


ARTICLE

DOI: 10.1038/s41467-018-05271-9

OPEN

# High-throughput discovery of organic cages and catenanes using computational screening fused with robotic synthesis

R.L. Greenaway<sup>1</sup>, V. Santolini<sup>2</sup>, M.J. Bennison<sup>1</sup>, B.M. Alston<sup>1</sup>, C.J. Pugh<sup>1</sup>, M.A. Little<sup>1</sup>, M. Miklitz <sup>2</sup>, E.G.B. Eden-Rump<sup>1</sup>, R. Clowes<sup>1</sup>, A. Shakil<sup>1</sup>, H.J. Cuthbertson<sup>1</sup>, H. Armstrong<sup>1</sup>, M.E. Briggs<sup>1</sup>, K.E. Jelfs<sup>2</sup> & A.I. Cooper<sup>1</sup>

Supramolecular synthesis is a powerful strategy for assembling complex molecules, but to do this by targeted design is challenging. This is because multicomponent assembly reactions have the potential to form a wide variety of products. High-throughput screening can explore a broad synthetic space, but this is inefficient and inelegant when applied blindly. Here we fuse computation with robotic synthesis to create a hybrid discovery workflow for discovering new organic cage molecules, and by extension, other supramolecular systems. A total of 78 precursor combinations were investigated by computation and experiment, leading to 33 cages that were formed cleanly in one-pot syntheses. Comparison of calculations with experimental outcomes across this broad library shows that computation has the power to focus experiments, for example by identifying linkers that are less likely to be reliable for cage formation. Screening also led to the unplanned discovery of a new cage topology—doubly bridged, triply interlocked cage catenanes.

<sup>1</sup>Department of Chemistry and Materials Innovation Factory, University of Liverpool, 51 Oxford Street, Liverpool L7 3NY, UK. <sup>2</sup>Department of Chemistry, Imperial College London, South Kensington, London SW7 2AZ, UK. Correspondence and requests for materials should be addressed to K.E.J. (email: [k.jelfs@imperial.ac.uk](mailto:k.jelfs@imperial.ac.uk)) or to A.I.C. (email: [aicooper@liverpool.ac.uk](mailto:aicooper@liverpool.ac.uk))

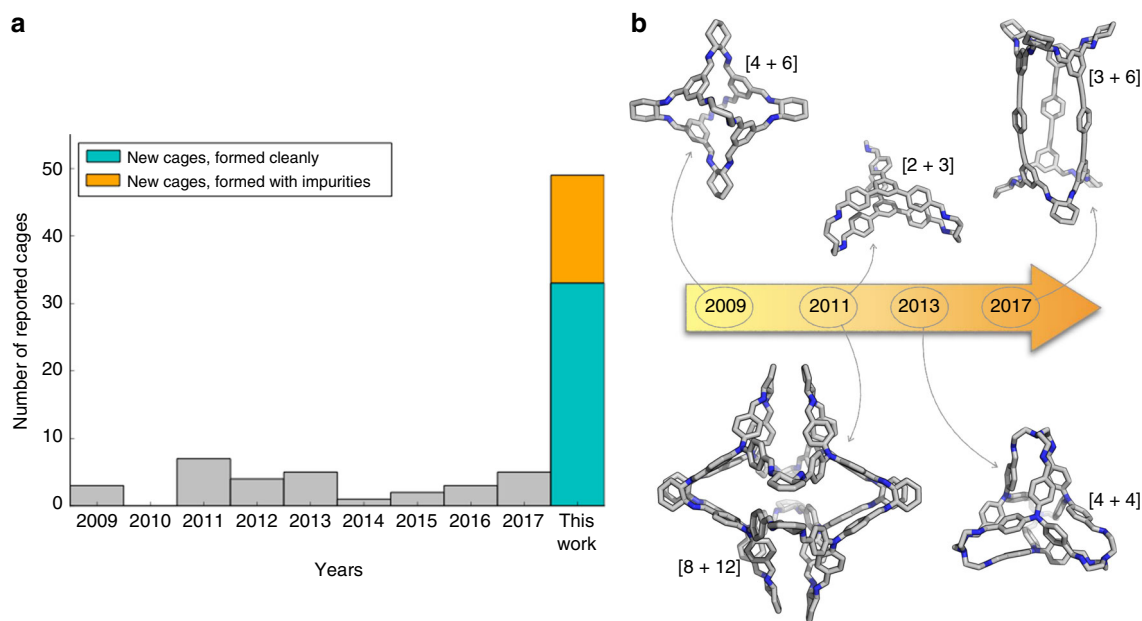
Supramolecular self-assembly is a powerful approach for generating complex organic and organometallic molecules, such as macrocycles<sup>1,2</sup>, cages<sup>3–6</sup>, catenanes<sup>7</sup>, rotaxanes<sup>8</sup>, molecular knots<sup>9</sup>, and molecular machines<sup>10,11</sup>. The targeted design of such molecules, however, can be very challenging, particularly as they become more elaborate<sup>12,13</sup>. Intuitive design strategies often fail as the system complexity increases and the number of possible self-assembled structures proliferates. Likewise, de novo computational design strategies are hampered by the size of such molecules and their conformational complexity<sup>14–16</sup>. As a result, the discovery of new self-assembled molecules can be relatively slow and labour intensive.

Porous organic cages (POCs) are a class of self-assembled molecules that show potential for molecular separation<sup>17</sup>, sensing<sup>18,19</sup>, and as building blocks for new materials, such as porous liquids<sup>20</sup>. However, despite significant interest in this area, the discovery rate for new cages has been relatively modest, reflecting the challenge in designing wholly new systems from scratch. New structures often involve small, iterative changes to known molecules. To relate this to our own activity, since 2009 we have published an average of three new cage molecules per year (Fig. 1a), with a variation on the cage topology every couple of years or so (Fig. 1b)<sup>4,21–24</sup>. Again, most of these molecules were discovered by making small, iterative structural changes to the precursors<sup>25–27</sup>. A bottleneck to advancing this area, as in many other areas of supramolecular assembly, is our ability to reliably design new self-assembled organic molecules. Pure in silico design strategies cannot yet cope with the complexity of supramolecular materials. Blind experimental screening approaches, while potentially rapid, are inefficient. Our strategy here is the close fusion of computational screening with automated synthesis and measurement protocols to create an integrated workflow for supramolecular materials discovery.

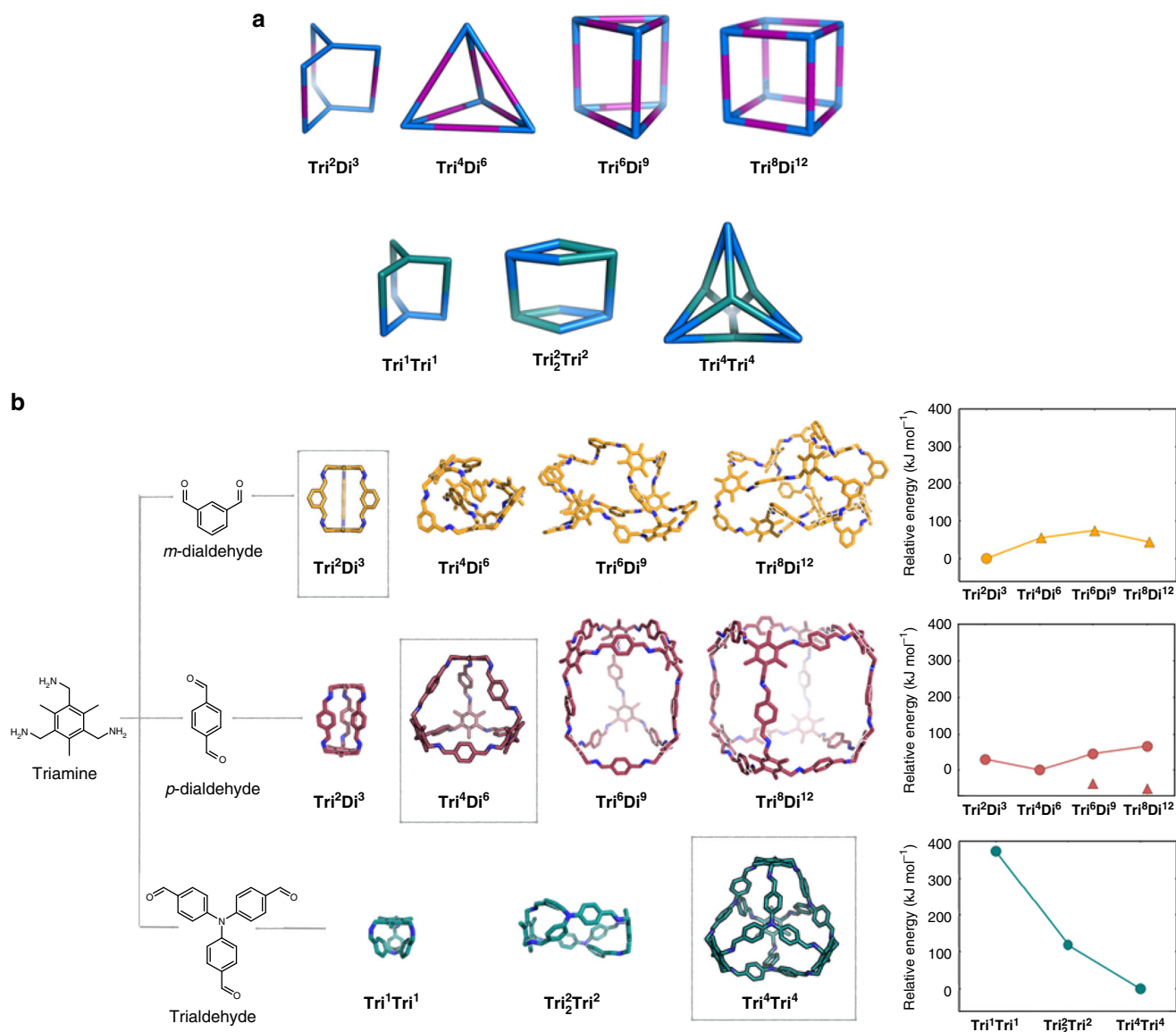
There have been significant advances in the use of computational design strategies for POCs and for porous materials in general<sup>16,28–31</sup>, along with a less widespread adoption in the use of high-throughput automation for materials synthesis<sup>32–35</sup>. Computation allows the prediction of the most likely cage

