Soft Matter

Accepted Manuscript

Volume 1 | Number 1 | Jan 2013 | Pages 1–100
Soft Matter
www.rsc.org/softmatter
ROYAL SOCIETY OF CHEMISTRY

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/softmatter

Hierarchical assembly may be a way to make large information-rich structures

Stephen Whitelam*
Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

Self-assembly in the laboratory can now yield ‘information-rich’ nanostructures in which each component is of a distinct type and has a defined spatial position. Ensuring the thermodynamic stability of such structures requires inter-component interaction energies to increase logarithmically with structure size, in order to counter the entropy gained upon mixing component types in solution. However, self-assembly in the presence of strong interactions results in general in kinetic trapping, so suggesting a limit to the size of an (equilibrium) structure that can be self-assembled from distinguishable components. Here we study numerically a two-dimensional hierarchical assembly scheme already considered in experiment. We show that this scheme is immune to the kinetic traps associated with strong ‘native’ interactions (interactions designed to stabilize the intended structure), and so, in principle, offers a way to make large information-rich structures. In this scheme the size of an assembled structure scales exponentially with the stage of assembly, and assembly can continue as long as random motion is able to bring structures into contact. The resulting superstructure could provide a template for building in the third dimension. The chief drawback of this scheme is that it is particularly susceptible to kinetic traps that result from ‘non-native’ interactions (interactions not required to stabilize the intended structure); the scale on which such a scheme can be realized therefore depends upon how effectively this latter kind of interaction can be suppressed.

I. INTRODUCTION

Molecular self-assembly in the laboratory is a promising way of making useful materials [1–4]. Self-assembly mediated by DNA [5], in particular, has been used to create a wide range of nanostructures [6–8], some of which can perform basic functions [9]. Recent work has demonstrated the self-assembly of DNA ‘brick’ nanostructures, which are solid, equilibrium structures in which each component or brick is of a distinct type and has a defined position [10, 11]. These structures self-assemble in solution because inter-component interactions, which are mediated by DNA basepairs, are stronger between components designed to be adjacent in the assembled structure than between components not designed to be adjacent in the assembled structure. A growing body of theoretical work [12–21] indicates that this principle of component-type complementarity might be used quite generally to create defined, multicomponent assemblies of e.g. colloids or other nanoscale building blocks. ‘Information-rich’ materials [22] of this kind have considerable technological potential [23, 24].

However, theoretical work also suggests that some obstacles must be overcome before we can self-assemble information-rich structures of arbitrary size (DNA brick nanostructures are about 1000 components in size [10, 11]). For one, the growth rate of a structure made of Q distinguishable components will likely be of order Q times less than that of the corresponding indistinguishable-particle structure, because only about 1 in Q inter-component collisions can promote growth [19]. For another, the energy scale of inter-component bonds must grow on the scale of $k_{\mathrm{B}}T \ln Q$ in order to render the desired structure thermodynamically stable, because bond energies must counter the entropy associated with permuting component types in solution [19, 20, 23] (see Appendix A). This energy scale is, for macroscopic objects with $Q \sim 10^{24}$, in excess of $50\,k_{\mathrm{B}}T$. Molecular self-assembly in the presence of interaction energies large on the scale of $k_{\mathrm{B}}T$ generally results in kinetic trapping, because strong bonds can prevent the correction of mistakes that happen when components undergoing Brownian motion collide randomly [26–32]. This is so even if the only interactions present are those designed to stabilize the intended structure – we will call these *native* interactions – so suggesting a limit to the scale on which information-rich structures may be self-assembled.

To overcome this limitation one must either seek to make defined *nonequilibrium* structures [33–36], or else engineer an assembly pathway that is immune to the kinetic traps associated with strong native inter-component interactions. This paper focuses on the second of these options.

We will focus on a 2D assembly scheme, a version of which has been used in experiment to create DNA nanotile lattices [25] (that scheme aimed to produce many copies of a small target structure; here we will be concerned with making one copy of a large target structure). In this scheme, each stage of assembly involves the formation of distinct squares, stabilized by four internal bonds. Each bond is in general mediated by multiple components. In the presence of native interactions only, each square can be formed only from four particular pieces. No matter the strength of the bonds between these four pieces, or the order in which bonds are made, no misbinding can happen during construction of the square. In other words, this scheme is immune to the kinetic traps associated with strong native inter-component in-

---

*swhitelam@lbl.gov

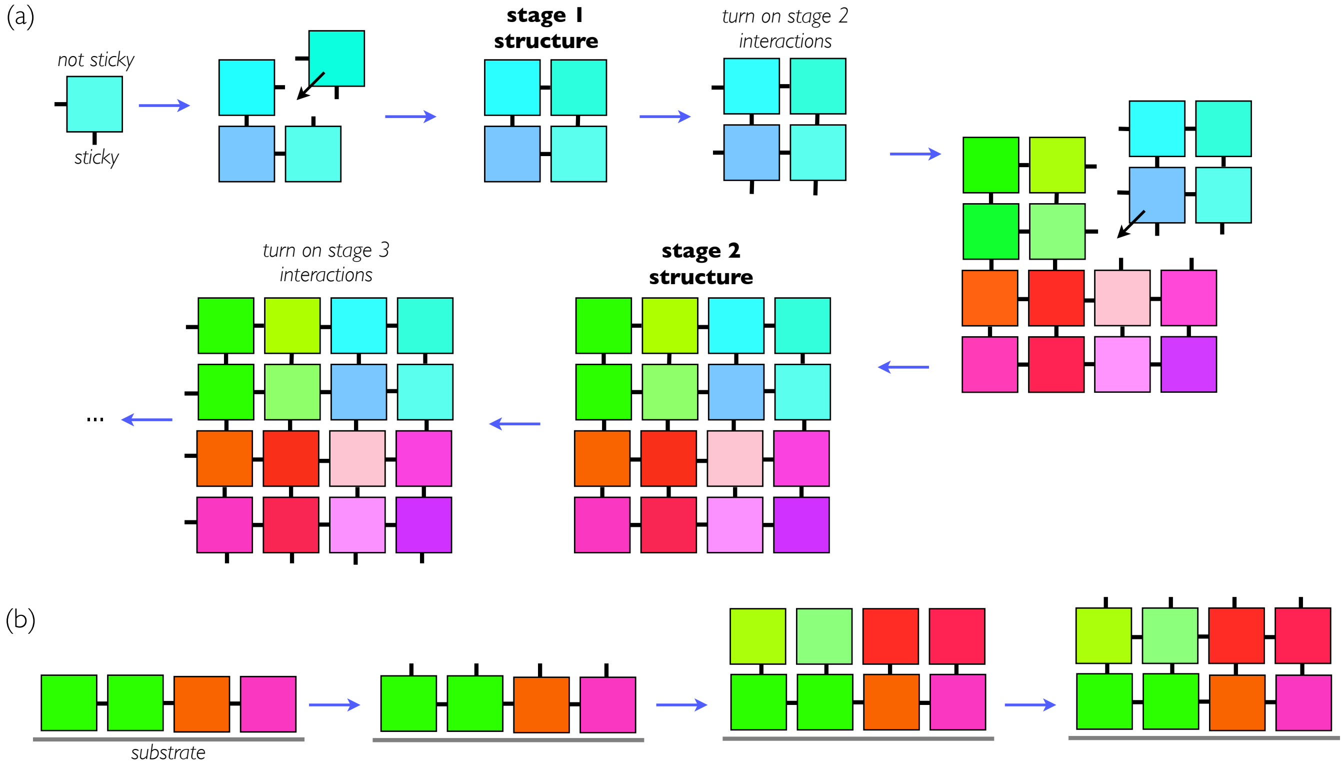


FIG. 1: Self-assembly of large, information-rich structures in thermal equilibrium will likely require strong inter-component interactions [19, 20, 23]. (a) The hierarchical assembly scheme demonstrated experimentally in Ref. [25] is immune to the kinetic traps associated with strong native interactions, i.e. interactions designed to stabilize the assembled structure. Each stage of assembly involves the formation of squares, mediated by four chemically-specific internal bonds. No matter the strength of these bonds or the order in which they form, kinetic trapping in the form of mis-binding cannot occur. (b) The resulting two-dimensional assembly could provide a template for building a three-dimensional structure, which could in principle also be done in a way that avoids kinetic traps associated with strong native interactions.

teractions. To enable such an assembly pathway one can either combine reactants in stages [25], or 'switch on' inter-component interactions in stages, as illustrated in Fig. 1(a). In the first stage of assembly, selected interactions between monomers promote the assembly of squares of size 4. One then turns on additional interactions to promote the formation of second-stage structures, squares of size 16, and so on, with stage n structures being of size 2^{2n}. This process can continue as long as structures can be brought into contact, so permitting the generation of an arbitrarily large information-rich array. This array could provide a template for building in the third dimension, with e.g. subsequent layers deposited one at a time [24]; see panel (b).

