Self-Assembly of Colloidal Spheres with Specific Interactions

Self-Assembly of Colloidal Spheres with Specific Interactions

A dissertation presented
by
Jesse Wronka Collins
to
The School of Engineering and Applied Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Applied Physics

Harvard University
Cambridge, Massachusetts

May 2014

Self-Assembly of Colloidal Spheres with Specific Interactions

Abstract

In this thesis, I discuss engineering colloidal particles to have specific, isotropic interactions and studying their cluster geometries in equilibrium. I discuss light scattering experiments showing that a highly specific protein, Dscam, is unstable against thermal aggregation. This result lead me to use DNA instead to control interparticle specificity. I coated 1-micron diameter polystyrene particles uniformly with DNA. I used fluorescence microscopy with oxygen-scavenging enzymes to observe these particles self-assembling in clusters. These experiments show that a packing of 6 spheres that is rarely seen in a single-component system is observed very often in an optimized 3-species system. Then I show experiments using the same 3 species but 9 total particles, finding that the equilibrium yields of the most likely cluster relative to other stable clusters are lower than at 6 particles. I conclude from these experiments that optimizing the assembly of an otherwise unlikely configuration may require nearly as many species as particles. Finally, I investigate the scalability of self-assembly of particles with isotropic and specific interactions theoretically. I use both exact and approximate partition functions to show that spheres with specific interactions can have energy landscapes with thermodynamically large numbers of strictly local minima relative to the number of their ground states. Compared to single-component systems, these systems of many different species may spend much more time in kinetic traps and never reach their ground states. Finally, I discuss briefly some directions for further study, including questions of how the results in this thesis may be related to protein folding and complex formation.

Contents

For my brother Jonathan,

who, one boyhood night, shared with me questions I continue to pursue: whether God is chemistry in our brains, and how society might become more dedicated to science

And how awkward is the human mind in divining the nature of things, when forsaken by the analogy of what we see and touch directly?

—LUDWIG EDUARD BOLTZMANN

If they were all one part, where would the body be? As it is, there are many parts, but one body. The eye cannot say to the hand, “I don’t need you!” And the head cannot say to the feet, “I don’t need you!” On the contrary, those parts of the body that seem to be weaker are indispensable

—1 CORINTHIANS 12:19-22

I’ve found a new awareness! I thought I was finished, but there are so many things I had never thought were possible! The more we study budō, the more fearful it becomes, but it is so magnificent in its unlimited vastness!

—RYU, STREET FIGHER ALPHA 3

Acknowledgements

In my life, and especially during my PhD, I have often, involuntarily and voluntarily, recalled memories of friends, colleagues, family and competitors in times when I wanted to give up. I have received direct and indirect forms of help, and it is not clear to me that I would have completed the research reported here without all the people in my life. Although I think of science as something that in principle can be performed by an individual in solitude, I do not have any evidence that any scientific knowledge can be found without at least having others to find it for, and who may, however subtly, influence one's direction of inquiry. Here I acknowledge explicitly a small subset of the people who have influenced and helped me along the way and whom I am able to recall voluntarily and write about in the short amount of time remaining before the May 16th deadline.

There are a few people who have not met me and who have not done research I am citing but whom nonetheless I consider invaluable to my completion of this project while maintaining something that might be broadly referred to as "sanity." I am very grateful to Robert F. Diggs, Wayne Coyne (with whom I did meet in the midst of writing this thesis), Thom Yorke and their respective musical groups. Their music and lyrics were especially important to me when my PhD training became particularly difficult and as it made me feel increasingly isolated from much of society. "Race for the Prize" made me feel like my chosen path was beautiful and appreciated even if I didn't end up making discoveries of the caliber I dreamed of. When I most wanted to give up with my lab work, I would listen to Wu-Tang and continue my experiments despite feelings of intense desperation. Along these lines I am also grateful to Noam Chomsky and John Dower, whose cultural views likewise made me feel part of a broader diaspora at the same time that my training made me feel far from much of culture. I am also grateful to developers of modern electronic communications technology for similar reasons. I must also thank Paul Zaloom, and those who helped him share his character, Beakman, with the world. Beakman embodies much of what I love about science, and I'm not sure I would have thought science worth pursuing without his influence.

I am grateful to members of my extended family for shaping me as a scientist in direct and indirect ways. Besides making me feel loved, my aunts, uncles, grandparents and cousins all contributed to my intellectual development. Many of my uncles and aunts on the Wronka side showed me how to love life and be an engineer, and I think it unlikely I would have applied to engineering schools were it not for them. My Uncle Phil has been a role model for giving my all to a cause much greater than myself. My Godfather, Mark, has taught me how to enjoy life and proudly maintain my inner child, something that I view as indispensible for making scientific progress. I also want to thank my Uncle Andrew for demonstrating how to cherish family memories while becoming independent. The combination of intellectualism and love practiced by all my Grandparents, both Josie and Walt, and Rose and Tom, and which they imparted on their families, opened the educational path I have taken. I think it a testament to my Grandpa Walt's influence that he, like me, understands that my work is about entropy without understanding what that means. I am very grateful to my cousins Sarah, Peter and Tommy for making me feel like I could be competitive about higher education and still be "cool." I thank my Uncle Rod for his questions about my work, such as "So did you figure it all out yet?", sharing with me the laughter, wonder and mystery that underlie my motivations. I am grateful to my Grandma Rose and Aunt Kandy and others for reminding me to decide to be happy every day, and to my Uncle Chris for also showing me how to stay happy and determined in face of adversity. Although Vijaya, Nath, Parimala and Krishna have been relatively recent additions to my family, their support of my scientific endeavors and discussions of my work and theirs

have been very meaningful to me. It is wonderful to have members of my family who have taken such a similar path in life and who hold PhD education and biochemical research in such high regard.

Many friends of mine have supported me throughout my scientific training. Joe Laws not only made me feel comfortable explaining my original plan for the self-assembly of colloidal spheres using biomolecules, he was excited about its potential and made me feel it was worth making sacrifices for. By going through Army Ranger school and leading troops in Afghanistan, Joe made me feel like my problems in lab were never something I should let get me down, and I thought of him often when I felt disinclined to push on. Jordy Clements, I think more than anyone else, has helped me feel comfortable speaking publically and honestly, an exercise which, though I did not expect it, proved essential to my scientific training. His continued interest in my work and its affect on my development has helped me better understand both my work and myself. Dean Ljubicic has been both a great friend and friendly competitor since we met at MIT. Besides both inspiring me with his own PhD work and always being there for conversation or down time, without Dean I would not have had many other formative friendships I've found among the brothers of Phi Delta Theta Mass Gamma, many of which influenced my decisions to pursue laboratory work and my PhD. Among my other fraternity brothers, Rick Henrikson and George Eng were particularly influential. I'd also like to thank Matt Molski and Kevin Bruffy, whose fierce competitive natures but also love of life have had a strong influence on my determination to complete my research project.

Although many school teachers have influenced me over the years, I must acknowledge a select few here. Al Rosen opened my eyes to the beauty and eloquence of physics for making sense of the world while teaching me that just because you're lecturing about physics doesn't mean you can't be funny too. Russell Brown first got me wondering about what defines "life" in an academic context, and although I did not expect this in high school, that question became and continues to be a primary scientific interest of mine. Eric Lander and the staff of 7.012 first showed me that the study of biology is about much more than memorization, and piqued my curiosity into the field. I am also grateful to whoever at MIT decided that 7.012 should be an institute requirement. Sudeep Agarwala, my 7.02 TA, introduced me, via a 1-3 star rating system, to which problems were of greatest import in the biology community, including the question of just how proteins fold, which would become a strong motivation for my thesis project. I am grateful to Mary and Paul LaRocca, Patrick Doyle and Marc Kastner for early exposure to laboratory work, and Dhananjay Dendukuri for his mentorship in independent scientific laboratory research when I was an undergraduate. Although he is the only non-science teacher I must acknowledge, Stephen Tapscott has perhaps had more influence than anyone on my approach to reading carefully scientific literature as well as formulating hypotheses to explain observations. His training in giving fair weights to equally valid interpretations of the same poetic text and any smidgen of his profound ability to find commonality in seemingly disparate topics seem to me at least as important to my scientific development as any other instruction, and perhaps of foremost importance.

Many friends in graduate school enabled my progress through it. Studying with Emily Gardel, Liza Morris, Keith Brown, Don Aubrecht, Michael Petralia and Adrian Podpirka enabled me to pass the first two years of coursework. Friends and flatmates from outside of my program provided a support network that was particularly valuable as I became more sequestered into my thesis work, and I am particularly grateful for the friendships of Stan Veuger, Jason Otterstrom and Kieran Wanderagula. I thank Sahand Hormoz for discussion and comments on a draft of mine on the more theoretical aspects of highly specific

spherical particle systems. Tom Kodger and Rodrigo Guerra provided advice on light scattering, chemical synthesis and many other technical aspects on my work from rotovaping to microfluidic design. Rodrigo Guerra's synthesis of poly-tert-Butyl per-fluorinated microspheres and W. Ben Roger's DNA-F108 coating of them and observations on their assembly proved a key turning point in finding a working system for my experiments. Ben Rogers provided the most direct help with my research. He provided descriptions of DNA-F108 chemistry and grafting protocols. We collaborated on the design of DNA sequences used for equilibrium assembly experiments and on fine-tuning the binding strengths of different species using his displacement strand technique. He also re-aligned the optical tweezers I used. I thank Ben, Laura Adams, Kosta Ladavac, and Zorana Zeravcic for advice on some of the political aspects of work in graduate school. Zorana's theoretical approach to the self-assembly problem influenced the decision to study 7-9 particle, 3 species systems experimentally. She provided computer renderings of many clusters that aided me in identifying them under microscopy.

In addition to Ben and Zorana, many other colleagues helped me on my way. Guangnan Meng's work on depletion-induced clusters provided a basis for my understanding of permutation entropy, and he must be credited with introducing Geomags into the lab as an invaluable research tool. Guangnan also taught me lessons in chemistry, microscopy and scientific programming. Dave Kaz and Ryan McGorty were role models for me in our young lab, and Dave Kaz originally built the optical tweezing set-up used for my cluster experiments. Thanks to Jerome Fung for working literally alongside me for many years, and for keeping me awake with his opera singing on more than a few late nights in lab. Nick Schade's quote wall kept lab fun and interesting. From Tom Dimiduk I learned to improve my passion for asking questions, and we have had many interesting discussions on topics tertiarily related to my thesis. I am very grateful to Becca Perry for helpful comments and questions on my experimental and theoretical work, including at least a partial explanation of the some of my strong binding limit results. Tricia Rogers and Brian Tilton assisted with fluorescent particle sorting experiments using my anisotropic, PDMS-glass well plates. It was also a privilege to work with Manhee Lee on the microfluidic work reported here. I am also grateful to David Weitz for the use of much of his laboratory equipment, for the work culture established by him and his group, and for an untold volume of coffee. I am indebted to Michael Brenner for collaborating with my advisor on this project, and for frequently meeting with me to discuss experimental direction and results over the home stretch, expediting the completion of my thesis while Vinny took sabbatical. Among those of other professors at Harvard, Michael's and David's multidisciplinary approaches to science have surely influenced my own.

When Vinny first described the idea to me of using biomolecules to control the cluster state of a set of microspheres, which could be visualized directly, he mentioned that I might learn something about protein folding, and I became very interested in the project right away, feeling like it was something I was especially suited to accomplish. I am indebted to him for recruiting me to Harvard (though I did not receive any other option), for originally sharing the idea of the project with me, and to him and Michael for discussing which cluster geometry to target, the octahedron. I am grateful to Vinny for occasional comments and questions on my presentations and work, for finally getting me to do optical tweezing instead of flow cytometry, for funding, and especially for trusting me to work largely on my own. I feel very lucky to have him as a role model.

Both my brother and sister have had a major impact on my path towards choosing my research project and during its course. My brother Jonathan strongly influenced my decision to apply to and ultimately

pursue education at MIT and has from a very early age shared my interests in science and technology while keeping me much more grounded and social than I would have been without him. Jon's determination in many arenas of his life inspired and continues to inspire me. I thank Jon for building and sharing big dreams with me as young boys, dreams which I was fortunate enough to keep alive in graduate school and which I still intend to realize. Kristen has provided me ample reason to stay committed to my work and, through our conversations, has refined my understanding of what exactly this activity called "science" really is. She has articulated human aspects of science that I have never heard articulated so clearly elsewhere, and helped me understand how privileged I am to experience science firsthand.

I've heard that whether or not one is to be a scientist is decided well before college or even secondary school. I thank my parents for cultivating my scientific interests from an early age. I am grateful to them for taking me to preschool at Children's Garden, where I remember doing experiments together and on my own to find out which objects and materials floated in water and honey, measuring rainfall at home, and observing plants and insects in our yard at 335 West Mountain Road, among other experiments. My parents have been wonderful role models, each in their own ways, and have given me much freedom to find my own path. When I was trying to decide what to study in college, I almost chose not to study physics, because it didn't seem clear to me how I could use it to make others happy, although I knew I would enjoy it myself. When my mom reminded me that I might need to make myself happy before I could make others happy, I decided to study physics, a decision leading ultimately to the PhD. Although I have much work to do to translate my research into a means of helping people, I believe the choices my parents helped me make were good.

Although I've thanked many people here, and there are many more I would like to thank for contributions to my personal and scientific growth, the person with whom I communicated the most during this period of my life is my wife, Prathima. Many of the reasons I have for thanking others are reasons I thank Prathima. She has shared big dreams with me, helped me understand my work and myself, advised me on professional and scientific questions, and been an inspiration, friend and friendly competitor. She maintained confidence that I would complete my research even when I had my doubts. She's helped me learn to communicate better. I've thought of her, and the hard and long hours she's put into being a surgeon and scientist, when I've wanted to quit on an experiment, then persevered. She's helped me relax and have fun outside of lab. She's provided material means of comfort. She's kept me healthy and made me laugh. Prathima, I don't know why I did not complain more about the difficulty of my scientific research over the past 8 years, but it might be because as long as I'm with you, and you're happy, so am I.

Chapter 1

Introduction

Many phenomena of scientific interest fall under the umbrella term "self-assembly." The term "self-assembly" has risen in usage frequency from $10^{-7}\%$ in 1960 to $4 \times 10^{-5}\%$ in 2008, according to the Google n-gram viewer[1]. Whitesides and Grzybowski write that definitions of "self-assembly" seem "limitlessly elastic" and the term itself "overused to the point of cliche."[2] They pose the broad and interesting question, "is anything not self-assembly?" but restrict their consideration to "processes that involve pre-existing components (separate or distinct parts of a disordered structure), are reversible, and can be controlled by proper design of the components." Whitesides and Grzybowski claim that "humans are attracted by the appearance of order from disorder" and that "living cells self-assemble, and understanding life will therefore require understanding self-assembly." I use "self-assembly" to refer to a thermal (non-driven) system's exploration of its configuration space in equilibrium (i.e. at fixed temperature). Although I suspect that understanding self-assembly may help us understand life, I am interested first in any marked differences to be found between the self-assembly phenomena of multicomponent systems and those of single or very few component systems. Living matter appears to involve the self-organization[1] of many different components on the molecular scale, perhaps in a way or at a level of complexity or specificity not observed in non-living matter.

Scientists have observed and engineered[3] chemically specific interactions in living systems. The first examples of molecular diversity and interaction specificity that I learned about are the antibodies generated by the mammalian immune system[4]. A phenomenon known as variable diverse joining (VDJ) recombination generates a vast number of different variable chains, which can be used by the immune system to identify, corral and dispose of potentially harmful pathogens (and lead to overreactions in some cases)[4]. We might also apply the concept of chemical specificity to understand the folding of individual proteins, including these antibodies themselves. The 1-dimensional sequence of amino acids appears to encode the 3-dimensional state of the protein. Proteins in known organisms at the time of this publication typically consist of about 20 different amino acids. Some amino acids interact with other amino acids preferentially. Positively charged amino acids may be attracted to negatively charged amino acids for example, but might interact only repulsively with other positively charged amino acids. Proteins, and complexes of proteins, may be thought of as systems of units of many different species interacting with each other with various degrees of chemical specificity.

[1] I use the term "self-organization" more loosely than "self-assembly." Whereas I use "self-assembly" to refer to exploration of configurations in thermal equilibrium, I use "self-organization" to refer either equilibrium or non-equilibrium exploration of configuration space, including cases in which the system's status as equilibrium or not is unknown.

More broadly, "interaction specificity" refers to the preference of particles to bond with particles of one or a few species, but less so or not at all with others. In contrast to a specific particle, which may favor particles of one or a few types, a non-specific particle can bond with particles of many different types or species. If particles of a given type bond with members of the same type, I call them "homophilic," and a set of more than one of these particles is non-specific. Particle diversity and specificity are thus entwined. The phenomenon of specific interactivity falls under the purview of condensed matter physics, and an important question is how can specificity constrain the structure of aggregrates of different types of particles. Interaction specificity is likewise important in biophysics, as many biological systems are mixtures of units interacting with various degrees of specificity. Basic questions in statistical mechanics may also tie in with questions about interaction specificity, such as how it might be related to the concept of particle distinguishability. Finally, interaction specificity is of interest to engineers. An understanding of how interaction specificity determines equilibrium structures and/or their assembly pathways may be useful in the design of self-assembling devices and systems.

I am interested in how specific interactions determine the equilibrium states of isotropic particles, as I am especially confused by things that are not spheres. Unfortunately, specifically interacting pairs of particles, even analogous macroscopic examples, such as jigsaw puzzle pieces, or locks and keys, are not spherical, at least in my experience prior to my work reported here. Specific interactions surely enable DNA bricks to assemble into unique structures[5], but the question remains whether the anisotropy of individual DNA bricks is required too. Theoretical studies[6, 7, 8, 9] suggest that it may be possible for a set of spherical particles to assemble into a unique cluster geometry if the number of unique species is not too low compared to the number of particles, and the interactions are optimized for that geometry. Experiments and theory on a single-species system of short-range, attractive colloidal spheres showed that the most likely equilibrium cluster geometries tended to be less symmetric than competing states[10, 11]. How would the self-assembly of isotropic but specifically attractive particles compare to isotropic, non-specific particles? Can a particular sphere packing, especially a packing that is unfavorable in a single-component system, form in high yield in a multi-component system? The problem then is how to synthesize, observe and make sense of a set of isotropic, specifically interacting particles.

To solve this problem, I first set out to design a system with isotropic and specific interactions. I looked for a biomolecule known to be highly specific and diverse. In addition, its interactions had to be reversible and, preferably, tunable by some control parameter, such as temperature. I also required the biomolecule to be coupled to the surface of microspheres, and the coupling chemistry to produce uniform surface coverage. So the biomolecule had to be highly specific, have controllable interaction strength, and had to be coupled to the surfaces of 1-micron particles uniformly.

The first biomolecule I studied was a family of related proteins known as Dscam, which stands for down-syndrome cell-adhesion molecule. Biologists have discovered that over 18,000 different highly specific isoforms of Dscam are generated via alternative splicing in flies[12]. They have also shown that these isoforms are highly specific. With few exceptions, a protein molecule of a given isoform binds to another molecule of the same isoform but not to a molecule of any other isoform[13]. I experimented with attaching Dscam to microspheres and performed light scattering studies of Dscam thermodynamics. My findings from these studies lead me to decide against using Dscam for my biomolecule-microsphere conjugate system.

Discouraged by my findings on Dscam, I studied and experimented with DNA instead to mediate

specific binding between microspheres. Prior physical studies of DNA-coated microspheres[14, 15] proved that DNA could be used to coat microspheres, that the interaction strengths between pairs of spheres could be tuned by temperature and salt concentration, and that the interactions could be modeled theoretically. I studied DNA design algorithms[16] and developed my own to computationally discover sets of DNA base pair sequences with specific interactions of comparable strengths. I designed a matrix of interactions to target a particle cluster, the octahedron. Finally, I experimented with different surface chemistries and observed small clusters and large aggregates to validate that the DNA-colloidal sphere system exhibited surface diffusion expected of uniform coverage.

With a highly specific and isotropic DNA-microsphere system in hand, I could at last approach the second major goal of my project: tweak the system so I could observe it in thermal equilibrium and study relative yields of different self-assembled structures. I colored particles with dyes and used fluorescence to distinguish each species. Fluorescence introduced a new problem, rapid and irreversible photo-induced elimination of the inter-particle attraction, which I would solve with an enzymatic oxygen-scavenging system. I varied temperature as well as salt and DNA concentrations to match interaction strengths of different species and observe equilibrium assembly on experimentally accessible timescales. Finally, I observed microspheres isothermally exploring different geometric configurations and tabulated the relative yields of these structures.

In the fifth year of my research, I made an unexpected experimental and conceptual discovery that opened up a theoretical avenue of research which I would pursue alongside my two original objectives. An ulterior motive for my work was the possibility of scaling up my specific sphere system, or one like it, to 10s, 100s, or more species of particles, and using it to self-assemble useful materials and devices. The discovery of a false minimum made me reconsider whether such scalability was physically possible. In what yield should I expect the true octahedron to assemble relative to its false minima? More broadly, as the number of species is increased, how does the number of unintended local minima scale relative to the number of targeted ground states? I explored this question by writing and solving partition functions, and by designing theoretical models based on my isotropic DNA-colloid system.

1.1 Background

Specific interactions may be vital to much self-assembling phenomena in natural and human-engineered biomolecular systems, but these often involve anisotropic building blocks, which may affect self-assembly differently than chemical specificity. Chemical specificity is exemplified by the antibodies generated by the mammalian immune system[17]. Each antibody differs in its amino acid sequence and 3-dimensional conformation. The adaptive immune system generates a vast diversity of antibodies to indentify foreign pathogens and clear them from the body.

An example of a human-engineered system of anisotropic biomolecules that assemble into particular structures are the DNA bricks of Ke et. al[5]. The authors designed and synthesized a 3D canvas of 459 DNA sequences. Each DNA-sequence of 32 nucleotides could bind via exact 8-base pair matches to 4 neighboring sequences. By carefully selecting subsets of the 459 sequences, the authors controlled the assembled geometry. They formed letters of the alphabet and many other 3D shapes, identifying structures via atomic force microscopy. They found yields usually in the range of 1 to 30%. They did not report details on what unintended structures may have competed with their intended states.

In-depth studies of the self-assembly of non-specific, isotropic, colloidal particles[10, 11] are of special relevance to this thesis. Meng et al.[10] experimentally investigated the clusters formed by small numbers (6-10) of microspheres. Using depletion forces induced by the osmotic pressure of a polymer, the micron-sized spheres were attracted to each other at short distances, near contact. The authors analyzed their experimental results with theory from statistical mechanics, treating each particle as identical but distinguishable. Their experimentally measured ensemble averages agreed well with calculations of equilibrium cluster probabilities taking into account vibrational and rotational entropies. The authors neglected any transient states, and compared only states that required the breaking of at least one contact to alter their connectivity. In a theoretical work, Holmes-Cerfon et al.[18] considered states intermediate to the local minima in Meng et al.'s work. Holmes-Cerfon's work describes the hills and planes, and connects the valleys, drawing a more complete picture of the energy landscape of these single-component systems of isotropic particles.

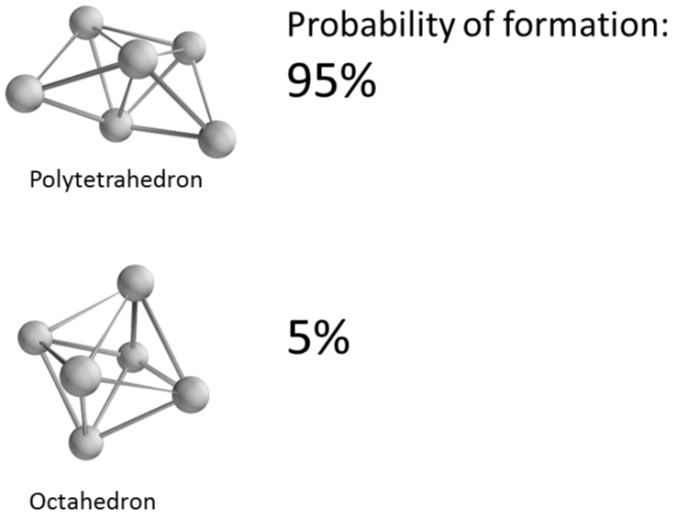


Figure 1.1: Meng et al. found the single-species polytetrahedron about 20 times as often as the octahedron.

The equilibrium results reported in this thesis may be best understood through comparison with Meng et al.'s findings[10] of equilibrium yields in a single-component system, and their dependence on the number of equivalent permutations of particles. Meng et al.'s N=6 system forms two different stable clusters, the polytetrahedron and the octahedron. Both these geometries have twelve point contacts, so we think of them as having the same potential energy. The authors observed, however, that the polytetrahedron formed about 20 times as often as the octahedron (see fig. 1.1). Much of this yield difference is accounted for by the number of rotationally unique arrangements of the particles. The number of arrangements of the particles that cannot be rotated into one another is 30 in the case of the octahedron, but 360 in the case of the polytetrahedron (see figures 1.2 and 1.3). In general, the number of distinguishable arrangements of less symmetric clusters are greater than the number of distinguishable arrangements

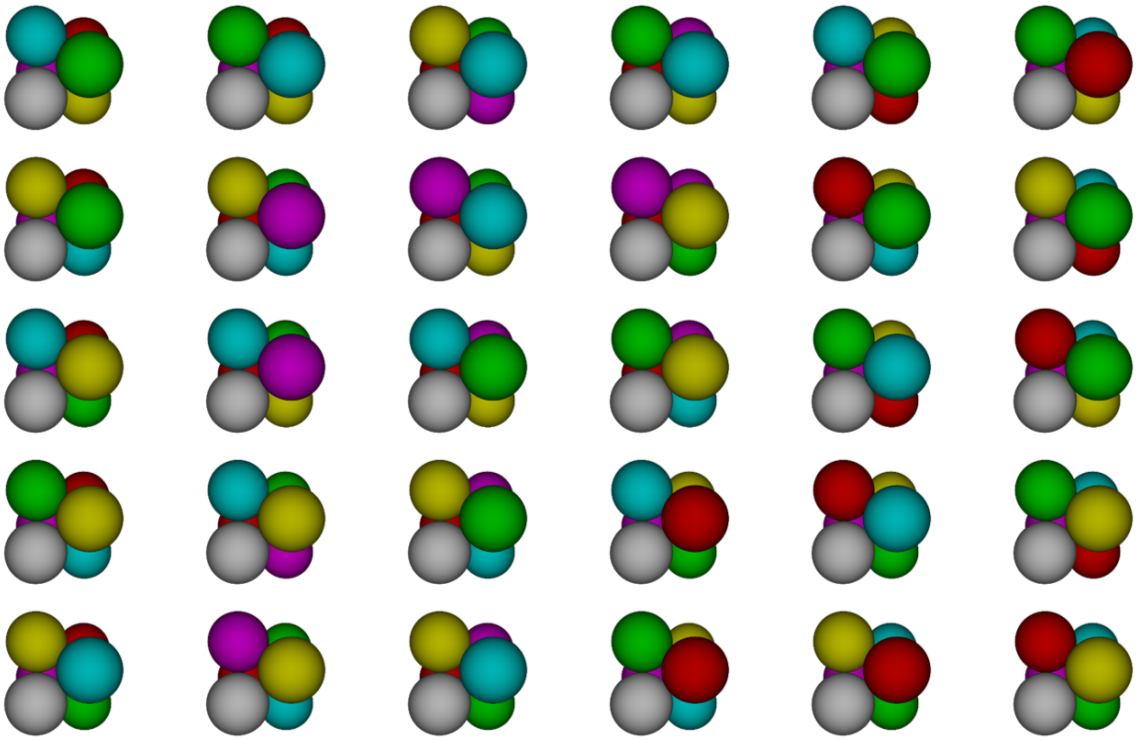


Figure 1.2: 30 arrangements of the single-species octhahedron that cannot be rotated into any other. Image courtesy of Guangnan Meng.

