Supporting Information

Zeravcic et al. 10.1073/pnas.1411765111

SI Text

SI Simulation Details

We use 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 ref. 3. Similar to refs. 4 and 5, our system consists of two types of particles: (*i*) solvent particles, modeled as standard DPD beads with soft repulsion, dissipative, and random interaction, and (*ii*) 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:

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i, \quad m_i \frac{d\mathbf{v}_i}{dt} = \mathbf{f}_i, \qquad \textbf{[S1]}$$

where $\mathbf{r}_i$ is the position vector of particle i, $\mathbf{v}_i$ is its velocity, and m_i is its mass. All of 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:

$$\mathbf{f}_i = \sum_{j \neq i} \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R + \mathbf{F}_{ij}^C, \qquad \textbf{[S2]}$$

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

$$\mathbf{F}_{ij}^D = -\gamma \omega^D(r_{ij})(\hat{\mathbf{r}}_{ij} \cdot \mathbf{v}_{ij})\hat{\mathbf{r}}_{ij}, \qquad \textbf{[S3]}$$

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

$$\mathbf{F}_{ij}^R = \left(1\big/\sqrt{\Delta t}\right)\sigma \omega^R(r_{ij})\theta_{ij}\hat{\mathbf{r}}_{ij}, \qquad \textbf{[S4]}$$

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:

$$\omega^D(r_{ij}) = \left(\omega^R(r_{ij})\right)^2, \sigma^2 = 2\gamma k_B T/m, \qquad \textbf{[S5]}$$

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

$$\omega_{ij} = \begin{cases} (1 - r_{ij}/r_{cut}) & \text{if } r_{ij} \leq r_{cut} \\ 0 & \text{otherwise.} \end{cases} \qquad \textbf{[S6]}$$

Because we simulate two types of particles, the interaction range will depend on which pair is interacting. The range of interaction r_{cut} equals (*i*) $r_{cut}^{CC} = 1.5D$ for two colloidal particles, where D is the diameter of colloids, (*ii*) $r_{cut}^{SS} = 0.5D$ for two solvent particles, and (*iii*) $r_{cut}^{CS} = 1.0D$ 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 $r^{CC} = 1.05D$:

$$\mathbf{F}_{ij}^{Ccc} = \begin{cases} 96\epsilon\left(\dfrac{1}{r_{ij}^{96}} - \dfrac{1}{r_{ij}^{48}}\right)\dfrac{\hat{\mathbf{r}}_{ij}}{rij} & \text{if } r_{ij} \leq r^{CC} \\ 0 & \text{otherwise,} \end{cases} \qquad \textbf{[S7]}$$

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:

$$\mathbf{F}_{ij}^C = a_{ij}\omega_{ij}^C \hat{\mathbf{r}}_{ij}, \qquad \textbf{[S8]}$$

where $\omega_{ij}^C = \omega_{ij}$ but with $r_{cut}^{CS} = 0.75D$, and a_{ij} is the repulsion parameter between particles i and j. Following ref. 2, we use $a_{ij} = 25 k_B T$ 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 three 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 with soft DPD forces the Lennard-Jones forces require a 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 $\Delta t_{LJ} = 0.0005$ 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 ref. 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 ref. 10 the authors showed that there are six rigid structures that can be assembled out of $N = 7$ particles (having 15 contacts), two of which are chiral enantiomers (see cluster images in Fig. 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 Fig. S1 *B*–*F*).

In Fig. S1 we show (*A*) absolute and (*B–F*) 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 1,000 different initial condition realizations, run at a fixed temperature T for $t_{run}=15{,}000$ 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 whether 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 that 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. The value of the interaction range $r^{CC}\sim 1.05D$ we use in all of our simulations roughly corresponds to the experimentally measured range of a DNA-coated $1-\mu$m particle.

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 ref. 11, we find maximal alphabets by choosing $n\le N_{coll}$ particle types such that (*i*) within the same type particles interact unfavorably 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 Fig. S2 *A* and *D*.

