Shape Prediction for Supramolecular Organic Nanostructures: [4+4] Macrocyclic Tetrapods

Michael E. Briggs, Kim E. Jelfs, Samantha Y. Chong, Catherine Lester, Marc Schmidtmann, Dave J. Adams and Andrew I. Cooper\*

Department of Chemistry and Centre for Materials Discovery, University of Liverpool, Crown Street, Liverpool, L69 7ZD, UK.

ABSTRACT Two [4+4] imine cages were synthesized with a unique macrocyclic tetrapod shape. The 3-dimensional structures of both molecules were predicted *a priori* by using conformer searching routines, illustrating a computational strategy with the potential to target new organic molecules with shape-persistent, intrinsic pores. These methods, in combination with crystal structure prediction, form part of a broader strategy for the computationally-led synthesis of functional organic solids.

\* Andrew I. Cooper,

Department of Chemistry and Centre for Materials Discovery, University of Liverpool, Crown Street, Liverpool, L69 7ZD, UK.

Tel: +44(0)1517943539. Fax: +44(0)1517942304. Email: [aicooper@liv.ac.uk](mailto:aicooper@liv.ac.uk)

Shape Prediction for Supramolecular Organic Nanostructures: [4+4] Macrocyclic Tetrapods

Michael E. Briggs, Kim E. Jelfs, Samantha Y. Chong, Catherine Lester, Marc Schmidtmann, Dave J. Adams and Andrew I. Cooper\*

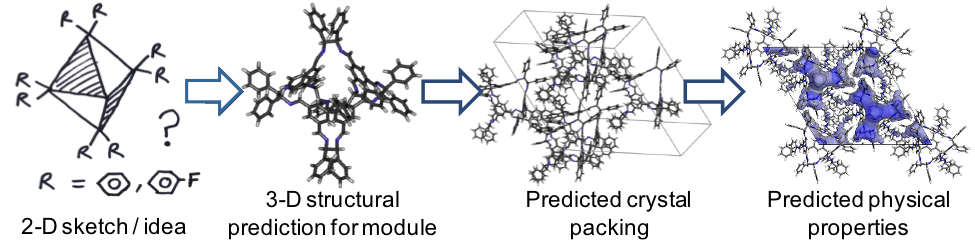
Department of Chemistry and Centre for Materials Discovery, University of Liverpool, Crown Street, Liverpool, L69 7ZD, UK.

\* aicooper@liv.ac.uk

ABSTRACT Two [4+4] imine cages were synthesized with a unique macrocyclic tetrapod shape. The 3-dimensional structures of both molecules were predicted *a priori* by using conformer searching routines, illustrating a computational strategy with the potential to target new organic molecules with shape-persistent, intrinsic pores. These methods, in combination with crystal structure prediction, form part of a broader strategy for the computationally-led synthesis of functional organic solids.

**INTRODUCTION**

Multicomponent self-assembly reactions are important in the synthesis of a wide range of complex molecular nanostructures.1-7 For organic systems, the use of reversible dynamic covalent chemistry (DCC) to synthesize macrocycles and molecular cages is well known: for example, via the cycloimidation of aldehydes and amines.7-10 Recently, a number of three-dimensional organic cages have been synthesized which can be either non-porous11-13 or permanently porous.14-17 Organic cage catenanes18 have also been prepared using DCC routes that can form up to 32 imine bonds in a single reaction.19 Some of these organic cages can exhibit surprisingly high surface areas in the solid state,16,20 in one case greater than 2000 m2 g-1,21 thus rivaling porous extended frameworks. The reversible nature of DCC reactions allows for error correction, and can afford products in high yields and purities within a one-pot, gram-scale synthesis.22 Molecular organic cages have been explored for gas storage and capture,15-17 in molecular separations,23,24 and in sensing.25 Also, unlike most nanoporous materials, porous molecular cages can be processed from homogeneous solutions into various formats.24,26,27 However, despite this range of interesting physical properties, the *targeted* synthesis of functional organic solids by dynamic self-assembly is still highly challenging, particularly as the constituent organic molecules become larger and more complex. This is because the computational prediction of three-dimensional structure, and hence function, for such materials requires fine precision in a sequence of interrelated calculation steps (Scheme 1).



**Scheme 1.** The steps required for *de novo* computational prediction, and hence ‘*in silico* design’, of physical properties for organic crystals, starting only with a conceptual molecular structure (left).

The first level of complexity lies in achieving the target molecular weight in the supramolecular self-assembly reaction. A given DCC reaction can usually produce, in principle, a number of molecular species in equilibrium with each other, each with a different molecular weight. Moreover, the course of the self-assembly reaction might be influenced by subtle changes in the structure of the reagents,15 as well as by experimental factors such as concentration, rate, order of reagent addition, and solvent.19 As such, even if a single molecular product is formed, rather than a mixture, it is not straightforward to predict, *a priori*, which product will be preferred in a new reaction. This makes the targeted synthesis of organic nanostructures inherently difficult, and we have therefore developed computational methods to calculate the relative formation energies for possible reaction products in cage-forming reactions.28

The prediction of three-dimensional molecular structure for self-assembled nanostructures can also be challenging, even when the molecular formula, and hence the molecular weight, of the molecule is known. Structural calculations for relatively small and rigid molecules are now fairly standard: for example, as part of crystal structure prediction (CSP) routines.29,30 CSP methods are typically based on global minimization of the lattice energy, and they often approximate the molecules as rigid in the initial calculations. The prediction of molecular structure is still challenging for larger species, such as large macrocycles or cages, that can comprise hundreds of atoms and possess significant conformational flexibility.13,15,31

Even in cases where both the molecular formula and the 3-D molecular structure are known, or can be accurately calculated,28 it still remains a major challenge to predict the crystal packing for large organic molecules,32,33 although we have recently achieved this for certain porous organic cages.16 Knowledge of the extended three-dimensional solid-state structure is, of course, essential to understand physical properties, such as porosity.

