66	107.31	123.76
4	69.99	114.95	166.32	201.20	107.11	122.29
5	69.99	112.66	165.78	200.66	107.10	122.25

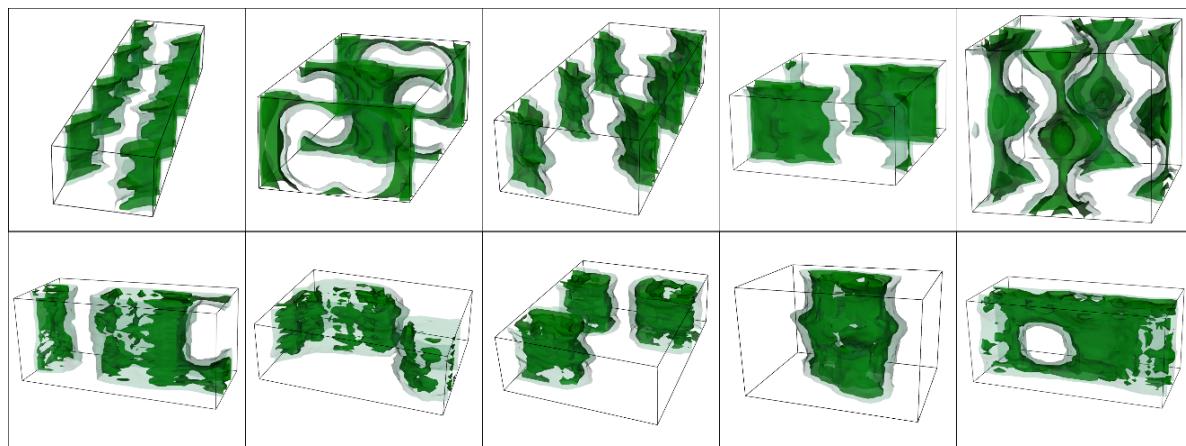


Fig. S16. Visualization of top performing energy shape (dataset 3, WC under 70 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

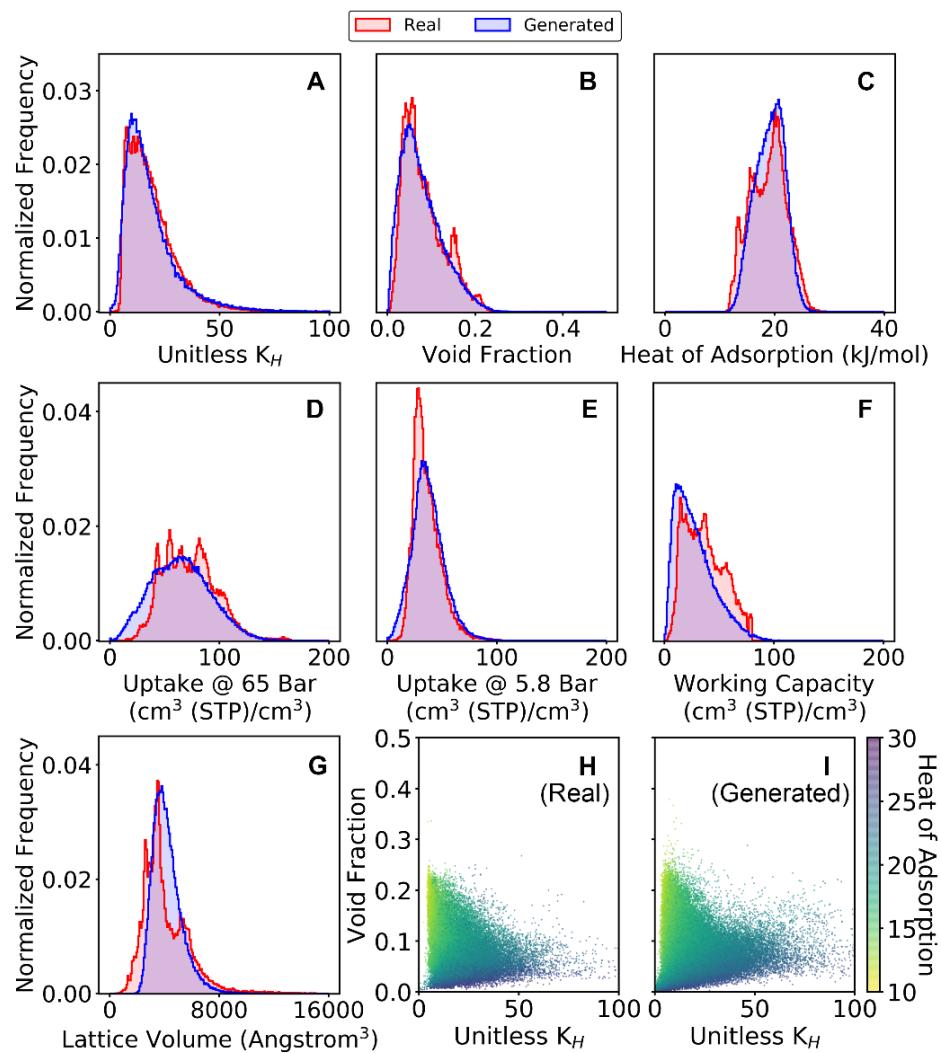


Fig. S17. Property distributions of dataset 4 (WC under  $80 \text{ cm}^3/\text{cm}^3$ )

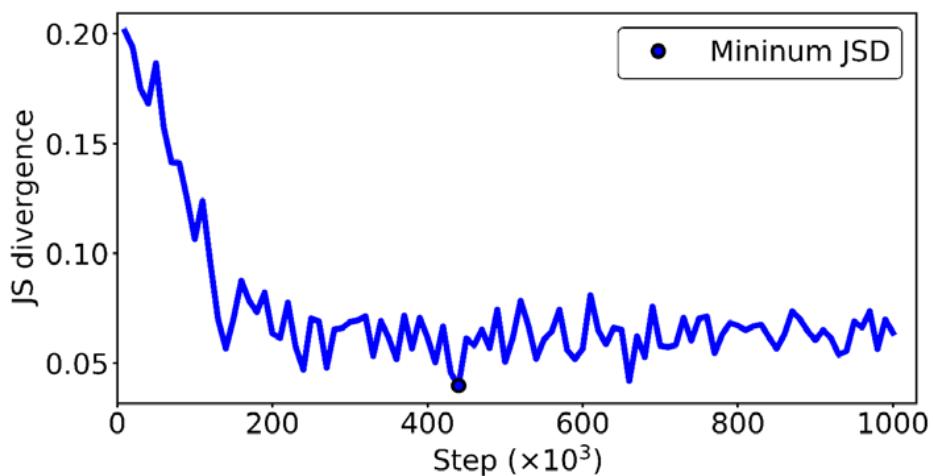


Fig. S18. JSD convergence of dataset 4 (WC under  $80 \text{ cm}^3/\text{cm}^3$ )

