# Rational design of self-assembly pathways for complex multicomponent structures

William M. Jacobs[1], Aleks Reinhardt, and Daan Frenkel[1]

Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

Edited by Athanassios Z. Panagiotopoulos, Princeton University, Princeton, NJ, and accepted by the Editorial Board April 7, 2015 (received for review February 3, 2015)

**The field of complex self-assembly is moving toward the design of multiparticle structures consisting of thousands of distinct building blocks. To exploit the potential benefits of structures with such "addressable complexity," we need to understand the factors that optimize the yield and the kinetics of self-assembly. Here we use a simple theoretical method to explain the key features responsible for the unexpected success of DNA-brick experiments, which are currently the only demonstration of reliable self-assembly with such a large number of components. Simulations confirm that our theory accurately predicts the narrow temperature window in which error-free assembly can occur. Even more strikingly, our theory predicts that correct assembly of the complete structure may require a time-dependent experimental protocol. Furthermore, we predict that low coordination numbers result in nonclassical nucleation behavior, which we find to be essential for achieving optimal nucleation kinetics under mild growth conditions. We also show that, rather surprisingly, the use of heterogeneous bond energies improves the nucleation kinetics and in fact appears to be necessary for assembling certain intricate 3D structures. This observation makes it possible to sculpt nucleation pathways by tuning the distribution of interaction strengths. These insights not only suggest how to improve the design of structures based on DNA bricks, but also point the way toward the creation of a much wider class of chemical or colloidal structures with addressable complexity.**

self-assembly | free-energy landscapes | nucleation | DNA nanotechnology

Recent experiments with short pieces of single-stranded DNA (1, 2) have shown that it is possible to assemble well-defined molecular superstructures from a single solution with more than merely a handful of distinct building blocks. These experiments use complementary DNA sequences to encode an addressable structure (3) in which each distinct single-stranded "brick" belongs in a specific location within the target assembly. A remarkable feature of these experiments is that even without careful control of the subunit stoichiometry or optimization of the DNA sequences, a large number of 2- and 3D designed structures with thousands of subunits assemble reliably (1, 2, 4, 5). The success of this approach is astounding given the many ways in which the assembly of an addressable structure could potentially go wrong (6–8).

Any attempt to optimize the assembly yield or to create even more complex structures should be based on a better understanding of the mechanism by which DNA bricks manage to self-assemble robustly. The existence of a sizable nucleation barrier, as originally proposed in refs. 1, 2, would remedy two possible sources of error that were previously thought to limit the successful assembly of multicomponent nanostructures: the depletion of free monomers and the uncontrolled aggregation of partially formed structures. Slowing the rate of nucleation would suppress competition among multiple nucleation sites for available monomers and give the complete structure a chance to assemble before encountering other partial structures. Recent simulations of a simplified model of a 3D addressable structure have provided evidence of a free-energy barrier for nucleation (9), suggesting that the ability to control this barrier should enable the assembly of a wide range of complex nanostructures. We therefore need to be able to predict how such a barrier depends on the design of the target structure and on the choice of DNA sequences. Until now, however, there have been no reliable techniques to predict the existence, let alone the magnitude, of a nucleation barrier for self-assembly in a mixture of complementary DNA bricks.

Here we show that the assembly of 3D DNA-brick nanostructures is indeed a nucleated process, but only in a narrow range of temperatures. The nucleation barrier in these systems is determined entirely by the topology of the designed interactions that stabilize the target structure. Controllable nucleation is therefore a general feature of addressable structures that can be tuned through the rational choice of designed interactions. We find that the reliable self-assembly of 3D DNA bricks is a direct consequence of their unusual nucleation behavior, which is not accounted for by existing theories that work for classical examples of self-assembly, such as crystal nucleation. We are thus able to provide a rational basis for the rather unconventional protocol used in the recent DNA-brick experiments by showing that they exploit a narrow window of opportunity where robust multicomponent self-assembly can take place.

## Structure Connectivity Determines Assembly

In constructing a model for the self-assembly of addressable structures, we note that the designed interactions should be much stronger than any attractive interactions between subunits that are not adjacent in a correctly assembled structure. The designed interactions

### Significance

Recent experiments have demonstrated that complex, three-dimensional nanostructures can be self-assembled out of thousands of short strands of preprogrammed DNA. However, the mechanism by which robust self-assembly occurs is poorly understood, and the same feat has not yet been achieved using any other molecular building block. Using a new theory of "addressable" self-assembly, we explain how the design of the target structure and the choice of interparticle interactions determine the self-assembly pathway, and, to our knowledge, for the first time predict that a time-dependent protocol, rather than merely a carefully tuned set of conditions, may be necessary to optimize the yield. With an understanding of these design principles, it should be possible to engineer addressable nanostructures using a much wider array of materials.

Author contributions: W.M.J., A.R., and D.F. designed research; W.M.J. and A.R. performed research; W.M.J., A.R., and D.F. analyzed data; and W.M.J., A.R., and D.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.Z.P. is a guest editor invited by the Editorial Board.

Data deposition: Supporting data related to this publication are available at github.com/wmjac/pygtsa.

[1]To whom correspondence may be addressed. Email: wjacobs@fas.harvard.edu or df246@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1502210112/-/DCSupplemental.

that stabilize the target structure can be described by a connectivity graph $G$, in which each vertex represents a distinct subunit and each edge indicates a correct bond. This graph allows us to describe the connectivity of the structure without specifying the geometric details and spatial organization of the building blocks. For structures constructed from DNA bricks, the edges of $G$ indicate the hybridization of DNA strands with complementary sequences that are adjacent in the target structure. An example 3D DNA-brick structure is shown along with its connectivity graph in Fig. 1 *A* and *B*.

In an ideal solution with exclusively designed interactions, the subunits assemble into clusters in which all allowed bonds are encoded in the connectivity graph of the target structure. To compute the free-energy difference between on-pathway clusters of a particular size and the unbonded single-stranded bricks, we must consider all of the ways in which a correctly bonded cluster with a given number of monomers can be assembled. These clusters can be described by the distinct "fragments" of the target structure, which correspond to connected subgraphs of the connectivity

