Size limits of self-assembled colloidal structures made using specific interactions

Classification: PHYSICAL SCIENCES, Physics

Size limits of self-assembled colloidal structures made using specific interactions

Zorana Zeravcic[1,2], Vinothan N. Manoharan[1,3] and Michael P. Brenner[1,2]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA
[2]Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, Massachusetts 02138, USA
[3]Department of Physics, Harvard University, 17 Oxford St. Cambridge, Massachusetts 02138, USA

Abstract

We establish size limitations for assembling structures of controlled size and shape out of colloidal particles with short ranged interactions. Through simulations we show that structures with highly variable shapes made out of dozens of particles can form with high yield, as long as each particle in the structure binds only to the particles in their local environment. To understand this, we identify the excited states that compete with the ground state structure, and demonstrate that these excited states have a completely topological characterization, valid when the interparticle interactions are short ranged. This allows complete enumeration of the energy landscape, and gives bounds on how large a colloidal structure can assemble with high yield. For large structures, the yield can be significant even with hundreds of particles.

Significance Statement

Nature uses hierarchical assembly to make complex structures, such as biomolecules, virus shells and microtubules, with high fidelity. Today a key challenge is to translate this process to artificial systems, which hinges on understanding the fundamental questions of efficiency and scalability of self-assembly. Although self-assembly has been studied for decades, the principles behind it and its fundamental and practical limits are still largely unknown. In this paper we establish size limitations for assembling structures of controlled size and shape out of colloidal particles with specific interactions. Inspired by simulations of structures with highly variable shapes and sizes, we develop an understanding of yield through a general theory of excited states that compete with the desired structure in assembly.

Size limits of self-assembled colloidal structures made using specific interactions

Zorana Zeravcic [*] [†], **Vinothan N. Manoharan** [*] [‡], **and Michael P. Brenner** [*] [†]

[*]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA,[†]Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, Massachusetts 02138, USA, and [‡]Department of Physics, Harvard University, 17 Oxford St. Cambridge, Massachusetts 02138, USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

We establish size limitations for assembling structures of controlled size and shape out of colloidal particles with short ranged interactions. Through simulations we show that structures with highly variable shapes made out of dozens of particles can form with high yield, as long as each particle binds only to the particles in their local environment. To understand this, we identify the excited states that compete with the ground state structure, and demonstrate that these excited states have a completely topological characterization, valid when the interparticle interactions are short ranged. This allows complete enumeration of the energy landscape, and gives bounds on how large a colloidal structure can assemble with high yield. The yield can be significant even with hundreds of particles.

inverse problem | scalability | local minima

Introduction

Nature uses hierarchical assembly of complicated building blocks to make highly functional structures like biomolecules, virus shells and microtubules, without any external influence and with high fidelity. Mimicking this would not only give more insight into biological mechanisms but would also help realize the dream of "bottom-up" assembly that has been a central theme of nanotechnology for many decades [1].

Like in biology, the information needed for assembling arbitrary macroscopic structures can be stored in the building blocks through the design of their interactions and interaction rules. Over the years great advances have been made by synthesizing new building blocks differing in geometry, composition and interactions [2, 3, 4, 5, 6, 7, 8, 9, 10], allowing for study of more complex objects. However, basic rules necessary for robust and efficient assembly of a desired structure in a scalable fashion and reasonable time scales are still not understood. A number of schemes for approaching this "inverse" statistical mechanics problem have been proposed [11, 12, 13], but a general framework and systematic studies are still missing. One of the essential underlying questions, having both practical and conceptual impact, is whether any desired macroscopic structure can be assembled with a high yield, out of a given set of building blocks. Or are there fundamental constraints limiting the structures that can be effectively built?

In this paper we address these general questions using the model system of DNA-coated particles, itself of considerable recent interest. We consider an isolated system of N spherical colloidal particles each of which is isotropically coated with DNA strands to control interparticle interactions. At the colloidal scale, such interactions have a range that is much shorter than the size of the particles. The use of DNA labeling to control binding specificity was originally pioneered for assembling nanoparticles [14, 15, 16, 17] into infinite crystals [18, 19, 20, 21, 22, 23, 24], where recently it was demonstrated that with two species with differing particle radii and DNA linker length a zoo of different crystal morphologies can be created [25]. Work at the colloidal scale has begun to bear fruit [26, 27, 17, 28, 29, 30, 13]. However the set of possible structures that could be coded is far more general, including structures of any shape and size, both rigid and flexible. For example, the number of clusters that can be assembled out of spherical particles with fixed size increases dramatically with particle number N, so that with only 10 particles there are 223 topologically distinct structures with at least $3N-6=24$ contacts [31, 32, 33].

Designing arbitrary complex structures requires using the specificity of interactions to make the desired target the energetic ground state. The most robust way of doing this is to make every particle in the target structure different, with interparticle interactions chosen to favor the desired local configuration in its target structure. The interactions between different particles are coded into an *interaction matrix* $\hat{I}$, specifying the interaction energy between every pair of particles.

We begin by asking how high the equilibrium yield can be when $\hat{I}$ is coded for an arbitrary large structure, using the simplest prescription in which every contact of the desired structure binds specifically, and every undesired contact does not bind. In this paper, the yield represents the probability of successful complete assembly of exactly N particles, in contrast to a common definition of yield as the percentage of particles from the bulk that assemble into copies of a desired structure. Numerical simulations using dissipative particle dynamics demonstrate that there is a temperature regime where high yield ($> 50\%$) assembly is possible, for a range of complex structures consisting of dozens of particles. This is striking, because as the number of particles grows, the number of competing states grows rapidly with N; such a high yield implies that these states are less competitive than naively expected. To understand why this is the case, we study the yield of an entire family of structures, the set of rigid clusters with $N \leq 9$ particles. We design $\hat{I}$ so each structure is the ground state, and numerically calculate the corresponding yield curves. For clusters, the yield degrades quickly with increasing N. However, we use the obtained insights to develop a complete description of the low-energy excited states that compete with the ground state, valid for asymptotically large structures. This description explains the high observed yield for large structures and points to the limits of equilibrium self-assembly with colloidal particles with short ranged interactions. Finally, we comment on the role and importance of kinetic effects.

Reserved for Publication Footnotes

Results

Designing Interactions. We study the assembly yield of arbitrary structures by choosing the interaction energy so that the desired structure is the ground state. This can be done uniquely for an isolated system of N spherical particles with isotropic interactions as follows: start with the adjacency matrix $\hat{A}$, which is the $N \times N$ matrix having an element $A_{ij} = 1$ if particles i and j are in contact and $A_{ij} = 0$ otherwise. We choose $\hat{I}$ directly from $\hat{A}$, by mapping non-zero elements of $\hat{A}$ to favorable interactions in $\hat{I}$ and zero elements to unfavorable interactions. Every contact in the desired structure has a bond energy $-\epsilon$ (favorable), while every other interaction has a higher energy ϵ (unfavorable). Setting all favorable interactions to have the same strength and setting all unfavorable interactions to have the same strength has been shown to optimize the equilibrium yield [34]. With this interaction matrix the different interactions between different particles are maximally specific. If some particles have identical sets of neighbors, their interactions are indistinguishable, so these particles are effectively of the same type. When the interaction matrix is reduced to show the interactions between the different particle types, it is called an *alphabet* [34], with the maximally specific interactions defining the *maximal alphabet*.

When a structure has a unique adjacency matrix, this procedure guarantees that the desired structure has the maximal number of contacts, and is therefore the unique ground state. But if a structure has no mirror symmetries, then its "chiral partner", obtained as the object's mirror reflection through an arbitrary mirror plane, cannot be made to coincide with the original object through proper rotations or translations. The chiral partners are therefore distinct assemblies of particles, though each particle shares the same neighbors in both (and therefore the chiral partners have the same $\hat{A}$). When a structure is built out of different types of particles, it generically has no mirror symmetries, even if the geometrical shape of the structure does.

Consequently, both chiral partners are ground states, and in this paper we identify both as being the desired structure. For equilibrium yield, this difference is not consequential, but we will see at the end of the paper that the simultaneous assembly of both chiral partners can lead to kinetic effects relevant for the yield.

The Assembly of Large Structures. To discover whether it is possible to assemble large structures with high yield, we use dissipative particle dynamics (DPD) [35, 36], and measure the equilibrium yield as a function of temperature. Our simulation contains N colloidal spheres of diameter D, with an interaction range of $1.05D$ (this range corresponds roughly to that of a DNA-coated $1\mu m$ particles [37]). The colloids are immersed into a DPD solvent of smaller particles. Colloids are modeled as $48-96$ Lennard-Jones spheres if they interact favorably, and with the repulsive part of the Lennard-Jones potential if they interact unfavorably. Simulations are run for a range of temperatures with a volume fraction of colloids $\phi_{coll} = 1/30$, and a larger volume fraction of solvent $\phi_{sol} \approx 0.2$. More details are given in the SI Text.

The complex structures include (i) a bipyramid with 44 particles; (ii) a bipyramid of 19 particles; (iii) a 19-particle chiral chain structure, and (iv) a 69-particle replica of Big Ben, with a crystalline base and a pyramidal top. Fig. 1a shows snapshots in the time evolution of the Big Ben assembly, starting with randomly distributed and thermalized particles. Assembly into the desired structure occurs with high yield, and this result prevails in most of the complex structures we have studied. Fig. 1b plots yield as a function of temperature T/ϵ for the four previously mentioned structures. Each data point is an ensemble average over ~ 100 different initial realizations, run at a fixed temperature T for a fixed time t_{run}. The yield is defined as the fraction of runs in the ensemble for which all of the bonds in the complete structure are observed at least once within a short time window ending at time t_{run}. This is still a conservative definition of yield, since if the structure with all the bonds is not observed in the time window, even due to a single particle bond missing, the structure is regarded as a failed assembly. (See SI Text for more details.)

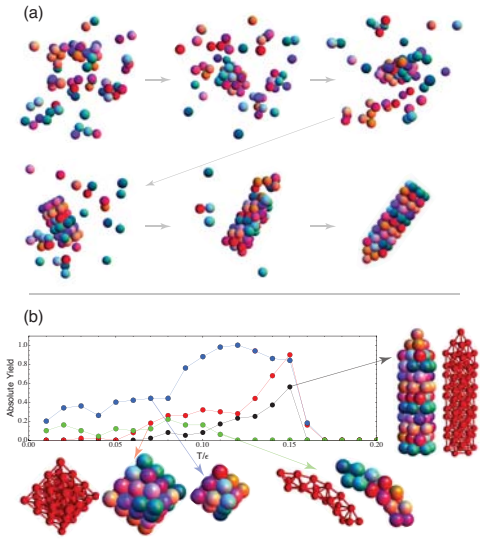

Fig. 1. (a) Snapshots in time of DPD simulation of the Big Ben assembly. (b) Absolute yield as a function of temperature T/ϵ for four larger structures described in the main text. Each data point is an ensemble average over 100 different initial condition simulations.

