

In silico design and assembly of cage molecules  
into porous molecular materials.

Marco Bernabei,<sup>\*a</sup> Raúl Pérez-Soto,<sup>a</sup> Ismael Gómez García<sup>ab</sup> and Maciej Haranczyk<sup>\*a</sup>

<sup>a</sup> IMDEA Materials Institute, C/Eric Kandel 2, 28906 - Getafe, Madrid, Spain

<sup>b</sup> Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain

\* Corresponding authors.

**Molecular Systems Design and Engineering, 2018, 3, 942-950**

## Abstract

Porous materials based on organic cage molecules are an exciting alternative to porous framework materials. Their modular chemistry offers significant structure tunability while their solution processability facilitate their assembly in functional structures in either crystalline or amorphous phase. The family of imine cages have been the most widely studied, with several analogues synthesized and tested in applications such noble gas separation, water purification and sensing. The crystals reported so far based on imine cages obtained by condensation of 1,3,5-triformylbenzene with several vicinal diamines (1,2-diamine) have rather small pore diameters. For example, the benchmark Covalent Cage 3 (CC3) crystal has pore limiting diameters of 3.65 Å, and may experience diffusion limitations hindering their technological applications. The results reported so far suggest that larger cavities are not supported in the crystal phase, unless they are stabilized by solvent molecules. In this work, we employ a density functional theory-refined crystal structure prediction approach to perform screening of various CC3 analogues to identify cages that can support stable low density porous structures. Based on analysis of the crystal energy landscapes, we present a number of synthetically accessible phases that support limiting diameters in the excess of 10 Å.

## 1. Introduction

Porous materials are critical to many industrial sectors, including petrochemicals, energy and water. In the effort to improve of the current technologies relying on porous materials, new classes of advanced porous materials are investigated. One of them is a class of porous molecular materials, in which discrete rigid molecules are assembled into 3-dimensional structures, either crystalline or amorphous.<sup>1</sup> Although porous molecular materials have been known for about five decades,<sup>2</sup> only the recently introduced classes of porous organic cage (POC) materials has offered structural tunability, through modular chemistry, that has allowed achieving porosity in par with advanced framework materials such as MOFs and COFs<sup>1</sup>. For example, for the cage-based material reported by Zhang et al. a surface area of 3758 m<sup>2</sup>/g was measured.<sup>3</sup>

One of the most studied families of POCs is a family of cages assembled by imine condensation reaction, i.e. imine cages<sup>2</sup>. Depending on the connectivity of a cage's building blocks, various cage topologies can be obtained,<sup>4</sup> including [3+6] tubular cages<sup>5</sup> and [2n+3n] cages structures with n=1,2,4.<sup>6,7,8,9,10,11,12</sup> In particular, several [4+6] imine cages synthesized in the group of Prof. Cooper since 2009, were obtained by condensation of a 1,3,5-triformylbenzene with several vicinal diamines. Chemical structures of these [4+6] cages are depicted in Figure 5-1. They exhibit an internal spherical pore with diameters spanning the range 5.3-5.8 Å and all but cage CC2 have tetrahedral symmetry. The synthesized cage molecules in Figure 5-1, due to their solution processability, can be fairly easily processed into nanoparticles, membranes and thin films<sup>13</sup>, exhibiting amorphous or crystalline 3D phases with different porosity patterns. Porosity in such POCs can emerge from (i) external voids created by inefficient packing, i.e. extrinsic porosity; (ii) a porous percolating network connecting the internal void through the cage windows or both (i) and (ii).

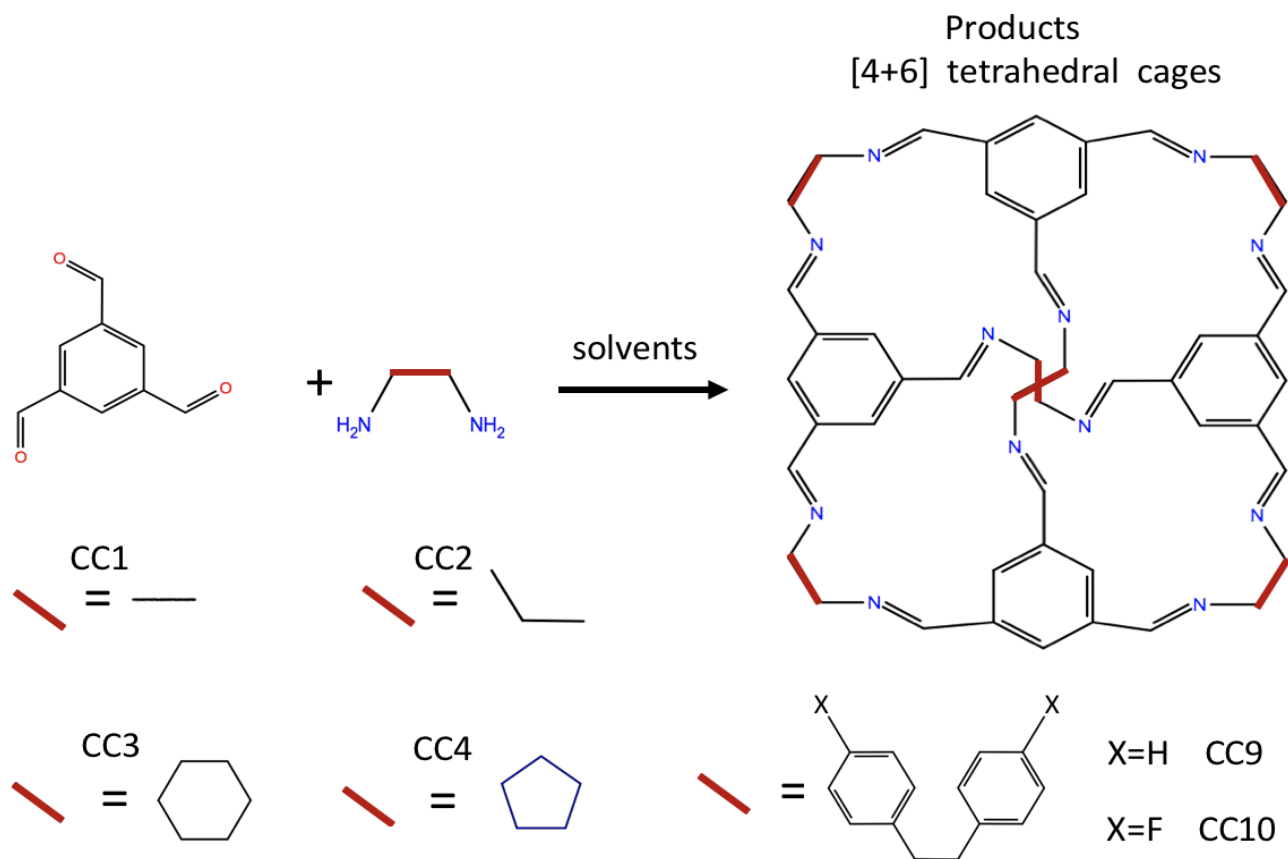
We summarized the porosity of desolvated stable crystal phases of several [4+6] imine cages in Table 5-1 by providing the corresponding pore limiting diameters (PLD), e.g. the diameter of the largest spherical probe that can diffuse through the material pores. All of these experimentally investigated phases exhibit small pores, with PLDs in the range of 1.67-3.97 Å. The porosity is strongly affected by the external cage functionalization, e.g. introduced by substituting the vicinal diamine linker. Also, the solvent employed during the crystallization process can steer towards various crystal polymorphs featuring different packing motifs and porosity upon solvent removal. For example, both polymorphs of CC4,<sup>14,16</sup>(CC4 $\alpha$  and CC4 $\beta$  in Table 5-1) were obtained by crystallization in different solvents which also led to polymorphism in the case of CC1<sup>15</sup> (CC1 $\alpha'$  and CC1 $\beta'$  in Table 5-1) and cage CC3<sup>16</sup>(CC3 $\alpha$  and CC3 $\beta$  in Table 5-1).

Although the structure of a cage, in particular the chemical functionalization of its building blocks, has a direct effect on the crystal packing and the resulting porosity, its relation is challenging both to uncover and to be expressed in the form of simple, empirical design rules. Computational techniques, in particular crystal structure prediction (CSP), can be employed to explore this relation for a particular cage material. The CSP procedure involves the generation of initial trial structures in several symmetry space groups and subsequent minimization of the free energy of the crystals.<sup>17</sup> The latter is usually approximated with the lattice energy expressed in terms of interatomic potential functions.<sup>18</sup> Typically, CSP results are presented as energy landscapes, i.e. plots of the minimized lattice energy as a function of the density of the predicted structures. Experimentally accessible structures are normally found in the leading edge of the energy landscape, defined as the set of predicted structures with the lowest energy at certain density<sup>19</sup>.

CSD Refcode	PLD (Å)	Density (g/cm <sup>3</sup> )
ELALAF (CC1β')	2.32	2.32
PUDWUH(CC1α')	1.67	1.67
PUDXAO (CC2)	3.90	3.90
PUDXES (CC3α)	3.65	3.65
PUDXES02 (CC3β)	3.41	3.41
OZECAY (CC4α)	3.57	3.57
OZECAY03 (CC4β)	3.97	3.97
GANDAC (CC9-R3)	2.19	2.19
GANDUW (CC10)	2.78	2.78

**Table 5-2.** Pore Limiting Diameter and density for the desolvated crystal structures formed by some of the [4+6] cages developed by Cooper's group. PLDs are computed with Zeo++ program. The crystal structures can be found in the Cambridge Structure Database (CSD) by searching the corresponding Refcode.

