Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2015, **11**, 8930

Self-assembly protocol design for periodic multicomponent structures†

William M. Jacobs*[a] and Daan Frenkel*[b]

Assembling molecular superstructures with many distinct components will allow unprecedented control over morphology at the nanoscale. Recently, this approach has been used to assemble periodic structures with precisely defined features, such as repeating arrays of pores and channels, using a large number of building blocks. Here we propose a predictive tool that allows us to optimize the nucleation and growth of unbounded, ordered structures. In what follows, we call these structures 'crystals', even though they may only be periodic in one or two dimensions. We find that the nucleation barriers and growth pathways for crystals consisting of many components exhibit generic features that are very different from those of simple crystals. To illustrate the very non-classical nature of the nucleation and growth of such structures, we study the formation of one and two-dimensional crystals with multicomponent unit cells. We find that, whilst the boundaries in the non-periodic dimensions significantly affect the stabilities of these crystals, the nucleation barriers are largely determined by the local connectivity of the associated bulk crystal and are independent of the number of distinct components in the unit cell. We predict that the self-assembly of crystals with complex morphologies can be made to follow specific pathways toward the target structure that successively incorporate key features of the three-dimensional target structure. In contrast with simple crystals, it is possible to tune the kinetics of nucleation and growth separately, thus minimizing defect formation. We show how control over self-assembly pathways can be used to optimize the kinetics of formation of extended structures with arbitrary nanoscale patterns.

Received 24th July 2015,
Accepted 15th September 2015

DOI: 10.1039/c5sm01841b

www.rsc.org/softmatter

I. Introduction

Experiments in complex self-assembly have demonstrated that nanostructures composed of hundreds or thousands of distinct components can be assembled robustly by designing highly specific interactions between the subunits. Such structures are 'addressable' in the sense that every subunit belongs in a specific location within the three-dimensional target structure.[1–6] Using designed interactions between putative neighbors, finite-sized structures with an impressive array of nanoscale features have been assembled by using single-stranded DNA 'bricks' as the molecular building blocks.[7–10] More recently, this approach has been used to assemble crystals with multicomponent unit cells that are periodic in only one or two dimensions.[11] These 'semi-infinite' crystals are promising for a wide variety of applications in which macroscopic materials with intricate, periodic nanoscale features are desirable.[12–14]

Addressable crystals are an exciting prospect because the use of many components offers direct control over the mechanism of self-assembly. In conventional crystals, thermodynamic stability and the kinetics of nucleation and growth are inextricably linked. This relationship between stability and kinetics is customarily described by the classical theory of nucleation, which predicts a single free-energy barrier separating a disordered fluid phase from a thermodynamically favored crystal phase.[15–17] Under conditions where the fluid phase is only mildly supersaturated, the rate-limiting free-energy barrier may result in nucleation that is impractically slow. Typically the only means of speeding up nucleation is to increase the supersaturation of the fluid, but this strategy often results in crystals with many defects.[5,18] Such limited control over the self-assembly pathway is one fundamental reason why growing high-quality crystals of large particles, such as proteins[19,20] and functionalized colloids,[21] is notoriously difficult. However, with the introduction of heterogeneous building blocks, it is now possible to decouple the nucleation and growth kinetics while tuning the stability of the crystal phase independently. The ability to optimize the kinetics at all stages of assembly is crucial for achieving high yields of intricate multicomponent crystals.

The recent experiments on DNA-brick crystals indicate that self-assembly only occurs in a narrow temperature window[11]

[a] *Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02139, USA. E-mail: wjacobs@fas.harvard.edu*
[b] *Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: df246@cam.ac.uk*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5sm01841b

and that it appears to be non-hierarchical. One might be tempted to assume that the nucleation of these extended structures is qualitatively similar to the (classical) nucleation of simple crystals, and therefore different from the highly non-classical nucleation that we reported in model studies of finite-sized addressable structures.[22–24] However, as we show below, the nucleation of addressable crystals can also be surprisingly non-classical. Moreover, the compositional heterogeneity of the multicomponent unit cell has profound effects on the dominant self-assembly pathways well into the growth phase. We predict non-hierarchical self-assembly at the level of the simplest periodic structures, but our study also shows that we can design hierarchical pathways that achieve controlled multi-step nucleation on the way to the completion of the entire crystal. Understanding how to design such pathways will allow us to engineer increasingly complex crystals using a wider variety of molecular building blocks.[25–27]

II. Free-energy landscapes of addressable crystals

In ref. 23, we introduced a theoretical method for predicting the free-energy landscape of a finite-sized, aperiodic addressable structure. We then demonstrated in ref. 24 that this landscape can be used to describe the mechanism of assembly, starting from a dilute solution, in a near-equilibrium protocol. This study showed that both the topology of the target structure and the choice of stabilizing bond energies control the nucleation behavior in finite-sized addressable structures, and that these features can be tuned to design robust pathways for self-assembly. Here we describe how this theoretical method can be used to design self-assembly pathways for addressable crystals with complex unit cells. In what follows, we consider the assembly of a target structure that is stabilized by specific, saturating bonds. To meet this requirement, the interactions between subunits should be short-ranged, and the subunit binding sites must only be capable of forming one bond at a time. We implicitly assume that the bond distances between all pairs of specifically bonded particles are similar and that the placement of the binding sites on the subunits is consistent with the geometry of the three-dimensional target structure. Our approach is thus applicable to a wide variety of molecular building blocks, including DNA bricks and anisotropic colloids with functionalized patches.

