Electronic Supporting Information

Topological design of porous organic molecules

Valentina Santolini, Marcin Miklitz, Enrico Berardo, Kim E. Jelfs

Table of contents

- 1. Computational Details
- 2. Comparison between computed and crystal structures.
- 3. Additional figures.
- 4. References.

1 Computational Details

1.1 Conformer searching

All the conformer search calculations were performed with the MacroModel package (version 10.3.015, Schrödinger, LLC, New York, NY, 2015-3) and the OPLS3 force field. The PES was sampled with the low mode search method, which is based on eigenvector following. Throughout, the number of steps (starting from a minimum of 20000) was increased until convergence was reached, *i.e.* the number of low energy conformers found was constant. The energy window for saved conformers was set to 25 kJ mol⁻¹ and the minimum and maximum moves as 3 Å and 10 Å, respectively. For the larger systems (all except the (Tritopic + Ditopic) family), a high temperature Molecular Dynamics (MD) was run instead, as a more effective method to identify the lowest energy conformations. MD simulations were run with MacroModel, with the NVE ensemble, a time step of 1 fs, a total duration of 100 ns and sampling and geometry optimising a structure every 10 ps. This procedure was carried out at both 500 and 1000 K to ensure sampling of the PES. When the simulations reached convergence, we analysed the subset of conformers found, taking into account the structures that lay within an energy window of 25 kJ mol⁻¹.

1.2 Inflation

When the lowest energy conformations of a cage lacked shape persistence (*i.e.* were collapsed without an internal void), we used our previously developed constrained MD approach to simulate the effect of solvent and search for metastable open conformations.¹ These simulations use the software PLUMED2² in combination with DL_POLY2.20³ and the all-atom OPLS forcefield.⁴ We were unable to use OPLS3 for these simulations as it is not available outside of commercial software, however all sampled structures were geometry optimised with OPLS3. The full details of these simulations are given in our earlier work,¹ but to summarise, simulations are carried out with a series of constraints of increasing size. The tens of thousands of sampled structures are geometry optimised and the lowest energy open conformations retained.

1.3 DFT methods

To get a more accurate energetic ranking of the conformations, we first DFT optimised the structures with the PBE functional⁵, the triple-zeta molecularly optimised basis sets⁶, the GTH-type pseudopotential,⁷ a plane wave grid cutoff of 350 Ry and the Grimme-D3 dispersion correction in CP2K.⁸ Secondly, we performed single point calculations on the PBE-optimised structures using the M06-2X functional⁹ and 6-311G* basis set in NWChem.¹⁰ The M06-2X functional was successfully used in the past on porous organic cages^{11,12}, and does not require additional dispersion corrections, as these are already embedded in the functional. All energies given in the main text are from the M06-2X/6-311G* calculations.

1.4 Comparison of relative energies

All the molecules discussed above are formed according to multiples of stoichiometric [m+n] ratios. When calculating the relative energies, the energy of one molecule is arbitrarily taken as a reference and all the others are calculated with respect to that molecule. Each molecular size is then divided by its formula unit so that values

are comparable. Thus, if we take the (Tritopic + Ditopic) family as an example: 4 cages are generated with increasing molecular sizes, [2+3], [4+6], [6+9], [8+12]. The energies of the latter three cages are divided respectively by 2, 3 and 4, so that their energies are comparable.

1.5 Relative formation energies

In a few later cases, we are comparing between cage molecules that do not contain multiples of a formula unit. In these cases, their relative formation energies are instead compared. All the cages generated for this work are based on the DCC imine formation reaction, where an aldehyde and an amine react to form an imine and a molecule of water. To calculate a formation energy per bond formed, the following equation was used:

$$E_{formation} = \frac{(E_{cage} + xE_{water}) - (mE_{aldehyde} + nE_{amine})}{xn}$$

where E_{cage} is the energy of the cage formed, E_{water} is the energy of the water produced in the reaction, $E_{aldehyde}$ is the energy of the aldehyde and E_{amine} is the energy of the imine. The number of aldehyde reactants is m, n is the number of amine precursors and xn is the number of imine bonds formed, equivalent to the number of water molecule produced in the reaction.

1.6 Void size calculation

We consider the cage void as the largest sphere that can be built from the centre of mass of the cage, without any of the atoms penetrating it. The void radius of the sphere is thus considered the distance between the centre of mass and the closest atom to it, taking into account the van der Waals radius of that atom.

1.7 Cage window size calculation

Window sizes were calculated for $\mathbf{Tet}_2^4\mathbf{Di}^8$, $\mathbf{Tet}^6\mathbf{Di}^{12}$ and $\mathbf{Tet}^8\mathbf{Di}^{16}$ molecules in order to see whether the size of the cage windows could be related to the different solvents used experimentally for their synthesis. The window sizes were calculated by the Python pyWINDOW package¹³ developed in our group, which works as follows:

- 1. Starting from the XYZ coordinates of a cage, the maximum diameter of the cage is defined as the distance between the two furthest atoms in the molecule obtained by calculating the euclidean distance matrix using the scikit-learn package.¹⁴ The calculated maximum diameter is then corrected for the appropriate van der Walls (vdW) radii.
- 2. A sphere, with the diameter equal to the maximum diameter of the cage and the centre of the sphere overlapping with the centre of mass of the cage is then defined. Using Vogel's method for a spiral distribution of points on a disc with the "golden angle" approach, adapted for a sphere using cylindrical coordinates¹⁵, a set of evenly distributed sampling points is projected on the

sphere's surface. The number of sampling points is dependent on the sampling sphere's surface area and is estimated according to this equation:

$$\log_{10}(sphere_surface_area) * 250 \tag{1}$$

The number 250 is chosen as it proved to give sufficient sampling.

