SUPPLEMENTARY INFORMATION

S1. FREE-ENERGY CALCULATIONS

In this section, we first review the theoretical method introduced in Ref. 1 for calculating a free-energy landscape from the connectivity graph that represents a target structure. We then describe how this approach can be applied to periodic addressable structures.

A. Review of the theoretical method

We model a solution of subunits with designed interactions in the grand-canonical ensemble, where the particles of each component type are held at a constant chemical potential. The set of designed interactions is encoded in the connectivity graph, G, that specifies the target structure. In such a graph, the vertices represent distinct subunits, while the edges indicate designed bonds. We can then write the grand partition function of all correctly bonded clusters as a sum over the connected subgraphs of G. This set of 'fragments' accounts for all possible on-pathway intermediates of the target structure. Each fragment, g, is in chemical equilibrium with the free monomers and can therefore be associated with the fugacity of the cluster that it represents[2],

$$z_g \equiv \exp\left[\beta \sum_{b\in\mathcal{E}(g)} \epsilon_b + \beta \sum_{v\in\mathcal{V}(g)} \mu_v + k_\mathrm{B}^{-1}\Delta S_\mathrm{r}(g)\right]. \quad \text{(S1)}$$

The sums in Eq. (S1) run over the fragment edges $\mathcal{E}(g)$ and vertices $\mathcal{V}(g)$, respectively; the magnitudes of the designed bond energies are given by $\{\epsilon_b\}$ and the chemical potentials by $\{\mu_v\}$. For simplicity, we shall assume here that all bond energies and all chemical potentials are the same, i.e., $\epsilon_b = \epsilon\,\forall b \in \mathcal{E}(G)$ and $\beta\mu_v = \ln\rho, \forall v \in \mathcal{V}(G)$, where ρ is the number density of each component in the reference fluid, although in general this need not be the case. Note that ρ corresponds to the volume fraction of each component in experiments. We furthermore assume that these clusters can be treated as ideal polymers. With this approximation, the final term in Eq. (S1), which accounts for the loss of rotational entropy, is

$$k_\mathrm{B}^{-1}\Delta S_\mathrm{r}(g) \equiv k_\mathrm{B}^{-1}\Delta S_\mathrm{c}(g) + k_\mathrm{B}^{-1}\Delta S_\mathrm{d}(g) \quad \text{(S2)}$$

$$k_\mathrm{B}^{-1}\Delta S_\mathrm{c}(g) \equiv -[V(g)-1]\ln q_\mathrm{c} \quad \text{(S3)}$$

$$k_\mathrm{B}^{-1}\Delta S_\mathrm{d}(g) \equiv -[V(g)-B(g)-1]\ln q_\mathrm{d}. \quad \text{(S4)}$$

The entropy difference ΔS_c accounts for the reduction of rotational degrees of freedom due to bond formation, and ΔS_d accounts for the reduction of the dihedral degrees of freedom resulting from the formation of closed loops. $B(g)$ is the number of bridges[3] of g, q_c is the maximum number of bonds formed by a monomer and q_d is the number of distinct dihedral orientations of a dimer. This approximation is appropriate in the dilute regime, where it is reasonable to ignore excluded volume interactions. The total number of distinct rotations of a monomer in the reference fluid is $q_\mathrm{r} \equiv q_\mathrm{c}q_\mathrm{d}$.

B. Calculation of (E,V)-sets

From Eq. (S1), it is clear that fragments with the same number of edges, E, and vertices, V, will have very similar thermodynamic properties. We can therefore avoid enumerating the exponentially large number of fragments and instead consider the average fugacities of sets of fragments, $h(E,V)$. We thus write the dimensionless grand partition function as

$$-\ln\Xi = -\sum_g z_g = -\sum_{E,V} |h(E,V)|\bar{z}_{E,V}, \quad \text{(S5)}$$

where $|h(E,V)|$ is the number of distinct fragments with E edges and V vertices and $\bar{z}_{E,V}$ is the average fugacity of a fragment in the set $h(E,V)$. With the assumption of equal bond energies and chemical potentials, this average fugacity is exactly

