J Nonlinear Sci
DOI 10.1007/s00332-016-9286-9

Journal of Nonlinear Science

Energy Landscapes for the Self-Assembly of Supramolecular Polyhedra

Emily R. Russell[1] · **Govind Menon**[2]

Received: 16 June 2015 / Accepted: 23 January 2016

Abstract We develop a mathematical model for the energy landscape of polyhedral supramolecular cages recently synthesized by self-assembly (Sun et al. in Science 328:1144–1147, 2010). Our model includes two essential features of the experiment: (1) geometry of the organic ligands and metallic ions; and (2) combinatorics. The molecular geometry is used to introduce an energy that favors square-planar vertices (modeling Pd^{2+} ions) and bent edges with one of two preferred opening angles (modeling boomerang-shaped ligands of two types). The combinatorics of the model involve two-colorings of edges of polyhedra with four-valent vertices. The set of such two-colorings, quotiented by the octahedral symmetry group, has a natural graph structure and is called the combinatorial configuration space. The energy landscape of our model is the energy of each state in the combinatorial configuration space. The challenge in the computation of the energy landscape is a combinatorial explosion in the number of two-colorings of edges. We describe sampling methods based on the symmetries

Communicated by Robert V. Kohn.

ERR acknowledges the support and hospitality of the Institute for Computational and Experimental Research in Mathematics (ICERM) at Brown University, where she was a postdoctoral fellow while carrying out this work. GM acknowledges partial support from NSF grant DMS 14-11278. This research was conducted using computational resources at the Center for Computation and Visualization, Brown University.

✉ Govind Menon
govind_menon@brown.edu

Emily R. Russell
emily_russell@post.harvard.edu

1 ICERM, Brown University, Box 1955, Providence, RI 02912, USA

2 Division of Applied Mathematics, Brown University, 182 George St., Providence, RI 02912, USA

Published online: 08 February 2016

of the configurations and connectivity of the configuration graph. When the two preferred opening angles encompass the geometrically ideal angle, the energy landscape exhibits a very low-energy minimum for the most symmetric configuration at equal mixing of the two angles, even when the average opening angle does not match the ideal angle.

Keywords Supramolecular chemistry · Self-assembly · Polyhedra · Molecular cages

Mathematics Subject Classification 52B05 · 70B15 · 92C40 · 92E10

1 Introduction

The central goal of the field of synthetic self-assembly is to create ordered assemblies of building blocks by mimicking the essentials of biophysical processes of self-assembly. In supramolecular chemistry, the building blocks are molecules, and a basic problem is to synthesize supramolecular cages, to shed light on the self-assembly of viral capsids (Liu et al. 2011; Takeda et al. 1999; Caspar and Klug 1962). Such cages may be used to contain or catalyze chemical reactions with great precision. For example, molecular flasks that stabilize highly reactive compounds, such as phosphorus, have been synthesized by self-assembly (Mal et al. 2009). While the chemistry of these examples differ, all of them share fundamental geometrical features. The cages are Platonic or Archimedean polyhedra, and their molecular design relies on the decomposition of the polyhedra into simpler geometric units, and a careful choice of molecules whose reactive units can be idealized as such units. For example, in the experiments reported in (Liu et al. 2011), the fundamental units are hexagonal 'molecular tiles' that assemble into (quasi)-truncated octahedra, each molecular tile forming one hexagonal face of the truncated octahedron.

Our primary interest in this article is to develop a mathematical model that sheds light on the self-assembly of organometallic molecular cages synthesized in Fujita's lab (Sun et al. 2010). These cages may be idealized as Platonic or Archimedean solids with four-valent vertices. They are realized in experiments by organometallic complexes consisting of four-valent palladium (Pd^{2+}) ions at vertices linked by 'boomerang-shaped' organic ligands. All such molecular cages have the general chemical formula $M_n L_{2n}$. The subscript of M indicates the number of palladium ions at the vertices, and the subscript of L denotes the number of organic ligand molecules making up the edges of the polyhedron. The symmetry of the vertices forces $n = 6$, 12, 24, 30 and 60. Several of these ligand polyhedra were observed in experiment, including the octahedron ($n = 6$), cuboctahedron ($n = 12$) and rhombicuboctahedron ($n = 24$) (Sun et al. 2010), shown in Fig. 1a–c.

A particularly interesting effect was observed when two different ligand molecules with differing bend angle were mixed. The concentration of the two different organic ligand molecules, types A and B say, served as a control parameter. Separately, these ligands formed cuboctahedra (Fig. 1b) and rhombicuboctahedra (Fig. 1c) of the form $M_{12}A_{24}$ and $M_{24}B_{48}$, respectively. In a mixture, a variety of supermolecules with the

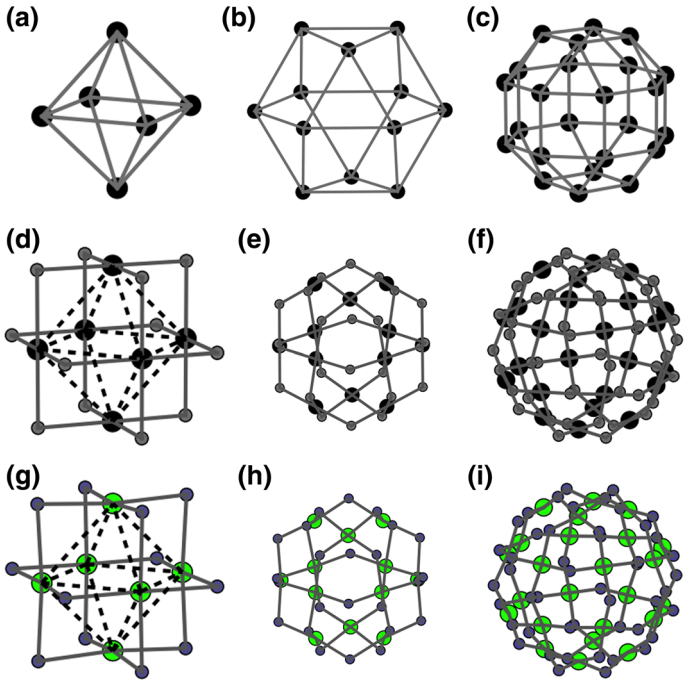

Fig. 1 Four-valent polyhedra with O_h symmetry. **a** Octahedron, **b** cuboctahedron, **c** rhombicuboctahedron. **d–f** Ideal supramolecular cage embeddings. Here, the Pd^{2+} ions are represented by *large black vertices*, while the 'boomerang'-shaped organic ligands are represented by *small gray vertices* ('elbows'), each with two arms; in **d**, the *straight edges* of the octahedron are included only for comparison. Note that the geometry at the ions is square planar. **g–i** Frustration. Minimum-energy embeddings when the preferred bend angle of the edges is $\theta_0 < \theta_*$. **g** $\theta_0 = 81°$, $\theta_* = 90°$, **h** $\theta_0 \approx 105.5°$, $\theta_* \approx 117.2°$, **i** $\theta_0 \approx 120.7°$, $\theta_* \approx 134.1°$. *Brightness of color* indicates the energy contribution from each vertex and each 'elbow.' Note that the geometry at the ions is deformed from square-planar (Color figure online)

general chemical formula $M_n A_{2n-m} B_m$, $0 \leq m \leq 2n$, and $n = 12$ or $n = 24$ are theoretically possible. Further, for $1 \leq m \leq 2n-1$ these polyhedral supermolecules typically have many distinct isomers, corresponding to different geometric arrangements of the A and B ligands. In experiment, as the concentration ratio was varied, a sharp transition was observed from a solution in which only cuboctahedra formed to one in which only rhombicuboctahedra formed. More precisely, while a solution consisting of all isomers $M_n A_{2n-m} B_m$ with $n = 12$ or $n = 24$ and $0 \leq m \leq 2n$ is possible in theory, in each experiment with a fixed concentration ratio of A and B ligands, the equilibrium solution consisted of only cuboctahedral isomers ($n = 12$) or only rhombicuboctahedral isomers ($n = 24$). The sharpness of this phase transition was described as *emergent behavior* by Fujita and co-authors (2010).