Table. S5. Top 5 Properties values of real and generated energy shapes (dataset 4, WC under 80 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	483.94	0.34	0.38	32.22	28.17
2	264.63	266.35	0.34	0.37	31.89	27.98
3	258.14	253.09	0.33	0.34	31.47	27.96
4	250.54	250.45	0.33	0.33	31.21	27.66
5	245.07	238.34	0.29	0.33	30.99	27.64
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	80.00	124.48	167.92	177.95	115.02	128.44
2	79.99	121.40	167.01	177.71	113.50	121.84
3	79.99	111.67	166.93	175.29	107.31	121.79
4	79.99	111.37	166.86	174.90	107.11	118.05
5	79.98	110.77	166.81	173.92	107.10	117.50

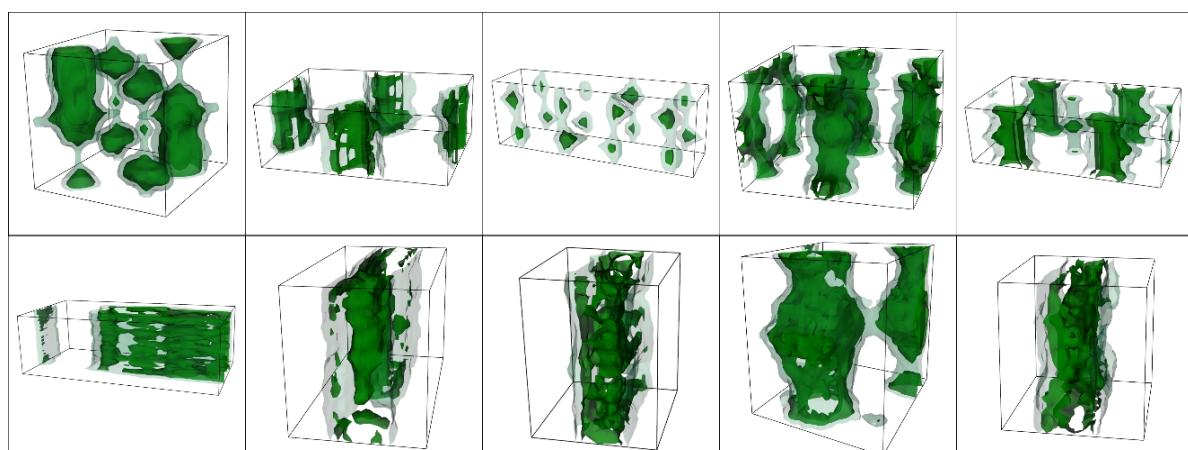


Fig. S19. Visualization of top performing energy shape (dataset 4, WC under 80 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

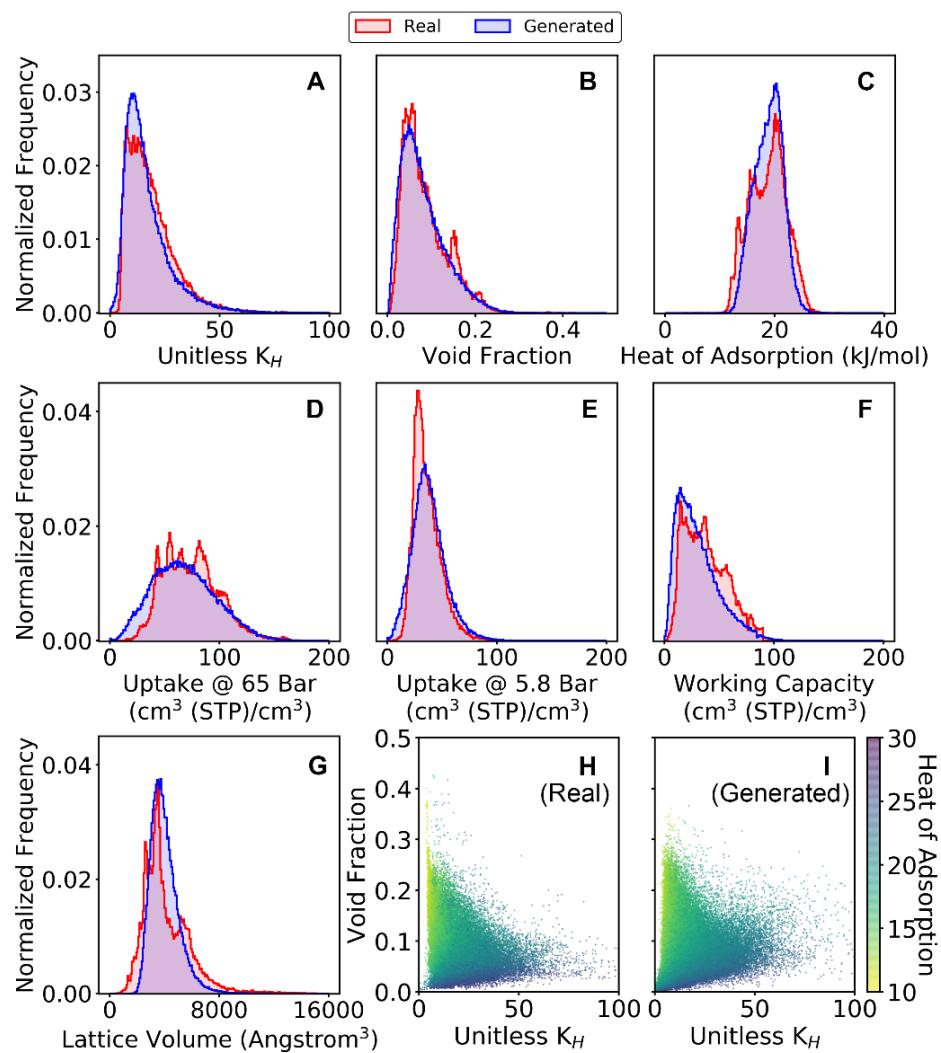


Fig. S20. Property distributions of dataset 5 (WC under  $90 \text{ cm}^3/\text{cm}^3$ )