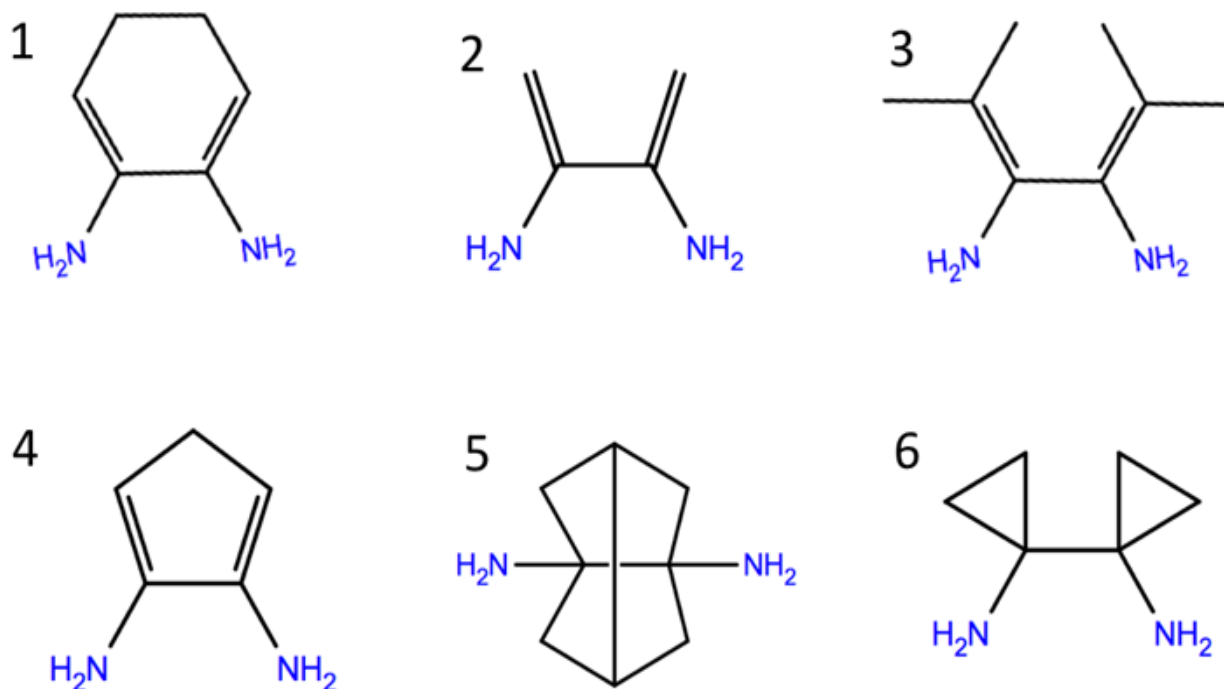
Due to the size of porous organic cages and their corresponding unit cell parameters, performing accurate crystal structure prediction study is a great challenge. So far, CSP based on lattice energy minimization for organic cages have been solely reported by Prof. Day's group<sup>5,7,14,20</sup>, and shed light on the experimentally observed trends in the crystallization behavior of several of imine cages w.r.t to changes in their structures. Also, a new polymorph of cage CC4 (CC4β in Table 5-1) was predicted and subsequently observed with experiments.



**Figure 5-1.** Schematic representation of the synthesis of several [4+6] imine cages<sup>4,5,6,7</sup>. The reactions were carried out using 1,3,5-triformylbenzene and several vicinal diamine linkers in different solvents.

The crystalline phases obtained by CSP can be characterized in terms of their porosity (PLD), thermal stability and other physical properties, e.g. guest-adsorption expressed as a gas uptake at certain conditions. Such characterization results can be combined with various structural descriptors and the relative stability information from CSP to produce informative plots, which are referred to the energy-structure-function maps (ESF maps, a term coined by Cooper and Day<sup>21</sup>). Such plots can be useful tools in materials discovery as they allow to identify screening targets, i.e. phases with desired properties, and estimate their synthesizability potential. Indeed, Cooper and Day used ESF maps to identify not-yet-discovered low-density phases ( $0.4 \text{ g/cm}^3$  and  $\text{PLD} \approx 20 \text{ \AA}$ ) of non-porous molecules

based on triptycene, spiro-biphenyl and pentyptycene cores with hydrogen-bonding functionalities.<sup>21</sup>



**Figure 5-2.** Final set of linkers employed to *in silico* synthesize new porous cages. Their corresponding PubChem Compound IDs are: 54142672 (1), 85839203 (2), 21282247 (3), 58399750 (4), 55283729 (5), 17942952 (6).

Herein, we present a POC discovery approach that combines *in silico* synthesis of novel imine cages, CSP and structure characterization. We demonstrate the approach with the discovery of several new [4+6] imine cage materials with unprecedented porosity as expressed by large PLDs. In particular, our screening targeted analogues of a robust CC3 material. We identified 6 achiral vicinal diamine linkers that are structurally compatible with the ones listed in Figure 5-1. The effect of chemical functionality and geometry of the new diamine linkers on the crystalline phases that they may form are studied by means of CSP. Two new low-density

phases, featuring very large pores with PLDs  $\approx 10\text{\AA}$  are identified and their stability is further assessed with extensive solid state DFT-D3 calculations and molecular dynamics simulations.

## 2. Computational methods

### 2.1 Database screening and 3D POCs molecular models.

We screened PubChem3D<sup>22</sup> database to identify candidate 1,2-diamine linkers for building in-silico [4+6] tetrahedral cages, which are analogues of the cages listed in Figure 5-1. The PubChem3D compound database contains 80 821 441 unique records of molecular structures with their 3D coordinates. The PubChem3D compounds were mined from patents, vendors list and scientific literature, and to some extent the set reflects all molecules ever considered by chemists. We filtered PubChem3D database to retain only vicinal diamine linkers (we will refer to them as 'linkers' onwards), obtaining a group of 6676 including the linkers in Figure 5-1. The latter was subsequently filtered to discard chiral with side flexible branches with rotatable bonds not suitable for our rigid-geometry based CSP approach. The remaining 318 linkers were used to *in silico* synthesize new [4+6] cage molecules by combining the SMILES strings encoding the building blocks. The 3D coordinates of the cages were then generated from the cage SMILES identifier by using the LigPrep tool of the Schrodinger's Material Science suite.<sup>23</sup> Next, the low(est) energy conformer search was carried out with a tool from the same suite coupled with the OPLS\_2005<sup>24</sup> force field. Further details about LigPrep and the conformational sampling procedure can be found in Ref 21. We retained only molecules with achiral linkers and whose lowest energy conformer had the requested tetrahedral symmetry. The molecular symmetry point group was verified with the proper tool of the Material Science suite. The final set of six linkers is presented in Figure 5-2 and the corresponding *in silico* synthesized porous cages are referred-to as M1-M6.

Axial chirality is an inherent property of the imine [4+6] tetrahedral cages and is determined by the chirality of the starting linker. As a consequence, for achiral linkers the lowest energy conformer state is degenerate and two isoenergetic enantiomers, respectively R and S, can be found (Figure 5-S1 in the ESI). In fact, in the case of cage CC1 spontaneous interconversion of the enantiomers R and S was experimentally observed.<sup>6,15</sup>

The lowest energy R-conformers of cages M1-M6 were subsequently relaxed at the density functional theory (DFT) level in vacuum, using the M06-2X functional and the 6-311G\*\* basis set implemented in the Jaguar software part of the Schrodinger's MaterialScience suite.<sup>23</sup> This level of the theory has been already successfully used for other organic porous cages.<sup>4,5</sup> The DFT-optimized conformers of M1-M6 confirmed the expected tetrahedral symmetry with four windows of equal size connecting the internal pore with the outer space.

The molecular porosity the DFT optimized M1-M6 was characterized with the largest cavity diameters (LCD) and the windows size (WS). The former is the size of the largest sphere that can occupy the internal cavity and the latter is the diameter of the largest spherical probe that can access the internal pore through the cage windows. Both LCDs and WSs were computed with our in-house tool<sup>25,26</sup>, which involves computational geometry approaches similar to ones we implemented to characterize porosity in periodic material structures in Zeo++ code.<sup>27</sup>

## 2.2 Crystal structure prediction (CSP) and porosity characterization

In order to investigate the crystal phases of cages M1-M6 a crystal structure prediction (CSP) study was executed. Since the lowest energy conformers of molecules M1-M6 can be found in two isoenergetic enantiomers, R and S respectively, the CSP was limited to the most common 6 space groups for enantiopure crystal structures ( $P1$ ,  $P2_1$ ,  $P2_12_12$ ,  $P2_12_12_1$ ,  $C222_1$ ,  $C2$ ) and 6 groups for racemic crystals ( $P2_1/c$ ,  $P\bar{1}$ ,  $C2/c$ ,  $Cc$ ,  $Pna2_1$ ,  $Pbca$ ) which

account for about 90% of the observed crystal structures for organic molecules. CSP calculations were performed using the UPACK<sup>28</sup> (Utrecht Crystal Package) program that perform global lattice energy minimization using molecular force fields.

Solvent-free initial candidate structures were generated using the random search technique as implemented in UPACK with  $Z'=1$  molecule in the asymmetric unit. During the random search, the molecule was kept rigid at its DFT geometry while Lennard-Jones and Coulomb intermolecular parameters taken from the OPLS\_2005 force field were used to initially minimize the lattice energy by optimizing the unit cell vectors, angles and relative molecular positions and orientations. A number of 5000 trial structures were generated in each space groups.

All crystal structures generated were further optimized using an improved model for the lattice energy. In particular, dispersion-repulsion interactions were modelled using the Buckingham potential with the FIT<sup>29</sup> parameters and isotropic electrostatic interactions described with the Coulomb potential using point charges. The latter were obtained by fitting to the molecular electrostatic potential (ESP) of the M06-2X/6-311G\*\* charge density calculated with Jaguar. All the interatomic interactions were calculated using the Ewald summation technique with a cutoff of 20 Å in real space. Equivalent structures resulting from different initial trial structures were grouped together by using a clustering algorithm as implemented in UPACK. The final structures were ranked w.r.t the energy of the global minimum.

Since molecules M1-M6 have tetrahedral symmetry, the structures generated in the initial space groups could belong to a space group with higher symmetry. The full symmetry was therefore determined with PLATON.<sup>30</sup> The porosity of each of the predicted structures was characterized by computing the pore limiting diameter (PLD). The calculation involved

our code Zeo++ in its high accuracy mode.<sup>37</sup> The collected results were used to build the energy-structure-porosity maps by augmenting the energy landscape with PLDs.

### **2.3 Refinement and validation.**

To further assess the accuracy of the energy ranking and stability predicted by our CSP approach, a subset of predicted structures was further optimized with a DFT-D3 periodic solid state approach with no symmetry constraints. Calculations were performed with the QuickStep module of CP2K<sup>31</sup> with PBE/TZVP-MOLOPT<sup>32</sup> functional and basis set, D3 Grimme dispersion correction<sup>33</sup> and an energy cutoff of 600 Ry.

The thermal stability of selected structures was further investigated by means of molecular dynamic (MD) simulations. Supercells of approximately 200 molecules, build by replicating in space the predicted unit cell, were submitted to MD simulations in the NPT ensemble at T=300K and P=1atm to verify if they maintain the structure over the period of the simulation (t=1ns). MD simulations were performed with the GROMACS<sup>34</sup> package, temperature and pressure were kept constant with the Berendsen thermostat and barostat respectively. The same force field as in the global minimization CSP step was adopted for nonbonded interactions and intramolecular terms were assigned with the OPLS\_2005 force field.

