Identifying and designing an information-driven approach for targeted colloidal self-assembly

Shannon Moran

February, 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)

Note: Total length must be less than 15 pages of text. Includes figures, excludes title page, list of references, and CV.

Contents

List of Figures

Preliminary Exam: Project Summary

Paragraph 1: Motivation

Paragraph 2: Where there are openings in the conversation about the role of information in self-assembly

Paragraph 3: 2 sentences each on the projects proposed

Paragraph 4: Concluding thoughts on future work

1 Introduction and motivation - 1 pg

When we think about the major challenges facing materials science, we are fundamentally faced with this idea of inversely designing materials. That is, I decide that I want to create a material that behaves your sweat-wicking shirt under one condition, stiffens under another, and when given a particular stimulus can reconfigure its structure. Currently, if I wanted to make a material like that for you, I'd naively take materials that have each of those properties and figure out how I could get them to work together. Or, I'd look for novel materials that have properties close to those of the material I want to make. We would call this designing the material.

This is inefficient. In the inverse design problem, we take the properties we want and create the materials that will give us those properties. Machine learning and materials science are coming together on the active front of this research. However, being able to predict, even perfectly, what can be made from existing materials by definition limits us to the set of materials that currently exist. This is a known challenge in materials science– how do we probabilistically explore phase space outside of phase space where we have data? While intriguing in its own right, that is not the topic of the thesis proposed here.

Instead, we might think about this from a fundamental physics point of view. If we want to make complex materials that have embedded stimuli responses, or assemble into a specific target structure, we must give the building blocks of such complex materials some amount of direction. We can think about this amount of direction as an amount of *information*.

This is not to say that we are looking to have building blocks act as storage devices, as in [1]. In that work, each building block is a cluster of multiple particles in whose arrangement can be stored a "high density" of information.

Similarly, in a recent proposal between our group and those of Marke Bathe (MIT), Mawgwi Bawendi (MIT), and Oleg Gang (Columbia), we proposed a biosynthetic, high-density storage structure composed of DNA nanocubes (Figure 2). Within these nanocubes, information could be stored in the different dies intercolated into the frame of the cube, into quantum dots placed into the frames, or even in the shape of the frames themselves. If we them move a level higher, we can imagine storing additional information in the order of these nanocubes relative to one another.

However, using these and solutions like them for high-density storage requires us being able to write, read, and store information into these formats. Fundamentally, these three challenges are predicated upon the ability to specifically place blocks where they need to go ("write"). Current methods include sonic and laser tweezers (manual), specific DNA interactions (energetic), or incremental addition (kinetic). How do we compare between these methods, though?

Here, I propose that the ability of building blocks to form a target structure can be distilled down into a concept of *information*.

This is not a new concept. In our group, we are comfortable with the concept that a target structure is the result of a minimization of free energy. In systems devoid of inter-particle forces, this then reduces down to a maximization of entropy.

Statistical mechanical "entropy" shares its name with information "entropy" in communications theory. While this directly came about because of the form similarity between the two, much energy since has been devoted to developing frameworks connecting the two. Jaynes, in the 1950s, spent two long articles trying to reconcile the two. Books, and multiple articles, have been dedicated to explaining why these concepts are similar.

While much time has been spent developing the theory, very little time has been spent directly leveraging this concept for embedding information in systems governed by statistical mechanical ensembles– such as colloidal-scale self-assembly.

Key line from Simons proposal: "A coherent framework of thermodynamic and non-equilibrium processes seen through information theoretic eyes could lead to new theories for encoding information in matter– which would allow for the design of novel materials and novel material behavioral control."

New outline:

- Self-assembly in materials science can lead to a variety of structures, complexity
- The future of materials science relies upon inverse design
- Inverse design requires understanding how the building blocks of a material inform its overall structure
- Thinking more simply about tailored self-assembly... we want to direct the behavior of a system
- We can think of this as giving the system some amount of information: binding preferences, kinetics, etc
- Lots of work has gone into trying to understand how to measure and then most efficiently provide systems with that direction (will detail in the background section)
- You know what does this really well, though? Proteins in nature
- Proteins conformationally change while binding; lots of complexity
- Here we use a simpler system of folding nets, which actually has much less complexity
- The overarching challenge is: what is the most efficient way of assembling a given structure?
- Such a question could motivate a career, so we will further limit this scope in the coming pages.