Given the substantial difficulties in predicting molecular weights, three-dimensional molecular structures, and crystal packings for large, self-assembled organic molecules, it can be argued that the *de novo* ‘design’ of functional organic solids, if not entirely an “illusion”,33 is certainly exceptionally challenging. Hence, we see a long-term need for computational strategies to underpin the rational design of functional molecular organic crystals, as already recognized for extended metal-organic frameworks (MOFs).34,35 This capability gap extends beyond the area of porous materials: indeed, the development of reliable predictive design methods covering all of the steps outlined in Scheme 1 would be relevant to any solid-state supramolecular organic materials.36

Previously, we showed that we could make accurate predictions for the crystal packings of a range of porous organic cages with up to 238 atoms per cage (formula mass = 793–1702 g mol‑1).16 The calculation of the molecular structures, prior to the crystal structure prediction step, was achieved using density functional theory (DFT) methods. This was facilitated by the relatively simple molecular topology of the cages and their rather rigid structures. By contrast, we have also synthesized much larger imine cages where these simple DFT methods are not feasible due to a combination of molecular size (512 atoms; 3573 g mol-1) and greater conformational flexibility.15 Large and flexible organic molecules such as these have complex conformational energy landscapes. Hence, simple ‘sketch and minimize’ strategies may not give the lowest energy structures because systems will become trapped in local energy minima that are associated, for example, with less stable isomers or conformers. Indeed, whilst chemical intuition alone might help us to predict simple shapes of structures with rigid fragments, it will not be a feasible approach for targeting large, complex shapes, as this will necessitate larger, often more flexible, precursors through lack of synthetic choice or by design to introduce functionality through flexibility. There is therefore a need to develop automated computational methods for broad and effective conformer searching for large, self-assembled organic nanostructures. These have the potential to not only predict shape, but to predict accurate conformations and to thus provide the basis for calculating properties *a priori*.

In this study, we exemplify such a strategy via two novel [4+4] macrocyclic imine cages. We show that the use of computation, and in particular automated conformer searches, gives much better structural insight than could be gained from intuition and chemical knowledge alone.

**EXPERIMENTAL SECTION**

**Conformer Searching:**

Atomistic calculations were performed using the OPLS all-atom forcefield37, which was originally parameterized for organic systems with a focus on conformer energetics, intermolecular energies, and thermodynamic properties. Geometry optimization was performed using the Polak-Ribiere Conjugate Gradient (PRCG) method within MacroModel38 and a convergence criterion of a gradient norm less than 0.05 kJ mol-1 Å-1. Conformer searches were performed using a low-mode39 sampling approach within MacroModel which follows the low frequency eigenvectors of the molecule. We used 10000 search steps with maximum and minimum move distances of 3 and 6 Å. All structures within an energy window of 50 kJ mol-1 of the lowest energy structure were retained, excluding mirror images. We applied a final refinement of the structures and energies by optimizing the resulting molecular structures with DFT-D3 calculations performed in CP2K40 with the PBE functional,41 TZVP-MOLOPT basis sets,42 GTH-type pseudopotential,43 a plane wave grid cutoff of 400 Ry, a cubic box of length 50 Å and the Grimme-D3 dispersion correction.44 This procedure was applied to any structure within 20 kJ mol-1 of the lowest energy conformer, which we have previously determined to be the maximum error in the relative energies for the OPLS force field. We have used this DFT setup in the past to reliably reproduce the structure and energetics of porous imine cages in both the solid state16 and as single molecules and dimer pairs.26 We have previously validated the conformer searching approach for other, known molecular imine cages, correctly reproducing molecular conformations as determined by single crystal X-ray diffraction structures.28

**Cage Synthesis:**

For **CC11**, a solution of a tripodal amine, tris(2-aminoethyl)amine (TREN) (89 mg, 0.61 mmol) in methanol (34 ml) was layered onto a solution of tris(4-formylphenyl)amine (200 mg, 0.61 mmol) in DCM (34 ml), and the biphasic reaction mixture was allowed to stand at ambient temperature without stirring (Scheme 2). After 11 days, the reaction mixture was sampled and analyzed by 1H NMR (CDCl3) (Figure S2) and MALDI-TOF mass spectrometry (Figure S3), which showed **CC11** to be the sole product. Although analysis of the reaction mixture suggested quantitative conversion of the starting materials to CC11 we were unable to isolate the bulk material as a dry solid, either by slow evaporation or by precipitation, due to decomposition of the material. 1H NMR (CDCl3): δ 7.79 (12H, s, N=CH), 7.09 (24H, d, *J* =8.5 Hz, Ar-H), 6.95 (24H, d, *J* =8.5 Hz, Ar-H), 3.80 (12H, d, *J*=11.4 Hz), 3.38 (12H, t, *J*=11.5 Hz), 3.07 (12H, t, *J*=12.6 Hz), 2.44 (12H, dd, *J*=12.9 and 3.6 Hz) ppm. 13C NMR (CDCl3) δ 161.2, 148.9, 132.1, 129.3, 124.2, 59.5, 55.2 ppm. MS (MALDI-TOF): 1687.9 ([M + H]+), 1749.9 ([M + Cu]+).

To produce single crystals of **CC11**, a solution of TREN (8.9 mg, 0.061 mmol) in methanol (3.5 ml) was carefully added to tris(4-formylphenyl)amine (20 mg, 0.061 mmol) so as not to disturb the solid aldehyde. The suspension was allowed to stand, whereupon the aldehyde slowly dissolved until colorless crystals were observed to have formed on a bed of undissolved aldehyde. With careful isolation, these crystals were found to be suitable for single crystal X-ray diffraction.

For **CC12**, a solution of TREN (15.9 mg, 0.109 mmol) in methanol (10.2 ml) was carefully layered onto a solution of tris-1-(1'-formyl-4-4'-biphenyl)amine (60 mg, 0.108 mmol) in DCM (15.3 ml) and the biphasic reaction mixture was allowed to stand at ambient temperature without stirring. After 13 days, an aliquot of the reaction mixture was taken and the solvent was evaporated under vacuum at rt. Analysis of the residue by 1H and 13C NMR (CDCl3) (figure S6) and MALDI-TOF mass spectrometry (figure S7) showed complete conversion of the starting materials to CC12. 1H NMR (CDCl3): δ 7.60 (s, 12 H, H-C=N-), 7.37 (d, J=8.9 Hz, 24 H, Ar-H), 7.33 (d, J=9.0 Hz, 24 H, Ar-H), 7.09 (d, J=8.4 Hz, 24 H, Ar-H), 6.74 (d, J=8.1 Hz, 24, Ar-H), 3.81 (d, J=9.2 Hz, 12 H), 3.38 (t, J=11.5, 12 H), 3.11 (t, J=11.8, 12 H), 2.37 (d, *J*=9.6 Hz, 12 H) ppm. 13C NMR (CDCl3) δ 162.0, 147.2, 142.5, 136.0, 135.3, 128.9, 127.8, 127.3, 124.6, 59.6, 55.2 ppm. MS (MALDI-TOF): 2600.8 ([M + H]+).

