Supplementary Information

Thermodynamic analysis of Xe/Kr selectivity in over 137,000 hypothetical metal-organic frameworks

Benjamin J. Sikora,[a] Christopher E. Wilmer,[a] Michael L. Greenfield[b] and Randall Q. Snurr[a]*

[a]Department of Chemical and Biological Engineering
Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

[b]Department of Chemical Engineering,
University of Rhode Island, Kingston, RI 02881, USA

* To whom correspondence should be addressed. E-mail: snurr@northwestern.edu

Table of contents

S1. List of Modular Building Blocks

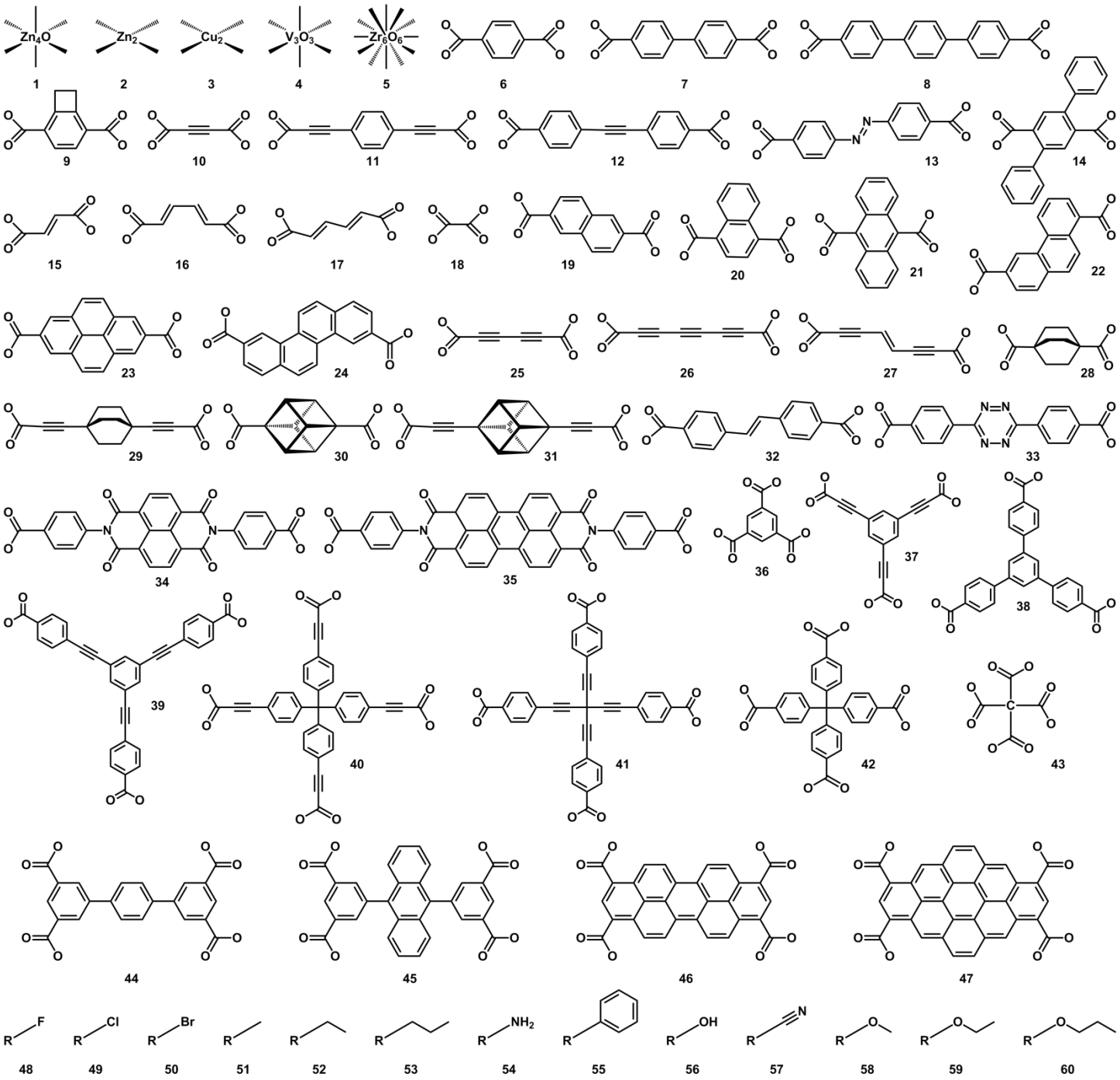

Figure S1. Full list of building blocks used to generate database of ~137,000 hypothetical MOFs. **1-5** are inorganic building blocks, **6-47** are organic building blocks and **48-60** are the functional groups. Building blocks **6-47** (except **18**) may be terminated with nitrogen atoms instead of carboxylic acid groups.