Key problem, taken from [10]: Self-assembly holds promise for creating new materials and devices because of its inherent parallelism, allowing many building blocks to simultaneously organize using preprogrammed interactions. An important trend in nanoparticle and colloid science is the synthesis of particles with unusual shapes and/or directional (??patchy??) interactions, whose anisotropy allows, in principle, assemblies of unprecedented complexity. However, patchy particles are more prone to long relaxation times during thermodynamically driven assembly, and there is no a priori way of predicting which particles might be good assembly candidates. Here we demonstrate a new conceptual approach to predict this information using sequences of intermediate clusters that appear during assembly. **Unfortunately, when an equilibrium solution or simulation of patchy particles fails to generate an ordered pattern it is not always obvious whether the culprit is thermodynamics or kinetics.** Recently there have been studies that attempt to quantify kinetic trapping through fluctuation-dissipation ratios (21,22) and through the interplay between specific and nonspecific interactions (3,5,23) but these methods do not provide predictive capabilities for thermodynamically stable structures. The fact that both thermodynamics and kinetics can prevent a system of particles from self-assembling is particularly trou- blesome for experimentalists that search parameter space via trial-and-error because experiments that fail to assemble do not provide information about how assembly might be improved.

We are ultimately searching for rational design of building blocks optimized for self-assembly that focuses on assembly pathway engineering: identifying the traps that occur as a system assembles so they may be circumvented. As systems self-assemble we hypothesize that the thermodynamically stable intermediate clusters that arise hold information about their ability to order. These sequences of intermediate clusters are assembly pathways and we propose a methodical analysis of them to predict the degree to which a system of building blocks will assemble a target pattern, which we refer to as the building block?s assembly propensity for the pattern. We foresee assembly pathway engineering proceeding as a collaboration among structural identification, kinetic measurements, and the assembly pathway analysis described here. [10]

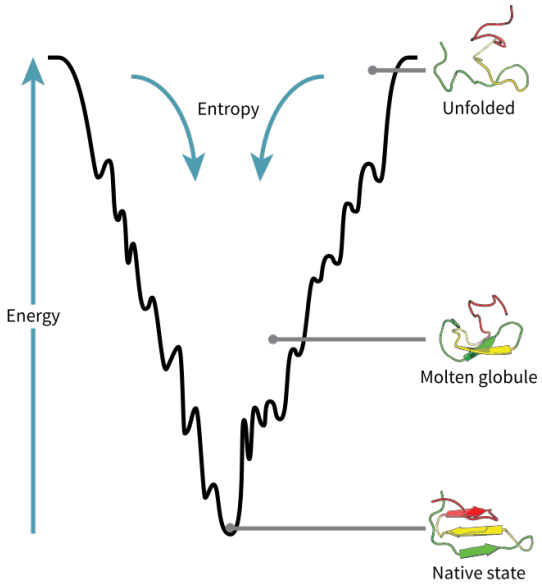

Figure 1: The diagram sketches how proteins fold into their native structures by minimizing their free energy. (Source: Wikipedia

2 Proposed Research

2.1 Model background

For our system, we use nets of the 5 Platonic shapes as described in [37]. These nets are implemented as rigid bodies with harmonic springs. Edges of each net "face" have non-specific attractive patches. Molecular dynamics (implemented in HOOMD-blue [31, 32]) is used to simulate assembly in the canonical (NVT) ensemble.

In a forth-coming study from the Glotzer lab using this system [37], the authors find that for the same target shape, compact nets with few leaves assemble most reliably (in agreement with [38]. By investigating the assembly behavior for shapes with fewer possible net constructions (ADD), they are then able to use these features to predict which nets will assemble for shapes with a wide variety of possible net constructions. 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) [39]. 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 know in studies of protein folding. In 1968, Cyrus Levinthal posed the following thought experiment, come to be known as "Levinthal's Paradox" [40]. Even at that time, proteins were known to assemble in less than 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 and "serve as nucleation points in the folding process".