In order to understand what drives this phase transition, we develop a 'minimal' model that combines the geometry of isomers with a phenomenological energy for deviations from the ideal bend angle. We then compute an energy landscape for isomers with the general chemical formula $M_n A_{2n-m} B_m$, $n = 6$, 12 and 24, $0 \leq m \leq 2n$. These computations reveal interesting features, such as the symmetries of minimum-

energy states. Further, they reveal the need to carefully account for symmetry in order to obtain a computationally tractable problem.

Our work suffers from certain limitations. First, the results presented here are zero-temperature calculations, so entropy and kinetics are ignored. Including these effects seems necessary to explain the emergence seen in Sun et al. (2010). We conclude this article with some remarks that indicate how entropic effects could be included by building on our computations. Second, from the standpoint of modeling, it should be noted that the physics of molecular cages is not as well understood as that of other model systems in self-assembly such as rigid colloidal clusters [see, e.g., Holmes-Cerfon et al. (2013) and the references therein]. The governing assumption behind our work is that physical models that mimic essential features of the experiments—in this case, the polyhedral geometry, a simple elastic energy and the combinatorics of isomers—serve to describe the system. These assumptions yield a model of appealing simplicity, but ignore the possibility that a detailed understanding of the chemical interaction of molecular cages is necessary to explain the experiments.

To the best of our knowledge, our work is the first attempt to develop a theoretical model for the experiments that is faithful to the underlying geometric combinatorics. Despite the apparent specificity of our model, the ideas presented here may be naturally adapted to the other experiments. More broadly, the development of a theoretical understanding of self-assembly is a deep fundamental exercise, rich in biological, mathematical and physical ideas. Our work is an instance of small, but growing, mathematical literature on synthetic self-assembly that emphasizes ideas from discrete geometry and statistical physics (Holmes-Cerfon et al. 2013; Pandey et al. 2011; Kaplan et al. 2014). These models typically include three aspects. The first is the identification of a configuration space that idealizes the set of intermediate structures that lie between the building blocks and the assembled product. The second is to understand the free energy landscape; in particular, to understand which of the states in the configuration space are energetically favorable. The third is a description of the kinetics of self-assembly. In this article, we focus on the first two aspects of this procedure, since (as will become clear) determining the energy landscape requires intensive computation, and is of independent interest.

The rest of this article is organized as follows. We describe the configuration space and energy in the next section. This is followed by a description of the combinatorial explosion and a symmetry-based sampling scheme. Finally, we describe our numerical results and conclusions.

2 The Configuration Space and Energy Function

We are interested in molecules of the form $M_n A_{2n-m} B_m$, with n vertices corresponding to the ions M, $2n-m$ ligands of type A and m ligands of type B. Our model involves: (1) simple geometric combinatorics, (2) a phenomenological energy function and (3) sampling schemes for fast computation. (1) For any m, the number of distinct isomers is the number of distinct two-colorings of the edges of the polyhedra modulo the action of the symmetry group of the polyhedron. (2) Our energy function penalizes the geometric distortion of the polyhedra caused by ligand molecules with differing

bend angles. (3) A calculation of the full energy landscape is impossible for $n = 12$ and $n = 24$ because of a combinatorial explosion. We use symmetry and heuristics based on $n = 6$ to sample the energy landscape.

2.1 Colorings and Configurations

More formally, let $\mathcal{P}$ be a polyhedron whose faces, edges and vertices are denoted $(\mathcal{F}, \mathcal{E}, \mathcal{V})$. We assume that $\mathcal{P}$ is one of the following: the octahedron; cuboctahedron; or rhombicuboctahedron. All these polyhedra, shown in Fig. 1, have regular four-valent vertices, and their symmetry group is the octahedral group, O_h, which is the direct product of S_4, the permutation group of four elements corresponding to permutations of the four pairs of opposing faces of the octahedron, and S_2, corresponding to reflection (Coxeter 1973). A two-coloring of edges, or coloring for short, is a map $c{:}\mathcal{E} \to \{0, 1\}$. The set of colorings is denoted C. It is convenient to view the colorings 'physically': we identify the polyhedron $\mathcal{P}$ with its standard embedding in space and paint the edges of $\mathcal{P}$ blue (0) or red (1), for ligands of type A and type B, respectively. We say that two colorings c_1 and c_2 are equivalent under O_h, written $c_1 \sim c_2$, if there exists $g \in O_h$ such that $gc_1 = c_2$. We define a *configuration* to be an equivalence class of colorings under the relation $\sim$. The quotient set C/O_h is called the *combinatorial configuration space*, or simply *configuration space*. It is the set of all distinct colorings modulo the symmetry of the polyhedron.

The enumeration of C/O_h is a classic problem in discrete group theory [we follow (Rotman 1995)]. Let $O_h c$ denote the orbit of a coloring c. We say that the *degeneracy* of a coloring c is the size of the orbit $|O_h c|$. Clearly, the degeneracy is the same for all colorings in the equivalence class $[c]$. Thus, it is meaningful to speak of the degeneracy of $[c]$. Let $o_c \subset O_h$ denote the stabilizer subgroup of $c \in C$. We call the size of this subgroup, $|o_c|$, the *symmetry number* of c, and denote it by s. By the orbit-stabilizer theorem and Lagrange's theorem, for every coloring $c \in C$,

$$|O_h c| \times |o_c| = |O_h| = 48. \tag{1}$$

Two colorings c_1 and c_2 are equivalent if and only if they lie on the same group orbit. Thus, the size of the configuration space, $|C/O_h|$, is simply the number of distinct orbits. Let C^g denote the set of colorings fixed by $g \in O_h$, i.e., $C^g = \{c \in C \,|\, gc = c\}$. By Burnside's lemma,

$$|C/O_h| = \frac{1}{|O_h|} \sum_{g \in O_h} |C^g|. \tag{2}$$

While this calculation of the size $|C/O_h|$ is well known, our interest lies in an explicit description of each equivalence class $[c] \in C/O_h$. We compute C/O_h explicitly for the octahedron and cuboctahedron as described below. We are unable to explicitly compute C/O_h for the rhombicuboctahedron because of a combinatorial explosion. Note that $|C/O_h|$ is bounded below by

$$|C/O_h| \geq \frac{|C|}{|O_h|} = \frac{2^{|\mathcal{E}|}}{48}. \tag{3}$$

The octahedron has 144 unique configurations (the lower bound is 85), and the cuboctahedron has 352,744 unique configurations (the lower bound is 349, 525). For the rhombicuboctahedron, the same bound yields $2^{48}/48 \approx 5, 864, 062, 014, 806$, which is too large to enumerate explicitly. Thus, one of the main computational challenges in our work is to obtain a realistic understanding of the energy landscape, despite the combinatorial explosion of $|C/O_h|$.