topology by studying the relative energies of different molecular assemblies or their formation energies<sup>36,37</sup>. Calculations can determine a priori whether a candidate cage structure will exhibit shape-persistence, and hence perhaps be porous<sup>22</sup>, and predict the effect of solvent scaffolding on molecular conformation<sup>38</sup>. Implicit solvation models, such as a polarisable continuum model (PCM) can be used to account for the electrostatic interactions between cages and the surrounding solvent, and have previously been found to stabilise the cage molecules relative to consideration in the gas phase<sup>37</sup>. However, there are also significant limitations to de novo computational design: for example, it is currently challenging to fully include the effect of solvent choice, particularly for large studies across a broad range of starting materials. Our strategy, therefore, was to identify the affordable elements of computational prediction that are of most value for fusion with high-throughput automated synthesis methods.

We report here a new family of organic cages discovered using this hybrid approach. A total of 78 possible imine cage combinations were investigated based on an array of three candidate amines and 26 candidate aldehydes. First, simulations were used to determine the most likely cage topology for each of the three types of amine–aldehyde combinations. Computation suggested clear topological preferences for some precursor combinations, as manifested by large energy differences, while the topological preferences for other combinations were less clear cut, suggesting the potential for mixed products. However, in this study none of the candidate reactions were discounted: rather, all combinations were trialled experimentally to test our predictions, like a training set. Overall, 33 cages were synthesised in pure form, and a further 16 cages could be identified along with side-products; this significantly exceeds the total number of cage molecules reported by our research group over the last 9 years (Fig. 1a). Successful experimental ‘hits’ were analysed with further simulations to identify candidate shape-persistent cages prior to scale-up and characterisation. We present single crystal X-ray diffraction for 18 of these structures, and there is excellent agreement with our a priori computational structure predictions. We also use post-experiment computation to explain why some systems formed unanticipated



**Fig. 1** High-throughput methods fused with computation provides a step-change in the discovery rate for new cage molecules. **a** Timeline shows the number of organic cage molecules reported by our research group using traditional, iterative experiments in period 2009–2017 (grey bars; total of 30 cages). The number of cage molecules synthesised in this new study is also shown: we achieved clean conversion to 33 cages (teal bar) plus partial formation of 16 others (orange bar). **b** Timeline showing variation in cage topologies studied by our group during the same period



**Fig. 2** Reaction schemes and computational modelling for three model covalent organic cages. **a** Scheme showing possible cage topologies; (top) candidate topologies for tritopic + ditopic reactions ( $\text{Tri}^x\text{Di}^y$ ) and (bottom) for tritopic + tritopic reactions ( $\text{Tri}^x\text{Tri}^y$ ). Tritopic precursors are shown in blue or cyan, ditopic precursors in purple. **b** Computationally derived structures for the possible topological combinations of the tritopic triamine linker ((2,4,6-trimethylbenzene-1,3,5-triyl)trimethanamine), with *m*-dialdehyde (isophthalaldehyde), *p*-dialdehyde (terephthalaldehyde), and trialdehyde (tris(4-formylphenyl)amine), respectively. The energetically preferred topologies have a grey box around them. The plots (right) show the relative internal energies calculated for the different topologies, plotted on a common energy scale. For topologies where both collapsed and open conformations are computed, the energy of the collapsed conformation is indicated by a triangle symbol; open conformations are indicated by circle symbols. For the *p*-dialdehyde, only open conformations are shown in the images, collapsed  $\text{Tri}^6\text{Di}^9$  and  $\text{Tri}^8\text{Di}^{12}$  topologies are shown in Supplementary Fig. 1.  $\text{Tri}^2\text{Di}^3$  cages are shown with orange carbons,  $\text{Tri}^4\text{Di}^6$  with maroon, and  $\text{Tri}^4\text{Tri}^4$  with teal. All nitrogens are blue and hydrogens are omitted

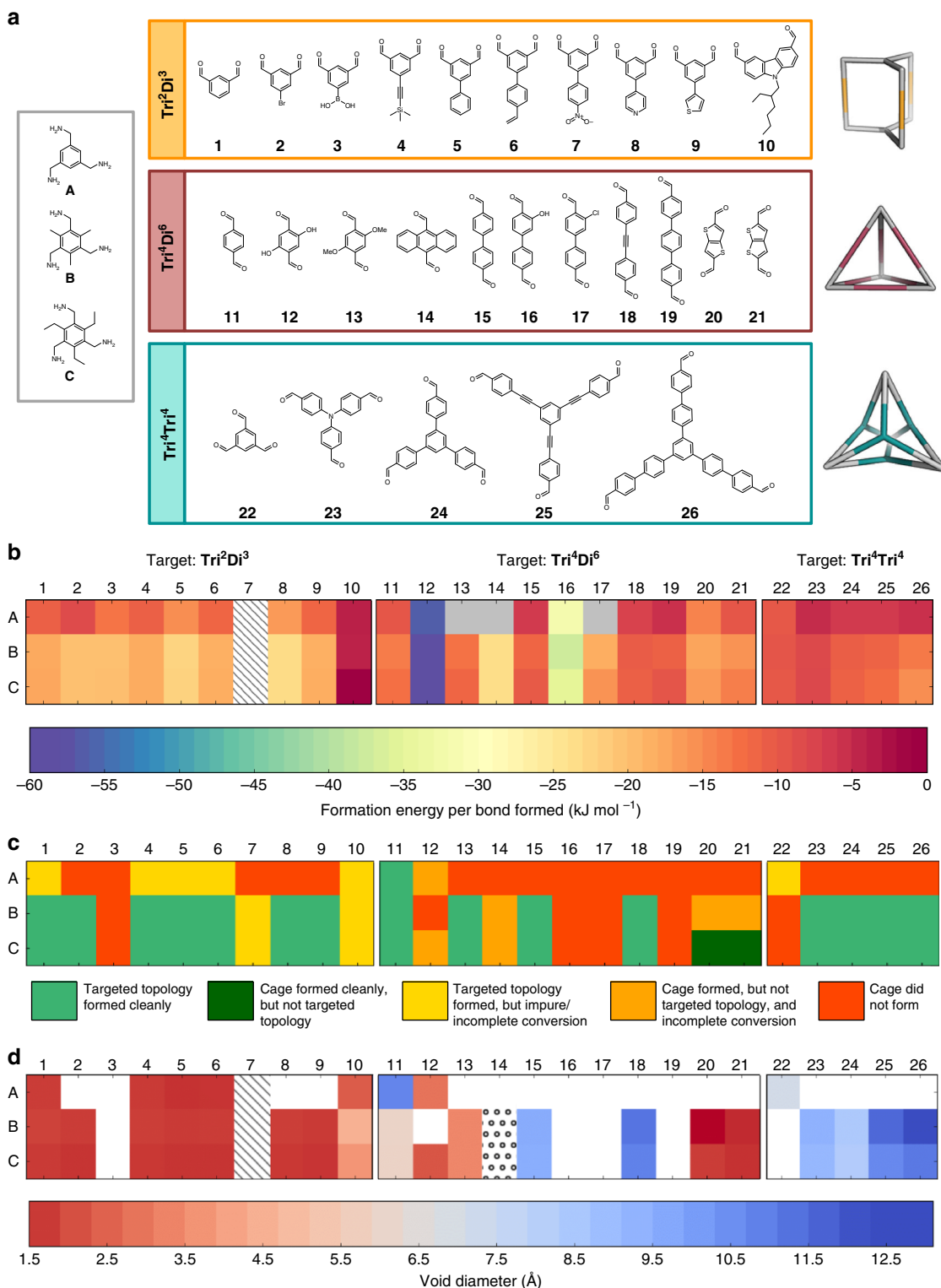
topologies or mixtures, such as a unique doubly bridged, triply interlocked cage catenane. This surprising result illustrates the power of high-throughput experimental screening to reveal unexpected molecular structures. Finally, we propose a more generalised workflow for the streamlined discovery of supramolecular materials based on what we learnt from this study.

## Results

**Computational prediction of model cage topologies.** Prior to any experiments, we used computation to assess the possible model cage topologies (see Supplementary Methods) that might be formed by the reaction of a triamine precursor, (2,4,6-trimethylbenzene-1,3,5-triyl)trimethanamine, with a representative

example of the three different aldehyde types; *meta* and *paradialdehydes*, and trigonal trialdehydes (Fig. 2), investigated in the high-throughput screen (Fig. 3a). These precursors are selected for the representative examples as the triamine is neither the most flexible, nor the most pre-configured, the *meta* and *para* dialdehydes the simplest linkers in the subsets, and the trigonal trialdehyde previously used to form tetrapods with a more flexible triamine linker<sup>23</sup>. We used our previous nomenclature for the cage topologies (Supplementary Note 1);<sup>36</sup> that is,  $\text{Tri}^x\text{Di}^y$  and  $\text{Tri}^x\text{Tri}^y$ , where the superscripts signify the number of tri-topic and di-topic precursors incorporated into the cage.