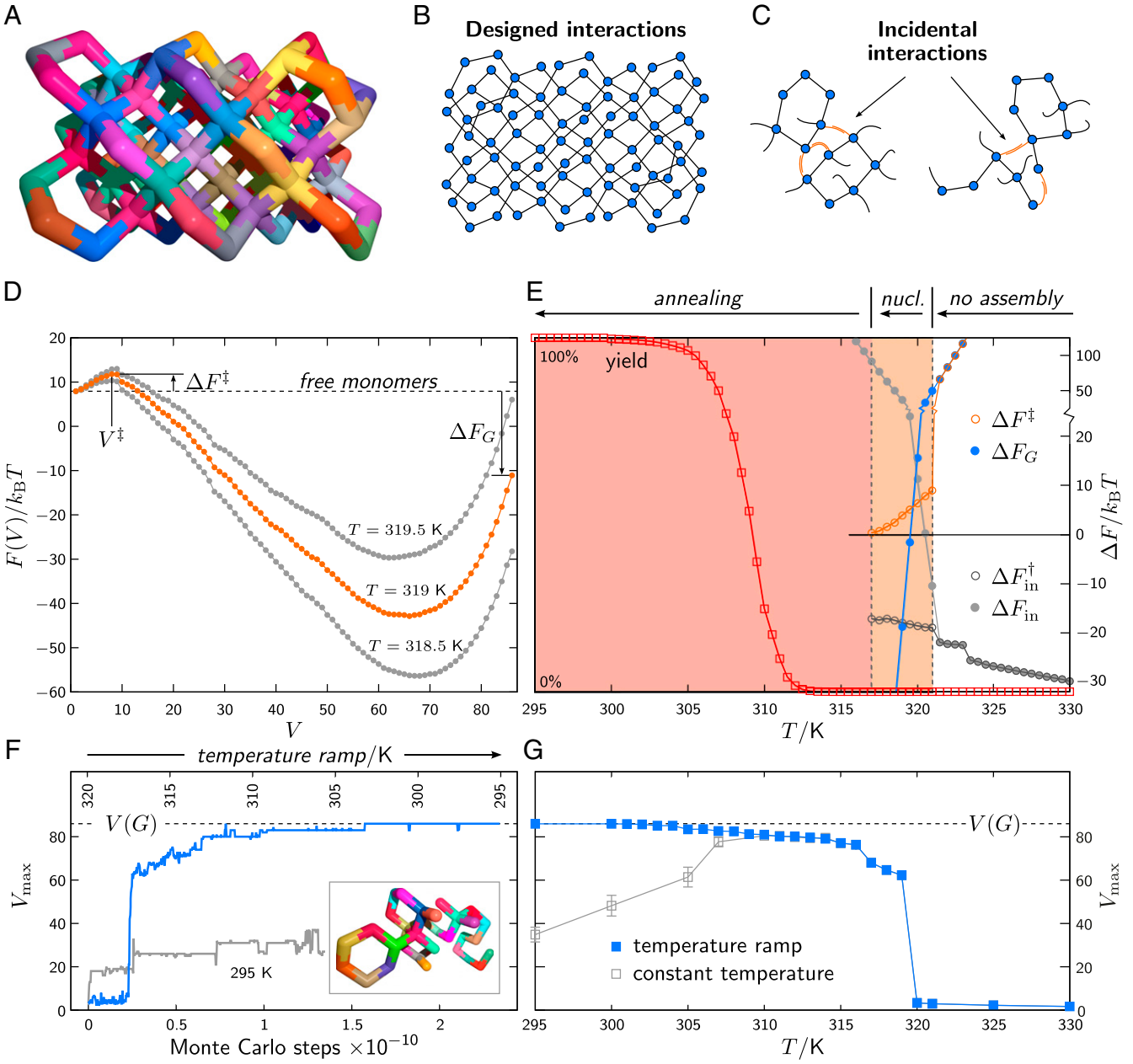



**Fig. 1.** Controlled nucleation is essential for robust self-assembly. (*A*) An example 86-strand DNA-brick structure and (*B*) its associated connectivity graph. (*C*) Incidental interactions between dangling ends, shown in orange, lead to incorrect associations between fragments. (*D*) Free energies of clusters of $V$ subunits with randomly chosen DNA sequences in units of $k_BT$, where $k_B$ is the Boltzmann constant and $T$ is the absolute temperature. The nucleation barrier $\Delta F^\ddagger$ and the target structure stability $\Delta F_G$ are strongly temperature-dependent. (*E*) The equilibrium yield with exclusively designed interactions and the nucleation barrier as a function of temperature. Also shown are the target structure stability and the free-energy difference between all off-pathway and on-pathway intermediates, in the presence, $\Delta F^\dagger_{in}$, and absence, $\Delta F_{in}$, of a nucleation barrier. The nucleation window is shown in orange. (*F*) Representative lattice Monte Carlo simulation trajectories with (blue) and without (gray) a temperature ramp. (*Inset*) A typical malformed structure. (*G*) The size of the largest correctly bonded stable cluster in lattice Monte Carlo simulations using a temperature ramp (blue) and in constant-temperature simulations initialized from free monomers in solution (gray).

graph. In a dilute solution with strong designed interactions, the numbers of edges and vertices are the primary factors determining the stability of a particular fragment. We therefore identify all possible assembly intermediates by grouping fragments into sets with the same number of edges and vertices and counting the total number of fragments in each set. This calculation yields the "density of states" of fragments of the target structure, which is an intrinsic property of the connectivity graph. We combine the density of states with information about the subunit geometries, monomer concentrations, and designed bond strengths to compute the free energies of the on-pathway clusters. We can also estimate the equilibrium probability of forming competing off-pathway structures, the overwhelming majority of which arises from undesired incidental interactions between subunits. For further discussion of the theory, please see *SI Text*, section S1 and ref. 10.

This theoretical approach is powerful because it can predict the free-energy landscape as a function of the degree of assembly between the monomers and the target structure. Furthermore, the predicted landscape captures the precise topology of the target structure, which is essential for understanding the assembly of addressable, finite-sized structures. In the case of DNA-brick structures, we can assign DNA hybridization free energies to the edges of the target connectivity graph to determine the temperature dependence of the free-energy landscape; for example, Fig. 1*D* shows the free-energy profile of the 86-strand DNA-brick structure with random DNA sequences at three temperatures. Our theoretical approach allows us to calculate the nucleation barrier $\Delta F^{\ddagger}$ by examining the free energies of clusters corresponding to fragments with exactly $V$ vertices. The critical number of strands required for nucleation is $V^{\ddagger}$: transient clusters with fewer than $V^{\ddagger}$ strands are more likely to dissociate than to continue incorporating additional strands. The presence of a substantial nucleation barrier therefore inhibits the proliferation of large, partially assembled fragments that stick together to form nontarget aggregates.

### Assembly Requires a Time-Dependent Protocol

Over a significant range of temperatures, we find that the free-energy profiles of DNA-brick structures exhibit both a nucleation barrier and a thermodynamically stable intermediate structure. The nucleation barrier is associated with the minimum number of subunits that must be assembled to complete one or more cycles, i.e., closed loops of stabilizing bonds in a fragment. For example, the critical number of monomers in the example structure at 319 K, $V^{\ddagger} = 8$, is one fewer than the nine subunits required to form a bicyclic fragment of the target structure. Under the conditions where nucleation is rate controlling, the minimum free-energy structure is not the complete 86-particle target structure, but rather a structure with only $V \simeq 65$ particles. This incomplete structure is favored by entropy, because it can be realized in many more ways than the unique target structure. Hence, the temperature where nucleation is rate controlling is higher than the temperature where the target structure is the most stable cluster. The existence of thermodynamically stable intermediates is a typical feature of DNA-brick structures and of complex addressable, finite-sized structures in general. [We note that thermodynamically stable partial structures have also been observed in previously reported simulations (9) of DNA-brick structures consisting of approximately 1,000 distinct subunits.]

This behavior is not compatible with classical nucleation theory (CNT), which predicts that, beyond the nucleation barrier, large clusters are always more stable than smaller clusters. As a consequence, in "classical" nucleation scenarios such as crystallization, there is a sharp boundary in temperature and concentration at which the largest-possible ordered structure, rather than the monomeric state, becomes thermodynamically stable. Typically, a simple fluid must be supersaturated well beyond this boundary to reduce the nucleation barrier, which arises due to the competition between the free-energy penalty of forming a solid–liquid interface and the increased stability due to the growth of an ordered structure (11–13). However, in the case of addressable self-assembly, and DNA bricks in particular, a nucleation barrier for the formation of a stable partial structure may exist even when the target structure is unstable relative to the free monomers.

