Perspective: Self-assembly of structures with addressable complexity

William M. Jacobs[1] and Daan Frenkel[2]

[1] *Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02139, USA*
wjacobs@fas.harvard.edu

[2] *Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*
df246@cam.ac.uk

(Dated: 4 February 2016)

ABSTRACT

The self-assembly of structures with 'addressable complexity' — where every component is distinct and is programmed to occupy a specific location within a target structure — is a promising route to engineering materials with precisely defined morphologies. Because systems with many components are inherently complicated, one might assume that the chances of successful self-assembly are extraordinarily small. Yet recent advances suggest otherwise: addressable structures with hundreds of distinct building blocks have been designed and assembled with nanometer precision. Despite this remarkable success, it is often challenging to optimize a self-assembly reaction to ensure that the intended structure is kinetically accessible. In this Perspective, we focus on the prediction of kinetic pathways for self-assembly and implications for the design of robust experimental protocols. The development of general principles to predict these pathways will enable the engineering of complex materials using a much wider range of building blocks than is currently possible.

I. INTRODUCTION

Most man-made materials that have been designed to form ordered, self-assembling structures are composed of only a handful of distinct components. For example, colloidal crystals rarely contain more than two distinct kinds of building blocks. While our ability to understand and control such materials is a remarkable achievement of soft matter science, the level of complexity that can be achieved is still limited by our ability to organize matter precisely at the smallest scales. The limited 'design space' of these few-component materials is obvious when compared with biological materials such as multi-protein or protein–RNA complexes, which are typically assembled from a large number of components and consequently have a huge diversity of form and function.

In the past few years, significant advances in the field of self-assembly have begun to rival the complexity of biological structures (Figure 1). Ground-breaking experiments performed by the Yin group have demonstrated that it is possible to design and self-assemble three dimensional objects out of hundreds of distinct components.[1–3] In these experiments, the building blocks are individual pieces of single-stranded DNA that can hybridize according to Watson–Crick base-pairing. In order to stabilize a particular three-dimensional structure, complementary DNA sequences are arranged to form highly specific bonds between DNA 'bricks' that are adjacent in the target structure.[4,5] Using 32-nucleotide bricks, structures with dimensions on the order of tens of nanometers have been assembled with intricate three-dimensional features, such as channels, pores and cavities. More recently, this strategy has been extended to construct periodic materials with equally intricate multicomponent unit cells.[6]

Meanwhile, DNA-origami experiments have rapidly progressed to the point where three-dimensional structures can be assembled with nearly 99% yield.[7–10] This approach also exploits base-pairing complementarity to stabilize a specific structure, although it differs from the DNA-brick experiments by incorporating a long 'backbone' strand that serves as a scaffold for self-assembly.[11] The Dietz group has recently shown that DNA origami, once assembled, can themselves be arranged into yet larger structures by designing specific interactions between the origami building blocks.[12] These experimental demonstrations promise to revolutionize the field of self-assembly. However, in order to realize the full potential of such materials and to move beyond systems that only contain DNA-based building blocks, we need to understand the factors that control the robust self-assembly of complex, multicomponent structures.

DNA-brick and DNA-origami structures represent a particularly important class of multicomponent structures: their remarkable complexity is due to the fact that the structures are completely 'addressable.'[13] Since every building block within an aperiodic addressable structure (or unit cell in a periodic addressable structure) is unique, it is possible to specify exactly where a particular subunit will be located within the target structure.[14,15] Addressable complexity therefore allows for precision at scales approaching the size of a single subunit. Furthermore, addressable complexity makes it possible to functionalize particular locations on the designed structure, enabling specificity of function as well as shape. It is not difficult to imagine how a completely synthetic enzyme might be assembled with this technology, once generalized to an appropriate set of chemical building blocks.

Addressable assembly offers a number of crucial experimental advantages over traditional techniques for synthe-
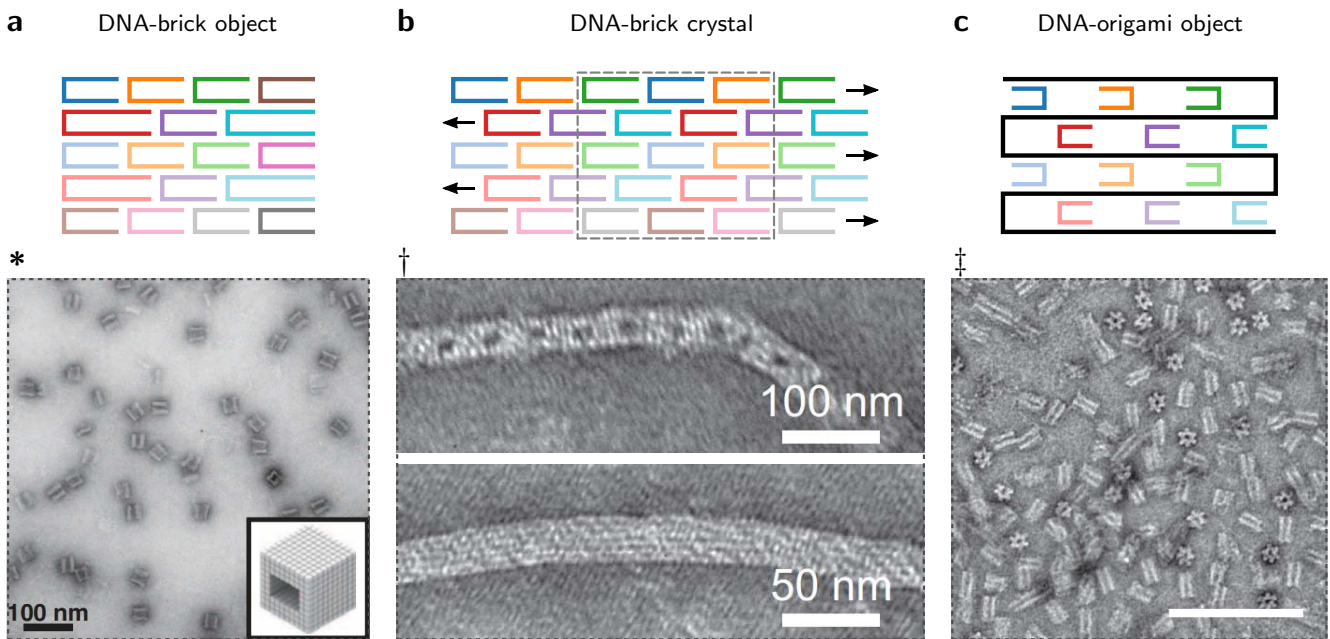
FIG. 1. **DNA-based addressable nanostructures.** (a) Finite-sized, aperiodic structures constructed from single-stranded DNA bricks. (b) Periodic DNA-brick structures in which every brick within the unit cell (indicated by the dashed rectangle in the schematic diagram) is unique. (c) DNA-origami structure in which the backbone scaffold (represented by the black line in the schematic diagram) is folded by single-stranded DNA staples. Representative TEM images are shown below. *From *Science* **338**, 1177 (2012). Reprinted with permission from AAAS. †Adapted by permission from Macmillan Publishers Ltd: *Nature Chemistry* **6**, 994, copyright 2014. ‡The scale bar indicates 200 nm. From *Science* **338**, 1458 (2012). Reprinted with permission from AAAS.

sizing and patterning materials. Because the information required for self-assembly is encoded locally by highly specific interactions between neighboring subunits, no templating is required to seed the growth of the target structure. Self-assembly can be carried out in a one-pot reaction, where the designed structure forms spontaneously from a solution in which all building blocks are simultaneously present. The local nature of the interactions means that it is also possible to construct an arbitrary design out of a pool of potentially reusable building blocks. For instance, experiments on DNA bricks have shown that a wide variety of structures can be assembled from the same set of strands, depending on which species are included in the solution:[2] shapes can be created by leaving out certain building blocks, rather than by adding more. This principle of modularity means that it is not necessary to redesign a new set of subunits from scratch in order to construct a completely different structure.[16]

A more subtle but equally important point is that addressable complexity makes it possible to design the *kinetics* of a self-assembly reaction. Designing a structure solely on the basis of its thermodynamic stability is typically insufficient to ensure high-yield self-assembly. In fact, it is far from obvious that multicomponent self-assembly should be successful at all. Conventional wisdom for a long time held that one-pot assembly techniques were doomed due to imperfect control over subunit stoichiometries and the many opportunities for in-

correct binding.[17,18] Whilst it is clear that these fears were warranted — attention to kinetics is essential for assembling multicomponent target structures without errors — it is nevertheless possible to design systems such that a robust pathway to the desired structure exists. From this point of view, the key achievement in experiments on DNA bricks and DNA origami is not the clever use of base-pairing to stabilize a particular structure, but rather the ability to guide the many subunits into place in a robust and controllable manner.

