Computationally guided synthesis of a hierarchical [4[2+3]+6] porous organic 'cage of cages'

Qiang Zhu,[1,2] Hang Qu,[1] Gokay Avci,[3] Roohollah Hafizi,[4] Chengxi Zhao,[1,5] Graeme M. Day,[4] Kim E. Jelfs,[3] Marc A. Little*[6] and Andrew I. Cooper*[1,2]

[1] Materials Innovation Factory and Department of Chemistry, University of Liverpool, 51 Oxford Street, Liverpool, L7 3NY UK.

[2] Leverhulme Research Centre for Functional Materials Design, University of Liverpool, 51 Oxford Street, Liverpool, L7 3NY UK.

[3] Department of Chemistry, Imperial College London, Molecular Sciences Research Hub, White City Campus, Wood Lane, London, W12 0BZ, UK.

[4] Computational Systems Chemistry, School of Chemistry, University of Southampton, SO17 1BJ, UK

[5] Key Laboratory for Advanced Materials and Joint International Research Laboratory of Precision Chemistry and Molecular Engineering, Feringa Nobel Prize Scientist Joint Research Center, Frontiers Science Center for Materiobiology and Dynamic Chemistry, Institute of Fine Chemicals, School of Chemistry and Molecular Engineering East China University of Science and Technology, Shanghai 200237, China.

[6] Institute of Chemical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS UK.

e-mail: m.little@hw.ac.uk; aicooper@liverpool.ac.uk

ABSTRACT

We report a two-step, hierarchical synthesis that assembles a trigonal prismatic organic cage into a more symmetrical, higher-order tetrahedral cage, or 'cage of cages'. Both the pre-formed [2+3] trigonal prismatic cage building blocks and the resultant tetrahedral **[4[2+3]+6]cage** molecule are constructed using ether bridges. This affords the **[4[2+3]+6]cage** molecule excellent hydrolytic stability that is not a feature of more common dynamic cage linkers, such as imines. Despite its relatively high molar mass (3001 g mol^{-1}), **[4[2+3]+6]cage** exhibits good solubility and crystallises into a porous superstructure with a surface area of 1056 m^2 g^{-1}. By contrast, the [2+3] building block is not porous. The **[4[2+3]+6]cage** molecule shows high CO_2 and SF_6 uptakes due to its polar skeleton. The preference for the **[4[2+3]+6]cage** molecule over other possible cage products can be predicted by computational modelling, as can its porous crystal packing, suggesting a broader design strategy for the hierarchical assembly of organic cages with synthetically engineered functions.

INTRODUCTION

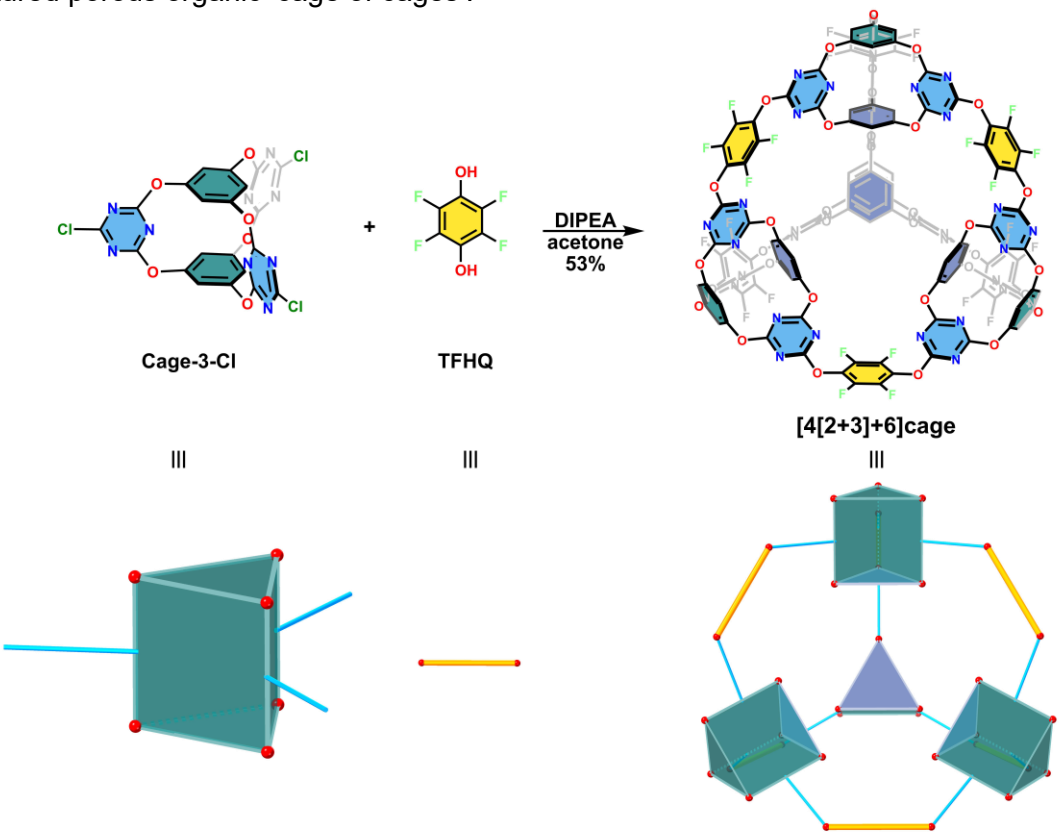
The chemical synthesis of complex organic molecules is part of our toolkit to access materials with unique structures and functions[1–5]. Supramolecular self-assembly is a powerful strategy to synthesise molecules comprising a number of separate precursors[6–8]; these assemblies can also be nanometres in size[9,10] or chemically interlocked[11,12]. However, obtaining the desired self-assembly outcomes for more complex molecules quickly becomes synthetically challenging, particularly when the bond-forming chemistry has low reversibility. This sets up a dichotomy: the more successful supramolecular reactions often lead to labile, unstable products, and this can limit the scope for applications. This can be tackled by careful tuning of precursor structure and functionality, such as molecular geometry, or by iterative optimization of the synthetic procedures, but the best reaction conditions are often not intuitively obvious.

Some of the earliest supramolecular systems were synthesised by condensing simple bidentate building blocks, such as ethylenediamine and triethylene glycol, to form cryptands and crown ethers, respectively[13]. These molecules inspired the synthesis of larger and more complex architectures. For example, Fujita introduced the concept of emergent behaviour in the assembly of large self-assembled macrocyclic products using carefully designed precursors[14]. Such supramolecular design strategies have allowed us to synthesise more complex self-assembled structures and, hence, to unlock new applications[2,15,16]. However, high structural complexity is often accompanied by increased synthetic challenges and significant unpredictability because of sensitivity to parameters such as the precise bond angles in the precursors[9,14,17].

Post-synthetic modifications have been used previously to enhance the porosity of organic cages[18,19], such as by hooping parts of the cage together[20]. More recently, we and others have used hierarchical assembly strategies to form topologically complex hydrogen-bonded organic frameworks (HOFs)[21,22] and covalently bonded materials, such as covalent organic frameworks (COFs)[23–26], using three-dimensional organic cages as the building blocks[27]. These studies have shown that cage-based building blocks can assemble into higher-order structures and increase the complexity of the resulting materials, for instance, by controlling network topology and interpenetration, while still offering a degree of structural predictability. In turn, this has afforded cage-based HOFs and 3-D cage-based COFs with properties such as guest-responsive structural flexibility[23] and self-healing behaviour[28]. However, to our knowledge, this hierarchical structuring approach has not been extended to the preparation of porous organic cage molecules[18,29]: that is, to synthesise larger porous cages from smaller organic cage precursors.

The use of organic cages as precursors to synthesise higher-order porous structures is attractive because it embeds cage molecules, with their own chemical complexity, into larger, hierarchical cages with the potential to create new functions while retaining useful properties such as solution processability[19,27,30]. For example, this strategy might produce porous materials with more sophisticated hierarchical porosities. To tackle this goal, we considered three criteria: (i) geometry – the cage precursors need geometries that can be arranged into a higher-order structure in a useful yield; (ii) chemical stability – the chemical bonding in the cages must not be too labile, both to impart stability for applications and also to avoid the dynamic scrambling that might occur, for example, in trying to construct an imine cage from another imine cage[31]; (iii) rigidity – the precursors need sufficient rigidity to direct chemical reactivity to the desired product and to ensure that the resultant hierarchical cage is shape persistent and retains its porous structure after removal of solvent from the voids.

