

# Designing an information-driven pathway approach for targeted self-assembly

Shannon Moran

February 21, 2018

This paper is submitted in partial fulfillment of the University of Michigan Chemical Engineering Department Preliminary Exam requirements.

## **Committee:**

Prof. Sharon Glotzer (Advisor)

Prof. Ronald Larson

Prof. Robert Ziff

Prof. Xiaoming Mao (Cognate: Physics)

## 1 Introduction

Self-assembly is a powerful tool for creating complex materials with tailored particle interactions. Systems order with interactions as simple as hard-particle excluded volume [1] and as complex as DNA-programmed origami [2]. While much research has been focused on understanding how to design highly-specific building blocks and predict assembled structure, relatively little work has been focused on optimizing specificity in self-assembly system design.

Put another way— **what is the minimal set of instructions needed to achieve targeted, self-assembled complexity?**

## 2 Background

Self-assembly is the process by which individual components arrange themselves into an ordered structure via free energy minimization [3]. This ordered structure is determined by the inter-particle interactions and assembly environment, and much recent work has focused on tuning these parameters to achieve target structures [4]. With advances in material synthesis techniques, complex building blocks can now be made with nano- and micro-precision. Correspondingly, there has been a focus in the literature to understand the role of building block properties on the bulk structure assembled.

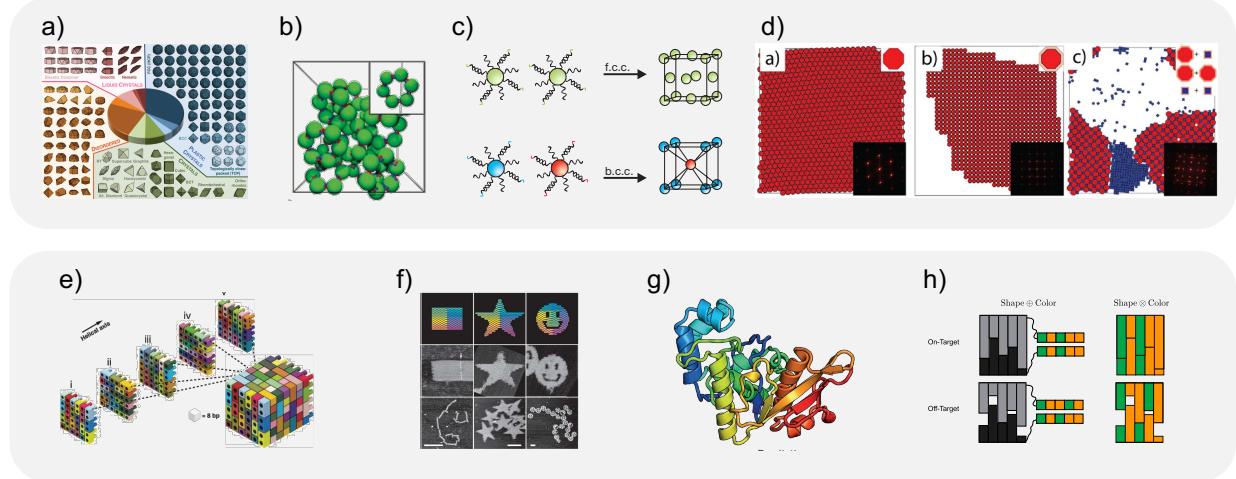
At the colloidal scale, seemingly simple additions of anisotropic interactions to building blocks can yield complex structures [5]. Hard particles with shape can assemble a rich variety of structures [1], even in the absence of explicit interactions between particles other than excluded volume. Later work described this as the effect of an effective entropic patchiness caused by shape interactions [6].

Even when specific interactions are added, they do not need to be complex for small changes to lead to very different structures. Adding explicit attractive patches to otherwise isotropic particles led to the formation of shapes, clusters, and rings, depending on the orientation of the patches [7]. Later, Millan *et al.* were able to access Archimedean tilings of regular polygons by incrementally adding attraction to guide assembly to target structures [8]. However, even non-specific directional attractive patches can cause particles (in one case, nanoplates forming a superlattice) to self-assemble structures differing from those accessible via entropic forces alone [9]. Experimentally, DNA-mediated particle interactions are now a powerful tool for tailoring crystal structure of a uniform building block [10].

The diversity of structures that can be accessed from simple interactions is impressive, and there is still much to be understood about how inter-particle interactions impact the structure assembled.

However, recent advances in synthesis of programmable DNA building blocks have now enabled significantly more complex interactions, resulting in self-assembly processes of addressable complexity. In addressable systems, each particle has a specific location (address) that it is “programmed” to assemble into.

Biologically, we are familiar with specifically-programmed assembly in DNA replication or protein folding [15]. Synthetically, though, the most well-known examples of addressable complexity are DNA-based. “One-pot” addressable self-assembly of DNA bricks uses the hybridization of complementary DNA sequences to construct addressable structures from a single pool of monomers [11]. Simulation results have shown that such one-pot self-assembly can succeed with highly simplified model subunits that lack the molecular details of DNA tiles, suggesting that similar design strategies should be possible in non-DNA-based synthetic systems [16].



**Figure 1: Self-assembly in the literature:** Top: Simple interactions between particles can lead to a rich variety of assembled structures. Systems shown: (a) ordered structures from hard polyhedra [1]; (b) isotropic particles with attractive patches [7]; (c) crystal structure tailored with DNA-mediated assembly [10]; (d) Archimedean tilings assembled from polygons with increasing bond specificity [8]. Bottom: Programmable interactions can enable addressable complexity. (e) DNA bricks assembled in a one-pot system with single-stranded DNA [11]; (f) DNA origami [12]; (g) de novo protein design [13]; (h) theoretical calculation of chemical and shape bond instruction capacity [14].

In addition to using programmed DNA interactions between particles, programmed self-assembly can also refer to design of DNA strands or protein primary structures themselves. DNA origami is perhaps the most well-known example of this [12, 2]. Given a target structure (such as the much-cited example of a smiley-face), a DNA strand can be programmed that will fold precisely into the designed target [12]. While some examples of protein design exist (e.g. [13]), it is possible that the additional challenge of designing structure from a set of 22 amino acid residues versus 4 nucleotide bases in DNA has limited its appeal as a designer system.

While much work has focused on designing interactions to tailor self-assembly, we must also note that simply having sufficiently-specific interactions does not guarantee a system will reach a target state. If assembly temperature is too high, bonds may be unable to stably form; if the assembly temperature is too low, the system may be unable to sufficiently sample configuration space to reach the target structure. Though the target maybe be the global free energy minimum, that does not mean that it can be reached kinetically. Temperatures and binding energies must be carefully tuned to enable systems to sufficiently sample possible configurations and escape kinetic traps while still enabling stability in the ground state (i.e. bonds do not spontaneously break). In a noteworthy example, the yield of the one-pot DNA self-assembly system studied in [11] was found via simulation to be highly sensitive to both temperature [16] and the annealing protocol used [17].

A collection of the self-assembly results discussed here and in the remainder of this proposal is shown in Figure 1.

## 3 Proposed Research

### Question to be addressed

While much research energy has been focused on understanding how simple interactions can lead to complex structures or how addressable structures can be programmed for self assembly, there is a middle ground with relatively sparse literature. How can we *optimize* the amount of self-assembly instruction given to a system to successfully assemble a target structure?

To assemble Archimedean tilings from regular polygons, Millan *et al.* incrementally applied attractive patches to particles until the target structure was assembled [8]. Their heuristics of increasing interaction specificity mirror the increasing specificity detailed above: hard interactions only, then symmetric patches, shape-specific patches, and edge-specific patches. The authors were able to demonstrate that a decision tree with just three nodes was sufficient to determine the level of instruction needed to reach a given target structure.