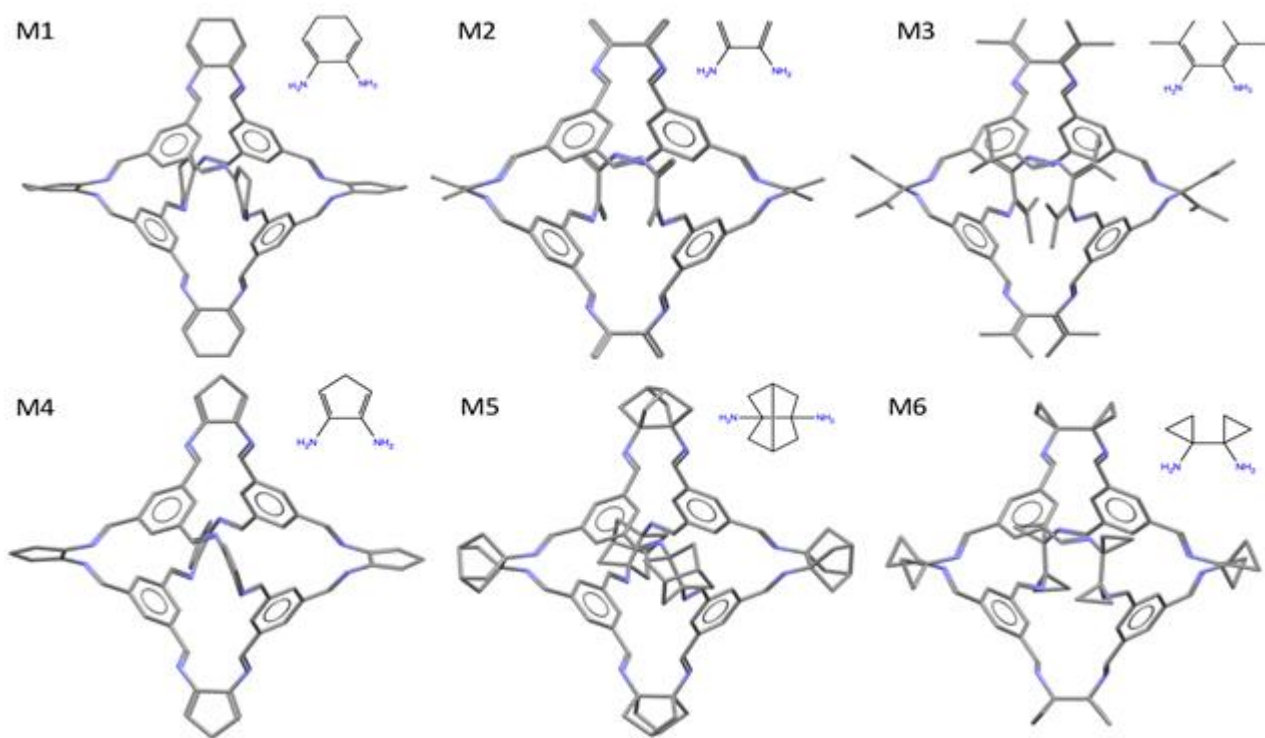
#### **Validation of the computational approach.**

Our CSP procedure was validated by predicting the crystal landscape of CC3, for which extensive prior experimental and CSP data is available. Our approach correctly predicts both racemic and enantiopure polymorphs with the correct energy ranking. The details of this validation are presented in the ESI section S2.

## 3. Results

### 3.1 New porous organic cages.

The set of six linkers obtained from our PubChem3D screening (Fig. 5-2) was used to build molecular models of six new porous organic [4+6] cages named M1-M6 (Fig. 5-3). M1-M6 have the same core molecular structure but different vertex functionalization. The final DFT-refined R-conformers of M1-M6 exhibit the correct tetrahedral symmetry and the corresponding largest cavity diameters (LCDs) and windows size (WSs) are collected in Table 5-S1. An important remark regards our assumption that our selected linkers with a 1,3,5-triformylbenzene (TFB) precursor ( see Figure 5-1) will lead to the formation of a shape persistent [4+6] cage. Predicting the formation of a cage from first principle is in fact still a great computational challenge and the final topology or the feasibility of the reaction cannot be simply guessed from the chemistry and geometry of the precursors. It has been in fact demonstrated that some reactions involving different diamines can lead to the formation of cages of different topology from the [4+6], e.g. [2+3] or [8+12].<sup>12,35</sup> Even if the formation of [2+3] cages it was proven not to be energetically favourable when vicinal diamines react with TFB<sup>35</sup>, it was experimentally observed that a particular diamine with the TFB precursor surprisingly led to the formation of a [8+12] cage in solution.<sup>12</sup> The latter however collapsed upon desolvation yielding a non -porous amorphous material.<sup>12</sup>



**Figure 5-3.** *In silico* designed porous organic cages studied in this work. Grey and blue atoms represent nitrogen and carbon, respectively. Hydrogen atoms are omitted for clarity. The LCD for M1, M2, M3, M4, M5 and M6 are 4.95 Å, 5.30 Å, 5.36 Å, 4.97 Å, 4.78 Å and 5.21Å respectively. All the molecules have tetrahedral symmetry and 4 windows of equal size.

### 3.2 Energy-structure-porosity maps.

The energy-structure-function maps for M1-M6, depicted in Fig. 5-4, present the relative lattice energy of each predicted phase, within the energy range of 70 kJ/mol w.r.t the global minimum, versus the corresponding phase density and its PLD value used as the point color determiner. These maps enable a comprehensive analysis of the M1-M6 predicted crystal phases as follows.

The lowest energy structures of M1-M6 phases, i.e. the structures with zero relative lattice energy, are referred to as  $\alpha$ -phases and depicted in Figure 5-S4. The  $\alpha$ -phases of M1-M6 are porous to  $H_2$  according to the corresponding value of the PLD larger than the size of a  $H_2$  probe of 2.18Å.

In the case of M4 and M5 phases, the global minimum is approximately 40 kJ/mol and 20 kJ/mol, respectively, more stable than the next respective most stable structure, while in the other cases hundreds of structures are found in an energy range of 40 kJ/mol. This number of structures in this energy range is often observed for organic molecules and cages, and suggest the possible presence of structural polymorphism, i. e. the existence of several crystal phases with different packing motifs and accessible with crystal engineering procedures.

The lowest energy predicted structures (i.e.  $\alpha$ -phases) are all racemates (R,S). Further inspection of the  $\alpha$ -phases (ESI Figure 5-S4) reveals the role of the linker in promoting certain packing arrangement over others

For planar and cyclic linkers such as M1 and M4, the global minimum is isostructural to the racemic  $\alpha$ -phase of CC3. These molecules show preference for a window-to-window packing with a tetrahedral porous network connecting the cage voids and PLDs equal to 3.1 Å and 3.14 Å respectively. Initially predicted in space group  $Ci$ , both structures show the full cubic symmetry  $Fd\bar{3}$ . In the case of M4 such preference is stronger since the second structure is found 40kJ/mol above the  $\alpha$ -phase.

A cubic  $Fd\bar{3}$  symmetry was also observed for the  $\alpha$ -phase of molecule M5. In this case however the globular-like diamine linker promotes an arene-to-arene packing. An interconnected porous network is still present but the PLD is 25% smaller when compared with the M1 and M4 equivalents.

The  $\alpha$ -phase of cage M2 shows a packing motif with molecules stacked along the  $a$ -axis of the unit cell with one window of a cage directed towards a neighbour arene face. The other three windows connect external voids with the internal pore of the cage but are partially occluded by the neighbors vertexes. The PLD is indeed 2.8 Å, smaller than the size of the

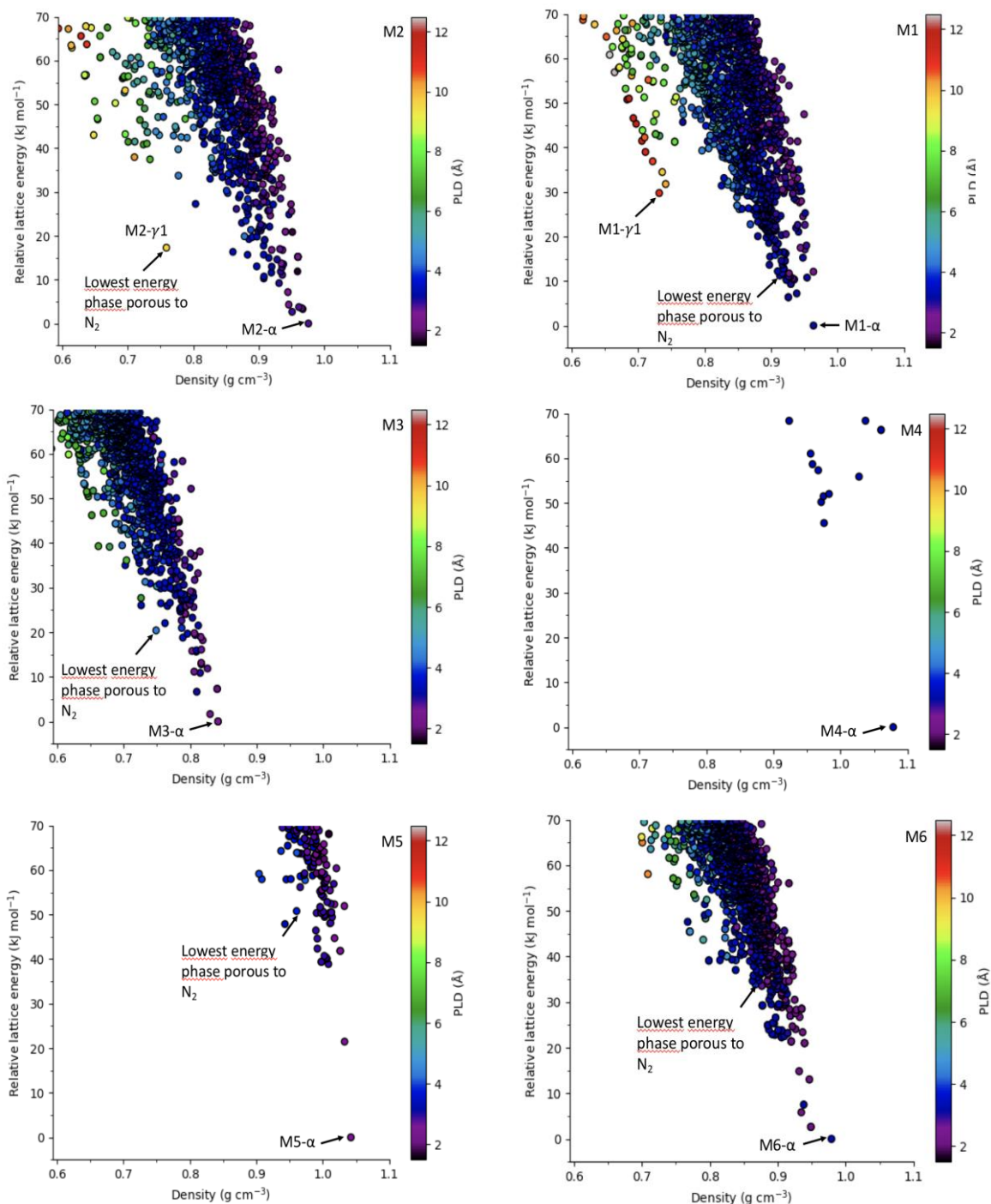
window. No special packing motif could be observed for the global minimum structures of molecule M3 and M6.

