

Direct observation and rational design of nucleation behavior in addressable self-assembly

Martin Sajfutdinow

Fraunhofer Institute for Cell Therapy and Immunology IZI,

Department of Diagnostics, Perlickstraße 1, 04103 Leipzig, Germany and

Faculty of Chemistry and Mineralogy, Leipzig University, Johannisallee 29, 04103 Leipzig, Germany

William M. Jacobs

Department of Chemistry and Chemical Biology, Harvard University,

12 Oxford Street, Cambridge, Massachusetts 02138, United States

Aleks Reinhardt

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

Christoph Schneider and David M. Smith

Fraunhofer Institute for Cell Therapy and Immunology IZI,

Department of Diagnostics, Perlickstraße 1, 04103 Leipzig, Germany

In order to optimize a self-assembly reaction, it is essential to understand the factors that govern its pathway. Here, we examine the influence of nucleation pathways in a model system for addressable, multi-component self-assembly based on a prototypical ‘DNA-brick’ structure. By combining temperature-dependent dynamic light scattering with coarse-grained simulations, we show how subtle changes in the nucleation pathway profoundly affect the yield of the correctly formed structures. In particular, we can increase the range of conditions over which self-assembly occurs by utilizing stable multi-subunit clusters that lower the nucleation barrier for assembling subunits in the interior of the structure. Consequently, modifying only a small portion of a structure is sufficient to optimize its assembly. Due to the generality of our coarse-grained model and the excellent agreement that we find with our experimental results, the design principles reported here are likely to apply generically to addressable, multi-component self-assembly.

Increasingly complex structures can now be created by self-assembly [1, 2], from nanostructures with tailored physicochemical properties, such as photonic crystals [3, 4], to quasicrystals [5–7]. In the limit where every subunit in a target structure is unique and bonds strongly with specific partners, such self-assembled structures are said to be ‘addressable’. Thus far, this degree of specificity has been demonstrated most convincingly by experiments on ‘DNA bricks’ [8], in which portions of single-stranded DNA molecules are designed to hybridize uniquely with complementary sequences on strands that occupy neighboring positions in the target structure. Modular nanostructures comprising thousands of distinct strands can be formed in this way, and because the location of each molecule in the target structure is precisely known, these structures can be functionalized at a nanometer length scale.

In addition to providing control over the geometry of the target structure, the use of addressable building blocks makes it possible to exert greater control over the *mechanism* of self-assembly [9]. Because each interaction between subunits can be individually tuned, addressable structures provide a useful platform for exploring the determinants of self-assembly pathways more generally [10]. Considerable progress has been made in this direction using computer simulations [11–20] and statistical mechanics [21, 22] to study coarse-grained models of addressable systems. In particular, coarse-grained modeling has predicted that nucleation barriers[23] are likely to play a particularly important role in addressable self-assembly, since in their absence, the large number of building blocks with similar bonding strengths can instead lead to widespread kinetic trapping and uncontrollable aggregation [13, 21]. These models have further shown that addressable systems often have highly non-classical nucleation barriers and well-defined critical nuclei [18, 21]. However, the microscopic nature of a self-assembly process is challenging to study experimentally. While it is possible to stop the reaction at a specific point along an annealing ramp [24, 25] for subsequent imaging [26,

27], this approach typically yields only qualitative insights. Furthermore, any partially formed structures must first be isolated before carrying out more detailed analyses, for example by using next-generation sequencing to examine defects in DNA nanostructures [28]. On the other hand, existing *in situ* methods only provide information on the kinetics of self-assembly by probing the interactions between pairs of subunits [24, 29–32]. As a result, the behavior of these interactions must then be extrapolated to describe the collective assembly of the complete structure.

In this work, we demonstrate that dynamic light scattering (DLS) can be used to track the collective assembly of addressable structures in greater detail. Unlike alternative *in situ* techniques, DLS provides us with a sensitive means of probing the complete distribution of cluster sizes throughout the course of the annealing protocol. Consequently, by applying DLS to DNA-brick self-assembly, we are able to analyze the nucleation process as a function of temperature and assembly time. Combining these results with extensive simulations, we show that it is possible to control the nucleation behavior rationally, with dramatic consequences for the yield of assembled structures. In particular, we demonstrate that a self-assembly mechanism can be optimized by altering the connections among a specific subset of subunits in order to modify the free-energy barrier for structure nucleation. The simplicity of our coarse-grained model suggests that these design principles are transferable to any multi-component system where the interactions between subunits can be programmed.

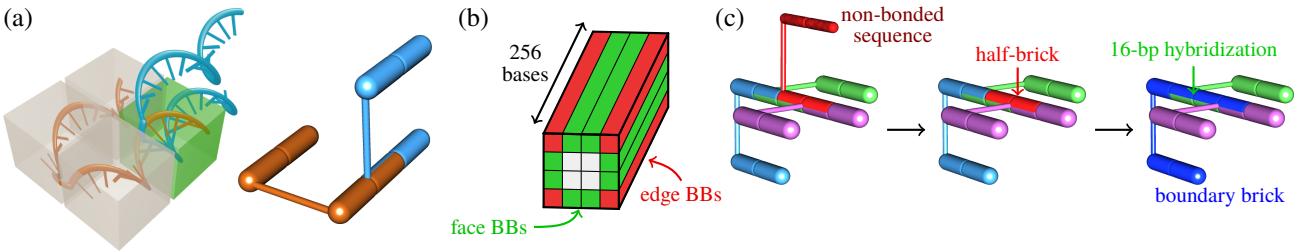


FIG. 1. (a) The bonding pattern between two DNA bricks in a strand and a cylindrical representation. Each molecule is partitioned into four domains (indicated by boxes), while the neighboring bricks are bonded through one pair of domains only (green box). (b) A schematic illustration of our target structure, highlighting the locations of the ‘edge’ and ‘face’ boundary bricks. (c) A schematic of the boundary brick set-up: the non-interacting DNA sequence at one of the outer surfaces of the target structure is removed and the remainder of that brick is fused with an adjacent brick.

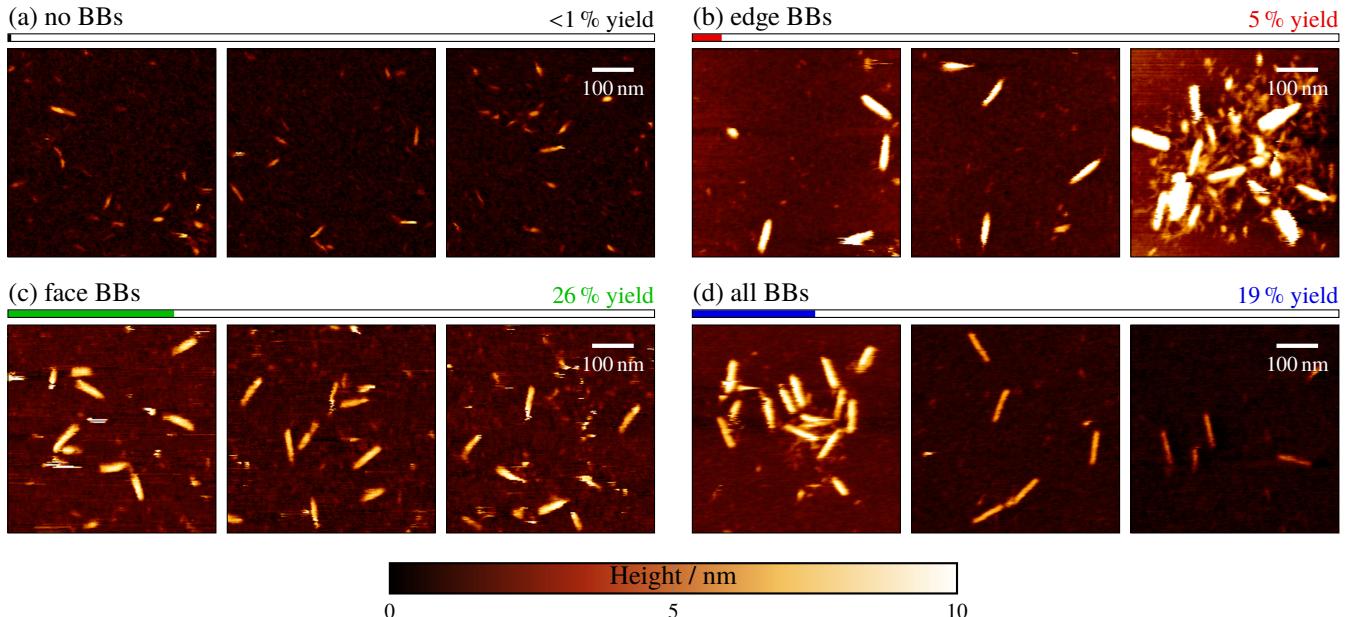


FIG. 2. AFM images for each system investigated (see Sec. SI-1.2). No purification was performed on these samples so that incompletely assembled structures can be clearly seen. Typical yields estimated via gel electrophoresis are shown for each system, relative to the total quantity of DNA strands in solution. For the edge-BB system in particular, multi-structure aggregates (shown in the right-most panel) are commonly observed.

RESULTS

Minor changes in nanostructure design strongly affect the yield and quality of self-assembly

As a model system, we examine the self-assembly of a 16-helix DNA cuboid. Following the canonical ‘DNA-brick’ design [8], the fundamental building blocks of this structure are 32-nucleotide (nt) ‘scaffold’ bricks. Each brick comprises four 8-base-pair (bp) domains that hybridize to connect adjacent helices (Fig. 1a). This cuboid is the smallest structure that ensures that bricks on opposite sides of the structure do not interact directly (Fig. 1b), while the high aspect ratio (4 helices \times 4 helices \times 256 bases) facilitates the identification of well-formed structures via AFM imaging.

To study the factors affecting the self-assembly yield, we designed variants of this cuboid by increasing the lengths of a small number of complementary domains, resulting in the formation of stable, multi-subunit clusters at high temperatures (see Sec. SI-2). This was achieved by varying the numbers and types of 48-nt ‘boundary bricks’ (BBs) at the exterior surface of the structure (Fig. 1c). In addition to a cuboid composed entirely of scaffold bricks (‘no BBs’), where the 16-bp half-

bricks at the exterior of the structure are left unconjugated, we designed variants with boundary bricks forming the corner helices (‘edge BBs’), connecting pairs of helices on the faces of the cuboid (‘face BBs’) or both (‘all BBs’). All variants of the cuboid structure self-assembled to some degree over the course of a 66 h linear annealing ramp (see Sec. SI-1.1). However, AFM imaging revealed striking differences in the quality of the assembled structures (Fig. 2). The all-BB, face-BB and edge-BB designs resulted in the assembly of many copies of structures with the expected aspect ratio, while designs without boundary bricks appeared to yield a negligible number of such structures.

DLS reveals collective assembly transitions

We next turned to DLS to obtain information on the size of structures as a function of temperature and assembly time. These measurements provide insight into the growth of clusters without requiring intercalating dyes or other perturbative additives to be introduced into the system. Because sub-micron-sized particles scatter visible light in the Rayleigh limit, where the scattering intensity scales as the sixth power of the particle size, DLS is highly sensitive to small populations of multi-strand

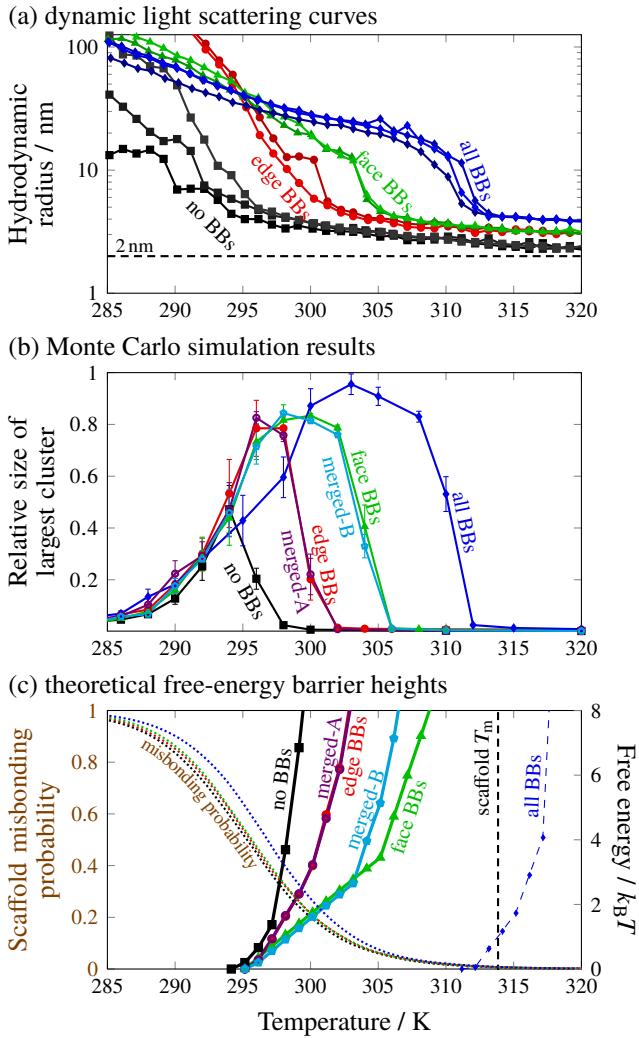


FIG. 3. (a) Hydrodynamic radii from dynamic light scattering following a 15.2 h annealing protocol, computed from intensity-weighted size distribution functions. Several experimental runs are shown for each system. (b) Size of largest correctly bonded cluster from Monte Carlo simulations. Each data point corresponds to an average over ten independent simulations in the long time limit, once nucleation has occurred. Error bars give the standard deviation in each case. The ‘merged’ curves refer to fixed edge dimers, as discussed in the text. (c) The height of the nucleation barrier as a function of temperature from theoretical calculations. The all-BB barrier refers to the nucleation of a network of boundary bricks, as discussed in the text. We also show the predicted scaffold melting temperature, T_m , and the median probability that a scaffold strand forms at least one mis-interaction in the absence of successful nucleation.

clusters. We used a time series of intensity measurements to determine the light-scattering auto-correlation function, from which a distribution of decay rates could be extracted. Assuming that the particles in the suspension do not interact with one another, a decay rate could be related to the diffusion coefficient of a spherical particle with an effective hydrodynamic radius R_h . Since determining the decay rate distribution from the auto-correlation function requires additional assumptions on the smoothness of the cluster-size distribution, we used multiple regularization methods to verify the robustness of our results (Fig. S1).

Carrying out DLS *in situ*, we measured the distribution of hydrodynamic radii of multi-strand clusters as a function of temperature over the course of a 15.2 h linear annealing protocol. We then calculated the intensity-weighted mean hydrodynamic radius, $\langle R_h \rangle$, from the particle-size distribution, excluding the

scattering intensity from any impurities in the system (Fig. S2a). When probing the extent of subunit dimerization with fluorescence annealing (see Sec. SI-2), we observed a prominent peak in the fluorescence response at relatively high temperatures due to the hybridization of boundary bricks; however, in DLS experiments, we did not observe any substantial change in the overall scattering intensity at temperatures above 315 K (Fig. S3). At these temperatures, the no-BB $\langle R_h \rangle \approx 2$ nm (Fig. 3a), which corresponds to a flexible 32-nt DNA strand (see Methods). Upon further cooling, we observed sharp increases in both $\langle R_h \rangle$ and the total scattering intensity for all three structures with boundary bricks. These transitions are reproducible across multiple annealing runs for the all-BB, face-BB structures and edge-BB structures (Fig. 3a). By contrast, the no-BB structures show large variations, and, in each run, $\langle R_h \rangle$ does not increase substantially until the annealing ramp reaches approximately 295 K.

Since our analysis of the DLS data clearly resolves two distinct populations in the full cluster-size distribution (Fig. 4a), it is possible to extract the mean hydrodynamic radius for each population by fitting a linear combination of gaussian functions at each temperature. A population of smaller particles, with mean R_h similar to the R_h of the full distribution above the transition temperature, is detectable over a temperature range of approximately 10 K below the initial transition for each structure (Fig. 4b). This population indicates that not all subunits are incorporated into complete structures at the initial transition temperature. Meanwhile, in the immediate vicinity of each initial transition temperature, the mean of the larger- R_h population matches the expected size of a fully assembled cuboid (~ 20 nm; see Methods), as confirmed by measurements of pre-assembled and purified all-BB structures (Fig. S4). The locations of these peaks remain nearly constant for at least 3 K below each initial transition.

We note that at lower temperatures, the intensity-weighted $\langle R_h \rangle$ increases beyond the size of a fully assembled structure. One possible explanation is that a fraction of the assembled structures aggregate, and $\langle R_h \rangle$ trends upwards due to the sixth-power dependence of the light scattering intensity on the particle size in the suspension. It is also likely that hydrodynamic interactions among proliferating large structures tend to shift the decay-rate distribution towards smaller effective diffusion constants. Furthermore, we cannot exclude the limited resolution of DLS, since the increased scattering intensity broadens the peak associated with the full-size structure and causes it to merge with aggregate and impurity peaks at temperatures below the initiation transition (Fig. S2b). Nevertheless, the ~ 20 nm population persists, since it can be resolved once more near 290 K (Fig. S2c).

Evidence for nucleation and growth

The bimodal size distributions obtained from our DLS measurements indicate that DNA-brick self-assembly is consistent with a nucleation-and-growth mechanism. Under the conditions where self-assembly occurs, scaffold strands are most likely to exist either in a monomeric state or as part of a nearly complete structure. By contrast, intermediate cluster sizes, which must be traversed in order to assemble a complete target structure, are thermodynamically unstable and, consequently, are unlikely to be observed. The size distributions that we extract from our DLS measurements therefore indicate that the formation of an unstable nucleus with an intermediate hydrodynamic radius is a relatively rare event (Fig. 4a,b).

