Optimizing the Free-Energy and Kinetic Accessibility for Systems of Self-Assembling Colloids

Anthony Trubiano

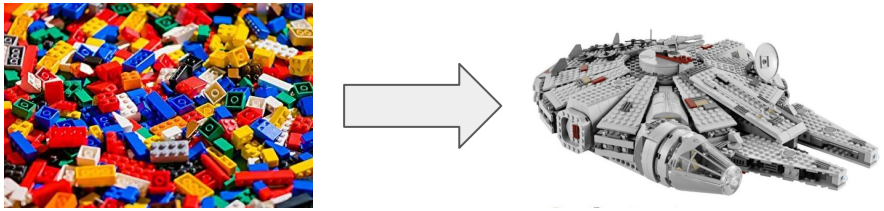
MSG Lunch Talk

May 7th, 2020

What is Self-Assembly?

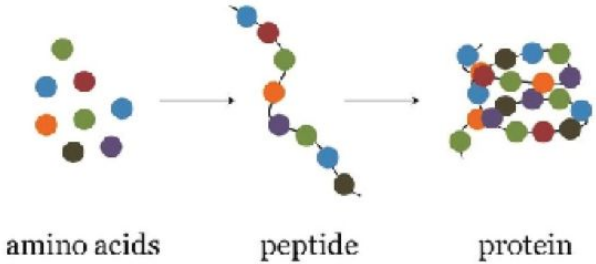
Self Assembly:

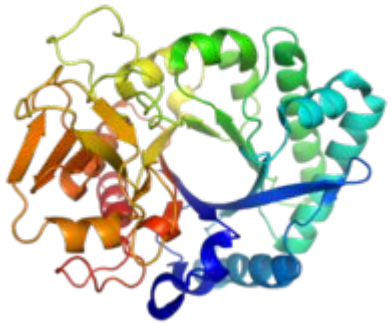
Spontaneous organization of a collection of individual units into a well-defined structure, without human (external) intervention.



Canonical Example:

Protein Folding.

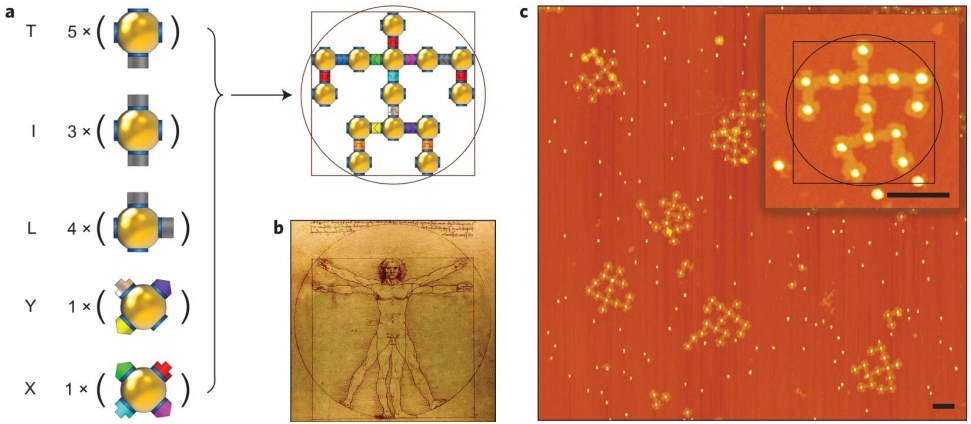




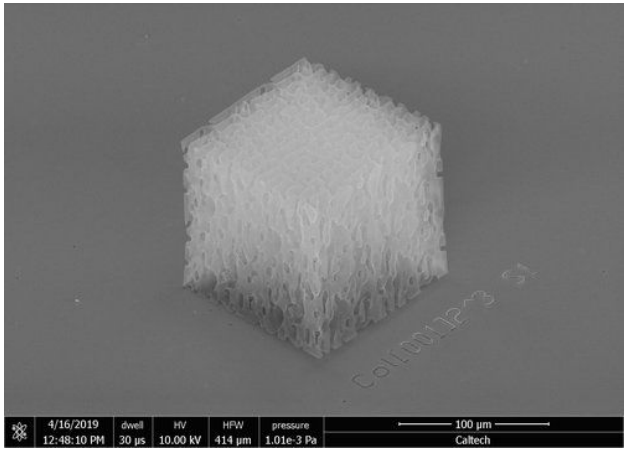
Huge State Space -> Levinthal's Paradox
198 bond angles, 3 stable configurations.

$$3^{198} \cong 10^{95}$$

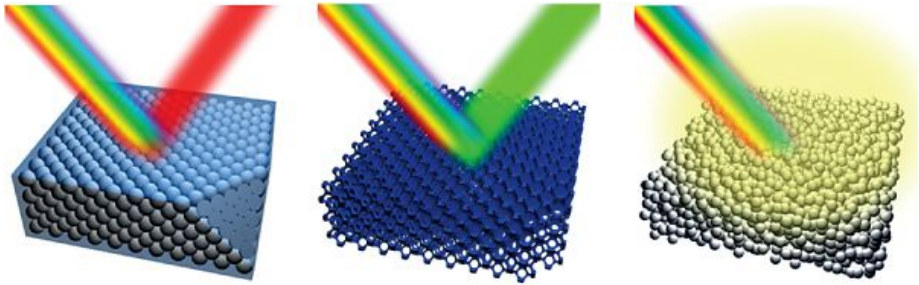
Reverse-Engineering Mother Nature



Oleg Gang. Nature, 2016.

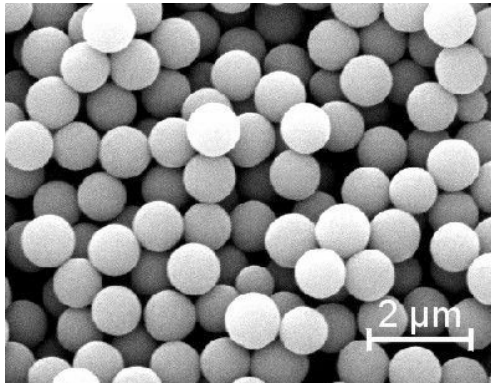


Greer Lab, CalTech 2020.