Further analysis of the energy-structure-porosity maps in Figure 5-4 reveals that the lowest energy phase porous to N<sub>2</sub>, is found in

the leading edge for all the systems but M4, for which none of the structures in a range of 70kJ/mol from the global minimum is found porous to nitrogen. We remind that experimental accessible structures are indeed often found in the leading edge of the energy landscape. The fraction of predicted structures at least porous to N<sub>2</sub> for molecules M1, M2, M3, M4, M5 and M6 is 37.5 %, 29.8 %, 55%, 0.0 %,8.8 % and 17.6 % respectively. The pore limiting diameters characterize the static porosity and do not account for the phenomenon known as “porosity without pores”, in which structures become dynamically porous due small flexible deformations of the cages induced by thermal fluctuations or to host-guest cooperative effects.<sup>36,37</sup>

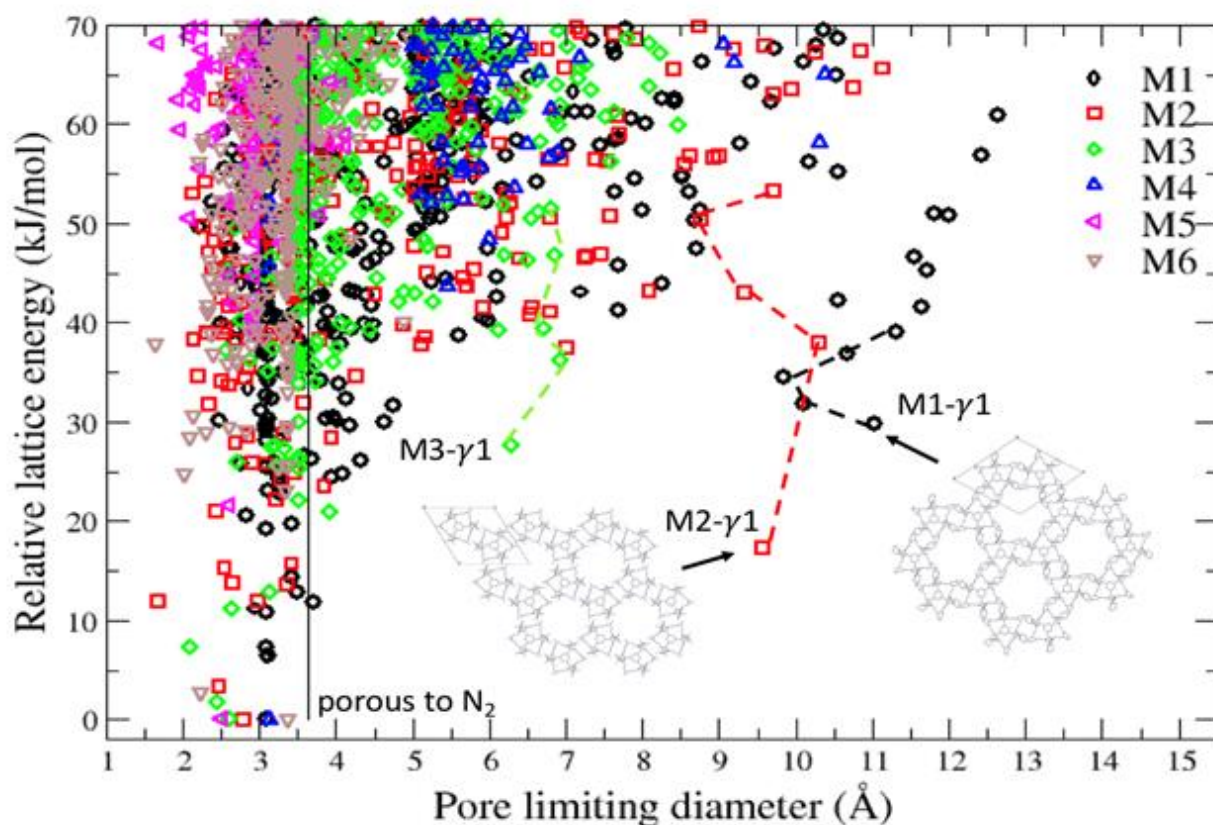
### **New highly-porous crystal phases.**

The energy landscapes of M1 and M2 present intriguing features, which have never-before been reported for imine cages. In particular, low-density racemic polymorphs emanate from the leading edges of the plots as local minima of highly-porous branches. We will refer to the latter as  $\gamma$ -branches. The structures belonging to the  $\gamma$ -branches are referred to as  $\gamma$ -phases. The minima of the M1- $\gamma$  and M2- $\gamma$  branches are labelled as M1- $\gamma$ 1 and M2- $\gamma$ 1 in Figure 5-4, with densities of 0.76 and 0.72 g/cm<sup>3</sup> respectively and large 1-dimensional porous channels with PLD around 10 Å.



**Figure 5-4.** Energy-structure-porosity maps for M1-M6. The lowest energy phases porous to N<sub>2</sub> are indicated. The  $\alpha$ -phases and the phases M1-  $\gamma$ 1 and M2-  $\gamma$ 1 are labelled and indicated. Each symbol is colour coded according to the PLD value of the corresponding crystal structure.

Their local minima character is confirmed when the landscape is plotted versus the PLD, represented in Figure 5-5 in which the five most stable  $\gamma$ -phases of both M1- $\gamma$  and M2- $\gamma$  branches are connected by dashed lines. By inspecting Figure 5-5 we notice that a new branch of porous structures, not visible from the plot in Figure 5-4, also appears in the case of M3 when the relative lattice energy is plotted versus the PLD. The corresponding local minima is called M3- $\gamma$ 1 and it's PLD takes the value of 6.29 Å.



**Figure 5-5.** Pore limiting diameter (PLD) as a function of the relative lattice energy of the corresponding predicted structures for molecules M1-M6. The vertical line at 3.64 Å correspond to the diameter of a N<sub>2</sub> probe. Dashed lines connect the  $\gamma$ -phases mentioned in the main text. The local minima of M1- $\gamma$ , M2- $\gamma$  and M3- $\gamma$  branches are labelled and indicated.

Both M1- $\gamma$ 1 and M2- $\gamma$ 1 show the full symmetry of  $P\bar{3}$ . In M1- $\gamma$ 1 and M2- $\gamma$ 1, cages are stacked in a window-to-arene packing motif with neighbour cages engaging C(*sp*<sup>3</sup>)H---N<sup>38</sup> and C(*sp*<sup>2</sup>)H--- $\pi$ <sup>39</sup> (ESI Figure 5-S5) weak hydrogen bonds respectively, thus promoting the

formation of 1-dimensional channels exhibiting zeolite-range porosity with PLD of ca. 10 Å. The M3- $\gamma$ 1 structure shows the  $P\bar{1}$  symmetry featuring 1-dimensional porous channels with PLD of ca 6.3 Å but no special packing motif is observed. Given that the M3- $\gamma$  branch is not as distinct as M1- $\gamma$  and M2- $\gamma$  as well as M3- $\gamma$  1 is less porous, this branch is not given further attention in this work.

### **Verification of CSP results.**

The main findings of the CSP approach based on classical force field and rigid molecule representations are further verified with solid-state DFT-D3 calculations as detailed in the computational methods section.

The five lowest energy structures w.r.t the predicted global minimum, i.e. rank 1 to 5, were further relaxed at DFT-D3 level. Comparison between the energy ranking at CSP and DFT-D3 level is presented in Table S4 in the ESI.

In the case of molecule M3, M4 and M5 the predicted  $\alpha$ -phases are confirmed to be global minima even at DFT-D3 level of the theory. The corresponding PLDs decrease of about 30-40 % in the case of M3- $\alpha$  and M4- $\alpha$  corresponding to a decrease in density of about 15 % in both cases. On the other hand, for M5- $\alpha$  the PLD increase of about 1 % being still porous to H<sub>2</sub> and the density is essentially constant.

Energy re-ranking is observed for the lowest energy structures of molecules M1, M2 and M6. Predicted structures rank 3 (M1), 5 (M2) and 5 (M6) become the new minima at the DFT-D3 level. The new DFT relaxed  $\alpha$ -phases are still H<sub>2</sub> porous in the case of M1 and M2. Interestingly, the new minimum of M6 is not racemic but enantiopure.

The predicted polymorphs M1- $\gamma$ 1 and M2- $\gamma$ 1 were also further optimized at the DFT-D3 level to assess their stability and investigate the possibility that they represent artefacts of the approximations used in the CSP approach. More precisely, the five most stable  $\gamma$ -phases of both M1- $\gamma$  and M2- $\gamma$  branches were selected in this task. They correspond to the

structures connected by, respectively, black and red dashed line in Figure 5-5, and are denoted as M1(2)- $\gamma$ X, where X=1...5 means increasing energy w.r.t the lowest energy phase. The CSP and DFT-D3 results for the  $\gamma$ -phases of M1(2) are collected in Table 5-3.

In the case of M1 we observe that M1- $\gamma$ 1 is still a local minimum of the highly-porous branch of the energy landscape at the DFT-D3 level and its symmetry is preserved. Ranking inversion is observed for phases M1- $\gamma$ 4 and M1- $\gamma$ 5. The overall highly-porous character of this branch of the landscape is maintained. A small decrease in the PLDs is observed as consequence of an increase of density at the DFT level. Similarly, M2- $\gamma$ 1 phase still preserves its local minima character after DFT-D3 and its symmetry. No energy re-ranking is observed except for the M2- $\gamma$ 5 which was indeed a saddle point of the predicted landscape and collapses to the high-density/low-porous region of the landscape after the DFT-D3 optimization. The PLDs of M2- $\gamma$ 1, M2- $\gamma$ 2, M2- $\gamma$ 3 and M2- $\gamma$ 4 experience a small decrease upon DFT-D3 relaxation.