Instead of focusing on minimum instructions, work on protein folding has focused on identifying key intermediate states formed during the folding process [18]. Jacobs *et al.* demonstrated that indeed, there are preferred transition states along pathways that lead to successful folding. In such systems, while the final structure is fully defined by the primary structure of the protein, sampling all possible configurations to find the final state would take an impossibly long time (this is known as “Levinthal’s paradox,” and will be further discussed later in the proposal). The folding of a protein introduces geometric considerations in which local native bonds preferentially form to scaffold the folding [19].

Thus, while proteins and patchy colloidal assembly appear to be very disparate approaches to this topic, combined they suggest an intriguing idea. Could a minimally sufficient set of non-specific interactions be combined with a constrained geometry (in the form of folding) to allow for self-assembly of a target structure?

### Proposed model

I propose studying this question using nets of the five Platonic solids as described in [20]. These nets can be implemented as rigid bodies connected by harmonic springs, with non-specific attractive patches on the edges of each net “face.” Molecular dynamics (implemented in HOOMD-blue [21, 22]) will be used to simulate assembly.

In a forth-coming study from the Glotzer lab using this system [20], the authors find that for the same target shape, compact nets with more leaves assemble most reliably (in agreement with [23]). By investigating the assembly behavior for shapes with fewer possible net constructions (i.e. tetrahedron, cube, and octahedron), they are then able to use these features to predict which nets will assemble for shapes with a wide variety of possible net constructions (i.e. dodecahedron and icosahedron). Additionally, they observe that reliable assembly is due to the formation of local native (that is, present in the final structure) bonds early in the simulation which in turn bring non-local contacts together to be folded next (i.e. cooperatively) [19]. They posit that as each net folds, it freezes out the fewest possible number of degrees of freedom, thereby maximizing the conformational entropy along the folding pathways.

This finding closely mirrors what is already known in studies of protein folding. In 1968, Cyrus Levinthal posed the following thought experiment, come to be known as “Levinthal’s Paradox” [24]. Proteins are known to fold on the order of a second— however, if proteins explored all their possible degrees of

freedom they would never be able to assemble. A polypeptide with 100 residues would have 99 peptide bonds, and therefore 198 different phi and psi bond angles. If each of those bond angles could be in one of three stable conformations, the protein would potentially have to sample  $3^{198}$  different conformations to reach its native configuration. Randomly sampling such a large number of configurations in search of the native configuration is not even possible in the life-time of the universe— and thus, Levinthal’s Paradox.

Levinthal posited that biological folding times could be explained by the formation of local, native contacts which would stabilize the structure. Indeed, over 20 years later, these small biases (on the order of just a few  $kT$ ) towards the native configuration were mathematically demonstrated to result in biologically realistic protein folding times [25, 26]. In analyses of real-world folding times, protein folding speeds correlate with the topology of the native protein. Fast folders usually have mostly local structure, such as helices and tight turns, whereas slow folders usually have more non-local structure, such as  $\beta$ -pleated sheets [27]. Further work on elucidating experimental protein kinetics detected the partially folded intermediates transition states in the folding process posited by Levinthal [28].

Given that we are able to replicate complex findings from protein folding in this simple system of a folding net, this suggests that we can continue to use this system as a toy model for developing theory of tailored self-assembly. In proteins, Levinthal posited that some local contacts may act as “nucleation points in the folding process”. Though the interactions of the edges in the folding nets are not specific like those in proteins, the same critical local contacts are observed in the folding process.

This suggests that the combination of geometry and attraction serve to make some local bonds in folding nets more effective than others. In turn, it may be possible to identify bonds that are more critical to assembly than others. If we are able to rank the importance of edge bonds on the yield of an assembly pathway, then it stands to reason that we can determine which interactions are the minimally sufficient set needed to guide assembly into a target structure.

## Project 1: Define measure of pathway information

### Background

Previously, work by Jankowski *et al.* measure the flux through and relative energies of transition states en route to a target pattern formation [29]. In that study, they sought to enable pathway design by identifying key transition “cluster” building blocks that a system would need to go through for the pathway to end in successful assembly. Here, we propose using this same approach but instead to identify key bonds that must form en route to a successfully-folded net (see Figure 3).

Next, we need an approach to quantifying how correlated a given bond is with successful assembly.

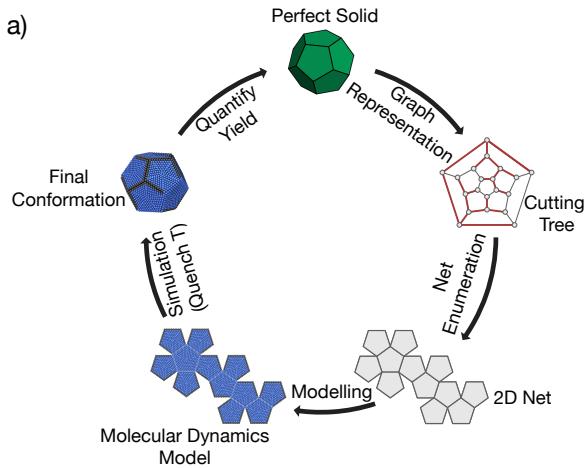


Figure 2: **Model, proposed work:** Process for developing nets from Platonic solids, from [20]

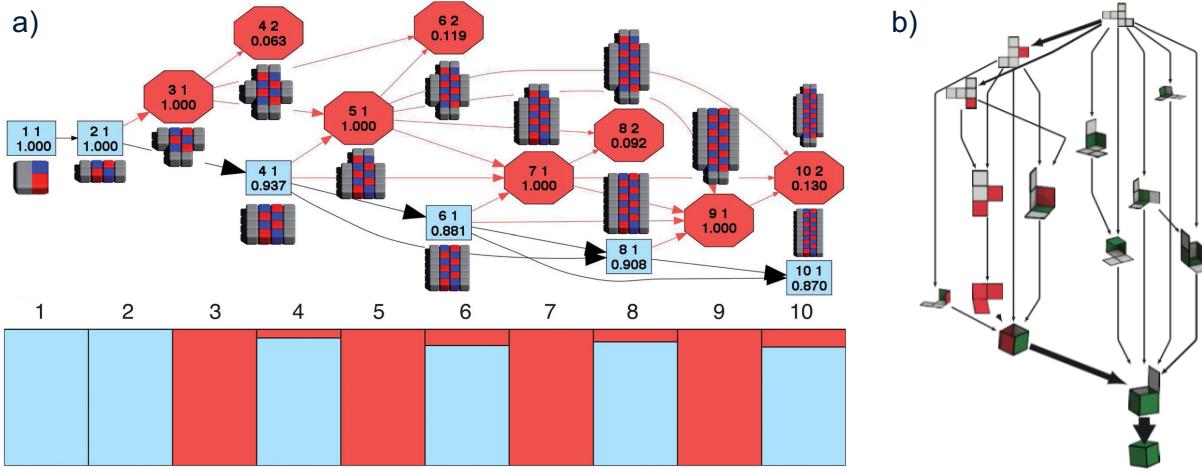


Figure 3: **Transition state analysis:** (a) Analysis of assembly pathways and relative populations of building blocks into target assembly patterns. [29]. Labels at each node indicate the size of a cluster, its energy level for its size (1 for lowest, 2 for second-lowest, etc.), and its probability compared to clusters of the same size. (b) Significant transition states and folding fluxes for net of a cube [20].

Huntley *et al.* have developed an analogous information-theory-driven approach for the specificity of lock and key bonds, which we can adapt for our uses [14].