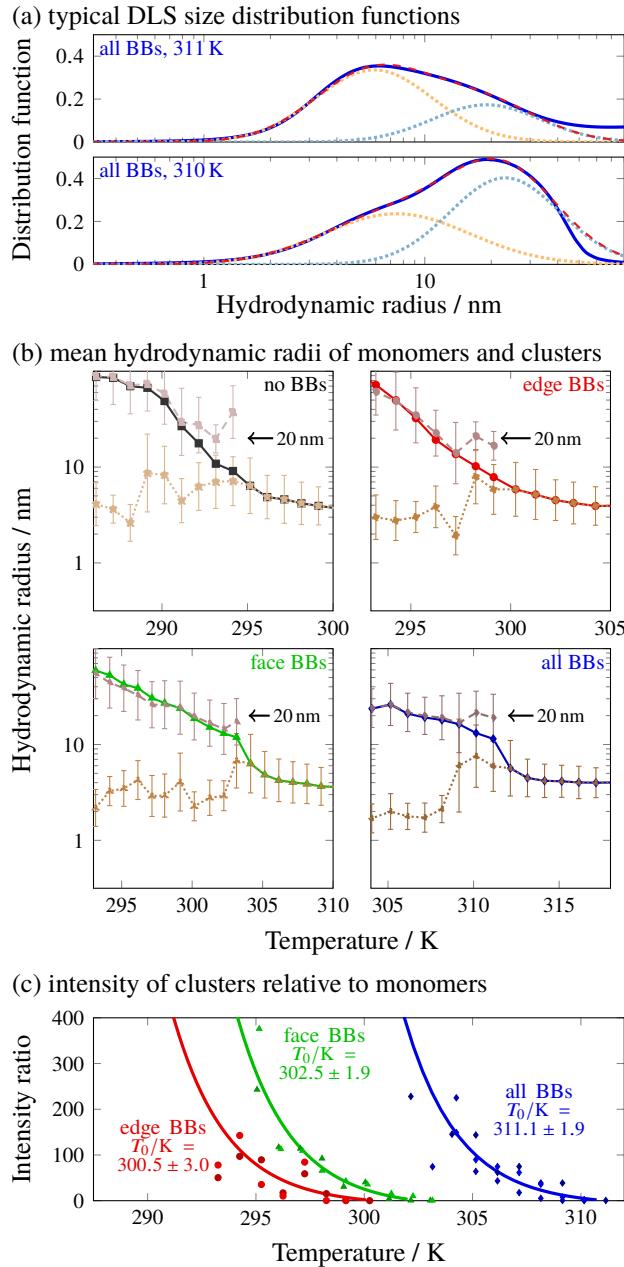


FIG. 4. (a) Representative size distribution functions computed from the DLS auto-correlation function (solid blue lines) at two temperatures. We also show a bimodal gaussian fit (dashed red lines) in each case, where the two populations are indicated by dotted lines. (b) The overall mean hydrodynamic radius and the mean hydrodynamic radius of the two gaussian fits for the four systems studied. Error bars indicate the widths of the fitted gaussian functions. (c) The ratio of the integrated intensities of the two peaks as a function of temperature, alongside an exponential fit $c \exp[-a(T - T_0)]$ for each system, with $1/a = 2.5$ K (see Methods).

To observe the self-assembly process in greater detail, we simulated the self-assembly of a coarse-grained model of DNA-brick cuboids using Monte Carlo dynamics that preserve the diffusion rates of various cluster sizes [13]. Considering a single copy of each cuboid, we calculated multiple self-assembly trajectories at constant temperature. Previous simulations of structures containing BBs [19] predicted variations in the nucleation behavior that agree with the relative ordering of the assembly temperatures for all four structural variants. However, in this work, we studied cuboid designs analogous to those used in our DLS experiments and chose the hybridization parameters to mimic the experimental conditions (see Methods). Remark-

ably, the highest temperatures at which nucleation occurs in these simulations (Fig. 3b) are in nearly quantitative agreement with the temperatures at which the all-BB, face-BB and edge-BB structures assemble in our experiments. It is important to note that, unlike in the experiments, all simulations were initialized from an unassembled solution with the total experimental monomer concentration at equimolar stoichiometry for each temperature. The simulated trajectories should thus only be compared to the initial transition in our DLS data, after which monomer depletion must be taken into account.

Our simulations indicate that the nucleation step determines the highest temperature at which self-assembly can occur [13, 21], while kinetic trapping arising from subunit misbonding and slow dynamics inhibits nucleation at low temperatures. This misbonding behavior is universal for all structures in our simulations (Fig. 3b) and is captured by a simple estimate of the probability of pairwise mis-interactions (Fig. 3c; see Methods). Consequently, the differences in nucleation behavior are primarily responsible for the varying widths of the temperature windows within which structures assemble.

Pre-formed clusters modify nucleation barriers

Evidence of high-temperature hybridization (see Sec. SI-2) suggests that stable dimers form in systems with BBs, in which the longer 16- or 24-bp domains bond either with scaffold bricks or with other BBs, at much higher temperatures than those at which nucleation occurs. This interpretation is also consistent with the increased mean R_h relative to the no-BB system in high-temperature DLS measurements (Fig. 3a). We therefore hypothesized that these pre-formed clusters play a key role in determining the nucleation rate. To test this idea, we ran simulations in which BB dimers were merged into single, permanently bonded units, mimicking the result of high-temperature hybridization in the experimental system. For example, simulations with merged edge BBs ('merged-A') showed that nucleation in this system is analogous to the edge-BB structure, indicating that these dimers form essentially to completion prior to structure nucleation (Fig. 3b).

This interpretation is also supported by free-energy calculations using a discrete combinatorial model (Fig. 3c). In this model [21, 22], we represent each distinct subunit type as a node in an abstract graph that describes the connectivity of the target structure (see Methods). Assuming that the number of subunits in a partially assembled cluster is a good reaction co-ordinate, this approach allows us to calculate the free-energy barrier to nucleation. In addition, by assuming that 16- and 24-bp domains hybridize completely at high temperatures, we can merge these bricks to explore the effects of the local subunit connectivity systematically. Comparison with simulations shows that this assumption is reasonable for the no-BB, edge-BB and face-BB structures, and that nucleation first occurs under conditions where the nucleation barrier is approximately $4 k_B T$. As in our simulations, we find that the face-BBs broaden the assembly window more significantly than edge-BBs do. This effect is crucial because it allows nucleation to begin at temperatures where misbonding is far less likely to occur.

Interestingly, we find that the assembly of the all-BB structure follows a three-step mechanism that is not well described by a one-dimensional free-energy landscape. Due to this limitation of our theory, the free-energy barrier for the all-BB system shown in Fig. 3c only pertains to the formation of an initial network of boundary bricks, a consequence of the fact that multiple 8-bp hybridizations can be formed between single pairs of pre-formed multimers in this system. These networks nucleate

prior to the incorporation of any interior scaffold strands, leading to extensive boundary-brick bonding and large cluster-size fluctuations in simulations at high temperatures (Fig. S5), in line with our high-temperature DLS measurements (Fig. 3a). Nucleation of the remainder of the structure then occurs in a separate assembly step, but only at temperatures slightly below the predicted scaffold-strand melting temperature, T_m (Fig. 3b,c).

Nucleation strongly affects self-assembly yield

Although the initial transitions in the DLS data are most striking, changes in the cluster-size distribution indicate that nucleation continues as annealing proceeds. By integrating the total scattering intensity associated with each peak in the cluster-size distribution, we determined that the intensity associated with the unassembled subunits falls off exponentially as the temperature decreases. Then, assuming that this intensity is proportional to the number density, we estimated the ratio of large- to small- R_h populations. The trends shown in Fig. 4c for the edge-BB, face-BB and all-BB structures follow the predictions of our free-energy calculations and nucleation theory, which indicate that the nucleation rate decreases with monomer depletion [9]. The measured ratios are consistent with the predicted functional form and temperature-scaling derived from the calculated nucleation barriers and the temperature-dependence of the hybridization free energies (see Methods). Since there must be some leftover subunits due to imperfect stoichiometry (measured to be approximately $\pm 10\%$), we do not expect the unassembled population to decay to zero in the experimental system. However, the associated intensity does not attain a constant level before the small- R_h peak falls below the detection range of the instrumentation.

The variations in nucleation behavior therefore provide an explanation for the extreme differences in yields among our structural variants, as well as the similarity between the ranking of the final yields and the order of the initial assembly transitions. At any given temperature, only a fraction of the potential structures ultimately form because nucleation slows as the reaction progresses [9]. Decreasing the temperature by means of an annealing protocol is therefore necessary to continue driving nucleation of additional structures. Structure designs that initially nucleate at higher temperatures benefit from a broader temperature range over which nucleation can occur, before reaching temperatures at which misbonding dominates.

Nucleation behavior and kinetic stability can be independently tuned

The differences among our cuboid variants do not affect the thermodynamic properties of the scaffold strands, which comprise the bulk of the structure. However, incorporating boundary bricks can, in principle, increase the kinetic stability of assembled structures. To examine this effect, we reversed the temperature ramp and used DLS to track the melting of pre-assembled structures. We observed sharp transitions in $\langle R_h \rangle$, corresponding to the complete melting of all structures in solution, but at considerably higher temperatures than the assembly transitions (Fig. 5a). Melting of the cuboid variants occurred in the reverse order of the assembly transitions, indicating that a strong bonding network of boundary bricks provides a kinetic barrier to disassembly. However, the most significant hysteresis was observed for structures that nucleate

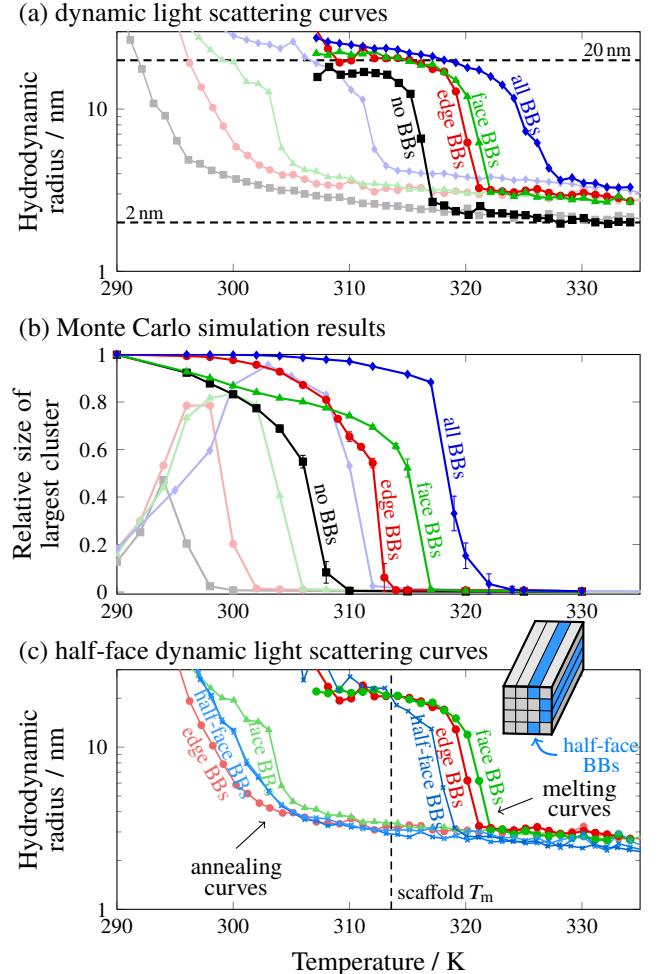


FIG. 5. The hysteresis in the size of the structure as a function of temperature as pre-assembled structures are heated in (a) experiments and (b) simulations. For comparison, the analogous cooling data from Fig. 3 are shown in faded colors. The scaffold melting temperature in the no-BB system is near 316 K, which is close to the predicted T_m , although the low-temperature R_h indicates that this structure is never fully formed. In (c), we show the annealing and melting curves for a half-face BB system, alongside a schematic illustration showing the BB locations. Importantly, the order in which the structures nucleate is different from the order in which they melt.

poorly, suggesting that the boundary bricks affect disassembly to a lesser extent than they affect nucleation.

Melting simulations of fully formed structures show similar trends (Fig. 5b). Analysis of the simulation trajectories reveals that scaffold bricks at the edges of the no-BB and face-BB structures disassemble first. The face-BB structures therefore lose bricks at lower temperatures than the edge-BB structures, although the face boundary bricks provide a larger barrier to complete disassembly. Disassembly occurs most abruptly in the case of the all-BB structures, with bricks initially dissociating from the unprotected ends of the structure. Nevertheless, these simulations are also consistent with a three-step disassembly mechanism, in which large networks of boundary-brick dimers persist for a few degrees above the apparent melting temperature (Fig. 5b).

To distinguish between the effects of nucleation and kinetic stability, we designed the ‘half-face-BB’ cuboid shown schematically in Fig. 5c. By incorporating face BBs on only one half of the structure, we predicted that we would see improved nucleation behavior, as with the full face-BB structure, but reduced kinetic stability. DLS measurements confirmed that

this structure initially nucleates at a temperature between the edge-BB and face-BB transitions (Fig. 5c), in agreement with our simulations and free-energy calculations (Fig. S6). The assembly yield is comparable to that of the edge-BB structure (Fig. S7), presumably because one half of the cuboid is not protected by boundary bricks and is thus more susceptible to aggregation from low-temperature misbonding.

While the assembly behavior is consistent with our expectations, DLS reveals that the half-face-BB structure melts *before* the edge-BB structure does. Melting occurs at a temperature closer to the T_m of the no-BB system, implying that the lack of boundary-brick protection on one face facilitates disassembly of the complete structure. Importantly, comparing the half-face-BB and no-BB systems highlights the crucial role of enhanced nucleation, as opposed to increased stability, in improving the yield. More generally, this example demonstrates that the nucleation behavior and thermal stability of DNA-brick nanostructures can be independently tuned and that their effects on the final yield can be differentiated.

Nucleation pathways are determined by the connectivity of pre-formed clusters

To identify the microscopic origin of the differing nucleation behaviors of the various designs, we calculated minimum-free-energy pathways using our theoretical model (Fig. 6a). For each structure, we determined the free-energy as a function of the number of independent subunits and the number of pre-formed dimers at a temperature where the nucleation barrier is approximately $5 k_B T$. The typical order in which dimers and scaffold strands are incorporated into a growing cluster is indicated by the minimum-free-energy nucleation pathways in Fig. 6a and illustrated in Figs S9–S11. We also show schematics of representative post-critical nuclei topologies. These nuclei are important because their formation is the rate-limiting step on each predicted nucleation pathway. However, because there are many topologically equivalent clusters within each structure, the free energy associated with each specific cluster depends on the hybridization free energies of the 8-bp domains involved.

These landscapes reveal crucial differences between the edge-BB and face-BB structures, which contain the same number of 48-nt BBs, and point to a key role of connections between the pre-formed dimers and interior scaffold strands. Because each subunit addition results in a loss of translational entropy, the free energy on a nucleation pathway only decreases when multiple 8-bp bonds are formed with a single subunit addition, forming a topologically closed cycle. In the absence of boundary bricks, the post-critical nucleus is a bicyclic cluster, comprising ten 8-bp bonds and nine subunits [21]. Yet by incorporating pre-formed dimers, fewer independent subunits are required to reach a post-critical nucleus. The 8-bp bonds can thus be weaker, leading to an elevated nucleation temperature. Despite the fact that the edge-BB and face-BB dimers have the same number of 8-nt domains for binding to other subunits, the topologies of the minimum-free-energy clusters in these structures are different: we find that the edge-BB structures require six subunits, including two BB dimers, to form a bicyclic cluster, while face-BB structures only require five subunits, including three dimers. Furthermore, simulation trajectories show that boundary-brick dimers comprise a larger fraction of the post-critical clusters in the face-BB structure than in the edge-BB structure (Fig. 6b,i–ii). We can also exclude the concentration of pre-formed dimers as the determining factor by comparing the half-face-BB and edge-BB structures, as

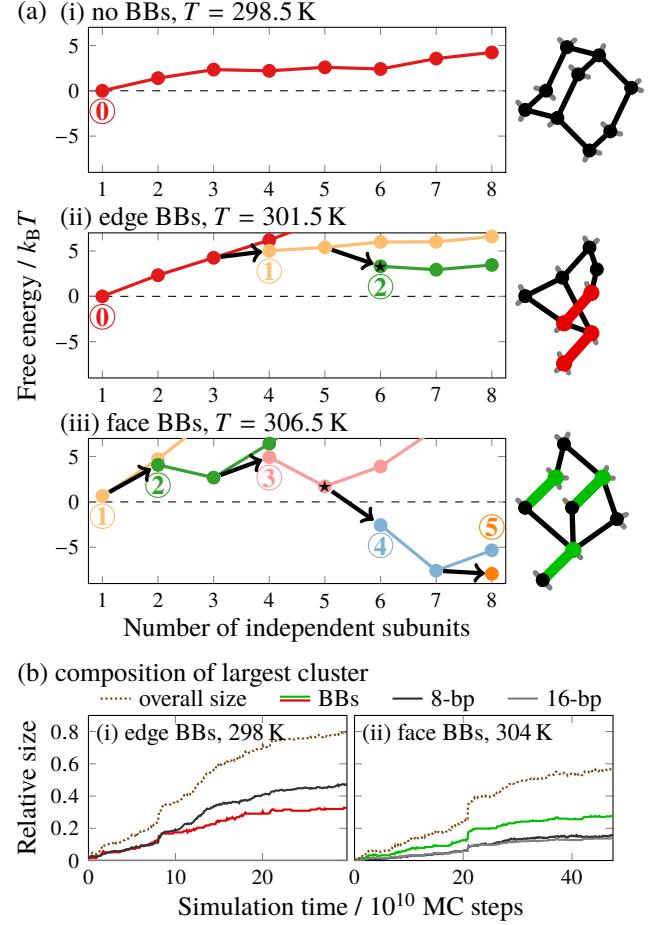


FIG. 6. (a) Free energy as a function of cluster size from theoretical calculations. The minimum-free-energy self-assembly pathway depends on the presence of pre-assembled multimers. Each step on the pathway incorporates an additional independent subunit, which may be either a single strand or a multimer formed by 16- and 24-bp hybridizations. The ringed colored numbers on each panel indicate the multimer count in the minimum-free-energy cluster. Arrows indicate multimer additions, while the solid lines follow pathways that grow via the incorporation of single-strand scaffold bricks and are colored for visual clarity. The free-energy landscapes are shown at temperatures at which the nucleation barrier is approximately $5 k_B T$. A schematic diagram of the connectivity of a representative post-critical nucleus, indicated with a star for the edge- and face-BB pathways, is also shown in each case. The no-BB post-critical cluster comprises 9 subunits and is just outside the range shown in the left-hand panel. (b) The relative size of the largest cluster in the system as a function of time for two individual Monte Carlo trajectories, and the fractions of boundary bricks and scaffold bricks in the system. The scaffold bricks are further divided into those that have only 8-bp hybridizations with neighboring bricks and those which also have 16-bp hybridizations. Trajectories start at roughly the point at which nucleation and growth first begin.

these systems contain the same number of pre-formed dimers but nucleate at significantly different temperatures.

