Research Article

# Out-of-oblivion Cage Molecules and their Porous Crystalline Phases

Ismael Gomez Garcia,[a,b] Marco Bernabei,[a] Raul Perez Soto,[a] Maciej Haranczyk[a,c]*

[a] IMDEA Materials Institute, C/Eric Kandel 2, 28906 - Getafe, Madrid, Spain

[b] Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain

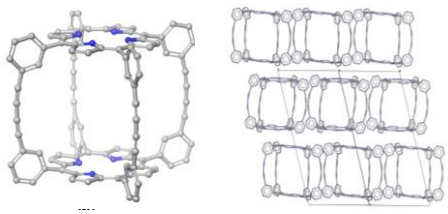
[c] Lawrence Berkeley National Laboratory, One Cyclotron Rd, Berkeley, CA 94720, USA

**Crystal Growth and Design, 2017, 17, 11, 5614–5619**

## Abstract

An automated molecular porosity detection approach was developed and applied to PubChem, a repository of ca. 94 million molecules, to discover intrinsically porous cage molecules, which, although previously considered by chemists, have remained in oblivion to the porous solids community as neither their crystal structures nor solid-state porosity have been previously reported. The effort led to identification of six such cage molecules reported over the span of the last two decades. The following crystal structure prediction effort suggests that these molecules can form stable low-energy porous crystalline phases. One of the identified lowest energy phases exhibit zeolite-range porosity with pore diameters of ca. 8Å and internal surface area of 1070 m2/g.

**Visual TOC**



Crystal structure prediction for six cage molecules, which are identified by a tailor-made screening approach, suggests that these molecules can form stable low-energy porous crystalline phases. One of the identified lowest energy phases exhibit zeolite-range porosity with pore diameters of ca. 8Å and internal surface area of 1070 m2/g.

## 1. Full work presentation

Materials composed of molecules exhibiting intrinsic porosity are rare. Examples of the latter include Noria[1] and triptycenetrisbenzimidazolone[2] as well as families of calixarines,[3] cucurbiturils,[4] cyclodextrins,[5] cryptophanes[6] and porous organic cages (POC).[7] Porous molecules can be classified as 'belts' or 'cages' depending whether they have, respectively, two or more windows on the perimeter of the internal pore. Unlike network polymers and framework materials, porous molecules can be synthesized first and then assembled into crystalline and amorphous solid state phases. Many properties of recently reported porous molecular materials are in par in those reported for framework materials such as metal organic frameworks. For example, Covalent Cage 3 (CC3) material exhibits high xenon/krypton selectivity[8] while a cage-based material reported by Zhang *et al.*[9] was reported to have the surface area of 3758 $m^2/g$. Porous molecular materials are subjects of intensive investigations, and a number of recent review articles summarizing these efforts are available.[10]

The introduction of a porous molecule to the porous solids community requires three steps: (i) the molecule has to be synthesized, (ii) the corresponding solid material (amorphous or crystalline) needs to be obtained, and (iii) its porosity needs to be confirmed and, preferably, further characterized by, for example, Brunauer-Emmett-Teller (BET) surface area measurement. In the case of recent rational synthesis efforts,[¡Error! Marcador no definido.,11] all three steps were completed, and these molecules can be easily identified by a routine literature search. A recent study by Milkitz *et al.*[11] relied on review articles to identify 41 intrinsically porous molecules, for which the corresponding crystal structures were available from the Cambridge Structure Database(CSD). In the case of molecules, for which only the first two steps were completed, their identification is more challenging. Although databases of experimental crystal structures are available, the crystals are often reported with solvents, which makes the structures non porous. Two studies of the

CSD[12,13] have either removed solvent or focused only on single-entity crystals, and used low density as a key to identify porous molecules. For example, the latter study of Evans and coworkers[¡Error! Marcador no definido.] has identified 481 porous molecular materials though the f raction of the number corresponding to intrinsically porous molecules have not been reported, i.e. the number of 481 also includes molecules with extrinsic porosity as well as ones that do not pack efficiently in the crystal structure.

In this study, we seek to identify porous molecules for which only step (i) has been completed. As the crystallographic information is not available in this case, these molecules scatter the scientific literature, patents and other databases in a form of structural formulas and the corresponding electronic formats. Fortunately, the scattered molecules are being collected into large repositories such as PubChem[14] or ChemSpider[15]. Our effort focuses on the larger of these two, i.e. PubChem (ca. 94 million compounds). The identification of the targeted molecules relies on a porosity detection approach, which is based on the limited information available in PubChem. Our search approach is therefore agnostic of the form, year and language in which the questioned porous molecules were originally introduced. By doing so, we can target molecules that could not be identified in previous searchers.

Our core algorithm for detection in intrinsic porosity requires 3D molecular model, e.g. Cartesian coordinates of atoms composing the molecule. It then identifies the void regions within the perimeter of the molecule, and characterizes them to determine if they correspond to intrinsic pores. The latter are defined on the basis of their exposure to the surroundings of the molecule. Our algorithm, described in detail in the Supporting Information (SI), can be highlighted as follows. First, a set of candidate voids for a given molecule is detected with the help of Voronoi tessellation, i.e. they are defined by Voronoi vertices. Then, for each of the vertices, we check whether it is an intrinsic void by calculating its Pore Exposure Ratio (PER, see Fig. 2-2). PER provides a measurement of the exposure of a Voronoi vertex is to the surroundings of the molecule, approximated by

a spherical perimeter, see Fig. 2-2. PER is defined as a ratio between the surface area of the largest fragment of the sphere exposed to the candidate void (Voronoi vertex), and the surface of the sphere. In practice, a spherical mesh is used to represent the sphere and the exposed regions are defined by a set of triangles formed within this mesh. By setting a PER threshold value, pore candidates can be identified as intrinsic, and in particular, cage-type. Visual analysis of a large set of molecules allowed setting the threshold empirically, i.e. PER below 0.36 identifies molecules with an internal pore.

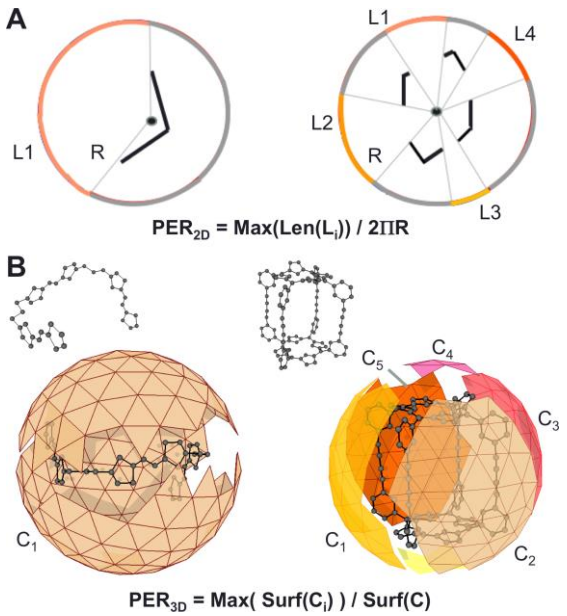



**Figure 2-1.** Intrinsic porosity detection with Pore Exposure Ratio. A) Method overview in 2D: A caricature of linear (left) and cage (right) molecules. A point (black dot) representing a void within the perimeter of the molecule (heavy black line) is partially exposed to the circle of the radius R surrounding the molecule. Colored regions in the circular boundary, L, correspond to openings that expose the internal void. Their size w.r.t the circular boundary, PER, represents how open is the space to access the point from the surrounding. Thus, too high PER values may indicate no actual cage is being observed (linear or close to linear objects). B) PER calculation in 3D: a linear (left) and cage (right) molecule. Spherical boundary mesh used in computing PER, i.e. PER is computed as the ratio of the surface of the biggest connected component, $C_x$, and the total surface area, S, being the sum of surface areas of all triangles forming the mesh. The illustrated linear molecule has PER of 0.81 while the cage molecule has a PER value of 0.15.

PubChem Compound Database (herein referred to as PubChem) stores 94 333 141 unique records of which only 80 828 441 have 3D coordinates available (referred to as PubChem3D).[16]

The coordinates of the molecules in PubChem3D were generated by the OpenEye Scientific's OMEGA tool using the MMFF94 force field.[17] We filtered PubChem database to discard: records with more than one molecule (mostly accompanied by solvents or salts), non-organic molecules, charged molecules, radicals, molecules without rings and lastly, molecules with more than 8 rotatable bonds in their ramifications ("flexible arms" increase degrees of freedom and render the corresponding crystal structure prediction computationally unfeasible (see the SI and references therein). For the remaining molecules that did not have a corresponding entry in Pubchem3D, we generated 3D structures using Ligprep tool from Schrodinger[18]. However, the latter step was executed only for those molecules with 48 or more atoms. The latter cutoff was based on the size of the smallest calixarene (i.e. calix[4]arene, 56 atoms) corrected for a 'safety margin' of 8 atoms. Technical details of the 3D generation with LigPrep are provided in the SI,

The resulting database of 1 258 975 molecules with 3D coordinates was characterized to predict their corresponding PER value. Out of that set, only 4 767 molecules were identified with PER below 0.36, corresponding to their cage-like nature. Upon further inspection, the set was further pruned to remove carbon nanostructures, e.g. buckyballs and buckybowls, belt-shaped molecules, cages with very small internal pores (i.e. diameters less than ca. 2.0Å) and molecules whose corresponding crystal structures were already deposited in the CSD. For the remaining structures, the search for the low energy conformers was carried out using Schrodinger's MacroModel suite with the OPLS_2005 force field[19] (see the SI for additional details). The latter search was performed in vacuum to exclude any solvation effects that could stabilize the internal pore. Finally, the lowest energy conformers were investigated again using the PER descriptor to confirm that they

were indeed porous. The final set of structures contained six entries, herein called M1 through M6, depicted in Fig 2-2. Although not in the scope of this work, we note that the above conformer search approach can be also performed in the presence of solvent, i.e. targeting molecular cages that are suitable for encapsulation and transport of molecular guests in liquid phases despite being non-porous in vacuum and (likely) desolvated crystalline phases.