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 [41, 42]. 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 β-pleated sheets [43].

Further work on elucidating experimental protein kinetics detected the partially folded intermediates transition states in the folding process posited by Levinthal [44]. Statistical mechanical models of protein folding uncovered that assembly does not follow single microscopic pathway. Instead, protein folding is characterized by funnel-shaped energy landscapes, with entropy increasing as proteins assemble with decreasing free energy as shown in Figure 1 [45].

However, low free energies of a target structure do not guarantee efficient assembly. One particularly clarifying way of seeing this is through disconnectivity graphs developed by David Wales and colleagues [22]. (Describe disconnectivity graphs here and how they're used.)

(Jacobs– worth talking about protein contact maps here?)

Given that we are able to replicate complex findings from protein folding in this simple system, this suggests that we can continue to use this system as a toy model for developing theory of tailored self-assembly. Levinthal posited that some local contacts may act as "nucleation points in the folding process". Can we identify similar critical bonds in foldings of nets? Is it possible to change the folding propensity of a "bad folding" net by seeding the net with a critical bond?

This second question follows from theories in origami. Why is it important that we find these most-critical self-assembly "seeds"? Origami theory is developed for a system a zero temperature– that is, developing theories to avoid needing to back-track along an assembly pathway (e.g. after encountering a local-minima) are desirable. In a class of NP-hard satisfiability (SAT) problems, careful seeding of the search for global minima has been shown to greatly reduce or even eliminate backtracking [56](10) before reaching the global minima [56](43). (Fun fact: this includes filling in the right boxes first in a game of Sudoku [56](43).)

Check [56](43) for good intro on origami and self-assembly.

2.2 Proposed projects

Nature is very good at already picking an optimized route through a free energy landscape [46].

Brain-dump of tasks I want to do:

- Can we adapt the binding mutual information idea from Brenner et al and apply it to nets, which have energetically non-specific but geometrically different binding opporunities?
- Can we identify what bonds are the "most important' for forming the finished structure? E.g. if I only want to add one specific bond, which is the best to add? How much of a difference can one bond make for a "bad" net? Can we borrow any other ideas from protein folding?
- Proteins fold by preferring local contacts– do some nets make
- Desmaine had that algorithm for minimizing the number of actuations needed to fold a specified shape
- For proteins, Jacobs demonstrated which bonds were critical to folding proceeding [46]

Methods:

1.
2. Make disconnectivity/funnel-type graphs for the nets studied by Paul

This requires that I can quantify:

- What it means for a net to have information: Specificity of its bonds
- Specificity of interactions
- Number of pathways available to the shape
- How narrow the energetic funnel

Thought experiment:

1. Let's say we're given a net that folds badly. We'll say that with non-specific interactions, the net would fold into the target structure with Y% probability at a given temperature. (We can get this from Paul's data)
2. We could then say, let's start these nets with one side in contact, either native or non-native contact
3. How does this biasing change the probability that this net will fold into the target?
4. Can we define some measure of how important that bond is to the net forming?
5. This is similar to Sudoku or other challenges where getting the right first move is really import
6. Challenge, though: What about having multiple connections?

1) Define a measure of pathway information.

- We already have ways of measuring how good a particular bond is
- Are there particular bonds/connections that are the most important to get correct to enable forming the desired final structure?
- With this metric, can we then define and minimize an information efficiency, e.g. the amount of bond specificity we need across the system to get a given success rate of assembly?
- Can we figure out which bonds/interactions are the most critical in an assembly process or an assembly pathway?

2) Use that measure of pathway information to design ideal pre-cursors for target structures.

- Pathway design using machine learning is a big open question
- We could find feature correlations, like Paul did
- There?s also an approach called ?Computable Information Density? published by some colleagues (Chaikin) on the arxiv last August. Basic idea is that you can (1) somehow represent your system as an array of information which you can (2) run though a compression algorithm and (3) the ?information? is just the length of that compressed information

3) Attempt to use machine learning to predict ideal pre-cursors for given target structures.

