

Supporting Information for: Eigencages: Learning a latent space of porous cage molecules

Arni Sturluson, Melanie T. Huynh, Arthur H. P. York, and Cory M. Simon*

School of Chemical, Biological, and Environmental Engineering. Corvallis, OR, USA.

E-mail: Cory.Simon@oregonstate.edu

S1 Distribution of cage molecule and cavity diameters

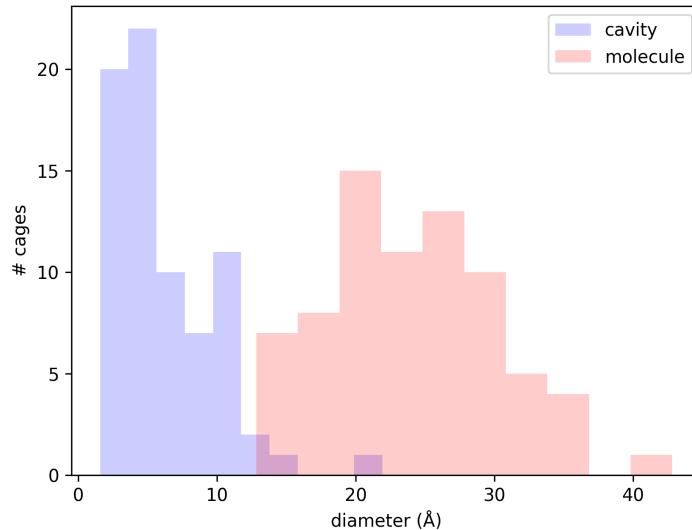


Figure S1.1: The distribution of cage molecule diameters (red) and cage cavity diameters (blue) among the 74 cages analyzed in this work. These diameters were computed with pywindow.¹

S2 Aligning the cages before scanning for void space

We compute the 3×3 symmetric moment of inertia matrix \mathbf{I} of the cage, defined relative to the center of mass, and find its eigendecomposition $\mathbf{I} = \mathbf{W}\Lambda\mathbf{W}^T$, with the eigenvalues arranged in a non-increasing order down the diagonal matrix Λ and the orthonormal eigenvectors in the columns of \mathbf{W} . The eigenvectors of the matrix \mathbf{I} are the principal axes of inertia, and the associated eigenvalues are the moments of inertia with respect to those axes. Then, we transform the coordinates of the centered cage by left-multiplying by the rotation matrix \mathbf{W}^T . The rotated cage then possesses a diagonal moment of inertia matrix with the moments of inertia arranged in a non-increasing order down the diagonal.

S2.1 Degenerate principal axes of inertia

Molecules that exhibit symmetry can possess degenerate moments and principal axes of inertia. As an example, consider sulfur trioxide (SO_3), a trigonal planar molecule, shown in Fig. S2.1. Its moment of inertia matrix \mathbf{I} , with Cartesian directions defined in Fig. S2.1 and the sulfur atom at the origin, is:

$$\mathbf{I} = m_O \ell_{S-O}^2 \begin{bmatrix} 3/2 & 0 & 0 \\ 0 & 3/2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad (\text{S1})$$

where m_O is the mass of an oxygen atom and ℓ_{S-O} is the sulfur-oxygen bond length. The eigenvalues of the moment of inertia matrix \mathbf{I} are $3/2$, $3/2$, and 3 times $m_O \ell_{S-O}^2$. The eigenvector (unique if unit norm enforced) associated with the largest eigenvalue, $3m_O \ell_{S-O}^2$, is $[0, 0, 1]$, indicating that the principal axis of inertia with the largest moment of inertia is pointing into the page with respect to Fig. S2.1. The remaining two principal axes of inertia must be orthogonal to $[0, 0, 1]$ and thus lay in the x - y plane. The moments of inertia about the remaining two principal axes of inertia, however, are equal, as the remaining

eigenvalues $3/2$, $3/2$ times $m_O \ell_{S-O}^2$ are degenerate. In fact, *any* vector laying in the x - y plane are eigenvectors of the matrix \mathbf{I} in eqn. S1 associated with eigenvalue $3m_O \ell_{S-O}^2/2$. This mathematically indicates that the two remaining principal axes of inertia for SO_3 are not unique.

Figure S2.1: Sulfur trioxide (SO_3) is a trigonal planar molecule with degenerate principal axes of inertia in the x - y plane.

As a consequence, the principal axes of inertia for a symmetric molecule such as SO_3 do not provide a means to uniquely align the molecule. Thus, e.g. to compare SO_3 with respect to a molecule exhibiting the same trigonal planar symmetry, such as boron trifluoride BF_3 , the principal axes of inertia cannot be used to align the two molecules along their bonds (aside from the principal axes of inertia associated with the largest moment of inertia, [001] in Fig. S2.1). With regards to its rotational dynamics, SO_3 behaves as a uniform disk.[?]

Several cage molecules exhibit a high degree of rotational symmetry, so as to behave akin to SO_3 , where the principal axes of inertia are not unique and thus cannot be used to consistently align them. For example, the three moments of inertia of CC1 (15869.8, 15869.6, 15869.4 amu- \AA^2) and of CC3 (32767.1, 32766.9, 32766.6 amu- \AA^2) are nearly equal, indicating that the rotational dynamics of CC1 and CC3 are akin to a sphere; therefore, using the principal axes of inertial to align CC1 and CC3 (i) theoretically will not provide a unique manner in which to align them and (ii) will be extremely sensitive to small perturbations in the structure. Note that CC3 and CC1 differ only in the vertices external to the cavity, but, as Fig. S2.2 shows, CC1 and CC3 are aligned irrationally on the basis of their computed principal axes of inertia due to the [near-]degeneracy of their axes of inertia. Another example that is highly related to the SO_3 example in Fig. S2.1 are B6 and B8, also shown in Fig. S2.2. Here, small differences in the three propeller-blade-like moieties that protrude from B6 and B8 result in significantly inconsistent alignments for comparing B6 and B8. B6 and B8 much resemble the SO_3 example in Fig. S2.1, as two principal axes of inertia are nearly degenerate and thus the principal axes in that plane (the plane of the page in Fig. S10.1) are very

sensitive to small perturbations in the structure.

Therefore, the principle axes of inertia are in many cases insufficient to determine an alignment of a porous cage molecule that is unique and robust to small differences. Fig. S2.3 displays the three moments of inertia with respect to the principal axes of inertia for each centered cage, normalized by the largest moment of inertia. Cages with uncolored bars, such as MC1 and MC4, display a significant difference between the three moments of inertia; therefore for these cages, the principal axes of rotation are a good way to consistently align these cages with similar cages. On the other hand, cages that have colored bars, such as CC1 and CC3, possess moments of inertia that are quite close to one another. We define consecutive moments of inertia to be *practically degenerate* (and color them in Fig. S2.3) when placing a single carbon atom at the periphery of the molecule in the direction of a principal axis changes the ranking of the principal axes of inertia. i.e., if the moments of inertia differ by less than $12r^2$ amu-Å, where r is the radius of the cage molecule, we decide that the principal axes of inertia do not provide a reasonable method to consistently align the cage with similar cages because small differences can drastically change the orientations.

We conclude that the set of cages with practically degenerate principal axes of inertia in Fig. S2.3 cannot be aligned using their principal axes of inertia because these alignments will not be robust to small changes in the structure.