Finally, note that the configuration space C/O_h inherits a natural graph structure from C. We say that two colorings c_m and c_{m+1} with m and $m+1$ 'red' edges are neighbors in C if they differ at a single edge. If c_m and c_{m+1} are neighbors, it is clear that gc_m and gc_{m+1} are also neighbors for each $g \in O_h$. Thus, every coloring in the orbit $O_h c_m$ is the neighbor of at least one coloring in the orbit $O_h c_{m+1}$ and it is natural to say that $[c_m]$ and $[c_{m+1}]$ are neighbors in C/O_h.

2.2 A Phenomenological Energy for Molecular Distortion

Each configuration $[c] \in C/O_h$ determines the unique combinatorial structure of a supramolecular conformation. An *energy landscape* is a map $F{:}C/O_h \to \mathbb{R}$.

We define a phenomenological energy by penalizing the distortion both of the edges and of the vertices from their preferred geometry. The edges are modeled as two straight arms, joined at an angle at the 'elbow' of the boomerang (Fig. 2a). Let $X \in \mathbb{R}^{3|\mathcal{E}|+3|\mathcal{V}|}$ denote the coordinates of the elbows of the ligands and of the vertices of the embedded polyhedron. We introduce energy terms which account separately for the distortion from a preferred length for each arm of an edge (E_{arms}); distortion from the preferred bend angle at each elbow (E_{elbow}); and distortion from square-planar geometry at each vertex (E_{vertex}):

$$G(X;[c]) = \sum_{\mathfrak{e}\in\mathcal{E}} \left(E_{\text{arms}}(\mathfrak{e}(X)) + E_{\text{elbow}}(\mathfrak{e}(X))\right) + \sum_{\mathfrak{v}\in\mathcal{V}} E_{\text{vertex}}(\mathfrak{v}(X)). \tag{4}$$

Here the notation $\mathfrak{e}(X)$ and $\mathfrak{v}(X)$ implies that the energy depends on the coordinates of each edge and vertex, respectively. Finally, the energy landscape $F([c])$ is defined by

$$F([c]) = \inf_X G(X;[c]), \tag{5}$$

where the infimum is taken over all embeddings of $\mathcal{P}$ into $\mathbb{R}^3$. In our numerical experiments, the energy G is minimized through a zero-temperature Monte Carlo annealing; for the largest system, the rhombicuboctahedron, the minimization takes about 100 s for a tolerance in the vertex positions of about 0.3 % of the arm length, or about 140 s for a tolerance of 0.1 % of the arm length.[1] This annealing does not reach the exact energy minimum, but gives an upper bound on the minimum energy which

[1] The tolerance is the width of the Gaussian distribution of random test displacements applied to the vertices in each dimension; we use successively smaller tolerances, and end the annealing process when the evolution of the total energy at the lowest tolerance has a sufficiently small average slope, the threshold slope set by trial and error to balance accuracy with efficiency.

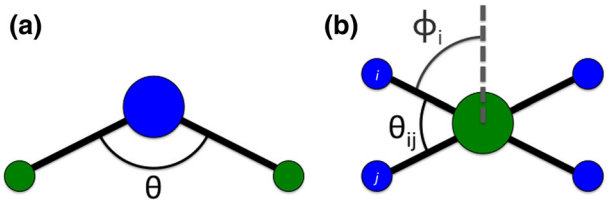

Fig. 2 Geometry of **a** edges, comprising an 'elbow' and two arms, and **b** four-valent vertices

we believe to be typically within a few percent of the actual infimum. This fact may be verified for the minimum-energy configurations shown in Fig. 1d–f.

We now describe the terms in the energy in local coordinates in more detail. We use a harmonic approximation for each energy term, that is, a term which is quadratic about the equilibrium point, as is standard practice in modeling of physical systems; physically accurate energies will have a more complex form, but this lowest-order approximation should capture the most interesting effects. We introduce an energy cost to deviations of the length of the arms from one:

$$E_{\text{arm}} = E_a \left((l_1 - 1.0)^2 + (l_2 - 1.0)^2 \right), \tag{6}$$

where l_1 and l_2 are the lengths of the two arms. We choose $E_a \gg E_e, E_p, E_v$ (defined below) to make the arms stiff; this has the effect that in general in the minimum-energy embedding, the arms are of length one and do not contribute significantly to the total energy.

We also introduce an energy cost to deviations of the angle of the elbow from its preferred opening angle:

$$E_{\text{elbow}} = E_e(\theta - \theta_0)^2 \tag{7}$$

with θ the actual opening angle and θ_0 the preferred opening angle (which depends on the type A or B). Thus, this term is the only term in the energy which depends on the *configuration* as well as on the embedding. The coefficient E_e sets the scale of the energy contribution of the edges. At each vertex, we introduce an energy cost for deviations from square-planar geometry (see Fig. 2b):

$$E_{\text{vertex}} = E_v \left[\sum_{\langle i,j \rangle} (\theta_{ij} - \pi/2)^2 + E_p \sum_i (\phi_i - \pi/2)^2 \right], \tag{8}$$

with the first sum over pairs of neighboring edges i and j and θ_{ij} the angle formed by the 'arms' of those edges; and the second sum over edges i and ϕ_i the angle formed between the arm of that edge and 'plane perpendicular,' defined as the average of the cross-products of the pairs of neighboring edges. The first sum pushes the edges to all be perpendicular to one another (square), while the second sum pushes the edges to be coplanar. The overall coefficient E_v sets the scale of the energy contribution of the vertices relative to the edges, while the coefficient E_p sets the scale of the energy contribution of the 'planar' aspect of the vertices relative to the 'square' aspect.

For each polyhedron, there is an ideal angle θ_* such that there is a zero-energy state when all edges have $\theta_0 = \theta_*$. We calculate these ideal angles analytically for each of the three polyhedra considered. Fixing one vertex at the origin, we write explicitly the positions **A**, **B**, **C**, **D** of its four neighbors in a regular polyhedron (ordered cyclically so that **B** and **D** are each closer to **A** than is **C**). We then solve for **E**, **F**, **G**, **H**, the positions of the elbows, which satisfy the following conditions:

$$\begin{aligned} |\mathbf{E}| &= |\mathbf{F}| = |\mathbf{G}| = |\mathbf{H}| \\ \mathbf{G} &= -\mathbf{E}\,,\ \mathbf{H} = -\mathbf{F} \\ \mathbf{E}\cdot\mathbf{F} &= \mathbf{E}\cdot\mathbf{H} = \mathbf{F}\cdot\mathbf{G} = \mathbf{G}\cdot\mathbf{H} = 0 \\ |\mathbf{A}-\mathbf{E}| &= |\mathbf{E}|\,,\ |\mathbf{B}-\mathbf{F}| = |\mathbf{F}|\,,\ |\mathbf{C}-\mathbf{G}| = |\mathbf{G}|\,,\ |\mathbf{D}-\mathbf{H}| = |\mathbf{H}| \end{aligned}$$

These relations ensure that the geometry at the vertex is square planar and that every arm of an edge is of the same length. Examining the triangle **OEA**, we see that the opening angle θ is then given by

$$\theta_* = \cos^{-1}\left(1 - \frac{|\mathbf{A}|^2}{2|\mathbf{E}|^2}\right) \tag{9}$$

Thus, if the preferred angle of every edge is set to θ_*, the embedding with this geometry at every vertex of the Platonic or Archimedean polyhedron will give a zero-energy configuration; these embeddings are shown in Fig. 1d–f. The values derived for the ideal angles of the three polyhedra are:

$$\begin{aligned} \theta_{*,\text{octahedron}} &= 90^\circ, \\ \theta_{*,\text{cuboctahedron}} &= \cos^{-1}\left(\frac{1-2\sqrt{2}}{4}\right) \approx 117.2^\circ, \\ \theta_{*,\text{rhombicuboctahedron}} &= \cos^{-1}\left(\frac{25-20\sqrt{2}}{33-20\sqrt{2}}\right) \approx 134.1^\circ. \end{aligned}$$

Note that these values for the cuboctahedron and rhombicuboctahedron correct the approximate values of 120° and 135°, respectively, given in Sun et al. (2010).