Is it possible to reproduce this success using building blocks other than single-stranded DNA? Computer simulations with simplified models suggest that the basic principles of addressable self-assembly may indeed be transferable to other building blocks.[19–21] This is important because assembling addressable structures out of DNA superstructures,[22] functionalized nanocolloids[23] or colloidal patchy particles[24–27] would allow us to access larger structures and to tailor materials to a wider variety of applications. Yet to do so, we will need to improve our understanding of the microscopic mechanisms at work in existing DNA-based systems. General principles can then be developed to guide the design of complex self-assembly reactions.

In this Perspective, we discuss the open challenges in predicting and controlling the self-assembly of systems with addressable complexity. We describe recent theoretical advances[21,28–33] that provide insights into the phys-
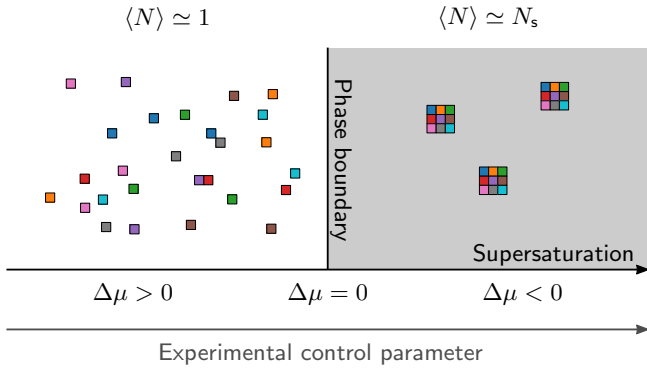
FIG. 2. **A phase boundary separates the unassembled and assembled states.** In general, the average value of an order parameter that discriminates between these two states changes abruptly across the phase boundary. In the schematic shown here, the average cluster size, $\langle N \rangle$, is equal to that of the complete structure, N_s, in the assembled phase. An experimental control parameter, such as temperature, is used to control the supersaturation, $\Delta\mu$, of the initially unassembled phase.

ical forces that determine the kinetic pathways for self-assembly and help to explain the success of DNA-brick and DNA-origami experiments. We also highlight new results that suggest how to optimize addressable structures and experimental protocols. We conclude by addressing promising future directions and a number of key questions in this rapidly evolving field.

II. DESCRIBING PATHWAYS FOR SELF-ASSEMBLY

Owing to the very large number of particles involved, self-assembly reactions are aptly described using the language of phase transformations. For all of the systems that we consider here — periodic and aperiodic addressable structures, as well as foldable polymers such as DNA origami — a mixture of particles spontaneously self-assembles into a highly organized configuration. Because this transformation is often highly cooperative (in the sense that many particles must interact in order for the self-assembly reaction to advance), the probability of finding the system in the assembled state changes sharply with the supersaturation (Figure 2). Small changes to an experimental control parameter, such as the temperature, can shift the equilibrium between the disordered and assembled states drastically, even if the binding curve for two monomers has a more gradual dependence on the control parameter. This effect contributes to the generic sensitivity of the experimental conditions where self-assembly can take place.

As in the case of monodisperse self-assembly, the disordered and assembled states of an addressable structure are separated by a large number of very low probability intermediate configurations. Yet multicomponent self-assembly can typically proceed via an enormous number

of distinguishable pathways through these intermediate states, making the determination of the mechanism of phase transformation more difficult. While the diversity of parallel routes may be daunting, it is nevertheless possible to describe the sequence of key collective events in a manner that enables mechanistic predictions and rational design.

For this purpose, an appropriately defined free-energy landscape is an essential tool for identifying the key events along the dominant pathway for phase transformation. Such a landscape plays the role of the familiar potential energy surface for simple reactions. However, the coordinates on a free-energy landscape are collective variables that measure the degree of assembly of the target structure along any of the (potentially numerous) parallel microscopic pathways. This approach reduces the complexity of a system with very many degrees of freedom to a small number of order parameters. Furthermore, a description in terms of free energies is much simpler than a chemical kinetics approach in which the equations describing the very large number of microscopic reactions are solved explicitly. Most importantly, a free-energy landscape can immediately reveal the kinetic bottlenecks and the collective behavior of a large number of reacting components. As in the case of a potential-energy surface, a complex self-assembly reaction is most likely to evolve via the lowest free-energy path. However, the two descriptions are very different: a low free-energy path may be favorable not because it corresponds to low potential-energy pathways, but because there are many distinct pathways. In the latter case, the self-assembly pathway on the free-energy surface is *entropically* favored. Because a self-assembly pathway must pass through many low-probability intermediate states, an unassembled system typically must be driven out of equilibrium in order for phase transformation to occur. The lowest free-energy path is a good approximation for describing the reaction dynamics provided that the conditions are suitably mild, such that the local degrees of freedom evolve faster than the order parameters that measure the progress of the self-assembly reaction.

It is convenient to relate a description of complex self-assembly to the standard theory for describing the dynamics of crystallization, classical nucleation theory (CNT)[34–36]. In Figure 3, we consider a system of identical particles that are initially in an unassembled state. The conditions are then changed, for instance by lowering the temperature, such that the assembled state becomes thermodynamically favored. If the unassembled phase is only moderately supersaturated beyond the coexistence boundary, it will not immediately convert to the assembled state. Instead, the now *metastable* unassembled phase will persist until a spontaneous fluctuation carries the system across the low probability intermediate states that separate the two phases. CNT predicts that this rate-limiting bottleneck for phase transformation can be described by a single order parameter that measures the size of the largest cluster in the system.
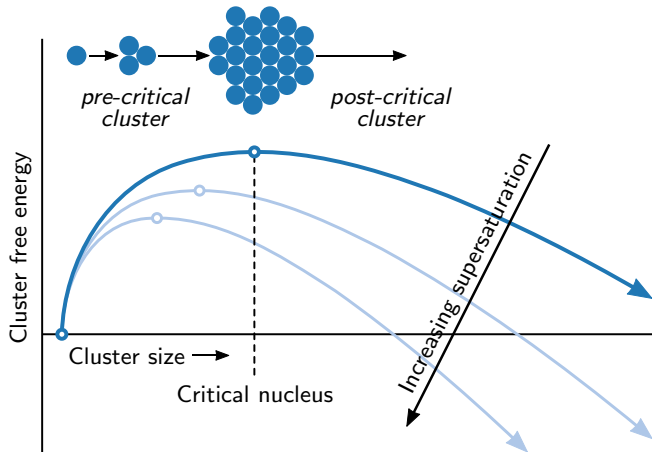
FIG. 3. **Classical nucleation theory predicts a nucleation barrier for the assembly of a simple crystal.** The free energy as a function of cluster size first increases due to the formation of an interface between a crystalline cluster and the surrounding solution before decreasing due to the thermodynamic stability of the bulk crystal phase. The competition between these surface and bulk free energies results in a free-energy barrier for nucleation. Both the height of the barrier and the size of the critical nucleus diminish with increasing supersaturation.

The existence of this bottleneck for crystallization is the consequence of a 'tug-of-war' between the two thermodynamic forces that antagonistically promote and oppose crystallization. On the one hand, the driving force for phase transformation is provided by the chemical potential difference between the two phases. Opposing this driving force is the free-energy cost to create an interface between a small crystalline cluster and the surrounding unassembled phase. The competition between these two forces gives rise to a free-energy barrier that must be overcome in order for a crystal to nucleate. CNT predicts that clusters smaller than the critical nucleus (the structure corresponding to the top of the barrier) are more likely to disassemble, while larger clusters are more likely to grow without bound. If nucleation events are sufficiently rare, then the nucleation rate is predicted to have an Arrhenius dependence on the free energy of the critical nucleus. Furthermore, CNT predicts that both the size of the critical nucleus and the height of the nucleation barrier shrink with increasing supersaturation (Figure 3). There are many cases where CNT is known to be inadequate, even for monodisperse systems.[37–42] However, if the free-energy landscape can be predicted, then this information can be used to determine whether nucleation and growth is a plausible mechanism for phase transformation.