The large PLD of M1- $\gamma$ 1 and M2- $\gamma$ 1 is due to the presence of large porous 1-D channels, extrinsic to the cages. Similar feature was experimentally observed for CC2 (CSD RefCode PUDXAO in Table 5-1) although its PLD (3.9 Å) is dwarfed by the PLDs reported here. Although the vertexes of the cages in M1- $\gamma$ 1 and M2- $\gamma$ 1, by occluding the windows of the neighbour cages, hinder the formation of guest-accessible path connecting the internal cage pores with the main 1-D channels, it was demonstrated for CC2 that cooperative mechanisms can promote the formation of such transitory connections thus improving the porosity of the material.<sup>37</sup>

Thermodynamic stability of M1- $\gamma$ 1 and M2- $\gamma$ 1 phases was investigated by the means of MD simulations (at T = 300K and P=1atm). Inspection of the snapshots at the beginning and at the end of 1ns simulations revealed only small fluctuations w.r.t the predicted

structures due to thermal effects. The overall symmetry and porosity of the structures were maintained (Figure 5-S6 in ESI).

To summarize, the presence of M1- $\gamma$ 1 and M2- $\gamma$ 1 phases in a separate low-density branch of the energy landscape, the results of our MD thermodynamic stability as well as their acceptably-low DFT-D3 energies w.r.t. lowest energy minima makes these highly-porous phases good candidates to be experimentally achievable by proper crystal engineering procedures<sup>21</sup>.

Structure	space group	CSP $\Delta E$ (kJ/mol)	CSP PLD( $\text{\AA}$ )	CSP density(g/cm <sup>3</sup> )	DFT-D3 $\Delta E$ (kJ/mol)	DFT-D3 PLD( $\text{\AA}$ )	DFT-D3 density(g/cm <sup>3</sup> )
M1- $\gamma$ 1	$P\bar{3}$	29.77	11.03	0.73	20.38(46.04)	10.28	0.78
M1- $\gamma$ 2	$P\bar{1}$	31.80	10.11	0.74	32.32(57.98)	9.90	0.77
M1- $\gamma$ 3	$P2_1/c$	34.48	9.83	0.73	35.89(61.55)	7.93	0.84
M1- $\gamma$ 4	$Cc$	36.90	10.67	0.72	43.35(69.01)	8.58	0.80
M1- $\gamma$ 5	$P2_1/c$	39.00	11.30	0.71	37.01(62.67)	9.74	0.79
Structure	space group	CSP $\Delta E$ (kJ/mol)	CSP PLD( $\text{\AA}$ )	CSP density(g/cm <sup>3</sup> )	DFT-D3 $\Delta E$ (kJ/mol)	DFT-D3 PLD( $\text{\AA}$ )	DFT-D3 density(g/cm <sup>3</sup> )
M2- $\gamma$ 1	$P\bar{3}$	17.29	9.56	0.76	34.44(43.24)	8.79	0.80
M2- $\gamma$ 2	$P2_1/c$	37.97	10.30	0.71	54.96(63.76)	9.39	0.76
M2- $\gamma$ 3	$P\bar{1}$	43.00	9.33	0.65	70.32(79.12)	9.10	0.67
M2- $\gamma$ 4	$C2/c$	50.30	8.77	0.68	74.34(83.14)	8.71	0.71
M2- $\gamma$ 5	$P2_1/c$	53.20	9.71	0.70	8.64(17.44)	1.70	1.04

**Table 5-3.** Comparison between CSP and DFT-D3 results for the first five  $\gamma$ -phases of the M1- $\gamma$  and M2- $\gamma$ branches.  $\Delta E$  are given w.r.t the energy of the  $\alpha$ -phases at CSP and DFT-D3 level. Additionally,  $\Delta E$  at the DFT-D3 level, are also given w.r.t. to the lowest energy phase identified at this level of theory and placed in () brackets.

## 4. Conclusions

We present a three-step approach enabling computational design of porous crystalline phases based on molecular cages. The steps are: (1) database screening to identify suitable building blocks corresponding to the desired chemistry, followed by *in silico* synthesis of cage molecules; (2) crystal structure prediction for the molecules of step (1) based on classical potential, followed by porosity characterization of the phases; (3) refinement of the identified critical phases at the DFT level, followed by the phase stability verification with molecular dynamics.

We demonstrate an application of the approach, in which structural analogues of the well know family of the Cooper [4+6] imine cages<sup>1,2</sup> are designed. In particular, our search of PubChem3D database of ca 80 million structures led six new vicinal diamines, and the corresponding six porous cages, which exhibit the expected tetrahedral symmetry. Their rigid structure was used as input for a crystal structure prediction study, followed by the porosity characterization of each predicted phase. Analysis of the resulting energy-structure-porosity maps revealed the existence of structures featuring rare high porosity, which manifested itself in pore limiting diameters in the excess of 10 Å. The stability of these new highly-porous polymorphs as local minima on the local energy surface was further confirmed with solid state DFT-D3 optimizations and molecular dynamics simulations.

Our computational approach constitutes the first step towards high-throughput screening of POCs for assembly of functional materials. We envision that in the near future our approach can be augmented with solvent stabilisation calculations, which can provide further insight on the experimental accessibility of the predicted  $\gamma$ -phases, as recently prototyped for cage CC1.<sup>40</sup>

## 5. Notes and references

### Conflicts of interest

There are no conflicts of interest to declare

### Acknowledgements

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. A. I. Cooper, ACS. Cent. Sci. 2017, 3, 544.
2. Hasell T.m A. I. Cooper Nature Reviews Materials 2016, 1. Available from: <http://dx.doi.org/10.1038/natrevmats.2016.53>
3. G. Zhang, O. Presly, F. White, I. M. Opper, M. A. Mastalerz Angew. Chem. Int. Ed. 2014, 53, 1516.
4. V. Santolini, M. Miklitz, E. Berardo, K. E. Jelfs Nanoscale 2017, 9, 5280.
5. A.G. Slater, M. A. Little, A. Pulido, S. Y. Chong, D. Holden, L. Chen et al. Nature Chemistry 2017, 9, 17.
6. T. Tozawa, J. T. Jones, S. I. Swamy, S. Jiang, D. J. Adams, S. Shakespear et al. Nature Materials 2009, 8, 973.
7. J. T. A. Jones, T. Hasell, Xiaofeng Wu, J. Bacsa, K. E. Jelfs, M. Schmidtman et al. Nature 2011, 474, 367.
8. T. Mitra, Xiaofeng Wu, R. Clowes, J. T. A. Jones, K. E. Jelfs, D. J. Adams et al. Chem. Eur. J. 2011, 17, 10235.

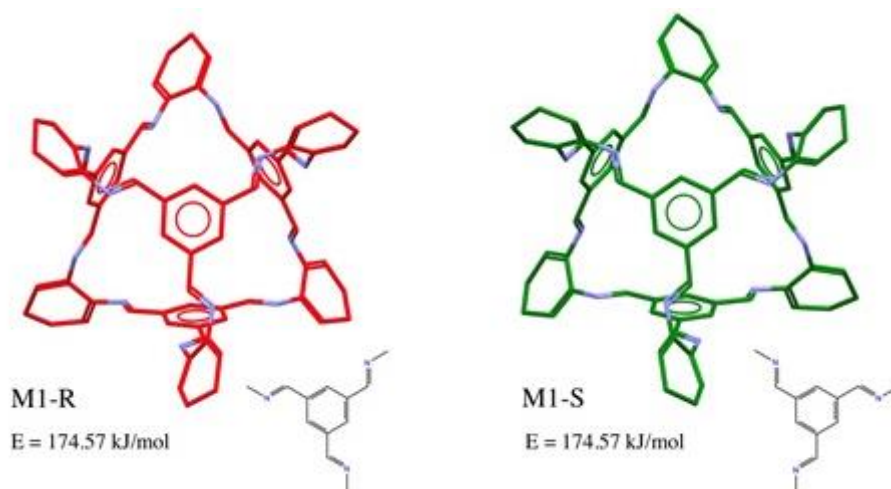
9. M. J. Bojdys, M. E. Briggs, J. T. A. Jones, D. J. Adams, S. Y. Chong, M. Schmidtmann, A. I. Cooper *J. Am. Chem. Soc.* 2011, 103, 16566.
10. M. Mastalerz, M. W. Schneider, I. M. Oppel, O. Presly *Angew. Chem. Int. Ed.* 2014, 50, 1046.
11. H. Ding, Y. Yang, B. Li, F. Pan, G. Zhu, M. Zeller, D. Yuan, C. Wang *Chem. Commun.* 2015, 51, 1976.
12. P. S. Reiss, M. A. Little, V. Santolini, S. Y. Chong, T. Hasell, K. E. Jelfs, M. Briggs, A. I. Cooper *Chem. Eur. J.* 2016, 22, 1.
13. T. Hasell, S. Y. Chong, K. E. Jelfs, D. J. Adams and A. I. Cooper *J. Am. Chem. Soc.* 2012, 134, 588.
14. E. O. Pyzer-Knapp, H. P. G. Thompson, F. Schiffmann, K. E. Jelfs, S. Y. Chong, M. A. Little, A. I. Cooper and G. M. Day. *Chem. Sci.* 2014, 5, 2235.
15. J. T. A. Jones, D. Holden, T. Mitra, T. Hasell, D. J. Adams, K. E. Jelfs, A. Trewin, D. J. Willock, G. M. Day, J. Bacsá, A. Steiner, A. I. Cooper, *Angew. Chem. Int. Ed.* 2011, 123, 775.
16. M. A. Little, Samantha Y. Chong, M. Schmidtmann, T. Hasell, A. I. Cooper, *Chem. Commun.* 2014, 50, 9465.
17. J. Nyman and G. M. Day, *CrystEngComm* 2015, 17, 5154.
18. G.M. Day, *Crystallogr. Rev.*, 2011, 17, 3.
19. A. J. Cruz-Cabeza, G. M. Day, W. Jones, *Chem. Eur. J.* 2009, 15, 13033.
20. J. D. Evans, K. E. Jelfs, G. D. Day, C. J. Doonan *Chem. Soc. Rev.* 2017, 46, 3286 and references therein.
21. A. Pulido, L. Chen, T. Kaczorowski, D. Holden, M. A. Little, S. Y. Chong, B. J. Slater, D. P. McMahon, B. Bonillo, C. J. Stackhouse, A. Stephenson, C. M. Kane, R. Clowes, T. Hassell, A. I. Cooper and G. M. Day, *Nature* 2016, 543, 657.

