Electronic supporting information

Predicting solvent effects on the structure of porous organic molecules

Valentina Santolini, Gareth A. Tribello, Kim E. Jelfs

This document includes:

1. Computational details
2. Results
3. Additional figures
4. References

1. Computational details

1.1. Example input file for PLUMED constrained MD simulations

Below is an example of the input file for the constrained MD simulations that were performed using DL_POLY2.20 and PLUMED2.0. First the centre of mass (COM) of the molecule is defined. The next section calculates the distances between the centre of mass and of the each atoms in the molecule and finds the minimum distance using the formula in the paper. The final section defines the restraint, which is applied on the minimum distance between the centre of mass and any atom. The restraint is gradually built up from 0 to 50000 kJ mol $^{-1}$ nm^{-2} over the first 20000 steps. The VERSE keyword ensures that the restraint is only applied to distances lower (L) than the position of the restraint. In the example given, the restraint is placed at 0.3 nm.

```
# This should be the COM of whole molecule
COM ATOMS=1-488 LABEL=com

# The difference between the largest and smallest distance
DISTANCES GROUPA=com GROUPB=1-488 MIN=BETA=10 LABEL=d1

MOVINGRESTRAINT ...
ARG=d1.min
STEP0=0 AT0=0.3 KAPPA0=0.0
STEP1=20000 AT1=0.3 KAPPA1=50000.0
VERSE=L
...  MOVINGRESTRAINT
```

1.2. Constrained MD calculations with a cylindrical probe

A cylindrical probe can be used instead of a spherical probe. In this case two positions are required in order to indicate the direction, $\mathbf{c}$, in which the long axis of the cylinder is oriented. In our work we used two centres of mass, $\mathbf{x}_a, \mathbf{x}_b$. The first of these was defined based on the position of three of the atoms atoms in the top of the molecule, while the second was defined by three atoms in the bottom of the molecule. The direction of the long axis of the cylinder is then:

$$\mathbf{c} = \frac{\mathbf{x}_a - \mathbf{x}_b}{|\mathbf{x}_a - \mathbf{x}_b|}$$

where the vertical bars in the denominator indicate that we are taking the norm of a vector. We can then calculate the distance, d, projected in the plane perpendicular to $\mathbf{c}$ between an atom at position $\mathbf{y}$ and the centre of the cylinder using the following cross product:

$$d = |(\mathbf{y} - \mathbf{x}_a) \times \mathbf{c}|$$

We then define the minimum distance projected in the plane perpendicular to $\mathbf{c}$ between the centre of the cylinder and all the atoms in the molecule by using a similar formula

to the one presented in our paper. Applying a restraint on this quantity corresponds to inflating a cylindrical probe inside the cage as opposed to a spherical one. The input file for doing this type of calculation with plumed is as follows:

```
# com1 is the COM of the top face of the cylinder, com2 of the other.  The
orientation is given by the vector from com1 to com2
COM ATOMS=3,25,39 LABEL=com1
COM ATOMS=49,69,81 LABEL=com2

# inplane distance between com1 and com2
d1:  INPLANEDISTANCES VECTORSTART=com1 VECTOREND=com2 GROUP=1-162 MIN=BETA=5

MOVINGRESTRAINT ...
ARG=d1.min
STEP0=0 AT0=0.5 KAPPA0=0.0
STEP1=20000 AT1=0.5 KAPPA1=50000.0
LABEL=m1
VERSE=L
...  MOVINGRESTRAINT
```

1.3. Energy minimisation calculations

We used MacroModel (version 9.6, Schrdinger, LLC, New York, NY, 2013) for geometry optimisations. A conjugate gradient minimisation algorithm was used with a convergence criterion of a remaining root mean square (RMS) force of 0.05 kJ mol^{-1} $Å^{-1}$.

1.4. Conformer search calculations for the collapsed structures

For conformer searching in Macromodel, we have previously used the low-mode (LMOD) approach[1] for porous organic imine cages.[2] However, due to the size of our systems and the fact that they contain many cycles, these calculations are extremely slow. Instead, we performed 1 ns MD simulations in Macromodel that we sampled every 1 ps, with the sampled structures being geometry optimised as in Section 1.3. For a full conformer search of these typically large molecules, we would recommend a larger number of steps, but we believe our calculations give an estimate of the energy difference to the higher energy open structures for each of the systems that serves our purposes. More extensive searches would only reveal a larger energy gap (as lower energy collapsed structures could be found).

1.5. Determination of spherical void size

In order to estimate the internal void size of conformations, we use a PERL script to determine the largest sphere that could fit within the centre of the molecule. First, the centre of mass of each molecule was calculated, and then the shortest distance between this centre of mass and any atom of the molecule was determined (equivalent to how the

constraint is applied for the MD simulations). This distance minus the van der Waals radius of the atom (C = 1.70 Å, H = 1.20 Å, and N = 1.55 Å) was then defined as being the radius of the largest sphere that can fit within the cage molecule. A molecule was then deemed to have an internal void (be 'open') if this radius was greater than a certain value, this varied across the different size cages as: 2.0 Å (**CC3** [8+12]), 2.5 Å (Mukherjee and Mastalerz cages), 3.5 Å (**CC8**) and 2.0 Å (**CC3** [6+9]). This is equivalent to the cage molecule being able to encapsulate small guests such as H_2, N_2, CO_2, Xe, and common organic solvents. Together with the OPLS energy, the void size was used to eliminate duplicate structures. Multiple duplicates being found for the low energy structures was also used to confirm thorough sampling of the low energy conformations.