A free-energy description of phase transformation has direct consequences for the development of effective experimental protocols. For instance, the presence of a nucleation barrier strongly affects the experimental conditions under which self-assembly can take place on a reasonable time scale. Likewise, the degree of supersaturation is correlated with the rate at which the target structure grows. This is important information, because high growth rates favor the formation of defects, even in simple crystals. At very high supersaturation, phase transformation is likely to take place via rapid spinodal decomposition instead of slow, orderly growth. The point at which the nucleation barrier disappears therefore sets a limit on the supersaturation at which a material can be assembled with few defects; beyond this point, diffusion plays a more important role than the interparticle interactions in determining the structure of the final material. This paradigm from conventional crystallization is likely to hold true for addressable structures, although the consequences for defects may be more significant: in addition to defects such as vacancies and dislocations that can be found in simple crystals, errors in the ordering of addressable subunits are likely to occur under harsh growth conditions and are more difficult to correct via thermal annealing.

Most importantly, the kinetic pathways that connect the unassembled and assembled phases determine the *robustness* of an experimental protocol: that is, whether a suitable pathway exists for variations in the experimental conditions. In some cases, a kinetically accessible pathway may not exist for all structure designs, or self-assembly may only be allowed in a very narrow window of experimental conditions. Being able to predict and control the range of conditions over which a material can be self-assembled therefore has considerable practical importance for experiments.

III. SPECIFIC BINDING IS INSUFFICIENT: LESSONS FROM DNA-COATED COLLOIDS

Colloidal self-assembly provides a prototypical system for studying phase transformations mediated by specific interactions.[43,44] Since their invention twenty years ago,[45] DNA-coated colloids (DNACCs) have received much attention due to the ease with which specific interactions can be engineered.[46–53] By grafting colloidal particles with DNA strands that expose single-stranded 'sticky ends,' DNACCs can be made to aggregate or crystallize through the hybridization of complementary sequences (Figure 4a–b). With control over the specific binding between colloids, it is possible to program a low-energy assembled state. Experiments with nano[54–60] and micron-sized[61,62] DNACCs have confirmed that a variety of crystal structures can be stabilized with one or two pairs of complementary sticky ends. Furthermore, recent experiments on micron-sized DNACCS indicate that crystallization under conditions of mild supersaturation is compatible with the predictions of classical nucleation theory.[62]

The self-assembly of truly multicomponent, colloidal structures would open the door to an entirely new class of functional nano-devices. However, at present, colloidal
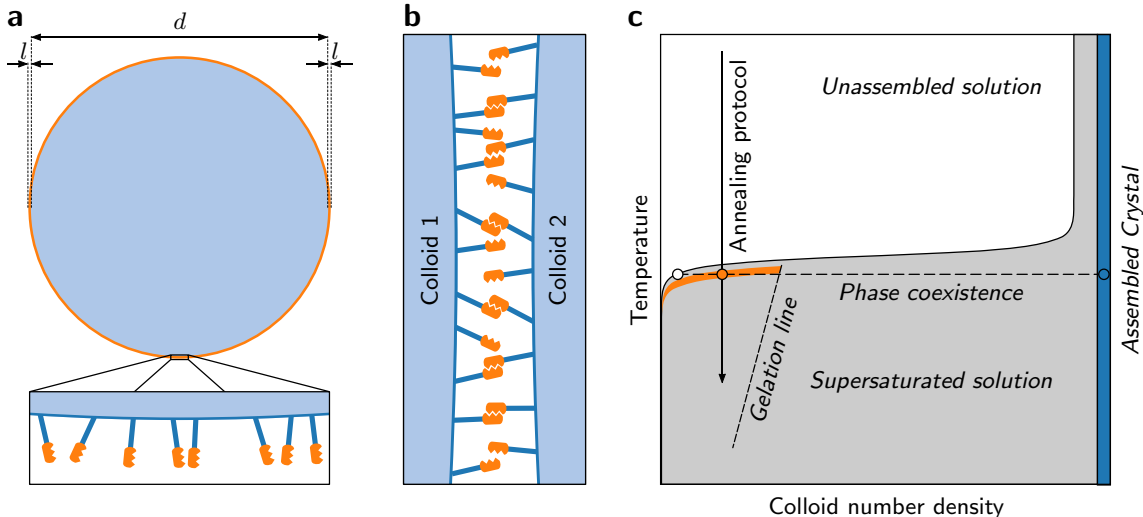
FIG. 4. **For DNA-coated colloids, specific binding does not ensure robust self-assembly.** (a) A schematic diagram of a micron-sized colloid. The grafted DNA strands with length l are much shorter than the particle diameter d. (b) The short-ranged attraction between two colloidal particles depends on the multivalent binding of complementary single-stranded DNA 'sticky ends' (orange tags). (c) This schematic phase diagram shows a narrow 'window' (orange region) for crystal nucleation and growth under conditions where the solution is supersaturated (gray region). When a low-density solution of initially unassembled colloids is cooled according to an annealing protocol, it enters a region of coexistence between the high-density crystal and low-density monomeric phases. Phase equilibrium is indicated by the dashed tie line connecting points on the boundaries of the coexistence region.

crystals contain very few distinct building blocks. While complex structures can in principle be designed,[63–65] they may not assemble for purely kinetic reasons. Controlling kinetic pathways in self-assembling materials with specific interactions is therefore crucial if colloidal self-assembly is to achieve the same level of complexity as DNA bricks or DNA origami.

Much of the difficulty in achieving high quality crystals is due to the narrow temperature window where crystallization can occur.[66,67] When two DNACCs approach within a distance comparable to the length of the grafted DNA, many sticky ends become available for binding.[68–72] As a result, the attractive interaction between a pair of colloids is extremely sensitive to temperature: typical binding curves are only about 1–2 °C wide. Consequently, it is often challenging to find experimental conditions where binding is strong enough to promote nucleation, yet weak enough to permit low-defect growth (Figure 4c). While this sensitivity is a generic combinatorial effect[73,74] that can be predicted quantitatively,[75,76] techniques for counteracting it have mostly relied on techniques that are specific to DNA hybridization. For instance, competition between partially complementary sequences[77,78] and the implementation of strand-displacement reactions with additional linker strands[79] have been shown to broaden the temperature window for crystallization. Increasing the DNA grafting density has also been shown to improve the kinetics of self-assembly by enhancing annealing at higher supersaturation.[62]

An important lesson is that the ability to design a low-energy structure does not guarantee that it can be self-assembled robustly. Especially with spherical particles that do not have a preferred direction for binding, the formation of disordered aggregates may be kinetically favored over the desired phase transformation. In special cases, there may be no viable pathway between a dilute unassembled solution and the target structure due to the presence of an intervening thermodynamically stable phase. For instance, simulations have shown that low-concentration solutions cannot crystallize if the unassembled phase does not coexist with the target crystal phase[80,81] or a liquid phase is favored instead.[82,83] These examples point to the importance of designing for both thermodynamic stability and kinetic accessibility in self-assembly reactions where we wish to control the precise placement of hundreds or thousands of distinct subunits.

IV. THEORETICAL APPROACHES TO ADDRESSABLE SELF-ASSEMBLY

In order to extend addressable self-assembly to a wider range of materials, it is essential that we find ways to control self-assembly pathways without relying on properties that are specific to DNA. With this goal in mind, we must ask, "What are the transferable features of DNA bricks and DNA origami?" Crucially, addressable self-assembly requires building blocks that support specific interactions, which are necessary to encode a unique target structure. Since correct bonds inevitably compete

with a large number of incorrect (or nonspecific) bonds, the designed bonds between putative neighbors must, on average, be much stronger than random interactions between subunits in the disassembled state.