For a system of  $N$  distinct locks ( $x_1, x_2, \dots, x_N \in X$ ) with unique partner keys ( $y_1, y_2, \dots, y_N \in Y$ ), we can calculate that binding between a lock  $x_i$  and key  $y_j$  will occur with probability  $p(x_i, y_j) = e^{-\beta E_{ij}}/Z$ , where  $Z$  is a normalization factor such that  $\sum_{i,j} p(x_i, y_j) = 1$  and  $E_{ij}$  is the binding energy.

In information theory, the mutual information quantifies how predictive the message received is of the signal that sent it. In the case of lock and key bonds, we can think about having a mutual information that is a measure of how predictive the identity of a lock  $x_i$  is of the identity of a key  $y_j$  bound to it. Specifically, we can write the mutual information  $I(X; Y)$  as:

$$I(X; Y) = \sum_{x_i \in X, y_j \in Y} p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (1)$$

where  $p(x_i)$  is the marginal distribution of  $x_i$ , representing the total probability of seeing  $x_i$  in a bound pair [and similarly  $p(y_j)$ ].

In the folding nets, instead of the state being a lock bonded with a key, we are instead measuring whether the presence of a bond is correlated to the net proceeding to folding (i.e. measuring the yield). That is, if bond  $x_i$  exists at a transition state, how predictive is this of the transition state proceeding to a successful folded final state?

#### Proposed work

First, our goal will be to define a metric that measures the likelihood of a bond being a part of a successful assembly pathway. The first approach to this will be taking pathways already enumerated in [20] and calculating their eventual yield in assembling the target structure, as done in [29]. With this correlation in hand, we can then calculate a first-pass mutual information measure between the existence of a bond in the structure and the assembled structure of the net.

This will provide an approximation of the ranking importance of the bonds formed in the assembly of a net. In case this approach is an over-simplification, we will need to find a way to account for cooperativity between bonds in the folding process. For example, one bond ( $x_1$ ) on its own may be a poor predictor that the structure will assemble, but when present with another bond  $x_2$  becomes highly predictive of successful assembly into the target structure. One key advantage of the nets system is that its possible configurations are countable. Thus, we could incrementally seed nets with combinations of edge bonds and measure their yield to fully characterize cooperativity between bonds in the folding pathway.

With this metric in hand, we can test whether “seeding” a poor-folding net with a critical bond will improve the folding yield. We will take a poor-folding net, seed it with a bonded edge we have estimated to be most critical to successful folding, and then allow system to proceed with folding. If our metric is good, we expect that we will be able to improve the yield through these bond additions. By incrementally adding critical bonds until a net folds, we will also be able to measure the minimum sufficient instruction needed for a given net to assemble at a given yield.

From this ratio of instruction (input) to measured yield (output), we can imagine calculating a type of “instruction efficiency”. Poor folding nets that require multiple bonds to be seeded to fold to a target structure would be said to be of low efficiency, while nets that folded the target structure with perfect yield even with no seeded bonds would be of high efficiency.

Ideally, we will be able to use these findings to predict which bonds will be critical in assembling a net we have not already characterized. Can we predict which bonds are going to be the most critical from heuristics we can develop based on mutual information between bonds and assembled structures in studied nets?

## Project 2: Develop energy landscapes for identifying kinetic barriers to assembly

### *Background:*

Seeding a structure with a critical bond, like those identified in Project 1, can help the assembly process avoid searching a local energetic minimum en route to the global minimum. In Project 1, we take a transition state flux-based probabilistic view to identify which bonds correlate with high yields. However, we also need to understand the underlying energetics to understand the mechanisms by which these bonds work. We would expect that critical bonds are those that prevent a configuration from exploring particularly deep (and thus energetically costly) local minima.

Disconnectivity graphs convert an energy landscape into a network of energy minima and transition states [30, 31, 32, 33]. This network reflects the connectivity between local energy basins. From disconnectivity graphs, we would expect to be able to identify large kinetic traps in the folding of these nets, and identify the location of transition states with bonds relative to those kinetic traps. Configurations that are likely to assemble reliably have few side-branches which have higher free energies than the target structure. Poor assembling configurations, on the other hand, have wide disconnectivity graphs, or deep local minima basins that are separated by an early transition state from the global minimum. Example disconnectivity graphs are shown in Figure 4.

### *Proposed work:*

First, we will develop disconnectivity graphs for this family of nets. David Wales’ group at the University of Cambridge has developed a code base that can be used to develop disconnectivity graphs, though

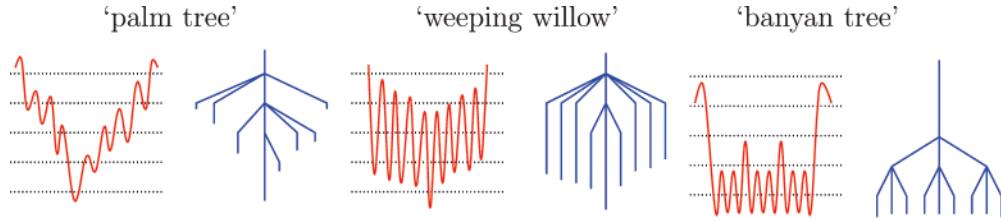


Figure 4: **Disconnectivity graphs:** One-dimensional potential energy functions (left) and the corresponding disconnectivity graphs (right) for representative energy landscapes [33]. The dotted lines indicate the energies at which analysis was performed to find local minima. Proteins are often characterized by the “palm tree” configuration and funneled energy landscape. For a net with a “willow tree” characteristic disconnectivity graph, we might expect that a seeding bond which is able to start the folding at the second major split from the top would lead to an increased folding yield.

applying it to new systems is non-trivial. A PhD graduate of Wales’ who was previously a post-doc in the Glotzer lab has agreed to assist with adapting this method to a folding system. With disconnectivity graphs in hand, we will then look to understand what they can tell us about (i) nets that fold well organically and (ii) the role of critical bonds in navigating assembly kinetics. Can we identify features of nets that correlate with “palm tree” style connective graphs, versus less assembly-friendly disconnectivity graph topologies?

First, we might expect that nets that fold well without assistance correspond to a narrower energetic funnel, like those seen in protein foldings. Identifying characteristics of nets that correspond with narrower energetic funnels would be a novel view of self-assembly in folding systems.

Second, we hypothesized that critical bonds would help the assembly avoid kinetic traps. We would expect these to be bonds that move the transition state from a node with multiple branches leading to locally deep minima, down to a node closer to the global minimum. We can also work in reverse. Given a disconnectivity graph, we can follow the transition states up the tree from the global minimum to find the most energetically favorable pathway from the starting net. As with nets with narrow energetic funnels, studying such pathways may give us further energetic insight into why some nets fold better than others.

While this proposed work seems straightforward, a variant this project has previously been attempted with a Glotzer lab member visiting the Wales group to develop disconnectivity graphs for our assembly work. The student found the code base to be fairly inflexible and difficult to apply to our systems. In this student’s judgement, it would almost be easier to re-write the disconnectivity graph generating code for our systems than adapt the version that exists for this folding problem.

Getting the analysis code working will be a non-trivial aspect of this work. However, having this tool on hand could have broad-reaching uses within our group for characterizing assembly pathways, which is an area of increasing interest among Glotzer lab PhD students.

### Project 3: Design pluripotent nets from minimal instructions

#### *Background*

Traditionally, the goal of self-assembly is to form one targeted structure from constituent components. However, systems which would self-assemble into different structures under different assembly conditions

could be of use in applications ranging from sensing to colloidal robotics. Such systems would be “pluripotent”, as they contain the instructions to form multiple different stable assemblies from the same starting materials.

Currently, the closest precursors to *colloidal* pluripotent materials are systems whose configuration changes when subjected to external fields. In one example, Han *et al.* developed hinged chains of electromagnetic blocks, whose hinges could be manipulated using a magnetic field to “fold” into different conformations, analogously to different secondary structures forming from an amino acid primary structure [34].