However, in experiment this value can only be chosen approximately, and the angle θ_0 is different from the ideal value. This causes frustration, as shown in Fig. 1g–i, and the minimum-energy state balances competing energies. In these examples, the total energy of the vertices is between 1.8 and 2.0 times the total energy of the edges. All of the energy contributions are necessary to observe frustration. This can be seen by counting the degrees of freedom. For example, if the planarity of the vertices is relaxed, the edge 'elbows' can rearrange themselves so that they satisfy their preferred angles at the expense of the square-planar geometry at the vertices. Only when all four energy terms are in competition do we have an interesting system. It is in this sense that we consider our energy model to be minimal.

Finally, we define the *average* preferred angle for mixtures of two ligands. The preferred angle used in Sun et al. (2010) is:

$$\bar{\theta}_0(c_A, c_B) = \frac{1}{c_A + c_B} \left(c_A \theta_0(A) + c_B \theta_0(B)\right), \tag{10}$$

where c_A and c_B denote the concentrations of ligands A and B. We will also use $\bar{\theta}_0$ with no arguments to mean

$$\bar{\theta}_0 \equiv \bar{\theta}_0(c, c) = \frac{1}{2} \left(\theta_0(A) + \theta_0(B)\right) \tag{11}$$

and $\bar{\theta}_0(m)$ in an isomer $M_n A_{2n-m} B_m$ to mean

$$\bar{\theta}_0(m) \equiv \bar{\theta}_0(2n - m, m) = \frac{1}{2n} \left((2n - m)\theta_0(A) + m\theta_0(B)\right) \tag{12}$$

3 Symmetries and Sampling Strategies

As discussed above, our goals are twofold: (1) to enumerate the configuration space C/O_h and (2) to calculate the energy landscape $F(C/O_h)$. The computational difficulty of these tasks depends strongly on the size of the polyhedron $\mathcal{P}$. When $\mathcal{P}$ is the octahedron, we are able to fully enumerate the configuration space and its graph structure, and compute the entire energy landscape. When $\mathcal{P}$ is the cuboctahedron, we fully enumerate the configuration space and its graph structure, but we sample the energies of only a fraction of the configuration space to compute a partial energy landscape. When $\mathcal{P}$ is the rhombicuboctahedron, even the configuration space is intractable. We enumerate only those configurations c with non-trivial symmetry number s, and sample the energy of only a fraction of these configurations. In what follows, we focus mainly on the sampling strategies used to obtain partial, but informative, descriptions of the configuration space and energy landscape for the two larger polyhedra.

The octahedron is the most straightforward. With only $2^{12} = 4096$ colorings to consider, we construct every coloring and calculate the equivalences under O_h to obtain a complete list of configurations. There are 144 configurations, of which 94 have non-trivial symmetry groups (see Table 1). We also calculate explicitly the graph structure, identifying the neighbor relations between configurations. We then calculate the energy of every one of the 144 configurations by solving the minimization problem (5).

We develop an algorithm to build the graph C/O_h for the cuboctahedron by induction on the number of 'red' edges, denoted m. There is a single configuration which has $m = 0$, that is, all edges are 'blue'; this provides the base case. As our induction step, we use the subset $(C/O_h)_m$ of configurations with m 'red' edges to construct the set $(C/O_h)_{m+1}$ as follows. For each element $[c] \in (C/O_h)_m$, we consider a representative coloring $c \in [c]$. This coloring has neighbors $\{c'\} \subset C_{m+1}$, obtained from c by changing the color of a single edge from 'blue' to 'red.' We compute the orbit $O_h c'$ for each neighbor c' and compare each element of $O_h c'$ against the configurations already seen in $(C/O_h)_{m+1}$. If the configuration has already been seen in $(C/O_h)_{m+1}$,

Table 1 Enumeration of configurations by symmetry number for each polyhedron

Symmetry number	Octahedral configurations	Cuboctahedral configurations	Rhombicuboctahedral configurations
48	2	2	4
24	0	1	4
16	2	2	28
12	2	6	29
8	6	23	644
6	4	35	716
4	21	211	46,991
3	1	44	16,015
2	56	5943	>40,648,786
1	50	346,477	$\approx 6 \times 10^{12}$
Total	144	352,744	$\approx 6 \times 10^{12}$

we make note that $[c]$ and $[c']$ are neighbors. If the configuration has not been seen before, we update our list of representatives in $(C/O_h)_{m+1}$ to include $[c']$ and make note of the neighbor relationship. This process ends when we find the single configuration with $m = 2n = 24$.[2] Thus, we do not enumerate every possible coloring, but we nonetheless identify all configurations and compute the connectivity of the graph C/O_h. We find a total of 352,744 configurations (see Table 1).

The computation of the energy for each configuration $[c]$ involves a numerical solution of the minimization problem (5). Rather than pursue this computation for each of the 352,744 configurations, we sample a fraction of the configurations and compute a partial energy landscape. We use a sampling strategy which is designed to identify both the lowest-energy configurations at each m, which are those of the greatest interest in determining the configurations present in a self-assembling physical system, and also the highest-energy configurations, allowing us to understand the scale of the energy differences which determine the relative proportions of configurations present, as well as giving us a more complete picture of the energy landscape. This strategy exploits two heuristic observations about the energy landscape of the octahedron:

1. For a given m, the highest- and lowest-energy configurations tend to have nontrivial symmetry number s;
2. High-energy configurations tend to have high-energy neighbors in the graph. Likewise, low-energy configurations have low-energy neighbors.

Our sampling strategy consists of two steps. First, we calculate the energy for all configurations with 'extreme' values of m, here defined as $m \leq 4$ or $m \geq 20$ (630 configurations total), and for all configurations with symmetry number $s \geq 3$ (324 configurations total, although there is overlap with the 630 extreme-m configurations).

[2] We could also use the symmetry between 'red' and 'blue' by ending the process when we have enumerated all configurations with $m = n = 12$, and finding the graph for $m > n$ by swapping the colors of every edge; we choose instead to continue the process to $m = 2n = 24$ in order to corroborate the correctness of our algorithm by comparing the results for m and $2n - m$. The time necessary is not prohibitive; the full calculation took about 12.5 h on a 2012 MacBook Pro laptop.

Finally, we calculate the energies of about 2 % (but at least 100) of the remaining configurations for each value of m such that $5 \leq m \leq 19$, chosen as follows. About $\frac{1}{3}$ to $\frac{1}{2}$ of these configurations are identified as the neighbors of the highest and lowest-energy configurations known at $m-1$ and $m+1$. The other roughly half of these configurations are sampled uniformly at random from the remaining configurations. In order to maximize the information we have about neighbors of high- and low-energy configurations, we carry out this calculation sequentially for different values of m, starting with $m = 5$ and $m = 19$ and working our way to less extreme values, ending with $m = n = 12$.