Another key feature of single-stranded DNA is that all bonds are saturating: that is, a single sequence can typically only hybridize with at most one other strand. Saturating bonds provide control over both subunit valence and, to a lesser extent, the directionality of interactions between subunits. Limited-valence bonding is essential for the self-assembly of complex structures to succeed at finite concentrations. In particular, reducing the valence of specific (boundary) subunits makes it possible to terminate the growth of a finite-sized structure. For the same reason, saturating bonds can be used to stabilize structures that are not convex (something that is impossible in, say, conventional crystallization). Directional, limited-valence interactions also reduce the number of competing structures that tend to lower the yield: when all subunits make directional interactions, it is typically necessary to form multiple incorrect bonds in order to incorporate a single subunit in an incorrect position. This is an example of error-correction by 'peer pressure.'[84,85] Limited-valence interactions also provide greater control over the kinetics of self-assembly, as we shall discuss below. Strictly speaking, saturating bonds are not a requirement in the case of foldable polymers, although DNA origami do indeed have this property.

Given these properties of the individual building blocks, we can apply a recently developed a theory[21,30,31] to predict the free-energy landscape of an addressable structure. Assuming that the designed interactions completely determine the pathways for self-assembly, we can relate features of the assembly pathways to properties of the target structure, including the geometry and concentrations of the subunits as well as the topology and bond strengths of the designed interactions. As a result, this theoretical approach is capable of making *detailed* predictions for *specific* target structures. In order to apply this method to describe DNA-based self-assembly, the individual bond strengths of complementary sequences can be predicted using standard thermodynamic models of DNA hybridization.[86,87]

In the following sections, we build on the principles of classical nucleation theory to describe the self-assembly of addressable structures. However, since addressable building blocks support the design of arbitrarily complex structures, these free-energy landscapes may not look classical at all! We focus here on mechanisms of nucleation and growth by monomer addition, since the probability of correctly merging clusters in a multicomponent system is typically very small.[19] The natural order parameter for defining the free-energy landscape is thus the number of assembled subunits. To illustrate a number of generic principles of addressable self-assembly, we sketch a series of simplified phase diagrams that are based on predicted free-energy landscapes (Figure 5). Where possible, we suggest analogies with conventional, and con-

ceptually simpler, systems that serve to guide our intuition: monodisperse crystallization, micelle formation, and the folding of small, two-state proteins. These comparisons point to rational design strategies for both structure and protocol optimization.

A. Crystals with complex unit cells

In an addressable crystal, every subunit within the unit cell is distinct. However, as unit cells are repeated periodically, an extended structure will contain many copies of the same building block. The compositional heterogeneity of the unit cell and the low valence of specific subunits make it possible to design crystals that are periodic in only one or two dimensions or that have non-convex repeating patterns. If all particle concentrations are roughly equivalent and all stabilizing interactions have similar strengths, then the phase diagram looks qualitatively like that of a simple crystal. In order to make a comparison with Figure 4c, we ignore the fact that all species are, in actuality, distinct and simply consider the total concentration of the subunits. As with simple crystals, the stability of the structure depends on both the interaction strengths and the chemical potentials of the subunits, which are very nearly proportional to the logarithms of the subunit concentrations.

The formation of the target phase is unlikely to occur precisely at the phase boundary, where the height of the nucleation barrier diverges. (The phase boundary is indicated by the solid line separating the stable unassembled phase from the supersaturated solution in Figure 4c.) Instead, a self-assembly reaction must take place at a finite supersaturation where the height of the nucleation barrier is reduced. Yet unlike conventional crystallization, nucleation in addressable systems can follow an enormous number of distinct microscopic pathways.[30] There is thus a diversity of structures at the top of the nucleation barrier: the critical nucleus does not correspond to a single unstable structure, but rather to an ensemble of structurally similar clusters. A free-energy analysis can identify the common features of these critical clusters — such as the total numbers of particles and inter-subunit bonds formed at the rate-limiting step along the lowest free-energy path — in addition to indicating how many parallel routes are possible.[31]

As in the case of DNACCs, the optimal conditions for crystallization typically reside within a narrow window inside the coexistence region. At lower supersaturation, nucleation is too slow, while at greater supersaturation, long annealing times due to the slow unbinding of incorrectly bonded particles impede the growth of the target structure. A nucleation event within this assembly window will allow the system to relax to equilibrium, at which point an assembled crystal will coexist with a low-concentration solution of free monomers (Figure 4c). A one-step self-assembly reaction is thus possible, since only a single nucleation event is required for phase transfor-
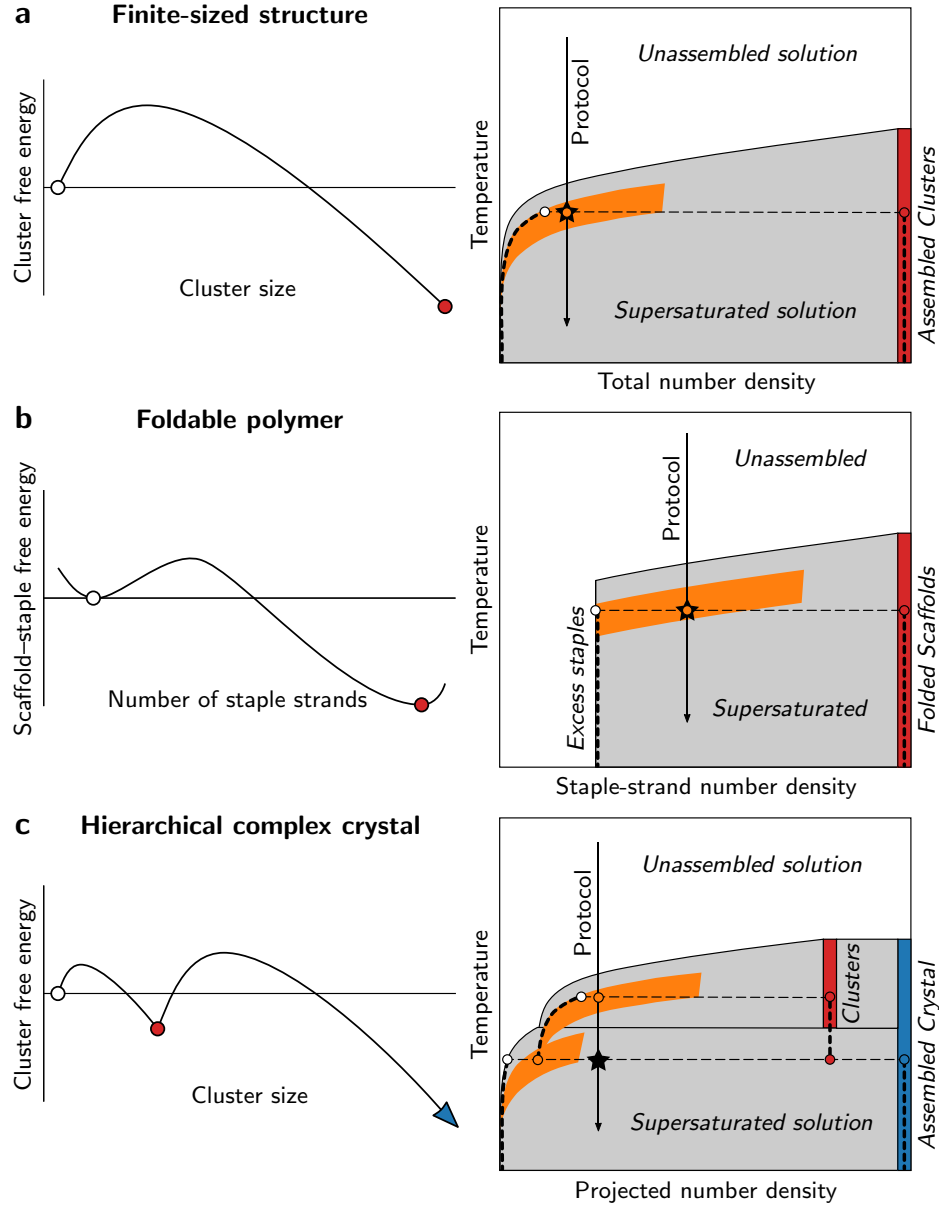
FIG. 5. **Predicted phase diagrams and free-energy profiles for three classes of addressable structures.** For each class of structures, the schematic free-energy surface on the left corresponds to the conditions marked by a star ($\star$) on the phase diagram. (a) In contrast with simple crystals, finite-sized structures require one nucleation event per copy. Consequently, the self-assembly reaction does not go to the equilibrium coexistence curve (solid black line), but rather stops at the boundary of the assembly window (orange region). Further cooling drives the formation of new assembled clusters (See Sec. IV B); this is indicated on the phase diagram by the heavy dashed curve. (b) The folding of nearly all polymer scaffolds in a solution can be carried out at constant temperature in the presence of excess staple strands. As with crystallization, nucleation is likely to require some degree of supersaturation below the coexistence curve (See Sec. IV C). (c) A similar phase diagram can be drawn for a complex crystal by projecting the number densities of the two types of particles: those that form stable clusters at intermediate temperatures, and those that link the clusters to form the complete crystal (See Sec. IV D). Following the indicated protocol, hierarchical self-assembly proceeds in two distinct stages: first cluster production, then crystallization within the second nucleation window. The heavy dashed curve indicates the number density of monomers in each phase as the cooling protocol progresses.

mation. (Multiple nucleation events may of course occur if the parent solution is quenched too deeply.) If the optimal nucleation window can be determined, it is possible to carry out self-assembly at a constant temperature. In this case, the reaction stops once the monomer concentration reaches the equilibrium coexistence curve, and the existence of defects in the assembled crystal is likely to depend on both the quench depth and the equilibrium

vacancy concentration. However, as in conventional crystallization, it may in practice be better to initiate nucleation at one temperature and then allow growth to proceed more slowly at a higher temperature, where defect formation is less likely.