For Figs. 3*A* and 5 of the main text we also used nonmaximal 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 with all of 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 (Fig. S1 *B–F*). However, the bonds between the colloids are short-lived, leading to small absolute yields of the ground states (Fig. 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 (Fig. 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 (Figs. 2 and 3 in the main text), we can identify the temperature range of the equilibrium regime, discussed in the 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 that 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, Fig. 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 *Temperature Regimes*), 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 Figs. 2, 3, and 5. Fig. 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 1,000 different initial condition realizations, run with $\phi_{coll}=1/30$ at a fixed temperature $T\in\{0.001,\ldots,0.200\}$ for $t_{run}=13{,}000$ for $N_{coll}=6$, $t_{run}=15{,}000$ for $N_{coll}=7$, $t_{run}=17{,}000$ for $N_{coll}=8$, and $t_{run}=19{,}000$ 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. The section *Maximum Yield in Simulations* describes how the maximal equilibrium yield, used for Figs. 3 and 5 of the main text, is extracted from the yield curves.

Simulation Details for Fig. 1. Fig. 1*B* 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 (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. However, 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}).

SI 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 $\exp[(S_m-S_0)/k_B]$ in an LM with m broken bonds (see Eq. **3** in the 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., Fig. 5) and for the illustrative examples of large structures (e.g., *Local Defects and Asymptotic Yield* and Fig. 6 in the main text).

In the 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 because 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, Holmes-Cerfon et al. (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 Eq. **3** in the 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 0.1–2 (the appropriate value for temperature without the s correction is 1), although the qualitative picture does not change. However, because 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 entropic contribution similar to that of an LM of the Big Ben replica, chiral chain, and so on, 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 PJ, Koelman JMVA (1992) Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhys Lett* 19(3):155–160.
2. Groot R, Warren P (1997) Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *J Chem Phys* 107:4423–4435.
3. Zeravcic Z, Brenner MP (2014) Self-replicating colloidal clusters. *Proc Natl Acad Sci USA* 111(5):1748–1753.
4. Dzwinel W, Yuen DA (2000) A two-level, discrete-particle approach for simulating ordered colloidal structures. *J Colloid Interface Sci* 225(1):179–190.
5. Whittle M, Travis KP (2010) Dynamic simulations of colloids by core-modified dissipative particle dynamics. *J Chem Phys* 132(12):124906.
6. Nikunen P, Karttunen M, Vattulainen I (2003) How would you integrate the equations of motion in dissipative particle dynamics simulations? *Comput Phys Commun* 153(3):407.
7. Español P, Warren P (1995) Statistical mechanics of dissipative particle dynamics. *Europhys Lett* 30(4):191.
8. Allen MP, Tildesley DJ (1987) *Computer Simulation of Liquids* (Oxford Univ Press, Oxford).
9. Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys* 76:637–649.
10. Meng G, Arkus N, Brenner MP, Manoharan VN (2010) The free-energy landscape of clusters of attractive hard spheres. *Science* 327(5965):560–563.
11. Hormoz S, Brenner MP (2011) Design principles for self-assembly with short-range interactions. *Proc Natl Acad Sci USA* 108(13):5193–5198.
12. Holmes-Cerfon M, Gortler SJ, Brenner MP (2013) A geometrical approach to computing free-energy landscapes from short-ranged potentials. *Proc Natl Acad Sci USA* 110(1):E5–E14.