For the rhombicuboctahedron, the combinatorial explosion in the number of configurations makes it intractable to even enumerate the approximately six trillion configurations. We focus instead on configurations with non-trivial symmetries, motivated by the heuristics above, and develop an algorithm to enumerate these 'symmetric' configurations, as follows.

For each non-identity element $g \in O_h$, we construct C^g, the set of colorings which are fixed by the action of g, as follows. We consider the orbits of the *edges* of the polyhedron under the subgroup $S_g = \langle g \rangle \subset O_h$ generated by g. In order for a coloring c to be fixed by g, i.e., $gc = c$, the edges in each orbit $S_g\mathfrak{e}$, $\mathfrak{e} \in \mathcal{E}$ must be of the same color. Thus, C^g is constructed as the set of two-colorings of edge orbits, $S_g\mathfrak{e}$. As every coloring with non-trivial symmetry number is fixed by at least one non-identity element, the union $\cup\{C^g \text{ s.t. } g \neq e\}$ then gives us exactly the symmetric colorings we are interested in. We then compare these colorings $/O_h$ to enumerate the configurations, that is, the equivalence classes of colorings, with $s \geq 3$ (64,431 configurations total; see Table 1). There are tens of millions of configurations with symmetry $s = 2$; we have not enumerated all of these, but have enumerated those with extreme values of m ($m \leq 8$ and $m \geq 40$), along with those at values of m relatively prime to 48 which have no higher-symmetry configurations ($m = 11, 13, 17, 19, 23, 25, 29, 31, 35, 37$).

In fact, we are able to shorten the calculation considerably by computing only a partial union over a subset of elements g; because we are interested ultimately in configurations, not colorings, we need only sure that at least one representative coloring of each equivalence class is present in the union. To select the elements g for which we compute C^g, we classify the forty-seven non-trivial elements of O_h into the following five subsets:

1. There are 20 elements of O_h with order $\geq$4. Each of these elements has a power g^a which is of order 2 or 3. Since any coloring fixed by g is also fixed by g^a, $C^g \subset C^{g^a}$, and we need only enumerate C^g for g of order 2 or 3 to get all configurations.
2. There are eight elements of O_h with order 3; while these fix different colorings, they generate the same equivalence classes in C/O_h. Thus, we need enumerate C^g for only one of these eight elements.
3. There are ten elements of O_h with order 2 which have the property that $|S_g\mathfrak{e}| = 2$ for all $\mathfrak{e} \in \mathcal{E}$. Of these, only three produce different configurations, while the remaining seven fix colorings which are equivalent to colorings already identified.
4. There are six elements of O_h with order 2 which have four orbits for which $|S_g\mathfrak{e}| = 1$, that is, these elements map four edges onto themselves. As with the

elements of order 3, these produce the same configurations, and so we enumerate C^g for only one.

5. There are three elements of O_h with order 2 which have eight orbits for which $|S_g \mathfrak{e}| = 1$, that is, these elements map eight edges onto themselves. These elements represent reflections across the plane bisecting a ring of square faces. Again, we enumerate C^g for only one of them.

Once the elements g are chosen, the computation of each C^g is automated, as is the processing of the union $\cup C^g$ to find just one representative coloring for each configuration. We then calculate energies for configurations sampled as follows. We calculate the energy for all configurations with symmetry $s \geq 6$ (1425 configurations total); about 2.5 % (but at least 40 at each m) randomly selected configurations with $s = 4$; about 2.5 % (but at least 40) randomly selected configurations with $s = 3$; and for $m \leq 8$, $m \geq 40$, or m relatively prime to 48, about 0.02 % (but at least 200) configurations with $s = 2$.[3] This gives us a total of roughly 10,000 energies calculated for each of the plots in the next section—a small fraction of the 6 trillion configurations, but a sampling which we believe captures the most interesting aspects of the energy landscape.

4 Results

In most of the examples below, we set the energy parameters $E_a = 10.0$; $E_p = 1.0$; and $E_e = E_v$ chosen to normalize the energy of the $m = 0$ configuration to 1.0, taking values $1.0 < E_e, E_v < 3.8$. In Sect. 4.5 only, we change this last condition and set $E_e \neq E_v$ as described later.

4.1 Energy Landscape at $\theta_* = \bar{\theta}_0$

In Fig. 3, we present a sample energy landscape when the average preferred angle is the same as the ideal angle, and $\theta_* = \bar{\theta}_0 = 1/2(\theta_0(A) + \theta_0(B))$ [this choice of parameters was used as a heuristic design principle in Sun et al. (2010); as discussed below the broad features of this energy landscape are insensitive to these choices of $\theta_0(A)$ and $\theta_0(B)$]. We observe that when m and n are held fixed, there is a strong variation in the energy of all isomers with chemical formula $M_n A_m B_{2n-m}$. For $m = n = |\mathcal{E}|/2$, these isomers include the configurations with minimum and maximum energy.

4.2 Extremal States Have High Symmetry

As in the example above, we typically find that the lowest-energy state is the most symmetric isomer with $m = n$. While this conclusion seems unsurprising, it must be recognized that there are also high-energy states with high-symmetry number!

[3] In Fig. 6a, we calculate the energy for configurations with $s = 3$ only for $m \leq 8$ and $m \geq 40$, and for a larger proportion of configurations with $s = 2$ for $m \leq 8$, $m \geq 40$, $m = 11, 13, 17, 31, 35, 37$.