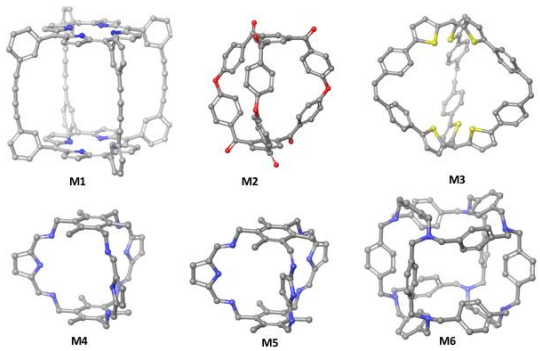



**Figure 2-2.** Porous organic molecules identified in this work. Gray, blue, red and yellow atoms represent carbon, nitrogen, oxygen and sulfur, respectively. Hydrogen atoms are omitted for clarity.

Molecules M1-M6 show different chemistries,[20] two of which were not highlighted in a recent review of porous molecules (M2, M6).[21] M1 belongs to the family of the porphyrin box cages. M2 includes ester and ketone functional groups in the structure. The cage M3 is a carbon-carbon cage that includes thiophene groups which clearly affect the shape of the cage forcing the sulfur atoms to point towards the center of the cage. M4 and M5 are, respectively, a symmetric and an asymmetric variation of the well known imine cages such as CC3, where the imine bond is one bond further from the aromatic ring. M6 is a tertiary amine cage.

In order to investigate whether M1-M6 can form crystalline porous materials, a crystal structure prediction study (CSP) was executed. The CSP was limited to 13 space groups (P21/c, P-1, P212121, P21, Pbca, C2/c, Pna21, Cc, Pca21, C2, P1, Pbcn, Pc) which

account for about 90% of the observed crystal structures for organic molecules. Within each group, 5000 random solvent-free configurations with one molecule in the asymmetric unit were generated and subsequently relaxed (see section SI for further details). The free energy of the crystal was approximated with a static lattice energy, defined as the sum of an intermolecular cohesive pair-wise potential energy, consisting of Lennard-Jones and Coulomb terms, and a combination of intramolecular energy terms which account for molecular flexibility. All calculations involved OPLS_2005 force field[i Error! Marcador no definido.] a nd were executed in the UPACK program.[22] The porosity of the crystal structure, i.e. pore limiting and largest cavity diameters (PLD and LCD, respectively) and the nitrogen surface area (SA) were characterized using Zeo++ program.[23,24] Nitrogen probe radius was set to 1.86 Å.

As mentioned above, a random search for possible solvent-free crystal structures and the subsequent minimization of the static lattice energy are at the core of the UPACK method for CSP calculations. CSP results are often presented as an energy landscape, i.e. a plot of the predicted lattice energy w.r.t the structure density for each predicted phase, which is a very useful crystal engineering tool. It does not, however, contain information on the phases' physical properties required to discover materials for a specific application. To address the latter, the energy-structure-function (ESF) maps[25] were recently introduced and successfully applied to guide the experimental discovery of porous molecular crystals with targeted properties.

If a large number of low energy distinct structures are found at the end of the CSP procedure, it might suggest the presence of a rich polymorphism, i.e. the existence of several crystal phases with different properties and/or porosity despite comparable values of the lattice energy. Recent calculations on over 1000 crystal structures of small organic molecules demonstrated that the free energy difference between pair of polymorphs seldom exceed the value of 7 kJ/mol.[26] However, molecules that can form host-guest

complexes and more generally crystallize in porous phases, show the tendency to pack less effectively giving rise to different polymorphs within and energy range of up to 50kJ/mol from the lowest predicted structure[27,¡Error! Marcador no definido.] and a relative energy difference t hat can exceed 20kJ/mol.[28] Moreover in some cases a reversible/irreversible switch between different polymorphs can be experimentally observed by exposure to different solvents.[29] Therefore, it is not surprising that experimental structures do not always correspond to the predicted lowest energy phase(s).

Our CSP approach was first verified by predicting the different solvent-driven polymorphs of cage molecule CC1 (see section in the SI for more exhaustive discussion). The non-porous polymorph CC1-α' and the $H_2$-porous polymorph CC1-β' were predicted approximately 7kJ/mol and 26 kJ/mol above the global minimum, respectively. A closer inspection of the energy-density plot (Fig. 2-S6) reveal that despite their energy difference both polymorphs lie in what is called the leading-edge of the landscape, defined as the set of structures with the lowest predicted energy at a certain density. Experimental accessible structures are indeed often identified in this region of the landscape.

We applied the sample CSP methodology to obtain possible experimentally accessible crystal phases for molecules M1-M6, and characterize their porosity. Fig. 2-3 shows the PLDs for the predicted structures for molecules M1-M6 within the energy range of 50kJ/mol w.r.t. the lowest predicted phase. The horizontal line at 2.32 Å represents the PLD of the experimental polymorph CC1-β of the imine cage CC1, which indeed is selectively porous to $H_2$. For molecule M1 we observe that 100% of the structure predicted in the 50kJ/mol range are at least porous to $H_2$. This result is mainly due to the rectangular box-shape of the molecule and the absence of phenyl, pyrrole or thiophene cyclic groups, present in the other molecules, which can partially reduce the size of the windows connecting intrinsic molecular pores and hence a decrease in the value of the corresponding PLD. The fraction of predicted structures in the 50kJ/mol range, with PLD

larger than 2.32 Å for molecules M2, M3, M4, M5 and M6 are, respectively, 2.6%, 16.2%, 41.3%, 9.4% and 9.4%.

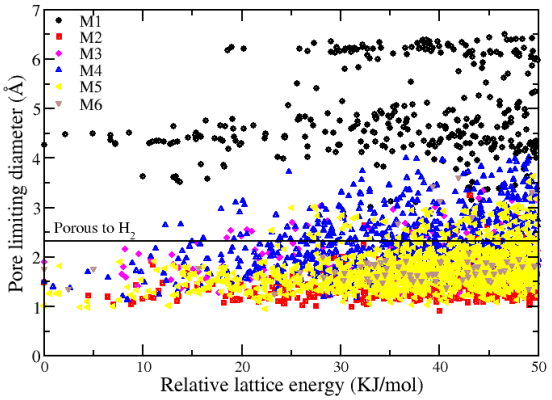



**Figure 2-3.** Pore limiting diameter (PLD) as a function of the relative lattice energy w.r.t. the corresponding lowest energy structures for molecules M1-M6. The horizontal line at 2.32 Å corresponds to the PLD of cage molecule CC1.

The lowest energy predicted structures with PLD larger than 2.32 are depicted in Figure 2-4, and are referred to as α-phases. As in may practical applications, it is required that the materials are also porous to larger guest species, we selected phases that compromise larger PLDs and low energy. They are referred to as β-phases, and are shown in Figure 2-5. The corresponding space group information and coordinates CIF files are collected in the SI. The porosity α- and β-phase polymorph were characterized in terms of PLDs, LCDs, and SAs, with the resulting highlighted in Figs. 2-4 and 2-5. Further void space analysis, including the void channel dimensionality and the visualizations of pore landscapes, are placed in the SI. The most porous of identified the lowest energy phases is assembled from molecule M1. αM1 has pores of 8.17 Å in diameter and the internal surface area of 1070 m2/g. These features make it comparable with inorganic zeolites.

The predicted energy landscape for molecules M1-M6 are depicted in Fig. 2-6. They resemble the typical ones observed for organic molecules, i.e. the leading-edge of the landscape as a function of density decreases almost monotonically. Except for the βM2 and

the αM5 structures, all the other α- and β-phase polymorphs belong to the leading-edge of the predicted crystal landscape. Thus, they constitute an example of possible crystal phases that could be experimentally accessible by crystal engineering procedures. Recrystallization or exposure to different solvent might lead to the discovery of other phases possibly comprised in Fig. 2-3 and Fig. 2-6.

Finally, it is important to comment on the synthetic feasibility of the identified six molecules. In the effort to confirm the previous reports on M1-M6, we followed their PubChem source information (see the SI). We were only able to directly identify the corresponding literature sources for M2 and M3. Neither of them reported the synthesis of these molecules (though a methyl-substituted M3 was synthesized). We could confirm, however, a previous synthesis of M4 though it required a rather extensive Web search, starting from a Google-based translation of the M4's PubChem-listed Japanese source article. Clearly, it can be noted that the discovery potential of the approach presented herein is significantly limited by the reference data included in the screened database, i.e. PubChem. This observation highlights the need for further development of community-available databases such as PubChem as well as tools used to mine the literature to create such databases.

