Supplementary Information for

Temperature Protocols to Guide Selective Self-Assembly of Competing Structures

Arunkumar Bupathy, Daan Frenkel, and Srikanth Sastry

Srikanth Sastry
E-mail: sastry@jncasr.ac.in

This PDF file includes:
- Supplementary text
- Figs. S1 to S6 (not allowed for Brief Reports)
- SI References
Supporting Information Text

1. Free Energy of Multi-Component Structures

We compute the free energy of our multi-component target structures using the formalism of Jacobs et al. (1, 2). To compute the free-energy difference between on-pathway aggregates of a particular size and the free monomers, we must consider all the ways in which a correctly bonded fragment can be assembled. Let \( D(n, b) \) denote the set of all correctly bonded fragments of size \( n \) monomers and \( b \) bonds of a given structure, and \( |D(n, b)| \) be the number of such fragments or the density of states. The free energy of an on-pathway aggregate of size \( n \) monomers can be written as

\[
\beta F(n) = -\ln \sum_b |D(n, b)| Z_{n,b},
\]

and the average fugacity \( Z_{n,b} \) of a fragment of size \( n \) monomers and \( b \) bonds is given by

\[
\ln Z_{n,b} = b\tilde{\epsilon}_{n,b} + n \ln \rho - (n - 1) \ln q_c.
\]

Here, \( \tilde{\epsilon}_{n,b} \) represents the mean bond energy of the fragment, \( n \ln \rho \) represents the loss in entropy due to bringing \( n \) monomers into contact, \( \rho \) is the density of free monomers in the reservoir, and the last term represents the loss in rotational entropy with \( q_c \) being the co-ordination number of the lattice. The mean bond energy is given by

\[
\beta \tilde{\epsilon} = \frac{1}{b} \ln \langle \exp \left( -\beta \sum_{i \in g} \epsilon_i \right) \rangle_{g \in D(n, b)}
\]

where \( \epsilon_i \) represent the energy of the bonds in a fragment \( g \) in the set \( D(n, b) \).

For each target structure, we make a connectivity graph \( G \) where the vertices of the graph represent the individual components and the edges represent the bonds between them. Connected sub-graphs of \( G \) represent correctly bonded fragments of the structure. We count the number of connected sub-graphs of size \( n \) vertices and \( b \) edges, statistically using the Wang-Landau flat-histogram algorithm (3). We start with a randomly initialized sub-graph \( g \) of \( G \), and a flat \( |D(n, b)| \equiv 1 \) for all \( n \) and \( b \). At each step, we either add a randomly chosen vertex that is adjacent to the given sub-graph \( g \) or remove a randomly chosen vertex from \( g \). The removal of a vertex that breaks the sub-graph \( g \) into disconnected parts (called cut-vertex or articulation point) is not allowed. The acceptance probability for transition from sub-graph \( g \) to \( g' \) is

\[
p(g \rightarrow g') = \min \left[ 1, \frac{|D(n(g), b(g))| n_{+}(g)}{|D(n(g'), b(g'))| n_{+}(g')} \right]
\]

where \( n_{+}(g) \) is the number of vertices in \( G \) that are adjacent to the sub-graph \( g \), and \( n_{-} \) is the number of non-cut-vertices in \( g \).

Each time a state \((n, b)\) is visited, we update \(|D(n, b)|\) by a multiplicative factor \( f > 1 \). We keep sampling the state space of connected sub-graphs, and updating \(|D(n, b)|\) until we sample all states with roughly the same frequency. This procedure is repeated with smaller values of \( f \), until we reach the desired level of accuracy. We then perform a biased Monte Carlo calculation using the density of states we just computed, with the acceptance probability of Eq. (4), so as to compute the average bond energy \( \tilde{\epsilon}_{n,b} \). The density of states need only be computed once. As a result, the free energy computations can be quite efficient, especially when parallelized, taking about 2.5s per target structure on a 32-core CPU operating at 3.8 GHz.