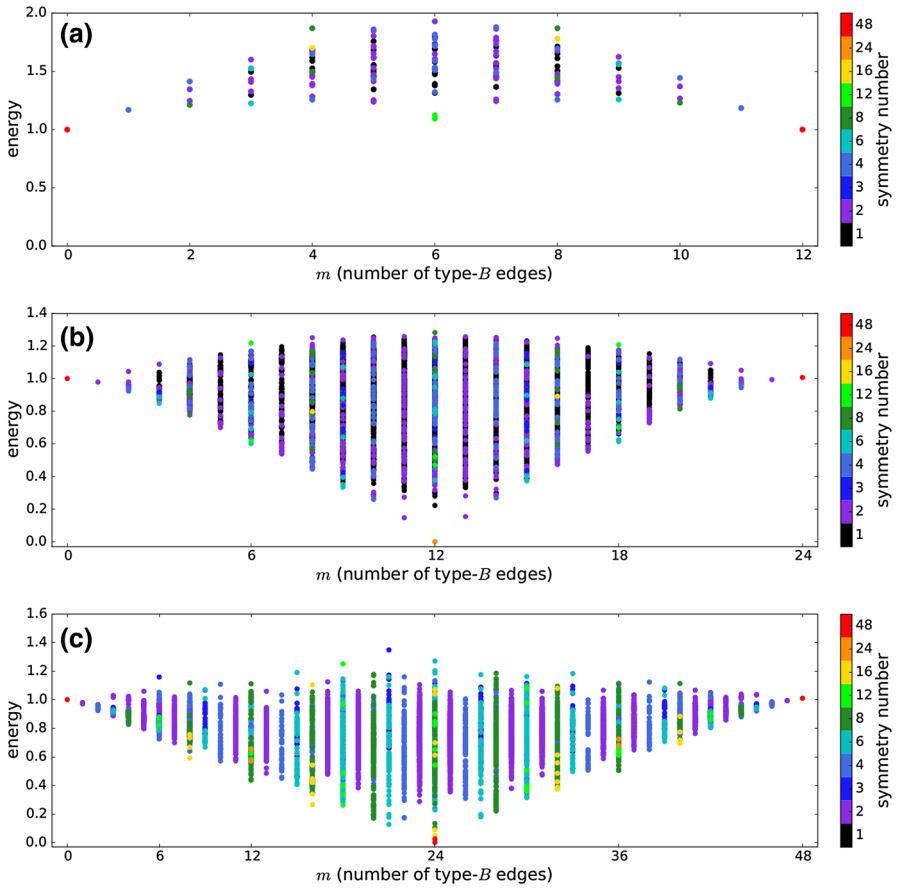

Fig. 3 Energy spectra. **a** The octahedron, **b** the cuboctahedron, and **c** the rhombicuboctahedron. The preferred angles are **a** 81° and 99° (ideal angle: 90°), **b** 105° and 129° (ideal angle: 117.2°), **c** 120.7° and 147.6° (ideal angle: 134.1°). For each polyhedron, the energies are normalized so that the configuration with $m = 0$ (that is, all edges have small preferred angles) has energy 1.0. The choice of energies calculated is described in the text. *Color* indicates the symmetry number of the configuration (Color figure online)

These ideas are illustrated in two extremal configurations for the cuboctahedron as shown in Fig. 4. Both the extremal energy configurations in Fig. 4 have reasonably large symmetry numbers, $s = 24$ and $s = 6$ and the lowest-energy configuration in Fig. 4 has an energy of less than 0.06 % of the normalized energy (and possibly zero energy). A heuristic explanation is as follows: at any vertex where two edges each of type A and B meet, the energy of the vertex is lower in configurations in which the edges of type A are adjacent to one another ("*cis* configuration") and opposite the edges of type B. This arrangement allows the energy to be reduced by canting the square plane of the ion. Configurations with many such opposed pairs tend to have high-symmetry numbers, and so the examination of highly symmetric configurations captures the low- energy states. If edges of type A are opposite one another—and in particular if four edges of the same type meet at a vertex—the vertex and the edges

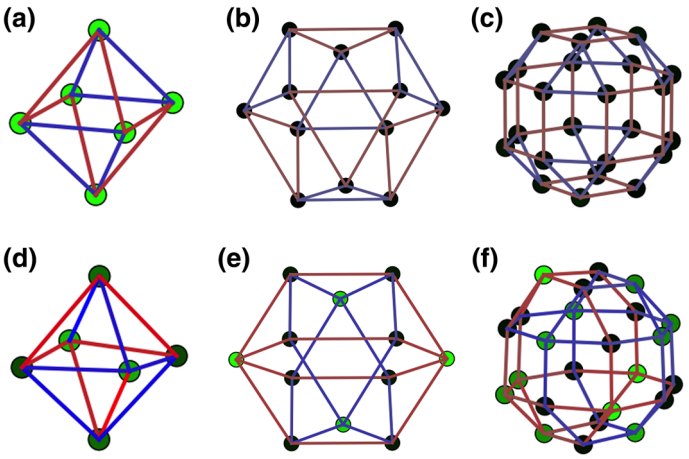

Fig. 4 Extremal energy configurations for $m = n$. The preferred angles are as in Fig. 3; blue edges are of type A with $\theta_0(A) < \theta_*$, while *red edges* are of type B with $\theta_0(B) > \theta_*$. *Brightness of the color* indicates the magnitude of the energy contribution. Note the deformation from a regular polyhedron; in particular, *blue edges* are shorter than *red edges*. **a–c** Minimum-energy configurations for the octahedron, cuboctahedron, and rhombicuboctahedron, **d–f** Maximum-energy configurations. Note the high-symmetry numbers: **a** $s = 12$, **b** $s = 24$, **c** $s = 48$, **d** $s = 2$, **e** $s = 6$, **f** $s = 6$. In the minimum-energy configurations, every vertex is adjacent to two edges of each type, and edges of different types lie opposite one another. For the cuboctahedron and rhombicuboctahedron, the maximum-energy configurations are entirely composed of high-energy 6- or 8-element closed rings which are either entirely of type A or entirely of type B; large energy contributions also come from the vertices at which four edges of type B meet (Color figure online)

have higher energy. By arranging symmetric configurations of such vertices, we obtain high-energy configurations for the whole system.

More generally, examination of any of the energy spectra suggests that the symmetric configurations tend to span or nearly span the range of energies at any value of m. This is clearest for the octahedron, where in every case, both the lowest- and highest-energy configurations at any m have symmetry number at least 2, and for $m \neq 5, 7$, the lowest-energy configuration has $s \geq 4$ and often the highest symmetry possible. Configurations with high-symmetry number of course have multiple vertices with the same arrangement, and so if these are favorable or unfavorable arrangements, the energy is correspondingly particularly low or particularly high. It is this observation which led us to emphasize symmetric configurations in sampling the energies of the cuboctahedron and the rhombicuboctahedron.

4.3 Low and High-Energy States are Connected

For the octahedron, we obtain a complete description of the configuration space as a graph, as shown in Fig. 5. We observe that low and high-energy configurations are usually connected to other low-energy and high- energy configurations, respectively. This feature of the energy landscape may be explained heuristically as follows. Starting from a low (resp. high)-energy configuration with several favorable arrangements of small-angle edges opposing large-angle edges, changing the preferred angle of a

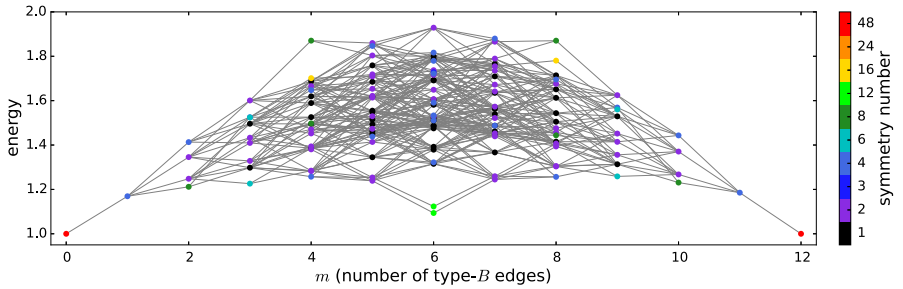

Fig. 5 Energy spectrum of the octahedron shown with connections between neighboring configurations. The range of the y-axis has been changed for clarity. Note that low-energy configurations tend to be connected to other low-energy configurations and likewise high-energy configurations to high- energy configurations

random edge will often leave most of these favorable arrangements unchanged, giving another low (resp. high)-energy configuration.

4.4 Asymmetry Around the Average Preferred Angle $\bar{\theta}_0$

We find that the shape of the energy landscape is robust as $\bar{\theta}_0$ varies. Specifically, when $\theta_0(A) < \theta_* < \theta_0(B)$, but $\theta_* \neq 1/2(\theta_0(A) + \theta_0(B))$, the energy spectrum 'tilts' as shown in Fig. 6. The lowest energy, however, is still found at even mixing, $m = n$, even though at this mixing the average angle does not match the ideal angle. This lowest energy is slightly higher than in the symmetric-angle case, but still well below the $m = 0$ and $m = |\mathcal{E}|$ configurations. We conclude that even if $\bar{\theta}_0 \neq \theta_*$, the energetically favorable arrangements of small-angle edges opposing large-angle edges allow for much of the frustration to be relieved.