1.6. Density functional theory (DFT) calculations of molecular orbitals

The molecular conformations of the solvated crystal structure for the Mukherjee cage and the lowest energy open conformation computed here were geometry optimised with Density Functional Theory (DFT). Gaussian09[3] was used with the B3LYP functional,[4] the 6-311++G* basis set and a convergence criterion of 10^{-8} Hartrees. This allowed us to compute the molecular orbitals for the conformations, shown in Figure S9.

1.7. Similarity index for comparison of molecular conformations

Comparison of molecular conformations was carried out using the Similarity Index function within Accelrys' Materials Studio v6.1, 2012 (BIOVIA). In all cases, we calculated the similarity index over all non-hydrogen atoms. The squared interatomic distance is calculated between all atom pairs in the molecules (regardless of distance), with equal weighting given to all atoms. A one-dimensional radial distribution function is then calculated and the difference between these then determines the similarity index, with the range of 0-100%, where 100% is a 'perfect' match.

1.8. Initial attempts to explore open molecular conformations

We first attempted to run MD simulations of a single molecule in the presence of a number of explicit solvent molecules, such as dichloromethane (a common solvent for these systems). We found no evidence of the porous molecule opening. In fact, if we started the MD from an open porous molecule, filled with solvent, the molecule collapsed so as to remove the void. This is despite the fact that we know the open structures to be local minima on the potential energy landscape.[5] Our next idea was to exploit the fact that the open forms of these cages are highly symmetric. We thus used steered MD[6] to drive the system towards configurations in which the faces of the cage were ordered around the porous molecule's centre of mass.[7] Remarkably we found at the end of these simulations that the molecule had adopted a configuration in which the faces were ordered, but that the structure was not open. Somehow the molecule always adopted a collapsed configuration in which some atoms occupied the pore region.

2. Results

System	non-H atoms	Collapsed	Inflated	Inflated void radius (Å)
CC3 [6+9]	128	187	353	2.6
CC3 [8+12]	168	220	320	4.1
CC8 [8+12]	272	395	508	4.6
Mastalerz [6+4]	170	1499	1657	5.2
Mukherjee [3+2]	93	794	860	3.6
CB6	72	2025	2129	2.8
CB10-ns	118	2977	3165	two cavities of 2.5
Cryptophane-A	66	399	854	2.46

Table S1: Comparison of the relative energy of the lowest energy collapsed and open structures for each system, calculated using OPLS and given in kJ mol^{-1}. The cryptophane-A collapsed conformation is taken from the SCXRD structure.

