

In silico design of supramolecules from their precursors: odd-even effects in cage-forming reactions

Kim E. Jelfs,^a Edward G. B. Eden,^a Jamie L. Culshaw,^a Stephen Shakespeare,^a Edward O. Pyzer-Knapp,^b Hugh P. G. Thompson,^b John Bacsá,^a Graeme M. Day,^{b,c} Dave J. Adams,^a Andrew I. Cooper^{a*}

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

^b Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.

^c Chemistry, University of Southampton, Southampton, SO17 1BJ, UK.

Supporting Information Placeholder

ABSTRACT: We synthesize a series of imine cage molecules where increasing the chain length of the alkanediamine precursor results in an alternation between [2+3] and [4+6] cage macrocycles. A computational procedure is developed to predict the thermodynamically preferred product and the lowest energy conformer, hence rationalizing the observed alternation and the 3-dimensional cage structures, based on the precursors.

The field of porous materials is largely dominated by extended networks, such as zeolites,¹ porous polymers² and metal-organic frameworks (MOFs).³ An alternative class of porous materials are discrete 'porous molecules',^{4,5} such as porous cages. Unlike networks, cage molecules assemble to form a molecular 'pore' in solution, prior to assembling in the solid state. Supramolecular cages can be metal-organic^{6,7} or wholly organic.⁸

Molecular porous materials offer certain potential advantages over extended networks and frameworks, such as solution processability.^{9,10} However, in terms of purposeful design of structure and function, porous molecular solids also introduce additional challenges.¹¹ In particular, while 3-D structure and topology are the overriding design parameters for frameworks,¹² *molar mass* is a key variable for discrete supramolecular assemblies, such as cages. Supramolecular cages may be prepared using either dynamic chemistry or irreversible reactions, but in both cases a given reaction can typically produce, in theory, a number of discrete molecular species, each with a different molecular weight. The course of the reaction can be influenced by subtle changes in the structure of the reagents, as well as by experimental factors such as concentration, the order of reagent addition, and the reaction solvent. It is therefore not straightforward to predict, *a priori*, precisely which product will be preferred in a new self-assembly reaction. As a result, the targeted synthesis of self-assembled

nanostructures is inherently difficult. For example, self-assembled metal-organic polyhedra can exhibit "emergent behavior", where small structural changes in ligands or templates lead to large and non-intuitive transitions to quite different geometrical cage products.^{13,14} Recently, Fujita *et al.* predicted empirically which metal-organic cage would be formed by determining the angle of the rigid ligands from density functional theory (DFT) calculations, and then matching this to bond angles in the various possible geometrical cage topologies.¹⁵ As yet, structures of organic cage compounds have not been predicted *a priori*, though this is a stated goal.^{4,16} The prediction of structure and size for new, hypothetical organic cages, and for other supramolecular organic species, is a necessary first step in *in silico* design of properties and function in organic solids. For example, while crystal structure prediction routines exist for the prediction of 3-dimensional crystal packings,¹⁷ and for predicting amorphous structures,¹⁸ at present, the structural formula of the molecule must be known beforehand from experiment. Unlike for MOFs, where an underlying network topology can be assumed for a given class of frameworks,¹² this uncertainty over molar mass for complex, self-assembled organic molecules precludes analogous computational strategies that screen large libraries of possible building blocks. Put simply: we cannot know for certain, *a priori*, how large the cage will be in a given, hypothetical reaction. The development of predictive methods to address this would be broadly useful in supramolecular organic chemistry, as well as in the specific field of porous solids. We could then apply computational methods to design structures, as was accomplished for zeolites through *in silico* design of organic templates.¹⁹