$$\bar{z}_{E,V} = \exp\left(E\beta\epsilon + V\beta\mu\right)\left\langle \exp[k_\mathrm{B}^{-1}\Delta S_\mathrm{r}(g)]\right\rangle_{g\in h(E,V)}, \quad \text{(S6)}$$

where $\langle\ldots\rangle_{g\in h(E,V)}$ averages over all fragments in $h(E,V)$. We must therefore determine which sets $h(E,V)$ are possible for a given target structure, and then count the number of fragments in each set. Once this 'density of states' of fragments is known, we can estimate the average rotational entropies by sampling uniformly from within each set.

Despite the enormous number of fragments in a typical addressable structure, it is possible to calculate the logarithm of the density of states, $\ln|h(E,V)|$, by applying the Wang–Landau flat-histogram method[4–6] to the state space of connected subgraphs of G. In this stochastic algorithm, we propose transitions between fragments that differ by a single edge, such that all visited graphs are connected subgraphs of G. Starting from a given fragment g, any edge in G that is adjacent to g may be added; however, edges whose removal would break g into two nontrivial subgraphs may not be deleted. In order for this algorithm to converge to the correct density of states, it is sufficient to require that the transitions between fragments obey detailed balance. Because this algorithm calculates the logarithm of the density of states, we can achieve an acceptable precision with far fewer iterations than there are distinct fragments. In other words, this stochastic approach exploits the degeneracy within the (E,V)-sets in order to make this combinatorial problem computationally tractable.

We note that the definition of the fragment fugacity provided in Eq. (S1) is the simplest formulation that is consistent with the ideal gas and ideal polymer approximations. We also place no restrictions on the connected

subgraphs of G, such as correlations in bond formation due to steric effects. These assumptions are appropriate for modeling molecular building blocks that are themselves floppy, like single-stranded DNA. However, if the building blocks are hard particles and the designed bonds do not allow dihedral rotations, then we can instead define the fragment set to be the *induced* connected subgraphs of G. In this case, we would modify the density-of-states algorithm to add or remove vertices instead of edges, and we would then choose $q_{\mathrm{d}} = 1$ such that $k_{\mathrm{B}}^{-1}\Delta S_{\mathrm{d}} = 0$ for all clusters. In general, performing calculations with the larger set of fragments is not expected to pose a problem for most systems, since maximally connected structures, which tend to have a very small number of bridges, dominate at equilibrium. It is also possible to improve the calculation of the ideal polymer entropy by including an additional term that accounts for the loop entropy of minimally connected clusters (i.e., few edges for many vertices). We have not included this term in this work because maximally connected structures dominate at equilibrium and, as a result, this additional complexity does not make a significant difference for the free-energy calculations.

C. Periodic target structures

In order to apply this theory to periodic addressable structures, we must account for the fact that the target structure, as well as the on-pathway clusters, may contain multiple copies of each component. The target structure can be described by an *infinite* graph having as many vertex colors[3] as there are distinct components. From this graph, we can define the unit-cell connectivity graph, G, in which every vertex is distinct. As shown in Figure 1c, we can also define a larger finite connectivity graph G' that contains an integer number of unit cells.

Unlike the case of a finite-sized structure, not every connected subgraph of G' corresponds to a physical cluster. In order to exclude off-pathway clusters, we only consider fragments that have fewer edges than the shortest invalid cycle[3], $E_{\max}$, of G'. We determine $E_{\max}$ by assuming that all vertices of G' are distinct and then finding the shortest path that connects any one of these vertices to its periodic image in the infinite connectivity graph of the target crystal. Although there are physically valid fragments in G' with more than E edges, counting these fragments using the density-of-states algorithm would lead to a systematic bias in $|h(E,V)|$. Instead, we consider fragments with more edges by performing calculations with a larger G'. Considering only fragments with $E < E_{\max}$ also ensures that the total number of fragments considered in the density-of-states algorithm is finite.