The relative internal energies were computed for each possible topology (Fig. 2b, right; Supplementary Note 2). Since the cages



are formed by reversible imine chemistry, one might expect thermodynamic products to be formed—that is, the topologies with the lowest relative internal energy. These calculations do not take into account the potential effect of reaction solvent. We chose not to include an implicit solvation model so as to explore only the effect of the internal energy of the cages' structures, especially since many of the reactions here were attempted in multiple solvents, each of which would have a differing dielectric constant. Ideally, we would want to geometry optimise each

structure with the implicit solvation model, to explore how the conformation adapts to the electrostatic influence of each solvent, however this would be prohibitively computationally expensive for the large number of systems here.

For cages involving the *m*-dialdehyde, an energetic preference of  $43 \text{ kJ mol}^{-1}$  per  $[2 + 3]$  unit was found for a  $\text{Tri}^2\text{Di}^3$  capsule topology (Fig. 2b, top row). The resulting cage was shape persistent; that is, it did not collapse. Larger candidate cage topologies (e.g.,  $\text{Tri}^4\text{Di}^6$ ) were all found to be higher in energy

**Fig. 3** Cage precursors, target cage topologies, and results of high-throughput computation and synthesis screening. **a** Three triamines (**A–C**) were combined with three types of aldehydes to target **Tri<sup>2</sup>Di<sup>3</sup>** capsules (**1–10**), **Tri<sup>4</sup>Di<sup>6</sup>** tetrahedrons (**11–21**), and **Tri<sup>4</sup>Tri<sup>4</sup>** tetrapods (**22–26**). **b** DFT formation energies per imine bond formed for the target cage topologies. These energies were not calculated for molecules that lacked shape persistence and collapsed (grey boxes). Cross-hatched squares denote three molecules excluded from the calculations because the force field was unreliable. **c** Results of the high-throughput synthesis screen. (i) light green squares—targeted cage topology was formed cleanly (31 examples); (ii) dark green squares—cage product was formed cleanly, but with topology that was not targeted (two examples); (iii) yellow squares—target topology formed but product impure or there was incomplete conversion (10 examples); (iv) orange squares—alternative topology formed and either impure or incomplete conversion (six examples); (v) red squares—no cage was formed (29 examples; 18 of these for triamine **A**). Cages that went against our targeted topology assumptions and formed an alternative topology are: **A12**, **C12**, **C14**, **B20**, **C20**, **B21**, and **C21** (**Tri<sup>2</sup>Di<sup>3</sup>** observed instead of targeted **Tri<sup>4</sup>Di<sup>6</sup>**) and **B14** (formed a mixture of **Tri<sup>2</sup>Di<sup>3</sup>** and **Tri<sup>4</sup>Di<sup>6</sup>**). **d** Void diameters as calculated from a priori computational models for cages that were formed in the subsequent reaction screen. White squares correspond to combinations where no cage was formed experimentally; squares filled with circles correspond to cages calculated to have no void

and non-shape-persistent, with the lowest energy conformations being collapsed with no internal void. This predicted preference for a **Tri<sup>2</sup>Di<sup>3</sup>** topology with (2,4,6-trimethylbenzene-1,3,5-triyl) trimethanamine agrees with previous studies on reactions of this amine with isophthalaldehyde<sup>39,40</sup>, and alternative *meta* dialdehydes<sup>41</sup>, which form **Tri<sup>2</sup>Di<sup>3</sup>** capsules.

For cages involving the *p*-dialdehyde (Fig. 2b, middle row), the **Tri<sup>4</sup>Di<sup>6</sup>** tetrahedron topology was predicted to be energetically favoured by 29 kJ mol<sup>-1</sup> per [2 + 3] unit with respect to other ‘open’ topologies, while the larger cage topologies collapsed to give lower energy forms (triangle points in Fig. 2b); their open conformations (circle points in Fig. 2b), which might be expected to dominate in solution<sup>38</sup>, were higher in energy than the uncollapsed **Tri<sup>4</sup>Di<sup>6</sup>** topology.

For cages that combined the trialdehyde with the tritopic amine (Fig. 2b, bottom row), three potential cage topologies were considered, **Tri<sup>1</sup>Tri<sup>1</sup>**, **Tri<sup>2</sup><sub>2</sub>Tri<sup>2</sup>**, and **Tri<sup>4</sup>Tri<sup>4</sup>**. The **Tri<sup>4</sup>Tri<sup>4</sup>** ‘tetrapod’ topology was predicted to be much more stable by 118 kJ mol<sup>-1</sup> per [1 + 1] unit.

The magnitude of the topological preference based on the relative energy differences for these three systems was ranked as **Tri<sup>4</sup>Tri<sup>4</sup>** (trialdehyde) > **Tri<sup>2</sup>Di<sup>3</sup>** (*m*-dialdehyde) > **Tri<sup>4</sup>Di<sup>6</sup>** (*p*-dialdehyde). That is, the energetic preference for the **Tri<sup>4</sup>Tri<sup>4</sup>** topology was predicted to be the most clear cut, suggesting a more reliable design basis for cages based on this topology, at least for this model reaction.

We next developed a robust synthesis method for these three model cages (Supplementary Note 3). The reaction conditions, which use only moderately high dilution and modest temperatures, were selected to be easily translated onto a robotic synthesis platform. Analysis of the cages by high-resolution mass spectroscopy (HRMS) confirmed that the experimental cage topologies agreed with the computational predictions; that is, for these three model systems at least, relative energies are a good predictor of cage topology preference.

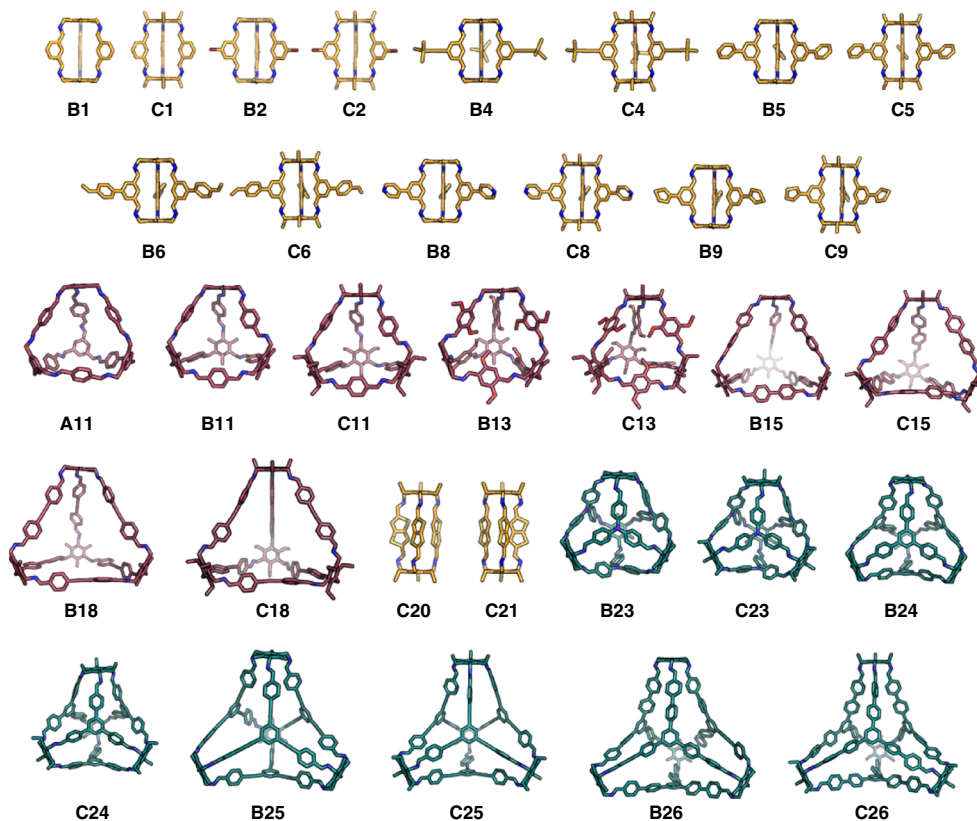
**High-throughput cage synthesis.** Using our optimised reaction conditions, we screened a much broader range of aldehydes (**1–26**) and amines (**A–C**) for cage formation using a synthesis robot (Fig. 3, Supplementary Fig. 2). For all combinations, the target cage topology was the one that was observed for the representative *m*-dialdehyde, *p*-dialdehyde, and trialdehyde in the model study, above. These three model combinations will be referred to henceforth as **B1**, **B11**, and **B23**, respectively (Fig. 3a).