In order for a specific three-dimensional crystal to be thermodynamically stable, the bonds between neighboring particles must be significantly stronger than any off-target interactions between subunits. We therefore consider a model crystal in which the building blocks only interact through 'designed' bonds. A simple example of a target structure is shown in Fig. 1a, where the designed bonds are indicated by lines connecting particles whose various colors indicate distinct component types. A single unit cell, which contains exactly one particle of each component, is highlighted. The connectivity of all particles in the perfectly formed crystal can be uniquely represented by a unit-cell graph, G, in which each vertex indicates a distinct component. The edges

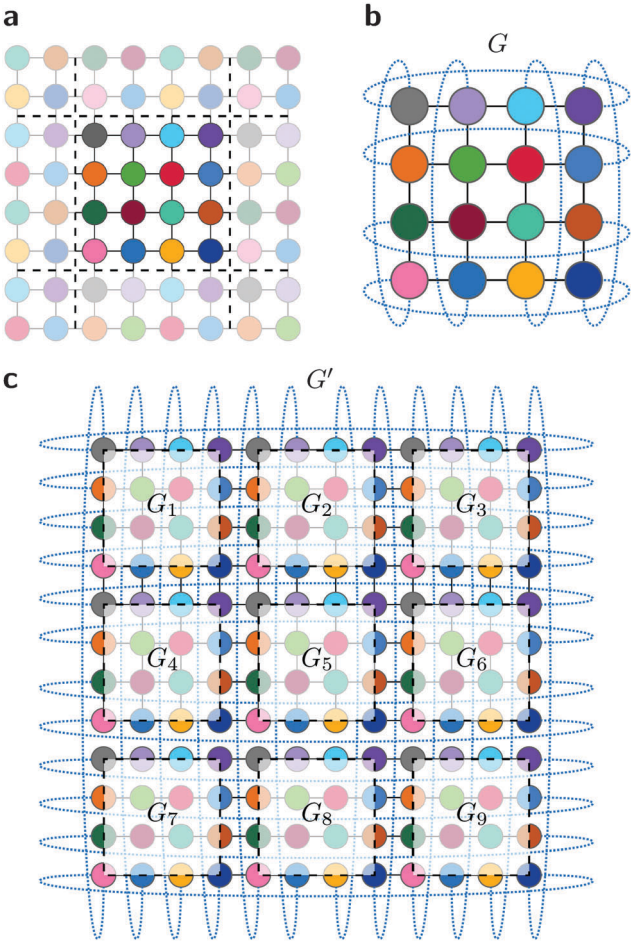

Fig. 1 A multicomponent crystal can be represented by a unit-cell connectivity graph that contains a single copy of each component. (a) A schematic of a two-dimensional crystal with a 16-component unit cell, with one such cell highlighted, and (b) its associated connectivity graph, G. All vertices of this graph are distinct, and the edges indicate designed interactions between components. Edges that connect a particle in the highlighted cell to the periodic image of its neighbor are shown here by dotted lines. (c) An example super cell, G', is composed of nine adjacent copies of the fundamental unit cell.

of G represent the designed bonds, either between particles that are physically adjacent within a single unit cell (solid lines in Fig. 1b) or between a particle and the periodic image of its neighbor in an adjacent unit cell (dotted lines). This distinction between types of edges depends only on how the graph is drawn and does not affect the thermodynamic properties of the crystal. Importantly, all topological properties of the infinite crystal are encoded in this unit cell.

The free energy per unit cell of the perfect crystal, relative to a dilute solution of free monomers, can be determined directly from the unit cell graph,

$$\Delta F_G \equiv -\sum_{b\in\mathcal{E}(G)} \varepsilon_b - \sum_{v\in\mathcal{V}(G)} \left(\mu_v - \beta^{-1}\ln q_r\right), \qquad (1)$$

where $\mathcal{E}(G)$ and $\mathcal{V}(G)$ are the edge and vertex sets, respectively, of the unit cell, $\beta \equiv 1/k_BT$ is the inverse temperature and k_B is the Boltzmann constant. The chemical potentials of the components,

$\{\mu_v\}$, and the magnitudes of the designed bond energies, $\{\varepsilon_b\}$, may, in principle, be different for each component v or bond b. The final term accounts for the loss of rotational entropy in a completely assembled crystal, where q_r is the rotational partition function of a monomer. Eqn (1) follows from the fact that the free energy of a macroscopic crystal must be extensive in the number of unit cells, and the equation $\Delta F_G = 0$ defines the phase boundary at which the defect-free crystal becomes stable. Although this suggests that self-assembly will only occur if $\Delta F_G < 0$, we shall see that it is possible to assemble stable, correctly bonded substructures under conditions where $\Delta F_G > 0$. In this regime, a structure with many missing components is substantially more stable than the defect-free target crystal. This entropic effect is much larger than in the case of simple crystals since not all subunits form the same number of designed bonds and, depending on the designed interactions, some particles may be bound much less strongly than others. However, this property of addressable structures is a strength, not a weakness: it means that, depending on the experimental conditions such as temperature and monomer concentrations, a given combination of building blocks can form a variety of structures with the same periodicity, but with very different patterns.

While a thermodynamic stability analysis is useful for estimating the phase boundaries of a multicomponent crystal, it does not predict whether a structure will assemble robustly. Thus, in order to study the kinetics of self-assembly, we calculate a free-energy landscape that takes into account all possible partial structures of the target crystal. These partial structures correspond to the connected clusters of the infinite target crystal. In a dilute solution, we can assume that the solution of correctly bonded clusters is ideal. The fugacity of a cluster represented by a connectivity graph g is then

$$z_g \equiv \exp\left[\beta \sum_{b\in\mathcal{E}(g)} \varepsilon_b + \beta \sum_{v\in\mathcal{V}(g)} \mu_v + {k_B}^{-1}\Delta S_r(g)\right]. \qquad (2)$$

The first two terms account for the designed interactions and the chemical potentials of the monomeric units that are incorporated in this cluster. The third term accounts for the loss of rotational entropy of the monomers upon association. (For further details, see ESI,† Section S1A.) In order to examine the effects of the connectivity of the unit cell, we shall initially assume that all bond energies are equal and that all monomeric units have the same chemical potential. The generalization to random heterogeneity in the bond energies is straightforward, and its effect on the free-energy landscape is discussed in ref. 23 and 24. However, the selective tuning of the strengths of individual bonds can lead to highly unusual free-energy landscapes for crystal self-assembly, with no counterpart in the self-assembly of crystals consisting of only a few components. This will be discussed in Section IIIC.

In order to predict the free-energy barrier for nucleation, we must account for all distinct clusters of the infinite target crystal. This includes clusters that span multiple unit cells and thus contain multiple particles of the same type. Because we can only explicitly consider finite unit cells in our calculations, we define a 'super' cell, G', that consists of multiple adjacent copies of the fundamental unit-cell, G. However, we must take care to exclude unphysically bonded clusters that are introduced by edges between periodic images of neighboring particles. We may thus only examine clusters with fewer edges than E_{max}, the length of the shortest path connecting a particle with its periodic image in a crystal of super cells. For example, in the nine-unit-cell super cell shown in Fig. 1c, some connected subgraphs with $E_{max} = 12$ percolate through the super cell and are therefore unphysical.