For organic cages that are synthesized by dynamic covalent chemistry,²⁰ the formation of the thermodynamic product, rather than a kinetic product,^{21,22} should in theory make computational predictions of structure more tractable. The challenge is two-fold: the total number of precursors that constitute the final product

must be predicted from many possible options. Indeed, for a dynamic $n+m$ condensation reaction, there are, in principle, an infinite number of integer multiples that can satisfy the underlying $n:m$ stoichiometry (i.e., $1[n+m]$, $2[n+m]$, ..., $x[n+m]$). An integral part of this prediction is the calculation of the most stable molecular conformation for each potential product, which is essential to generate meaningful energy rankings. This is challenging with large organic compounds, because models are difficult to construct and it is hard to determine the energetically preferred conformation. In practice, the molar mass of organic cage compounds has been observed to depend upon not only the precursors, but also the solvent²³ and reaction conditions, thus adding a further complexity to any computational strategy.

Previously, we and others have synthesized a range of porous imine cages by the condensation reaction of trialdehydes and diamines in $n[2+3]$ reactions to form $[2+3]$,^{24,25} $[4+6]$,^{8,17} and $[8+12]$ ²⁶ cages. The variation in cage size is a typical example of ‘emergent behavior’, since the trends observed are not simply intuitive. For instance, a small change in the diamine structure from 1,2-cyclopentane diamine to 1,2-cyclohexane diamine directs the product from a highly porous $[4+6]$ cage¹⁷ to a non-porous and much larger $[8+12]$ cage.²⁶ We have not yet observed any clear dependence on the choice of reaction solvent, but Warmuth *et al.* demonstrated variation between three different imine cage sizes that depend on solvent choice as well as ligand geometry.²³

Here, we synthesize a series of imine cages with increasing alkanediamine chain length (see reaction scheme, Fig. 1), and examine the effect on the cage size. We then use computational methods to rationalize the observed trend, and show that it is possible to predict the size and 3-D structure of these cages accurately.

Previously, the $[4+6]$ cage, **CC1**, was prepared via the condensation reaction of 1,3,5-triformylbenzene with 1,2-ethylenediamine.⁸ The most common of two conformers observed for **CC1** has tetrahedral symmetry and can exist as one of two helical enantiomers, **CC1-R** and **CC1-S** (Fig. S1). Here, we carried out the same reaction with a set of alkanediamines of increasing carbon chain length: 1,3-propanediamine, 1,4-butanediamine, and 1,5-pentanediamine (see ESI, Section 1). All reactions were carried out following the same methods used to form **CC1**,²⁷ with slight modifications in reaction times to optimize conversion. The products from the reactions with 1,3-diaminopropane and 1,4-diaminobutane (**CC-propane** and **CC-butane**, respectively) could not be isolated from their reaction mixtures. All attempts to do so led to the formation of insoluble material, most likely a cross-linked imine polymer. However, it was possible to characterize both **CC-propane** and **CC-butane** *in situ* by using an internal standard to calculate yields by NMR. **CC-pentane** could be isolated without decomposition and was recrystallized from methanol.

To examine the relative energies of the potential $n[2+3]$ combinations and the associated conformers of

these cages, models for each precursor combination were built in both a $[2+3]$ and $[4+6]$ ratio, assuming initially a similar geometry to that observed in earlier cages synthesized by our group. The OPLS all-atom force field²⁸ was used, since this was parameterized for organic systems with a focus on conformer energetics, intermolecular energies, and thermodynamic properties. OPLS was found to perform well at reproducing previously reported imine cage structures and energetics (see ESI, Section 2.2 & Fig. S6 for validation). Next, we used the MacroModel conformer searching tool to search for all the low-energy conformers of each structure, using the low-mode sampling method,²⁹ followed by refinement of the lowest energy conformations using dispersion-corrected DFT. The conformer generation and refinement was coupled with an automated procedure to check for conformations with an internal ‘pore’, defined here as a spherical void with a radius of >2.5 Å.