To meet these three criteria, we chose a trigonal prismatic [2+3] ether-bridged cage molecule, **Cage-3-Cl**, as the polyhedral building block to construct a hierarchical 'cage of cages' (Scheme 1). The preconfigured rigid geometry and excellent chemical stability of **Cage-3-Cl** allowed this [2+3] cage to assemble with tetrafluorohydroquinone (**TFHQ**) into an organic cage compound, **[4[2+3]+6]cage**. To our knowledge, this is the first example of a hierarchically structured porous organic 'cage of cages'.



Scheme 1. Synthetic route for the [4[2+3]+6]cage molecule. The triangular prism and the yellow sticks in the lower scheme represent **Cage-3-Cl** and **TFHQ**, respectively.

RESULTS AND DISCUSSION

Nucleophilic aromatic substitution (S_NAr) reactions have been reported to undergo reversible covalent bond formation when using electron-poor aromatic compounds[32–34], while still leading to stable molecular products. Reversible error-correction is important for the formation of complex molecules that must self-sort during the reaction from a variety of possible products. Although the S_NAr reaction has been used in the synthesis of ether-bridged cages, most tend to be [2+3] or [2+4] cage products with small intrinsic cavities[35–37], with the exception of a larger [4+6] ether-linked cage reported by Carrillo *et al*.[32]. One possible reason for the lack of larger cages synthesised via S_NAr chemistry is the less predictable orientation of the ether bridges, as compared to the imines and boronate esters where larger cages are more commonplace[10,38–42].

Previous investigations by our group and others have demonstrated that **Cage-3-Cl** has a highly symmetrical and rigid triangular prism geometry both in solution and in the solid-state[21,36]. This geometry makes **Cage-3-Cl** an ideal building block for forming higher-order cage molecules, such as molecular barrels[20]. The three residual Cl atoms exhibit high reactivity[43,44], which is essential for forming ether bridges. We selected **TFHQ** as the linear bridge between **Cage-3-Cl** molecules because the fluorine atoms might afford extra barriers to restrict the rotation of the ether bridges, and might improve the solubility of the resulting cage-cage molecules[36,45].

To explore the available bond angles and the relative flexibility of the ether bridges in possible hierarchical cage products, we performed molecular dynamics (MD) and density functional theory (DFT) calculations. Models were constructed with the supramolecular toolkit (*stk*) software[46] to predict the most likely reaction products. As shown in Figure 1, the [4[2+3]+6] stoichiometry is predicted to form a stable, shape-persistent cage structure that exhibits a much lower energy than alternative [2[2+3]+3] and [8[2+3]+12] topologies. The [2[2+3]+3] topology has by far the highest relative energy (660.8 kJ mol^{-1}) due to its highly strained geometry. The [8[2+3]+12] topology has higher relative energy (24.04 kJ mol^{-1}) than the [4[2+3]+6] cage, which suggests that the [4[2+3]+6] topology is the thermodynamically favoured product, although we stress that these calculations do not include any solvent effects. As such, the [8[2+3]+12] topology might also be accessible under other synthesis conditions, whereas we predict that the [2[2+3]+3] topology is not. The *cis-trans* configurations of the ether bridges in the hypothetical [8[2+3]+12]cage can result in various positional configurations; all of these structural conformers were predicted to have relative energies that were between 24.0–229.1 kJ mol^{-1} higher than the [4[2+3]+6]cage, indicating a strong preference for the [4[2+3]+6] product (see Section 1, Figure S1-4).

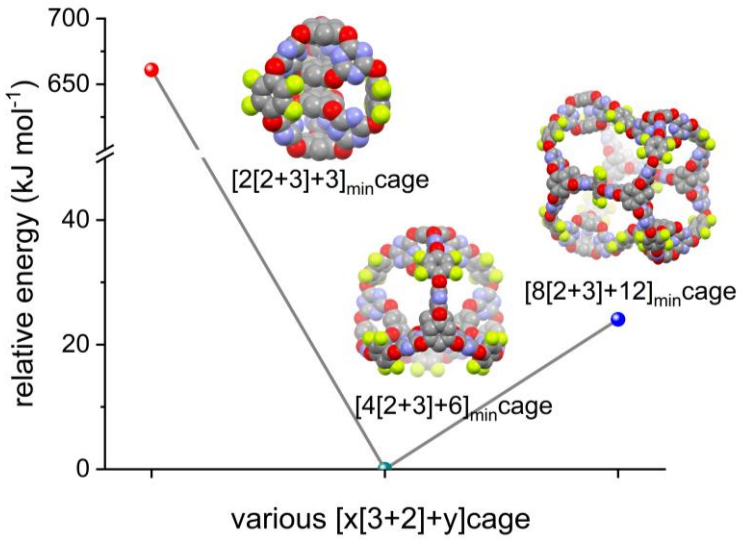


Figure 1. Relative density functional theory (DFT) energies for minimum energy configurations for [x[2+3]+y] cages. x = number of **Cage-3-Cl** cages, y = number of **TFHQ** linkers. Atom colour: C: grey, N: blue, O: red, and F: green. H atoms are omitted for clarity. Note the break in the energy scale for the highly strained [2[2+3]+3]cage.

These simulation results suggested that it might be possible to synthesise **[4[2+3]+6]cage** via the S_NAr reaction between **Cage-3-Cl** and **TFHQ** (Scheme 1). We, therefore, attempted the reaction experimentally, and screened a range of conditions where we varied the reagent concentration, solvent, and base (Table S1). From these experiments, we found that the reaction in acetone in the presence of the acid scavenger, *N*,*N*-diisopropylethylamine (DIPEA), afforded a new product with the highest yield of 53% after purification. The ^{1}H NMR spectrum for the purified reaction product from the acetone reaction with DIPEA showed two singlets at 7.09 and 6.85 ppm, which we assigned to the two aromatic protons in the [3+2] cage (H_a and H_b, Figure 2a, S5). The presence of two singlets indicates different environments, which we attribute to one of the protons being more shielded. However, apart from this splitting of the aromatic proton singlet in **Cage-3-Cl**, the NMR indicated that the resulting product had high symmetry in solution. In the ^{13}C NMR spectrum, we observed three signals in the 174.5–173.1 ppm range (Figure 2b, S6), which we assigned to the triazine ring carbon atoms. We attribute the characteristic splitting, observed at 142.5 and 140.0 ppm with a coupling constant of 250 MHz, to the coupling between the carbon and fluorine atoms in the **TFHQ** linker (Figure 2b, S6). We also confirmed the presence of these fluorinated aromatic rings by ^{19}F NMR spectroscopy, observing a singlet at -155.62 ppm (Figure S7), indicating that the fluorine atoms were symmetrically equivalent in solution. We also used matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry to analyse the reaction

product. We found an ion with a mass (M+H) at 3002.0756 (Figure 2c, S8, S9), which matched well with the theoretical value of 3002.0871, indicating the formation of **[4[2+3]+6]cage**.