Because the total number of distinct clusters grows exponentially with E_{max}, directly enumerating these clusters quickly becomes intractable. We instead group the clusters into sets with the same numbers of edges, E, and vertices, V, since these topological properties are the primary determinants of the cluster fugacity given in eqn (2). We determine the number of clusters in each (E,V)-set and then estimate $\langle\exp[{k_B}^{-1}\Delta S_r(g)]\rangle_{E,V}$ in order to calculate the average fugacity of each set of clusters. (See ESI,† Section S1B for further details of these calculations.) Finally, we calculate the free energy of a cluster composed of a given number of particles, regardless of the identities of the particles involved,

$$F(V) \equiv -k_B T \ln \sum_g \mathbf{1}[|\mathcal{V}(g)| = V] z_g, \qquad (3)$$

where V is the number of particles in the cluster and the indicator function $\mathbf{1}[\cdot]$ evaluates to unity if its argument is true and to zero otherwise. Eqn (3) defines the free-energy landscape as a function of the degree of assembly of the target crystal.

Besides providing a practical algorithm for calculating the free-energy landscape of an addressable crystal, our approach provides an important insight into the thermodynamics of these structures. By computing the total number of distinct connected clusters in a super cell of size $|G'|$, we are overcounting every distinct cluster by a factor $|G'|/|G|$. It is then clear from eqn (3) that all free-energy differences are in fact independent of our definition of the unit cell, assuming that the unit cell does not possess any rotational symmetries (see ESI,† Section S1C). We therefore find that the number of distinct components in the unit cell does not affect the thermodynamics of self-assembly, including both the height of the nucleation barrier, $\Delta F^{\ddagger} \equiv \max[F(V)] - F(1)$, and the unit-cell stability, ΔF_G. All effects related to the use multiple distinct components instead originate from the non-periodic dimensions or any spatially heterogeneous motifs within the unit cell, as we shall see in the following sections. We discuss some practical aspects of this observation in Section IV.

III. Results

We now apply this theory to examine the stability, nucleation and growth of semi-infinite crystals composed of four-coordinated DNA bricks.[22] We consider model structures built from a minimal repeating unit of $4 \times 4 \times 4$ particles; we therefore define $a = 4$ to be the length of one minimal repeating unit in all dimensions. We also fix the chemical potential of every component to be $\mu = k_B T \ln \rho$,

where the per-component number density in the reference fluid is $\rho = 1 \times 10^{-5}$. While the results that we present here are specific to model DNA-brick crystals, these examples highlight many generic features of addressable multicomponent crystals.

A. Nucleation and stability of semi-infinite structures

We first consider two cases of three-dimensional, periodic DNA-brick structures: 'columns', which are periodic in the z dimension, and 'slabs', which are periodic in both the x and y dimensions. Schematics of these structures, as well as example three-dimensional models of tetrahedrally coordinated DNA bricks, are shown in Fig. 2a. Unsurprisingly, we find that both the stability boundaries and lines of constant nucleation-barrier height for these structures shift toward stronger bonds (or, alternatively, to lower temperatures) relative to the bulk crystal, *i.e.*, a structure in which all lattice sites are occupied (Fig. 2b). These boundaries converge to the bulk values as $l \to \infty$. We observe similar trends for both columns and slabs having the same number of particles, la or za, respectively, in the non-periodic dimensions. For these structures, we predict a single nucleation barrier under conditions where the target structure is thermodynamically stable. As a result, clusters consisting of more particles than the critical nucleus are expected to grow without bound. This prediction is consistent with existing experiments on DNA-brick crystals, which found an optimal temperature window for assembly with randomly assigned DNA sequences.[11]

The free-energy landscapes shown in Fig. 2c and d indicate that the kinetics of nucleation and growth are affected differently by the design of the unit cell. We note that for every cluster size $V > 1$, the cluster free energy of a semi-infinite structure, eqn (3), is determined by a subset of the connected clusters of the bulk lattice. In particular, the removal of certain maximally connected substructures, which form the greatest number of bonds for a given number of particles and are associated with