B. Finite-sized, aperiodic structures

In the case of finite-sized, aperiodic structures, the assembled phase is a solution of discrete clusters. This presents the potential problem of 'monomer starvation:' if most of the available subunits are tied up in small partial structures, then the complete assembly of any single copy of the target structure may be slowed dramatically. Without free monomers in solution, error-free growth might proceed via Ostwald ripening, whereby monomers detach from one cluster and reattach to a larger cluster, or the merging of precisely compatible partial structures. However, simulations indicate that such mechanisms are unlikely to succeed.[19] It is also essential to avoid aggregation as a result of interactions between partial structures, through either correct or incorrect bonds, under conditions where the timescale for unbinding is comparable to that of cluster diffusion. Great care must therefore be taken to avoid the rapid proliferation of partially assembled structures. It is even conceivable that vulnerable partial structures might be protected by the introduction of weakly binding 'chaperones.' To our knowledge, such an approach has not yet been explored.

These considerations suggest that nucleation should be the rate-limiting step along the assembly pathway, both to prevent the proliferation of incompatible partial structures and to promote assembly via monomer addition. Following the crystallization paradigm, Figure 5a shows an optimal self-assembly window, where rate-limiting nucleation and conditions of mild supersaturation promote error-free growth of the target structure. However, unlike the case of a periodic structure, the conditions for optimal assembly shift as clusters are produced. The reason is that one nucleation event is required per copy of the structure, but the nucleation rate depends strongly on the concentration of the available monomers. If we again ignore the fact that there are many distinct components, then the production of assembled structures at constant overall concentration can be modeled by analogy with micelle formation. At the critical micelle concentration, an equilibrium between cluster disassembly and nucleation is reestablished after clusters are initially produced by a supersaturated solution. The equilibrium population of clusters can therefore be calculated at any point within the coexistence region by means of a common-tangent construction (Figure 6). This calculation shows clearly that the nucleation barrier *increases* as the mixture equilibrates. (The common tangent construction is not precisely correct in a multicomponent system since the clusters may have different compositions, but the analogy with micelle formation is useful nonetheless.) In
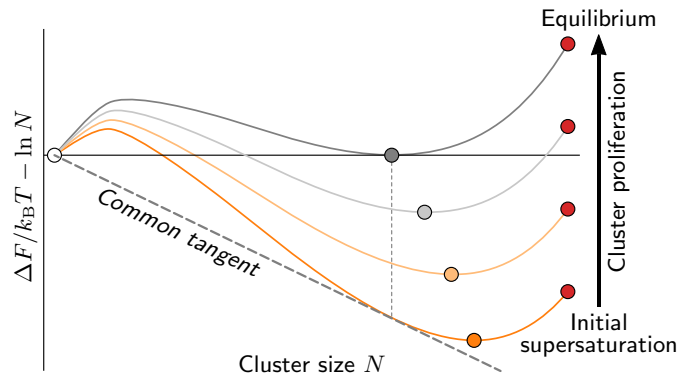

FIG. 6. **The proliferation of (partially assembled) clusters drives the free-energy landscape toward the equilibrium free-energy profile.** Four schematic free-energy profiles (accounting for the cluster free energy, ΔF, and the number of monomers within each cluster, N) are shown at decreasing values of the supersaturation, assuming that the bond strength is held constant. A common-tangent construction can be used to estimate the equilibrium free-energy profile and the average cluster size at a fixed total concentration. The height of the nucleation barrier increases as the pool of available monomers is depleted; as a result, the reaction may stop before reaching equilibrium. In an addressable structure, differing cluster compositions result in deviations from this picture, although the effect of cluster proliferation is qualitatively unchanged.

an addressable structure with hundreds of components, the nucleation barrier may be extremely large at low supersaturation. As a result, it is likely that such a system will be unable to produce the equilibrium number of clusters on an experimental timescale: with both the temperature and the total concentrations of all species held constant, the assembly window closes before all possible target structures have nucleated.

An important consequence is that the experimental conditions must be changed continuously in order to maintain a constant supersaturation. An optimal self-assembly protocol must therefore involve some form of time-dependent temperature control (or other suitable change in the solution conditions[88]). On the phase diagram in Figure 5a, we indicate this by showing that the monomeric fraction does not immediately go to the equilibrium phase boundary, but slowly approaches it at low temperature. Prior knowledge of the nucleation rate and its concentration dependence could be used to design an annealing protocol that stays within the optimal assembly window as a reaction progresses.

A further complication is that, in addressable self-assembly, the complete structure may not always be the most stable cluster.[21] This non-classical behavior originates from the greater mixing entropy of distinguishable partial structures, which can dominate the free-energy landscape at moderate supersaturation. Stable partial structures are a generic feature of finite-sized structures, since the subunits closest to the surface of the structure tend to make fewer connections, particularly in the case

of low-valence building blocks. This effect is exacerbated by heterogeneity in the inter-subunit bond strengths or component concentrations. As a result, a finite-sized structure may require a *multi-step* protocol in order to complete the self-assembly reaction: first, nucleation and growth of a stable partial cluster at one temperature, followed by the complete assembly of the target structure at a lower temperature. An example of a free-energy landscape requiring such a protocol is shown in Figure 7.

C. Foldable polymers

The free-energy-based approach sketched above can be extended to model the self-assembly of polymers with specific interactions. As with naturally occurring two-state proteins, where specific residue–residue contacts encode a thermodynamically stable native state, the number of staple-mediated contacts between backbone binding sites can serve as a suitable order parameter for describing the thermodynamics of folding.[89] This choice is particularly apt when all bonds are saturating, as in the case of DNA origami.

The role of the backbone in constraining the possible folding pathways is a key difference between scaffold-based self-assembly and crystallization. The free-energy barrier that separates the unassembled and assembled states (Figure 5b) arises primarily from the reduction of the backbone's configurational entropy as opposed to the binding of the staple strands. Consequently, certain binding sites, due to their locations on the backbone, may be considerably more likely than others to contribute to the critical nucleus. The presence of a free-energy barrier also results in significant hysteresis in the folding and unfolding temperatures, as observed in experiments on DNA origami.[7,90] Such behavior resembles the familiar two-state kinetics of small proteins in the presence of a denaturant,[91] where specific residues are particularly important for the folding kinetics.[92,93] Following this analogy, it is reasonable to expect that the rate of folding will be increased by stabilizing the 'native' state. At higher supersaturation, kinetic arrest or misbinding is likely to hinder correct assembly. Folding is thus likely to occur within an optimal supersaturation window.

As in the case of scaffold-free aperiodic structures, one nucleation event is required per copy of the target structure. Yet since all interactions are mediated by the backbone scaffold, there is little risk of monomer starvation. The assembly of DNA origami can therefore be performed with an excess of staple strands, such that the depletion of strands does not significantly affect the thermodynamic driving force for assembly (Figure 5b).