22. E. E. Bolton, J. Chen, S. Kim, L. Han, S. He, W. Shi, V. Simonyan, Y. sun, P. A. Thiessen, J. Wang, B. Yu, J. Zhang, S. H. Bryant, *Journal of Cheminformatics*. 2011, 3, 32.  
URL: <https://pubchem.ncbi.nlm.nih.gov/release3d.html>
23. Material Science Suite 2017-2, Schrödinger, LLC, New York, NY, 2017.
24. J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Fetls, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio and R. M. Levy, *J. Comp. Chem.* 2005, 26,1752.
25. I. Gomez Garcia, M. Bernabei, R. Perez Soto and M. Haranczyk, *Crystal Growth & Design*. 2017, 17, 564.
26. I. Gomez Garcia, M. Bernabei, M. Haranczyk *Journal of Chemical theory and Computation* 2018, submitted.
27. M. Pinheiro, R.L. Martin, C.H. Rycroft, M. Haranczyk, *CrystEngComm* 2013, 15 7531; T. F. Williams, C. H. Rycroft, M. Kazi, J. C. Meza, M. Haranczyk, *Mesoporous Materials* 2012, 149, 134; URL: [www.zeoplusplus.org](http://www.zeoplusplus.org).
28. B. P. Van Eijck, J. Krooj, *Acta Cryst. B* 2000, 56, 535.
29. D. S. Combes, S. L. Price, D. J. Willock, M. Leslies, *J. Phys. Chem*, 1996, 100, 7352.
30. A. L. Speck, PLATON, a multipurpose crystallographic tool, Utrecht University, Utrecht, The Netherlands, 2001.
31. J. Van de Vondede, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, 167, 103.
32. J. Van de Vondede and J. Hutter, *J. Chem. Phys.*, 2007, 127, 114105.
33. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, 132, 154104.
34. D. Van Der Spoel, E. Lindhal, B. Hess, G. Groenho, A. E. Mark and H. J. C. Berendsen, *J. Comp. Chem.* 2005, 26, 1701.
35. K. E. Jelfs, E. G. B. Eden, J. L. Culshaw, S. Shakespeare, E. O. Pyzer-Knapp, H. P. G. Thompson *J. Am. Chem. Soc.*, 2013, 135, 9307.

36. P. K. Thallapally, T. B. Wirsig, L. J. Barbour and J. L. Atwood, *Chem. Commun.* 2005, 35, 4420
37. D. Holden, S. Y. Chong, L. Chen, K. E. Jelfs, T. Hasell and A. I. Cooper *Chemical Science* 2016, 7, 4875.
38. M. Mazik, D. Blaser and R. Boese, *Tetrahedron Letters*. 2000, 41, 5827.
39. H. Suezawa, S. Ishihara, Y. Umezawa, S. Tsuboya and M. Nishio, *Eur. J. Org. Chem.* 2004, 2004, 4816.
40. D. P. McMahon, A. Stephenson, S. Y. Chong, M. A. Little, J. T. A. Jones, A. I. Cooper, G. M. Day *Faraday Discuss.*, 2018, Advance Article. DOI: 10.1039/C8FD00031J

## 6. Supporting information

### 6.1 Porosity characterization of the lowest energy conformers of molecules M1 -M6.



**Figure 5-S1.** Lowest energy conformers of cage M1. Both enantiomers M1-R and M1-S have the same energy and tetrahedral symmetry. Nitrogen atoms are colored in blue, carbon atoms in red and green for M1-R and M1-S respectively and hydrogen atoms are omitted for clarity

	LCD (Å) DFT- M06- 2X	Window size (Å) DFT- M06-2X
CC3- R	5.50	3.77
M1- R	4.94	3.20
M2- R	5.31	3.52
M3- R	5.36	3.60
M4- R	4.97	3.24
M5- R	4.78	3.06
M6- R	5.21	3.48

**Table 5-S1.** Largest cavity diameters and window size for the lowest energy R-conformers of cage CC3 and molecules M1-M6 after DFT optimizations

## 6.2 CSP validation: the benchmark covalent cage CC3

Cage CC3 is probably the most studied cage in the imine cage family with a LCD of 5.50 Å and four windows of equal size of 3.75 Å. The enantiomer CC3-R was synthesized<sup>1</sup> from a mixture of 1,3,5-triformylbenzene and chiral (*R,R*)-1,2-diaminocyclohexane (see Figure 5-1 in the manuscript for a schematic representation). Similarly, the energetically equivalent enantiomer CC3-S can be made by a mixture of 1,3,5-triformylbenzene and chiral (*S,S*)-1,2-diaminocyclohexane. Applications of porous materials based on cage CC3 in its homochiral polymorph CC3-R(S)- $\alpha$ , are for example: Xenon/Krypton separation<sup>2</sup>, enantiomeric and molecular size separation<sup>3</sup>, water desalination<sup>4</sup>.

A 3D model of CC3 molecule was initially generated with LigPrep starting from a 1D SMILES corresponding to the enantiomer CC3-R(S) and submitted to a conformational search procedure.

The experimentally observed enantiomer CC3-R(S) was correctly predicted as the lowest energy conformer and further optimized in vacuum at the M06-2X/6-311G\*\* level of theory.

The DFT optimized molecular geometry was used as input for the CSP study. Homochiral crystal phases based on CC3-R were searched in the most common 6 space groups for enantiopure crystal structures (*P1*, *P2<sub>1</sub>*, *P2<sub>1</sub>2<sub>1</sub>2*, *P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>*, *C222<sub>1</sub>*, *C2*). Similarly, racemic co-crystals CC3-(R,S) were generated in the most common 6 space groups for racemates (*P2<sub>1</sub>/c*, *P $\bar{1}$* , *C2/c*, *Cc*, *Pna2<sub>1</sub>*, *Pbca*). More details on the CSP procedure are provided in the method section in the main text.

The resulting energy landscape, i.e. the plot of relative lattice energy of CC3 as a function of density, is shown in Figure 5-S2.

The energy landscape of CC3 has been already investigated by Day's group employing a Monte Carlo simulated annealing for initial structure generation and a subsequent final

minimization procedure was carried out with the DMACRYS software for rigid molecules, which make use of anisotropic atom-atom model potentials.<sup>5</sup>

Our results are in quantitative agreement with the results of Day's group even using an isotropic Coulomb potential for the electrostatic interactions. Only 12 structures are found in a 50 kJ/mol range from the global minimum, and five of them, color-filled circles in Figure 5-S2, feature a tetrahedral porous network connecting the cavities of the cages.

The global minimum, black filled point in Figure 5-S2, corresponds to the racemic co-crystal cubic phase CC3-(R,S) $\alpha$ . The latter is not available in the Cambridge structure database since the structure and unit cell parameters were refined from PXRD experiments<sup>6</sup> therefore, a direct geometric comparison with the predicted phases is not possible.

The homochiral phase CC3-R $\alpha$ , yellow point in Figure 5-S2, is correctly predicted as the global minimum of the structures generated in the six enantiopure space groups. The latter, initially predicted in the space group  $P2_12_12_1$  shows the full experimental symmetry  $F4_132$  and is a good geometrical match to the experimental structure PUDXES in Table 5-1.

Due to the large size of the imine cages our approach is limited to generating structures with  $Z'=1$  molecules in the asymmetric unit, therefore the polymorph PUDXES02 of CC3, also known as CC3-R $\beta$  phase (see Table 5-1) was not predicted since it was experimentally observed in the space group  $P3$  with  $Z'=3$ .

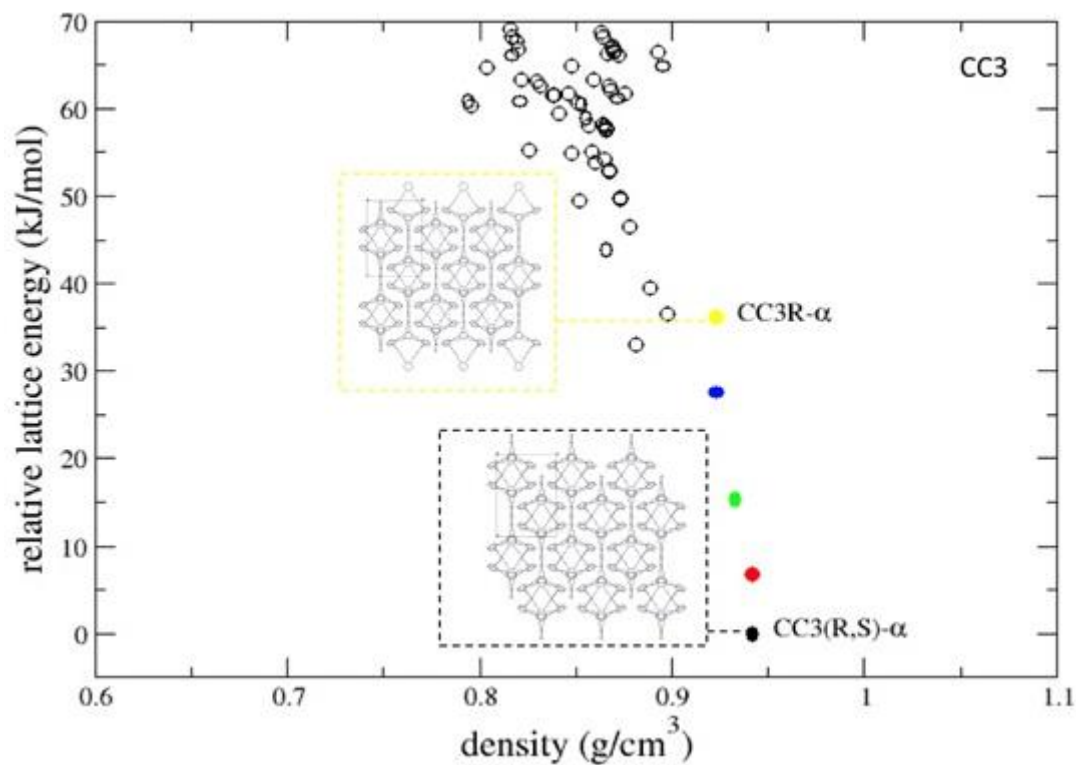
Although our CSP approach does not account for solvent inclusion, the structures of solvated polymorphs can still be predicted as local minima in the energy landscape even if they incur structural rearrangement upon solvent removal.

The solvated framework of PUDXES02, i.e. CC3-R $\beta$ -solvated with CSD-Refcode NODVIN, was experimentally observed in space group  $R3$  with  $Z'=1$  cage in the asymmetric unit.

The structure corresponding to the CC3-R $\beta$ -solvated framework (not shown in Figure 5-S2) showed indeed the correct full  $R3$  symmetry and was predicted to be a high energy structure found approximately 50 kJ/mol above the CC3-R $\alpha$  phase.

Structure overlay with experimental structures and comparison of the cell parameters for the predicted phases CC3-R $\alpha$  and CC3-R $\beta$ -solvated are presented in Figure 5-S3 and Table 5-S2, respectively.