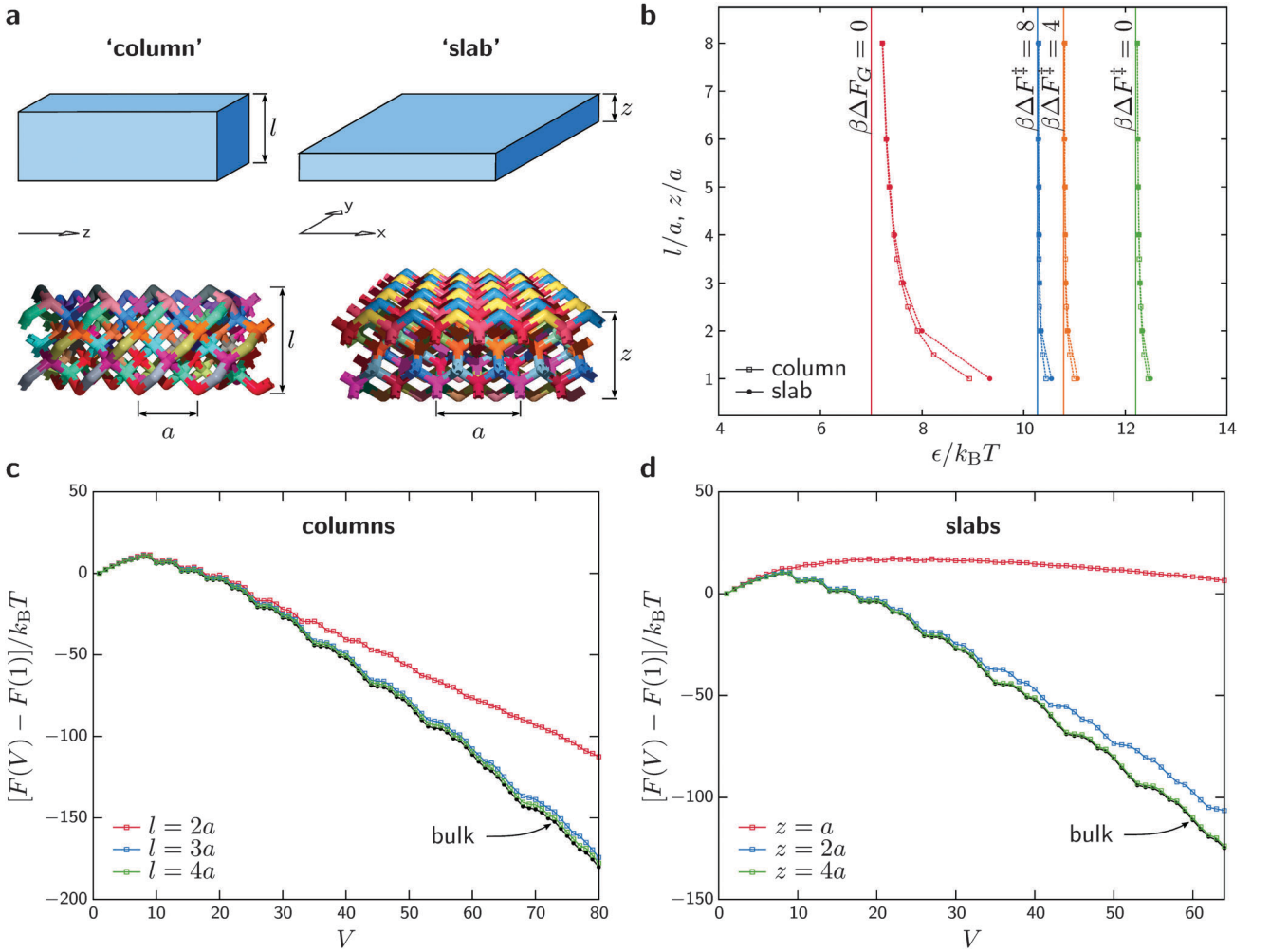

Fig. 2 Non-periodic boundaries affect the stabilities of DNA-brick crystals to a greater extent than they affect the nucleation behavior. (a) Schematic diagrams of 'columns' and 'slabs', which are periodic in the z and x–y dimensions, respectively. Example three-dimensional DNA-brick structures are shown below, with the minimal repeating length, a, indicated. (b) Stability boundaries and lines of constant nucleation-barrier height; the equivalent values for the bulk crystal are indicated by solid lines. (c and d) The free-energy landscapes of select columns and slabs, assuming $\beta\varepsilon = 10$. As l and z increase, these landscapes converge to the bulk free-energy landscape. The $z = a$ slab has a planar connectivity graph and thus exhibits drastically different nucleation behavior.

the local minima of $F(V)$, drastically alters the landscape. This effect can be seen in the divergence of the free-energy profiles of the $l = 2a$ and $z = 2a$ structures from the bulk profile in Fig. 2c and d. A more extreme example is the $z = a$ slab (Fig. 2d), which has a planar connectivity graph, and, as a result, is missing the strongly connected substructures that give rise to the step-like behavior of the bulk free-energy profile. These examples show that the thermodynamic driving force for self-assembly, $\partial F/\partial V$, can be significantly weaker for semi-infinite structures than for the bulk crystal under equivalent conditions. However, the free energies of small clusters, which determine the nucleation barrier, converge much more rapidly to the bulk values (Fig. 2b). This is because the non-classical nucleation barriers shown in Fig. 2c and d are primarily determined by the appearance of certain stable motifs, which may span multiple adjacent unit cells and are only weakly affected by the non-periodic boundaries of a semi-infinite structure. These observations imply that semi-infinite structures can be self-assembled at lower supersaturation than the corresponding bulk crystal without significantly affecting the nucleation rate. Simple crystals contain fewer defects when they are grown at low supersaturation, and recent simulation results[5] have demonstrated that the same holds true for the assembly of addressable structures. The free-energy landscapes of semi-infinite structures are thus more favorable than those of bulk crystals for achieving defect-free self-assembly. In Section IIIC, we shall show how this feature can be exploited in the design of a hierarchical self-assembly pathway.

B. Growth pathways in structures with complex morphologies

One consequence of the preferred assembly of extended structures that incorporate only a subset of the components of the target crystal is that the cluster size, while still an unambiguous measure of the progress of 'correct' assembly, does not measure the degree of assembly of the defect-free target structure. As an example, we consider two crystals from a class of structures with non-convex unit cells in Fig. 3. (An analysis of a variety of such structures is presented in ESI,† Section S2.) These structures exhibit mesoscale features that are larger than a single building block but smaller than the entire unit cell. Here the relevant mesoscale features are an interior channel (in the case of the column in Fig. 3a), and periodically repeating rectangular pores (in the case of the slab in Fig. 3b). In the case of the hollow column, we can now distinguish between large clusters that comprise only a single face of the column and clusters that completely enclose the interior channel as intended (see Fig. 3a). Using our theoretical approach, we can calculate the probability that a cluster of a given size completes a full loop around the interior channel. (For details of this calculation, see ESI,† Section S3.) Similarly, for the case of the perforated slab, we can calculate the probability that a cluster of a given size completely