In the case that the preferred angles do not encompass the ideal angle—i.e., either both $\theta_0(A), \theta_B(0) > \theta_*$ or both $\theta_0(A), \theta_B(0) < \theta_*$—the shape of the energy spectrum changes. In the case $\theta_* < \theta_0(A) < \theta_0(B)$, the lowest-energy state occurs for $m = 0$ and the energy increases roughly linearly with m; an example for the cuboctahedron is shown in Fig. 6. Isomers with a fixed m still exhibit a range of energies, but the variation withs m is similar to the m-dependent energy contribution.

4.5 Varying the Energy Ratios

There is little variation in the shape of the energy spectrum as the parameters E_{edge}, E_{vertex}, and E_{planar} (we fix $E_{\text{arm}} = 10.0 > E_{\text{edge,vertex,planar}}$) are changed. Figure 6c shows the energy spectrum for a rhombicuboctahedron with the preferred angles symmetrically encompassing the ideal angle, but with the edge energy scale E_{edge} twice the vertex energy scale E_{vertex} (E_{planar} is kept at 1.0). This is nearly indistinguishable in overall shape from Fig. 3c, with $E_{\text{edge}} = E_{\text{vertex}}$.

We also observed that the ratio between the energy of a rhombicuboctahedron with only one ligand type, $M_{24}A_{48}$, and a cuboctahedron with the same ligand type

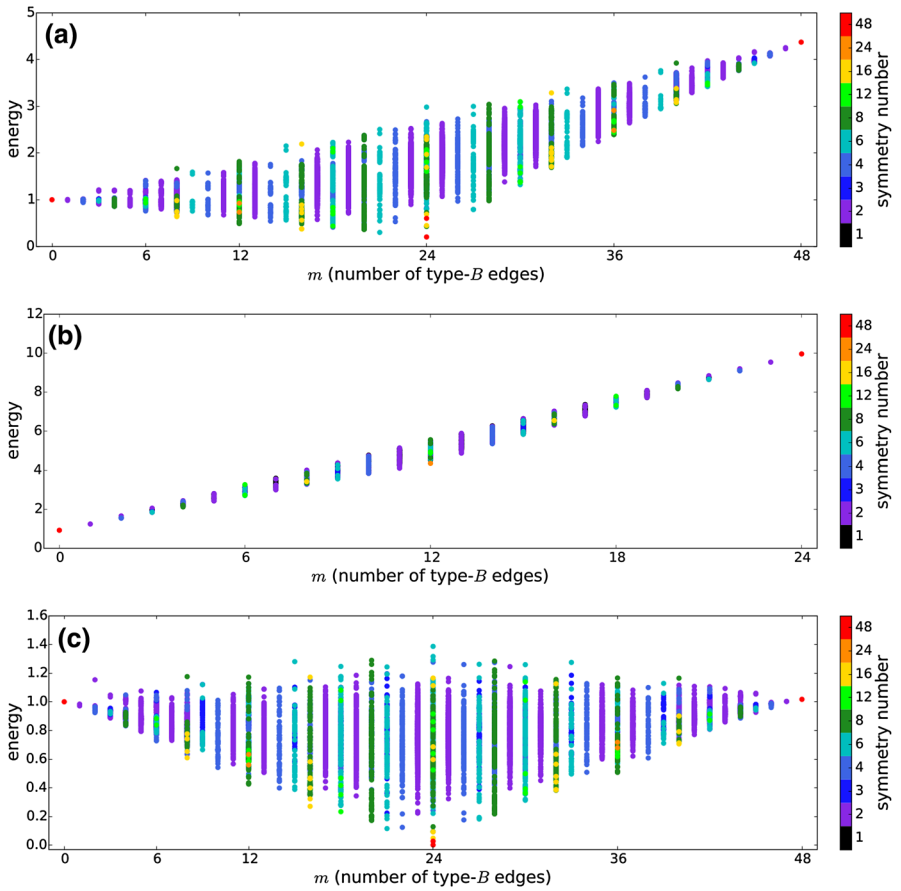

Fig. 6 Variation of the energy landscapes with model parameters. **a** Asymmetric preferred angles. Partial energy spectrum for the rhombicuboctahedron with preferred angles 127° and 149°; the average angle matches the ideal angle (134.1°) for $m \approx 16$. These angles are chosen to match those of the two organic ligands reported in Sun et al. (2010). Note that the lowest-energy configuration is still at $m = n = 24$. **b** Partial energy spectrum for the cuboctahedron with the same preferred angles as in (**a**); in this case, both preferred angles are larger than the ideal angle (117.2°). Note that the energy increases rapidly with added type-B edges. **c** Partial energy spectrum for the rhombicuboctahedron with the same preferred angles as in Fig. 3c, but with $E_e = 2E_v$. Note that the shape of the energy landscape is almost unchanged, with perhaps a very slight increase in the maximum energies. For each energy landscape, the overall energy scale is chosen so that the configuration with $m = 0$ has energy 1.0

and parameters, $M_{12}A_{24}$, is nearly constant over a much larger range of energy ratios, varying both $E_{\text{edge}}/E_{\text{vertex}}$, and E_{planar} (data not shown). Thus, the ratios of the energy parameters in this system seem to be relatively unimportant; the competition between the edge energies and vertex energies results in an embedding which balances those energies in such a way that the shape of the energy spectrum is roughly independent of the ratios; this shape is almost entirely determined by the preferred angles. The overall energy scale, of course, will be important in any system in which temperature plays a role.

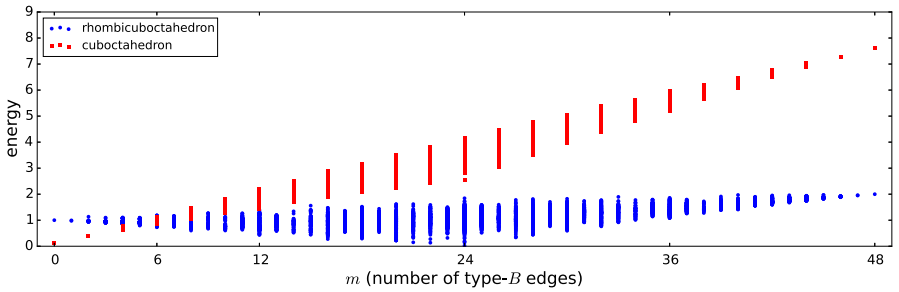

Fig. 7 Crossover from cuboctahedron to rhombicuboctahedron. Partial energy spectra of cuboctahedron and rhombicuboctahedron plotted together; *blue points* give the rhombicuboctahedron and *red* the cuboctahedron (Color figure online)

4.6 Crossover Between Polyhedra

Finally, we compare two different polyhedra to one another as shown in Fig. 7. For the cuboctahedron, we scale both the number of colored edges and the energy by two in order to compare energy per molecule. The preferred angles are 121.5° and 151.8°; both angles are larger than the ideal angle for the cuboctahedron, and the average angle matches the rhombicuboctahedral ideal angle for $m \approx 20$. Note that there is a crossover; for $m \lesssim 6$, the cuboctahedron is energetically favored, while for $m \gtrsim 6$, the rhombicuboctahedron is energetically favored.

5 Discussion

The main contributions in this work are:

1. The development of a 'minimal' mechano-combinatorial model for molecular cages inspired by experiments (Sun et al. 2010).
2. The development of a symmetry-based computational scheme to determine minimum-energy states at zero temperature.