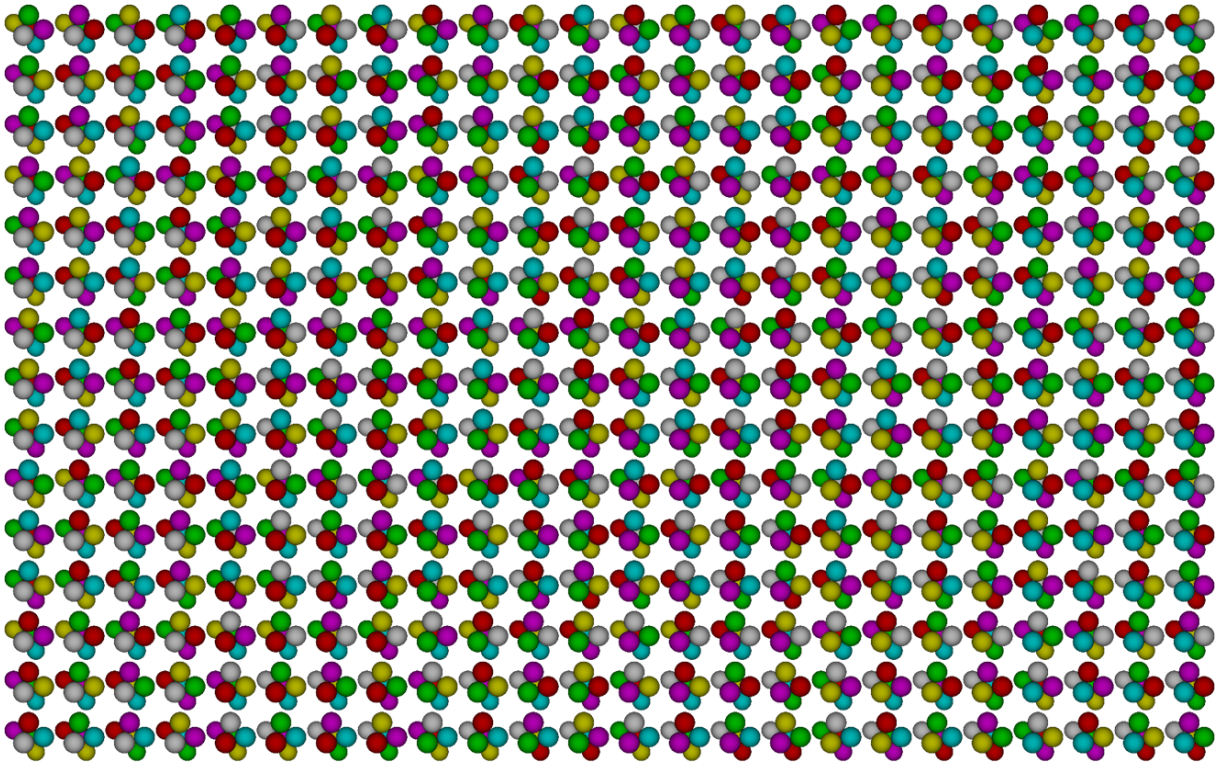


Figure 1.3: 360 arrangements of the single-species polytetrahedron that cannot be rotated into any other. Image courtesy of Guangnan Meng.

of more symmetric clusters, and the ratio of these numbers account for much of the associated yield differences in single-component systems[10].

Various groups have theorized about multi-component systems of chemically specific, isotropic particles, and their assembly into configurations that might be expected to be unlikely for non-specific particles. Lukatsky et al.[6] studied a 2D system of isotropic particles of 4 different chemical species interacting via nearest and next-nearest neighbor interactions, noting the similarity to nanoscale spheres coated with DNA sequences, linked by soluble strands. They considered a set of interactions between the four particles that favor a particular square lattice and found that, if the interactions in the intended tiling are energetically strong compared to any other interactions, their tiling self-assembles in Monte Carlo simulations.

Others have studied theoretical models of specific spheres arranging into 3D clusters. Licata and Tkachenko, in one study[7], perform a Langevin dynamics simulation of a finite number of beads attached via specific springs. The beads start in an initial, random state and over time the springs pull the beads together. They find that whether the beads assemble into the lone optimal configuration depends on the repulsive term in their interaction potential. If the repulsion is soft, the beads do not jam, but if the repulsion is hard sphere-like, they tend to jam in a configuration other than single energy-minizing configuration. In another theoretical study[19], Licata and Tkachenko consider clusters of spheres coated with DNA that link to soluble DNA sequences. They consider mixtures of many spheres and soluble sequences. In their model, each soluble sequence can bind a finite number of spheres, and they argue that the spheres would minimize their second moment about each soluble linker. The soluble sequence has N sticky portions, each one specific for a particular species of spherical particles. They found that the intended clusters, up to labelling degeneracy, could form in high yield for a range of concentrations of particles and soluble strands. Hormoz and Brenner studied theoretically[8] the self-assembly of an 8-sphere cluster. They choose a highly symmetric cluster that formed in less than 1% of the time in experimental studies of a single-component system[10]. They report the cluster equilibrium probability as a function of alphabet size (number of species), number of favorable interactions, and standard deviation of the distribution of energies of excited states. They find their target tetrahedral cluster assembles in higher yield at 8 species compared to 2 species, and that the yield approaches 70% compared to the distribution of excited states. Zeravcic and Brenner report[9] a specificity matrix making a 6-particle octahedron cluster the ground state, using 3 different species. They devise and simulate the self-replication of this cluster.

Tkachenko theoretically studied the hierarchical assembly of nanoparticles coated with a small number of DNA sequences[20]. He considered the angular and positional displacements of neighboring spheres as they depend on the size of the DNA sequences that bridge them. He performs a rigidity analysis, comparing local constraints, depending on average coordination and particle numbers, to the total degrees of freedom, which depends instead on dimension and particle number. He thus finds a lower bound on the number of different particle types required for assembly of the target. He then analyzes the thermodynamics and, considering a sequential assembly scheme, also derives bounds on the particle concentration and width of the distribution of different DNA melting temperatures. Halverson and Tkachenko simulated systems of spheres with directional and specific interactions[21], similar to the system Tkachenko considered theoretically alone [20]. They studied assembly of soft cubes of 6 particles as a function of number of different particle types and number of different specific interactions per particle.

Experimental studies of the assembly of DNA-coated colloids stretches back at least to the nanoparticle aggregates made by Mirkin et al.[22] and the small clusters made by Alivisatos et al.[23] in 1996. Alivisatos's work on clusters was likely non-equilibrium, as were subsequent studies on DNA-coated nanoparticles[24, 25] and microparticles [26, 27, 28]. Others studies of DNA-coated microspheres cover temperature-reversibility[29], defect particles with slightly mismatched DNA incorporating into a crystal of particles of even more slightly mismatched DNA[30], and crystal-crystal phase transformations[31]. The kinetics of a DNA-coated mircrosphere diffusing along a 2D DNA-coated surface [32] were studied as well as kinetics of binding and unbinding of a pair of DNA-coated particles in a line-optical tweezer [33]. The thermodynamics of DNA-microsphere binding have been theorized and compared with 2D aggregation experiments [34] as well as with particle tracking measurements of the pairwise interaction potential[14, 15].

In addition to the theory on isotropic and specific particles, and the experimental work on specific, DNA-functionalized spheres, an experimental observation of the self-assembly of isotropic and specific particles is needed. Are DNA-coated microspheres truly isotropic, or are their surfaces heterogeneously coated? If the number of species is comparable to the number of particles, might an experimental system self-assemble into a structure that is unlikely in a fewer- or single-component system? How do the energy landscapes of highly specific particles compare to those of less or non-specific particles? Could self-assembly of N particles of N different species of DNA- or Dscam-coated spheres become a scalable method for engineering useful technologies?

I set out to answer these questions by, first, engineering an isotropic, highly specific, biomolecule-coated microparticle system and, second, studying this system exploring its phase space in equilibrium. To achieve the first goal, I studied the physical chemistry of Dscam dimerization using light scattering techniques. Then I explored the affect of different surface chemistries on the isotropy and surface diffusion of DNA-coated microspheres. Next, I made sure the system could be imaged in fluorescence, and one species of particle could be discerned from another. Finally I observed this system transitioning through various cluster geometries in equilibrium. I measured the amount of time the cluster spent in various geometries and studied how the yields depended on the number of species relative to the number of different particles. I briefly studied non-equilibrium organization of specific particles and applied basic statistical mechanics to ask whether, in theory, many specific spheres could condense into their ground states without being arrested in a plethora of geometrically frustrated conformations.

Chapter 2

Dscam Studies

2.1 Introduction: Molecular Diversity and Specificity of Dscam

Dscam is a fascinating example of molecular diversity and recognition specificity. The gene for it has 3 hypervariable domains, one with 12 different possible base pair sequences, another with 48 and the last with 33[12]. One exon of each of the 12, 48 and 33 possibilities is transcribed into an mRNA that encodes a protein. ELISA-based assays suggest that proteins of each of the $> 95\%$ of the $12 \times 48 \times 33 = 19008$ different possible isoforms exhibit homophilic dimerization, but little or no heterophilic binding[13]. This diversity and molecular specificity is required for proper growth of the brains of flies[12]. Dscam is a transmembrane protein, and its highly variable domains are exposed on the outside of the neuron cell membrane. Thousands of different isoforms are required for these neurons to exhibit self-avoidance[35], i.e. grow axons that form connections to different neurons as opposed to themselves.

My advisor drew my attention to Dscam, as its extraordinary molecular diversity and specificity might be harnessed for conferring colloidal spheres many different but highly specific interparticle interactions. We visited Dietmar Schmucker at his office and lab at the Dana Farber Cancer Institute and discussed our plan for studying self-assembly of colloidal spheres. A molecular biologist named Roland Puettman-Holgado gave me samples of microspheres that had been coated with a Dscam isoform, and which aggregated, perhaps because of the isoform-isoform specific homophilic binding. Dscam seemed like a promising candidate to control the specific interactions of our microspheres.

To determine if I could use Dscam to study the self-assembly of isotropic but specific particles, I studied the thermal reversibility of the Dscam dimer with light scattering. Light scattering off proteins can be used to indirectly measure the protein size[36, 37]. An experiment showing that Dscam-coated beads aggregated was promising, but it did not prove the bead interactions were truly specific, that they were homogenously covered, nor that the bonds were reversible. In a pilot experiment, I used dynamic light scattering[38] to study the effect of heat on the bead aggregate size, and found that the aggregates did not dissociate. I turned my attention to the protein itself, because the beads and Dscam were bound to each other by an IgG-Protein A interaction, which may have contributed to irreversible aggregation, but could possibly be avoided with another coupling chemistry. So I studied Dscam in different salt conditions, with both static and dynamic light scattering, in an attempt to understand and control the thermodynamics of its specific interactions.

2.2 Methods

2.2.1 Protein Expression, Purification and Quantitation

I received recombinant yeast strains, one expressing isoform 1.30.34 and another expressing isoform 7.27.25, from Dietmar Schmucker's lab at Dana Farber. The Dscam gene expression was controlled by a methanol-inducible promoter. After growing the yeast at 32° C with a laboratory shaker in a glass flask (covered with alumninum foil but with a 2-3 holes poked through) for 1 or more days, I induce Dscam expression by addition of a small quantity of methanol to the growth media. After giving the yeast at least a few hours or overnight to express the protein and secrete it into the extracellular media, I remove the yeast and media from the glass flasks and centrifuge the cells at 1000g or more to pellet them. Then I pour off the supernatant for purification.

To start the purification of protein from supernatant, I first vacuum-filter the supernatant through a 0.2 micron PES filter. This filtered supernatant could be stored at 4° C for weeks or longer awaiting the next step. I purify the Dscam from the filtered extracellular medium using a Nickel or Cobalt resin. The recombinant protein was engineered to have a 6xHIS tag, which coordinates onto the metal ions attached to the resin. I wash the resin thoroughly and elute with low pH buffer. The eluent fractions containing the highest quantity of protein are typically visibly yellow-brown.

To quantify the concentration of protein purified, I use UV/Vis spectroscopy. I ran one Bradford assay to obtain the absorption coefficient of the protein, which was a value typical of proteins. After each purification, I obtain the absorption at 260 nm. I often obtained concentrations in the 1-10 gram/L range, in a volume of about 0.5-1.5 milliliters (mls).

2.2.2 Light Scattering Methods

To perform a light scattering experiment on Dscam, I dilute the stock solution, in either deionized water or phosphate buffered saline (pH 7.5, 10 mM) with added NaCl, to a desired concentration in a glass vial. Then I insert the vial in a bath of toluene, or other fluid that matches the refractive index of the glass vial. The light scattering instrument I use sends a laser reflecting off some mirrors, through a beam splitter at which point the alignment and intensity of half the split beam is measured, and the other half is focused into the index matching fluid and sample container. The experimental setup is shown and depicted schematically in figure 2.1.

The wavelength of light I use is 532 nm, which is much larger than the size of the protein. Electron micrographs depict Dscam to be less than 10 nm in its longest dimension[39]. I also make sure the sample is dilute enough that there is no multiple scattering. This regime of low concentration, and size ratio in which the scattering wavelength is much larger than the scatterer, is known as Rayleigh scattering.

I simultaneously obtain two distinct kinds of data from the same measurement: static and dynamic. Static light scattering is a measurement of the average intensity of the scattered light. Static light scattering measurements must be taken of both the sample and the buffer without the protein. The protein-free scattered intensity must be subtracted from the protein solution's scattered intensity. Dynamic light scattering is a measurement of the autocorrelation of the intensity of the scattered light over a range of "waiting" or "lag" times[38]. I show a sample trace of the raw data and how it can be broken down into both static and dynamic measurements in figure 2.2.

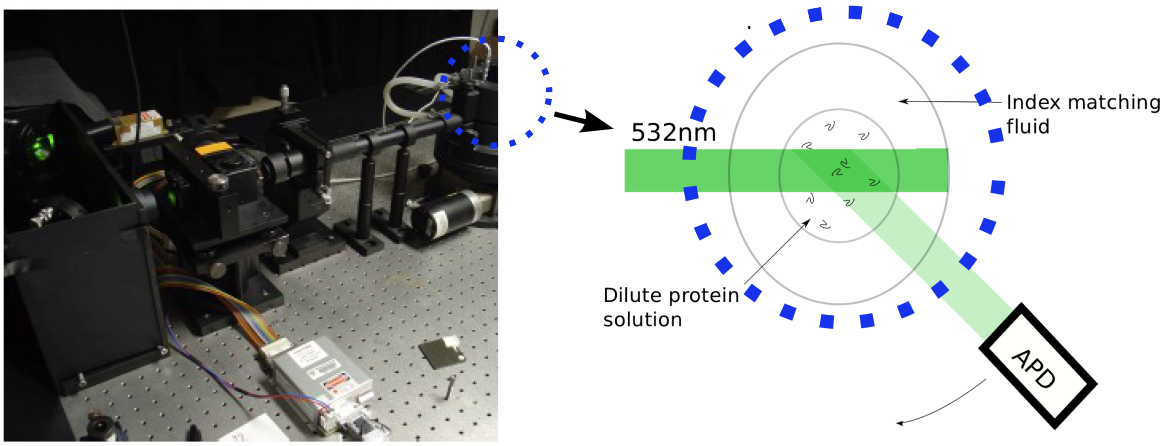


Figure 2.1: Light scattering instrument and schematic

The light scattering instrument is equipped with a fast correlating board so that correlations in intensity can be measured/calculated down to about 25 nanoseconds. The diffusion of the proteins in the scattering volume randomizes their positions relative to each other. Some of the arrangements of the proteins lead to more constructive interference or more destructive interfence of the laser light than others[38]. In arrangements that produce more destructive interefence, the intensity dips slightly below the average intensity measured in static light scattering, and it measures above the average when the arrangement produces more constructive interface than average.

2.2.3 Static Light Scattering Data Analysis: Zimm Plots

The time-averaged intensity of scattered light measured in a static light scattering experiment can be used to gain physical information about individual protein molecules and interactions between them[36]. In this Rayleigh scattering regime, in which the particles of interest are much smaller than the wavelength of light, the scattered intensity is proportional to the number of scatterers times the square of the mass[36]. I measure the scattered light intensity at a variety of angles at each in a series of different concentrations of protein. I extrapolate a value for scattered intensity at zero degrees, i.e. in the forward scattering direction. Then I find a line that minimizes the least squared residuals to these extrapolated intensity values. This method is known as a Zimm plot[40], and from it one gains a measurement of the molecular weight and 2nd viral coefficient.

2.2.4 Dynamic Light Scattering: Cumulants and Constrained Regularization

The autocorrelation of the scattered intensity typically starts out at a value indicating perfect correlation at very short lag times, and decays to lower or perfectly uncorrelated value and longer lag times. The lag time at which the autocorrelation decays to $1/e$ of the initial value is often used as a characteristic lag time. There are a variety of ways to extract information about the particle hydrodynamic radius from the

autocorrelation function. The choice of method may be influenced by any a priori information about the sample[38].

If the sample is expected to have a monomodal size distribution, it may be possible to fit the autocorrelation function with a cumulant expansion[38]. A 2nd cumulant expansion may be useful for obtaining information about a distribution that is monomodal but polydisperse. I used a 2nd cumulant fit to infer the information about the protein diffusivity at a variety of detection angles. This fit gives a value known as gamma, which is related to the diffusivity times the square of the q (scattering) vector[38]. Extrapolating to zero angle (zero q-vector) should give $\gamma = 0$ in the case of spherical solute, and a non-zero value for anisotropic molecules. The slope gives the diffusivity. I relate the diffusivity to the effective hydrodynamic radius using the Einstein-Stokes-Smulochowski relationship. I repeat the procedure at different protein concentrations.

If the sample is expected to have a multimodal size distribution, I use constrained regularization to fit the correlation function[38, 41]. The inverse problem of finding a size distribution given the exponentially decaying correlation function is "ill-conditioned." There may be many different solutions that fit the data equally well. Constrained regularization is a computational method that minizimes not only the least squares residuals to the data, but also another parameter, such as the number of peaks or the smoothness of them.

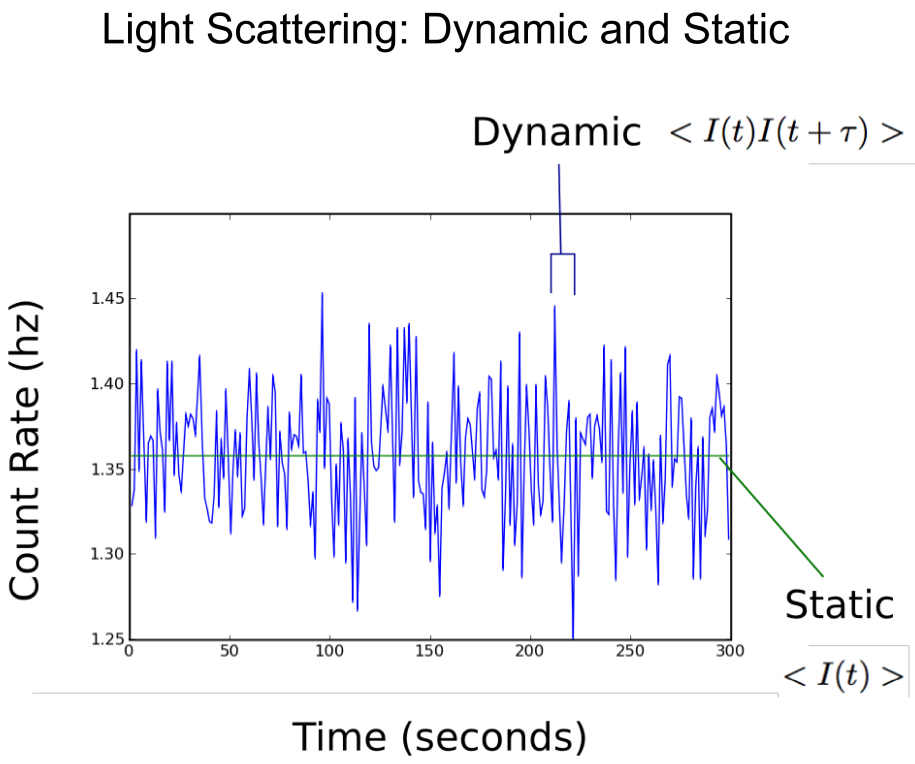


Figure 2.2: Photon count rate raw data. Dynamic light scattering measures the autocorrelation function of this data, down to 25 nanosecond (ns) lag times, whereas static light scattering measures the average intensity.

2.3 Results

2.3.1 Static Light Scattering

My first Zimm plot covered a wide range of concentrations, and was intriguingly non-linear (see Fig. 2.3). I report the data as $\frac{Kc}{R}$, which is inversely proportional to scattered intensity. K is a material constant, c is the protein concentration, and R is the Rayleigh ratio[36]. The higher the value of $\frac{Kc}{R}$, the more the protein scatters, and the higher the molecular weight and/or the more associated it is. Scattered intensity of Dscam in PBS and Dscam in water were nearly identical at high protein concentration, and protein-protein interactions were neglible. Scattered intensity of Dscam in PBS vs. water differed in slope and magnitude at low concentrations.

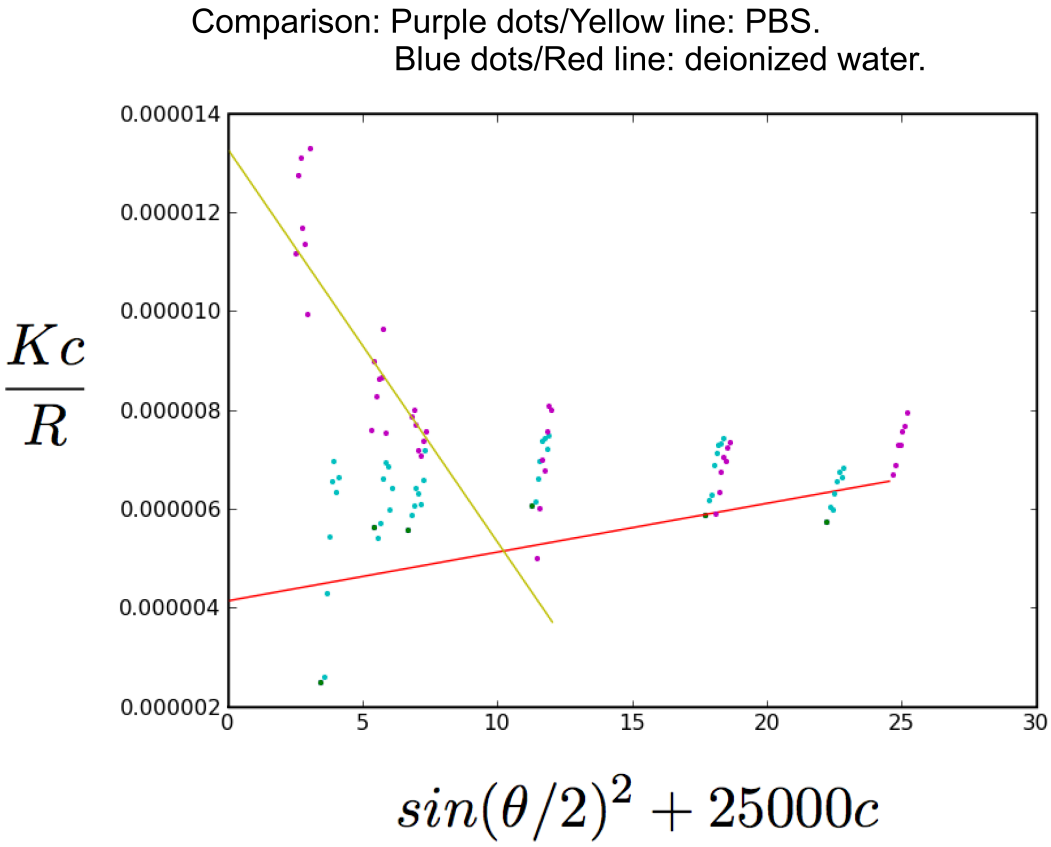


Figure 2.3: Dscam Isoform 1.30.34 IgG1-8 light scattering data. Each cluster of points corresponds to data taken at one concentration and various angles (45-120°). $\frac{Kc}{R} \propto \frac{1}{I}$. Low scattered intensity suggests monomers, and high intensity suggests dimers.

Ionic Strength	Molecular weight (kDa)	Hydrodynamic radius (nm)
175	103	7
25	114	7
0	140	9

Table 2.1: Summary of molecular weights of Dscam IgG1-8 1.30.34 obtained via static light scattering and hydrodynamic radius obtained via dynamic light scattering.

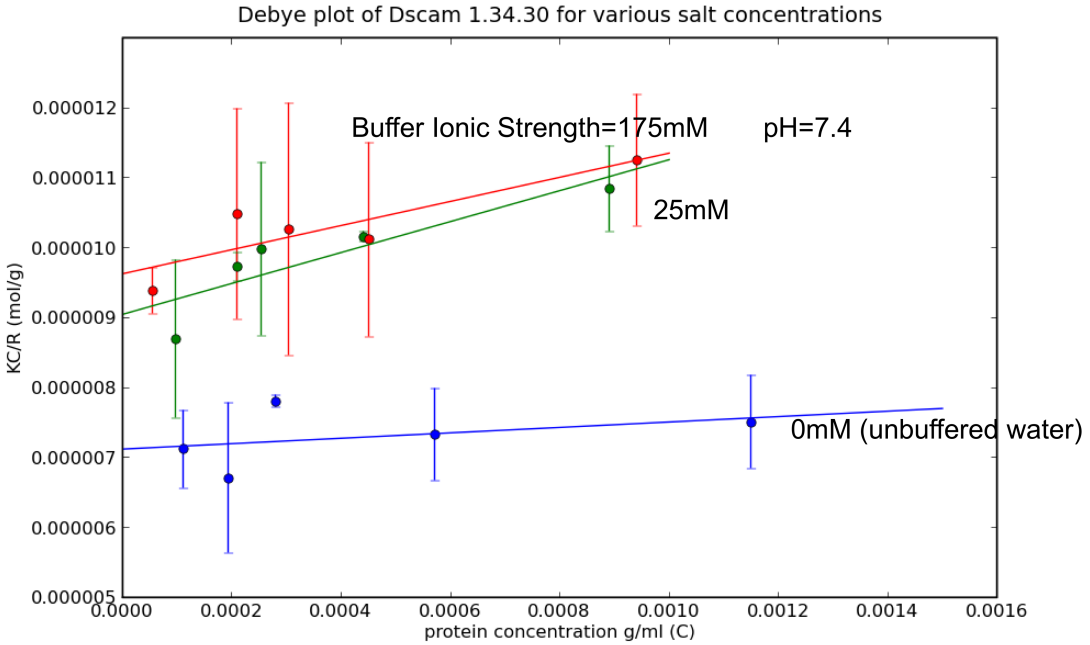


Figure 2.4: Salt and/or pH buffer decreases scattered light intensity from Dscam IgG1-8 1.30.34, but increases the second virial coefficient. Fits are to $\frac{Kc}{R} = \frac{1}{M_w} + 2A_2c$. Molecular weight of the monomer is 110 kDa.

After the preliminary concentration series, I gathered more data points at the lowest concentrations where the slopes were linear. I also tested two different salt solutions, one with 175 mM NaCl, the other with 25 mM NaCl. As shown in figure 2.4 and table 2.3.1, the 25 mM and 175 mM salt solutions scatter nearly identically, and they tend to scatter more than the salt-free solution. By extrapolation to zero concentration, it is possible to obtain an estimate for the protein molecular weight. The values I obtained via light scattering compare well to the known protein molecular weight of 110kDa. The extrapolations for 175 mM gives 103 kDa, 25 mM gives 114 kDa, and the deionized water solution gives a slightly larger value of 140 kDa.