Our simulations exhibit several regimes as a function of temperature T, with a glassy regime at low T and an equilibrium regime at high T (see SI Text). At the highest temperatures (at $T/\epsilon \gtrsim 0.16$) the system is in equilibrium, but the bonds between the colloids are short lived, leading to small absolute yields of the ground states. The most striking fea-

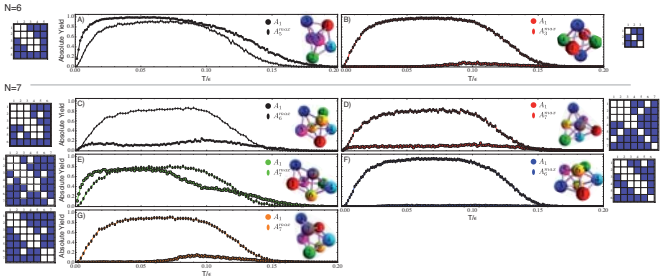

Fig. 2. Absolute yield as a function of temperature T/ϵ for maximal alphabet vs. identical particles, for all the clusters with $N = 6$ (panels A-B) and $N = 7$ particles (panels C-G). Matrices next to the panels are the maximal alphabet interaction matrices. When particles in a cluster have identical sets of neighbors, they are effectively of the same type, making the alphabet smaller than N (panels A, B, C and F). In general, the yield curves of the clusters with designed interactions outperform the ones with identical interactions. The only exception is the $N = 6$ polytetrahedral cluster shown in A. When all the particles are identical, this cluster appears $\sim 96\%$ of the times, competing with the highly symmetric octahedron (panel B), and without any kinetic traps. Although introduction of specific interactions eliminates competition with the second ground state, it also introduces multiple kinetic traps that impact the yield in most of the temperature range.

ture of the yield curves in Fig. 1b is that the maximum yield is so high, despite the large number of particles in the desired state; this implies that the number of equilibria that are competing with the ground state is relatively small.

Clusters. To uncover the landscape of equilibria that compete with the ground state, we examine a simpler problem, the assembly of small clusters of particles with at least $3N-6$ contacts. Here the complete set of structures is known for $N \leq 11$ particles [32, 33]. Clusters of identical particles have a degenerate ground state when $6 \leq N \leq 9$ [38, 31], with yields predominantly determined by the rotational entropy, suppressing highly symmetric clusters. Using the prescription for particle interactions ($\hat{I}$) described above, we simulate the yield as a function of temperature for each cluster with $N \leq 9$. Fig. plots yield as a function of temperature T/ϵ for all the ground state clusters with $N = 6$ and 7 particles. As above, each data point is an ensemble average over different initial condition realizations, run at a fixed temperature T for a fixed time t_{run} (see SI Text). Using the positions of particles we form an adjacency matrix and use its eigenvalues to uniquely identify the assembled structure at t_{run}.

The panels of Fig. compare yield curves of given clusters for identical particles, with those when interactions are determined by $\hat{I}$. The yield improvements are dramatic, with the most enhancement occurring for symmetrical clusters, where the rotational entropy penalty is lifted (see panels B, F and G of Fig.). The yield curves from the clusters simulation exhibit the same phenomenology as those in Fig. 1. By comparing time and ensemble averages, we show that the equilibrium regime extends down to $T/\epsilon \sim 0.1$. Below $T/\epsilon \sim 0.1$, the relaxation time of clusters becomes comparable to t_{run} and the results are strongly influenced by kinetic effects. See SI Text and SI Figs. S1 and S2 for more details.

Fig. a shows how the maximum equilibrium yield Y_{max} (SI Text) depends on N, for maximal alphabets of clusters with $6 \leq N \leq 9$ and 26 out of 223 rigid clusters with $N = 10$. Fig. a also includes all non-maximal alphabets (SI Text) for $N = 6, 7, 8$. These are alphabets that uniquely encode for a given cluster as the ground state, but have a smaller number of different particle types. The maximal alphabets give the highest yield, as previously predicted [34]. The maximum yield monotonically decreases with growing N. For each N, the yield is determined by the geometry of the clusters: Fig. b shows that the yield of clusters increases with decreasing second moment, i.e., with increasing symmetry.

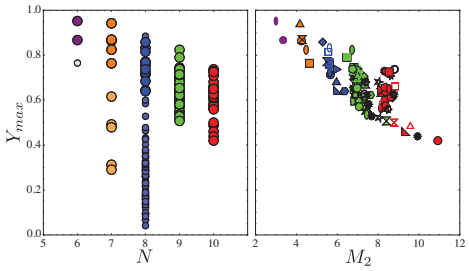

Fig. 3. A) Maximal equilibrium yield Y_{max} extracted from simulations (see SI Text) vs. the cluster size N, for all the alphabets of all the $N = 6, 7$ and 8 clusters, for maximal alphabets of all the $N = 9$ clusters, and for maximal alphabets of a subset of the $N = 10$ clusters. Big data points correspond to maximal alphabets, and small to all the other alphabet sizes. In general, maximal alphabets give the biggest yield. B) Maximal equilibrium yield Y_{max} extracted from simulations vs. the anisotropy of the cluster measured by the second moment M_2 of the corresponding cluster. $M_2 = \sum_{i=1}^{N} |\mathbf{r}_i|^2$, where $\mathbf{r}_i$ is the position of i-th particle with respect to center of mass. Only maximal alphabets are included. Color corresponds to the number of particles in a cluster (purple — $N = 6$, orange — $N = 7$, blue — $N = 8$, green — $N = 9$ and red — $N = 10$).

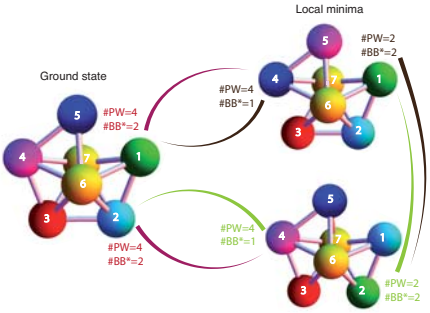

Fig. 4. Energy landscape for a $N = 7$ cluster designed using maximal alphabet. Only the lowest energy LMs are shown, both missing one bond compared to the ground state. Both can be obtained by permuting two particles in the ground state. $\#BB^*$ is the minimal number of bonds that need to be broken for transition between different states. $\#PW$ is the number of distinct pathways by which the transition can be achieved. For example, to transition from the ground state to the top local minimum one needs to break at least two bonds. One of the four pathways is to break bonds between particle pairs red — purple and front-yellow — purple, and then smoothly exchange positions of purple and blue particles before reconnecting the purple with the front-yellow.

What determines the equilibrium yield?. Consider N particles with a fixed alphabet that determines the ground state cluster C. Vibrationally and rotationally excited states of C preserve the cluster's structure, without breaking interparticle bonds. This means that our simulations would identify these states as C too. Hence, the partition function that describes the ground state is $Z_C = \frac{1}{\sigma_C} Z_C^0 Z_C^{vib} Z_C^{rot} \equiv \frac{1}{\sigma_C} Z_C^0 e^{S_0/k_B}$, where σ is the symmetry number and $Z_C^{(0)}$ the partition function given by the potential energy of the geometrical configuration C. Z_C^{vib} and Z_C^{rot} are the vibrational and rotational partition functions, respectively; these are both entropic, with S_0 the corresponding total entropy.

The states that compete for yield with the cluster C are the low-energy excited states. In particular, a local minimum (LM) state is a stable configuration of N particles and must

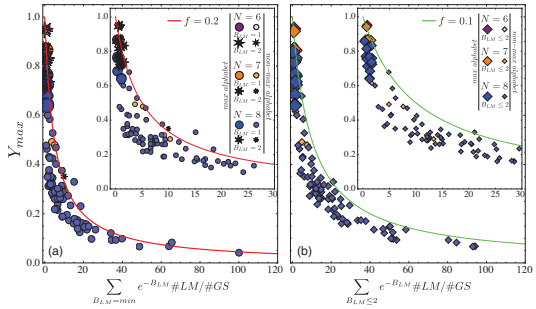

Fig. 5. Maximal equilibrium yield Y_{max} extracted from simulations (see SI Text) vs. the theoretical detrimental contribution of low-energy local minima (LMs), which depends on the number ($\#LM$) and energy (B_{LM}) of the LMs, see Eq. [1]. The factor $\#GS$ accounts for the chirality of ground states by taking values 1 (non-chiral designed cluster) or 2 (chiral pair). All alphabets of all $N = 6, 7, 8$ clusters are shown, with large symbols denoting maximal alphabets. Lines present the upper bound on yield from Eq. [1], which has the LM entropy, assumed equal for all LMs, as single variable parameter f chosen freely here, and we set $\beta\epsilon \equiv 1$ to represent the relevant equilibrium regime. (a) Only the lowest LMs are retained in the calculation of the detrimental contribution, having minimal number of broken bonds B_{LM} that takes values 1 or 2 depending on the cluster and alphabet. Inset is a zoom-in of the main panel. (b) All low lying LMs up to $B_{LM} = 2$ are retained. Inset is a zoom-in of the main panel.

have at least one particle bond less than C. Each LM is characterized by the number of broken bonds compared to C, B_{LM}, each bond costing an energy ϵ. As an example, Fig. shows the energy landscape with the two lowest energy local minima that arise for the maximal alphabet of one of the $N = 7$ clusters. Each of these local minima have $B_{LM} = 1$. Kinetic landscapes of this type for a few of clusters with $N = 6$ and 7 show that both the number of LM and B_{LM} is quite variable between different cluster geometries (SI Figs. S3-S5).

The partition function of the j^{th} local minimum is given by $Z^j_{LM} = \frac{1}{\sigma_{LM}} Z^{0,j}_{LM} e^{S_j/k_B}$, consisting of energetic and entropic parts, S_j being the entropy. For floppy structures, the entropy includes the freedom to explore the entire set of motions consistent with the imposed bond constraints. These entropies can be calculated asymptotically in the limit of vanishing interaction range and identical particles: In this limit they are roughly proportional to the number of missing bonds and depend on the geometry [39].