3. Additional figures

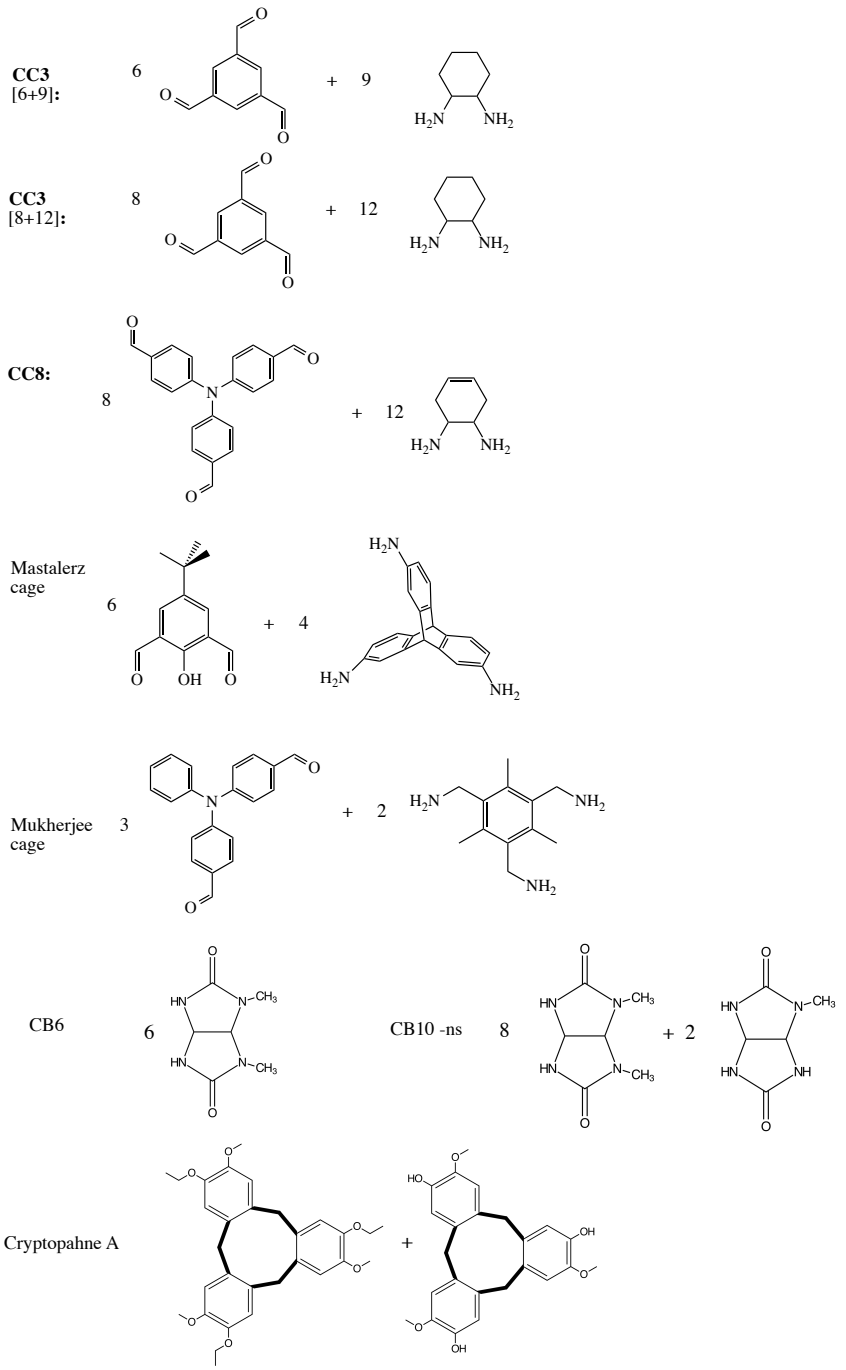

Figure S1: The synthesis of each of the porous organic molecules discussed here. For the Mukherjee cage, the reaction is followed by a reduction to convert the imines to amines.

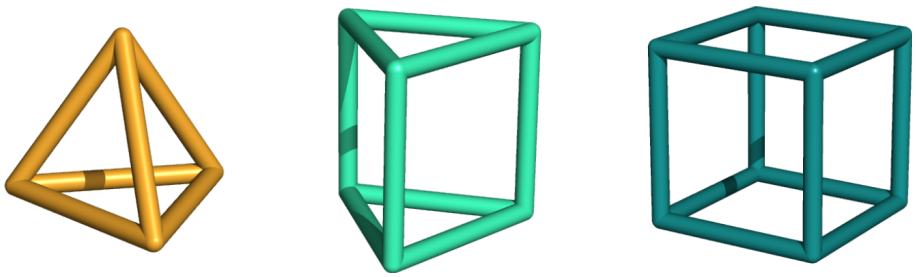

Figure S2: The underlying topology of the different molecular cage systems studied here; (left) tetrahedron - Mastalerz cage; (centre) triangular prism - **CC3** [6+9] cage and (right) cube - **CC8** and **CC3** [8+12].

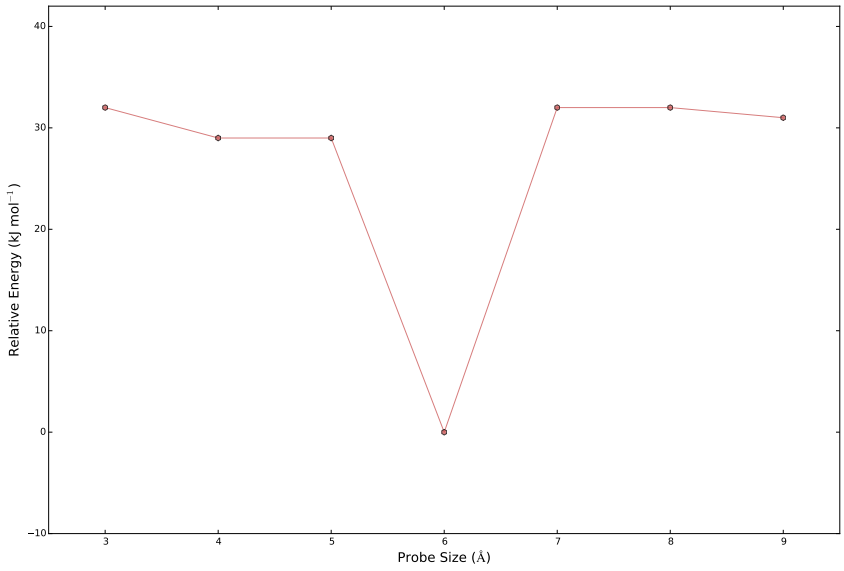

Figure S3: Plot of the lowest energy conformation against constraint size for **CC8**.