2. Effect of Maximizing the Shared Bonds

We discuss here the effect of maximizing the number of shared bonds between the two targets structures that we consider. In other words, we try to construct the two targets such that most components have the same set of neighbors in both the targets. We start with an initial composition where both targets are made of the same structures that we consider. Let \( M \) denote the set of all correctly bonded fragments of \( n \) monomers and \( b \) bonds of a given structure, and \( |M(n, b)| \) be the number of such fragments or the density of states. The free energy of an on-pathway aggregate of size \( n \) monomers can be written as

\[
\beta F(n) = -\ln \sum_b |M(n, b)| Z_{n,b},
\]

and the average fugacity \( Z_{n,b} \) of a fragment of size \( n \) monomers and \( b \) bonds is given by

\[
\ln Z_{n,b} = b\tilde{\epsilon}_{n,b} + n \ln \rho - (n - 1) \ln q_c.
\]

Here, \( \tilde{\epsilon}_{n,b} \) represents the mean bond energy of the fragment, \( n \ln \rho \) represents the loss in entropy due to bringing \( n \) monomers into contact, \( \rho \) is the density of free monomers in the reservoir, and the last term represents the loss in rotational entropy with \( q_c \) being the co-ordination number of the lattice. The mean bond energy is given by

\[
\beta \tilde{\epsilon} = \frac{1}{b} \ln \langle \exp \left( -\beta \sum_{i \in g} \epsilon_i \right) \rangle_{g \in D(n, b)}
\]

where \( \epsilon_i \) represent the energy of the bonds in a fragment \( g \) in the set \( D(n, b) \).

For each target structure, we make a connectivity graph \( G \) where the vertices of the graph represent the individual components and the edges represent the bonds between them. Connected sub-graphs of \( G \) represent correctly bonded fragments of the structure. We count the number of connected sub-graphs of size \( n \) vertices and \( b \) edges, statistically using the Wang-Landau flat-histogram algorithm (3). We start with a randomly initialized sub-graph \( g \) of \( G \), and a flat \( |D(n, b)| \equiv 1 \) for all \( n \) and \( b \). At each step, we either add a randomly chosen vertex that is adjacent to the given sub-graph \( g \) or remove a randomly chosen vertex from \( g \). The removal of a vertex that breaks the sub-graph \( g \) into disconnected parts (called cut-vertex or articulation point) is not allowed. The acceptance probability for transition from sub-graph \( g \) to \( g' \) is

\[
p(g \rightarrow g') = \min \left[ 1, \frac{|D(n(g), b(g))| n_{+}(g)}{|D(n(g'), b(g'))| n_{+}(g')} \right]
\]

where \( n_{+}(g) \) is the number of vertices in \( G \) that are adjacent to the sub-graph \( g \), and \( n_{-} \) is the number of non-cut-vertices in \( g \).

Each time a state \((n, b)\) is visited, we update \(|D(n, b)|\) by a multiplicative factor \( f > 1 \). We keep sampling the state space of connected sub-graphs, and updating \(|D(n, b)|\) until we sample all states with roughly the same frequency. This procedure is repeated with smaller values of \( f \), until we reach the desired level of accuracy. We then perform a biased Monte Carlo calculation using the density of states we just computed, with the acceptance probability of Eq. (4), so as to compute the average bond energy \( \tilde{\epsilon}_{n,b} \). The density of states need only be computed once. As a result, the free energy computations can be quite efficient, especially when parallelized, taking about 2.5s per target structure on a 32-core CPU operating at 3.8 GHz.
Fig. S1. Panel (A) shows a composition where the number of shared bonds between the two targets $S_1$ and $P_1$ have been maximized, leading to large chunks of one target being reused by the other as indicated by the numbered, outlined regions. There are some boundary species in either target that have non-zero interactions, as they are internal in the other. These are replaced with additional components to make the boundaries inert, as shown in the bottom of panel (A). Panel (B) shows the typical aggregates obtained with the protocol 1 shown in Fig. 3(A) of the main text. Panel (C) shows that the chimeric aggregate ($C_1$) has the same nucleation barrier as that of $P$. But since it is more stable than $P$, most aggregates turn into chimeras.