While this work was inspired by Sun et al. (2010), it does not as yet provide an explanation for the emergence seen in experiments. The inclusion of physical effects that are ignored here, such as entropic effects and kinetics, should shed more light on the process of self-assembly. We first discuss certain interesting features of the energy landscape that our work has revealed. We then describe how our computations could be used as a basis for the inclusion of entropic effects. The computational challenge of including kinetics in a related model is described in the companion paper (Johnson and Menon 2015).

First, let us discuss some important features of the zero- temperature energy landscape. We have found a strong variation in the energy of isomers—it is not just the number of each type of edge, but their arrangement on the polyhedron that determines the energy of the system. When the preferred angles encompass the ideal angle for the polyhedron, the lowest energies are achieved by high-symmetry configurations in which edges meet at *cis* configurations at a vertex. For some choices of parameters, we

see a shift in minimum-energy states from cuboctahedron to rhombicuboctahedron. We also note a striking analogy to another experiment in Fujita's lab, in which ligands with different lengths but similar angles were mixed (Sun et al. 2014). For a sufficiently large ratio between the lengths, $r \approx 2$, they observed a mixed yet sorted structure of the form $M_{12}A_{12}B_{12}$ with a well-defined structure in which longer and shorter ligand molecules were found opposing one another. Our work shows that similarly mixed yet sorted structures are favored when ligands with different angles are mixed, although the energy differences are small enough that the sorting is not as complete as in Sun et al. (2014).

Let us now turn to a discussion of entropy. In the experiments, polyhedra with a range of m self-assemble in a single solution (Sun et al. 2010). Examination of our energy spectra show first of all that there is a gap between the lowest-energy state and the next lowest, so that at sufficiently low temperature, only the lowest-energy state should be observed. This does not take into account the determination of m from the bulk ligand concentration; however, the lower envelope of the energy spectrum is, for the parameters we have studied, concave down, meaning that for $p \neq 1/2$ the system forms as many polyhedra as possible with $m = n$, the lowest-energy state, and forms pure polyhedra when one type of ligand is in excess. That configurations with $m \neq 0, n, 2n$ are seen indicating that higher-energy states are thermally accessible due to the entropic considerations. Let us now indicate how our work can be extended to include these effects.

We distinguish between two approaches : (1) a full computational study of the free energy landscape at positive temperature; (2) a reduced model that builds on our study, but includes translational, vibrational and rotational entropy. Study (1) requires the construction of a Markov Chain Monte Carlo scheme that efficiently explores the energy landscape at finite temperature to approximate the full partition function of the system. This approach is completely different from the main thrust of our paper, and we do not pursue it. On the other hand, as in colloidal systems of rigid spheres, it is possible to focus on the entropic contribution to the relative probability of different energy minima, rather than integrating over all possible configurations. We work in analogy with the recent review (Cates and Manoharan 2015, Section3) for colloidal systems of rigid spheres. The authors contrast the relative probability of two isomers (the octahedron and tri-tetrahedron, respectively), by approximating the full partition function by a product $Z = Z_{\mathrm{T}} Z_{\mathrm{R}} Z_{\mathrm{V}}$ that accounts for the translational (T) free energy, rotational (R) free energy and vibrational (V) free energy, respectively. The translational free energy depends on the internal energy of different configurations and is the same for the two clusters. However, their rotational and vibrational energies are distinct and may be computed by classical formulas [Eqs. (3) and (4) in Cates and Manoharan (2015)] that require a computation of the vibrational frequencies of normal modes of the cluster, moments of inertia and symmetry factors of different states. These ideas have a natural extension to our system: the minimum-energy state for a fixed combinatorial type is the analog of a rigid cluster (e.g., the octahedron and tri-tetrahedron are two clusters with the same number of particles, but distinct combinatorics). Thus, the computational task of augmenting our model to (approximately) include entropy involves determining the vibrational modes and moments of inertia for each combinatorial type, by linearizing about the minimum-energy states that have

been determined in this paper. The vibrational modes may be determined numerically by computing the eigenvalues of the stiffness matrix obtained by linearizing the elastic energy at the minimizing geometry, for each choice of parameter values used here. This task is sufficiently computationally intensive that we do not pursue it here, though we hope to address it in future work.

We have also not taken into account kinetics. We have considered the configuration space as a graph by drawing connections between configurations which are related by the change of preferred angle of a single edge; however, we do not expect this to be a realistic model for the ways in which edges dynamically arrange themselves. The kinetics are crucial to establishing whether the system can reach its equilibrium distribution and on what timescale, or whether non-equilibrium states become trapped.

In conclusion, the work presented here provides insight into the importance of configuration and symmetry in the determination of the free energies of self-assembling molecular cages with a large number of isomers. It illustrates the importance of exploiting symmetry to reduce the size of otherwise combinatorially intractable problems. Finally, the zero-temperature calculations serve as the starting point for approximate calculations at positive temperatures as outlined above.

References

Caspar, D.L.D., Klug, A.: Physical principles in the construction of regular viruses. In: Chovnick, A. (ed.) Cold Spring Harbor Symposia on Quantitative Biology vol. 27, pp. 1–24. Cold Spring Harbor Laboratory Press, Long Island, NY (1962)

Cates, M.E., Manoharan, V.N.: Testing the Foundations of Classical Entropy: Colloid Experiments. arXiv:1507.04030 (2015)

Coxeter, H.S.M.: Regular Polytopes. Dover Publications, New York (1973)

Holmes-Cerfon, M., Gortler, S.J., Brenner, M.P.: A geometrical approach to computing free-energy landscapes from short-ranged potentials. Proc. Natl. Acad. Sci. **110**, E5–E14 (2013)

Johnson, D., Menon, G.: The Building Game: From Enumerative Combinatorics to Conformational Diffusion. Preprint (2015)

Kaplan, R., Klobušickỳ, J., Pandey, S., Gracias, D.H., Menon, G.: Building polyhedra by self-assembly: theory and experiment, Artif. Life **20**, (2014)

Liu, Y., Hu, C., Comotti, A., Ward, M.D.: Supramolecular Archimedean cages assembled with 72 Hydrogen bonds. Science **333**, 436–440 (2011)

Mal, P., Breiner, B., Rissanen, K., Nitschke, J.R.: White phosphorus is air-stable within a self-assembled tetrahedral capsule. Science **324**, 1697–1699 (2009)

Pandey, S., Ewing, M., Kunas, A., Nguyen, N., Gracias, D.H., Menon, G.: Algorithmic design of self-folding polyhedra. Proc. Natl. Acad. Sci. **108**, 19885–19890 (2011)

Rotman, J.J.: An Introduction to the Theory of Groups. Graduate Texts in Mathematics, vol. 148, 4th edn. Springer-Verlag, New York (1995)

Sun, Q.F., Iwasa, J., Ogawa, D., Ishido, Y., Sato, S., Ozeki, T., Sei, Y., Yamaguchi, K., Fujita, M.: Self-assembled $M_{24}L_{48}$ polyhedra and their sharp structural switch upon subtle ligand variation. Science **328**, 1144–1147 (2010)

Sun, Q.F., Sato, S., Fujita, M.: An m12(l1)12(l2)12 cantellated tetrahedron: a case study on mixed-ligand self-assembly. Angew. Chem. Int. Ed. **53**, 13510–13513 (2014)

Takeda, N., Umemoto, K., Yamaguchi, K., Fujita, M.: A nanometre-sized hexahedral coordination capsule assembled from 24 components. Nature **398**, 794–796 (1999)