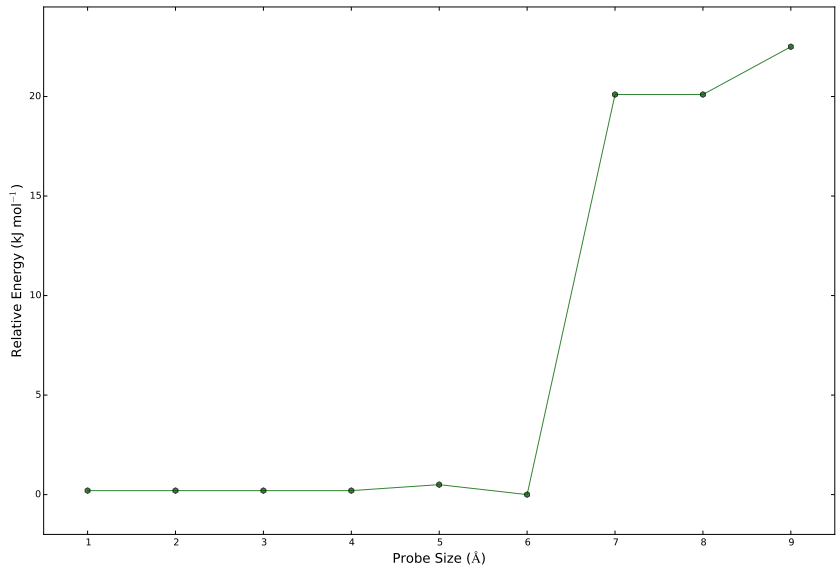

Figure S4: Plot of the lowest energy conformation against constraint size for the Mastalerz cage.

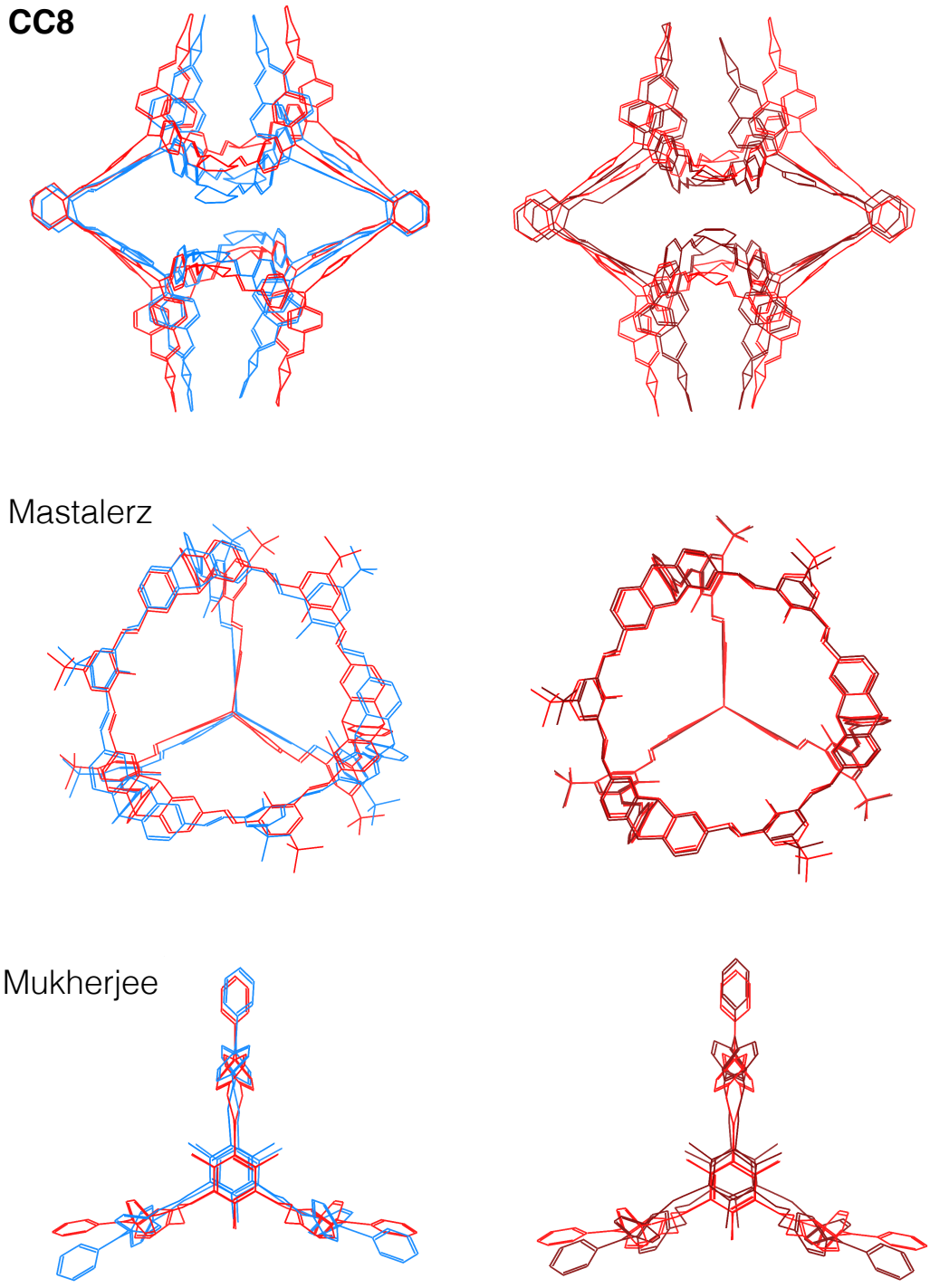

Figure S5: Overlays of **CC8**, Mastalerz and Mukherjee cages, (left) the computed open structure (blue) and the crystal structure (red) and (right) a comparison of the crystal structure (red) and the OPLS-minimised crystal structure (dark red).