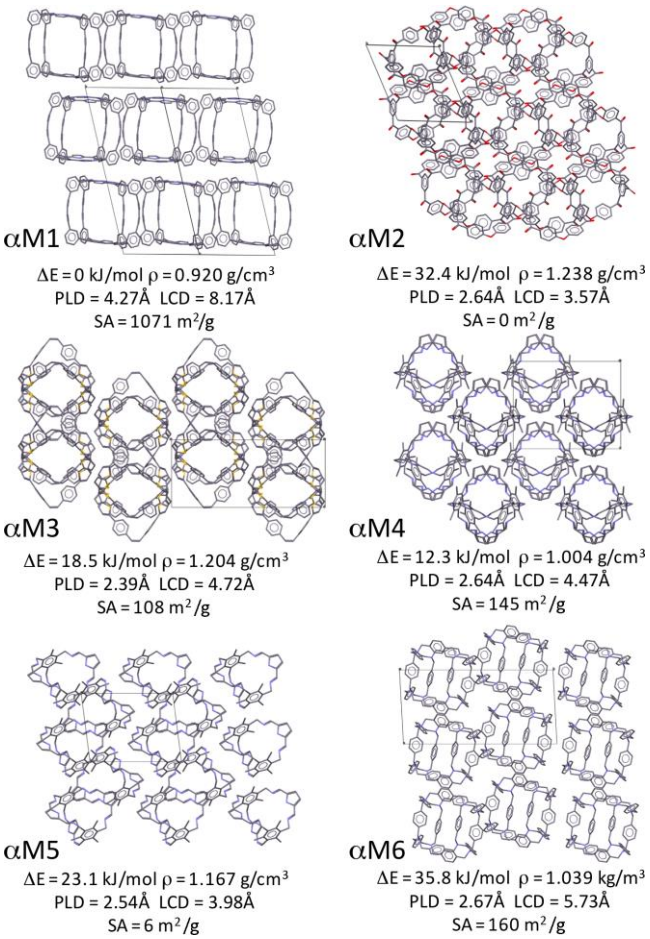



**Figure 2-4.** Lowest energy and porous (PLD>2.32Å) crystalline structures for M1-M6. The information in the captions includes the lattice energy w.r.t the lowest energy phase (ΔE), density (ϱ), PLD, LCD and the nitrogen surface area (SA).

In the article, we presented new porous materials based on cage molecules, which have stayed out of the sight of the community, i.e. ones that have been introduced and/or synthesized in the past but for which the porosity of the corresponding solid forms has not been discussed. Identification of these cages was facilitated by the development of molecular porosity characterization tools and approaches. Application of this methodology to PubChem repository of a ca. 94 million molecules led to identification of six molecules fulfilling our oblivion criteria, e.g. in many cases their PubChem entries point to national conferences, non-English journals or commercial labs, and they do not have any corresponding entry in the current version Cambridge Structure Database. Through

computational crystal structure prediction, we observed that the identified six molecules can form low-energy polymorphs with different porosity and therefore are good candidates to form porous molecular crystals. Our effort highlighted the potential of tailor-made material informatics tools to mine chemical databases to identify rare building blocks of porous cage materials.

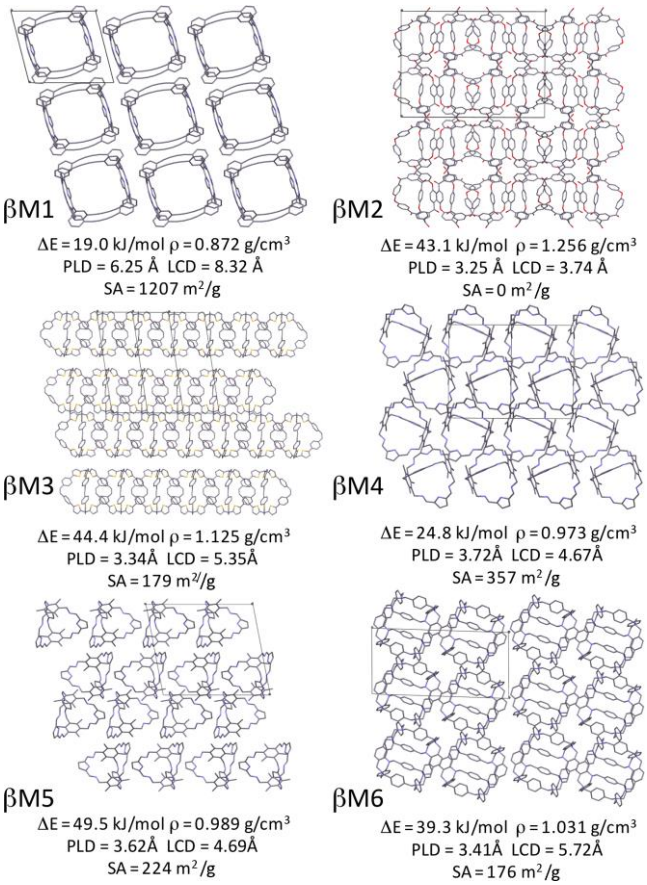



**Figure 2-5.** Examples of M1-M6 phases that combine low lattice-energy (ΔE<50 kJ/mol) and large PLDs.

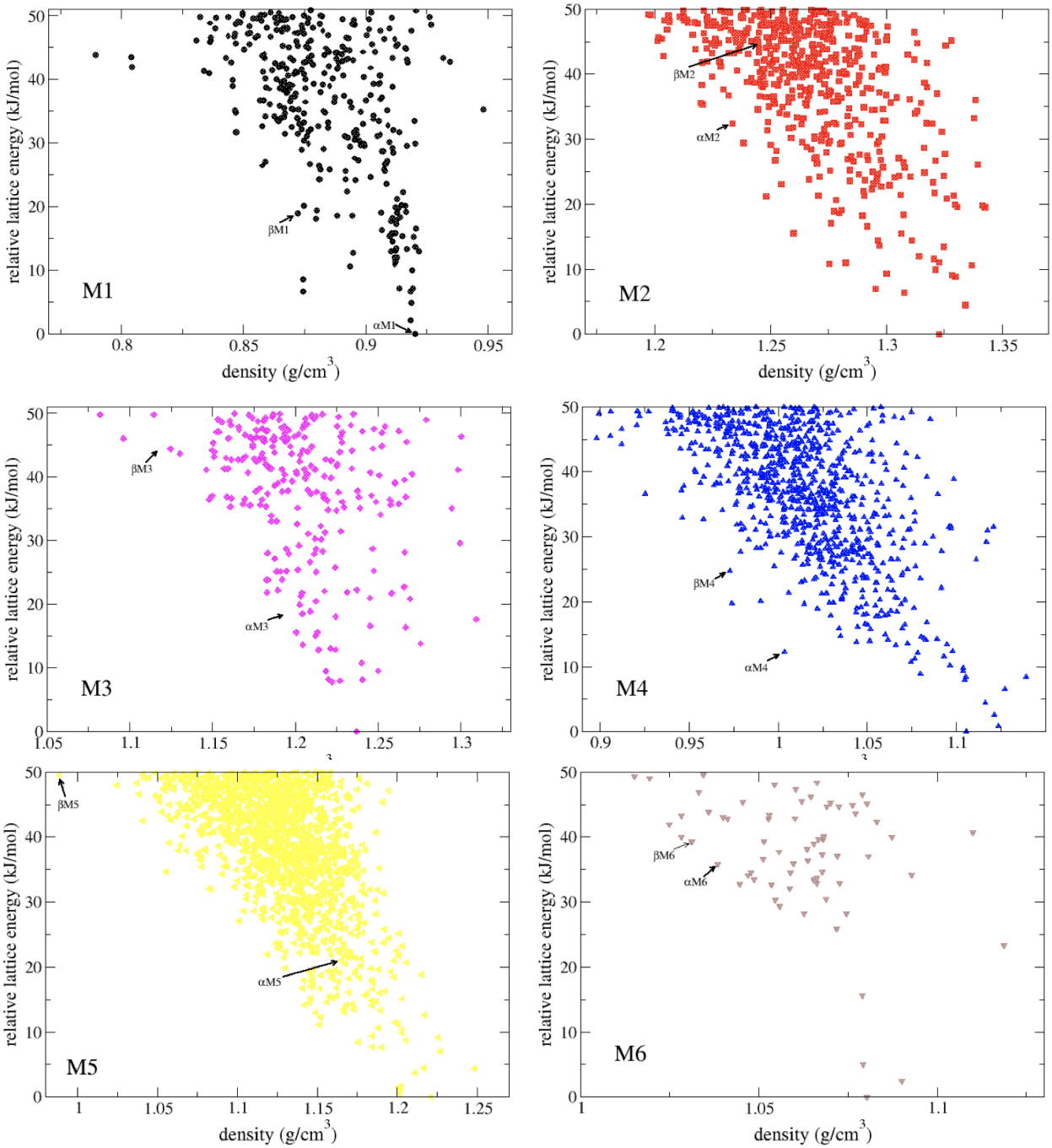



**Figure 2-6**. Energy landscape of molecules M1-M6. The α-phases and the β-phases are labeled and indicated. Except for the predicted βM2 and αM5, the identified α-phases and the β-phases belong to the leading-edge of the predicted landscape.

## 2. Notes and References

**Supporting Information**.

The following files are available free of charge.

A detailed description of the molecular porosity detection algorithm, the implementation of database screening approach and the crystal structure prediction algorithm (PDF).

Predicted crystal structures of α and β phases (compressed CIF).

All predicted phases for molecules M1-M6 (compressed CIF).

**Corresponding Author**

maciej.haranczyk@imdea.org

**Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Acknowledgement**

Authors acknowledge support from the Spanish Ministry of Economy and Competitiveness (RYC-2013-13949) and resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**References**


1. Kudo, H.; Hayashi, R.; Mitani, K.; Yokozawa, T.; Kasuga, N. C.; Nishikubo, T. Molecular Waterwheel (Noria) from a Simple Condensation of Resorcinol and an Alkanedial. *Angew. Chem. Int. Ed.* **2006**, *45*, 7948–7952.

2. Mastalerz, M.; Oppel, I. M. Rational Construction of an Extrinsic Porous Molecular Crystal with an Extraordinary High Specific Surface Area. *Angew. Chem. Int. Ed.* **2012**, *51*, 5252–5255.

3. Dalgarno, S. J.; Thallapally, P.K.; Barbour, L. J.; Atwood; J. L. Engineering Void Space in Organic van der Waals Crystals: Calixarenes Lead the Way *Chem. Soc. Rev.* **2007**, *36*, 236-245.

4. Pan, S.; Mandal, S.; Chattaraj, P. K. Cucurbit[6]uril: A Possible Host for Noble Gas Atoms. *J. Phys. Chem. B* **2015**, *119*, 10962–10974.

5. Christoforides, E.; Mentzafos, D.; Bethanis, K. Structural Studies of the Inclusion Complexes of the (+)- and (−)-Borneol Enantiomers in α- and β-Cyclodextrin *J. Incl.Phenom. Macrocycl. Chem.* **2015**, *81*, 193–203.