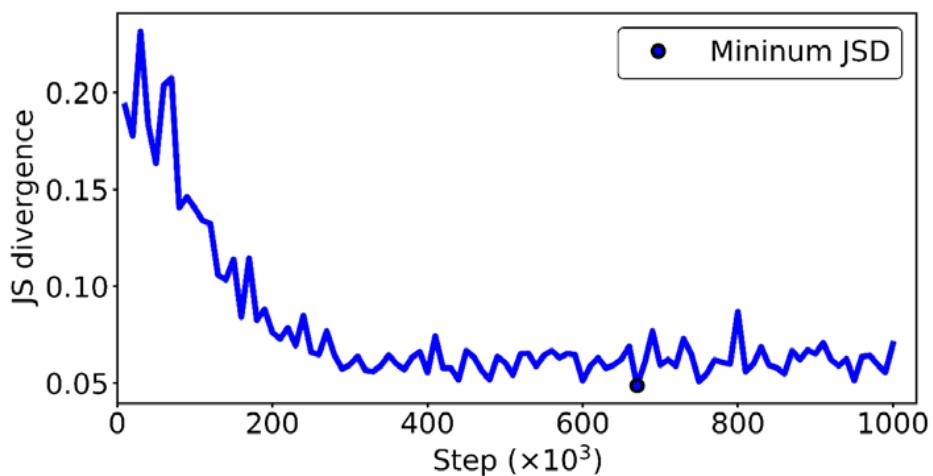


Fig. S21. JSD convergence of dataset 5 (WC under  $90 \text{ cm}^3/\text{cm}^3$ )

Table. S6. Top 5 Properties values of real and generated energy shapes (dataset 5, WC under 90 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	454.23	0.55	0.37	32.22	28.82
2	264.63	211.68	0.47	0.36	31.89	28.65
3	258.14	139.08	0.46	0.35	31.47	28.55
4	250.54	129.67	0.43	0.34	31.21	28.43
5	245.07	126.73	0.42	0.34	30.99	28.17
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	90.00	130.02	231.68	216.01	143.90	130.59
2	90.00	128.08	167.92	201.16	115.02	129.74
3	89.99	126.77	167.01	199.24	113.50	120.59
4	89.98	124.59	166.93	191.91	107.31	119.90
5	89.97	122.51	166.86	190.43	107.11	119.23

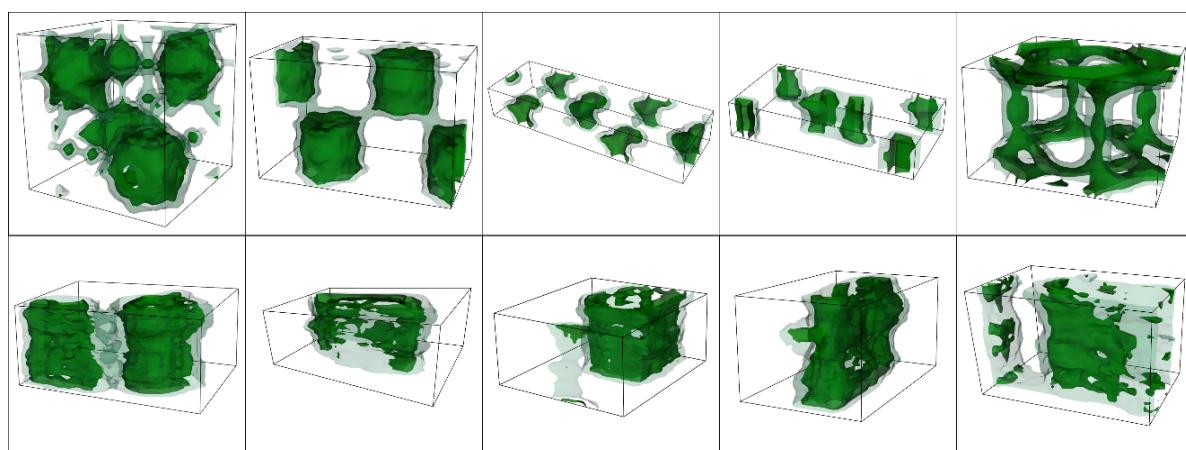


Fig. S22. Visualization of top performing energy shape (dataset 5, WC under 90 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

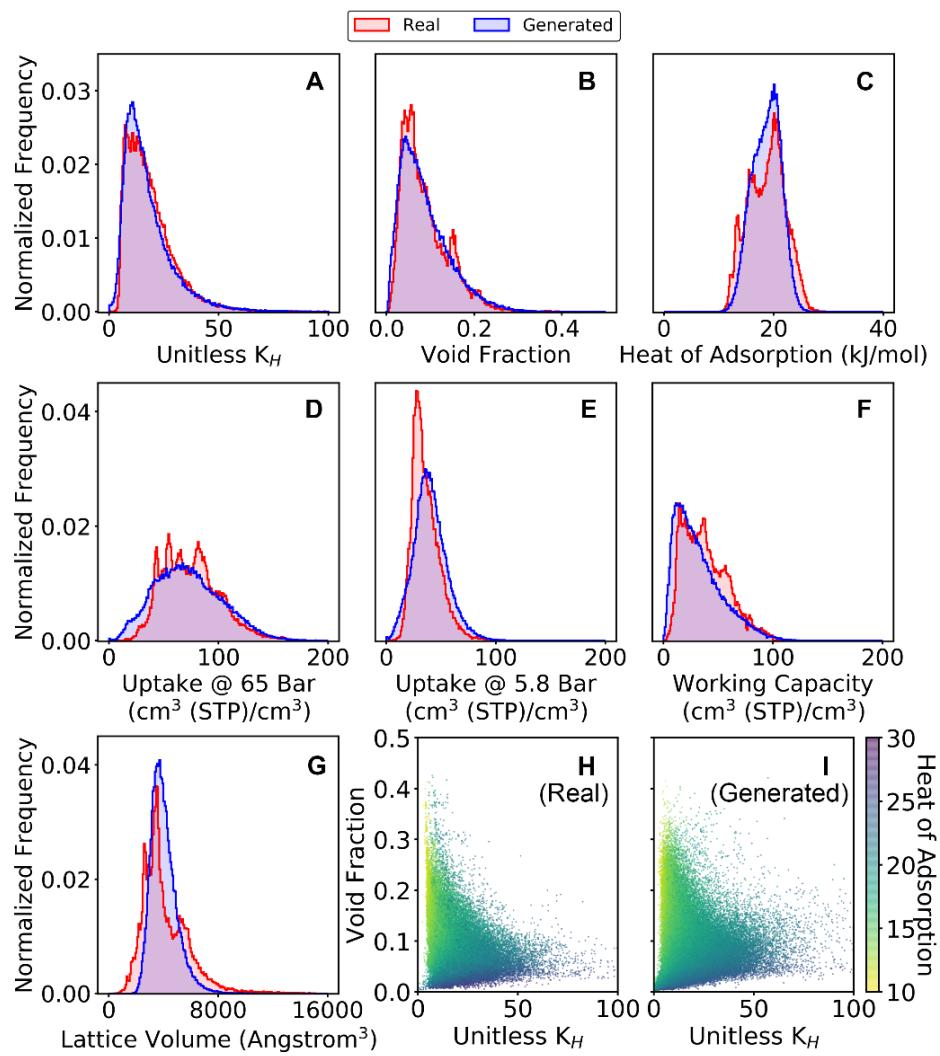


Fig. S23. Property distributions of dataset 6 (WC under  $100 \text{ cm}^3/\text{cm}^3$ )