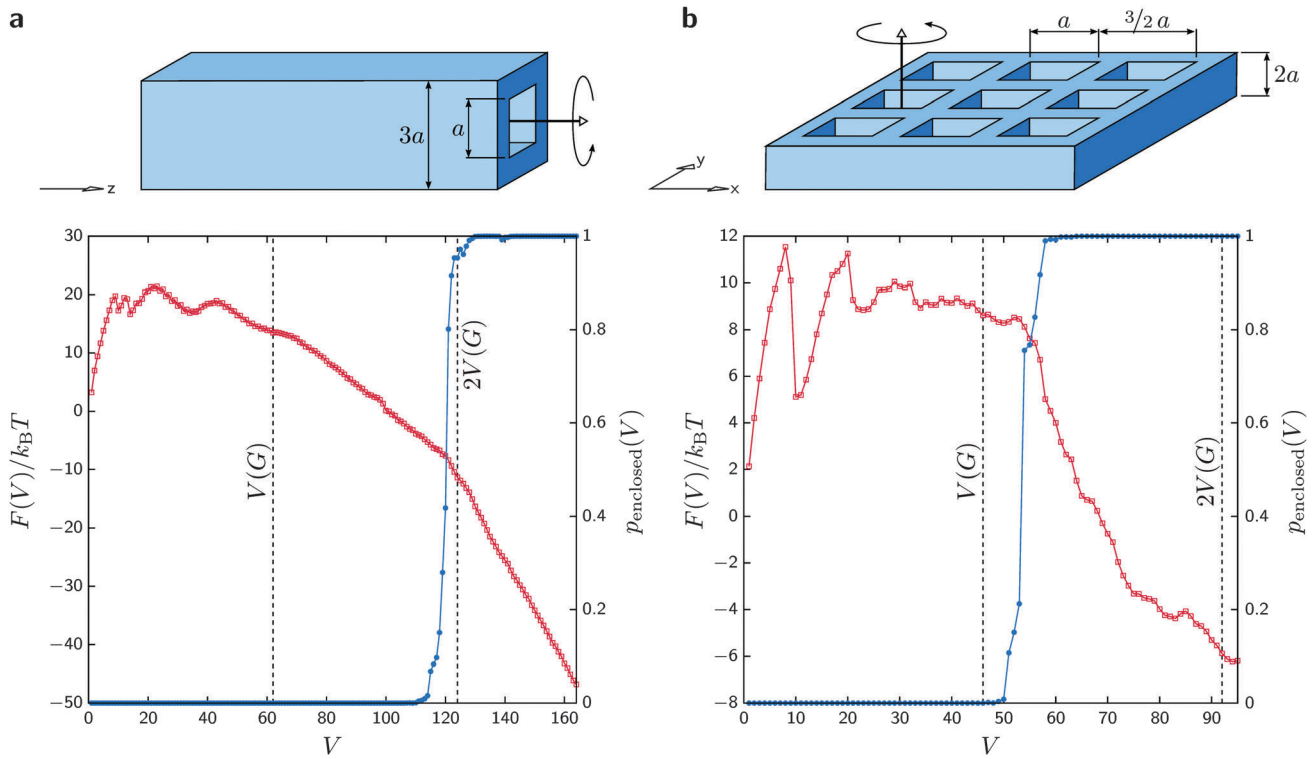

Fig. 3 The dominant growth pathways of DNA-brick crystals with complex morphologies are highly reproducible. (a) The free-energy landscape of a $3a \times 3a$ column with an $a \times a$ interior channel ($\beta\varepsilon = 9.6$), and the probability, $p_{\text{enclosed}}(V)$, that a cluster of V subunits encloses this channel, as indicated on the three-dimensional model. (b) The free-energy landscape of a $\frac{3}{2}a \times \frac{3}{2}a \times 2a$ unit-cell slab with a periodic array of $a \times a$ pores ($\beta\varepsilon = 10.5$), and the probability that a cluster of V subunits encloses at least one pore. The column is periodic in the z dimension, while the slab is periodic in the x and y dimensions. For both structures, the sharp increase in $p_{\text{enclosed}}(V)$ is accompanied by a decrease in the slope of $F(V)$ but is unrelated to the number of components in the fundamental unit cell, $V(G)$.

encloses at least one rectangular pore. The results of these calculations, which are relatively insensitive to the precise value of the bond strength, are shown in Fig. 3.

Our calculations show that the probability of completing each mesoscale feature increases sharply beyond a particular cluster size. Strikingly, we find that the dominant near-equilibrium assembly pathways are highly reproducible, despite the enormous number of potential on-pathway clusters. Furthermore, in both examples, the completion of these features occurs far beyond the initial nucleation barrier and deep into the growth phase of the crystal (*i.e.*, the regime where the free energy is steadily decreasing with increasing cluster size). There is also a significant change in the slope of $F(V)$ accompanying the completion of these features, confirming that the enclosing of a channel or a pore significantly enhances the thermodynamic driving force for subsequent growth. We stress that the cluster size at which a channel or pore is enclosed is unrelated to the number of components in the unit cell, and is in fact is considerably larger than the number of particles contained in a single unit cell. This example illustrates that, in order to predict the dominant growth pathway, it is necessary to consider maximally connected substructures that extend over multiple adjacent unit cells.

C. Design of hierarchical assembly pathways

The free-energy profile in Fig. 3b shows that several metastable clusters may be encountered during the self-assembly of a periodic structure consisting of unit cells with a complex topology. A free-energy landscape with these features is indicative of hierarchical self-assembly: since some of these clusters are separated by significant free-energy barriers, the growth of the target structure is likely to proceed *via* a sequence of discrete, partially assembled structures. By selectively stabilizing specific substructures, it is even possible to design a free-energy landscape with a globally stable phase that differs from the complete crystal in either composition or cluster size. The selective stabilization of on-pathway clusters thus enables the specification of experimental protocols in which the growing cluster is guided from one thermodynamically stable phase to the next. Below, we demonstrate how knowledge of the free-energy landscape of a structure with a bimodal bond-energy distribution can be used to design such hierarchical assembly pathways.

For both of the structures examined in the previous example, we choose a subset of bonds that stabilize maximally connected substructures, which correspond roughly to the local minima in homogeneous-bond-energy landscapes. These bonds, which are highlighted in Fig. 4a and b, are assigned bond energies $-\varepsilon_B$, while all other bonds are assigned $-\varepsilon_A$. We apply eqn (1) to predict the phase boundaries separating stable substructures, and then use the complete free-energy landscape to map the detailed phase diagrams shown in Fig. 4e and f. (For details, see ESI,† Section S4.)

In the case of the self-assembling column (Fig. 4a, c and e), assembly may proceed *via* compositionally distinct phases. By choosing ε_B to be significantly larger than ε_A, we assemble a structure consisting primarily of type-B bonds that does not enclose the interior channel (phase II). On the other hand, we predict that crystals assembled in the hatched region of the phase diagram will enclose this channel, as shown previously in the case $\varepsilon_A = \varepsilon_B$ (phase III). We also note that the height of the dominant nucleation barrier depends primarily on the strongest bonds, whereas the phase boundaries depend in general on both ε_A and ε_B. This means that rapid nucleation occurs closer to the phase boundary, *i.e.*, at lower supersaturation, in phase II, where only part of the complete structure is stable. We therefore expect to assemble the target crystal with fewer defects by following an annealing protocol that passes slowly through phase II before entering phase III.