- [55]: Nonlinear Machine Learning of Patchy Colloid Self-Assembly Pathways and Mechanisms out of the Furguson group

2.3 Collaboration: Synthetic biology memory

The below was a response to an NSF call for proposals for a Semiconductor Synthetic Biology.

We proposed using DNA-mediated assembly to store information in nanoparticle arrays.

Specifically, this is an example of addressable complexity, then trying to engineer how to get the particles to where they should go in the most energetic and information/complexity-efficient manner possible.

From intro: In Aim 1, we will investigate monomeric block formation, exploring the self-assembly of arbitrary geometric DNA objects with incorporated optical elements that can be manufactured as information carriers, while allowing for superstructure formation through DNA-sequence barcoding. We will explore static assembly of 1D arrays of such DNA nanoparticles integrated with Memory Blocks (DNAMB) for encoding bitstream information that can be read out by fluorescence

and electron microscopy. In Aim 2, we will explore 2D and 3D assembly, investigating techniques to algorithmically assemble and read out digital 2D and 3D information using optical and tomographic methods. In Aim 3, we will use molecular decision computing to assemble distinct, alternative lattices based on specific external signals. These results will offer the ability to encode and decode arbitrary datasets in ultra-dense molecular hard-drives, with environmental sensing and recording.

Aim 2. Dense, programmable molecular memory in 2D and 3D bit module lattice assemblies Overview & Rationale. Nanoparticle self-assembly depends on a balance of interaction forces, entropic effects, and system kinetics37,53,69-73. We can leverage these properties to direct self-assembly of shaped DNA nanoparticle into 2D and 3D arrays by controlling the position and valency overhangs that provide connectivity between DNA nanoparticles. Wireframe structure of DNA particle is highly suitable for encapsulation of memory blocks (e.g. Au NP, QDs, fluorescent dyes) and creation of DNAMB, a pixel in 2D or 3D arrays. To achieve information storage capabilities, it is required to investigate how the connectivity properties of DNAMB can be translated into their designed arrangement in the information- storing arrays. To self-assemble these systems into 2D and 3D ultra-dense data blocks, we will investigate the minimum interaction specificity needed to direct self-assembly into high fidelity ordered 2D and 3D arrays. We will also explore information retrieval from these arrays in 2D and 3D within pixels consisting of 1x1, 2x2, 4x4, etc., nanoparticle block arrays. In addition, we will establish methods for generating robust memory arrays that can preserve information under extreme conditions.

Proposed research, Aim 1: Computationally, Glotzer and colleagues will develop a bit module interaction model to study the role of the DNA linkages on DNA cage self-assembly. Specifically, previous work on modeling solid particles with DNA-facilitated attraction37 will be extended to model the DNA cages that will be experimentally made by Gang, and it will consider realistic features of nanoparticle systems82,83. With this model in place, we can then extend the framework of digital alchemy, which treats particle properties as a thermodynamic variable, to particle interactions (here, DNA linkages)54. In this way, we can inversely design ideal DNA cages (e.g. shape, patchy interactions) that will robustly assemble a target structure. We will seek to balance site specificity without being overly unique?that is, design the highest information interactions that will allow for the minimum amount of linkage specificity for directing self-assembly84,85. This computational framework for the inverse design of bit packages that will assemble a given structure will enable high- throughput screening of particles of interest and serve as the basis for complex hierarchical structure and array assembly in the remainder of Aims 2 and 3.

Sub-Aim 2.2. Hierarchical assembly logic for higher-dimensional information storage Overview. In Sub-aim 2.1, we explored approaches to assembling DNA frames into target 2D and 3D assemblies. Next, we precisely order ?bit modules??that is, DNA cages carrying functional particles? into arrays of discrete information. Toward this end, we design modules that carry the minimal information needed to reach target arrangements through a combination of particle anisotropy and DNA linkers. DNA computing groups have previously used DNA linkers to self-assemble complex 2D patterns6 and 3D shapes9. Here we extending these approaches to realize hierarchical 3D nanoparticle assembly design so that pixelated images act as dense data storage units. We will explore several complementary approaches to hierarchical assembly engineering, including sequential nanoparticle addition and ?one-pot assembly?, each of which will be explored together with inverse computational design of self-assembly pathways and particle geometries to achieve a robust assembly of designed arrays. Using the optical characterization strategies from Sub-aim 2.1, we will decode the information encoded in the structure, and probe sources of error and information loss in the self-assembly and read-out processes.