regions respectively. We identify one of the chimeric aggregates, and compute the free energy curve for the chimera. The free energies of $S_1$, $P_1$ and the chimera $C_1$ are shown in the panel (C) of Fig. S1. The important point to note is that the chimera has a nucleation barrier that is comparable to $P_1$, which we are trying to grow. But the chimera is more stable than $P_1$. So in most cases we observe the chimera rather than the desired structure. Furthermore, we cannot melt the chimeric aggregate through temperature cycling, as that would destroy any nuclei of $P_1$.

3. Spurious aggregation from bonded loops of incompatible components

In the main text, we show how chimeric aggregation can occur if the two targets share the same bond, i.e., a pair of neighboring components. However, even with no shared bonds, there is a possibility of incompatible components forming bonded loops. Such erroneous loops are harder to correct as they require the breaking of two bonds. We illustrate this in Fig. S2. On the left we show two example structures $T_1$ and $T_2$ with different internal arrangements such that there are no common bonds between them. On the right, we show a partially complete aggregate that is compatible with $T_1$, except for one block as highlighted by the broken red rectangle. This incompatible piece is stable against removal as it is held together by two bonds. The complete loop of bonds that stabilizes the incompatible monomer is not present in either $T_1$ or $T_2$, but the individual bonds are. In general, any such loop irrespective of its size can present an opportunity for erroneous aggregation. But the most important loops are those made of four nearest neighbors, as longer loops are much less likely to form. If present in large numbers such loops could also cause chimeric aggregation. However, this is only an issue when the number of stored structures is large or the component library is small. We therefore do not optimize the target compositions against such incompatible loops.

Fig. S2. Two example structures $T_1$ and $T_2$ that differ in their internal arrangement of components such that they have no common bonds. The rightmost panel shows an aggregate that is compatible with $T_1$ except for one monomer. The loop of bonds that stabilizes the incompatible monomer is not present in either $T_1$ or $T_2$, but the individual bonds are.
4. Tuning the free energy landscapes manually

In the main text, we considered the generic case in which one of the target structure has a lower free energy barrier to nucleation whereas the other structure has a lower free energy minimum upon complete aggregation. The manner in which the design of aggregates in such a case could be optimised was described. Here, we demonstrate that target structures can indeed be prepared to satisfy such an assumption. To do so, we start by considering the initial structures considered in the main text, with free energy curves as shown in Fig. 1(B), displaying the same free energy barrier for both structures at the lower temperature shown, and with the S structure having a lower free energy upon complete formation. The free energy barrier of the P structure can be lowered by selecting a region at its centre, which is about the size of the critical nucleus (\( \sim 4 \times 4 \)) and increasing the strength of the bonds within the region, or by increasing the concentration of the species in that region. We need not optimize for the free energy minima since the free energy minimum of P is already higher than S. This is illustrated in Fig. S3 where the panels A and B show the effect of varying the bond strengths and panels C and D show the effect of varying the chemical potentials. Note that the free energies computed using the procedure outlined in Sec. 1 assumes that intermediate sized clusters of a given size \( n \) are equivalent regardless of composition. This assumption is not strictly valid in the present case, and therefore a more careful computation of the free energy is necessary if the precise values of the barrier height are of interest. We do not pursue such a calculation here, since our purpose is simply to demonstrate that the free energy profiles can be modified in desired ways by tuning the interaction energies or chemical potentials.

Since the bonds that constitute the \( 4 \times 4 \) region in P are not shared by the two structures, varying the bond strength of the corresponding bonds does not affect the bonds in S. On the other hand, since the same components are shared between the two structures, increasing the concentration of the species constituting the \( 4 \times 4 \) region in P will also lower the free energy curves of S, but to a smaller degree near the free energy barrier, as these components are not spatially correlated in S.