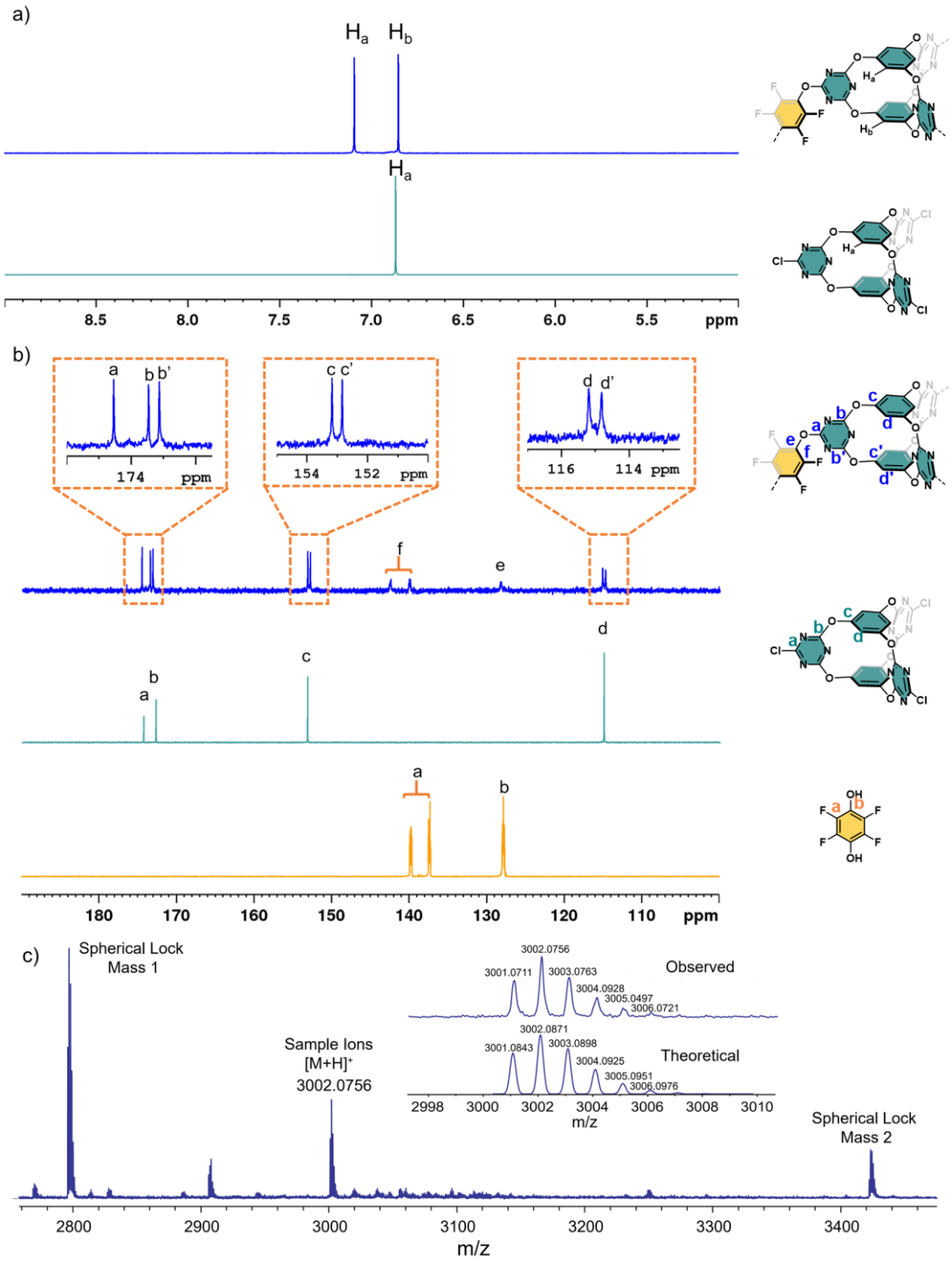


Figure 2: NMR spectra and molecular structures of [4[2+3]+6]cage, Cage-3-Cl and TFHQ: (a) ^{1}H NMR (400 MHz, acetone-d_6) spectra: **Cage-3-Cl** (green, bottom) and **[4[2+3]+6]cage** (blue, top). (b) ^{13}C NMR (100 MHz, dioxane-d_8) spectra: **TFHQ** (yellow, bottom), **Cage-3-Cl** (green, middle) and **[4[2+3]+6]cage** (blue, top) Insets, zoom-ins of boxed regions. (c) High-resolution MALDI-TOF spectra of **[4[2+3]+6]cage**. Two internal calibrants (SphericalTM) with *m*/*z*'s = 2979 and 3423 that bracketed the ion of interest were used to limit the *m*/*z* error to ± 5 ppm.

We next grew crystals for single-crystal X-ray diffraction analysis to confirm the structure of the **[4[2+3]+6]cage** molecule. Slow evaporation of a mixture of acetone/ethanol afforded single crystals suitable for X-ray analysis using synchrotron radiation (Figure S10, Table S2). The synchrotron single crystal structure, which we refined in the monoclinic $P2_1$ space group, revealed that the **[4[2+3]+6]cage** molecule adopts a tetrahedral topology, where four **Cage-3-Cl** cage molecules serve as the vertices and six **TFHQ** molecules are located as the edges (Figure 3a). The interior and the exterior aryl caps of the **Cage-3-Cl** cage molecules form a core-shell structure, defining an inner and outer truncated tetrahedron with edge lengths of 6.4 and 13.7 Å, respectively (Figure 3b). We also calculated the electrostatic potentials for the **[4[2+3]+6]cage** molecule, which showed that the centre of the **[4[2+3]+6]cage** molecule is surrounded by aromatic rings, affording π-π interactions for any guest molecules within the cage (Figure 3c, see section 1 of the ESI).

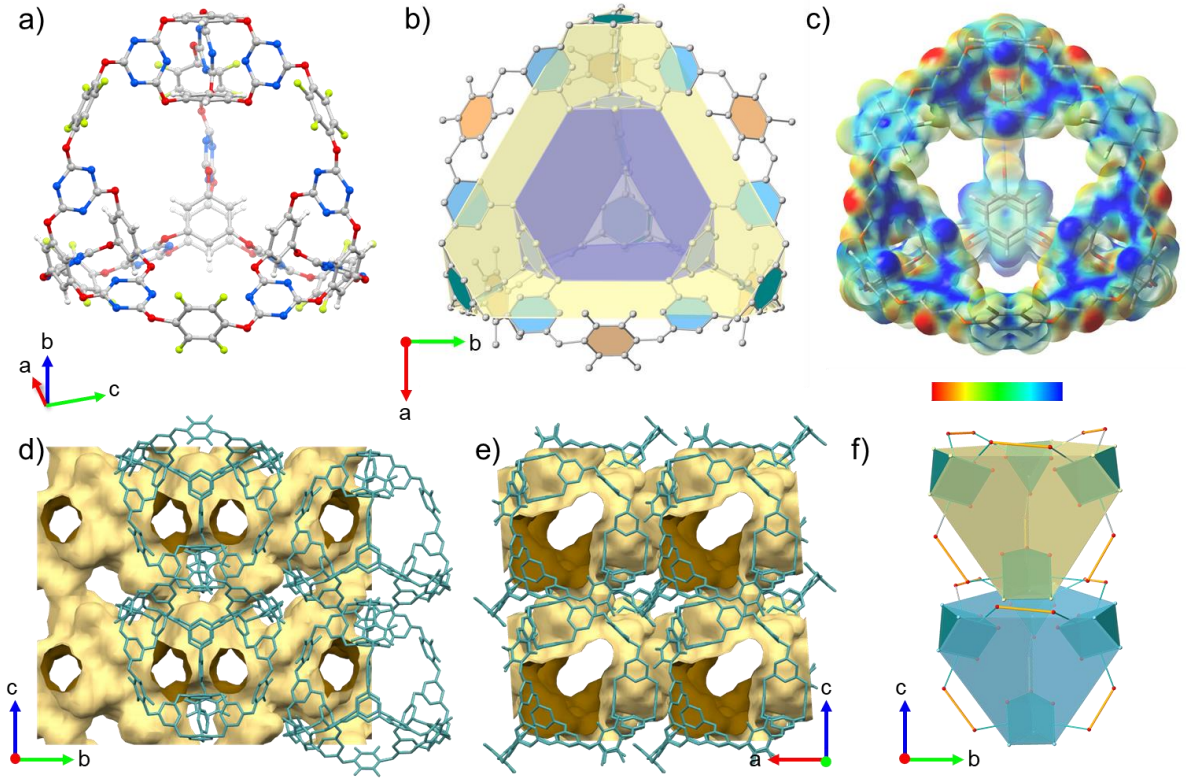


Figure 3: Crystal structure of [4[2+3]+6]cage: (a) Structure of an individual **[4[2+3]+6]cage** molecule. Atom colours: C: grey; H: white; N: blue; O: red; F: green. (b) Representation of the **[4[2+3]+6]cage** molecule using two truncated tetrahedra on the inner and outer aryl caps of the [2+3] **Cage-3-Cl** cage molecules. For clarity, all atoms here are coloured grey. (c) Electrostatic potential maps of the **[4[2+3]+6]cage** molecule. The red and blue surfaces represent negative and positive regions of potential, respectively. (d, e) Pore channels in the extended **[4[2+3]+6]cage** crystal structure as viewed along the *a*-axis and *b*-axis, respectively. For clarity, H atoms are omitted in Figures b, e, and f. The yellow surfaces in d and e represent the contact surface as measured using a 1.2 Å diameter probe. (f) Scheme explaining the window splitting in the **[4[2+3]+6]cage** crystal structure along the *a*-axis; the window of the lower blue cage is partially occluded by the aryl face of the upper, yellow cage.