2.3.2 Dynamic Light Scattering

My DLS results gave hydrodynamic radii in the range of 7-9 nm (see table 2.3.1). These results are consistent with electron micrographs of Dscam IgG1-8[39]. The raw data, cumulant fit and hydrodynamic radius can be seen in figure 2.5. The hydrodynamic radius was 9 nm is deionized water, but 7 nm in PBS.

In another set of experiments, I investigated the thermal stability of Dscam. I fixed the protein concentration and experimentally acquired the intensity autocorrelation function after heating the sample to 320K. For the first 20 minutes, the correlation function did not appear to change. After 20 minutes, the decay of the correlation function visibly decreased in steepness. I used a constrained regularization method to try find the best fits to these correlation functions[41]. As shown in figure 2.6, the distributions I found using this method were bimodal.

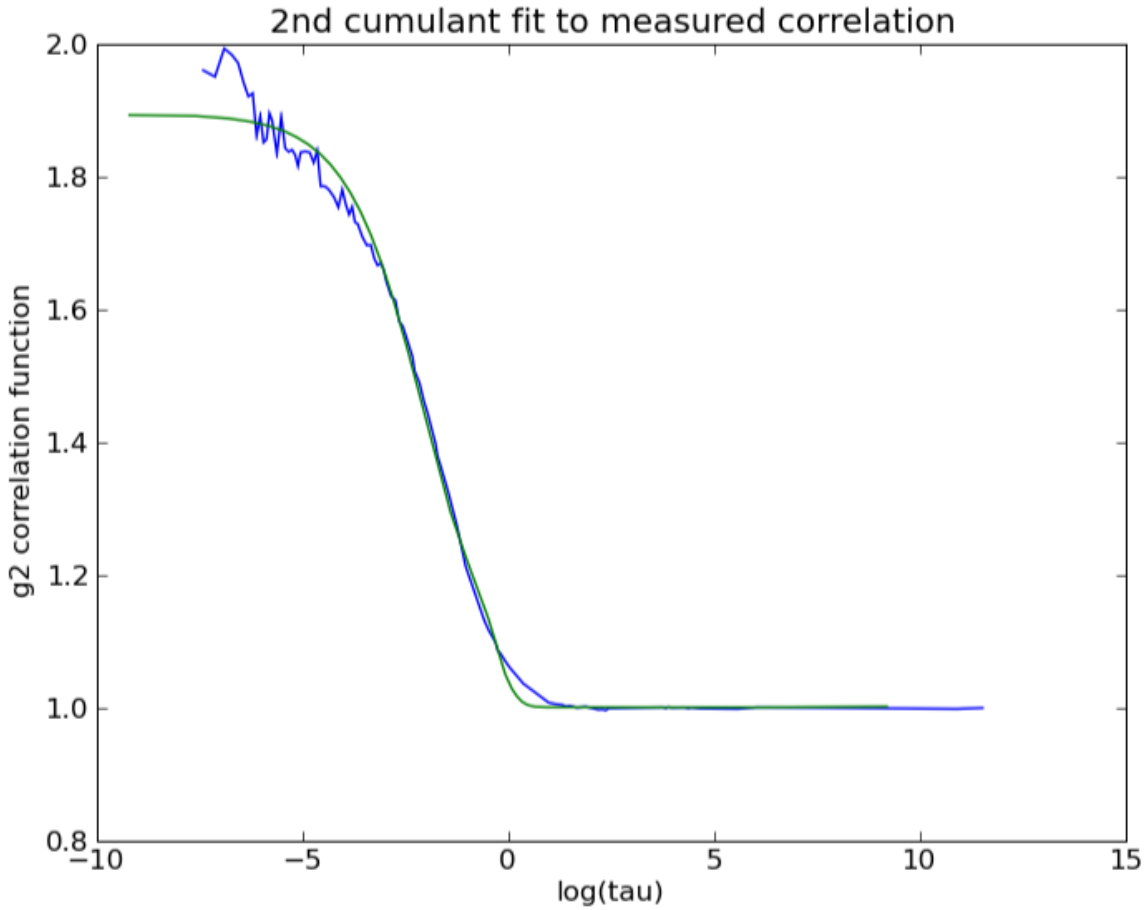


Figure 2.5: Example intensity-intensity correlation function (blue) and 2nd cumulant fit (green).

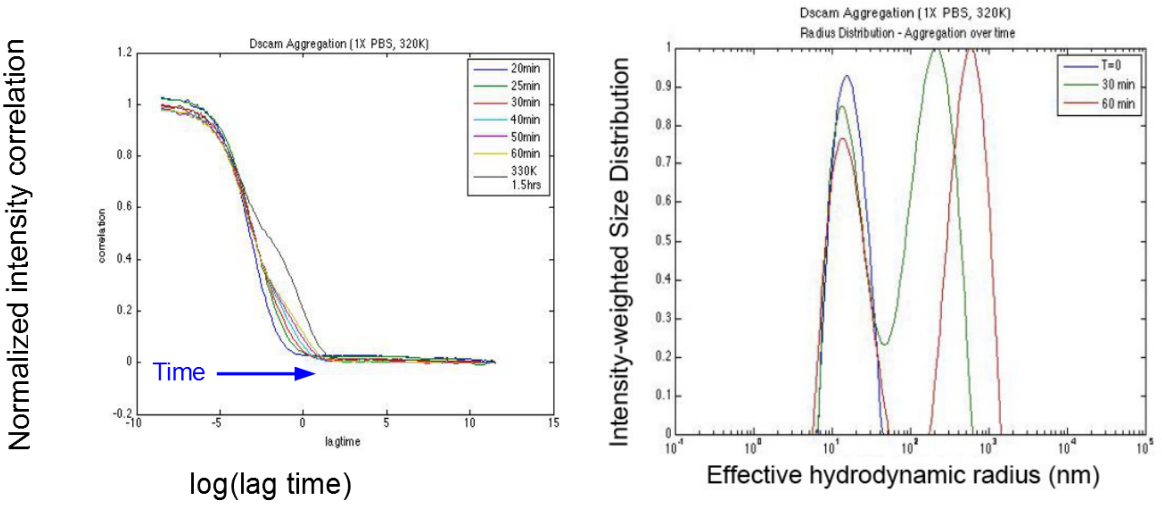


Figure 2.6: Left: Time series of autocorrelation functions at high temperature, showing longer tails after longer periods of heating. Right: Constrained regularization shows that the protein forms 100-1000 micron sized aggregates upon heating after 10s of minutes.

2.4 Discussion: Dscam Dimers and Dropping Them

2.4.1 Static Light Scattering

The scattered intensity of light from the proteins changes dramatically as the concentration of the protein changes. If the system is in equilibrium between associated and dissociated states, we might expect the scattered intensity to be non-linear over a range of concentrations. The slope of the $\frac{Kc}{R}$ vs. concentration curve is related to the interactions between solute molecules, to their second virial coeffcient[36]. A possible explanation of the data is that the protein is a dimer at lowest concentrations in deionized water but a monomer in low concentrations in the salt solution. At highest concentrations the equilibrium shifts to dimers in both high salt and no salt solutions. Ionic bonds, between positively charged residues on one protein molecule and negatively charged residues on the other, and vice versa, may hold the dimer together. It appears that salt concentration may be used to control Dscam dimerization.

2.4.2 Dynamic Light Scattering

To explain the temperature-based DLS measurement, I hypothesizee that the proteins were denaturing and forming large aggregates, perhaps by binding non-specifically through various hydrophobic domains that were exposed by the heat. This suggests that very large aggregates were forming and growing over time, and fewer of the proteins remained in their native folded states. Perhaps further study at more moderate temperature is needed to say for sure, but these results did not bode well for the use of Dscam as a means of mediating colloidal self-assembly.

Seeing these light scattering results, I decided against using Dscam to control colloidal self-assembly. Salt is not a very useful control parameter compared to temperature, as it would require sample chambers capable of dialysis. Temperature is much easier to use, but Dscam is unstable at moderate to high temperatures. Moreover, each coupling chemistry I looked into did not seem like it would lead to a uniform coating of protein on each bead. Any covalent coupling chemistry I found would target more than one residue along the protein backbone, making the surface non-uniformly coated. The proteins would have different attachment sites and orientations. Although I considering looking into other coupling strategies, the thermodynamic results were especially disconcerting. I decided not to use Dscam to mediate microsphere-microsphere interactions. Instead, I decided to use DNA.

Chapter 3

DNA-coated Microspheres for Studying Specific Interactions

3.1 Introduction

Based on work by others[14, 15], DNA-coated microspheres seemed an obvious choice after Dscam proved to be thermally unstable. I designed DNA-sequences using various computational approaches (see Appendix C), such that the interactions were specific but thermodynamically similar. I grafted DNA to F108 pluronic surfactant (BASF) and coated polymeric microspheres with this DNA-F108 surfactant molecule (see Appendix D). Next, I had to confirm that the system really behaved like isotropic particles, and that each species could be distinguished.

3.1.1 Surface Diffusion and Fluorescence Stability

First, I required such particles to "roll" or "slide" (ie. laterally diffuse) along each others surfaces, which itself would require the particles to be uniformly and densely coated with DNA. If the particles were coated heteorogeneously, such motion might be inhibited. The particles would not really be isotropic. They could get stuck in patches of DNA denser than the average surface density. A parameter that I had experimented only slightly with, the microsphere surface groups, turned to be of great importance towards achieving rapid surface diffusion over wide temperature windows.

Second, the particles would have to be dyed different colors without affecting the interactions. I found, however, that aggregates of my fluorescently dyed particles would irreversibly break apart after exposure from a fluorescent light source. To understand this problem, I first confirmed that it did not occur in particles without dye. Then I measured how many seconds of lamp exposure it took for the clusters to break apart into individual particles. Then I asked whether the total time until disaggregation depended on how short the bursts of exposure were. I found that the total exposure time until disaggregation was the same whether it was broken into short, ~ 0.5 second, bursts or continuous illumination of minutes. These results were consistent with a hypothesis of indirect DNA damage via oxygen radicals excited by the fluorophores.

In this chapter, I describe how I settled on a chemical system of DNA-microspheres that diffuse freely across each others' surfaces, and which is compatible with fluorescence imaging while maintaining colloidal stability and specific interactions. I controlled the surface chemical groups of the colloidal

spheres, studying both carboxylate-modified latex (CML) and sulfate. I directly imaged the surface diffusion of bound particles. Then I studied the stability of clusters of dyed particles in fluorescence with and without a glucose-oxidase/catalase oxygen-scavenging enzyme system.

3.2 Methods and Materials

3.2.1 Bulk Experiments

I perfomeded these bulk experiments in collaboration with W. Ben Rogers. We suspend one or two species of beads at 0.1-0.5 % v/v in 1X Tris-EDTA (TE) at various NaCl (EMD, 99.0%) or $MgCl_2$ (Mallinckrodt, 99.1%) concentrations. We allow all particles to aggregate and then observe them using an inverted microscope (Nikon TE-2000). In some cases, for instance in single-component systems, we observe differences in assembly over the course of the first few hours. We may give the samples about 12 hours to fully equilibrate before we take images. We take still micrographs using a Nikon camera or video using an Edmund Optics USB camera.

3.2.2 Small Cluster Experiments

We suspend beads at about 0.01% v/v in 1X TE and NaCl or $MgCl_2$. Usually we can easily find four particle chains, i.e. A-B-A-B, or can observe the binding of single particle to a trimer, making the open A-B-A-B configuration. We then observe the open four particle chain, watching for alterations in bond angle (analogous to cis-trans alteration in butene) and in some cases, loop closure. We record video using an Edmund Optics USB camera.

3.2.3 Preparing Samples for Fluorescent Imaging

When exciting dyed beads with laser light or a mercury arc lamp, we observe that fluorescently dyed DNA-micropheres, initially bound to each other, irreversibly dissociate due to the photo-excitation. To reduce or eliminate the effect of the light source on the interactions of the DNA-coated beads, I suspend the beads with the glucose and glucose oxidase/catalase mixture commonly used in single molecule experiments[42, 43, 44]. I prepare stock solutions of glucose oxidase (20 mg/ml) (Sigma) mixed with catalase (3.5 mg/ml) (Spectrum) in 10 mM Tris pH 8.0, 125 mM NaCl. I prepare stock solutions of D(+) glucose (Sigma 99.5%) at 450 mg/ml in DI H_2O. I store these stock solutions in 20-100 microliter (ul) aliquots at -20 $^\circ$ C. Just prior to the microsphere experiment, we mix 1 ul of glucose oxidase/catalase stock with 8 ul of 1X TE salt solution, 0.1 of 0.4 CML Opti-link (Thermo) polysytrene beads, and finally add 1 of glucose stock. CML polystyrene beads clump together in the presence of the enzymes. I think these CML beads preferentially adsorb the enzymes, slowing or stopping the enzymes from interacting with the DNA-F108 coated sulfate beads, which would cause them to stick irreversibly to each other or the chamber walls. We finally add 1 ul of this glucose oxidase/catalase/glucose/CML mixture per 50 ul of DNA-coated microsphere solution.

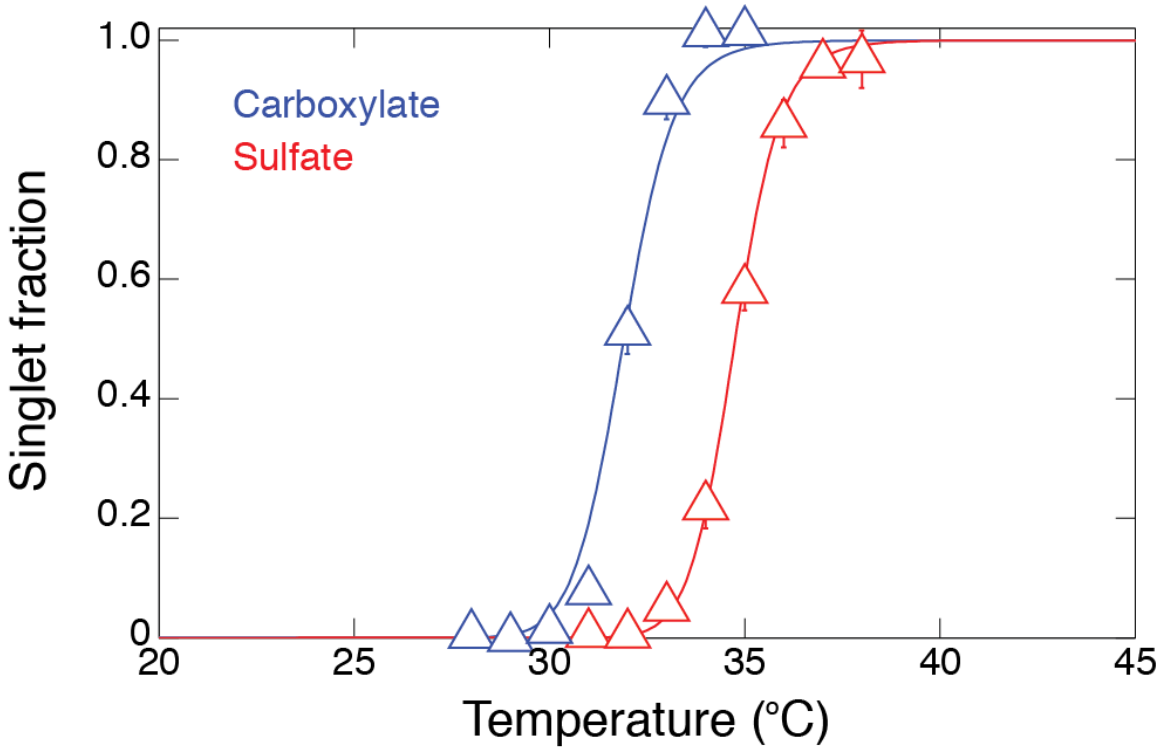


Figure 3.1: Melting Curves of DNA-coated, carboxylate-modified and sulfate beads. Figure courtesy of W. Ben Rogers.

3.3 Results

The CML and sulfate beads each have similar melting profiles and temperatures. The temperature at which half the CML beads are in singlets is 32$\pm$0.1° C and that temperature is 35$\pm$0.1° C for sulfate. Both sets of beads exhibit a sharp transition from all particles being bound to all particles being unbound over a window of 3$\pm$0.1° C. This similarity shows that the thermodynamics of binding does not depend on the chemical moieties (carboxylate brush or sulfate) at the particle surface. Based on these results, the adsorption yield of the DNA-F108 to the surface likely does not depend strongly on the surface groups. The melting temperature also represents an upper bound for temperatures at which surface diffusion may be observed, as particles tend not to stay in contact at higher temperatures.

We then examine in detail large aggregates of 100s-1000s of beads, noting any differences in their features which may be signatures of differences in surface diffusion (see figure 3.6). These aggregrates assemble over 12-24 hours at 22 $\pm$0.1° C. At this temperature, beads are very unlikely to unbind. Working at such a low temperature should make any differences in surface diffusion more apparent than at higher temperatures, where beads might move along each others surfaces by unbinding completely and diffusing then rebinding.

We find that CML form loosely connected networks. Out of 130 in-focus 1 micron CML particles, I count about 7 particles with only a single bond and 24 with only two bonds. The mode of the CML contact distribution is 3 and the mean is 3.5 (see figure 3.7). It is clear that many of these particles have not diffused along each others' surfaces to find positions in contact with at least 3 other particles. Although we do not observe new contacts forming, we observe brownian movement of the thin chains of particles within the network.

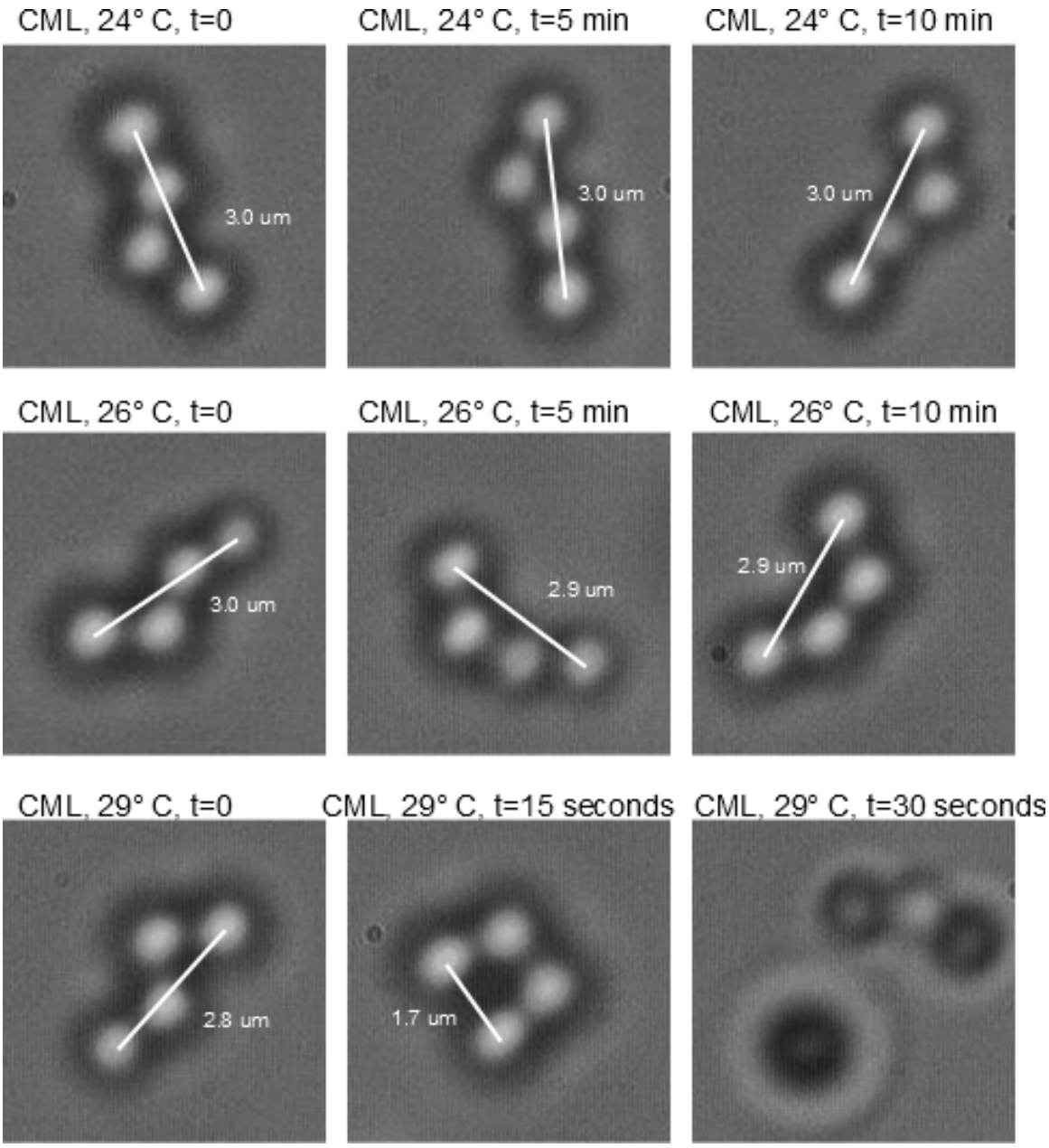


Figure 3.2: a-c, CML at 24,26 and 29 ° C

Compared to CML, aggregates of sulfate beads are compact. Out of 121 sulfate beads, we do not observe any with only one contact. The mode of the sulfate particle contact distribution is 6 and the mean is 4.9 (see figure 3.7). Although the entire aggregate might show some slight brownian motion, the beads within the network appear to be fixed relative to each other, not fluctuating as in the thin chains of the CML aggregates.

To determine the range of temperatures over which surface diffusion is observable, we perform microscopy on 4 particle chains with temperature control. We find the chains in an open A-B-A-B conformation. We pay special attention to the distance between the centers of the particles at ends of the chain. This distance can in principle vary from a maximum of 3 particle diameters to a minimum of 1.

A small cluster of 4 CML beads reveals local rearrangement over a narrow, $1 \pm 0.1^\circ$ C, temperature range. Over ten minutes, at 22, 24, and 26 ° C, no definite change in chain conformation is apparent:

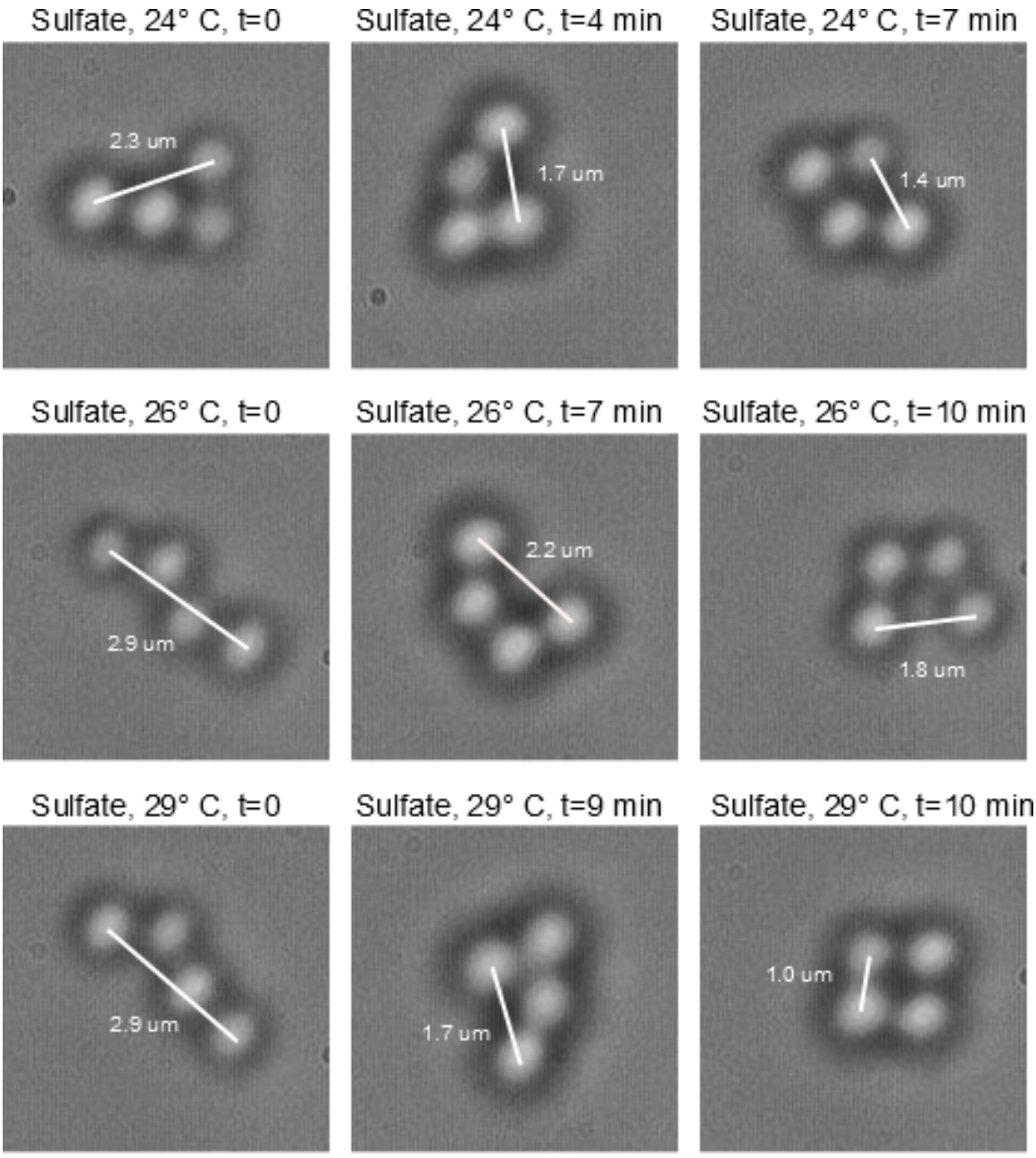


Figure 3.3: a-d, sulfate at 24,26 and 29 ° C

the distance between the particles at the ends of the chain appears unchanged at 2-3 full diameters. At 29 ° C, rapid local rearrangement does occur, as the chain of 4 particles moves from an extended state to a compact state. However, before a 4th bond can form, a particle detaches from the chain. At this temperature, bond breaking events accompany local rearrangement (see fig. 3.2). These results are repeatable. We have observed many other chains of CML beads making no change in bond angles over days at similar temperatures.

In comparison, chains of 4 sulfate beads alter their conformations over a wide range of temperatures, from 24 to 29 ° C. Over ten minutes at 22 ° C, no definite change in chain conformation appears: the centers of the particles at the ends of the chain appear 2-3 full particle diameters away from each other at all time points. At 24, 26, and 29 ° C, however, the 4-sphere sulfate chains rearrange. The center-to-center

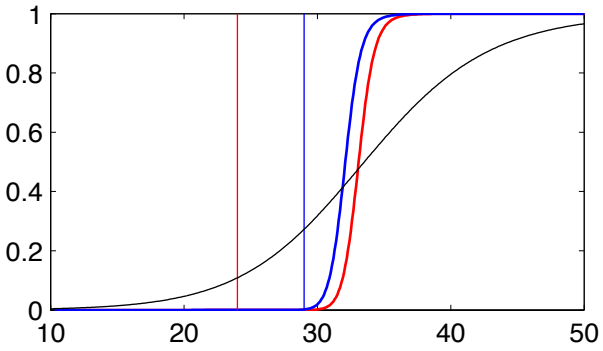


Figure 3.4: Melting curves with vertical lines at temperature of onset of observed surface diffusion, sulfate in red, CML in blue. Predicted DNA-DNA melting curve in black. Figure courtesy of W. Ben Rogers.

distance of the spheres on the ends reduces from 2-3 particle diameters to 1-2 diameters in 10 minutes or sooner. At 29 $^\circ$ C, the spheres on opposite ends of the chain stick to each other and the cluster remains in the compact state thereafter. At these temperatures, we do not observe any bond breaking events over the ten minutes (see fig. 3.3). These results are repeatable. We have observed many other chains of sulfate beads changing bond angles and forming new bonds over minutes at similar temperatures.