Our synthetic experiments showed a clear alternation in the cage size with increasing alkanediamine chain length. Diamines with an even number of carbon atoms all formed a $[4+6]$ cage, while those with an odd number formed a $[2+3]$ cage, as confirmed in each case by mass spectrometry (ESI Section 1 and Table 1 for the molecular ion masses). Single crystal X-ray diffraction structures for the $[4+6]$ cage **CC1**⁸ and, here, for the new **CC-pentane** cage, are also consistent with this odd/even alternation rule.

Odd-even effects are well known for n -alkanes and their derivatives,³⁰ with alternating patterns observed in properties such as melting point.³¹ An odd-even effect arising from the chain conformation is also observed for n -alkane assembly into monolayers on surfaces³² and in the formation of different network topologies,³³ but to our knowledge this is the first example of this effect in macrocyclic organic cage compounds.

The $[2+3]/[4+6]$ cage alternation rule was observed for all cages in this series across a range of reaction solvents, including dichloromethane, ethyl acetate, methanol, and ethanol. This insensitivity to solvent suggests that the alternating preference for $[2+3]$ and $[4+6]$ cage structures is intrinsic, and outweighs any influence of solvent stabilization energies. With the exception of **CC1**, which can be either porous or non-porous depending on its solid-state polymorph,³⁴ all of the other cages synthesized here were found to be non-porous.

These initial screening calculations were designed to identify the most likely reaction product formed in solution, rather than the solid state structure. Therefore, we only compared the energies of $[4+6]$ conformations that possess an internal cavity, ignoring those that are “collapsed”, and that hence lack a pore. We chose this approach because the cages are formed initially in solution with solvent molecules inside the cages that can act as scaffolds. For example, in a previous study for a cage molecule that collapses, we observed a cage cavity in the solvate structure²⁶ which was lost upon desolvation:

indeed, that material contained so much solvent that it was akin to a ‘solution-phase’ single crystal structure.

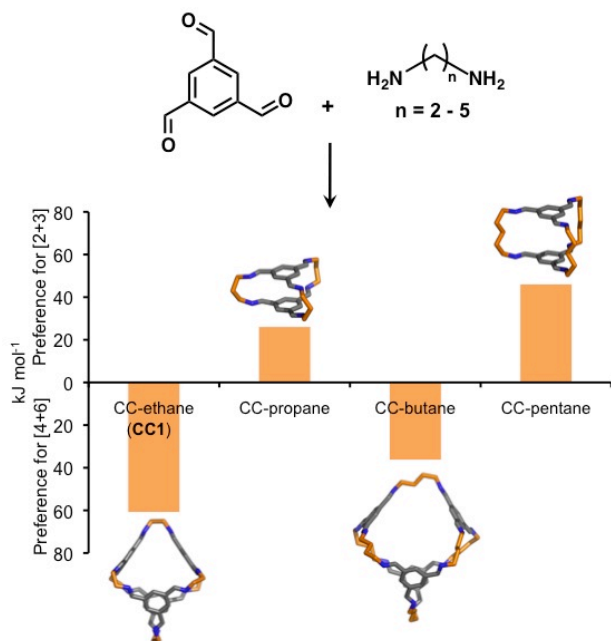


Figure 1. Alternation of cage size with increasing alkane chain length, as predicted by calculations, agrees with experiment (calculated structures shown here). Alkane chains in orange, hydrogens omitted.

Here, we defined a ‘pore’ as being present if the spherical radius of the cavity was larger than 2.5 Å, which is sufficient to encapsulate any of the solvents used here (ESI, Section 2.5). We note that if collapsed conformations are included, then a collapsed conformation of a [4+6] cage is the most energetically favored structure for $n=3$ and higher. This suggests a scaffolding effect of the solvent in directing the synthesis.