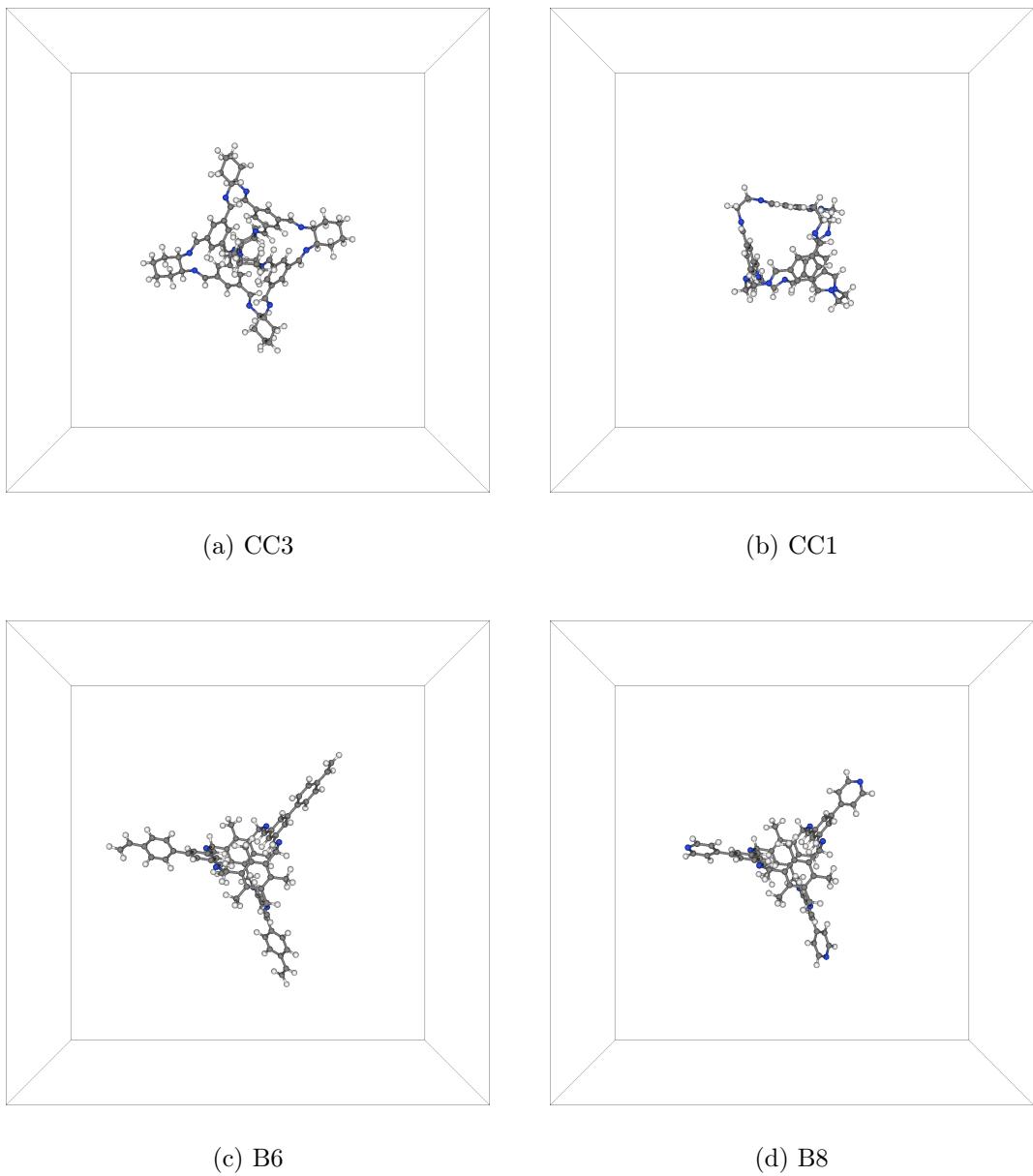


Figure S2.2: Cases where the computed principal axes of inertia do not provide reasonable alignments of cages. (a, b) The rotational dynamics of CC1 and CC3 are akin to a sphere; the three moments of inertia for each CC1 and CC3 are computed to be approximately equal (CC1: 15869.8, 15869.6, 15869.4 amu- \AA^2 ; CC3: 32767.1, 32766.9, 32766.6 amu- \AA^2). Shown are the two cages aligned with their computed principal axes of inertia. Despite differing only in the their vertices at the periphery of the cage, the two molecules are aligned inconsistently/irrationally. (c, d) Shown are cages B6 and B8 aligned with their principal axes of inertia. B6 and B8 have two nearly degenerate axes of inertia (B6: 39371.2, 29141.2, 28881.8 amu- \AA^2 ; B8: 27874.4, 23253.6, 23190.7 amu- \AA^2). As a result, despite only subtle differences in the three propeller-blade-like moieties that protrude from their cores, the molecules are not aligned consistently.



Figure S2.3: Moments of inertia of the cages about their principal axes of rotation. Each panel corresponds to a different cage. The three bars are the three moments of inertia about the principal axes of rotation (the eigenvalues λ_i of the moment of inertia matrix), normalized by the largest moment of inertia. Bars are colored if two consecutive moments of inertia differ by less than $12r^2$ amu-Å, where r is the radius of the cage molecule. Red: all moments of inertia are practically degenerate; blue, green: the first two and last two, respectively, moments of inertia are practically degenerate. For cages with colored bars, the principal axes of inertia do not provide a reasonable method to consistently align the cage with similar cages because small differences can drastically change the orientation.

S3 The data matrix \mathbf{A}



Figure S3.1: An attempt at visualizing the structure of the $c \times g^3$ data matrix \mathbf{A} . The color depicts the magnitude of the entry of the matrix. This is an “attempt” because $g^3 \gg c$, so the data matrix \mathbf{A} is wide and difficult to visualize without a longer page. Zoom in to see some structure. We remark that $10879/15625$ of the columns (representing pixels in the cage void space images) are all zeros.

S4 The singular values of \mathbf{A}

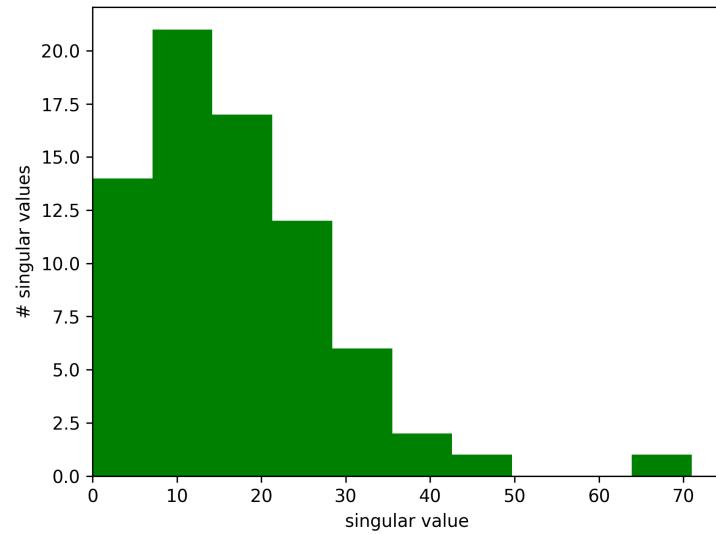


Figure S4.1: The distribution of the set of singular values $\sigma_1, \sigma_2, \dots, \sigma_{74}$ of \mathbf{A} .

S5 Relative error from approximating \mathbf{A} as \mathbf{A}_ν

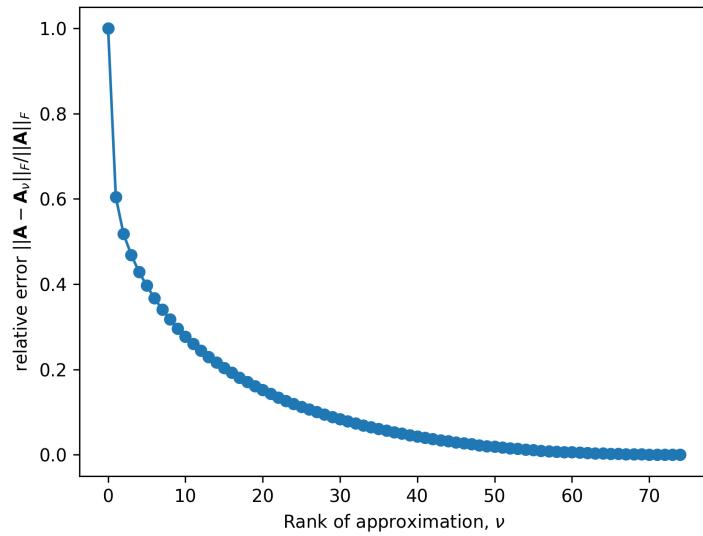


Figure S5.1: The relative error in approximating the data matrix \mathbf{A} with a lower-rank- ν approximant \mathbf{A}_ν given in eqn. 6. The formula for the relative error as a function of ν , related to the singular values of \mathbf{A} , is given in eqn. 7.

S6 Highlighting remaining clusters in latent cage space

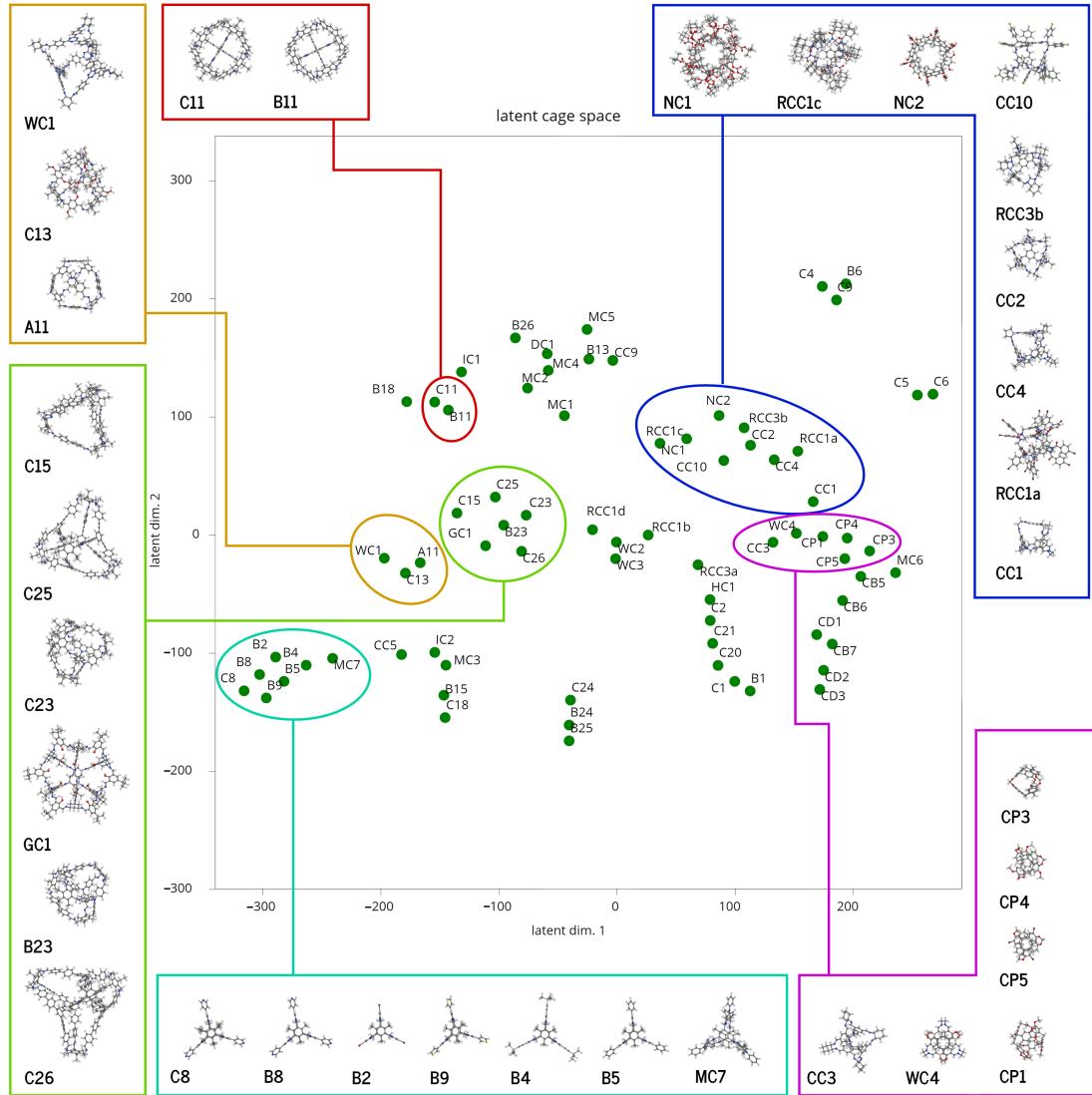


Figure S6.1: The latent space of cages $\mathbf{U}_\nu \Sigma_\nu$, embedded into 2D by t-SNE.^{2,3} Salient clusters are highlighted. See Fig. 7 for the remaining clusters.