An experiment to assemble such a structure requires a protocol: first nucleation at a relatively high temperature, and then further cooling to complete the formation of the target structure. [DNA-brick structures have also been assembled at constant temperature by changing the solution conditions during the course of the experiment (4). For simplicity, we assume that the experimental control parameter is the temperature, but clearly other methods of altering the DNA hybridization free energies during the assembly process may also be suitable.] This behavior can be seen in Fig. 1*E*, where we identify a narrow temperature window in which there is a significant yet surmountable nucleation barrier. Unlike CNT, the nucleation barrier does not diverge as the temperature is increased. Instead, there is a well-defined temperature above which all clusters have a higher free energy than the free monomers. As the temperature is lowered further, the nucleation barrier disappears entirely before the equilibrium yield, defined as the fraction of all clusters that are correctly assembled as the complete target structure, increases measurably above zero. The equilibrium yield tends to 100% at low temperatures, because we have thus far assumed that only designed interactions are possible. Therefore, because of the presence of stable intermediate structures, it is typically impossible to assemble the target structure completely at any temperature where nucleation is rate controlling.

To examine the importance of a nucleation barrier for preventing misassembly, we estimate the free-energy difference between off-pathway aggregates and all on-pathway intermediates, $\Delta F_{\text{in}}$, by calculating the probability of incidental interactions between partially assembled structures (10). From the connectivity graph of the example DNA-brick structure, we can calculate the total free energy of aggregated clusters by considering all of the ways that partially assembled structures can interact via the dangling ends of the single-stranded bricks, as shown in Fig. 1*C*. We also estimate this free-energy difference in the case of slow nucleation, $\Delta F^{\dagger}_{\text{in}}$, by only allowing one of the interacting clusters in a misassembled intermediate to have $V > V^{\ddagger}$.

The above analysis supports our claim that a substantial nucleation barrier is essential for accurate self-assembly. Our calculations show that even with very weak incidental interactions, incorrect bonding between the multiple dangling ends of large partial structures prevents error-free assembly at equilibrium, because $\Delta F_{\text{in}} > 0$. The presence of a nucleation barrier slows the approach to equilibrium, maintaining the viability of the correctly assembled clusters.

These theoretical predictions are confirmed by extensive Monte Carlo simulations of the structure shown in Fig. 1*A*. In these simulations, the DNA bricks are modeled as rigid particles that move on a cubic lattice, but otherwise the sequence complementarity and the hybridization free energies of the experimental system are preserved (9). Using realistic dynamics (14), we simulate the assembly of the target structure using a single copy of each monomer. In Fig. 1*F*, we compare a representative trajectory from a simulation using a linear temperature ramp with a trajectory from a constant-temperature simulation starting from free monomers in solution. We also report the largest stable cluster size averaged over many such trajectories in Fig. 1*G*. Nucleation first occurs within the predicted nucleation window where $\Delta F^{\ddagger} \simeq 8\, k_{\text{B}}T$. At 319 K, the size of the largest stable cluster coincides precisely with the predicted average cluster size at the free-energy minimum in Fig. 1*D* [Incomplete assembly at a constant temperature has also been observed in experiments (15) as well as in our simulations of more complex structures, such as a rectangular slab with a raised checkerboard pattern.] Intermediate

structures assembled via a temperature ramp continue to grow at lower temperatures, whereas clusters formed directly from a solution of free monomers become arrested in conformations that are incompatible with further growth (Fig. 1*F*, *Inset*). In agreement with our theoretical predictions, the simulation results demonstrate that a time-dependent protocol is essential for correctly assembling a complete DNA-brick structure.

### Coordination Number Controls the Nucleation Barrier

In the modular assemblies reported in ref. 2, the maximum coordination number of bricks in the interior of the structure is 4. However, one can envisage other building blocks, such as functionalized molecular constructs or nanocolloids, that have a different coordination number. To investigate the effect of the coordination number on the nucleation barrier, we compare the free-energy profile of a 48-strand DNA-brick structure with those of two higher-coordinated structures (Fig. 2*A*): a simple cubic structure with coordination number $q_c = 6$ and a close-packed structure with $q_c = 12$. (For a discussion of 2D structures, see *SI Text*, section S4.) In Fig. 2*B*, we show the free-energy profiles at 50% yield assuming identical bond energies within each structure.

One striking difference between the $q_c = 4$ structure and the higher-coordinated examples is the stability of the target at 50% yield. In the DNA-brick structure, the target structure coexists in nearly equal populations with a partial structure that is missing a single cycle. In the structures with higher coordination numbers, however, the target has the same free energy as the free monomers at 50% yield. Intermediate structures are therefore globally unstable at all temperatures, as predicted by CNT.

A second point of distinction among these structures lies in the relative stability of intermediate cluster sizes. Whereas the DNA-brick structure assembles by completing individual cycles, the cubic structure grows by adding one face at a time to an expanding cuboid. [This equilibrium growth pathway is similar to that described in ref. 16.] With $q_c = 12$, the greater diversity of fragments with the same number of vertices smooths out the free-energy profile near the top of the nucleation barrier. The fitted black line in Fig. 2*B* shows that the assembly of this structure does in fact obey CNT (*SI Text*, section S2).

The differences among these free-energy profiles originate from the topologies of the connectivity graphs of the example structures. The most important determinant of the nucleation behavior is simply the number of vertices required to complete each additional cycle in the target connectivity graph, which is controlled by the maximum coordination number of the subunits. [Although the coordination number also affects the rotational entropy in lattice-based simulations, altering the rotational entropy in our theoretical method has the same effect as changing the monomer concentration (*SI Text*, section S3) and thus does not affect the shape of the free-energy profile.] Our findings imply that controlled self-assembly of 3D addressable structures is unlikely to be achieved straightforwardly using subunits with coordination numbers higher than 4. In higher-coordinated structures, which are well described by CNT, it would be necessary to go to high supersaturation to find a surmountable nucleation barrier; however, such an approach is likely to fail due to kinetic trapping (17). However, in DNA-brick structures, the nucleation barrier is surmountable at low supersaturation and is relatively insensitive to the size of the target structure (Fig. 2*C*). The reliable self-assembly of large DNA-brick structures is thus a direct consequence of the small number of bonds made by each brick.

### Heterogeneous Bond Energies Improve Kinetics

Recent publications have argued that equal bond energies should enhance the stability of the designed structure (18) and reduce errors during growth (19). By contrast, we find that the kinetics of DNA-brick assembly are actually worse if one selects DNA sequences that minimize the variance in the bond energies. Our observation is consistent with the successful use of random DNA sequences in the original experiments with DNA bricks (1, 2). Here again, the nucleation behavior is responsible for this unexpected result.