However, folding architectures have been widely explored as a *macroscopic* means of using a pluripotent starting material (often a sheet with seeded creases for folding) and accessing a variety of structures.

In studies of origami folding algorithms, Desmaine and colleagues developed a fold-finding algorithm to move between multiple folded configurations of a 4-box pleated sheet, while minimizing the number of actuators (folds), groups, or edges [35]. Trolley *et al.* demonstrate that such heat-sensitive seams could be used to make macro-scale self-folding origami that could be triggered by uniform heating, rather than targeted stimulus of the folds [36]. And in an impactful paper from early 2017, a robust example of algorithmically-designed 3D reconfigurable materials based on space-filling tessellations demonstrated the ability to move from one structure to another without undergoing an intermediate “unfold” step [37].

In Projects 1 & 2, we will have rigorously characterized methods of folding in a colloidal system. Here in Project 3, we will seek to apply these methods to creating colloidal-scale pluripotent materials.

#### *Proposed work*

We will start with a pluripotent folding precursor. The simplest system, based on Projects 1 & 2, would be to take nets of the Platonic solids and have two target states: one of the folded state, and the other a local minimum made the preferred state through the folding design process. For example, the lowest free energy configuration for some octahedra nets is actually an open boat-like structure which has higher rotational entropy than the assembled octahedron [20].

Unlike in the previous projects, we cannot use non-specific bonds throughout the structure and “seed” bonds along critical edges prior to letting assembly proceed. Instead, we will “seed” using specific attractive patches along the critical edges for one of the target structures. When we want to begin folding towards a target structure, we will “turn on” the attractive patches to drive those critical bonds to form.

Our previous work developing the information capacity of net/bond combinations will allow us to define the minimum set of bonds we will need to specify to fold each target state. In doing so, in specifying bonds for the first structure, we should ideally have un-specified bonds that can be used to form the second. Of course, it is possible that given the geometry of the nets, some edges will need to be seeded in both the target structures. For the sake of this theoretical demonstration, we can have three types of specific patches on the edges: (1) specific only with those patches needed in configuration 1, (2) specific only in configuration 2, and (3) specific in both.

For simplicity, assembly would transition between pluripotent states through the unfolded state. Depending on the success of elucidating kinetic barriers and transition states in Project 2, an additional challenge for this Project could be designing pathways that avoided unfolding completely, or expanding to more complex starting materials like the 4-box pleated sheet studied by [35].

Turning bonds “on/off” would experimentally still rely upon assembly conditions, though we can imagine this change in attraction happening in response to some change in solvent conditions (e.g. a sensing response). Developing a variant of the system that could be synthesized in lab would be a significant step forward in programmable colloidal self-assembly. Working with collaborators to test this model in the lab would very likely extend the reach of the theories developed here. (Note that this is not proposed as part of the thesis work, but would be a stretch-goal.)

## 4 Previous work

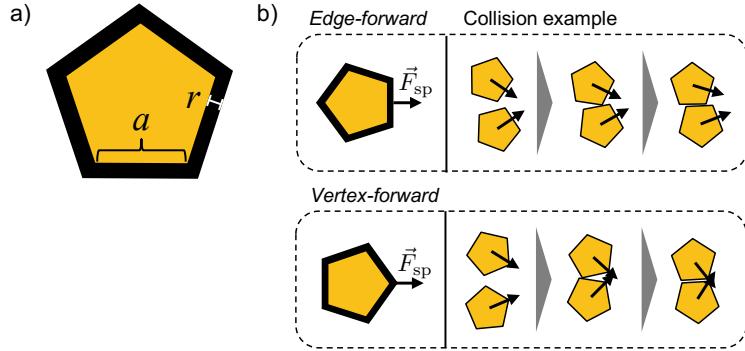
The research proposed above utilizes a different system from my work to date. While the proposed research looks at the interplay of folding geometry and non-specific attractions on equilibrium assembly, I previously studied the role of hard particle shape interactions in an actively driven system. Consistent in this research is an interest in understanding, and characterizing, the role of particle interactions on self-assembly. Below is excerpted from an in-preparation paper detailing the role of anisotropy on self-assembly in active systems [38].

### Background

Active matter has been a field of rapidly expanding interest and research activity over the last decade [39, 40, 41, 42]. Vicsek’s pioneering work showed that collections of point particles with alignment rules displayed rich collective behavior, including phase separation [43]. However, theoretical work seeking to describe the collective behavior of bacteria demonstrated that phase separation was not reliant upon explicit alignment rules [44]. Giant number fluctuations characteristic of phase separation in these systems of isotropic particles were found to lead to phase separation based on density-dependent slowing in a phenomena now known as “motility-induced phase separation” (MIPS) [45]. This phase separation in isotropic systems has been described as athermal phase separation [46], kinetic steady-state balancing of particle fluxes [47, 48], classical nucleation [49, 50], and balancing of collision theory timescales [51]. Importantly, this same phase separation predicted by theory has been observed in experiments, which confirm the activity-dependent formation of “active crystals” at low system densities [52, 53].

However, in real-world systems (e.g. bacteria) particles are rarely isotropic. Simulations of rods with varying aspect ratios have been shown to display a rich variety of collective motion dependent on shape and system density [54, 55]. Drawing from the example of anisotropic swimmers such as bull sperm and *Chlamydomonas*, Wensink *et al* observed that shape and direction of the translational driving force relative to the shape (referred to as “force offset” in this paper) allowed for differing modes of collective motion and onset of critical behavior [56]. Similarly, gear shaped “spinners” with differing directions of a rotational driving force were shown to phase separate through competing steric interactions (repulsive for opposite spinners, and attractive for like spinners) [57, 58, 59]. In a study of active dumbbells, Suma *et al* noted that particle anisotropy allowed for stabilization of cluster rotation, a phenomenon not seen in clusters of isotropic particles [60]. Finally, a study of active squares uncovered an “oscillatory” activity/density regime in which large clusters would break up and re-form at steady state [61].

Why is a description of the role of active particle anisotropy needed? We can view this spontaneous “phase separation” and organization as a type of non-equilibrium self-assembly [62]. It is well established that in the absence of attractive forces in equilibrium self-assembly, shape alone is sufficient to determine the minimum free-energy structure [63, 1, 64, 65]. While self-assembling, fluctuations allow a system to randomly sample system configurations until it finds the one with the minimum free energy. However,



**Figure 5: Model system:** (a) Shape anisotropy is studied with a family of regular polygons of side number  $n = 3 - 8$ , characterized by side length  $a$ . Particles interact through a purely repulsive WCA potential of radius  $r = 1$ . The dimensions of all shapes are set such that  $\frac{n \cdot a}{2\pi r} = 0.9$ . (b) Force anisotropy is implemented through the direction of the self-propelling force, which propels the shape either edge- or vertex-forward. A key feature of this system is that collisions of anisotropic particles can sustain dimer (and larger n-mer) translational and/or rotational motion. Illustrative collisions are provided for each force director.

sometimes the global free energy minimum is kinetically difficult to reach, and the system can become kinetically trapped in a metastable state. Activity provides a driving force which can help anneal a system out of these kinetic traps [66, 67] or stabilize system configurations that would be unstable without the additional driving force [68]. Understanding how particle anisotropy combined with an active force director will impact the collective motion, even in the form of a general heuristic, would open the doors to studying non-equilibrium self-assembly tailored through particle anisotropy.

To date, the differing behavior between active disks and active shapes has been explained in system-specific terms. Here, we study a system of active shapes to systematically understand and describe the role both shape and force anisotropy play in the collective behavior of translationally-driven active systems using the model shown in Fig. 4. Shapes are implemented using the discrete element method [69] implemented in HOOMD-blue [21, 22]. We use Langevin dynamics to model the movement of the particles studied here, and take care to select a sufficiently small mass that the system is effectively Brownian, in line with the expected dynamics of bacteria. Full simulation parameters can be found in the Methods section of the in-preparation paper.