The lowest energy combination calculated for each molecule was found to be [4+6] for **CC1** (-61 kJ mol^{-1}) and **CC-butane** (-36 kJ mol^{-1}), and [2+3] for **CC-propane** (-26 kJ mol^{-1}) and **CC-pentane** (-46 kJ mol^{-1}) (Figure 1). The energies given in parentheses here are expressed as relative energy per [2+3] unit with respect to the alternative cage size. Hence, these calculations match the experimental observation of an odd-even effect. Table S3 gives the relative energies for all the low energy conformers, and Figure S9 gives the structures for the lowest energy conformer for each stoichiometry. The energetically preferred ‘inflated’ structures are all shown in Figure 1. For **CC1**, only one low-energy conformer (defined as lying within 20 kJ mol^{-1} of the minimum energy conformer, and ignoring mirror images) was found for the hypothetical [2+3] combination. However, two conformers, one with T and one with C_3 symmetry, were found for the [4+6] case, as found previously by experiment (Fig. 2).³⁴ For the longer chain diamines, there were often several low-energy conformers, typically differing only by small torsional differences in the alkane chains. For both of the [2+3] cages, the imine groups of the two aryl faces are aligned in the lowest

energy [2+3] conformation, such that they have a C_3 axis running through the centers of the arene rings.

So far, we have only been able to determine the single crystal structures of the **CC1** molecule (including both the T and C_3 symmetry conformers)^{8,34} and, of the new molecules here, a methanol solvate of **CC-pentane**. An overlay of molecular geometry found in these crystal structures and the lowest-energy calculated conformers are shown for these three systems in Figure 2. While one might have expected some differences between theory and experiment due to crystal packing forces, it is clear that the correct conformations have been determined by the calculations in each case. Indeed, there is excellent agreement, with root mean square displacements of atomic positions, excluding hydrogens, in the range 0.116–0.243 Å. Hence, the calculated conformations for the **CC-propane** and **CC-butane** molecules, for which we do not yet have crystal structures, can be considered as predictions that might be validated experimentally in the future (structures are provided in ESI).

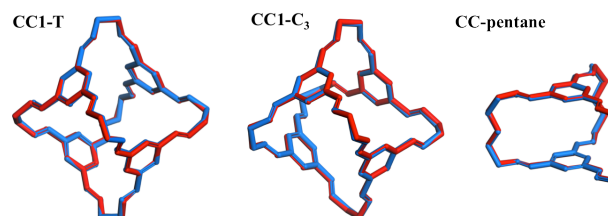


Figure 2. Overlay of X-ray structure (blue) and predicted structure (red) for **CC1-T**, **CC1-C₃**, and **CC-pentane**. **CC1-T** = desolvated crystal structure⁸; **CC1-C₃** = DCM solvate;²⁸ **CC-pentane** = methanol solvate prepared in this study.

As for the origin of this odd/even alternation, the [2+3] **CC1** cage is energetically disfavored because the ethane linker is simply too short to separate the aryl cage faces without unfavorable distortion of the arenes. For cages with longer alkane chains, it appears that the self-assembly reaction is directed towards structures that allow a staggered anti-conformation in the alkane chain, resulting in the nitrogen atoms being staggered gauche with respect to each other in the [4+6] molecules, or in an eclipsed configuration in the [2+3] molecules (Table 1).

Table 1. Mass spectrometry data and amine conformation for the lowest-energy computed conformer for each cage

Diamine (cage)	m/z	Molar ratio	Amine conformation	
			[2+3]	[4+6]
1,2-ethane (CC1) ⁸	793.4	[4+6]	Staggered gauche	Staggered gauche
1,3-propane (CC-propane)	439.3	[2+3]	Eclipsed	Synclinal
1,4-butane (CC-butane)	961.6	[4+6]	Non-linear	Staggered gauche
1,5-pentane (CC-pentane)	523.5	[2+3]	Eclipsed	Syn-periplanar

An examination of the various conformations allows us to comment on the shape-persistence of these molecules. All of the [4+6] molecules, with the exception of **CC1**, have low-lying collapsed conformations, possibly because of the easy rotation about alkane chain conformations (Fig. S9). This explains the experimental observation that the other [4+6] cage, **CC-butane**, is non-porous, and is not stable to desolvation. The comparison of computed energies for ‘porous’, inflated conformers and alternative non-porous, ‘collapsed’ conformers gives us a potential methodology to screen *in silico* for new intrinsically porous cages in the future. Indeed, one would have predicted from these calculations, *a priori*, that **CC1** is a good candidate for intrinsic porosity while **CC-butane** is not.

