Temperature protocols to guide selective self-assembly of competing structures

Arunkumar Bupathy[a], Daan Frenkel[b], and Srikanth Sastry[a,1]

[a]Theoretical Sciences Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560064, India; and [b]Centre for Computational Chemistry, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

Edited by David Weitz, Department of Physics Division of Engineering and Applied Science, Harvard University, Cambridge, MA; received October 21, 2021; accepted January 17, 2022

Multicomponent self-assembly mixtures offer the possibility of encoding multiple target structures with the same set of interacting components. Selective retrieval of one of the stored structures has been attempted by preparing an initial state that favors the assembly of the required target, through seeding, concentration patterning, or specific choices of interaction strengths. This may not be possible in an experiment where on-the-fly reconfiguration of the building blocks to switch functionality may be required. In this paper, we explore principles of inverse design of a multicomponent, self-assembly mixture capable of encoding two competing structures that can be selected through simple temperature protocols. We design the target structures to realize the generic situation in which one of the targets has the lower nucleation barrier, while the other is globally more stable. We observe that, to avoid the formation of spurious or chimeric aggregates, the number of neighboring component pairs that occur in both structures should be minimal. Our design also requires the inclusion of components that are part of only one of the target structures. We observe, however, that to maximize the selectivity of retrieval, the component library itself should be maximally shared by the two targets, within such a constraint. We demonstrate that temperature protocols can be designed that lead to the formation of either one of the target structures with high selectivity. We discuss the important role played by secondary aggregation products in improving selectivity, which we term "vestigial aggregates."

self-assembly | programmable matter | directed assembly

Self-assembly is a fundamental manufacturing mechanism of Nature. Many mesoscale cellular structures required for biological functionality, such as membranes, microtubules, actin fibrils, ribosomes, etc., are formed through self-assembly, often driven by nonequilibrium forces (1, 2). Even though the cytoplasm contains thousands of molecular components, the various cellular structures self-assemble with remarkable precision and may even share components (3). However, in rare cases, they can misassemble, leading to impaired function or even diseases (4, 5). Mechanisms for controlling synthesis in the cell, such as molecular chaperones and compartmentalization of enzymatic action (6–8), are widely studied.

When designing self-assembling systems, one must contend with unwanted "off-target" structures. It is therefore important to understand what causes misassembly and how it can be avoided, both at the design stage and during assembly. On the other hand, an ability to reuse the same building blocks to assemble different structures can be extremely useful to create smart materials that can change their functionality in response to an external stimulus. Often, such materials are designed to change their shape and functionality through conformational or morphological changes of their building blocks (9–15). However, it is also possible to have multiple structures that differ in the spatial arrangement of their building blocks. Multicomponent mixtures not only allow for such a design (16–20), but can also form finite structures with arbitrary complexity (21, 22). Experiments with DNA bricks have shown a way of self-assembling complex structures with hundreds of distinct components (23, 24). Such an addressable assembly, where each component and its position in the target structure are uniquely defined, is made possible because of the complementarity of the DNA hybridization process.

Here, we investigate the generic problem of designing two competing target structures of distinct shape, a feature that we show involves new nontrivial challenges. By designing the location of components and the strength of their interactions, we show that the nucleation behavior of the target structures can be tuned such that either of them can be assembled using distinct time-varying temperature protocols. Further, such a design shows that to avoid chimeric aggregation, the neighborhood of the individual components in the two structures should be maximally different. We show that the design of the competing structure requires the inclusion of components that form part of only one of the two structures. However, in order to maximize the selectivity of self-assembly, the component library should be maximally shared by the targets, within the constraints imposed by the design.

We begin by noting that even with a single target structure, multicomponent systems assemble quite differently from classical nucleation, due to the fact that the components need to bind in certain specific ways to form the correct structure (25, 26). Numerical evidence and theoretical analyses have shown that multicomponent self-assembly proceeds via a nonclassical nucleation process (27–30). For successful assembly, the

Significance

Designing functional materials with switchable properties is a highly sought-after goal in materials science. We investigate the computational design of a multicomponent self-assembly system that can be guided into forming one of two predefined competing structures through simple temperature protocols. Our investigation reveals design constraints that should be observed in order to avoid formation of spurious or chimeric aggregates and to maximize selectivity of the target structure. We demonstrate that temperature protocols can be designed that lead to the formation of either one of the target structures with high selectivity. We discuss the important role played by secondary aggregation products, which we term "vestigial aggregates."

Author contributions: A.B., D.F., and S.S. designed research; A.B. performed research; A.B., D.F., and S.S. analyzed data; and A.B., D.F., and S.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: sastry@jncasr.ac.in.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2119315119/-/DCSupplemental.