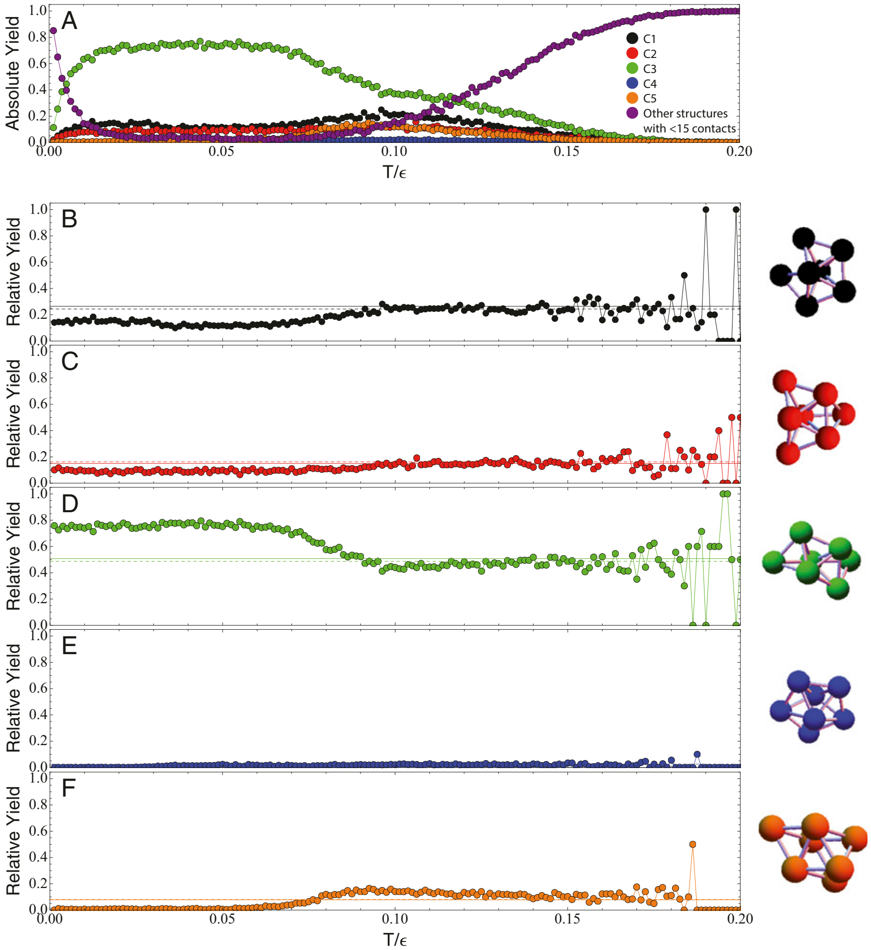


Fig. S1. (*A*) Absolute yield as a function of temperature T/ϵ, measured from simulations as explained in the text. (*B–F*) Relative yield of the five rigid clusters as a function of temperature T/ϵ. Horizontal lines are referent values obtained from the experiments and partition function calculations. The onset of the equilibrium regime is at $T/\epsilon \simeq 0.1$.

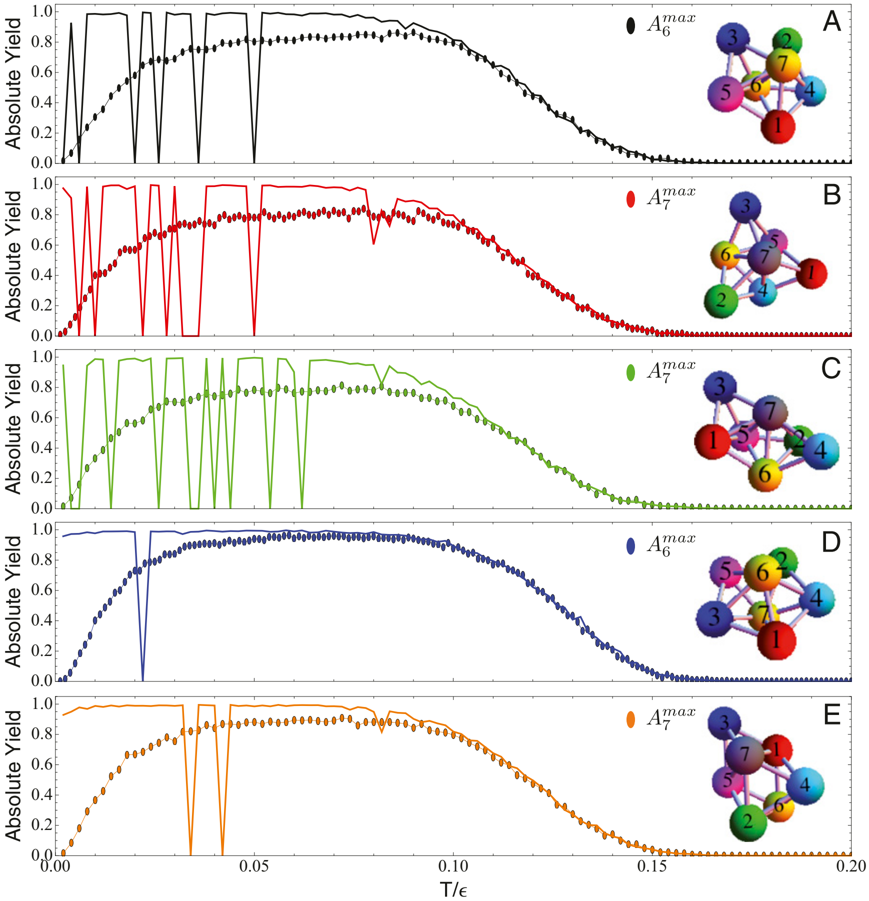


Fig. S2. (*A*–*E*) Absolute yield as a function of rescaled temperature T/ϵ, measured from simulations as explained in the text. Data points are ensemble average results and solid lines are time average results. Comparison of the two averages for each cluster reveals that the equilibrium regime extends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