The chief drawback of this scheme it that it is, like other forms of hierarchical assembly [37–39], particularly susceptible to kinetic traps caused by 'undesigned' or *non-native* interactions, by which we mean interactions that are not required to stabilize the intended structure. Such interactions include those between components *not* designed to be adjacent in the assembled structure, or interactions between intended neighbors that promote binding incompatible with the target structure. Non-native interactions might arise in experiment because of 'accidental' complementarity between DNA sequences, or because of nonspecific van der Waals attractions. 'Non-hierarchical' many-component self-assembly can proceed in the presence of such interactions [10, 18], provided that the difference in energy scales between the set of native and non-native interactions is large enough [19, 20]. But attractive non-native interactions are tolerated poorly by a stage-by-stage hierarchical scheme, in which the effective building block size is ever-increasing, because undesirable interactions between two clusters increase in strength in proportion to their surface area. Therefore, above some size scale, kinetic trapping of a hierarchical assembly scheme in the presence of attractive non-native interactions would seem to be inevitable.

In view of these considerations, the hierarchical scheme is likely to be inferior to a non-hierarchical scheme if non-native interactions are not intentionally suppressed. Indeed, the authors of Ref. [25] found for DNA tiles that the 'square-upon-square' hierarchical assembly scheme was inferior to a non-hierarchical scheme (in which all component types were combined directly), even when used to make a relatively small 4 × 4 array. Despite this find-

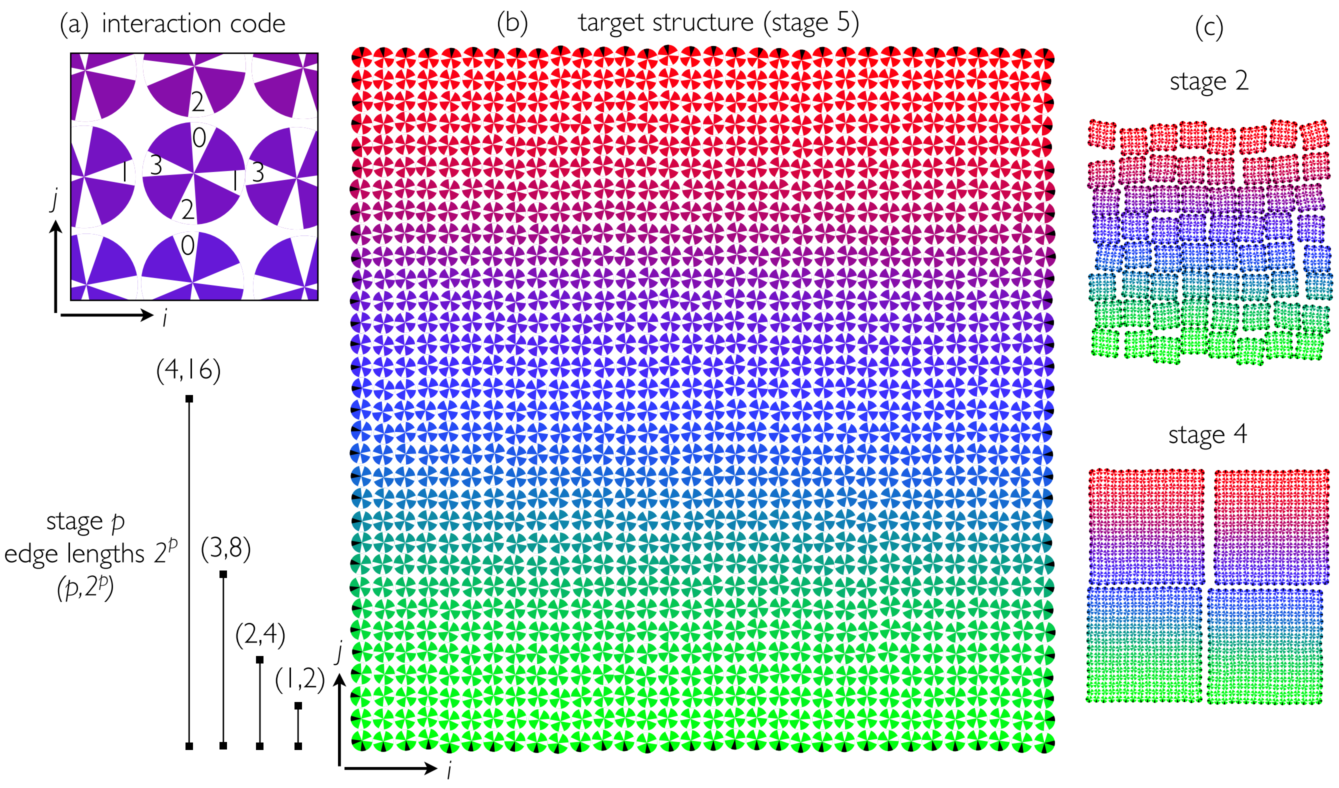


FIG. 2: Simulation model of the idea sketched in Fig. 1. (a) Disc geometry and patch-patch native interaction rules. (b) Target structure for self-assembly, a 'stage 5' square of size $2^{10} = 1024$. All discs in the structure are of a distinct type, as indicated by their color. (c) Maximally-bonded configurations of stage 2 and stage 4.

ing, we argue here that the hierarchical scheme can in principle be used to build larger information-rich structures than can the non-hierarchical scheme (which cannot tolerate strong native interactions), *if* one can suppress non-native interactions. Perhaps this could be done by combining DNA-mediated interactions with some form of inter-particle repulsion (we discuss this point further in the conclusions section). If so, the stage-by-stage hierarchical assembly pathway offers a potential route to the assembly of large information-rich structures in thermal equilibrium.

In Section II we introduce an off-lattice particle-based computer model that can be used to study stage-by-stage hierarchical assembly. In Section III we show that simulations of this model confirm that stage-by-stage assembly results in error-free formation of a desired structure, even in the presence of arbitrarily strong native interactions. We show that the same scheme fares poorly in the presence of attractive non-native interactions. We also discuss the efficiency of a conventional self-assembly process in the presence of irreversible or reversible native interactions. We conclude in Section IV.

II. MODELING STAGE-BY-STAGE ASSEMBLY

The idea of assembling structures in a stage-by-stage manner has been studied theoretically in the DNA tile literature [40–42], where the possibility of extending stage-by-stage assembly to large scales was noted. The particular case of square-upon-square assembly sketched in Fig. 1(a) has been demonstrated in experiments in which DNA tiles formed many distinct 4×4 arrays [25]. Here we study this process numerically, using off-lattice molecular simulation, in order to demonstrate that it can in principle be used to avoid kinetic traps associated with very strong native interactions. For the reasons discussed in Section I, such avoidance is a necessary feature of a process designed to make large information-rich structures in thermal equilibrium. We shall also show that this scheme is particularly vulnerable to kinetic trapping caused by non-native interactions.