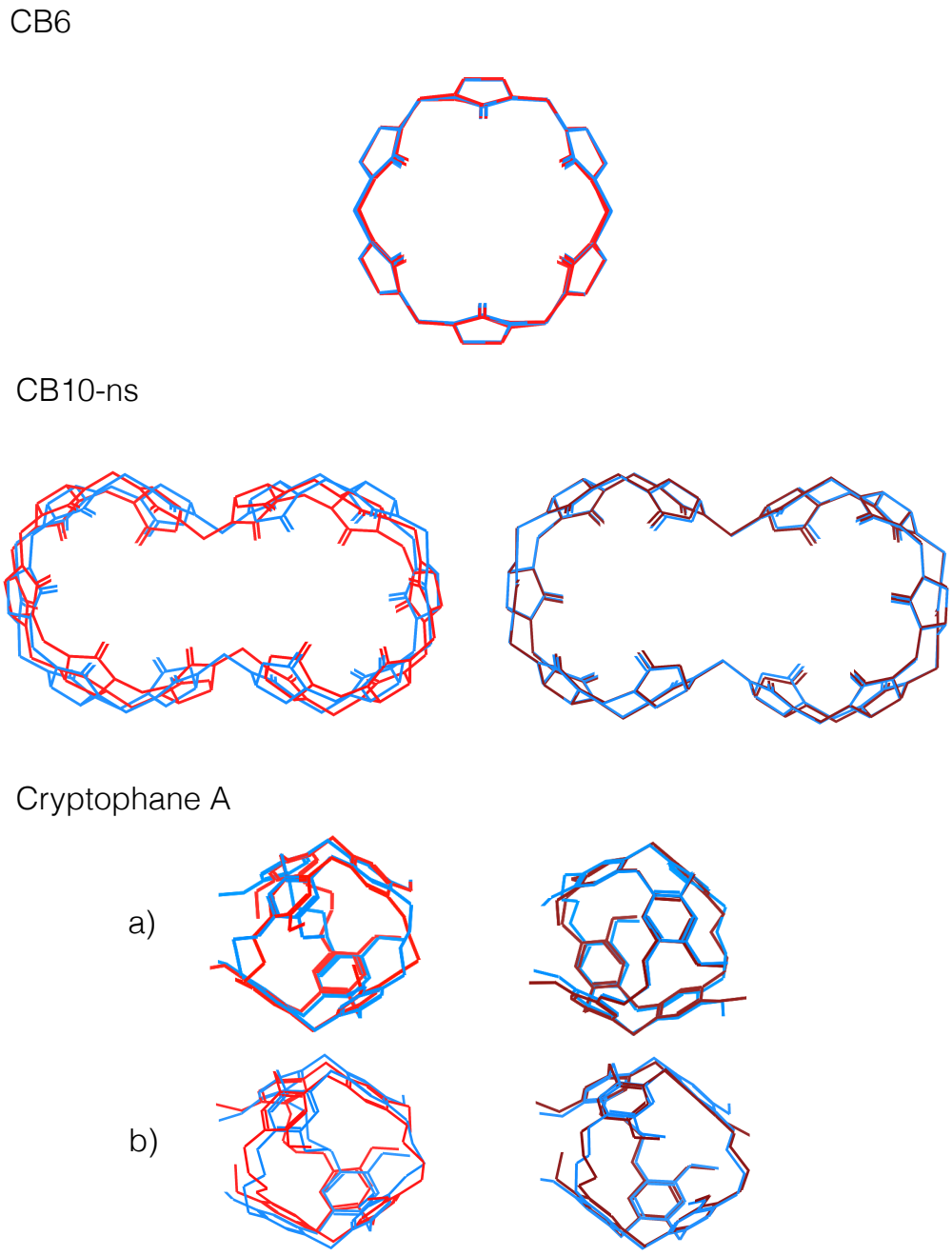

Figure S6: Overlays of CB6 computed open structure (blue) and the crystal structure (red) on top, CB10-ns and cryptophane-A, (left) the computed open structure (blue) and the crystal structure (red) and (right) a comparison of the computed open structure (blue) and the OPLS-minimised crystal structure (dark red). In the cryptophane section two computed structures, inflated with probes of different size, are superposed to a) the crystal structure loaded with Xe, and b) the crystal structure loaded with water.