With a complete enumeration of the set of local minima, the equilibrium yield of the ground state cluster is given by $Y^{eq}_C = \frac{Z_C}{Z_C + \sum_j Z^j_{LM}}$. To go further, we make the simplifying approximation that each of the local minima with the same number of bonds broken has the same entropy, so that $Z^j_{LM}/Z_C \approx \frac{\sigma_C}{\sigma_{LM}} \cdot e^{-B_j \beta\epsilon} \cdot f(B_j)$, with $\beta = (k_B T)^{-1}$ and where $f(m) \equiv \exp\left[(S_m - S_0)/k_B\right]$ accounts for the entropic free energy lost from breaking m bonds. The partition function then becomes

$$Y^{eq}_C = \frac{1}{1 + \sum_m f(m) N_m e^{-\beta m \epsilon}}, \qquad [1]$$

where N_m is the number of local minima with the number of broken bonds $B_j = m$, and we set the σ factors to one temporarily for simplicity of presentation. The maximum yield is determined by the balance between N_m and the exponential penalty of higher m. The dependence of N_m on m is a purely geometrical problem, since the landscape of local minima depends only on the geometry of the structure being assembled.

Note that when a designed cluster C has two chiralities, we identify both as desired ground states, thereby doubling Z_C, i.e., introducing factor 1/2 in the sum over m.

Clusters. For the clusters, we determine N_m by completely enumerating the local minima for any given alphabet and cluster: We consider all possible arrangements of the particle labels of the given alphabet on a complete list of clusters having the same number of particles as the given cluster, and carefully remove any duplicates. With these local minima we can check the correlation between the maximum yield measured in simulations (see SI for the definition) and the number and type of local minima. Fig. a plots the maximum yield of all alphabets of all $N = 6, 7, 8$ clusters, as a function of the number of lowest lying local minima; for the clusters the minimum number of bonds broken equals 1 or 2. The yield correlates strongly with the number of lowest energy LMs. Fig. b also considers the LMs with one additional broken bond; the correlation improves only slightly, implying that the yield of these small clusters is determined by the competition between the ground state and the lowest lying local minima.

In these plots we also show the prediction from Eq.[1], with the symmetry number factors reinstated. We used the entropic free energy loss $f(m) \equiv const$ (see SI Text) as the only free parameter, obtaining a good agreement with data. The value of $\beta\epsilon$ in the curves is set to one, representing the regime of equilibrium with fluctuating bonds, appropriate for the simulation temperatures which gave the maximum yields.

Large structures. Superficially, the observations of high yields for large structures (Fig. 1) are even more curious in light of the cluster results, since the extrapolated yield from Fig. a would correspond to a negligible yield for Fig. 1b. This discrepancy suggests that the dominant low-energy local minima change as N increases. To understand this, we enumerate the local minima for large structures, and determine which local minima are competing with the ground state. Within our prescription for specifying $\hat{I}$, the low-energy local minima have a simple mathematical structure. They are obtained by permuting particles in the structure that share at least one neighbor. Using such permutations, we enumerate the LMs for each of the three structures shown in Fig. 1b (see SI). Fig. shows the rapid increase of number of LMs with growing number of broken bonds m in the large structures. We then compute the predicted yield from Eq. [1] by including only LMs with $m \leq M$, and see that the yield quickly converges as M is increased (we fix $\beta\epsilon \equiv 1$, while $f(m) \equiv 1$, see SI Text). This shows that yield is dominated by low lying LMs, including some with more than the minimal number of broken bonds. A special case is the chiral chain, which only has $m = 1, 2$ LMs due to its quasi-one-dimensional shape.

Local Defects and asymptotic yield. For each of the large structure examples above, the relevant low lying LMs feature permutations of only nearest neighbor particles. We note that in case of small clusters, all LMs are obtained by permutations of two particles (see SI Text and SI Fig. S6). Considering an arbitrary structure, the interactions given by $\hat{I}$ imply that the permutation of far away particles i and j would break all their bonds with the rest of structure.

With these observations, we define a *local defect* as a permutation of two particles i, j that are in contact or share at least one neighbor. The energy of such a defect is determined by the local environment of particles i, j: The number of broken bonds is

$$b_{\text{local defect}} = \#NN(i) + \#NN(j) - 2 \cdot \#NN(i,j)\ , \qquad [2]$$

where $\#NN(i)$ is number of nearest neighbors to particle i, and $\#NN(i,j)$ the number of nearest neighbors shared by i and j, including the bond between i and j.

When both i, j are positioned deeply inside the bulk of the structure, we will call it a *bulk defect*. Bulk defects tend to have high energies as there are many nearest neighbors in the bulk. *Surface defects* correspond to either or both of i, j on the surface of the structure; these typically have fewer broken bonds and lower energy (see SI Fig. S7). Continuing the classification, the structure might have ridges and sharp apexes, leading to *line* and *point* defects, respectively. Any low-energy local minimum is obtained as a configuration of a particular set of local defects. We neglect configurations where defects overlap, because as N grows the number of such configurations is negligible compared to the number of configurations

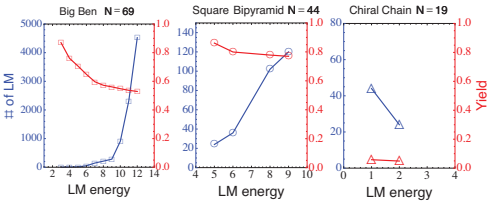

Fig. 6. Equilibrium yield and number of local minima for three large structures. The number of local minima (blue) for a fixed LM energy (which equals the number of broken bonds) is calculated by enumerating all possible LMs obtained by permuting nearby particles in the designed structure and counting the number of broken bonds. The equilibrium yield (red) is calculated using Eq. [1], where we include all LMs up to a fixed LM energy on the horizontal axis. In the equation we put for simplicity the entropy factors $f(m) \equiv 1$ (see SI Text), and choose $\beta\epsilon \equiv 1$ to represent the equilibrium regime relevant for the simulations throughout this paper.

with well-separated defect locations. In this limit, the energy of a configuration of defects is just the sum of the defects' individual energies.

We can now use these ideas to understand how the yield depends on the size of a structure. As illustrations, we consider two particular examples, a cube structure, and a chain structure.

Cube. We consider a cube with face-centered-cubic arrangement of particles, and with sides of $L = 10$ particles. A bulk local defect costs $b_3 = 16$ bonds according to Eq. [**2**]. There are $N_3 \simeq L^3 = 1000$ positions in the cube where such a defect can be located. Analogously, a surface, line and point defects cost $b_2 = 10$, $b_1 = 6$, $b_0 = 3$ bonds, having $N_2 \simeq 6L^2 = 600$, $N_1 \simeq 12L = 120$, $N_0 \simeq 8$ positions available in the cube structure, respectively. (We simplify by assuming there is only a single version of every defect type, e.g., all possible surface defects have the same value b_2, etc.) Using this input we can find the number of LMs contributing to each term in Eq. [**1**]: for example LMs with $B_{LM} = 9$ can be created out of three point defects, each with $b_0 = 3$. There are $\binom{N_0}{3} = 56$ such LMs, since the point defects can occur at any three out of eight corners of the cube; also, such a $B_{LM} = 9$ LM could consist of one point and one line defect, and there are $\binom{N_0}{1}\binom{N_1}{1} = 960$ such LMs. In total there are 1016 LMs, giving a contribution to the denominator of Eq [**1**] of $1016 \cdot e^{-9} = 0.13$ (as explained above, we fix $f(m) \equiv 1$ and $\beta\epsilon \equiv 1$). This lowers the yield to 72%. Simple counting shows that LMs with one or two point or line defects dominate and the total yield is about 50%. This is a surprisingly large yield for a 1000 particle structure.

As structure size increases, the yield decreases. The same argument as above implies that if we consider a much larger cube, e.g., with $L = 90$ and $\sim 10^6$ particles, we find a yield of $\sim 1\%$, dominated by surface and line defects. Although both cube examples have large $N \gg 1$, the yield varies between a high value and negligible value.

Linear Chain. For the example of a linear chain with length $L = 20$, we would typically have $b_0 = 1$, $N_0 = 2$ for (end-)point defect, and $b_1 = 2$ for a line defect having $N_1 \simeq L = 20$ available positions. Two line defects give the biggest contribution $\binom{N_1}{2}e^{-2\cdot b_1}$ which lowers yield to 22%, while adding contributions from LMs with one or three line defects, with or without one point defect, is enough to converge close to the final, relatively low yield of 5%. This is in accord with our above simulation results showing low chain yield.

Arbitrary structure. Finally we consider an arbitrary large structure that has a large volume-to-surface, surface-to-edge and edge-to-corner ratios, and abundant contacts so that it is rigid. We call this a 'bulky' three-dimensional structure. Then we can roughly estimate the input data and use the theory to qualitatively distinguish outcomes of considerable yield and negligible yield. Consider a structure of linear dimension L (measured in particles), which has point, line, surface, bulk defects labeled by the corresponding dimensionality $d = 0, 1, 2, 3$ respectively, each defect type costing b_d broken bonds and, according to spatial dimensionality, having $N_d \simeq L^d$ possible locations.

Consider contributions to the yield from LMs consisting of x_d defects of type d. For simplicity we do not consider LMs that contain different types of defects at the same time. First we demand that the yield stays considerable, i.e., the LM contribution to denominator of Eq. [**1**] stays much smaller than unity, i.e., $\binom{N_d}{x_d} \cdot \exp(-b_d) \ll 1$. This demand leads to

$$\exp(b_d) \gg L^d, \qquad [3]$$

where we assumed that the defects are dilute, $x_d \ll L^d$, while system is large $L^d \gg 1$; with these assumptions the value of x_d drops out. Intuitively, the inequality says that yield stays considerable if the defect cost b_d is high compared to available system size L^d, for every defect type d. The high cost of defects is consistent with our assumption of their diluteness in the dominant LMs.

If however the defects are energetically cheap compared to system size, i.e., the condition in Eq. [**3**] is violated, the yield is significantly diminished due to LM contributions. In this case of energetically cheap defects, it can also happen that the defect number in relevant LMs becomes large, violating the diluteness assumption. (Notice that this regime can never happen with point defects $d = 0$.) Due to high density of defects our basic approximation of LMs as configurations of non-interacting defects fails, but clearly we can conclude that the yield is negligible.

It is clear that this analysis can be applied to 'non-bulky' structures too, since we focused on each defect type separately. For instance, the above example of a chain is representative of quasi-one-dimensional structures and is a special case of the arbitrary structure where surface and bulk defects are absent. Additionally, one can consider structures for which the scaling of defect numbers with structure dimension is not trivial; for example, in a planar fractal-like structure the edge-length and the number of line-defects would scale as a non-integer power of linear system size. Overall our analysis gives a rough but general and simple understanding of equilibrium yield limitations based on local defects.

Chirality and Kinetic Effects. Our above theoretical estimates give yields in reasonable quantitative agreement with simulations, in which we indeed observe that LM configurations occur in the assembly process. However, simulations also reveal that some suppression of yield is due to chirality.

In accordance to our analysis of chirality, when separate parts of the assembling structure nucleate independently, they randomly acquire one of the two chiralities, leading to an inherent kinetic effect: Formed pieces with opposite chirality can never properly join into the structure. Instead, the pieces can weakly connect, e.g., along one of their edges (see SI Fig. S8). The detrimental contribution to yield from these effects could dominate the contribution from low lying LMs, and further analysis of such kinetic effects should be valuable.