Based on these findings, we hypothesized that by changing the local connectivity of the edge BBs, we might be able to reproduce the enhanced nucleation behavior of the face-BB structure. To test this hypothesis directly, we ran simulations of the edge-BB system in which we explicitly merged each edge dimer with one of its neighboring scaffold strands in the target structure (Fig. S8). The only difference between this structure ('merged-B') and a normal edge-BB dimer is that this additional connection to the interior scaffold strands, which would otherwise need to form spontaneously during the nucleation process and thus entail a loss of translational

entropy, has been fixed in place. This modification leads to post-critical nuclei that comprise five independent subunits, resulting in assembly behavior that is nearly analogous to that of the face-BB structure (Fig. 3b,c). Thus, although this particular modification cannot be directly achieved experimentally using DNA bricks, our simulations and theory both show that the addition of a single connection to the interior of the structure can alter the nucleation behavior significantly.

Design rules for enhanced nucleation

Based on our experimental, simulation and theoretical findings, we propose four design rules for enhancing the nucleation behavior and assembly yield in addressable systems:

1. The key determinant of the structure yield is the separation between the initial nucleation and misbonding temperatures. While the misbonding temperature is set by the pairwise interactions and the subunit concentrations, the nucleation temperature can be tuned through rational structure design. By contrast, changes to the subunit interactions that uniformly affect both correct and incorrect bonds are unlikely to improve the yield.
2. Altering the ‘valency’ of specific subunits to create multi-step pathways, for example by forming boundary-brick dimers at high temperatures, is a viable strategy for controlling nucleation, because it can change the number of independent subunits in the critical nucleus. On the other hand, tuning individual bond strengths is a less effective strategy for selecting a specific nucleation pathway, since the number of parallel pathways grows superextensively with the size of the target structure.
3. Controlling the topology of the critical nuclei is crucial. It may not be optimal simply to add high-valency subunits, as in the case of the edge BBs. Instead, efficient nucleation requires that the critical nuclei contain many stabilizing bonds but few subunits, favoring the formation of free-energy-reducing topological cycles earlier in the nucleation pathway. This is achieved in the case of the face-BB and merged-B structures by maximizing the number of bonds between the pre-formed dimers and the interior scaffold bricks.
4. Only a small portion of a structure needs to be optimized to achieve enhanced nucleation behavior. For example, comparison of the half-face-BB and no-BB systems shows that modifying fewer than 20 % of the subunits drastically raises the initial nucleation temperature and markedly improves the yield.

DISCUSSION

By combining dynamic light scattering with a coarse-grained theoretical model, we have shown that the ultimate yield of correctly assembled structures is largely determined by the nucleation pathway. As a specific example, we have investigated the role of nucleation kinetics in addressable self-assembly by modifying the bonding characteristics of specific subunits at the boundaries of a DNA-brick nanostructure. We have shown that the location and design of the altered subunits determine the free-energy landscape for self-assembly and control the temperature at which nucleation first becomes feasible. Moreover, the nearly quantitative agreement between the predictions of a

coarse-grained model and our experimental results allows us to rationalize these striking effects on the self-assembly behavior.

Taken together, our experiments and modeling establish essential design principles for improving the self-assembly of addressable nanostructures. In a typical annealing protocol, structures have a limited time and temperature window in which to form: at high temperatures, a large free-energy barrier inhibits nucleation, while at low temperatures, self-assembly is limited by kinetic arrest. The key to successful self-assembly is to increase the width of the temperature window over which nucleation can occur, thereby maximizing the thermodynamic segregation between the critical nucleation step and detrimental misbonding. This can be achieved by stabilizing the critical nucleus, which allows self-assembly to proceed when the sub-unit interactions are still relatively weak. To demonstrate this principle with DNA bricks, we have shown that the increased valency of boundary-brick dimers, which assemble at temperatures much higher than those at which nucleation can occur, lowers the free-energy barrier to nucleation by decreasing the entropic cost of forming a critical number of stabilizing bonds. However, this strategy only works if the high-valency subunits are optimally connected to the remainder of the structure, as evidenced by the difference in nucleation behavior between the edge-BB and face-BB structures. More generally, our results show that it is possible to use a relatively small number of high-valency subunits to design the nucleation pathway rationally and suggest that this approach is not necessarily limited to manipulating bricks at the boundaries of a structure.

Our experiments represent the first time that three-dimensional structure nucleation has been probed explicitly in the context of addressable self-assembly. This advance has been enabled by our use of dynamic light scattering, which allows us to observe structure growth, rather than merely the fraction of inter-subunit bonds that are formed at a given temperature. This distinction is particularly evident in the system evaluated here, where the initial nucleation temperature does not necessarily correlate with the maximal increase in DNA base-pairing. Furthermore, the cluster-size distributions that we obtain from dynamic light scattering resolve the populations of unincorporated bricks, complete structures and aggregates, making it possible to track the evolution of these species throughout the course of an annealing protocol. These measurements have provided explicit experimental evidence that DNA bricks self-assemble via a nucleation-and-growth mechanism and have verified more general theoretical predictions regarding the relationship between the design of addressable structures and their nucleation kinetics.

The excellent agreement between the predictions of our theoretical model and our experimental results demonstrates that our coarse-grained approach captures the fundamental physics of addressable self-assembly. This agreement gives us confidence that our theory and simulations can be used to guide rational design strategies for complex self-assembly, not only in the context of DNA bricks specifically, but — precisely because of the generality of the models used — also for optimizing addressable systems more broadly. We anticipate that the principles established here will therefore guide efforts to design the nucleation behavior of colloidal systems such as supramolecular and nanoparticle lattices [33–35], protein nanostructures [36] and DNA-origami-based systems with programmable interactions [37]. For example, analogous pre-nucleation clusters could be constructed by forming high-temperature bonds between caged nanoparticles [38] or by directly introducing a small population of dumbbell-like subunits. Alternatively, the connectivity of specific subunits could be altered by changing the arrangement of directional

patches on colloidal particles [39]. As we have demonstrated here, successful implementation will require knowledge of the effects of such modifications on the critical nucleus for structure assembly, which dictates the optimal design strategy for any specific system.

MATERIALS AND METHODS

In the Supplementary Information, we describe how we chose the DNA sequences for the strands for each system studied. We also provide complete details of the annealing protocols, the conditions used in AFM and gel electrophoresis, and the protocols used when obtaining fluorescence and light scattering data in the Extended Methods. Supporting data are available at the University of Cambridge Data Repository, doi:10.17863/cam.XXXX.

Structure annealing

Structures were assembled using a strand concentration of 153 nm per sequence in a buffer of 15 mM MgCl₂, 0.5 mM EDTA and 5 mM Tris at pH 8. Strands in the reaction mixture were denatured at 90 °C for 10 min and then gradually cooled via either (i) a 15.2 h protocol (reciprocal cooling rate 12 min K⁻¹) or (ii) a 66 h protocol (reciprocal cooling rate 52 min K⁻¹).

Atomic force microscopy

Samples from annealing protocol (ii) were immobilized for 10 min on poly-L-ornithine coated mica discs and imaged in liquid in intermittent contact mode using a BioLever Mini cantilever and JPK Nanowizard 3 AFM.

Agarose gel electrophoresis

Structures were analyzed via gel electrophoresis on a gel made from 2 wt% agarose in 0.5 × TBE and 10 mM MgCl₂. Electrophoresis was performed at 80 V and 4 °C for 2 h. The gel was post-stained with ethidium bromide and the yield was estimated using GelBandFitter software [40].

Fluorescence annealing

Annealing protocol (i) was used for fluorescence annealing experiments (see Sec. SI-2) with 10 nm SYBR green I solution [41] added to the reaction mixture. The fluorescence signal was measured, as a function of time and temperature, with an ABI Prism 7900HT-Fast Real Time PCR system at 488 nm.

Static and dynamic light scattering

Using annealing protocol (i), light scattering measurements were performed in the last 2 min of each temperature step. 20 μL samples were filled into ZEN2112 quartz cuvettes (Malvern), covered by molecular biology grade mineral oil, and sealed with a plastic lid that was further fixed with tape. Light scattering was measured using a Malvern Zetasizer NanoZSP apparatus at an angle of 173°. For DLS, the intensity auto-correlation function was computed from 12 measurements at

10 s intervals. Cluster-size distributions were determined from the auto-correlation data using multiple regularization methods to verify their robustness (see Fig. S1).

Reference cluster-size calculations

The hydrodynamic radius of a freely jointed chain is $R_h = (3\pi N \ell^2 / 128)^{1/2}$ [42], where N is the number of segments and ℓ is the length of a segment. To estimate the hydrodynamic radius for single-stranded DNA, we use a typical Kuhn length of $\ell = 4.45$ nm and length per DNA base of $d = 0.676$ nm [43]. A 32-nt brick comprises $N \approx 32 \times d/\ell = 4.9$ Kuhn segments and hence $R_h \approx 2.7$ nm. Since the quantities ℓ and d we used here do not correspond to the temperatures and salt concentrations used in our experiments, this calculation only provides us with a very rough estimate of the magnitude of R_h .

For cylindrical structures, the translational diffusion coefficient is given by [44]

$$D_{tr} = \frac{k_B T [\ln(L/d) + \gamma]}{3\pi\eta L}, \quad (1)$$

where k_B is the Boltzmann constant, T is the absolute temperature, η is the viscosity of the medium, L is the cylinder length, d is the cylinder diameter, and γ is an end-effect correction given by $\gamma = 0.312 + 0.565(d/L) + 0.1(d/L)^2$. Assuming that the hydrodynamics of a cylinder are well approximated by a sphere with hydrodynamic radius R_h , we can equate this diffusion coefficient to that of a sphere using the Stokes–Einstein–Smoluchowski equation,

$$D_{tr} = \frac{k_B T}{6\pi\eta R_h}, \quad (2)$$

leading to an approximate hydrodynamic radius of $R_h = (L/2)[\ln(L/d) + \gamma]^{-1}$. Our cuboid target structure can be treated as a cylinder with approximate dimensions $L \approx 86$ nm and $d \approx 15$ nm. In DLS experiments, we therefore expect our target structure to have a hydrodynamic radius of $R_h \approx 20$ nm.

Monte Carlo simulations

We performed lattice Metropolis Monte Carlo simulations of DNA brick self-assembly using a simple coarse-grained potential [13, 17, 19]. Every DNA brick was represented as a ‘patchy particle’ with four patches corresponding to its four domains, each of which was assigned a specific unique sequence, chosen randomly but with the constraint that patches that are bonded in the target structure have complementary DNA sequences. The interaction energies correspond to the hybridization free energies of these sequences obtained from the SantaLucia parameterization [45]. When computing these hybridization free energies, we used a salt correction [46] corresponding to salt concentrations of [Na⁺] = 0 and [Mg²⁺] = 0.015 M. In the simulations reported here, we used a system of 550 bricks in a box with lattice parameter $a\sqrt{3}$, where $a\sqrt{3}$ is the shortest possible distance between any two particles. Assuming typical brick dimensions of $(a\sqrt{3})^3 \approx 2.5$ nm × 2.5 nm × 2.7 nm [8, 18], this set-up corresponds to a concentration of 153 nm. We accounted for boundary bricks by imposing rigid bonds between dimers (or, in certain cases, larger multimers) of these patchy particles that would be merged into a single boundary brick in experiment [19]. Particles connected in this way remain at a fixed distance and dihedral angle to one another throughout the

simulation. Non-interacting patches on the outside of the target structure were assigned a poly-T sequence in order to passivate them.

Free-energy calculations

All free-energy calculations were carried out using the abstract-graph model described in Ref. 21. The free energy of a particular cluster g , comprising a set of subunits $\mathcal{V}(g)$, is

$$\frac{F_g}{k_B T} = - \sum_{\substack{i,j \in \mathcal{V}(g) \\ j \in \mathcal{E}(i)}} \frac{\epsilon_{ij}}{2} - (N_g - 1) \ln \frac{\rho}{q_{\text{rot}}} + (N_g - B_g - 1) \ln q_{\text{dih}}, \quad (3)$$

where ρ is the strand concentration, N_g is the number of subunits in the cluster, $\mathcal{E}(i)$ indicates the set of strands that are neighbors of strand i in the target structure, and the dimensionless bond strengths are $\epsilon_{ij} = \ln[\exp(-\Delta G_{ij}/k_B T) - 1]$. We determined ΔG_{ij} for each pair of complementary sequences i, j in the experimental systems using the SantaLucia parameterization described above. Each subunit was assumed to have $q_{\text{rot}} = 4$ rotational degrees of freedom, and each single bond was assumed to have $q_{\text{dih}} = 3$ dihedral degrees of freedom; these values were chosen to match the Monte Carlo simulations. B_g refers to the number of ‘bridges’ in the graph g [21]. The cluster free energy as a function of the number of correctly bonded subunits is

$$F(N) = -k_B T \ln \sum_g \mathbf{1}(N_g = N) \exp(-F_g/k_B T), \quad (4)$$

where $\mathbf{1}(\cdot)$ is the indicator function. $F(N)$ was calculated using the efficient Monte Carlo approach described in Ref. 21. Similarly, in Fig. 6a, the cluster free energy was calculated as a function of the total number of subunits and the number of pre-formed dimers.

The melting temperature, T_m , of an infinite lattice of scaffold strands was estimated based on the mean of the 8-bp scaffold-strand hybridization free energies by solving the equation

$$(z/2)\overline{\Delta G(T_m)} = k_B T_m \ln(\rho/q_{\text{rot}} q_{\text{dih}}), \quad (5)$$

where $z = 4$ is the lattice co-ordination number.

Misbonding calculations

We used a two-state model (i.e. bonded or not bonded) to calculate the probability that a strand forms at least one misinteraction, assuming that no domains are correctly hybridized. We found the longest complementary subsequence for each pair of strands i and j that are not neighbors in the target structure and calculated the associated hybridization free energy, $\Delta G_{\min,ij}$. [In cases where there are multiple complementary subsequences $\{s\}$ of the same length for a given pair of strands, we calculated the Boltzmann-weighted sum, $\Delta G_{\min,ij} = -k_B T \ln \sum_s \exp(-\Delta G_s/k_B T)$.] We then computed the probability that a strand i forms a misinteraction, $p_{\text{mis},i}$,

$$p_{\text{mis},i} = \frac{Z_{\text{mis},i}}{1 + Z_{\text{mis},i}}, \quad (6)$$

$$Z_{\text{mis},i} = \sum_{j \notin \mathcal{E}(i)} \rho \exp(-\Delta G_{\min,ij}/k_B T). \quad (7)$$

When computing the probability of scaffold-strand misbonding, the index i represents a scaffold strand, while the index j runs

over all strands in the system. Although this approach is only approximate, it captures the competition between designed and incorrect bonding that is seen in the MC simulations (Fig. 3b,c) remarkably well.

Cluster population ratios

We assume that annealing is slow, so that nucleation is always rate-limiting. We can write the nucleation barrier height as

$$F^\dagger = -n^\dagger k_B T \ln \rho_m + \varepsilon(T) E^\dagger + C, \quad (8)$$

where n^\dagger is the number of independent subunits in the critical nucleus, E^\dagger is the number of 8-bp bonds in the critical nucleus, and C is a constant that accounts for the (effective) number of parallel nucleation pathways, as well as the rotational entropy terms. The bond energy, ε , is a decreasing function of temperature, while the per-species monomer concentration, ρ_m , also decreases as the reaction progresses. Initially, we have ρ_T of each species. For simplicity, let us assume that, given this initial monomer concentration, the barrier is infinitely high above some critical temperature T_0 . Nucleation begins once $T \leq T_0$, where F^\dagger is finite. (In reality, nucleation can begin as soon as the target structure, or any large cluster, becomes thermodynamically stable. However, the nucleation rate is proportional to $\exp(-F^\dagger/k_B T)$, so the highest barrier that can be crossed depends on the cooling rate.)