In order to apply the density-of-states algorithm, we must temporarily assume that all vertices of G' are distinct. As a result, the density-of-states calculation counts physically identical fragments multiple times, and thus systematically overestimates $|h(E,V)|$. For example, if there are $M \equiv |G'|/|G|$ copies of the unit cell and all fragments with more than two vertices are asymmetric (i.e., have no non-trivial automorphisms), then every fragment of G' will be overcounted exactly M times. This asymmetry condition is achieved by all addressable structures that have $E_{\max} > 2$. If $E_{\max} = 1$, then the connectivity graph would represent a single-component crystal; likewise, if $E_{\max} = 2$, then the graph would represent a binary crystal. In both of these cases, some fragments have additional symmetries that are not accounted for in the present density-of-states algorithm.

Clearly, the free-energy landscape must be independent of the number of unit cells included in G'. This is verified in Figure S1a, where we plot $F(V)$ for a bulk crystal without correcting for the overcounting of frag-

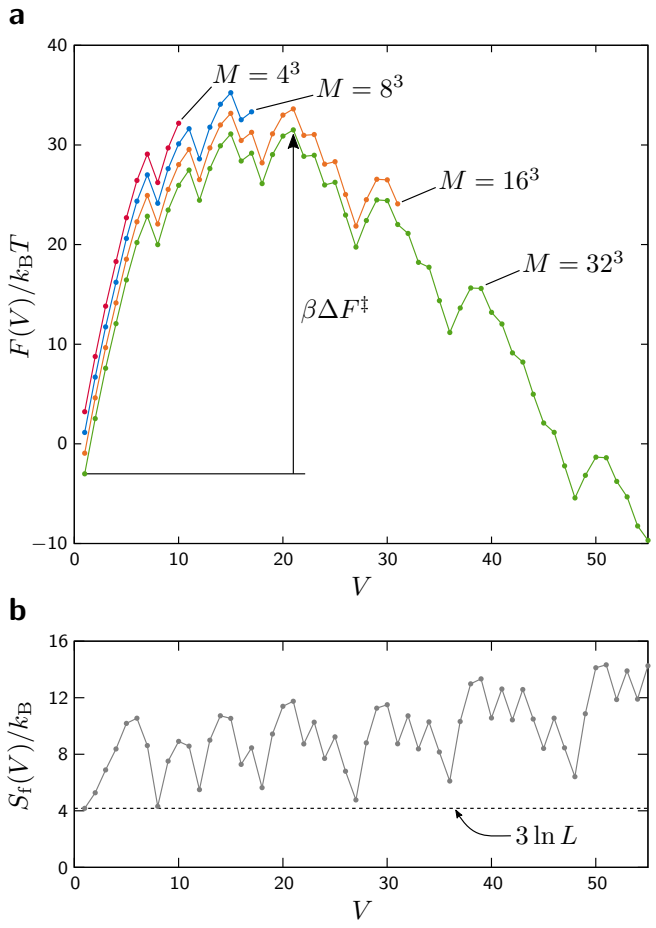

FIG. S1. The free-energy landscape of a multicomponent simple-cubic crystal is insensitive to the number of unit cells considered in the calculation. (a) Each free-energy profile is computed using M copies of a $L \times L \times L$ unit cell, where $L = 4$, by summing over all fragments with $E < E_{\max} = M^{1/3}L$. Here, we have not corrected for fragment overcounting; as a result, the constant difference between these profiles is $\beta\Delta F = -\ln \Delta M$. The height of the nucleation barrier, $\Delta F^{\ddagger}$, is independent of M. (b) The fragment entropy of clusters with V particles (see Sec. S1 D).

ments. Conversely, the free-energy landscape must not be affected if we instead consider the number of unit cells M to be fixed and the number of distinct components within each unit cell to be variable. We can therefore calculate the free-energy landscape of a given crystal *once* using a sufficiently large connectivity graph. Assuming that G' contains an integer number of unit cells, we can then specify the number of components in the periodic dimensions by adding a constant to Eq. (3) such that $F(1) = -\mu - k_{\mathrm{B}}T \ln V(G)$.