The building blocks we used are shown in concise form in Figure S1. The inorganic building blocks **2** and **3** (referred to sometimes as paddlewheels) are able to coordinate to nitrogen containing compounds (e.g., pyrazine). Included in this study, but not shown in Figure S1, are the N analogous building blocks terminated by nitrogen atoms instead of carboxylic acid groups.

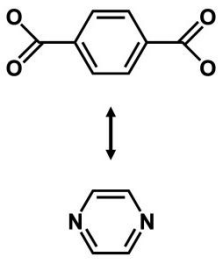

Figure S2. Building blocks in Figure S1 are shown with terminal carboxylate groups; however, every such building block also exists with a nitrogen terminated group, as well, for coordinating to paddlewheels. There is an exception for building block **18**, which can only be terminated in carboxylate groups.

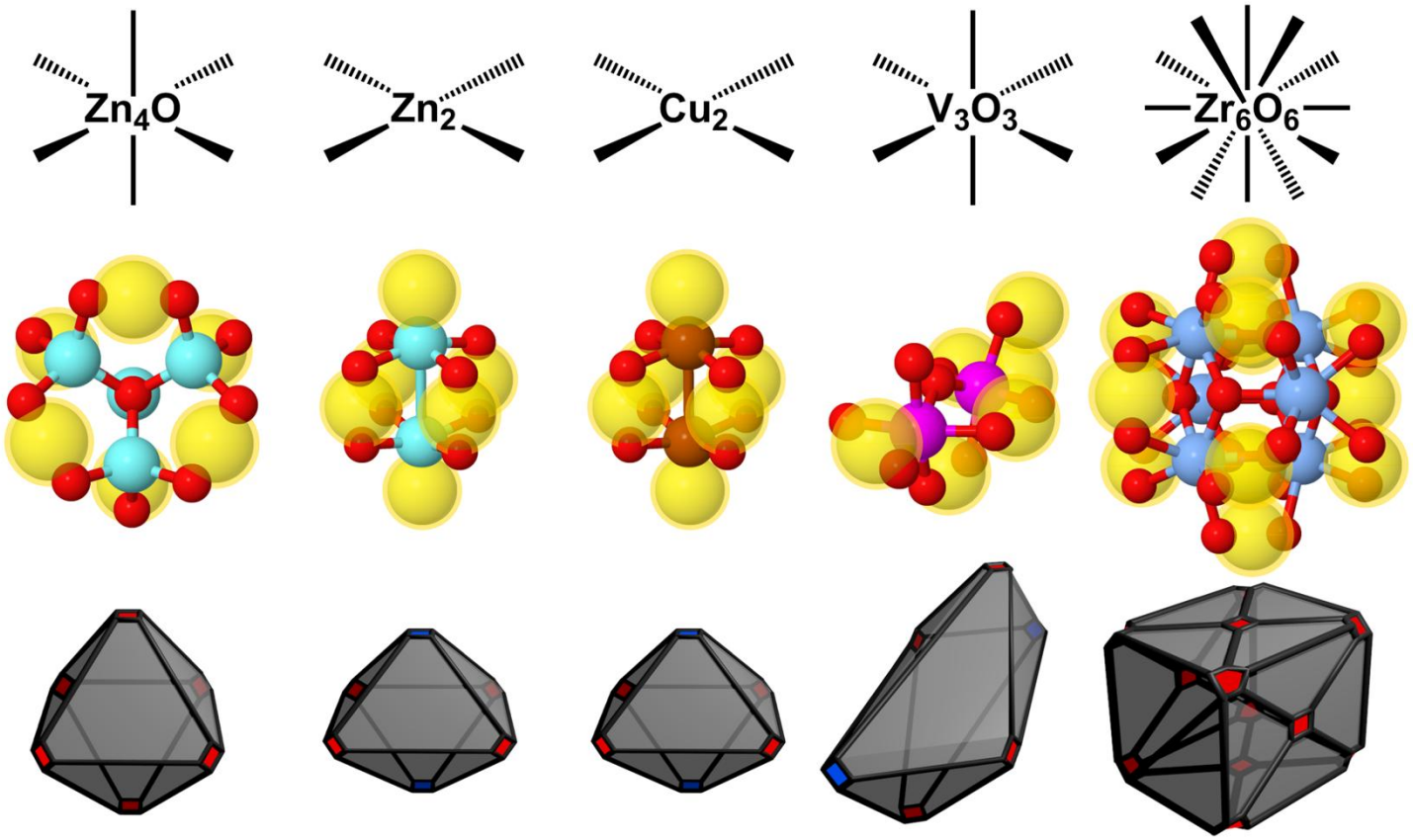

Figure S3. 3D structures of the metal-containing building blocks. Top row shows 2D schematic. Middle row shows 3D atomistic structure with connection points indicated by translucent yellow spheres. Oxygen atoms from carboxylic acid groups on connecting ligands are shown already connected to the metal building blocks. The bottom row shows the representative geometry of the metal cluster, with colored patches indicating connection sites (red and blue patches are used to distinguish between connection sites with different chemistries required for bonding).