Our model of stage-by-stage assembly comprises 1024 hard discs of radius a on a smooth, two-dimensional substrate. Each disc is of a distinct type, one of 1024 possibilities, labeled (i, j), with $i, j = 0, 1, \dots, 31$. Discs are decorated with 4 sticky patches, each of opening angle $\pi/6$, arranged in a regular way (i.e. neighboring patch bisectors are separated by an angle $\pi/2$), so as to allow self-assembly of a square lattice structure. As shown in

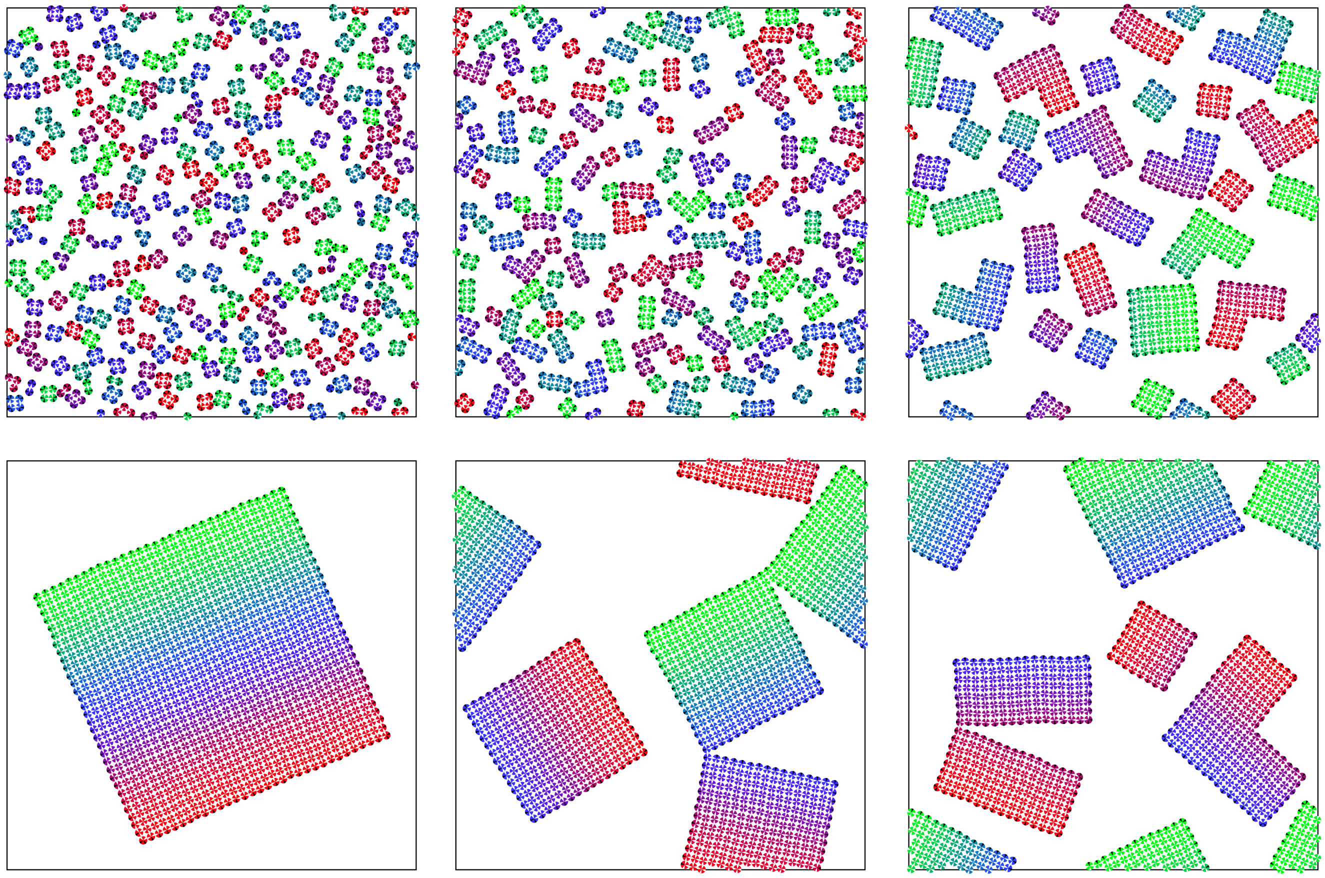


FIG. 3: A time-ordered series of snapshots from a single trajectory of the stage-by-stage hierarchical assembly procedure described in Section II, modeled on the experiments reported in Ref. [25]. The procedure results in error-free assembly of the target structure (see Fig. 2). Times of snapshots, clockwise from top left, are (in millions of MC sweeps) 5, 15, 50, 117, 163, 332.

Fig. 2(a), patches p are numbered 0, 1, 2, 3 in a clockwise fashion. Discs possess attractive pairwise interactions that are square-well in both angle [43] and range. If the centers of two discs (i,j) and (i',j') are separated by a distance d that satisfies $2a < d \leq 2a + a/5$, and if the line joining disc centers cuts through one patch on each disc, then discs receive a pairwise energetic reward of $-E\, k_{\rm B}T$, where

$$E = f(i,j,p;i',j',p')\epsilon_{\rm native} + (1-f(i,j,p;i',j',p'))\,\epsilon_{\rm non-native}. \tag{1}$$

Here $\epsilon_{\rm native} > 0$ and $\epsilon_{\rm non-native} > 0$ are the energy scales of native and non-native interactions. Unless otherwise stated we took $\epsilon_{\rm native} \to \infty$ and $\epsilon_{\rm non-native} = 0$. The function f depends on the disc- and patch identities involved in the pairwise contact: p is the patch number of disc (i,j) involved in the pairwise contact, and p' is the patch number of disc (i',j') involved in the pairwise contact. f is chosen so that the structure shown in Fig. 2(b) is the thermodynamically stable one. This structure, the target for self-assembly, is a square lattice in which disc type (i,j) is found at the intersection of the the ith column and jth row of the lattice, numbered from the bottom left. That is, the bottom left disc is of type (0,0), and the top right disc is of type (31,31). Disc types are colored so that the target structure possesses a green-to-blue-to-red color gradient. To stabilize this structure we require $f(i,j,p;i',j',p')$ to be 1 if patch 1 on disc (i,j) interacts with patch 3 on disc $(i+1,j)$, or if patch 0 on disc (i,j) interacts with patch 2 on disc $(i,j+1)$; these are 'native' interactions. Otherwise f is 0, indicating a

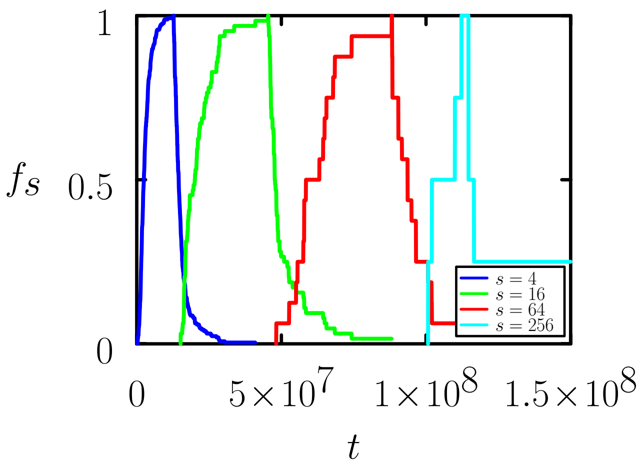


FIG. 4: For the trajectory shown in Fig. 3, we plot as a function of time t the fraction f_s of the system's mass contained within (natively-bonded) clusters of size s. The completion of each stage of assembly corresponds to the point at which each line reaches the value unity, at which point the interactions required to promote the next stage of assembly are automatically switched on. Assembly went to completion at time $t \approx 3 \times 10^8$.

'non-native' interaction.

In equation form these requirements read

$$f(i,j,p;i',j',p') = \delta(i+1,i')\delta(j,j')\delta(p,1)\delta(p',3) \\ +\delta(j+1,j')\delta(i,i')\delta(p,0)\delta(p',2), \quad (2)$$

where $\delta(a,b)$ is 1 if $a = b$, and 0 otherwise. In pictures, in general, patches are shown white when engaged in a native contact, and black when unbound or engaged in a non-native contact.

To allow assembly of the target structure in a hierarchical manner we further distinguish disc native interactions by 'stage'. Interactions up to stage n allow the self-assembly of distinct squares of 2^{2n} discs. Thus, interactions of stage 1 can promote the formation of squares of 4 discs (of 256 possible types); stage 1 and 2 interactions allow the formation of 64 distinct square 16-mers; stage 1, 2, and 3 interactions allow the formation of 16 distinct square 64-mers; stage 1, 2, 3 and 4 interactions allow the formation of 4 distinct square 256-mers; and stage 1,2,3,4 and 5 interactions allow the formation of one square of size 1024. Maximally-bonded configurations that result from interactions of all stages up to 2, 4 and 5 are shown in Fig. 2(b,c).