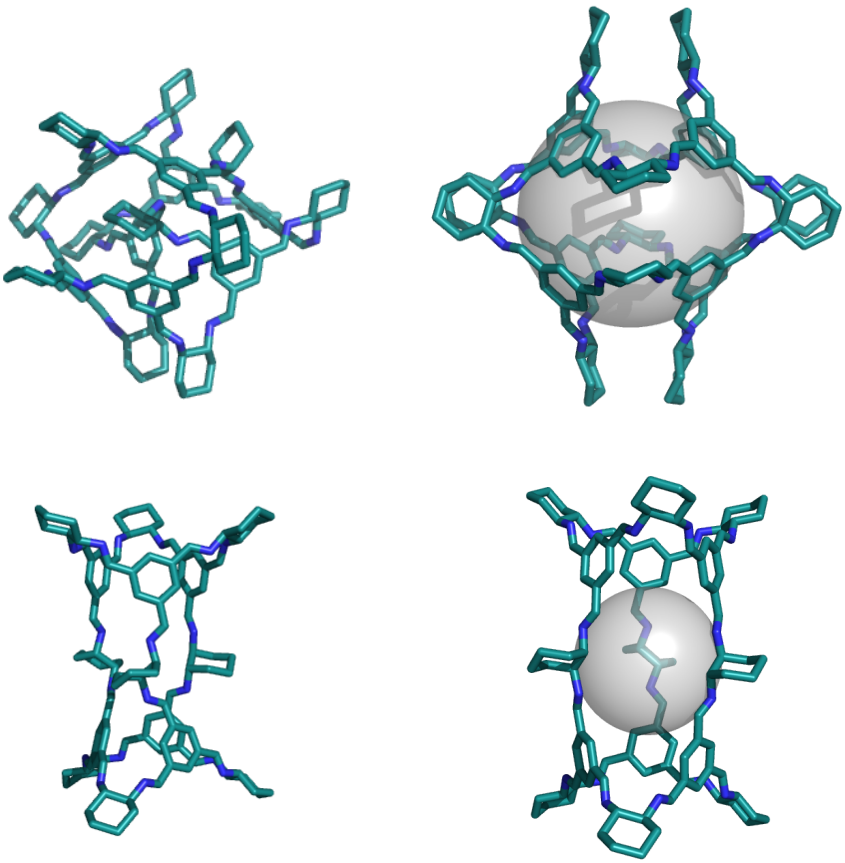

Figure S7: Collapsed (left) and inflated (right) structures of **CC3** [8+12] (top) and [6+9] (bottom).

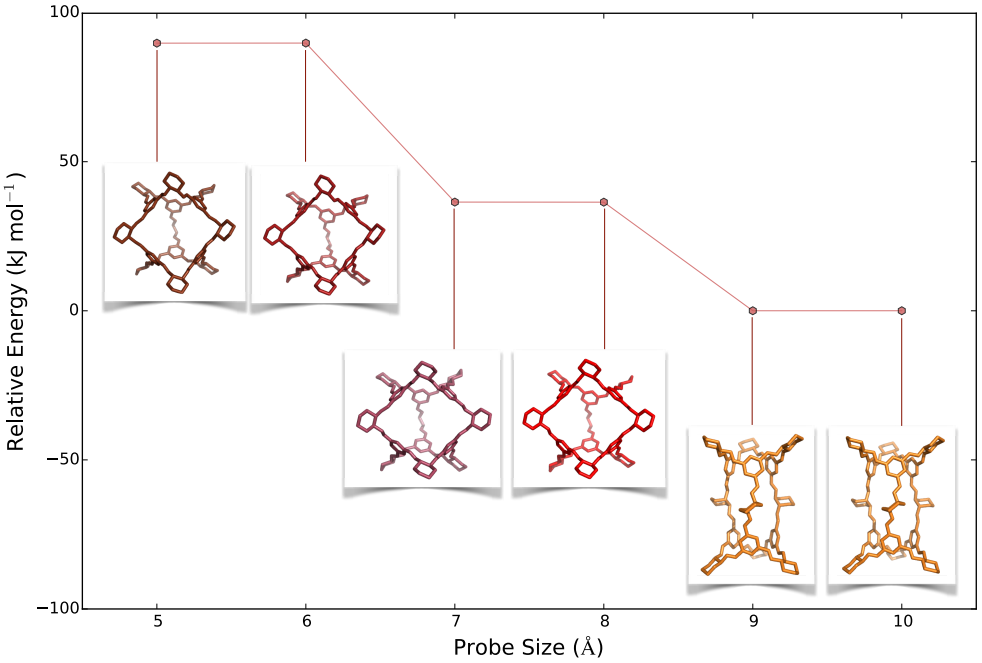

Figure S8: Plot of the lowest energy conformation against constraint size for **CC3** [6+9] for 300 K simulations.

