Optimization of Non-Equilibrium Self-Assembly Protocols Using Markov State Models

Anthony Trubiano, Michael Hagan

Department of Physics, Brandeis University

Motivation

Self-assembling systems are typically designed to optimize the equilibrium stability of a target structure; however, equilibrium is not always reached on experimentally accessible timescales. Such systems may get stuck in long-lived metastable states, called kinetic traps, resulting in yields much smaller than the equilibrium estimate [1].

An increasingly studied approach to this problem is to introduce some time-dependent driving force to the system to drive it out of equilibrium and help facilitate formation of the target state. This method has seen success in simulations [2] and experiment [3], but the choice of time-dependent protocol falls on the user. Here we use tools from optimal control theory and Markov State Model (MSM) analysis to compute protocols that are optimal for a given target state.

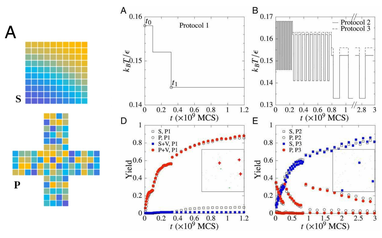


Image: Bupathy et al,. PNAS (2022) [2]

We illustrate our approach on a model for T=4 icosahedral capsid growth, showing improved assembly yields and rates with the computed protocol. These techniques could be used within IRG1 to improve yields of capsids (Fraden lab) and tubules (Rogers lab) constructed from triangular sub-units, as well as in the Rogers lab for their experiments with DNA coated nano-particles.

Example: T=4 Icosahedral Capsids

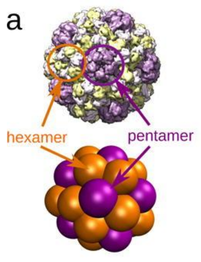
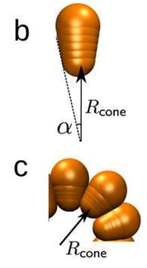
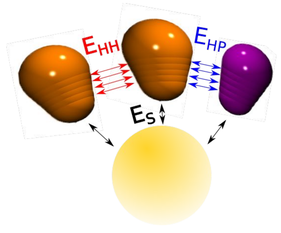


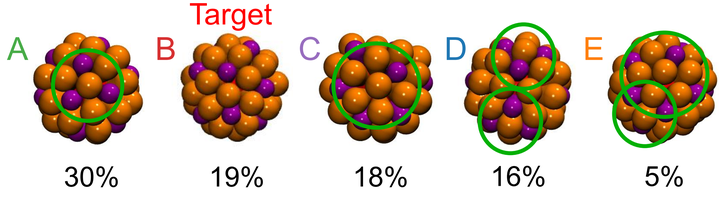
Image: Lazaro et al., Soft Matter (2018) [4]





Schematic of interactions between the two subunit types and the spherical nanoparticle.

Our model is adapted from [4], which considered assembly of conical subunits on curved nanoparticles. We use subunits of different sizes to represent hexamers and pentamers, and let the hexamer-hexamer and hexamer-pentamer interaction strength vary. We use HOOMD to simulate their assembly on the surface of a spherical nanoparticle for 800,000 non-dimensional time units.



For the best choice of interaction strength parameters, the target T=4 capsid has a yield of around 20%. It is more likely, however, that the system ends up in a configuration consisting of one or more *defects*. **Can we achieve more efficient target assembly by computing an optimal time-dependent protocol for the interaction strengths?**

Markov State Models

Markov State Models (MSMs) are a powerful tool for coarse-graining the dynamics of complex molecular systems into a form that is tractable to analyze [5]. The state space is discretized into a set of macro-states, and the dynamics are encoded by estimating the probability to transition between each of these states after a lag time, τ.

For our example, we define discrete states according to the number of each type of subunit attached to the nanoparticle, as well as a measure of the T=4 icosahedral symmetry of their configuration.

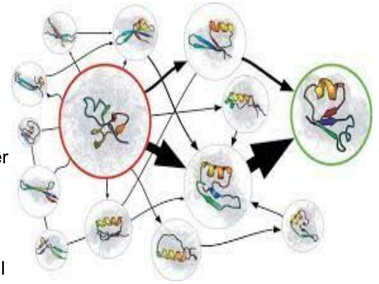
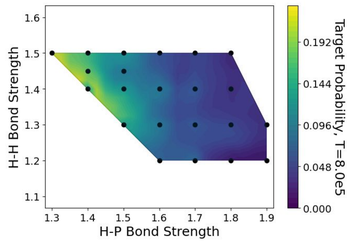


Image: Voelz et al,. JACS (2010) [5]



We construct local MSMs at each node in the feasible 2D parameter space (left) by sampling trajectories. Radial basis function interpolation is used to evaluate the transition probabilities away from the nodes. The interpolants can be used to build an estimate of the system transition matrix for non-sampled parameter values.