This theory predicts a complex free-energy landscape with many local minima for the elementary case of an infinite simple-cubic crystal (Figure S1a). These metastable cluster sizes correspond to a relatively small number of maximally connected fragments (Figure S1b), which have the largest number of edges for a given number vertices. The clusters associated with these fragments are stabilized by the formation of one or more cycles upon the incorporation of a single additional particle.

D. Fragment entropy

We define the 'fragment entropy' of all clusters containing V particles to be

$$S_{\mathrm{f}}(V) \equiv -k_{\mathrm{B}} \sum_g p(g|V) \ln p(g|V), \tag{S7}$$

where

$$p(g|V) \equiv \frac{\mathbf{1}[|\mathcal{V}(g)| = V]\, z_g}{e^{-\beta F(V)}} \tag{S8}$$

is the equilibrium probability of observing a cluster with connectivity graph g among all clusters with V subunits. Eq. (S7) can be rewritten in terms of quantities that are calculated directly in our theory,

$$k_{\mathrm{B}}^{-1} S_{\mathrm{f}}(V) = -\frac{\sum_E (E\beta\epsilon + V\beta\mu + k_{\mathrm{B}}^{-1}\Delta S_{\mathrm{r}})|h(E,V)|\bar{z}_{E,V}}{\sum_E |h(E,V)|\bar{z}_{E,V}} - \beta F(V). \tag{S9}$$

This entropy is a measure of the number of distinct fragments that contribute to the total free energy $F(V)$ of all clusters containing V particles. S_{f} tends to increase with V, although it decreases dramatically where specific maximally connected substructures dominate the free-energy landscape (Figure S1b).

The fragment entropy contributes to the thermodynamic driving force for self-assembly. In a finite-sized structure, S_{f} must go to zero upon complete assembly. Yet in an addressable structure with at least one periodic dimension, S_{f} can grow without bound. Clearly, at large cluster sizes, S_{f} should be extensive with respect to the number of (partially assembled) unit cells.

S2. DNA-BRICK CRYSTALS WITH COMPLEX MORPHOLOGIES

In this model of DNA-brick crystals, we assume that all structures are constructed on a tetrahedrally coordinated cubic lattice with geometric constants $q_{\mathrm{c}} = 4$ and $q_{\mathrm{d}} = 3$. The $a \times a \times a$ minimal repeating unit on this lattice corresponds to the 4 helix $\times$ 4 helix $\times$ 32 base-pair DNA-brick motif in Ref. 7. Unlike these experiments, however, we do not modify the connectivity of the subunits at the boundaries of the structure.

In Figure S2, we examine the free-energy landscapes of a class of non-convex addressable crystals. We consider columns with a single $h \times h$ channel and slabs with $h \times h$ pores embedded in repeating $L \times L$ unit cells. Despite the greater complexity of these structures, self-assembly with homogeneous bond energies is similar in many respects to that of the convex unit cells shown in Figure 2. We observe additional local minima in $F(V)$ but do not find any conditions where there are globally stable intermediate structures. As a result, we predict that phases with distinct finite cluster sizes or compositions cannot be thermodynamically favored when all designed bonds have equal strengths.

S3. PROBABILITY OF FORMING SPECIFIC CYCLES

In this section, we describe a method for unambiguously determining whether a particular fragment makes a complete loop around a channel or a pore in an addressable crystal. We discuss this method first in the case of a column with a single interior channel, and then in the case of a slab with an array of pores.