To produce single crystals of **CC12**, a solution of TREN (5.2 mg, 0.036 mmol) in methanol (3.4 ml) was carefully layered onto a solution of tris-1-(1'-formyl-4-4'-biphenyl)amine (20 mg, 0.036 mmol) in DCM (3.4 ml) and the biphasic reaction mixture was allowed to stand at ambient temperature without stirring. After 11 days crystals of **CC12** were observed to have formed. Suitable single crystals could only be isolated by encapsulating them in polyethylene glycol.

**Characterization:**

NMR. Solution 1H and 13C NMR spectra were recorded in deuterated chloroform at 400.13 MHz using a Bruker Avance 400 NMR spectrometer.

MALDI. Mass spectra were recorded on a Waters Micromass M@LDI-TOF mass spectrometer with Waters MassLynx software. -Cyano-4-hydroxy-cinnamic acid (Sigma) was used as the matrix. The Pulse and Source Voltages were 3400V and 16000V respectively.

Powder X-ray Diffraction. Variable temperature powder X-ray diffraction data (VT-PXRD) were collected with the multi-analysing crystal (MAC) detector on the I11 beamline at Diamond Light Source (λ = 0.825556 Å) and using the Cyberstar hot air blower with a ramp rate of 5 K/min. The sample was held in a 1 mm diameter special glass capillary and spun to improve powder averaging. Analysis of the powder diffraction patterns was carried out using TOPAS-Academic45 software. See Supporting Information for details.

Single Crystal X-ray Diffraction. Single crystal X-ray data was measured on a Rigaku MicroMax-007 HF rotating anode diffractometer (Mo-Kα radiation, λ = 0.71073 Å, Kappa 4 circle goniometer, Rigaku Saturn724+ detector). Empirical absorption corrections using equivalent reflections were performed with the program SADABS;46 the structures were solved with the program SHELXD and refined with SHELXL47 using the OLEX2 GUI.48 Non H atoms in the molecules were refined anisotropically, H atoms were fixed to geometric positions using the riding model; most non H atoms of the solvent molecules were largely refined isotropically, solvent H atoms were omitted.



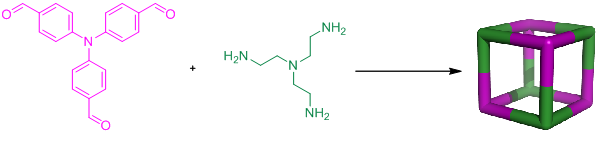


**Scheme 2.** Synthesis of cages **CC11** and **CC12**.

**RESULTS**

**Cage Design and Synthesis:** All of the imine cage molecules reported so far by our group have been prepared by the cycloimidation reaction of trialdehydes with aliphatic diamines: that is, the underlying aldehyde-to-amine reaction stoichiometry is 2:3. This has allowed the synthesis of [2+3],28,49 [4+6],14,16,28,50 and [8+12] cages.15 Self-assembly reactions with 2:3 stoichiometry are also common for metal-organic cages.55 Different geometrical shapes and topologies can be accessed by exploring different reaction stoichiometries, and here we explored the synthesis of two new cage-like molecules, **CC11** and **CC12**, via the cycloimidation reaction of a triamine with two homologous trialdehydes (Scheme 2); that is, an overall stoichiometry of 1:1, assuming complete imine condensation. A tripodal amine, tris(2-aminoethyl)amine (TREN), was used which was previously reported to form small [2+3] molecular cages when reacted with dialdehydes.51

There are many potential [n+n] combinations that would satisfy a 1:1 stoichiometry, not including the possibility of cross-linked polyimine formation. However, trial reactions and, in particular, mass spectrometry and 1H NMR measurements (Supporting Information, Figs. S3, S4, S6 & S7) suggested the major product in both of these cycloimination reactions was a [4+4] cage52 (observed molar masses for **CC11** and **CC12** = 1686 and 2599 g mol-1). Given the size, complexity, and probable conformational flexibility of these molecules, it is hard to make intuitive predictions as to their most likely 3-dimensional structures, although the approximately trigonal triarylamine node would seem to preclude the idealized ‘cube’ structure to which these [4+4] cages are topologically equivalent (Scheme 3).

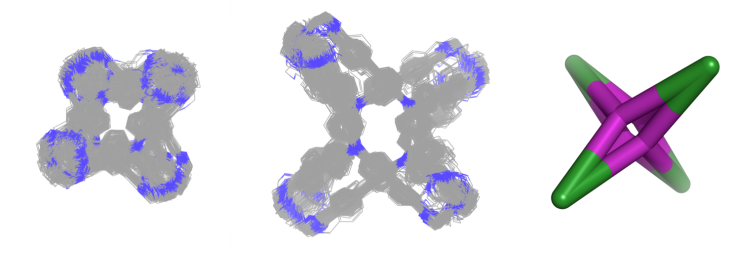


**Scheme 3.** The two [4+4] cages **CC11** and **CC12** (only scheme for **CC11** is shown here) are both topologically equivalent to a cube.

Moreover, as detailed below, it proved difficult to isolate crystals that were suitable for structure determination by single crystal X-ray diffraction. As such, we were confronted with a common predicament in supramolecular chemistry: chemical characterization, in this case NMR and mass spectrometry, suggests an intriguing self-assembled structure, but crystal growth and structure determination prove challenging.

**Structure Prediction and Conformer Searching:** Structural models for the two new imine cages, **CC11** and **CC12**, were constructed manually, assuming an underlying [4+4] topology as shown in Scheme 3. Immediately upon geometry optimization, the conformation of both molecules became tetrapodal in nature, with the tertiary amine nitrogens of the TREN linker forming the four apexes of the macrocyclic ‘tetrapods’. This suggests, in the gas phase at least, where no solvent is included, that no energetic barriers exist to prevent the ‘collapse’ of the inflated cage. To explore whether there were alternative conformations, perhaps possessing an internal void, we then performed extensive conformer searching to find low energy conformers for each of the two molecules.