The probability of the target state can be extracted for any parameter set and time, either to inform experimental setups or help infer unknown system parameters.

Optimization Problem

The evolution of the probability distribution over coarse-grained states, $\vec{p}$, is given by the Forward Kolmogorov Equation as

$$\vec{p}^{\,n+1} = \vec{p}^{\,n} P(\theta_n) \qquad \vec{p}^{\,0} = \vec{p}_0$$

Here, P is the transition matrix evaluated at the system parameters at time step n, represented by θ_n. With the probabilities constrained by this equation, we can pose an optimization problem to maximize the probability of a target set of states, B, at the final time, N, by computing

$$p^* = \max_{\{\theta_k\}_{k=1}^N} p_B^N \quad \text{where} \quad p_B^N = \sum_{i \in B} p_i^N$$

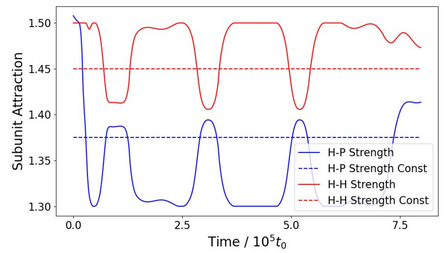
Derivatives of this probability are inefficient to compute directly, so we instead use an adjoint method. The adjoint problem is the Backward Kolmogorov Equation, given by

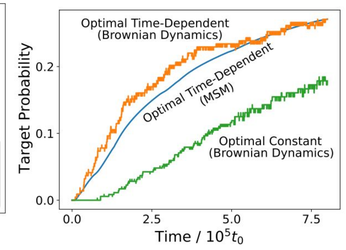
$$\vec{F}^{\,n} = P(\theta_n)\vec{F}^{\,n+1} \quad \text{where} \quad F_i^k = E[1_B(X_N)|X_k = i]$$

This equation is solved backward in time for the adjoint variable, $\vec{F}$, which represents the conditional expectation of the indicator function for the target set, where X_k denotes the state of the system at time k. Derivatives can now be efficiently computed using the expression on the left. We use this to perform gradient descent over the protocol path. A number of penalty functions can also be included; for example, to ensure smooth protocols that are physically realizable in an experiment.

$$\frac{\partial p_B^N}{\partial \theta_k} = \vec{p}^{\,k} \frac{\partial P_k}{\partial \theta_k} \vec{F}^{k+1}$$

Optimization Results

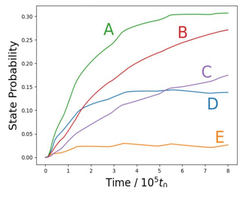


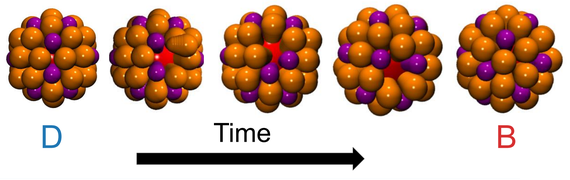


The computed optimal protocol begins with parameters that give fast but low yield assembly, then transitions to parameters that give slower, but higher yield, assembly.

To test our method, we sample 250 trajectories according to this protocol, and find that the estimated yield matches well with the MSM prediction. By comparing to simulations using the best constant protocol, the optimal protocol indeed achieves a higher yield at the final time, as well as a faster average assembly rate.

One of the ways this protocol boosts the target yield is by ratcheting the system out of some of the defect states. This can be seen below as a decrease in the probability of some states over time. Below, we also show snapshots along the transition pathway from a state with two defects into the target state.





Acknowledgements

We would like to thank Eric Vanden-Eijnden for the idea for applying optimal control principles to self-assembling systems modeled by MSMs, and Miranda Holmes-Cerfon for helpful discussions on the method.

We acknowledge support from NIH R01GM108021 and the Brandeis NSF MRSEC, Bioinspired Soft Materials, DMR-2011846. We also acknowledge computational support from NSF XSEDE computing resources allocation TG-MCB090163 and the Brandeis HPCC which is partially supported by the NSF through DMR-MRSEC 2011846 and OAC-1920147.

References

[1] Trubiano and Holmes-Cerfon, Thermodynamic Stability versus Kinetic Accessibility: Pareto Fronts for Programmable Self-Assembly, Soft Matter (2021)
[2] Bupathy et al., Temperature Protocols to Guide Selective Self-Assembly of Competing Structures, PNAS (2022)
[3] Basagoiti et al., Design Rules for Two-DImensional Colloidomer Folding, Bulletins of the APS (2022)
[4] Lazaro et al., Self-Assembly of Convex Particles on Spherocylindrical Surfaces, Soft Matter (2019)
[5] Voelz et al., Molecular Simulation of Ab Initio Protein Folding for Millisecond Folder NTL9(1-39), JACS (2010)