6. Joseph, A. I.; Lapidus, S. H.; Kane, C. M.; Holman, K. T. Extreme Confinement of Xenon by Cryptophane-111 in the Solid State. *Angew. Chem. Int. Ed.* **2015**, *54*, 1471–1475.

7. Tozawa, T.; Jones, J.T. A.; Swamy, S.I.; Jiang, S.; Adams, D.J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S.Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A.M. Z.; Steiner, A.; Cooper, A.I. Porous Organic Cages. *Nature Materials* **2009**, *8*, 973-978.

8. Chen, L.; Reiss, P. S.; Chong, S.Y.; Holden, D.; Jelfs,K.E.; Hasell, T.; Little, M.A.; Kewley, A.; Briggs, M. E.; Stephenson, A.; Thomas, K. M.; Armstrong, J. A.; Bell, J.; Busto, J.; Noel, R.; Liu, J.; Strachan, D. M.; Thallapally, P.K.; Cooper A. I. Separation of Rare Gases and Chiral Molecules by Selective Binding in Porous Organic Cages. *Nature Materials* **2014,** *13*, 954–960.

9. Zhang, G.; Presly, O.; White, F.; Oppel, I. M.; Mastalerz, M. A Permanent Mesoporous Organic Cage with an Exceptionally High Surface Area. *Angew. Chem. Int. Ed.***2014,** *53*, 1516–1520.

10. Cooper, A.I. Porous Molecular Solids and Liquids. *ACS Cent. Sci.* **2017**, *3*, 544–553 and references therein.

11. 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. *J. Phys. Chem. C* **2017**, *121*, 15211-15222.

12. Msayib, K. J.; Book, D.; Budd, P. M.; Chaukura, N.; Harris, K. D. M.; Helliwell, M.; Tedds, S.; Walton, A.; Warren, J. E.; Xu M.; McKeown, N. B. Nitrogen and Hydrogen Adsorption by an Organic Microporous Crystal. *Angew. Chem. Int. Ed.* **2009**, *48*, 3273–3277.

13. Evans, J. D.; Huang, D. M.; Haranczyk, M.; Thornton, A. W.; Sumby, C. J.; Doonan, C. J. Computational Identification of Organic Porous Molecular Crystals. *CrystEngComm* **2016**, *18*, 4133–4141.

14. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S.H. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44*, D1202-13.

15. http://www.chemspider.com/ Accessed July 13, 2017

16. https://pubchem.ncbi.nlm.nih.gov/ Accessed February 14, 2017.

17. Bolton, E. E.; Chen, J.; Kim, S.; Han, L.; He, S.; Shi, W.; Simonyan, V.; Sun, Y.; Thiessen, P. A., Wang, J., Yu, B.; Zhang, J., Bryant, S. H. PubChem3D: a new resource for scientists. *J. Chemoinformatics* **2011**, *3*, 32.

18. Materials Science Suite 2017-2, Schrödinger, LLC, New York, NY, 2017.

19. Banks**,** J. L., Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Fetls, A. K.; Halgren, T. A.; Mainz, D.T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M.P.; Zhang, L.Y.; Berne, B. J.; Friesner, R.A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comp. Chem.* 2005, *26*, 1752-80.

20. PubChem Compound IDs: 101377082 (M1), 102021452 (M2), 102210546 (M3), 102263757 (M4), 102333795 (M5), 16148678 (M6).

21. Hassel, T.; Cooper, A. I. Porous Organic Cages: Soluble, Modular and Molecular Pores. *Nat. Rev. Mater.* 2016, *1*, 16053.

22. Van Eijck, B. P.; Krooj, J. Structure Predictions Allowing More than One Molecule in the Asymmetric Unit. *Acta Cryst.* **2000,** *B56*, 535-542.

23. Willems, T.F.; Rycroft, C.H.; Kazi, M.; Meza, J.C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Materials* **2012**, *149*, 134-141.

24. http://www.zeoplusplus.org; accessed July 13, 2007.

25. Pulido , A.; Chen, L.; Kaczorowski, T.; Holden, D.; Little, M. A.; Chong, S. Y.; Slater, B.J.; McMahon, D. P.; Bonillo, B.; Stackhouse, C. J.; Stephenson, A.; Kane, C. M.; Clowes, R.; Hasell, T.; Cooper,A. I.; Day, G.M. Functional Materials Discovery using Energy–Structure–Function Maps. *Nature* **2016,** *543*, 657-664.

26. Nyman, J.; Day, G. M. Static and Lattice Vibrational Energy Differences between Polymorphs. *CrystEngComm* **2010,** *17*, 5154-5165.

27. Cruz-Cabeza, A.; Day, G. M.; Jones, W. Predicting Inclusion Behaviour and Framework Structures in Organic Crystals. *Chem. Eur. J.* **2009,** *15*, 13033-13040.

28. Pyzer-Knapp, E. O.; Thompson, H. P. G.; Schiffmann, F.; Jelfs, K.; Chong, S.; Little, M. A.; Cooper, A. I.; Day, G. Predicted Crystal Energy Landscapes of Porous Organic Cages. *Chem. Sci.* **2014**, *5*, 2235-2245.

29. Jones, J. T. A.; Holden, D.; Mitra, T.; Hasell, T.; Adams, D. J.; Jelfs, K.E.; Trewin, A.; Willock, D. J.; Day, G. M.; Bacsa, J.; Steiner, A.; Cooper, A.I. On–Off Porosity Switching in a Molecular Organic Solid. *Angew. Chem.* **2011,** *123*, 775-779.

# 3. Supporting information

## 3.1 Molecular porosity detection - Algorithm and its implementation

**General overview.** Four important steps of the molecular porosity detection and characterization approach are depicted in Figure 2-S1. They are as follows:

1. The algorithm starts with a given molecule, which is specified as a list of atoms with their Cartesian coordinates. The molecule can be preprocessed to remove atoms not relevant to the analysis, e.g. hydrogen atoms;

2. Voronoi tessellation is executed using the atoms of step (1), and as a result the Voronoi network is generated (built from vertices and edges of the individual Voronoi cells). The Voronoi network provides a graph representation of the void space around the atoms, i.e. the Voronoi nodes highlight the position of cavities formed by sets of 4 atoms as the nodes are positioned at the center of spheres that circumscribe the quadruplets of atoms;

3. For each node, Pore Exposure Ratio (PER) is calculated. It measures the exposure of a given Voronoi node to the surroundings of the molecule, integrating both geometrical and topological information about the molecule's structure;

4. Finally, a threshold PER value is used to determine if the Voronoi nodes correspond to internal or external cavities.

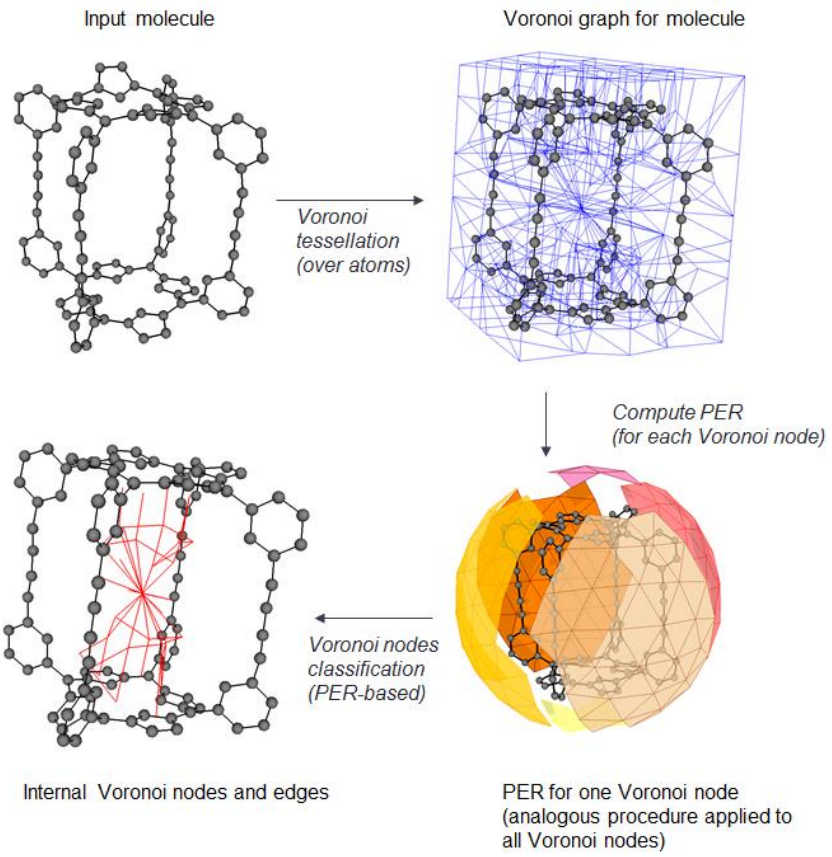



**Figure 2-S1.** Outline of the pore detection algorithm: In order to detect the pores of a molecule (top left), empty space is first characterized with the help of Voronoi tessellation (top, right). Then, each Voronoi node's PER is computed (bottom, right), and those below a given threshold are classified as internal. A map of internal porosity is obtained (bottom, left), as a subset of the initial Voronoi graph.

**Identification of void space: Voronoi tessellation.** For characterization of the empty space around and within the molecule, Voronoi graph of atoms is constructed. This graph places nodes and edges in regions where the distances to nearest atoms are all equal, dividing the space in several regions (one per atom) called Voronoi cells. In the interior of each cell, distance from any point to the associated atom is always smaller than to any other atom of the molecule. Cell boundaries correspond to shared regions, where points are at equal distance to two or more atoms. An example Voronoi tessellation is depicted in Figure 2-S2.