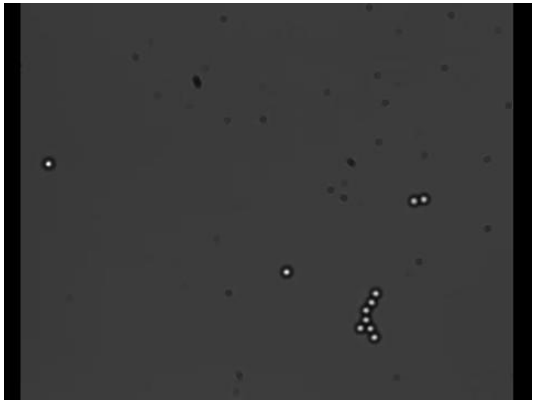


The Search for Design Principles - A Simpler Model

- We'll look at colloids as a model system.
  - Size range: 100nm to a few microns
  - Short ranged interactions
- Advantages:
  - Relatively simple
  - Can be directly studied experimentally
  - Can exhibit interesting behavior



Model Problem:
The folding of a linear chain of N (6 or 7) colloids, modeled as 2D disks of unit diameter.
(Movie: Ellen Klein, Manoharan Lab)



Simulating the Linear Colloidal Chain

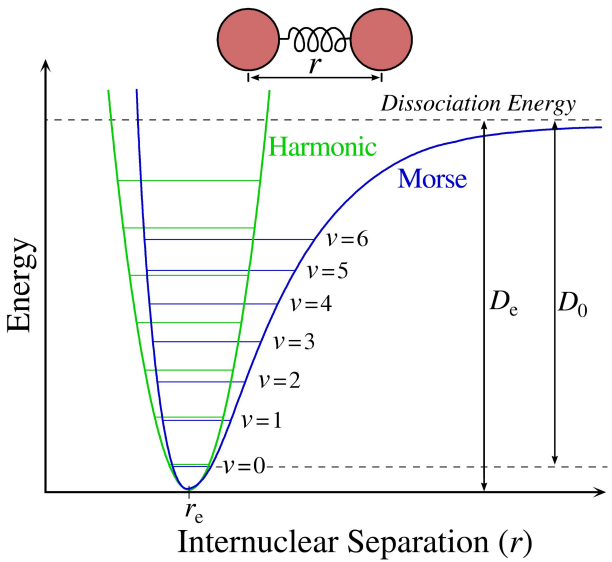
Brownian Dynamics Simulation

$$X'(t) = -\frac{1}{\gamma}\nabla U(X) + \sqrt{\frac{2}{\beta\gamma}}R(t)$$

$$U_M(r) = E\left(e^{-2\rho(r-d)} - 2e^{-\rho(r-d)}\right)$$

$$\kappa_M = \frac{e^E}{\sqrt{2E\rho^2}}$$

(Interaction Strength)



A Numerical Experiment - 6 Particle Folding

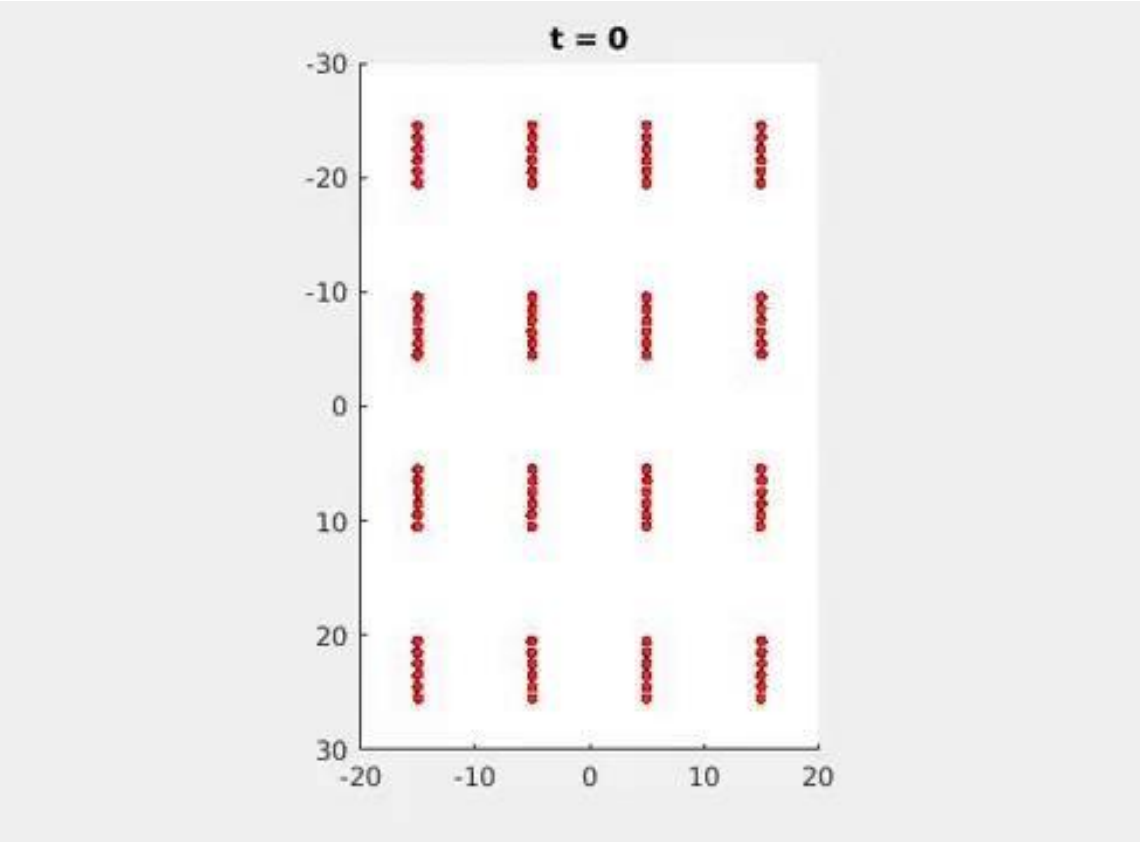
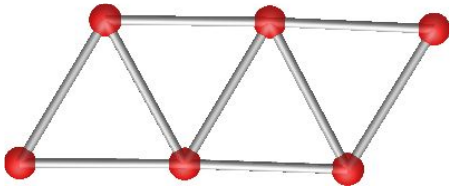
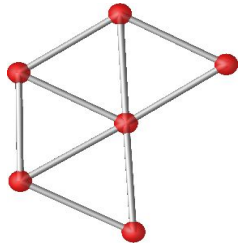
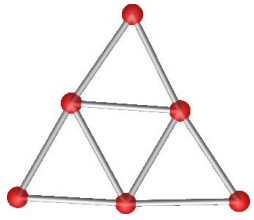
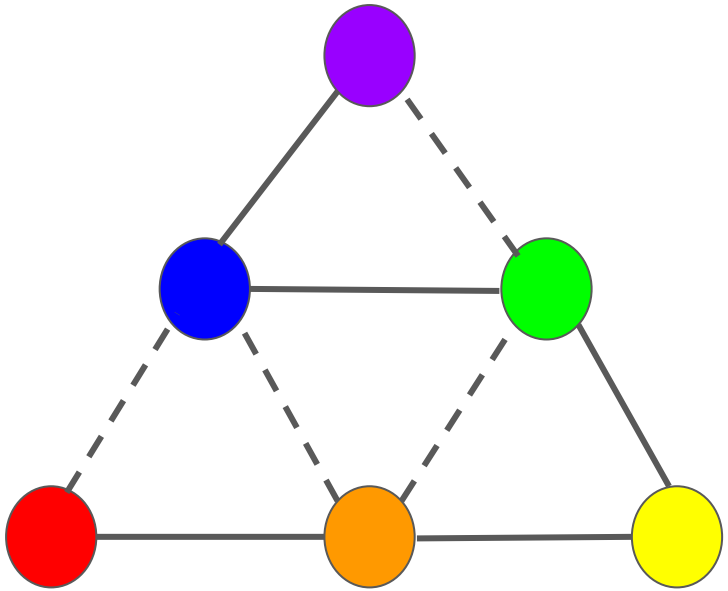