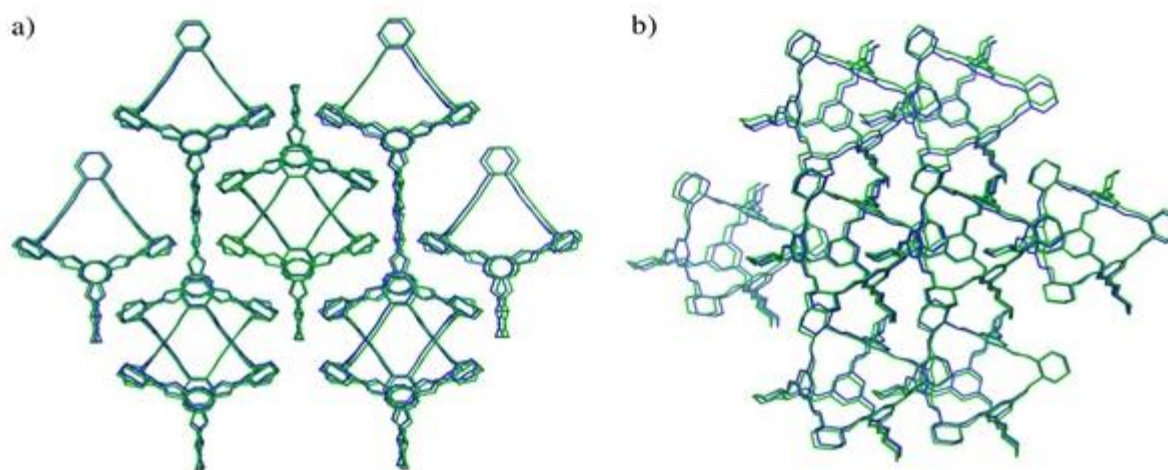
Finally, to validate the accuracy of our energy model in predicting the landscape of porous organic cages we made use of DFT-D3 solid state periodic calculations to fully relax the lowest energy structures of the landscape. Results for cage CC3 are listed in Table 5-S3. Although the relative energy of the phase CC3-R $\alpha$  increase of about 50 % after DFT-D3 relaxation, the global energy ranking is preserved and porosity, characterized by the PLD, is maintained.



**Figure 5-S2.** Crystal energy landscape of CC3. Energies are w.r.t the global minimum. Both homochiral and racemic polymorphs are correctly predicted as the lowest energy structures in the enantiopure and racemic space groups, respectively.

		Cell axes length (Å)			Cell angles			Density (g/cm <sup>3</sup> )
Material	Method	a	b	c	$\alpha$	$\beta$	$\gamma$	
CC3-R $\alpha$ ( <i>F</i> <sub>4132</sub> )	Exp.	24.8 0	24.8 0	24.80	90.0 0	90.00	90.00	0.97
	CSP	25.2 4	25.2 4	25.24	90.0 0	90.00	90.00	0.92
CC3-R $\beta$ - solv ( <i>R</i> <sub>3</sub> )	Exp.	25.6 3	25.6 3	11.18	90.0 0	90.00	120.00	0.87
	CSP	26.3 6	26.3 5	11.39	90.0 3	90.01	120.00	0.81

**Table 5-S2.** Cell parameters from our CSP study compared to the experimental values for CC3-R $\alpha$  (CSD-Refcode PUDXES) and CC3-R $\beta$ -solvated (CSD-Refcode NODVIN). Both CC3-R $\alpha$  and CC3-R $\beta$ -solvated were initially predicted in space groups P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> and P1 respectively. Full symmetry was revealed after analysis with PLATON.

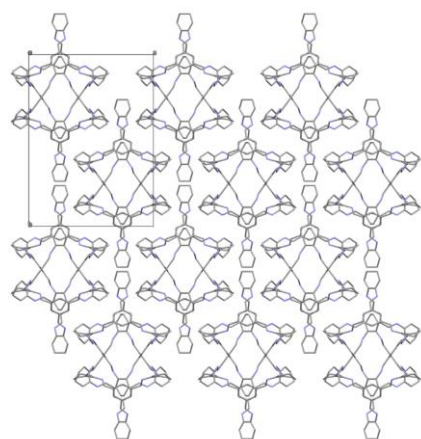


**Figure 5-S3.** Structural Overlay of cluster of 15 molecules for experimental (blue) and predicted (green) CC3-R $\alpha$  (Panel a) and CC3-R $\beta$ -solvated (Panel b). Solvent molecules in panel b are omitted for clarity.

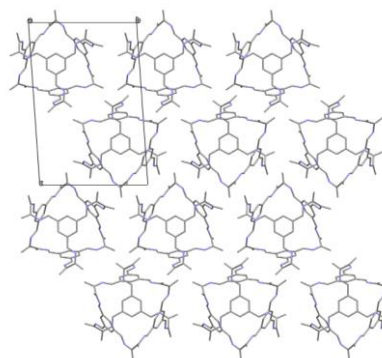
Structure	space group	CSP $\Delta E$ (kJ/mol)	CSP PLD(Å)	CSP density(g/cm <sup>3</sup> )	DFT-D3 $\Delta E$ (kJ/mol)	DFT-D3 PLD(Å)	DFT-D3 density(g/cm <sup>3</sup> )
CC3-(R,S) $\alpha$	P2 <sub>1</sub> /c	0.00	3.74	0.94	0.00	4.05	0.99
CC3-(R,S)	C <sub>2</sub> /c	6.71	3.74	0.94	7.86	3.93	0.99
CC3-(R,S)	Pna2 <sub>1</sub>	15.40	3.74	0.93	22.60	4.00	0.98
CC3-(R,S)	C <sub>2</sub> /c	27.51	3.74	0.92	30.93	3.86	0.99
CC3-R $\alpha$	F4 <sub>1</sub> 32	36.21	3.74	0.92	53.38	3.89	0.95

**Table 5-S3.** Comparison between CSP and DFT-D3 results for cage CC3.  $\Delta E$  are given w.r.t the energy of the global minimum found in CSP

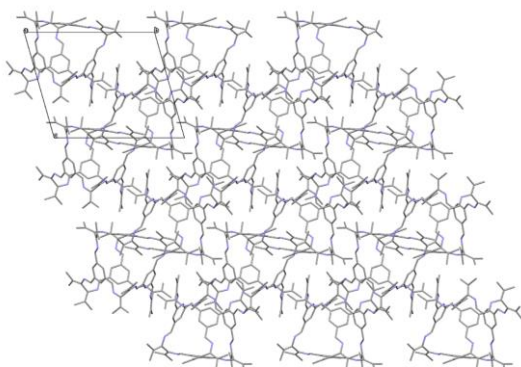
### 6.3 Predicted $\alpha$ -phases of molecules M1-M6.



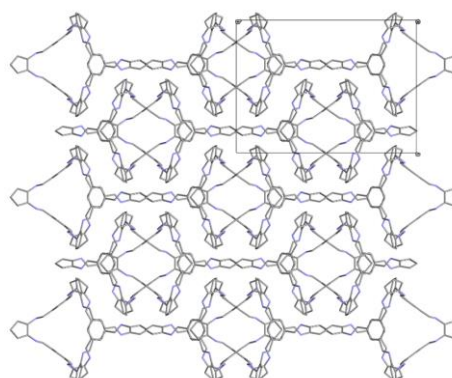
M1- $\alpha$   $\Delta E = 0.0$  kJ/mol PLD = 3.10 Å



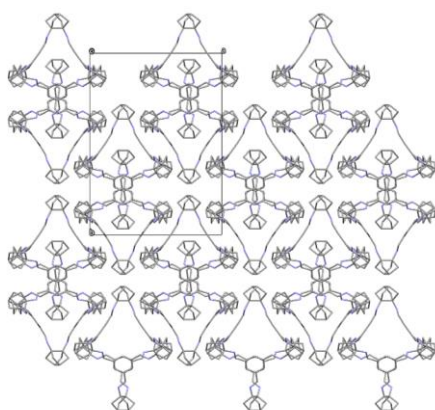
M2- $\alpha$   $\Delta E = 0.0$  kJ/mol PLD = 2.80 Å



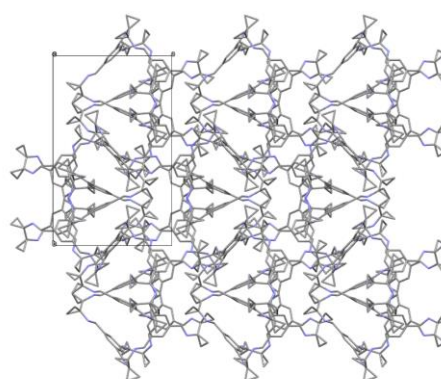
M3- $\alpha$   $\Delta E = 0.0$  kJ/mol PLD = 2.61 Å



M4- $\alpha$   $\Delta E = 0.0$  kJ/mol PLD = 3.13 Å



M5- $\alpha$   $\Delta E = 0.0$  kJ/mol PLD = 2.51 Å



M6- $\alpha$   $\Delta E = 0.0$  kJ/mol PLD = 3.36 Å

**Figure 5-S4.** Lowest energy predicted structures for M1-M6. The information in the captions includes the relative lattice energy ( $\Delta E$ ) and the PLD.

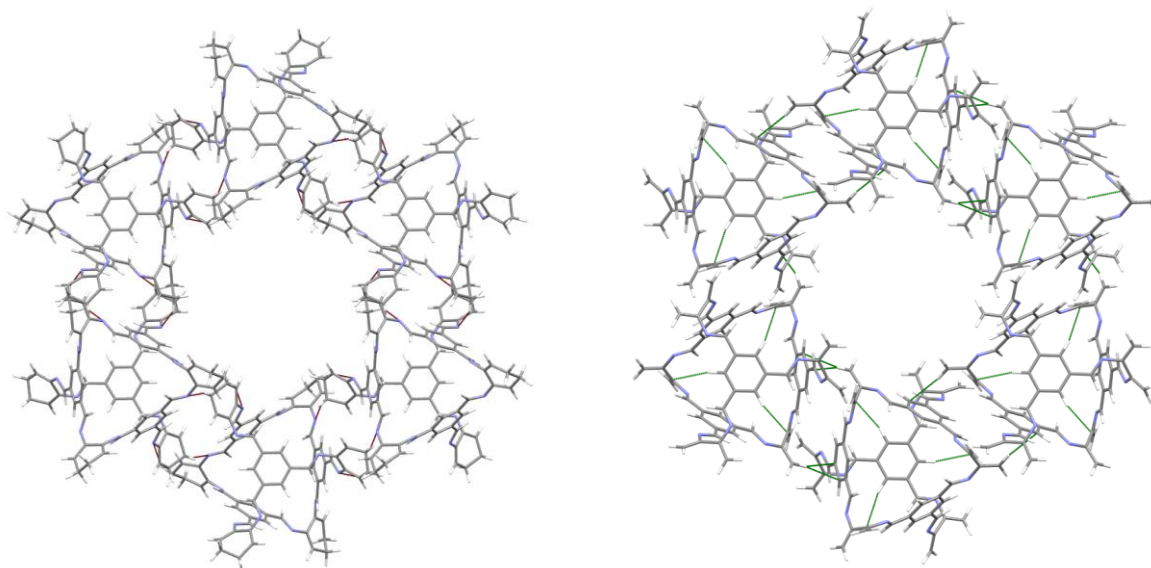
#### 6.4 Comparison between CSP and DFT-D3 results for the first five predicted structures of molecules M1-M6.