Discussion

To summarize, we have demonstrated through numerical simulations that high yield aggregates of coated colloidal spheres can be created with specific, short ranged, interactions. Strikingly, our simulations indicate that high yield structures form with dozens of particles. We developed a theoretical framework for understanding this result, based on the fact that the low-energy local minima competing with the designed ground state consist of configurations in which particles in the ground state structure swap places. For example, in 'bulky' (as defined above) structures of hundreds of particles it is the surface defects which are most detrimental to yield. The scalings implied by these calculations indicate that high yield bulky structures can form from $N \sim 1000$ particles with specific interactions. This represents a fundamental limit for the complexity of structures that can be robustly built out of purely equilibrium interparticle interactions.

Our focus on maximally specific interactions not only enables the local defect analysis, but also prevents transitioning between different structures without breaking any bonds. Still, in non-rigid structures, global floppy modes (which do

not change the bond network) could influence the yield, and we leave this question for future study.

We note that there are technological challenges with implementing the high yield DNA coated colloid schemes outlined here: Our maximal interaction specificity construction requires a different DNA strand to mediate the interaction for every contact, e.g. since a particle in a bulk crystal has of order 6 nearest neighbors, that many different types of strands per particle are required. Although the practical limit of how many different types of strands per particle can be used is much higher than what we require [40], the density of strands is not high enough yet to avoid kinetic effects [40]. Nonetheless, it is possible to implement the basic schemes outlined here with non-maximal alphabets, in which the number of different strands on each particle is less than the number of contacts. If carefully chosen, a non-maximal alphabet uniquely identifies a target structure — though having more low-energy excited states, leading to a smaller yield. We have included such non-maximal alphabets in our simulations of clusters (Fig.), and the yields can still be significantly higher than with non-specific interactions.

There are other opportunities to further increase yield by removing the assumption of equilibrium interactions, which was the basis of our analysis. For example, recent work [41] has shown that in a system with a fixed number of building blocks, kinetic effects can be critical for achieving successful assembly. One example of non-equilibrium design that is natural for colloidal assembly is to allow some of the bonds to be irreversible. Any irreversible bond that does not limit pathways out of local minima will increase the yield of the ground state. The assembly of complex systems in biology suggests other ways of beating the equilibrium threshold, including (i) the possibility of using error correction, by allowing energy consuming reactions to bias toward the correctly formed structure; and (ii) including allosteric interactions, in which the binding energy of a particle depends on the set of particles that it binds to. Determining how best to implement these schemes with colloid mediated DNA interactions is an important topic for future research.

Materials and Methods

A detailed description of our simulations together with one simulation movie (Movies S1) is included in the supplementary information.

ACKNOWLEDGMENTS. The first observation of a local minimum in a colloidal system was made by Jesse Collins in an experimental study of octahedral clusters, and we thank him for sharing his results. We also thank Sahand Hormoz and Natalie Arkus for early discussions and Yu Qin for running some of the computations. This research was funded by the George F. Carrier Fellowship, the National Science Foundation through the Harvard Materials Research Science and Engineering Center (DMR-0820484), the Division of Mathematical Sciences (DMS-0907985), and by Grant RFP-12-04 from the Foundational Questions in Evolutionary Biology Fund. M.P.B. is an investigator of the Simons Foundation. V.N.M. acknowledges support from an Alfred P. Sloan research fellowship.

1. Whitesides, G. M & Grzybowski, B. (2002) Self-assembly at all scales. Science 295:2418–2421.
2. Whitesides, G, Mathias, J, & Seto, C. (1991) Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. Science 254:1312–1319.
3. Grzybowski, B. A & Whitesides, G. M. (2002) Directed dynamic self-assembly of objects rotating on two parallel fluid interfaces. The Journal of Chemical Physics 116:8571–8577.
4. Glotzer, S & Solomon, M. (2007) Anisotropy of building blocks and their assembly into complex structures. Nature Materials 6:557–562.
5. Mitragotri, S & Lahann, J. (2009) Physical approaches to biomaterial design. Nature Materials 8:15–23.
6. Akcora, P et al. (2009) Anisotropic self-assembly of spherical polymer-grafted nanoparticles. Nature Materials 8:354–359.
7. Yin, P, Choi, H. M. T, Calvert, C. R, & Pierce, N. A. (2008) Programming biomolecular self-assembly pathways. Nature 451:318–322.
8. Ke, Y, Ong, L. L, Shih, W. M, & Yin, P. (2012) Three-Dimensional Structures Self-Assembled from DNA Bricks. Science 338:1177–1183.
9. van Blaaderen, A. (2003) Chemistry: colloidal molecules and beyond. Science 301:470–471.
10. Li, F, Josephson, D. P, & Stein, A. (2011) Colloidal assembly: the road from particles to colloidal molecules and crystals. Angewandte Chemie International Edition 50:360–388.
11. Rechtsman, M. C, Stillinger, F. H, & Torquato, S. (2005) Optimized Interactions for Targeted Self-Assembly: Application to a Honeycomb Lattice. Physical Review Letters 95:228301–228305.
12. Torquato, S. (2009) Inverse optimization techniques for targeted self-assembly. Soft Matter 5:1157–1173.
13. Halverson, J. D & Tkachenko, A. V. (2013) DNA-programmed mesoscopic architecture. Physical Review E 87:62310–062317.
14. Mirkin, C. A, Letsinger, R. L, Mucic, R. C, & Storhoff, J. J. (1996) A DNA-based method for rationally assembling nanoparticles into macroscopic materials. Nature 382:607–609.
15. Alivisatos, A. P et al. (1996) Organization of'nanocrystal molecules' using DNA. 382:609–611.
16. Valignat, M, Theodoly, O, & Crocker, J. (2005) Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids. PNAS 102:4225–4229.
17. Biancaniello, P, Kim, A, & Crocker, J. (2005) Colloidal Interactions and Self-Assembly Using DNA Hybridization. Physical Review Letters 94:058302–058307.
18. Park, S. Y et al. (2008) DNA-programmable nanoparticle crystallization. Nature 451:553–556.
19. Nykypanchuk, D, Maye, M. M, van der Lelie, D, & Gang, O. (2008) DNA-guided crystallization of colloidal nanoparticles. Nature 451:549–552.
20. Kim, A, Biancaniello, P, & Crocker, J. (2006) Engineering DNA-mediated colloidal crystallization. Langmuir 22:1991–2001.
21. Lukatsky, D, Mulder, B, & Frenkel, D. (2006) Designing ordered DNA-linked nanoparticle assemblies. Journal of Physics: Condensed Matter 18:S567.
22. Zhang, C et al. (2013) A general approach to DNA-programmable atom equivalents. Nature Materials 12:741–746.
23. Knorowski, C, Burleigh, S, & Travesset, A. (2011) Dynamics and statics of DNA-programmable nanoparticle self-assembly and crystallization. Physical review letters 106:215501–215505.
24. Auyeung, E, Li et al. (2014) DNA-mediated nanoparticle crystallization into wulff polyhedra. Nature 505:73–77.
25. Macfarlane, R. J, Lee, B, Jones, M. R, Harris, N, Schatz, G. C, & Mirkin, C. A. (2011) Nanoparticle Superlattice Engineering with DNA. Science 334:204–208.
26. McGinley, J. T, Jenkins, I, Sinno, T, & Crocker, J. C. (2013) Assembling colloidal clusters using crystalline templates and reprogrammable DNA interactions. Soft Matter 9:9119–9128.
27. Martinez-Veracoechea, F. J, Mladek, B. M, Tkachenko, A. V, & Frenkel, D. (2011) Design rule for colloidal crystals of DNA-functionalized particles. Physical review letters 107:045902–045907.
28. Theodorakis, P. E, Dellago, C, & Kahl, G. (2013) A coarse-grained model for DNA-functionalized spherical colloids, revisited: Effective pair potential from parallel replica simulations. The Journal of Chemical Physics 138:025101.
29. Di Michele, L & Eiser, E. (2013) Developments in understanding and controlling self assembly of DNA-functionalized colloids. Physical Chemistry Chemical Physics 15:3115–3129.
30. Feng, L, Pontani, L.-L, Dreyfus, R, Chaikin, P, & Brujic, J. (2013) Specificity, flexibility and valence of DNA bonds guide emulsion architecture. Soft Matter 9:9816–9823.
31. Arkus, N, Manoharan, V. N, & Brenner, M. P. (2009) Minimal Energy Clusters of Hard Spheres with Short Range Attractions. Physical Review Letters 103:118303–118307.
32. Arkus, N, Manoharan, V. N, & Brenner, M. P. (2011) Deriving Finite Sphere Packings. SIAM Journal on Discrete Mathematics 25:1860–1901.
33. Hoy, R. S, Harwayne-Gidansky, J, & O'hern, C. S. (2012) Structure of finite sphere packings via exact enumeration: Implications for colloidal crystal nucleation. Physical Review E 85:051403–0514018.
34. Hormoz, S & Brenner, M. (2011) Design principles for self-assembly with short-range interactions. PNAS 108:5193–5199.
35. Hoogerbrugge, P. J & Koelman, J. M. V. A. (1992) Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. Europhysics Letters 19:155–160.
36. Groot, R & Warren, P. (1997) Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. The Journal of Chemical Physics 107:4423–4435.
37. Rogers, W. B & Crocker, J. C. (2011) Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling. PNAS 108:15687–15692.
38. Meng, G, Arkus, N, Brenner, M. P, & Manoharan, V. N. (2010) The Free-Energy Landscape of Clusters of Attractive Hard Spheres. Science 327:560–563.
39. Holmes-Cerfon, M, Gortler, S. J, & Brenner, M. P. (2013) A geometrical approach to computing free-energy landscapes from short-ranged potentials. PNAS 110:E5–E17.
40. Wu, K.-T et al. (2012) Polygamous particles. Proceedings of the National Academy of Sciences 109:18731–18736.
41. Reinhardt, A & Frenkel, D. (2014) Numerical evidence for nucleated self-assembly of DNA brick structures. Physical Review Letters 112:238103.

Supplementary Information for Size limits of self-assembled colloidal structures made using specific interactions

Zorana Zeravcic [*] [†], **Vinothan N. Manoharan** [*] [‡], **and Michael P. Brenner** [*] [†]

[*]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA,[†]Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, Massachusetts 02138, USA, and [‡]Department of Physics, Harvard University, 17 Oxford St. Cambridge, Massachusetts 02138, USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Simulation details

We use Dissipative Particle Dynamics (DPD) techniques [1, 2] to simulate self-assembly of structures show in Figs. 1 and 2, and all clusters discussed in Figs. 3 and 5 of the main text. Our specific simulation setup was first introduced in [3]. Similar to [4, 5], our system consists of two types of particles: 1) Solvent particles, modeled as standard DPD beads with soft repulsion, dissipative and random interaction; 2) Colloidal particles, which are larger DPD beads, that have the conservative force between two colloids replaced with $48-96$ Lennard-Jones interaction.