Table of Results - Ground State Yields

State	Simulation	10000 BD	Theory (Eq)
	10/16 (63%)	76%	49%
	5/16 (32%)	22%	40%
	1/16 (6%)	2%	11%

How Do We Make a Triangle?

By Designing the Interactions (Zorana Zeravcic, 2014)



Optimization methods exist to determine these interactions, **BUT** no existing method addresses the issue of kinetic accessibility!

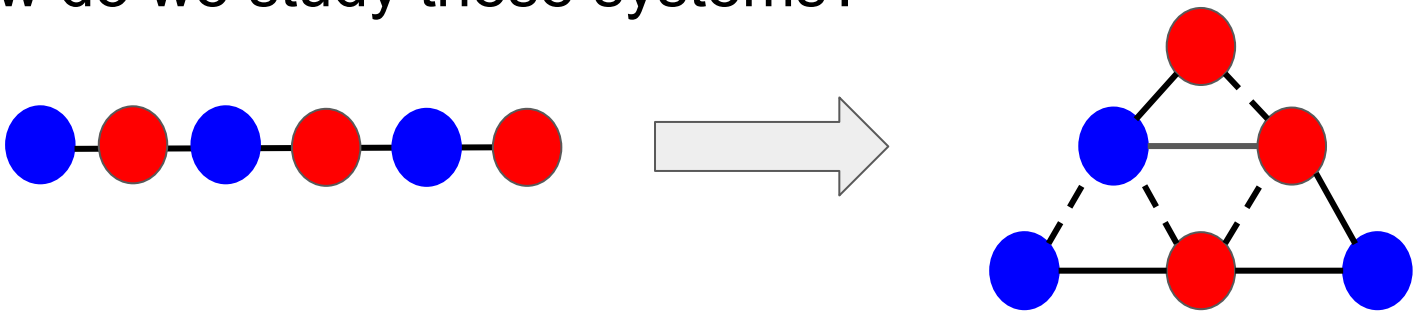
Having 6 particle types and 9 independent interactions is:

1. Expensive
2. Complicated
3. Not Scalable

How Do We Make a Triangle, Efficiently?

Question: What is the minimal modification to our system that allows us to design a structure (e.g. Triangle) that self-assembles with high yield and high rate?

Proposal: As a first step, let's try the simplest extension; 2 particle types, A and B, with 3 interactions. How do we study these systems?