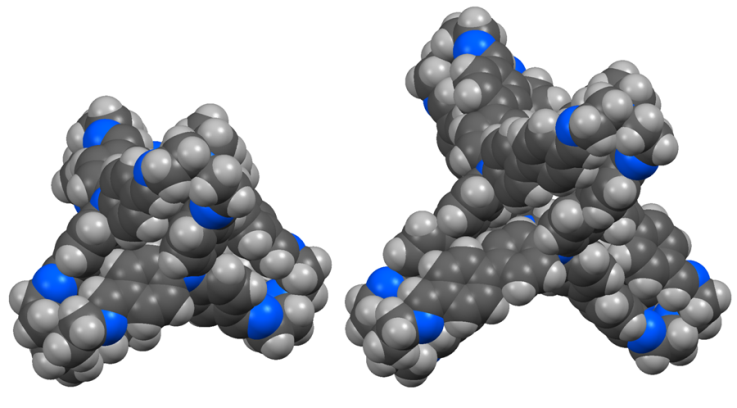
The results from the OPLS energy-ranked conformer search showed that all low energy conformations of these cage structures were broadly tetrapodal in nature, as shown in Figure 1 by an overlay of all conformations found within 50 kJ mol-1 of the global minimum (147 conformations for **CC11** and 266 conformations for **CC12**, excluding mirror images). The differences between these conformations involve small rotations about the phenyl rings on the tetrapod ‘arms’, and rotations about the Carene-Carene-Cimine-Nimine bond. These facile torsional rotations are sufficient to explain the inward folding of the molecule from an idealized, cube-like structure. The five lowest energy conformations of **CC11** and **CC12**, which fall within a 5 kJ mol-1 range in the DFT calculations, differ only in rotations of the phenyl rings. In keeping with this, **CC12** exhibits disorder in the rotation of the phenyl rings in the single crystal structure of its solvate.



**Figure 1**. Left & center: An overlay of all conformations found to lie within 50 kJ mol-1 of the global minimum for **CC11** (left) and **CC12** (right) from the OPLS initial conformer search (wireframe representations). Carbons are shown in grey and nitrogens in blue; hydrogens are omitted for clarity. Right: Scheme showing the underlying topology of the predicted [4+4] tetrapod structures.

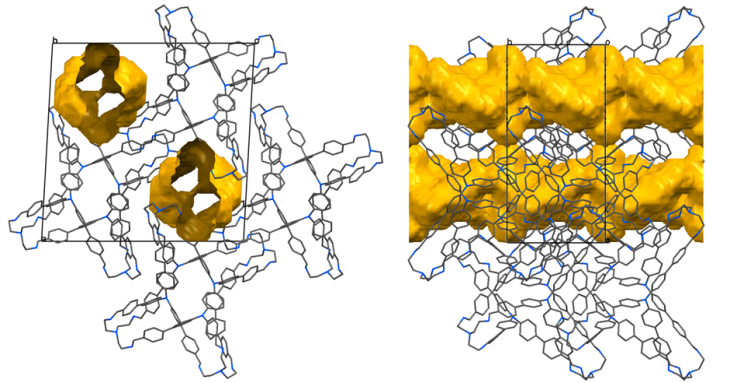
**Crystal Structure Determination:** Initially, all attempts at isolating crystalline forms of **CC11** failed, whether this was done by slow evaporation, by slow precipitation, or by careful solvent exchange. In all cases, the material underwent a rapid color change upon removal from the mother liquor and became irreversibly insoluble, suggesting that the material rapidly polymerizes upon spontaneous desolvation. We eventually succeeded in growing single crystals of the [4+4] adducts directly by carrying out the reactions in pure methanol, but these crystals were also found to degrade rapidly upon removal from the reaction solvent. The colorless crystals were finally stabilized by encapsulating them in perfluoropolyether oil prior to removal from the reaction solvent, and we obtained single crystal X-ray data at 100 K for the solvate of **CC11** (Figure 2, left).

The unit cell for **CC11** was identified as triclinic *P-1* with 8 molecules per unit cell, of which 4 molecules were found to be crystallographically independent. It is possible that the crystal packing is directed by the solvent, and that other solvents could give more regular packing motifs, although as yet no other crystal modifications have been discovered. A large amount of disordered MeOH and water was also observed in the voids between the cages. Interestingly, no solvent molecules were observed to be located within the cage itself, unlike other shape-persistent imine cages that have permanent voids which are filled with solvent upon product isolation.13,17,20 The encapsulated crystals degraded within the protective oil coating upon desolvation at ambient temperature. The cage molecules are broadly tetrapodal in shape, in agreement with the conformer predictions, with the TREN moiety forming the cap at the end of each of the arms of the tetrapod. To our knowledge, this is the first example of a [4+4] reaction between a trialdehyde and a triamine, and the first example of a macrocyclic organic ‘tetrapod’.



**Figure 2.** Single crystal X-ray structures of ‘tetrapods’ (*left*) **CC11** and (*right*) **CC12** in space-filling representation. One molecule of **CC11** is displayed here: there are four crystallographically independent molecules in the structure (Z´ = 4).