The dynamics of each particle in our simulation is governed by Newton's equations of motion:

$$\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}=\mathbf{v}_i,\quad m_i\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t}=\mathbf{f}_i. \qquad [1]$$

where $\mathbf{r}_i$ is the position vector of particle i, $\mathbf{v}_i$ is its velocity and m_i is its mass. All the particles in our simulation have equal mass $m=1$ (unit of mass in our simulation). The force acting on particle i is composed out of three parts:

$$\mathbf{f}_i=\sum_{j\neq i}\mathbf{F}^D_{ij}+\mathbf{F}^R_{ij}+\mathbf{F}^C_{ij}, \qquad [2]$$

where $\mathbf{F}^D$ is the dissipative force, $\mathbf{F}^R$ is the random force, and $\mathbf{F}^C$ is the conservative force. The dissipative force is:

$$\mathbf{F}^D_{ij}=-\gamma\omega^D(r_{ij})(\hat{\mathbf{r}}_{ij}\cdot\mathbf{v}_{ij})\hat{\mathbf{r}}_{ij}, \qquad [3]$$

where γ is the viscosity coefficient, $\mathbf{v}_{ij}=\mathbf{v}_i-\mathbf{v}_j$ is the relative velocity of particles i and j, $r_{ij}=|\mathbf{r}_i-\mathbf{r}_j|$ is the distance between the centers of particles i and j, $\hat{\mathbf{r}}_{ij}=\mathbf{r}_{ij}/r_{ij}$ is a unit vector and ω^D is a distance dependent weight function. The random force is:

$$\mathbf{F}^R_{ij}=(1/\sqrt{\Delta t})\sigma\omega^R(r_{ij})\theta_{ij}\hat{\mathbf{r}}_{ij}, \qquad [4]$$

where σ is the noise strength, ω^R is a distance dependent weight function, Δt is the simulation time step and θ_{ij} is a random variable. Instead of taking θ_{ij} to be a variable with a Gaussian distribution and unit variance, as is standard in DPD simulations, in this work we use $\theta_{ij}=\sqrt{3}(2\cdot\zeta-1)$ where ζ is a uniformly distributed random number $\zeta\in U(0,1)$ [6]. This choice makes the simulation very efficient and the results are basically indistinguishable from those calculated using Gaussian numbers. To ensure momentum conservation, in DPD algorithms $\theta_{ij}=\theta_{ji}$.

One of the two weight functions ω^D and ω^R can be chosen arbitrarily, therefore fixing the other weight function [1, 2, 7]. To ensure that the system has Gibbsian equilibrium the viscosity and noise have to be related by a fluctuation dissipation theorem; this leads to the following relations:

$$\omega^D(r_{ij})=(\omega^R(r_{ij}))^2,\ \ \sigma^2=2\gamma k_BT/m, \qquad [5]$$

where k_B is the Boltzman constant and T the temperature. In our simulations we use $\gamma=10$. As is practiced in DPD simulations $\omega^R_{ij}=\omega_{ij}$, where ω_{ij} is a simple (soft) weighting function that vanishes at some interaction range r_{cut}:

$$\omega_{ij}=\begin{cases}(1-r_{ij}/r_{cut}) & \text{if } r_{ij}\leq r_{cut}\\ 0 & \text{otherwise.}\end{cases} \qquad [6]$$

Since we simulate two types of particles, the interaction range will depend on which pair is interacting. The range of interaction r_{cut} equals: 1) $r^{CC}_{cut}=1.5D$ for two colloidal particles, where D is the diameter of colloids, 2) $r^{SS}_{cut}=0.5D$ for two solvent particles and 3) $r^{CS}_{cut}=1.0D$ for the interaction between a colloidal and a solvent particle. For the same reason, the conservative forces will differ depending on which particles are interacting. The colloid-colloid conservative force is modeled by interaction of $48-96$ Lennard-Jones spheres with a unit diameter $D=1$ and a short interaction range of $r^{CC}=1.05D$:

$$\mathbf{F}^{Ccc}_{ij}=\begin{cases}96\epsilon\left(\frac{1}{r_{ij}^{96}}-\frac{1}{r_{ij}^{48}}\right)\frac{\hat{\mathbf{r}}_{ij}}{rij} & \text{if } r_{ij}\leq r^{CC}\\ 0 & \text{otherwise,}\end{cases} \qquad [7]$$

where ϵ is the bond strength (energy scale in our simulations). The conservative force between two solvent (DPD) particles and between a solvent particle and a colloid is a soft repulsion:

$$\mathbf{F}^C_{ij}=a_{ij}\omega^C_{ij}\hat{\mathbf{r}}_{ij}, \qquad [8]$$

where $\omega^C_{ij}=\omega_{ij}$ but with $r^{CS}_{cut}=0.75D$, and a_{ij} is the repulsion parameter between particles i and j. Following [2], we use $a_{ij}=25k_BT$ for both solvent-solvent and solvent-colloid interactions and $\rho_{sol}=N_{sol}/V=3$ for the solvent number density, where the volume of the simulation box V follows from the colloid volume fraction ϕ_{coll} as $V=(N_{coll}\pi D^3)/(6\,\phi_{coll})$. Our simulation box has periodic boundaries in all 3 dimensions, and we use the linked-cell algorithm to speed up the calculations [8].

To integrate equations of motion we use the standard velocity-Verlet algorithm [9]. One of the advantages of DPD simulations is the use of large time steps in the integration. However, compared to soft DPD forces, the Lennard-Jones forces require significantly shorter time step for accurate integration. Therefore to exploit the advantageous efficiency of DPD in our simulation, we use the multiple time step algorithm [8] with $\Delta t=0.035$ for DPD forces and

Reserved for Publication Footnotes

$\Delta t_{LJ} = 0.0005$ for Lennard-Jones forces, where time is measured in the units of $(D/2)\sqrt{m/\epsilon}$.

Calibration of simulation. The values for interaction ranges we use are a result of calibration of our simulation based on the experimental results of [10]. In that work, the authors experimentally study spontaneous self-assembly of identical colloidal particles into small clusters and measure their yields. With the use of calibrated interaction ranges (values quoted above), our simulation reproduces the experimentally observed yields for all clusters of $N = 6, 7$ and 8 identical particles.

To explain the calibration procedure, we start by considering our simulation results for $N = 7$ identical particles. In [10] the authors showed that there are 6 rigid structures that can be assembled out of $N = 7$ particles (having 15 contacts), two of which are chiral enantiomers (see cluster images in Figure S1(b-f). Experimentally measured relative yields of the five ground states in equilibrium are shown to be consistent with the statistical mechanics calculations of their partition functions (dashed and solid horizontal lines in Figure S1(b-f).

In Figure S1 we show (a) absolute and (b) relative yields as a function of temperature T (in units of bond strength ϵ) for $N = 7$ particles, obtained from our calibrated simulations. Each data point is an ensemble average of 1000 different initial condition realizations, run at a fixed temperature T for $t_{run} = 15000$ time steps Δt. For these simulations, we use $\phi_{coll} = 1/30$ and $N_{coll} = 7$. At the end of a simulation run, we identify if a formed structure is one of the rigid clusters: Using the positions of colloids we form an adjacency matrix $\hat{A}$ (a $N_{coll} \times N_{coll}$ matrix, with an element $A_{ij} = 1$ if particles i and j are in contact and $A_{ij} = 0$ otherwise), calculate its eigenvalues and compare them with the referent values for polytetrahedron and octahedron. The eigenvalues of an adjacency matrix uniquely determine the cluster which is formed by the particles.

It is important to note that the colloid-colloid conservative interaction range r^{CC} must not be too large, because that would allow bonds between particles comprising a cluster that are geometrically impossible when particles are hard spheres. This interaction range is experimentally measured to be $r^{CC} \sim 1.05D$, and we use this value in all our simulations.

Designing the structures. To ensure that the desired cluster (or structure) is the only ground state that can be assembled out of N_{coll} particles, we need to introduce specific interactions between particles. In the example of clusters, following [11], we find maximal alphabets by choosing $n \leq N_{coll}$ particle types such that: *i)* Within the same type particles interact unfavourably and *ii)* with other types particles interact favorably. Generally all particles are different $n \equiv N_{coll}$ except in very special cases when multiple particles have the exact same set of neighbors, which effectively makes their interaction rules, and thus types, indistinguishable, see clusters in Figure S2(a) and (d).

For Figures 3(a) and 5 of the main text we also used non-maximal alphabets for $N_{coll} = 6, 7$ and 8, which were constructed using a straightforward and exhaustive enumeration procedure: All possible $N_{coll} \times N_{coll}$ interaction matrices are compared to all the adjacency matrices of N_{coll} clusters to find which ones encode a cluster uniquely. Such interaction matrices are the sought alphabets where identical rows and columns represent particles of the same type.

Temperature regimes. In the main text we emphasized that in general our simulations exhibit several regimes as a function of temperature T with a glassy regime at low T and an equilibrium melting regime at high T.

In the calibration simulations of identical particles explained in the previous section, above $T/\epsilon \gtrsim 0.16$ the relative yields match the equilibrium calculations (Figure S1(b-f). However the bonds between the colloids are short lived, leading to small absolute yields of the ground states, Figure S1(a). With a given ensemble size, this also leads to statistical noise in relative yields. The equilibrium regime extends down to $T/\epsilon \sim 0.1$, where the noise in the relative yields is small. Below $T/\epsilon \sim 0.1$, the relaxation time of clusters becomes comparable to t_{run} and the results are strongly influenced by kinetic effects.

In our cluster simulations with designed interactions, we observe the same temperature behavior. The equilibrium regime is confirmed by comparing the time and ensemble averages of the absolute yields (Figure S2).

The simulations of large structures exhibit similar temperature behavior; however, the melting of structures occurs rapidly at temperatures that vary with structure size and geometry. By comparing ensemble averages with different t_{run} we observe that the extent of kinetic regime depends on structure size and geometry as well, but always covers the whole range $T/\epsilon < 0.1$.

Finally we note that in our simulation setup, a dimer completely dissolves for $T/\epsilon > 0.25$. Although one might naively expect for this to occur at $T/\epsilon \gtrsim 1$, the stability of bonds is also influenced by the vibrational frequency in the Lennard-Jones potential well.

Maximum yield in simulations. Here we give the definition of the maximum yield Y_{max} in simulations.