In equation form we distinguish interactions that participate in different stages of assembly by modifying Eq. (2) to read

$$f(i,j,p;i',j',p') = \delta(i+1,i')\delta(j,j')\delta(p,1)\delta(p',3)\prod_{n=1}^{5}\mathcal{M}_n(i') + \delta(j+1,j')\delta(i,i')\delta(p,0)\delta(p',2)\prod_{n=1}^{5}\mathcal{M}_n(j'), \quad (3)$$

where $\mathcal{M}_n(i) = G_n$ if i is a multiple of 2^{n-1}, and is unity otherwise. The parameter G_n, which we control externally, is either 0, signaling that stage n interactions are 'turned off', or 1, signaling that stage n interactions are 'turned on'. We shall restrict ourselves to the case in which $G_n = 1$ implies $G_m = 1$ for all $m < n$, e.g. if stage 3 interactions are 'turned on' ($G_3 = 1$), then so too are stage 1 and stage 2 interactions ($G_1 = G_2 = 1$).

We did simulations of 1024 discs, one of each type, in the NVT ensemble. Simulation boxes were square, with periodic boundaries, of size such that the disc hard-core packing fraction was 32%. To demonstrate the fact that error-free hierarchical assembly can occur in the presence of strong interactions we took $\epsilon_{\rm native} \to \infty$, so that contacts, once made, could not be broken. Disc structures therefore do not dissociate, although they possess internal flexibility on account of the finite range and angular specificity of disc-disc interactions. Discs were initially randomly dispersed and oriented, subject to there being no hard-core overlaps. We used the virtual-move Monte Carlo algorithm [44] described in the appendix of Ref. [45] to move discs according to an approximation of overdamped motion. Translation and rotation of interacting clusters of discs occurs under this dynamical scheme, and the rate of cluster diffusion can be controlled to a degree by the user. We chose to reject moves of clusters of hydrodynamic radius R so that, approximately, translational and rotational cluster diffusion constants scaled as $D_{\rm trans}(R) \propto R^{-1}$ and $D_{\rm rot}(R) \propto R^{-3}$ (see the SI of [46]). Different choices may be appropriate for different types of surface. We chose the basic scale of disc displacement so that an isolated disc will diffuse a characteristic length equal to its own diameter in 33.3 Monte Carlo sweeps.

To promote stage-by-stage assembly we began simulations with only stage 1 interactions turned on, i.e. we set $G_1 = 1$ and $G_n = 0$ for $n > 1$. When the first stage of assembly was complete, i.e. when the simulation box contained 256 distinct square 4-mers, we turned on stage 2 interactions, i.e. we set $G_2 = 1$ (with the condition $G_1 = 1$ unchanged). Stage 2 interactions allow the 256 square 4-mers to assemble hierarchically into 64 distinct square 16-mers. When stage 2 of self-assembly was complete we turned on stage 3 interactions, and so on until stage 5 of self-assembly was complete. For comparison, we also carried out hierarchical simulations in

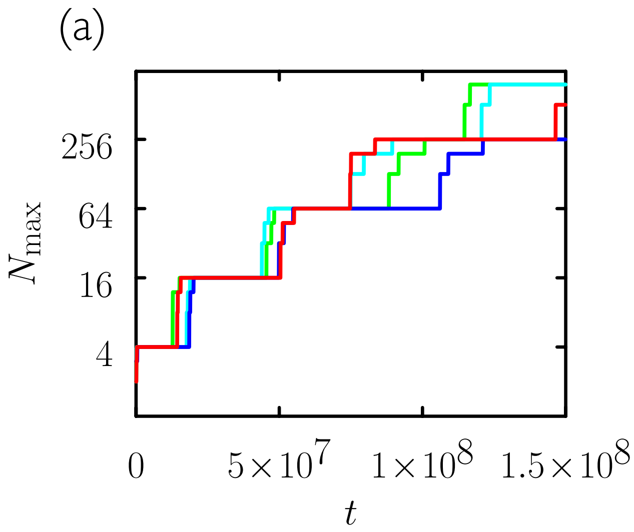


FIG. 5: (a) Size $N_{\rm max}$ of the largest natively-bonded cluster, shown for 4 independent trajectories of the stage-by-stage hierarchical assembly scheme. (b) Mean (symbols) and standard deviation (error bars) of the time $t_{\rm fin}$ required to finish each stage of assembly (the times between the peaks of adjacent curves in a plot such as Fig. 4), averaged over 8 independent trajectories. The blue and red lines correspond to cases in which non-native interaction strengths were set to zero and to $k_{\rm B}T$, respectively.

which non-native patch-patch interactions (those that return zero on the right-hand side of Eq. (3)) were nonzero (Fig. 5(b), Fig. 6, Fig. 7), and we did 'non-hierarchical' simulations in which all native interactions were turned on from the start of the simulation, i.e. we set $G_n = 1$ for $n = 1, 2, 3, 4, 5$ (Fig. 8, Fig. 9, Fig. 10).

III. SIMULATION RESULTS

Stage-by-stage assembly works in the presence of arbitrarily strong native interactions. In Fig. 3 we show snapshots of simulation configurations, ordered clockwise from the top left, from a single dynamic trajectory of the stage-by-stage assembly procedure described in Section II. The scheme works – the result is the low-energy, 1024-particle structure with no errors – even though native interactions are unbreakably strong (non-native interactions are absent). Assembly is a diffusion-limited process, and no mis-bound configurations exist. This scheme is also unaffected by the 'monomer starvation' kinetic trap seen in the study of viral capsid self-assembly [27, 47] or in models of DNA brick assembly if one aims to produce multiple copies of a target structure [39]: all pieces have a prescribed set of binding partners, and so it is not possible to use up the monomer supply by making (say) a large number of trimers that cannot be completed. One needs to wait long enough for components to find their native partners, but the result is guaranteed to be free of error.

In Fig. 4 we show, as a function of time (number of Monte Carlo sweeps), the fraction f_s of the system's mass contained within natively-bonded clusters of size s. The completion of each stage of assembly corresponds to the point at which the individual lines reach the value unity, at which point the interactions required to promote the next stage of assembly are automatically switched on. From trajectory to trajectory we found that that time required to complete each stage varied, the more so as assembly progressed. This variability can be seen in Fig. 5(a), which shows for 4 independent trajectories the size of the largest (natively-bonded) cluster as a function of time. In panel (b) we show the mean and standard deviation of the time required to complete each stage of assembly, averaged over 8 independent trajectories. Stage completion time generally increases as assembly progresses, because larger clusters diffuse more slowly than smaller ones, even though the effective number of component types diminishes (see Appendix B).

Stage-by-stage assembly is vulnerable to kinetic traps associated with non-native attractions. We have shown that the stage-by-stage assembly scheme works in the presence of strong native interactions. However, its major weakness is that it is susceptible to kinetic traps associated with non-native interactions. The red line Fig. 5(b) corresponds to simulations done in the presence of non-native interactions of energy $k_{\rm B}T$, meaning that such an attraction exists between every patch-pair in the system that does not participate in a native interaction. Native contacts are again unbreakably strong. At stage 4 there is a slight slowing of assembly, relative to the case of zero non-native interaction, because of brief mutual adhesion of large clusters.

For stronger non-native interactions we found that assembly is arrested prior to stage 4. In Fig. 6(a) we show, as a function of time, the size $N_{\rm max}$ of the largest