To demonstrate the difference between random DNA sequences and sequences chosen to yield monodisperse bond energies, we consider the relatively simple nonconvex DNA-brick structure shown in Fig. 3*A*. This 74-brick structure, constructed by removing the interior strands and two faces from a cuboidal structure, assembles roughly face-by-face when using random DNA sequences. The relevant nucleation barrier, as predicted theoretically in Fig. 3*B* and confirmed with Monte Carlo simulations in Fig. 3*C*, is the completion of the third face. With monodisperse bond energies and an equivalent mean interaction

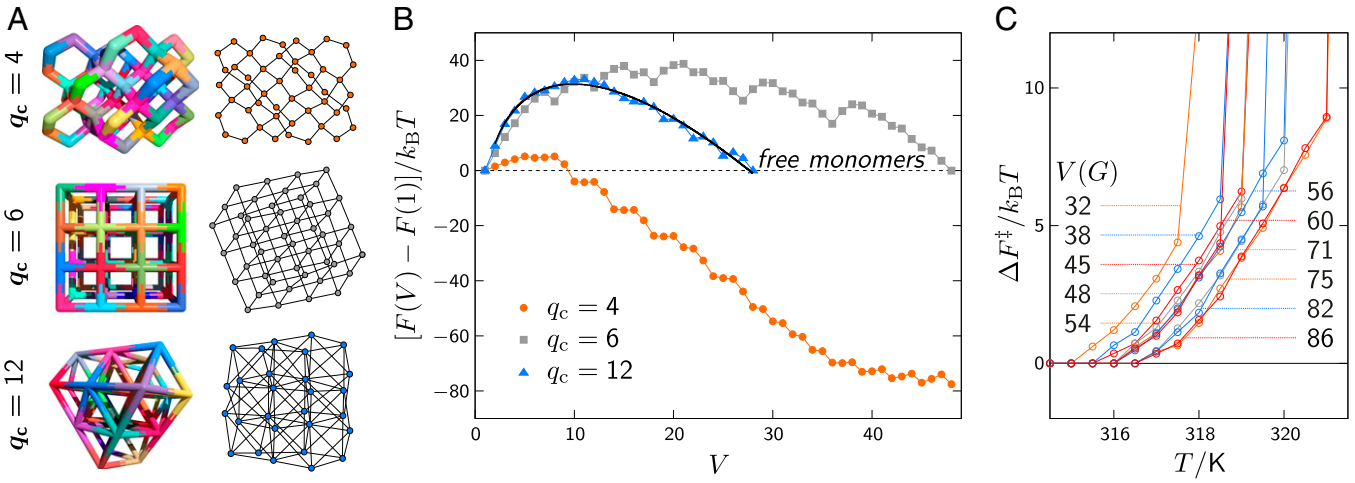



**Fig. 2.** Dependence of the nucleation barrier on the coordination number of the target structure. (*A*) Example structures with tetrahedral coordination ($q_c = 4$), octahedral coordination ($q_c = 6$), and close-packed coordination ($q_c = 12$), along with their associated connectivity graphs. (*B*) Free-energy profiles of these three structures at 50% yield assuming identical bond energies within each structure. The black line shows the fit of classical nucleation theory to the nucleation barrier of the $q_c = 12$ structure. (*C*) The dependence of the nucleation barrier on the total number of strands, $V(G)$, in DNA-brick structures with randomly chosen DNA sequences. The nucleation temperature does not increase monotonically with $V(G)$ in these roughly cuboidal structures because surface effects are considerable.

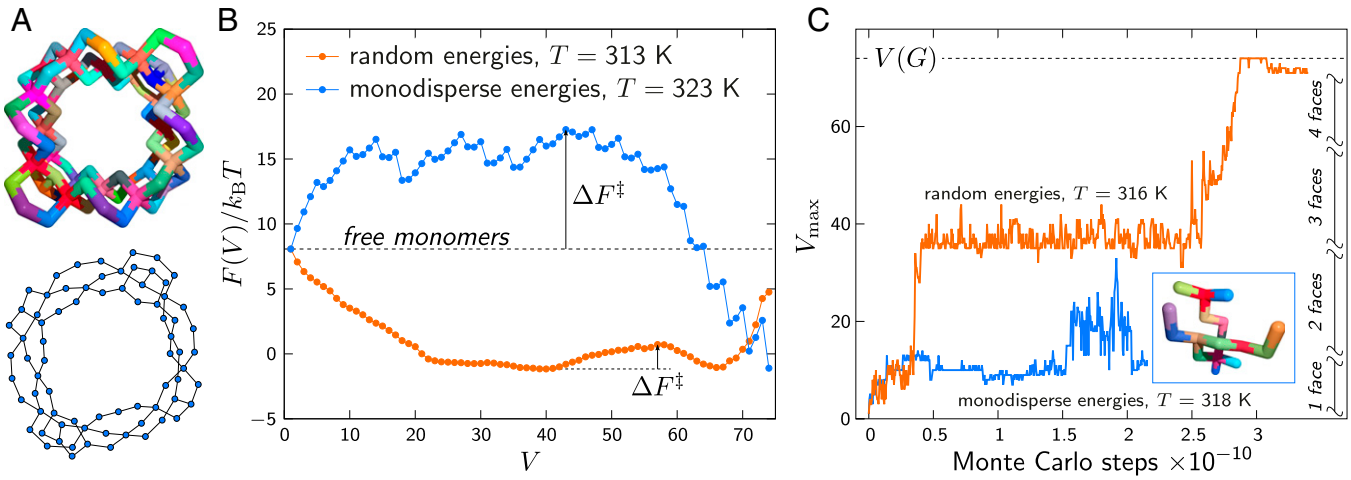



**Fig. 3.** Random DNA sequences can improve the kinetics of self-assembly. (*A*) A nonconvex, 74-strand DNA-brick structure and its associated connectivity graph. (*B*) Free-energy profiles and effective nucleation barriers $\Delta F^{\ddagger}$, with both random and monodisperse designed interaction strengths. The temperatures are chosen such that the mean hybridization free energies of these two distributions are equal. (*C*) Representative lattice Monte Carlo simulation trajectories with heterogeneous and monodisperse bond energies. Correct assembly proceeds by assembling one face of the structure at a time. The temperature of the monodisperse system must be lowered to achieve nucleation, but the largest cluster quickly becomes kinetically trapped (*Inset*).

strength, a much larger nucleation barrier appears before the first face forms. Attempts to reduce this nucleation barrier by increasing the mean bond energy result in kinetic trapping and arrested growth. Despite promising fluctuations in the largest cluster size in the simulation trajectory with monodisperse energies, multiple competing nuclei appear, and the largest cluster remains poorly configured for further assembly (Fig. 3*C*, *Inset*).

The use of sequences with a broad distribution of hybridization free energies results in a more suitable nucleation barrier because such a distribution selectively stabilizes small and floppy intermediate structures. This is a statistical effect: because there are far fewer ways of constructing a maximally connected fragment with a given number of monomers, the chance that randomly assigned sequences concentrate the strongest bonds in a compact fragment is vanishingly small in a large structure. As a result, the dominant nucleation pathways no longer need to follow the maximally connected fragments. The use of a broad distribution of bond energies therefore tends to reduce nucleation barriers, because unstable fragments near the top of a barrier contain fewer cycles and are thus affected more significantly by the variance in the bond-energy distribution.

## Design Principles for (DNA) Nanostructures

The insights provided by our predictive theory allow us to understand the general principles underlying the unexpected success of DNA-brick self-assembly. Slow, controlled nucleation at low supersaturation is achieved for large structures because each brick can only make a small number of designed connections. Because of an appreciable nucleation barrier that appears in a narrow temperature window, monomer depletion does not pose a significant problem for one-pot assembly. Surprisingly, complex structures with randomly selected complementary DNA sequences experience enhanced nucleation, making larger intermediate structures kinetically accessible at higher temperatures.