Proposed research. Assembly of encoded 3D arrays can be approached in two strategies, or a combination thereof (Fig. 2). In an entirely ?one-pot? assembly of a 3D array, all modules are linked with a large binding sequence set that has been fully computationally defined. Such an

approach requires an enormous number of unique binding sequences, and even if fully defined, can run into high error rates when considering the assembly and packing of large (as compared with molecular assembly) and charged modules and/or materials. A second approach using sequential binding based on module groups of similar binding layouts requires less sequence diversity and can be automated using robotic liquid handling. However, this is vastly more process- and time-intensive than one-pot assembly. This approach represents hierarchical assembly, whereby 1D structures (?strings?) would be formed from the modules, 2D planes formed from the 1D libraries, and finally 3D encoded arrays from stacking of selected planes. An optimal assembly process that balances fully-encoded organization with direct addition of binding components would offer a hybrid approach of hierarchical assembly with sequential addition of groupings of computationally defined structures. Each of these strategies will be explored in this aim, using a combination of high-throughput, structure-based computational modeling and experiments.

The Glotzer group will extend their digital alchemy framework to probe diverse DNA linkage sequences and conjugation designs to realize specific, targeted inter-particle interactions. In addition, they will explore the roles of these interactions on the kinetics of array assembly to enable pathway design [10] into desired arrays while avoiding undesirable ?side products?. In this way, we will explore computationally the interplay between the two extremes of one-pot and sequential assembly, and identify which combinations provide lowest assembly error while minimizing both assembly time and the number of required unique binding sequences. The Gang group will employ a home-built robotic system for automatic synthesis and assembly of DNAMB; that will allow establishing practical methods for creation of large number of diverse blocks required for the hierarchical assembly. While such approaches have been applied to molecular systems, they have not yet been realized for DNA frames integrating inorganic NP. To implement complementary pathway design strategies, Gang will fabricate DNA frames with thermally differentiated inter-vertex hybridizations to promote highly specific assembly path during thermally-driven self- assembly. For example, DNAMB strings will be assembled at higher temperatures, and planar and 3D arrays assembled at lower and lowest temperatures, respectively. We will use SAXS and tomography methods to reveal the pathway-controlled assembly process. The computational design of frames and pathways will be performed jointly between the Bathe, Glotzer, and Gang labs.

3 Time table

See Figure 3 for key tasks and milestones through 2020, based on the projects outlined in the above sections.

4 Conclusions and potential impact

Ultimately, tailoring self-assembly and addressable complexity are driven by the question:

How do I give particles in my system sufficient information to find their proper position in a target whole?

In systems targeting particular structures, tailored attraction (experimentally via single-stranded DNA or theoretically via an attractive potential) and repulsion can be used to tailor the structure of particle assembly. In systems targeting addressable complexity with DNA origami and protein folding, such direction can be given with exact detail through genetic code. Work by Brenner and Manoharan has sought to quantify the amount of specificity bonds in such systems can contain, and the limits on target system complexity given differing bond informations.

However, missing in this conversation is a measure of the importance of each individual interaction on the emergent behavior of the whole system. I intend to contribute in part during the remainder