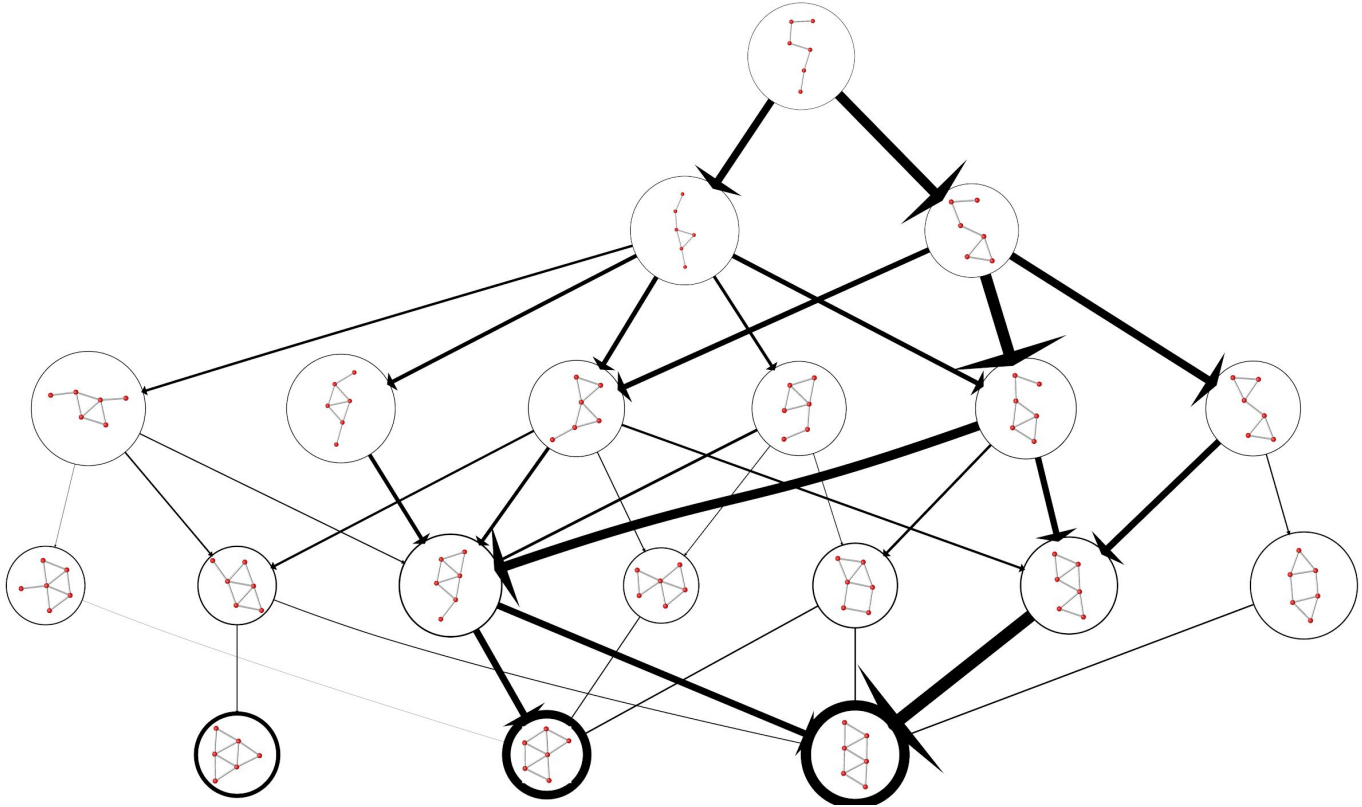
Outline

1. Discuss the model we've developed to study these folding problems.
    a. Should address both free energy and kinetic accessibility.

2. Use the model to address design questions posed.
    a. How do we form that pesky triangle?

3. How accurate/appropriate is the model?
    a. Compare predictions to experiment.

Continuous Time Markov Chain (CTMC) Model



3 Ingredients:

1. Forward Rates
2. Equilibrium Probabilities
3. Backward Rates

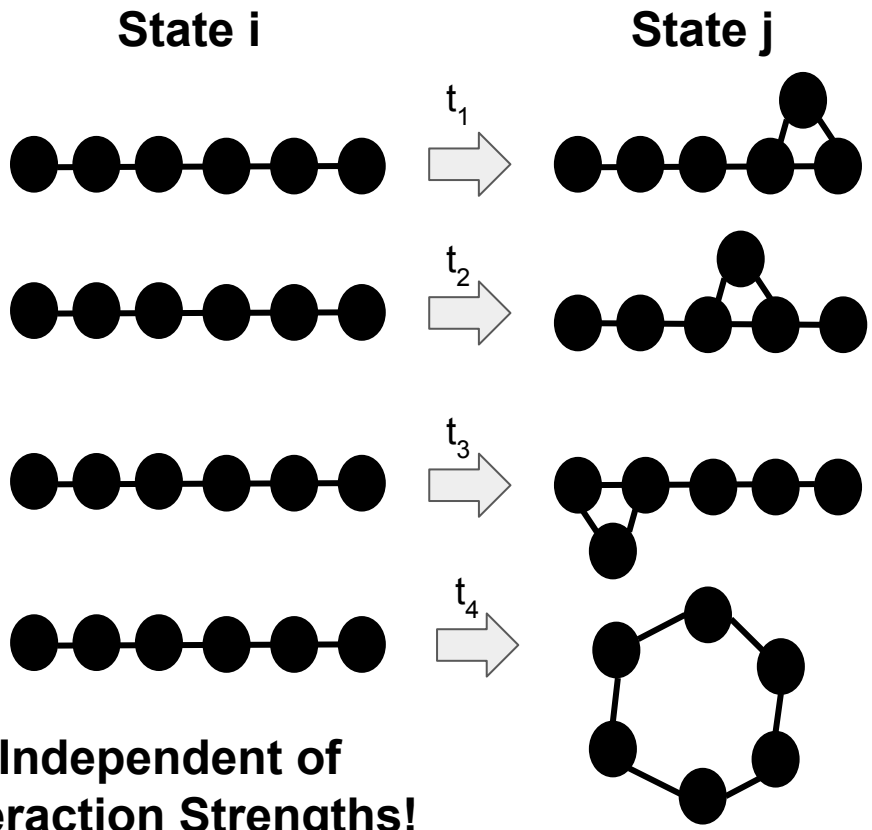
1. Forward Rate Estimation

- Use Reflecting Brownian Dynamics to sample the exit time out of state i, and keep counts of the number of times state j is visited.

Mean First Exit Time $\hat{\tau} = \frac{1}{N}\sum_{j=1}^{N} t_j$

Transition Probability $\hat{P}_{ij} = \frac{c_j}{N}$

Transition Rate $\hat{T}_{ij} = \frac{\hat{P}_{ij}}{\hat{\tau}}$



Independent of Interaction Strengths!

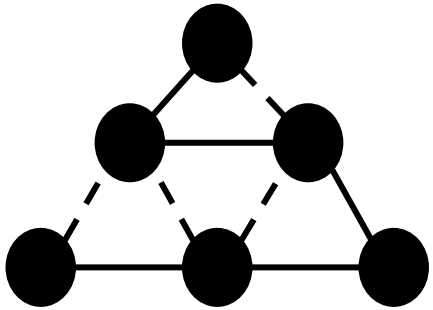
2. Equilibrium Probability Estimation

- Use Brownian Dynamics (or another sampler) to generate a very long trajectory.
  - Estimate the equilibrium probability by simply counting how much time is spent in each state.
- Depends on the choice of interaction strength, k.
  - Choosing smaller values explores the state space much quicker.
  - Can be quickly computed for other choices via a re-weighting scheme.