![Figure S3](image_url)

**Fig. S3.** Panel A shows the strength of the interactions in the two structures S and P. A \( 4 \times 4 \) region in the core of P is given stronger bonds. This reduces the nucleation barrier of P relative to S as shown in panel B. Panel C shows the free monomer concentration of the component species that constitute the two structures, where the components at the core of P are enriched. Panel D shows the free energy curves as a result of the concentration pattern.

5. Computing the Nucleation Rates

We need to know the nucleation rates of the target structures to identify the length of the simulation protocols. We follow Debenedetti (4) to derive the nucleation rates for the multicomponent self-assembly system. Consider the system in a supercooled state, where the gaseous phase is metastable. As a result of density fluctuations, small nuclei are constantly being formed and destroyed. We assume that the concentration of these nuclei is small, and that they grow or shrink by one particle
at a time. We can then write the difference between the rate at which \( n \)-particle nuclei are formed and the rate at which they are destroyed, due to single particle events as

\[ J(n) = f(n-1)P(n-1)\gamma(n-1) - f(n)P(n)\alpha(n) \]  \[ (6) \]

where \( f(n) \) is the concentration of \( n \)-particle nuclei; \( P(n) \) represent their perimeter; \( \gamma(n-1) \) is the flux per unit time and length (area in 3d) of single particles onto an \((n-1)\)-particle nucleus; and \( \alpha(n) \) is the flux of single particles leaving an \( n \)-particle nucleus. Of the two rates \( \gamma \) can be computed, but \( \alpha \) is not known. Assuming that an equilibrium distribution of nuclei can be established in the bulk metastable phase, we can write

\[ C(n-1)P(n-1)\gamma = C(n)P(n)\alpha \]  \[ (7) \]

where we have assumed that the fluxes are independent of the size of the nuclei, and \( C(n) \) is the equilibrium concentration of nuclei of size \( n \). Using (7) in (6), we get

\[ J(n) = \gamma P(n-1)C(n-1) \left[ \frac{f(n-1)}{C(n-1)} - \frac{f(n)}{C(n)} \right] \]  \[ (8) \]

For a time invariant propagation of clusters we should have

\[ \frac{\partial f(n,t)}{\partial t} = J(n) - J(n+1) = 0, \]  \[ (9) \]

and is established when \( J \) becomes independent of \( n \). Rearranging (8), and summing from \( n = 2 \) to \( n = \Lambda \), where \( \Lambda \) is much larger than the critical nucleus size, we obtain

\[ J \approx \frac{f(1)}{C(1)} \sum_{n=1}^{\Lambda} \frac{1}{\rho P(n)C(n)} \left[ \gamma P(n)C(n) \right]^{-1} \]  \[ (10) \]

where we have made use of the fact that in the bulk metastable phase the equilibrium and actual concentration of single particles are nearly the same and that at the onset of the phase transformation the concentration of sufficiently large nuclei is vanishingly small.

The equilibrium concentration \( C(n) = \rho^\text{tot} \exp[-\Delta F(n)/k_B T] \), where \( \rho^\text{tot} \) is the total monomer concentration, and \( \Delta F(n) = F(n) - F(1) \) which we already compute. Here we have assumed that all \( M \) components are present in equal concentration \( \rho \) in the mixture, such that \( \rho^\text{tot} = M\rho \). However, in our multi-component system not all incoming particles can form bonds with the nucleus. This is because each block in a given target structure is a distinct species with specific and directional bonds. Therefore, of the total flux \( \gamma^\text{tot} \) of incoming particles, only a fraction \( (P(n)/M) \) can form bonds with an \( n \)-particle nucleus, and with a probability \( 1/AP(n) \) since the bonds are specific and directional. So the effective value of \( \gamma \) is

\[ \gamma = \frac{1}{4M} \gamma^\text{tot} \]  \[ (11) \]