Published February 14, 2022.

theories also predict a protocol that allows for slow nucleation followed by completion of growth at a lower temperature (30, 31). Due to incidental interactions between components, there may also be numerous undesired ways in which they can aggregate, which increases the probability of formation of undesired structures. Thus, the designed interactions should be made sufficiently strong, so as to offset this entropic effect (32, 33). Some studies suggested that a narrow distribution of designed interactions is required for error-free self-assembly (32, 33). However, other studies showed that variable bond strengths may improve the kinetics and diminish the competition between fragments that are incorporated at an early stage (30, 34). To avoid the formation of off-target structures, it is important that the self-assembling structure can anneal during growth. This implies that the assembly should take place under thermodynamic conditions, where the growth is almost reversible (30, 35). As a consequence, the range of thermodynamic parameters within which self-assembly can be made error-free is significantly reduced (24, 27, 36, 37).

Focusing now on strategies to design multiple target structures from the same building blocks, the two most pertinent questions are: 1) how to design the targets while avoiding misaggregation; and 2) how to guide their self-assembly into specific target structures. The simplest examples of distinct targets that form from the same building blocks are objects that have the same shape, but differ in the spatial arrangement of building blocks (16–18). Components that are neighbors in any of the multiple targets are assigned attractive interactions. If the interactions are of equal strength, then the retrieval of any specific target from the mixture requires a target-specific seeding procedure or concentration pattern. Selective retrieval by strengthening a few bonds specific to the desired target has also been attempted (16, 18), in which case the system always favors the formation of one structure over others.

Self-Assembly Targets

To explore the strategy for designing different structures from the same building blocks, we consider two structures (square [S] and plus [P]), defined on a two-dimensional square lattice (Fig. 1*A*). Each target is composed of $N = 100$ square blocks, with four distinct interacting edges. Each component is represented with a distinct color, but the four possible orientations are not shown for the sake of clarity. To begin with, we consider S and P to be made of the same set of 100 distinct components. The labeling of the blocks is arbitrary and is done in sequential order in the S structure. These components are placed randomly in the P structure with random orientations.

Given the two targets, we need to specify interactions between the edges of the components, such that they are stable. Representing by $k, l \in \{1, 2, 3 \ldots, 4M\}$ a pair of component edges, where M is the size of the component library, the interaction matrix I that encodes both targets has the form

$$I_{kl} \begin{cases} < 0, & \text{if the edges } k \text{ and } l \text{ are in} \\ & \text{contact in S or P} \\ = 0, & \text{otherwise.} \end{cases} \quad [1]$$

Initially, we assign equal strengths $-\epsilon$ to all the bonds and compute their free-energy curves as a function of the aggregate size, using the method of Jacobs et al. (28, 30), which is outlined in *SI Appendix*, section 1. In Fig. 1*B*, we show the free-energy curves for two temperatures, with all components having the same concentration $\rho = 7.5 \times 10^{-5}$ in the mixture, which corresponds to having three copies of each component in a 200×200 lattice. While this is a relatively small system, it is sufficient to demonstrate the important features of our design.

The differences in the connectivity of the two structures naturally give rise to differences in their free-energy landscapes. From Fig. 1*B*, it is clear that we can devise a temperature protocol to selectively retrieve target S, since it can nucleate at the higher temperature shown where P is unstable, and the growth can be completed at the lower temperature (28, 30). However, we cannot define a protocol for the formation of P without the simultaneous formation of S. There could also be other spurious or chimeric aggregates, as explained below, and, in general, it is not possible to predict such aggregates a priori. However, we show below that we can avoid spurious aggregation by an appropriate choice of interactions and composition of the target structures. In the following sections, we discuss how to choose the target compositions and our scheme for tuning the interactions that allows selective retrieval of either structure through different temperature protocols.

Choosing the Target Compositions

Although it may be possible to design the structures with fewer components than N, we do not do so, as larger libraries are better for kinetic accessibility of the structures (33, 38). Because the components are shared, the neighboring components with attractive interactions of a given component can be different in the two targets. Such promiscuous interactions could lead to assembly errors because an exposed edge of a growing aggregate has multiple candidates for attachment. A possible strategy to minimize the number of aggregation paths due to promiscuous interactions is to maximize the number of components that share

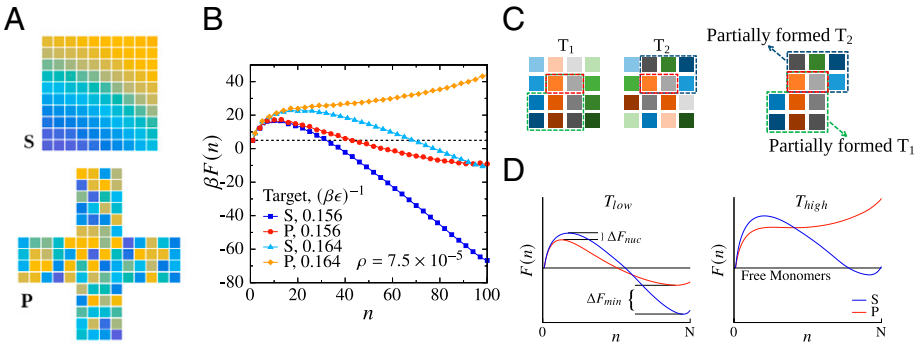

Fig. 1. *A* shows the two structures that we use as design targets in this study. Each target is made of 100 distinct components—square blocks with four interacting faces and an orientation (not depicted for the sake of clarity). We shall change the composition of the targets later in our design. *B* shows the free-energy landscapes of the two targets at $k_BT/\epsilon = 0.156$ and 0.164 when all designed bonds are of equal strength ϵ. (*C*) Illustration showing two example structures, T_1 and T_2, that differ in their internal arrangement of components, except for one common bond, as highlighted by the broken red rectangle. In *C*, *Right*, a possible chimeric aggregate that is part T_1 and part T_2 is shown. (*D*) Schematic showing a scenario where one target has the lower barrier, while the other has the lower free-energy minimum. As the temperature is increased, the structure with the lower barrier will become unstable first.