We next study the effect of fluorescence illumination on the beads. We use differently colored fluorescent dyes to distinguish each type of bead. Does the fluorescent dye or light source affect the surface diffusion of the beads? We repeat the experiment involving the A-B-A-B chain of sulfate beads, but image the particles with the fluorecent light source.

Without oxygen scavenging enzymes, clusters of the fluorescent beads illuminated by a Mercury arc lamp quickly break apart. After this "fluorescence-induced melting" occurs, the particles fail to bond each other again. Even after a day of waiting, the beads remain as singlets. In the experiment shown in Fig. 3.5 (top), the cluster melts in less than 1 minute of arc lamp exposure. The closer the particles are to

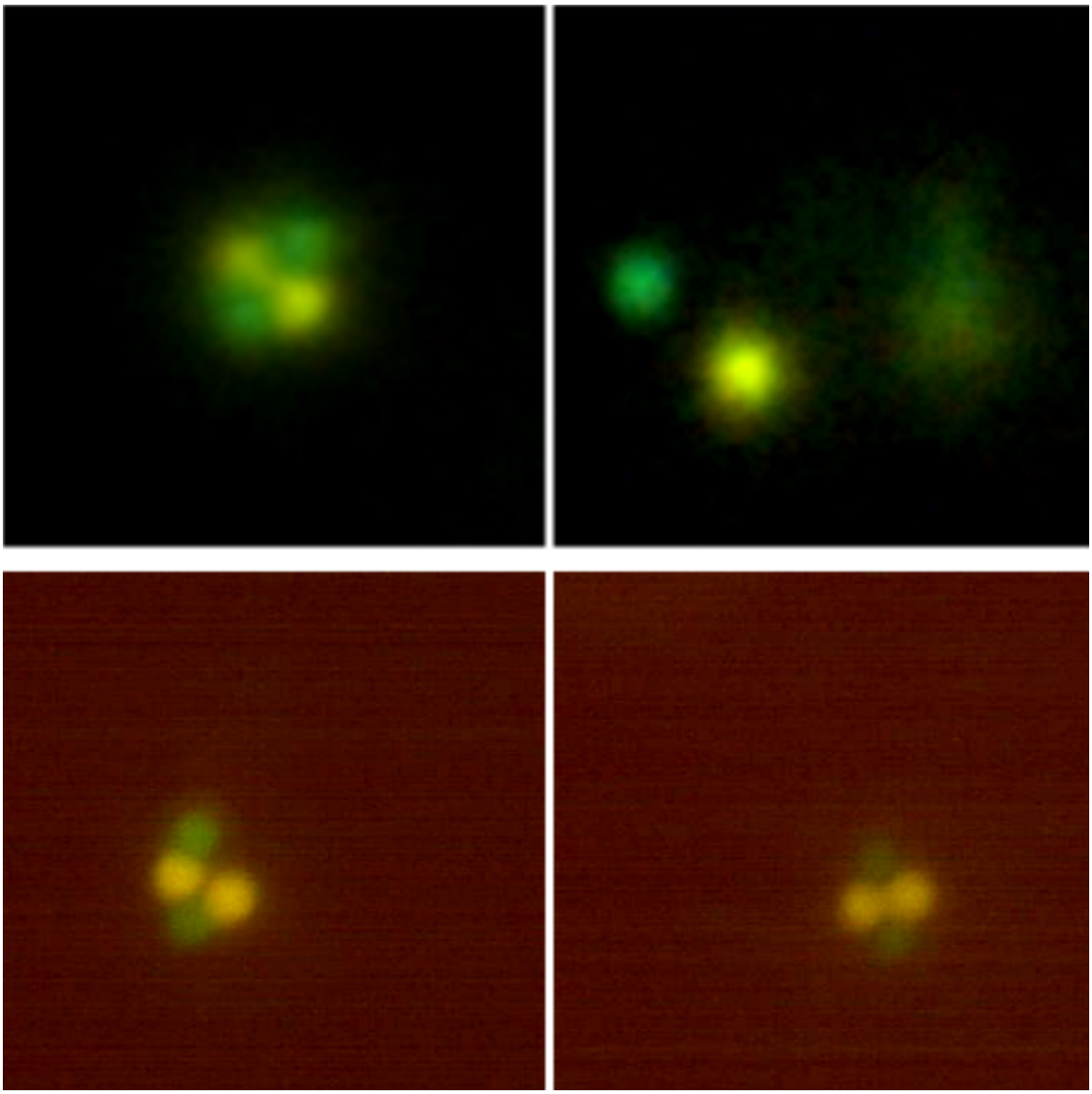


Figure 3.5: Top: without enzymes, a cluster irreversibly breaks apart in 10 seconds of mercury arc lamp exposure. Bottom: with enzymatic oxygen-scavenging system, a cluster is stable for over 10 minutes of mercury arc lamp exposure, and in fact forms a new contact over that time.

their melting temperature, the shorter the time until they break apart. For example, within a degree of the melting temperature, melting occurs within seconds, and at 10 degrees below the melting temperature, the melting occurs after 2 minutes. We also observe this irreversible "melting" in clusters of beads containing the Nile red fluorophore over minutes under halogen lamp exposure (12 V 100W).

With the oxygen scavenging enzymes, the clusters are stable for at least 10s of minutes of illumination and self-assemble just as without enzymes. We show in Fig. 3.5 (bottom) a 4-particle chain exposed consistently for approximately 20 minutes. This cluster is at 22 ° C and locally rearranges from the open 4-sphere chain into the closed loop, apparently forming the fourth and final bond. The fluorescence from the beads are distinguishable for at least 20 minutes of constant exposure.

We repeated the 4-particle cluster experiment with sulfate beads, but changed length of the sticky end from 6 bp to 8 bp, to test the hypothesis that a longer sticky end slows surface diffusion. We do not

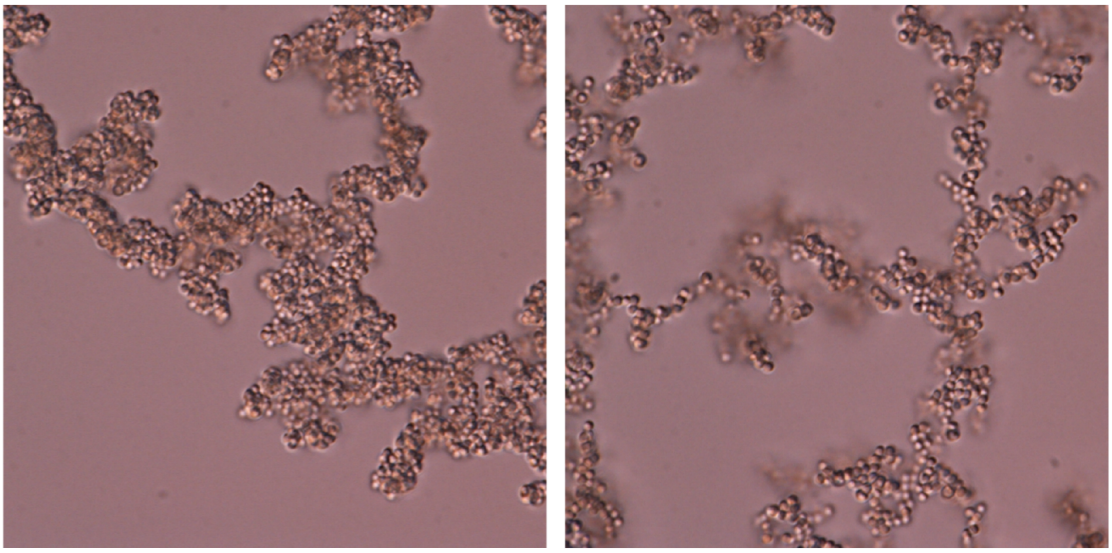


Figure 3.6: Morphology of large aggregates of beads depends on surface groups, sulfate on left, CML on right

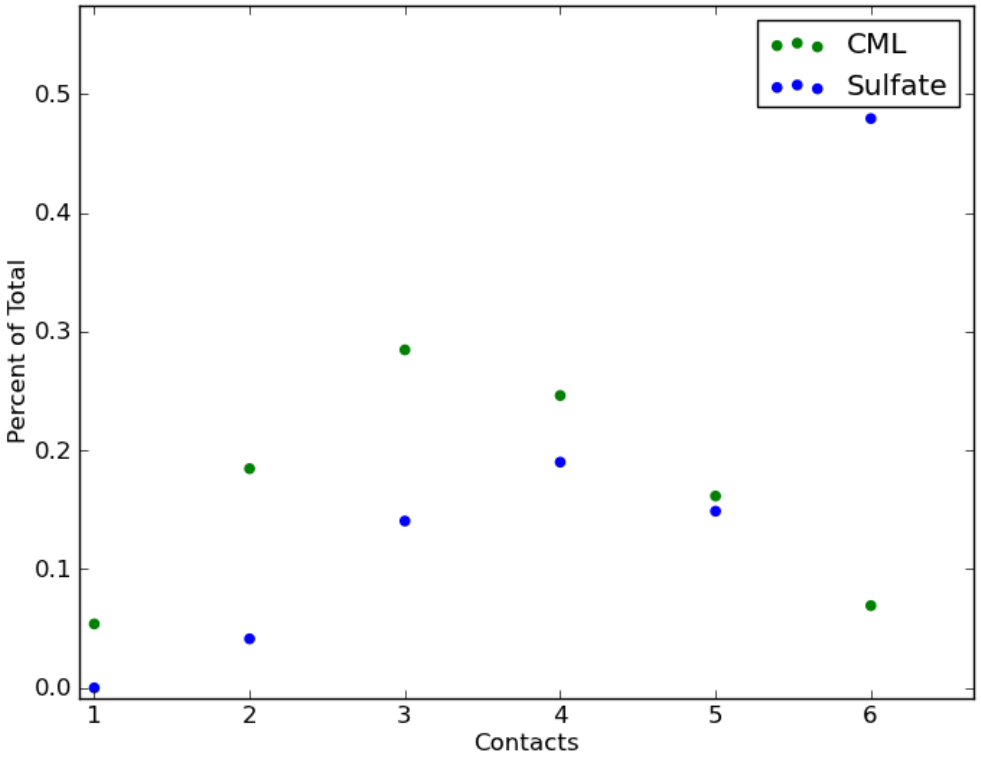


Figure 3.7: Histograms of sulfate and CML contact numbers

observe any surface diffusion over ten minutes of the same beads interacting via 8 bp sticky ends, which also have a much lower (stronger) free energy of binding. We looked for such lateral diffusion over the same range of temperatures, from 22 to 29 °, as the 6 bp beads. The 8 bp beads have a much higher

melting temperature than the 6 bp beads, at 60 $\pm 1^\circ$ C.

3.4 Discussion

We interpret these results in the context of a simple model for the surface diffusion, based on the lifetimes of DNA bridges formed between particles and the diffusivity of the particles themselves. We expect the surface diffusivity to saturate at no faster than the typical lateral diffusivity of a sphere such as ours near a solid surface, about $0.85kT/(6\pi\eta r)$[45], where k is Boltzmann's constant, T is temperature, η is kinematic viscosity of the medium and r is the radius of the sphere. We consider two timescales, t_p, the typical time for a 1 micron particle to diffuse laterally over 10 nanometers (nm), or roughly the length of the surface bound DNA sequence, and t_{dna}, the typical lifetime of single DNA duplex. For 1 micron beads, $t_p \sim 1$ milliseconds (ms), nearly independent of temperature. The DNA duplex lifetime depends strongly on temperature, salt concentration, and base pair length, but has been measured by Wallace et al.[46] to be: $0.1\,\mathrm{ms} < t_{dna} < 100\,\mathrm{ms}$.

For a given DNA-sequence, we try to maximize surface density and salt concentration, each of which increases the particle-particle melting temperature[34, 47]. The higher the temperature, the shorter is t_{dna}[46]. The higher the working temperature of the experiment, the more likely the condition $t_{dna} < t_p$ is met. Even if the beads are homogenously coated with DNA, we expect the DNA duplex lifetime to limit the rate of surface diffusion at conditions where $t_{dna} > t_p$. Our 6-bp sequences and salt conditions are similar to the conditions of Wallace et al.'s experiments. Wallace et al. performed experiments at 100 mM NaCl and 20 mM $MgCl_2$, ionic strenghts weaker and stronger than ours at 250 mM NaCl. At 24° C, they found[46] lifetimes of 1 (100 mM NaCl) and 25 (20 mM $MgCl_2$) ms, and at 29°C, 0.5 and 6 ms. Our beads thus are likely to be limited by the average duplex DNA opening rate at low temperatures. Their motion might not be rate-limited by the individual DNA duplex at higher temperatures.

The difference in surface diffusivities of DNA-F108 coated CML and sulfate beads seems most likely to be related to differences in their surface properties. Sulfate surface groups are very small. Such beads are typically thought to be stabilized by electrostatic repulsion[48]. The CML surface, however, is thought to be brush-like, and may have larger and/or more defects than the sulfate surface[49]. DNA-F108 may be less likely to adhere on regions of the particle surface occupied by the carboxylate chains than on adjacent regions, if any, resulting in a more heterogeneous coating than the DNA-F108 coating of sulfate beads. Based on Xu et al.'s work[32], we would expect the beads with the more heterogeneous surface to get stuck in dense patches for long times in comparison with the more uniformly covered surface.

Based on the oxygen-scavenging results, I hypothesize that when exposed to excitation wavelengths, the dye enters an excited electronic state, and transfers this energy into singlet oxygen or hydroxyl radicals, which in turn oxidize DNA, breaking the polymer apart. After enough time, enough of the bonds have been broken so that the beads themselves cease to bind to each other. This mechanism could explain the irreversibility of the process. The enzymes, combined with glucose, would prevent this damage by removing all the oxygen from the system prior to exposure.

The oxygen-scavenging enzymes enable observation of the colloidal spheres over long timescales, which is needed for the time-averaged study of a single set of particles transitioning through their configuration space. Such a time-averaged study would meet the second goal of my experimental project, to observe and study the self-assembly of isotropic, specific particles. Alternatively, I could have attempted an ensemble-averaged study, and to that end I had fabricated anisotropic microwell plates and collaborated

on a microfluidic approach to solve the problem of co-localizing many different sets of particles of just the right species (see Appendix B), but these approaches could have used further refinement. With specific and isotropic particles, whose species could be indentified over long times, I could finally fine-tune their interaction strengths and study the structures they explore in equilibrium.

Chapter 4

Equilibrium Self-Assembly

4.1 Introduction

With my first major research goal met, it was time to go after the second goal: observe and make sense of the self-assembly of a set of highly specific, isotropic particles. I engineered a system 3 different species. Particles of each color stick to particles of other colors, but not particles of the same color. A specificity matrix describes these interactions but a connectivity matrix describes actual contacts between pairs of particles. I chose a specificity matrix that enables every bond in the octahedron, but no other bonds (see Fig. 4.3 A and Appendix E). The specificity matrix is homologous to the connectivity matrix of the target octahedron. We targeted the octahedron because it is not favorable in single-component experiments[10], but expected our specificity matrix to make it the only configuration that maximizes the number of bonds.

Our original hypothesis was the octahedron would form in high yield in equilibrium, and we did not know of any other stable structures that might compete with it. By controlling the specificity, we were confident the octahedron would be the configuration with the lowest potential energy. Before we had true equilibrium self-assembly, however, I had discovered, in preliminary experiments, a stable 10-bond cluster that could compete with the ocathedron. Which cluster would have a lower free energy, i.e. which would we observe more frequently in equilibrium? Once we had a self-assembling system, we decided to explore another question. How does the yield of the most likely clusters, especially any cluster that is the only one to maximize the number of bonds, depend on the number of species relative to the number of particles? Here I report my studies of this 3-species system at N=6 and N=9 particles (for N=7 and N=8, see Appendix G).

4.2 Methods

4.2.1 DNA sequences

Here are the sequences coupled to the polystyrene surfaces (IDT, /5AmMC6/ stands for a '5 Amino modifier at the end of a 6 carbon atom chain):

1, /5AmMC6/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT GTA GGG TTC;

1', /5AmMC6/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT GAA CCC;

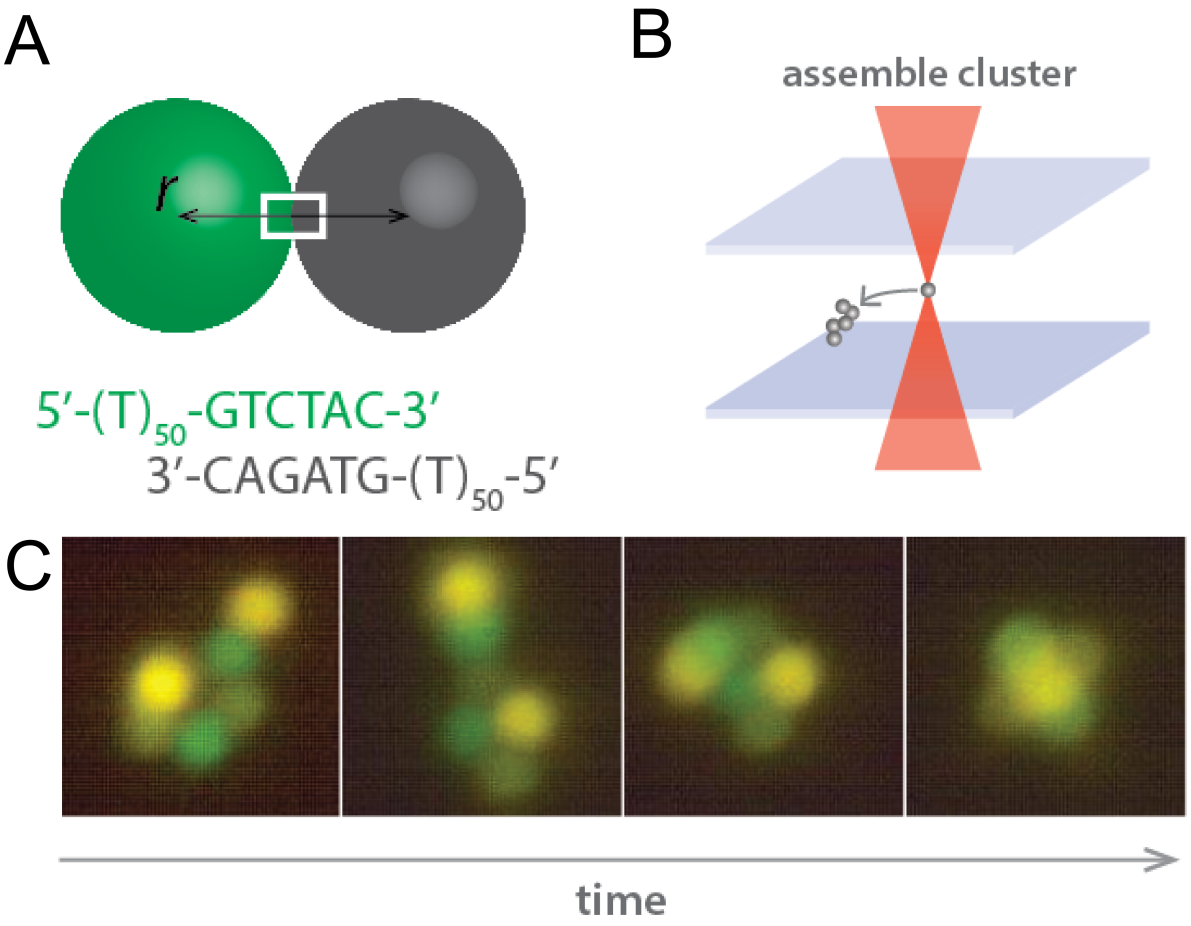


Figure 4.1: Experimental Set-up. A.) Schematic of DNA-coated colloidal spheres B.) I assemble the initial cluster using optical tweezers. C.) As the temperature increases from the quenched initial temperature to the high temperature required for equilibration, the cluster forms new contacts. Panels A. and B. are courtesy of W. Ben Rogers.

2, /5AmMC6/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT CAT CAC TCC;

2', /5AmMC6/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT GGA GTG;

3, /5AmMC6/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT CTT GTC TGG;

3', /5AmMC6/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT CCA GAC.

We also ordered these short "displacer" sequences to control finely the melting temperatures: D1, CCCTAC; D2, GTGATG; D3, GACAAG.

I ordered DNA from IDT at 3M concentration suspended in pH 8.0 Tris EDTA buffer. I reacted DNA with F108 and coated beads with this surfactant following the protocol in Appendix D. The relative amounts of DNA-F108 added to each bead type to match their thermodynamic profiles are summarized in table 4.2.1. In addition to tuning the relative amounts of DNA-F108 on the particles, I added displacer sequences to further refine and match the three melting profiles. The displacer concentrations in the

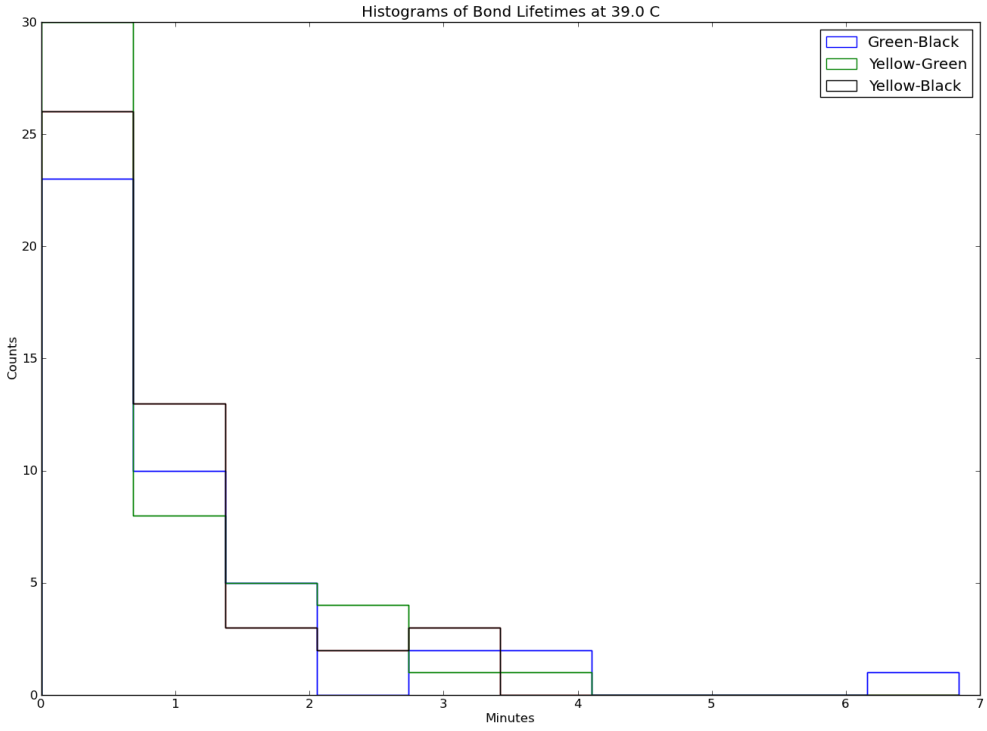


Figure 4.2: Histograms of Bond Lifetimes. Bond lifetime for each of the 3 different bond types were measured using fluorescent microscopy. The average lifetimes are 1 minute for Green-Gray, 45 seconds for Yellow-Gray and 45 seconds for Green-Yellow.

experiment were 0 uM D1, 50 uM D2, and 50 uM D3.

Color	Dye	ul DNA-F108 stocks added
Yellow	Nile Red	1.9 ul 2, 2.1 ul 3'
Green:	BP 546	2.4 ul 3, 1.6 ul 1'
Gray:	DCM	1.9 ul 1, 2.1 ul 2'

Table 4.1: Bead colors, dyes and relative amounts of different DNA-F108 used for speciation

4.2.2 Tuning Interaction Strengths

I tune the specificity and strength of the interparticle interactions so that the spheres may explore multiple configurations in equilibrium. The coarse sphere-sphere interactions are summarized in a specificity matrix (Fig.4.3 A), which illustrates the set of pairwise interactions between particles. I adjust the surface densities of DNA-coatings and add soluble competitive DNA sequences so that each of the 3 distinct favorable interactions is about equal in magnitude at the observation temperature, 39.0°C (Fig. 4.3 B). In equilibrium, the spheres continually form and break DNA-DNA contacts with each other, and transition through different cluster geometries (Fig. 4.3 C-E). I count the average lifetime of each of the 3 different bond times, and report these in a histogram (Fig. 4.2). Each of the bond lifetimes is on average 45-60 seconds.

4.2.3 Initiating an Equilibrium Experiment and Recording Data

After the proper number and types of particle have been brought together by optical tweezing in dilute solution (Fig. 4.1 B), I increase the temperature to 39.0°C. As the temperature increases, the cluster changes configuration and forms new contacts (Fig. 4.1 C). I do not count the cluster configuration toward equilibrium yield until 15 minutes have passed since the 39.0°C setpoint has been measured by a platinum thermistor coupled to the sample by thermal paste. After the temperature reaches its setpoint, equilibrium binding and unbinding may be observed (Fig. 4.3 C-E). Every minute, I expose the sample to fluorescence for approximately 10 seconds to obtain particle identification data in the recorded video. I image and record video in brightfield at all other times. I used optical tweezers to keep extra particles away from the cluster of interest. If a particle detaches from the cluster completely, and does not return after a short time, I use optical tweezing to return the particle to the cluster, and count the cluster state only after waiting a time characteristic of cluster rearrangement (about 1 minute).

4.3 Results

At N=6, I observed the true octahedron 16 times as often as I observed 3 metastable, 10-favorable contact states combined (Fig. 4.4 C). These metastable sructures can appear identical to each other and to the octahedron in brightfield, but are each clearly distinguishable in fluorescence microscopy. (Fig.4.4 A-B). I observed the true octahedron in 130 instances each separated by at least a minute, but observed the strictly local minima only 2 or 3 times each. I neglect any observations that cannot be ascribed to any of these 4 states.

At N=9, I observed a rigid, 21-favorable contact cluster (Fig 4.5 A) over 2 times as frequently as any other state, but less frequently than I observe 8 metastable 20-favorable contact states combined (Fig 4.5 B). These 8 states are geometrically distinct from the 21 contact state. Each of these 8 states fall into one of two disjoint sets, with each member of the same set having the same appearance in brightfield (equivalently, the same connectivity up to colors). Members of one of the sets, a set of 3, each contain an octahedron as a subcomponent. Each member of the other set of 5 lack any octahedral subcomponent. Although I also observe some difficult to distinguish 18- and 19-favorable contact structures that are locally stable, I do not report this data here. If considering only these 20 and 21 contact states, I observe the 21-favorable contact state 31% of the time, the set of 3 similar clusters 31% of the time, and the set of 5 similar clusters 38% of the time.