For comparison with theoretical prediction, we are interested in maximal equilibrium yield. In the case of clusters (Figures 2 and 3 in the main text), we can identify the temperature range of the equilibrium regime, Section above, starting roughly at $T/\epsilon \sim 0.1$. The highest yield in this regime occurs at the lowest temperature; as the temperature grows we observe increasing bond fluctuations which are detrimental to the yield. We note that for clusters, entering the glassy regime below $T/\epsilon < 0.1$ the yield roughly levels off, so the extracted maximal equilibrium yield can be identified as maximal for all temperatures within the yield error.

For large structures, Figure 1 in the main text, the absolute yield curves have a pronounced peak that occurs roughly in the range of temperatures $T/\epsilon \in \{0.10, 0.16\}$. Although we cannot quantitatively precisely identify the equilibrium regime (see Section xx), we observe that for all structures at temperature $T \simeq 0.1$ the bonds switch from mostly frozen to fluctuating. Therefore we consider the peak yield as being in equilibrium regime.

Simulation details for Figures 2, 3 and 5. Figure 2 of the main text shows yield curves as a result of simulations of self-assembly of small clusters with $N \in \{6, \ldots, 10\}$ particles having designed interactions. Each data point in a yield curve is an ensemble average of 1000 different initial condition realizations, run with $\phi_{coll} = 1/30$ at a fixed temperature $T \in \{0.001, \ldots, 0.200\}$ for: $t_{run} = 13000$ for $N_{coll} = 6$, $t_{run} = 15000$ for $N_{coll} = 7$, $t_{run} = 17000$ for $N_{coll} = 8$, and $t_{run} = 19000$ for $N_{coll} = 9$ and $N_{coll} = 10$ time steps Δt. At the end of t_{run} a cluster is identified using the eigenvalues of the adjacency matrix that is constructed from the relative positions of the particles. Section below describes how the maximal equilibrium yield, used for Figures 3 and 5 of the main text, is extracted from the yield curves.

Simulation details for Figure 1. Figure 1b of the main text shows yield curves as a result of simulations of self-assembly of large arbitrary structures with $N \in \{19, 44, 69\}$ particles having interactions designed to assemble: a square bipiramid (two sizes), a chiral chain and a replica of Big Ben (see main text). Each data point in a yield curve is the fraction of successful assemblies in the ensemble of 100 different initial condition realizations, run with $\phi_{coll} = 1/30$ at a fixed temperature $T \in \{0.001, \ldots, 0.200\}$ for: $t_{run} = 10^5$ for $N_{coll} = 19$ and $t_{run} = 2 \cdot 10^6$ for other structures. To define a successful assembly we consider the time window of duration $\sim 5\% \cdot t_{run}$ ending at t_{run}. Within this time window we check the

formed bonds in ~ 10 regularly spaced time frames. If at least in one of the frames we observe the completely formed structure, the run is considered a successful assembly. For temperatures $T/\epsilon \gtrsim 0.1$ bonds fluctuate on time scale much shorter than the time window. On the other hand, the time scale to transition between local minima and ground states is longer than the time window. The case of the chiral chain $N_{coll} = 19$ is special, because the quasi one dimensionality allows transitions to local minima by breaking only two bonds. The chain is therefore more similar to the small clusters, so we only consider one timeframe, i.e., the state observed at t_{run}.

Entropic free energy loss due to broken bonds

Here we discuss our approximations of the function $f(m)$, which represents the entropic free energy loss $\exp\left[(S_m - S_0)/k_B\right]$ in an LM with m broken bonds (see Eqn.[3] of main text). The focus of this paper is not on such entropic effects, however, we do need to assign a value to $f(m)$ in the prediction of yield for the clusters (e.g., Figs.5 and 6 of main text) and for the illustrative examples of large structures (e.g., section on local defects in main text).

In analysis of cluster yields (Fig.5), the function $f(m)$ appears in the definition of horizontal axis variable which is common for all clusters. The precise value of f can vary slightly from cluster to cluster since their LMs vary in geometry and therefore in entropy. In lack of any such detailed knowledge, we use the zero-th order approximation $f(m) = const$, with $const$ a free parameter, which is in the end validated by the overall match of the observed data and the theoretically predicted trend for the yields.

As explained in the main text, Ref. [12] found that in the limit of vanishing interaction range and identical particles, the entropy loss is proportional to the number of missing bonds; in other words, the approximation $f(m) = \exp(s \cdot m)$, with s a constant, could be appropriate in that limit.

We have tested this exponential approximation for a range of s values, but in the case of clusters the limited range $m = 1, 2$ and the inherent noise in the data precluded any significant departures from the match of data and theory we present in Fig.5 using the zero-th order approximation. We can therefore conclude that the latter approximation is suitable for the presented analysis.

In the yield predictions for big structures (Fig.6) there is a wider range of broken bonds m so one might expect larger significance of $f(m)$. For instance, if we assume the exponential approximation, the s effectively raises the temperature (see Eqn.[3] of main text), and we have found that our predictions of yield and of the type of defects which dominate in the relevant LMs are both somewhat sensitive to temperature values in the arbitrarily chosen range 0.1 to 2 (the appropriate value for temperature without the s correction is 1), although the qualitative picture does not change. However, since calculating the entropic loss in the LMs or estimating the appropriate value for s in the exponential approximation are complicated challenges on their own, our zero-th order approximation is in the end validated only by the fact that the Big Ben replica and chiral chain simulations are consistent with the theoretical predictions for bulky and linear structures.

An important point is that a local defect and its entropic contribution, i.e., the vibrational and rotational motions allowed by the missing bonds in the defect, depend only on the local particle neighborhood. This means that an LM of an arbitrary large structure will have similar entropic contribution as an LM of the Big Ben replica, chiral chain, etc., having the same number and types of local defects. We therefore have no reason to assume that the approximation for $f(m)$ will fail for structures having more particles or more complicated shapes than the Big Ben replica and other structures simulated in this paper.

1. Hoogerbrugge, P. J & Koelman, J. M. V. A. (1992) Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. Europhysics Letters 19:155–160.
2. Groot, R & Warren, P. (1997) Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. The Journal of Chemical Physics 107:4423–4435.
3. Zeravcic, Z & Brenner, M. P. (2014) Self-replicating colloidal clusters. Proceedings of the National Academy of Sciences 111:1748–1753.
4. Dzwinel, W & Yuen, D. A. (2000) A two-level, discrete-particle approach for simulating ordered colloidal structures. Journal Of Colloid And Interface Science.
5. Whittle, M & Travis, K. P. (2010) Dynamic simulations of colloids by core-modified dissipative particle dynamics. The Journal of Chemical Physics 132:124906.
6. Nikunen, P, Karttunen, M, & Vattulainen, I. (2003) How would you integrate the equations of motion in dissipative particle dynamics simulations? Computer Physics Communications 153:407.
7. Español, P & Warren, P. (1995) Statistical Mechanics of Dissipative Particle Dynamics. Europhysics Letters (EPL) 30:191.
8. Allen, M. P & Tildesley, D. J. (1987) Computer simulation of liquids. (Oxford University Press).
9. Swope, W. C, Andersen, H. C, Berens, P. H, & Wilson, K. R. (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. The Journal of Chemical Physics 76:637–649.
10. Meng, G, Arkus, N, Brenner, M. P, & Manoharan, V. N. (2010) The Free-Energy Landscape of Clusters of Attractive Hard Spheres. Science 327:560–563.
11. Hormoz, S & Brenner, M. (2011) Design principles for self-assembly with short-range interactions. PNAS 108:5193–5199.
12. Holmes-Cerfon, M, Gortler, S. J, & Brenner, M. P. (2013) A geometrical approach to computing free-energy landscapes from short-ranged potentials. PNAS 110:E5–E17.

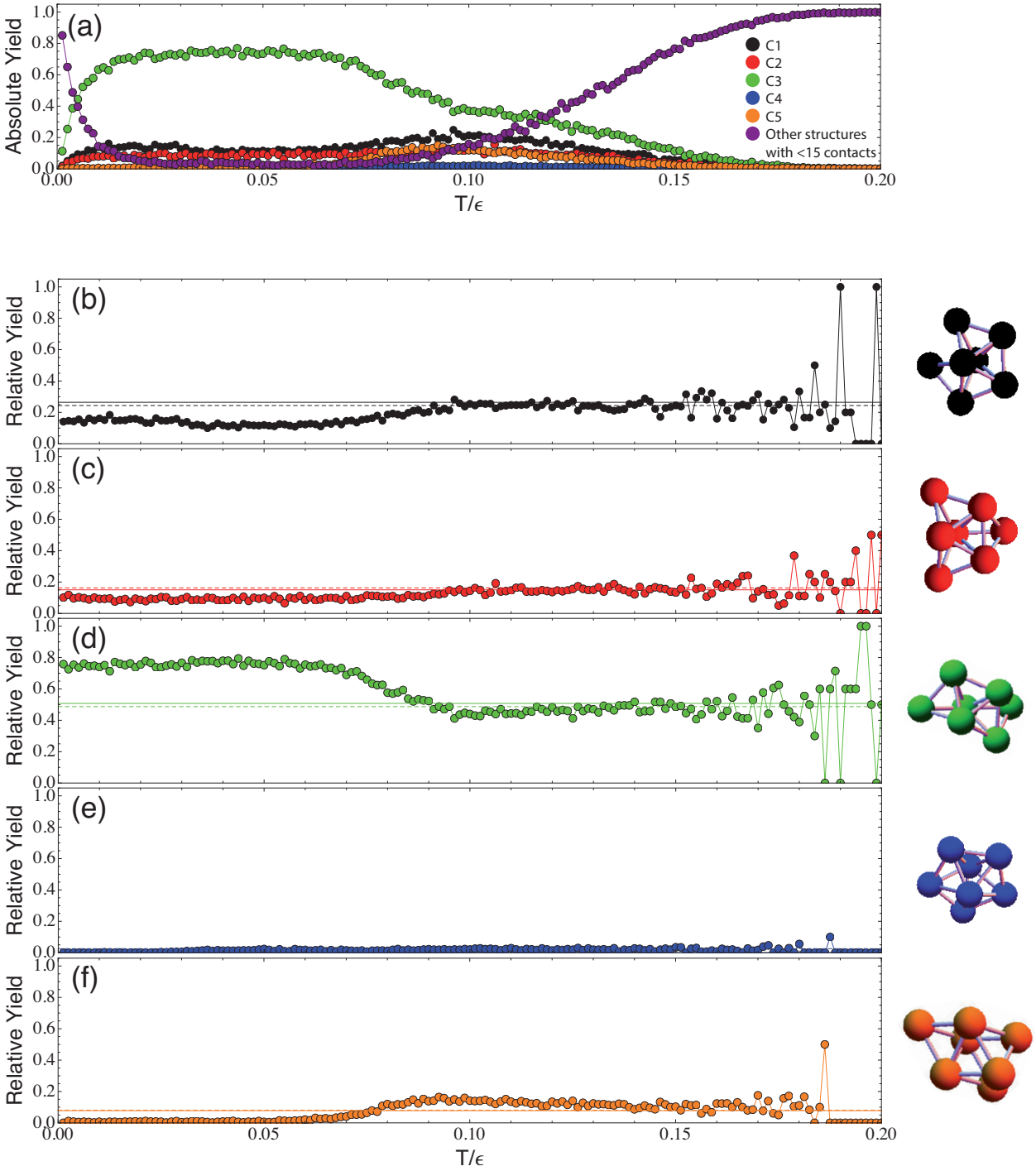

Figure S1: (a) Absolute yield as a function of temperature T/ϵ measured from simulations as explained in the text above. (b-f) Relative yield of the five rigid clusters as a function of temperature T/ϵ. Horizontal lines are referent values obtained from the experiments and partition function calculations. The onset of the equilibrium regime is at $T/\epsilon \simeq 0.1$.