the share the same neighborhood in the two structures. However, such a strategy, while reducing the number of aggregation paths, causes chimeric structures to be more stable. We illustrate this in Fig. 1*C*, which shows two structures, T_1 and T_2, each made of the same set of 16 distinct components, which differ in their internal arrangement of components. There is one pair of nearest-neighbors that are shared by both the structures indicated by dashed red rectangles. In Fig. 1 *C*, *Right*, a possible chimeric aggregate is shown. The (stable) chimera is made of parts of T_1 and T_2 that are held together by two bonds. This shared motif acts as a stronger glue between the incompatible pieces than individual shared sites may be. This point is further detailed in *SI Appendix*, sections 2 and 3. In *SI Appendix*, section 2, we demonstrate the effect of maximizing the number of shared bonds on the self-assembly of our example structures S and P. We discuss additional considerations relevant to avoiding spurious aggregation in *SI Appendix*, section 3.

We therefore minimize the number of shared bonds between the two structures. Additionally, the boundaries of the two targets need to be inert to prevent any aggregation beyond completion of the desired structures. This is partly achieved by using the same set of components for the boundaries of each structure. However, since their boundaries can be of unequal lengths, there will be some blocks that have nonzero interactions. To avoid this, we choose a slightly larger component library to begin with, so that the extra components serve as additional boundary blocks. Before we proceed with the implementation, we discuss how to tune the interactions for protocol guided retrieval.

Tuning the Interactions for Targeted Retrieval

Considering that a temperature range is available in which the target structures can form through a nucleation process, we generically have the possibilities that the free-energy barrier and the free-energy minimum for one structure is higher than for the other or that one of the structures has a lower free-energy barrier while having a higher free-energy minimum when fully formed. In the former case, no obvious protocol exists to selectively form the structure with a higher free-energy barrier and minimum, whereas for the latter, such a protocol can be devised, as we discuss below. We thus consider the latter scenario. Further, as we describe, such a scenario can be realized through different approaches.

The free-energy curves for the situation we consider are shown schematically in Fig. 1*D*. At low temperatures, both the structures are stable. The structure P has the lower nucleation barrier, but the structure S is globally more stable (or vice versa). As the temperature is increased, the structure P becomes unstable, while the structure S remains stable or metastable. At T_{low}, where P has a sizeable, but surmountable, nucleation barrier, we would nucleate P with a higher probability than S. If the nucleation barrier of S is sufficiently high, then we would rarely nucleate S so that we can grow P with a high degree of selectivity. On the other hand, at sufficiently lower temperatures, both S and P will be able to nucleate (possibly along with other, disordered aggregates). Once they nucleate, we can then make P (as well as disordered aggregates) dissociate back into the mixture by increasing the temperature until P becomes unstable, thus selectively retrieving S.

We can achieve such a scenario by tuning the strengths of the individual bonds between the components (or the chemical potentials of the components, which we do not pursue here). Nevertheless, both approaches are illustrated in *SI Appendix*, section 4. This is possible as long as there are enough bonds in either structure. Experimentally, the tuning of bond strengths can be achieved by varying the size or the number of attractive patches, in the case of patchy colloids, or by varying the strand lengths, in the case of DNA bricks. Specifically, we require the product $\Delta F_{nuc} \times \Delta F_{min}$ to be negative and their magnitudes such that: 1) There is a sizeable difference in their nucleation rates; and 2) S remains stable for a sufficiently higher window of temperatures than P. In *Optimized Target Compositions and Free Energies*, we perform the optimization of the target composition and the interaction strengths for our example structures.

Optimized Target Compositions and Free Energies

In our example, target S has a boundary of length 36 blocks, and P has a boundary of length 50 blocks. Using the boundary components from S to form the boundary of P, we would need an additional 14 components to make the boundary of P fully inert, thus increasing the size of our component library M to 114. Consequently, only 86 components are shared by both S and P, and each target has 14 components that are unique to it.