4.4 Discussion

Although the N=6 and N=9 systems both have a single dominant cluster, my observations suggest their configuration spaces and energy landscapes differ not just quantitatively, but qualitatively. How can we make sense of the fact that the octahedron occurs 94% of the time, but the N=9 conjoined octahedra only occurs 31% of the time, relative to other stable states? Both are the states that maximize the number of contacts, but the conjoined octahedra state has only one more favorable contact than the next-lowest-lying minima. Moreover, there are in principle 9 minima with only one fewer favorable contact in that N=9 system. Another way of comparing the N=6 and N=9 systems is to analyze the number of bonds that must be broken to transition into and out of the two ground states. The 21-favorable contact state can be

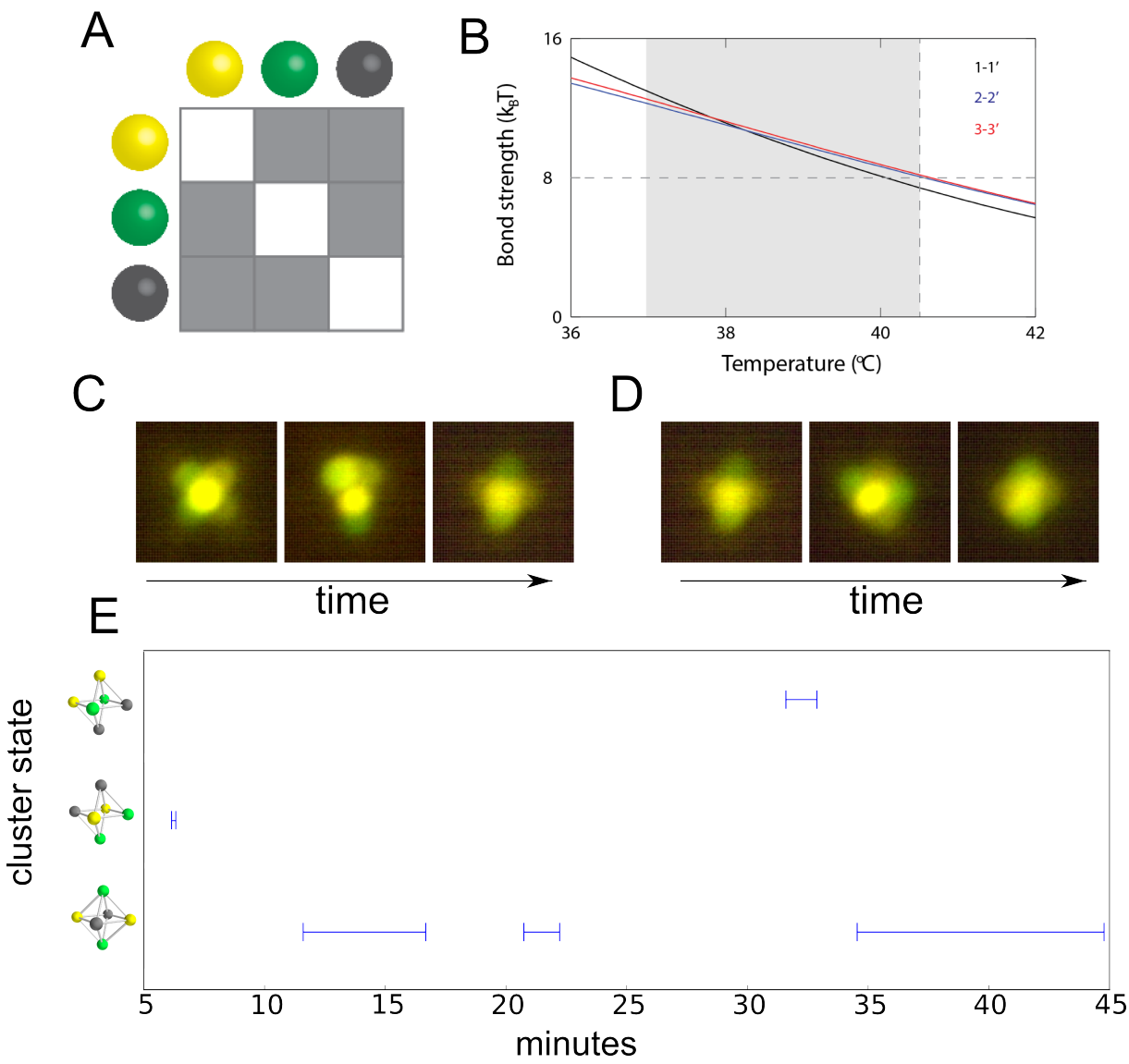


Figure 4.3: Specific, Equilibrium Interactions. A) Interaction Matrix. Particles of each type (color) bind only particles of a different type (color). Gray squares indicate favorable interactions, white indicates purely repulsive interactions. B) Bond strength versus temperature calculated from measured singlet fractions for each pair of attractive species. C) As might be expected for equilibrium binding, a cluster loses at least one green-yellow contact and transitions from one locally stable state to another over 6 minutes. D) The same cluster in C) loses a green-gray contact but regains it over 3 minutes. E) As bonds break and form, particles explore various configurations, as expected for equilibrium binding of a system with multiple stable energy-minimizing states. Panels A and B courtesy of W. Ben Rogers.

rearranged into one of the stable 20 contact geometries by breaking only 2 bonds, instead of the 3 required to reconfigure the N=6 ground state into a strictly local minimum. In terms of the free energy landscape, the N=9 system might have one deepest well, but it has many wells that are almost as deep, or, neglecting color, a couple wells that are almost as deep but much wider. Differences in strengths of bonds between different species would affect the N=6 and N=9 systems differently. Because the 10-contact minima at N=6 consist of a strict subset of the bonds of the true octahedron, any difference in bond strengths cannot enable them to be lower in total potential energy than the true octahedron. At N=9, however, a substantial

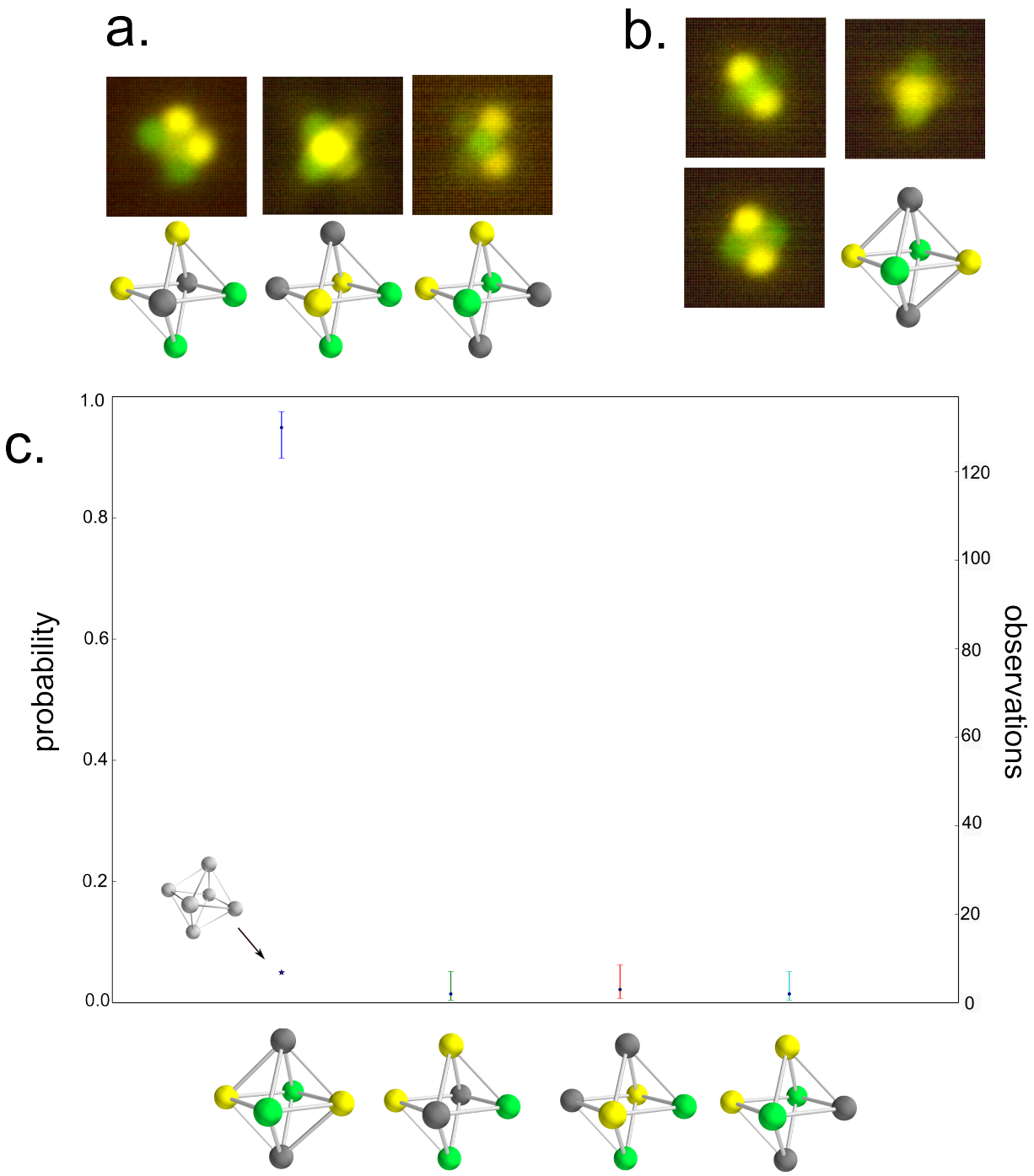


Figure 4.4: 6 particle experiment A) Fluorescent micrographs and rendering of 3 distinct 10 contact, strictly local minima. B) Fluorescent micrographs of octahedron along each of its 3 distinguishable fourfold-symmetry axes, and rendering. C) Probabilities and number of observations of each of the 4 distinguishable stable structures. Observations are considered distinct here if they are separated by at least 1 minute. Error bars for probability correspond to Wilson score 95% confidence intervals. The probabiliy of a single-species octahedron relative to the single-species tritetrahedral structure is shown for comparison [10].

difference in bond strengths of different bond types could mean that one or more of the 20-favorable contact stable structures actually have lower potential energy than the 21-favorable contact state.

I experimentally implemented a specificity matrix meeting a design constraint: each particle is capable

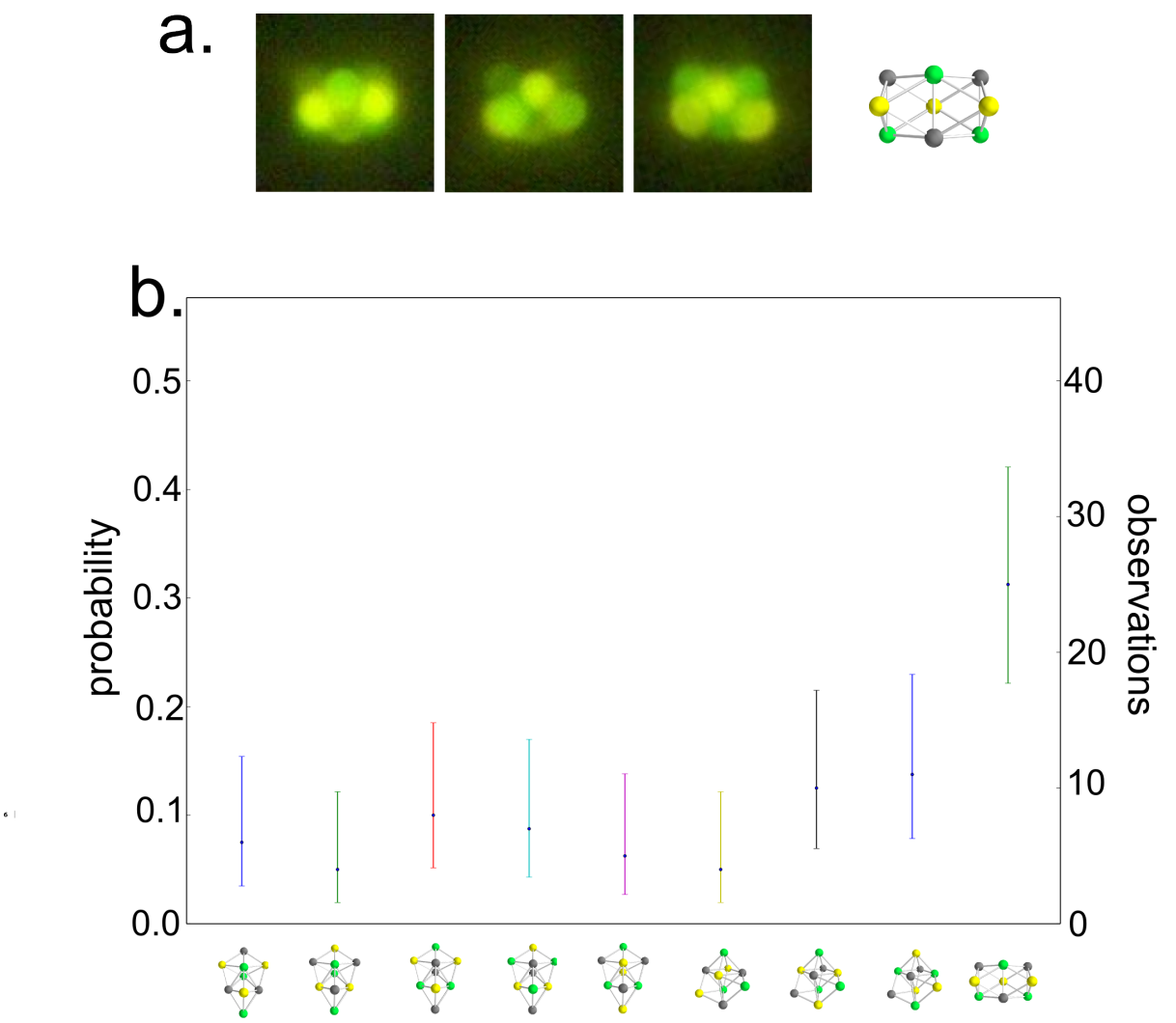


Figure 4.5: 9 particle experiment A) Fluorescent micrographs of 9 particle 21 favorable contact cluster along each of 3 distinct, central twofold symmetry axes and rendering. B) Probabilities and number of observations of locally stable clusters with at least 20 favorable contacts. The error bars reflect the Wilson score 95% confidence interval about the actual probability. Observations must be separated by at least 1 minute.

of bonding with the nearest neighbors in the target structure, but not with any particle besides those particles. We found that for a small number of particles, this condition was sufficient to enable the equilibrium formation of the target geometry in high yield, even though this geometry is entropically unfavorable in a non-specific system. This design constraint is not met by our 9 particle system. In that system, particles can bond with particles besides those they neighbor in the maximum bond state. Both the N=6 and N=9 systems, however, exhibit non-rigid states that are nonetheless locally stable, unlike comparable non-specific systems. A remaining question is whether self-assembly is necessary for organization of a unique ground state. Could a non-equilibrum assembly process work just as well or better (see Appendix F)? Experimentally, I only investigate small numbers of spheres. I had to wonder, how scalable is our self-assembly method? If we had 10s or 100s or species, would it work just as well?

Chapter 5

Theory: Scaling Relations and Statistical Mechanics

5.1 Introduction

As I mentioned in the introduction to this thesis, I started a theoretical study on the side while working towards my first experimental goal. After I discovered the 10-bond local minimum, I became interested in characterizing it and discovering rules for finding other local minima (see Appendix A). I wondered whether or not I could expect the octahedron to self-assemble in high yield relative to its strictly local minima. I wondered too how scalable self-assembly via specific interactions might be. Even if I were to find that the octahedron assembled in high yield, what could I say about systems of more particles and species? Given enough species and control over their interactions, in theory, any cluster geometry of any number of particles could be made to be the unique ground state in potential energy. However, that says nothing about how long it would take to self-assemble. What if it took forever? If so, the approach of using many species to self-assemble arbitrary structures might not seem very useful.

My approach to this question of kinetics starts with equilibrium statistical mechanics. I write partition functions that describe systems of many different particles that have a well-defined ground state. From these partition functions, I calculate the ground state probability as function of the number of particles and species of the system. For these models, it is always theoretically possible to make the ground state have high yield by lowering the temperature or making certain bonds stronger. That introduces practical issues though. Lowering the temperature or strengthening certain bonds increases the bond lifetimes and decreases transition rates, which increase the time spent in kinetic traps. In the case of one exactly solvable model, I also address the question of how the expected number of transitions before entering the ground state grows with the number of particles and species. The more particles and species, the lower the temperature must be for the ground state to form in equilibrium, the slower the transition rates are, and the more time the system spends in local minima.

Compared to single-component systems, many specific spheres may be disordered in equilibrium or take an unusually long time to reach their ground states. An implication of this theory is that, unlike the few particle cases I studied experimentally, self-assembly of large numbers of DNA-coated spheres may not assemble into target structures on reasonable timescales. I use the phrase “permutation glass” to refer to these systems of many different species with slow kinetics and/or equilibrium disorder. Finally, I ask

how separating the particles into two different systems might affect their self-assembly.

5.2 Swapping Particles and Enumerating Local Minima in Non-Specific vs. Specific Systems

In a non-specific, single-component system, a cluster is as stable before a swap as after the swap (Fig. 5.1).

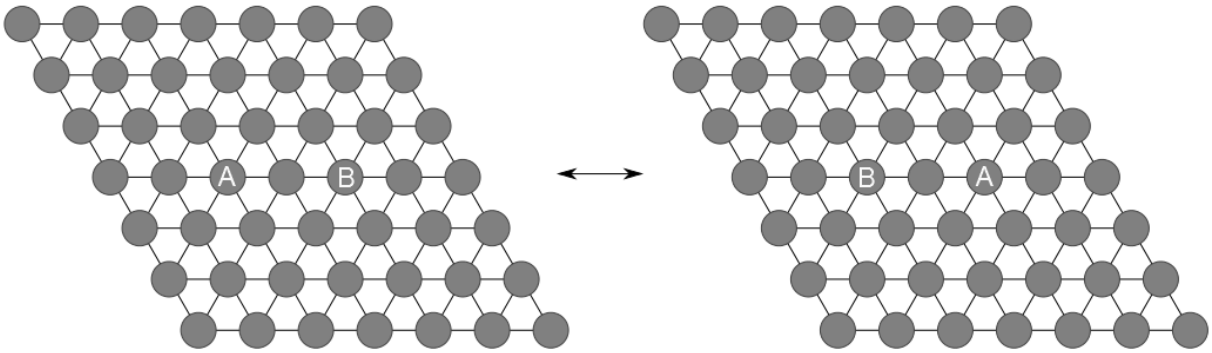


Figure 5.1: A set of non-specific particles in their ground state does not lose bonds when two of them swap positions

In a system of many different spheres, however, the connectivity and energy of the system be different after a swap compared to before the swap. Consider a system of N unique spheres in their lowest energy state. I assume that each particle can bond with the nearest neighbors in this state, but not with any particle besides those particles. Consider starting in the ground state and swapping any pair of particles. After the swap, the total energy is higher and the number of bonds is lower (Fig. 5.2).

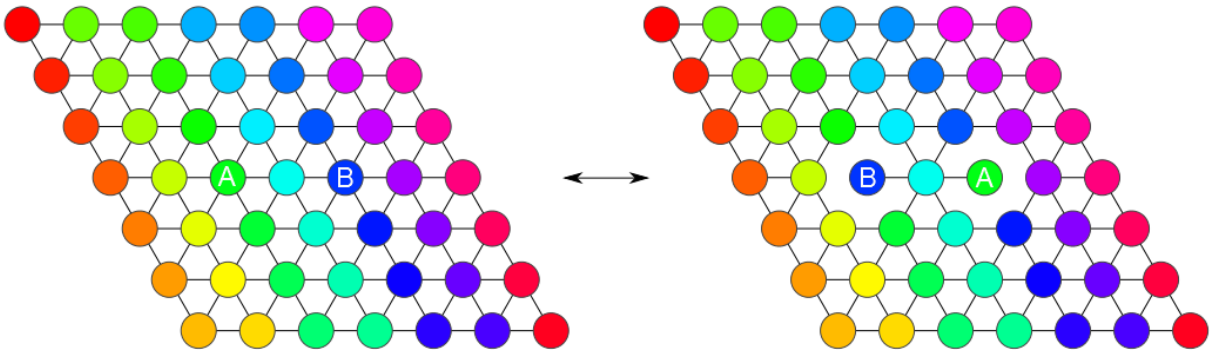


Figure 5.2: A set of maximally specific particles in their ground state loses bonds when two particles swap positions

A set of highly specific spheres can form cages in ways that a set of non-specific spheres cannot. A non-specific sphere caged by other non-specific spheres will tend to form bonds with the caging particles, but a specific sphere could be trapped without the ability to bond with any of the surrounding spheres. The number of ways of forming such cages can be very large. A cage size can be finite and small, on the order of $\sim 6-10$ spheres depending on dimensionality. A large cluster or lattice of N specific spheres could have nearly as many cages as spheres. Consider any set of cages. The number of ways to fill those cages with the remaining $M < N$ particles is $M!$. Let's say only one of those $M!$ arrangements is the

ground state, and again, each particle likes to bind only those particles it neighbors in the ground state, not any particle except those particles. In other words, the particles are maximally specific. Not only are all $M! - 1$ other ways to rearrange the caged particles in their cages local minima, but about $M!/e$ of these rearrangements have no fixed points (see Fig. 5.3). In these derangements, each of the permutated spheres retains at most 1 bond, and fewer if the inner layers of no two adjacent cages share any particles. As the number of highly specific particles increases but a lattice or close-packed cluster remains the ground state, the number of ways to rearrange them into cages diverges rapidly.

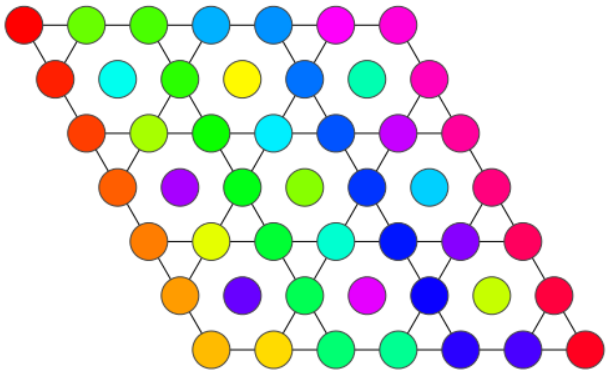


Figure 5.3: One of $\sim \frac{9!}{e}$ >100,000 local minima in which each of nine particles caged here forms at most 1 bond.

5.3 Partition Functions of Systems with Broken Permutation Symmetry

5.3.1 Exactly Solvable Model: Quasi 1D System

Imagine chemically patterning a 1D stripe of single-strand (ss) DNA molecules on a glass surface, binding them to the surface their 3' end, exposing a sequence of nucleotides I call A at the 5' end. Place N_S DNA-coated beads of N_S different species near the stripe, where each bead has been labelled with a ss sequence A*, denoting the complementary sequence to A, as well as 2 other types of DNA. I choose the DNA sequences to satisfy the following rules: particle i binds to particle $i + 1$ and $i - 1$, but no other particles. I reserve the index $i = 1$ for the stripe on the surface. I assume that the particles maximize their bonds by lining up in an open (as opposed to closed) chain (but I will show the closed chain behaves equivalently). The ground state is the N particle chain bound along the DNA stripe (see Fig. 5.4 and 5.5).

What are the local minima of this system? Any arrangement of the N_S spheres other than the ground state ordering (and its reverse ordering if I choose to count it) is a local minimum; at least one particle must break the AA* bond, leave the stripe, and reenter between another pair of particles. I write the partition function for this system,

$$Z_{1D}(N) = e^{-N\beta U_0} \sum_{k=0}^{N} A(N,k) e^{-k\beta U},$$

where U_0 is the energy of a stripe-particle bond, k is the number of nearest neighbor pairs that share a bond (and thus the number of bonds), and $A(N,k)$ is the number of arrangements that have exactly

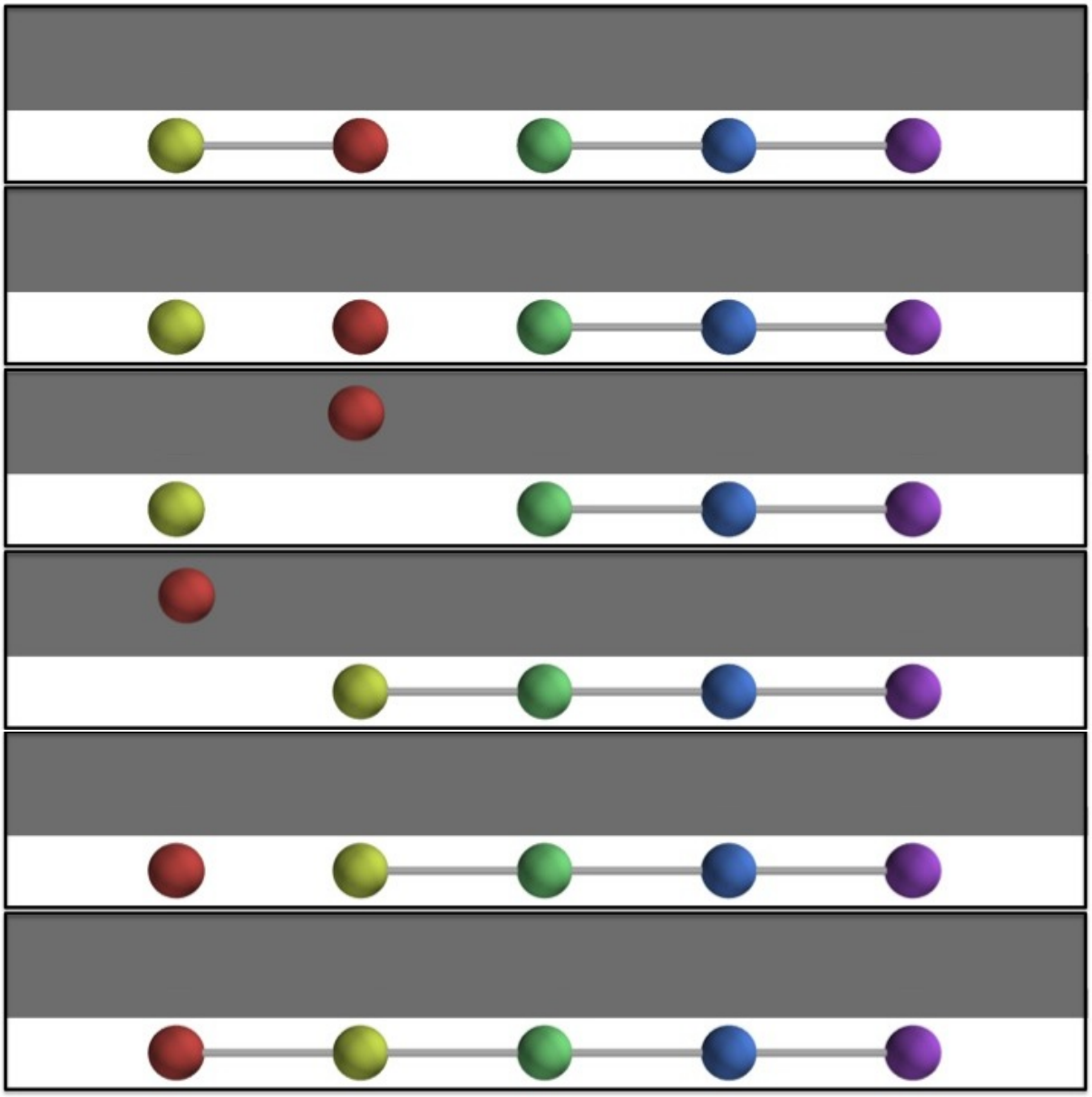


Figure 5.4: A sequence of steps (going from top of the page towards bottom) taking a local minimum (5+4=9 bonds) of 5 particles to their ground state (5+5=10 bonds). The white stripe represents a surface coated with DNA that is complementary to the colorful particles. The particles are coated with DNA such that they can maximize their inter-particle bonds only in the linear arrangement red-yellow-green-blue-purple. Because the stripe is thin relative to particle size, the particles much debond from the white stripe then diffuse along the uncoated grey surface in order to rearrange. Here the local minimum first breaks a (red-yellow) bond along the chain, then the red particle breaks its bond with the white stripe and diffuses along the uncoated grey surface while the yellow particle diffuses along the white DNA coated stripe. The yellow particle bonds with the green particle forming a 4-particle chain. The red particle binds the white stripe on the opposite side of the yellow particle and finally binds with the yellow particle again forming the ground state.

k out of the N possible pairs of neighbors. For chains, there is no $k = N$ state. The total number of

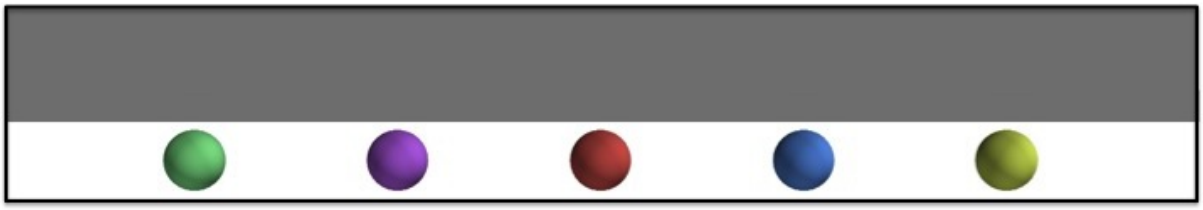


Figure 5.5: Zero bond local minimum. There are N!/2 ways to arrange the beads on the stripe, each of which corresponds to a basin in the energy landscape. For high N, most of these arrangements have very few (≤ 2) bonds along the chain (but each arrangement has all N particles bound to the stripe). Here is an arrangement of 5 spheres lacking any bonds along the chain.

arrangements is equal to the sum of $A(N,k)$ over all k.

$$\sum_{k=0}^{N} A(N,k) = \begin{cases} (n-1)!/2 & \text{open chains} \\ n!/2 & \text{closed chains} \end{cases}$$

This sums to the number of cycles of N distinguishable objects in the case of bracelets, and the number of permutations of N distinguishable objects in the case of chains (each divided by 2 to take into account dihedral symmetry). The vast number of local minima motivates further study of this partition function. As there is only one ground state, $A(N,N) = 1$, but $\sim N!$ strictly local minima, this system meets the definition of a permutation glass. I solve this system analytically using some results in combinatorics.