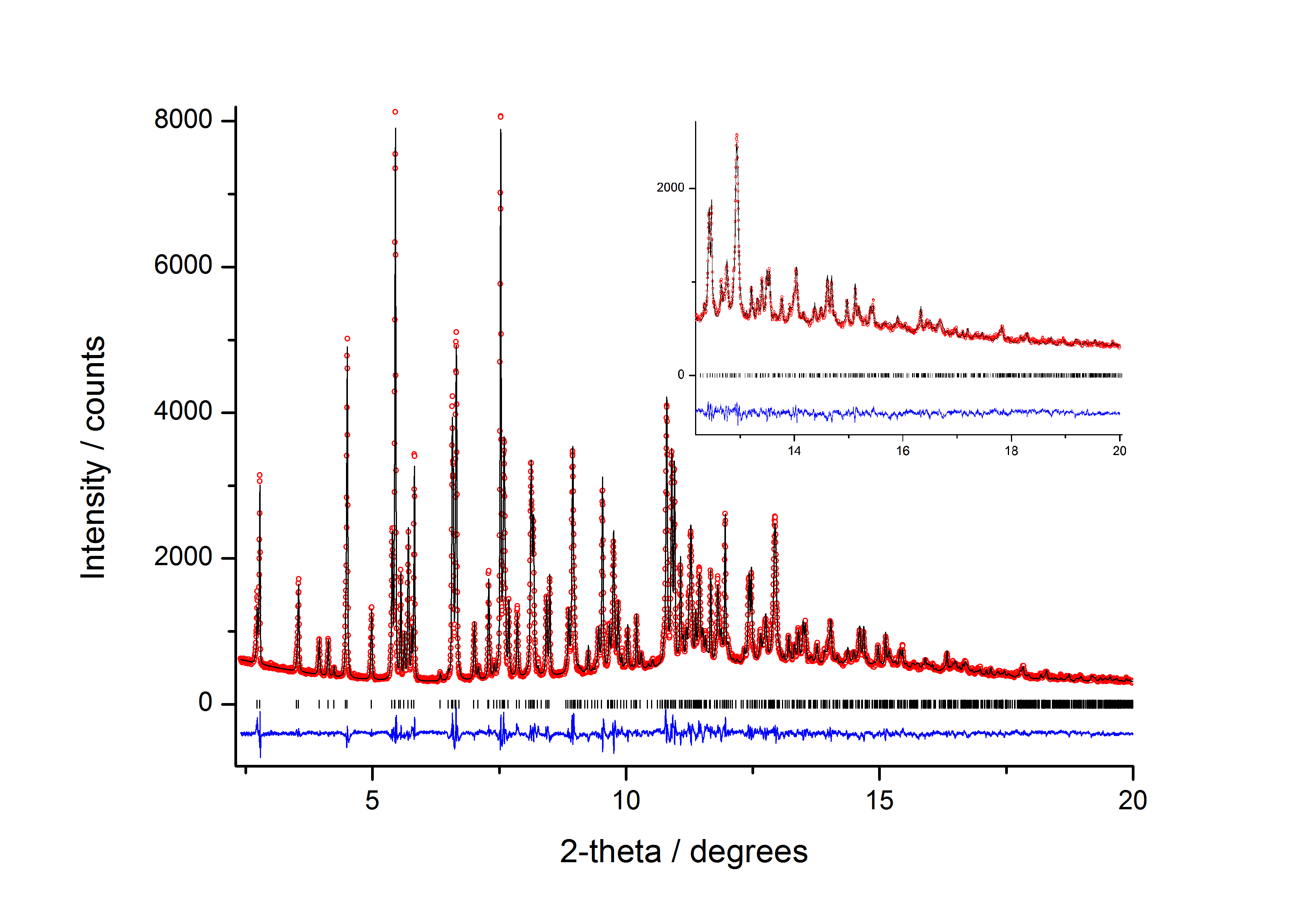
In contrast to **CC11**, we were able to crystallize **CC12** by slightly altering the ratio of the dichloromethane to methanol in the reaction mixture. Single crystals of **CC12** were unstable to desolvation, and were observed to transform to a fine but still polycrystalline yellow powder when removed from the reaction mixture. No crystals of air-dried **CC12** could be isolated that were suitable for single crystal X-ray diffraction. Uniquely, a polyethylene glycol (PEG) oil was found, unlike all the other protective fluids investigated, to protect the **CC12** crystals from decomposition. The [4+4] cages were again found to be tetrapodal in shape, with each of the four arms capped by the TREN moiety. The unit cell was identified as monoclinic with space group *P2/n*. We believe that the PEG oil—and possibly the diethylene glycol component in that oil—replaces the original reaction solvents (dichloromethane and methanol) in the large, 1-dimensional solvent-accessible channels in the crystals (Figure 3), thus preventing degradation of the structure. For example, the PEG-containing crystal was slowly (over several hours) heated to 350 K; at about 275 K it began to show some signs of decomposition, but no further deterioration was observed afterwards. The crystal was then left at 350 K for approximately one hour, after which it was cooled to 100 K to obtain a dataset of the partially desolvated material. This revealed that the dichloromethane molecules were removed from the original structure, while the channel content—the PEG oil—still remained.



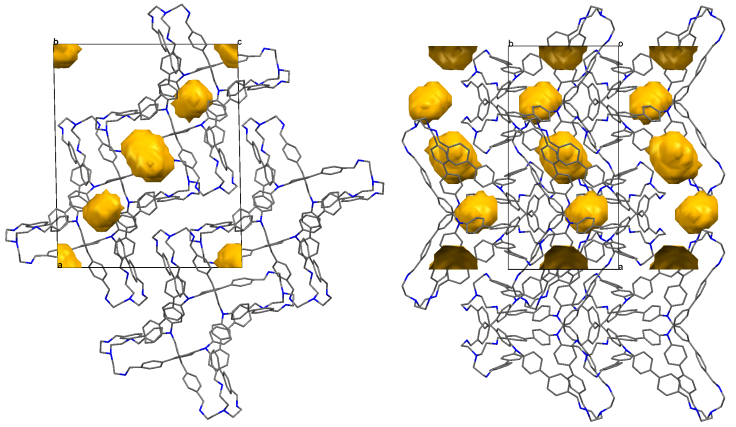
**Figure 3**. Packing diagrams for **CC12**, viewed along the crystallographic *b*-axis (*left*) and along the *c*-axis (*right*); solvent-accessible channels (probe radius 1.82 Å) shown after *in silico* removal of the refined solvent content. We believe that these channels, shown here in yellow, are filled by the involatile PEG-containing oil, and that this stabilizes the crystal structure against desolvation.

Non-encapsulated crystals of **CC12** that were removed from the reaction solvent and exposed to air for a few seconds rapidly cracked, but were found by powder X-ray diffraction (PXRD) to remain polycrystalline. Variable temperature PXRD of the air-dried microcrystalline powder held in an unsealed glass capillary was used to monitor the change in diffraction upon loss of the dichloromethane guest (Figure S8; Table S1). Upon heating to 363 K, the constant position and relative intensities of the diffraction peaks indicated that desolvation was complete and the profile collected at this temperature was used to determine the structure of desolvated **CC12** directly from the powder diffraction data.

The PXRD profile was indexed on the basis of the first 40 observable peaks (up to *d* ≈ 4.9 Å) to give a monoclinic cell and assigned the space group *P2/n*, consistent with 0.5 molecules of **CC12** (178 atoms, 100 non-H atoms) in the asymmetric unit. Lattice parameters (*a* = 22. 3391(1), *b* = 13.3802(1), *c* = 27.0670(2) Å, *ß* = 91.1259(6)°, *V* = 8088.8(1) Å3) and peak shape parameters were extracted by Le Bail53 fitting and direct space structure solution was achieved using the simulated annealing routine implemented in TOPAS-Academic. Refinement of the structure (Figure 4) was carried out using geometric restraints to afford the final crystal structure for desolvated **CC12** (Figure 5).