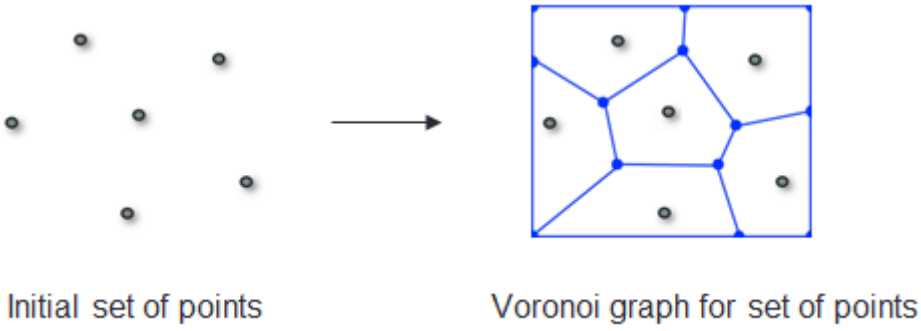



**Figure 2-S2.** 2D Voronoi tessellation. Left, a set of points (grey). Right, 2-dimensional Voronoi tessellation constructed around the points within a rectangular boundary. Blue lines represent Voronoi edges, blue dots represent Voronoi nodes.

The algorithm starts with a set of points $S$ (in our implementation, the set of points is the set of molecule's atoms). For each point $s$ in $S$, a Voronoi cell is computed as the region of space that is closer to that point than to any other in $S$. That is:

$$x \text{ in Voronoi\_cell}(s) \text{ if } dist(x, s) \leq dist(x, s') \text{ for every } s' \text{ in } S\text{-}\{s\}$$

where $dist(x, s)$ stands for the distance between point $x$ and atom $s$. The points on the boundary of the Voronoi cell are those at the exact same distance to the corresponding neighbor, thus being part of another point cell (see Figure 2-S2). This process constructs a graph in space that covers regions of maximal separation between atoms. Thus, is to be expected to represent voids around and within the molecule.

To take into account atom radii, the radical Voronoi tessellation was used following our past experience in developing Zeo++ software tool.[1] This technique changes the distance function to add atomic radii information. When using radical Voronoi tessellation, cells are given by the equation:

$$x \text{ in Voronoi\_cell}(s) \text{ if } dist^2(x, s) - rad^2(s) \leq dist^2(x, s') - rad^2(s') \text{ for every } s' \text{ in } S\text{-}\{s\}$$

Where $rad^2(s)$ stands for the square radius of atom $s$, and $dist^2(x, s)$ stands for the square of the distance from the atom $s$ to the point $x$.

Several open source implementations of Voronoi tessellation are available and well described in the scientific literature.[2] In our work, we used Voro++ library.[3]

**Identification of internal regions: Pore Exposure Ratio.** To identify the internal space of a molecule, we introduce the Pore Exposure Ratio (PER) descriptor, which characterizes how exposed a given point is with respect to a defined surrounding, i.e. spherical perimeter around the molecule. A point buried inside a molecule will be fully encapsulated by the molecule's atoms, therefore, its PER will be zero. A point in a shallow, "on-surface", molecular cavity will be exposed to the molecule's surroundings and its PER will approach a value close to 1. In our approach, we calculate PER only for points defined by the Voronoi nodes as they highlight centers of cavities.

The workflow of the algorithm is depicted in Figure 2-S3. Given a point of space, $x$, a spherical grid (it can be also seen as a mesh of triangles) $G$ is constructed around $x$ (see Figure 2-S3, top), centered at the point, and with the radius equal to 1.5 times the largest value of the distance between $x$ and each of the atoms (thus $G$ completely encapsulates the entire molecule). The algorithm decides whether $x$ is an intrinsic pore of the molecule by studying its accessibility from the outer sphere without crossing any chemical bonds or atoms. To do this, it modifies the outer grid, erasing some of its triangles, according to the following criterion: for each triangle $T$ in $G$, a tetrahedron $P$ formed by $T$ and the point $x$ is constructed (see Figure 2-S3, right). If $P$ intersects any bond of the molecule, $T$ is erased, being kept otherwise. As a result, a subgrid with one or more connected components is obtained. Pore Exposure Ratio is then computed as the ratio among the biggest connected component's (BCC) surface and the total surface of the grid $G$. I.e:

$$PER(x) = S(BCC)/S(G)$$

The algorithm for computation of PER can be described in terms of pseudocode as follows:

```
PoreExposureRatio(Point)
      G = spheric_grid_around(Molecule, Point)
      GridSurface = surface(G)
      Subgrid = erase_triangles(Molecule, Point, G)
      CCs = connected_components(Shadow)
      for (i in 1 to number(CCs))
          SurfacesList[i] = surface(CCs[i])
      end for
      SurfaceMax = max(SurfacesList)
      Rate = SurfaceMax/GridSurface
```

The functions are defined as follows:

`spheric_grid(Molecule, Point)`: Constructs a spherical grid around the point and the molecule. The grid is centered at the point, and has radius equal to 1.5 times the maximum of the distances from the point to the atoms. Thus, it surrounds the molecule. To make it regular, grid is constructed using Vogel's method.[4]

`surface(Grid)`: Computes the surface of a set of triangles, by computing each triangle surface individually and then summing them all.

`erase_triangles(Molecule, Point, G)`: Computes subgrid for the given point and molecule, erasing triangles from the grid surface if there is a bond in between them and the reference point. This procedure is depicted and explained in Figure 2-S3, right side.

`connected_components(Grid)`: For a given set of triangles in space, computes the set of connected components of those triangles. Two triangles are considered to be connected if they share an edge. Result is returned in form of a linked list.

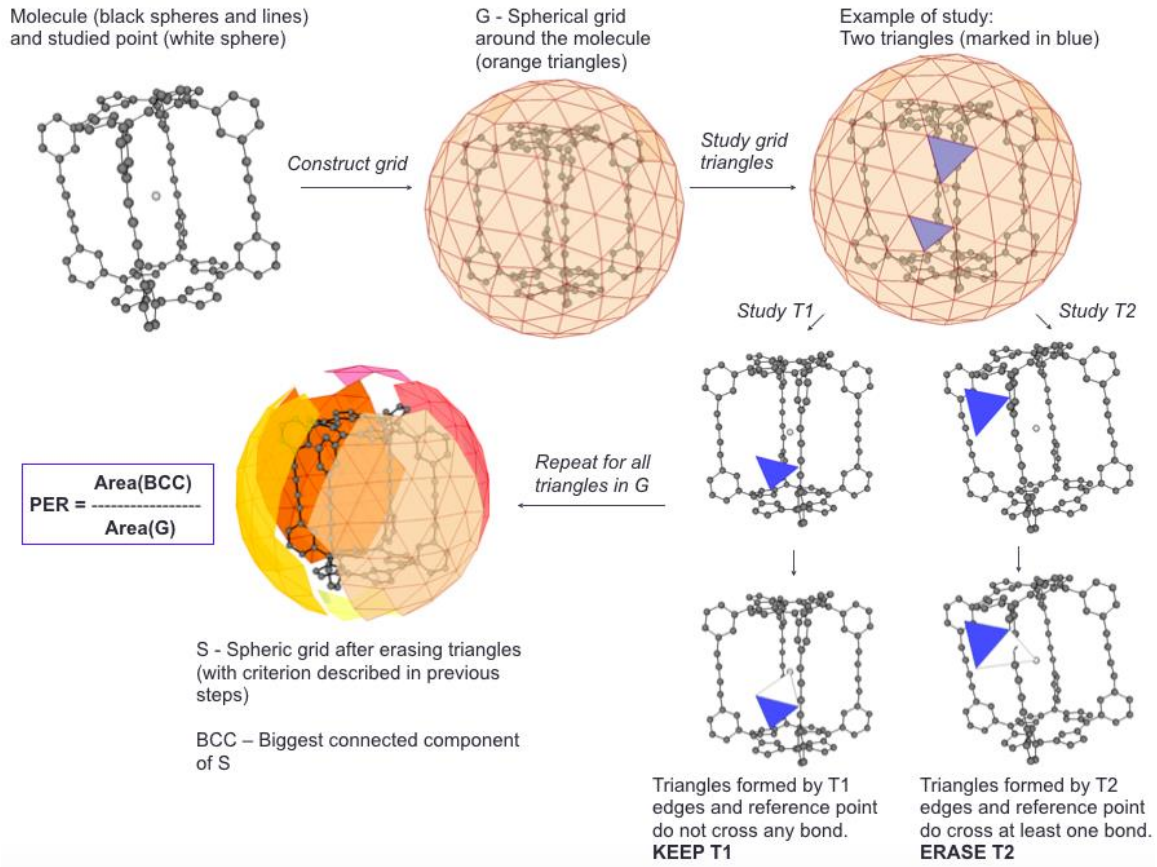



**Figure 2-S3.** Pore Exposure Ratio algorithm description. PER is computed for a given point of space with respect to the given molecule (top, left). First, a spherical grid is built centered at the point and surrounding the molecule (top, center). Then, every triangle of the grid is studied separately. Two cases are shown in the figure: one triangle that is kept (*T*1) and one that is erased (*T*2). Resulting object is a grid with some missing triangles (bottom, left). Connected component with highest surface is chosen, and *PER* is computed as the ratio of that surface and total grid surface.

**Pore identification with help of PER and Voronoi tessellation:** In order to identify molecule's pores, previously described methods are combined. The algorithm for such identification can be described in terms of pseudocode in the following way:

```
V = voro_tessellation(Molecule)
for (v_node in V)
      PER[v_node] = pore_exposure_ratio(v_node)
      if (PER[v_node] < PER_Threshold)
          v_node.type = internal
      else
          v_node.type = external
      end if
end for
```

A molecule is considered to have internal porosity (i.e., being a porous molecular cage) if it has at least one internal Voronoi node. The threshold for determination porosity is selected empirically (see next Section).