An explicit formula for the multiplicities, $A(N,k)$ in the chain case is given in [50], which considers the number of "neighbors who remain neighbors after rearrangements." The cyclic case has been referred to as "the dinner table problem" [51]. I am interested in the limiting case of large N which has been shown to be equivalent for open and closed chains (actually purely 1D chains) [50, 51],

$$\lim_{N\to\infty} \frac{A(N,k)}{N!} = \frac{2^k}{k!} e^{-2}.$$

The $A(N,k)$ fall on a poisson distribution of mean 2, so the vast majority of the states of very high energy, such as the state in Fig. 5.5. Let's consider only the $k = 0$ states, of which there are $N!e^{-2}$. Considering just these states and the ground state, $k = N$, it is clear that $\frac{\Omega_{k=0}}{\Omega_{k=N}} = e^{-2}N!$, $\Delta S = ke^{-2}\ln N! \approx kN\ln N$, and $\Delta S/N \propto \ln N$. If not restricting the stripe length, the strictly local minima are dilute.

I wrote a python program using NetworkX[52] to display the energy landscapes of the first few N of this quasi-1D system (Figs. 5.6- 5.8, Appendix I).

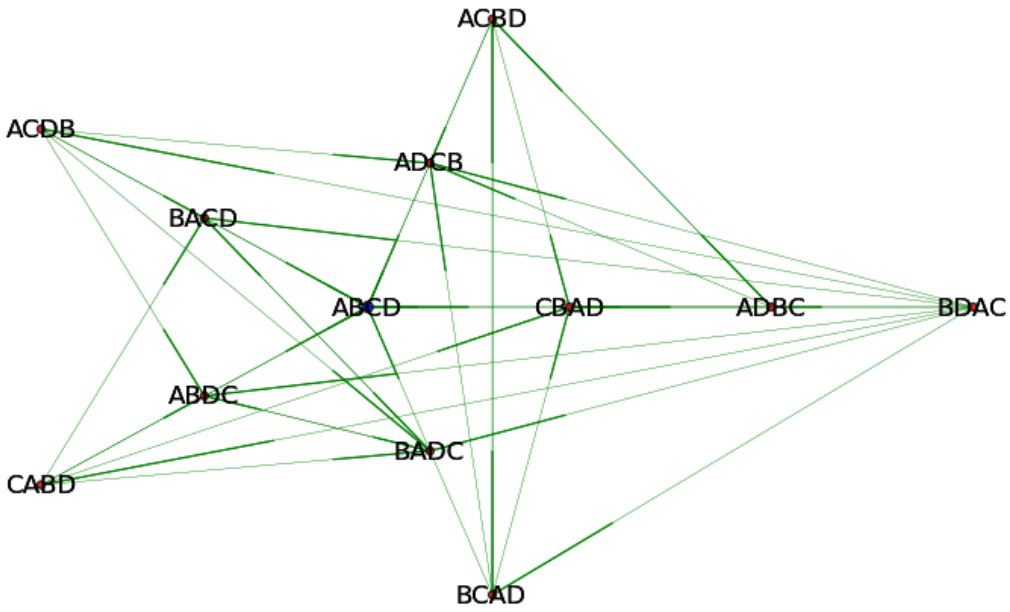


Figure 5.6: Energy landscape of the N=4 system. Red nodes represent local minima. The single blue node in center represents the lone ground state. Edges drawn from one node to another end in a thicker green line segment if a transition requiring only one bond breakage can be made from one minima to the other. The minima are drawn on shells, with the innermost shell corresponding to one missing bond, the second shell missing two bonds, and so on.

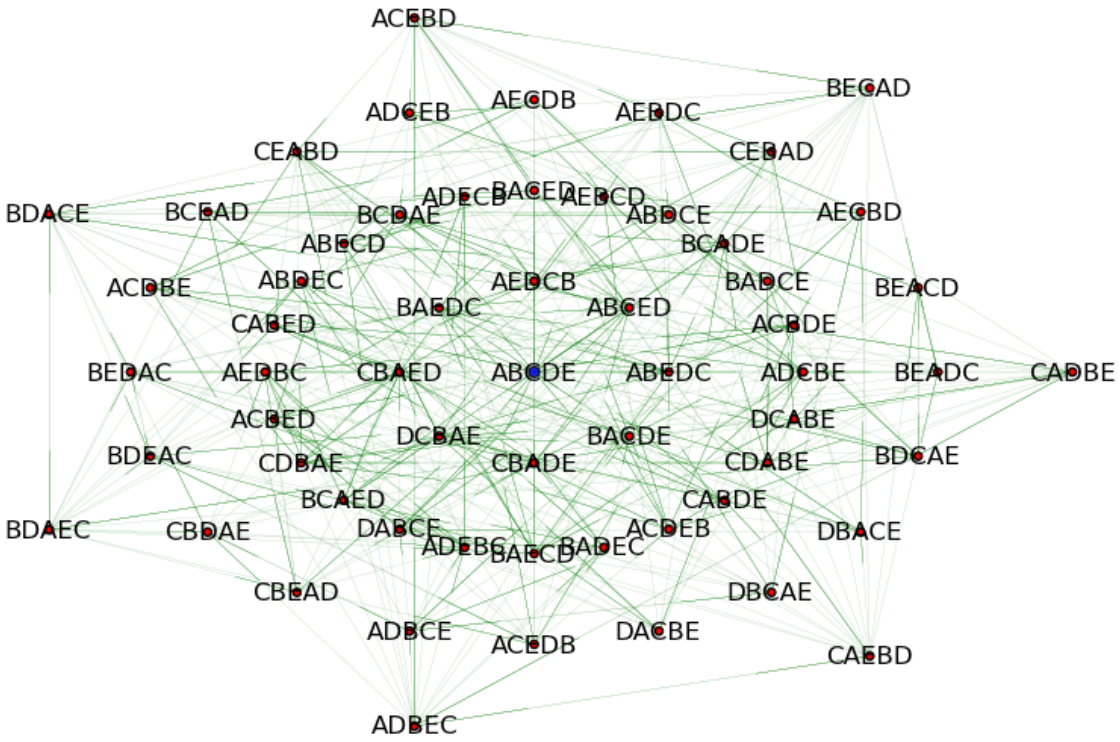


Figure 5.7: Energy landscape of N=5 system. Red nodes represent local minima. The single blue node in center represents the lone ground state. Edges drawn from one node to another end in a thicker green line segment if a transition requiring only one bond breakage can be made from one minima to the other. The minima are drawn on shells, with the innermost shell corresponding to one missing bond, the second shell missing two bonds, and so on.

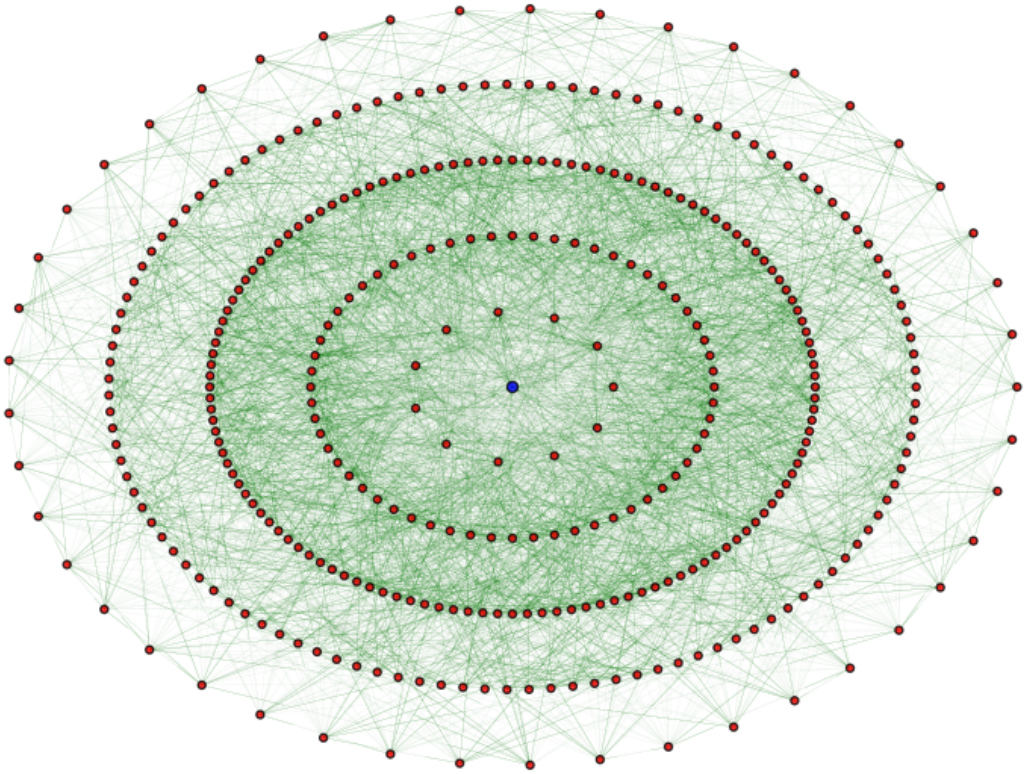


Figure 5.8: Energy landscape of the N=6 system. Red nodes represent local minima. The single blue node in center represents the lone ground state. The minima are drawn on shells, with the innermost shell corresponding to one missing bond, the second shell missing two bonds, and so on.

Expected Number of Transitions Grows with System Size

The typical self-assembly time can be estimated as the product of the typical number of transitions before reaching the ground state times the typical time spent in each local minima before transitioning to the next. Elsewhere in this chapter I write an expression showing how the transition rates may be very slow. The exactly solvable model here, however, provides an opportunity to consider how the typical number of transitions might grow quickly with the number of species and particles. I calculate the pagerank[52] of the ground state on the energy landscapes of 4 to 8 particles of this quasi-1D model. The inverse of this pagerank is proportional to the expected number of local minima visited on the way to this ground state (see Fig. 5.9).

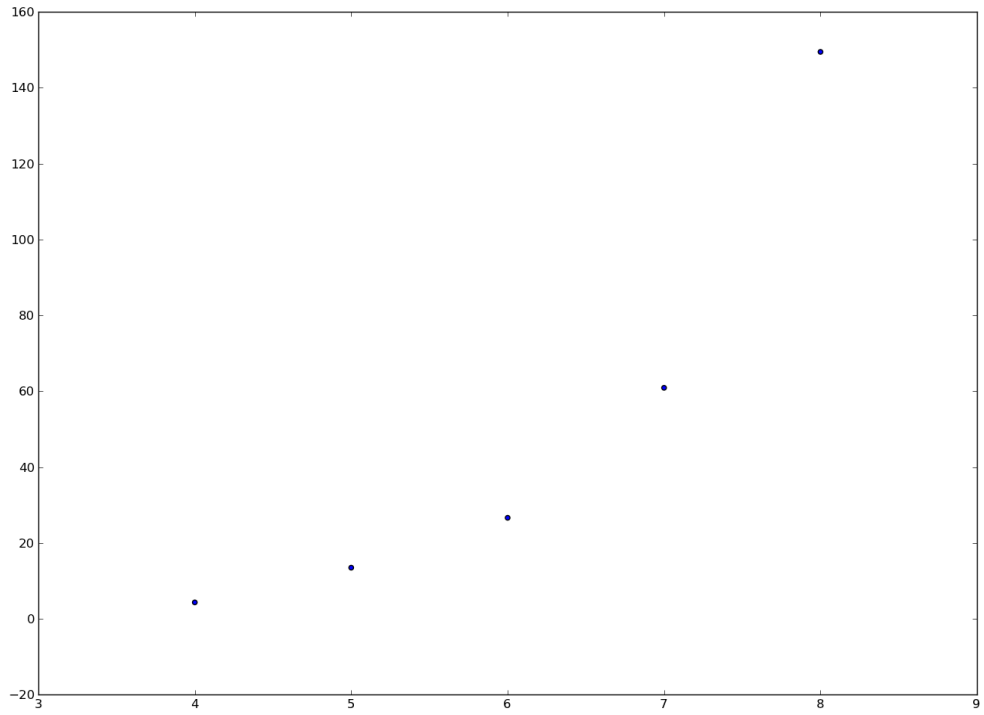


Figure 5.9: The expected number of transitions before finding the ground state calculated from the energy landscape graphs for 4-8 particles in the quasi-1D system. The calculation neglects any transitions that require breaking more than one bond.

5.3.2 2-state Approximation Method

For more complex systems, I approximate the partition function using 2-state models. These partition functions show how the large number of strictly local minima relative to ground states means systems of many species can require lower temperatures and/or longer times to equilibrate into a global minimum than systems of few species. The Boltzmann probability of the ground state may be lower than expected in comparison with a similar total particle number but much fewer species system. I use this 2-state approximation to obtain an upper bound on the ground state probability.

The actual partition function can be separated into two terms, the term for the ground state and a sum over all excited states,

$$Z = \Omega_0 e^{-U_0\beta} + \sum_{i>0} \Omega_i e^{-U_i\beta},$$

where Ω_i represents the multiplicity of state i, U_i is the potential energy of state i, β is the "thermodynamic beta" (i.e. $\frac{1}{kT}$), and $i = 0$ indicates the ground state. By definition, $\forall i > 0 : U_0 < U_i$ and $\forall i : U_i \leq 0$. Because the sum over excited states can be difficult for me to calculate, I consider only one of its addends. I write a two-state partition function,

$$Z_{2s} = \Omega_0 e^{-U_0\beta} + \Omega_E e^{-U_E\beta},$$

considering just one excited state with multiplicity Ω_E and energy U_E. Because all the terms in Z positive,

$$Z_{2s} \leq Z.$$

The ground state probability as calculated via the 2-state model is at least the actual probability:

$$P_{0,2s} = \frac{\Omega_0 e^{-U_0\beta}}{Z_{2s}},$$

$$P_0 = \frac{\Omega_0 e^{-U_0\beta}}{Z},$$

$$P_{0,2s} \geq P_0.$$

$P_{0,2s}$ gives an upper bound for P_0.

In general, the ground state probability for a two state model can be written as

$$P_0 = \frac{1}{1 + \frac{\Omega_E}{\Omega_0} e^{-(U_E - U_0)\beta}}.$$

This approximation also gives us an upper bound on the temperature normalized energy difference between the ground state and excited states required for the ground state to be most favorable in equilibrium. U_0 is negative and less than all U_i, and β is generally positive, running from 0 (infinitely high temperature) to infinity (absolute zero). For low enough $\Delta U_{2s} = U_0 - U_E$, the ground state will be favorable, its probability approaching 100%. Similarly, for low enough temperature (high β), the ground state occupancy approaches 100%. The $\Delta U \beta$ that a 2-state model predicts for a given ground state probability is typically an overestimate. The full model would require a more negative value of the temperature normalized energy difference to predict the same ground state probability.

5.3.3 Comparison of Non-specific and Specific Systems

Earlier in this section I compared visually a system of N non-specific spheres with a more colorful system of N highly specific spheres. The gray non-specific spheres could bind with any other gray sphere. Each specific sphere however, was allowed to bind with its nearest neighbors in the single ground state, but not with any other spheres.

I write a two-state partition function for sticky hard spheres (non-specific) as $Z_{ns} = N! e^{-nN\beta u} + N! e^N e^{-mN\beta u}$ where u is the bond energy, n is the number of bonds per particle in the FCC/HCP lattice (i.e. 6 in 3D), m is some bond per particle number for non-lattice packings such that $m < n$, $N!$ is the permutation degeneracy in both cases, and e^N is a bound on the number of non-isomorphic sphere packings[53]. I have neglected vibrational entropies and symmetry numbers for simplicity. This partition function is characteristic of an extensive thermodynamic entropy difference between excited states and ground states.

I write a two-state partition function for N specific spheres as $Z_s = e^{-nN\beta u} + \frac{N!}{e} e^{-mN\beta u}$, where u is the bond energy, n is the number of bonds per particle in the ground state (i.e. 6 in 3D FCC/HCP), m is some bond per particle number for highly caged states such that $m < n$, $\frac{N!}{e}$ is approximately the number of permutations with no fixed points, or derangements.[1] I am assuming cages consist of c particles, with

[1] A more precise form of this 2-state partition function would take into account the number of different ways to choose the cages and would include a statement about how the cages are formed, e.g. by forming a snub trihexagonal tiling from the spheres.

$N >> c$, so there are $\frac{N}{c}$ cages, and $\frac{N}{c} \sim N$.[2] Unlike the partition function for N non-specific spheres, this partition function can be shown to be superextensive in the entropy.

From these two partition functions, we can see that the entropy difference between the entropy of the excited states and the entropy of the ground state differs greatly for non-specific particles and specific particles. The number of microstates of excited states of non-specific spheres is $\Omega_{E,ns} \approx N!e^N$ and the number of microstates corresponding to the ground state is $\Omega_{0,ns} \approx N!$. The entropy of non-specific sphere excited states is $S_{E,ns} = k \ln \Omega_{E,ns} \approx k \ln(N!e^N)$, and the entropy difference between excited states and ground states in the non-specific case is $\Delta S_{ns} = S_{E,NS} - S_{0,ns} \approx k \ln(N!e^N) - k \ln(N!) = kN$. $\Delta S_{ns}/N$ is extensive. It does not depend on the number of particles. The entropy difference between excited states and ground states of the many species, highly specific system is $\Delta S_s = S_{E,s} - S_{0,s} \approx k \ln(N!) - k \ln(1) = k \ln(N!) \approx kN \ln(N)$. $\Delta S_s/N \propto \ln(N)$ is superextensive:

$$\lim_{N \to \infty} \frac{\Delta S_s}{N} \to \infty.$$

5.3.4 3D Cluster: Omniphilic Particles Control Order/Disorder

My experiments focus on the self-assembly of 3D clusters, and a set of multple species of spheres with a unique ground state. How would addition of an "omniphilic" $i = 1$ species to a cluster of highly specific particles alter the clusters' thermodynamic stability? To answer this question let's again turn to a two-state picture.

The ground state 3D cluster has an average coordination number n, and potential energy U for each bond in the "fully specified" N_C particle cluster giving a total potential energy of $U_{cluster} = nUN_C$. Here N_C is the number of colorful (highly specific) spheres, and N_S is the total number of species, including the omniphilic species, denoted by species index $i = 1$. Let's add some $i = 1$ particles into the system and consider both the ground state, which is the 3D cluster with its surface coated by species 1, and the local minima in which a cage of m species 1 beads surrounds each of the N_C colorful particles. Thus we require at least mN_C species 1 beads.

The energy of the local minima I have described is mU_1N_C where U_1 is the bond energy of between a member of species 1 and a member of any other species. To calculate the energy of the coated ground state cluster, we need an estimate for the number of species 1 beads that can cover the cluster surface. I approximate the cluster as a glob-like sphere of radius R and volume $V_C = N_C V_p$ where $V_p = 4/3\pi r^3$ is the size of a single particle of radius r. I take the number of coating particles also of radius r to fill a single-particle thick shell on the surface of the glob such that $N_{shell} = \frac{4\pi R^2 r}{4/3\pi r^3} \propto R^2/r^2 \propto N_c^{2/3}$. So only about $U_1 N_C^{2/3}$ bonds can form at the surface in addition to the $U_{cluster}$ given above (see Fig. 5.10 for an illustration).

We'd like the multiplicity of each state to evaluate equilibrium probabilities. The cluster state has $\Omega_{gs} \sim N_C^{2/3}!$ degeneracy by permutating the $i = 1$ species particles that coat the surface. If there are N_{cages} ways to build N_C cages from mN_C beads, then there are $N_{cages}N_C!$ local minima of this type (as there are $N_C!$ ways to fill each set of cages, see Fig. 5.11). Actually there are about $\Omega_{lm} = \frac{(mN_C)!}{m!^{N_C}}$ ways

[2]The assumption $N/c \sim N$, made here for convenience, is delicate. The ratio of the functions e^N and $\frac{N}{c}!$ is non-monotonic, and even small c may require $N \sim ce^{c+1}$ species before the maximally specific multi-component system's permutation entropy compares to the single-component system's configurational entropy of non-isomorphic packings. However, another way to arrive at a similar model partition function is to assume that each particle can bond with any other particle besides those which it neighbors in the ground state, albeit with a weaker bond. Such a system could be engineered by coating each of the colloidal sphere species with DNA of the same weak binding sequence and its complement.

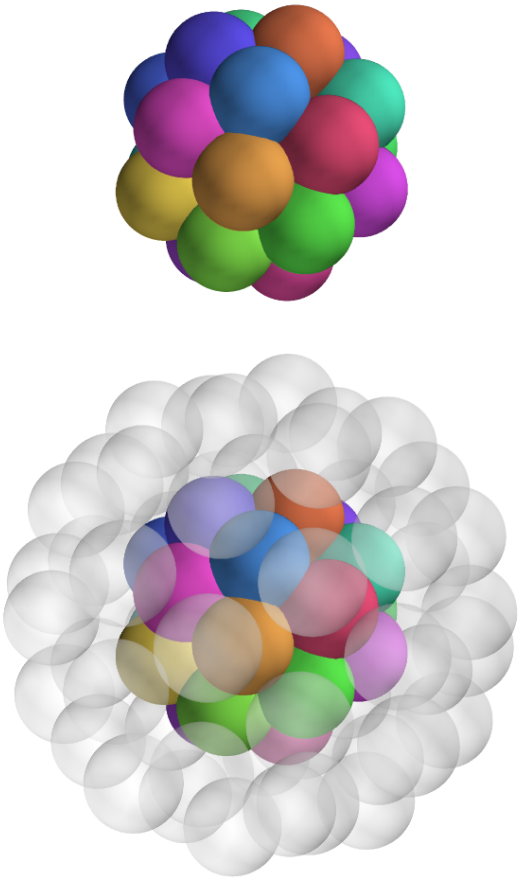


Figure 5.10: 3D ground state cluster without (above) and with (below) non-specific permutable species.

to arrange this type of local minimum, up to the symmetry group for m spheres caging a central sphere. Finally I write the Boltzmann ratio of the local minima to the ground states:

$$\frac{P_2}{P_1} = \frac{\frac{(mN_C)!}{m!^{N_C}} e^{-mN_C\beta U_1}}{N_C^{2/3}! e^{-nN_C\beta U - N_C^{2/3}\beta U_1}}.$$

Saying that P_1 is a ground state is equivalent to making the constraint on the bond strengths such that $m/n|U_1| < |U|$, neglecting the contributing of the surface states, which act to stabilize the ground state as I have defined it. Otherwise the system does not have a well-defined ground state and I expect disordered structures to prevail at all temperatures. Neglecting the energy terms, and solving for the entropy I find:

$$\Delta S = \ln \Omega_{lm} - \ln \Omega_{gs} = \ln(mN_C)! - N_C \ln m! - \ln N_C^{2/3}!,$$

$$\lim_{N_C \to \infty} \frac{\Delta S}{N_C} \to m \ln N_C.$$

The surface states are not sufficiently numerous to contribute in the thermodynamic limit. The more species, the stronger the interactions must be for the ground state to be stable, but the stronger the interactions, the longer it may take to transition out of metastable states. The non-specific sphere concentration acts like a disorder/order parameter (see Fig. 5.12, right). A system equilibrated into the ground state may take a long time to melt if its interactions are very strong even if the non-specific sphere

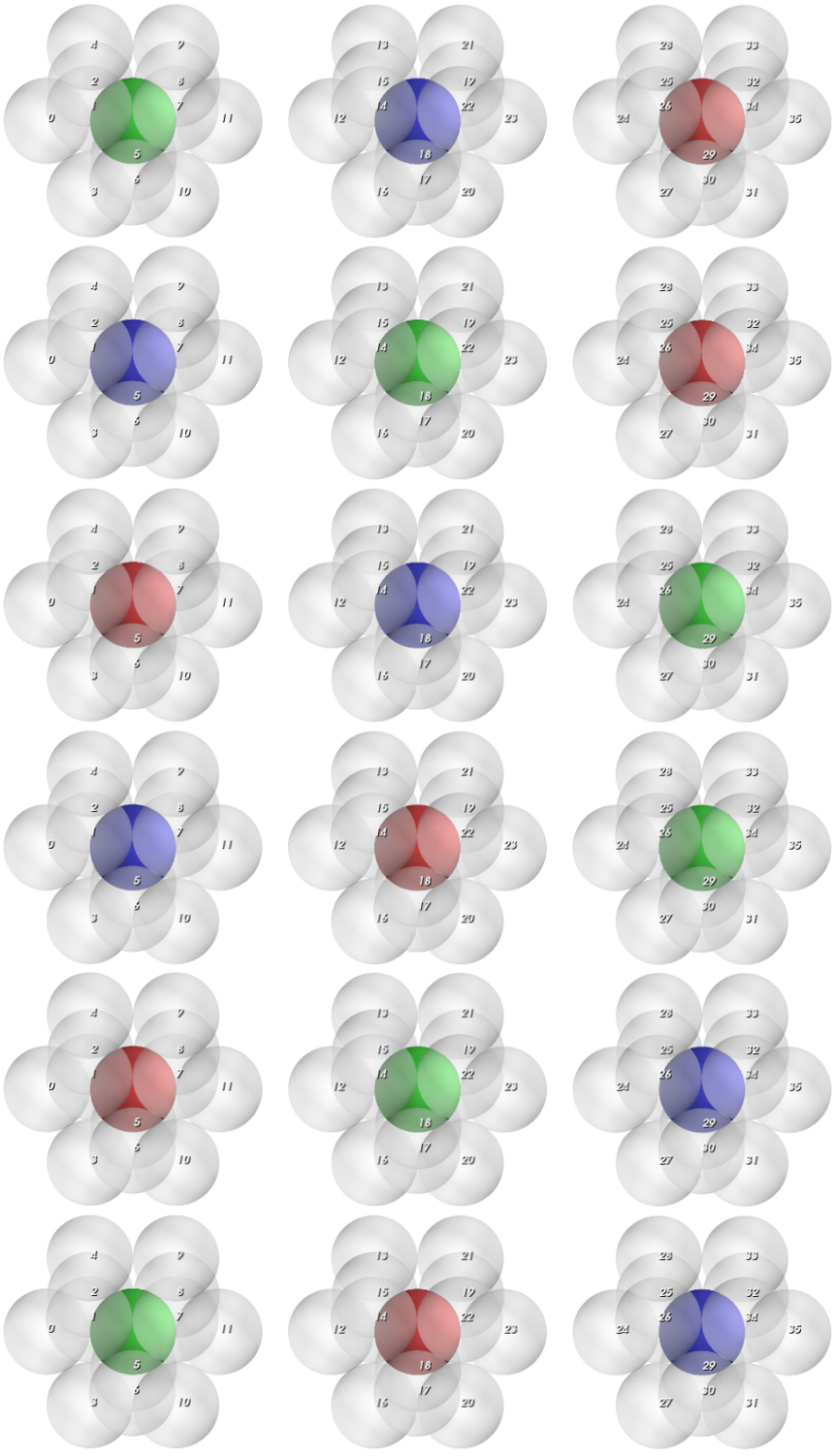


Figure 5.11: 3! distinct but structurally similar local minima for the 3D cluster. I labelled non-specific white spheres with numbers to show that they are distinguishable. The 3 12-particle cages in each of the 3! distinct minima are the same, as the number labels show. There is an additional set of similar 3! local minima for each way to construct the 3 12-sphere cages.

cages are introduced at sufficiently high concentration to shift the equilibrium to disorder, as will be shown more explicitly in a following section.