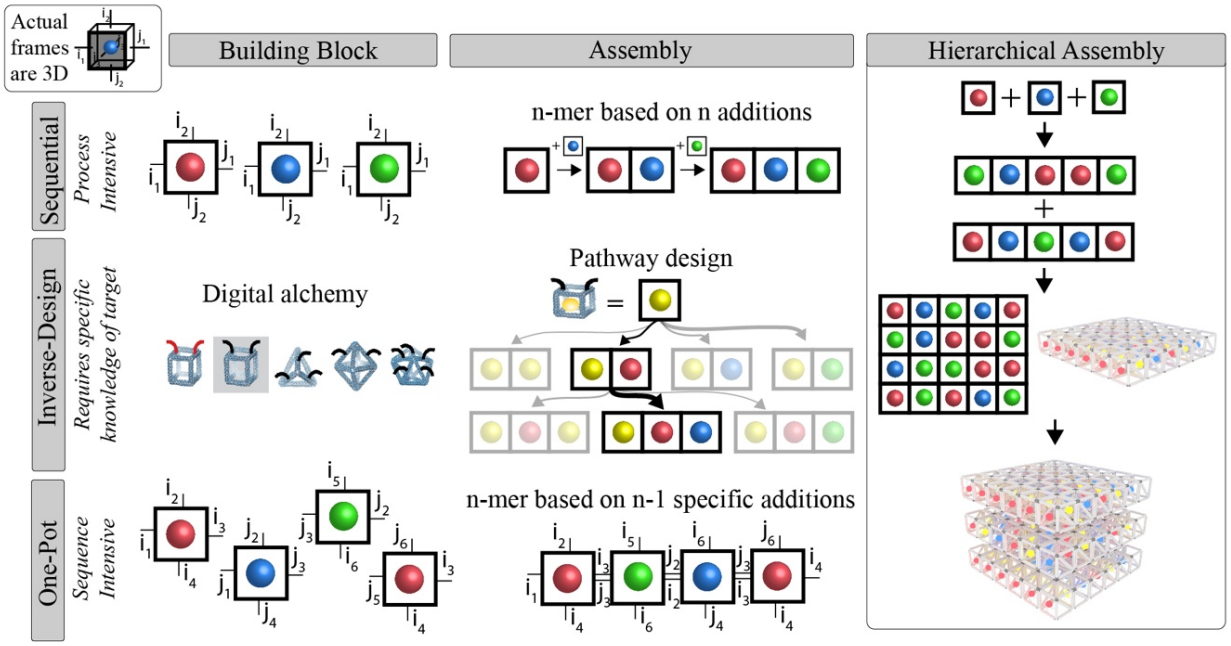

Figure 2: (Lifted from SemiSynBio Proposal) Strategies for building information encoded 1D, 2D and 3D arrays. Sequential operations are very deterministic and can be carried out by automated robotic equipment, but even so are heavily process intensive and require many individual assembly steps. One-pot systems can be fully computationally defined, though in practice are heavily sequence intensive and be subjected to errors more readily than in molecular- scale systems when accounting for kinetic and thermodynamics of packing larger objects and materials. A hierarchical assembly methodology offers a hybrid approach of both strategies, where a sequential addition of structures preformed in a one-pot setup provide the desired 3D material organization.

of my PhD. Instead of simply observing emergent behavior as an outcome of specifically-designed behavior of individuals, we could instead engineer such behavior as a quantifiable outcome of the interaction of an information-rich network of agents. Having such knowledge may allow us to embed multiple "states" in a material, such as the minimum folds to move between two target origami structures [57]. Being able to measure, and ultimately control, the assembly directions stored in a starting material is a critical step to enabling materials by design and advancing materials science.

			2017				2018				2019				2020	
Project	**Status**	**Description**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**
Active shapes	100%	Data collection														
		Analysis														
		Writing														
		Submit for publication														
Defining information		Object 1														
		Object 2														
		Object 3														
		Object 4														
Applying information		Object 1														
		Object 2														
		Object 3														
		Object 4														
Machine learning		Object 1														
		Object 2														
		Object 3														
		Object 4														
Thesis		Data meeting														
		Defense														

Figure 3: Key milestones and tasks from Preliminary Exam through target defense date.

References

[1] Carolyn L. Phillips, Eric Jankowski, Bhaskar Jyoti Krishnatreya, Kazem V. Edmond, Stefano Sacanna, David G Grier, David J. Pine, and Sharon C. Gltozer. Digital colloids: reconfigurable clusters as high information density elements. *Soft Matter*, 10:7468–7479, 2014.

[2] 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.

[3] 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.