With the previous classification of internal Voronoi nodes, the pore size of a porous molecule can be determined as well as the molecule's PER can be assigned. This process can be described in the following pseudocode:

```
GetInternalPore(VoronoiGraph)
      InternalNodes =
get_internal_nodes(VoronoiGraph)
      N = node_with_max_radius(InternalNodes)
      MoleculePoreSize = N.radius
      MolecularPER = N.PER
      return(PoreSize, MoleculePER)
```

Above, the radius of a given Voronoi node is the minimum of the distances from that node to the atoms minus the atom radius. This value is returned directly from the Voronoi library (Voro++).

**Comparison of Pore Exposure Ratios for selected molecules. Threshold selection.** In order to establish a threshold value for the PER, many molecules were examined. In Figure 2-S4, few example of molecules are presented along with the PER values associated to either their pores (if any) or the center of the box (for non-porous molecules). Porous molecules in Figure 2-S4 have PER values between 0.1 and 0.31. On the other hand, PER values for non-porous molecules are much larger, i.e. above 0.75. By inspecting more than 300 molecules from PubChem, a PER threshold of 0.36 was chosen. We note here that it is quite challenging to capture the tremendous shape diversity of possible porous molecules with just one descriptor. The selected value is, therefore, rather conservative to favor the selection of porous molecules at the cost of increase of the false positives during the PubChem screening (i.e. filtering of the PubChem database and collected in the "PER prediction set" (see Figure 2-S5)).

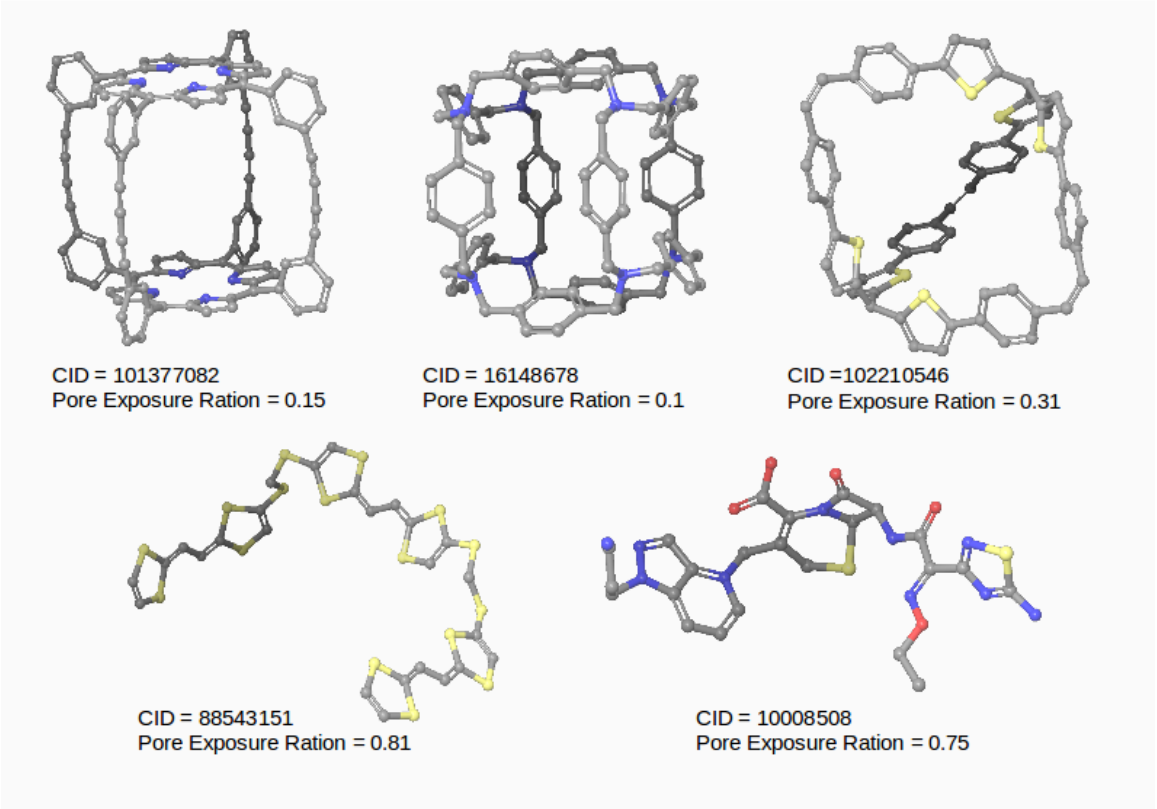



**Figure 2-S4. Examples of PER for selected molecules.** Several molecular schemes are shown, all with their PubChem CIDs. Top three molecules are identified as cage molecules while bottom two molecules are non porous molecules. PER values are computed for the center of the box containing the molecule and for the position of the largest pore for non-porous and porous molecules, respectively.

## 3.2 Screening of PubChem database

The PubChem Compound Database (PubChem) was downloaded in the SDF data format from the PubChem FTP server. Molecules in each SDF file were converted to the SMILES format by using the OpenBabel Software to facilitate processing.[5] The resulting set of 94 333 141 compounds was filtered to remove molecules that were irrelevant to our search for cage molecules. The filtering consisted of 6 steps, which are outlined below and were implemented in Python. They are also and depicted in Figure 2-S5.

Step 1. **Single Molecule filter**: Each compound with more than one molecule was identified by detecting a "." in the corresponding SMILES sequence and discarded.

Step 2. **Organic Molecule filter**: In order to reduce our set to only organic molecules we used the same criteria described in the PubChem3D release notes,[6] i.e. molecules with at least one match of the following SMARTS pattern "[!#1;!#6;!#7;!#8;!#9;!#14;!#15;!#16;!#17;!#35;!#53]" in their SMILES were labeled as non-organic and discarded.

Step 3. **Neutral Molecule filter**: The total charge of a compound was deduced directly from the SMILES sequence. Molecules with and absolute charge larger than 0 were labeled as charged and discarded.

Step 4. **Non-radical Molecule filter**: Radicals were identified using the SMARTS pattern "[X5v4,X5v3,X5v2,X5v1, X4v3,X4v2,X4v1,X3v2,X3v1,X2v1]". Those molecules with at least one match of the former pattern were labeled as radical and discarded.

Step 5. **Core-based Molecule filter**: Here, we refer to the "core" of a molecule as a substructural unit that can encapsulate internal space due to its topology. The "core" is a subgraph of the molecular graph that fulfills either of the following: 1) it is a non-planar graph or 2) it is a non-planar graph after adding a universal vertex. Molecules with at least one "core" were identified using an algorithm of Boyer and Myrdov,[7] implemented in the graph-tool module of Python.[8]

Molecules without a "core" were labeled and discarded.

Step 6. **Molecule with more than 8 rotatable bonds filter**: First, the largest core (from Step 5) in a molecule is detected. Then, all the atoms and their connecting bonds that emanate from the previously identified largest core substructure are identified. Those bonds were considered to be rotatable if they were single bonds and were not part of a small ring structure (we consider a ring as small if its size is smaller than 8). Bonds involving hydrogen atoms were not included in the counting of rotatable bonds.

Molecules with more than 8 rotatable bonds emanating from the largest core were labeled and discarded. This step allowed us to keep only possible good candidate for which crystal structure prediction (see below and main text) calculations were computationally feasible.

The above 6-steps of sequential filtering reduced the PubChem set to 1 417 637 molecules. Among them, 308 018 molecules did not have an entry in PubChem3D database, hence their 3D coordinates were not available in the original PubChem database. For those structures, 3D coordinates were generated only for those molecules with 48-or-more atoms. The cutoff of 48 was chosen based on the size of the smallest calix[4]arene (a basket molecule with and internal pore and extensively investigated in the field of guest-host chemistry and porous materials).

The final set of compounds suitable for the internal porosity analysis contained 1 258 975 molecules.

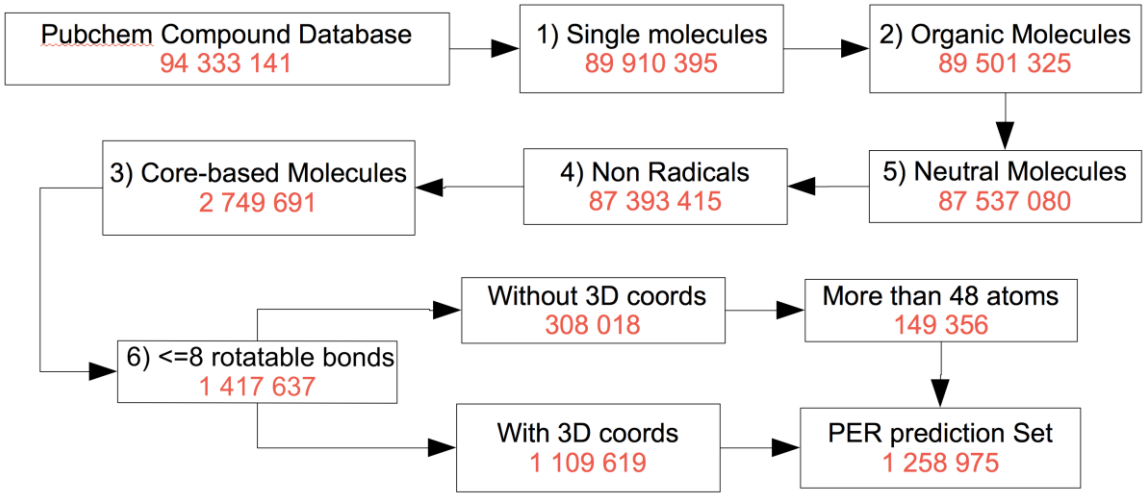



**Figure 2-S5.** PubChem database screening flow-chart. A 6-steps sequential filtering protocol is applied to the PubChem data base, searching for candidates for porous organic molecules. The initial number of 94 333 141 compounds was reduced to a final set of 1 258 975 molecules that were further analyzed with our molecular porosity detection algorithm.