S2. Crystal Generator Algorithm

Building block details. We have described our algorithm recently elsewhere.[1] In order to recombine building blocks into crystals, additional topological and geometrical information is manually assigned to each building block (see Figure S4). The topological information takes the form of numbered connection sites so that the generation algorithm can interpret instructions such as "connect building block **2,** site **3,** to building block **10,** site **1**". Additionally, this information is used as part of the algorithm's termination criteria; only when every connection site has been connected is a single MOF generation complete. The geometrical information takes the form of three "pseudo-atoms" and a list of angles for every connection site in the building block. The pseudo-atoms each possess a coordinate in 3d space, as well as a label (here referred to arbitrarily as *R*, *G*, or *B*). The purpose of the pseudo-atoms is to unambiguously specify the relative orientation of two connected building blocks. Specifically, given two connection sites *X* and *Y* and their corresponding pseudo-atoms R_X, G_X, B_X, and R_Y, G_Y, B_Y, the building blocks are oriented correctly when the coordinates of R_X equal R_Y, the coordinates of G_X equal G_Y, and the vector R_XB_X is anti-parallel to the vector R_YB_Y. Finally, if there are multiple "correct" orientations (for example, phenyl rings in a linear chain experience multiple energy minima of equal depth as a function of their relative orientations along the chain axis), the list of angles specifies alternate orientations, equivalent to rotating the pseudo-atoms about the *RB* axis by the specified angle.

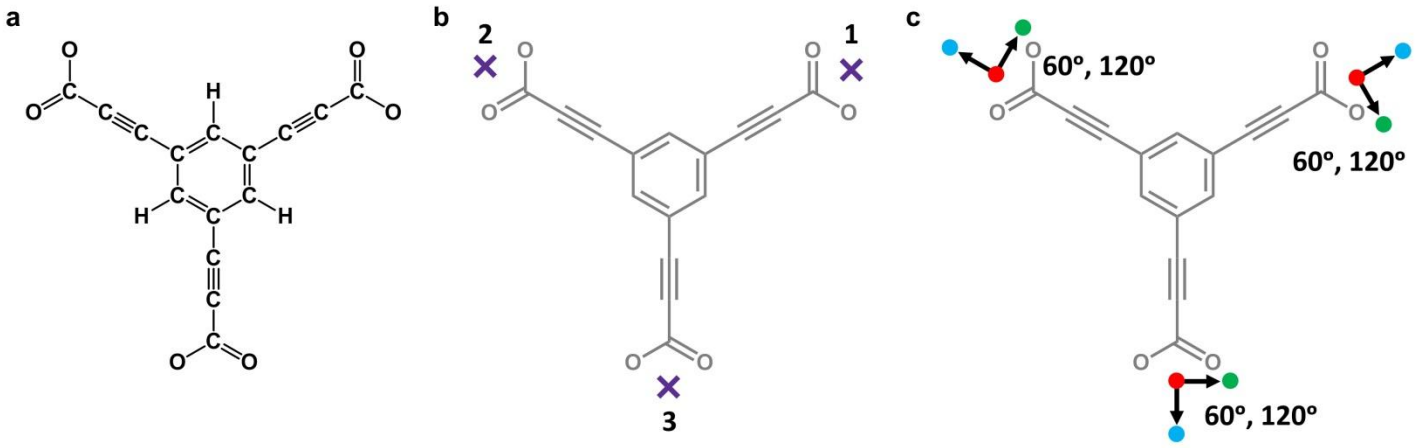

Figure S4. Encoded in the building blocks are the **(a)** atom composition and geometry, **(b)** topological information via numbered connection sites and **(c)** geometrical information via pseudo-atoms (colored as red, green and blue dots for *R*, *G* and *B* pseudo-atoms respectively) and lists of angles for alternative orientations.

Generation algorithm. The algorithm enumerates all possible combinations of building blocks in all possible arrangements. This is possible because the building blocks are numbered, and also because all possible arrangements of building blocks can be written as enumerable strings. For example, the string "1-2-3-1-2-3" means:

- "**1**-2-3-1-2-3" → place a building block of type 1 anywhere
- "1-**2**-3-1-2-3" → select a building block of type 2 (not yet placed anywhere)
- "1-2-**3**-1-2-3" → select its connection site 3
- "1-2-3-**1**-2-3" → connect the selected building block to the 1st building block placed
- "1-2-3-1-**2**-3" → connect the selected connection site to the 2nd connection site on the 1st building block
- "1-2-3-1-2-**3**" → rotate the selected building block using the 3rd angle listed at the selected connection site

This string of six integers completely describes the connection between two building blocks, and thus the specification of a complete crystal is achieved by catenation of many such strings into, for lack of a better term, a "superstring" (e.g., "1-2-3-1-2-3"-"4-5-6-4-5-6"-"1-1-1-2-2-2"-"2-1-2-1-2-1"). This superstring may encode the construction of a complete crystal composed of dozens of types of building blocks, but in this work we applied a constraint such that no superstring contains more than four types of building blocks at once. Specifically, we considered superstrings that specified at most only one inorganic, one functional-group, and two organic types of building blocks simultaneously. Note that the constraint is a limit on the maximum number of types of building block, and a particular superstring may contain fewer types.