If the monomers are diffusing with a diffusion constant \( D \) and considering a square nucleus of size \( n = 4r^2 \), we can approximate \( \gamma^\text{tot} = \rho^\text{tot}4\pi^2/t \approx \rho^\text{tot}8DDd \) and \( P(n) \approx 4\sqrt{n} \) which gives

\[ J \approx \frac{8DDd\rho^\text{tot}2}{M} \left[ \sum_{n=1}^{\Lambda} \frac{\exp[\Delta F(n)/k_B T]}{\sqrt{n}} \right]^{-1} \]  \[ (12) \]

In Fig. S4 we show the nucleation times measured from MC simulations as well as the estimates from Eq. (12) scaled by constant factors. The nucleation times in simulations are measured as \( 2 \times \tau_{fp}(n^*) \), where \( \tau_{fp}(n^*) \) is the time at the inflexion point of the mean first-passage time data. While the temperature dependence from MC simulations and the calculation above agree, the calculated rates are different by scale factors which are different for \( S \) and \( P \), being \( \approx 50 \) and \( \approx 90 \) respectively. We use the scaled analytical values to extrapolate the nucleation rates to arbitrary temperatures and free monomer concentrations.

### 6. Scaling the temperature with monomer concentration

Since we perform simulations in the canonical ensemble, growth of structures depletes free monomers. The rate at which the monomers are consumed can be written as

\[ \frac{d\rho}{dt} = -k\rho \]  \[ (13) \]

where \( \rho \) is the free monomer concentration of the components in the mixture, and the kinetic factor

\[ k = Ae^{-\beta F(n^*)} \]  \[ (14) \]

where \( F(n^*) \) is the nucleation barrier of the targeted structure, with a critical nucleus of size \( n^* \) at an inverse temperature \( \beta \).

We want \( k \) to remain constant, even while \( \rho \) changes so that

\[ \rho = \rho_0 e^{-kt} \]  \[ (15) \]
where $\rho_0 = \rho(0)$ is the initial monomer concentration. For this to be true, we require
\[
\beta(t_1)F(n^*, \rho(t_1)) = \beta(t_2)F(n^*, \rho(t_2)) = \text{const.} \quad [16]
\]
If we assume
\[
\frac{\beta(t_1)F(n^*, \rho(t_1))}{\beta(t_2)F(n^*, \rho(t_2))} = \frac{\beta(t_1)F(N, \rho(t_1))}{\beta(t_2)F(N, \rho(t_2))} \quad [17]
\]
and from Eq. (1) and (2) we can write
\[
\beta(t)F(N, \rho(t)) = \beta(t)E(N) + N\ln \rho(t) - (N - 1)\ln q_c
= \beta(t)E(N) + N[-kt + \ln \rho_0] - (N - 1)\ln q_c
= \beta(0)F(N, \rho_0)
= \beta(0)E(N) + N\ln \rho_0 - (N - 1)\ln q_c \quad [18]
\]
This is only true if $\beta(t)E(N) - Nkt = \beta(0)E(N)$ which gives
\[
\beta(t) = \beta(0) + \frac{N}{E(N)}kt = \beta(0) + \frac{N}{E(N)}\ln \frac{\rho_0}{\rho(t)} \quad [19]
\]
7. Effect of sharing a smaller fraction of the component library

In Fig. S5 (A), we show two structures $S_2$ and $P_2$, which are of the same shape and size ($N = 100$) as those in the main text, but they share a smaller fraction of the component library. Specifically, they only share 50 components (solid squares) which are internal. The components that compose each of their boundaries (open squares) are unique to either structure. Thus the total component library is of size 150. The permutation of components is optimized as described in the main text, and we tune the free energy curves of $S_2$ and $P_2$, as described in the main text. For retrieving $P_2$, we use the same protocol 1 as shown in Fig. 3(A) of the main text, and protocol 4 shown in panel (C) for $S_2$. The differential consumption of monomers by the either target lowers the free energy of the other structure. The free energies of the different structures, at times $t_0$ and $t_1$ of protocol 1 (Fig. 3 (A), main text) are shown in panel (B). The free energy barrier of $S_2$ is lower than that of $P_2$, the targeted structure, at time $t_1$.