The use of a temperature ramp plays a more crucial role than previously thought. Cooling the DNA-brick solution slowly is not just a convenient way of locating good assembly conditions, as in the case of conventional crystals; rather, it is an essential nonequilibrium protocol for achieving error-free assembly of finite-sized structures. The explanation of slow nucleation and fast growth that was originally proposed in refs. 1, 2 is therefore incomplete: fast growth allows the DNA bricks to assemble into a stable, on-pathway intermediate that must be annealed at lower temperatures to complete the target structure. For nucleation to be rate controlling, the monomers must diffuse sufficiently rapidly for the solution of precritical clusters to reach equilibrium. With this assumption, our results suggest that the growth of supercritical clusters proceeds by monomer addition and that, as a result, incorrect assembly is rare. However, successful assembly also relies on the slow diffusion of large intermediates, whose rate of diffusion changes approximately inversely with the cluster radius (20). Remaining out of equilibrium throughout the assembly protocol is necessary to avoid the aggregation of these partial structures, which we predict to be the globally favored thermodynamic state.

Our approach also suggests how to improve the design of DNA-brick nanostructures beyond the random selection of uniformly distributed DNA sequences. For a given target structure, it is easy to tune the nucleation barrier by adjusting the statistical distribution of bond energies. Complementary DNA sequences can then be assigned to the structure to achieve the desired distribution of hybridization free energies. Furthermore, with an understanding of the origin of the nucleation barrier in a particular structure, it is possible to optimize the annealing protocol rationally to increase the yield of the target assembly. Our approach also provides a means of systematically investigating how local modifications to the coordination number through the fusing of adjacent strands affect the nucleation behavior of DNA-brick structures (21).

The theoretical method used here greatly simplifies the quantitative prediction of nucleation barriers and intermediate structures with widespread applications for controlling the self-assembly of biomolecular or synthetic building blocks. Addressable self-assembly holds great promise for building intricate 3D structures that are likely to require optimization on a case-by-case basis. Because our predictive theory is sensitive to the details of a particular target structure, performing these calculations for nanostructures of experimental interest will enable the precise engineering of assembly properties at the design stage. In order for potential users to perform such experimental protocol design, we provide a user-friendly software package online at github.com/wmjac/pygtsa.

## Methods

**DNA Hybridization Free Energies.** We compute the hybridization free energies of complementary 8-nucleotide DNA sequences using established empirical formulas (22, 23) assuming salt concentrations of $[Na^+]$ = 1 mol dm$^{-3}$ and $[Mg^{2+}]$ = 0.08 mol dm$^{-3}$. For the calculations with monodisperse bond

energies, we use the sequences provided in ref. 19. The strengths of incidental interactions are estimated based on the longest attractive overlap for each pair of noncomplementary sequences. In calculations of the equilibrium yield and free-energy profiles, we report the average thermodynamic properties using 1,000 randomly chosen complete sets of DNA sequences. See *SI Text*, section S5 for further details.

**Lattice Monte Carlo Simulations.** Constant-temperature lattice Monte Carlo simulations are carried out using the virtual move Monte Carlo algorithm (14) to produce physical dynamics. Rigid particles, each with four distinct patches fixed in a tetrahedral arrangement, are confined to a cubic lattice. A single copy of each required subunit is present in the simulation box with $62 \times 62 \times 62$ lattice sites. Complete details are given in ref. 9. For comparison with the results of these simulations, the theoretical calculations reported here assume the same dimensionless monomer concentration, $\rho = 62^{-3}$, lattice coordination number, $q_c = 4$, and fixed number of dihedral angles, $q_d = 3$ (*SI Text*, section S1).

**ACKNOWLEDGMENTS.** 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. W.M.J. acknowledges support from the Gates Cambridge Trust and the National Science Foundation Graduate Research Fellowship under Grant DGE-1143678.

1. Wei B, Dai M, Yin P (2012) Complex shapes self-assembled from single-stranded DNA tiles. *Nature* 485(7400):623–626.
2. Ke Y, Ong LL, Shih WM, Yin P (2012) Three-dimensional structures self-assembled from DNA bricks. *Science* 338(6111):1177–1183.
3. Rothemund PW, Winfree E (2000) *The Program-Size Complexity of Self-Assembled Squares* (ACM, New York), pp 459–468.
4. Zhang Z, Song J, Besenbacher F, Dong M, Gothelf KV (2013) Self-assembly of DNA origami and single-stranded tile structures at room temperature. *Angew Chem Int Ed Engl* 52(35):9219–9223.
5. Wei B, et al. (2013) Design space for complex DNA structures. *J Am Chem Soc* 135(48): 18080–18088.
6. Rothemund PW, Andersen ES (2012) Nanotechnology: The importance of being modular. *Nature* 485(7400):584–585.
7. Gothelf KV (2012) Materials science. LEGO-like DNA structures. *Science* 338(6111): 1159–1160.
8. Kim AJ, Scarlett R, Biancaniello PL, Sinno T, Crocker JC (2009) Probing interfacial equilibration in microsphere crystals formed by DNA-directed assembly. *Nat Mater* 8(1):52–55.
9. Reinhardt A, Frenkel D (2014) Numerical evidence for nucleated self-assembly of DNA brick structures. *Phys Rev Lett* 112(23):238103.
10. Jacobs WM, Reinhardt A, Frenkel D (2015) Communication: Theoretical prediction of free-energy landscapes for complex self-assembly. *J Chem Phys* 142(2):021101.
11. Gibbs JW (1878) On the equilibrium of heterogeneous substances. *Am J Sci* 16(96):441–458.
12. Oxtoby DW (1992) Homogeneous nucleation: Theory and experiment. *J Phys Condens Matter* 4(38):7627–7650.
13. Sear RP (2007) Nucleation: Theory and applications to protein solutions and colloidal suspensions. *J Phys Condens Matter* 19(3):033101.
14. Whitelam S, Geissler PL (2007) Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles. *J Chem Phys* 127(15):154101.
15. Sobczak JPJ, Martin TG, Gerling T, Dietz H (2012) Rapid folding of DNA into nanoscale shapes at constant temperature. *Science* 338(6113):1458–1461.
16. Shneidman VA (2003) On the lowest energy nucleation path in a supersaturated lattice gas. *J Stat Phys* 112(1-2):293–318.
17. Whitelam S, Jack RL (2015) The statistical mechanics of dynamic pathways to self-assembly. *Annu Rev Phys Chem* 66(1):143–163.
18. Hormoz S, Brenner MP (2011) Design principles for self-assembly with short-range interactions. *Proc Natl Acad Sci USA* 108(13):5193–5198.
19. Hedges LO, Mannige RV, Whitelam S (2014) Growth of equilibrium structures built from a large number of distinct component types. *Soft Matter* 10(34):6404–6416.
20. Einstein A (1905) Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann Phys* 322(8):549–560.
21. Li W, Yang Y, Jiang S, Yan H, Liu Y (2014) Controlled nucleation and growth of DNA tile arrays within prescribed DNA origami frames and their dynamics. *J Am Chem Soc* 136(10):3724–3727.
22. SantaLucia J, Jr, Hicks D (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* 33(1):415–440.
23. Koehler RT, Peyret N (2005) Thermodynamic properties of DNA sequences: Characteristic values for the human genome. *Bioinformatics* 21(16):3333–3339.