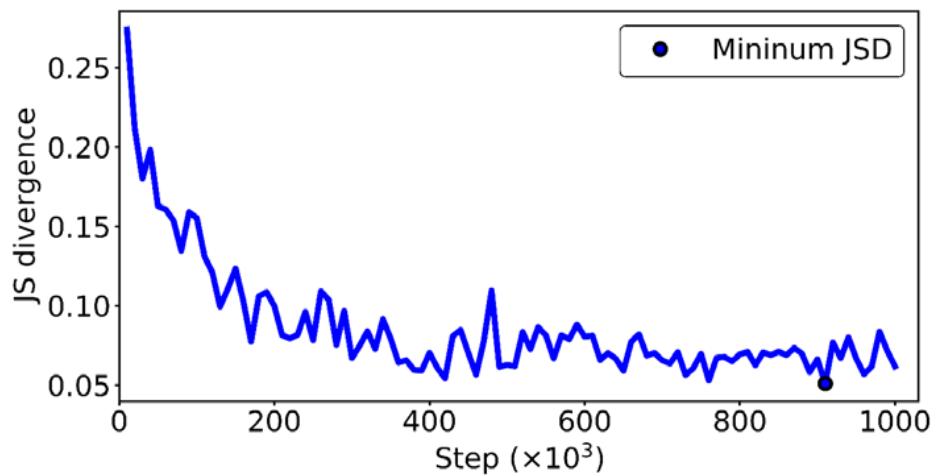


Fig. S24. JSD convergence of dataset 6 (WC under  $100 \text{ cm}^3/\text{cm}^3$ )

Table. S7. Top 5 Properties values of real and generated energy shapes (dataset 6, WC under 100 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	239.14	0.55	0.42	32.22	27.42
2	264.63	180.74	0.47	0.41	31.89	27.41
3	258.14	178.16	0.47	0.40	31.47	26.92
4	250.54	169.82	0.46	0.40	31.21	26.85
5	245.07	160.96	0.46	0.40	30.99	26.66
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	99.99	149.58	231.68	209.56	143.90	123.24
2	99.96	139.96	177.13	208.35	115.02	121.50
3	99.96	138.78	170.84	204.82	113.50	118.79
4	99.95	134.59	167.92	201.47	107.31	116.36
5	99.95	134.23	167.47	199.99	107.11	114.81

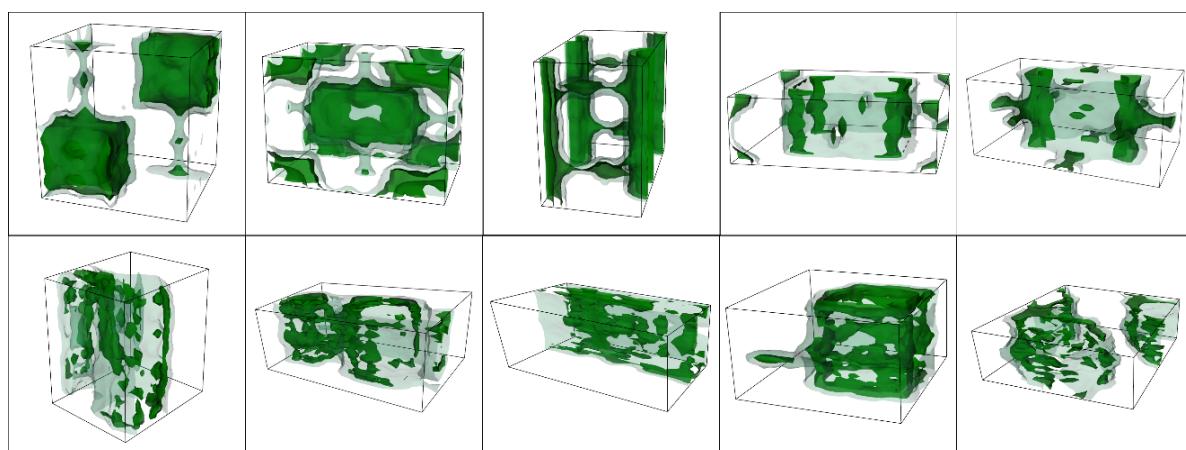


Fig. S25. Visualization of top performing energy shape (dataset 6, WC under 100 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

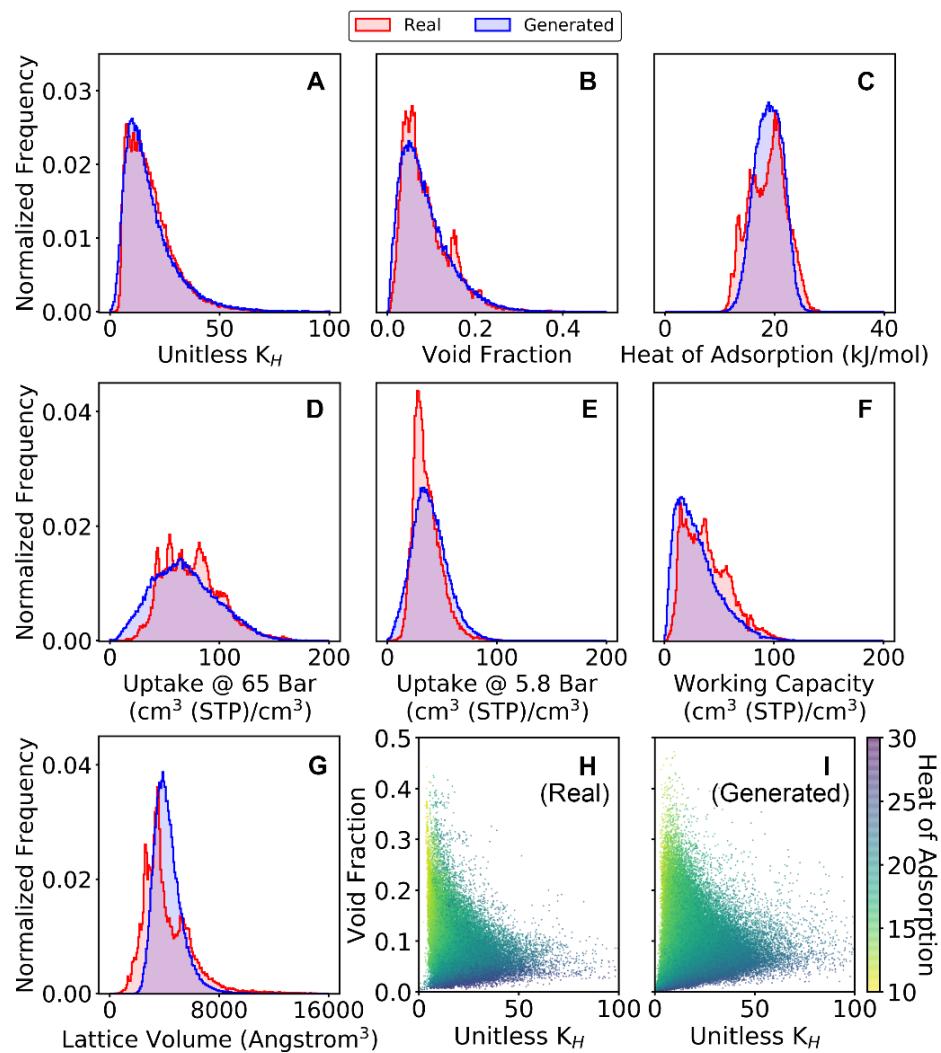


Fig. S26. Property distributions of dataset 7 (WC under  $110 \text{ cm}^3/\text{cm}^3$ )