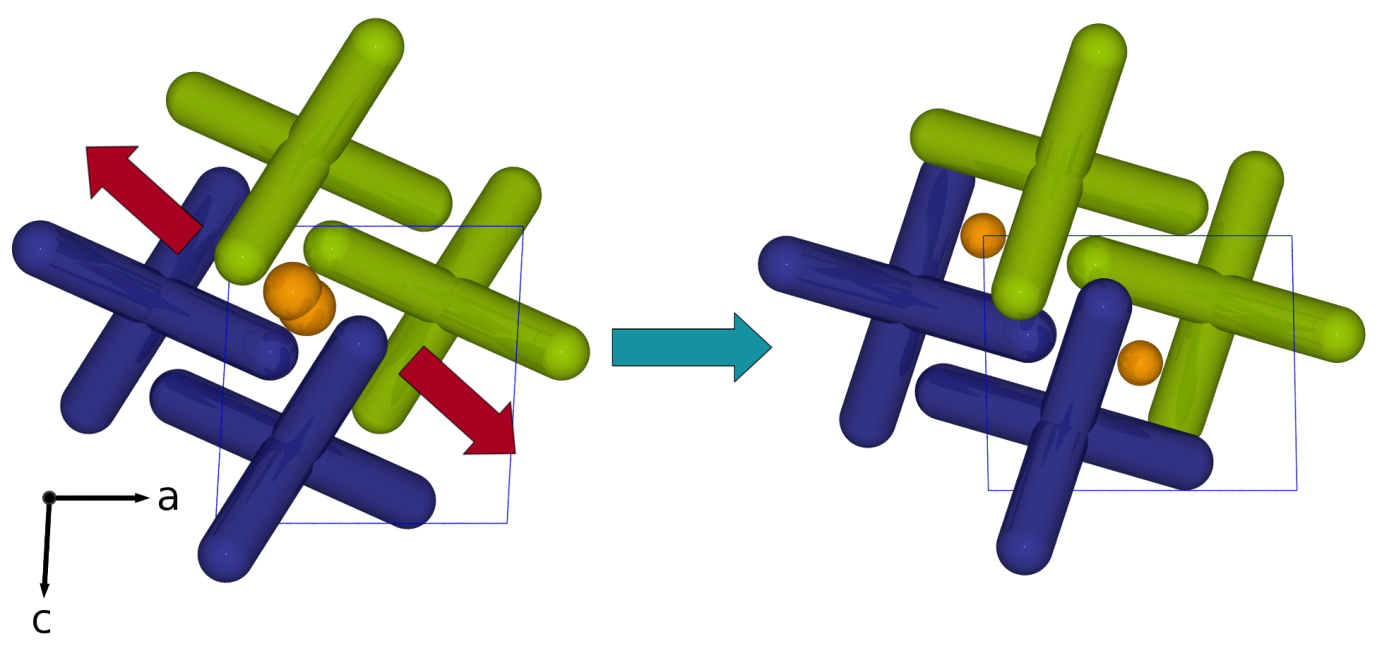


**Figure 4.** Final observed (circles), calculated (solid line) and difference (below) X-ray powder diffraction profile (** = 0.825556 Å) for the Rietveld refinement of desolvated **CC12** (*R*wp = 5.08 %, *R*p = 3.91 %, *R*Bragg = 1.60 %, **2 = 1.88). Reflection positions are also marked.



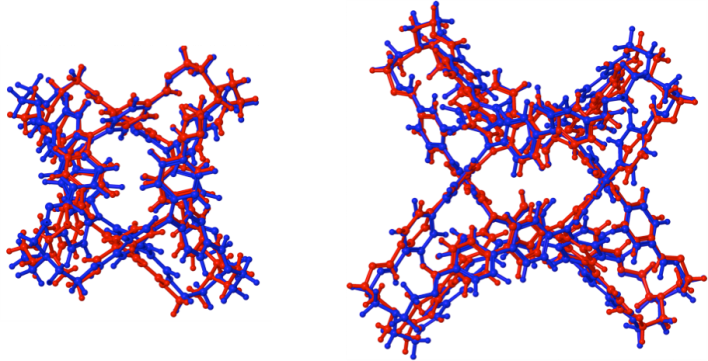
**Figure 5.** Crystal structure of desolvated **CC12** as determined from PXRD data viewed along the (*left*) *b*-axis and (*right*) *c*-axis, showing the isolated voids (probe radius 1.82 Å). Hydrogen atoms are omitted for clarity.

Difference Fourier maps showed no significant electron density corresponding to residual solvent. The distances between tertiary N atoms forming the tetrapod apexes is indicative of this closed structure (N3–N4 = 7.11(2) and 10.19(2) Å, *c.f.*, 10.263(5) and 10.906(5) Å in PEG-stabilized **CC12**). This creates columns of isolated, closed voids along the b-axis between **CC12** molecules, rather than the interconnected 1-D channels observed in the presence of PEG. The propagation of a π–π interaction (Ph59–64...Ph59–64 at -x, -y, 2-z = 3.94 Å) generates layers of **CC12** molecules parallel to (101), which stack to form the three-dimensional structure, stabilized by weak van der Waals interactions. The conformation adopted by the phenyl rings also creates a small intrinsic pore within the cage itself. The desolvated **CC12** phase therefore shows a significantly denser packing mode where the 1-D solvent-accessible channels (Figure 3) observed in the PEG-encapsulated crystals are no longer present. These channels, formed by the intersection of four discrete **CC12** molecules, are closed by the opposing movement parallel to [101] of layers of tetrapods, as illustrated schematically in Figure 6.



**Figure 6:** Scheme showing proposed mechanism of desolvation to form polycrystalline **CC12** by air-drying. Adjacent layers of **CC12** molecules move in opposing directions to close the 1-D channels observed in the presence of PEG fluid (*left*). The denser, desolvated **CC12** phase shows isolated voids inside the cage molecules (Fig. 5) and disconnected voids formed within layers of **CC12** parallel to the (101) plane (*right*).

**Comparison between Theoretical and Experimental Structures:** A key aim of this study was to test our ability to predict, using automated conformer searching, the most likely structures for large, complex organic molecules where the most stable conformation is not intuitively obvious. Figure 7 shows overlays of the experimental crystal structures and the lowest-energy computed conformations for both **CC11** and **CC12**. There is an excellent level of prediction for both of these complex molecules. The conformations with respect to the imine bonds are all correct: differences between theory and experiment lie mainly in small rotations of the phenyl torsions which, as already mentioned, are also manifested as disorder in the crystal structure for the **CC12** system.



**Figure 7:** Overlay of the experimental molecular conformation of **CC11** (left) and **CC12** (right) taken from crystal structures (red) and the lowest energy calculated conformations (blue). For **CC11** one of the four crystallographically-independent molecules in the solvate crystal structure is shown (see figure S1 for the other 3 overlays): for **CC12** (*Z*´ = 0.5), the conformation determined from the desolvated crystal structure is shown.

We next examined whether alternative shapes, such as a distorted cube, might be found to lie higher in energy and, hence, be accessible by templating using an appropriate solvent or template. To do this, the conformer search was repeated with larger maximum moves allowed (up to 20 Å), and all structures (26233 in total) lying within 1000 kJ mol-1 of the global minimum were retained. Automated analysis of the size of the internal void for these structures showed that all structures in this large energy window maintain an ‘collapsed’ tetrapodal conformation, with a very small internal void (up to 7 Å3) compared to the much larger void (>2000 Å3) that would be expected for an expanded, shape-persistent [4+4] ‘porous’ cage. As such, these simulations would suggest, *a priori*, that (i) this [4+4] combination is strongly predisposed to form a tetrapodal cage, irrespective of solvent, and (ii) this combination of building blocks has little promise for the construction of shape-persistent cages with large, internal cavities.