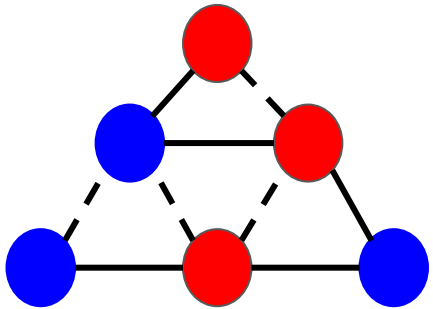
Movie: Miranda Holmes-Cerfon

Re-weighting Examples

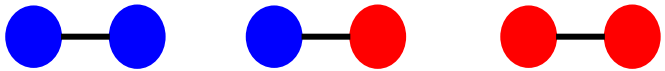
$$\kappa_1 = 2, \quad \pi_\Delta(\kappa_1) = 0.03 \qquad \pi_j(\kappa_2) = \pi_j(\kappa_1) * \left(\frac{\kappa_2}{\kappa_1}\right)^{b(j)}$$



$$\pi_\Delta(\kappa_2) = 0.03 * \left(\frac{\kappa_2}{2}\right)^4$$



$$\pi_\Delta(\vec{\kappa_2}) = 0.03 * \left(\frac{\kappa_{AA}}{2}\right)\left(\frac{\kappa_{AB}}{2}\right)\left(\frac{\kappa_{BB}}{2}\right)^2$$



3. Backwards Rates and Detailed Balance

- Estimating the rate of bond breakage is hard. What can we do instead?

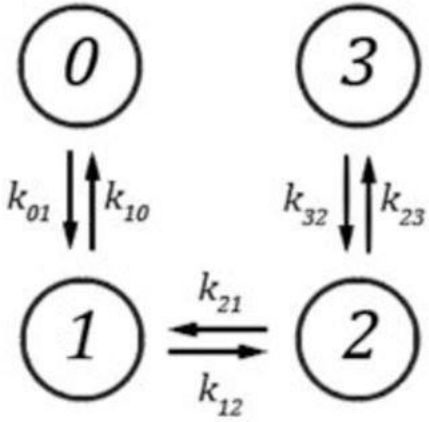
- Detailed Balance: In equilibrium, each process is in balance with its reverse process.

- Mathematically,

$$\pi_i T_{ij} = \pi_j T_{ji}$$

- Re-arranging,

$$T_{ji} = \frac{\pi_i}{\pi_j} T_{ij}$$



CTMC Model Outputs

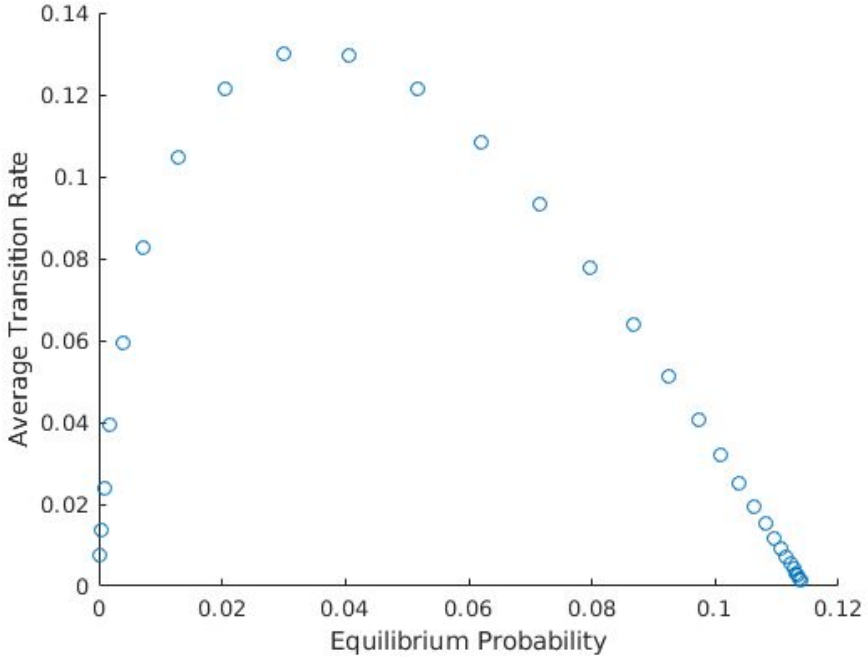
1. Equilibrium Probabilities for each ground state
   a. Must be estimated once for one parameter set.
   b. Re-weighting scheme allows fast computation for any parameter set.
2. Average transition rates to each ground state.
   a. Can be computed using the transition rate matrix, T, by solving a linear system.

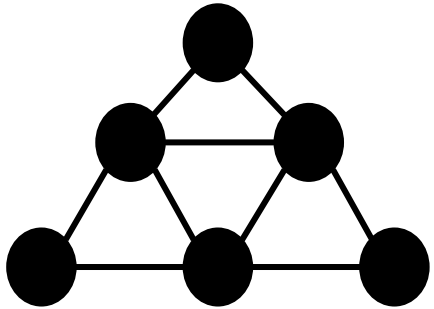
Results for k=2

State	Eq. Prob	Rate
Trapezoid	12.5%	0.54
Chevron	10.1%	0.42
Triangle	2.9%	0.09

$$T_G\vec{\tau} = -1, \quad \vec{\tau}(G) = 0$$

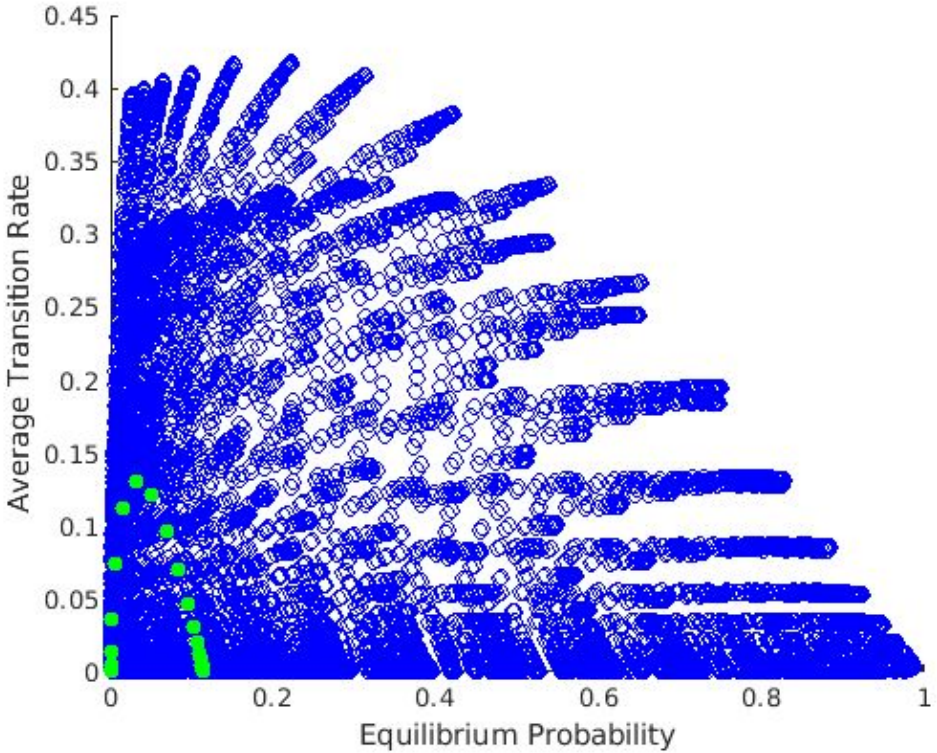
Model Output - Triangle State, Identical Interactions