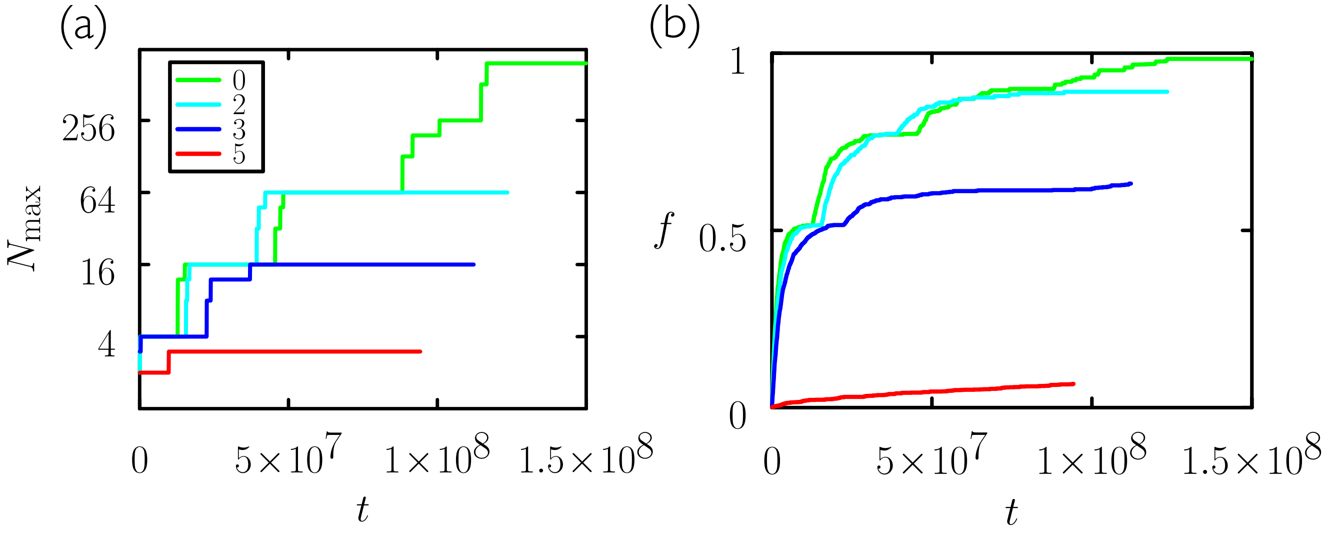


FIG. 6: (a) Size $N_{\max}$ of the largest natively-bonded cluster in the system from 4 trajectories run in the presence of attractive non-native interactions of $0\,k_{\mathrm{B}}T$, $2\,k_{\mathrm{B}}T$, $3\,k_{\mathrm{B}}T$, and $5\,k_{\mathrm{B}}T$. In all cases, native contacts are unbreakably strong. Assembly in the presence of non-native interactions becomes very slow at some point, the earlier the stronger the interaction. (b) Fraction $f \leq 1$ of possible native contacts made at time t, for the same 4 trajectories.

natively-bonded cluster in the system, from independent trajectories run in the presence of attractive non-native interaction energies of 0, $2\,k_{\mathrm{B}}T$, $3\,k_{\mathrm{B}}T$, and $5\,k_{\mathrm{B}}T$. In all cases, native contacts are unbreakably strong. Assembly in the presence of non-native interactions becomes very slow at some point, the earlier the stronger the interaction. In panel (b) we show the fraction $f \leq 1$ of possible native contacts that have been made at time t for all trajectories; these fractions show behavior similar to that of $N_{\max}$.

The cause of this dynamical slowdown can be seen clearly in configuration snapshots. In Fig. 7(a) we show a time-ordered series of snapshots from a simulation run in the presence of attractive non-native contacts of energy $3\,k_{\mathrm{B}}T$. The square-upon-square hierarchical assembly process is subverted by a second organizational process, that of square 4-mers into a structure in which cluster types are arranged essentially at random (compare the color pattern to that of the target structure). As a result, 4-mers combine to form natively-bonded higher-order structures only slowly. In panel (b) we show pictures of structures assembled in the presence of attractive non-native contacts of energy $2\,k_{\mathrm{B}}T$, $3\,k_{\mathrm{B}}T$, and $5\,k_{\mathrm{B}}T$. These pictures show that assembly of compositionally-random structures occurs from smaller building blocks as non-native contact energy decreases. Non-native interactions of energy $5\,k_{\mathrm{B}}T$ cause a compositionally-nonspecific assembly of monomers, so preventing even stage 1 of the hierarchical assembly process from happening.

'Conventional' self-assembly does not work in the presence of arbitrarily strong native interactions. The hierarchical assembly procedure's chief virtue is its ability to tolerate arbitrarily strong native interactions. This ability is not common in molecular self-assembly, and is not shared by a self-assembly process involving the same set of components whose native interactions are all turned on from the start of the simulation. A typical result of this process is shown in Fig. 8. Even though only native interactions operate, if these form in the wrong order – which, inevitably, they do – then the target structure cannot assemble because of steric obstructions.

However, conventional self-assembly done in the presence of reversible native interactions can result in the target structure with few errors. By running a large number of simulations with different values of the native interaction energy, we found that, for native interaction energy of 8.6 $k_{\mathrm{B}}T$, nucleation happened after some delay but was still rapid enough to be seen in direct simulation. Data from two such simulations are shown in Fig. 9 and Fig. 10. Some mistakes result when multiple nucleation events occur and the resulting large clusters fuse in a way that blocks certain binding sites. But assembly is largely successful, reiterating the results of Refs. [18] and [10]: nucleation and growth of an 'information-rich' structure of about 10^3 components can occur in the presence of reversible interactions of fixed strength.

IV. CONCLUSIONS

Ensuring the thermodynamic stability of a structure made from Q distinguishable components requires the energy scale of inter-component interactions to grow in proportion to $k_{\mathrm{B}}T \ln Q$. But self-assembly done in the presence of strong interactions results in general in ki-

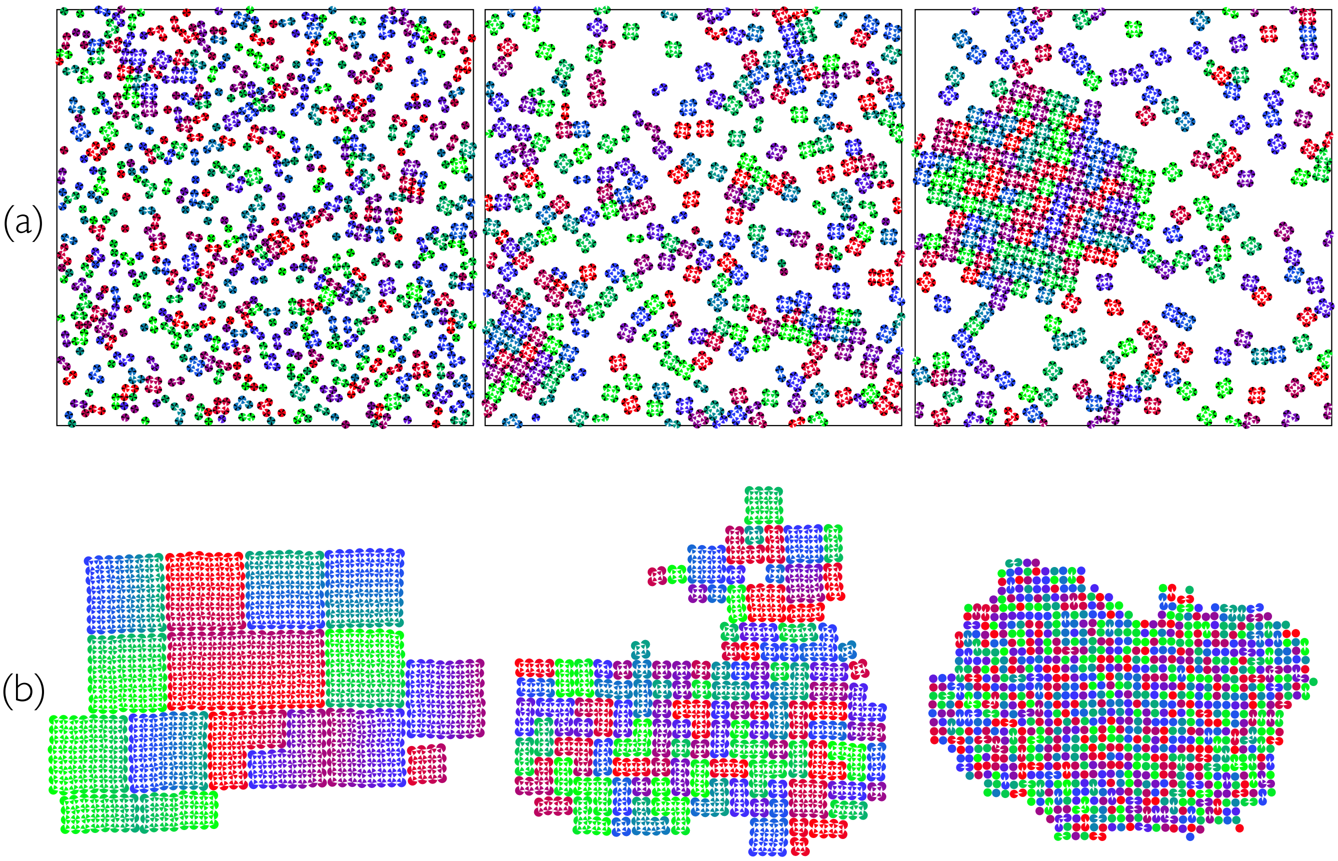


FIG. 7: (a) A time-ordered series of snapshots (time increases left to right) from a simulation run in the presence of attractive non-native contacts of energy $3\,k_{\mathrm{B}}T$. The square-upon-square hierarchical dynamics is overcome by a second organizational process that sees 4-mers assemble into compositionally-disordered structures. As a result, the hierarchical assembly procedure all but stops. (b) Structures assembled in the presence of attractive non-native contacts of energy $2\,k_{\mathrm{B}}T$, $3\,k_{\mathrm{B}}T$, and $5\,k_{\mathrm{B}}T$ (the only patches shown are those engaged in native contacts).