Nucleation will proceed at a given temperature until ρ_m decreases to a point where F^\dagger is again insurmountable. Denoting this critical barrier height by F_0^\dagger , we can relate the final monomer concentration at any temperature to the initial concentration at the critical temperature,

$$F_0^\dagger - C = -n^\dagger k_B T \ln \rho_m(T) + \varepsilon(T) E^\dagger \quad (9)$$

$$= -n^\dagger k_B T_0 \ln \rho_T + \varepsilon(T_0) E^\dagger, \quad (10)$$

so that

$$\frac{\rho_T}{\rho_m} = \exp\left\{-\frac{E^\dagger}{n^\dagger} \left[\frac{\varepsilon(T)}{k_B T} - \frac{\varepsilon(T_0)}{k_B T_0} \right]\right\}. \quad (11)$$

The ratio of the scattering intensities originating from each population is

$$\frac{I_c}{I_m} = \frac{R_{h,c}^6}{N R_{h,m}^6} \left(\frac{\rho_T - \rho_m}{\rho_m} \right) \quad (12)$$

$$= \frac{R_{h,c}^6}{N R_{h,m}^6} \left(\exp\left\{-\frac{E^\dagger}{n^\dagger} \left[\frac{\varepsilon(T)}{k_B T} - \frac{\varepsilon(T_0)}{k_B T_0} \right]\right\} - 1 \right), \quad (13)$$

where N is the number of distinct subunits in the target structure. Assuming that the hydrodynamic radii of the monomers and clusters remain constant, and because $\varepsilon/k_B T$ is a nearly linear function of T in the range of interest (Fig. S12), we expect the intensity ratio to have a functional form

$$\frac{I_c}{I_m} = \text{const} \times \exp[-a(T - T_0)], \quad (14)$$

where $a = (E^\dagger/n^\dagger)(d\beta\varepsilon/dT)$ and $\beta = 1/k_B T$. Using a linear fit to the energies as a function of temperature at temperatures of interest (Fig. S12), $d\beta\varepsilon/dT \approx 0.34 \text{ K}^{-1}$. From the theoretical free-energy profiles (Fig. 6), we know that for edge BBs, $E^\dagger/n^\dagger = 7/6$, whilst for face BBs, the ratio is $6/5$. Hence we can estimate that $1/a \approx 2.5 \text{ K}$.

To calculate the intensity associated with each peak in the DLS data, we first fitted a sum of gaussians to the distribution function, $f(R_h)$. We then numerically integrated the peak associated with the gaussian function f_g , according to

$$I_{c/m} = \int_0^{\infty} \min[f_{g,c/m}(R_h), f(R_h)] d \ln R_h. \quad (15)$$

ACKNOWLEDGMENTS

We thank Daan Frenkel for helpful discussions. This work was supported by the Engineering and Physical Sciences Research Council [Programme Grant EP/I001352/1], the European Regional Development Fund [100185665], Fraunhofer Attract Funding [601683] and the National Institutes of Health [Grant F32GM116231].

-
- [1] S. Whitelam and R. L. Jack, *Annu. Rev. Phys. Chem.* **66**, 143 (2015).
- [2] D. Frenkel, *Nat. Mater.* **14**, 9 (2015).
- [3] J. Zhang, Z. Sun, and B. Yang, *Curr. Opin. Colloid In.* **14**, 103 (2009).
- [4] E. Sowade, T. Blaudeck, and R. R. Baumann, *Cryst. Growth Des.* **16**, 1017 (2016).
- [5] J. I. Urgel, D. Écija, G. Lyu, R. Zhang, C.-A. Palma, W. Auwärter, N. Lin, and J. V. Barth, *Nat. Chem.* **8**, 657 (2016).
- [6] A. Reinhardt, J. S. Schreck, F. Romano, and J. P. K. Doye, *J. Phys.: Condens. Matter* **29**, 014006 (2017).
- [7] P. F. Damasceno, S. C. Glotzer, and M. Engel, *J. Phys.: Condens. Matter* **29**, 234005 (2017).
- [8] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, *Science* **338**, 1177 (2012).
- [9] W. M. Jacobs and D. Frenkel, *J. Am. Chem. Soc.* **138**, 2457 (2016).
- [10] L. Cademartiri and K. J. M. Bishop, *Nat. Mater.* **14**, 2 (2015).
- [11] R. Schulman and E. Winfree, *SIAM J. Comput.* **39**, 1581 (2010).
- [12] J. Zenk and R. Schulman, *PLoS ONE* **9**, e111233 (2014).
- [13] A. Reinhardt and D. Frenkel, *Phys. Rev. Lett.* **112**, 238103 (2014).
- [14] Z. Zeravcic, V. N. Manoharan, and M. P. Brenner, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15918 (2014).
- [15] J. Madge and M. A. Miller, *J. Chem. Phys.* **143**, 044905 (2015).
- [16] J. Madge and M. A. Miller, *Soft Matter* **13**, 7780 (2017).
- [17] A. Reinhardt, C. P. Ho, and D. Frenkel, *Faraday Discuss.* **186**, 215 (2016).
- [18] A. Reinhardt and D. Frenkel, *Soft Matter* **12**, 6253 (2016).
- [19] H. K. Wayment-Steele, D. Frenkel, and A. Reinhardt, *Soft Matter* **13**, 1670 (2017).
- [20] D. J. Wales, *J. Chem. Phys.* **146**, 054306 (2017).
- [21] W. M. Jacobs, A. Reinhardt, and D. Frenkel, *J. Chem. Phys.* **142**, 021101 (2015); *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6313 (2015).
- [22] W. M. Jacobs and D. Frenkel, *Soft Matter* **11**, 8930 (2015).
- [23] The term ‘nucleation’ in the context of DNA self-assembly is occasionally used to refer to the initial thermodynamically disfavored formation of a few base pairs of a double strand, which is then followed by zipping [31]. We use the term in the more traditional sense to mean the formation of a cluster of the target structure.
- [24] J.-P. J. Sobczak, T. G. Martin, T. Gerling, and H. Dietz, *Science* **338**, 1458 (2012).
- [25] J. M. Majikes, J. A. Nash, and T. H. LaBean, *Nanoscale* **9**, 1637 (2017).
- [26] J. Song, J.-M. Arbona, Z. Zhang, L. Liu, E. Xie, J. Elezgaray, J.-P. Aime, K. V. Gothelf, F. Besenbacher, and M. Dong, *J. Am. Chem. Soc.* **134**, 9844 (2012).
- [27] T. Kato, R. P. Goodman, C. M. Erben, A. J. Turberfield, and K. Namba, *Nano Lett.* **9**, 2747 (2009).
- [28] C. Myhrvold, M. Baym, N. Hanikel, L. L. Ong, J. S. Gootenberg, and P. Yin, *Nat. Commun.* **8**, 14698 (2017).
- [29] X. Wei, J. Nangreave, S. Jiang, H. Yan, and Y. Liu, *J. Am. Chem. Soc.* **135**, 6165 (2013).
- [30] S. Jiang, F. Hong, H. Hu, H. Yan, and Y. Liu, *ACS Nano* **11**, 9370 (2017).
- [31] A. V. Pinheiro, J. Nangreave, S. Jiang, H. Yan, and Y. Liu, *ACS Nano* **6**, 5521 (2012).
- [32] X. Wei, J. Nangreave, and Y. Liu, *Acc. Chem. Res.* **47**, 1861 (2014).
- [33] R. J. Macfarlane, B. Lee, M. R. Jones, N. Harris, G. C. Schatz, and C. A. Mirkin, *Science* **334**, 204 (2011).
- [34] M. Huang, C.-H. Hsu, J. Wang, S. Mei, X. Dong, Y. Li, M. Li, H. Liu, W. Zhang, T. Aida, W.-B. Zhang, K. Yue, and S. Z. D. Cheng, *Science* **348**, 424 (2015).
- [35] H. Lin, S. Lee, L. Sun, M. Spellings, M. Engel, S. C. Glotzer, and C. A. Mirkin, *Science* **355**, 931 (2017).
- [36] J. B. Bale, S. Gonon, Y. Liu, W. Sheffler, D. Ellis, C. Thomas, D. Cascio, T. O. Yeates, T. Gonon, N. P. King, and D. Baker, *Science* **353**, 389 (2016).
- [37] T. Gerling, K. F. Wagenbauer, A. M. Neuner, and H. Dietz, *Science* **347**, 1446 (2015).
- [38] W. Liu, M. Tagawa, H. L. Xin, T. Wang, H. Emamy, H. Li, K. G. Yager, F. W. Starr, A. V. Tkachenko, and O. Gang, *Science* **351**, 582 (2016).
- [39] Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck, and D. J. Pine, *Nature* **491**, 51 (2012).
- [40] M. I. Mitov, M. L. Greaser, and K. S. Campbell, *Electrophoresis* **30**, 848 (2009).
- [41] H. Zipper, H. Brunner, J. Bernhagen, and F. Vitzthum, *Nucleic Acids Res.* **32**, e103 (2004).
- [42] I. Teraoka, *Polymer solutions: an introduction to physical properties* (Wiley, New York, 2002).
- [43] Q. Chi, G. Wang, and J. Jiang, *Physica A* **392**, 1072 (2013).
- [44] J. García de la Torre and V. A. Bloomfield, *Q. Rev. Biophys.* **14**, 81 (1981).
- [45] J. SantaLucia Jr and D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415 (2004).
- [46] R. T. Koehler and N. Peyret, *Bioinformatics* **21**, 3333 (2005).
- [47] S. W. Provencher, *Comput. Phys. Commun.* **27**, 213 (1982).
- [48] S. Hansen, *Eur. Biophys. J.* **46**, 1 (2017).
- [49] W. Wang, T. Lin, S. Zhang, T. Bai, Y. Mi, and B. Wei, *Nucleic Acids Res.* **44**, 7989 (2016).
- [50] Y. Ke, L. L. Ong, W. Sun, J. Song, M. Dong, W. M. Shih, and P. Yin, *Nat. Chem.* **6**, 994 (2014).
- [51] S. Giglio, P. T. Monis, and C. P. Saint, *Nucleic Acids Res.* **31**, e136 (2003).
- [52] D. Suh and J. B. Chaires, *Bioorg. Med. Chem.* **3**, 723 (1995).
- [53] K. M. Ririe, R. P. Rasmussen, and C. T. Wittwer, *Anal. Biochem.* **245**, 154 (1997).

SUPPLEMENTARY INFORMATION

1 EXTENDED METHODS

1.1 Annealing protocols

Sequences for the target cuboid structure are given in Sec. SI-4. For the all-BB system, sequences were taken from Ref. 8 (Table S8), and DNA strands were appropriately decoupled to split the relevant boundary bricks for the face-BB, edge-BB and no-BB systems. All sequences were purchased from Eurofins Genomics in 100 μm stocks in ddH₂O, and then pooled using a Tecan Genesis Workstation 150 liquid handling robot. We used a strand concentration of 153 nM in 1 \times assembly buffer, i.e. a solution of 15 mM MgCl₂, 0.5 mM EDTA and 5 mM Tris, pH 8. The strand solution was denatured at 90 °C for 10 min and then gradually cooled. We used two linear cooling protocols: (i) in the 15.2 h protocol, the reciprocal cooling rate was 12 min K⁻¹, and (ii) in the 66 h protocol, it was 52 min K⁻¹. The annealed samples were stored at 4 °C. Prior to end-point DLS measurements (Fig. S4), samples were supplemented with 2.5 mM EDTA to reduce high molecular weight contamination.

1.2 Atomic force microscopy

Samples were prepared following the 66 h annealing protocol. A freshly cleaved mica disc was coated with 100 μL of 0.5 wt% poly-L-ornithine solution for 5 min and rinsed three times with 1 \times assembly buffer. In order to be able to image samples in liquid mode, an acrylic glass ring was glued by Thin Pour (Reprorubber) onto a slide to surround the mica disc and form a fluid cell. For each sample, 1.5 pmol per brick was deposited on the coated mica for 10 min. Afterwards, the cell was filled with 1 \times assembly buffer and imaged using the JPK Nanowizard 3 atomic force microscope and a BioLever Mini cantilever in intermittent contact mode in liquid. Images were recorded with a target amplitude of 15 nm.

1.3 Agarose gel electrophoresis

Assembly of DNA brick structures was confirmed by non-denaturing agarose gel electrophoresis. Samples (300 fmol per brick) were analyzed on a gel made from 2 wt% agarose in 0.5 \times TBE and 10 mM MgCl₂. Electrophoresis was performed at 80 V and 4 °C for 2 h. The gel was post-stained with 0.5 $\mu\text{g mL}^{-1}$ ethidium bromide solution and scanned in using the Intas GDS gel set instrument for structure visualization. To estimate structure yield, the band intensity was approximated by fitting densitometry profiles with an SQP algorithm to gaussian functions using the GelBandFitter software [40]. The mass of the structure fractions was estimated via a 1 kb standard (GeneRuler, ThermoFischer Scientific) and related to the total mass loaded of 850 ng.

1.4 Fluorescence annealing

In fluorescence annealing experiments, the same conditions as in the 15.2 h annealing protocol were used, except that 10 nM SYBR green I solution [41] was added to the strand mixture. SYBR green I in buffer solution was analyzed as a negative

control. Samples were placed on a MicroAmp Fast Plate 96-well tray and sealed with adhesive film. The plate was loaded onto the ABI Prism 7900HT-Fast Real Time PCR system, with dye excitation effected by an argon ion laser at 488 nm. The fluorescence signal was detected at 525 nm every 8.5 s and averaged over time at each temperature, and its derivative with respect to temperature was computed numerically. The data were smoothed out via a gaussian filter with standard deviation of 1.5 K.

1.5 Static and dynamic light scattering

The same conditions as in the 15.2 h annealing protocol were used and the measurement was performed in the last 2 min of the 12 min cooling step. 20 μL samples were filled into ZEN2112 quartz cuvettes (Malvern), covered by molecular biology grade mineral oil, and sealed with a plastic lid that was further fixed with tape. Light scattering was measured using a Malvern Zetasizer NanoZSP apparatus at an angle of 173°. The viscosity of the samples was determined at five temperatures spanning the region of interest and fitted to $\eta/(10^{-5} \text{ Pa s}) = 1.78 \times \exp[617/(T/\text{K} - 138.5)]$. The refractive index was measured to be 1.331. For dynamic light scattering, the intensity auto-correlation function was computed from 12 measurements at 10 s intervals.

When analyzing DLS data for solutions comprising a range of particle sizes, the inverse Laplace transform used to obtain a particle size distribution from the intensity auto-correlation function is not uniquely determined [47], and the choice of fitting functions and parameters can significantly affect the final result [48]. We have therefore computed a number of fits to the distribution of hydrodynamic radii using several regularization methods, including CONTIN and maximum entropy constraints, to verify that the data are robust.

2 MONITORING STRAND HYBRIDIZATION

Fluorescence measurements

We monitored the progress of domain hybridization during the annealing protocol via fluorescence, using SYBR green I as a double-stranded DNA probe [24]. We observed a dominant maximum in the fluorescence derivative between 335 K and 350 K for all structures with boundary bricks (Fig. S0a), indicating a significant amount of base-pairing at relatively high temperatures. However, as we discuss in the main text, no complete structures are capable of forming at these temperatures.

Comparison with theoretical annealing curves suggests that the assembly of boundary-brick structures is a two-step process. To demonstrate this, we show in Fig. S0b the temperature derivative of the equilibrium number of base pairs in a solution of monomers and dimers [45, 46], assuming that stable misbonding between non-complementary domains cannot occur. The high-temperature transitions correspond to the hybridization between pairs of boundary bricks (where continuous 24-bp segments are hybridized) or between one scaffold strand and one boundary brick (with 16-bp hybridized segments). Consequently, the assembly of the full structure must occur in the presence of these pre-formed clusters. These calculations also indicate that the fluorescence-signal contributions from each domain length overlap significantly, since the domain melting

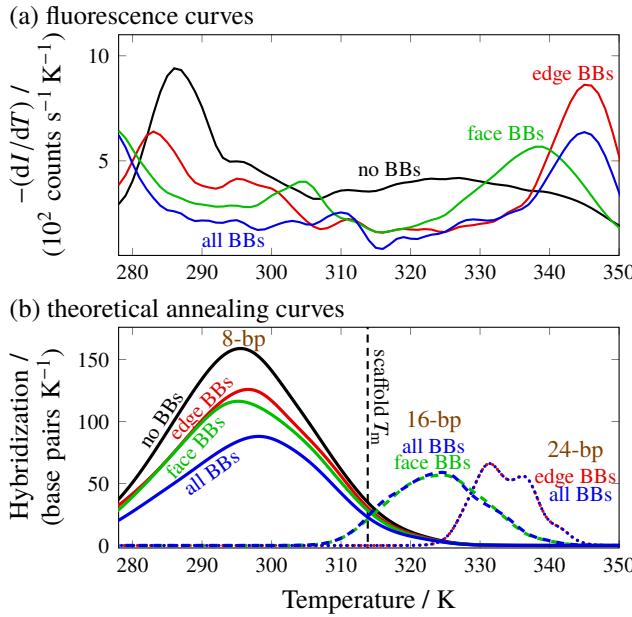


FIG. S0. (a) Derivative of the fluorescence signal I with respect to the temperature obtained from a 15.2 h annealing protocol. (b) The corresponding theoretical annealing curves. In agreement with the experimental fluorescence data, contributions from 16- (dashed lines) and 24-bp (dotted lines) hybridizations dominate at higher temperatures. We also predict the melting temperature T_m of a periodic lattice of scaffold strands for the no-BB structure, which approximates the highest temperature at which the cores of the structures are thermodynamically stable.

temperatures vary widely according to their specific sequences, and each hybridization reaction tends to occur over a broad ($\gtrsim 10$ K) range of temperatures. In particular, the theoretical annealing curves predict a broad maximum associated with the 8-bp domains near 295 K.