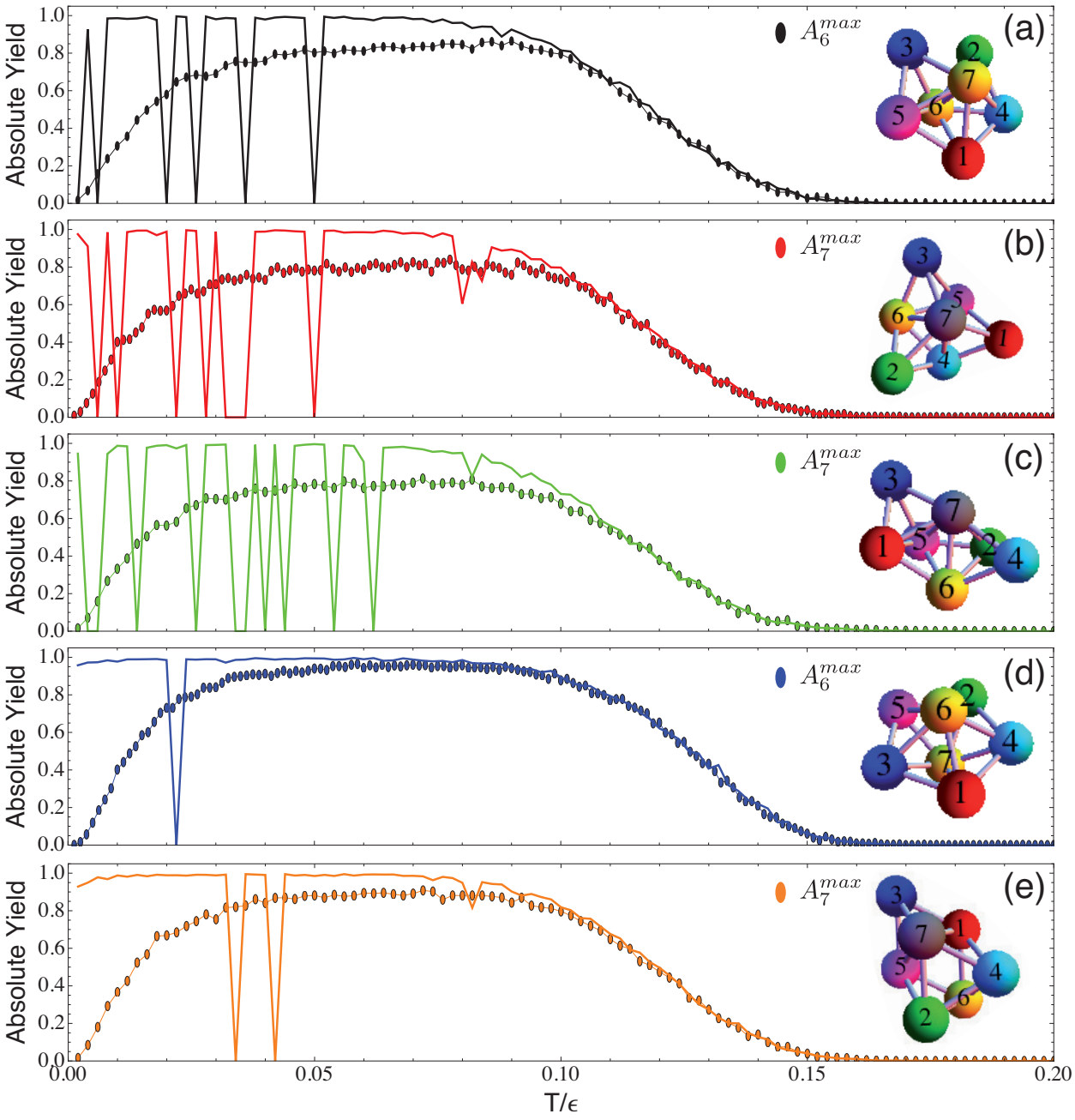

Figure S2: (a-e) Absolute yield as a function of rescaled temperature T/ϵ measured from simulations as explained in the text above. Data points are ensemble average results and solid lines are time average results. Comparison of the two averages for each cluster reveals that the equilibrium regime extends above $T/\epsilon \simeq 0.1$.

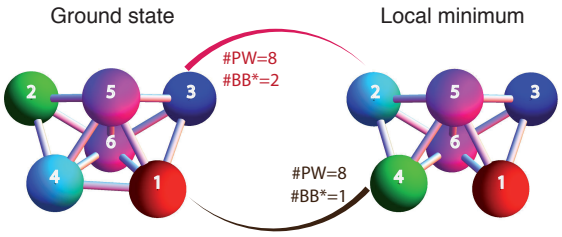

Figure S3: Energy landscape for a $N = 6$ cluster designed using maximal alphabet. Only the lowest energy LM is shown. It is missing one bond compared to the ground state and it can be obtained by permuting two particles in the ground state (see Fig. S6 cluster C1). $\#BB$ is the minimal number of bonds that need to be broken for transition between different states. $\#PW$ is the number of distinct pathways by which the transition can be achieved. For example, to transition from the ground state to the local minimum one needs to break at least two bonds. One of the eight pathways is to break bonds between particle pairs red — cyan and front-purple — cyan, and then smoothly exchange positions of cyan and green particles before reconnecting the cyan with the front-purple.

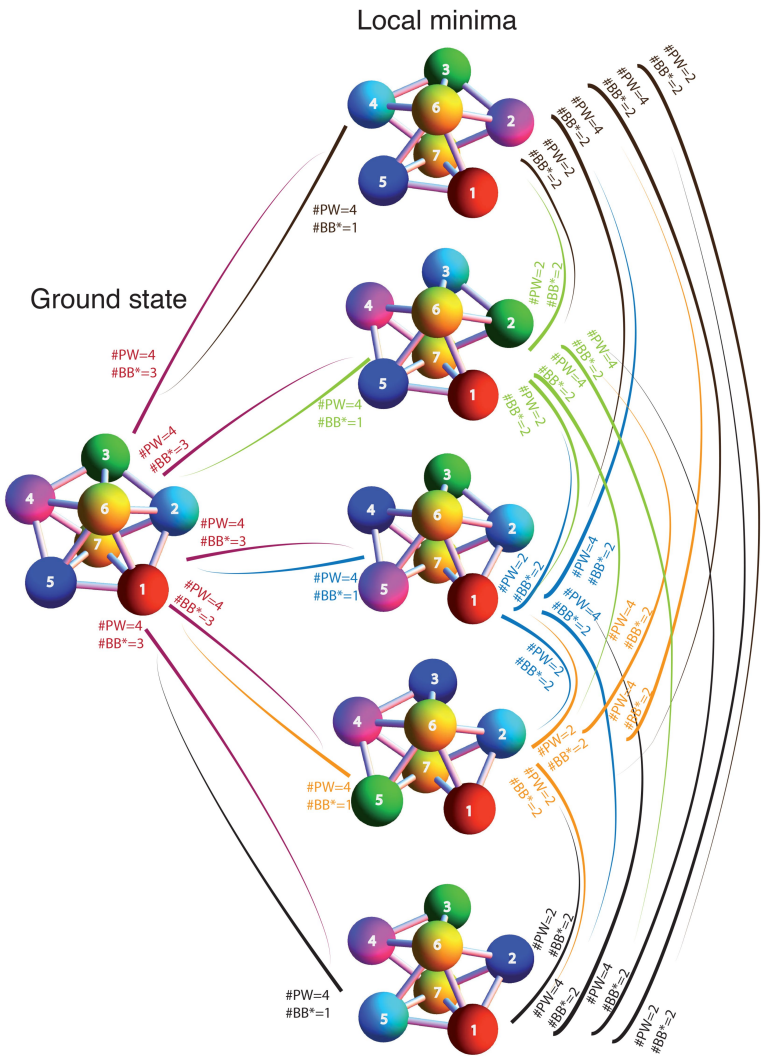

Figure S4: Energy landscape for a $N = 7$ cluster designed using maximal alphabet. Only the lowest energy LMs are shown, all missing two bonds compared to the ground state. All five minima can be obtained by permuting two particles in the ground state (see Fig. S6 cluster C4). $\#BB$ is the minimal number of bonds that need to be broken for transition between different states. $\#PW$ is the number of distinct pathways by which the transition can be achieved. For example, to transition from the ground state to the top local minimum one needs to break at least three bonds. One of the four pathways is to break bonds between particle pairs red — cyan, blue — purple and front-yellow — red, and then smoothly exchange positions of red and blue particles before reconnecting the red with the front-yellow.