# Supporting Information

Jacobs et al. 10.1073/pnas.1502210112

## SI Text

**S1. Theoretical Free-Energy and Equilibrium Yield Calculations.** We introduce the theoretical free-energy calculations by way of the simple example structure shown in Fig. S1; a thorough explanation of this theoretical method is presented in ref. 1. The connectivity graph $G$ represents all designed bonds between adjacent subunits in the target structure. These bonds and subunits are indicated by the edges and vertices of $G$, respectively. Because $G$ represents an addressable structure, every vertex is distinct. From this graph we are able to determine all of the relevant thermodynamic properties of the intermediate and target structures in a near-equilibrium assembly protocol.

The connected subgraphs (fragments) of $G$ represent all possible on-pathway clusters (i.e., clusters with at most one particle of each component and only designed bonds) of the target structure. To make the subsequent computations tractable, we group these fragments into sets, $h(E,V)$, in which all fragments have precisely $E$ edges and $V$ vertices. In the example structure, the fragments in each of these sets all have the same unlabeled connectivity graphs; we can therefore show all sets of fragments of $G$ by removing the component labels in Fig. S1*A*. The number of distinct subgraphs in each set is shown below each graph. We also list the relevant topological properties of each fragment. In general, the unlabeled connectivity graphs of fragments in the same set are not isomorphic. However, the fugacities of the bonded clusters that they represent are very similar, which justifies their collective treatment as a single set in the free-energy calculations.

In the dilute limit, the dimensionless grand potential $-\ln\Xi$ of the on-pathway clusters can be written in terms of a sum over all sets of fragments,

$$Z_{\mathrm{id}} \equiv \ln\Xi = \sum_{E,V} |h(E,V)|\, \bar{z}_{E,V}. \qquad \textbf{[S1]}$$

The average fugacity of the fragments in each set $\bar{z}_{E,V}$ depends on the topologies of the fragment graphs as well as the geometry of the subunits and the solution conditions. Ignoring excluded volume interactions, we can approximate $\ln\bar{z}_{E,V}$ as

$$\ln\bar{z}_{E,V} = E\beta\tilde{\epsilon}_{E,V} + V\ln\rho - (V-1)\ln q_{\mathrm{c}} - \left(V - \overline{B}_{E,V} - 1\right)\ln q_{\mathrm{d}}, \qquad \textbf{[S2]}$$

where $k_{\mathrm{B}}\ln q_{\mathrm{c}}$ is the rotational entropy of a monomer, $k_{\mathrm{B}}\ln q_{\mathrm{d}}$ is the dihedral entropy of an unconstrained dimer, and $\rho$ is the dimensionless concentration. The mean dimensionless dihedral entropy of a fragment in this set is $\overline{B}_{E,V}\ln q_{\mathrm{d}} \equiv \ln\langle q_{\mathrm{d}}^{B(g)}\rangle_{g\in h(E,V)}$, where $B(g)$ is the number of bridges (2) in the fragment $g$. This quantity is shown for each of the sets of fragments in Fig. S1*A*. The exponentially weighted mean bond energy within each set, $\beta\tilde{\epsilon}_{E,V}$, is

$$\beta\tilde{\epsilon}_{E,V} \equiv \left\langle \frac{1}{E}\ln\left\langle \exp\sum_{b\in\mathcal{E}(g)}\beta\epsilon_b \right\rangle_{g\in h(E,V)} \right\rangle_{\mathrm{seq}}, \qquad \textbf{[S3]}$$

where $\epsilon_b$ is the absolute value of the hybridization free energy of bond $b$, $\beta\equiv 1/k_{\mathrm{B}}T$ is the inverse temperature, and $\mathcal{E}(g)$ is the edge set of fragment $g$. The inner average runs over all fragments in the set $h(E,V)$ with fixed bond energies $\{\epsilon_b\}$. In the case of DNA-brick structures, the outer average samples DNA sequences so that each complete set of bond energies for the target structure is chosen independently from the same distribution of hybridization free energies.

The free energy of a correctly bonded cluster of $V$ monomers is thus

$$\beta F(V) \equiv -\ln\sum_{E} |h(E,V)|\, \bar{z}_{E,V}, \qquad \textbf{[S4]}$$

because we do not distinguish among equally sized clusters with varying compositions. This definition is appropriate for studying nucleation, as any subset of monomers has the potential to serve as a nucleation site. An example free-energy profile is shown in Fig. S1*B*. In ref. 1, we showed that the number of linearly independent cycles in a fragment, $C\equiv E-V+1$, is an important determinant of the stability of the corresponding clusters. In the example structure, only the target structure contains a cycle; as a result, the free energy in Fig. S1*B* does not decrease until the final bond is formed.

The equilibrium yield $\eta$ is defined as the fraction of all on-pathway clusters in solution that are correctly assembled into the target structure,

$$\eta \equiv \frac{\langle N_G\rangle}{\sum_g\langle N_g\rangle} = \frac{z_G}{Z_{\mathrm{id}}}, \qquad \textbf{[S5]}$$

where $\langle N_G\rangle$ is the grand-canonical average number of copies of fragment $g$ in solution and $z_G$ is the fugacity of the target structure. In the case of structures with higher coordination numbers, a few edges may be removed from the connectivity graph without allowing any subunit to disassociate or rotate. For these structures, we replace $z_G$ in Eq. **S5** with a sum over the fugacities of all fragments that enforce the correct geometry of the target structure.

Lastly, the free-energy difference between all off-pathway and on-pathway clusters is estimated to be

$$\beta\Delta F_{\mathrm{in}} \equiv -\ln Z_{\mathrm{in}} + \ln Z_{\mathrm{id}}, \qquad \textbf{[S6]}$$

where $Z_{\mathrm{in}}$ is obtained from a high-temperature expansion of dimers of fragments $g$ and $g'$ that associate via weak incidental interactions with strength $\beta w$,

$$Z_{\mathrm{in}} \simeq \sum_{g,g'} z_g z_{g'} \sum_{\lambda}^{\min\left(A_g,A_{g'}\right)} \binom{A_g}{\lambda}\binom{A_{g'}}{\lambda}\lambda!\frac{2^{-\delta_{gg'}}\lambda\rho\beta w}{q_{\mathrm{c}}q_{\mathrm{d}}^{2\lambda-1}}. \qquad \textbf{[S7]}$$

The sum over $\lambda$ accounts for the many possible combinations of interactions between the $A(g)$ and $A(g')$ nonbonded "dangling ends" of the fragments $g$ and $g'$, respectively. In terms of the connectivity graph, $A(g)$ is the number of edges of $G$ that are adjacent to $g$. These adjacent edges are indicated by short gray lines in Fig. S1*A*. To apply Eq. **S7**, we calculate the average number of adjacent edges $\overline{A}_{E,V}\equiv\langle A_g\rangle_{g\in h(E,V)}$ for each set of fragments and rewrite $Z_{\mathrm{in}}$ as a sum over all pairs of fragment sets. It is also possible for an off-pathway cluster to form in the absence of incidental interactions by incorporating multiple particles of a single component via designed bonds; however, this precludes the completion of one or more cycles of the target connectivity graph and thus greatly reduces the probability of forming such a structure at equilibrium. In a dilute solution at low supersaturation, aggregation via incidental

interactions is the more significant practical limitation for complex self-assembly due to the combinatorial explosion of misinteractions in Eq. **S7**.