S7 Henry coefficient calculations

We describe more details here of the Henry coefficient calculations used to obtain the data in Fig. 8. We model the energetics of the interaction of a gas atom (Xe, Kr, He) with the atoms of the cage as pairwise additive and with 12-6 Lennard-Jones potentials. The ϵ_i and σ_i parameters of the Lennard-Jones potentials for atom i - atom i interactions are taken from the Universal Force Field⁴ (cutoff radius 14 Å). Geometric mixing rules are applied for cross-interactions.

We placed an isolated cage molecule in an empty simulation box and took it as rigid. If $\mathbf{x} \in \mathbb{R}^3$ is the position of a gas atom in the simulation box, the molecular model described above gives us the potential energy of the gas molecule, $U(\mathbf{x})$.

Consider the isolated porous cage molecule immersed in a bath of gas; the simulation box is drawn around the porous cage molecule. Henry's law models the density of gas in the simulation box at thermodynamic equilibrium, ρ , as $\rho = KP$, with K the Henry coefficient and P the pressure of the gas. Henry's law is valid only at low surface coverage.

The Henry coefficient of a gas in the simulation box including the isolated cage molecule is given as:

$$K = \beta \frac{1}{|\Omega|} \int_{\Omega} e^{-\beta U(\mathbf{x})} d\mathbf{x}, \quad (\text{S2})$$

where $\beta = (k_B T)^{-1}$ is the thermodynamic beta (T temperature, k_B Boltzmann constant) and the integral is over the simulation box Ω (which has volume $|\Omega|$). Note that if $U(\mathbf{x}) = 0$, Henry's law recovers the ideal gas law $\rho = \beta P$, where $|\Omega|$ is the volume of the simulation box.

We computed the average in eqn. S2 via Monte Carlo integration, i.e. Widom particle insertions,⁵ using 1000 Monte Carlo insertions per volume (Å³) of the simulation box. We used `PorousMaterials.jl v0.1.1`⁶ to compute the Henry coefficient.

S7.1 Comparing to experimental data

By considering only an isolated cage molecule in a simulation box, as opposed to many cage molecules packed together to form a solid, we account for only the influence of the intrinsic porosity to the cage molecule on the adsorption. This is an approximation, as we are neglecting the influence of the extrinsic porosity that arises from how the cage molecules pack together to form a bulk solid. When the cage molecules pack together, the adsorption sites on the exterior of the molecule or at the cage windows can be modified depending on how the cage molecules pack together. However, when the interior of the cage molecule is the dominant adsorption site in the bulk solid, this approach may be a viable method to predict the Henry coefficient of a bulk porous cage solid.

To evaluate our crude method of considering an isolated cage molecule, Fig. S7.1 compares experimentally measured xenon and krypton adsorption isotherms in noria⁷ (NC2) and CC3⁸ to the resulting Henry's law with the Henry coefficient obtained from our simulations. Fig. S7.1 shows this method yields a very good prediction of experimental Xe and Kr Henry coefficients in noria⁷ (NC2), but underestimates the Henry coefficients in CC3.⁸ See Patil et al.⁷ for more discussion.

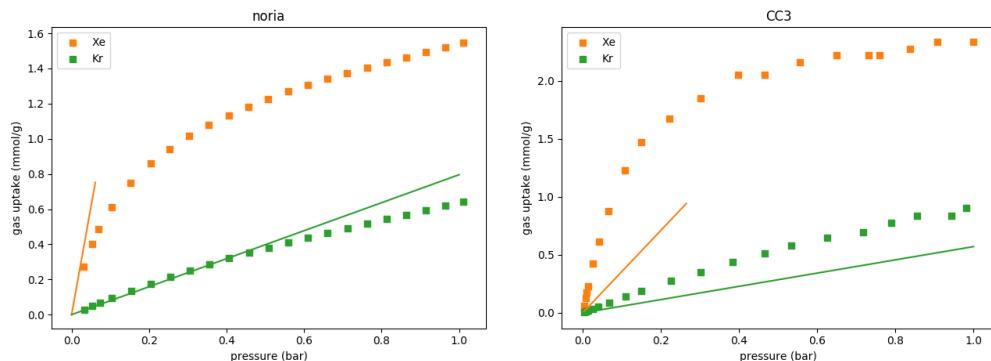


Figure S7.1: Comparison of experimental xenon and krypton adsorption isotherms at 298 K in noria⁷ and CC3⁸ to the simulated adsorption. Points show experimental data and the lines show Henry's law with the Henry coefficient taken from our simulations. The simulated Henry coefficient of He was subtracted to account for the empty space in the simulation box.

S8 Relationship of latent representation and cage molecule and cavity diameters

Fig. S8.1 shows that the first component of the learned latent representation of a cage is strongly correlated with its cavity diameter. Fig. S8.2 depicts how the 2D embedding (by t-SNE) of the $\nu = 23$ dimensional latent representations is correlated with the molecule diameter and cavity diameter of a porous organic cage molecule.

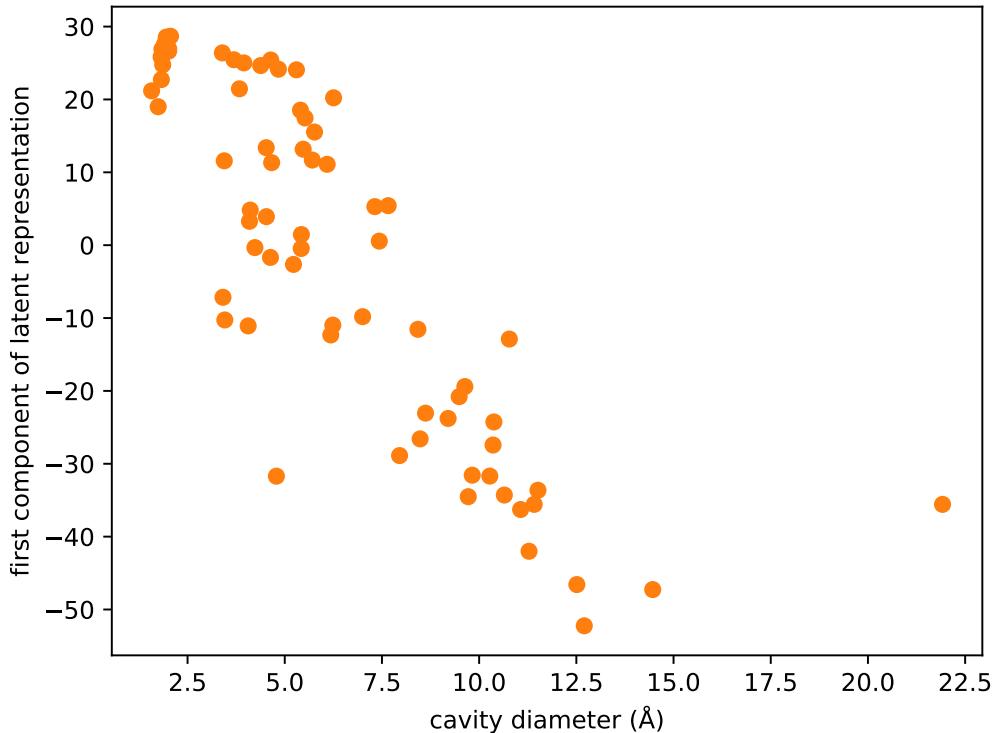


Figure S8.1: The first component of the latent representation of a cage is strongly correlated to its cavity diameter. The cavity diameter was computed by pywindow.¹ In the reconstruction of the 3D void space image of a given cage, the first component of the latent representation (in the first column of $\mathbf{U}_\nu \Sigma_\nu$) is the weight on the first eigencage \mathbf{v}_1 in the reconstruction of the 3D void space image via eqn. 8. That the first eigencage is a good descriptor of pore size is suggested by the radial symmetry of the core of the first eigencage in Fig. 5 that is added to the approximately radially symmetric average cage $\bar{\mathbf{c}}$ to reconstruct the 3D void space image via eqn. 8.