## Results

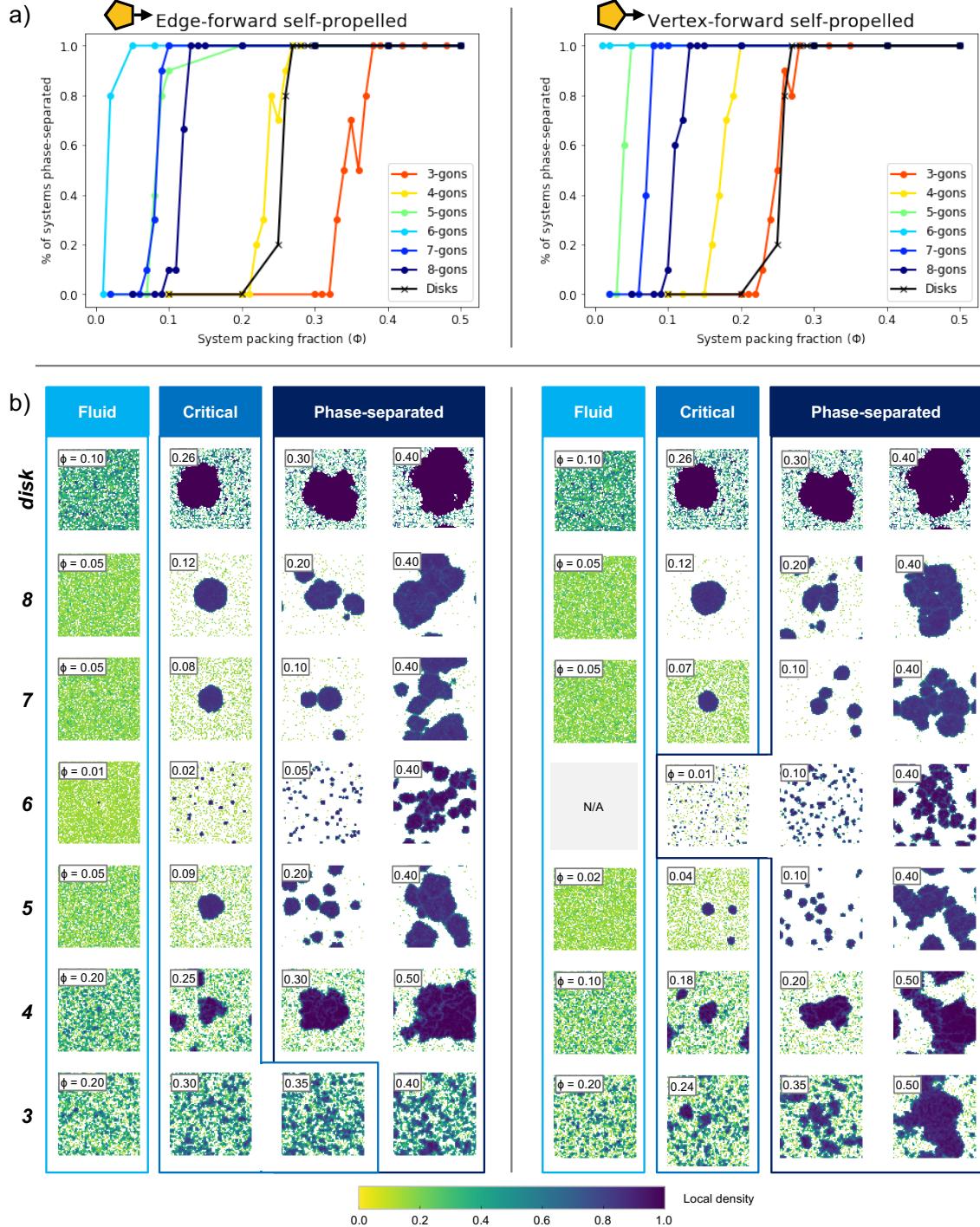
### Critical density

We characterize phase separation in terms of a “critical system density” at which the system separates into a sparse gas phase and dense clusters (based on a local density calculation detailed in the Methods section of the in-preparation paper). We would expect that particle anisotropy should depress critical density relative to that seen in disks [61]. If we hypothesize that clustering follows the same mechanism as an equilibrium phase transition, we would expect the nature of this “transition” to mirror that of these shapes in equilibrium. Specifically, work by Anderson *et al.* suggested that shapes beyond 7-gons display a phase transition from fluid to hexatic to solid analogous to the transition seen in disks. Naively,

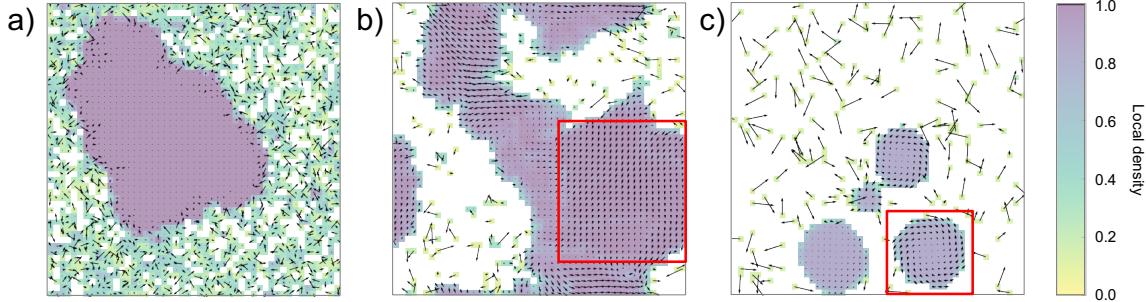
we might then expect that the critical density will decrease with decreasing  $n$ , and would expect phase separation behavior for  $n \geq 7$  to be indistinguishable from disks.

As shown in Figure 6A, we do observe a change in critical density relative to that seen in disks. However, critical density does not correlate linearly with the number of particle sides, and is not uniformly depressed by the introduction of shape. Instead, we see 6-gons phase separate at extremely low packing fractions (even at  $\Phi = 0.01$  in vertex-forward models). Next, 5-, 7-, and 8-gons phase separate at moderately low packing fractions. Finally, 3- and 4-gons phase separate at similar or even higher critical densities than those of disks.

At first glance, this trend appears to correlate with (1) the free volume in the cluster and (2) the ability of the shape's densest packing to stabilize shear. Particles in the cluster attempt to pack into their shape's densest packing [70] (results not shown in Prelim Report). For 6-gons, this results in a honeycomb lattice, while 5-, 7-, and 8-gons assemble into lattices with interstitial free space. While 3- and 4-gons can completely tile space, their densest packings are not able to stabilize shear forces in the cluster. Unlike clusters of disks, which cannot sustain translational or rotational motion, clusters of shapes can maintain momentum. This directional movement enables both stabilized grains and shear stresses between grains not seen in disks. Consequently, clusters of 3- and 4-gon are fundamentally less stable than their higher- $n$  counter parts. This explanation also explains Prymidis *et al*'s discovery of an "oscillatory" regime in their phase diagram [61]. We would expect to see such a regime in any shape whose densest packing is prone to shear planes.



**Figure 6: Critical density and nucleation behavior:** (a) Ten replicates are studied at each  $n$ -gon and system packing fraction, and the fraction of replicates that phase-separate at steady state is determined as described in the Methods. We see that the transition to phase separation is an abrupt one versus system packing fraction. (b) Representative snapshots for the fluid ( $\leq 0.10$  of systems phase separated), critical ( $> 0.1$  and  $< 0.9$ ), and phase-separated ( $\geq 0.9$ ) regimes. A distinctive feature of phase separation in systems of anisotropic particles is the formation of multiple stable clusters. Large clusters at high packing fractions are formed by the collision of these multiple clusters.