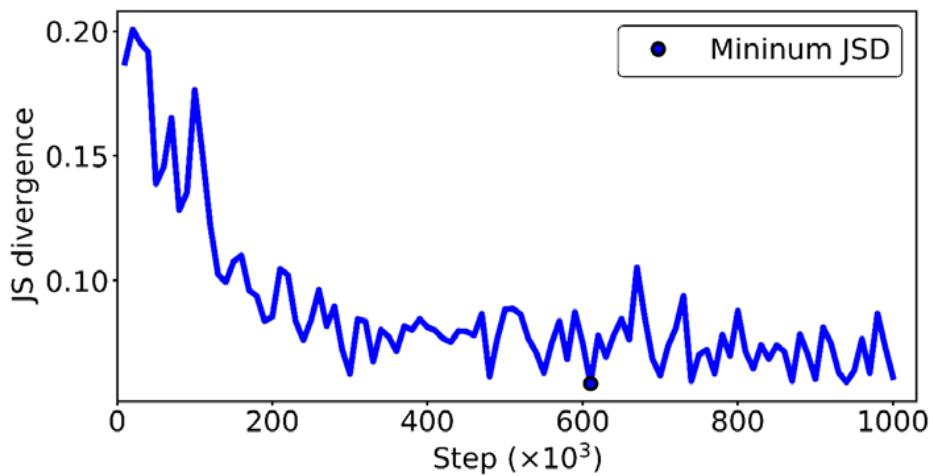


Fig. S27. JSD convergence of dataset 7 (WC under  $110 \text{ cm}^3/\text{cm}^3$ )

Table. S8. Top 5 Properties values of real and generated energy shapes (dataset 7, WC under 110 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	405.44	0.55	0.53	32.22	27.44
2	264.63	244.01	0.51	0.47	31.89	27.17
3	258.14	240.65	0.48	0.47	31.47	27.03
4	250.54	183.82	0.47	0.46	31.21	27.00
5	245.07	169.68	0.47	0.46	30.99	26.98
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	109.99	160.35	231.68	211.16	143.90	133.37
2	109.98	149.98	184.47	209.95	115.02	132.20
3	109.97	144.34	178.98	209.30	113.50	132.00
4	109.86	141.09	178.64	203.97	107.31	128.86
5	109.63	137.59	178.30	202.95	107.11	127.90

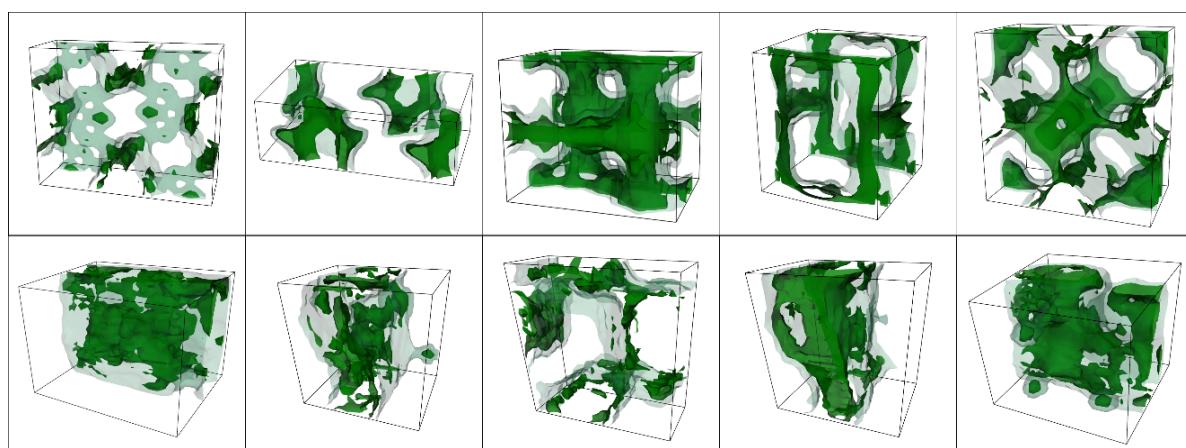


Fig. S28. Visualization of top performing energy shape (dataset 7, WC under 110 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