These computational methods could be automated in the future to efficiently screen hypothetical cage molecules prior to synthesis. To do this for large numbers of precursor combinations, as for MOFs,¹² it would first be necessary to build algorithms for the automated construction of starting models for the various $x[n+m]$ combinations, probably by approximating, in the first instance, to the relevant Platonic or Archimedean solids.⁶ Combined with methods for crystal structure prediction, which have been successfully applied to porous molecular crystals,^{17,35} this forms, in principle, a complete workflow for the computational prediction of 3-D structure and properties for supramolecular organic materials. The generalization of this strategy will be extremely challenging: for example, not all systems will be so insensitive to solvent stabilization effects, which are not considered in the isolated molecule calculations used in the present work. Also, without substantial development, these methods cannot routinely assess more complex topologies, such as catenanes,³⁶ where the partner cage in the catenane is undoubtedly competing with reaction solvent as a scaffold, or template. Nonetheless, these findings are the first step toward the long-term goal of a joined up methodology for computationally-guided design of organic crystalline porous solids, starting from the constituent chemical precursors.

Supporting Information. Experimental procedures, data, crystal structure, computational methods and predicted structures can be found in the supporting information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author aicooper@liv.ac.uk

ACKNOWLEDGMENT

We thank EPSRC for financial support under grant EP/H000925/1. A.I.C. is a Royal Society Wolfson Research Merit Award holder. This work made use of the facilities of HECToR through our membership of the UK’s HPC Materials Chemistry Consortium (EPSRC EP/F067496).

REFERENCES

(1) Cundy, C.; Cox, P. *Chem. Rev.* **2003**, *103*, 663.