**Figure 7: Cluster displacement:** Shown are the particle displacement fields for simulations at steady state, laid over a map of local densities. (a) Clusters of disks have no net motion, with particle motion limited to the cluster boundaries and gas phase. (Shown is a system of disks at  $\Phi = 0.3$ ). In contrast, clusters of shapes are able to convert particle translational forces to net motion (red insets). Such clusters display both (b) net translational motion (shown for  $n = 4$ ,  $\Phi = 0.5$ , vertex-forward) and (c) net rotational motion (shown for  $n = 7$ ,  $\Phi = 0.1$ , edge-forward).

## Collective behavior

Prior studies of driven shapes have observed that clusters are able to sustain rotational [60] and translational motion. We also observe this behavior, as seen in Figure 7. As shown in Fig 7A, clusters of disks do not have rotational or translational motion. The only net motion within the cluster is at the boundaries, where a balance of particle fluxes in/out characterize the steady state configuration [48].

However, to the best of our knowledge, other studies have not observed a key consequence of this ability: the formation of multiple small, stable clusters as precursors to bulk phase separation. As seen in Fig. 6, phase separation in 5-, 7-, and 8-gons is characterized by the formation of one or few clusters which nucleate and grow. In 6-gons, cluster nucleation is so favorable that we see the nucleation of many small clusters even in the critical regime. For 3- and 4-gons, this picture is less clear but still points towards the formation of multiple small clusters as pre-cursors to phase separation.

In viewing videos of the evolutions of these systems (available during Preliminary Oral Exam), we see that at intermediate densities, these small clusters are stable. This can also be seen in Figure 7C, in which clusters are spinning in place. At higher densities in systems of shapes, multiple small clusters form and translationally collide with one another to create larger clusters; evidence of these collisions can be seen in grain boundaries in large clusters at steady state.

In contrast, systems of isotropic particles cannot have multiple clusters collide, as clusters cannot sustain the motion required for a collision. Instead, while multiple clusters may nucleate, a final system cluster is caused by either (1) smaller clusters dissolving and one large cluster dominating or (2) multiple smaller clusters merging into one another as they grow. A steady state of multiple small clusters would be highly unlikely in a system of isotropic particles, as clusters in such systems are only stabilized by particles being self-propelled into the cluster.

This suggests that particle anisotropy allows these systems to pursue a different mechanism of nucleation. Importantly, this mechanism of cluster formation does not appear to be achievable in disks. Redner *et al.* investigated the addition of attraction to active isotropic particles at a constant packing fraction of  $\Phi = 0.4$  [47]. At high activity and low attraction, that study replicated the formation of one large

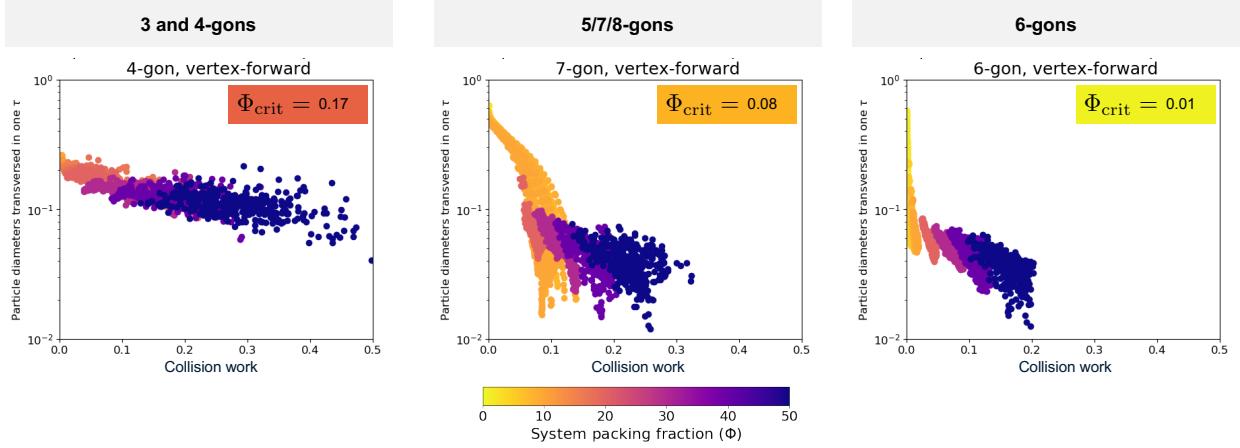


Figure 8: **Collision efficiency:** Shown are the evolution of particle velocity versus work done by inter-particle collisions in a system over the course of simulation. Each system packing fraction is composed of ten replicates, with data snapshots taken approximately every  $125\tau$ . Statepoints shown are representative of behavior in each group of  $n$ -gons.

cluster seen in the literature. At lower activities and high attraction, they were able to access a gel state; however, at no point in the activity/attraction phase space were multiple, clearly-separated stable clusters observed.

### How anisotropy impacts the collective behavior

While it is clear that shape and force anisotropy work together to impact the critical density and nucleation behavior, an explanation from existing theories is less clear. A kinetics-based theory developed in [50] treats cluster formation as a balance of in/out fluxes of particles at the boundary of a cluster. This balance relies upon the assumption that particles can escape clusters when the direction of their active force diffuses above the horizon of the cluster (either solo or as part of a multi-particle escape event). However, such diffusion is significantly sterically hindered in clusters of shapes. While these assumptions work well in systems of disks where they correctly predict the formation of a large cluster which nucleates and grows in cases of phase separation, it is not clear how to tailor this theory to the behavior of systems of anisotropic particles.

Alternatively, we can take a collision-theory based approach as described in [51], which treats phase separation as the result when the average collision time ( $\tau_C$ ) exceeds that of the average particle's free time between collisions ( $\tau_F$ ). Calculating the  $\tau_F$  of the particles and assuming a constant  $\tau_C$  for all shapes, we would expect to find 3-gons with the lowest critical density, with 8-gons the highest, not in line with our simulation results. This confirms our intuition that the mechanism by which shapes depress the critical density is by increasing the collision time.

We posit that this increased collision time is the result of “more effective” collisions between some shapes versus others. In the context of MIPS, we can say that “more effective” collisions are those that translate inter-particle forces—collisions—to greater decreases in particle velocity. That is, for a given collision, collisions between some particle shapes will lead to a steeper decreases in particle velocity than others. This also re-affirms our decision to study emergent behavior versus system density, a good proxy for the likelihood of inter-particle collisions that are required for phase separation to occur.

We can measure the net work done by all inter-particle collisions in the system as  $W = \frac{1}{2} \sum_i \sum_{j \neq i} \vec{F}_{ij} \cdot \vec{r}_{ij}$ , where where  $\vec{F}_{ij}$  and  $\vec{r}_{ij}$  are the pairwise forces and distance between particles, respectively. To study the efficiency of collisions for different shapes, we plot velocity versus the system's "collision work" in Figure 8. The velocity is shown as a percentage of the terminal velocity a particle would reach if it experienced no collisions.

We see that the evolution of velocity versus work done by collision follows very different patterns for the three groups of shapes. Velocity quickly decreases with very little work for 6-gons, while velocities of 3 and 4-gon systems change little with collisions. As might be expected from earlier results, 5/7/8-gons fall somewhere in between.

As presented, this data shows the time evolution for systems at each density, from start of the simulation (higher velocity) through steady state (lower velocity). Next steps will be to probe the relaxation time from this data, which we expect to be shape-dependent given the obvious difference in velocity/work relationships in the critical regime for each group of shapes. We are hopeful this will provide a definitive explanation for the role of shape in these systems.