The interior of the cage core exhibits electron-poor character because of the V-shaped electron-deficient clefts formed by the triazine rings of **Cage-3-Cl** and the fluorine-decorated aromatic rings. This environment might be useful for selective guest molecule separation[47–49]. In the extended crystal structure of this cage-of-cages, the asymmetric cell contains one **[4[2+3]+6]cage** molecule, which assembles into a porous supramolecular structure by interacting with 12 neighbouring **[4[2+3]+6]cage** molecules through van der Waals forces (Figure S11). Two of the windows in the **[4[2+3]+6]cage** molecule are narrowed into smaller channels by the **Cage-3-Cl** vertices from neighbouring cage molecules (Figure 3d, f, S12), yielding 3D interconnected pore channels (Figure 3d, e). Using Zeo++[50], we calculated that the pore limiting diameter of the **[4[2+3]+6]cage** crystal structure was 6.4 Å and the largest cavity diameter was 8.9 Å (Table S3, Figure S13–15), suggesting that the structure is microporous. From these calculations, we also determined that voids in the **[4[2+3]+6]cage** crystal structure that is accessible to a 1.65 Å CO_2 probe occupy 32.0% of the unit cell volume (Table S3).

There was strong agreement between the predicted structure for the [4[2+3]+6]cage molecule and the molecule observed in the crystal structure (Figure 4). This validates the theoretical predictions, and the close match between the CSP-predicted structure and experimental crystal structure adds confidence in the crystal structure refinement (Figure S17). The root mean squared displacement (RMSD) was calculated as 0.5 Å with a maximum distance between atoms of 1.4 Å. However, the experimental displacement parameters are large due to disorder in the crystal structure (Figure S11a). Further attempts to synthesise the larger [8[2+3]+12] product by varying the reaction conditions were unsuccessful, based on MALDI-TOF analysis of the resulting products (Table S1, Figure S8), in line with the molecular stability predictions (Figure 1).

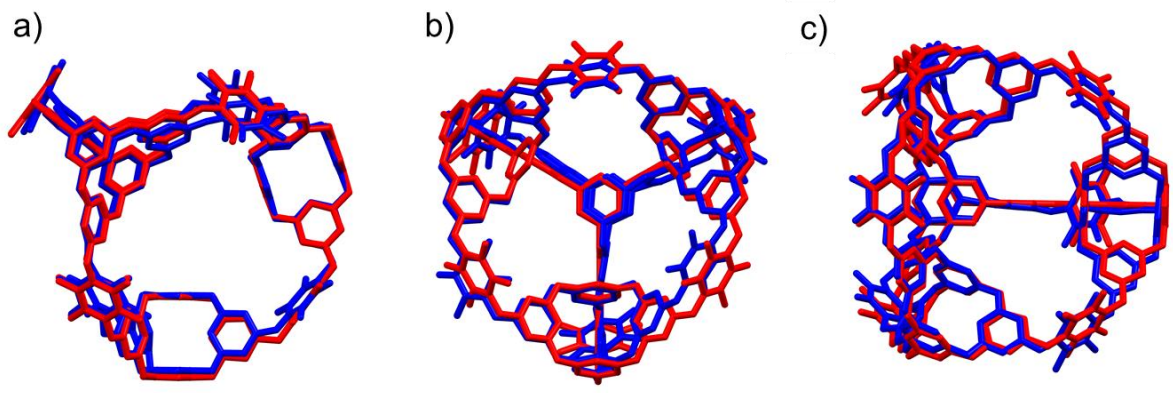


Figure 4: Comparison between predicted [4[2+3]+6]cage model and experimental crystal structure. The predicted structure (red) overlaid with the single crystal X-ray

diffraction structure (blue) is shown as viewed along the (a) *a*, (b) *b,* and (c) *c*-crystallographic axes.

In principle, catenation of this cage is possible, given its large intrinsic voids (>10 Å diameter), as observed for significantly smaller imine cages[11]. However, we saw no evidence for catenated cage side-products, either by NMR or by MALDI-TOF characterisation.

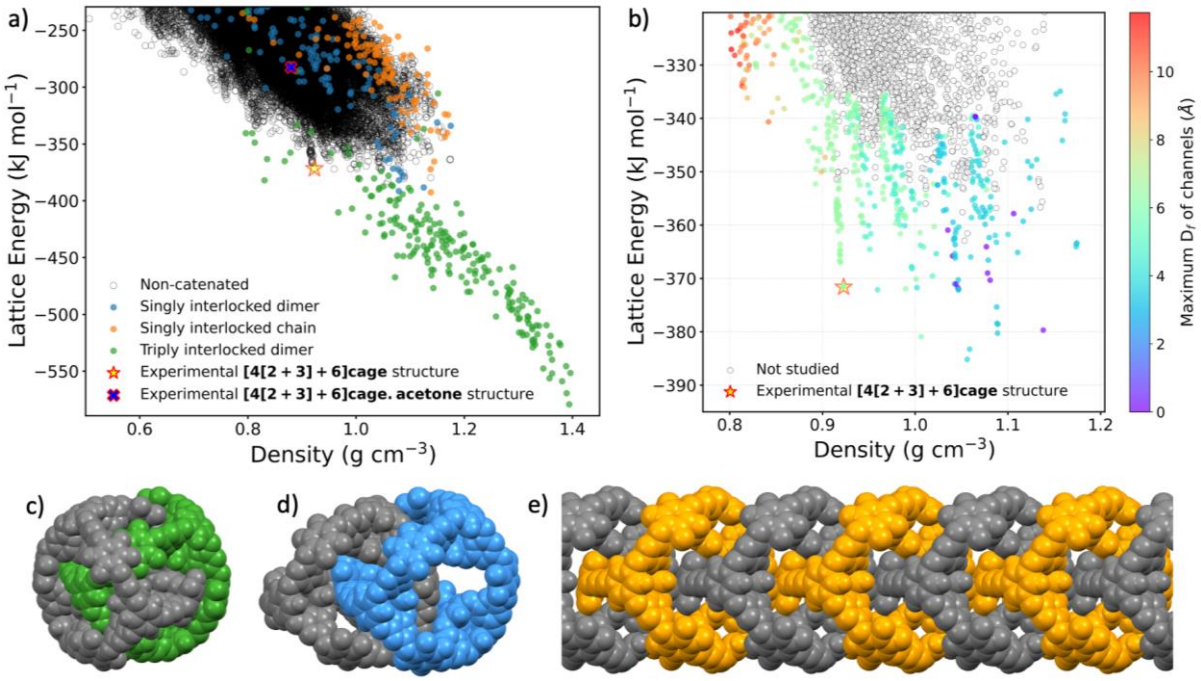


Figure 5: Crystal structure prediction for [4[2+3]+6]cage. (a) Computational crystal energy landscape of **[4[2+3]+6]cage** with colour-coded categorisation based on catenation type: discrete, non-catenated cages (uncoloured circles), triply interlocked cage dimers (green circles), singly interlocked cage dimers (blue), and singly interlocked 1-D cage chains (orange). The yellow star and blue cross represent the predicted structures matching the experimentally observed **[4[2+3]+6]cage** crystal structure and **[4[2+3]+6]cage·acetone** solvated structure, respectively. (b) Energy landscape after removal of the catenated structures, with color-coding based on the diameter of the largest sphere capable of freely moving within the crystal structure's channel(s). Channels are found based on their ability to accommodate a CO_2 molecule. D_f = 0 corresponds to no channel being found. Atomic structures depicted for examples of (c) a triply interlocked cage dimer, (d) a singly interlocked cage dimer, and (e) a singly interlocked 1-D cage chain.

We next used crystal structure prediction (CSP) to explore the solid-state packing of these hierarchical cages. The lattice energy landscape was explored using quasi-random sampling of the crystal packing space with the Global Lattice Energy Explorer (GLEE)[51]. Initial trial structures were generated from rigid molecules and subjected to lattice energy minimisation

employing an empirically parametrized potential with atomic multipole electrostatics[52] (see SI Section 4, Tables S4–5, and Figures S16–25 for full details).