Needless to say, not every superstring corresponds to a meaningful crystal. In fact, even the shorter six character strings can be invalid, such as "1-2-3-1-2-3," which leads to a nonsensical connection of two inorganic building blocks (building blocks #1 and #2 in Figure S1). Here the generator would skip to the next arrangement, namely, "1-2-3-1-2-4", and so-forth. The generation procedure is summarized in Figure S5. In this example, if "1-2-3-1-2-3"-"4-5-6-4-5-6" was the n^{th} arrangement, then "1-2-3-1-2-3"-"4-5-6-4-5-7" would be called the $(n+1)^{th}$ arrangement. The number of all possible arrangements for a set of a building blocks is n_{max}. In our screening procedure, if no logical MOF structure could be generated in the first 64,000 arrangements of the chosen building blocks, then the arrangement superstring was incremented by a large random value (e.g., "1-2-3-1-2-3"-"4-5-6-4-5-7" might jump to "4-1-3-2-4-4"-"1-3-4-2-2-5", thus trying a radically different arrangement of building blocks). Note that, a new arrangement superstring does not need to preserve the number of building blocks used by the previous superstring. The only variable that was fixed when a new arrangement string was generated was the constraint on building block types. For each set of building block constraints (denoted by *i*, *j*, *k* and *m* in Figure S5), five arrangement strings were attempted, yielding between zero and five hypothetical MOF structures. If the n^{th} arrangement resulted in a successful MOF, the $(n+1)^{th}$ arrangement was subsequently considered. If no MOF structure could be found after 5 random increments, then the next set of building block constraints was chosen (see Figure S5).

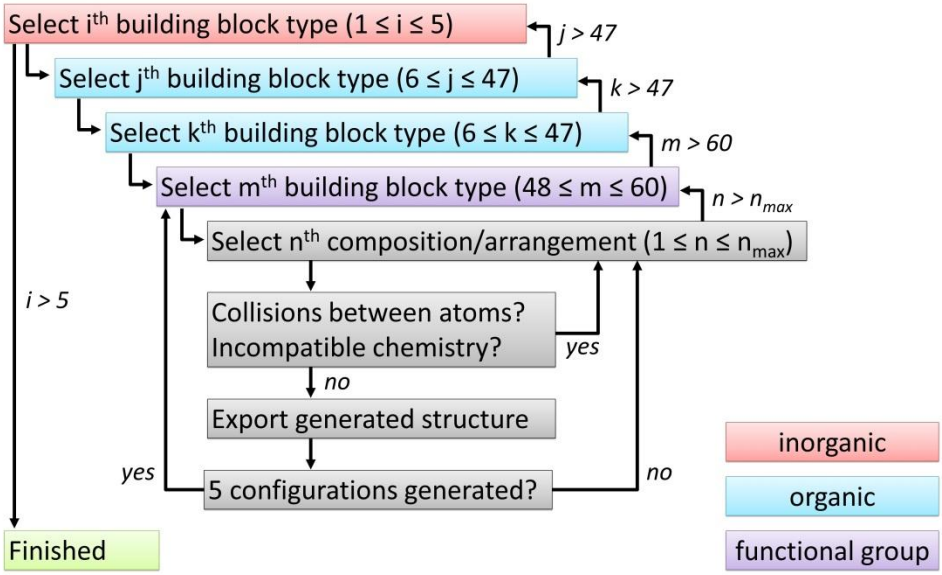

Figure S5. A flowchart depicting how hypothetical MOFs are enumeratively generated from a library of building blocks. The upper and lower limits of *i*, *j*, *k*, and *m* refer to the numbered building blocks in Figure S1. In the "select nth composition/arrangement" step, the total number and arrangement of building blocks is encoded in an enumerable string. In the particular library of building blocks we used, functional groups could be connected in any location where a hydrogen atom is otherwise bonded to a carbon atom, provided no atomic collisions occur. In the following "collisions between atoms" step, structures were "colliding" if any two atoms were closer than one angstrom apart. This distance was used so as not to discard potentially interesting MOFs due slight structural errors introduced in the generation process.

S3. Database of Hypothetical Metal-Organic Frameworks – Available Online

The full database of hypothetical crystals is available online at hmofs.northwestern.edu. Every crystal structure may be downloaded in the CIF format. Pore-size distributions and simulated powder x-ray diffraction patterns may be downloaded in the comma separated value (CSV) format (despite the format name, data is separated using white space characters). The reader is encouraged to unearth structure-property relationships we have not recognized in the main text (of which there are certainly many, for example, see Figure S6[1]).