We explored two additional analogues of triamine **B**, both with less (**A**) and more (**C**) alkyl substitution on the arene ring; this was designed to introduce different degrees of flexibility and structural pre-configuration. For the targeted **Tri<sup>2</sup>Di<sup>3</sup>** analogues of **B1**, we explored ten different *m*-dialdehydes with various substituents at the five position (**1–9**), plus a carbazole dialdehyde, **10** (Fig. 3a, top panel). For the targeted **Tri<sup>4</sup>Di<sup>6</sup>** analogues of **B11**, we studied 11

linear dialdehydes (**11–21**) containing different functionalities and degrees of rotational flexibility (Fig. 3a, middle panel). For the **Tri<sup>4</sup>Tri<sup>4</sup>** analogues of **B23**, we studied five tritopic aldehydes (**22–26**) with various lengths and flexibility (Fig. 3a, bottom panel). In total, these three amines permuted with 26 aldehydes gave 78 reaction combinations.

Prior to the high-throughput experimental screen, models for all but three of the 78 target cages were constructed in silico and analysed computationally (Supplementary Note 4). Cages containing the di-topic linker **7** were omitted because the nitro-group was poorly described by the forcefield<sup>42</sup>. We also simplified **10** by removing the external alkyl substituents. We assessed the formation energies of the hypothetical cages and screened them for conformation, size, and shape persistence, the latter to identify cages with voids that are stable to desolvation (see Supplementary Tables 1–2). For each possible outcome, we searched for the low energy conformations and then minimised those structures using DFT calculations at the PBE+D3/TZVP level, before calculating the formation energy per imine bond formed, as shown in Fig. 3b. This energy normalisation using formation energies, rather than relative energies per sub-unit, allows us to directly compare the energies of the cages regardless of their size or topology. We chose the PBE+D3/TZVP level of theory as a compromise between accuracy and the ability for us to run the large number of calculations required in a high-throughput screen. Taken in isolation, we did not expect these formation energies to be generally predictive of experimental ‘hits’, due to the exclusion of factors, such as solvation effects and phase behaviour, coupled with the potential inaccuracies of DFT formation energies. Rather, we calculated these formation energies to establish whether they might play a more statistical role in guiding high-throughput screening studies in combination with other computational and experimental information.

The calculated formation energy per imine bond was found to be negative in all cases, although some hypothetical cages were much more energetically favoured than others. The calculated energy per imine bond ranged from ~-1 to -58 kJ mol<sup>-1</sup>, and there was good correlation between cages with a less energetically favoured formation energy (dark red in Fig. 3b) and cages that did not form experimentally (red in Fig. 3c). Outliers for this trend include **A16**, **B16**, and **C16**, which had favourable formation energies (average -34.7 kJ mol<sup>-1</sup>), but yet were not observed. This could be due to factors not considered in determining the formation energy of the complete cage alone, for instance, a “kinetic bottleneck” in the reaction pathway to that molecule<sup>37,43</sup>. On the other hand, the formation energy for **B** and **C** combined with precursors from **23–26** was generally only weakly favoured (average formation energy -11.0 kJ mol<sup>-1</sup>), but yet these molecules were all formed experimentally. In this instance, these large, open molecules may be particularly stabilised, relative to their precursors, by the influence of solvent<sup>37</sup>.



**Fig. 4** The 33 cages that formed cleanly in this high-throughput study. **Tri<sup>2</sup>Di<sup>3</sup>** cages are shown with orange carbons, **Tri<sup>4</sup>Di<sup>6</sup>** with maroon, and **Tri<sup>4</sup>Tri<sup>4</sup>** with teal. Remaining atom colouring is as follows; oxygen (red), bromine (brown), boron (pink), silicon and sulphur (yellow), and nitrogen (blue). Hydrogens are omitted. All structures are derived from computational predictions. For cages **C20** and **C21**, **Tri<sup>2</sup>Di<sup>3</sup>** cages were formed as shown, rather than the predicted **Tri<sup>4</sup>Di<sup>6</sup>** cages. All other cages formed the predicted, targeted topology. Supplementary Fig. 4 shows structures for cages that were identified by solution analysis, but that were either impure or showed incomplete conversion of the starting materials. Supplementary Fig. 3 shows the computed structures for cages that did not form by experiment (red squares in Fig. 3c)

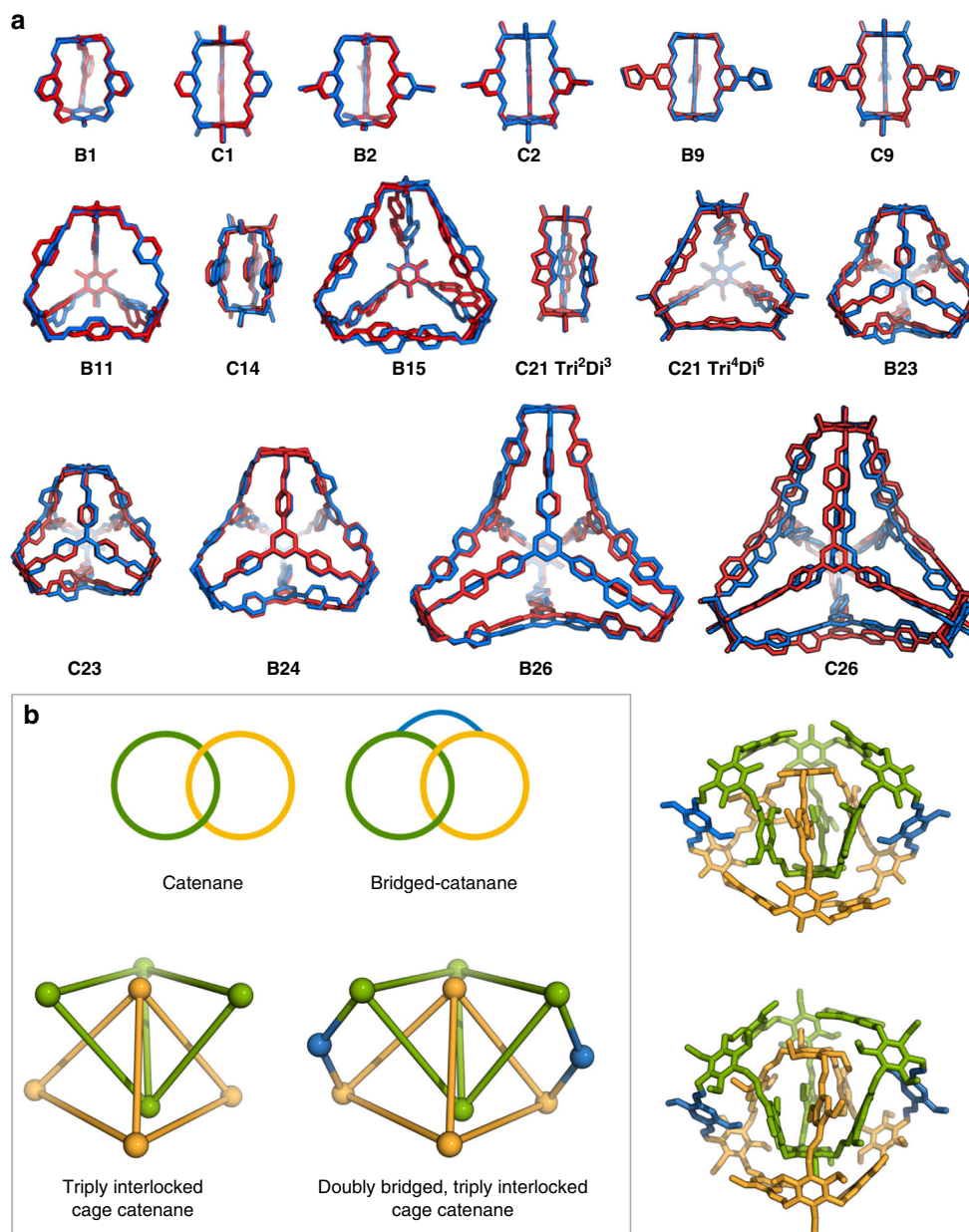
With the exception of three of the **Tri<sup>4</sup>Di<sup>6</sup>** cages (**A13**, **A14**, and **A17**, grey squares in Fig. 3b, structures in Supplementary Fig. 3), all of the hypothetical cages were predicted to be shape persistent in their target topologies. The average formation energy per imine bond for the three series of cages built from amines **A**, **B**, and **C** was predicted to be  $-12.5$ ,  $-17.9$ , and  $-17.3$   $\text{kJ mol}^{-1}$ , respectively. A priori computation suggests, therefore, that triamines **B** and **C** might be better choices for imine cage formation than triamine **A**, possibly due to the increased flexibility and lack of pre-organisation in triamine **A**<sup>44</sup>. This was reflected by experiment: the failure rate for triamine **A** (red, yellow, and orange squares in top row, Fig. 3c; only one system, **A11**, gave clean cage conversion) is much higher than for triamines **B** and **C**.

The **Tri<sup>2</sup>Di<sup>3</sup>** cages (1–10) were found to be generally similar in energy, with the exception of cages containing aldehyde **10**, which were less favoured; we did not observe clean cage formation for **10** with any of the three triamines (Fig. 3c). **Tri<sup>4</sup>Tri<sup>4</sup>** cages (22–26) were in general somewhat less energetically favoured than **Tri<sup>2</sup>Di<sup>3</sup>** cages (average formation energies per imine bond of  $-10$  and  $-16$   $\text{kJ mol}^{-1}$ , respectively), although the experimental ‘hit rate’ for these two topologies across the array was rather similar (Fig. 3c). For the **Tri<sup>4</sup>Di<sup>6</sup>** (11–21) cages, the picture was more complicated with a wide variety of formation energies; for example, cages formed from aldehyde **12** were predicted to be particularly favoured, most likely due to intramolecular H-bonding<sup>41,45</sup>, although this did not translate into experiment (Fig. 3c).