[4] Pablo F. Damasceno, Michael Engel, and Sharon C. Glotzer. Predictive self-assembly of polyhedtra into complex structures. *Science*, 337:453–457, 2012.

[5] 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.

[6] Zheni Zhang and Sharon C. Glotzer. Self-assembly of patchy particles. *Nano Letters*, 4(8):1407–1413, 2004.

[7] Sharon C. Glotzer and Michael J. Solomon. Anisotropy of building blocks and their assembly into complex structures. *Nature*, 6:557–562, 2007.

[8] Eric Jankowski and Sharon C. Glotzer. A comparison of new methods for generating energy-minimizing configurations of patchy particles. *The Journal of Chemical Physics*, 131(104104), 2009.

[9] Eric Jankowski and Sharon C. Glotzer. Calculation of partition functions for the self-assembly of patch particles. *The Journal of Physical Chemistry B*, 115:14321–14326, 2011.

[10] Eric Jankowski and Sharon C. Glotzer. Screning and designing patchy particles for optimized self-assembly propensity through assembly pathway engineering. *Soft Matter*, 8(2852), 2012.

[11] 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.

[12] Carlos Ernesto Castro, Fabian Kilchherr, Do-Nyun Kim, Enrique Lin Shiao, Tobias Wauer, Philipp Wortmann, Mark Bathe, and Hendrik Dietz. A primer to scaffolded dna origami. *Nature Methods*, 8(3):221–229, 2011.

[13] Pengfei Wang, Stavros Gaitanaros, Seungwoo Lee, Mark Bathe, William M. Shih, and Yonggang Ke. Programming self-assembly of dna origami honeycomb two-dimensional lattices and plasmonic metamaterials. *Journal fo the American Chemical Society*, 138:7733–7740, 2016.

[14] Remi Veneziano, Sakul Ratanalert, Kaiming Zhang, Hao Zhang, Fei adn Yan, Wah Chiu, and Mark Bathe. Designer nanoscale dna assemblies programmed from the top down. *Science*, page 10.1126/science.aaf4388, 2016.

[15] Po-Ssu Huang, Scott E. Boyken, and David Baker. The coming of age of *de novo* protein design. *Nature*, 537:320–327, 2016.

[16] Chad A. Mirkin, Robert L. Letsinger, Robert C. Mucic, and James A. Storhoff. A dna-based method for rationally assembling nanoparticles into macroscopic materials. *Nature*, 382:607–609, 1996.

[17] Sung Yong Park, Abigail K. R. Lytton-Jean, Byeongdu Lee, Steven Weigand, George C. Schatz, and Chad A. Mirkin. Dna-programmable nanopartilce crystallization. *Nature*, 451:553–556, 2008.

[18] Matthew N. O'Brien, Hai-Xin Lin, Martin Girard, Monica Olvera de la Cruz, and Chad A. Mirkin. Programming colloidal crystal habit with anisotropic nanoparticle building blocks and dna bonds. *Journal fo the American Chemical Society*, 138:1456214565, 2016.

[19] Wenyan Liu, Jonathan Halverson, Ye Tian, Alexei V. Tkachenko, and Oleg Gang. Self-organized architectures from assorted dna-framed nanoparticles. *Nature Chemistry*, 8:867–873, 2016.

[20] Dmytro Nykypanchuk, Matthew M. Maye, Daniel van der Lelie, and Oleg Gang. Dna-guided crystallization of colloidal nanoparticles. *Nature*, 451:549–552, 2008.

[21] Aleks Reinhardt and Daan Frenkel. Numerical evidence for nucleated self-assembly of dna brick structures. *Physical Review Letters*, 112(23), Jun 2014.

[22] David J. Wales. Atomic clusters with addressable complexity. *The Journal of Chemical Physics*, 146(054306), 2017.

[23] Juan M. R. Parrondo, Jordan M. Horowitz, and Takahiro Sagawa. Thermodynamics of information. *Nature Physics*, 11:131–139, 2015.

[24] Arieh Ben-Naim. *A Farewell to Entropy: Statistical Thermodynamics Based on Information*. World Scientific, 5 Toh Tuck Link, Singapore 596224, 2008.