## 5 Conclusions and potential impact

The proposed research seeks to fill a gap in the self-assembly literature: what is the minimal amount of instruction we need to give a system for it to assemble a target structure? Recent work has shown that the interaction specificity seen in DNA-mediated addressable assembly is not specific to DNA systems [16], and that synthetic systems can replicate interaction specificity achieved by nature [14]. While addressable assembly from a synthetic DNA analog has not been shown yet, an experimental realization is not far along the research horizon. However, a core question that will need to be answered is: how much instruction will this system need to provide to successfully assemble its target structure?

The work in this PhD research proposal seeks to provide an approach and metric to answer that question. See Figure 9 for key tasks and milestones.

Project	Status	Description	2018				2019				2020	
			Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
Project 0 <i>Active shapes</i>	100%	Data collection										
	90%	Analysis										
	75%	Writing										
	0%	Submit, revisions										
Project 1 <i>Information metrics</i>	0%	Data collection										
	0%	Analysis										
	0%	Writing										
	0%	Submit, revisions										
Project 2 <i>Energy landscapes</i>	0%	Develop code										
	0%	Data collection										
	0%	Analysis										
	0%	Writing										
Project 3 <i>Pluripotent materials</i>	0%	Submit, revisions										
	0%	Data collection										
	0%	Analysis										
	0%	Writing										
Thesis	0%	Submit, revisions										
	0%	Data meeting										
	0%	Defense										

Figure 9: **Timeline:** Key milestones and tasks from Preliminary Exam through target defense date.

## References

- [1] Pablo F. Damasceno, Michael Engel, and Sharon C. Glotzer. Predictive self-assembly of polyhedra into complex structures. *Science*, 337:453–457, 2012.
- [2] Erik Winfree, Furong Liu, Lisa A. Wenzler, and Nadrian C. Seeman. Design and self-assembly of two-dimensional dna crystals. *Nature*, 394(539-544), 1998.
- [3] George M. Whitesides and Bartosz Grzybowski. Self-assembly at all scales. *Science*, 295(5564):2418–2421, 2002.
- [4] Michael A. Boles, Michael Engel, and Dmitri V. Talapin. Self-assembly of colloidal nanocrystals: From intricate structures to functional materials. *Chemical Reviews*, 116(11220-11289), 2016.
- [5] Sharon C. Glotzer and Michael J. Solomon. Anisotropy of building blocks and their assembly into complex structures. *Nature*, 6:557–562, 2007.
- [6] Greg van Anders, N. Khalid Ahmed, Ross Smith, Michael Engel, and Sharon C. Glotzer. Entropically patchy particles: Engineering valence through shape entropy. *ACS Nano*, 8(1):931–940, 2014.
- [7] Zheni Zhang and Sharon C. Glotzer. Self-assembly of patchy particles. *Nano Letters*, 4(8):1407–1413, 2004.
- [8] Jaime A. Millan, Daniel Ortiz, Greg van Anders, and Sharon C. Glotzer. Self-assembly of archimedean tilings with enthalpically and entropically patchy polygon. *ACS Nano*, 8(3):2918–2928, 2014.
- [9] Xingchen Ye, Jun Chen, Michael Engel, Jaime A. Millan, Wenbin Li, Quozhong Xing, Joshua E. Collins, Cherie R. Kagan, Ju Li, Sharon C. Glotzer, and Christopher B. Murray. Competition of shape and interaction patchiness for self-assembling nanoplates. *Nature Chemistry*, 5:466–473, 2013.
- [10] Sung Yong Park, Abigail K. R. Lytton-Jean, Byeongdu Lee, Steven Weigand, George C. Schatz, and Chad A. Mirkin. Dna-programmable nanoparticle crystallization. *Nature*, 451:553–556, 2008.
- [11] Yonggang Ke, Luvena L. Ong, William M. Shih, and Peng Yin. Three-dimensional structures self-assembled from dna bricks. *Science*, 338(6111):1177–1183, 2012.
- [12] Paul W. K. Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440(16):297–302, 2008.
- [13] Po-Ssu Huang, Scott E. Boyken, and David Baker. The coming of age of *de novo* protein design. *Nature*, 537:320–327, 2016.
- [14] Miriam H. Huntley, Arvind Murugan, and Michael P. Brenner. Information capacity of specific interactions. *Proceedings of the National Academy of Sciences*, 113(21):5841–5846, May 2016.
- [15] Ken A. Dill. Folding proteins: finding a needle in a haystack. *Current Opinion in Structural Biology*, 3:99–103, 1993.
- [16] Aleks Reinhardt and Daan Frenkel. Numerical evidence for nucleated self-assembly of dna brick structures. *Physical Review Letters*, 112(23), Jun 2014.
- [17] William M. Jacobs, Aleks Reinhardt, and Daan Frenkel. Rational design of self-assembly pathways for complex multicomponent structures. *Proceedings of the National Academy of Sciences*, 112(20):6313–6318, May 2015.

- [18] William M. Jacobs and Eugene I. Shakhnovich. Structure-based prediction of protein-folding transition paths. *Biophysical Journal*, 111:925–936, 2016.
- [19] Ken A. Dill, Klaus M. Fiebig, and Sun Hue Chan. Cooperativity in protein-folding kinetics. *Proceedings of the National Academy of Sciences*, 90:1942–1946, 1993.
- [20] Paul M. Dodd, Pablo F. Damasceno, and Sharon C. Glotzer. Polyhedra nets fold cooperatively at high temperature. *unpublished*, 2018.
- [21] Joshua A. Anderson, Chris D. Lorenz, and A. Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics*, 227(10):5342–5359, May 2008.
- [22] Jens Glaser, Trung Dac Nguyen, Joshua A. Anderson, Pak Lui, Filippo Spiga, Jaime A. Millan, David C. Morse, and Sharon C. Glotzer. Strong scaling of general-purpose molecular dynamics simulations on gpus. *Computer Physics Communications*, 192:97–107, 2015.
- [23] Anum Azam, Timothy G. Leong, Aasiyeh M. Zarafshar, and David H. Gracias. Compactness determines the success of cube and octahedron self-assembly. *PLoS One*, 4(2):e4451, 2009.
- [24] Cyrus Levinthal. How to fold graciously. *University of Illinois Bulletin: Mössbauern Spectroscopy in Biological Systems Proceedings*, 67(41):22–24, 1969.
- [25] Robert Zwanzig, Attila Szabo, and Biman Baghi. Levinthal’s paradox. *Proceedings of the National Academy of Sciences*, 89:20–22, 1992.
- [26] Leandro Martínez. Introducing the levinthal’s protein folding paradox and its solution. *Journal of Chemical Education*, 91:1918–1923, 2014.
- [27] Kevin W. Plaxco, Kim T. Simons, and David Baker. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*, 277:985–994, 1998.
- [28] Ken A. Dill, S Banu Ozkan, Thomas R. Weikl, John D. Chodera, and Vincent A. Voelz. The protein folding problem: when will it be solved? *Current Opinion in Structural Biology*, 17:342–346, 2007.
- [29] Eric Jankowski and Sharon C. Glotzer. Screening and designing patchy particles for optimized self-assembly propensity through assembly pathway engineering. *Soft Matter*, 8(2852), 2012.
- [30] Lewis C. Smeeton, Mark T. Oakley, and Roy L. Johnston. Visualizing energy landscapes with metric connectivity graphs. *Journal of Computational Chemistry*, 35:1481–1490, 2014.
- [31] Oren M. Becker and Martin Karplus. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *The Journal of Chemical Physics*, 106(1495), 1997.
- [32] David J. Wales. Atomic clusters with addressable complexity. *The Journal of Chemical Physics*, 146(054306), 2017.
- [33] David J. Wales and Tetyana V. Bogdan. Potential energy and free energy landscapes. *Journal of Physical Chemistry B*, 110:20765–20776, 2006.
- [34] Koohee Han, C. Wyatt Shield IV, Nidhi M. Diwakar, Buvnesh Bhari, Gabriel P. López, and Orlin D. Velev. Sequence-encoded colloidal origami and microbot assemblies from patchy magnetic cubes. *Science Advances*, 3(e1701108), 2017.