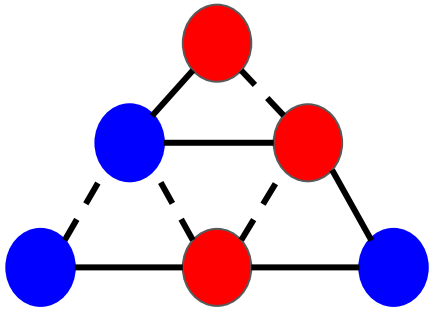




- Interaction strengths vary from 0 to infinity.
- Equally spaced samples in log space.

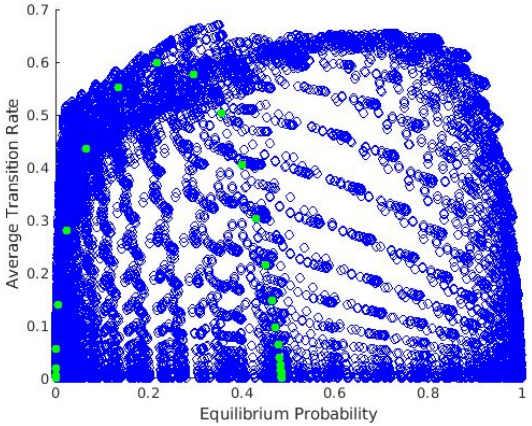
Model Output - Triangle State, Multiple Interactions

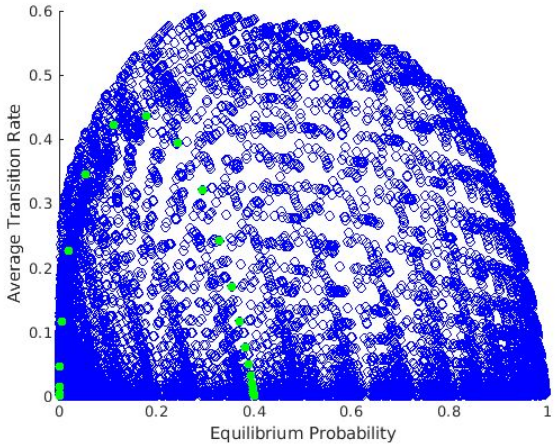


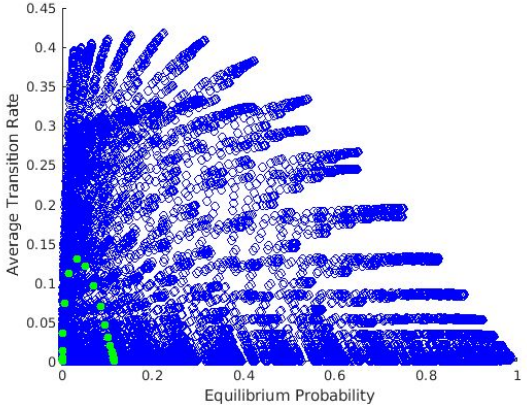


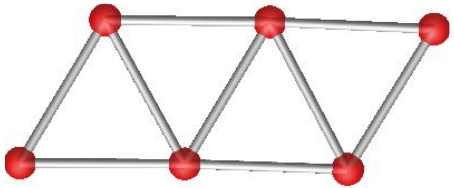
- 3 interaction strengths vary from 0 to infinity, independently.
- Equally spaced samples in log space.