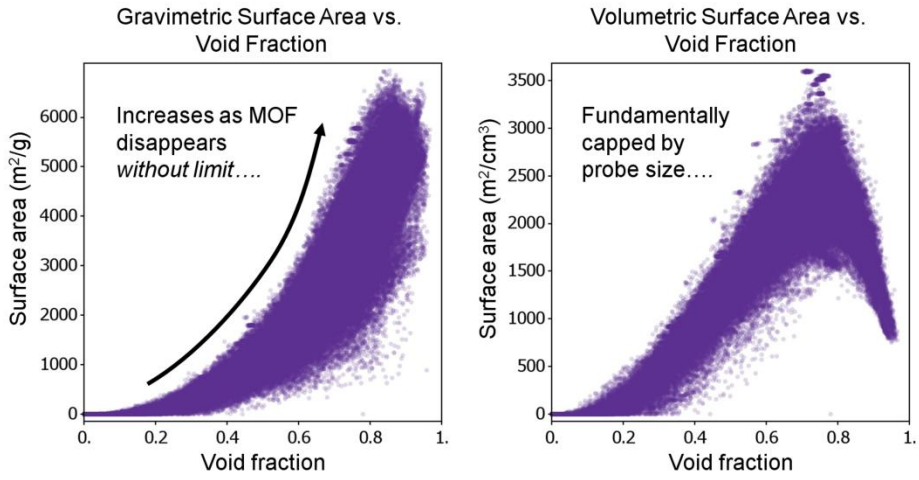

Figure S6. Two structure-property relationships[1] found in our database that are purely geometric in nature. Importantly, this data suggests that volumetric surface area is fundamentally capped to a value determined by the probe size used, whereas gravimetric surface area shows no obvious limit.

S4. Grand Canonical Monte Carlo Simulations - Details & Benchmarking

Atomistic grand canonical Monte Carlo (GCMC) simulations were performed to estimate the adsorption isotherms of Xe/Kr mixtures in all of the hypothetical MOFs. Interaction energies between non-bonded atoms were computed through the Lennard-Jones (LJ) potential:

$$\mathcal{V}_{ij} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

where r_{ij} is the distance between atoms i and j and ε_{ij} and σ_{ij} are the LJ well depth and diameter, respectively. LJ parameters between atoms of different types were calculated using the Lorentz-Berthelot mixing rules (i.e., geometric average of well depths and arithmetic average of diameters). LJ parameters for framework atoms were taken from the Universal Force Field (UFF).[2] Krypton and xenon are modeled as single spheres, with parameters taken from Hirschfelder et al.[3] and Talu and Myers,[4] respectively.

Table S1. Lennard-Jones parameters for xenon, krypton and framework atoms in all hypothetical MOFs.

Atom type	LJ σ (Å)	LJ ε/k_B (K)
C	3.43	52.83
O	3.12	30.19
H	2.57	22.14
N	3.26	34.72
F	2.997	25.16
Cl	3.517	114.23
Br	3.73	126.30
Zn	2.46	62.40
Cu	3.114	2.516
V	2.80	8.05
Zr	2.783	34.72
Xe	4.10	221.0
Kr	3.636	166.4

All GCMC simulations of Xe/Kr adsorption included an *M*-cycle equilibration period followed by an *M*-cycle production run, where *M* was 1000, 5000, or 10,000 as described in the main text (see Figure 4). A cycle consists of *N* Monte Carlo steps; where *N* is equal to the number of molecules (which fluctuates during a GCMC simulation). All simulations included random insertion, deletion, translation and identity change moves. Atoms in the MOF were held fixed at their crystallographic positions. An LJ cutoff distance of 12.0 Å was used for all simulations. A 2x2x2 unit cell of every crystal was used for the simulations. Xe/Kr adsorption was simulated at three pressures, 1.0, 5.0 and 10 bar, at 273 K for all crystals. Fugacities needed to run the GCMC simulations were calculated using the Peng-Robinson equation of state.[5]

In order to determine how many GCMC cycles were needed to obtain sufficiently accurate results for Xe/Kr selectivity screening, we compared simulation data on three MOFs, HKUST-1,[6] ZIF-8[7] and IRMOF-1[8] at 1 bar and 273 K at various numbers of GCMC cycles (see Figure S7).