Surprisingly, the CSP landscape for **[4[2+3]+6]cage** (Figure 5) showed catenated structures, along with the non-catenated cage that was observed experimentally, even though the discrete **[4[2+3]+6]cage** molecule was used for the CSP calculations. Three distinct catenations were identified in the predicted crystal structures: triply interlocked cage dimers (Figure 5c), singly interlocked cage dimers (Figure 5d), and singly interlocked 1-D cage chains[12,53] (Figure 5e). The details of the methods used for catenation detection are provided in SI Section 4 and Figures S18–20. All sampled structures within a 197 kJ mol^{-1} energy window from the global energy minimum were found to be catenanes (Figures S21–22), indicating a strong thermodynamic preference over the non-catenated cages observed by experiment. To verify the relative energies calculated using the rigid-molecule, force field approach, a selection of catenated and non-catenated predicted structures were re-evaluated using periodic DFT, which confirmed this greater thermodynamic stability (see SI Section 4 for full details).

While the CSP study did not explicitly target catenated structures, the sampled catenated configurations suggest that triply interlocked catenanes (green points, Figure 5a), in particular, might be much more thermodynamically stable in the solid state. This echoes previous findings for [4+6] imine cages, where discrete cages were found to transform into triply interlocked catenanes upon exposure to acid, suggesting that the individual cages were the kinetic rather than the thermodynamic product[11]. The absence of catenanes in our experiments might be explained by the much lower reversibility of the ether bonding in the **[4[2+3]+6]cage** molecule, which is not accounted for in the CSP calculations. Prompted by these solid-state CSP results, we also explored the relative thermodynamic stability of catenanes at the molecular level. DFT calculations of catenane dimers showed that the energy difference between the molecular equivalent non-catenated **[4[2+3]+6]cage** dimer and trimer fragments retrieved from the global lowest energy CSP, and the corresponding triply interlocked catenane molecular fragment was 373.7 kJ mol^{-1} and 324.7 kJ mol^{-1} respectively, reaffirming strong thermodynamic favour towards the catenane structures.

When we remove the catenated structures from the CSP plot (Figure 5b, Figure S23), this reveals the observed experimental structure positioned at the bottom of a low-density ‘spike’ in the energy landscape, approximately 13.6 kJ mol^{-1} higher than the global energy minimum for non-catenated cages. The predicted crystal structure reproduces the geometry of the experimentally-determined **[4[2+3]+6] cage** crystal structure accurately (Figure S17), confirming that the crystal structure determined by X-ray diffraction corresponds to a low-energy local minimum in lattice energy. The colour-coding in this ‘non-catenated’ crystal

structure landscape represents the diameter of the largest sphere capable of unrestricted movement within the crystal structure channels. Channel dimensions are determined based on their capacity to accommodate a CO_2 molecule with a kinetic radius of 1.65 Å (Figures S24–25). In the landscape depicted in Figure 5b, void analysis has been restricted to structures within 20 kJ mol^{-1} of the low energy edge of the energy-density distribution of structures. Except for a very small number of predicted structures (the purple points, Figure 5b), all investigated structures, including the synthesised structure, show potential for CO_2 uptake. That is, CSP suggests that **[4[2+3]+6]cage** has an intrinsic propensity to be porous in the majority of its potential crystalline packing modes.

Molecular crystals exhibiting permanent porosity in the solid state are attractive for practical applications such as gas capture, separation, and catalysis[18,54]. One successful approach that we and others have developed is to form porous organic crystals by synthesizing cages with prefabricated shape-persistent cavities that are retained after solvents are removed during activation[18,54,50]. Our calculations revealed that the ether bridges in the **[4[2+3]+6]cage** skeleton appeared to be relatively rigid, suggesting shape persistence. We therefore investigated the porosity in the **[4[2+3]+6]cage** crystals using gas sorption analysis. We activated the **[4[2+3]+6]cage** crystals by first exchanging the ethanol and acetone crystallisation solvents with diethyl ether or *n*-pentane, which we chose because of their low surface tensions. Then, we removed any residual solvent from the crystals under a dynamic vacuum at room temperature. Subsequent powder X-ray diffraction (PXRD) analysis revealed that the **[4[2+3]+6]cage** crystals retained some crystallinity after being activated using these conditions (Figure S26). The **[4[2+3]+6]cage** crystals activated via the diethyl ether solvent exchange route appeared more crystalline, and this sample was used for the subsequent gas sorption experiments described here.

Nitrogen sorption isotherms recorded at 77 K and found that the crystalline **[4[2+3]+6]cage** exhibits a Type-I N_2 sorption isotherm with a relatively high Brunauer-Emmett-Teller (BET) surface of 1056 $m^2 g^{-1}$ (Figures 6a, S27–29), consistent with a microporous solid and the pore size distribution plot calculated using Zeo++[51] (Table S3, Figure S13). We found that crystalline **[4[2+3]+6]cage** has a CO_2 uptake capacity of 3.98 mmol g^{-1} at 1 bar and 273 K (Figure 6b, S30). This CO_2 uptake is high compared to other porous organic crystalline materials, such as COFs[55], at comparable temperatures and pressures, and is one of the highest CO_2 uptakes reported to date for a porous organic cage (Table S6)[56,57]. The calculated isosteric heat of adsorption of CO_2 on crystalline **[4[2+3]+6]cage** ranges between 21.1 to 23.2 kJ mol^{-1} (Figure S31), which indicates a strong affinity between the adsorbed CO_2 gas and polar **[4[2+3]+6]cage** crystal pores, rationalizing this high uptake capacity. In addition, we found that crystalline **[4[2+3]+6]cage** has a high SF_6 capacity of 3.21 mmol g^{-1} at 1 bar and 273 K

(Figure S32). The calculated isosteric heat of adsorption of SF_6 on crystalline **[4[2+3]+6]cage** ranges between 29.2 to 29.5 kJ mol^{-1}, which again indicates a strong affinity between adsorbed SF_6 gas molecules and the **[4[2+3]+6]cage** crystal pores (Figure S33). Analysis of the **[4[2+3]+6]cage** powder after the gas sorption isotherms by PXRD analysis indicated that the material remained crystalline during these measurements (Figure S34).

We also uncovered a second crystal structure of the **[4[2+3]+6]cage** molecule during this study, referred to as **[4[2+3]+6]cage·acetone**, which crystallised from slow evaporation of an acetone-d_6 solution (Figure S35). **[4[2+3]+6]cage·acetone** crystallised in the cubic space group $I\bar{4}3m$ (a = 23.2901(15) Å, V = 12633(2)Å^3, Table S7) with the ether-bridged cage adopting a perfect tetrahedral geometry in the structure (Figure S36). The **[4[2+3]+6]cage·acetone** lost crystallinity rapidly after being removed from the acetone-d_6 solvent and cracked (Figure S35). We therefore performed single crystal analysis by sealing a solvated crystal in a borosilicate capillary containing residual acetone-d_6 solvent. However, due to the poorer crystal stability of **4[2+3]+6]cage·acetone**, we did not investigate its solid-state properties further. The instability of this form was further investigated through computational geometry optimization of the crystal structure. Employing the FIT+DMA force field for rigid body geometry optimisation of the structure after solvent removal, the structure distorted significantly from its cubic lattice, adopting a monoclinic form, in keeping with the observed experimental instability. Details can be found in section S8 of ESI. The relaxed structure, denoted by a blue cross in the landscape of Figure 5a, is situated 103 kJ mol^{-1} above the global energy minimum on the landscape of non-catenated structures. This energy difference underscores the crucial role of solvent stabilisation in the synthesis of this solvated structure, and can also help to rationalise why this tetrahedral molecular structure was not predicted using gas phase (*i.e.*, solvent-free) DFT calculations (Figure 4).