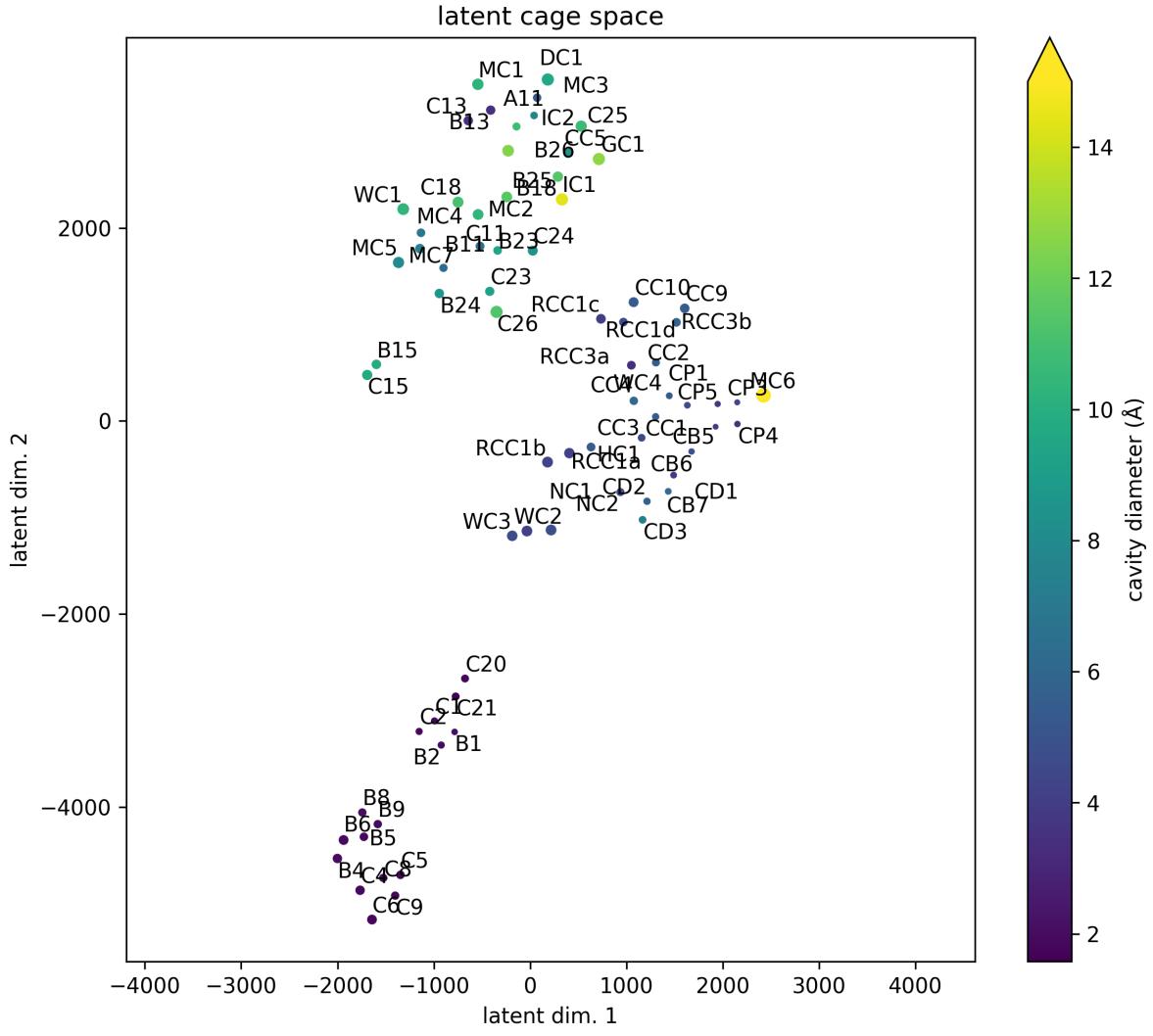


Figure S8.2: The 2D embedding (by t-SNE) of the learned latent representation of the 3D void space images contained in the rows of $\mathbf{U}_\nu \boldsymbol{\Sigma}_\nu$. Here, the diameter of points represents molecule diameter; color represents cavity diameter, both computed by `pywindow`.¹ Note that neighbors in latent space tend to have similar cavity and molecule diameters.

S9 Relationship of latent representation and number of windows to enter cavity

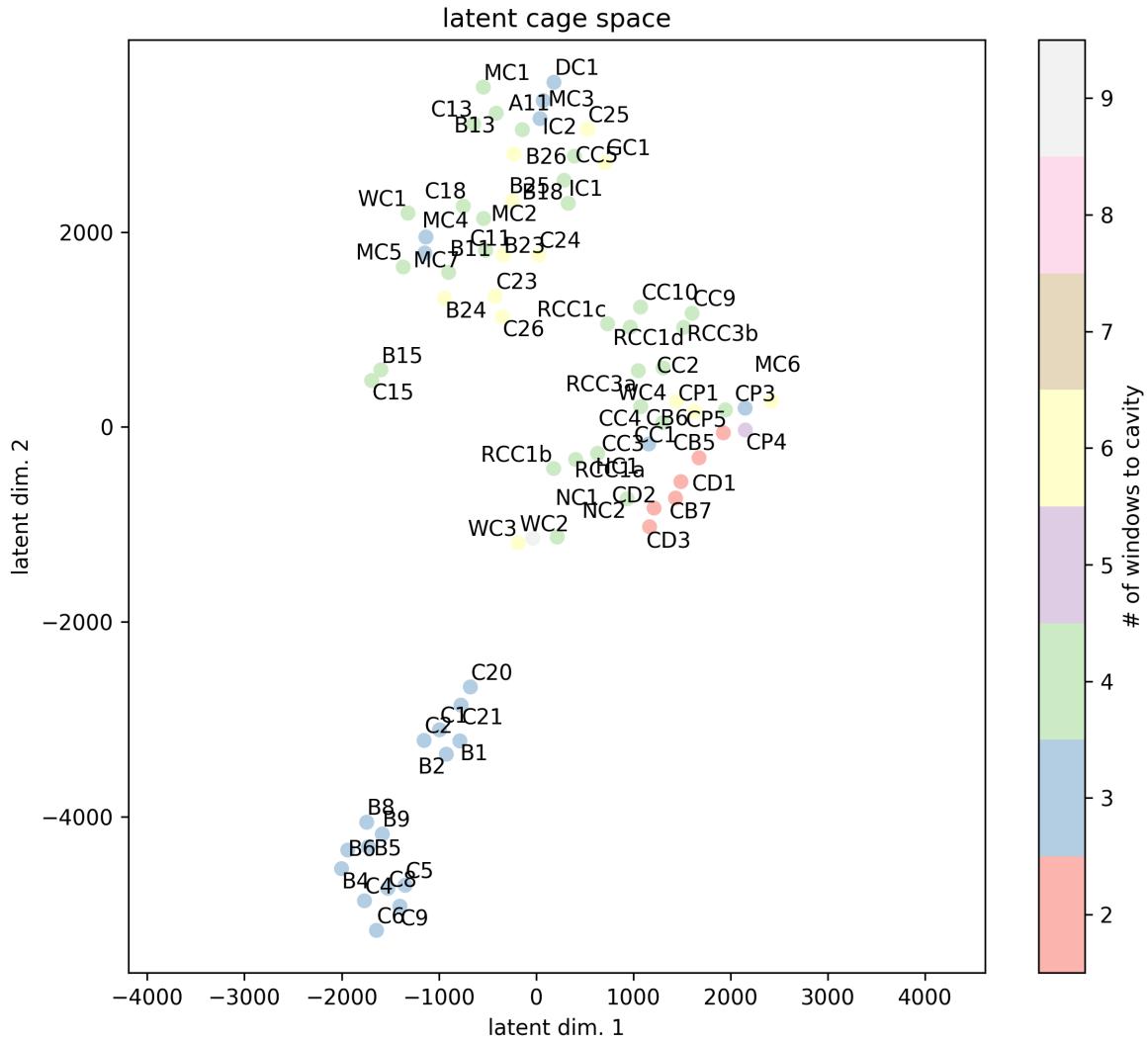
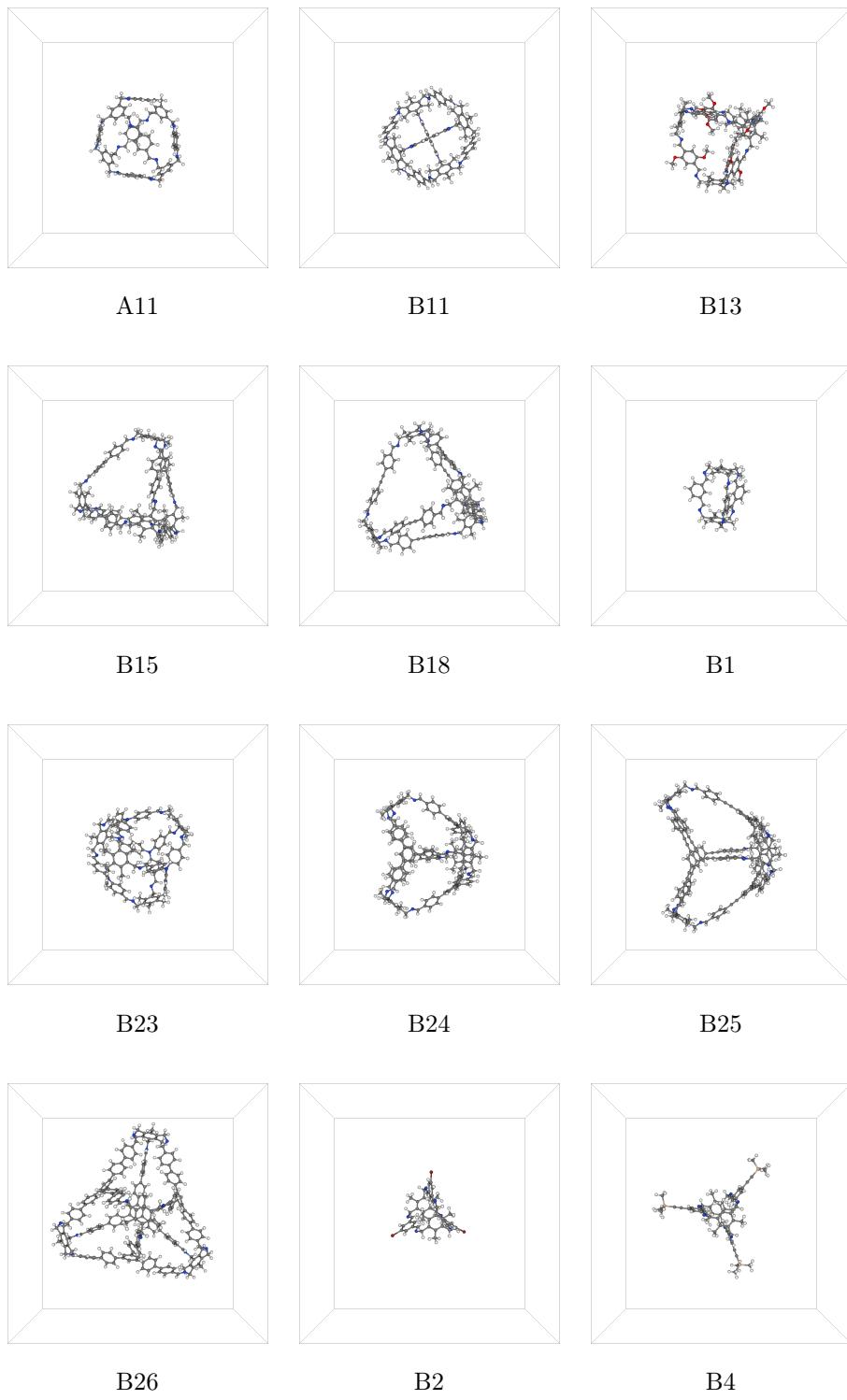
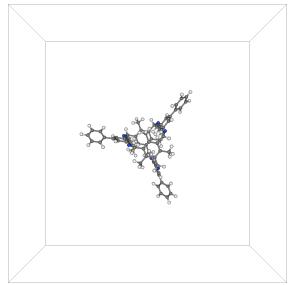


Figure S9.1: The 2D embedding (by t-SNE) of the learned latent representation of the 3D void space images contained in the rows of $\mathbf{U}_\nu, \Sigma_\nu$. Here, the color represents the number of windows possessed by the cage molecule. The gas molecules enter the cavity of the cage through the windows. The number of windows is computed by `pywindow`.¹ Note that cages within clusters in latent space have the same number of windows.