In an addressable structure as large as the ones considered in the main text, it is not possible to enumerate all fragments as we have done in Fig. S1. It is however possible to determine the number of fragments in each set $h(E,V)$ to arbitrary precision by performing a stochastic calculation in the fragment state space. This procedure, which is described in ref. 1, need only be done once for a given connectivity graph. Furthermore, the reciprocal of the now-determined density of states, $|h(E,V)|^{-1}$, can be used as a biasing potential for subsequent Markov chain Monte Carlo sampling of the fragment state space; we can therefore ensure that all sets $h(E,V)$ are sampled, on average, with equal frequency. As a result, we can rapidly calculate the fragment-set averages $\overline{B}_{E,V}$ and $\overline{A}_{E,V}$ to statistical precision.

**S2. Classical Nucleation Theory.** Classical nucleation theory predicts a free-energy barrier for the nucleation of a stable, ordered structure from an unstable, fluid phase. The height of the barrier and the size of the critical nucleus vary with the degree of supersaturation of the fluid phase (3, 4). Assuming spherical nuclei, the classical prediction for the free-energy difference between a nucleus of the ordered phase containing $V$ monomers and the bulk fluid phase is

$$\Delta F(V) = -\Delta\mu V + \gamma\left(36\pi\rho_{\mathrm{ord}}^{-2}\right)^{1/3}(V-V_0)^{2/3}, \qquad \textbf{[S8]}$$

where $\Delta\mu$ is the bulk free-energy difference per particle between the fluid phase and the ordered phase, $\gamma$ is the free-energy cost per unit area of forming an interface between the two phases, and $\rho_{\mathrm{ord}}$ is the number density of the ordered phase. In Fig. 2*B*, the black line shows the fit to Eq. **S8** with $V_0 = 1$.

**S3. Concentration Dependence.** In the main text, we report all results of the theoretical calculations assuming a dimensionless concentration of $62^{-3}$ for comparison with the lattice Monte Carlo simulations. Changing this concentration shifts both the equilibrium yield and the nucleation barrier linearly with $\ln\ \rho$. The concentration dependence of the nucleation barrier of the example 86-strand structure at several temperatures is shown in Fig. S2. Although we have assumed equal concentrations of all monomers, polydispersity in the monomer concentrations can be easily incorporated into the theoretical treatment in much the same way as the distributions of designed interaction energies.

**S4. Assembly of 2D Structures.** We find that the shapes of the free-energy profiles of 2D structures are also strongly affected by their coordination numbers. In Fig. S3, we show the free-energy profiles at 50% yield of two similarly sized 2D structures with coordination numbers $q_c = 3$ and $q_c = 4$. The behavior of the 2D structure with $q_c = 3$ is similar to that of the 4-coordinated 3D structures examined in the main text. The 2D structure with $q_c = 4$ exhibits the same face-by-face assembly as 3D structures with octahedral coordination; in the 2D case, however, coexistence at 50% yield occurs between the target structure and fragments with the same number of monomers but fewer bonds.

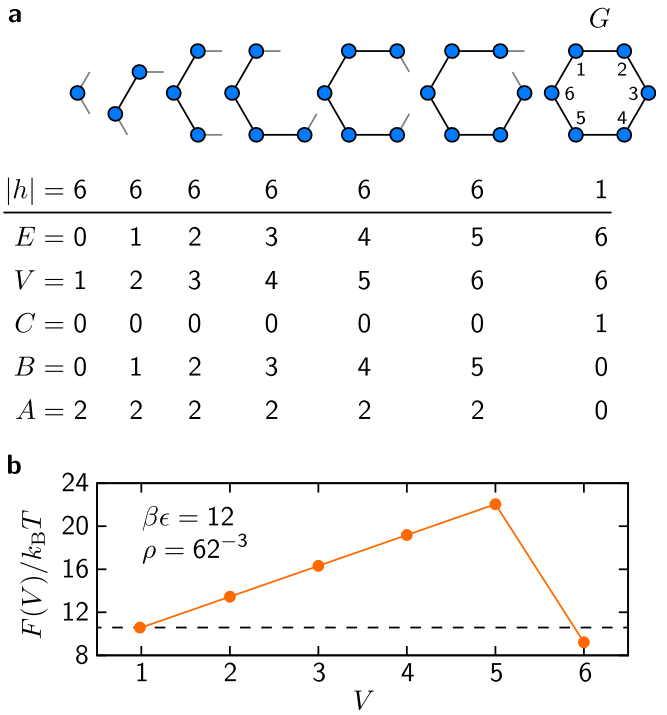
Two-dimensional DNA-tile structures with $q_c = 4$ have been successfully assembled (5). In general, the nucleation barriers in 2D structures are much lower than in 3D structures with a similar number of monomers. Consequently, a lower supersaturation is required in order for the target structure to become kinetically accessible, and kinetic trapping is therefore less likely to interfere with accurate assembly. Nevertheless, these calculations suggest that lower coordinated 2D structures, such as the hexagonal lattice pictured in Fig. S3*A*, might assemble more robustly in experiments.

**S5. Hybridization Free-Energy Distributions.** DNA hybridization free energies are strongly temperature-dependent (6, 7). In Fig. S4*A*, we compare the two hybridization free-energy distributions used in the main text. The mean and the variance of the hybridization free energies of 8-nucleotide sequences are shown for both the case of randomly chosen sequences and the case of sequences selected to yield monodisperse bond energies (8). Designed interactions occur between complementary sequences, whereas incidental interactions are calculated based on the most attractive overlapping regions of two noncomplementary sequences.

In the lattice Monte Carlo simulations, all monomers on adjacent lattice sites experience a weak 100 K repulsion at all temperatures. In calculations involving designed interactions, this repulsion is subtracted from the mean interaction strength. When estimating incidental interactions, we ignore associations between noncomplementary sequences that have a maximum attractive interaction of less than 100 K. The means and variances reported in Fig. S4*A* are thus calculated based on the fraction of pairs of noncomplementary sequences that attract more strongly than 100 K; this fraction is shown in Fig. S4*B*.

These incidental interaction distributions are clearly approximate and are defined to match the lattice Monte Carlo simulations. Nevertheless, the choices made in defining these distributions have a negligible effect on the calculated values of $Z_{\mathrm{in}}$ (1) and are irrelevant to the prediction of nucleation barriers and equilibrium yields. To apply this theoretical method to an experimental system, the designed interaction distributions should be recalculated in accordance with the experimental solution conditions.


1. Jacobs WM, Reinhardt A, Frenkel D (2015) Communication: Theoretical prediction of free-energy landscapes for complex self-assembly. *J Chem Phys* 142(2):021101.
2. Bondy JA, Murty USR (1976) *Graph Theory with Applications* (Macmillan, London), Vol 6.
3. Oxtoby DW (1992) Homogeneous nucleation: Theory and experiment. *J Phys Condens Matter* 4(38):7627–7650.
4. Sear RP (2007) Nucleation: Theory and applications to protein solutions and colloidal suspensions. *J Phys Condens Matter* 19(3):033101.
5. Wei B, Dai M, Yin P (2012) Complex shapes self-assembled from single-stranded DNA tiles. *Nature* 485(7400):623–626.
6. SantaLucia J, Jr, Hicks D (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* 33(1):415–440.
7. Koehler RT, Peyret N (2005) Thermodynamic properties of DNA sequences: Characteristic values for the human genome. *Bioinformatics* 21(16):3333–3339.
8. Hedges LO, Mannige RV, Whitelam S (2014) Growth of equilibrium structures built from a large number of distinct component types. *Soft Matter* 10(34):6404–6416.