A similar effect is observed for $S_2$ as well (Fig. 3(F), main text). As the simulation proceeds, and two copies of $S_2$ are formed, the free energy curve of $P_2$ is shifted below $S_2$ at the temperature ($k_B T/\epsilon = 0.152$) where we expect $P_2$ to be unstable. This leads to the formation of the undesired structure, and lowers the yield of the desired one as shown by the open symbols in panel (D) and (E) of Fig. S5. The increase in temperature required to melt $P_2$ also destabilizes $S_2$, and hence the yield of $S_2$ is limited, as shown by the solid symbols in Fig S5(E). Also note that the melting temperature for $P_2$ in simulations can be slightly higher ($k_B T/\epsilon = 0.158$) than that predicted from the free energy curves ($k_B T/\epsilon = 0.156$, as shown in Fig. 3(F) of the main text).

With additional vestigial aggregates shown in panel (A) constructed from the non-shared components of $S$, the selectivity of $P_2$ can be improved significantly. These vestiges are constructed so that they do not share any bonds with $S_2$. But note that their boundaries can have non-zero interactions, as some of them are internal in $S_2$. These aggregates nevertheless buffer the differential depletion of components and improve the yield of $P_2$ as shown by solid symbols in Fig. S5(D).
Fig. S5. Panel (A) shows two structures S₂ and P₂ that share only 50 components (solid squares). The open squares of green-blue and orange-red shades are unique to either. The vestiges are constructed from the non-shared components (open blocks) of S₂. In the absence of vestiges, differential consumption of components by either target lowers the free energy of the other, as shown in panel (B) at two different times along protocol 1 of Fig. 3(A) in the main text. This lowers the yield of P₂ and a significant fraction of S₂ is also nucleated as shown by the open symbols in panel (D). With the vestiges, the yield of P₂ is significantly increased as shown by the solid symbols. Using protocol 4 shown in panel (C), which is constructed expecting all components to be consumed equally, a significant fraction of P₂ is never melted during the temperature cycling (open circles, panel (E)). Increasing the temperature (protocol 5, panel (C)) to melt fragments of P₂ also melts the nuclei of S₂, thus reducing the yield of S₂ as shown by solid symbols in panel (E).

8. Snapshots of the completed assemblies

In Fig. S6 we show typical snapshots of the assemblies formed upon further cooling the system to \((\beta \epsilon)^{-1} = 0.11\) and simulating for an additional \(10^9\) MCS, at the end of protocols 1 and 3 shown in Fig. 3 of the main text. At the end of these extended simulations we observe 280 perfect copies of P for protocol 1, and 254 perfect copies of S for protocol 3, out of the expected maximum of 288 copies of either (3 copies \(\times\) 96 independent runs). These fractions are higher than the yield we observed at the end of protocols 1 and 3 (see Fig. 3 of the main text), which we defined as the fraction of particles that have assembled into either structures. This indicates that the fraction of correctly assembled blocks at higher temperatures is a good indicator of the fraction of perfectly assembled structures that form upon annealing to lower temperatures.
Fig. S6. Snapshots of the assemblies formed upon cooling the system to $(\beta\epsilon)^{-1} = 0.11$ and simulating for an additional $10^9$ MCS, at the end of protocols 1 (panel A) and 3 (panel B) shown in Fig. 3 of the main text.

References

1. WM Jacobs, A Reinhardt, D Frenkel, Communication: Theoretical prediction of free-energy landscapes for complex self-assembly. *The J. Chem. Phys.* **142**, 021101 (2015).
2. WM Jacobs, A Reinhardt, D Frenkel, Rational design of self-assembly pathways for complex multicomponent structures. *Proc. Natl. Acad. Sci.* **112**, 6313–6318 (2015).
3. F Wang, DP Landau, Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **86**, 2050–2053 (2001).
4. PG Debenedetti, *Metastable Liquids: Concepts and Principles*. (Princeton University Press, Princeton, New Jersey), (1996).