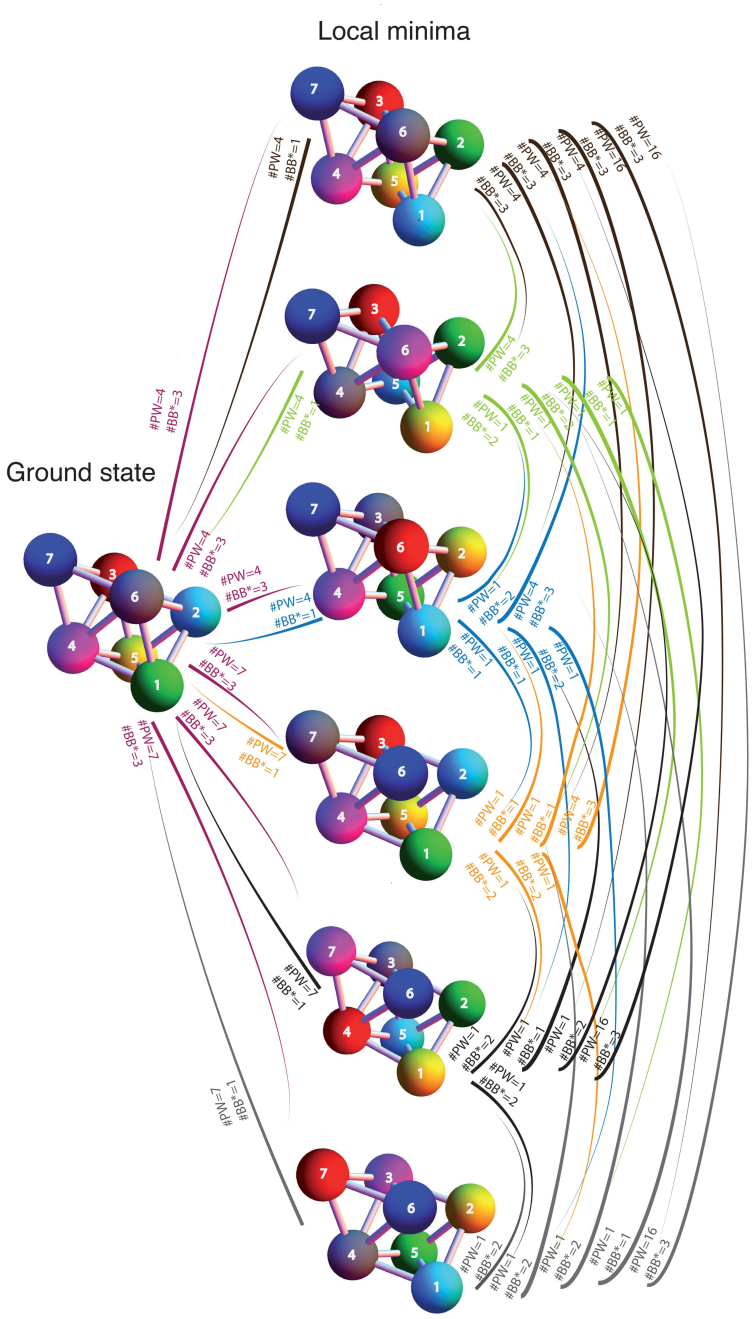

Figure S5: Energy landscape for a $N = 7$ cluster designed using maximal alphabet. Only the lowest energy LMs are shown, all missing two bonds compared to the ground state. All six minima can be obtained by permuting two particles in the ground state (see Fig. S6 cluster C5). $\#BB$ is the minimal number of bonds that need to be broken for transition between different states. $\#PW$ is the number of distinct pathways by which the transition can be achieved. For example, to transition from the ground state to the top local minimum one needs to break at least three bonds. One of the three pathways is to break bonds between particle pairs red — cyan, green — purple and gray — green, and then smoothly exchange positions of green and cyan particles before reconnecting the green with the gray. This cluster has a mirror (chiral) pair, that also has six lowest energy LMs (mirror images of the ones shown in this figure). We note that out of the total number of pathways written next to each of the transitions from bottom three local minima to the ground state, three pathways can lead to the chiral partner of the ground state.

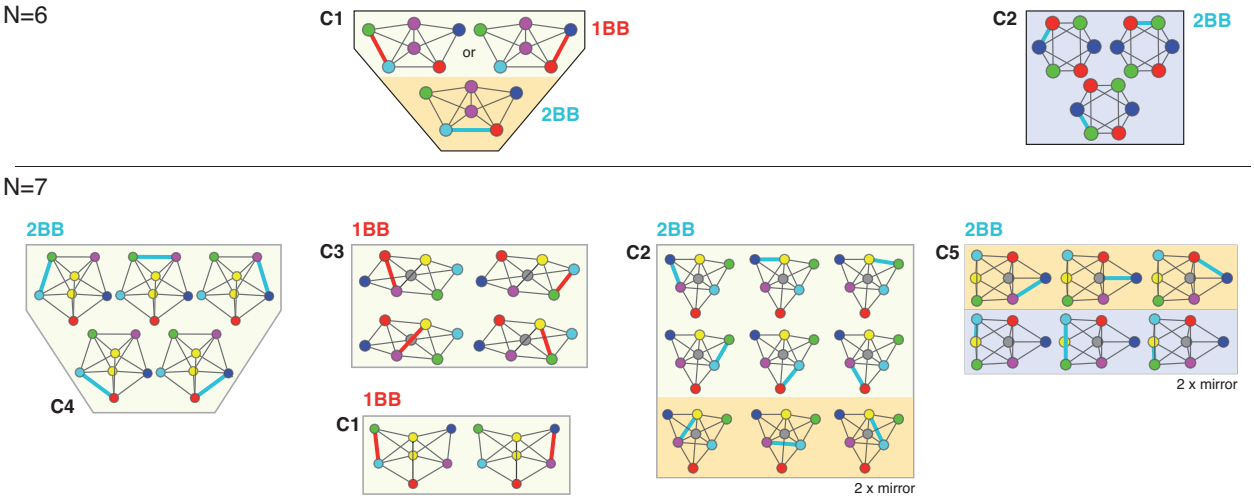

Figure S6: All low energy local minima of clusters designed using maximal alphabets can be obtained by permuting two neighboring particles — here we present all the clusters for $N_{coll} = 6, 7$ with their unique maximal alphabets. Thick bonds (red or blue) mark the particle pair whose permutation gives a local minimum having n broken bonds (red = 1 and blue = 2 bonds). The note "2×mirror" means additional LMs can be obtained by the same permutation from the mirrored (chiral) pair of the ground state. The background colors identify particular types of zero energy motion in the floppy local minimum. These persist in different clusters because the clusters share substructures.

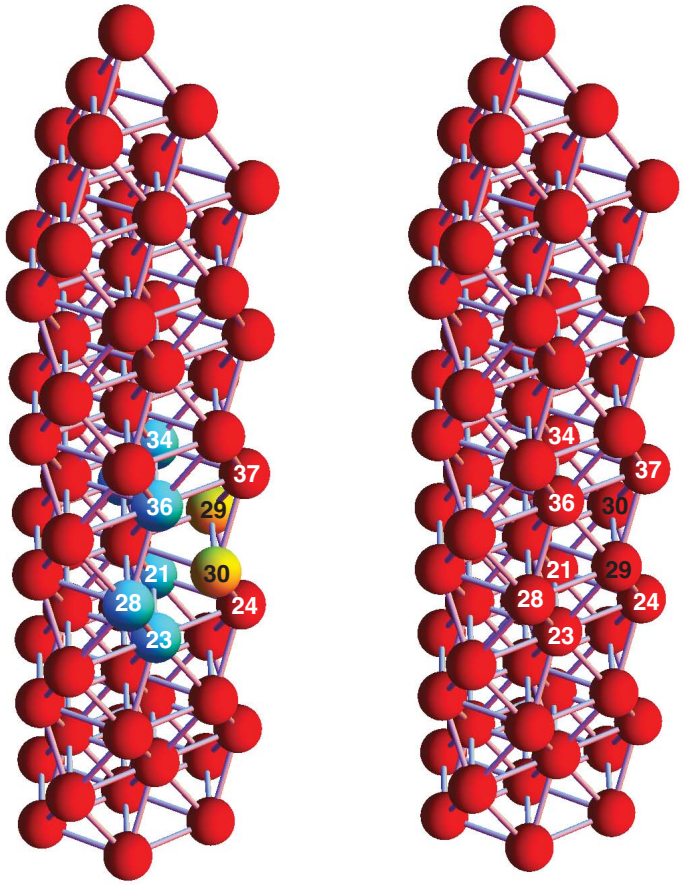

Figure S7: An example of a local minimum state obtained by creating a single surface defect in the Big Ben replica. The Big Ben is designed using maximal alphabet with all the particles being of different types, but we do not distinguish them by color. (Left) Full Big Ben structure with all the bonds. (Right) A single surface defect. The defect is created by permuting two nearest neighbor particles on the surface (particles 29 and 30 colored yellow). The total number of broken bonds is 6 (Eqn. 4 of the main text). Particles that have lost a bond are colored in cyan.

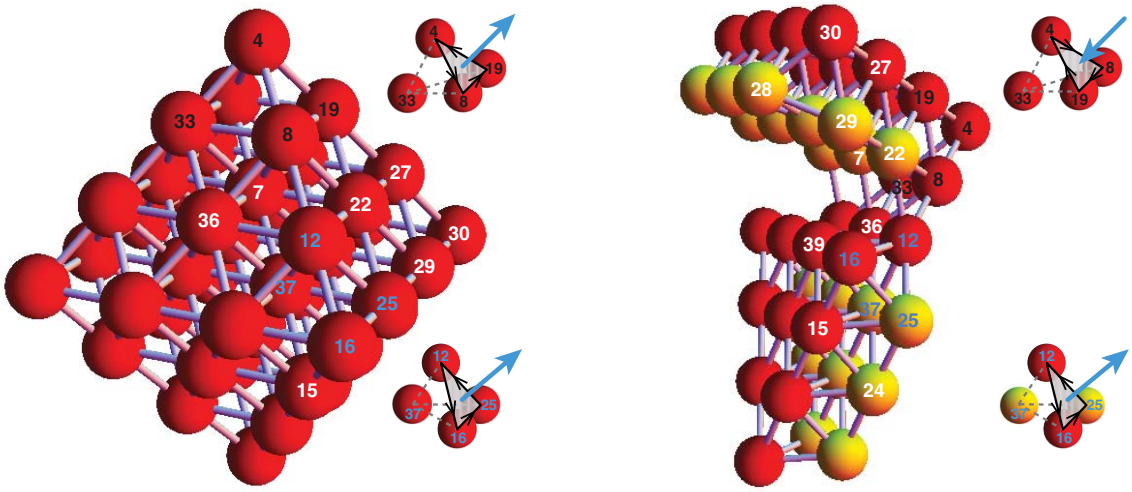

Figure S8: Kinetic effects due to chirality mismatch in the example of square bipyramid made out of $N_{coll} = 44$ particles. Although all the particles are of different types we do not distinguish them by color. (Left) Completely assembled square bipyramid. (Right) State from a simulation in which two parts of the structure with opposite chiralities are in contact. Particles colored yellow have missing bonds, and these two yellow layers should be adjacent in the bypiramid, however, due to chirality mismatch this is impossible. To see this, consider the four-particle substructures shown. The arbitrarily chosen ordering (4 — 8 — 19 and 12 — 16 — 25) defines a direction (blue arrow) that can be away from ("+") or towards ("−") the fourth particle (33 and 37). All signs ± are reversed by any mirror operation, specifically, they are opposite for any four-particle substructure in the two chiral versions of the bypiramid. The relative sign of the two chosen substructures (4 — 8 — 19 — 33 and 12 — 16 — 25 — 37) is changed (compare Left and Right) so the two bypiramid halves (Right) have opposite chirality; the yellow layers cannot bond due to all the mismatched triangles in them.

Video S1: Visualization of a self-assembly simulation of the Big Ben replica with $N_{coll} = 69$ particles that are all of different type.