- 3. For each sampling point distributed on the sphere, a vector connecting the centre of mass of the cage and the point is defined. The largest included sphere, which corresponds to the distance between a point on the vector and the closest atom from the cage molecule corrected by an appropriate vdW radius, is calculated. This procedure is repeated for points along the vector in 0.1 Å increments and results in a set of the largest included spheres along the vector path.
- 4. The vectors with a set of the largest included spheres that do not cross any atom of the molecule are clustered by using the density-based spatial clustering algorithm from the scikit-learn package. The number of distinct clusters determines the number of windows found in the cage molecule.
- 5. Next, for each cluster a single sampling vector with the set of the largest diameters for the largest included spheres along its path is chosen and a circular plane perpendicular to this vector is generated. The diameter of this plane is defined and it corresponds to the necking of a cage's window.
- 6. The point at which the sampling vector passes through this spherical plane of a window is a rough estimate of the window centre. Next, an optimisation of the centre of the window plane is performed using minimize feature of SciPy package.¹⁶ The XYZ coordinates of the window centre are used as variables and the diameter of the window is the output of a two step minimization function, resulting in more accurate XYZ coordinates for the centre of the window plane.
- 7. As a result the number of molecular windows in the molecule along with the window circular diameter are obtained.

The results obtained are listed in Table S1.

	N^o of triangu-	N^o of squared	Window Size Aver-
	lar windows	windows	age (Å)
	(TW)	(SW)	
${f Tet}_2^4{f Di}^8$	4	2	5.6 (TW), 1.8 (SW)
${f Tet}^{ar 6}{f Di}^{12}$	8	-	5.2 (TW)
$\mathbf{Tet}^{8}\mathbf{Di}^{16}$ - less inflated	8	2	4.4 (TW), 6.3 (SW)
$\mathbf{Tet}^{8}\mathbf{Di}^{16}$ - more inflated	8	2	4.5 (TW), 8.5 (SW)

Table S1: Information about the window sizes of $\mathbf{Tet}_2^4\mathbf{Di}^8$, $\mathbf{Tet}^6\mathbf{Di}^{12}$ and $\mathbf{Tet}^8\mathbf{Di}^{16}$.

2 Comparison between computed and crystal structures.



Figure S1: A) Overlay of computed (blue) and crystal structure (red) of Tri⁴Di⁶; B) Overlay of computed (blue) and crystal structure (red) of Tri⁴Tri⁴.



Figure S2: **CC8 - Tri⁸Di**¹², **A)** Overlay of computed (blue) and crystal structure (red), **B)** Overlay of computed (blue) and minimised crystal structures (dark red).



Figure S3: CC5 - Tri⁴Di⁶, A) Overlay of computed (blue) and crystal structure (red), B) Overlay of computed (blue) and minimised crystal structures (dark red).

3 Additional figures.



Figure S4: Two different conformers of $\mathbf{Tet}^{8}\mathbf{Di}^{16}$, the less inflated and lower in energy on the left, and the most inflated and higher in energy on the right.

References

- V. Santolini, G. A. Tribello and K. E. Jelfs, Chem. Commun. (Cambridge, U. K.), 2015, 51, 15542–15545.
- [2] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, Comput. Phys. Commun., 2014, 185, 604–613.
- [3] W. Smith, C. W. Yong and P. M. Rodger, Mol. Simul., 2002, 28, 385–471.
- [4] W. Jorgensen, D. Maxwell and J. TiradoRives, J. Am. Chem. Soc., 1996, 118, 11225–11236.
- [5] J. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett., 1996, 77, 3865–3868.
- [6] J. VandeVondele and J. Hutter, J. Chem. Phys., 2007, 127, 114105.

- [7] S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B*, 1996, **54**, 1703–1710.
- [8] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, 167, 103–128.
- [9] Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **120**, 215–241.
- [10] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus and W. A. De Jong, *Comput. Phys. Commun.*, 2010, **181**, 1477–1489.
- [11] A. Burgun, P. Valente, J. D. Evans, D. M. Huang, C. J. Sumby and C. J. Doonan, *Chem. Commun. (Cambridge, U. K.)*, 2016, **52**, 8850–8853.
- [12] M. Saleh, H. M. Lee, K. C. Kemp and K. S. Kim, ACS Appl. Mater. Interfaces, 2014, 6, 7325–7333.
- [13] M. Miklitz and K. E. Jelfs, *Manuscript in preparation*, 2016.
- [14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. of Machine Learning Research*, 2011, **12**, 2825–2830.
- [15] H. Vogel, Math. Biosci., 1979, 44, 179–189.
- [16] http://www.scipy.org/.