In the case of the perforated slab (Fig. 4b, d and f), a large difference in the two bond energies leads to multiple nucleation barriers, as well as phases in which finite-sized clusters are the lowest-free-energy structures. In the region marked phase III, we find that a fluid of 30-particle substructures is thermodynamically stable. This phase is separated from the completely unassembled fluid (phase I) by a narrow region of stability (phase II) of 27-particles clusters, which correspond to 30-particle clusters missing one of their four 'arms'. Upon equilibration in both phases II and III, the free-energy landscapes indicate that a proliferation of clusters with type-B bonds will coexist with a mostly monomeric solution of the remaining unassembled components. In the hatched region of phase IV, the complete crystal is globally stable, although the first 30-particle substructure is kinetically separated from further growth by a large secondary nucleation barrier, $\Delta F^{\ddagger\ddagger}$. As a result, we predict that assembly in this region will proceed *via* a series of nucleation events, each having a well defined nucleation barrier. If this region is entered from phase III, then these nucleation events will correspond to the assembly of type-A 'linker' substructures between pre-assembled type-B clusters. It is important to point out that comparable behavior is not observed upon stabilizing the complementary substructure, such that $\varepsilon_A > \varepsilon_B$. Instead of monodisperse 24-particle clusters connected by type-A bonds, we find a stable phase (phase V) corresponding to a mixture of various substructures, including some that incorporate type-B bonds. This mixed phase arises because, in this case, the selected substructure is poorly stabilized by a low density of designed interactions, and, as a result, corresponds to a low-probability cluster on the homogeneous-bond-energy landscape. Thus, while it is in principle possible to stabilize any part of the unit cell with a judicious choice of bond energies, maintaining control over a multistep assembly pathway requires attention to the intrinsic stabilities of competing substructures.

IV. Discussion

Using a simple, graph-based approach, we have studied the factors than control the robust assembly of extended multicomponent structures consisting of many distinct building blocks. Below, we briefly summarize our main findings and then discuss some of the implications.