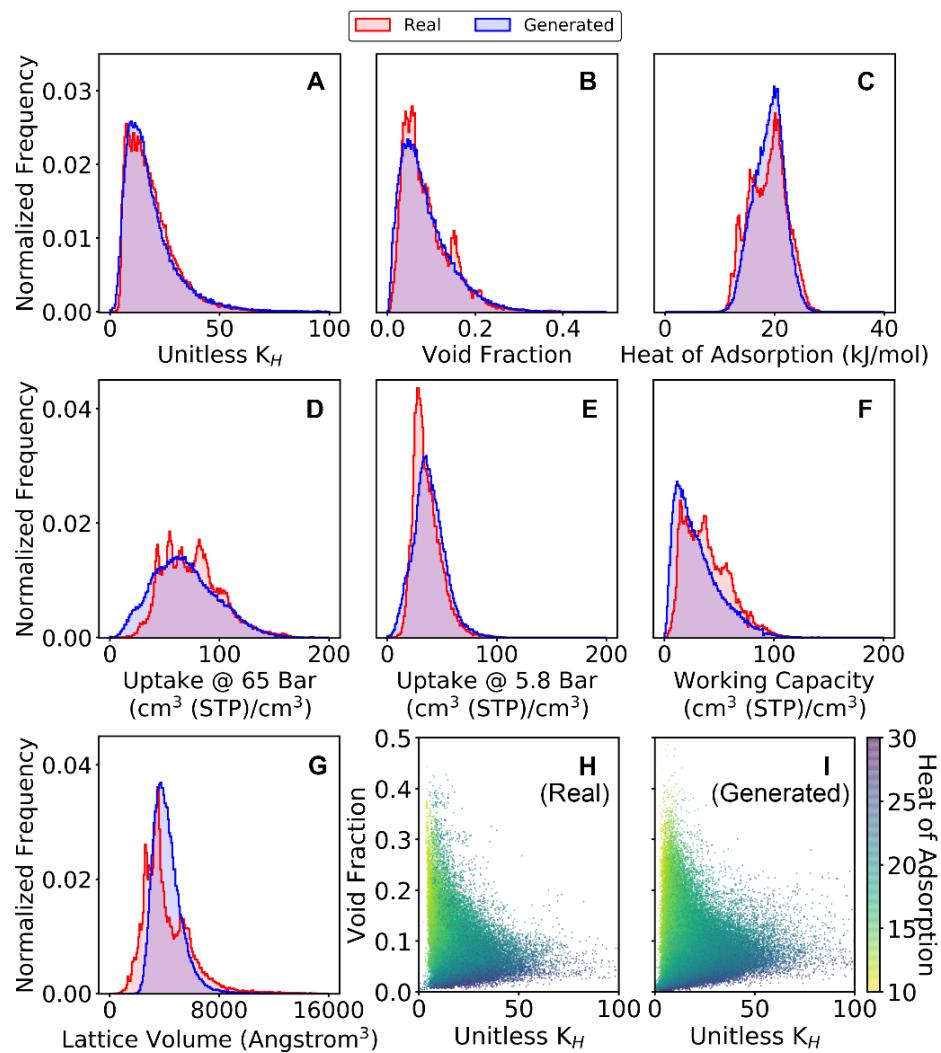


Fig. S29. Property distributions of dataset 8 (WC under  $120 \text{ cm}^3/\text{cm}^3$ )

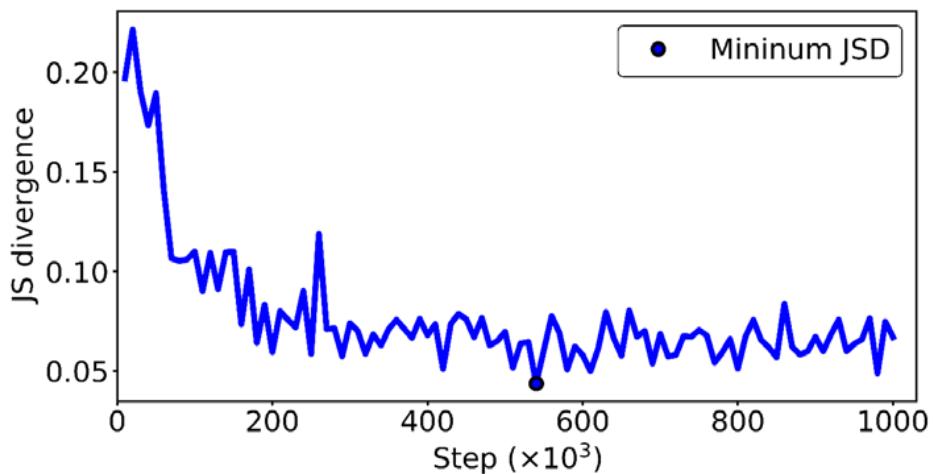


Fig. S30. JSD convergence of dataset 8 (WC under  $120 \text{ cm}^3/\text{cm}^3$ )

Table. S9. Top 5 Properties values of real and generated energy shapes (dataset 8, WC under 120 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	299.28	0.55	0.49	32.22	28.57
2	264.63	264.45	0.51	0.48	31.89	28.41
3	258.14	227.51	0.48	0.48	31.47	28.06
4	250.54	201.54	0.47	0.46	31.21	27.98
5	245.07	191.08	0.47	0.46	30.99	27.94
Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		
Rank	Real	Generated	Real	Generated	Real	Generated
1	119.97	168.70	231.68	224.05	143.90	152.91
2	119.96	163.73	184.47	218.67	115.02	122.04
3	119.74	151.96	182.66	211.88	113.50	120.50
4	119.59	151.87	182.62	208.13	107.31	116.86
5	119.50	149.42	178.98	207.91	107.11	114.91

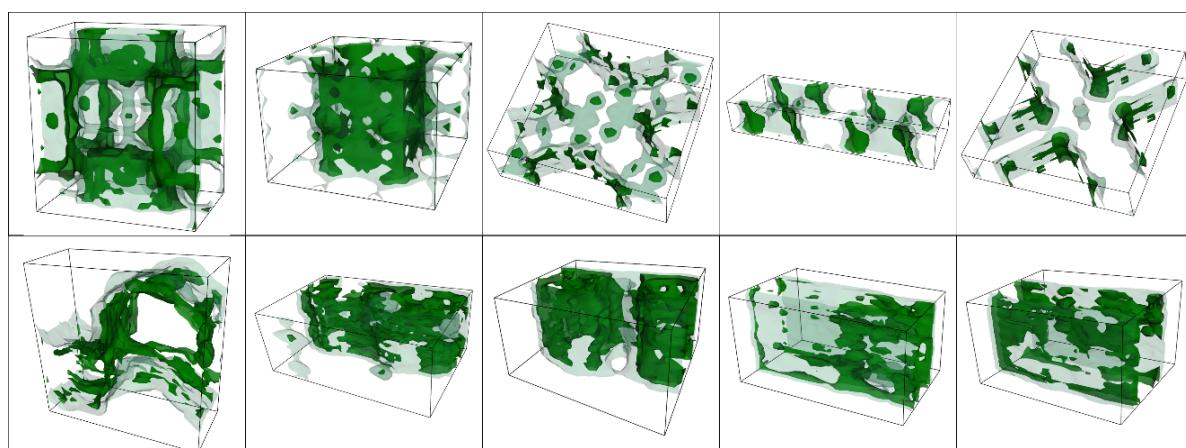


Fig. S31. Visualization of top performing energy shape (dataset 8, WC under 120 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