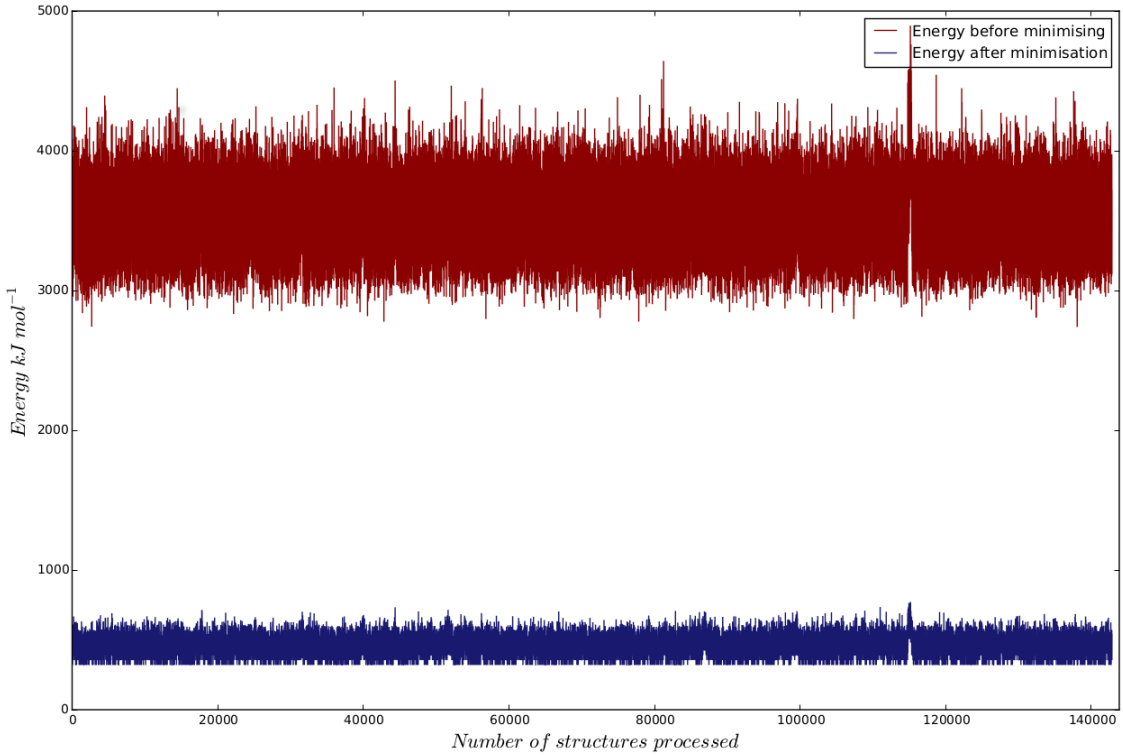

Figure S9: The OPLS energies of sampled **CC3** [8+12] structures from a constrained MD simulation at 300 K with a constraint radius of 5 Å. Points in red are the energies as sampled and points in blue are after geometry optimisation. The average decrease in energy upon geometry optimisation is 3066 kJ mol^{-1}.

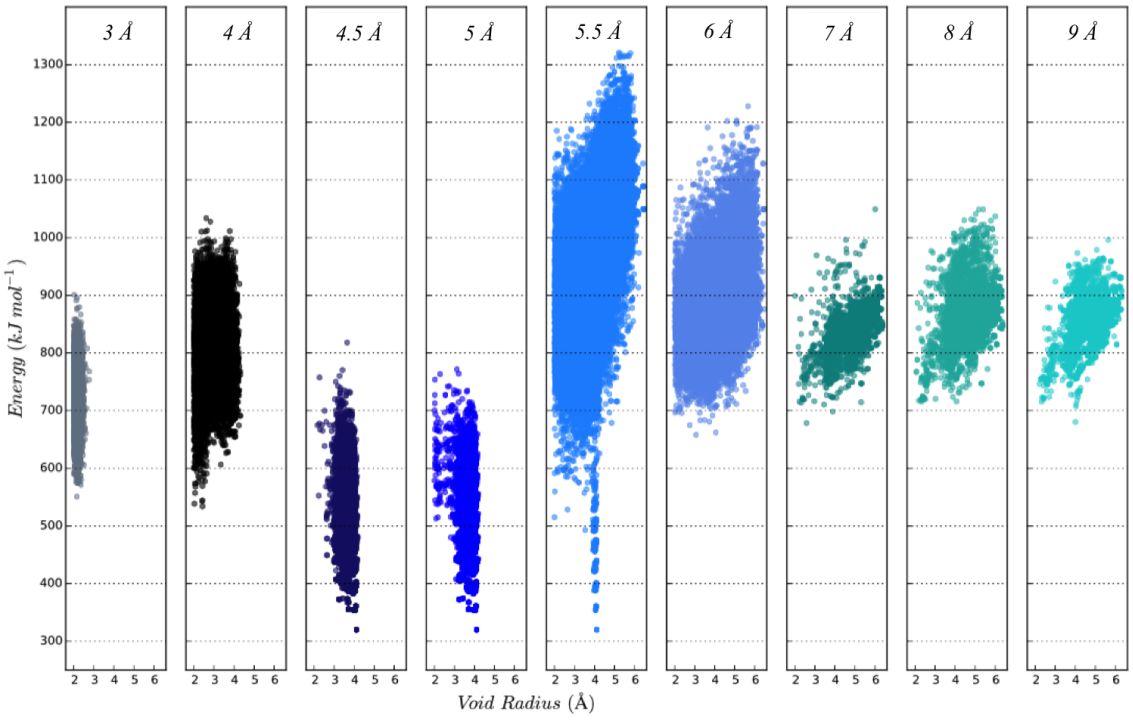

Figure S10: The OPLS geometry optimised energies and final void sizes of **CC3** [8+12] structures from constrained MD simulations at 300 K with a range of different constraint radii (as labelled at top). Each sub-plot shows a separate simulation run with a different constraint radii. Note that for the void radius plotted, the van der Waals radius of the nearest atom is subtracted (C = 1.70 Å, H = 1.20 Å, and N = 1.55 Å), thus these values are typically smaller than the constraint sizes applied.