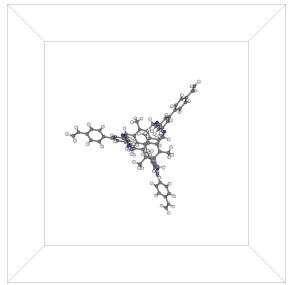
S10 Visualization of the cage structures

Larger images of the cages in Fig. 1 are shown in Fig. S10.1.

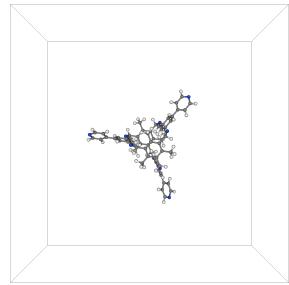




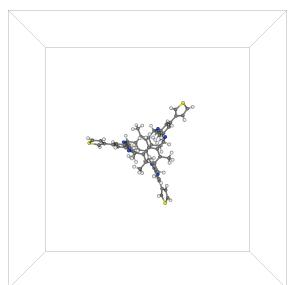
B5



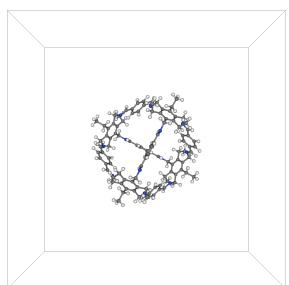
B6



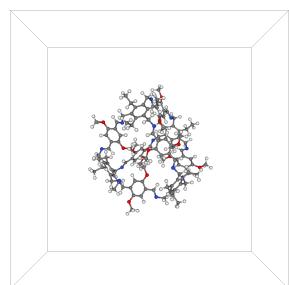
B8



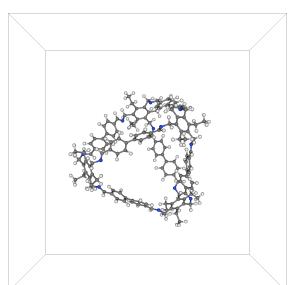
B9



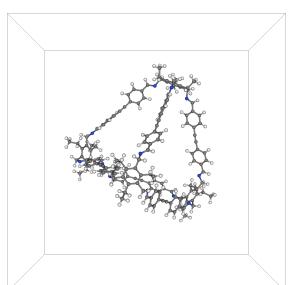
C11



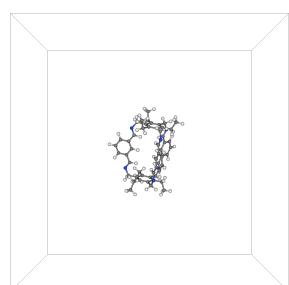
C13



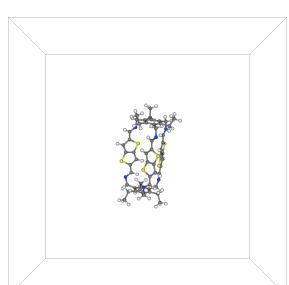
C15



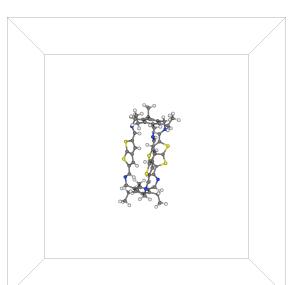
C18



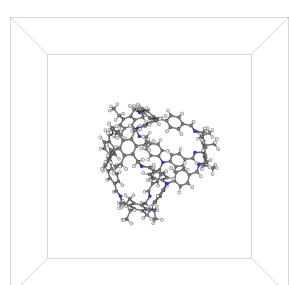
C1



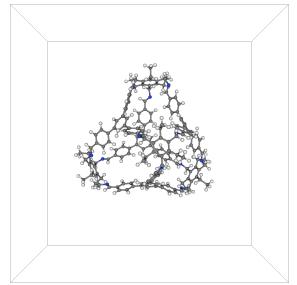
C20



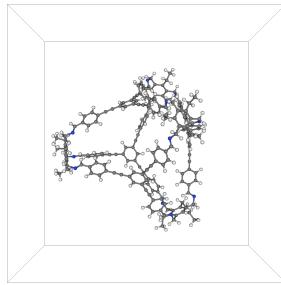
C21



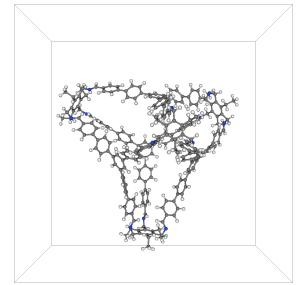
C23



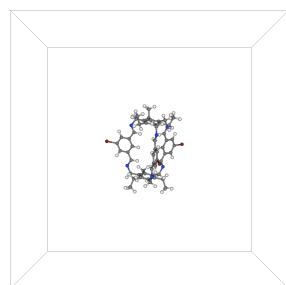
C24



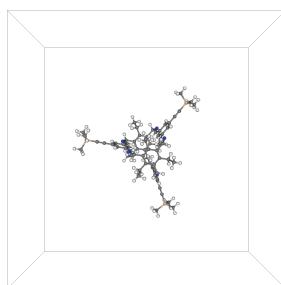
C25



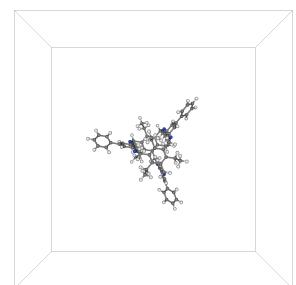