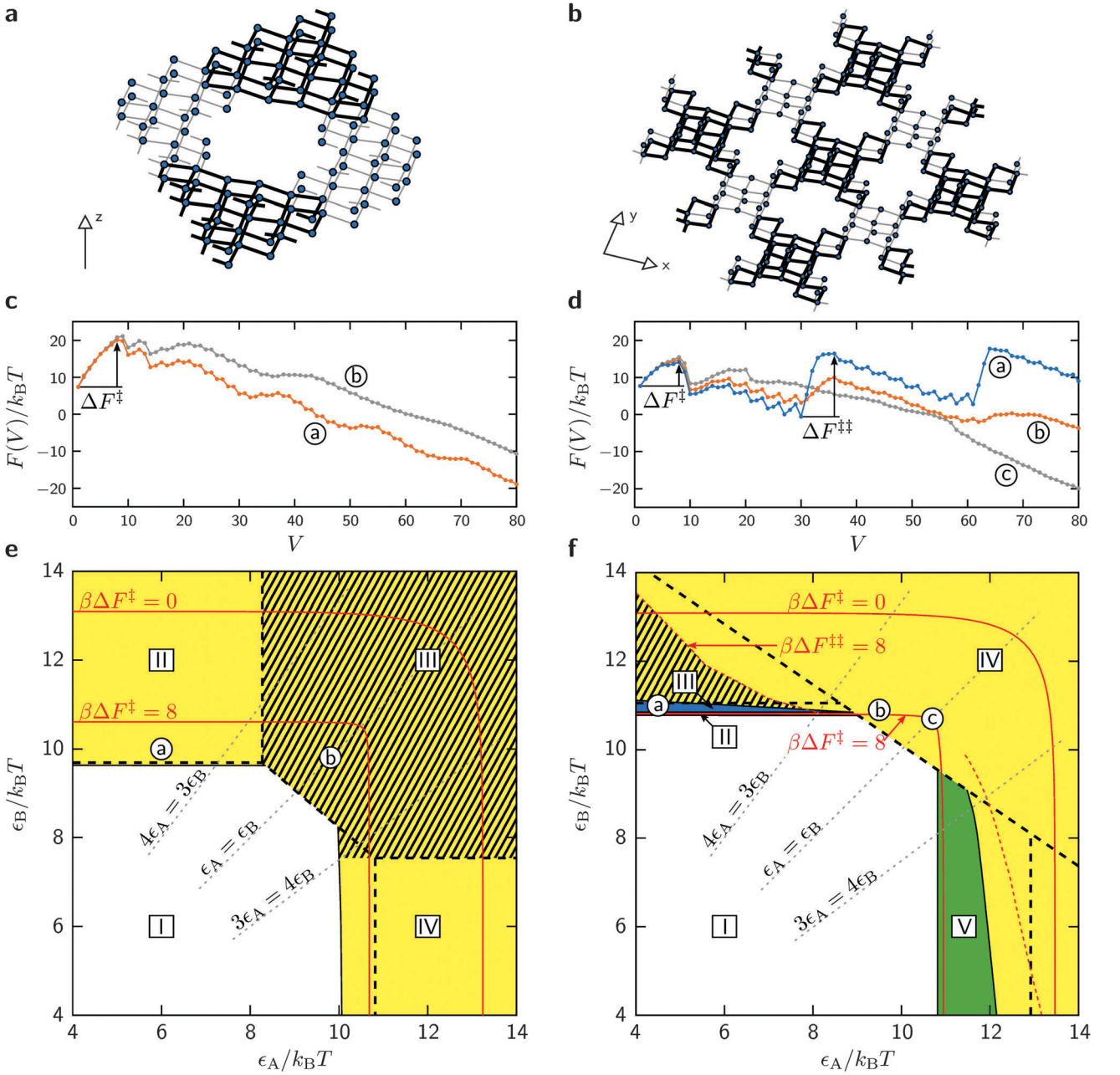

Fig. 4 Hierarchical self-assembly pathways can be designed by selectively strengthening bonds in an addressable crystal. (a and b) Connectivity graphs representing the structures examined in Fig. 3 with specific substructures highlighted; unterminated edges continue into adjacent unit cells. Darkened edges are assigned bond energies $-\varepsilon_B$, while all other edges are assigned bond energies $-\varepsilon_A$. (c and d) Representative free-energy landscapes showing both hierarchical and non-hierarchical assembly pathways, corresponding to the labeled points in the phase diagrams below. Depending on the bond energies, a secondary nucleation barrier, $\Delta F^{\ddagger\ddagger}$, may appear following the assembly of the stabilized substructure. (e and f) Phase diagrams showing the conditions under which various substructures are globally stable, and the predictions of a simple stability analysis (dashed lines). Lines of constant nucleation-barrier height are shown in red. In (e), the phases are as follows: I, unassembled fluid; II, infinite cluster with type-B bonds; III, complete crystal; and IV, infinite cluster with type-A bonds. In (f), the phases are: I, unassembled fluid; II, 27-particle cluster; III, 30-particle cluster; IV, complete crystal; and V, mixed clusters with fewer than 24 particles.

• We find that classical nucleation theory fails to describe the self-assembly of addressable multicomponent crystals. In particular, in addressable structures, nucleation and growth can be tuned independently.

• Magic-number clusters play a key role in the nucleation of addressable structures and can even be made stable.

• In addressable crystals, the thermodynamics of self-assembly is insensitive to the number of distinct components used to construct a periodically repeating structure, provided that this number is large enough to specify the geometry of the fundamental unit cell.

• Depending on the conditions, the same set of building blocks can form a variety of patterns that are substructures of

the complete lattice. Periodic structures of incompletely formed unit cells may be entropically favored over 'perfect' lattices.

• None of these findings are specific to structures with DNA-mediated bonds. These observations apply equally to addressable crystals that are periodic in all three dimensions, such as multicomponent zeolite-like structures.

Critical nuclei or magic clusters

In the case of the DNA-brick lattice considered here, the nucleation barrier is highly non-classical because the most stable intermediate structures contain specific ('magic') numbers of particles and are not spherical. This non-classical behavior is due to the fact that nucleation proceeds preferentially *via* maximally connected clusters that have the lowest free energy for a given number of particles. In particular, in structures with a low coordination number, the number of distinct, maximally connected clusters is small. Moreover, the initial increase in the free energy of pre-critical clusters is a result of the loss of translational entropy due to the formation of designed bonds and need not be proportional to the surface area of the cluster, as assumed by classical nucleation theory.

Choosing an optimal number of building blocks

In practice, not all interactions between subunits can be made perfectly specific. As a result, in a multicomponent solution with possible unintended interactions, a rate-limiting nucleation barrier may not be sufficient to guarantee on-pathway assembly. Misassembly due to aggregation is thus likely to limit the number of distinct components that can be accommodated in an addressable structure. In our calculations, the total concentration of monomers in the reference fluid is proportional to the number of distinct components in the unit cell. In order to incorporate additional components while maintaining a constant supersaturation, the total concentration of subunits in the solution must be increased, enhancing the probability of amorphous aggregation *via* incidental interactions. Alternatively, fixing the total concentration while increasing the designed bond energies may not be straightforward, due to the physical limitations of the molecular building blocks and the slower kinetics that are typically associated with stronger bonds.[18] These considerations suggest that a robust design should use the smallest number of building blocks that is consistent with the target structure.

Not one but many patterns

Although the assembly of the fully-formed target structure only occurs if the solution is supersaturated with respect to the defect-free, periodically repeated unit cell, there is always a regime where infinite structures of incompletely formed unit cells are thermodynamically stable, even though $\Delta F_G > 0$. Such structures are entropically favored due to the enormous number of distinct ways of partially assembling a semi-infinite crystal. These infinite substructures are similar to simple crystals with point defects, since the thermodynamic stability of the crystal is increased by leaving some components in each unit cell unassembled; however, this effect is much more significant in addressable periodic structures, in particular those with complex morphologies and heterogeneous bond energies. Accounting for this entropic effect is crucial for the design of hierarchical assembly pathways, since a stability analysis based on the complete unit cell significantly underestimates the region in which an infinite structure is thermodynamically favored (Fig. 4f). The ability to design a variety of periodic patterns using different mixtures of the same set of building blocks is a unique feature of structures with addressable complexity: this makes it possible to 'tune' the shape of the repeating pattern, such as an array of holes with controllable shapes, simply by changing the composition of the monomeric solution. The possibility to create such structures is directly related to the fact that partially assembled structures may be thermodynamically stable well before the perfect structure that contains all building blocks becomes the favored structure.

Robust self-assembly pathways

Our calculations suggest several means of speeding up robust self-assembly and controlling the assembly mechanism in complex, ordered structures. Because equivalent nucleation barriers can be achieved at lower supersaturation in semi-infinite crystals, it is likely that defect-free annealing is easier to achieve in complex structures than in the corresponding bulk crystal. We have demonstrated that the topology of the connectivity graph not only controls the nucleation barrier, but also determines the assembly pathway deep into the growth phase. Clusters spanning multiple unit cells, and thus containing more than one copy of some components, are required to assemble the key three-dimensional features at equilibrium, despite the fact that these features can all be formed, by definition, within a single unit cell. The dominant growth pathways for structures with complex morphologies correspond to predictable features of the free-energy landscape, implying that these pathways are designable as well.

Steering assembly to proceed *via* intermediate structures

We have also shown how to design a multi-stage assembly mechanism by exploiting the heterogeneity inherent in a multicomponent crystal. With control over the stabilities of partial structures and the nucleation barriers between them, one can propose a time-dependent protocol to guide assembly through several intermediate phases before arriving at the final crystal structure. Such a protocol might be as simple as changing the temperature, with the heterogeneous bond energies held constant. More exotic pathways might also be programmed using specific interaction free-energies that are themselves temperature-dependent, such as the hybridization free energies of complementary DNA strands.[28,29] Alternatively, we can optimize the assembly conditions to achieve step-wise assembly without using a time-dependent protocol by designing a free-energy landscape with multiple sequential nucleation barriers. While it is always possible to leave some subunits of a crystal completely disassembled by tuning the designed bond energies independently, careful attention to the underlying free-energy landscape is necessary to preferentially nucleate a particular finite-sized structure. Our example calculations demonstrate that it is advantageous to strengthen bonds that coincide with an intrinsically stable

substructure in order to establish a pathway to a target region of the phase diagram.