Overall, the high-throughput robotic synthesis screen (Supplementary Fig. 5, Supplementary Table 3) yielded a total of 33 cages

(Fig. 4), where both <sup>1</sup>H NMR spectroscopy and HRMS indicated clean cage formation without additional work up (green squares in Fig. 3c). There were a further 16 reactions (yellow/orange squares in Fig. 3c) where partial cage formation was observed (Supplementary Fig. 4). In some cases—for example, cage **C20**—the <sup>1</sup>H NMR spectrum indicated clean cage formation, but there was no clear mass ion in the HRMS. For such systems, diffusion NMR was particularly useful because it provides an alternative and rapid method to determine the size (Supplementary Tables 4–9), and hence the topology, of the cages directly from the reaction mixtures<sup>46</sup>. We calculated the average size and maximum diameters of the cages from the energy-minimised computed conformations (Supplementary Note 5) and compared these values to the experimentally determined solvodynamic diameters (Supplementary Tables 10–11), allowing us to determine the topology of the cages. There was strong agreement between the measured size and the computed average size (Supplementary Fig. 6), which provides a rapid and direct measure of the success of our predictions without the need to obtain single crystal structures.

We next computed the cavity sizes for cages that were formed by experiment (Fig. 3d) by calculating the largest sphere that could be placed in the central molecular cavity (Supplementary Tables 1, 2, and 12). If we assume that a diameter of more than 2.89 Å (the kinetic diameter of H<sub>2</sub>)<sup>47</sup> is required for the cages to host any guest at all, then 21 cages in this screen (27% of the 78 combinations) met this criteria. Most **Tri<sup>2</sup>Di<sup>3</sup>** cages had cavities predicted to be smaller than 1.8 Å in diameter, making them too small to host guests. The **Tri<sup>4</sup>Di<sup>6</sup>** cages had larger cavities with diameters ranging from 2.3 to 11.2 Å, with the smaller cavities



**Fig. 5** Crystal structures of cages and bridged cage catenanes. **a** Comparison of modelled and experimental cage structures—overlays of available single crystal X-ray structures for each cage (red) with the computed lowest energy conformation (blue). **b** Crystal structures of the two unexpected [8 + 12] covalently bridged, triply interlocked cage catenanes formed by re-equilibration of targeted cages, **B13** and **C13**, and comparison of the unique topology with previous catenanes, bridged catenanes, and triply interlocked cage catenanes. The [8 + 12] bridged cage catenane is triply interlocked (symmetry equivalents coloured in yellow and green), but uniquely, it has two different shaped windows and is linked by two covalent bridges (blue). Hydrogens and external functional groups have been removed for clarity. For the full crystal structures see Supplementary Figs. 12 and 22

typically resulting from functionality present on the aldehyde penetrating into the internal cavity, as for **B13** and **C13** (Fig. 4). The largest predicted cavities across the library were calculated for the **Tri⁴Tri⁴** cages, with **B26** having the largest cavity diameter (12.3 Å).

**Comparison of computed structures and crystal structures.** For cages where we could obtain single crystal X-ray structures (Supplementary Figs. 7–27), we found an excellent geometric match with the predicted low-energy conformation obtained by a priori computation (Fig. 5a, Supplementary Table 13, and Supplementary Fig. 28), even though these calculations do not take into account crystal packing forces or solvent effects. The average

root-mean-square-deviation (RMSD) between the crystal structure and the simulations was 0.53, 1.07, and 1.26 Å for the **Tri²Di³**, **Tri⁴Di⁶**, and **Tri⁴Tri⁴** cages respectively. In all cases, the RMSD decreases upon geometry optimisation of a molecule extracted from the experimental crystal structure (to 0.24, 0.96, and 1.15 Å respectively), which could be indicative of crystal packing only having a minor influence on the structures.

**C21**, which formed a **Tri²Di³** cage in the high-throughput synthesis screen, gave a product mixture upon scale-up and direct recrystallisation of the reaction solution: in this case, a crystal structure of both a **Tri²Di³** (Supplementary Fig. 24) and a **Tri⁴Di⁶** (Supplementary Fig. 25) cage was obtained. The formation of different cage topologies from the same precursors

by simply changing reaction conditions has been observed before<sup>48,49</sup>. When the reaction was carried out in chloroform at reflux in the high-throughput screen, the **Tri<sup>2</sup>Di<sup>3</sup>** cage dominated. By contrast, the scaled-up reaction in dichloromethane at reflux gave a mixture of the **Tri<sup>2</sup>Di<sup>3</sup>** and **Tri<sup>4</sup>Di<sup>6</sup>** cages (Supplementary Figs. 29 and 30). **C21** illustrates the potential sensitivity of the cage topology to the precise experimental conditions.

**Discovery of a unique bridged cage catenane.** A unique and unpredicted cage topology was also discovered during recrystallisation studies: a covalently bridged, triply interlocked cage catenane (Fig. 5b). This topology can be formed upon crystallisation by dynamic re-equilibration of two of the **Tri<sup>4</sup>Di<sup>6</sup>** cages, **B13** and **C13**. They comprise an [8 + 12] stoichiometry and are both triply interlocked (catenated) and di-covalently linked (bridged); this topology seems to be stabilised by offset  $\pi$ - $\pi$  stacking interactions between the two sets of dimethoxy substituted aromatic groups, which are stacked <4 Å apart in the crystal structure (Supplementary Figs. 12 and 22). The formation of simpler interlocked cage catenanes has been observed before, either upon recrystallisation or by direct synthesis using an acid catalyst<sup>50,51</sup>. Here, we found that the targeted, non-catenated **Tri<sup>4</sup>Di<sup>6</sup>** cages **B13** and **C13** re-equilibrated to form covalently bridged cage catenanes during recrystallisation. Additional simulations rationalised the formation of these unique structures (Supplementary Note 6); the bridged catenated [8 + 12] species was calculated to be more stable (>17 kJ mol<sup>-1</sup> per imine bond or 215 kJ mol<sup>-1</sup> per [4 + 6] unit) in comparison to the parent **Tri<sup>4</sup>Di<sup>6</sup>** cage, **B13**, which was initially formed in solution, as targeted.

**Comparison of experiment with computational predictions.** In the 33 cases where a single cage product was formed cleanly with good conversion, 31 cages (94%) were formed with the predicted, targeted topology. Of these, two cages (**B13** and **C13**) were found to re-equilibrate to form catenanes. Computed formation energies were found to be useful at a coarse grain level—for example, in suggesting correctly that reactions with triamine **A** were less promising than for triamines **B** and **C**—but they are not an unambiguous predictor of whether or not a given cage can be synthesised. For example, cages with precursor **10** were found to have the *least* favourable formation energies, with relatively high strain around the imine bonds, and yet **A10**, **B10**, and **C10** were all formed by experiment, albeit with residual aldehyde starting material and potential oligomeric side products. Cages comprising aldehyde **12** were found to have particularly favourable formation energies, yet **B12** was not formed under these reaction conditions and both **A12** and **C12** were formed (in impure form) in a **Tri<sup>2</sup>Di<sup>3</sup>** topology rather than **Tri<sup>4</sup>Di<sup>6</sup>**.

Some precursor combinations went against our targeted topology assumptions and formed an alternative topology. All of these examples were found in the **Tri<sup>4</sup>Di<sup>6</sup>** series; where **Tri<sup>2</sup>Di<sup>3</sup>** and **Tri<sup>4</sup>Tri<sup>4</sup>** topologies were targeted, all cages gave the predicted topology. For **A12**, **C12**, **C14**, **B20**, **C20**, **B21**, and **C21**, a **Tri<sup>2</sup>Di<sup>3</sup>** cage was observed instead of the targeted **Tri<sup>4</sup>Di<sup>6</sup>**; **B14** formed a mixture of both of these topologies (Supplementary Fig. 31). Thus, of the 33 reactions predicted to form **Tri<sup>4</sup>Di<sup>6</sup>** cages, only 9 formed the predicted **Tri<sup>4</sup>Di<sup>6</sup>**; 7 formed **Tri<sup>2</sup>Di<sup>3</sup>**, 16 did not form a cage, and 1 formed a **Tri<sup>2</sup>Di<sup>3</sup>/Tri<sup>4</sup>Di<sup>6</sup>** mixture. The energetic preference for the **Tri<sup>4</sup>Di<sup>6</sup>** topology in the example system, **B11**, was the smallest of the three systems investigated; 29 kJ mol<sup>-1</sup> per [2 + 3] unit, compared to 43 kJ mol<sup>-1</sup> per [2 + 3] unit, and 118 kJ mol<sup>-1</sup> per [1 + 1] unit for **Tri<sup>2</sup>Di<sup>3</sup>** and **Tri<sup>4</sup>Tri<sup>4</sup>** topologies, respectively. This may rationalise the poor

predictability for the **Tri<sup>4</sup>Di<sup>6</sup>** systems; the lack of a strong driving force for a particular topology might also help to explain why so many of the **Tri<sup>4</sup>Di<sup>6</sup>** targeted reactions failed to produce any cage product at all (red squares, Fig. 3c, middle panel).