### 3.3 Crystal structure prediction

**Approach.** Crystal structure prediction (CSP) calculations were performed using a random search technique implemented in the UPACK (Utrecht Crystal Package) program suite.[9] UPACK provides hypothetical crystal structures by approximating the free energy of the crystal with a static lattice energy and minimizing it by using a molecular force field (FF). The latter is defined as the sum of an intermolecular potential, consisting of van der Waals and Coulomb terms, and a combination of intramolecular energy terms, which account for molecular flexibility.

We employed the widely used transferable OPLS_2005 force field[10] to model the lattice energy for the systems considered in this study. In the OPLS_2005-FF framework, the

cohesive energy is modeled as the sum of Lennard-Jones and Coulomb terms while intramolecular interaction include stretching, bending and torsional terms. OPLS_2005 FF is an enhanced version of the original OPLS force field,[11] and is developed by the Schrodinger[12] company to provide a bigger coverage of organic functionality. In particular, intramolecular parameters have been fit to reproduce the conformational energies derived at a higher level of quantum theory for a large set of organic molecules.

As stated above, static lattice energy minimization is at the core of the UPACK method for crystal structure prediction simulations. Energy minimization steps are normally performed by a combination of the steepest descendent method followed by a certain number of conjugated gradient minimization steps. Intermolecular interactions are truncated at a certain cut-off based on the charged groups distance and the Ewald summation method is employed for Coulomb and Lennard-Jones $r^{-6}$ dispersion terms.

A full CSP study with UPACK was carried out in three steps:

**Step 1**. In the first step, we generated candidate structures (5000 are usually sufficient) by means of a random search method combined with a rough energy minimization (using a cut-of 8Å for intermolecular interactions). Initial structures are generated for a set of selected space groups with Z'=1 molecules in the asymmetric unit cell.

The number of degrees of freedom for the random search comprehends up to 6 variables for the unit cell plus 6 degrees of freedom that describe global translation and rotation of the molecule in the unit cell. For molecules with rotatable bonds further degrees of freedom are considered during the random search to account for different molecular conformations.

The total number of degrees of freedom, D, cannot be arbitrarily large for the random search method to be successful. Indeed, it was empirically noticed¡Error! Marcador no definido. that the r andom search method started to fail when the total number of random variables is around 20.

**Step 2**. During this step, a minimization of the static lattice energy with a larger cut-off of 20.0Å w.r.t all the degrees of freedom was used to further refine the structures generated in Step 1.

**Step 3**. Equivalent structures resulting from different initial orientations of Step 1 were grouped together by the means of a clustering algorithm based on interatomic radial distribution functions as implemented in UPACK. The final set of unique structures was finally ranked according to their lattice energy relative to the lowest minimum found.

*Validation.* In the last few years, OPLS_2005 FF was successfully employed to investigate the lowest energy conformers of shape persistent molecules with large cavities and their ability to form guest inclusion complexes in permanent porous liquids.[13] In this work, we adopt OPLS_2005 parameters to predict the packing behavior of such cage-like molecules in the crystal phase, and as it is the first such application, we have carefully validated it. We tested our aforementioned procedure by performing the CSP calculations for the well-known porous imine cage CC1. CC1 cage molecule is one of a series of porous organic shape-persistent cages with large internal cavities synthesized in the group of Prof. Cooper of University of Liverpool, UK. These molecules have been of great interest especially for their capacity to tune porosity when exposed to different solvents.[14]

The energy landscape of cage CC1 has been already investigated in two independent CSPs works employing different approaches such as Monte Carlo simulated annealing and quasi-random sampling for generating the initial trial set of structures.[15] In both cases the final minimization procedure was carried out with the DMACRYS crystal structure modeling software for rigid molecules, which employs anisotropic atom-atom model potentials.[16] In both studies, the two experimentally observed polymorphs CC1-α' (non-porous) and CC1-β'($H_2$ selectively porous) were found in the low energy region of the phase landscape.

Currently, all the triclinic, monoclinic and orthorhombic space groups are available to be searched in UPACK but in this work we limit our CSP calculations to a subset of 13 space groups (P21/c,P-1,P212121,P21,Pbca,C2/c,Pna21,Cc,Pca21,C2,P1,Pbcn,Pc), which account for about 90% of the observed crystal structures for organic molecules with Z'=1 molecules in the asymmetric cell. For a group with internal symmetry, the generated structures can actually belong to a space group with higher symmetry. In such cases, the full symmetry was determined by post-processing the structures with the program

PLATON.[17] Finally, the interatomic interactions cut-offs during the first and the second stages of our procedure were set equal to 8Å and 20Å respectively.

The obtained potential energy-phase landscape for cage CC1 is shown in Figure 2-S6.

Both polymorphs are found by the UPACK search. Polymorph CC1-α' is correctly predicted in the space group P21/c (see Table 2-S1). On the other hand, polymorph CC1-β' was predicted in the P1 space group, which reduced to higher symmetry R3 after analysis with PLATON. A closer inspection of the energy landscape reveal that α' and β' structures lie approximately 7kJ/mol and 26 kJ/mol above the global minimum, respectively. More precisely, UPACK predicted energy difference between the two polymorphs is $\Delta E = E(\alpha') - E(\beta') = -19.88$ kJ/mol. This energy difference, its value and sign, is within generally acceptable difference from the corresponding -9.57 kJ/mol value obtained after further relaxation of the structures by means of dispersion-corrected solid state DFT-D calculations performed with QuantumEspresso.[18] A difference of 20 kJ/mol or more between the experimentally recognized polymorphs was predicted for other organic porous molecular materials,[15,19] and its order of magnitude is well maintained between anisotropic atom-atom model potentials and DFT-D approaches.

Our predicted structures for cage CC1 present comparable geometrical match to the experimental polymorphs, see Figure 2-S7 and 2-S8, when compared with the DMACRYS results of reference 15. Comparisons for unit cell parameters, root mean square displacement $RMSD_{30}$ for a cluster of 30 molecules, and values of the pore limiting diameter (PLD) are resumed in Table 2-S1.

Motivated by the results for cage CC1, we decided to apply UPACK prediction procedure to study the porosity properties of the predicted structures for the molecules selected in

this work. In particular we will focus on the porosity properties of all the structures predicted in a 50kJ/mol range above the global minimum.

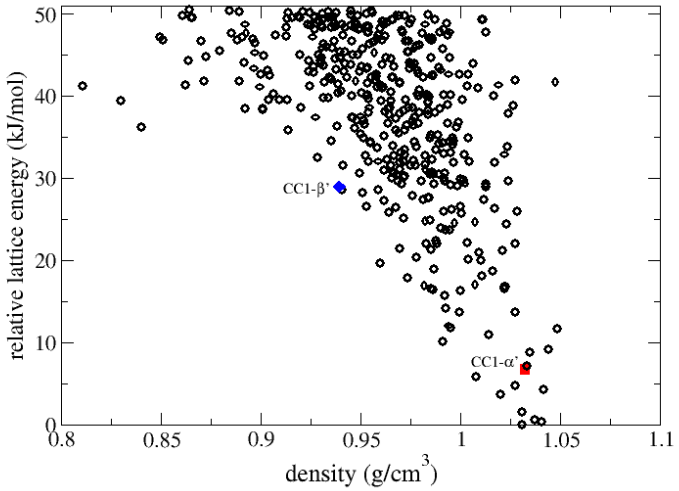



**Figure 2-S6.** Energy-density plot for porous cage CC1. Energy density map for cage CC1 as predicted by UPACK with OPLS_2005 force field. Each point corresponds to a unique structure. Predicted structures that geometrically match the experimental polymorphs CC1-α' and CC1-β' are labeled and highlight with a filled red and blue point, respectively. Geometrical parameters are given in Table 2-S1.

| | | Cell axes length (Å) | | | Cell angles | | | Density (g/cm$^3$) | $RMSD_{30}$ (Å) | PLD (Å) |
|---|---|---|---|---|---|---|---|---|---|---|
| Material | Structure | a | b | c | α | β | γ | | | |
| CC1 **α'**(P21/c) | Exp. | 12.810 | 10.910 | 36.810 | 90.00 | 97.49 | 90.00 | 1.033 | | 1.675 |
| | UPACK + OPLS_2005 | 12.725 | 10.865 | 37.689 | 90.00 | 101.72 | 90.00 | 1.032 | 0.749 | 1.495 |
| | Quasi-random +rigid DMACRYS | 13.425 | 11.156 | 37.761 | 90.00 | 94.45 | 90.00 | 0.934 | 0.812 | 2.025 |
| | | | | | | | | | | |
| CC1 **β'**(R3) | Exp. | 21.015 | 21.015 | 10.491 | 90.00 | 90.00 | 120.00 | 0.988 | | 2.318 |
| | UPACK + OPLS_2005 | 21.119 | 21.119 | 10.730 | 90.00 | 90.00 | 120.00 | 0.953 | 0.191 | 2.469 |
| | Quasi- | 21.623 | 21.623 | 10.851 | 90.02 | 90.02 | 119.98 | 0.899 | 0.603 | 2.405 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | random+rigid DMACRYS | | | | | | | | | |

**Table 2-S1.** Geometrical parameters from our CSP study with UPACK-OPLS_2005 FF, compared to experimental values and CSP results from reference 15. Polymorph CC1-β' was predicted in P1 space group but showed the full R3 space group symmetry after analysis with PLATON. RMSD over a cluster of 30 molecules where computed with the RMS module implemented in the GROMACS package.[20] Pore limiting diameters (PLD) were evaluated with the Zeo++ code.[¡Error! Marcador no definido.,21]