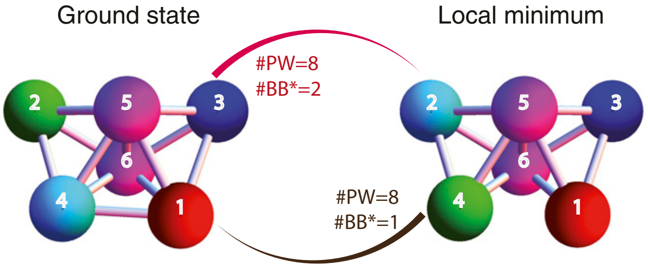


Fig. S3. Energy landscape for a $N = 6$ cluster designed using maximal alphabet. Only the lowest-energy LM is shown. It is missing one bond compared with the ground state and it can be obtained by permuting two particles in the ground state (Fig. S6, cluster C1). *#BB** is the minimal number of bonds 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.

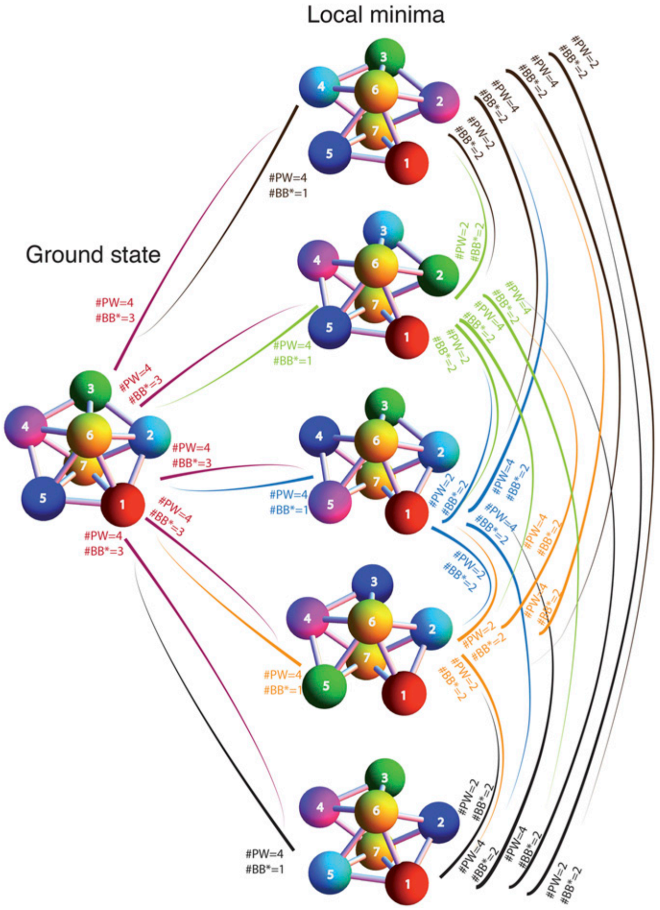


Fig. S4. Energy landscape for a $N=7$ cluster designed using maximal alphabet. Only the lowest-energy LMs are shown, all missing two bonds compared with the ground state. All five minima can be obtained by permuting two particles in the ground state (Fig. S6, cluster C4). #*BB** is the minimal number 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.