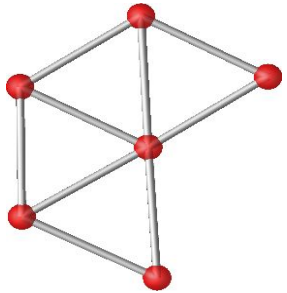
Scatter Diagrams - All Ground States

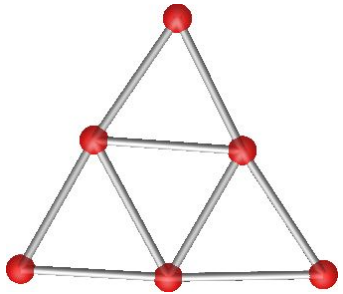




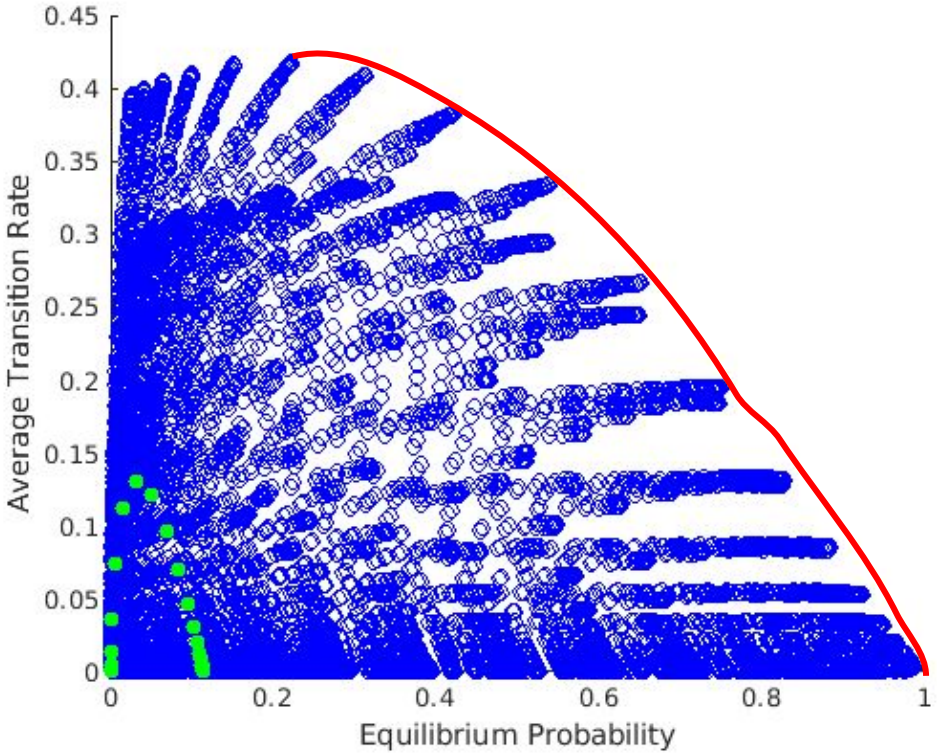








Finding “Optimal” Design Parameters

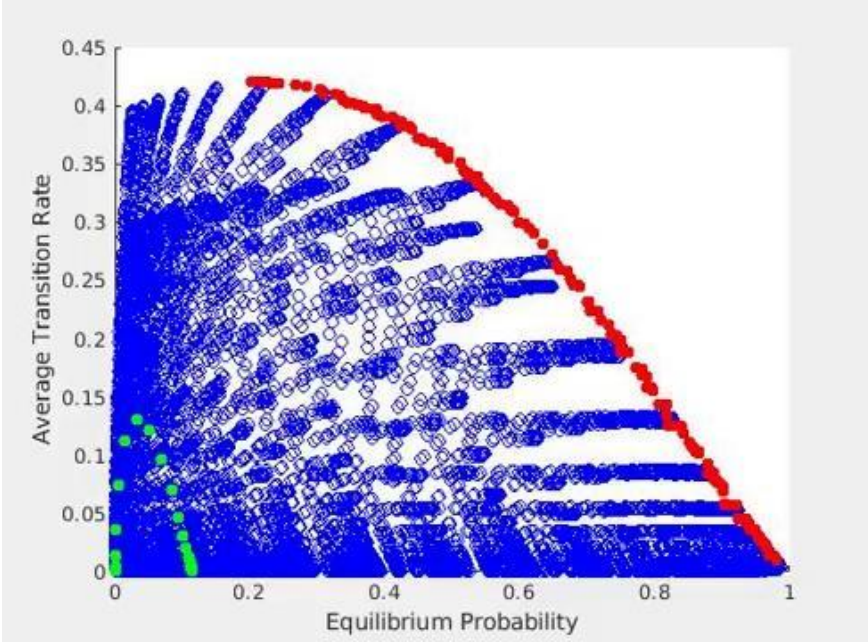


- Want to maximize both rate and probability.
  - Multi-objective optimization (MOO)
  - Solutions are not unique, unless each objective is given a weight.
- Pareto Optimality
  - No objective can be increased without decreasing another.
- How do we find the Pareto front?

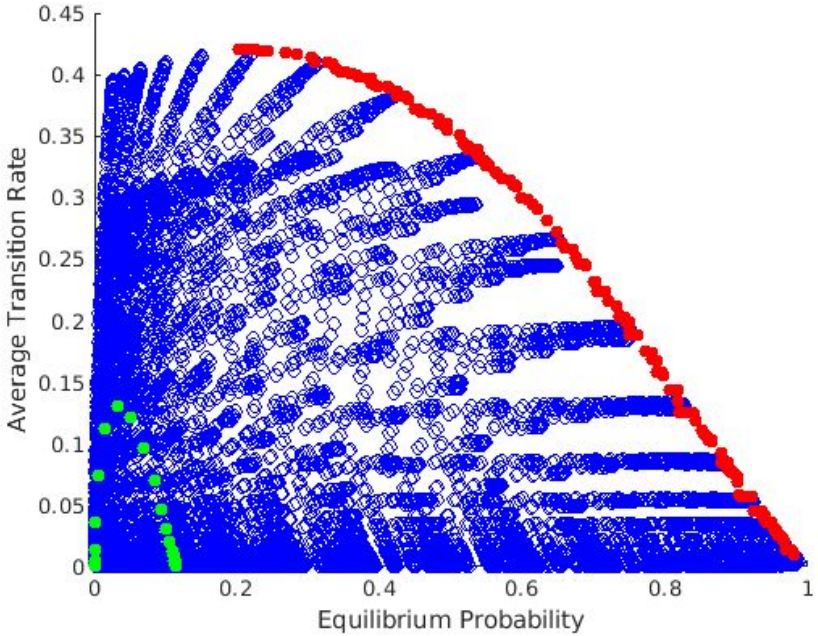
Genetic Algorithms for Multi-Objective Optimization

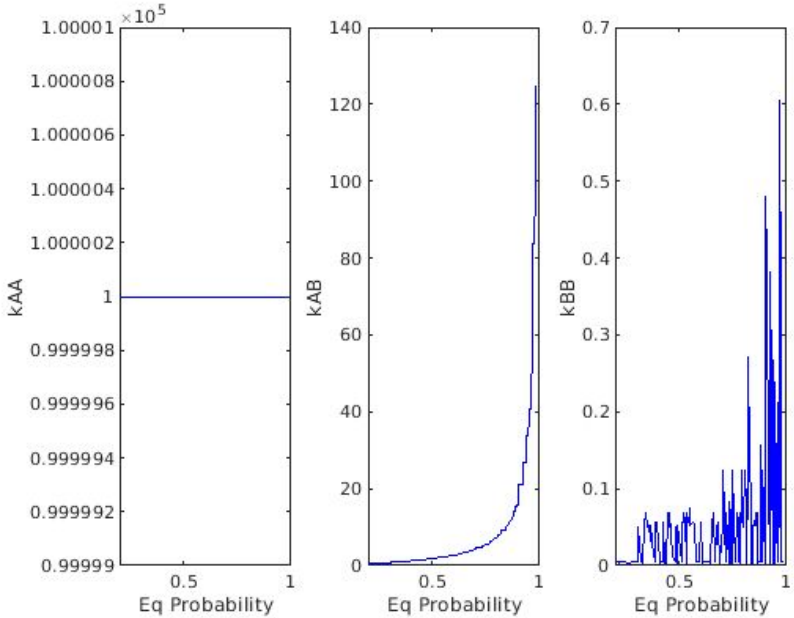
Survival of the fittest.