Structure	space group	CSP $\Delta E(\text{kJ/mol})$	CSP PLD( $\text{\AA}$ )	CSP density( $\text{g/cm}^3$ )	DFT-D3 $\Delta E(\text{kJ/mol})$	DFT-D3 PLD( $\text{\AA}$ )	DFT-D3 density( $\text{g/cm}^3$ )
M1- $\alpha$	$Fd-3$	0.00	3.10	0.96	0.00	3.12	1.00
M1-2	$P\bar{1}$	6.37	3.10	0.93	-8.80	2.73	1.00
M1-3	$P\bar{1}$	7.22	3.08	0.94	-25.66	2.88	1.07
M1-4	$P\bar{1}$	9.35	3.08	0.93	-4.42	2.53	1.00
M1-5	$P\bar{1}$	10.29	3.23	0.93	-2.27	2.80	1.00
Structure	space group	CSP $\Delta E(\text{kJ/mol})$	CSP PLD( $\text{\AA}$ )	CSP density( $\text{g/cm}^3$ )	DFT-D3 $\Delta E(\text{kJ/mol})$	DFT-D3 PLD( $\text{\AA}$ )	DFT-D3 density( $\text{g/cm}^3$ )
M2- $\alpha$	$P\bar{1}$	0.00	2.80	0.97	0.00	2.09	1.04
M2-2		2.66	2.88	0.95	0.71	2.48	1.03
M2-3	$P2_1/c$						
	$P\bar{1}$	3.30	2.47	0.97	-6.34	2.06	1.05
M2-4	$P\bar{1}$	3.52	2.60	0.97	-6.44	2.04	1.05
M2-5		3.74	2.82	0.96	-8.80	2.55	1.03
	$P2_1/c$						
Structure	space group	CSP $\Delta E(\text{kJ/mol})$	CSP PLD( $\text{\AA}$ )	CSP density( $\text{g/cm}^3$ )	DFT-D3 $\Delta E(\text{kJ/mol})$	DFT-D3 PLD( $\text{\AA}$ )	DFT-D3 density( $\text{g/cm}^3$ )
M3- $\alpha$	$P\bar{1}$	0.00	2.61	0.84	0.00	1.70	0.98
M3-2	$P2_1/c$	1.65	2.44	0.83	37.09	1.64	0.92
M3-3	$P2_1/c$	6.66	3.07	0.80	46.64	2.44	0.89
M3-4	$P\bar{1}$	7.30	2.09	0.84	31.56	1.80	0.95
M3-5	$C_2/c$	10.90	3.01	0.81	45.01	2.17	0.90
Structure	space group	CSP $\Delta E(\text{kJ/mol})$	CSP PLD( $\text{\AA}$ )	CSP density( $\text{g/cm}^3$ )	DFT-D3 $\Delta E(\text{kJ/mol})$	DFT-D3 PLD( $\text{\AA}$ )	DFT-D3 density( $\text{g/cm}^3$ )
M4- $\alpha$	$Fd-3$	0.00	3.14	1.08	0.00	3.30	1.09
M4-2	$P\bar{1}$	45.54	3.14	0.97	49.40	3.13	1.01
M4-3	$P\bar{1}$	50.22	3.13	0.97	17.63	3.27	1.09
M4-4		51.50	3.12	0.97	51.51	3.20	1.04
M4-5	$P2_1/c$						
	$P2_1/c$	52.03	3.13	0.98	13.78	3.84	1.09
	$P2_1/c$						
Structure	space group	CSP $\Delta E(\text{kJ/mol})$	CSP PLD( $\text{\AA}$ )	CSP density( $\text{g/cm}^3$ )	DFT-D3 $\Delta E(\text{kJ/mol})$	DFT-D3 PLD( $\text{\AA}$ )	DFT-D3 density( $\text{g/cm}^3$ )
M5- $\alpha$	$Fd-3$	0.00	2.51	1.04	0.00	1.45	1.22
M5-2	$C_2/c$	21.48	2.59	1.03	21.62	1.81	1.18

M5-3	$P2_1/c$	38.90	2.95	1.00	38.22	2.84	1.07
M5-4	$P\bar{1}$	39.40	2.94	0.99	38.96	2.84	1.06
M5-5	$P\bar{1}$	40.38	2.94	1.00	39.06	2.82	1.06
Structure	space group	CSP $\Delta E$ (kJ/mol)	CSP PLD(Å)	CSP density(g/cm <sup>3</sup> )	DFT-D3 $\Delta E$ (kJ/mol)	DFT-D3 PLD(Å)	DFT-D3 density(g/cm <sup>3</sup> )
M6- $\alpha$	$P2_1/c$	0.00	3.37	0.98	0.00	3.13	1.04
M6-2	$P\bar{1}$	2.62	2.22	0.95	7.45	1.93	1.00
M6-3	$P\bar{1}$	5.83	2.28	0.93	23.24	2.00	0.98
M6-4		7.52	3.30	0.94	22.77	3.03	0.98
M6-5	$R\bar{3}$	13.04	2.19	0.95	-8.80	1.96	1.07

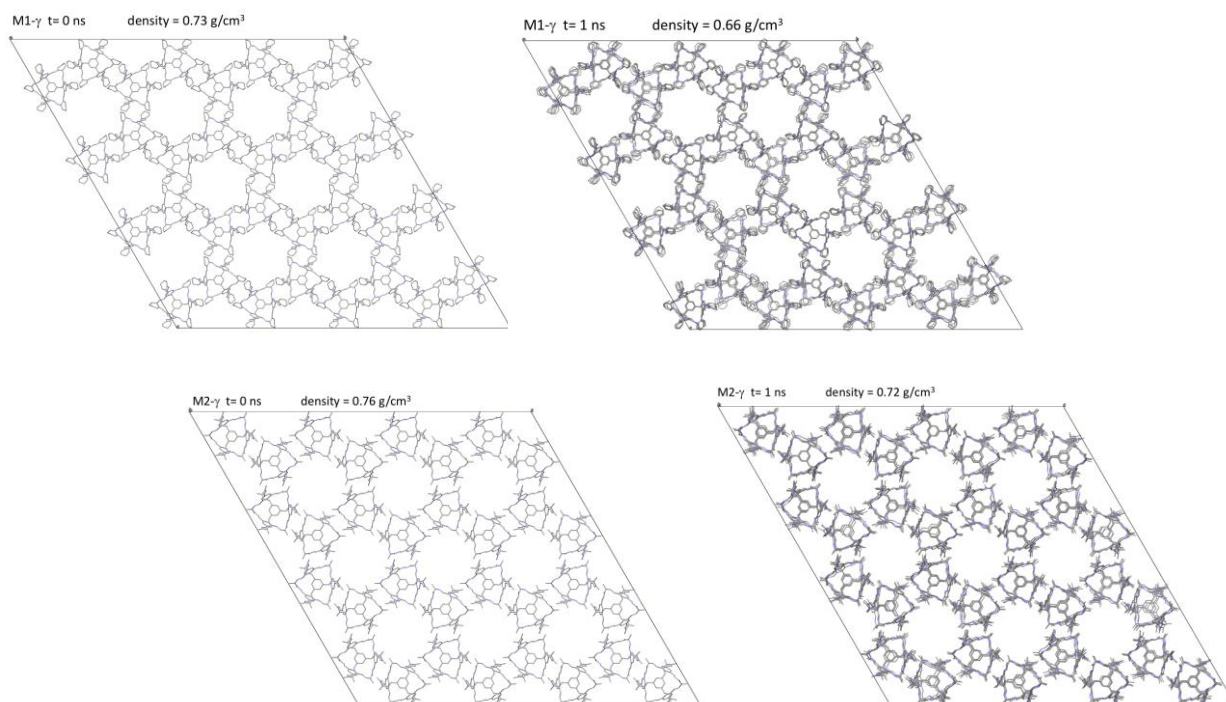
**Table 5-S4.** Comparison between CSP and DFT-D3 results for the first five structures w.r.t. the predicted global minimum.  $\Delta E$  are given w.r.t the energy of the the  $\alpha$ -phases at CSP and DFT-D3 level.

### 6.5 H-bond patterns in M1- $\gamma$ and M2- $\gamma$ .



**Figure 5-S5.** Left panel: Small cluster of the structure M1- $\gamma$ . Intermolecular C(sp<sup>3</sup>)H...N bonds are displayed as red lines. Right panel: Small cluster of the structure M2- $\gamma$ . Intermolecular C(sp<sup>2</sup>)H... $\pi$  bonds are displayed as green lines.

## 6.6 MD simulations to assess the thermal stability of M1- $\gamma$ and M2- $\gamma$ .



**Figure 5-S6.** *Top panels:* Projection of a supercell of structure M1- $\gamma$  at the beginning (left) and at the end (right) of a 1ns MD simulation. *Bottom panels:* Projection of a supercell of structure M2- $\gamma$  at the beginning (left) and at the end (right) of a 1ns MD simulation

## 6.7 References.

1. T. Tozawa, J. T. Jones, S. I. Swamy, S. Jiang, D. J. Adams, S. Shakespear et al. *Nature Materials* 2009, 8, 973.
2. L. Chen, P. S. Reiss, S. Y. Chong, D. Holden, K. E. Jelfs, T. Hasell et al., *Nature Materials*, 2014, 13, 954.
3. A. Kewley, A. Stephenson, L. Chen, M. E. Briggs, T. Hasell and A. I. Cooper, *Chemistry of Materials* 2015, 27, 3207.; T. Mitra, K. E. Jelfs, M. Schmidtman, A. Ahmed, S. Y. Chong, D. J. Adams and A. I. Cooper, *Nature Chemistry*, 2013, 5, 276.
4. X. Kong and J. Jiang, *Phys. Chem. Chem. Phys.*, 2017, 19, 18178.; X. Kong and J. Jiang, *J. Phys. Chem. C*, 2018, 122, 1732.

5. E. O. Pyzer-Knapp, H. P. G. Thompson, F. Schiffmann, K. E. Jelfs, S. Y. Chong, M. A. Little, A. I. Cooper and G. M. Day. *Chem. Sci.* 2014, 5, 2235 and references therein.
6. J. A. T. Jones, T. Hasell, X. Wu, J. Bacsá, K. E. Jelfs, M. Schmidtman et al., *Nature*, 2011, 474, 367.