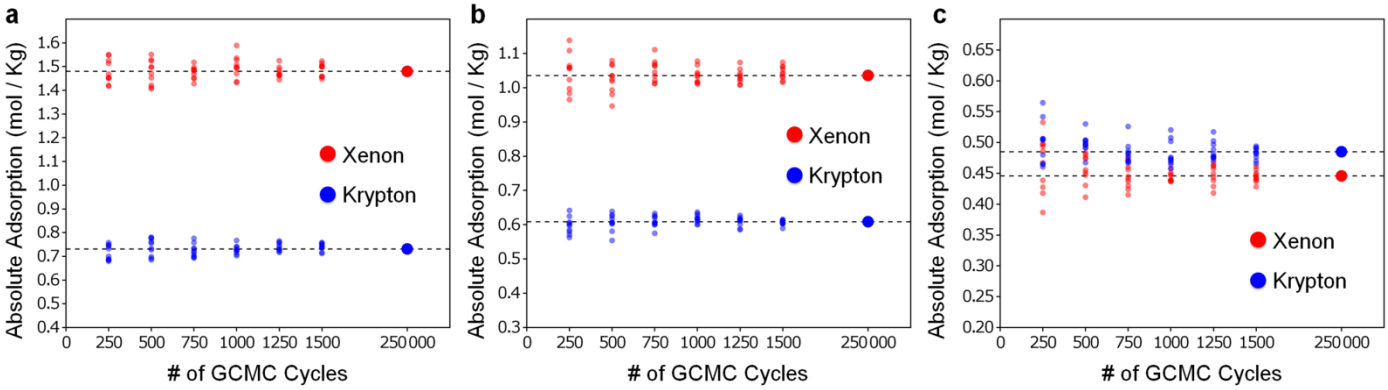

Figure S7. GCMC simulations of Xe/Kr mixtures at 1 bar at varying levels of quality for **(a)** HKUST-1, **(b)** ZIF-8 and **(c)** IRMOF-1. Each dot corresponds to one simulation run, and for every number of GCMC cycles chosen, 10 simulation runs were performed.

S5. Geometric Analysis – Framework Atom Details

The geometric analysis of the hypothetical MOF database requires that each framework atom be treated as a hard sphere. In doing so, the occupied volume (volume occupied by the framework atoms) of every tetrahedron created by the Delaunay tessellation can be calculated. Table S2 lists the hard sphere diameters obtained from the Cambridge Structural Database (CSD).[9]

Table S2. Hard sphere diameters for xenon, krypton and framework atoms in all hypothetical MOFs. Values marked with an asterisk (*) were not available in the CSD and the σ values from UFF were used instead.

Atom type	van der Waals D (Å)
C	3.40
O	3.04
H	2.18
N	3.10
F	2.94
Cl	3.50
Br	3.70
Zn	2.78
Cu	2.80
V	2.80*
Zr	2.783*

S6. Plots Displaying Complete Data Sets

The trade-off between selectivity and adsorption capacity at different pressures is conveyed via overlapping data sets in the main text (Figure 5). The individual data sets, unobstructed and on a linear scale, are shown below.