I have used cages to phrase this argument. I need some kind of geometric constraint such as cages

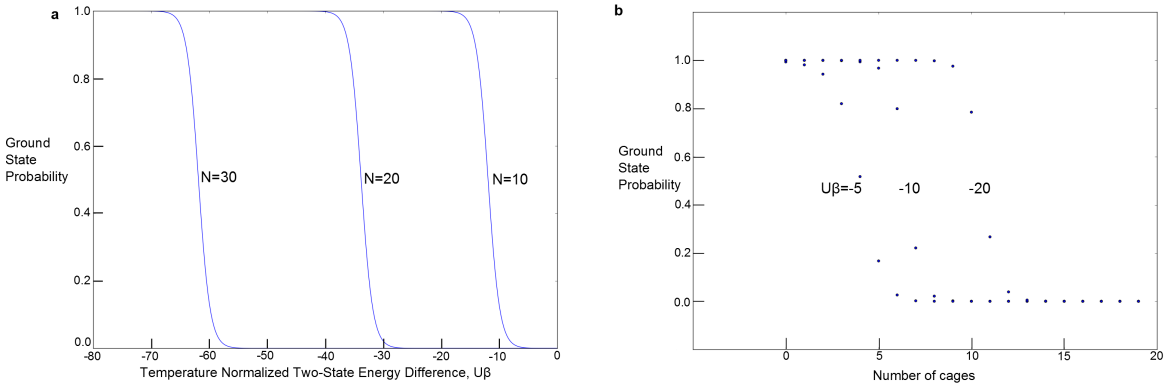


Figure 5.12: 3D model plots. Left: The ground state probability as a function of dimesionless energy/temperature for different numbers of species and particles. 30 species requires about twice as strong bonds or half as high temperature for the ground state to dominate compared to 20 species. Right: The ground state probability as a function of the number of cages (of omniphilic spheres) for different values of dimensionless binding energy/temperature. The number of species is fixed at 20. Sufficiently high number of cages can shift the equilibrium to disorder, but if the cluster is in the ground state and the binding strength is high, it might take a long time for a set of cages introduced into the system to unravel the cluster.

to define the local minima strictly. If I take $m = 1$, I don't have geometric constraints preventing the colorful beads from binding each other. The number of ways to map each of the specific spheres to each of the non-specific spheres is again a factorial in particle number, but a soft or zero-energy deformation can take such spheres into their "globular" ground state. The situation might be different if the particles are physically or chemically anisotropic.

Chain Model

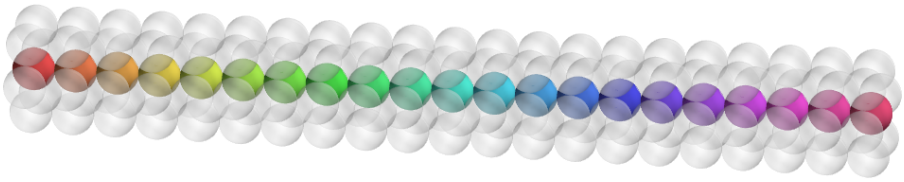


Figure 5.13: Even if a hard constraint prevents this colorful chain from breaking, the permutation entropy of the white spheres prevent the chain from folding into its ground state. Consider a set of rings as shown. Any given set of N rings map to N colorful spheres in N! different ways. A relatively small additional contribution to the permutation entropy comes from the number of different ways to construct N rings.

Another local minimum is one in which the N_C particles maintain the connectivity of a chain (Fig. 5.13). I could enforce this constraint by making certain bonds stronger in the specificity matrix. Now instead of cages, consider rings of $l < m \quad i = 1$ particles. Like the cages, any set of rings can be placed around the particles in $N_C!$ different ways. I expect a similar permutation glass; one with "folded" and "unfolded" states. The rings geometrically prevent any bonds besides those along the chain backbone from forming.

5.3.5 Boltzmann Probabilities and Generalized Permutation Glass

I have described some model systems with sufficiently numerous local minima to make normally extensive variables such as entropy grow faster than linearly with particle number N, even in the large N limit. Similarly, the binding energy required for the Boltzmann probability of the ground state configuration to outweigh the Boltzmann probability of any other configuration depends also on N even in the thermodynamic limit. The more particles and species, the lower the temperature, or the stronger the binding energy, required for the ground state to dominate the distribution.

In these permutation glasses, the normalized energy needed for 50% of the members of the statistical ensemble to occupy the ground state configuration, $\beta U_{50\%}$, generally goes as the number of particles to a power between 0 and 1 times the natural logarithm of the particle number:

$$\beta U_{50\%} \propto N^{\gamma} \ln N,\ 0 < \gamma < 1$$

For example, let's take a partition function nearly equivalent to that of the 2D permutation glass case. In this system, the difference between the potential energy of the ground state and the local minima is a constant, C (where $C = 1$ in our particular model). Here I show that $\gamma = 1$ for such a system. I write the partition function,

$$Z_{\gamma=1} = e^{-N\beta U} + N! e^{-(N-C)\beta U}$$

where C is a positive finite constant, $(0 < C < N)$. The Boltzmann probability for the ground state follows:

$$P_{gs} = \frac{1}{1 + N! e^{C\beta U}}$$

Setting $P_{gs} = .5$, i.e. the Boltzmann probability of the ground state is as high as that of all the other states combined, I find the following:

$$\beta U_{50\%} = -\ln(N!)/C \approx -N \ln(N)/C$$

This result says that the magnitude of the binding energy at which half the ensemble is in the ground state grows with $N^{\gamma} \ln N$ where $\gamma = 1$. As the system size increases linearly, the temperature required to shift the system into its equilibrium ground state decreases faster than linearly. Would such a system at infinite particle number and zero temperature ever reach its ground state?

I analyze of the quasi-1D and 3D permutation glass models similarly. If I neglect all the local minima besides the zero energy states in the quasi-1D system, I write a partition function $Z_{1D} = e^{-N\beta U} + N!$ and find $\gamma = 0$. The actual system may be slightly different, but I certainly expect $\gamma < 1$ as the quasi-1D energy states are skewed heavily towards zero and nearly zero bond configurations. Now I consider the 3D permutation glass and assume that the energy terms linear in N_C exactly cancel, i.e. $mU_1 = nU$. This leaves only the contribute of the surface states and $\gamma = 1/3$. I conclude that the exact thermodynamic behavior of permutation glasses can depend on the distribution of the higher energy states and the dimensionality.

5.4 Scaling of Bond Lifetimes with Number of Species

In this section I discuss how the typical lifetime for a bond must increase with number of species for the ground state to dominant the partition function. The lifetime of a bond is associated with the energy of the

interaction, with stronger bonds having longer typical lifetimes. One way to relate bond lifetime to bond energy is via the Arrhenius approximation,

$$t = \tau e^{-\beta u},$$

where t is the average bond lifetime, τ is a microscopic timescale, and $u < 0$ is the energy of a single bond, which must be overcome by a thermal fluctuation for the bond to be broken. According to this theory, the typical bond lifetime scales as an exponential in the bond energy and the inverse temperature.

5.4.1 Many Highly Specific Spheres

I return to the case of N specific spheres, with partition function $Z_s = e^{-nN\beta u} + \frac{N!}{e}e^{-mN\beta u}$. The ground state probability is

$$P_0 = \frac{1}{1 + \frac{N!}{e}e^{-(m-n)N\beta u}}.$$

Let's call $\theta = u\beta$ and $P_0(\theta_{50\%}) = .5$, so $\theta_{50\%}$ describes the set of u and β 2-tuples at which the ground state is predicted to occur 50% of the time in equilibrium. Solving for $\theta_{50\%}$, I find

$$\theta_{50\%} = \frac{\ln(\frac{N!}{e})}{N(m-n)}.$$

The bond lifetime is a function of θ. The bond lifetime at which ground state is predicted to have 50% yield can be written as

$$t_{50\%} = \tau e^{-\theta_{50\%}}$$

As in the drawings above, $m-n$ is near negative unity, so let's take it to be just that as an approximation. I find that

$$t_{50\%}(N) \approx \tau N.$$

The bond lifetime required for majority ground state occupancy diverges as the number of species (and particles) increases. I do not expect single-component, non-specific systems to have a bond lifetime required for majority ground state occupancy to depend on number of particles. In that case, I find

$$\theta_{50\%} = \frac{1}{m-n} \approx -1$$

and

$$t_{50\%} \approx \tau,$$

independent of the number of particles.

In this example of N differently colored specific particles, each of which binds only to its nearest neighbors in the one ground state, and to no other particles, the bond lifetime for equilibrium assembly $\tau \propto N$. Systems may be conceived that require bond lifetimes for ground state dominance to depend much more strongly with N, such as systems in which the energy difference between the ground state and one or many excited states is arbitrarily small, and systems with omniphilic particles. In the above models, I've considered one way of partitioning the particles and considered the $\sim N!$ ways to permute the particles in that partition, but there are also many ways to partition the particles.

5.4.2 3D Cluster Example

In this section I show how the typical bond lifetime needed for ground state dominance of the partition function might diverge very rapidly with the number of species (and particles) for the 3D model mixture of specific spheres and an omniphilic species. The two-state approximate partition function for this system is

$$Z = N_{ss}^{2/3}!e^{-(nN_{ss}^{2/3}E_{ss}+mN_{ss}E_{ss})\beta} + N_{ss}!e^{-mN_{ns}E_{ns}\beta},$$

where N_{ss} is the number of specific (colorful) spheres, N_{ns} is the number of non-specific spheres, n is a typical coordination number of the non-specific particles at the surface of the ground state globule, m is a coordination number of each of the specific spheres in their ground state globule as well as the coordination number of the omniphilic particles caging the specific particles, E_{ss} is the energy of a specific sphere-specific sphere bond, and E_{ns} is the energy of a non-specific sphere-specific sphere bond ($E_{ss} < 0, E_{ns} < 0$). The probability of the ground state is

$$P_{gs} = \frac{1}{1 + \frac{N!}{N^{2/3}!}e^{(-mN_{ns}E_{ns}+nN_{ss}^{2/3}E_{ns}+mN_{ss}E_{ss})\beta}},$$

and the total potential energy is

$$\beta(nN_{ss}^{2/3}E_{ns} + mN_{ss}E_{ss} - mN_{ns}E_{ns}).$$

I neglect the contribution of the surface term $N^{2/3}E_{ns}$ to the total energy for $N \gg N^{2/3}$. For the globule of colorful particles to be the ground state, this constraint must be satisfied: $mN_{ss}E_{ss} - mN_{ns}E_{ns} < 0$, i.e. $N_{ss}E_{ss} < N_{ns}E_{ns}$. I assume $\Delta U = N_{ss}E_{ss} - N_{ns}E_{ns} = \epsilon < 0$, i.e. the energy difference between ground state and excited states, $\Delta U = \epsilon$, is small, negative and independent of N. Again I call $\theta = \Delta U\beta$ and rewrite the probability as a function of θ,

$$P_{gs}(N, \theta_{50\%}) = 0.5 = \frac{1}{1 + \frac{N!}{N^{2/3}!}e^{-\theta_{50\%}}},$$

and solve for $\theta_{50\%}$:

$$\theta_{50\%} = -\ln \frac{N!}{N^{2/3}!}.$$

As earlier, I use the Arrhenius assumption to relate the bond lifetime to the energy depedence, and in turn to the scaling with system size:

$$t_{50\%} = \tau e^{-\theta_{50\%}} \propto \frac{N!}{N^{2/3}!}.$$

In this model mixture of specific and omniphilic spheres, with fixed and small energy difference ϵ between ground and excited states, the bond lifetime for the ground state to dominate the probability diverges rapidly with system size.

5.5 Slow Diffusion at Low Temperatures

Lowering the temperature to increase the ground state occupancy not only increases bond lifetimes, it also slows down the diffusion of the particles.

The particle diffusivity, D, is related to temperature by the Stokes-Einstein-Smoluchowski[54, 55] relation for spheres,

$$D = \frac{k_B T}{6\pi\eta r},$$

where η is the viscosity and r is the radius of the sphere. Let's assume we are again studying the 3D cluster of spheres with non-specific additives as in the previous section. I fix the particle bond energy ΔU, but lower the temperature, T to $T_{50\%}$, to theoretically achieve majority ground state occupany. Recasting the relation for $\Theta_{50\%}$ from the previous section in terms of $T_{50\%}$, we see that temperature may be related to system size

$$\frac{\Delta U}{k_B T_{50\%}} = -\ln \frac{N!}{N^{2/3}!}.$$

Solving for D in terms of N, I find

$$D = \frac{\Delta U}{6\pi\eta r \ln \frac{N^{2/3}!}{N!}}.$$

The more particles and species, the slower the diffusivity at the temperature required for majority ground state occupancy.

5.6 Transition Rates

Transition rates from one stable state to the next can be thought of as roughly the inverse of the bond lifetimes. The longer the bond lifetime, the slower the transition rate. One equation for the transition rate which is slightly different from the Arrhenius equation is the Eyring equation[56]. In this equation, the rate, k, is similar to the rate from the Arrhenius equation multiplied by the temperature:

$$k \propto T e^{\theta}.$$

As I've shown, the temperature for ground state majority occupancy depends on the number of particles and species, so

$$k_{50\%} \propto T_{50\%} e^{\Delta U T_{50\%}}.$$

Substituting in from my expressions for $T_{50\%}$ and $\theta_{50\%}$ from the previous sections, I find

$$k_{50\%}(N) \propto \frac{N^{2/3}!}{N! \ln \frac{N^{2/3}!}{N!}}.$$

Returning the to Arrhenius approximation instead, I find

$$k_{50\%} \propto \frac{N^{2/3}!}{N!}.$$

In either case, the transition rate goes to zero as N increases. This poses a problem for self-assembly. Although the true equilibrium state of the system is its ground state, the number of transitions per unit time goes to zero. It can take forever for the system to hop from one local minimum to the next.

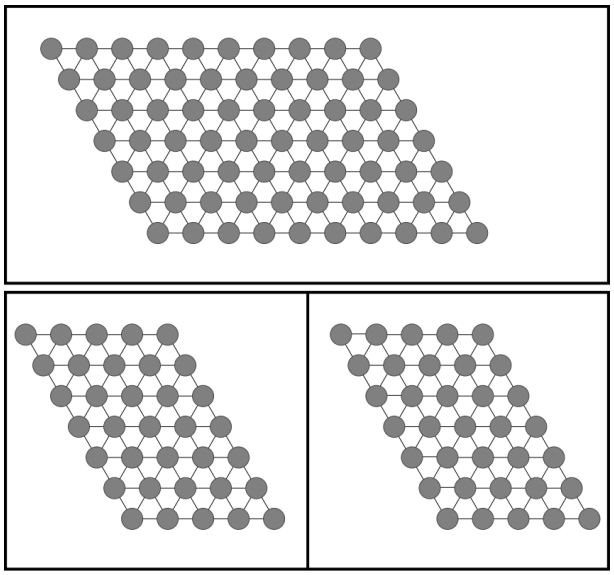


Figure 5.14: Non-specific spheres in one closed system (top) and partitioned between two closed systems (bottom). The number ratio of local minima to global minima here does not depend on whether the system is split between two closed systems or is in only one.

5.7 Entropy of Local Minima in 1 vs. 2 Closed Systems

The large entropy difference between excited states and ground states in systems of many specific spheres carries implications for how such particles condense when assembly takes place in 1 closed system versus 2 or more closed systems.[3] Here I compare the entropy differences between excited and ground states of a non-specific system in 1 then 2 closed systems. Then I repeat the same analysis for a highly specific set of N particles of N different species to show how the number ratio of local minima to ground states is lower if these particles are partitioned into two separate closed systems than in only one.

As earlier in this section, I use a 2-state approximation to model a set of distinguishable, non-specific particles, each of which bonds with any other particle in the system with equal strength. N particles can form the ground state in $\Omega_0 = N!$ different arrangements, and can form local minima packings in $\Omega_E = e^N N!$ different arrangements. If all N particles are in the same closed system,

$$\Delta S_1 = S_E - S_0 = k \ln \frac{\Omega_E}{\Omega_0} = kN,$$

where ΔS_1 is the entropy difference between S_E, the entropy of the excited states, and S_0, the entropy of the ground state. If the system is divided between two closed systems, one with N_A and another with N_B particles ($N = N_A + N_B$), the entropy difference between excited states and ground state is calculated similarly:

$$\Delta S_A = S_{A,E} - S_{A,0} = kN_A,$$

and

$$\Delta S_B = S_{B,E} - S_{B,0} = kN_B,$$

where ΔS_A is the difference between $S_{A,E}$, the entropy of excited states in closed system A, and $S_{A,0}$, the entropy of the ground state in closed system A, and similary for closed system B. The difference

[3] This section applies to isolated systems as well as closed systems.

between the total entropy difference in the 2 closed system case and the 1 closed system case is

$$\Delta\Delta S_{ns} = \Delta S_A + \Delta S_B - \Delta S_1 = 0.$$

Introducing a non-interacting barrier into a system of single-component, interacting collodial spheres, making two systems from one, does not change the overall entropy of their local minima relative to the entropy of ground states.

A set of many highly specific spheres, however, has a lower total entropy of local minima relative to entropy of lowest energy states if portioned into two closed systems as opposed to one. N distinguishable, highly specific spheres can form their ground state in only $\Omega_0 = 1$ arrangement, but can form strictly local minima configurations in about $\Omega_E \approx N!$ different arrangements.[4] If all N particles are in the same closed system,

$$\Delta S_1 = S_E - S_0 = k \ln \frac{\Omega_E}{\Omega_0} = k \ln N!,$$

where ΔS_1 is the entropy difference between S_E, the entropy of the excited states, and S_0 the entropy of the ground state. If the system is divided between two closed systems, one with N_A and another with N_B particles ($N = N_A + N_B$), the maximum[5] entropy difference between excited states and ground state is calculated similarly:

$$\Delta S_A = S_{A,E} - S_{A,0} = k \ln N_A!,$$

and

$$\Delta S_B = S_{B,E} - S_{B,0} = k \ln N_B!,$$

where ΔS_A is the difference between $S_{A,E}$, the entropy of the excited states in closed system A, and $S_{A,0}$, the entropy of the ground state in closed system A, and similarly for closed system B. The difference between the total entropy difference in the 2 closed system case and the 1 closed system case is

$$\Delta\Delta S_{ss} = \Delta S_A + \Delta S_B - \Delta S_1 = k \ln N_A! + k \ln N_B! - k \ln N! < 0.$$

Introducing a non-interacting barrier into a system of N-component, highly specific, interacting colloidal spheres, making two systems from one, decreases the overall entropy of local minima relative to the entropy of ground states. Although there are many different partitions of a set of N specific particles from one container into two containers, any partition decreases the total number of local minima relative to the number of lowest energy states.

[4] Systems that can be approximated this way include the Quasi-1D system above and closed-packed systems of N particles of N different species that have one global potential energy minimizing arrangement but can be trapped either via caging or less-preferential bonding in $\sim N!$ other arrangements.

[5] There are many ways to partition the N spheres into two compartments. Some of these partitions of the set will not have a single well-defined ground state, but will have multiple states with the same lowest energy. Said energy would be higher than the energy of the ground state of a partition of the same number of particles, but who consist of a connected subset of the full N particle ground state. For N_A particles, the number of stable excited states divided by the number of lowest energy states is in any case no more than $N_A!$.

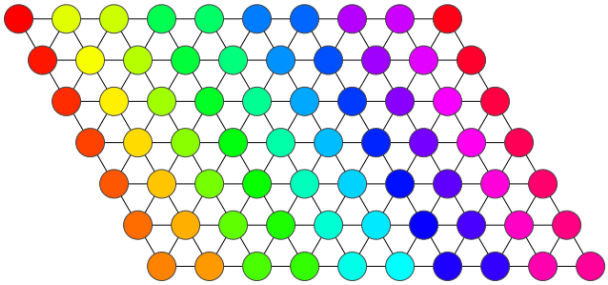


Figure 5.15: A 70 sphere ground state.

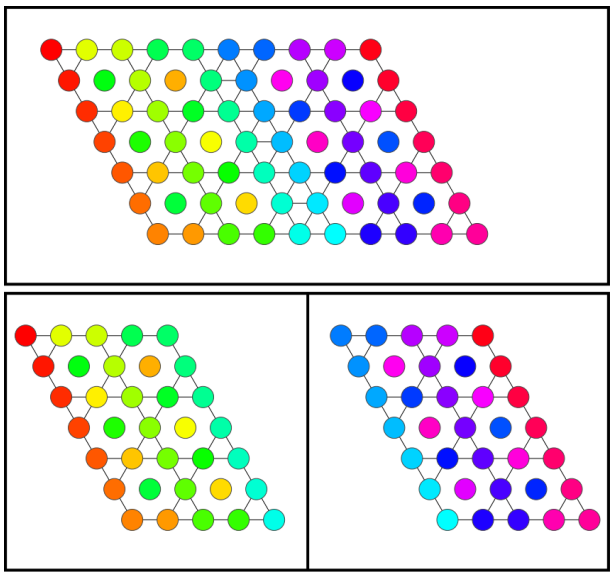


Figure 5.16: Specific spheres in one closed system (top) and partitioned between two closed systems (bottom). The total number of local minima in the system partitioned into 2 closed systems is strictly less than the number of local minima if the particles are all in the same closed system.

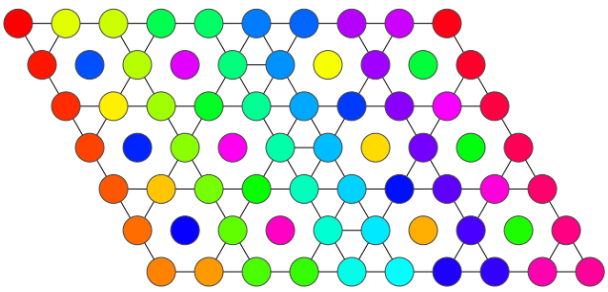


Figure 5.17: One of the minima that is possible in a single closed system, but not in the divided system above.

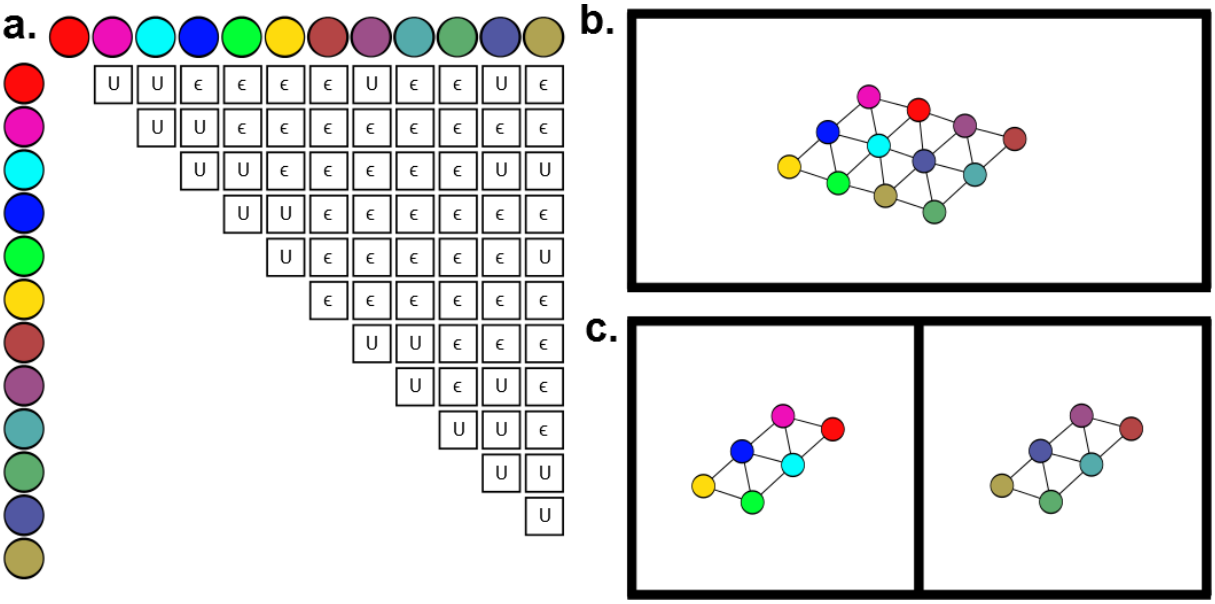


Figure 5.18: N specific particles with less preferential interactions have a lower number ratio of local energy minimizing states to lowest energy states if split between two separate containers than if in one. a) Specificity Matrix. U represents the binding strength of the most preferential interactions, and ϵ represents the weaker binding energy of the less preferential interactions, $U < \epsilon < 0$. b) Ground State. Black bonds represent the most preferential interactions of energy U. $\sim N!$ local minima are possible in one container. c) Split between two containers, the total number of possible local minima is not more than $\sim 2(\frac{N}{2}!)$.

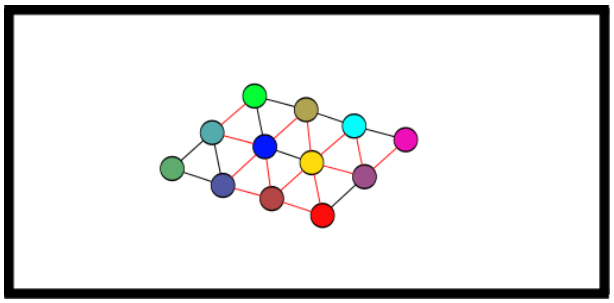


Figure 5.19: One of the minima that is possible in a single closed system, but not in the divided system above. Red bonds represent less preferential bonds of energy ϵ, where $U < \epsilon < 0$

Chapter 6

Conclusion

From the experimental results, I conclude that a set of specific particles assembles into an entropically unfavorable configuration to maximize the number of specific bonds. Such assembly is higher in yield if the number of species is comparable to the number of particles, and if the interaction matrix is an exact match for the target connectivity matrix, as borne out by comparison of the N=6 and N=9 experiments. For small numbers of specific particles, equilibrium assembly of a target structure is observable on laboratory timescales.