netic trapping, suggesting a limit to the size of a self-assembled structure built from distinguishable components. Here we have used simulation to show that the stage-by-stage assembly of squares demonstrated in experiment in Ref. [25] can work even in the presence of arbitrarily strong interactions, as long as those interactions are the ‘native’ ones required to stabilize the target structure. This property is not shared by most other forms of self-assembly, and is required if one wishes to make large, information-rich structures in equilibrium.

Note that we have assembled a single structure of size 2^{10} to demonstrate that error-free hierarchical assembly can occur even in the face of arbitrarily strong interactions. We stopped at stage 5 of the procedure because simulating larger systems is too time-consuming, but the hierarchical procedure can be used in principle to assemble a single “information-rich” structure of any size. One way to see this is to consider a visual proof by induction. Fig. 3 shows the hierarchical assembly of a 2^{10}-particle structure from monomer building blocks. Imagine now that we have 2^{10} such reactions occurring in parallel. Provided that disc-disc native interactions are chosen in the appropriate manner, combining the results of all reactions would lead to the hierarchical assembly of a 2^{20}-particle structure: imagine a time-series like that shown in Fig. 3 in which the smallest discs shown are not monomers but the 2^{10}-particle assemblies that result from the first set of reactions. And so on. Size limitations then result from practical considerations.

One such consideration is that the stage-by-stage scheme is particularly susceptible to kinetic trapping caused by attractive non-native interactions, and so these must be suppressed if the scheme is to be used to make large objects. With DNA-mediated interactions alone, as noted in Ref. [23], it seems difficult to guarantee strong native interactions and very weak non-native ones, on account of accidental basepair complementarity. Choosing complementary sequences at random, one would expect two ‘non-complementary’ 4-letter sequences of length M to be ‘accidentally’ complementary in $M/4$ places on average. To suppress non-native interactions it may be necessary to equip components with repulsive interactions intermediate in energy scale between fully-complementary interactions and accidentally-

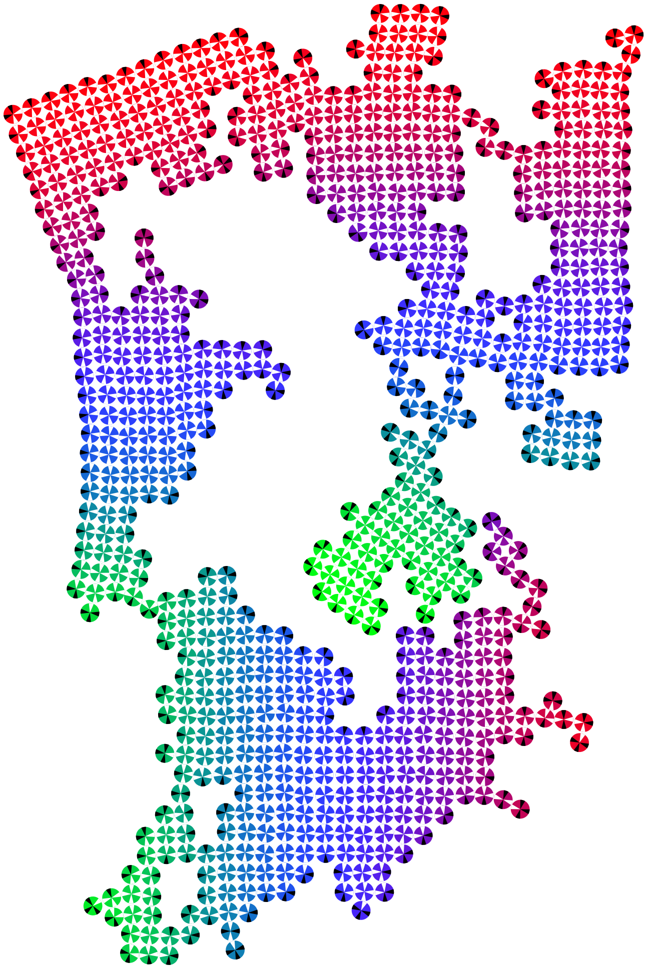


FIG. 8: Non-stage-by-stage self-assembly results in kinetic trapping in the presence of strong native interactions. Here we show a structure that results from self-assembly in the presence of the same set of interactions use to produce Fig. 3; here, however, all stages of interaction were 'switched on' from the start.

complementary ones.

If this can be done, then in principle the scheme presented in Ref. [25] and simulated here can be used to make large objects. However, technical challenges exist, of which four are as follows (more detailed discussions of several of these points can be found in the DNA tile literature [25, 40–42]). First, one must be able to 'switch on' interactions selectively. Perhaps if native interactions correspond to a certain number of fully-complementary basepairs, then selected strands could initially be 'turned off' or 'protected' by partially-complementary single strands. At the required stage of the process the system temperature could be raised slightly, so as to allow thermal dissociation of protector strands only. Protector strands of subsequent stages must be successively more strongly bound, and all must be much less strongly bound than native inter-component bonds. Second, one must be able to monitor the progress of assembly with sufficient precision to know when all the structures in the current stage have assembled (or wait long enough to be sure that this has happened); premature activation of the next stage's bonds could cause kinetic trapping. Perhaps 'smarter' ways to vary inter-component interactions without requiring explicit intervention by the user can be developed, by using novel feedback schemes [48] or other nonequilibrium controls of assembly [49–51]. Third, large clusters diffuse slowly, particular when bound to a surface, and so it is likely that stirring the solution or shaking the surface will be required to allow assembly to proceed beyond a certain lengthscale (shaking may also help break up kinetic traps caused by residual non-native interactions). Fourth, this scheme confronts the challenges inherent to any scheme involving a large number of component types, such as difficulties of synthesis (the number of DNA labels must increase with system size, for instance; see the SI of Ref. [25]) and issues of long timescales. That said, Ref. [25] shows that square-upon-square assembly reactions can be run in parallel: one could in principle combine the outcome of K separate hierarchical processes as the starting point for the next stage of assembly, thereby achieving K-fold speed-up relative to the case of sequential self-assembly. Table I compares conventional and hierarchical assembly schemes.

We note finally that the scheme proposed here cannot be applied (without introducing the idea of twist-sensitive, protein-like interactions [52, 53]) to two-dimensional objects free to move in three-dimensional space: flip over a square and it can bind with its intended partner in a non-ideal way. A similar observation applies to three-dimensional structures. Therefore, we speculate that building 3D information-rich structures may be best done by using a hierarchical scheme to generate large 2D surface-bound assemblies, and then building layer by layer on top of them.

Acknowledgments

This work was done at the Molecular Foundry at Lawrence Berkeley National Laboratory, supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02–05CH11231.