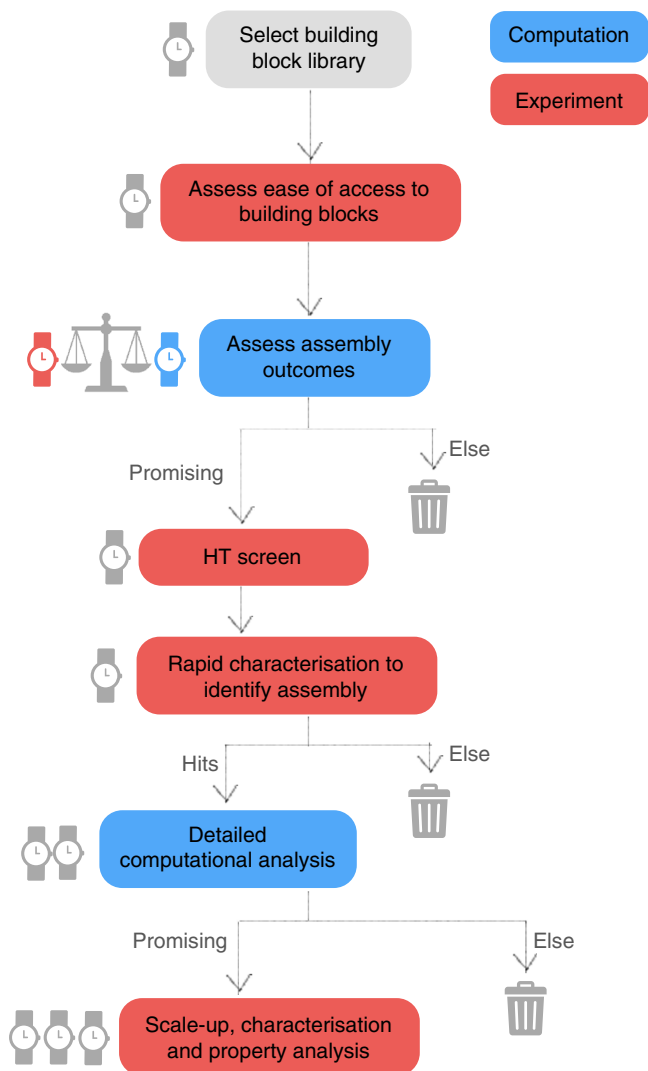
To try to explain the unexpected **Tri<sup>2</sup>Di<sup>3</sup>** topologies (**A12**, **C12**, **B14**, **C14**, **B20**, **C20**, **B21**, and **C21**), the energies for these outcomes were calculated and compared to the targeted **Tri<sup>4</sup>Di<sup>6</sup>** topology (Supplementary Tables 12 and 14, and Supplementary Fig. 31). For **B21** and **C21**, a preference was found for the observed **Tri<sup>2</sup>Di<sup>3</sup>** topology (by 24 and 35 kJ mol<sup>-1</sup> per [2 + 3] unit, respectively); in all the remaining cases, an energetic preference was calculated for the **Tri<sup>4</sup>Di<sup>6</sup>** topology rather than the observed **Tri<sup>2</sup>Di<sup>3</sup>** topology. This preference was often small: around 5 kJ mol<sup>-1</sup> per [2 + 3] unit for **A12** and **C20**. We cannot therefore rationalise the observed topologies for **A12**, **C12**, **B14**, **C14**, **B20** and **C20** based upon the relative internal energies of the isolated cage molecules alone. It is possible that the entropy of the systems, solvent interactions, the stability and solubility of reaction intermediates or competing reaction pathways play a decisive role. Taken together, our results suggest that an energetic preference of at least 30 kJ mol<sup>-1</sup> per repeat unit is required to be confident of a reliable computational prediction of a particular cage topology.

**Integrated experimental-computational workflow.** While previous studies have compared computation and experiment for a few closely related cage molecules<sup>36,37,52,53</sup>, this new study is unique because we make this comparison across a broad array of 78 supramolecular systems. Based on the lessons learned, we propose a more generalised workflow for supramolecular materials discovery that couples automated synthesis closely with computational studies (Fig. 6). One goal of this hybrid approach is to remove as many inferior targets as early as possible, to save time and cost. By applying computation at the appropriate stages, we can greatly streamline the materials discovery process and target more effective reaction combinations: for example, removal of the ‘unreliable’ triamine **A** from the array based on formation energy calculations (Fig. 3b) would have reduced the number of reactions by one third while missing only one of the 33 reactions (**A11**) that were found by experiment to give clean cage conversion (Fig. 3c). Likewise, the high success rate for the target **Tri<sup>4</sup>Tri<sup>4</sup>** topologies (80% if one excludes triamine **A**; Fig. 3c) was anticipated by our relative energy calculations that showed a more pronounced energy penalty for alternative cage topologies (Fig. 2b).

**Discussion.** In conclusion, the combination of computational modelling with robotic automation and high-throughput characterisation techniques has led to the discovery of a range of new organic cages. By investigating 78 building block combinations, 33 cages, 32 of which are new, were synthesised cleanly on a robot synthesiser in one-pot reactions without work up, with 16 other combinations showing evidence of cage formation along with side products or unreacted starting material. The scale-up of these experimental ‘hits’ was investigated, and crystallisation studies were conducted; this yielded 18 single crystal structures and led to the serendipitous discovery of two covalently bridged cage catenanes.

From a computational perspective, it is beneficial to have a strong energetic preference towards a topological outcome, as this increases the chance of the target cage being formed. This study shows how smaller energetic preferences can be overridden by other experimental factors, such as solvent stabilisation. Our rule of thumb is that a computed energetic preference of at least 30 kJ mol<sup>-1</sup> per repeat unit is required for confident predictive design.





**Fig. 6** Proposed workflow for the fusion of high-throughput automated synthesis with computation for the accelerated discovery of supramolecular materials. Experimental stages are shown in red and computational stages in blue. Approximate time cost for each step is indicated (1 clock = quick; 3 clocks = time-intensive). In this first study, all building block combinations were investigated to validate the approach, but the future goal is to remove low-value targets as early as possible, thus saving time and cost. First, an initial building block library is selected, and their ease of synthesis is assessed. Then, a judgement is made on a sliding scale as to whether a very rapid computational assessment of properties is conducted (e.g., test for shape-persistence in assumed topology), up to a full analysis of multiple assembly outcomes, detailed energetic assessment of the likely outcome and full property assessment. The basis for this judgement revolves around the synthetic effort in obtaining a given building block. For building blocks that are commercially available at low cost, there is less benefit for full computational assessment of the outcomes, but for building blocks that require custom, multi-step synthesis, there is great value in only embarking on synthesis if computation suggests there is a strong likelihood of forming supramolecular products of interest. After the synthetic screen, rapid assessment of the reaction outcomes (e.g., via mass spectrometry and diffusion NMR) are used to identify where ‘hits’ have occurred and what the reaction outcome is. For ‘hits’, more detailed computational analysis is performed and only those that are promising (e.g., possessing shape persistent voids for porous solids) are selected for scale-up and property measurements

False negatives are also a possibility in high-throughput screening: for example, in the case of **C22** there was no indication of the targeted  $\text{Tri}^4\text{Tri}^4$  cage in the  $^1\text{H}$  NMR spectrum arising from the robotic screen. Coupled with the formation of an insoluble precipitate, this suggested the formation of oligomers or polymer. However, Mastalerz et al. recently reported the successful synthesis of this cage under different reaction conditions, albeit in only 27% yield and with a considerable amount of polymer formation too (68%)<sup>44</sup>. In the future, this might be addressed by using higher capacity synthesis robots and by investigating a broader range of reaction solvents.

Such high-throughput studies covering a broad range of conditions stand to benefit particularly from computational guidance; for example, by focussing the range of precursors that is investigated. Coupled over many cycles, this hybrid discovery approach promises to yield a large database of successful and failed reactions, along with associated computational properties, that should fuel machine learning approaches in the future<sup>54</sup>. Moreover, this methodology could be combined with fast crystal structure prediction methods to predict the likely packing motifs for new cages<sup>27,55</sup>, thus allowing us to weight the selection of precursors toward systems that are likely to deliver materials with specific solid-state properties, such as adsorption selectivity for a particular gas. Furthermore, this integrated experimental-computational workflow could be applicable to other supramolecular assemblies, such as metal-organic cages, and is not limited to reversible imine chemistries.

Applications of these new cages are currently under investigation: for example, after purification some of the hits from this study demonstrate gas uptakes that are comparable to other porous cages in the literature; **B11** and **B23** demonstrate  $\text{H}_2$  uptakes of 2.31 and 3.15  $\text{mmol g}^{-1}$  (77 K, 1 bar), respectively (Supplementary Figs. 32–34). We are also considering some of these molecules as tectons in other supramolecular assemblies, such as hydrogen-bonded organic frameworks or other extended crystalline frameworks<sup>56</sup>.

## Methods

**General synthetic and analytical methods.** See Supplementary Methods for further details.

**Procedures for the synthesis of precursors.** The synthesis and characterisation of triamines **A**, **B**, and **C**, is described in the Supplementary Methods. Aldehydes **1**, **3**, **10**, **11**, **13–15**, and **20–23** were commercially available, and the synthesis and characterisation of aldehydes **2**, **4–9**, **12**, **16–19**, and **24–26** is included in the Supplementary Methods. For spectra of novel precursors see Supplementary Figs. 35–42.