We start with an initial assignment such that S is made of components 1 to 100, and P is made of components 15 to 114. We then iterate on the internal permutation of components and their orientations so as to minimize the cost function

$$C = n_{shared-bonds} + n_{active-boundary} + f \times \sum_{t\in\{S,P\}} \sum_{\substack{(\alpha,\beta)\\ shared}} r^t_{\alpha\beta}, \quad [2]$$

where $n_{shared-bonds}$ is the number of shared bonds between the two targets, $n_{active-boundary}$ is the number of boundary components that have nonzero interactions, and $r^t_{\alpha\beta} = |\mathbf{r}^t_\alpha - \mathbf{r}^t_\beta|$ is the separation between a pair of components α and β in target t, where α and β are shared by both targets. $f \ll 1$ is a weight factor to ensure that the first two terms are minimized with higher priority. At each step, we swap the identity of two random sites within a target or perform random rotation of a randomly chosen component in one of the targets and accept the move with probability $\exp(-\gamma\Delta C)$, with $\gamma = 0.05^{-1}$ and the weight $f = 0.0002$ for our specific example. The third term ensures that the shared components form compact cores in either target. The reason for this shall be clear later, when we introduce the idea of vestigial aggregation as a chemical buffer.

The composition of the targets after such an optimization is shown in Fig. 2*A*. The open green/white and orange/white blocks are unique to either target, while the other components are shared. The choice of a larger library ensures that there are no interacting boundary edges, and, in our specific case, there are no shared bonds between the two targets either; i.e., the first two terms of Eq. **2** are zero in the optimized composition. Note that the composition shown here is not unique and that there may be many such equivalent arrangements.

Once the compositions are fixed, we tune the bond strengths. This is done by initializing the strength of all designed bonds to $-\epsilon$ and performing random updates on the individual bonds and accepting them with probability $\exp(-\gamma\Delta D)$, where γ is the inverse optimization temperature (0.03^{-1} in our case) and $D = \Delta F_{nuc} \times \Delta F_{min}$. The value of ΔF_{nuc} and ΔF_{min} are computed at a fixed temperature $k_B T/\epsilon = 0.156$ such that with all bonds at equal strength, both structures have a nucleation barrier of about $10 k_B T$. For the free-energy calculations, we took the free-monomer concentration $\rho \equiv \rho_\alpha = 7.5 \times 10^{-5}$ for all component species α, which is the value we use in our Monte Carlo (MC) simulations. We tuned the bond strengths until $|\Delta F_{nuc}| \approx 3 k_B T$, which gave $|\Delta F_{min}| \approx 42 k_B T$. This corresponds to about 40 MC updates per bond. The distribution of bond strengths for both S and P after the optimization are shown in Fig. 2*B*. While this method works well, it is not optimal for targeting the individual magnitudes of ΔF_{nuc} and ΔF_{min}, and one might consider a different cost function. Further, we emphasize that the free-energy curves can also be optimized in other ways, as discussed in *SI Appendix*, section 4.

In Fig. 2*C*, we show the free-energy curves after the optimization of the bond strengths. With the tuned interactions, the

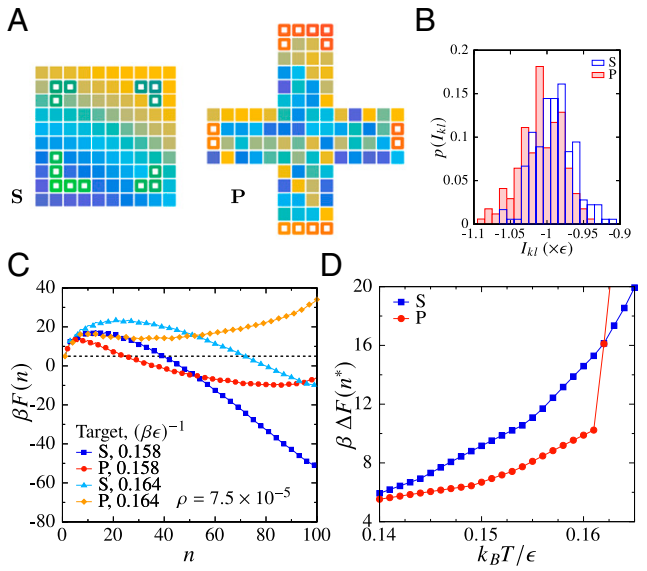

Fig. 2. *A* shows the target structures with optimized internal composition. The colors of the blocks represent the component type. The open squares in green and orange shades are unique to either target, while the other components are shared. *B* shows the distribution of bond strengths for the two structures after they have been tuned, as described in the text. *C* shows their free-energy curves after the optimization. *D* shows the nucleation barriers of the two structures versus temperature after the tuning.

rate at which the nucleation barriers increase with temperature is different for the two structures, as shown in Fig. 2*D*. Hence, their nucleation times also grow apart with temperature, and this feature is crucial for our design. At $k_B\,T/\epsilon \approx 0.163$, P becomes unstable.

Retrieving the Targets through Temperature Protocol

For target P, the simulation temperature is chosen such that the estimated nucleation time of S is greater than P by a factor of at least 10. The estimation of the nucleation times is described in *SI Appendix*, section 5. For target S, we start at a lower temperature, where the nucleation time of S is no more than four times that of P. We then ramp up to a temperature where P becomes unstable, which can be slightly higher than predicted by the free-energy curves, and wait until the nuclei of P are dissolved. This is repeated multiple times to improve the yield of S, similar in spirit to kinetic proofreading mechanisms (39, 40).