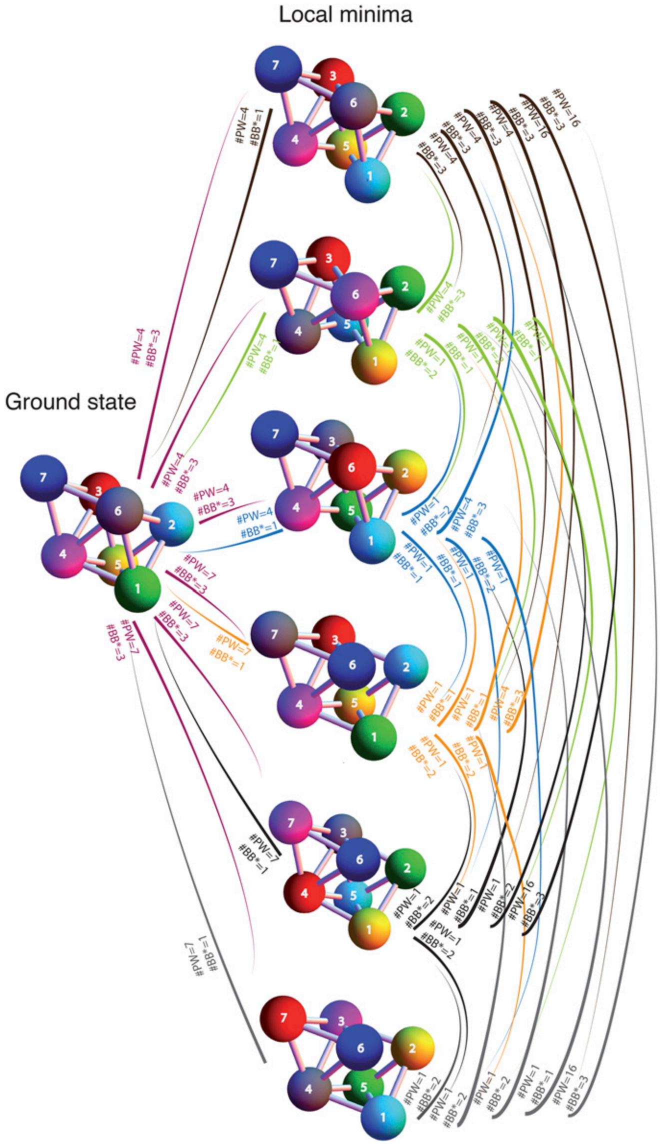


Fig. S5. Energy landscape for a $N = 7$ cluster designed using maximal alphabet. Only the lowest-energy LMs are shown, all missing two bonds compared with the ground state. All six minima can be obtained by permuting two particles in the ground state (Fig. S6, cluster C5). #*BB** is the minimal number 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.

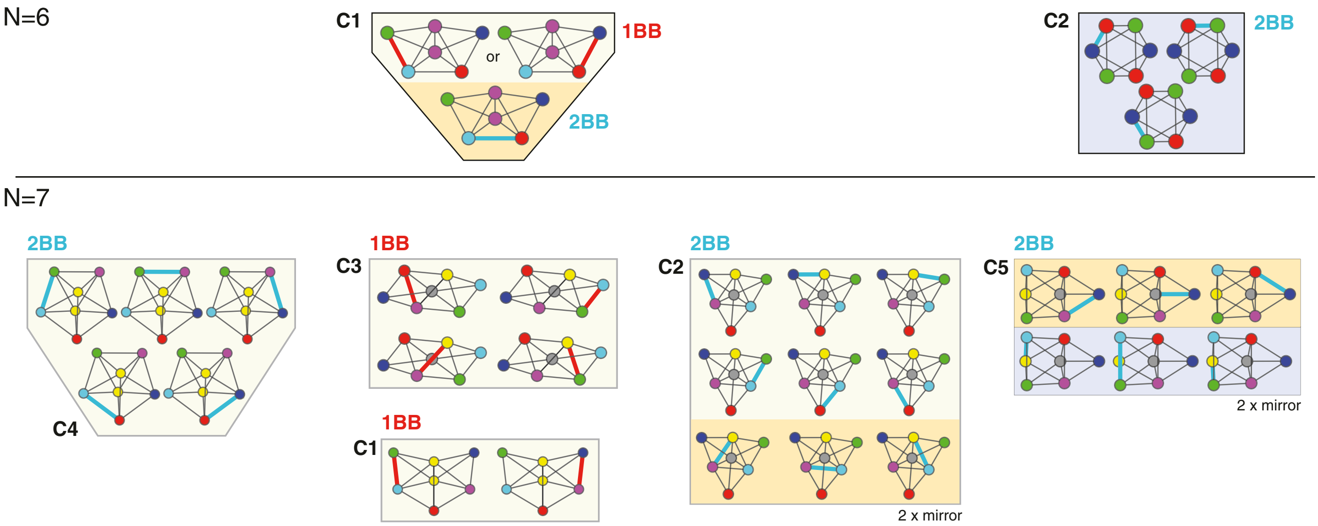


Fig. S6. All low-energy local minima of clusters designed using maximal alphabets can be obtained by permuting two neighboring particles—here we present all of the clusters for $N_{coll} = 6,7$ with their unique maximal alphabets. Thick bonds (red or blue) mark the particle pair whose permutation gives a local minimum having *n* broken bonds (red, one bond and blue, two 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.