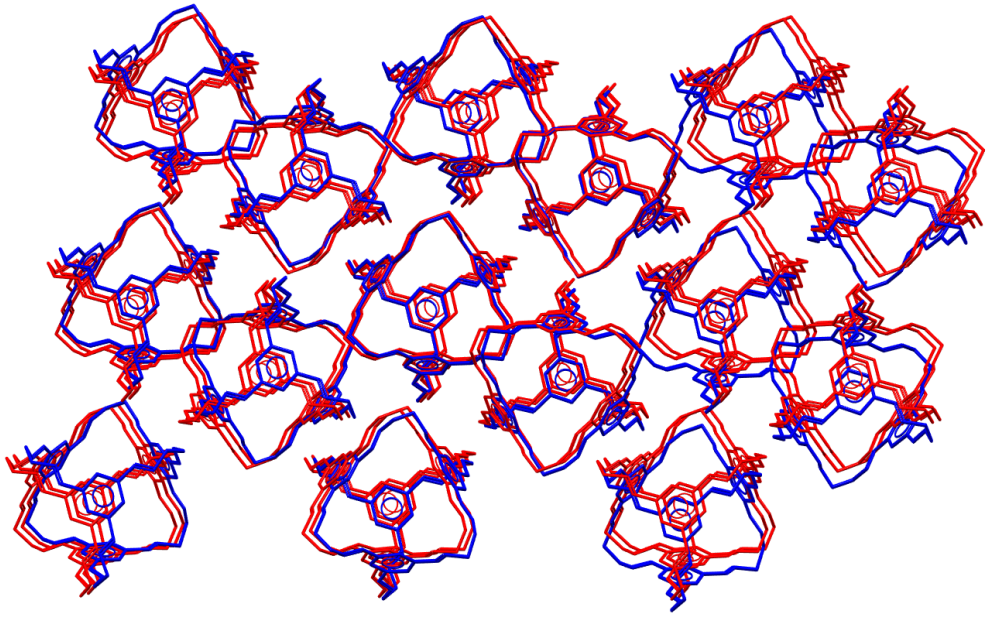


**Figure 2-S7.** Structural comparison for experimental and UPACK-predicted CC1-α' structures. Structural overlay of a cluster of 30 molecules for experimental (blue) and predicted (red) CC1-α' structures. The root mean square displacement, computed without taking into account the hydrogen atoms, takes a value of $RMSD_{30}$=0.749 Å.

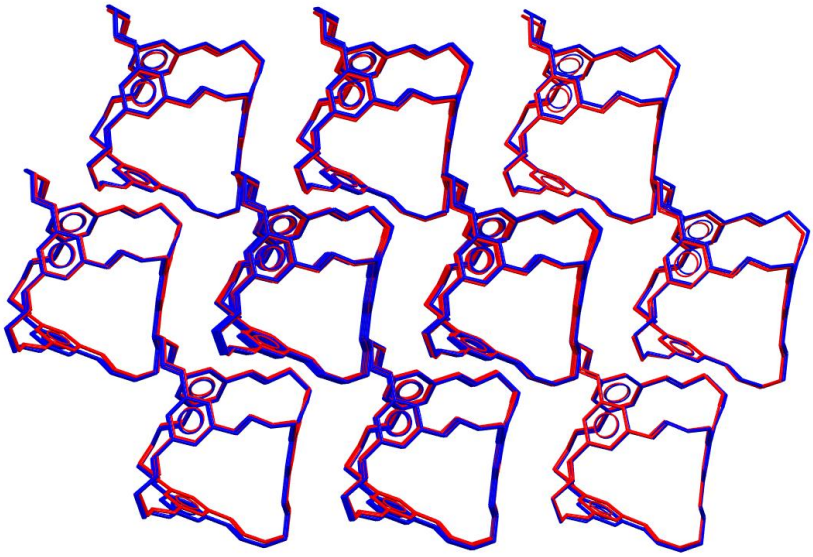


**Figure 2-S8.** Structural comparison for experimental and UPACK-predicted CC1-β' structures. Structural overlay of a cluster of 30 molecules for experimental (blue) and predicted (red) CC1-β' structures. The root mean square displacement, computed without taking into account the hydrogen atoms, takes a value of $RMSD_{30}$=0.191 Å.

### 3.4 Source information for the identified molecules M1-M6

The PubChem Compound Database contains validated chemical information provided to describe substances in PubChem Substance database. The latter contains descriptions of samples from a variety of sources. By inserting the corresponding CID (Compound-Index) in the web page https://pubchem.ncbi.nlm.nih.gov/# is possible to retrieve information about the source, the deposit date and possible scientific publications that make reference to a specific compound object of study. In the table below we resume such information for molecules M1-M6, respectively.

| Molecule | PubChemID | PubChem Source | Reference | Source/reference language |
|---|---|---|---|---|
| M1 | 101377082 | The Japanese Science and Technology Agency. | Oniki, Y. and Kenta, G., Structural and Organic Chemistry Discussion Paper Preliminary Proceedings, Volume 35th, Page 377 (2005). | Japanese |
| M2 | 102021452 | The Japanese Science and Technology Agency | Colquhoun, H. M., Arico, F., Williams, D. J. One-step syntheses of very large cage-type molecules from aromatic sub-units. Chem. Comm., 2574-2575 (2001). | English |
| M3 | 102210546 | The Japanese Science and Technology Agency | Kurata, H., Haruki, K.m Oda, M. Stilbene-Extended Trithienylmethanophanes: A novel C3-symmetric cage molecule capable of accommodating and organic molecule as well as three Silver(I) ions. Chemstry Letters 34, 484-485 (2005). | English |
| M4 | 102263757 | The Japanese Science and Technology Agency | Iwasaky, Y., Nagata, K. Inomata, T. Ozawa, T. Proceedings from the Symposium on Complexation Chemistry, Volume 64th, Page 127 (2014). | Japanese |
| M5 | 102333795 | The Japanese Science and Technology Agency | Publications:Yonaga, S., Tange, K., Hatanaka, T., Funahashi, Y. Proceedings from the Symposium on Complexation Chemistry, Volumet 64th, Page 332 (2014). | Japanese |
| M6 | 16148678 | Thomson-Pharma | | |

**References**


1. Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous and Mesoporous Materials* **2012**, 149, 134-141.

2. (a) Jamin, C.; Pion, S.; Teillaud, T. 3D Triangulations. In CGAL User and Reference Manual. [Journal] // CGAL Editorial Board, 4.10 edition. - 2017. (b) Rycroft C. H. Voro++: a three-dimensional Voronoi cell library in C++ .Paper LBNL-1430E. January 23rd, 2009.

3. Rycroft, C. H. Voro++ 2010. http://math.lbl.gov/voro++/about.html

4. Vogel, H. Mathematical *Biosciences* **1979,** *44***,** 179-189.

5. O'Boyle, N.M. ; Banck, M. ; James, C.A. ; Morley, C. ; Vandermeersch, T. ; Hutchison, G. R. J. Cheminformatics **2011**, *3*, 1-14.

6. National Center for Biotechnology Information. *PubChem3D* Release Notes. 2011 March.

7. Boyer, J.M. ; Myrvold, W. J. *Journal of Graph Algorithms and Applications* **2004**, *8*, 241-273.

8. Peixoto,T.P . *The graph-tool python library*. Figshare. 2014. DOI: 0.6084/m9.figshare.1164194

9. Van Eijck, B. P.; *Krooj*, J. *Acta Cryst. B* **2004,** *56*, 535-542.

10. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Fetls, A. K.; Halgren, T. A.; Mainz, D.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.;

Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. *J. Comp. Chem.* 2005, *26*, 1752.

11. Jorgensen, W. L.; Maxwell, D.; **Tirado**-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

12. **Materials Science Suite 2017-2**, Schrödinger, LLC, New York, NY, 2017.
(a) Jelfs, K. E.; Eden, E. G. B.; Culshaw, J. L.; Shakespeare, S.; Pyzer-Knap, E.O.; Thompson, H. P.G.; Bacsa, J.; Day, G M.; Adams, D. J.; Cooper, A. I. J. Am. Chem. Soc. 2013, 135, 9307-9310; (b) Giri, N.; Del Popolo, M. G.; Melaugh, G.; Greenway, R.L.; Ratzke, K.; Koschine, T.; Pison, L.; Costa-Gomes, M.F.; Cooper, A.I.; James, S. L. Nature Letters 2015, 527, 216-220 .

13. Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. Nature Materials 2009, 8, 973-978; (c ) Jones, J. T. A.; Holden, D.; Mitra, T.; Hasell, T.; Adams, J.D.; Jelfs, K.E.; Trewin, A.; Willock, D. J.; Day, G. M.; Bacsa, J.; Steiner, A.; Cooper, A.I. Angew. Chem. 2011, 123, 775-779.

14. (a) Pyzer-Knapp, E. O.; Thompson, H. P. G.; Schiffmann, F.; Jelfs, K. E.; Chong, S. Y.; Little, M. A.; Cooper, A. I.; Day, G. M. Chem. Sci. 2014, 5, 2235−2245; (b) Case, D. H.; Campbell, J. E.; Bygrave, P. J.; Day, G. M. J. Chem Theory Comput. 2016, 12, 910-924.

15. Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Phys. Chem. Chem. Phys. 2010, 12, 8478−8490.

16. Speck, A. L., PLATON, a multipurpose crystallographic tool, Utrecht University, Utrecht, The Netherlands, 2001.

17. Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G.L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; Fabris, S.; Fratesi, G.; de Gironcoli, S.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A.P.; Smogunov, A.; Umari, P.; Wentzcovitch, R.M. J. Phys. Condens. Matter 2009, 21, 395502 .

18. Pulido , A.; Chen, L.; Kaczorowski, T.; Holden, D.; Little, M. A.; Chong, S.Y.; Slater, B.J.; McMahon, D. P.; Bonillo, B.; Stackhouse, C. J.; Stephenson, A.; Kane, C. M.; Clowes, R.; Hasell, T.; Cooper, A. I.; Day, G.M. Nature 2016, 543, 657-664.

19. Van Der Spoel, D.; Lindhal, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. J. Comp. Chem. 2005, 26, 1701-1718.

20. www.zeoplusplus.org, accessed July 13, 2017.