- (2) Budd, P. M.; Mckeown, N. B. *Polym. Chem.* **2010**, *1*, 63.
 (3) Férey, G. *Chem. Soc. Rev.* **2008**, *37*, 191.
 (4) Mastalerz, M. *Angew. Chem. Int. Ed.* **2010**, *49*, 5042.
 (5) Tian, J.; Thallapally, P. K.; McGrail, B. P. *Cryst. Eng. Comm.* **2012**, *14*, 1909.
 (6) Seidel, S. R.; Stang, P. J. *Acc. Chem. Res.* **2002**, *35*, 972.
 (7) Lu, Z.; Knobler, C. B.; Furukawa, H.; Wang, B.; Liu, G.; Yaghi, O. M. *J. Am. Chem. Soc.* **2009**, *131*, 12532.
 (8) Tozawa, T. *et al. Nat. Mater.* **2009**, *8*, 973.
 (9) Budd, P. M.; Ghanem, B. S.; Makhseed, S.; Mckeown, N. B.; Msayib, K. J.; Tattershall, C. E. *Chem. Commun.* **2004**, 230.
 (10) Hasell, T.; Chong, S. Y.; Jelfs, K. E.; Adams, D. J.; Cooper, A. I. *J. Am. Chem. Soc.* **2012**, *134*, 588.
 (11) Desiraju, G. R. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2311.
 (12) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nature Chem.* **2011**, *4*, 83.
 (13) Sun, Q. F.; Iwasa, J.; Ogawa, D.; Ishido, Y.; Sato, S.; Ozeki, T.; Sei, Y.; Yamaguchi, K.; Fujita, M. *Science* **2010**, *328*, 1144.
 (14) Riddell, I. A.; Hristova, Y. R.; Clegg, J. K.; Wood, C. S.; Breiner, B.; Nitschke, J. R. *J. Am. Chem. Soc.* **2013**, *135*, 2723.
 (15) Bunzen, J.; Iwasa, J.; Bonakdarzadeh, P.; Numata, E.; Rissanen, K.; Sato, S.; Fujita, M. *Angew. Chem. Int. Ed.* **2012**, *51*, 3161.
 (16) Jelfs, K. E.; Cooper, A. I. *Curr. Opin. Solid State & Mater. Sci.* **2013**, DOI: 10.1016/j.cossms.2012.12.001.
 (17) Jones, J. T. A. *et al. Nature* **2011**, *474*, 367.
 (18) Abbott, L. J.; Mcdermott, A. G.; Del Regno, A.; Taylor, R. G. D.; Bezzu, G.; Msayib, K. J.; Mckeown, N. B.; Siperstein, F. R.; Runt, J.; Colina, C. M. *J. Phys. Chem. B* **2013**, *117*, 355.
 (19) Lewis, D.; Willock, D.; Catlow, C.; Thomas, J.; Hutchings, G. *Nature* **1996**, *382*, 604.
 (20) Lehn, J. *Chem-Eur. J.* **1999**, *5*, 2455.
 (21) Avellaneda, A.; Valente, P.; Burgun, A.; Evans, J. D.; Markwell-Heys, A. W.; Rankine, D.; Nielsen, D. J.; Hill, M. R.; Sumbly, C. J.; Doonan, C. J. *Angew. Chem. Int. Ed.* **2013**, *52*, 3746.
 (22) Cram, D. *Science* **1983**, *219*, 1177.
 (23) Liu, X.; Warmuth, R. *J. Am. Chem. Soc.* **2006**, *128*, 14120.
 (24) Jiang, S.; Bacsá, J.; Wu, X.; Jones, J. T. A.; Dawson, R.; Trewin, A.; Adams, D. J.; Cooper, A. I. *Chem. Commun.* **2011**, *47*, 8919.
 (25) Schneider, M. W.; Opper, I. M.; Mastalerz, M. *Chem-Eur. J.* **2012**, *18*, 4156.
 (26) Jelfs, K. E.; Wu, X.; Schmidtman, M.; Jones, J. T. A.; Warren, J. E.; Adams, D. J.; Cooper, A. I. *Angew. Chem. Int. Ed.* **2011**, *50*, 10653.
 (27) Lydon, D. P.; Campbell, N. L.; Adams, D. J.; Cooper, A. I. *Synth. Commun.* **2011**, *41*, 2146.
 (28) Jorgensen, W. L.; Maxwell, D.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
 (29) Kolossvary, I.; Guida, W. *J. Am. Chem. Soc.* **1996**, *118*, 5011.
 (30) Kishore, K.; Bharat, S.; Kannan, S. *J. Chem. Phys.* **1996**, *105*, 11364.
 (31) Badea, E.; Gatta, G. D.; D’Angelo, D.; Brunetti, B.; Rečková, Z. *J. Chem. Thermodynamics* **2006**, *38*, 1546.
 (32) Espeau, P.; Reynolds, P. A.; Dowling, T.; Cookson, D.; White, J. W. *Faraday Trans.* **1997**, *93*, 3201.
 (33) Mukherjee, G.; Biradha, K. *Cryst. Growth. Des.* **2011**, *11*, 924.
 (34) Jones, J. T. A.; Holden, D.; Mitra, T.; Hasell, T.; Adams, D. J.; Jelfs, K. E.; Trewin, A.; Willock, D. J.; Day, G. M.; Bacsá, J.; Steiner, A.; Cooper, A. I. *Angew. Chem. Int. Ed.* **2010**, *123*, 775.
 (35) Cruz-Cabeza, A. J.; Day, G. M.; Jones, W. *Chem-Eur. J.* **2009**, *15*, 13033.
 (36) Hasell, T.; Wu, X.; Jones, J. T. A.; Bacsá, J.; Steiner, A.; Mitra, T.; Trewin, A.; Adams, D. J.; Cooper, A. I. *Nature Chem.* **2010**, *2*, 750.