Analysis of fluorescence data has previously been used to distinguish between single- and multi-step assembly mechanisms for DNA tile systems with varying domain lengths. For example, a similar step-wise assembly process was seen in DX-tile structures comprising short (10- and 11-bp) and long (21-bp) hybridizations, and the presence of two distinct maxima in the fluorescence derivative was interpreted as evidence of hierarchical assembly [49]. By contrast, fluorescence measurements of DNA-brick crystallization using equal-length domains exhibited no evidence of hierarchical self-assembly [50]. In our measurements, there appear to be multiple local maxima in the annealing curves at temperatures below $T_m \approx 315$ K, the highest temperature at which our theoretical calculations predict that a lattice of scaffold strands will be thermodynamically stable. However, these signals are significantly weaker than the higher-temperature hybridizations which dominate the fluorescence signal. Interpreting the lower-temperature maxima is additionally hindered by several known sources of bias, including high background signals [41] and the preferential binding of SYBR green I to GC-rich sequences [51]. Furthermore, the intercalating SYBR green I probes distort the double-helical structure of DNA molecules [52], which increases their melting temperatures [53] and precludes a quantitative analysis.

Hybridization calculations

All hybridization calculations were carried out using the SantaLucia parameterization and the solution conditions described in the Methods section of the main text. In this section, we consider a two-state model (i.e. bonded or not bonded) for each domain and examine the simple case where pairs of strands hybridize to form dimers, but not larger multimers. We denote the hybridization free energy between complementary domains on a pair of strands i and j by ΔG_{ij} . The equilibrium probability that a strand i is correctly hybridized with its putative neighbor strand j is

$$p_{ij}(T) = \frac{\rho \exp(-\Delta G'_{ij}/k_B T)}{1 + \rho \exp(-\Delta G'_{ij}/k_B T)}, \quad (16)$$

where ρ is the dimensionless strand number density, T is the absolute temperature, and we assume that all species are present in equal concentrations. The hybridization free energies are written as $\Delta G'_{ij}$ to indicate that we use the longest complementary subsequence of strands i and j , which, due to the random sequence design, is occasionally longer than the intended domain length. To calculate the total change in base-pairing during an annealing protocol (Fig. S0b), we took the temperature derivative of the ensemble average of correctly formed base pairs,

$$\text{Hybridization} = -\frac{d}{dT} \sum_{\substack{i < j \\ j \in \mathcal{E}(i)}} l_{ij} p_{ij}(T), \quad (17)$$

where l_{ij} is the length of each hybridizing domain.

3 SUPPLEMENTAL FIGURES

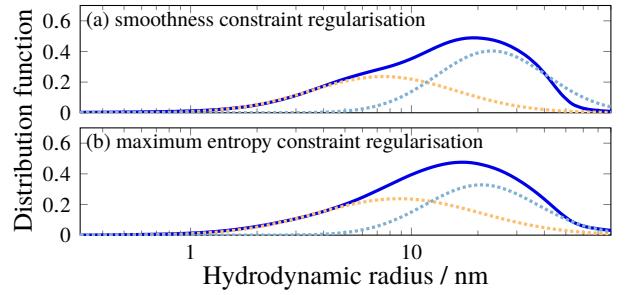


FIG. S1. Example size distribution functions (solid blue lines) for the all-BB system at 310 K determined from the intensity auto-correlation function using two regularization methods. We used (a) a smoothness constraint functional [47] with a smoothing exponent [48] of $8/5$ and (b) a maximum entropy constraint with a gaussian prior distribution [48]. We also performed the analysis with a range of additional smoothing exponents and via entropy maximization with a uniform prior distribution. The various approaches agree well, although the agreement is not perfect for individual data points. Nevertheless, the trends for the distributions and the gaussian fits (dotted lines) are the same regardless of the regularization procedure used. We used a smoothness constraint regularization (corresponding to panel (a)) for the data we reported in the main text. We found that no significant differences arise when performing the full analysis with the maximum entropy constraint, indicating that the data are robust to the choice of regularization method.

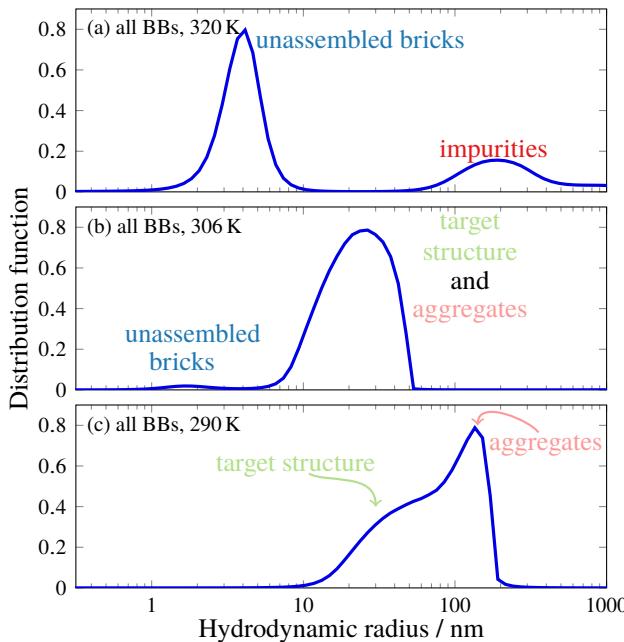


FIG. S2. Size distribution functions for the all-BB system from DLS experiments at several temperatures. At low temperatures [(b)], it is difficult to resolve the ‘target’ structures from larger aggregates, and the mean size reported in the main text therefore steadily increases. However, at even lower temperatures [(c)], where aggregates become sufficiently large, it again becomes possible to resolve the target structure (at approximately 25 nm). At high temperatures [(a)], by contrast, a peak can be resolved at very large hydrodynamic radius, although the overall scattering intensity is very low. This large-size peak is likely to be due to dust; it gradually disappears as the system is cooled and the overall scattering intensity increases and becomes dominated by the forming structures.

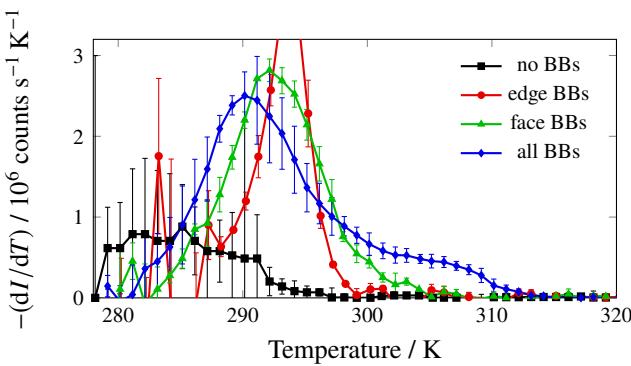


FIG. S3. Static light scattering curves for self-assembly following the 15.2 h annealing protocol.

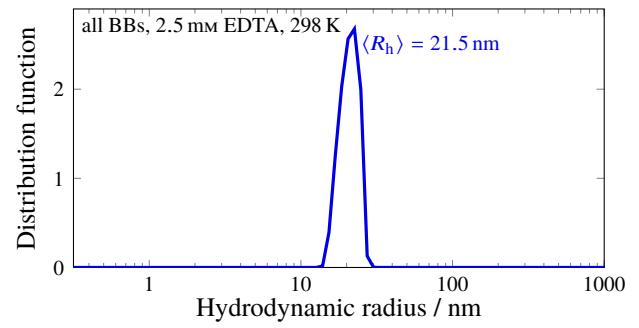


FIG. S4. Size distribution function from DLS experiments for the all-BB system to which 2.5 mM EDTA was added. Addition of EDTA to a solution at the end of the annealing protocol reduces the high molecular weight contamination and sharpens both the DLS distribution function and the gel electrophoresis band associated with the target product. The resulting solution can thus be said to be ‘purified’.

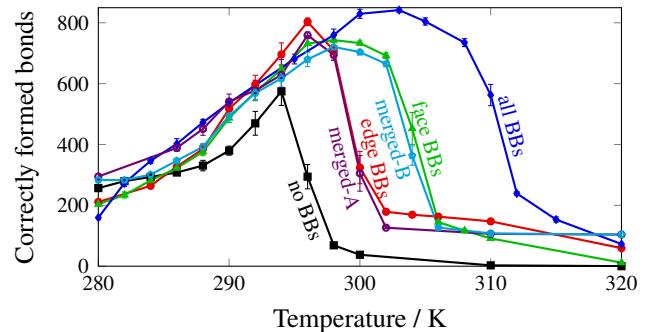


FIG. S5. The number of correctly formed bonds in the system as a function of temperature from Monte Carlo simulations. Each data point corresponds to an average over ten independent simulations in the long time limit once nucleation has occurred. Error bars give the standard deviation in each case.

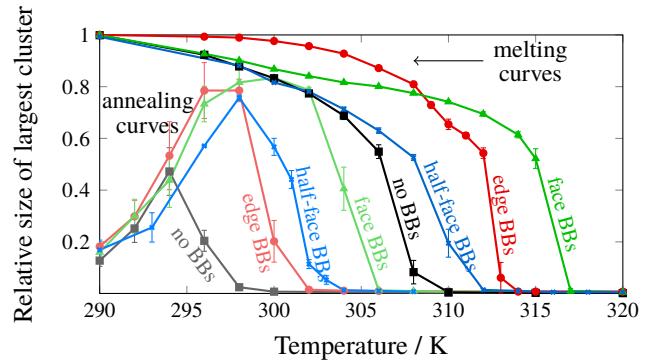


FIG. S6. Melting and annealing curves for the half-face BB structure in Monte Carlo simulations.

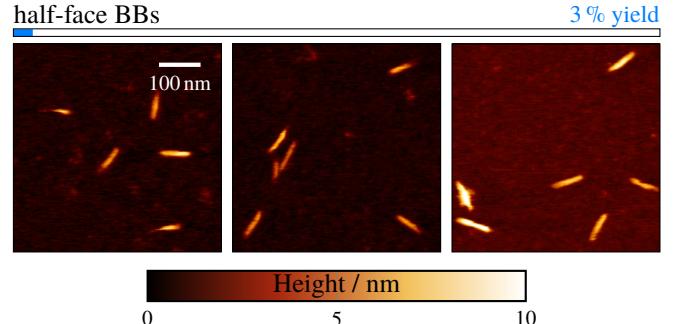


FIG. S7. AFM images for the half-face BB structure.

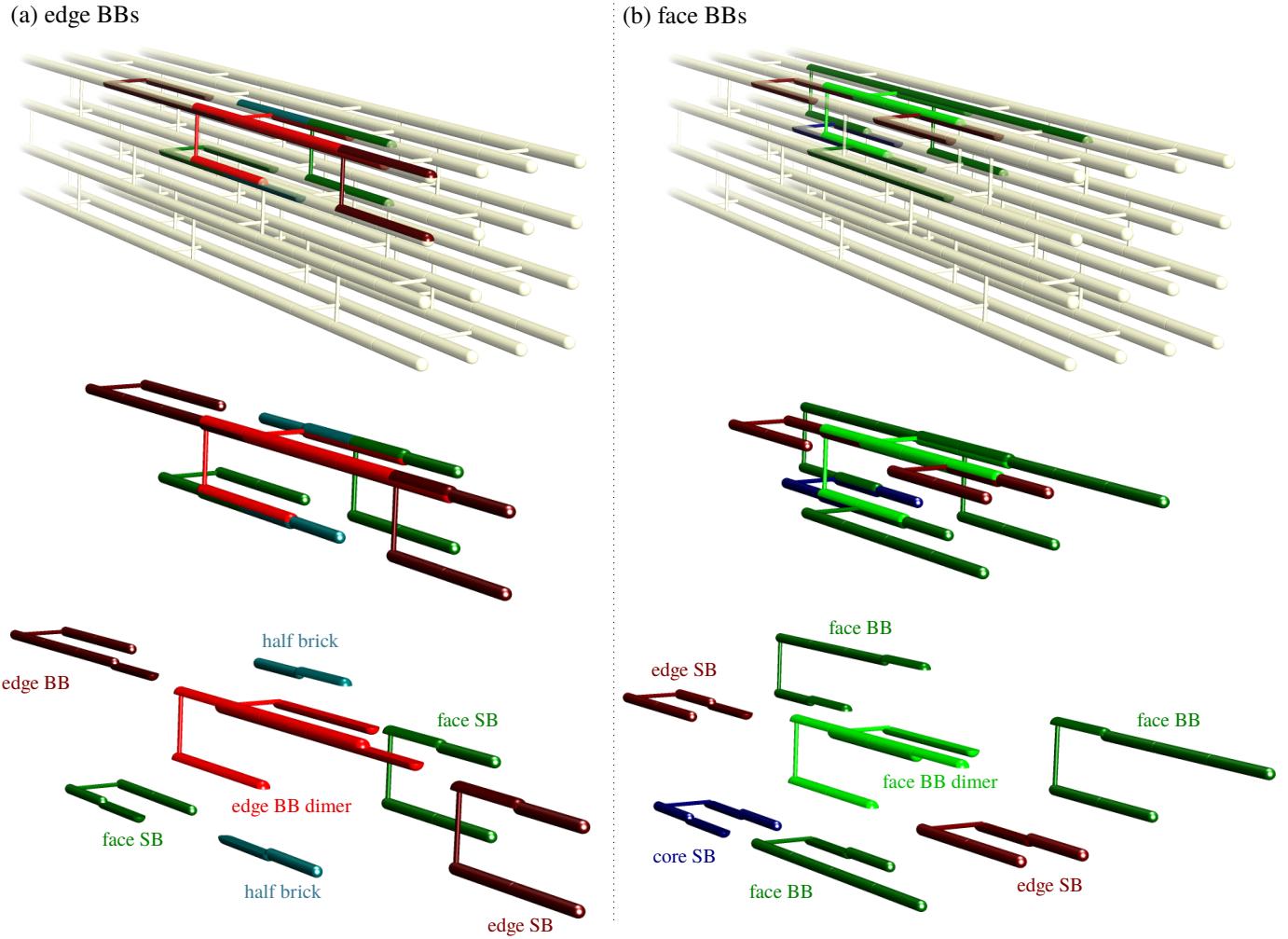


FIG. S8. Boundary brick dimers and their nearest neighbors in a schematic representation. In the top panel, the location within the target structure is shown. In the bottom panel, the neighbors are shown spread out and labeled to make their identification clearer. ‘SB’ stands for a 32-nt ‘scaffold brick’. The edge-BB system’s nucleation properties were also investigated by merging some bricks, as described in the main text. In particular, the ‘merged-A’ building block corresponds to the edge BB dimer shown in red. The ‘merged-B’ building block corresponds to the edge BB dimer and one of the face SBs shown in dark green. Either one of these face SBs could have been chosen, as both of them have direct connections to core strands. In our simulations, the face SB that is merged with the edge BB dimer is the one whose center of mass is nearer the cuboid’s principal axis in the target structure.