| $\lvert h\rvert = 6$ | 6 | 6 | 6 | 6 | 6 | 1 |
|---|---|---|---|---|---|---|
| $E = 0$ | 1 | 2 | 3 | 4 | 5 | 6 |
| $V = 1$ | 2 | 3 | 4 | 5 | 6 | 6 |
| $C = 0$ | 0 | 0 | 0 | 0 | 0 | 1 |
| $B = 0$ | 1 | 2 | 3 | 4 | 5 | 0 |
| $A = 2$ | 2 | 2 | 2 | 2 | 2 | 0 |

**Fig. S1.** Simple example of a connectivity graph and free-energy calculation. (*A*) The connectivity graph *G* represents six distinct particles that are bonded via designed interactions. To the left of *G* are the unlabeled connectivity graphs of all on-pathway intermediate clusters. The number of distinct fragments in each set $|h(E, V)|$ and the relevant topological properties of the fragments in these sets are shown below their corresponding connectivity graphs. These properties are *E*, number of edges; *V*, number of vertices; *C*, number of cycles; *B*, number of bridges; and *A*, number of adjacent edges. (*B*) The free-energy profile as a function of the cluster size *V*, assuming that all bonds have equal strengths, i.e., $\beta\epsilon_b = \beta\epsilon\ \forall b \in \mathcal{E}(G)$. In this calculation, the geometric constants are assumed to be $q_c = 4$ and $q_d = 3$.

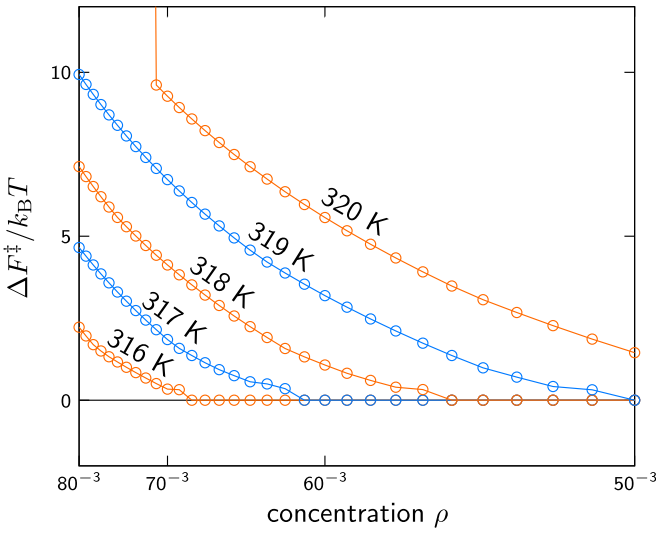



**Fig. S2.** Dependence of the height of the nucleation barrier $\Delta F^{\ddagger}$ on the dimensionless concentration $\rho$ for the example 86-strand structure shown in Fig. 1*A*.

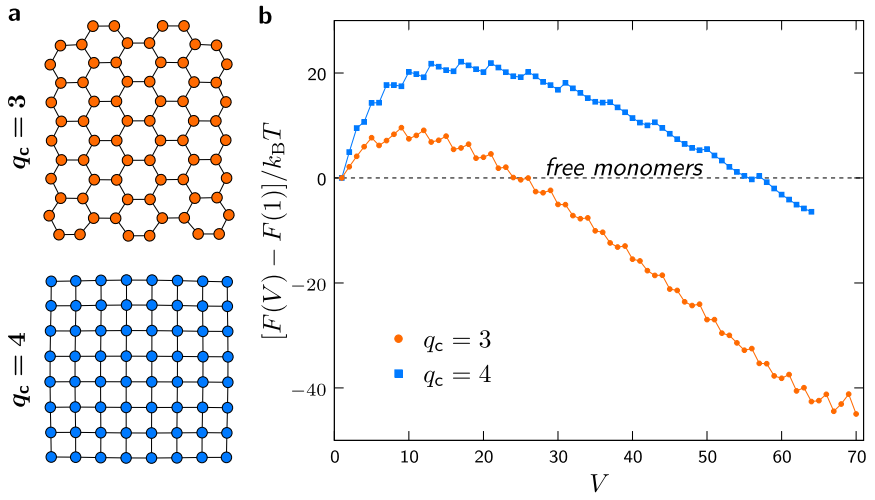



**Fig. S3.** Dependence of the nucleation barrier on the coordination number of 2D structures. (*A*) Connectivity graphs of 2D structures with coordination numbers $q_c = 3$ and $q_c = 4$. (*B*) The free-energy profiles of these structures at 50% yield assuming identical bond energies within each structure. For both structures, we assume $q_d = 2$.

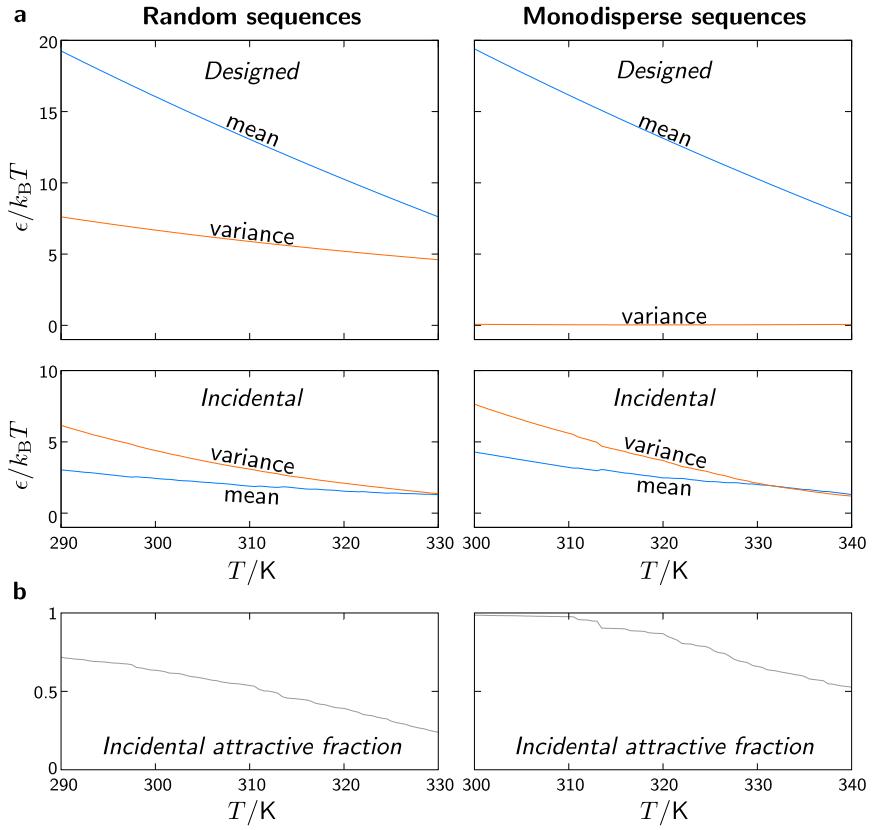



**Fig. S4.** DNA hybridization free-energy distributions. (*A*) The mean and variance of the hybridization free-energy distributions with randomly chosen DNA sequences (*Left*) and with sequences selected to yield monodisperse bond energies (*Right*). Designed interactions refer to complementary DNA sequences, whereas incidental interactions report the hybridization free energy of the most attractive overlapping region of two noncomplementary sequences. Note that the means of the designed interactions of these two distributions are shifted by ∼10 K. (*B*) The fraction of pairs of noncomplementary sequences that attract more strongly than 100 K.