Appendix A: Note on the consequences of particle distinguishability

Imagine that our goal is to promote the self-assembly, from solution, of a solid, equilibrium structure made of Q components. Components generally lose translational and rotational entropy upon going from solution to a solid structure, and so the energy scale $\epsilon/(k_{\rm B}T)$ of inter-component bonds in the solid structure must be large enough that a monomer's energetic gain upon solidifica-

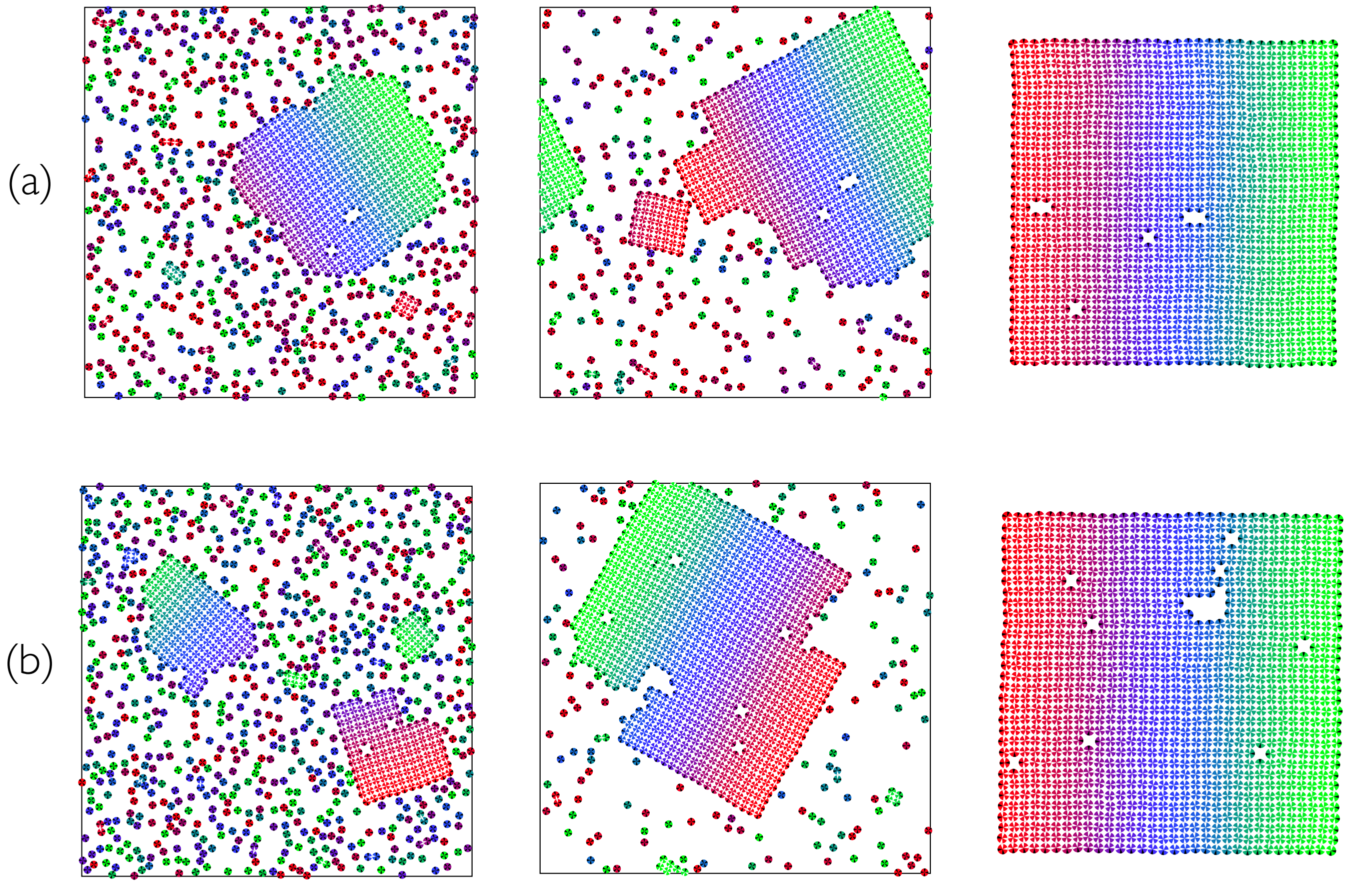


FIG. 9: Non-stage-by-stage self-assembly results in the target structure with only a few errors if reversible interactions are used. Here we show time-ordered snapshots from two simulations (one in (a), one in (b)) run with native interactions of energy $8.6\,k_{\rm B}T$ (and with no non-native interactions). Some mistakes result when multiple nucleation events occur and the resulting large clusters fuse.

TABLE I: Pros and contras of assembly schemes

conventional	stage-by-stage
Intolerant of strong native interactions	Tolerant of strong native interactions
Tolerant of weak non-native interactions	Intolerant of weak non-native interactions
No need for user intervention	Requires user intervention
Reaction not easily parallelizable	Reactions can be run in parallel

tion exceeds temperature times the entropic cost of its removal from solution. If components are distinguishable, and if the solid structure of interest conforms to one particular arrangement of components, then upon solidification one *also* loses an entropy $Q^{-1}k_{\rm B}\ln Q! \approx k_{\rm B}(\ln Q - 1)$ per particle, because the number of accessible configurations of the solution phase is $Q!$ times that of its distinguishable counterpart. The inter-component bond strength $\epsilon/(k_{\rm B}T)$ must therefore grow on the scale of $\ln Q$ in order to ensure the stability of a distinguishable structure built from Q components. The average bond energy per particle must be at least $k_{\rm B}T/(z/2) \times \ln Q$ greater than in the corresponding single-component case, where z is the number of contacts made by each particle. For Q macroscopic, of order 10^{24}, this excess bond energy exceeds $27\,k_{\rm B}T$ when $z = 4$ (appropriate to the case modeled here or the experiments of Ref. [10]), and is about $9\,k_{\rm B}T$ for $z = 12$.

A similar result can be obtained by considering the microscopic dynamics of growth. Say that in the indistinguishable case the rate of arrival of a component at a given point on a solid structure is $R_{\rm arrive}$, which is proportional to component concentration. Say that the rate of departure of that particle from the structure is $R_{\rm depart} \propto \exp(-\beta z \epsilon)$, where $\beta \equiv 1/(k_{\rm B}T)$ and z is the number of bonds made by the particle. In general, the

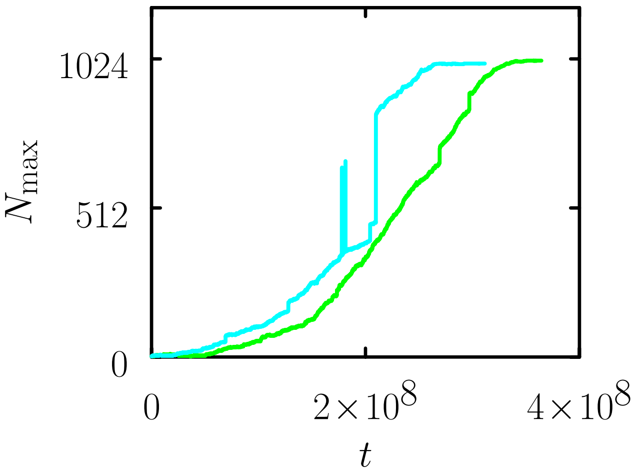


FIG. 10: Size $N_{\max}$ of the largest natively-bound cluster in the system, for the two trajectories shown in Fig. 9. The green (resp. cyan) line here corresponds to the top (resp. bottom) panel in Fig. 9. The final yield of assembly is slightly lower for the case of the cyan/bottom trajectory.

arrival rate must exceed the departure rate in order to ensure stability of the structure with respect to dissolution. Now in the *distinguishable* case the arrival rate, for fixed total component concentration, is $\approx R_{\rm arrive}/Q$, because only about 1 in Q encounters involves the component designed to fit at that point on the structure (now we imagine for simplicity that the number of components in solution is very large, and does not diminish upon binding). So upon going from indistinguishable to distinguishable the rate of arrival of material has dropped by a factor of Q. To guarantee structural stability, therefore, the rate of departure of material must drop by a similar factor, i.e. $R_{\rm depart} \to R_{\rm depart}/Q$. Hence stability requires that $\beta z \epsilon$ must increase in proportion to $\ln Q$.

We tested this expectation by building the 1024-particle target structure and simulating it (in the same square simulations boxes that we used for dynamic simulations) in the presence of native interactions only, and in the presence of native and nonnative interactions of equal strength. In the former case particles make energetic bonds only if they adopt one particular position in the structure, while in the latter case the energy of the system is invariant under particle permutation, or if particles are rotated by multiples of $\pi/2$. We therefore expect the extra pair energy (in units of $k_{\rm B}T$) required to stabilize the structure in the native-interaction case to be $\Delta\epsilon = \frac{1}{2}Q^{-1}\ln\left(4^Q Q!\right) \approx \frac{1}{2}\left(\ln 4 + \ln Q - 1\right) \approx 3.66$ (for $Q = 1024$). By performing simulations for different values of binding energies we determined that in the presence of native interactions the structure melted for bond energy values ϵ between 7.3 and 7.1, while in the presence of native and nonnative interactions of equal strength the condensed structure melted for values of ϵ between 3.8 and 3.6. Although simulations of this nature do not give us a precise measure of structural stability, the difference between these two 'melting points' is 3.5 ± 0.2, consistent with our expectation of 3.66.