Outlook

We expect that the approach to hierarchical assembly described here will open up new routes to more intricate three-dimensional structures. Although hierarchical assembly is often associated with a greater tendency for aggregation,[30–32] we speculate that kinetic traps might be avoided by preventing large, fully assembled substructures from interacting directly: the presence of other building blocks that link the substructures together is essential. In our example of multi-step assembly, the growth pathway must cross a secondary nucleation barrier involving linker subunits that assemble reversibly. With control over the bond strengths in the linker region, in addition to the overall structure topology, it is straightforward to apply the predictive tool presented here to optimize the kinetics at this stage of assembly as well.

Experimentally, this strategy for hierarchical assembly is currently realizable using DNA bricks. In this system, the hybridization free energies between subunits can be specified independently by rationally designing the complementary DNA sequences. Unlike the single optimal assembly temperature reported in ref. 11, we expect that an assembly protocol that passes through a stable cluster phase before forming the complete crystal would exhibit two distinct transition temperatures. Furthermore, the strategies described here for lowering the nucleation barrier near the fluid-crystal phase boundary are expected to reduce the hysteresis in the observed freezing and melting temperatures. As a result, we expect that such strategies will broaden the temperature range over which high quality crystals can be assembled. These predictions can also be tested with physically realistic Monte Carlo simulations using the model described in ref. 22 and 24. Through the analysis of dynamic assembly trajectories, simulations of this type could be used to validate the predicted growth pathways in structures with complex morphologies.

Finally, we stress once more that the general principles described in this paper are in no way limited to DNA bricks. These principles could also apply to other forms of multicomponent molecular assembly, or to the assembly of multicomponent colloidal structures, provided that the building blocks have been designed to reproduce the functionality assumed in our model.

Acknowledgements

We would like to thank Aleks Reinhardt for numerous discussions and helpful comments on the manuscript. This work was carried out with support from the European Research Council (Advanced Grant 227758) and the Engineering and Physical Sciences Research Council Programme Grant EP/I001352/1. Supporting data related to this publication are available at https://github.com/wmjac/pygtsa.

References

1 E. Winfree, F. Liu, L. A. Wenzler and N. C. Seeman, *Nature*, 1998, **394**, 539.
2 P. W. K. Rothemund and E. Winfree, *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, ACM, 2000, pp. 459–468.
3 A. V. Tkachenko, *Phys. Rev. Lett.*, 2011, **106**, 255501.
4 J. D. Halverson and A. V. Tkachenko, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2013, **87**, 062310.
5 L. O. Hedges, R. V. Mannige and S. Whitelam, *Soft Matter*, 2014, **10**, 6404.
6 A. Murugan, J. Zou and M. P. Brenner, *Nat. Commun.*, 2015, **6**, 6203.
7 B. Wei, M. Dai and P. Yin, *Nature*, 2012, **485**, 623.
8 Y. Ke, L. L. Ong, W. M. Shih and P. Yin, *Science*, 2012, **338**, 1177.
9 Z. Zhang, J. Song, F. Besenbacher, M. Dong and K. V. Gothelf, *Angew. Chem., Int. Ed.*, 2013, **52**, 9219.
10 B. Wei, M. Dai, C. Myhrvold, Y. Ke, R. Jungmann and P. Yin, *J. Am. Chem. Soc.*, 2013, **135**, 18080.
11 Y. Ke, L. L. Ong, W. Sun, J. Song, M. Dong, W. M. Shih and P. Yin, *Nat. Chem.*, 2014, **6**, 994.
12 D. Frenkel, *Nat. Mater.*, 2015, **14**, 9.
13 L. Cademartiri and K. J. M. Bishop, *Nat. Mater.*, 2015, **14**, 2.
14 N. Geerts and E. Eiser, *Soft Matter*, 2010, **6**, 4647.
15 J. W. Gibbs, *Am. J. Sci.*, 1878, **16**, 441.
16 D. W. Oxtoby, *J. Phys.: Condens. Matter*, 1992, **4**, 7627.
17 R. P. Sear, *J. Phys.: Condens. Matter*, 2007, **19**, 033101.
18 S. Whitelam and R. L. Jack, *Annu. Rev. Phys. Chem.*, 2015, **66**, 143.
19 A. George and W. W. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1994, **50**, 361.
20 P. R. ten Wolde and D. Frenkel, *Science*, 1997, **277**, 1975.
21 A. J. Kim, R. Scarlett, P. L. Biancaniello, T. Sinno and J. C. Crocker, *Nat. Mater.*, 2008, **8**, 52.
22 A. Reinhardt and D. Frenkel, *Phys. Rev. Lett.*, 2014, **112**, 238103.
23 W. M. Jacobs, A. Reinhardt and D. Frenkel, *J. Chem. Phys.*, 2015, **142**, 021101.
24 W. M. Jacobs, A. Reinhardt and D. Frenkel, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 6313.
25 S. C. Glotzer and M. J. Solomon, *Nat. Mater.*, 2007, **6**, 557.
26 Y. Tian, T. Wang, W. Liu, H. L. Xin, H. Li, Y. Ke, W. M. Shih and O. Gang, *Nat. Nanotechnol.*, 2015, **10**, 637.
27 Y. Zhang, S. Pal, B. Srinivasan, T. Vo, S. Kumar and O. Gang, *Nat. Mater.*, 2015, **14**, 840.
28 J. SantaLucia Jr and D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.*, 2004, **33**, 415.
29 R. T. Koehler and N. Peyret, *Bioinformatics*, 2005, **21**, 3333.
30 T. K. Haxton and S. Whitelam, *Soft Matter*, 2013, **9**, 6851.
31 S. Whitelam, 2015, arXiv preprint arXiv:1505.07501.
32 S. H. Park, C. Pistol, S. J. Ahn, J. H. Reif, A. R. Lebeck, C. Dwyer and T. H. LaBean, *Angew. Chem.*, 2006, **118**, 749.