In the case of the column, we make a cut through one wall of the channel, as indicated by the orange stripe in Figure S3a, and identify the edges of the connectivity graph G' that correspond to bonds that intersect this cut in the three-dimensional structure. A fragment g encloses the channel if and only if there is at least one cycle of g that makes an odd number of crossings of this cut. We therefore check whether an arbitrarily chosen cycle basis of the fragment g includes at least one cycle that traverses an odd number of the identified edges of G'. For this particular fragment, the indicator function $\mathbf{1}_{\mathrm{cycle}}(g) = 1$ if g contains at least one such cycle; otherwise, $\mathbf{1}_{\mathrm{cycle}}(g) = 0$. We can then determine the

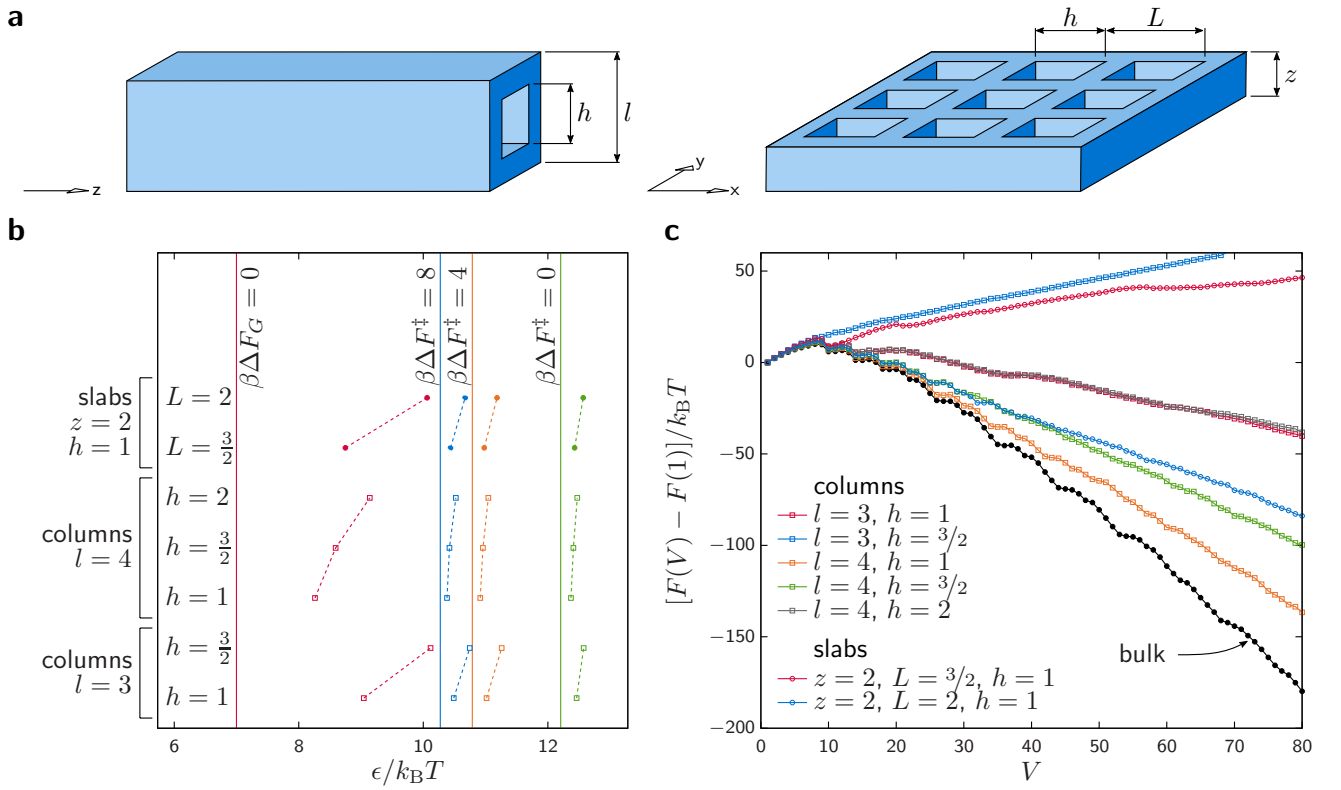

FIG. S2. The free-energy profiles of DNA-brick crystals with complex morphologies may differ significantly from the bulk profile. Nevertheless, there are no conditions where finite-sized clusters with homogeneous bond energies are globally stable. (a) Schematics of a column with a $h \times h$ channel (left) and a slab with a periodic array of $h \times h$ pores within $L \times L \times z$ unit cells (right). (b) The stability boundaries and nucleation behavior of selected structures. (c) The free-energy profiles of the selected structures, assuming $\beta\epsilon = 10$. All lengths l, z, L and h are given in units of the minimal repeating unit, a.