C26



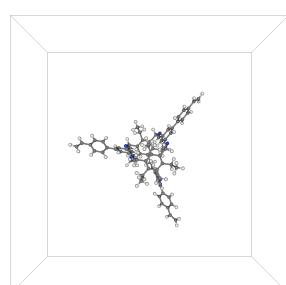
C2



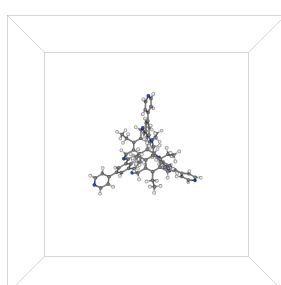
C4



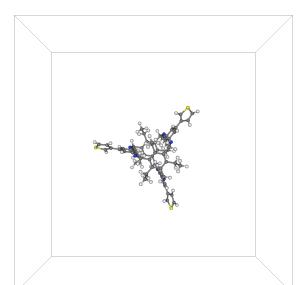
C5



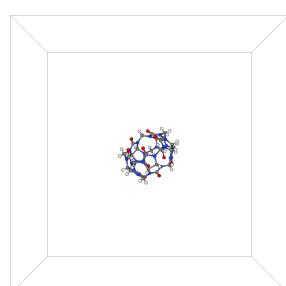
C6



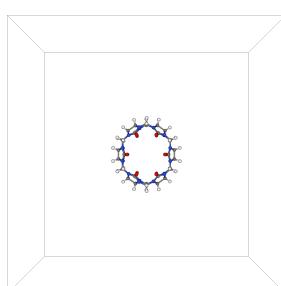
C8



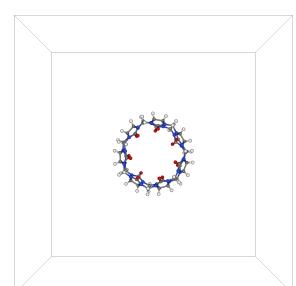
C9



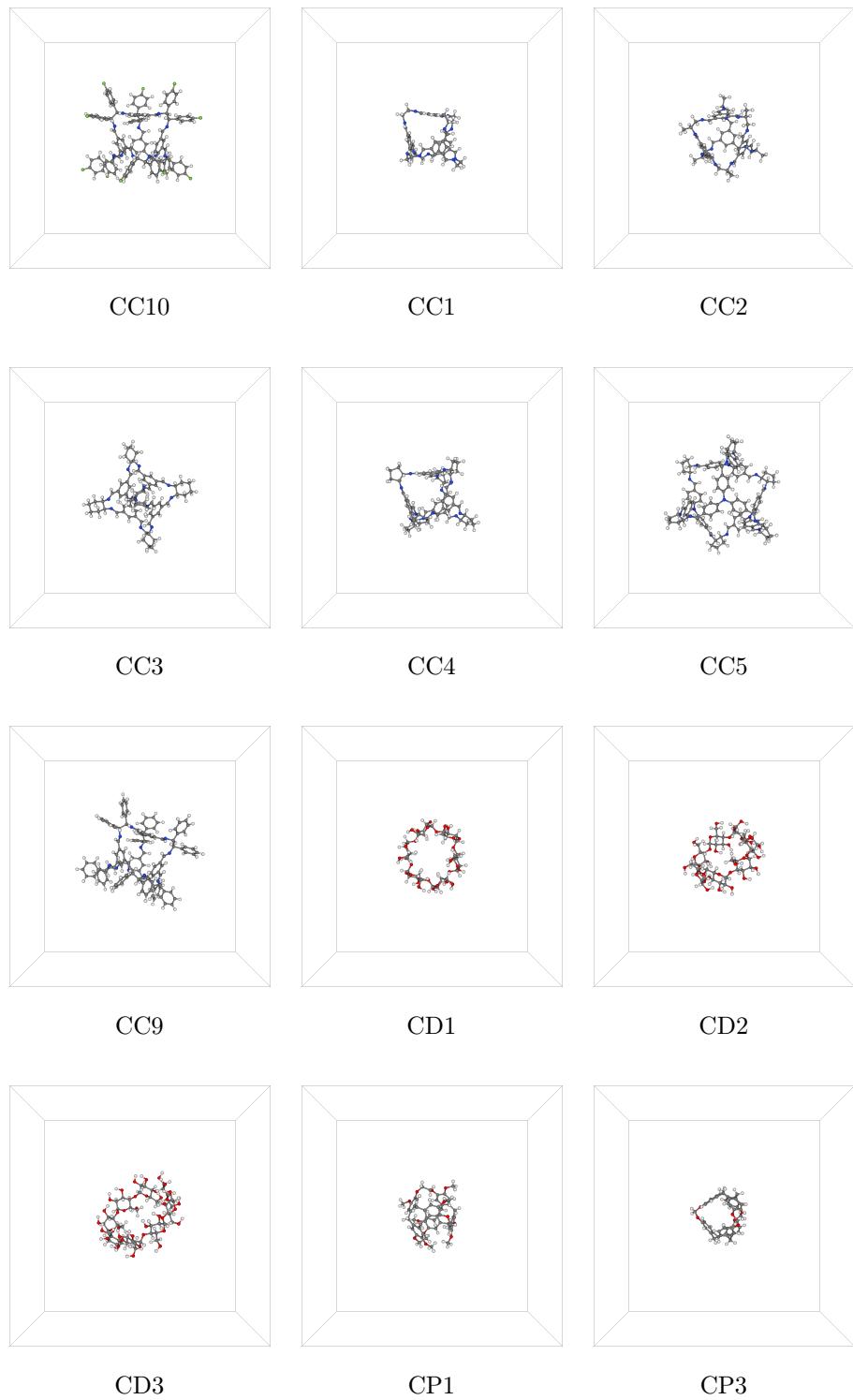
CB5

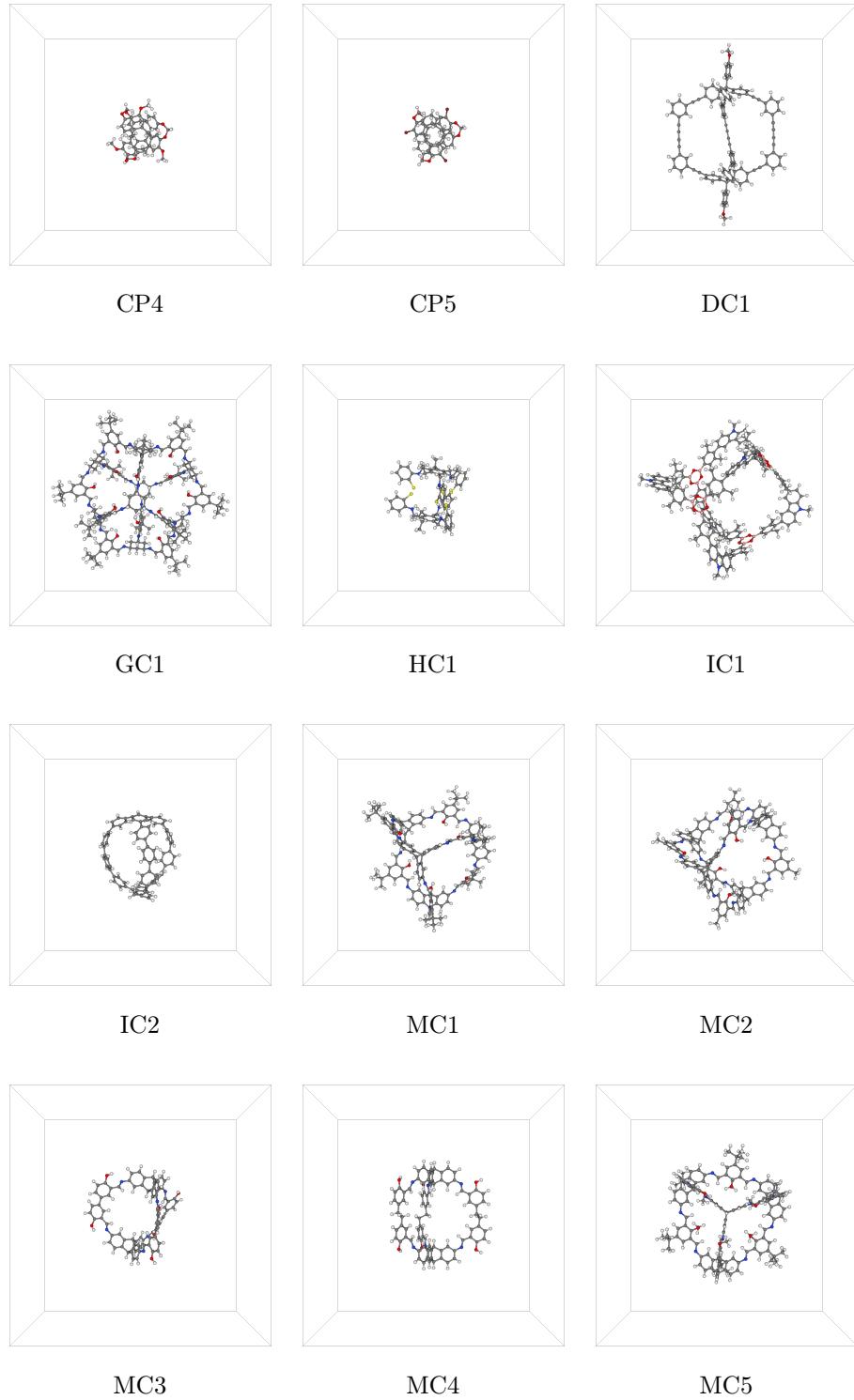


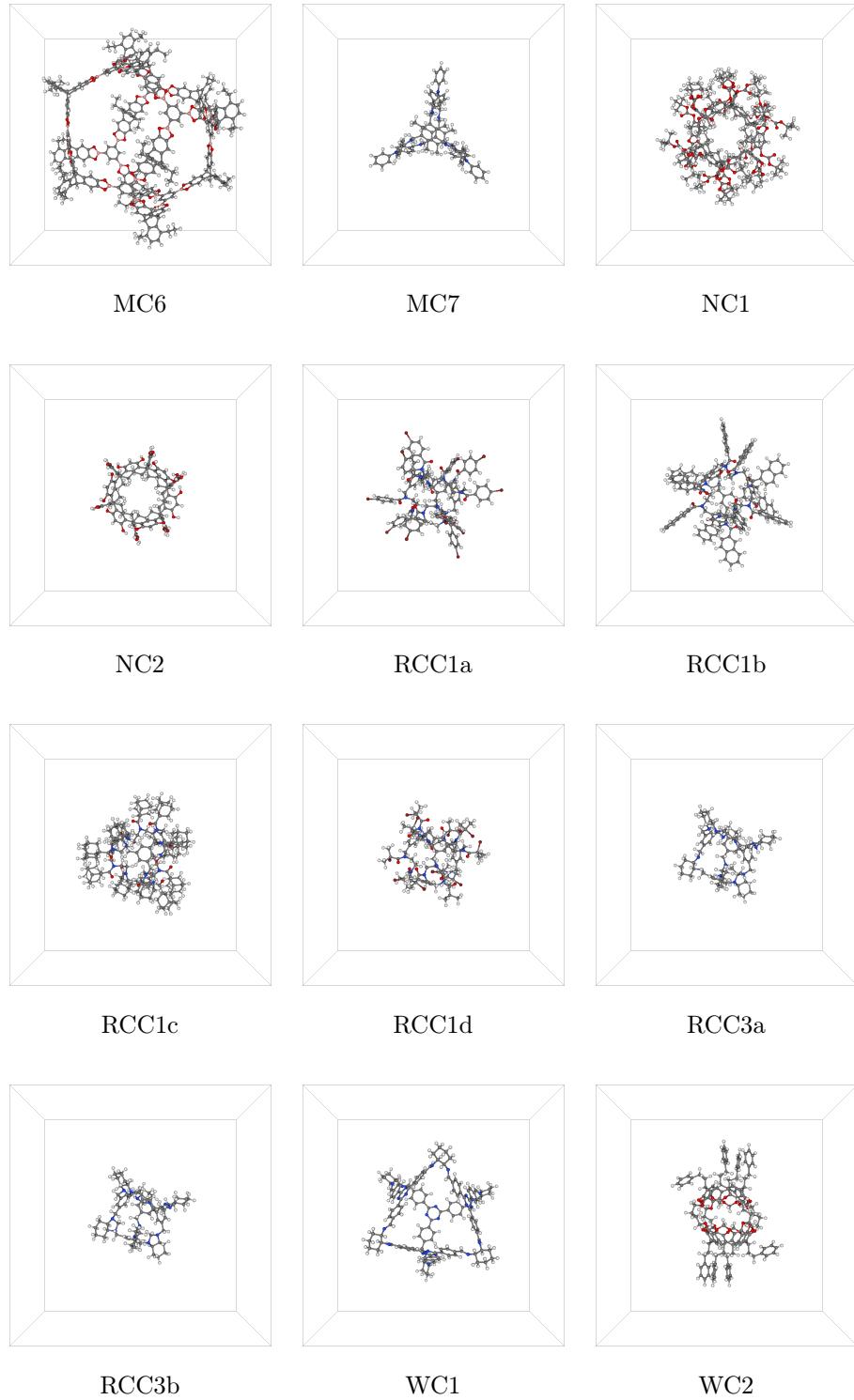
CB6



CB7







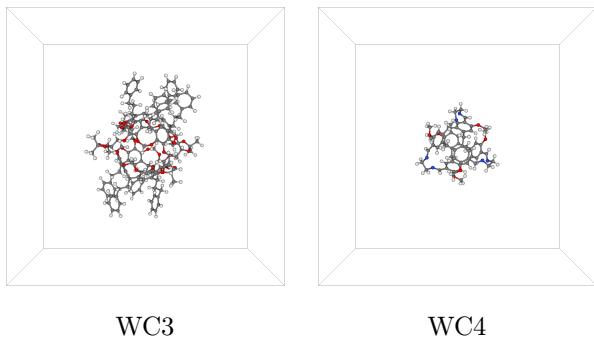


Figure S10.1: Visualizations of the structures of all 74 porous organic cage molecules analyzed in this study. The .xyz files of these cages are from Miklitz et al. and Greenaway et al.;^{9,10} see these references for the naming conventions. The cages are visualized in their centered and aligned states in the $[-20, 20]^3$ Å snapshot box in preparation for taking the 3D cage cavity image.