D. Hierarchical multicomponent crystals

There is no reason to limit addressable self-assembly to pathways that can also be found in conventional materials. By appropriately tuning the subunit interactions and the component stoichiometries, it is, in principle, possible to construct more exotic pathways. For instance, addressable complexity enables hierarchical approaches to self-assembly, whereby a multi-step pathway passes through one or more stable or metastable phases en route to the intended structure.[31,94,95]

One example of such a scheme is shown in Figure 5c. As described in Ref. 31, two types of bonds can be introduced to stabilize a particular substructure within a complex multicomponent crystal. This results in two classes of subunits: those that can form the favored bonds, and those that cannot. In Figure 5c, we choose a projection of the concentrations of these two subsets of particles in such a way that we can draw a two-dimensional phase diagram. Making appropriate choices for the bond strengths, we can now design a cluster phase that becomes stable before the complete crystal. The pathway to the final structure now looks like a combination of those described previously: partial equilibration due to a proliferation of clusters that coexist with the remaining unassembled particles, followed by a second nucleation window where these clusters reorganize to form a crystal. The modularity demonstrated by experiments on DNA bricks can also be viewed within this framework, where a myriad of distinct substructures are realized by changing the component concentrations within a high-dimensional phase diagram.

V. OPTIMIZING STRUCTURE DESIGNS

Improving our ability to predict the free-energy landscapes of addressable structures holds the key to broadening the range of conditions over which self-assembly can occur. Such predictions are essential because they provide a rational basis for crafting pathways that are robust to variations in experimental conditions. Several recent theoretical advances suggest that practical engineering tools based on these principles may be within reach.

Theoretical analysis and simulations have shown that these landscapes are strongly affected by the topology of the designed interactions that stabilize the target structure. In particular, structures with limited-valence subunits tend to exhibit non-classical nucleation barriers, where 'magic-number' clusters play an important role.[21,30,31] As a result, the size of the critical nucleus may be constant over a range of conditions. The dependence of the nucleation rate on the supersaturation has also been found to vary with the subunit valence, with lower valences typically leading to faster nucleation under milder growth conditions.[21,96] Knowledge of the nature of the transition state is particularly informative for predicting pathways to more topologically complex structures. For instance, non-convex structures are more likely to have critical nuclei that occur late along the dominant assembly pathway, closer to the assembled
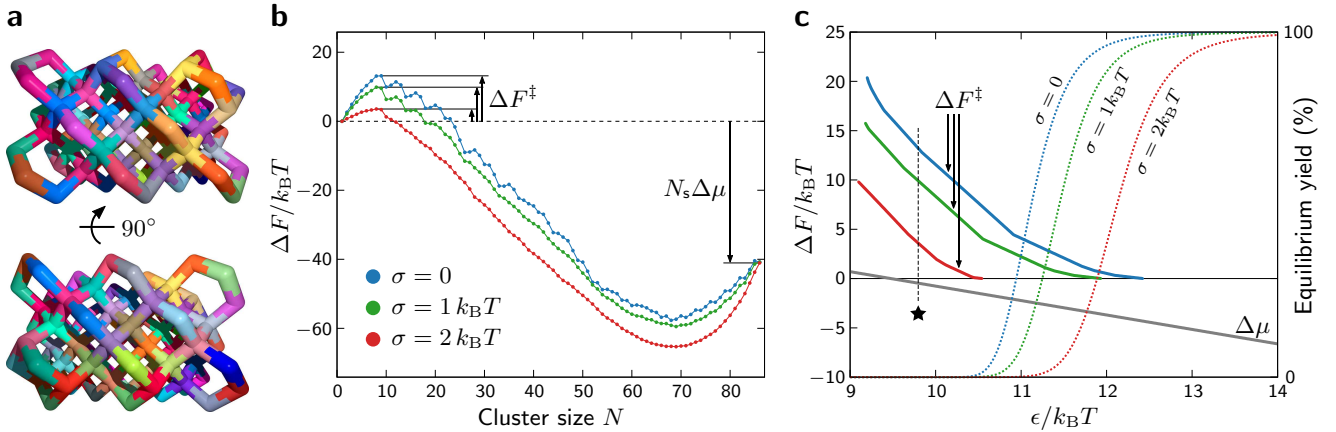
FIG. 7. **Bond-energy heterogeneity can be used to tune the nucleation barrier of a finite-sized addressable structure.** (a) Two views of an 86-particle structure (adapted from Ref. 21). (b) Increasing the variance, σ^2, of the randomly assigned bond energies reduces the height of the nucleation barrier, $\Delta F^{\ddagger}$. In contrast, the thermodynamic stability of the complete structure, $N_{\mathrm{s}}\Delta\mu$, is fixed by the mean bond energy, ϵ. As a result, the supersaturation and the nucleation kinetics can be tuned separately. (c) Given a particular choice of σ, the nucleation barrier can be controlled by changing the temperature, T. The color code in this panel is as in panel **b**. The solid curves describe the dependence of the nucleation barrier on the inverse temperature. The dashed curves indicate the yield of the target structure as a function of the inverse temperature. Notably, the conditions where nucleation is rate-limiting, where the supersaturation is mild, and where the equilibrium yield is high do not overlap, indicating the need for a multi-step assembly protocol. The free-energy profiles in panel **b** were generated at the value of $\epsilon/k_{\mathrm{B}}T$ corresponding to the position of the star ($\star$) in panel **c**.

state. Such non-universality in transition-state structures is well known from studies of protein folding,[97,98] and is thus likely to be the case for scaffold-based addressable structures as well.

In addition to 'rewiring' subunit bonds to alter the bonding topology, the use of addressable building blocks makes it possible to tune the bond energies independently. Directing self-assembly reactions in this manner can be a challenging optimization problem when a very large number of parallel microscopic pathways are present. Nevertheless, it is possible to make predictions at a statistical level as to how randomly distributed bonds will affect the nucleation and growth kinetics of a particular structure (Figure 7).[30] Rather than attempting to design the strength of every bond individually, one can often tune the statistical distribution of bond strengths to achieve rapid, but still rate-limiting, nucleation under mild growth conditions. For non-convex structures, some degree of bond-energy heterogeneity may in fact be necessary for self-assembly to be viable at all (Figure 8).[21] However, simulations[20] also show that excessive heterogeneity increases the concentration of defects during the growth phase, since the presence of a few very strong bonds tends to increase the probability of kinetic trapping. When the standard deviation of the bond-energy distribution is large enough to be comparable to the mean bond strength, a subset of particles may not be able to attach to a growing cluster at moderate supersaturation. These considerations must all be taken into account when designing an addressable structure.

The ability to tune the stabilities of various substructures will also enable the design of materials with controllable growth kinetics. In this case, each stable intermediate phase can serve as a template for the next step in a hierarchical assembly pathway. The fact that we can control the interactions between every pair of particles in an addressable structure means that the nucleation kinetics of each transition can be designed as well.[31] For example, control over the kinetics at intermediate stages could be used to reduce the probability of aggregation, which often limits the effectiveness of hierarchical approaches.[94,95,99] The strategy shown in Figure 5d is but one possibility for achieving such carefully controlled pathways experimentally.