probability that a cluster of size V encloses the channel by calculating

$$p_{\text{enclosed}}(V) \equiv \sum_g \mathbf{1}_{\text{cycle}}(g) p(g|V) \tag{S10}$$
$$= \sum_E \left[\frac{\langle \mathbf{1}_{\text{cycle}}(g) e^{k_B^{-1}\Delta S_d(g)} \rangle_{g \in h(E,V)}}{\langle e^{k_B^{-1}\Delta S_d(g)} \rangle_{g \in h(E,V)}} \right] |h(E,V)| \bar{z}_{E,V} \Big/ \sum_E |h(E,V)| \bar{z}_{E,V}.$$

A similar strategy can be applied in the case of a perforated slab by making two such cuts, as shown in Figure S3b. This construction is used to determine whether the fragment g encloses either of the two neighboring pores indicated in Figure S3b. We must therefore repeat this test for every pair of neighboring pores. We then apply Eq. (S10), except in this case, $\mathbf{1}_{\text{cycle}}(g) = 1$ indicates that any one of the pores in G' is enclosed by a cluster with the connectivity graph g.

S4. CALCULATION OF HIERARCHICAL ASSEMBLY PHASE DIAGRAMS

In this section, we describe how the phase diagrams shown in Figures 4e and 4f can be constructed from the theory presented in Secs. II and S1. We address two complementary approaches: predictions based on the thermodynamic stability of the unit cell, Eq. (1), and more detailed calculations using the complete free-energy landscape. For simplicity, we assume that self-assembly takes place at constant chemical potential, and we label the resulting phases according to which (sub)structures are

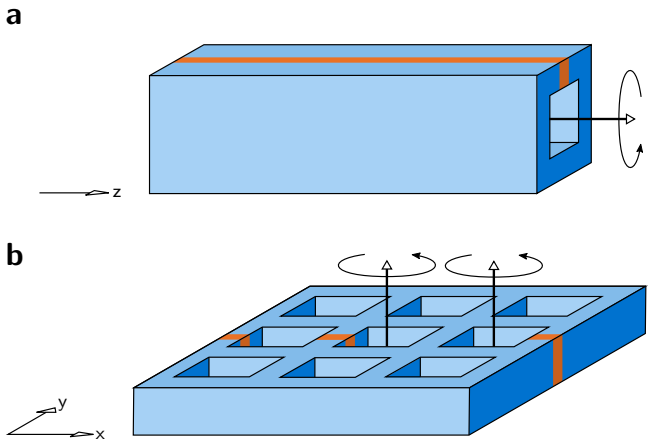

FIG. S3. Schematic diagrams showing the 'cuts,' indicated by orange stripes, required for cycle detection. (a) Any cycle with an odd number of crossings of the orange stripe encloses the interior channel. (b) Any cycle with an odd number of crossings of either of the orange stripes encloses one of the two indicated pores.

most stable. In an experiment with finite monomer concentrations, however, a monomeric fluid will always coexist with these stable (sub)structures. In such a case, the phase diagrams shown in Figures 4e and 4f report the phases that will form first, starting from a supersaturated fluid; the assembly reaction will then cease before all particles have assembled into the most stable (sub)structure as a result of monomer depletion.