S11 Assessing effects of flexibility in cages

To evaluate how sensitive our methods are to flexibility, four cages (**CC2**, **CC3**, **CC4** and **CC5**) were allowed to fluctuate at 298 K via a molecular dynamics (MD) simulation. The simulations were performed using GULP v5.0¹¹ in the NVT ensemble using the leapfrog Verlet algorithm with the Nose-Hoover thermostat following an optimization process. The cage-specific force field (CSFF) was used to model the bond, angle and torsion potentials¹² where missing parameters were obtained from the polymer consistent force field (PCFF). The CSFF was specifically developed to properly describe organic porous cages such as the ones we are working with in this paper. Both coulombic and Lennard Jones (12-6) potentials were used for interactions between atoms separated by 2 or more atoms. Partial charges were assigned using the parameters provided by the (PCFF)^{13 12}. Due to the nature of the porous organic cages, no periodic boundaries were used in these simulations. The equilibration time was 50 ps while the production time was 1 ns with a timestep of 0.5 fs.

Aligned cages were used for input coordinates rather than going through the process of re-aligning them, due to the excessive computational time required. During the MD simulation 400 snapshots of each fluctuating cage were gathered and analyzed. To determine the validity of the simulations, the window diameter was calculated for each snapshot in **CC3** using `pywindow` (Fig S11.1). Holden et al. did a similar analysis of **CC3** window diameter during a MD simulation with the before-mentioned CSFF force field, and our result seem to be in good agreement with theirs.¹²

For us to visualize how the fluctuating cages compare in latent space, we need a dimension reduction algorithm that learns a mapping function, which t-SNE does not do. If t-SNE was used, we would have to retrain it with the inclusion of the fluctuating cages, which would result in overfitting to the large amount of fluctuating cages. A simpler choice is mapping the cages onto the two main principal components using principal component analysis (PCA). Principal components are linearly independent and lie along the axes that contribute the most towards the variance. PCA and singular value decomposition (SVD) are closely related

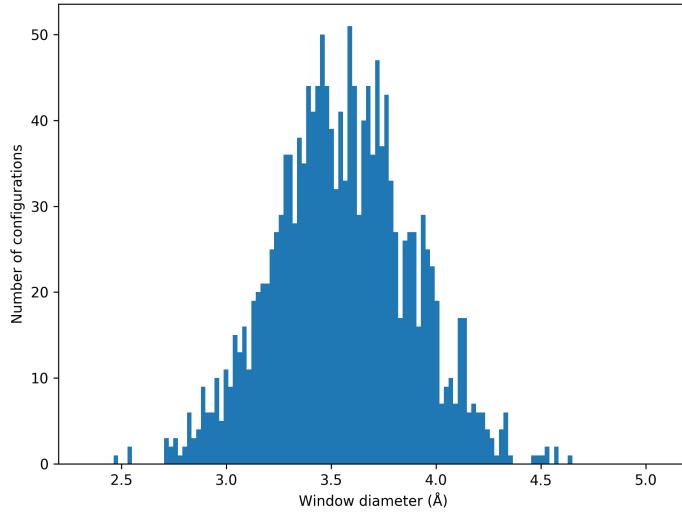


Figure S11.1: The window diameter was calculated using `pywindow` for **CC3** after the molecular dynamic simulation described above. All four windows of **CC3** are included.

in that we can use the SVD decomposition to map the data onto the two main principal components. By forming and arranging the data from the fluctuating cages the same way as the original cages and subtracting the mean of the original dataset we can use \mathbf{V}_ν to map the new matrix $\hat{\mathbf{A}}$, containing the information about the fluctuating cages, to a 2-dimensional space \mathbf{Y} by choosing $\nu = 2$, where \mathbf{V}_ν is the orthonormal vector, containing the eigencages that we formed using the original data.

$$\hat{\mathbf{A}}\mathbf{V}_2 = \mathbf{Y} = \mathbf{U}_2\Sigma_2 \quad (\text{S3})$$

Figure S11.2 shows the fluctuating snapshots in addition to the original dataset after being mapped to the main 2 principal components. For all four cages, after fluctuating they move away from the original location and occupy a latent basin. This is probably due to the optimization process which happened prior to the NVT MD simulation. In the optimization the static structure is relaxed, such that the potential energy of the cage reaches a (local) minimum. Regardless, the figure does provide valuable insight, showing that each cage is able to fluctuate due to thermal fluctuations or induced stress. Instead of thinking about

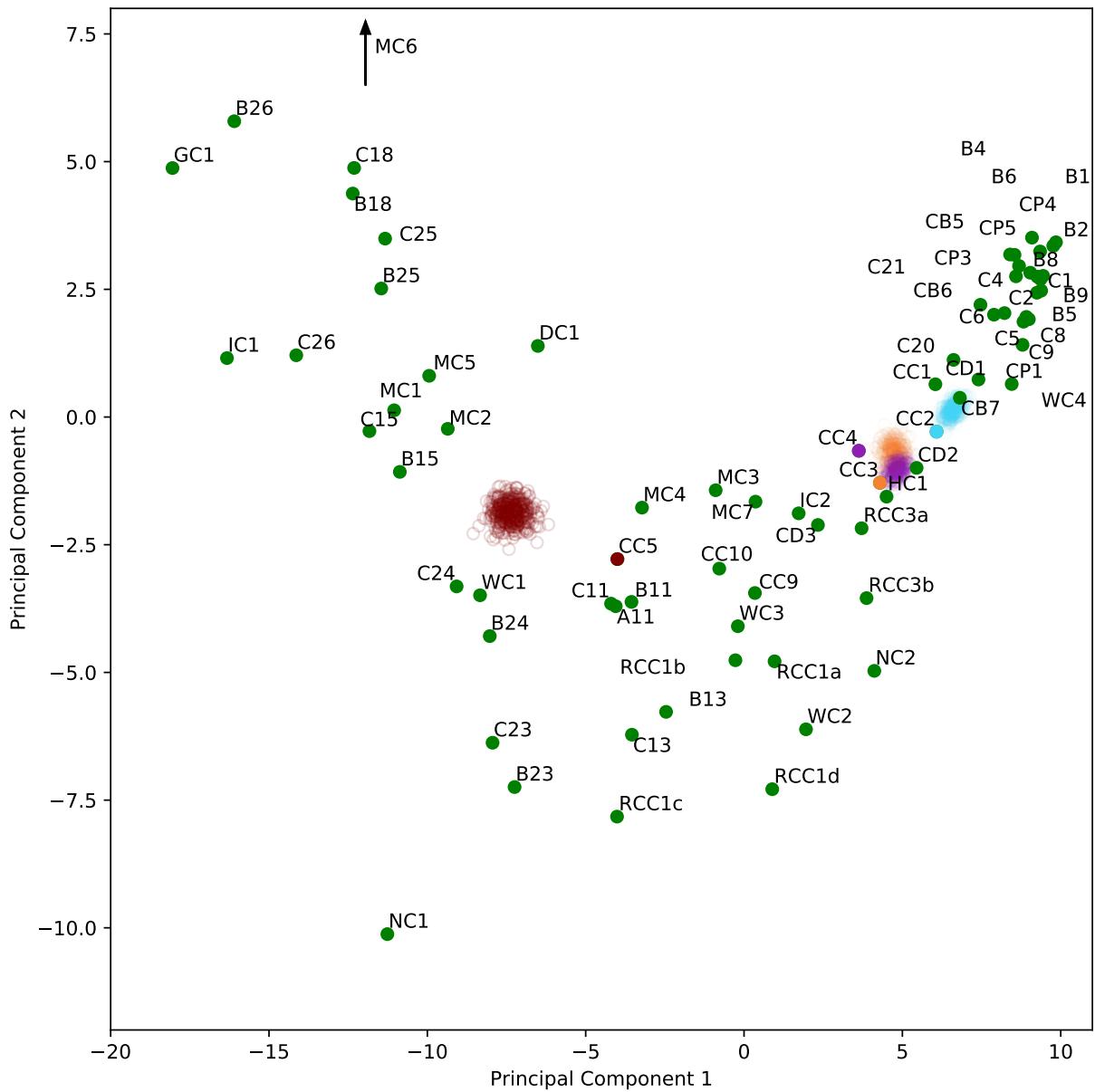


Figure S11.2: The cages mapped to the two main principal components. **CC2** (cyan), **CC3** (orange), **CC4** (purple) and **CC5** (maroon) were allowed to fluctuate in a molecular dynamic simulation. Each fluctuating cage resides in a latent basin, where the size is determined by the different geometrical configurations it's able to sample. **MC6** has been left out of the plot for visualization reasons.

each cage as a point in latent space, it moves around in a basin that follows a distribution determined by its' geometry. Out of the four selected cages, **CC5** is the largest and most flexible, leading to more configurations being explored. This seems to lead to the cage occupying a larger basin in the two-dimensional representation. Comparing the results to **CC2**, **CC3** and **CC4**, they occupy a much smaller basin due to the smaller size and more rigid structure.