We perform simulations with a fixed number of particles, instead of with a constant chemical potential, using the virtual move MC (VMMC) method (41, 42). For selective retrieval of S, either protocol will work equally well. However, under conditions where P can be selectively nucleated, the assembled structure is not complete (Fig. 2*C*), and lowering the temperature to complete the assembly, under constant chemical potential conditions, leads to the nucleation of S as well. While there may be other ways of addressing this issue, we consider closed systems—which may be easier to realize experimentally—and devise a suitable protocol for such a setup.

Since we simulate the system in the canonical ensemble and allow for multiple copies of each target, the depletion of free monomers with time shifts the free-energy curves up. We compensate for this by lowering the simulation temperature. The temperature shift can be approximated by requiring that the free energy (divided by the temperature) of the targeted structure remains invariant with depletion (*SI Appendix*, section 6) and is given by

$$\beta(t) = \beta(0) + \frac{N}{E(N)} \ln \frac{\rho(0)}{\rho(t)}, \qquad [3]$$

where $\beta(t)$ and $\rho(t)$ represent the inverse temperature and the monomer concentration at time t, and $E(N)$ is the potential energy of the fully grown targeted structure, which is of size N. Note that we consider all component species corresponding to the targeted structure to be consumed at the same rate.

Fig. 3 *A* and *B* show the temperature protocols used to retrieve the targets P and S, respectively. We perform the simulations on 200×200 lattices with three copies of each component type. Correspondingly, both the protocols have three sections, each lower in temperature than the previous section, so as to ensure that the free energy of the targeted structure remains invariant with monomer depletion, according to Eq. **3**. We infer the monomer densities from the nucleation rate estimates (*SI Appendix*, section 5) to adapt the temperature of the protocols with time, and the lengths of the sections are chosen such that the corresponding structures have sufficient time to nucleate. The depletion of monomers, and hence the temperature shift, is discrete in the present simulations given the small system size, but a more continuous variation, according to Eq. **3**, will apply for larger system sizes, such as those relevant experimentally.

Nonshared Components Are Detrimental to Selective Retrieval. Fig. 3 *D* and *E* show the yields of the two targets (open symbols) obtained with their corresponding protocols. The yield is computed as the fraction of particles that have assembled into copies of a given target structure, and the results are averaged over thermal fluctuations in a window of 10^5 MC steps and 96 independent runs. The oscillations in the yield of P seen in Fig. 3*E* correspond to the formation and melting of P, during a temperature cycle.

Although the yields of the targeted structures are high (>80%), there is a small fraction of the other structure that is formed. As the targets only partly share the component library, targeting either leads to differential depletion of components. This lowers the free energy of the structure that is not being targeted. This is shown in Fig. 3*C* for the case when P is targeted. At time t_1 indicated in Fig. 3*A*, two-thirds of the shared components are consumed, and the difference in the free-energy barriers between S and P is lower than the initial value. This would not be the case if all components were depleted equally. As a result, some copies of S are nucleated, as shown by the open squares in Fig. 3*D*. This effect is more pronounced the fewer the fraction of components shared by the structures (*SI Appendix*, section 7).

A similar effect occurs when targeting S, where nuclei of P are not completely melted as shown by the open circles in Fig. 3*E*. Thus, a fraction of the nuclei of P are never recycled to form S. This can be rectified by slightly increasing the melting temperature, as shown by the dashed lines (protocol 3) in Fig. 3*B*. The corresponding yields are shown in filled symbols in Fig. 3*E*. However, if the differential depletion leads to a significant lowering of the free energy of P, then the increase in temperature required to dissolve P might also dissolve S. This would happen if the two structures shared fewer components. We illustrate such a case in Fig. 3*F* and discuss further in *SI Appendix*, section 7. These results demonstrate that good design requires the components to be maximally shared between the two structures.

Vestigial Aggregation as a Buffer for Differential Depletion. In protocol 1, the nucleation barrier to S decreases as aggregation proceeds, owing to the relative increase in the concentration of the nonshared components. However, we can mitigate this by programming attractive interactions between the nonshared components of S (the open green blocks in Fig. 2*A*), such that they form an additional or vestigial aggregate V faster than S can nucleate, thus acting as a buffer.

This is possible because the components from which we construct V are not consumed by P, whereas a majority of the species that constitute S are. Combined with the decreasing temperature,