A. Predictions based on the stability of the unit cell

For a given unit-cell connectivity graph G, we assume that all edges are assigned bond energies $-\epsilon_{\mathrm{A}}$ or $-\epsilon_{\mathrm{B}}$. We define E_{A} to be the number of edges in G with bond energy $-\epsilon_{\mathrm{A}}$, and E_{B} to be the number of edges with bond energy $-\epsilon_{\mathrm{B}}$. We further define V_{A} to be the number of vertices of G that are adjacent to edges with bond energy $-\epsilon_{\mathrm{A}}$, and V_{B} to be the number of vertices that are adjacent to edges with bond energy $-\epsilon_{\mathrm{B}}$. Note that $V_{\mathrm{A}} + V_{\mathrm{B}} > V(G)$. Using Eq. (1), the phase boundary for the complete crystal simplifies to

$$E_{\mathrm{A}}\beta\epsilon_{\mathrm{A}} + E_{\mathrm{B}}\beta\epsilon_{\mathrm{B}} + V(G)[\beta\mu - \ln q_{\mathrm{r}}] = 0. \qquad \text{(S11)}$$

Each substructure is independently stable if

$$E_{\mathrm{A,B}}\beta\epsilon_{\mathrm{A,B}} + V_{\mathrm{A,B}}(\beta\mu - \ln q_{\mathrm{r}}) > 0. \qquad \text{(S12)}$$

These equations are used to predict the boundaries of the unassembled fluid phase.

Because the structure shown in Figure 4b cannot grow into an infinite crystal without forming bonds of both types, this analysis suggests that Eq. (S11) should, in this case, define the phase boundary of the complete crystal. This predicted phase diagram is indicated by dashed lines in Figure 4d. However, the structure shown in Figure 4a can grow without bound using only one type of bond. We therefore calculate the phase boundary of the complete crystal assuming that one of the stabilized substructures, A or B, has already been assembled,

$$E_{\mathrm{A,B}}\beta\epsilon_{\mathrm{A,B}} + [V(G) - V_{\mathrm{B,A}}](\beta\mu - \ln q_{\mathrm{r}}) > 0. \qquad \text{(S13)}$$

The hatched region in Figure 4c indicates where the complete crystal is predicted to be stable.

B. Calculations using the complete free-energy landscape

With two different bond energies, we must now consider sets of fragments with the same number of edges of each type, E_{A} and E_{B}. Following Eq. (S6), the average fugacity in each set is

$$\bar{z}_{E_{\mathrm{A}},E_{\mathrm{B}},V} \equiv \exp\left(E_{\mathrm{A}}\beta\epsilon_{\mathrm{A}} + E_{\mathrm{B}}\beta\epsilon_{\mathrm{B}} + V\beta\mu\right) \times \left\langle \exp[k_{\mathrm{B}}^{-1}\Delta S_{\mathrm{r}}(g)]\right\rangle_{g\in h(E_{\mathrm{A}},E_{\mathrm{B}},V)}. \qquad \text{(S14)}$$

Likewise, we must now count the total number of fragments in each set, $|h(E_{\mathrm{A}}, E_{\mathrm{B}}, V)|$. We use these quantities to determine the globally stable cluster sizes on the free-energy landscape, and in this way we calculate the coexistence curves shown in Figures 4c and 4d. For the phases where the largest stable cluster size is infinite (yellow regions in the phase diagrams), the coexistence curves are approximated from free-energy landscapes calculated up to a maximum cluster size $V = 80$. We note that the assembled phases always encroach further into the unassembled fluid than predicted by the stability analysis. Similarly, it is also likely that the extent of the hatched region in Figure 4c, which is based solely on the stability analysis, is slightly underestimated.

[1] W. M. Jacobs, A. Reinhardt, and D. Frenkel, J. Chem. Phys. **142**, 021101 (2015).
[2] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, London, 2006).
[3] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, Vol. 6 (Macmillan, London, 1976).
[4] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[5] R. E. Belardinelli and V. D. Pereyra, J. Chem. Phys. **127**, 184105 (2007).
[6] R. E. Belardinelli and V. D. Pereyra, Phys. Rev. E **75**, 046701 (2007).
[7] Y. Ke, L. L. Ong, W. Sun, J. Song, M. Dong, W. M. Shih, and P. Yin, Nat. Chem. **6**, 994 (2014).