**DISCUSSION**

This study shows that systematic conformer searching can accurately predict the low-energy three-dimensional structures of large organic cages that have significant conformational flexibility. We believe that these calculations are useful in a predictive sense because the time required for the conformer searches (< 24 hours on a single core) is significantly shorter than the laboratory time required for preparation, crystallization, and structure solution for molecules of this type. For example, these particular [4+4] combinations are revealed by computation to be poor choices for the generation of shape-persistent cages with permanent internal voids. Clearly, the methods could also be used in the reverse sense to identify combinations of building blocks that lead to cage structures where a stable, shape-persistent void represents the conformational energy minimum. We therefore see this method as a useful component in a broader strategy for the *in silico* design of functional organic solids, as outlined in Scheme 1.

There are several challenges that remain in terms of broader practical implementation. Notably, the manual construction of the initial ‘starting’ models for nanostructures formed by multicomponent condensation is itself a somewhat laborious and time-consuming task, and we are currently building automated, geometry-based methods for constructing these systems, much as has been done for metal-organic frameworks.34,35 This would allow automated molecule construction and conformer searching for large libraries of hypothetical materials using, for example, aldehydes and amines as the building blocks.

The approach also has some inherent limitations: for example, the calculations are performed in the gas phase, and will not therefore take account of crystal packing forces, although this could be incorporated in a subsequent crystal structure prediction step for molecules that provoke preliminary interest. Likewise, while tetrapods **CC11** and **CC12** do not exhibit ‘solvent-inflated’ structures, it is likely that other molecules will crystallize as solvates with molecular structures that are quite different from the low-energy, desolvated forms: indeed, we have already observed this for a large [8+12] imine cage.15 However, given that our primary interest is in porous organic crystals, a methodology that predicts the lowest energy desolvated form is of broad utility. Indeed, the type of ‘virtual porosity’ referred to by Barbour,54 which results from deleting guests from a crystal structure, is of less practical interest and will not be predicted using these gas phase calculations. It is also known that some macrocyclic cages can form catenanes under certain conditions, and this methodology would not predict such structures without significant extension: indeed, the automated construction of, for example, models for triply-interpenetrating catenanes18 is a significantly more challenging prospect than the construction of models for discrete, non-catenated cages, where starting structures could be approximated, in the first instance, to Platonic and Archimedean solids.55

It is also important to stress that intrinsic porosity in a single, isolated molecule does not ensure porosity in the solid state: for example, **CC1** is intrinsically porous but its most stable polymorph packs in such a way that these intrinsic voids are disconnected.14 In principle, this too can be addressed via crystal structure prediction methods, as we already demonstrated for smaller [4+6] cages.16

**CONCLUSIONS**

We have synthesized two new [4+4] imine cages with unique macrocyclic ‘tetrapod’ structures. These cages have little or no intrinsic void volume in the solid state, and this can be predicted *de novo* by systematic conformer searches that reveal no low-energy structures with a permanent pore. That is, calculations tell us, with good confidence, that there is no prospect with these building blocks of forming anything like the ‘porous organic cube’ illustrated in Scheme 3. There is excellent agreement between the lowest energy predicted structures and experiment for both [4+4] tetrapods, as well as for a range of [4+6] imine cages that we synthesized in previous studies. Coupled with algorithms that automate the initial structure generation, this methodology will form a basis in the future for screening large numbers of hypothetical organic cages with respect to their potential for intrinsic porosity. Importantly, we believe that computation here is much more powerful than simple ‘intuition’: in this case, it gives an unambiguous prediction that non-porous ‘tetrapods’ will be formed, and that no other conformers are thermodynamically competitive. Indeed, the lowest-energy predicted structures (Fig. 7) are remarkably close to those characterized by single crystal X-ray diffraction. By contrast, chemical intuition alone suggested that an idealized cube (Scheme 3) was unlikely—but it told us little more than that—and these macrocyclic tetrapods were a synthetic surprise. In fact, automated conformer searching readily anticipates the formation of these structures, based solely on knowledge of the structural formula for the molecule.

In the longer term, we aim to combine these methods with crystal structure prediction16 to create a more general approach to the computationally-targeted design of organic solids, as outlined in Scheme 1. The broad utility of this would not be limited to porous materials.

ASSOCIATED CONTENT

**Supporting Information**. NMR spectra, mass spectrometry, crystallography and computer simulation details. This material is available free of charge via the Internet at http://pubs.acs.org

AUTHOR INFORMATION

Corresponding Author

\* aicooper@liv.ac.uk

Author Contributions

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

Funding Sources

Funded through support by the Engineering & Physical Sciences Research Council (EPSRC, grant EP/ H000925/1).

ACKNOWLEDGMENT

A.I.C. is a Royal Society Wolfson award holder. We thank Dr G. M. Day for useful discussions and advice and Mr Mark Prescott for mass spectrometry analysis. This work made use of the facilities of HECToR through our membership of the UK’s HPC Materials Chemistry Consortium (EPSRC EP/F067496). We also thank Diamond Light Source for access to beamline I11 (EE7040) that contributed to the results presented here and Drs. J. E. Parker and A. R. Lennie for their assistance during the experiment.

REFERENCES

(1) Lehn, J. M. *Angew. Chem.-Int. Edit. Engl.* **1990**, *29*, 1304–1319.

(2) Seidel, S. R.; Stang, P. J. *Acc. Chem. Res.* **2002**, *35*, 972–983.

(3) Fyfe, M. C. T.; Stoddart, J. F. *Acc. Chem. Res.* **1997**, *30*, 393–401.

(4) Fujita, M.; Oguro, D.; Miyazawa, M.; Oka, H.; Yamaguchi, K.; Ogura, K. *Nature* **1995**, *378*, 469–471.

(5) Kay, E. R.; Leigh, D. A.; Zerbetto, F. *Angew. Chem. Int. Edit.* **2007**, *46*, 72–191.

(6) Nitschke, J. R. *Acc. Chem. Res.* **2007**, *40*, 103–112.

(7) Mastalerz, M. *Angew. Chem. Int. Edit.* **2010**, *49*, 5042–5053.

(8) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. *Angew. Chem. Int. Edit.* **2002**, *41*, 898–952.

(9) Jin, Y.; Jin, A.; McCaffrey, R.; Long, H.; Zhang, W. *J. Org. Chem.* **2012**, *77*, 7392–7400.