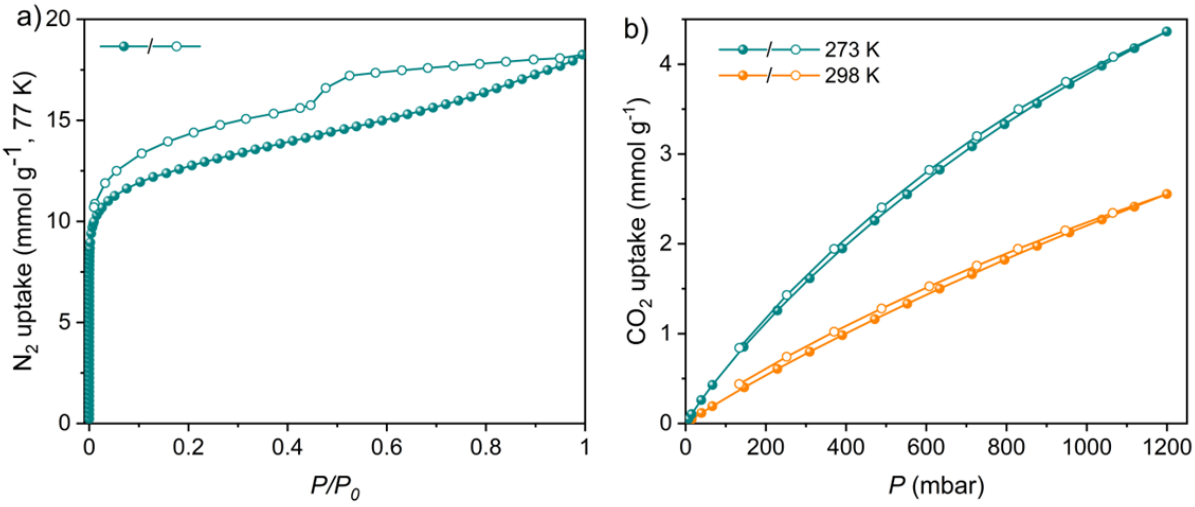


Figure 6: Gas sorption isotherms for [4[2+3]+6]cage. a) N_2 at 77 K; b) CO_2 at 273 K (cyan) and 298 K (orange). Closed and open symbols are adsorption and desorption isotherms, respectively.

For practical applications, gas sorption capacity is not the only criterion. For example, most CO_2 capture applications involve wet or humid gas streams, and hence water stability is important. Many porous organic cage materials are unstable to water, such as imine cages and (particularly) boronate ester cages. We, therefore, explored the hydrolytic stability of the **[4[2+3]+6]cage** molecule by immersing the synthesised crystals in water for 12 days. Subsequent analysis of the sample by ^{1}H NMR spectroscopy revealed that **[4[2+3]+6]cage** remained chemically intact under these conditions (Figure S38). PXRD analysis of the same sample also revealed that the **[4[2+3]+6]cage** crystals retained their crystallinity under these conditions (Figure S39). Hence, both the chemical and crystal structure of **[4[2+3]+6]cage** molecule appear to have good hydrolytic stability.

CONCLUSIONS

In summary, we report the assembly of a new type of porous organic cage—a 'cage of cages'—that was synthesised using a two-step hierarchical self-assembly strategy. As far as we know, this is the first example of using organic cages to synthesise more complex porous organic cages. In this first example, we demonstrate the strategy by assembling four trigonal cages into a larger tetrahedral cage. The resulting **[4[2+3]+6]cage** molecule exhibits excellent stability in water, and crystals of the **[4[2+3]+6]cage** shows permanent porosity and a high surface area of 1056 $m^2 g^{-1}$. The abundance of polar atoms in the cage cavity endows it with high CO_2 and SF_6 uptake capacity. The good solubility of **[4[2+3]+6]cage** in acetone indicates it has the potential to be used as a building block for even more complex structures, such as porous cage cocrystals. More broadly, this illustrates a strategy for hierarchical molecular assembly using computation as a guide to assess the most likely reaction products. For example, it might be possible in the future to design analogous systems where the [2+3] cages contribute discrete, prefabricated porosity into a higher-order, hierarchically porous crystal.

This study also showcases the use of computational design in supramolecular synthesis, both at the molecular level (Figure 4) and in the solid state (Figure 5). It is notable that triply interlocked cage catenane dimers emerged as the most stable predicted crystal packings (Figure 5a). Such catenanes were not observed in experiments, most likely because they are kinetically disfavoured, but they are nonetheless synthetically plausible because analogous structures have been formed using more reversible [4+6] imine cage-forming reactions[11]. Less obviously, infinite 1-D catenated cage chains are also produced in these simulations (Figure 5e), and in some cases these structures are predicted to have similar lattice energies than the

experimentally-observed non-catenated cage (Figure 5a). This highlights how *a priori* structure predictions have the power to suggest non-intuitive new materials, although it is unclear how one might design a kinetic pathway to these chain structures, even though analogous structures have been observed for less complex macrocycles[53].

METHODS

Molecular simulations

Both **Cage-3-Cl** and cage-of-cages models were constructed in Tri2Di3, Tri4Di6, Tri8Di12 topologies using the stk software[46]. All cages were annealed with a molecular dynamics simulation at 700 K for 50 ns with a timestep of 0.5 fs after a 100 ps equilibration time with OPLS4 force field as implemented in the Macromodel Suite[58]. Five hundred random configurations from the total MD duration were sampled and energy minimised, with the lowest energy configuration selected for Density Functional Theory (DFT) calculations. DFT calculations were performed with CP2K 2023.1[59] version software using the Generalised Gradient Approximation (GGA) theory with the PBE functional[60] and def2-TZVP basis sets[61]. A planewave cut-off value of 400 Ry and a relative cut-off value of 100 Ry were parameterised to obtain converged energy levels and dispersion interactions were accounted for with Grimme's DFT-D3 approach[62].

The geometries of the [4[2+3]+6]cage were then fully optimised by means of the hybrid M06-2X functional9 in Gaussian16[63]. The def2-SVP basis set[64,65] was applied for all atoms. No symmetry or geometry constraint was imposed during optimisations. The optimised geometries were verified as local minima on the potential energy surface by frequency computations at the same theoretical level[63].

Synthesis of [4[2+3]+6]cage

To synthesise **[4[2+3]+6]cage**, DIPEA (61 µL, 0.35 mmol) was dissolved in acetone (25 mL) and purged with N_2 for 10 minutes. To the acetone solution, a mixture of **Cage-3-Cl** (58.7 mg, 0.1 mmol) and tetrafluorohydroquinone (**TFHQ**, 27.3 mg, 0.15 mmol) in acetone (6 mL) was added dropwise over 3 hours under a N_2 atmosphere. After the addition was complete, the reaction was stirred at room temperature for 36 hours. The solvent was then removed by rotary evaporation, and the crude product was purified by column chromatography using acetone/CH_2Cl_2 (10% *vol*/*vol* acetone) as eluent to afford **[4[2+3]+6]cage** as a white solid in 53% isolated yield: 40 mg (0.013 mmol). ^{1}H NMR (400 MHz, acetone-d_6): δ 7.09 (s, 12H, H_b), 6.85 (s, 12H, H_a); ^{19}F NMR (376 MHz, acetone-d_6): δ -155.62; ^{13}C NMR (100 MHz, dioxane-d_8): δ 174.5, 173.5, 173.1, 153.2, 152.8, 142.5, 140.1, 140.0, 128.3, 115.2, 114.8. MALDI-TOF $[M+H]^+$, $[C_{120}H_{24}F_{24}N_{36}O_{36}+H]^+$: Calc. 3002.0871, Found 3002.0756.

Crystal structure prediction

CSP involves the following general steps: (i) molecular geometry optimisation; (ii) trial crystal structure generation; (iii) local lattice energy minimisation of trial structures; and (iv) duplicate removal.

The geometry of the molecular cage was optimised at the B3LYP/6-311G(d,p) level using the Gaussian09 software[66], and the resulting geometry was kept fixed throughout the subsequent steps. Trial crystal structures are generated using the Global Lattice Energy Explorer (GLEE) code[51]. Subsequently, these trial structures undergo lattice optimisation while preserving the rigidity of the molecular cage. For this task, we employ an empirically parametrised intermolecular atom-atom exp-6 potential coupled with atomic multipole electrostatics. The force field parameters are acquired from the FIT force field[67,68]. Atom-centred multipoles up to hexadecapole on each atom were derived from the electron density through distributed multipole analysis (DMA), and partial charges (used in early stages of optimisation) were fitted to the molecular electrostatic potential generated by these multipoles[69,70]. The overall model is denoted as FIT+DMA.