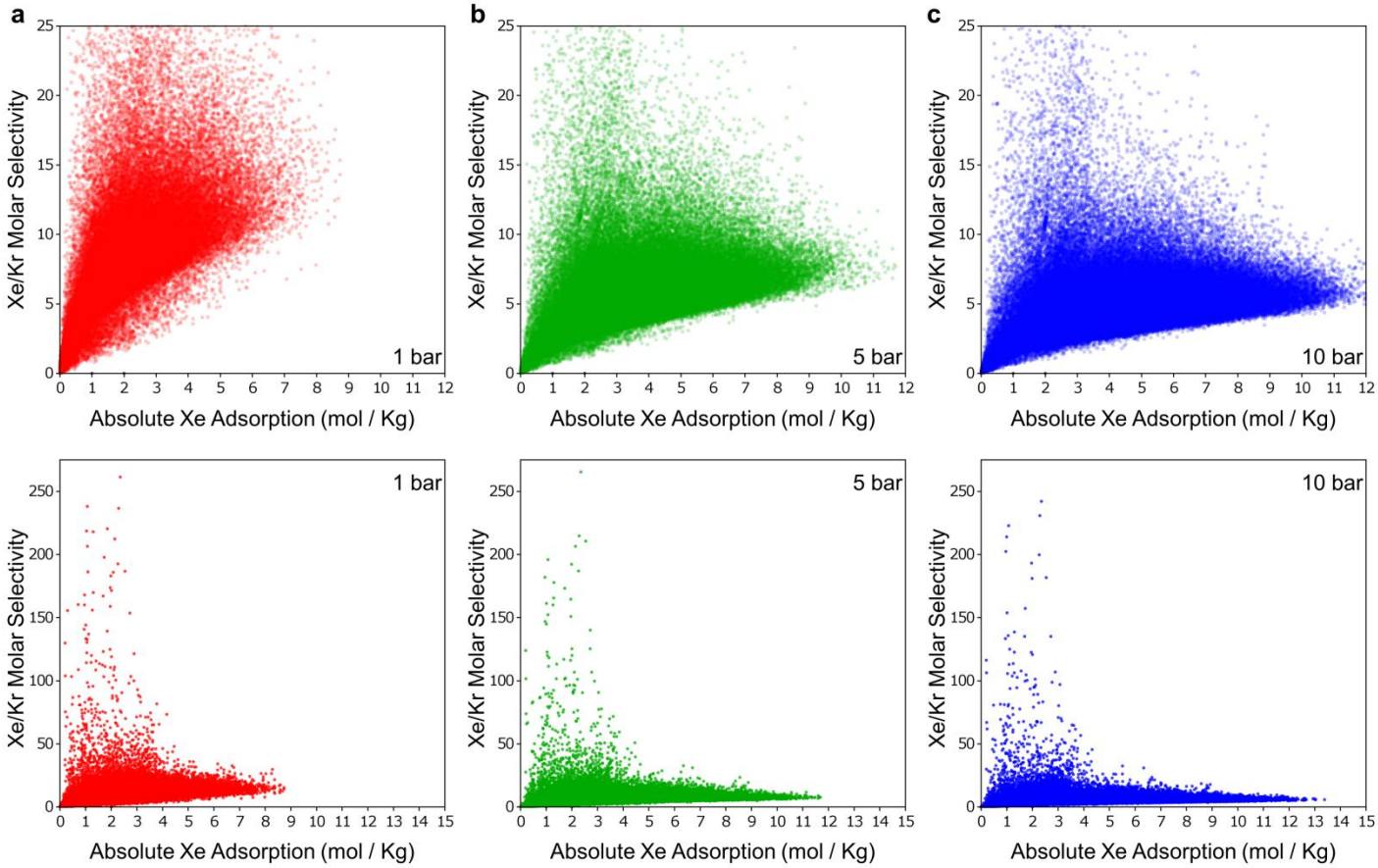

Figure S8. GCMC simulations of Xe/Kr mixtures, showing plot regions chosen for (**top**) clarity and for (**bottom**) completeness at **(a)** 1 bar, **(b)** 5 bar and **(c)** 10 bar showing selectivity versus absolute adsorption.

In the main text, Figures 6 and 7 do not show the entire data set because the plot regions were chosen for clarity. Below, the same data are plotted with larger plot regions that show the entire data sets.

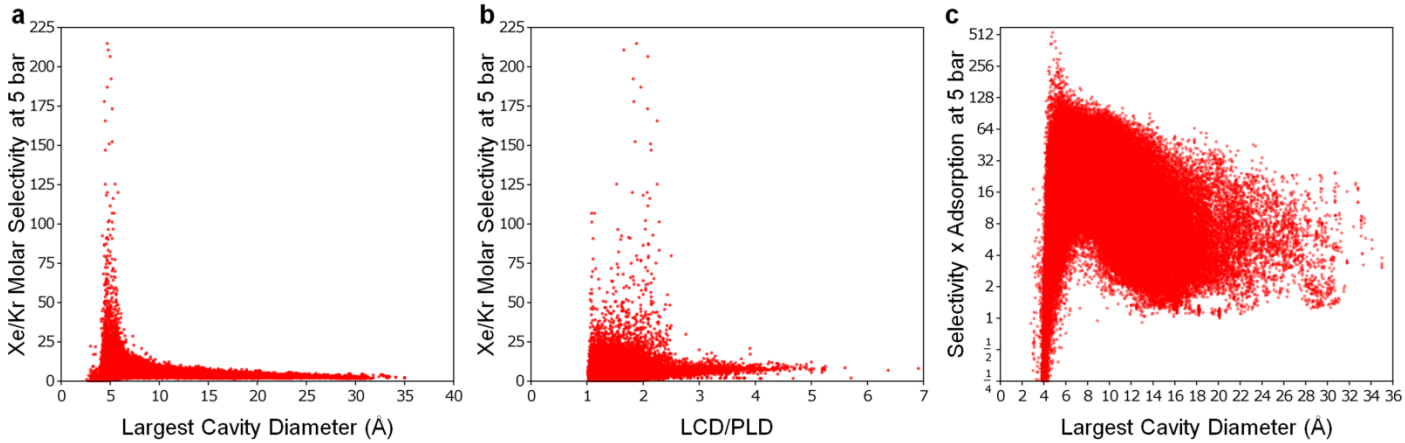

Figure S9. GCMC simulations of Xe/Kr mixtures, showing plot regions chosen for completeness. The graphs correspond to **(a)** Figure 6a, **(b)** Figure 6b, and **(c)** Figure 7 in the main text.

S7. Accessibility of Frameworks

GCMC simulations allow for the insertion of Xe and Kr atoms into pores regardless of whether those void spaces are accessible from the outside of the crystal. This opens the possibility of false-positives when identifying promising candidates based on GCMC computations alone – namely, a MOF may appear to be highly selective for xenon when in fact those atoms could never diffuse into the structure. However, the geometric analysis of the hypothetical database found that in the vast majority of cases (see Figure S10) structures contained *percolating* accessible volumes (the voids form a network through which a probe is able to travel unhindered) equal to their accessible volumes (the space that can fit an inserted probe molecule). Thus, in these cases, cavities with volumes that are accessible by Monte Carlo insertion are also accessible kinetically. In the spirit of this screening procedure, the top candidates should be analyzed in more detail.