S12 Exploring accessibility of cages

To assess if the porous organic cages were able to contain a Xe atom, a grid was constructed such that the distance between each adjacent datapoint was 0.1 Å. The potential energy was calculated at each point using a 12-6 Lennard Jones potential. Force field parameters were taken from the Universal Force Field (UFF) and geometric mixing rules were applied. When the potential energy grid was ready, we assigned an integer value to each point, -1 if the energy was above a certain energy value (here $15 k_B T$ where k_B is the Boltzmann constant and the temperature was taken to be $T = 298$ K), which corresponds to a non-occupiable point, and 1 if the energy was below that value. Next we see what occupiable points are connected and relabel the values in the grid to correspond to the newly-formed segments. If the center of the cage is non-occupiable or belongs to a different segment than the area outside of the cage, we deem that cage inaccessible. Otherwise we claim that a Xe atom can be adsorbed within the cage, given that it can overcome an energy barrier of $15 k_B T$.

This method was applied to a static image of all the porous organic cages, not taking into consideration that the pore can "breath", where the window diameter can fluctuate significantly.⁹ Our results can be seen in Table S1, where the inaccessible cages can be seen with their maximum window diameter (calculated with `pywindow`). For these cages, the adsorption is mainly taking place outside of the pore cavity. Miklitz et al.⁹ conducted a similar analysis where they used molecular dynamics to observe the pore breathing and they calculated a pore-limiting envelope (PLE). While our work focused on a static image of each cage, their calculation also looked at a series of images generated during a MD simulation. Even so, the static analyses are in good agreement. Most of the cages listed as having a void diameter too small or that could barely fit Xe were found in our analysis. A notable exception was **RCC3a**, which has the smallest void diameter according to Miklitz et al. The openness of **RCC3a** is most likely the reason it is not listed as inaccessible in our work. Our analysis has more differences to the dynamic analysis performed by Miklitz et al. Their dynamic PLE analysis shows that a subset of the cages are open to Xe for some percentage

of time. Noteworthy differences in our analysis and theirs is **RCC3b** and **CB6**, which are open to Xe 1.1% and 4.5% of the time, respectively. Other cages such as **CC3**, **CC2** and **NC2** were open to Xe roughly a tenth of the time. Here we are restricted to the static cages. Like previously mentioned, we assumed a Xe atom could overcome a $15 k_{\text{B}}T$ energy barrier, which might be a reason for this discrepancy.

Table S1: A list of the inaccessible cages alongside their maximum window diameter (calculated with `pywindow`)

| Cage | Maximum window diameter [Å] |
|-------|-----------------------------|
| CB5 | 2.334 |
| CP1 | 2.392 |
| CP3 | 2.679 |
| CP4 | 1.495 |
| CP5 | 1.474 |
| HC1 | 3.594 |
| RCC1a | 2.016 |
| RCC1b | 2.180 |
| RCC1c | 2.786 |
| RCC1d | 2.578 |
| WC2 | 2.682 |
| WC3 | 1.578 |
| WC4 | 3.281 |

Figuring out where the adsorption takes place in the accessible cages is another issue. A grid was overlaid on the cages and the energy calculated like previously. Next the local minima was found in the resulting energy matrix by sampling each element of the grid and ensuring no lower values were adjacent to it. Positive local minima were discarded. The void radius was calculated by taking the 2-norm of the vector between the center of mass of the cage and the closest atom and correcting it with the van der Waals radius of the closest atom. On figure S12.1 the energy corresponding to each local minima is plotted against the distance from the center of mass of each cage.

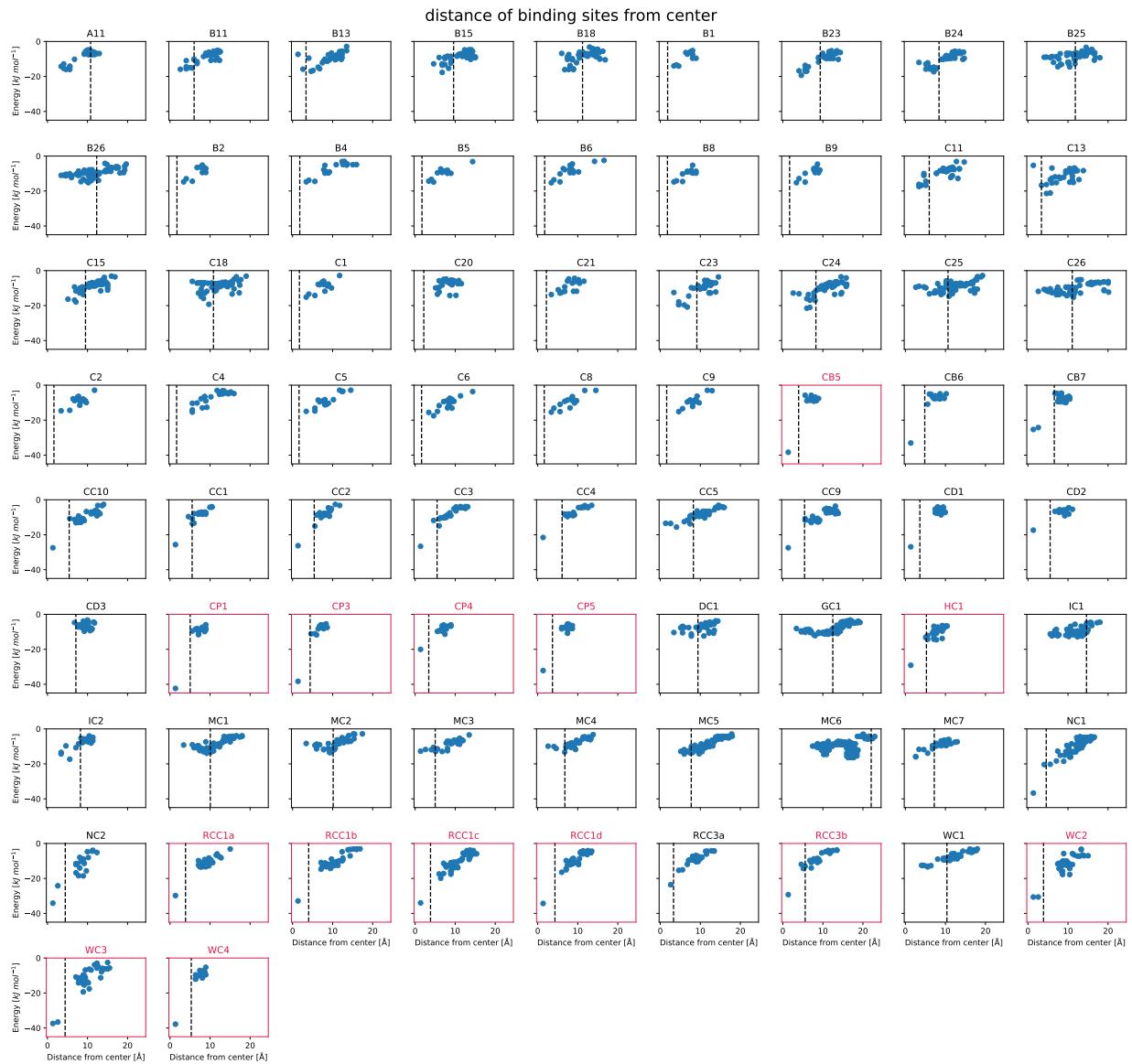


Figure S12.1: Energy of each local minima [kJ mol^{-1}] plotted against the distance from the center of mass of each cage [\AA]. The red frames correspond to the cages deemed inaccessible by our methods. The void diameter is plotted as a dotted line to give a better idea what's inside and outside of the cage.

S13 Safety statement

As this research is all computational in nature, no unexpected or unusually high safety hazards were encountered.

References

- (1) Miklitz, M.; Jelfs, K. *pywindow: Automated Structural Analysis of Molecular Pores.* *ChemRxiv* **2018**,
- (2) Maaten, L. v. d.; Hinton, G. Visualizing data using t-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579–2605.
- (3) Wattenberg, M.; Viégas, F.; Johnson, I. How to Use t-SNE Effectively. *Distill* **2016**,
- (4) Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard III, W. A.; Skiff, W. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.
- (5) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Elsevier, 2001; Vol. 1.
- (6) PorousMaterials.jl. 2018; <http://dx.doi.org/10.5281/zenodo.1400839>.
- (7) Patil, R. S.; Banerjee, D.; Simon, C. M.; Atwood, J. L.; Thallapally, P. K. Noria: A Highly Xe-Selective Nanoporous Organic Solid. *Chemistry—A European Journal* **2016**, *22*, 12618–12623.
- (8) Chen, L. et al. Separation of rare gases and chiral molecules by selective binding in porous organic cages. *Nature Materials* **2014**, *13*, 954.
- (9) Miklitz, M.; Jiang, S.; Clowes, R.; Briggs, M. E.; Cooper, A. I.; Jelfs, K. E. Computational screening of porous organic molecules for xenon/krypton separation. *The Journal of Physical Chemistry C* **2017**, *121*, 15211–15222.
- (10) Greenaway, R.; Santolini, V.; Bennison, M.; Alston, B.; Pugh, C.; Little, M.; Miklitz, M.; Eden-Rump, E.; Clowes, R.; Shakil, A.; Cuthbertson, H.; Armstrong, H.;

- Briggs, M.; Jelfs, K.; Cooper, A. High-throughput discovery of organic cages and catenanes using computational screening fused with robotic synthesis. *Nature Communications* **2018**, *9*, 2849.
- (11) Gale, J. D.; Rohl, A. L. The General Utility Lattice Program (GULP). *Molecular Simulation* **2003**, *29*, 291–341.
- (12) Holden, D.; Jelfs, K. E.; Cooper, A. I.; Trewin, A.; Willock, D. J. Bespoke Force Field for Simulating the Molecular Dynamics of Porous Organic Cages. *The Journal of Physical Chemistry C* **2012**, *116*, 16639–16651.
- (13) Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase ApplicationsOverview with Details on Alkane and Benzene Compounds. *The Journal of Physical Chemistry B* **1998**, *102*, 7338–7364.