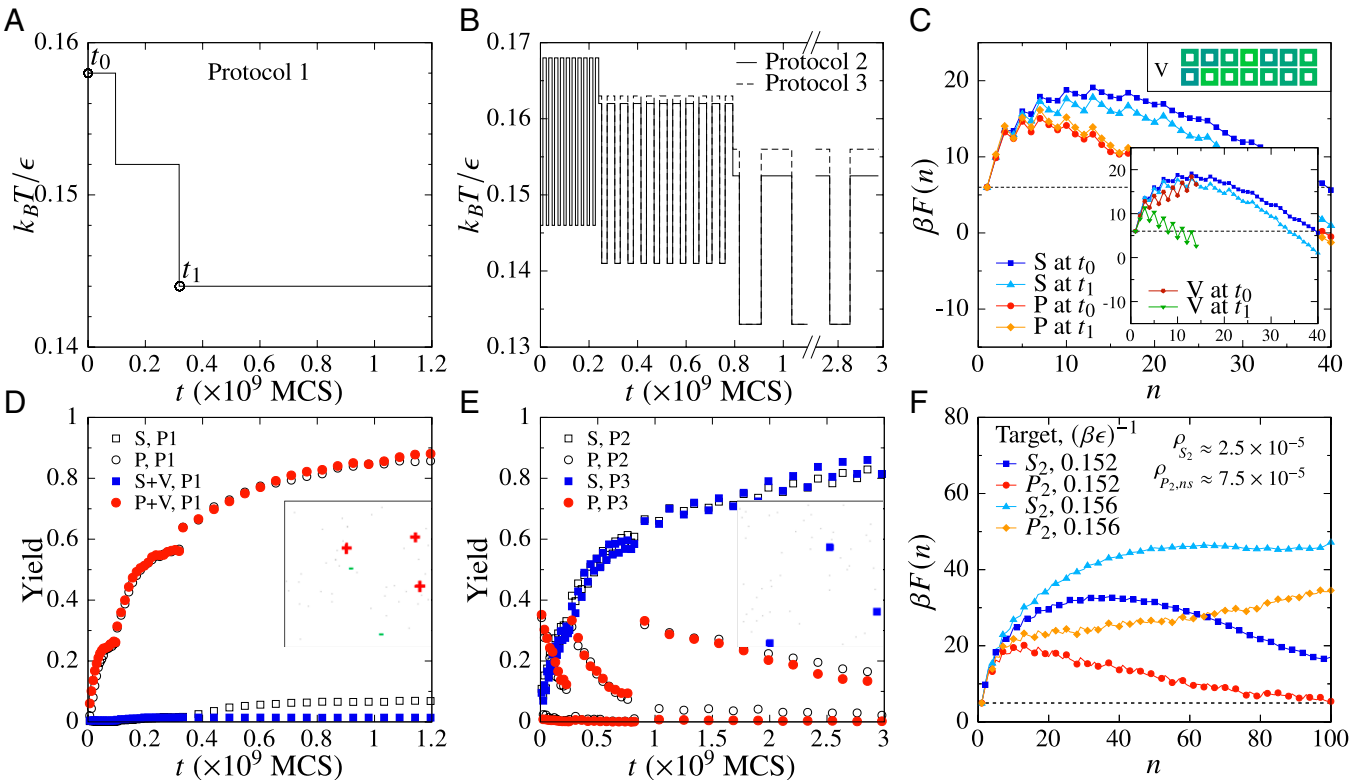

Fig. 3. *A* and *B* show the protocols used to retrieve the targets P and S, respectively. Simulations were done on 200 × 200 lattices with three copies of each component. Both protocols have three sections, each lower in temperature than the previous, to compensate for monomer consumption (Eq. **3** and text). As the targets only partly share components, growth of either differentially lowers the free energy of the other. *C* shows this for protocol 1. The nucleation barrier of S is lower at time t_1 than at t_0. Hence, a fraction of S is also formed, as shown by the open symbols in *D*. Similarly for protocol 2, the free energy of P is lowered, and it no longer melts completely, as shown by the open symbols in *E*. We create a vestigial aggregate V (*C*, top right corner) from the nonshared components of S (open green squares, Fig. 2*A*), with interactions such that it stabilizes faster than S (*C*, *Inset*). V acts as a buffer, improving selective retrieval of P, as shown by the filled symbols in *D*. For targeting S, a slight increase in the melting temperature (protocol 3, *B*) ensures that P is fully melted, as shown by filled symbols in *E*. *D* and *E*, *Insets* show typical configurations obtained at the end of protocols 1 and 3, where aggregates matching the targets P, S, and V are shaded in red, blue, and green, respectively. *F* shows a case when two structures S_2 and P_2 share only one-third of the component library (*SI Appendix*, section 7). When two copies of S_2 are formed, the free-monomer concentration of the components of S_2 is $\rho_{S_2} \approx 2.5 \times 10^{-5}$, and that of the nonshared components of P_2 is $\rho_{P_2,ns} \approx 7.5 \times 10^{-5}$. This stabilizes P_2 at the expected melting temperature ($k_BT/\epsilon = 0.152$). The increased temperature required to melt P_2 ($k_BT/\epsilon = 0.156$) also destabilizes S_2.

this allows us to choose the bond strengths, such that at the start of protocol 1, V is unstable, and, with time, it stabilizes with a much lower barrier than S. Such an aggregate is shown in the top right corner of Fig. 3*C*, and its free-energy curves at times t_0 and t_1 are shown in Fig. 3 *C*, *Inset*. It is constructed so as to not share bonds with any other structure composed of the vestigial components, to avoid chimeric aggregation, and with all bonds having strength -1.11ϵ. However, its edges have nonzero interactions, as it is constructed from components internal to S. Also note that such a scheme is not effective for S, nor is it needed, as any such aggregate (as also P) will be melted during temperature cycling.