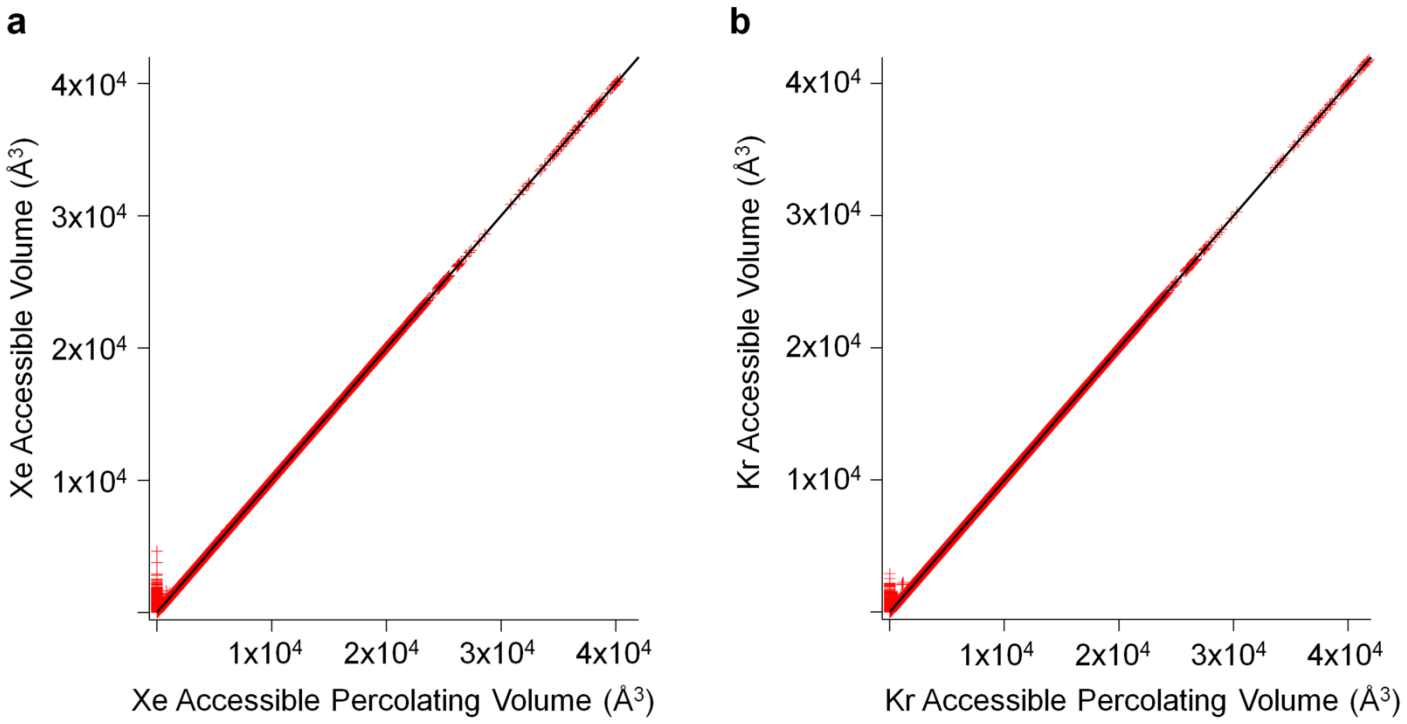

Figure S10. Parity plots comparing accessible volume against accessible percolating volume. Since the volume is dependent on the probe size, Xe **(a)** and Kr **(b)** were looked at separately. Units of volume are expressed per repeating unit cell.

References

(1) Wilmer, C. E.; M. Leaf; Lee, C.-. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2011**, doi:10.1038/nchem.1192.
(2) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.
(3) Hirschfelder, J. O.; Curtiss, C. F.; Bird, R. B. *The Molecular Theory of Gases and Liquids*; Revised.; Wiley-Interscience, 1964.
(4) Talu, O.; Myers, A. L. *Colloid. Surface. A* **2001**, *187-188*, 83-93.
(5) Frost, H.; Düren, T.; Snurr, R. Q. *J. Phys. Chem. B* **2006**, *110*, 9565-9570.
(6) Chui, S. S.-Y.; Lo, S. M.-F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. *Science* **1999**, *283*, 1148-1150.
(7) Park, K. S.; Ni, Z.; Côté, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10186-10191.
(8) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. *Nature* **1999**, *402*, 276-279.
(9) Cambridge Structural Database (CSD) www.ccdc.cam.ac.uk/products/csd.