- Initialize a population with randomly sampled parameters.
- Sort population from best to worst.
- Perform natural selection.
  - Best 10% live
  - Best 50% mate, then die
    - Randomly pass parameters to children
    - Can mutate
- Repeat until converged. Evolution!



Extracted Parameters - Triangle State





Future Directions

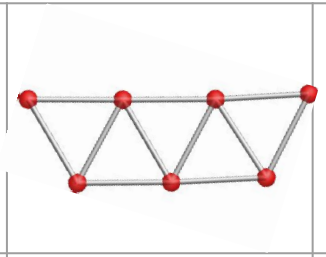
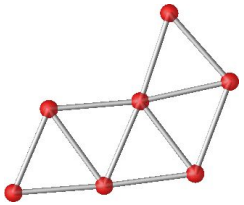
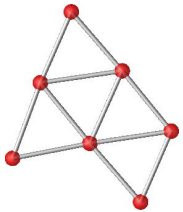
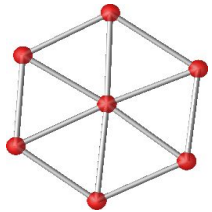
1. Combinatorial Optimization
    a. Is there a better ordering than ABABAB? Can we determine this efficiently?
    b. Mixed continuous/discrete multi-objective optimization?
2. Non-equilibrium folding
    a. Can we do even better by introducing a time dependent folding scheme?
        i. Temperature ramping/control
3. Extend to other systems
    a. Lattice Proteins (HP model)
    b. Full Protein

How Good is the Model?

Sources of Error:

1. Sampling Error
    a. Rates - Estimated error bars < 1% for all forward transitions.
    b. Equilibrium Probabilities - Estimated error bars < 3% in worst case.
2. Assumptions
    a. The continuous dynamics can be well-approximated by a Markov State Model.
        i. Does the Markov property hold? Exactly? Approximately? Not at all?
        ii. Work by Tony Lelièvre suggests the Quasi-Stationary Distribution (QSD) is the right object to construct the CTMC model with respect to.
            1. A future direction is to compare the QSD and stationary distribution.

Experimental Validation

State				
Brownian Dynamics Yield	61%	34%	3%	3%
Model Yield	40%	38%	12%	8%
Experimental Yield (Klein, Manoharan, 2019)	44%	42%	9%	5%

Our model is *significantly* closer to the experiment than BD simulation!

What’s going on?

- We are using BD to sample the transition rates for the CTMC model. Why does our model give different results than BD?
  - The only difference is, our model uses BD trajectories that are *equilibrated* to be in the stationary distribution.

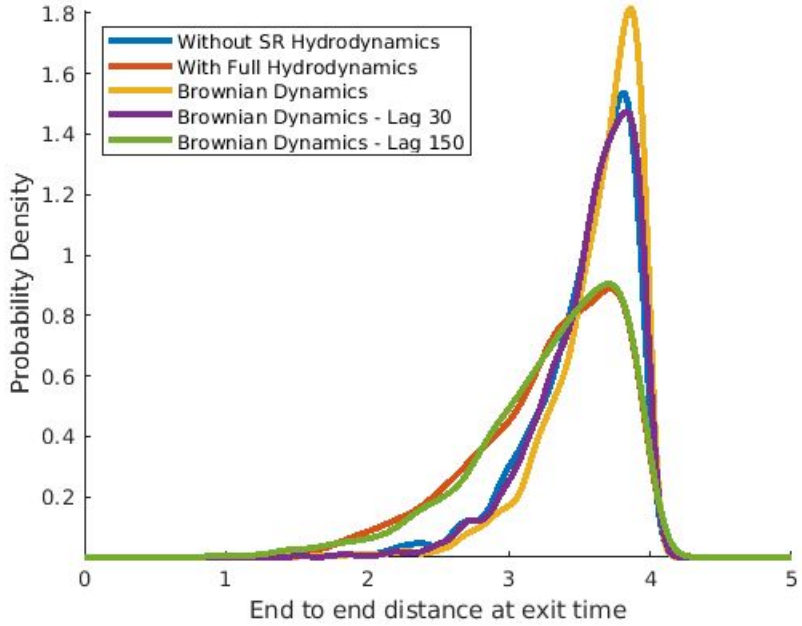
- Why does equilibration lead to a more accurate model?
  - Hypothesis: Equilibration is capturing the effect of hydrodynamics.

- Let’s test this with Hydrodynamics simulations
  - Rigid Multi-Blob Code from Aleks & Brennan

Hydrodynamics Simulations

- 2 types of run
  - Short and Long Range Hydrodynamics
  - Just Long Range Hydrodynamics

- Compare trajectories from HD simulations to Brownian Dynamics (with and without equilibration)
  - Ex: Distribution of an order parameter at the first hitting time.



Conclusion

- Constructed a model to simultaneously study both the free energy and kinetics of a self assembling system of colloids.

- Showed how the model could be used to come up with optimal design parameters.
  - Showed the existence of the “Pareto Front for self-assembly”

- Examined the validity of the model.
  - Hypothesized that equilibration in the CTMC model was acting to capture the effects of hydrodynamics.

Reweighting Scheme for Changing Interactions

Equilibrium Density is the Boltzmann Distribution

$$\rho(x;\kappa) = \frac{1}{Z} e^{-\beta H(x;\kappa)}$$

Integrate over configurations, x, consistent with state j.

$$\pi_j(\kappa) = \frac{1}{Z} \int_{x \in S_j} e^{-\beta H(x;\kappa)} dx$$

Do the *complicated* integral. (See Kallus, Holmes-Cerfon 2016)

$$\pi_j(\kappa) = \frac{1}{Z} c_j \kappa^{b(j)}$$

$$\pi_j(\kappa_2) = \frac{1}{Z_2} c_j \kappa_2^{b(j)} * \left(\frac{\kappa_1}{\kappa_1}\right)^{b(j)} = \frac{Z_1}{Z_2} \pi_j(\kappa_1) * \left(\frac{\kappa_2}{\kappa_1}\right)^{b(j)}$$