The same considerations also suggest that the growth rate of a distinguishable structure will be a factor of Q less than that of its indistinguishable counterpart, i.e. $R_{\rm arrive} - R_{\rm depart} \to Q^{-1}(R_{\rm arrive} - R_{\rm depart})$ upon going from indistinguishable to distinguishable.

Appendix B: Stage completion times

Let us assume that collisions between squares are uncorrelated events. During stage n of assembly these events occur with mean time $\tau(n) \propto 1/D(n-1)$, where $D(n-1) \propto 2^{-n}$ for translation-limited events, and $D(n-1) \propto 2^{-3n}$ for rotation-limited events. To go from stage $n-1$ to stage n of assembly there must occur of order $2^{10-2(n-1)} \propto 2^{-2n}$ events. So the mean time to complete stage n of assembly scales, very roughly, as 2^{-n} if assembly is translation-limited, and 2^{n} if assembly is rotation-limited. The data of Fig. 5(b) lie between these extremes but appear to generally increase with n, suggesting that assembly is dominated, especially at late stages, by the rotational component of diffusion.

---

[1] S. C. Glotzer and M. J. Solomon, Nature Materials **6**, 557 (2007).
[2] J. V. Barth, G. Costantini, and K. Kern, Nature **437**, 671 (2005).
[3] G. M. Whitesides and B. Grzybowski, Science **295**, 2418 (2002).
[4] A. J. Blake, N. R. Champness, P. Hubberstey, W.-S. Li, M. A. Withersby, and M. Schröder, Coordination chemistry Reviews **183**, 117 (1999).
[5] N. C. Seeman, Annual Review of Biophysics and Biomolecular Structure **27**, 225 (1998).
[6] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, Nature **394**, 539 (1998).
[7] B. Wei, M. Dai, and P. Yin, Nature **485**, 623 (2012).
[8] K. Fujibayashi, R. Hariadi, S. H. Park, E. Winfree, and S. Murata, Nano Letters **8**, 1791 (2007).
[9] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. Oliveira, et al., Nature **459**, 73 (2009).
[10] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, Science **338**, 1177 (2012).
[11] Y. Ke, L. L. Ong, W. Sun, J. Song, M. Dong, W. M. Shih, and P. Yin, Nature Chemistry (2014).
[12] K. Fujibayashi, D. Y. Zhang, E. Winfree, and S. Murata,

Natural Computing **8**, 589 (2009).

[13] R. Schulman, S. Lee, N. Papadakis, and E. Winfree, in *DNA Computing* (Springer, 2004), pp. 108–125.

[14] H.-L. Chen and A. Goel, in *DNA Computing* (Springer, 2005), pp. 62–75.

[15] E. Winfree and R. Bekbolatov, in *DNA Computing* (Springer, 2004), pp. 126–144.

[16] P. W. Rothemund, N. Papadakis, and E. Winfree, PLoS biology **2**, e424 (2004).

[17] J. D. Halverson and A. V. Tkachenko, Physical Review E **87**, 062310 (2013).

[18] A. Reinhardt and D. Frenkel, Physical Review Letters **112**, 238103 (2014).

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

[20] A. Murugan, J. Zou, and M. P. Brenner, Nature Communications **6** (2015).

[21] W. M. Jacobs, A. Reinhardt, and D. Frenkel, The Journal of Chemical Physics **142**, 021101 (2015).

[22] R. N. Zuckermann, Peptide Science **96**, 545 (2011).

[23] D. Frenkel, Nature Materials **14**, 9 (2015).

[24] L. Cademartiri and K. J. Bishop, Nature Materials **14**, 2 (2015).

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

[26] R. L. Jack, M. F. Hagan, and D. Chandler, Physical Review E **76**, 021119 (2007).

[27] M. F. Hagan and D. Chandler, Biophysical Journal **91**, 42 (2006).

[28] D. Rapaport, Physical Review Letters **101**, 186101 (2008).

[29] A. W. Wilber, J. P. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, The Journal of Chemical Physics **127**, 085106 (2007).

[30] E. Zaccarelli, Journal of Physics: Condensed Matter **19**, 323101 (2007).

[31] P. J. Lu, E. Zaccarelli, F. Ciulla, A. B. Schofield, F. Sciortino, and D. A. Weitz, Nature **453**, 499 (2008).

[32] S. Whitelam and R. L. Jack, Annual Review of Physical Chemistry **66**, 143 (2015).

[33] E. Sanz, C. Valeriani, D. Frenkel, and M. Dijkstra, Physical Review Letters **99**, 55501 (2007).

[34] B. Peters, The Journal of Chemical Physics **131**, 244103 (2009).

[35] R. T. Scarlett, J. C. Crocker, and T. Sinno, The Journal of Chemical Physics **132**, 234705 (2010).

[36] S. Whitelam, L. O. Hedges, and J. D. Schmit, Physical Review Letters **112**, 155504 (2014).

[37] G. Villar, A. W. Wilber, A. J. Williamson, P. Thiara, J. P. Doye, A. A. Louis, M. N. Jochum, A. C. Lewis, and E. D. Levy, Physical Review Letters **102**, 118106 (2009).

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

[39] J. Madge and M. A. Miller, The Journal of Chemical Physics **143**, 044905 (2015).

[40] E. D. Demaine, S. Eisenstat, M. Ishaque, and A. Winslow, in *DNA Computing and Molecular Programming* (Springer, 2011), pp. 100–114.

[41] M. J. Patitz and S. M. Summers, Natural Computing **9**, 135 (2010).

[42] Z. Abel, N. Benbernou, M. Damian, E. D. Demaine, M. L. Demaine, R. Flatland, S. D. Kominers, and R. Schwelle, in *Proceedings of the twenty-first annual ACM-SIAM symposium on discrete algorithms* (Society for Industrial and Applied Mathematics, 2010), pp. 1045–1064.

[43] N. Kern and D. Frenkel, The Journal of Chemical Physics **118**, 9882 (2003).

[44] S. Whitelam and P. L. Geissler, The Journal of Chemical Physics **127**, 154101 (2007).

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

[46] T. K. Haxton and S. Whitelam, Soft Matter **8**, 3558 (2012).

[47] D. Endres and A. Zlotnick, Biophysical Journal **83**, 1217 (2002).

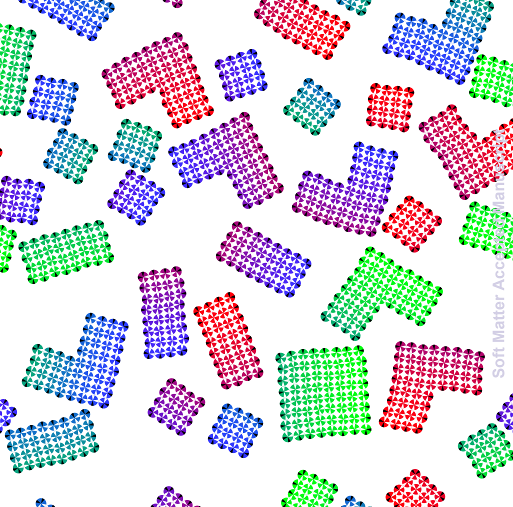
[48] D. Klotsa and R. L. Jack, The Journal of Chemical Physics **138**, 094502 (2013).

[49] M. Byun, W. Han, B. Li, X. Xin, and Z. Lin, Angewandte Chemie International Edition **52**, 1122 (2013).

[50] W. Han, M. He, M. Byun, B. Li, and Z. Lin, Angewandte Chemie **125**, 2624 (2013).

[51] B. Li, W. Han, B. Jiang, and Z. Lin, ACS Nano **8**, 2936 (2014).

[52] R. V. Mannige and C. L. Brooks, Proceedings of the National Academy of Sciences **106**, 8531 (2009).

[53] E. D. Levy, J. B. Pereira-Leal, C. Chothia, and S. A. Teichmann, PLoS Computational Biology **2**, e155 (2006).