**Optimisation of model cage reactions.** Terephthalaldehyde **11** (9.7 mg, 0.072 mmol, 3.0 eq.) and (2,4,6-trimethylbenzene-1,3,5-triyl)trimethanamine **B** (2.0–3.0 eq.) were dissolved in  $\text{CDCl}_3$  (1–13 mL) and stirred at room temperature or 60 °C for 1–6 days (see Supplementary Table 15). The formation of insoluble precipitate was monitored, alongside reaction progress by  $^1\text{H}$  NMR spectroscopy, to determine the optimal cage formation conditions (see Supplementary Fig. 43). The optimised conditions were then investigated for scalability with the  $\text{Tri}^4\text{Di}^6$  cage **B11**, before confirming that the conditions were translatable to both a  $\text{Tri}^2\text{Di}^3$  (**B1**), and  $\text{Tri}^4\text{Tri}^4$  (**B23**) topology—see Supplementary Methods for the synthesis and characterisation details, and Supplementary Figs. 44–49 for the corresponding spectra.

**High-throughput synthetic screen general method.** Synthetic screening was carried out on a Chemspeed Accelerator SLT-100 synthesiser platform by liquid dispensing stock solutions of the precursors (2.5–5  $\text{mg mL}^{-1}$ , Supplementary Tables 16–19) into jacketed reactor vessels and heating for 3 days at 65 °C with vortexing (800 rpm)—for ease of analysis the reactions were carried out in deuterated chloroform. The reactions were analysed directly, prior to isolation, using high-throughput techniques including  $^1\text{H}$  NMR spectroscopy and high-resolution mass spectrometry (HRMS) —see Supplementary Figs. 50–72. For each targeted topology ( $\text{Tri}^2\text{Di}^3$ ,  $\text{Tri}^4\text{Di}^6$ , and  $\text{Tri}^4\text{Tri}^4$ ), a number of precursor combinations, which afforded clean cage formation, were selected, and diffusion NMR conducted, allowing the stoichiometry and sizes of the formed cages to be determined. For

combinations that indicated cage formation, the reaction solutions were then filtered, where necessary, to remove any insoluble precipitate (see Supplementary Table 20), and the solvent removed under reduced pressure in a Combindancer evaporator. The resulting isolated solids were analysed by powder X-ray diffraction (PXRD); all samples were amorphous on direct isolation from the reaction solvent (Supplementary Figs. 73–76). The stability to isolation was also investigated by re-dissolving and re-analysing the samples by  $^1\text{H}$  NMR spectroscopy—in some instances, the solid either did not fully re-dissolve suggesting some instability to isolation, or decomposition was visible after comparing the pre-isolation and post-isolation spectra—for a representative worked example of the high-throughput characterisation see Supplementary Note 7, and for full details and a summary of the results see Supplementary Methods and Supplementary Tables 21–26.

**Representative procedure for the scale-up of the new cage series, as exemplified by B11.** For ease of isolation, the solvent used in the scale-up reactions was switched from chloroform to DCM where possible, as this allowed a simpler solvent exchange to hexane with the resulting precipitated cages collected by filtration. A solution of terephthalaldehyde **11** (194 mg, 1.45 mmol, 6.0 eq.) and (2,4,6-trimethylbenzene-1,3,5-triyl)trimethanamine **B** (250 mg, 1.21 mmol, 5.0 eq.) dissolved in DCM (260 mL) was heated at 40 °C for 2 days. The reaction was allowed to cool to room temperature and the mixture filtered to remove any insoluble precipitate before the addition of hexane (200 mL). The DCM was carefully removed in vacuo to give a colourless precipitate which was collected by filtration and dried in vacuo to yield pure cage product **B11** (224 mg, 0.16 mmol, 65%). IR ( $\nu_{\text{max}}$  (cm $^{-1}$ )) 3391 (br), 2868, 1636, 1566, 1449, 1374, 1315, 1218;  $^1\text{H}$  NMR (500 MHz, CDCl $_3$ )  $\delta_{\text{H}}$  8.30 (t,  $J = 1.6$  Hz, 12H), 7.75 (s, 24H), 4.93 (s, 24H), 2.37 (s, 36H);  $^{13}\text{C}$  NMR (101 MHz, CDCl $_3$ )  $\delta_{\text{C}}$  160.4, 138.3, 137.7, 133.2, 128.5, 59.2, 16.4; HRMS (ES $^+$ ) calc. for **Tri $^4$ Di $^6$**  cage C $_{96}$ H $_{96}$ N $_{12}$  1417.7912; found  $[\text{M} + \text{H}]^+$  1418.8056,  $[\text{M} + 2\text{H}]^{2+}$  709.9087 and  $[\text{M} + \text{H} + \text{Na}]^{2+}$  720.8998.

The synthesis and characterisation data of the other cage hits are described in the Supplementary Methods—for the corresponding spectra, see Supplementary Figs. 77–114. See Supplementary Methods for crystal structure refinement details and notes for those cages and catenanes where single crystal X-ray structures could be obtained.

**Computational modelling.** To determine the intrinsic structural preference for a topology in the representative set of three systems (**B1**, **B11**, **B23**), models of the most likely topologies were assembled: for **B1** and **B11**—**Tri $^2$ Di $^3$** , **Tri $^4$ Di $^6$** , **Tri $^6$ Di $^9$**  and **Tri $^8$ Di $^{12}$** , and for **B23**—**Tri $^1$ Tri $^1$** , **Tri $^2$ Tri $^2$**  and **Tri $^4$ Tri $^4$** . Conformer searches were carried out for each structure using high-temperature MD simulations, followed by geometry optimisation, all with OPLS3 $^{42}$ . Low-energy structures were then geometry optimised with DFT calculations at the PBE+D3/TZVP level $^{57,58}$  and relative energies compared. The same computational approach was used to find the low-energy conformations for the complete set of high-throughput cages, assuming their targeted topology was realised. Precursor 7 was not included and precursor 10 was simplified by removing alkyl groups. The void diameters of cages were calculated by consideration of the distance between the centre of mass of a cage and the edge of the closest atom; the maximum cage diameter as the largest distance between the edges of any two atoms. The calculation of the average molecular diameter is described in the supporting information. For the set of molecules that unexpectedly formed **Tri $^2$ Di $^3$**  rather than the targeted **Tri $^4$ Di $^6$**  topology, the **Tri $^2$ Di $^3$**  topology was modelled to compare the relative energies of the two outcomes. To compare the relative energy of the bridged catenated structure to the related **Tri $^4$ Di $^6$**  topology, the structure of a single [8 + 12] molecule was taken from the SC-XRD structure and geometry optimised at the DFT level.

**Data availability.** Supplementary crystallographic files, which include structure factors, have been deposited with the Cambridge Crystallographic Data Centre (CCDC) as deposition numbers 1827867–1827887. These data files can be obtained free of charge from [http://www.ccdc.cam.ac.uk/data\\_request/cif](http://www.ccdc.cam.ac.uk/data_request/cif).

All relevant data including full synthetic, characterisation, crystallographic, and computational details are available in the Supplementary Information. Structures for all the low-energy computed cages are also supplied as Supplementary Data 1.