[25] Guangnan Meng, Natalie Arkus, Michael P. Brenner, and Vinothan N. Manoharan. The free-energy landscape of clusters of attractive hard spheres. *Science*, 327:560–563, 2010.

[26] William M. Jacobs, Aleks Reinhardt, and Daan Frenkel. Communication: Theoretical prediction of free-energy landscapes for complex self-assembly. *The Journal of Chemical Physics*, 142(021101), 2015.

[27] 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.

[28] Ken A. Dill. Folding proteins: finding a needle in a haystack. *Current Opinion in Structural Biology*, 3:99–103, 1993.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] John D. Weeks, David Chandler, and Hans C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics*, 54(12):5237–5247, Jun 1971.

[35] Carolyn L. Phillips, Joshua A. Anderson, and Sharon C. Glotzer. Pseudo-random number generation for brownian dynamics and dissipative particle dynamics simulations on gpu devices. *Journal of Computational Physics*, 230(19):7191–7201, Aug 2011.

[36] Shannon E. Moran, Isaac R. Bruss, and Sharon C. Glotzer. Leveraging anisotropy for tailored self-assembly of active matter. *in preparation*, 2018.

[37] Paul M. Dodd, Pablo F. Damasceno, and Sharon C. Glotzer. Polyhedra nets fold cooperatively at high temperature. *unpublished*, 2018.

[38] 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.

[39] 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.

[40] Cyrus Levinthal. How to fold graciously. *University of Illinois Bulletin: Mössbaun Spectroscopy in Biological Systems Proceedings*, 67(41):22–24, 1969.

[41] Robert Zwanzig, Attila Szabo, and Biman Baghi. Levinthal's paradox. *Proceedings of the National Academy of Sciences*, 89:20–22, 1992.

[42] Leandro Martìnez. Introducing the levinthal's protein folding paradox and its solution. *Journal of Chemical Education*, 91:1918–1923, 2014.

[43] 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.

[44] 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.

[45] Ken A. Dill and Sun Chan Hue. From levinthal to pathways to funnels. *Nature*, 4(1):10–19, 1997.

[46] William M. Jacobs and Eugene I. Shakhnovich. Structure-based prediction of protein-folding transition paths. *Biophysical Journal*, 111:925–936, 2016.

[47] Paul W. K. Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440(16):297–302, 2008.

[48] Joseph O'Rourke. *How to Fold It: The Mathematics of Linkages, Origami, and Polyhedra*. Cambridge University Press, 32 Avenue of the Americas, New York, NY 10013-2473, USA, 2011.

[49] Joseph O'Rourke and Erik D. Demaine. *Geometric Folding Algorithms: Linkages, Origami, Polyhedra*. Cambridge University Press, 32 Avenue of the Americas, New York, NY 10013-2473, USA, 2007.

[50] Gregory R. Bowman and Vijay S. Pande. Protein folded states are kinetic hubs. *Proceedings of the National Academy of Sciences*, 107(24):10890–10895, 2010.

[51] Frank Noè, Christof Sch̀utte, Eric Vanden-Eijnden, Lothar Reich, and Thomas R. Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences*, 106(45):19011–19016, 2009.

[52] Matthew R. Perkett and Michael F. Hagan. Using markov state models to study self-assembly. *The Journal of Chemical Physics*, 140(214101), 2014.

[53] Philipp Metzner, Sch̀utte, and Eric Vanden-Eijnden. Transition path theory for markov jump processes. *Multiscle Modelling and Simulation*, 7(3):1192–1219, 2009.

[54] E Weinan and Eric Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual Reviews of Pysical Chemistry*, 61:391–420, 2010.

[55] Andrew W. Long and Andrew L. Ferguson. Nonlinear machine learning of patchy colloid self-assembly pathways and mechanisms. *The Journal of Physical Chemistry B*, 118:4228–4244, 2014.

[56] Menachem Stern, Matthew Pinson, and Arvind Murugan. The difficulty of folding self-folding origami. *arXiv*, 1703.04161v1, 2017.

[57] 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.