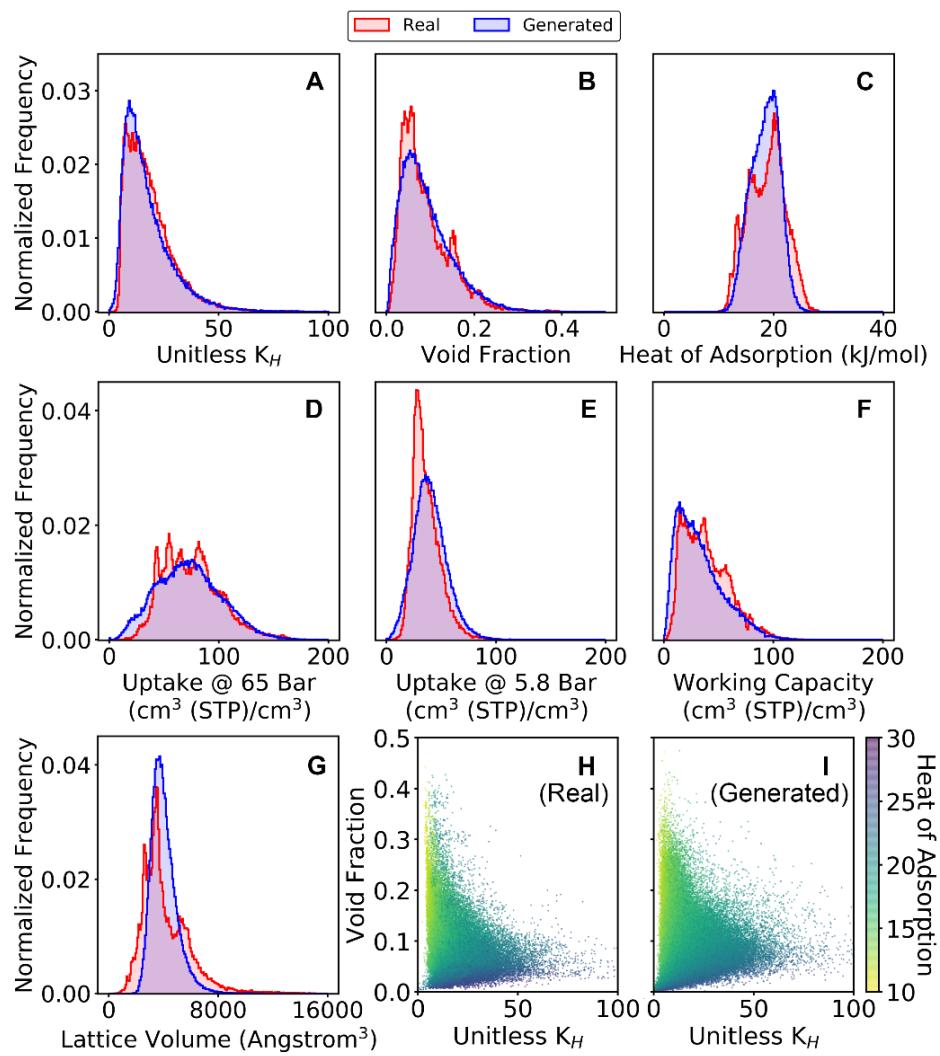


Fig. S32. Property distributions of dataset 9 (WC under  $130 \text{ cm}^3/\text{cm}^3$ )

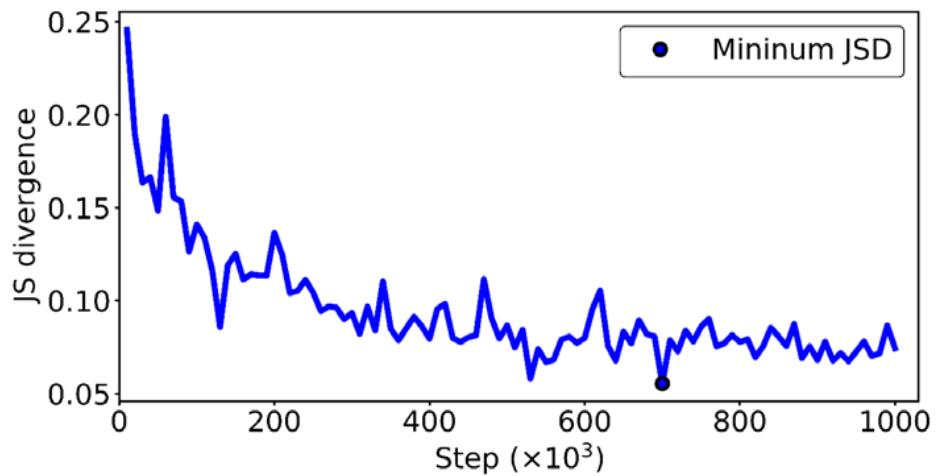


Fig. S33. JSD convergence of dataset 9 (WC under  $130 \text{ cm}^3/\text{cm}^3$ )

Table. S10. Top 5 Properties values of real and generated energy shapes (dataset 9, WC under 130 cm<sup>3</sup>/cm<sup>3</sup>)

Rank	Unitless K <sub>H</sub>		Void Fraction		Heat of Adsorption (kJ/mol)	
	Real	Generated	Real	Generated	Real	Generated
1	398.00	274.67	0.55	0.47	32.22	29.89
2	264.63	197.84	0.51	0.47	31.89	28.54
3	258.14	177.42	0.48	0.45	31.47	27.89
4	250.54	172.44	0.47	0.43	31.21	27.45
5	245.07	171.87	0.47	0.43	30.99	27.16
		Working Capacity (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )		CH <sub>4</sub> Uptake @ 65bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )	CH <sub>4</sub> Uptake @ 5.8bar (cm <sup>3</sup> (STP)/ cm <sup>3</sup> )	
Rank	Real	Generated	Real	Generated	Real	Generated
1	129.99	154.47	232.76	204.00	143.90	131.05
2	129.69	150.68	231.68	203.17	115.02	127.29
3	129.13	147.03	202.03	202.80	113.50	120.87
4	129.01	145.12	198.01	196.97	112.10	119.37
5	128.99	143.21	192.21	196.24	107.31	119.31

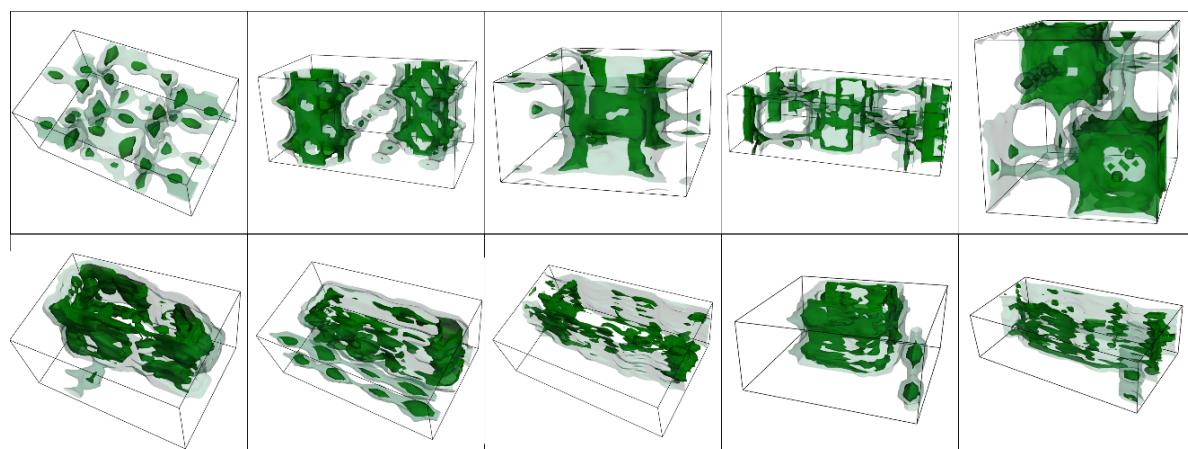


Fig. S34. Visualization of top performing energy shape (dataset 9, WC under 130 cm<sup>3</sup>/cm<sup>3</sup>). First row is real energy shapes and the second row is generated energy shapes.