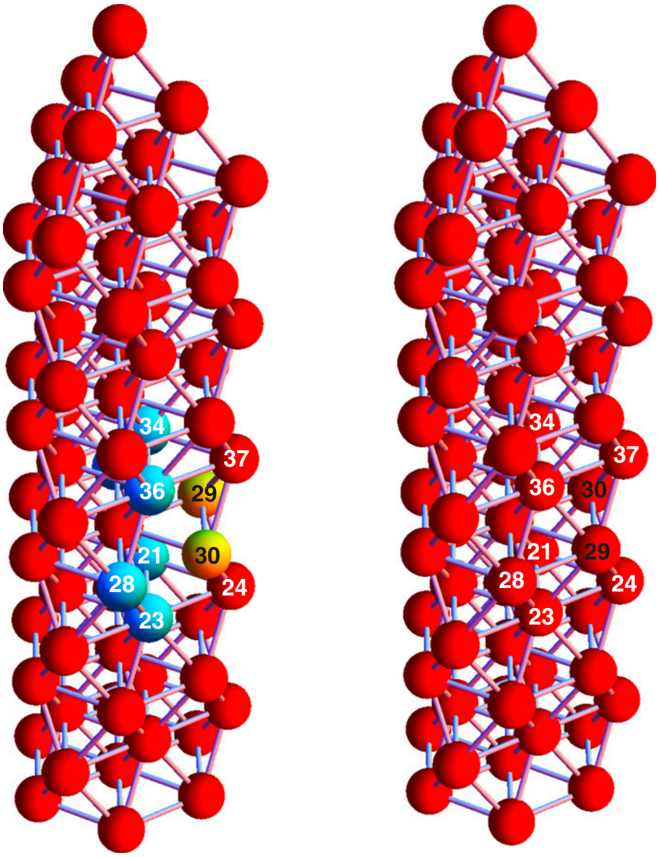


Fig. S7. An example of a local minimum state obtained by creating a single surface defect in the Big Ben replica. The Big Ben is designed using maximal alphabet with all of the particles being of different types, but we do not distinguish them by color. (*Left*) Full Big Ben structure with all of the 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 six (Eq. **4** in the main text). Particles that have lost a bond are colored in cyan.

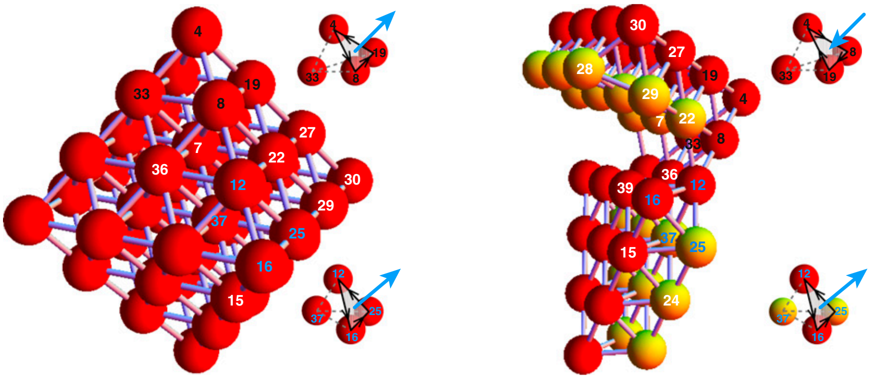
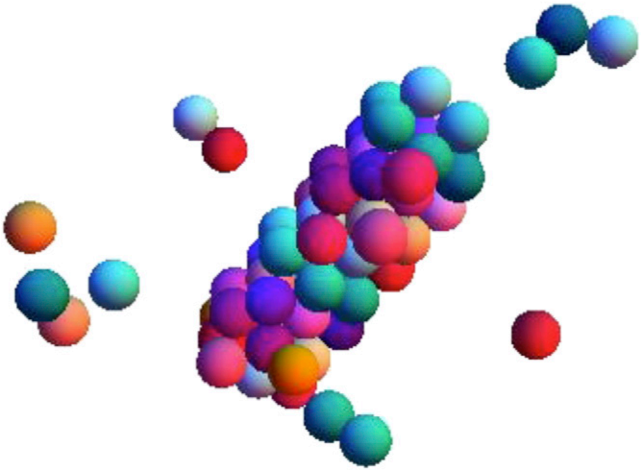


Fig. S8. Kinetic effects owing to chirality mismatch in the example of square bipyramid made out of $N_{coll} = 44$ particles. Although all of the particles are of different types we do not distinguish them by color. (*Left*) Completely assembled square bipyramid. (*Right*) State from a simulation in 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, owing 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 toward ("−") 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 owing to all of the mismatched triangles in them.



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

Movie S1