Received: 25 May 2018 Accepted: 21 June 2018

Published online: 20 July 2018

## References

- Liu, Z., Nalluri, S. K. M. & Stoddart, J. F. Surveying macrocyclic chemistry: from flexible crown ethers to rigid cyclophanes. *Chem. Soc. Rev.* **46**, 2459–2478 (2017).
- Kondratuk, D. V. et al. Two vernier-templated routes to a 24-porphyrin nanoring. *Angew. Chem. Int. Ed.* **51**, 6696–6699 (2012).
- Fujita, M. et al. Self-assembly of ten molecules into nanometre-sized organic host frameworks. *Nature* **378**, 469–471 (1995).
- Tozawa, T. et al. Porous organic cages. *Nat. Mater.* **8**, 973–978 (2009).
- Avellaneda, A. et al. Kinetically controlled porosity in a robust organic cage material. *Angew. Chem. Int. Ed.* **52**, 3746–3749 (2013).
- Zhang, G., Presly, O., White, F., Oppel, I. M. & Mastalerz, M. A permanent mesoporous organic cage with an exceptionally high surface area. *Angew. Chem. Int. Ed.* **53**, 1516–1520 (2014).
- Gil-Ramirez, G., Leigh, D. A. & Stephens, A. J. Catenanes: fifty years of molecular links. *Angew. Chem. Int. Ed.* **54**, 6110–6150 (2015).
- Xue, M., Yang, Y., Chi, X., Yan, X. & Huang, F. Development of pseudorotaxanes and rotaxanes: from synthesis to stimuli-responsive motions to applications. *Chem. Rev.* **115**, 7398–7501 (2015).
- Fielden, S. D. P., Leigh, D. A. & Woltering, S. L. Molecular knots. *Angew. Chem. Int. Ed.* **56**, 11166–11194 (2017).
- Erbas-Cakmak, S., Leigh, D. A., McTernan, C. T. & Nussbaumer, A. L. Artificial molecular machines. *Chem. Rev.* **115**, 10081–10206 (2015).
- Kassem, S. et al. Artificial molecular motors. *Chem. Soc. Rev.* **46**, 2592–2621 (2017).
- Fujita, D. et al. Self-assembly of tetravalent Goldberg polyhedra from 144 small components. *Nature* **540**, 563–566 (2016).
- Danon, J. J. et al. Braiding a molecular knot with eight crossings. *Science* **355**, 159–162 (2017).
- Jelfs, K. E. & Cooper, A. I. Molecular simulations to understand and to design porous organic molecules. *Curr. Opin. Solid State M* **17**, 19–30 (2013).
- Young, N. J. & Hay, B. P. Structural design principles for self-assembled coordination polygons and polyhedra. *Chem. Commun.* **49**, 1354–1379 (2013).
- Evans, J. D., Jelfs, K. E., Day, G. M. & Doonan, C. J. Application of computational methods to the design and characterisation of porous molecular materials. *Chem. Soc. Rev.* **46**, 3286–3301 (2017).
- Zhang, J. H. et al. Homochiral porous organic cage with high selectivity for the separation of racemates in gas chromatography. *Anal. Chem.* **87**, 7817–7824 (2015).
- Brutschy, M., Schneider, M. W., Mastalerz, M. & Waldvogel, S. R. Direct gravimetric sensing of GBL by a molecular recognition process in organic cage compounds. *Chem. Commun.* **49**, 8398–8400 (2013).
- Acharyya, K. & Mukherjee, P. S. A fluorescent organic cage for picric acid detection. *Chem. Commun.* **50**, 15788–15791 (2014).
- Giri, N. et al. Liquids with permanent porosity. *Nature* **527**, 216–220 (2015).
- Jiang, S. et al. Selective gas sorption in a [3+2] ‘propeller’ cage crystal. *Chem. Commun.* **47**, 8919–8921 (2011).
- Jelfs, K. E. et al. Large self-assembled chiral organic cages: synthesis, structure and self-persistence. *Angew. Chem. Int. Ed.* **50**, 10653–10656 (2011).
- Briggs, M. E. et al. Shape prediction for supramolecular organic nanostructures: [4+4] Macrocyclic tetrapods. *Cryst. Growth Des.* **13**, 4993–5000 (2013).
- Slater, A. G. et al. Reticular synthesis of porous molecular 1D nanotubes and 3D networks. *Nat. Chem.* **9**, 17–25 (2017).
- Bojdys, M. J. et al. Supramolecular engineering of intrinsic and extrinsic porosity in covalent organic cages. *J. Am. Chem. Soc.* **133**, 16566–16571 (2011).
- Hasell, T. et al. Controlling the crystallisation of porous organic cages: Molecular analogs of isoreticular frameworks using shape-specific directing solvents. *J. Am. Chem. Soc.* **136**, 1438–1448 (2014).
- Jones, J. T. A. et al. Modular and predictable assembly of porous organic molecular crystals. *Nature* **474**, 367–371 (2011).
- Colon, Y. J. & Snurr, R. Q. High-throughput computational screening of metal-organic frameworks. *Chem. Soc. Rev.* **43**, 5735–5749 (2014).
- Fraux, G. & Coudert, F. X. Recent advances in the computational chemistry of soft porous crystals. *Chem. Commun.* **53**, 7211–7221 (2017).
- Garcia, I. G., Bernabei, M., Soto, R. P. & Haranczyk, M. Out-of-oblivion cage molecules and their porous crystalline phases. *Cryst. Growth Des.* **17**, 5614–5619 (2017).
- Nazarian, D., Camp, J. S., Chung, Y. G., Snurr, R. Q. & Sholl, D. S. Large-scale refinement of metal-organic framework structures using Density Functional Theory. *Chem. Mater.* **29**, 2521–2528 (2016).
- Potyrailo, R. et al. Combinatorial and high-throughput screening of materials libraries: review of state of the art. *ACS Comb. Sci.* **13**, 579–633 (2011).
- Blanc, F. et al. Dynamic nuclear polarization NMR spectroscopy allows high-throughput characterisation of microporous organic polymers. *J. Am. Chem. Soc.* **135**, 15290–15293 (2013).
- Banerjee, R. et al. High-throughput synthesis of zeolitic imidazolate frameworks and application to CO $_2$  capture. *Science* **319**, 939–943 (2008).
- Kelty, M. L. et al. High-throughput synthesis and characterisation of nanocrystalline porphyrinic zirconium metal-organic frameworks. *Chem. Commun.* **52**, 7854–7857 (2016).

36. Santolini, V., Miklitz, M., Berardo, E. & Jelfs, K. E. Topological landscapes of porous organic cages. *Nanoscale* **9**, 5280–5298 (2017).
37. Zhu, G. et al. Formation mechanisms and defect engineering of imine-based porous organic cages. *Chem. Mater.* **30**, 262–272 (2017).
38. Santolini, V., Tribello, G. A. & Jelfs, K. E. Predicting solvent effects on the structure of porous organic molecules. *Chem. Commun.* **51**, 15542–15545 (2015).
39. De Rycke, N., Marrot, J., Couty, F. & David, O. R. P. Synthesis and characterisation of hexasubstituted azacryptands. *Tetrahedron Lett.* **51**, 6521–6525 (2010).
40. Arunachalam, M., Ravikumar, I. & Ghosh, P. A new hexaaza bicyclic cyclophane with dual binding sites. *J. Org. Chem.* **73**, 9144–9147 (2008).
41. Acharyya, K. & Mukherjee, P. S. Hydrogen-bond-driven controlled molecular marriage in covalent cages. *Chem. Eur. J.* **20**, 1646–1657 (2014).
42. Harder, E. et al. OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **12**, 281–296 (2016).
43. Moneypenny, T. P. et al. Product distribution from precursor bite angle variation in multitopic alkyne metathesis: evidence for a putative kinetic bottleneck. *J. Am. Chem. Soc.* **140**, 5825–5833 (2018).
44. Mastalerz, M., Lauer, J., Zhang, W. S., Rominger, F. & Schroder, R. Shape-persistent [4+4] imine cages with a truncated tetrahedral geometry. *Chem. Eur. J.* **24**, 1816–1820 (2018).
45. Mastalerz, M. One-pot synthesis of a shape-persistent endo-functionalised nano-sized adamantoid compound. *Chem. Commun.* **0** 4756–4758 (2008).
46. Avram, L. & Cohen, Y. Diffusion NMR of molecular cages and capsules. *Chem. Soc. Rev.* **44**, 586–602 (2015).
47. Robeson, L. M. Correlation of separation factor versus permeability for polymeric membranes. *J. Membr. Sci.* **62**, 165–185 (1991).
48. Stackhouse, Pugh, C. J. et al. Cage doubling: Solvent-mediated re-equilibration of a [3+6] prismatic organic cage to a large [6+12] truncated tetrahedron. *Cryst. Growth Des.* **18**, 2759–2764 (2018).
49. Liu, X. & Warmuth, R. Solvent effects in thermodynamically controlled multicomponent nanocage syntheses. *J. Am. Chem. Soc.* **128**, 14120–14127 (2006).
50. Hasell, T. et al. Triply interlocked covalent organic cages. *Nat. Chem.* **2**, 750–755 (2010).
51. Zhang, G., Presly, O., White, F., Opper, I. M. & Mastalerz, M. A shape-persistent quadruply interlocked giant cage catenane with two distinct pores in the solid state. *Angew. Chem. Int. Ed.* **53**, 5126–5130 (2014).
52. Jin, Y., Jin, A., McCaffrey, R., Long, H. & Zhang, W. Design strategies for shape-persistent covalent organic polyhedrons (COPs) through imine condensation/metathesis. *J. Org. Chem.* **77**, 7392–7400 (2012).
53. Jelfs, K. E. et al. In silico design of supramolecules from their precursors: Odd–even effects in cage-forming reactions. *J. Am. Chem. Soc.* **135**, 9307–9310 (2013).
54. Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
55. Pyzer-Knapp, E. O. et al. Predicted crystal energy landscapes of porous organic cages. *Chem. Sci.* **5**, 2235–2245 (2014).
56. Cooper, A. I. Porous molecular solids and liquids. *ACS Cent. Sci.* **3**, 544–553 (2017).
57. Perdew, J., Burke, K. & Ernzerhof, M. Generalised gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
58. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154–104(2010).

## Acknowledgements

We acknowledge funding from the European Research Council under FP7 (RobOT, ERC Grant Agreement no. 321156), the Engineering and Physical Sciences Research Council (EPSRC) (EP/N004884/1, EP/P005543/1), including the Materials Chemistry Consortium (EP/L000202/1) for time on the UK supercomputer, ARCHER. K.E.J. thanks the Royal Society for a University Research Fellowship. B.M.A. is affiliated to The Leverhulme Research Centre for Functional Materials Design at the University of Liverpool and would like to acknowledge their support in this research. We thank the Advanced Light Source, supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract no. DE-AC02-05CH11231, and thank Ming Liu, Andrew Stephenson, Simon J. Teat, Laura McCormick for their assistance during these experiments. We thank Diamond Light Source for access to beamlines I19 (MT15777) and Sarah Barnett for her assistance during this experiment. We thank Stephen Moss and the MicroBioRefinery for assistance with QTOF-MS measurements.

## Author contributions

R.L.G., K.E.J. and A.I.C. conceived the project. R.L.G., M.J.B., M.E.B., A.S., and H.A. prepared the precursors, and R.L.G., M.J.B., A.S. and H.C. carried out the high-throughput synthesis and characterisation of cage molecules. B.M.A. configured the Chemspeed platform, including designing the workflow architecture, setup, and programming for the high-throughput screens. E.G.B.E. carried out the diffusion NMR measurements. R.L.G. and M.J.B. scaled up the cage formations, and R.L.G., C.J.P., M.J.B. and M.A.L. crystallised the cage molecules. M.A.L. and C.J.P. interpreted the X-ray data. R.C. carried out the sorption measurements. K.E.J. led the computational modelling, with V.S. carrying out the modelling, with the exception of the computational calculation of solvodynamic diameters being carried out by M.M.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-05271-9>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018