### S2-3 Interpolation between Energy Shapes

The interpolation between two energy shapes is performed to show that the generator does not simply learn the one-to-one mapping of noise to sample. In general, the interpolation between two samples is performed with linear interpolation in noise space. However, in this work, we use the trigonometric interpolation given by

$$\mathbf{z} = \cos \theta \cdot \mathbf{z}_1 + \sin \theta \cdot \mathbf{z}_2, \quad (20)$$

where  $\mathbf{z}$  is interpolated noise,  $\mathbf{z}_1$  is the noise source of an energy shape 1,  $\mathbf{z}_2$  is the noise source of an energy shape 2 and  $\theta$  is a value within  $[0, \pi/2]$ . We used this interpolation scheme due to the following reason. The generator function  $G$  learns the mapping  $\mathbf{z}$  to  $\mathbf{X}$  from  $P_z$  and  $P_X$  respectively. Therefore, if the value of  $\mathbf{z}$  is not from  $P_z$ , there is no guarantee that the generated sample is proper. In the case of interpolation, the linearly interpolated noise does not follow the noise distribution  $P_z$  because there is no guarantee that the linear interpolation between two samples from the same distribution follows the original distribution. For example, the distribution of the sum of two independent multivariate normal distribution  $N(\mu_1 \mathbf{I}, \sigma_1^2 \mathbf{I})$  and  $N(\mu_2 \mathbf{I}, \sigma_2^2 \mathbf{I})$  is  $N((\mu_1 + \mu_2) \mathbf{I}, (\sigma_1^2 + \sigma_2^2) \mathbf{I})$ . In our case, we use the standard normal distribution as noise. So if we use trigonometric interpolation given as above, the distribution of interpolated noise becomes  $N(\mathbf{0} + \mathbf{0}, (\cos^2 \theta + \sin^2 \theta) \mathbf{I})$ , which is the same as standard normal distribution. Also, for that reason, we did not use optimization over  $z$  space for any purpose (e.g., finding the maximum of some property). Because the  $\mathbf{z}$  obtained from the optimization can be far from the distribution  $P_z$ . This means that the obtained energy shapes from the noise  $\mathbf{z}$  can be highly abnormally point because the  $\mathbf{z}$  does not follow  $P_z$ . The interpolation result of randomly selected three energy shapes are shown in Fig. S35.

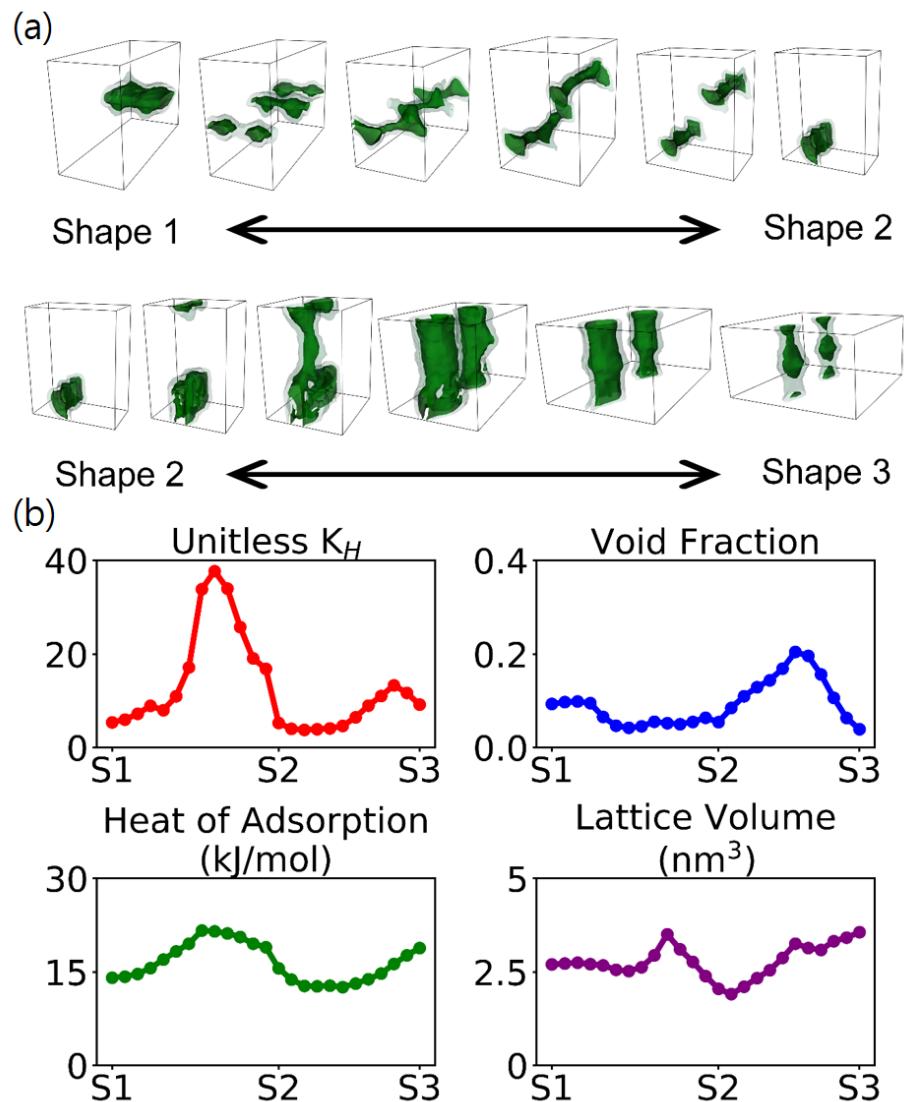


Fig. S35. Interpolation result of three structures. (a) Visual representation of the interpolation, (b) The change of properties between structures.

It was shown that the energy shape and the properties are changed continuously. It should be pointed out that the value of properties vary continuously, but the values are not monotonically changed nor lie in between the values of two energy shapes.

## References

- S1. D. Paik, M. Haranczyk, J. Kim, *J. Mol. Graph. Model.* 2016, **66**, 91-98.
- S2. J. Kim, B. Smit, *J. Chem. Theory Comput.* 2012, **8**, 2336-2343.
- S3. E. Garcia-Perez *et al.*, *Adsorption* 2007, **13**, 469-476.
- S4. A. Radford, L. Metz, S. Chintala, 2015, <https://arxiv.org/abs/1511.06434>.
- S5. D.P. Kingma, J. Ba, 2014, <https://arxiv.org/abs/1412.6980>.
- S6. M. Abadi *et al.*, 2016, <https://arxiv.org/abs/1603.04467>.