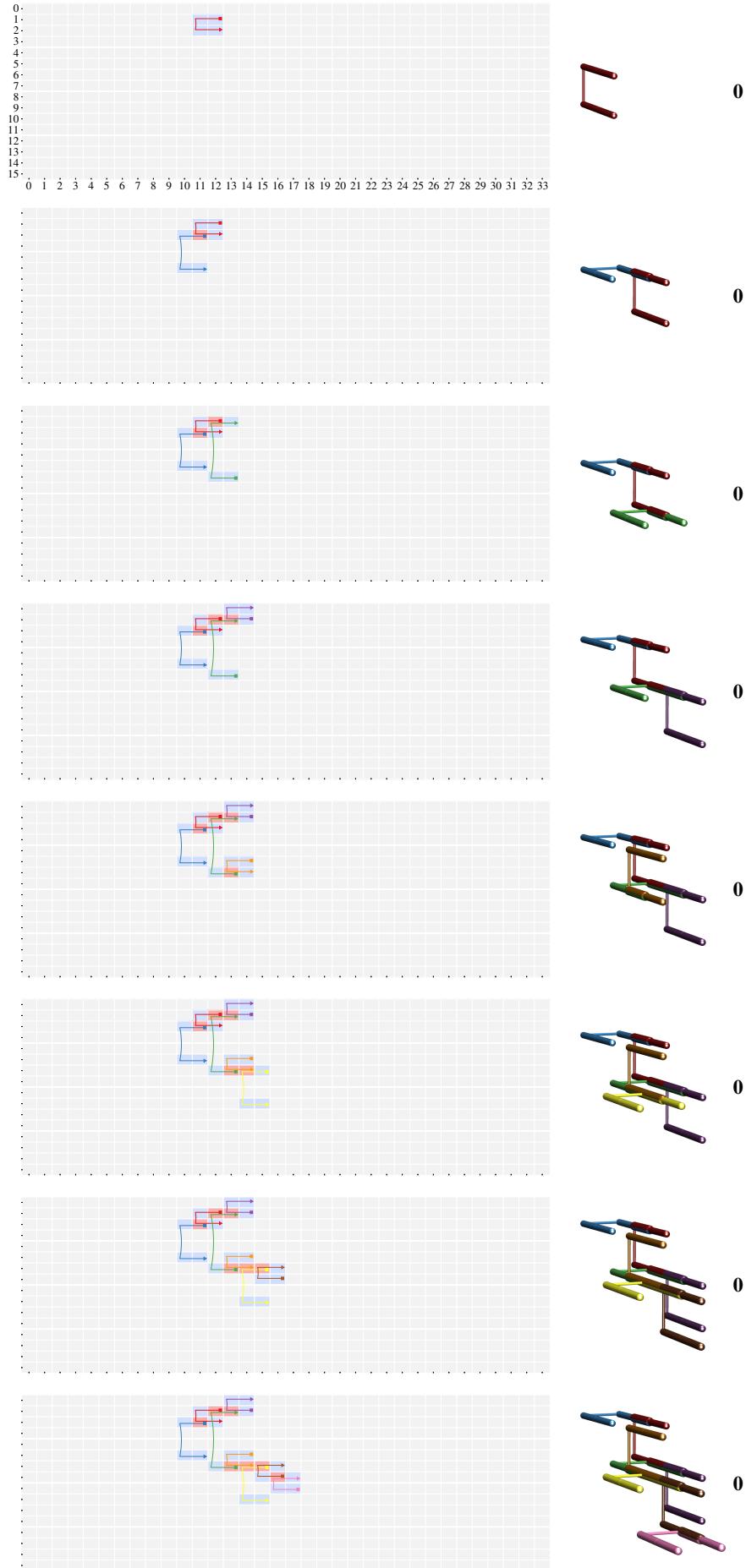


FIG. S9. Example low free-energy nucleation pathway for the no-BB system in two representations: a Cadnano-style connectivity diagram and a three-dimensional rendering. In the latter, DNA brick domains are represented by cylinders. Non-bonded domains are represented by smaller cylinders, while where two DNA bricks are bonded, larger multicolored cylinders are used. Each new monomer or multimer added to the cluster along the nucleation pathway is colored in a different hue. The bold number to the right of each structure indicates the number of multimers in the structure.

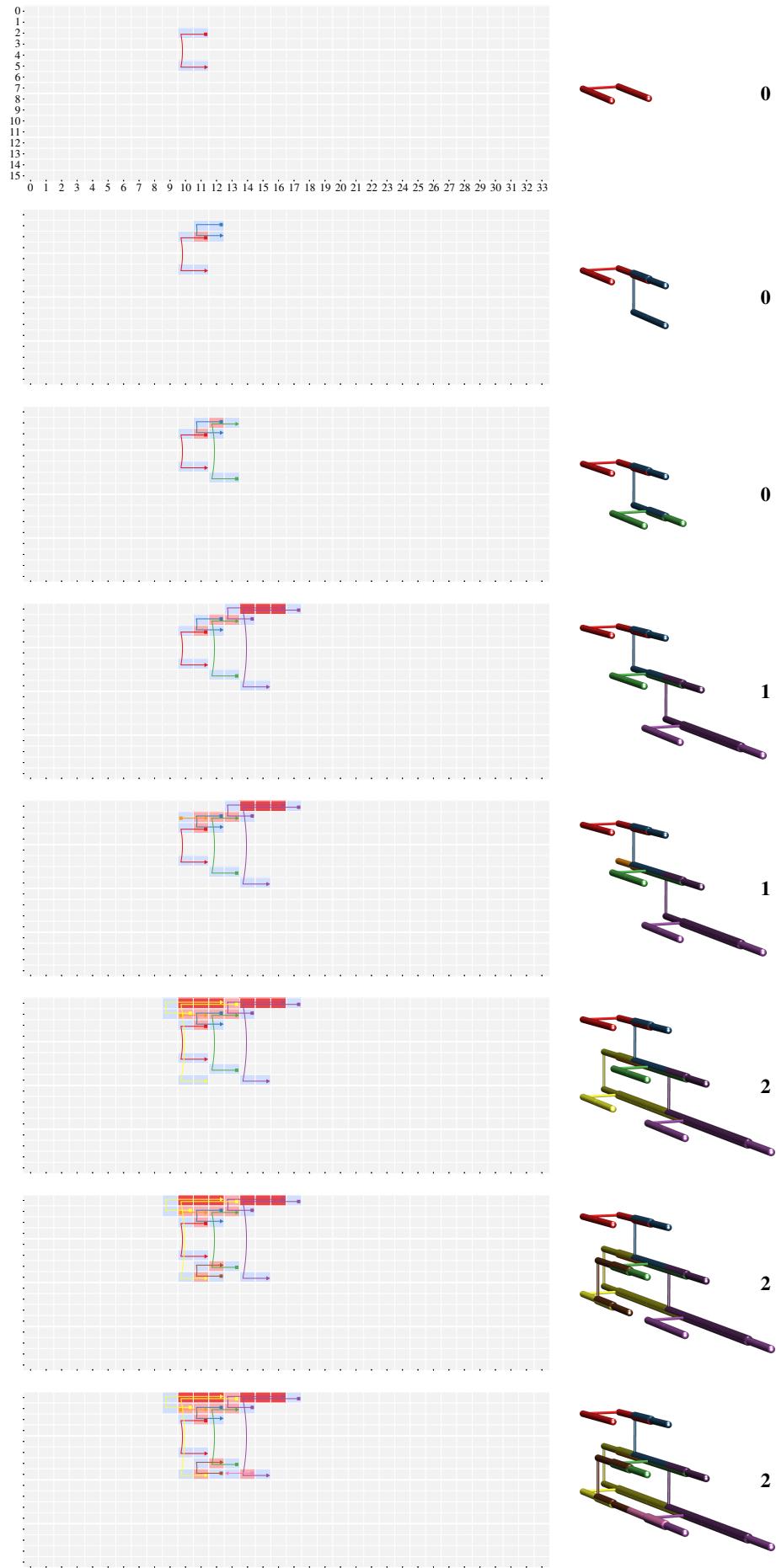


FIG. S10. Example low free-energy nucleation pathway for the edge-BB system in two representations, as in Fig. S9.

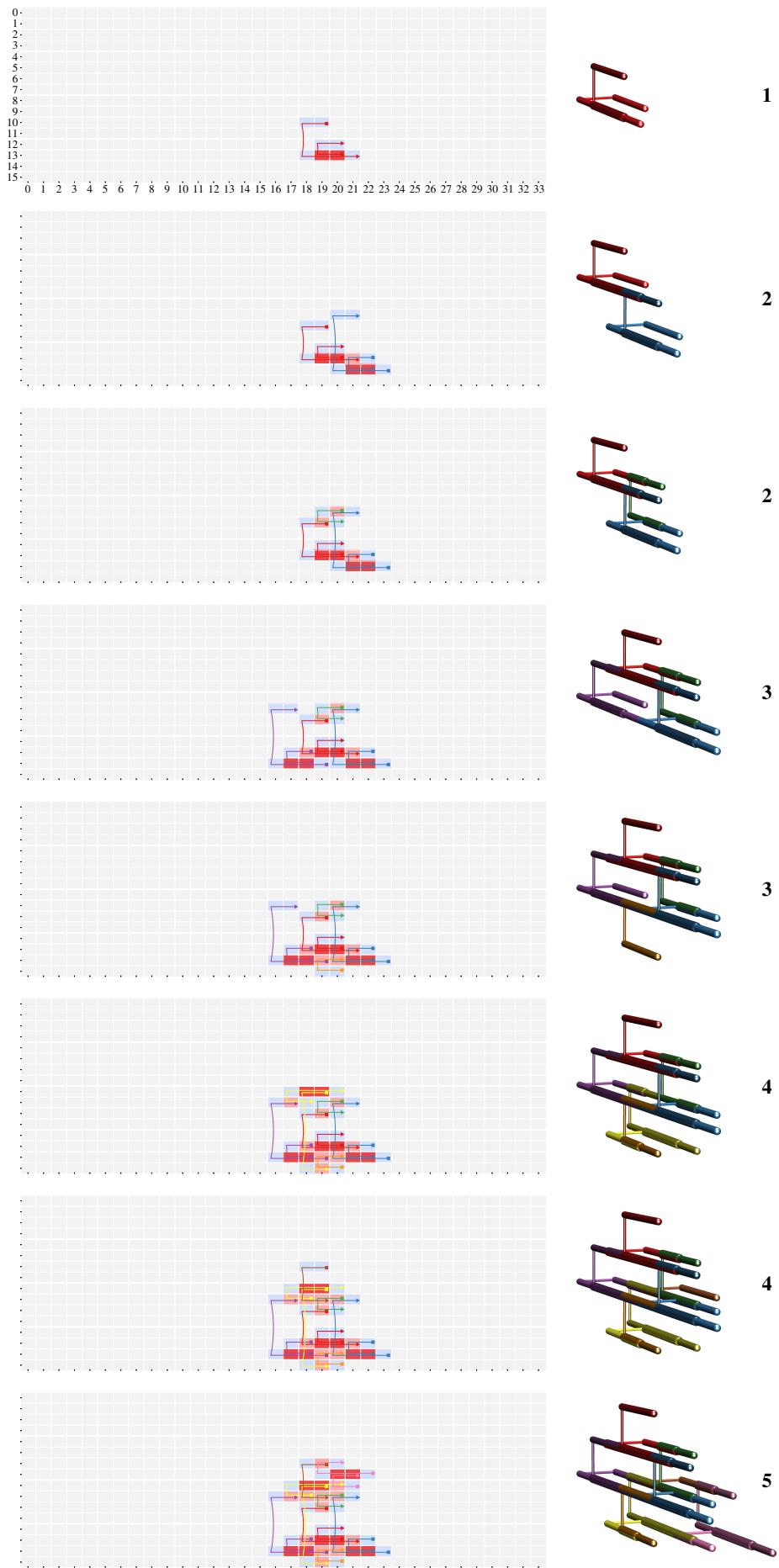


FIG. S11. Example low free-energy nucleation pathway for the face-BB system in two representations, as in Fig. S9.

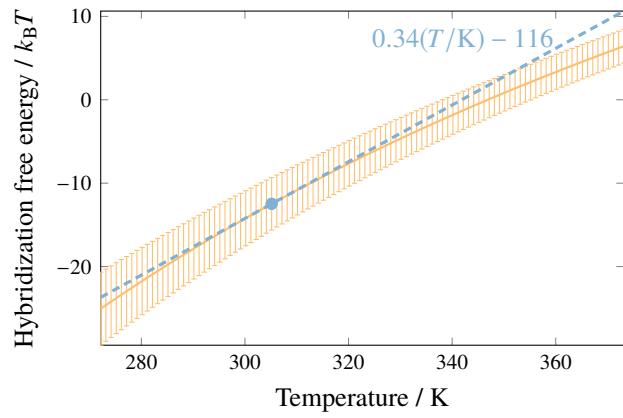


FIG. S12. The mean hybridization free energies of 8-bp interactions as a function of temperature for the no-BB structure, computed via the SantaLucia thermodynamic model [45], with error bars reflecting the standard deviation. The tangent to the curve at 305 K is also shown, demonstrating that the hybridization free energy is well described by a linear function over the region of interest (295 K to 315 K).

4 DNA BRICK SEQUENCES

The following sequences comprise our library of DNA bricks.

ID	Sequence
1	TTCTTAAA TTTTTTTT
2	TTGGCTGA TTTTTTTT
3	TTTTTTTT CGCGCTAA
4	TTTTTTTT CCAACAGG
5	TATGGTGA TTTTTTTT TTTTTTTT TGGCACCC
6	ATCCGAAG TTTTTTTT TTTTTTTT CGCGGACA
7	CGGTGTC TTTTTTTT TTTTTTTT GGGGATGA ATCCTTC AACATCTC
8	TCTGATAT TTTTTTTT TTTTTTTT GCAGTGCC CCTACTCT CACCTTC
9	ATATCAGA GAGCACAG
10	TCCGGAT GGGTGCRA
11	AGAGCTG GAGATGAT
12	GCTTCGCG CCGTGTGG
13	TAATAAA TGTCGGCG GATCAGGC ATTTCGAC
14	AGTTAGCA TTACCGGG GTACCTCTG GACCGCTA
15	GGAAAGGT TCATCCCC TTAAAGAA TGATCGCA CCAGGTTA AGTGGCTC
16	TACATCTC GCTCGTC ATCACCCTA ATCCACCA CCCATGCA GAAAGATA
17	GTGGCTT CTGTCCTC ATCCCTG CTCCGAA
18	TCCCTAA GTCGAATT GAGATGTA AGTTTCAC
19	ACCTTTA ATCATCTC CAGACTCT TTTTTTTT TTTTTTTT GAGCCGAT
20	GGGAGATT TAGCGTCG CAGGTCTA TTTTTTTT TTTTTTTT GGACGAGC
21	CACTTTT GAGATGAT GAAACATC TTTGGTTG
22	AAGGAAT GTGAAACT CAATGATA TGACGGAT
23	ACATCTG GAAAGGTG AGAGTGG GGCAGTGC TCAGCCAA ACAGTGGG
24	CTTACACT TTCCGGAG CAAGGGAT ATCCGTC CTCGGAT TCTCTGGC
25	TATCATGG CCGAGAGA TGTCAACT TGATGACG CAGGAGT CTCCTAGG
26	TGATGGG TGGTGGAT CGCGAACG TGTGTGA ATGATCGG TCAATGAT
27	ACAATGGT ATCTTTG AAGGCCAC GTAAACCG TAATTGG AGCCCGGC
28	ACCTGGT CGTCATAC TAAAGGGT AGAAAGTA GTAGGTTA CGCGTGGG
29	GCTTTAA TATCTTC GGTGGGG GAGTAAGG
30	GTTCAGTG ATCCGTC ACCATTGT TATTGGCC
31	CACCAAGA GAGCCACT TAACCTGG TGCGATCA ATCCCAGA CACGTAAA
32	CCACAAAG TCTTGTAT AAAAGATC TGCGAGAT CTTTACGC GCTTGAAC
33	TTCCCTT CAACAAA AGATGTC CCCACTGT TTATATA CAAAGAT
34	CCGAAATT CGGTTTAC CGAGATGT TTACTTGC ACTTGGGA TTAGGATC
35	CTTGCCTG GGCCAATA CTTTGTG TCCACTAT
36	CACTCCG CCTAGGAG CGGTCCTG CTGTCAGG
37	GCGTAAAG ATCTCGCA TTCTGGT CAAAAGT AATCATAC TCACGGT
38	TTACAGCAG CCTTACTC CGCGAACC TTTACGTG TTTAGGGA ATACAAGA
39	GTGCTCTT GAGCGGTG TAAAGAGC CTCGCTTC AAGTCCGT GTCGGTAG
40	AAAACCTT ACATATGA CGGATCAT TCACGACA AATCTCCC GCTAGATG
41	GGGAGCTG GCCGGGCT ATTACTTT TGCCATAT
42	GTGTTTC ATAGTGA AAAGCAC GTTAACTC
43	GTAGTTA CCCAGCGG TACCTAC TTACTTCT AGTGTAAAG CAACCCAT
44	CGACTCAA CCTCACAG CGAGACGG CATCTAGC ATTACCTT CACCGCTC
45	GCCGACTC GTTCAGC ACTTATTAT CTGGCTAT
46	ACTGAGGC GAGTTAAC TTATGTC TAACCTGC
47	TCATGTGG GATCTTAA TCCCAAGT GCAAGTAA AAAAGGAA TTATGAGA
48	GATCACAA ATATGGCA AAAAGTAT ATGGGTTG CACTAACAC TATAGCAA
49	GAACATATA TTGCTATA TCGGACTG TGGCGCTA CCTACACC AGACATGC
50	ACGGACTT GAAAGGGG GAAGTTTG GCGACTCA TTGCTGAG TTGACGCG
51	ATATAGGT CTGACACC CAGCTCCC AGGTAAAGT GTATTATC GCGTCAT
52	GGTGTAGG TACGCCA TAAACTAC TTTTCCAC GTTGCCTT CGACCGAT
53	GGGGGGT CTACCGAC ATATTGCT TTGCCCCT
54	TGTTGAGG CGAGGTTT ACCTATAT AGCCATTAT
55	AGCTTGGG AACCGTGA GTATGATT AGTTTGG CTGCTGAA TGTCCGTG
56	CGACCACT TGACACCC GAGTCGGC CTGACAGA CATCTTT GGGGTCTG
57	AGCCTACG ATAGCACG ATAATAGT TCTCTATA CAGGAAG GGTGTCAG
58	GATAATAC ACTTACCT CGCACATG AGGTCTAC CAGAAACT ATGTACCG
59	GCGGAGAG AAATGGCT ATGTTGCG ACAATATG
60	CGTCAGC GCGATCTG GATCCAGC TCAGATT
61	AAAAGATG TCTGTAG CCCAACTG TAGCATAA AACTGTTG TGTCTTAT
62	GTCAAAGT AAGGGCAAA AGCAATAT CACGGACA GGAACAC GGGGTCA
63	GCGTATGC TGGAAATG AAACCTTC TACGTTT GCGATGTT GATGGAA
64	TTAGTTT CGCTGCAA CTAGCAA TGAGTCG TTGAGTCG GTGGTAA
65	ATGCGCTA ATTGACGC GACTTTGT GTAATCTC
66	CATACACA CATATTGT CGATACGG CCTGGTTT
67	CTGAGCGT ATCGTCGG AGGGCAAC GTGGAAA TTGTGATC CCCTCCCA
68	AATAGGCT ATTCAAGA GCTGGAGC TGCGCGAG GCACGTAT GTCTTGAG
69	TAGGCTA AAATCTG CGTGGATC TTATCCAC GCCTCACT CCATTCTA
70	ACCATCGC AAAAGCTA AAACATAA TGACGTT GAGGTGAT AGCTAAAG
71	GCGGCTGG ACGACCCC GCGGTGCG ACAGCTGA
72	GTAATGCG AAACCAAG AGCCTATT CATCAAGG
73	GATATCCA CGGTACAT AGTTCTG TGAGACCT CGTAGGCT AGAGCGGT
74	TATTAAGT TTTCGGCG TGAGCGAT CAATGACCT TATATGG ACACACCA
75	GATCTGA GAGATTCAC ACAAAAGTC TGGGAGGG CCTCAACA TCTTGAAT
76	ATACGTGC CTCGGCGA ACGCTCA CGGGTCTG TAGGGTGA CGTATTAG
77	CAAATCCC TTCCCATC AAATACGA CTAAGCCG
78	CGTTGAC CCTTGATG ACTTAATA TTCAGGCT
79	CGTTGTC ATAGGACA CAACAGTT TTATGCTA ACTTGGAC CGGGCATC
80	GCAGGGTG AAAGTGA CGACGGCG GCGCAGGG CTCGACGG TTAATTG
81	ACTCTGT TCAGTCGT CGCACGCC ACCGCTCT CTCTCCG CGGGCAA
82	CCATATGA GGTGATIG TGGATATC TCCGTGCA GTAAACGA TAGGTTTC

Table continues ...