In Fig. 3*D*, we show the yields of the two structures with protocol 1 and in the presence of additional interactions encoding the vestige V (filled symbols). The aggregate V acts as a buffer improving selective retrieval of P. For protocol 3, the presence of the additional interactions corresponding to V do not interfere with the nucleation of S. This is because we constructed S such that the components that form V (open green blocks, Fig. 2) are not at the core of S. With the updated protocol 3 and the vestigial structure, we find both good yields and selectivity. Note that at the end the protocols shown in Fig. 3, the assemblies are not complete. However, they are completed upon further cooling, as shown in *SI Appendix*, section 8. Finally, we note that the vestigial aggregation works well to buffer differential depletion, even if a smaller fraction of components are shared (*SI Appendix*, section 7), when targeting P.

Summary

We have investigated the design of a multicomponent mixture that is capable of forming more than one aggregate and the ability to selectively target one of these structures through a suitable temperature protocol. Our results reveal that developing protocols for such addressable self-assembly crucially involves the design of the two structures, so as to prevent spurious or chimeric aggregates and to maximize the selectivity of retrieval. We consider selective assembly of either one of two competing structures and the generic situation, wherein one of them has a lower free-energy barrier to nucleation, whereas the other structure has lower free energy upon complete assembly. We show that such a situation can be realized through an appropriate choice of interaction strengths of components. We show that the avoidance of spurious aggregation pathways favors a design where individual components have different bonding partners in the two structures. Our design also requires inclusion of some components that are not shared by both target structures, but we find that selectivity of assembly is maximally achieved when the components shared by both structures are maximized. Our results highlight the role played by secondary aggregation products, which we term "vestigial aggregates." We demonstrate that with

such design, protocols of temperature variation can be defined that selectively assemble the desired structures with high selectivity and yield. We believe these results provide valuable guidance for further experimental investigations of addressable self-assembly of multiple targets in multicomponent self-assembling systems.

Materials and Methods

Lattice Model for Self-Assembly. We model our self-assembly mixture on a two-dimensional lattice of size $L \times L$, where each site can be empty or occupied by one of the M possible components, which are square blocks with four distinct interacting edges and orientations. Let $\rho_i = n_i/L^2$ be the concentration of the component i. We denote the interaction strength between two component edges $k, l \in \{1, 2, 3, \ldots, 4M\}$ by I_{kl}. Given a configuration of the system, its potential energy is given by

$$E = \sum_{\langle kl \rangle} I_{kl}, \qquad [4]$$

where the sum is performed over all pairs of component edges that are in contact. We simulate the self-assembly dynamics using the VMMC method (41), following Whitelam and Geissler (42), with the translational diffusion of clusters scaling as n_C^{-1} and the rotational diffusion scaling as I_C^{-1}, where n_C and I_C represent the size of the cluster and the moment of inertia of the cluster about the rotation axis.

Data Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We acknowledge support from the UK-India Education and Research Initiative and the Department of Science and Technology, India, under Grants IND/CONT/G/16-17/104, DST/INT/UK/P-149/2016, the Thematic Unit of Excellence on Computational Materials Science, and the National Supercomputing Mission facility (Param Yukti) at the Jawaharlal Nehru Centre for Advanced Scientific Research for computational resources. S.S. acknowledges support through JC Bose Fellowship JBR/2020/000015 from the Science and Engineering Research Board, Department of Science and Technology, India.