VI. OPTIMIZING SELF-ASSEMBLY PROTOCOLS

The foregoing analysis, based on schematic phase diagrams, explains why some classes of addressable structures can be assembled under constant conditions, while others require more complex experimental protocols. Going beyond such qualitative descriptions, we can apply a detailed understanding of the free-energy landscape of a target structure to make precise predictions for protocol optimization. For instance, an understanding of the nucleation kinetics of a particular structure could be used to derive a temperature ramp for producing clusters at a constant rate. Such a strategy is attractive because it could balance the trade-off between maximizing the yield and minimizing both the time required for the reaction and the likelihood of aggregation along the way. In cases where a pathway to the complete structure does not exist in the nucleation-limited regime, knowledge of a target structure's free-energy landscape could be used
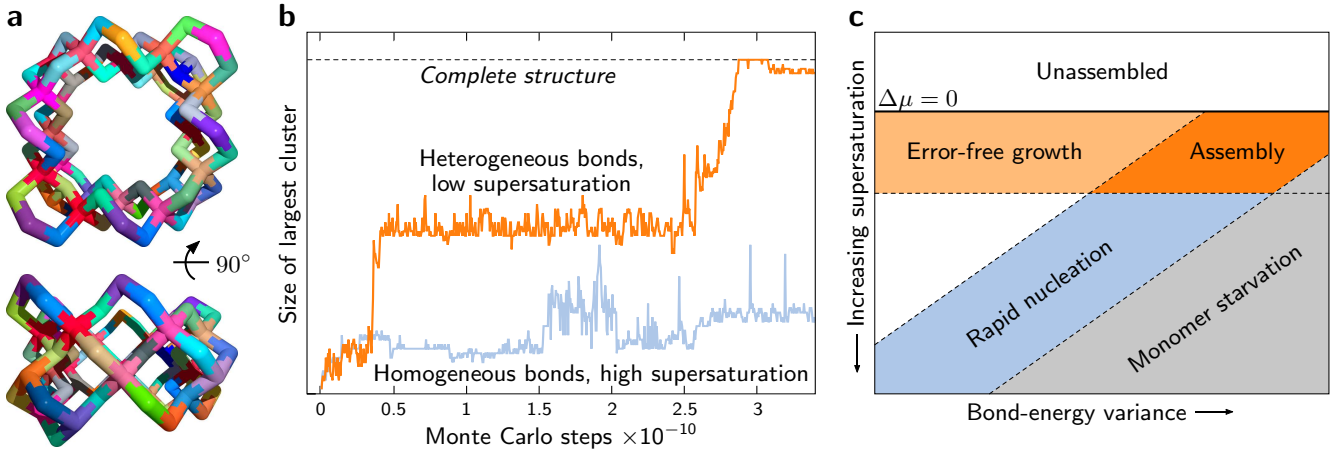
FIG. 8. **Bond-energy heterogeneity facilitates self-assembly.** (a) Two views of a non-convex 74-particle structure (adapted from Ref. 21). (b) Monte Carlo simulations[21] demonstrate that successful assembly requires heterogeneous bonds and low supersaturation; assembly does not occur with equal bond energies at any supersaturation. (c) This observation can be explained by considering the window where the conditions for rapid, yet rate-limiting, nucleation and the conditions for error-free growth overlap. Here the standard deviation of the bond-energy distribution is assumed to be considerably smaller than the mean bond strength; with a broader distribution of bond energies, defects or incomplete assembly are likely to occur (See Sec. V). Notably, the bond-energy variance is typically a *structural* design parameter, while the degree of supersaturation is a *protocol* design parameter.

to determine the number of stages required for a multi-step self-assembly protocol. In particular, our ability to execute hierarchical protocols will likely rely on the accurate prediction of equilibrium phase diagrams and nucleation kinetics. Nevertheless, in order for structure-specific landscapes and phase diagrams to guide the design of optimal protocols, free-energy-based models will need to be properly calibrated against experiments. Recent improvements in the simulation of DNA-based self-assembly[100–103] are making progress toward this goal.

VII. OUTLOOK

The optimization of kinetic pathways is a central challenge in the quest to self-assemble structures with ever greater complexity. Particular emphasis should be placed on designing pathways that are robust to variations in experimental protocols, since this will make complex self-assembly accessible for practical applications. For this purpose, the theoretical framework that we describe here is well suited for predicting the mechanism of phase transformation and the rate-limiting steps in a self-assembly reaction. Nevertheless, the utility of this approach for optimizing experimental protocols has yet to be tested beyond a small number of simulation studies; at present, there is limited experimental analysis of nucleation rates and mechanisms of assembly in DNA-based nanotechnology. Close collaboration between theory and experiment will be necessary to drive further progress and to validate proposed design strategies.

The field of complex self-assembly has seen extraordinarily rapid progress in recent years, but we are only beginning to realize the full potential of addressable DNA-brick and DNA-origami structures. We anticipate that these successes will soon be extended to a greater variety of building blocks and that this technology will radically change the way that complex materials are assembled in the future.

[1] B. Wei, M. Dai, and P. Yin, Nature **485**, 623 (2012).
[2] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, Science **338**, 1177 (2012).
[3] B. Wei, M. Dai, C. Myhrvold, Y. Ke, R. Jungmann, and P. Yin, J. Am. Chem. Soc. **135**, 18080 (2013).
[4] N. C. Seeman, J. Theor. Biol. **99**, 237 (1982).
[5] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, Nature **394**, 539 (1998).
[6] Y. Ke, L. L. Ong, W. Sun, J. Song, M. Dong, W. M. Shih, and P. Yin, Nature Chemistry **6**, 994 (2014).
[7] J.-P. J. Sobczak, T. G. Martin, T. Gerling, and H. Dietz, Science **338**, 1458 (2012).
[8] S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. M. Shih, Nature **459**, 414 (2009).
[9] X.-c. Bai, T. G. Martin, S. H. Scheres, and H. Dietz, Proc. Natl. Acad. Sci. U.S.A. **109**, 20012 (2012).
[10] K. F. Wagenbauer, C. H. Wachauf, and H. Dietz, Nat. Comm. **5**, 3651 (2014).
[11] P. W. K. Rothemund, Nature **440**, 297 (2006).
[12] T. Gerling, K. F. Wagenbauer, A. M. Neuner, and H. Dietz, Science **347**, 1446 (2015).

[13]P. W. K. Rothemund and E. Winfree, in *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing* (ACM, 2000) pp. 459–468.

[14]D. Frenkel, Nat. Mater. **14**, 9 (2015).

[15]L. Cademartiri and K. J. M. Bishop, Nat. Mater. **14**, 2 (2015).

[16]A. Murugan, Z. Zeravcic, M. P. Brenner, and S. Leibler, Proc. Natl. Acad. Sci. U.S.A. **112**, 54 (2015).

[17]P. W. K. Rothemund and E. S. Andersen, Nature **485**, 584 (2012).

[18]K. V. Gothelf, Science **338**, 1159 (2012).

[19]A. Reinhardt and D. Frenkel, Phys. Rev. Lett. **112**, 238103 (2014).

[20]L. O. Hedges, R. V. Mannige, and S. Whitelam, Soft Matter **10**, 6404 (2014).

[21]W. M. Jacobs, A. Reinhardt, and D. Frenkel, Proc. Natl. Acad. Sci. U.S.A. **112**, 6313 (2015).

[22]Y. Tian, T. Wang, W. Liu, H. L. Xin, H. Li, Y. Ke, W. M. Shih, and O. Gang, Nat. Nanotech. **10**, 637 (2015).

[23]Y. Zhang, F. Lu, K. G. Yager, D. van der Lelie, and O. Gang, Nat. Nanotech. **8**, 865 (2013).

[24]S. C. Glotzer and M. J. Solomon, Nat. Mater. **6**, 557 (2007).

[25]L. Feng, R. Dreyfus, R. Sha, N. C. Seeman, and P. M. Chaikin, Adv. Mater. **25**, 2779 (2013).

[26]S. Angioletti-Uberti, P. Varilly, B. M. Mognetti, and D. Frenkel, Phys. Rev. Lett. **113**, 128303 (2014).

[27]N. Dorsaz, L. Filion, F. Smallenburg, and D. Frenkel, Faraday Disc. **159**, 9 (2012).

[28]A. V. Tkachenko, Phys. Rev. Lett. **106**, 255501 (2011).

[29]J. D. Halverson and A. V. Tkachenko, Phys. Rev. E **87**, 062310 (2013).

[30]W. M. Jacobs, A. Reinhardt, and D. Frenkel, J. Chem. Phys. **142**, 021101 (2015).

[31]W. M. Jacobs and D. Frenkel, Soft Matter (2015).

[32]A. Murugan, J. Zou, and M. P. Brenner, Nat. Comm. **6** (2015).

[33]J. Madge and M. A. Miller, J. Chem. Phys. **143**, 044905 (2015).

[34]J. W. Gibbs, Am. J. Sci. **96**, 441 (1878).

[35]D. W. Oxtoby, J. Phys. Cond. Matter **4**, 7627 (1992).

[36]R. P. Sear, J. Phys. Cond. Matter **19**, 033101 (2007).

[37]S. Whitelam and R. L. Jack, Ann. Rev. Phys. Chem. **66**, 143 (2015).

[38]P. R. ten Wolde and D. Frenkel, Science **277**, 1975 (1997).

[39]S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, Soft Matter **5**, 1251 (2009).

[40]T. K. Haxton, L. O. Hedges, and S. Whitelam, Soft Matter (2015).

[41]S. Whitelam, Phys. Rev. Lett. **105**, 088102 (2010).

[42]S. Whitelam, R. Schulman, and L. Hedges, Phys. Rev. Lett. **109**, 265506 (2012).

[43]D. Frenkel, Science **296**, 65 (2002).