ID	Sequence
83	CCAGTAAAT AGCCTGAA CACCCCTGC ATCGACGG
84	GCTAGCAT CTCAACAGC CTGGGAAT TAGGTAAG
85	CCGTCGAG CCGTGC GACAAACG CATATCCT GTTACTT AGGGATGG
86	TCTGGCGC CGGCCTAG TCGTATTG GATGCCGG TGTGTATG CTCACTTT
87	GAAAGATC ATAACGTT GGGATTG TTCCGCT ATAGGTAAC CTGGGTTT
88	GTCAATTG CTTTAGGT ATCACCTC AAACGTC AAGCCTA TAGCCCTA CGGATCAA
89	GAATTATA TTGGTGTG CCATATCA ACAACCC
90	GGGGCAGA CGCTCGAT GATCTTTC GGACCGAG
91	GGCTTGCA CTAACTG TCACCCCTA CAGACGG TCAGGATC ATTTTGGA
92	TTTGAAA TCTACCA TATCCAG TGATCCG CGCATTAC AACGTTAT
93	CGGCCGC CAAATTAAG TAGCTGAT TAGCTTC
94	AGCTATTG CTGGTCC AATTGAG CATGTCCC
95	GAAAGAAGA GAAACCTA CTGTTTAC TGACCGA ACAGAGT GCTGGCAT
96	TGGCCCG GGGTTTG TGATATGG TCAAATAT GTCAACCG GAAACAC
97	CTGAACT GTGTTTC ATGCTAGC GCCACAG ACTACTTA AGACTTAG
98	TTACCTAT AGGCGGA AAATTGAC ATTCCTAC AATCGTA CGGTTCT
99	CTTCTACC TCTTAATG TATAATC ATCCTGTG GATTACGG CGGCCCTAG
100	TAAGTAGT CGTGTGG TGCAAGCC CTCACCTG ACCTCTTA GCACAATA
101	CTCAACAT AAACCCAG CCGTACGG TGAACTT
102	GTGTAAGA GGGACATG GGTAGAAG GAAACTCTG
103	ATTACCGT CCATCCCT AAGTAACAG AGGATATG GCGCCAGA CCACAATG
104	GCCGATGG AAAACAGC CGGGCCG GGAAGAC CTTTACGC CAAAGGAG
105	TGCAACGG GAAACGTA ATGCACTA ATGCCACG ATTACTGG CATTAAAGA
106	CGGTAATC CAACGGAT TCTTGTCT ATCTTAGT CCCCAAGT GCTACATG
107	CATGGAG CAGAGTTC CCATCGC AAAATTGA
108	AACCGCGC CTAACTG TTAGAATC ACCTTAA
109	GCCTAAAG GTCTCTT CAGGTTAAT CGGGGGT TACGAAC TAAACGCAC
110	CTCAGTTT AGAGTCA CCGTCACG CATTGTGG TTCGCCG GTCGTTT
111	CGACCTCT AGATCCAA TTGTTAGG AGCTAACAA GATATTTC AAGGACAC
112	TCTTCAAA AAACACCG TCACGATT TGAGAAAT TTTACAAA GACGGACA
113	ATGTTCT CTAAGCCG AGTAAACA ATCAAGGA
114	GAAGAGGT TCAATTGTT AGAGGTG AGGCGAGT
115	GTAGGCTC TATTGTC TAAGAGGT CAGGTGAG CGGGCCCA ATTGTACCC
116	ACGGAAAC CCTTACCA CGCCGGT GATGTGCC TGCACAAA GCAAGCCG
117	GTGGCGG CTTAACGGT GATTCTAA TGTCCTC ATAGAGT GTGGATCT
118	GAAATATC TGTAGCT TTTGAAGA AGGCTCG TCATTCCG TCCCCAA
119	ACTCCGGC CTCTCTG TATATTAG CTTACCA
120	GACCCGAC ATCGCTC GTTCCGT GCGTTCTT
121	GGGATTCG CAGTTCG ACTTGGG ACTAAGAT CGTTGCGA ATTTTCG
122	GTTACCGC TATAATTA AGAACATT CGGGTTCG GGGAAATT AGTACAAAG
123	GGCGAGTT TCCCTGAT TGTTACT GTTACAAAT TCTTACAC GGTTAGGA
124	TTTCGCA CGCACATC GAGGCTAC CGGTCTAT AGCAGAAAT GTTAA
125	CACCTACC GTGTCCTC GTCTCTG CGGTCTT
126	ATATTACA AAGAACGC CGGTAAC ATCACATA
127	CGAGTCGT GTGGTGTG AGTTGTA CACCCCG AACTGAG GAGAACAT
128	CCTAGCGC TAAAGGAT CGCCGGAT AGACGTGC CACCGAG CTATGAG
129	CGTCGCTC TTGGTAAG CTAAATATA CGCAAAAT CTTCCATC TAATTATA
130	CAATTCCC CGAACCCG CGAACATC AGGACGAT CTCTTCG GGTGCTAC
131	ACGTGGTC TATGTGAT CGCTCTGG TAAACCTC
132	CGGACACA CGGCTCTG TGCGACCA AGTTTAT
133	CTCGCTG CGACGTCG ACTACGTC TCGAACGA GCTCGGT AGTCTCAA
134	ATTTCGGC GGAACGGG ACAGGACCT ATGTTCTC ACCTCTTC AGGTCTTA
135	GCCAGACT TTACGGC GGTAGGT TAAGACCT CGCCCTGA CGTAGAAC
136	TAGGCCG CTTGGGG CGGAATGT CGGAGGCT TCCGCCAC CCCTGACA
137	GAGCTCAC CTGTTACT CTATCTC TTACCAAGG
138	GGGGATGC GAGGGTATAG AGTCTGGC AACGGGTT
139	TCCAATC TTATACCC ATTCTGCT ATAGACGG AACTGCGC GTCACTGC
140	AGGGATTG AGTACCAAC TTGTCCTC AGATGCGA TATGACAG GTGAGCAC
141	TTCCGCGC ATAAAATCT TGTCGCA TGTCAGGG GTCGGGTC GCGTAA
142	TCAGGGCG AGGTCTTA CGGCCGCTA CGTCTTAA TAACGTGC CTGATCAC
143	GGACTGTT CTACATAG TGACTTGG ACGAGGTT
144	TCAGCGC AACCCGGT CAATCCCT ATGCCGTT
145	GCATCGGC GTAGCACC CGAACGAG ATCGTCTC AGACGACG GATCTCTT
146	ATGACTCC AGGAGAGG GTAGCTC CTACGTT TGACGAG AATACCGT
147	ATAACCAT CTCGGTGA AGGATAAG CGACTGAC TGTAAAT TGGGTACT
148	CTGTCATA TGCCATCT GATTGGG GAGTCAGC GCAATAAA GAAACTGC
149	TTCAATTG CTGTCAG CGATGTT TGCCACA
150	CCTACGCC GAAACGGT CGGAGTCAT CTGGCAT
151	AGTTTCTT TTGAGACT ACCCGAGC TCGTCGA GCGAAAAT CAAGGAGC
152	AGTTAGAT CTCATGCC AACAGTCC GATATGG CCCAGTGA GCTCGAC
153	CGCTTAAA AACCTCTG CCAAGTC AGGAGATC GACCACGT CCTCTCT
154	CTTCGTC CAACGAGT CGCGATGC AGCACCA TATTGAC GTCTTGG
155	AAGTGTAA ATGCGAG ATCTAACT TGATTTT
156	CCCGACCT GTGTCAC CGACGGAG AGGCTTCC
157	TCTCTGG CGCTTAC AGGAACAT CGGAGGGC GAAGATAG AGAGCATA
158	TTAAGATG TGTCGCA CGAACATCG CGTCCTTG CGATCCGC TCGCATGG
159	AAGAATGG TCTACATC GTAATGAA GGATAGAT AAGCAGAG CATTITGA
160	TTCAGCGG GTATCTG CGACGTTA AATAGACG GCGCGAA GACCTCCG
161	AAGAGCAA ACGGTATT ATCCGGT CACAGCGC
162	ACATGAGG AAAAATCA CCATTCTT GTGCGCAT
163	TTACCTCT GCACTTTC TTATTGTC CTGACTC ATGGTTAT CCCTTCCC
164	TTACCAAGC GTGCGGG AGGTGGGG ACAGCAGT ACCCATGG CGCACAC
165	TATTTCTT GGAAGGCT TCTCTG CGGAGGTG GCCGCTGA GATGTAGA
166	CTCTGCTT ATCTATCC CGCTGAA AGGATTAT TCAATAAT GCACCTGC
167	CTCCCTGC TGTCGAG CGACGGAC CACTCTG
168	TTCTAGCC ACTGGCAC GCTGGTAA TCTGGGCT
169	CAGCTGAA CGCAAGAC GTCAAATAA TGCGCT TTTAACGCG TTTAGAGC
170	ATTCAAGA TCGAGACCC TTGCTCTT GGACGTC CCATCAGA TGGAGCTG

Table continues ...

ID	Sequence
171	TGGACCAC GCGCTGTG AACGGGAT GGGAAAGG GGCCTAGG GGGCGCAC
172	CCATGGGT ACTGCTGT GAGGGTAA CACGTTGG ACTCCTAC TGCCCCGC
173	TTAACAA TCAAAATG ATATTGTA GCGAGATA
174	CTACATT AGCCAGA TTCTGAAT CGCGGGCT
175	AAACTACTG TATGCTCT CTATCTTC GCCCCCTCG CATCTTAA ATATAACT
176	CTAGTAAC CGACCGTT GACAGGGA CCACCGCA TCTTCTC GTATATCA
177	CTAGGACC CAGAAAGT GTCGCTGG GGTCTAAA TAACACTT GGTCTCGA
178	TCTGATGG AGACGCC TCAAGCTG CTGCCAG TGATTAGC GCTTAGGG
179	CATTCTAT AGCCCCGG GTTACTAG CGGCCCTTG
180	GATGATCT GTGGTGC CACAATTG GCTGGAAG
181	GAGAAAGA TGCGGTGG CAGTAGTT GTGATTAG AATTGCT AAACCGGA
182	TATCCGCC TATCTCGC TACAATAT AGTTATAT CCTCATGT AACGGCTG
183	ATTAAGAT GTGGAATG TTGTTGAA CATACTCC TCATTAGG GTCCACCG
184	TTCCAAAC GCAGGTGC ATTATGTA ATAATCCT AGGAAATA CCACTGCC
185	AGGGTCA CAGCTCCA AAGCCTGT CGAGGCCT
186	GACTCTCA CAAGGGC AGCTTAAAT TAGTAAGG
187	TGTTTTAAG GCGGGCA GTAGGAGT CCAACGTG GTGGTCCA CGGGACAG
188	CTTTCTG CTTCCAGG AAATTGTC GGCAGTGG GGCTAGAA CATTCCAC
189	CTGGGGG TGATATAC CTCACTAA ACTCCTTC
190	GTCCTATC CCTCTACTA CGCGTACA CACCAAGA
191	AACCTACC CCCTAACG GCTAACATC CTGGGCAG GGTCTTAG GTTGCAGA
192	TTACAGGG ACGGCTCG ACAGGCTT CGTGGCCG AAATGTAG CCATGAAT
193	TGTACCGC ATTATGAG AGATCATC CATTGACG TCGTTAGC TTTTTTTT
194	CCTAATGA GGAGTAGT GTTGGAA TACCTGCT CCTGAATG TTTTTTTT
195	CTCTTACG CTCAATAG TGAACCCCT TACTATCG
196	AGAGTTCT ACCAGCGA AGAGACTC GCAGCCCT
197	CGTAACGA CGTCAATG TAAAAACCA ACAACCGA
198	CATTCAAG AGCAGGTA CAGAAAGG GAACCGAC
199	TTTTTTTT CGATAGTA GTTAGGTT TTTTTTTT
200	TTTTTTTT AGGGCTGC CCCCAAGG TTTTTTTT
201	TTTTTTTT CGGTGGAC AGAACCTCT TTTTTTTT
202	TTTTTTTT TCTTGGT CGTAGGAG TTTTTTTT
203	TTTTTTTT CGGGTTGT CGCTGTAA TTTTTTTT
204	TTTTTTTT GTCGGTGC GATAGGAC TTTTTTTT
205	TTTTTTTT TCCGGTTT AGCGGAATT CTAATCAC GGGGATA TGCGTGGT
206	TTTTTTTT GAAGGAGT TTAGTGAG TCTGCAAC ATAGAATG CTATTGAG
207	CAAGACT TTTTTTTT TTTTTTTT GAGGGAT
208	CAGGTAC TTTTTTTT TTTTTTTT GGACGAGC
209	TCTGATAT TTTTTTTT TTTTTTTT GCACTGCC
210	CGCTGATC TTTTTTTT TTTTTTTT GGGGATGA
211	CAAGGAT ATCGGCTC CTTCGGAT TCTCTGGC
212	TACATCTC GCTCGTCC TACCCATA ATCCACCA
213	AGAGTAGG GGCGATGC TCAGCCAA ACAGTGGG
214	GGAAGGAT TCATCCCC TTAAGAA TGATCGCA
215	TATCATTG GCGCAAGA TGCTAACT GTATGACG
216	TGATGGG TGTTGGAT CGCGGAAC TGCTGTGA
217	AGATGTTT CCACTGT TTATATTA CAAAGAT
218	TAACCTGG TGCGATCA ATCCCGGA CACGTAAG
219	ACCTGGTC CGTCATAC TAAAAGGT AGAAGTAA
220	CCGATCAT TCACGACA AATCTCCC GCTAGATG
221	AACTAGGT ATCTTTG AAGCGCAC GTAAACCG
222	CCGCAACC TTTACGTG TTTAGGGG ATACAAGA
223	TACCTCTC TTACTTCT AGTGTAAAG CAACCCAT
224	CCAGACCG CATCTAGC ATTACCTT CACCGCTC
225	CCGAAATTG CGGGTTAC CGAGATGT TTACTTGC
226	CCACAAAG TCTTGTAT AAAGACTG TGCGAGAT
227	AAAGTAAT ATGGGTTG CACTAAAC TATAGCAA
228	GTGCTCTT GAGCGGTG TTAAGAGC CTCGCTTC
229	TCCCAAGT GCAAGTAA AAAAGGAA TTATGAGA
230	GCCTAAAG ATCTCGA TTCTGGT CCAAAACT
231	GAACATAAT TTGCTATA TCGGACTG TGGGGTGA
232	ACGGACTT GAAGCGAC GAAGTTT GCGACTCA
233	ATAATAGT TCTCATAA CAGGCAAG GGTGTCA
234	GTATGATT AGTTTGG CTGCTGAA TGTCTGTG
235	GGTGTAGG TACGCCA TAAACTAC TTTTCCAC
236	CTCAGCA TGAGTCG TTGAGTCG GTGGATAA
237	ATAATGGT CTGACACC CAGCTCCC AGGTAAGT
238	AGCAATAT CACGGACA GGAAACAC GGGTGTCA
239	AGGCAAC GTGAAAAA TTGTGATC CCCTCCCA
240	CTCGGATC TTATCCAC GCCTCAGT CCATTCA
241	GATAATAC ACTTACCT CCACATGA AGGTCTAC
242	CGACGATT TGACACCC GAGTCGGC CTGACAGA
243	ACAAAGTC TGGGGGG CCTCAACA TCTTGAAT
244	GCCTATGC TGAAATGG AACCTTCC TACGTTTT
245	AGTTCTG GTAGACCT CGTAGGCT AGAGCGGT
246	AAAAGATG TCTGTAG CCCAAGCT TAGCATAA
247	AATAGGCT ATTCAAGA GCTGGAGC TGCGCGAG
248	ACCATGCG AAAAGCTA AAACCTAA TGACGTTT
249	CGCACGGC ACCGGCTCT CTCTCCGC CGGGCATC
250	CAACAGTT TTATGCTA ACTTGGAC CGGGTCTG
251	ATACGTGC CTCGGCA ACGCTAG CGGGTCTG
252	ATCACCTC AAACGCTA TAGCCCTA CGGATCAA
253	TATTAAGT TTTCGGCG TGAGGGAT CAATGACC
254	TCTGTTTT GATGCCCG TGTTGATG CTCACTTT
255	TCACCCCTA CAGACCGG TCAGGATC ATTTTGGG
256	ATTCCCAAG TTGATCCG CGCATTAC AACGTTAT
257	CCATATGA GGTCAATT TGAGATATC TCCGTGCA
258	GCAGGGTG AAAGTGAG CCAGCCGC CGCGAGGG

Table continues ...