The search for space groups involves sampling the 10 most common space groups for organic crystals along with four trigonal space groups (143, 144, 145, and 146), each with one molecule in the asymmetric unit. A quasi-random method is employed to search these selected space groups separately, and valid structures are lattice energy minimised using the DMACRYS software[52] in a two-stage protocol. The first stage involves FIT+DMA with partial charges, followed by the second stage with multipole electrostatics. More details can be found in the supporting information.

DATA AVAILABILITY

The authors declare that the data supporting the findings of this study are available within the paper, its supplementary information files, and the Cambridge Crystallographic Data Centre (deposition numbers: 2303319 for **[4[2+3]+6]cage** and 2326368 for **[4[2+3]+6]cage·acetone**). The crystal structure prediction data can be found at: https://doi.org/10.5258/SOTON/D2929.

SUPPLEMENTARY INFORMATION

Supplementary Information: Figures S1–39, Tables S1–8, Schemes S1, synthetic procedures and methods, molecular simulation, NMR, MALDI-TOF MS, powder X-ray diffraction, single crystal X-ray diffraction, crystal structure prediction, and gas sorption analysis.

Supplementary Data 1: Simulated structures of **[2[3+2]+3]cage**, **4[3+2]+6]cage**, **[8[3+2]+12]cage**, **[8[3+2]+12]cage_1**, **[8[3+2]+12]cage_2** and **[8[3+2]+12]cage_3**.

Supplementary Data 2: X-ray crystallographic data of **[4[2+3]+6]cage** (CIF)

Supplementary Data 3: Video showing the single crystal structure of **[4[2+3]+6]cage** (mp4)

Supplementary Data 4: X-ray crystallographic data of **[4[2+3]+6]cage·acetone** (CIF)

AUTHOR INFORMATION

Qiang Zhu — Department of Chemistry and Materials Innovation Factory, University of Liverpool, Liverpool L7 3NY, U.K.; Leverhulme Research Centre for Functional Materials Design, University of Liverpool, Liverpool L7 3NY, U.K.

Hang Qu — Department of Chemistry and Materials Innovation Factory, University of Liverpool, Liverpool L7 3NY, U.K.

Gokay Avci — Department of Chemistry, Imperial College London, Molecular Sciences Research Hub, White City Campus, Wood Lane, London, W12 0BZ, U.K.

Roohollah Hafizi — Computational Systems Chemistry, School of Chemistry, University of Southampton, SO17 1BJ, U.K.

Chengxi Zhao — Department of Chemistry and Materials Innovation Factory, University of Liverpool, Liverpool L7 3NY, U.K.; Key Laboratory for Advanced Materials and Joint International Research Laboratory of Precision Chemistry and Molecular Engineering, Feringa Nobel Prize Scientist Joint Research Centre, Frontiers Science Centre for Materiobiology and Dynamic Chemistry, Institute of Fine Chemicals, School of Chemistry and Molecular Engineering East China University of Science and Technology, Shanghai 200237, China.

Graeme M. Day — Computational Systems Chemistry, School of Chemistry, University of Southampton, SO17 1BJ, U.K.

Kim E. Jelfs — Department of Chemistry, Imperial College London, Molecular Sciences Research Hub, White City Campus, Wood Lane, London, W12 0BZ, U.K.

Marc A. Little — Institute of Chemical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS U.K.

Andrew I. Cooper — Department of Chemistry and Materials Innovation Factory, University of Liverpool, Liverpool L7 3NY, U.K.; Leverhulme Research Centre for Functional Materials Design, University of Liverpool, Liverpool L7 3NY, U.K.;

AUTHOR CONTRIBUTIONS

Q.Z. led the experimental work and the synthesis and characterization of the materials. A.I.C., K.E.J., and M.A.L. conceived the idea and modelling strategy with Q.Z. and supervised the project. H.Q. and M.A.L. conducted the single crystal X-ray diffraction analysis and solved the structure. G. A., K.E.J., and C.Z. performed the molecular simulations. R.H. and G.M.D. performed the crystal structure prediction, and G.M.D. supervised this part of the project. Q.Z., G.A., R.H., G.M.D., K.E.J., M.A.L. and A.I.C. analysed the data and prepared the manuscript. All authors discussed the results and contributed to the manuscript.

FUNDING

The authors received funding from the Engineering and Physical Sciences Research Council (EPSRC, EP/V026887/1) and the Leverhulme Trust via the Leverhulme Research Centre for Functional Materials Design. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant CoMMaD No 758370). AIC thanks the Royal Society for a Research Professorship (RSRP\S2\232003).

NOTES

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A.I.C. thanks the Royal Society for a Research Professorship. C.Z. acknowledges the China Scholarship Council for financial support (No.202106745008). R.H. acknowledges the Iridis 5 High Performance Computing facility, and associated support services at the University of Southampton. Via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/R029431 and EP/X035859), this work used the Archer2 HPC facility. We acknowledge Dr Ann Hunter for performing the MALDI-TOF analysis at the National Mass Spectrometry Facility (NMSF) at Swansea University and Diamond Light Source for access to beamlines I19 (CY30461).

REFERENCES

1. Lehn, J. M. Toward self-organization and complex matter. *Science* **295**, 2400–2403 (2002).

2. McTernan, C. T. *et al.* Beyond platonic: how to build metal-organic polyhedra capable of binding low-symmetry, information-rich molecular cargoes. *Chem. Rev.* **122**, 10393–10437 (2022).

3. Guillerm, V. & Eddaoudi, M. The importance of highly connected building units in reticular chemistry: thoughtful design of metal-organic frameworks. *Acc. Chem. Res.* **54**, 3298–3312 (2021).

4. Montà-González, G. *et al.* Purely covalent molecular cages and containers for guest encapsulation. *Chem. Rev.* **122**, 13636–13708 (2022).

5. Lutz, J. F. *et al.* From precision polymers to complex materials and systems. *Nat. Rev. Mater.* **1**, 16024 (2016).

6. Acharyya, K. & Mukherjee, P. S. Organic imine cages: molecular marriage and applications. *Angew. Chem. Int. Ed.* **58**, 8640–8653 (2019).

7. Deng, H. *et al.* Multiple functional groups of varying ratios in metal-organic frameworks. *Science* **327**, 846–850 (2010).

8. Pullen, S. & Clever, G. H. Mixed-ligand metal-organic frameworks and heteroleptic coordination cages as multifunctional scaffolds - a comparison. *Acc. Chem. Res.* **51**, 3052–3064 (2018).

9. Fujita, D. *et al.* Self-assembly of tetravalent Goldberg polyhedra from 144 small components. *Nature* **540**, 563–566 (2016).

10. Koo, J. *et al.* Gigantic porphyrinic cages. *Chem* **6**, 3374–3384 (2020).

11. Hasell, T. *et al.* Triply interlocked covalent organic cages. *Nat. Chem.* **2**, 750–755 (2010).

12. Benke, B. P. *et al.* Dimeric and trimeric catenation of giant chiral [8 + 12] imine cubes driven by weak supramolecular interactions. *Nat. Chem.* **15**, 413–423 (2023).

13. Core concepts in supramolecular chemistry and nanochemistry. Steed, J. W., Turner, D. R. & Wallace, K. J. John Wiley & Sons, Ltd: Chichester. 2007. *J. Am. Chem. Soc.* **129**, 14524–14524 (2007).

14. Sun, Q. F. *et al.* Self-assembled M24L48 polyhedra and their sharp structural switch

upon subtle liqand variation. *Science* **328**, 1144–1147 (2010).

15. Fujita, D. *et al.* Protein stabilization and refolding in a gigantic self-assembled cage. *Chem* **7**, 2672–2683 (2021).

16. Chakrabarty, R. *et al.* Supramolecular coordination: self-assembly of finite two- and three-dimensional ensembles. *Chem. Rev.* **111**, 6810–6918 (2011).

17. Santolini, V. *et al.* Topological landscapes of porous organic cages. *Nanoscale* **9**, 5280–5298 (2017).

18. Yang, X. *et al.* Porous organic cages. *Chem. Rev.* **123**, 4602–4634 (2023).

19. Mastalerz, M. Porous shape-persistent organic cage compounds of different size, geometry, and function. *Acc. Chem. Res.* **51**, 2411–2422 (2018).