1. G. M. Whitesides, B. Grzybowski, Self-assembly at all scales. *Science* **295**, 2418–2421 (2002).
2. G. M. Whitesides, M. Boncheva, Beyond molecules: Self-assembly of mesoscopic and macroscopic components. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4769–4774 (2002).
3. S. Kühner *et al.*, Proteome organization in a genome-reduced bacterium. *Science* **326**, 1235–1240 (2009).
4. P. T. Lansbury, Jr, Evolution of amyloid: What normal protein folding may tell us about fibrillogenesis and disease. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3342–3344 (1999).
5. F. Chiti, C. M. Dobson, Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
6. J. P. Hendrick, F. U. Hartl, Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.* **62**, 349–384 (1993).
7. A. Jaiman, M. Thattai, Golgi compartments enable controlled biomolecular assembly using promiscuous enzymes. *eLife* **9**, e49573 (2020).
8. A. Yadav, Q. Vagne, P. Sens, G. Iyengar, M. Rao, Glycan processing in the Golgi–optimal information coding and constraints on cisternal number and enzyme specificity. arXiv [Preprint] (2020). https://arxiv.org/abs/2005.08740 (Accessed 21 October 2021).
9. T. D. Nguyen, E. Jankowski, S. C. Glotzer, Self-assembly and reconfigurability of shape-shifting particles. *ACS Nano* **5**, 8892–8903 (2011).
10. T. Gong *et al.*, Thermally activated reversible shape switch of polymer particles. *J. Mater. Chem. B Mater. Biol. Med.* **2**, 6855–6866 (2014).
11. J. W. Yoo, S. Mitragotri, Polymer particles that switch shape in response to a stimulus. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11205–11210 (2010).
12. S. Whitelam *et al.*, The impact of conformational fluctuations on self-assembly: Cooperative aggregation of archaeal chaperonin proteins. *Nano Lett.* **9**, 292–297 (2009).
13. V. M. Batista, M. A. Miller, Crystallization of deformable spherical colloids. *Phys. Rev. Lett.* **105**, 088305 (2010).
14. W. Lewandowski, M. Fruhnert, J. Mieczkowski, C. Rockstuhl, E. Górecka, Dynamically self-assembled silver nanoparticles as a thermally tunable metamaterial. *Nat. Commun.* **6**, 6590 (2015).
15. S. Sacanna *et al.*, Shaping colloids for self-assembly. *Nat. Commun.* **4**, 1688 (2013).
16. A. Murugan, Z. Zeravcic, M. P. Brenner, S. Leibler, Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 54–59 (2015).
17. W. Zhong, D. J. Schwab, A. Murugan, Associative pattern recognition through macro-molecular self-assembly. *J. Stat. Phys.* **167**, 806–826 (2017).
18. G. Bisker, J. L. England, Nonequilibrium associative retrieval of multiple stored self-assembly targets. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E10531–E10538 (2018).
19. R. D. Barish, R. Schulman, P. W. K. Rothemund, E. Winfree, An information-bearing seed for nucleating algorithmic self-assembly. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6054–6059 (2009).
20. M. Nguyen, S. Vaikuntanathan, Design principles for non-equilibrium self-assembly. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14231–14236 (2016).
21. J. D. Halverson, A. V. Tkachenko, DNA-programmed mesoscopic architecture. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **87**, 062310 (2013).
22. Z. Zeravcic, V. N. Manoharan, M. P. Brenner, Size limits of self-assembled colloidal structures made using specific interactions. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15918–15923 (2014).
23. Y. Ke, L. L. Ong, W. M. Shih, P. Yin, Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177–1183 (2012).
24. L. L. Ong *et al.*, Programmable self-assembly of three-dimensional nanostructures from 10,000 unique components. *Nature* **552**, 72–77 (2017).
25. L. Cademartiri, K. J. Bishop, Programmable self-assembly. *Nat. Mater.* **14**, 2–9 (2015).
26. D. Frenkel, Order through entropy. *Nat. Mater.* **14**, 9–12 (2015).
27. A. Reinhardt, D. Frenkel, Numerical evidence for nucleated self-assembly of DNA brick structures. *Phys. Rev. Lett.* **112**, 238103 (2014).
28. W. M. Jacobs, A. Reinhardt, D. Frenkel, Communication: Theoretical prediction of free-energy landscapes for complex self-assembly. *J. Chem. Phys.* **142**, 021101 (2015).
29. W. M. Jacobs, D. Frenkel, Self-assembly of structures with addressable complexity. *J. Am. Chem. Soc.* **138**, 2457–2467 (2016).
30. W. M. Jacobs, A. Reinhardt, D. Frenkel, Rational design of self-assembly pathways for complex multicomponent structures. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6313–6318 (2015).
31. W. M. Jacobs, D. Frenkel, Self-assembly protocol design for periodic multicomponent structures. *Soft Matter* **11**, 8930–8938 (2015).
32. L. O. Hedges, R. V. Mannige, S. Whitelam, Growth of equilibrium structures built from a large number of distinct component types. *Soft Matter* **10**, 6404–6416 (2014).
33. S. Hormoz, M. P. Brenner, Design principles for self-assembly with short-range interactions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5193–5198 (2011).
34. J. Madge, D. Bourne, M. A. Miller, Controlling fragment competition on pathways to addressable self-assembly. *J. Phys. Chem. B* **122**, 9815–9825 (2018).
35. C. G. Evans, E. Winfree, Physical principles for DNA tile self-assembly. *Chem. Soc. Rev.* **46**, 3808–3829 (2017).
36. J. Madge, M. A. Miller, Design strategies for self-assembly of discrete targets. *J. Chem. Phys.* **143**, 044905 (2015).
37. M. Sajfutdinow, W. M. Jacobs, A. Reinhardt, C. Schneider, D. M. Smith, Direct observation and rational design of nucleation behavior in addressable self-assembly. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5877–E5886 (2018).
38. A. Trubiano, M. Holmes-Cerfon, Thermodynamic stability *versus* kinetic accessibility: Pareto fronts for programmable self-assembly. *Soft Matter* **17**, 6797–6807 (2021).
39. J. J. Hopfield, Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4135–4139 (1974).
40. J. Ninio, Kinetic amplification of enzyme discrimination. *Biochimie* **57**, 587–595 (1975).
41. S. Whitelam, E. H. Feng, M. F. Hagan, P. L. Geissler, The role of collective motion in examples of coarsening and self-assembly. *Soft Matter* **5**, 1251–1262 (2009).
42. S. Whitelam, P. L. Geissler, Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles. *J. Chem. Phys.* **127**, 154101 (2007). Correction in: *J. Chem. Phys.* **128**, 219901 (2008).