(10) Belowich, M. E.; Stoddart, J. F. *Chem. Soc. Rev.* **2012**, *41*, 2003–2024.

(11) Liu, X. J.; Liu, Y.; Li, G.; Warmuth, R. *Angew. Chem. Int. Edit.* **2006**, *45*, 901–904.

(12) Christinat, N.; Scopelliti, R.; Severin, K. *Angew. Chem. Int. Edit.* **2008**, *47*, 1848–1852.

(13) Skowronek, P.; Warzajtis, B.; Rychlewska, U.; Gawronski, J. *Chem. Commun.* **2013**, *49*, 2524–2526.

(14) 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. *Nat. Mater.* **2009**, *8*, 973–978.

(15) Jelfs, K. E.; Wu, X.; Schmidtmann, M.; Jones, J. T. A.; Warren, J. E.; Adams, D. J.; Cooper, A. I. *Angew. Chem. Int. Edit.* **2011**, *50*, 10653–10656.

(16) Jones, J. T. A.; Hasell, T.; Wu, X.; Bacsa, J.; Jelfs, K. E.; Schmidtmann, M.; Chong, S. Y.; Adams, D. J.; Trewin, A.; Schiffman, F.; Cora, F.; Slater, B.; Steiner, A.; Day, G. M.; Cooper, A. I. *Nature* **2011**, *474*, 367–371.

(17) Mastalerz, M.; Schneider, M. W.; Oppel, I. M.; Presly, O. *Angew. Chem. Int. Edit.* **2011**, *50*, 1046–1051.

(18) Hasell, T.; Wu, X.; Jones, J. T. A.; Bacsa, J.; Steiner, A.; Mitra, T.; Trewin, A.; Adams, D. J.; Cooper, A. I. *Nat. Chem.* **2010**, *2*, 750–755.

(19) Liu, X.; Warmuth, R. *J. Am. Chem. Soc.* **2006**, *128*, 14120–14127.

(20) Avellaneda, A.; Valente, P.; Burgun, A.; Evans, J. D.; Markwell-Heys, A. W.; Rankine, D.; Nielsen, D. J.; Hill, M. R.; Sumby, C. J.; Doonan, C. J. *Angew. Chem. Int. Edit.* **2013**, *52*, 3746–3749.

(21) Schneider, M. W.; Oppel, I. M.; Mastalerz, M. *Chem. Eur. J.* **2012**, *18*, 4156–4160.

(22) Lydon, D. P.; Campbell, N. L.; Adams, D. J.; Cooper, A. I. *Synth. Commun.* **2011**, *41*, 2146–2151.

(23) Mitra, T.; Jelfs, K. E.; Schmidtmann, M.; Ahmed, A.; Chong, S. Y.; Adams, D. J.; Cooper, A. I. *Nat. Chem.* **2013**, *5*, 276–281.

(24) Bushell, A. F.; Budd, P. M.; Attfield, M. P.; Jones, J. T. A.; Hasell, T.; Cooper, A. I.; Bernardo, P.; Bazzarelli, F.; Clarizia, G.; Jansen, J. C. *Angew. Chem. Int. Edit.* **2013**, *52*, 1253–1256.

(25) Brutschy, M.; Schneider, M. W.; Mastalerz, M.; Waldvogel, S. R. *Adv. Mater.* **2012**, *24*, 6049–6052.

(26) Hasell, T.; Chong, S. Y.; Jelfs, K. E.; Adams, D. J.; Cooper, A. I. *J. Am. Chem. Soc.* **2012**, *134*, 588–598.

(27) Hasell, T.; Zhang, H.; Cooper, A. I. *Adv. Mater.* **2012**, *24*, 5732–5737.

(28) Jelfs, K. E.; Eden, E. G. B.; Culshaw, J. L.; Shakespeare, S.; Pyzer-Knapp, 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.

(29) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J. S.; Valle, Della, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M. A.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; de Streek, van, J.; Wolf, A. K.; Schweizer, B. *Acta Crystallogr., Sect. B: Struct. Sci.* **2009**, *65*, 107–125.

(30) Price, S. L. *Acc. Chem. Res.* **2009**, *42*, 117–126.

(31) Jelfs, K. E.; Cooper, A. I. *Curr. Opin. Solid State Mater. Sci.* **2013**, *17*, 19–30.

(32) Desiraju, G. R. *Angew. Chem. Int. Edit.* **1995**, *34*, 2311–2327.

(33) Jansen, M.; Schön, J. C. *Angew. Chem. Int. Edit.* **2006**, *45*, 3406–3412.

(34) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, *4*, 83–89.

(35) Wilmer, C. E.; Farha, O. K.; Bae, Y. S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, *5*, 9849–9856.

(36) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sanchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. *J Phys Chem Lett* **2011**, *2*, 2241–2251.

(37) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(38) MacroModel, version 9.9, Schrödinger, LLC, New York, NY, 2011.

(39) Kolossvary, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011–5019.

(40) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(41) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(42) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105.

(43) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(44) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104–19.

(45) Coelho, A. A. *TOPAS Academic*, version 4.1, 2007; http://www.topas-academic.net.

(46) Sheldrick, G. M, University of Göttingen, Germany 2008.

(47) Sheldrick, G. M. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2008**, *64*, 112–122.

(48) Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H. *J. Appl. Crystallogr.* **2009**, *42*, 339–341.

(49) Jiang, S.; Bacsa, J.; Wu, X.; Jones, J. T. A.; Dawson, R.; Trewin, A.; Adams, D. J.; Cooper, A. I. *Chem. Commun.* **2011**, *47*, 8919–8921.

(50) Bojdys, M. J.; Briggs, M. E.; Jones, J. T. A.; Adams, D. J.; Chong, S. Y.; Schmidtmann, M.; Cooper, A. I. *J. Am. Chem. Soc.* **2011**, *133*, 16566–16571.

(51) Ravikumar, I.; Lakshminarayanan, P. S.; Suresh, E.; Ghosh, P. *Inorg. Chem.* **2008**, *47*, 7992–7999.

(52) A [2+2] catenane would also give the same molar masses, however the relatively simple 1H NMR suggested a non-catenated product.

(53) Le Bail, A.; Duroy, H.; Fourquet, J. L. *Mater. Res. Bull.* **1988**, *23*, 447–452.

(54) Barbour, L. J. *Chem. Commun.* **2006**, 1163–1168.

(55) Chakrabarty, R.; Mukherjee, P. S.; Stang, P. J. *Chem. Rev.* **2011**, *111*, 6810–6918.