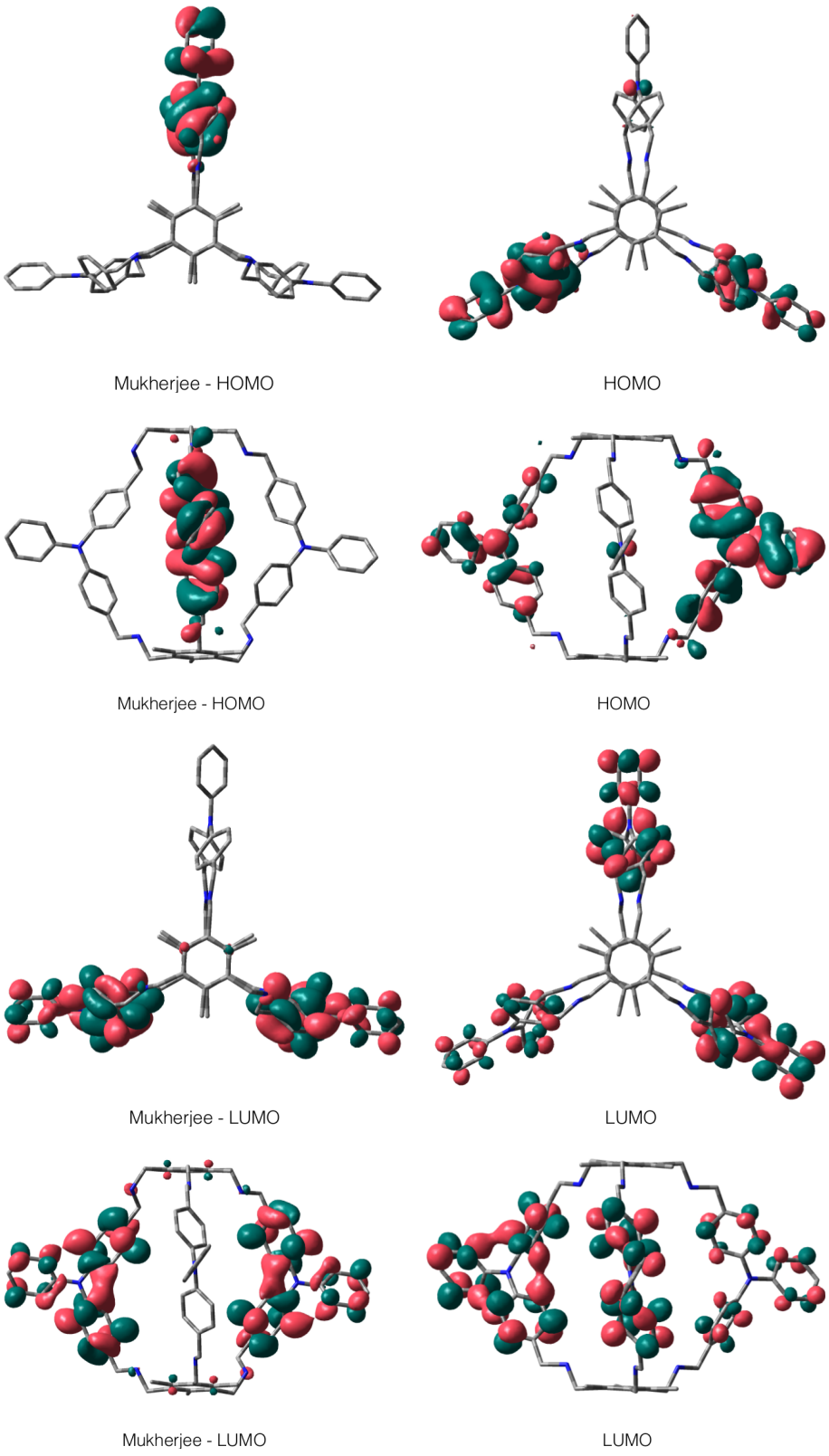

Figure S11: Computed molecular orbitals for the Mukherjee cage, from the solvate crystal structure (left) and the low energy computed structure (right). The top two rows show the HOMO from two perspectives, and the bottom two rows show the LUMO from two perspectives.

4. References

[1] I. Kolossvary and W. C. Guida, *J. Am. Chem. Soc.*, 1996, **118**, 5011–5019.

[2] K. E. Jelfs, E. Eden, J. L. Culshaw, S. Shakespeare, E. O. Pyzer-Knapp, H. P. G. Thompson, J. Bacsa, G. M. Day, D. J. Adams and A. I. Cooper, *J. Am. Chem. Soc.*, 2013, **135**, 9307.

[3] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, . Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian09 Revision D.01*, Gaussian Inc. Wallingford CT 2009.

[4] A. Becke, *Journal of Chemical Physics*, 1993, **98**, 5648–5652.

[5] K. E. Jelfs, X. Wu, M. Schmidtmann, J. T. A. Jones, J. E. Warren, D. J. Adams and A. I. Cooper, *Angew. Chem. Int. Ed.*, 2011, **50**, 10653.

[6] H. Grubmuller, B. Heymann and P. Tavan, *Science*, 1996, **271**, 997.

[7] P. J. Steinhardt, D. R. Nelson and M. Ronchetti, *Phys. Rev. B*, 1983, **28**, 784.