ID	Sequence
259	TGATATGG TCCAAAAT GTGCAACG GGAACAC
260	GAAAGATC ATAACGTT GGGATTG TTCCGGCT
261	TCGTTAC TGACCGA ACGAGAGT GCTGGCAT
262	CGCTGAC CCGTGGC GACAACCC CATATCCT
263	CTCGAATT GTGTTCC ATGCTAGC GCCACACG
264	TTACCTAT AGGGCGA AAATTGAC ATTTCCTAC
265	ATCGACTA ATGCCAGC ATTACTGG CATTAAGA
266	AAAGTAAC AGGATATG CGGCCAGA CCACAAATG
267	TAAGTAGT CGTGTGGC TGCAAGCC CTACCTG
268	TCACGATT GTAGAAAT TTACAAGA GACGGACA
269	CTTCTACC TCTTAATG TATAATTG ATCCTGTG
270	CGCTCACG CATTGTTG TTCCGGCG GTCGTTTT
271	TAAGAGT CAGGTGAG CGGGCCCA ATTGTAC
272	GATTCTAA TGTCCTG AATAGAGT TGGAATCT
273	CCGTAATC CACAGGAT TCTTGTCT ATCTTAGT
274	CGCCATGG AAAACGAC GCGGGCCG GGAAGGAC
275	TGTTTACT GTGACAAT TCTTACAC GGGTAGGA
276	CGACCTCT AGATCCCA ATGTTAGG AGCTAAC
277	ACTTGGGG ACTAAGAT CGTTGCCA ATTTTCG
278	GCCTAAAG GTCTCTC ACGGTAAT CGGGGTG
279	ACGGAAC TCTTACCC CGCGGGT GATGTGCC
280	AAATAATC TGTTAGT TTGAAAGA AGGCTCCG
281	CTAATATA GCGAAAAT CTTCATG TAATTATA
282	AGTTGTA CACCCCGG AAACGTAG GAGAACAT
283	TTTTCGCA CGCACATC GAGCTTAC CGCTCTAT
284	CGGAATGA CGGAGGCC TCCGCCAC CCTTGACA
285	GTTACCGC TATAATTG AGAACATT CGGGTTCG
286	ACAGGACG ATGTTCTC ACCTCTC AGGTCTTA
287	ATTCTGCT ATAGACGG AACTGCCG GTCACTGC
288	TGTTGCA TGTCAGGG GTCCGGTC GCGTAAAG
289	CAATTCCC CGAACCGG CGAACATC AGGAGAT
290	CCTAGCGC TAAGACCT GCGGGAGT AGACGTG
291	AGGATAAG CGACTGAC TGTAAATAT TGGGTACT
292	GCCAGACT TTACGGC GTTGGTGTG TAAGACCT
293	CGAACGAG ATCGTCTT AGACGACG GATCTCTT
294	CTCGCTG CGACGTCT ACGACTCG TCGAACGA
295	AGGGATTG AGTACCCCA TGTGTCCT AGATGCCA
296	TCAGGGCG AGGTCTTA GGCGGCTA CGCTTATT
297	CCAAGTCA AGGAGATC GACCACTG CCTCTCTT
298	ACCCGAGC TGTTGCA GCCAAAAT CAAGGACG
299	CTGTCATA TGCCATCT GATTGGAA GAGTCAG
300	GCACGTTA ATAGACG GCGCGGAA GACCTCCG
301	ATGACTCC AGGAGAGG GTGAGCTC CTACGTG
302	AAGCATCG CGTCCTTG GCATCCG TCGCATGG
303	TTTATTGCT CGTGGACTC ATGGTTAT CCCTTCCC
304	TCTCGTC CGGAGGTC CGCCGCTG GATGTAGA
305	CTTCGTC CAACGCTG CGGCATGC AGCACCCA
306	AGTTAGAT CCATGCGA AACAGTCC GTATAGCG
307	AACGGGAT GGGAAAGGG GGCCTAGG GGCCGAC
308	AAGAATGG TCTACATC GTAATGAA GGATAGAT
309	GTCAAAAT TGCGTGT TTAAACGC TTAGAGC
310	TCACTGGG CGCTATAC AGGAAACT CGAGGGGC
311	TTACCAAGC GTGCGGGC AGGTGGGG ACAGCACT
312	CTCTGCTT ATCTATCC CGCGTGAAG AGGATTAT
313	GTGCTGG GTGCTAAA TAACACTT GGCTCGA
314	CTATCTC GCCCCCTG CATCTTAA ATATAACT
315	CCATGGGT ACTGCTGT GAGGGTAA CACGTTGG
316	ATTATGTC ATATCTT AGGAAATA CCACCTGCC
317	ATTCAAGA TGCGAGCC TTGCTCTT GGACGCTT
318	TACAATAT AGTTATAT CCTCATGT AACGGT
319	GTAGGAGT CCAACGTG GTGGTCCA CGGGGAGC
320	AAATATGG CGCAGTGG GGCTAGAA CATTCCAC
321	TCTGATGG AGACGTCC TCAAGCTG CTGCCAG
322	CTAGTAAC CGACCGT CGACAGGG CCACCCGA
323	ACAGGCTT CGTCGGCC AAATGTAG CCATGAAT
324	ATTAAGCT GTGGAATG TTGTTGAA CATACTCC
325	GCTAATCA CTGGCGAG GTTCTAG GTTGCAGA
326	GAGAAAGA TGCGGTGG CAGTAGTT GTGATTAG
327	TGTACCGC ATTATGAG AGATCATC CATTGAGC
328	CCTAATGA GGAGTAGT GTTGGAAA TACCTGCT
329	TTAGTGAG TCTGCAAC ATAGAATG CTATTGAG
330	ACGCAATT CTAATCAC GGCGGATA TGCGTGGT
331	ACCTTTAAT ATCATCTC
332	GGGGAGTT TAGCGGTG
333	CCTACTCT CACCTTTC
334	ATCCTCTC AACATCTC
335	CTTACACT TTGGGGAG
336	CCCATGCA GAAAAGATA
337	ACATCTCG GAAAAGGTG
338	CCAGGTTA AGTGGCTC
339	GACCAAGT CTCCCTAG
340	ATGATCGG TCATATGT
341	TTCCCTTT CAACCAA
342	CACCAAGA GAGGCACT
343	GTAGGTA CCGCTGGG
344	AAAACCTTC ACATATGA
345	TAATTCCG AGGCCGGC
346	TTCAGCAG CCTTACTC

Table continues ...

ID	Sequence
347	GTAGTTA CCCAGCG
348	CGACTAA CCTCACAG
349	ACTTGGG TTAGGATC
350	CTTACCG GCTTGAC
351	GATCACAA ATATGGCA
352	AACTCCGT GTCGGTAG
353	TCATGTGG GATCCTAA
354	AATCATAC TCACGGTT
355	CCTACACC ACAGATGC
356	TTGCTGAG TTGACGC
357	AGCCTACG ATAGCCAG
358	AGCTTGGG AACCGTGA
359	GTGCCCT CGACCGAT
360	TTAGTTT CGGTCAA
361	GTATTATC GCGTCAA
362	GTCCAAGT AAGGGCAA
363	CTGAGCGT ATCGGTCG
364	TAGGGCTA AAATCTGA
365	CAGAAACT ATGTACCG
366	CATCTTT GGGGTGCT
367	GATCCTGA GAGATTAC
368	GGCATGGT GATGGGAA
369	GATATCCA CGGTACAT
370	AACTGTG TGTCTTAT
371	GCACGTAT GTCTTGAG
372	GAGGTGAT ACGTAAG
373	ACTCTCGT TCAGTCGT
374	CGTTTGTG ATAGGACA
375	TAGGGTGA CGTATTAG
376	GTCAAATT CTTAACGT
377	TCATATGG ACACACCA
378	TCTGGCGC CGGCTTAG
379	GGCTTGCA CTAATACG
380	TTGTAAA CTTACCTA
381	GTAACACGA TAGGTTTC
382	CTGACGG TTAATTG
383	TGGGCCCC GGGTTTGT
384	ATAGGTAA CTGGGTTT
385	AACCAAGA GAAACCTA
386	GTTACTT AGGGATGG
387	ACTACTTA AGACTTAG
388	AATCGTGA CGGTTTCT
389	TCGCAACG GAAGGTA
390	ATTACCGT CCATCCCT
391	ACCTCTTA GCACAATA
392	TCTCAAA AGAACCG
393	GATTACGG CGGCTTAG
394	CTCAGTTT AAGATTCA
395	GTAGGCTC TATTGTGC
396	GTGGCGGA CCTAAGGT
397	CCCCAAGT GCTACATG
398	CTTACGG CCAAGGAG
399	GCCGAGTT TCCTTGAT
400	GATATTTC AAGGACAC
401	GGCATTG GATGTAGC
402	TACGAAC CAACGCAC
403	TGCGAAAA GCAAGCCG
404	TCATTCCG TCCCCCAA
405	CGTCGCTT TGGGTAAAG
406	CGAGTCGT GTGCGTTG
407	AGCAGAAT GGTATAAA
408	TAGCCGCC TTGGGGGA
409	GGGAATTG AGTACAAG
410	ATTTTGGC GGAACGGG
411	TCCCAATC TTATATCC
412	TCCCGCGC ATAAAAC
413	CTCGTTCG GGTGCTAC
414	CACCGAG CTATGTAG
415	ATAACCAT CCTGGTGA
416	CGCCCTGA CGTAGAAC
417	GCATGCCG GTAGCACC
418	GCTGGGT AGTCTAA
419	TATGACAG GTGAGCAC
420	TAACGTG CTGATCAC
421	GGCTTAAA AACCTCGT
422	AGTTTCTT TTGAGACT
423	GCAATAAA GAAACTGC
424	TCAGCGGG GTGATCAG
425	TGACGAAG AATACCGT
426	TTAACGATG TGTGGCAA
427	TTACCTCT GCAGTTTC
428	TATTTCCTT GGAAGCT
429	TATTTGAC GTCTTGCG
430	CCCACTGA GCTCGACA
431	TGGACAC GCGCTGTG
432	AACCGAGG CATTTTGA
433	CAGCTTGA CGCAAGAC
434	GAAGATAG AGAGCAT

Table continues ...

ID	Sequence
435	ACCCATGG CGCACCCAC
436	TCAATAAT GCACCTGC
437	CTAGGACC CAGAAGTG
438	AACTACTG TATGCTCT
439	ACTCCTAC TGCCCCGCC
440	TTCCCTAAC GCAGGTCG
441	CCATCAGA TGGAGCTG
442	TATCCGCC TATCTCGC
443	TGTTTTAA GGGGGGCA
444	CTTTCTG CTTCAGC
445	TGATTAGC GCTTAGGG
446	TCTTCTC GTATATCA
447	TTACAGCG ACGCCTCG
448	TCATTAGG GTCCACCG
449	AACCTACC CCCTAAGC
450	AATTGCGT AAACCGGA
451	TCGTTACG TTTTTTTT
452	CCTGAATG TTTTTTTT
453	TTTTTTTT GAAGGAGT
454	TTTTTTTT TCCGGTTT

The following bricks are used in all structures investigated:

1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 17, 18, 21, 22, 29, 30, 35, 36, 41, 42, 45, 46, 53, 54, 59, 60, 65, 66, 71, 72, 77, 78, 83, 84, 89, 90, 93, 94, 101, 102, 107, 108, 113, 114, 119, 120, 125, 126, 131, 132, 137, 138, 143, 144, 149, 150, 155, 156, 161, 162, 167, 168, 173, 174, 179, 180, 185, 186, 189, 190, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204

In addition to the bricks common to all structures, the following bricks are used for each class of structure studied.

The **no-BB** system (330 bricks in total):

207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454

The **edge-BB** system (268 bricks in total):

8, 15, 19, 23, 26, 28, 31, 34, 37, 40, 43, 47, 50, 52, 55, 58, 61, 64, 67, 70, 73, 76, 79, 82, 85, 88, 91, 95, 98, 100, 103, 106, 109, 112, 115, 118, 121, 124, 127, 130, 133, 136, 139, 142, 145, 148, 151, 154, 157, 160, 163, 166, 169, 172, 175, 178, 181, 184, 187, 191, 194, 205, 208, 210, 211, 212, 215, 217, 221, 222, 224, 226, 227, 228, 231, 233, 237, 238, 240, 242, 243, 244, 247, 249, 253, 254, 256, 258, 259, 260, 263, 265, 269, 270, 272, 274, 275, 276, 279, 281, 285, 286, 288, 290, 291, 292, 295, 297, 301, 302, 304, 306, 307, 308, 311, 313, 317, 318, 320, 322, 323, 324, 327, 329, 332, 334, 335, 336, 339, 341, 345, 346, 348, 350, 351, 352, 353, 355, 357, 361, 362, 364, 366, 367, 368, 371, 373, 377, 378, 380, 382, 383, 384, 387, 389, 393, 394, 396, 398, 399, 400, 403, 405, 409, 410, 412, 414, 415, 416, 419, 421, 425, 426, 428, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 446, 447, 448, 451, 452, 453, 454

The **face-BB** system (268 bricks in total):

7, 16, 20, 24, 25, 27, 32, 33, 38, 39, 44, 48, 49, 51, 56, 57, 62, 63, 68, 69, 74, 75, 80, 81, 86, 87, 92, 96, 97, 99, 104, 105, 110, 111, 116, 117, 122, 123, 128, 129, 134, 135, 140, 141, 146, 147, 152, 153, 158, 159, 164, 165, 170, 171, 176, 177, 182, 183, 188, 192, 193, 206, 207, 209, 213, 214, 216, 218, 219, 220, 223, 225, 229, 230, 232, 234, 235, 236, 239, 241, 245, 246, 248, 250, 251, 252, 255, 257, 261, 262, 264, 266, 267, 268, 271, 273, 277, 278, 280, 282, 283, 284, 287, 289, 293, 294, 296, 298, 299, 300, 303, 305, 309, 310, 312, 314, 315, 316, 319, 321, 325, 326, 328, 330, 331, 333, 337, 338, 340, 342, 343, 344, 347, 349, 353, 354, 356, 358, 359, 360, 363, 365, 369, 370, 372, 374, 375, 376, 379, 381, 385, 386, 388, 390, 391, 392, 395, 397, 401, 402, 404, 406, 408, 411, 413, 417, 418, 420, 422, 423, 424, 427, 429, 433, 434, 436, 438, 439, 440, 443, 445, 449, 450, 452, 454

The **half-face-BB** system (299 bricks in total):

7, 16, 20, 32, 38, 39, 44, 45, 56, 62, 63, 69, 80, 86, 87, 92, 104, 110, 111, 117, 128, 134, 135, 141, 152, 158, 159, 165, 176, 182, 183, 188, 192, 193, 207, 209, 211, 213, 214, 215, 216, 217, 218,

219, 220, 221, 223, 225, 227, 229, 230, 231, 232, 233, 234, 235, 236, 237, 239, 241, 243, 245, 246, 247, 248, 249, 250, 251, 252, 253, 255, 257, 259, 261, 262, 263, 264, 265, 266, 267, 268, 269, 271, 273, 275, 277, 278, 279, 280, 281, 282, 283, 284, 285, 287, 289, 291, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 305, 307, 309, 310, 311, 312, 313, 314, 315, 316, 317, 319, 321, 323, 325, 326, 327, 328, 329, 330, 331, 333, 335, 337, 338, 339, 340, 341, 342, 343, 344, 345, 347, 349, 351, 353, 354, 355, 356, 357, 358, 359, 360, 361, 363, 365, 367, 369, 370, 371, 372, 373, 374, 375, 376, 377, 379, 381, 383, 385, 386, 387, 388, 389, 390, 391, 392, 393, 395, 397, 399, 401, 402, 403, 404, 405, 406, 407, 408, 409, 411, 413, 415, 417, 418, 419, 420, 421, 422, 423, 424, 425, 427, 429, 431, 433, 434, 435, 436, 437, 438, 439, 440, 441, 443, 445, 447, 449, 450, 451, 452, 453, 454

The **all-BB** system (206 bricks in total):

7, 8, 15, 16, 19, 20, 23, 24, 25, 26, 27, 28, 31, 32, 33, 34, 37, 38, 39, 40, 43, 44, 47, 48, 49, 50, 51, 52, 55, 56, 57, 58, 61, 62, 63, 64, 67, 68, 69, 70, 73, 74, 75, 76, 79, 80, 81, 82, 85, 86, 87, 88, 91, 92, 95, 96, 97, 98, 99, 100, 103, 104, 105, 106, 109, 110, 111, 112, 115, 116, 117, 118, 121, 122, 123, 124, 127, 128, 129, 130, 133, 134, 135, 136, 139, 140, 141, 142, 145, 146, 147, 148, 151, 152, 153, 154, 157, 158, 159, 160, 163, 164, 165, 166, 169, 170, 171, 172, 175, 176, 177, 178, 181, 182, 183, 184, 187, 188, 191, 192, 193, 194, 205, 206