20. Wang, Q. Q. *et al.* Molecular barrel by a hooping strategy: synthesis, structure, and selective CO_2 adsorption facilitated by lone pair-π interactions. *J. Am. Chem. Soc.* **139**, 635–638 (2017).

21. Zhu, Q. *et al.* Analogy powered by prediction and structural invariants: computationally led discovery of a mesoporous hydrogen-bonded organic cage crystal. *J. Am. Chem. Soc.* **144**, 9893–9901 (2022).

22. Han, B. *et al.* Postsynthetic metalation of a robust hydrogen-bonded organic framework for heterogeneous catalysis. *J. Am. Chem. Soc.* **141**, 8737–8740 (2019).

23. Zhu, Q. *et al.* 3D Cage COFs: A dynamic three-dimensional covalent organic framework with high-connectivity organic cage nodes. *J. Am. Chem. Soc.* **142**, 16842–16848 (2020).

24. Ji, C. *et al.* Tunable cage-based three-dimensional covalent organic frameworks. *CCS Chem.* **4**, 3095–3105 (2022).

25. Ma, J.-X. *et al.* Cage based crystalline covalent organic frameworks. *J. Am. Chem. Soc.* **141**, 3843–3848 (2019).

26. Kory, M. J. *et al.* Gram-scale synthesis of two-dimensional polymer crystals and their structure analysis by X-ray diffraction. *Nat. Chem.* **6**, 779–784 (2014).

27. Giri, A. *et al.* Cavitand and molecular cage-based porous organic polymers. *ACS Omega* **5**, 28413–28424 (2020).

28. Zhu, Q. *et al.* Soft hydrogen-bonded organic frameworks constructed using a flexible organic cage hinge. *J. Am. Chem. Soc.* **145**, 23352–23360 (2023).

29. Zhang, G. & Mastalerz, M. Organic cage compounds-from shape-persistency to function. *Chem. Soc. Rev.* **43**, 1934–1947 (2014).

30. Das, S. *et al.* Porous organic materials: strategic design and structure-function correlation. *Chem. Rev.* **117**, 1515–1563 (2017).

31. Jiang, S. *et al.* Porous organic molecular solids by dynamic covalent scrambling. *Nat. Commun.* **2**, 207 (2011).

32. Santos, T. *et al.* Dynamic nucleophilic aromatic substitution of tetrazines. *Angew. Chem. Int. Ed.* **60**, 18783–18791 (2021).

33. Terrier, F. *Modern Nucleophilic Aromatic Substitution* (Wiley, 2013). doi:10.1002/9783527656141.

34. Katz, J. L. *et al.* Synthesis of oxacalixarenes incorporating nitrogen heterocycles: Evidence for thermodynamic control. *Org. Lett.* **8**, 2755–2758 (2006).

35. Katz, J. L. *et al.* Single-step synthesis of D_{3h}-symmetric bicyclooxacalixarenes. *Org. Lett.* **7**, 3505–3507 (2005).

36. Wang, D. X. *et al.* Versatile anion-π interactions between halides and a conformationally rigid bis(tetraoxacalix[2]arene[2]triazine) cage and their directing effect on molecular assembly. *Chem. Eur. J.* **16**, 13053–13057 (2010).

37. Wang, Z. *et al.* Multicolor tunable polymeric nanoparticle from the tetraphenylethylene cage for temperature sensing in living cells. *J. Am. Chem. Soc.* **142**, 512–519 (2020).

38. Zhang, G. *et al.* A permanent mesoporous organic cage with an exceptionally high surface area. *Angew. Chem. Int. Ed.* **53**, 1516–1520 (2014).

39. Hong, S. *et al.* Porphyrin boxes: rationally designed porous organic cages. *Angew. Chem. Int. Ed.* **54**, 13241–13244 (2015).

40. Ivanova, S. *et al.* Isoreticular crystallization of highly porous cubic covalent organic cage compounds**. *Angew. Chem. Int. Ed.* **60**, 17455–17463 (2021).

41. Hähsler, M. & Mastalerz, M. A giant [8+12] boronic ester cage with 48 terminal alkene units in the periphery for postsynthetic alkene metathesis. *Chem. Eur. J.* **27**, 233–237 (2021).

42. Ono, K. *et al.* Self-assembly of nanometer-sized boroxine cages from diboronic acids. *J. Am. Chem. Soc.* **137**, 7015–7018 (2015).

43. Kory, M. J. *et al.* Facile synthesis and theoretical conformation analysis of a triazine-

based double-decker rotor molecule with three anthracene blades. *Chem. Eur. J.* **20**, 6934–6938 (2014).

44. Luo, N. *et al.* Exploiting anion–π interactions for efficient and selective catalysis with chiral molecular cages. *Angew. Chem. Int. Ed.* **60**, 20650–20655 (2021).

45. Mooibroek, T. J. & Gamez, P. The s-triazine ring, a remarkable unit to generate supramolecular interactions. *Inorganica Chim. Acta* **360**, 381–404 (2007).

46. Turcani, L. *et al.* stk: A python toolkit for supramolecular assembly. *J. Comput. Chem.* **39**, 1931–1942 (2018).

47. Itoh, Y. *et al.* Ultrafast water permeation through nanochannels with a densely fluorous interior surface. *Science* **376**, 738–743 (2022).

48. Yoshizawa, M. *et al.* Diels-Alder in aqueous molecular hosts: unusual regioselectivity and efficient catalysis. *Science* **312**, 251–254 (2006).

49. Yamashina, M. *et al.* An antiaromatic-walled nanospace. *Nature* **574**, 511–515 (2019).

50. Willems, T. F. *et al.* Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **149**, 134–141 (2012).

51. Case, D. H. *et al.* Convergence properties of crystal structure prediction by quasi-random sampling. *J. Chem. Theory Comput.* **12**, 910–924 (2016).

52. Price, S. L. *et al.* Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Phys. Chem. Chem. Phys.* **12**, 8478–8490 (2010).

53. Loots, L. & Barbour, L. J. An infinite catenane self-assembled by π···π interactions. *Chem. Commun.* **49**, 671–673 (2013).

54. Tozawa, T. *et al.* Porous organic cages. *Nat. Mater.* **8**, 973–978 (2009).

55. Ozdemir, J. *et al.* Covalent organic frameworks for the capture, fixation, or reduction of CO_2. *Front. Energy Res.* **7**, 77 (2019).

56. Kunde, T. *et al.* Porous organic compounds – small pores on the rise. *Eur. J. Org. Chem.* **2021**, 5844–5856 (2021).

57. Kunde, T. *et al.* A porous fluorinated organic [4+4] imine cage showing CO_2 and H_2 adsorption. *Chem. Commun.* **56**, 4761–4764 (2020).

58. Lu, C. *et al.* OPLS4: Improving force field accuracy on challenging regimes of

chemical space. *J. Chem. Theory Comput.* **17**, 4291–4300 (2021).

59. Kühne, T. D. *et al.* CP2K: An electronic structure and molecular dynamics software package -Quickstep: efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103 (2020).

60. Perdew, J. P. *et al*. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

61. VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **127**, 114105 (2007).

62. Grimme, S. *et al*. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).

63. Frisch, M. J. *et al.* Gaussian 16 revision a. 03. 2016; gaussian inc. *Wallingford CT* **2**, 4 (2016).

64. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).

65. Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **8**, 1057–1065 (2006).

66. Frisch, M. J. *et al.* Gaussian09 Revision D.01, Gaussian Inc. Wallingford CT. *Gaussian 09 Revision C.01* at (2010).

67. Coombes, D. S. *et al.* Role of electrostatic interactions in determining the crystal structures of polar organic molecules. A distributed multipole study. *J. Phys. Chem.* **100**, 7352–7360 (1996).

68. Beyer, T. & Price, S. L. Dimer or catemer? Low-energy crystal packings for small carboxylic acids. *J. Phys. Chem. B* **104**, 2647–2655 (2000).

69. Stone, A. J. Distributed multipole analysis: stability for large basis sets. *J. Chem. Theory Comput.* **1**, 1128–1132 (2005).

70. Ferenczy, G. G. Charges derived from distributed multipole series. *J. Comput. Chem.* **12**, 913–917 (1991).