[44]V. N. Manoharan, Science **349**, 1253751 (2015).

[45]C. A. Mirkin, R. L. Letsinger, R. C. Mucic, J. J. Storhoff, *et al.*, Nature **382**, 607 (1996).

[46]N. Geerts and E. Eiser, Soft Matter **6**, 4647 (2010).

[47]P. E. Theodorakis, N. G. Fytas, G. Kahl, and C. Dellago, Condensed Matter Physics **18**, 22801 (2015).

[48]A. V. Tkachenko, Phys. Rev. Lett. **89**, 148303 (2002).

[49]D. Lukatsky and D. Frenkel, Phys. Rev. Lett. **92**, 068302 (2004).

[50]D. Lukatsky and D. Frenkel, J. Chem. Phys. **122**, 214904 (2005).

[51]D. Lukatsky, B. Mulder, and D. Frenkel, J. Phys. Condens. Mat. **18**, S567 (2006).

[52]R. Dreyfus, M. E. Leunissen, R. Sha, A. V. Tkachenko, N. C. Seeman, D. J. Pine, and P. M. Chaikin, Phys. Rev. Lett. **102**, 048301 (2009).

[53]C. Knorowski, S. Burleigh, and A. Travesset, Phys. Rev. Lett. **106**, 215501 (2011).

[54]D. Nykypanchuk, M. M. Maye, D. van der Lelie, and O. Gang, Nature **451**, 549 (2008).

[55]S. Y. Park, A. K. R. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, and C. A. Mirkin, Nature **451**, 553 (2008).

[56]H. Xiong, D. van der Lelie, and O. Gang, J. Am. Chem. Soc.

[57]H. Xiong, D. van der Lelie, and O. Gang, Phys. Rev. Lett. **102**, 015504 (2009).

[58]R. J. Macfarlane, B. Lee, H. D. Hill, A. J. Senesi, S. Seifert, and C. A. Mirkin, Proc. Natl. Acad. Sci. U.S.A. **106**, 10493 (2009).

[59]R. V. Thaner, Y. Kim, T. I. Li, R. J. Macfarlane, S. T. Nguyen, M. Olvera de la Cruz, and C. A. Mirkin, Nano Lett. **15**, 5545 (2015).

[60]Y. Zhang, S. Pal, B. Srinivasan, T. Vo, S. Kumar, and O. Gang, Nat. Mater. **14**, 840 (2015).

[61]P. L. Biancaniello, A. J. Kim, and J. C. Crocker, Phys. Rev. Lett. **94**, 058302 (2005).

[62]Y. Wang, Y. Wang, X. Zheng, É. Ducrot, J. S. Yodh, M. Weck, and D. J. Pine, Nat. Comm. **6**, 7253.

[63]N. A. Licata and A. V. Tkachenko, Phys. Rev. E **74**, 040401 (2006).

[64]S. Hormoz and M. P. Brenner, Proc. Natl. Acad. Sci. U.S.A. **108**, 5193 (2011).

[65]Z. Zeravcic, V. N. Manoharan, and M. P. Brenner, Proc. Natl. Acad. Sci. U.S.A. **111**, 15918 (2014).

[66]A. J. Kim, R. Scarlett, P. L. Biancaniello, T. Sinno, and J. C. Crocker, Nat. Mater. **8**, 52 (2008).

[67]W. B. Rogers and J. C. Crocker, Proc. Natl. Acad. Sci. U.S.A. **108**, 15687 (2011).

[68]B. M. Mladek, J. Fornleitner, F. J. Martinez-Veracoechea, A. Dawid, and D. Frenkel, Phys. Rev. Lett. **108**, 268301 (2012).

[69]B. M. Mladek, J. Fornleitner, F. J. Martinez-Veracoechea, A. Dawid, and D. Frenkel, Soft Matter **9**, 7342 (2013).

[70]M. E. Leunissen and D. Frenkel, J. Chem. Phys. **134**, 084702 (2011).

[71]M.-P. Valignat, O. Theodoly, J. C. Crocker, W. B. Russel, and P. M. Chaikin, Proc. Natl. Acad. Sci. U.S.A. **102**, 4225 (2005).

[72]N. A. Licata and A. V. Tkachenko, Phys. Rev. E **74**, 041408 (2006).

[73]F. J. Martinez-Veracoechea and D. Frenkel, Proc. Natl. Acad. Sci. U.S.A. **108**, 10963 (2011).

[74]F. J. Martinez-Veracoechea and M. E. Leunissen, Soft Matter **9**, 3213 (2013).

[75]S. Angioletti-Uberti, P. Varilly, B. M. Mognetti, A. V. Tkachenko, and D. Frenkel, J. Chem. Phys. **138**, 021102 (2013).

[76]P. Varilly, S. Angioletti-Uberti, B. M. Mognetti, and D. Frenkel, J. Chem. Phys. **137**, 094108 (2012).

[77]S. Angioletti-Uberti, B. M. Mognetti, and D. Frenkel, Nat. Mater. **11**, 518 (2012).

[78]B. M. Mognetti, M. Leunissen, and D. Frenkel, Soft Matter **8**, 2213 (2012).

[79]W. B. Rogers and V. N. Manoharan, Science **347**, 639 (2015).

[80]F. J. Martinez-Veracoechea, B. M. Mladek, A. V. Tkachenko, and D. Frenkel, Phys. Rev. Lett. **107**, 045902 (2011).

[81]F. J. Martinez-Veracoechea, B. Bozorgui, and D. Frenkel, Soft Matter **6**, 6136 (2010).

[82]F. Smallenburg and F. Sciortino, Nat. Phys. **9**, 554 (2013).

[83]L. Rovigatti, F. Smallenburg, F. Romano, and F. Sciortino, ACS Nano **8**, 3567 (2014).

[84]R. Schulman, C. Wright, and E. Winfree, ACS Nano **9**, 5760 (2015).

[85]K. Fujibayashi, D. Y. Zhang, E. Winfree, and S. Murata, Natural Computing **8**, 589 (2009).

[86]J. SantaLucia Jr and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).

[87]R. T. Koehler and N. Peyret, Bioinformatics **21**, 3333 (2005).

[88]Z. Zhang, J. Song, F. Besenbacher, M. Dong, and K. V. Gothelf, Angew. Chem. Int. Ed. Engl. **52**, 9219 (2013).

[89]J. N. Onuchic and P. G. Wolynes, Curr. Opin. Struct. Biol. **14**, 70 (2004).

[90]K. E. Dunn, F. Dannenberg, T. E. Ouldridge, M. Kwiatkowska, A. J. Turberfield, and J. Bath, Nature **525**, 82 (2015).

[91]A. R. Fersht, A. Matouschek, and L. Serrano, J. Mol. Biol. **224**, 771 (1992).

[92]N. V. Dokholyan, L. Li, F. Ding, and E. I. Shakhnovich, Proc.

Natl. Acad. Sci. U.S.A. **99**, 8637 (2002).

[93] M. Vendruscolo, N. V. Dokholyan, E. Paci, and M. Karplus, Phys. Rev. E **65**, 061910 (2002).

[94] S. H. Park, C. Pistol, S. J. Ahn, J. H. Reif, A. R. Lebeck, C. Dwyer, and T. H. LaBean, Angew. Chem. **118**, 749 (2006).

[95] S. Whitelam, Soft Matter **11**, 8225 (2015).

[96] D. Frenkel, A. Reinhardt, and C. P. Ho, Faraday Disc. (2015).

[97] A. R. Fersht, Curr. Opin. Struct. Biol. **7**, 3 (1997).

[98] E. I. Shakhnovich, Curr. Opin. Struct. Biol. **7**, 29 (1997).

[99] T. K. Haxton and S. Whitelam, Soft Matter **9**, 6851 (2013).

[100] J. P. Doye, T. E. Ouldridge, A. A. Louis, F. Romano, P. Šulc, C. Matek, B. E. Snodin, L. Rovigatti, J. S. Schreck, R. M. Harrison, *et al.*, Phys. Chem. Chem. Phys. **15**, 20395 (2013).

[101] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **134**, 085101 (2011).

[102] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. Doye, and A. A. Louis, J. Chem. Phys. **137**, 135101 (2012).

[103] T. E. Ouldridge, Mol. Phys. **113**, 1 (2015).

TABLE OF CONTENTS GRAPHIC