- [35] Byoungkwon An, Nadia Benbernou, Erik D. Demaine, and Daniela Rus. Planning to fold multiple objects from a single self-folding sheet. *Robotica*, 29(01):87–102, Jan 2011.
- [36] Michael T. Tolley, Samuel M. Felton, Shuhei Miyashita, Daniel Aukes, Daniela Rus, and Robert J Wood. Self-folding origami: shape memory composites activated by uniform heating. *Smarter Materials and Structures*, 23(094006), 2014.
- [37] Johannes T. B. Overvelde, James C. Weaver, Chuck Hoberman, and Katia Bertoldi. Rational design of reconfigurable prismatic architected materials. *Nature*, 541:347–352, 2017.
- [38] Shannon E. Moran, Isaac R. Bruss, and Sharon C. Glotzer. Leveraging anisotropy for tailored self-assembly of active matter. *in preparation*, 2018.
- [39] Sriram Ramaswamy. The mechanics and statistics of active matter. *Annual Review of Condensed Matter Physics*, 1(1):323–345, Aug 2010.
- [40] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, and R. Aditi Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85(3):1143–1189, Jul 2013.
- [41] Clemens Bechinger, Roberto Di Leonardo, Hartmut Löwen, Charles Reichhardt, Giorgio Volpe, and Giovanni Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88(4), Nov 2016.
- [42] M. C. Marchetti, Y. Fily, S. Henkes, A. Patch, and D. Yllanes. Minimal model of active colloids highlights the role of mechanical interactions in controlling the emergent behavior of active matter. *Current Opinion in Colloid & Interface Science*, 21:34–43, 2016.
- [43] T. Vicsek, E. Czirók, Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6), 1995.
- [44] M. E. Cates, D. Marenduzzo, I. Pagonabarraga, and J. Tailleur. Arrested phase separation in reproducing bacteria creates a generic route to pattern formation. *Proceedings of the National Academy of Sciences*, 107(26):11715–11720, 2010.
- [45] M. E. Cates and J. Tailleur. When are active brownian particles and run-and-tumble particles equivalent? consequences for motility-induced phase separataion. *Europhysics Letters*, 101(20010), 2013.
- [46] Yaouen Fily and M. Cristina Marchetti. Athermal phase separation of self-propelled particles with no alignment. *Physical Review Letters*, 108(23), Jun 2012.
- [47] Gabriel S. Redner, Aparna Baskaran, and Michael F. Hagan. Reentrant phase behavior in active colloids with attraction. *Physical Review E*, 88(1), Jul 2013.
- [48] Gabriel S. Redner, Michael F. Hagan, and Aparna Baskaran. Structure and dynamics of a phase-separating active colloidal fluid. *Physical Review Letters*, 110(5), Jan 2013.
- [49] David Richard, H. Loewen, and Thomas Speck. Nucleation pathway and kinetics of phase-separating active brownian particles. *Soft Matter*, 12:5257–5264, 2016.
- [50] Gabriel S. Redner, Caleb G. Wagner, A. Baskaran, and Michael F. Hagan. Classical nucleation theory description of active colloid assembly. *Physical Review Letters*, 117(148002), 2016.
- [51] Isaac R. Bruss and Sharon C. Glotzer. Phase separation of self-propelled ballistic particles. *arxiv*, (1712.01983), 2017.

- [52] J. Palacci, S. Sacanna, A. P. Steinberg, D. Pine, and P. M. Chaikin. Living crystals of light-activated colloidal surfers. *Science*, 339:936–940, 2013.
- [53] Alexander P. Petroff, Xiao-Lun Wu, and Albert Libchaber. Fast-moving bacteria self-organize into active two-dimensional crystals of rotating cells. *Physical Review Letters*, 114(15), Apr 2015.
- [54] H. H. Wensink and H. Loewen. Emergent states in dense systems of active rods: from swarming to turbulence. *Journal of Physics: Condensed Matter*, 24(464130), 2012.
- [55] Yingzi Yang, Vincent Marceau, and Gerhard Gompper. Swarm behavior of self-propelled rods and swimming flagella. *Physical Review E*, 82(3), Sep 2010.
- [56] H. H. Wensink, V. Kantsler, R. E. Goldstein, and J. Dunkel. Controlling active self-assembly through broken particle-shape symmetry. *Physical Review E*, 89(1), Jan 2014.
- [57] Nguyen H.P. Nguyen, Daphne Klotsa, Michael Engel, and Sharon C. Glotzer. Emergent collective phenomena in a mixture of hard shapes through active rotation. *Physical Review Letters*, 112(7), Feb 2014.
- [58] Syeda Sabrina, Matthew Spellings, Sharon C. Glotzer, and Kyle J. M. Bishop. Coarsening dynamics of binary liquids with active rotation. *Soft Matter*, 11(43):8409–8416, 2015.
- [59] Matthew Spellings, Michael Engel, Daphne Klotsa, Syeda Sabrina, Aaron M. Drews, Nguyen H. P. Nguyen, Kyle J. M. Bishop, and Sharon C. Glotzer. Shape control and compartmentalization in active colloidal cells. *Proceedings of the National Academy of Sciences*, 112(34):E4642–E4650, Aug 2015.
- [60] A. Suma, G. Gonnella, D. Marenduzzo, and E. Orlandini. Motility-induced phase separation in an active dumbbell fluid. *Europhysics Letters*, 108(56004), 2014.
- [61] V. Prymidis, S. Samin, and L. Filion. State behaviour and dynamics of self-propelled brownian squares: a simulation study. *Soft Matter*, 2016.
- [62] Stephen Mann. Self-assembly and transformation of hybrid nano-objects and nanostructures under equilibrium and non-equilibrium conditions. *Nature Materials*, 8(10):781–792, Sep 2009.
- [63] L. Onsager. The effects of shape on the interaction of colloidal particles. *Annals New York Academy of Sciences*, pages 627–659, 1949.
- [64] V. N. Manoharan. Colloidal matter: Packing, geometry, and entropy. *Science*, 349(6251):1253751–1253751, Aug 2015.
- [65] G. van Anders, D. Klotsa, N.K. Ahmed, Michael Engel, and Sharon C. Glotzer. Understanding shape entropy through local dense packing. *Proceedings of the National Academy of Sciences*, 111(45):E4812–E4821, 2014.
- [66] B. van der Meer, M. Dijkstra, and L. Filion. Removing grain boundaries from three-dimensional colloidal crystals using active dopants. *Soft Matter*, 12(25):5630–5635, 2016.
- [67] S. A. Mallory and A. Cacciuto. Activity-assisted self-assembly of colloidal particles. *Physical Review E*, 94(022607), 2016.
- [68] Jie Zhang, Jing Yan, and Steve Granick. Directed self-assembly pathways of active colloidal clusters. *Angewandte Chemie International Edition*, 55(17):5166–5169, Mar 2016.

- 
- [69] Matthew Spellings, Ryan L. Marson, Joshua A. Anderson, and Sharon C. Glotzer. Gpu accelerated discrete element method (dem) molecular dynamics for conservative, faceted particle simulations. *Journal of Computational Physics*, 334:460–467, 2017.
  - [70] Steven Atkinson, Yang Jiao, and Salvatore Torquato. Maximally dense packings of two-dimensional convex and concave noncircular particles. *Physical Review E*, 86(3), Sep 2012.