The equilibrium assembly of large numbers of specific spheres, however, may not be possible on the same timescales, as shown by my theoretical studies. These systems have a much higher number of distinguishable local minima relative to ground states than single- and few-component systems. Such systems exhibit a kind of paradox: they may require strong interactions for the ground state to be most likely in equilibrium, but the large number of local minima and long bond lifetimes may prevent the system from reaching that ground state over easily observable timescales. Compartmentalizing these systems into 2 or more different closed systems decreases the overall entropy of their local minima relative to the entropy of lowest energy states.

I decided to study colloidal spheres with specific interactions as a way to understand how proteins fold, among other reasons. Although I think I have learned much about how colloidal spheres with specific interactions assemble, many questions remain, both in the assembly of these isotropic systems, and in how they may or may not be analogous to biological molecules.

6.1 Lingering Questions

- Can a ground state in which each particle can deform to a position above and below planes defined by each triplet of its nearest neighbors have other local minima?
- For any highly specificed set of microspheres, assuming the system may be allowed to maximize its bonds before the next particle is added, is there always an order of particle additions that will lead to the ground state in 100% yield in the strong binding limit?
- How might compartmentalization of a set of specific particles into 2 or more closed systems help them assemble?

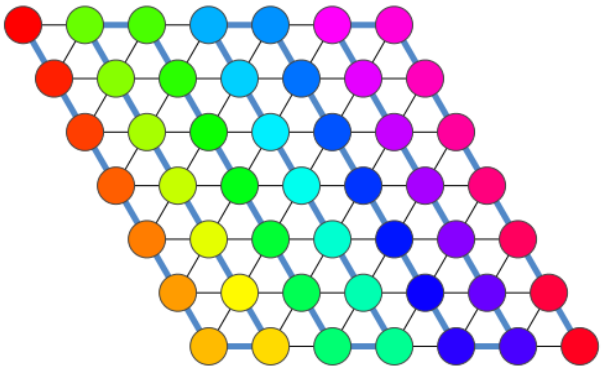


Figure 6.1: A path of irreversible bonds (blue) prevents many local minima from being possible.

- Can hard sphere glasses be thought of as having broken permutation symmetry? If a crystal of hard spheres were divided into N unique clusters and each cluster was solidified or sintered, could the clusters recrystallize at the same temperature and timescales as the free sphere system?
- Do proteins fold into true free energy minima, or is their binding strong enough that they may fold into metastable states and stay there? How many free energy minima can proteins have, and how many can they access in equilibrium?
- Can proteins be effectively modeled by coarse-graining over certain 2-,3- or n-letter words of amino acids? Is there a specificitiy matrix for such amino-acid words?
- How did homochirality evolve, and does it enable the proteins to fold along the nearly the same pathway every time, and avoid many kinetic traps?
- How do chaperones aid in protein folding? How does molecular crowding influence protein folding and complex formation? How do water molecules influence folding, and does the number ratio of water molecules to amino acids shift the protein folding equilibrium? Do less specific proteins shift the equilibrium of complexes of more specific proteins? How does the equilibrium thermodynamics described here relate to the scale-free structure of the protein connectome? Do hubs on the protein connectome regulate the assembly of complexes of more specific proteins?
- How does the increased timescale required for large numbers of specific particle assembly relate, if at all, to living things? Is chemical specificity necessary for life as we know it to exist? How might the concepts here apply to systems much larger than biological molecules, such as organisms and social networks?
- How general are the concepts of specific interactions, and do they apply to anything smaller than biological molecules, such as fundamental particles? Is the omniphilicility of a fundamental particle necessary for it to slow other particles down? ...

Bibliography

[1] Jean-Baptiste Michel et al. “Quantitative Analysis of Culture Using Millions of Digitized Books”. en. In: *Science* 331.6014 (Jan. 2011). PMID: 21163965, pp. 176–182. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1199644. URL: http://www.sciencemag.org.ezp-prod1.hul.harvard.edu/content/331/6014/176 (visited on 04/29/2014).

[2] George M. Whitesides and Bartosz Grzybowski. “Self-Assembly at All Scales”. en. In: *Science* 295.5564 (Mar. 2002). PMID: 11923529, pp. 2418–2421. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1070821. URL: http://www.sciencemag.org.ezp-prod1.hul.harvard.edu/content/295/5564/2418 (visited on 04/29/2014).

[3] Sandra R. Whaley et al. “Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly”. In: *Nature* 405.6787 (June 2000), pp. 665–668. ISSN: 0028-0836. DOI: 10.1038/35015043. URL: http://www.nature.com.ezp-prod1.hul.harvard.edu/nature/journal/v405/n6787/full/405665a0.html (visited on 04/29/2014).

[4] William K. Purves et al. *Life: The Science of Biology*. Vol. 6. W. H. Freeman, 2000, pp. 358–373. ISBN: 0716738732. URL: http://www.amazon.com/Life-Science-William-K-Purves/dp/0716738732%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0716738732.

[5] Yonggang Ke et al. “Three-Dimensional Structures Self-Assembled from DNA Bricks”. en. In: *Science* 338.6111 (Nov. 2012). PMID: 23197527, pp. 1177–1183. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1227268. URL: http://www.sciencemag.org.ezp-prod1.hul.harvard.edu/content/338/6111/1177 (visited on 02/14/2014).

[6] D B Lukatsky, B M Mulder, and D Frenkel. “Designing ordered DNA-linked nanoparticle assemblies”. In: *Journal of Physics: Condensed Matter* 18.18 (May 2006), S567–S580. ISSN: 0953-8984, 1361-648X. DOI: 10.1088/0953-8984/18/18/S05. URL: http://iopscience.iop.org/0953-8984/18/18/S05/pdf/0953-8984_18_18_S05.pdf (visited on 02/14/2014).

[7] Nicholas Licata and Alexei Tkachenko. “Errorproof programmable self-assembly of DNA-nanoparticle clusters”. In: *Physical Review E* 74.4 (Oct. 2006). ISSN: 1539-3755. DOI: 10.1103/PhysRevE.74.041406. URL: http://link.aps.org/doi/10.1103/PhysRevE.74.041406.

[8] Sahand Hormoz and Michael P. Brenner. “Design principles for self-assembly with short-range interactions”. en. In: *Proceedings of the National Academy of Sciences* 108.13 (Mar. 2011). PMID: 21383135, pp. 5193–5198. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1014094108. URL: http://www.pnas.org/content/108/13/5193 (visited on 02/17/2014).

[9] Zorana Zeravcic and Michael P. Brenner. “Self-replicating colloidal clusters”. en. In: *Proceedings of the National Academy of Sciences* (Jan. 2014). PMID: 24449887, p. 201313601. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1313601111. URL: http://www.pnas.org/content/early/2014/01/17/1313601111 (visited on 02/14/2014).

[10] Guangnan Meng et al. “The Free-Energy Landscape of Clusters of Attractive Hard Spheres”. en. In: *Science* 327.5965 (Jan. 2010). PMID: 20110500, pp. 560–563. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1181263. URL: http://www.sciencemag.org.ezp-prod1.hul.harvard.edu/content/327/5965/560 (visited on 02/16/2014).

[11] “Real-space studies of the structure and dynamics of self-assembled colloidal clusters - Faraday Discussions (RSC Publishing)”. In: *Faraday Discussions* (2012). URL: http://pubs.rsc.org.ezp-prod1.hul.harvard.edu/en/Content/ArticleLanding/2012/FD/c2fd20061a#!divAbstract (visited on 04/30/2014).

[12] Schmucker et al. “Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity”. In: *Cell* 101.6 (2000). ISSN: 0092-8674.

[13] Woj M. Wojtowicz et al. “Alternative Splicing of Drosophila Dscam Generates Axon Guidance Receptors that Exhibit Isoform-Specific Homophilic Binding”. English. In: *Cell* 118.5 (Mar. 2004). PMID: 15339666, pp. 619–633. ISSN: 0092-8674. DOI: 10.1016/j.cell.2004.08.021. URL: http://www.cell.com/article/S0092867404007962/abstract (visited on 04/28/2014).

[14] Paul L. Biancaniello, Anthony J. Kim, and John C. Crocker. “Colloidal Interactions and Self-Assembly Using DNA Hybridization”. In: *Physical Review Letters* 94.5 (Feb. 2005), p. 058302. DOI: 10.1103/PhysRevLett.94.058302. URL: http://link.aps.org/doi/10.1103/PhysRevLett.94.058302 (visited on 07/09/2013).

[15] W. Benjamin Rogers and John C. Crocker. “Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling”. en. In: *Proceedings of the National Academy of Sciences* 108.38 (Sept. 2011). PMID: 21896714, pp. 15687–15692. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1109853108. URL: http://www.pnas.org/content/108/38/15687 (visited on 07/09/2013).

[16] Robert M. Dirks et al. “Paradigms for computational nucleic acid design”. en. In: *Nucleic Acids Research* 32.4 (Mar. 2004). PMID: 14990744, pp. 1392–1403. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkh291. URL: http://nar.oxfordjournals.org.ezp-prod1.hul.harvard.edu/content/32/4/1392 (visited on 04/29/2014).

[17] Joshua Lederberg. “Genes and Antibodies”. In: *Science* 129.3364 (June 1959), pp. 1649–1653. ISSN: 0036-8075. URL: http://www.jstor.org/stable/1756525 (visited on 02/25/2014).

[18] Miranda Holmes-Cerfon, Steven J. Gortler, and Michael P. Brenner. “A geometrical approach to computing free-energy landscapes from short-ranged potentials”. en. In: *Proceedings of the National Academy of Sciences* 110.1 (Jan. 2013). PMID: 23248296, E5–E14. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1211720110. URL: http://www.pnas.org/content/110/1/E5 (visited on 02/14/2014).

[19] Nicholas Licata and Alexei Tkachenko. “Self-assembly of DNA-coded nanoclusters”. In: *Physical Review E* 74.4 (Oct. 2006). ISSN: 1539-3755. DOI: 10.1103/PhysRevE.74.040401. URL: http://link.aps.org/doi/10.1103/PhysRevE.74.040401.

[20] Alexei V. Tkachenko. “Theory of Programmable Hierarchic Self-Assembly”. In: *Physical Review Letters* 106.25 (June 2011), p. 255501. DOI: 10.1103/PhysRevLett.106.255501. URL: http://link.aps.org/doi/10.1103/PhysRevLett.106.255501 (visited on 03/05/2014).

[21] Jonathan D. Halverson and Alexei V. Tkachenko. “DNA-programmed mesoscopic architecture”. In: *Physical Review E* 87.6 (June 2013), p. 062310. DOI: 10.1103/PhysRevE.87.062310. URL: http://link.aps.org/doi/10.1103/PhysRevE.87.062310 (visited on 03/05/2014).

[22] Chad A. Mirkin et al. “A DNA-based method for rationally assembling nanoparticles into macroscopic materials”. en. In: *Nature* 382.6592 (Aug. 1996), pp. 607–609. DOI: 10.1038/382607a0. URL: http://www.nature.com/nature/journal/v382/n6592/abs/382607a0.html (visited on 07/11/2013).

[23] A. Paul Alivisatos et al. “Organization of ’nanocrystal molecules’ using DNA”. In: *Nature* 382.6592 (Aug. 1996), pp. 609–611. DOI: 10.1038/382609a0. URL: http://www.nature.com.ezp-prod1.hul.harvard.edu/nature/journal/v382/n6592/abs/382609a0.html (visited on 07/11/2013).

[24] Qian Chen et al. “3D Motion of DNA-Au Nanoconjugates in Graphene Liquid Cell Electron Microscopy”. In: *Nano Letters* 13.9 (Sept. 2013), pp. 4556–4561. ISSN: 1530-6984. DOI: 10.1021/nl402694n. URL: http://dx.doi.org/10.1021/nl402694n (visited on 02/24/2014).

[25] Alexander J. Mastroianni, Shelley A. Claridge, and A. Paul Alivisatos. “Pyramidal and Chiral Groupings of Gold Nanocrystals Assembled Using DNA Scaffolds”. In: *Journal of the American Chemical Society* 131.24 (June 2009), pp. 8455–8459. ISSN: 0002-7863. DOI: 10.1021/ja808570g. URL: http://dx.doi.org/10.1021/ja808570g (visited on 07/30/2013).

[26] Carissa M. Soto, Amritha Srinivasan, and Banahalli R. Ratna. “Controlled Assembly of Mesoscale Structures Using DNA as Molecular Bridges”. In: *Journal of the American Chemical Society* 124.29 (July 2002), pp. 8508–8509. ISSN: 0002-7863. DOI: 10.1021/ja017653f. URL: http://dx.doi.org/10.1021/ja017653f (visited on 07/11/2013).

[27] Kun-Ta Wu et al. “Polygamous particles”. en. In: *Proceedings of the National Academy of Sciences* 109.46 (Nov. 2012). PMID: 23100534, pp. 18731–18736. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1207356109. URL: http://www.pnas.org/content/109/46/18731 (visited on 07/11/2013).

[28] Nicholas B. Schade et al. “Tetrahedral Colloidal Clusters from Random Parking of Bidisperse Spheres”. In: *Physical Review Letters* 110.14 (Apr. 2013), p. 148303. DOI: 10.1103/PhysRevLett.110.148303. URL: http://link.aps.org/doi/10.1103/PhysRevLett.110.148303 (visited on 07/11/2013).

[29] Marie-Pierre Valignat et al. “Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids”. en. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.12 (Mar. 2005). PMID: 15758072, pp. 4225–4229. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0500507102. URL: http://www.pnas.org/content/102/12/4225 (visited on 07/09/2013).

[30] Anthony J. Kim et al. “Probing interfacial equilibration in microsphere crystals formed by DNA-directed assembly”. en. In: *Nature Materials* 8.1 (Jan. 2009), pp. 52–55. ISSN: 1476-1122. DOI: 10.1038/nmat2338. URL: http://www.nature.com/nmat/journal/v8/n1/abs/nmat2338.html (visited on 07/09/2013).

[31] Marie T. Casey et al. “Driving diffusionless transformations in colloidal crystals using DNA hand-shaking”. en. In: *Nature Communications* 3 (Nov. 2012), p. 1209. DOI: 10.1038/ncomms2206. URL: http://www.nature.com/ncomms/journal/v3/n11/full/ncomms2206.html (visited on 07/09/2013).

[32] Q Xu et al. “Subdiffusion of a sticky particle on a surface”. eng. In: *Physical review letters* 106.22 (June 2011). PMID: 21702635, p. 228102. ISSN: 1079-7114.

[33] W. Benjamin Rogers, Talid Sinno, and John C. Crocker. “Kinetics and non-exponential binding of DNA-coated colloids”. en. In: *Soft Matter* 9.28 (June 2013), pp. 6412–6417. ISSN: 1744-6848. DOI: 10.1039/C3SM50593F. URL: http://pubs.rsc.org/en/content/articlelanding/2013/sm/c3sm50593f (visited on 07/09/2013).

[34] Rémi Dreyfus et al. “Simple quantitative model for the reversible association of DNA coated colloids”. eng. In: *Physical review letters* 102.4 (Jan. 2009). PMID: 19257481, p. 048301. ISSN: 0031-9007.

[35] Daisuke Hattori et al. “Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms”. In: *Nature* 461.7264 (Oct. 2009), pp. 644–648. ISSN: 0028-0836. DOI: 10.1038/nature08431. URL: http://www.nature.com.ezp-prod1.hul.harvard.edu/nature/journal/v461/n7264/abs/nature08431.html (visited on 05/01/2014).

[36] Milton Kerker. *The Scattering of Light, and other Electromagnetic Radiation*. New York, Academic Press, 1969.

[37] Pierre Aymard, Dominique Durand, and Taco Nicolai. “The effect of temperature and ionic strength on the dimerisation of -lactoglobulin”. In: *International Journal of Biological Macromolecules* 19.3 (Oct. 1996), pp. 213–221. ISSN: 0141-8130. DOI: 10.1016/0141-8130(96)01130-0. URL: http://www.sciencedirect.com/science/article/pii/0141813096011300 (visited on 04/30/2014).

[38] Bruce J. Berne, Robert Pecora, and Physics. *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics (Dover Books on Physics)*. Dover Publications, 2000. ISBN: 0486411559. URL: http://www.amazon.com/Dynamic-Light-Scattering-Applications-Chemistry/dp/0486411559%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0486411559.

[39] Rob Meijers et al. "Structural basis of Dscam isoform specificity". In: *Nature* 449.7161 (Sept. 2007), pp. 487–491. ISSN: 0028-0836. DOI: 10.1038/nature06147. URL: http://www.nature.com.ezp-prod1.hul.harvard.edu/nature/journal/v449/n7161/full/nature06147.html (visited on 04/28/2014).

[40] Bruno H. Zimm. "Molecular Theory of the Scattering of Light in Fluids". In: *The Journal of Chemical Physics* 13.4 (Apr. 1945), pp. 141–145. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1724013. URL: http://scitation.aip.org.ezp-prod1.hul.harvard.edu/content/aip/journal/jcp/13/4/10.1063/1.1724013 (visited on 05/01/2014).

[41] Stephen W. Provencher. "CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations". In: *Computer Physics Communications* 27 (Sept. 1982), pp. 229–242. ISSN: 0010-4655. DOI: 10.1016/0010-4655(82)90174-6. URL: http://adsabs.harvard.edu/abs/1982CoPhC..27..229P (visited on 05/01/2014).

[42] Rajendra Uppoor and Paul J. Niebergall. "-D(+) Glucose-Glucose Oxidase-Catalase for Use as an Antioxidant System". In: *Pharmaceutical Development and Technology* 1.2 (Jan. 1996), pp. 127–134. ISSN: 1083-7450, 1097-9867. DOI: 10.3109/10837459609029887. URL: http://informahealthcare.com/doi/abs/10.3109/10837459609029887 (visited on 07/12/2013).

[43] K. Sakata-Sogawa et al. "Direct measurement of DNA molecular length in solution using optical tweezers: detection of looping due to binding protein interactions". en. In: *European Biophysics Journal* 27.1 (Jan. 1998), pp. 55–61. ISSN: 0175-7571, 1432-1017. DOI: 10.1007/s002490050110. URL: http://link.springer.com/article/10.1007/s002490050110 (visited on 07/12/2013).

[44] Markita P. Landry et al. "Characterization of Photoactivated Singlet Oxygen Damage in Single-Molecule Optical Trap Experiments". In: *Biophysical Journal* 97.8 (Oct. 2009), pp. 2128–2136. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2009.07.048. URL: http://www.sciencedirect.com/science/article/pii/S0006349509013149 (visited on 07/12/2013).

[45] Luc P. Faucheux and Albert J. Libchaber. "Confined Brownian motion". In: *Physical Review E* 49.6 (June 1994), pp. 5158–5163. DOI: 10.1103/PhysRevE.49.5158. URL: http://link.aps.org/doi/10.1103/PhysRevE.49.5158 (visited on 04/30/2014).

[46] Mark I. Wallace et al. "Non-Arrhenius kinetics for the loop closure of a DNA hairpin". en. In: *Proceedings of the National Academy of Sciences* 98.10 (May 2001). PMID: 11320222, pp. 5584–5589. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.101523498. URL: http://www.pnas.org/content/98/10/5584 (visited on 04/30/2014).

[47] Phillip H. Rogers et al. “Selective, Controllable, and Reversible Aggregation of Polystyrene Latex Microspheres via DNA Hybridization”. In: *Langmuir* 21.12 (June 2005), pp. 5562–5569. ISSN: 0743-7463. DOI: 10.1021/la046790y. URL: http://dx.doi.org/10.1021/la046790y (visited on 04/30/2014).

[48] Allison M. Yake et al. “Fabrication of Colloidal Doublets by a Salting OutQuenchingFusing Technique”. In: *Langmuir* 22.22 (Oct. 2006), pp. 9135–9141. ISSN: 0743-7463. DOI: 10.1021/la061339n. URL: http://dx.doi.org/10.1021/la061339n (visited on 04/30/2014).

[49] David M. Kaz et al. “Physical ageing of the contact line on colloidal particles at liquid interfaces”. In: *Nature Materials* 11.2 (Dec. 2011), pp. 138–142. ISSN: 1476-1122, 1476-4660. DOI: 10.1038/nmat3190. URL: http://www.nature.com.ezp-prod1.hul.harvard.edu/nmat/journal/v11/n2/fig_tab/nmat3190_F2.html (visited on 04/30/2014).

[50] David P. Robbins. “The probability that neighbors remain neighbors after random rearrangements”. In: *American Mathematical Monthly* 87.2 (1980), pp. 122–124. ISSN: 0002-9890 (print), 1930-0972 (electronic).

[51] Bengt Aspvall and Frank Liang. “The Dinner Table Problem”. In: *Stanford Dept. of Computer Science Report, Dec. 1980.* (). URL: http://i.stanford.edu/pub/cstr/reports/cs/tr/80/829/CS-TR-80-829.pdf (visited on 04/29/2014).

[52] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. “Exploring network structure, dynamics, and function using NetworkX”. In: *Proceedings of the 7th Python in Science Conference (SciPy2008)*. Pasadena, CA USA, Aug. 2008, pp. 11–15.

[53] Natalie Arkus, Vinothan N. Manoharan, and Michael P. Brenner. “Minimal Energy Clusters of Hard Spheres with Short Range Attractions”. In: *Physical Review Letters* 103.11 (Sept. 2009), p. 118303. DOI: 10.1103/PhysRevLett.103.118303. URL: http://link.aps.org/doi/10.1103/PhysRevLett.103.118303 (visited on 04/23/2014).

[54] A. Einstein. “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen”. en. In: *Annalen der Physik* 322.8 (Jan. 1905), pp. 549–560. ISSN: 1521-3889. DOI: 10.1002/andp.19053220806. URL: http://onlinelibrary.wiley.com.ezp-prod1.hul.harvard.edu/doi/10.1002/andp.19053220806/abstract (visited on 05/02/2014).

[55] M. von Smoluchowski. “Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen”. en. In: *Annalen der Physik* 326.14 (Jan. 1906), pp. 756–780. ISSN: 1521-3889. DOI: 10.1002/andp.19063261405. URL: http://onlinelibrary.wiley.com.ezp-prod1.hul.harvard.edu/doi/10.1002/andp.19063261405/abstract (visited on 05/02/2014).

[56] Henry Eyring. “The Activated Complex in Chemical Reactions”. In: *The Journal of Chemical Physics* 3.2 (Feb. 1935), pp. 107–115. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1749604. URL: http://scitation.aip.org.ezp-prod1.hul.harvard.edu/content/aip/journal/jcp/3/2/10.1063/1.1749604 (visited on 05/04/2014).

[57] Manhee Lee et al. “Synchronized reinjection and coalescence of droplets in microfluidics”. en. In: *Lab on a Chip* 14.3 (2014), p. 509. ISSN: 1473-0197, 1473-0189. DOI: 10.1039/c3lc51214b. URL: http://xlink.rsc.org/?DOI=c3lc51214b (visited on 04/28/2014).

[58] Robert M. Dirks et al. “Thermodynamic Analysis of Interacting Nucleic Acid Strands”. en. In: *SIAM Review* 49.1 (Jan. 2007), pp. 65–88. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/060651100. URL: http://authors.library.caltech.edu/7706/ (visited on 04/29/2014).

[59] Joseph N. Zadeh et al. “NUPACK: Analysis and design of nucleic acid systems”. en. In: *Journal of Computational Chemistry* 32.1 (Jan. 2011), pp. 170–173. ISSN: 1096-987X. DOI: 10.1002/jcc.21596. URL: http://onlinelibrary.wiley.com.ezp-prod1.hul.harvard.edu/doi/10.1002/jcc.21596/abstract (visited on 04/29/2014).

[60] Anthony J. Kim, Vinothan N. Manoharan, and John C. Crocker. “Swelling-Based Method for Preparing Stable, Functionalized Polymer Colloids”. In: *Journal of the American Chemical Society* 127.6 (Feb. 2005), pp. 1592–1593. ISSN: 0002-7863. DOI: 10.1021/ja0450051. URL: http://dx.doi.org/10.1021/ja0450051 (visited on 04/30/2014).

Appendix A

N=6 Local Minimum Details

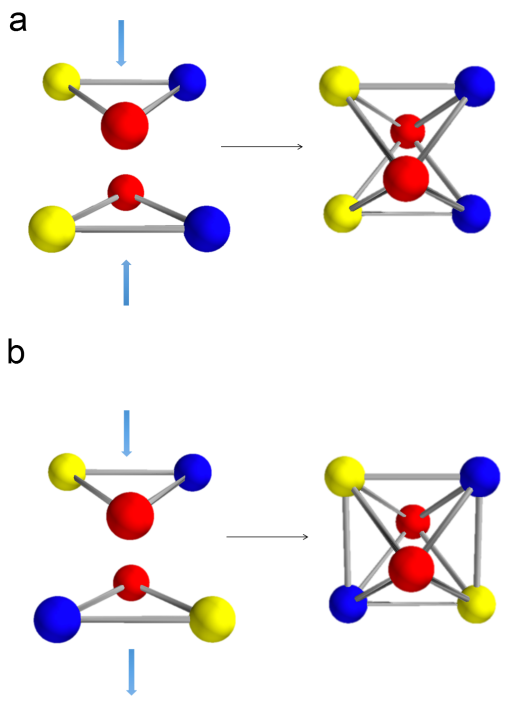


Figure A.1: This local minimum can be thought of as two trimers with opposite relative chirality.

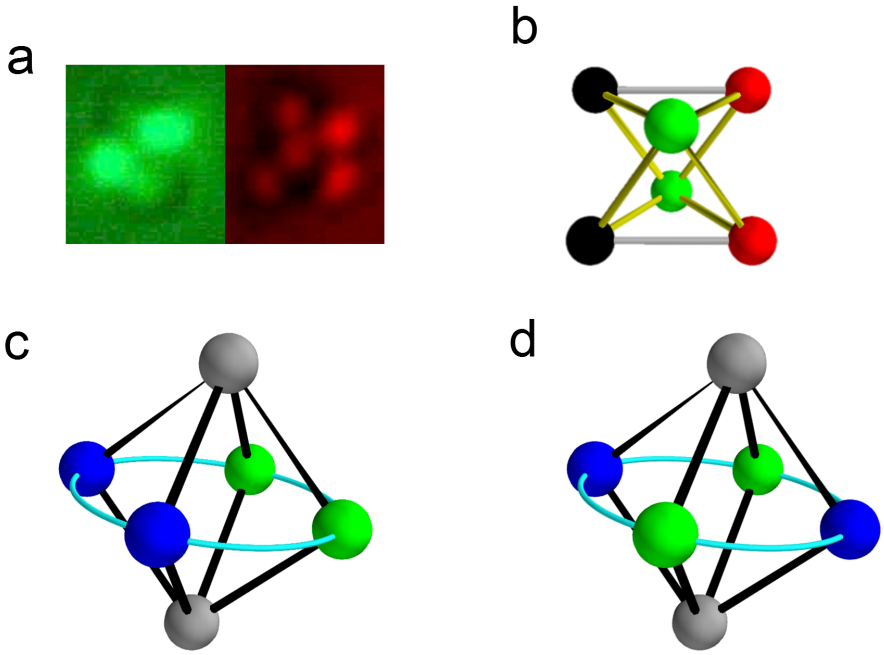


Figure A.2: a.) Local Minimum b.) Any one of the 8 bonds could be broken to allow the structure to switch into the ground state. c.) This 8 bond structure is in the basin of the local minium. d.) This 8 bond structure is in the basin of the ground state.

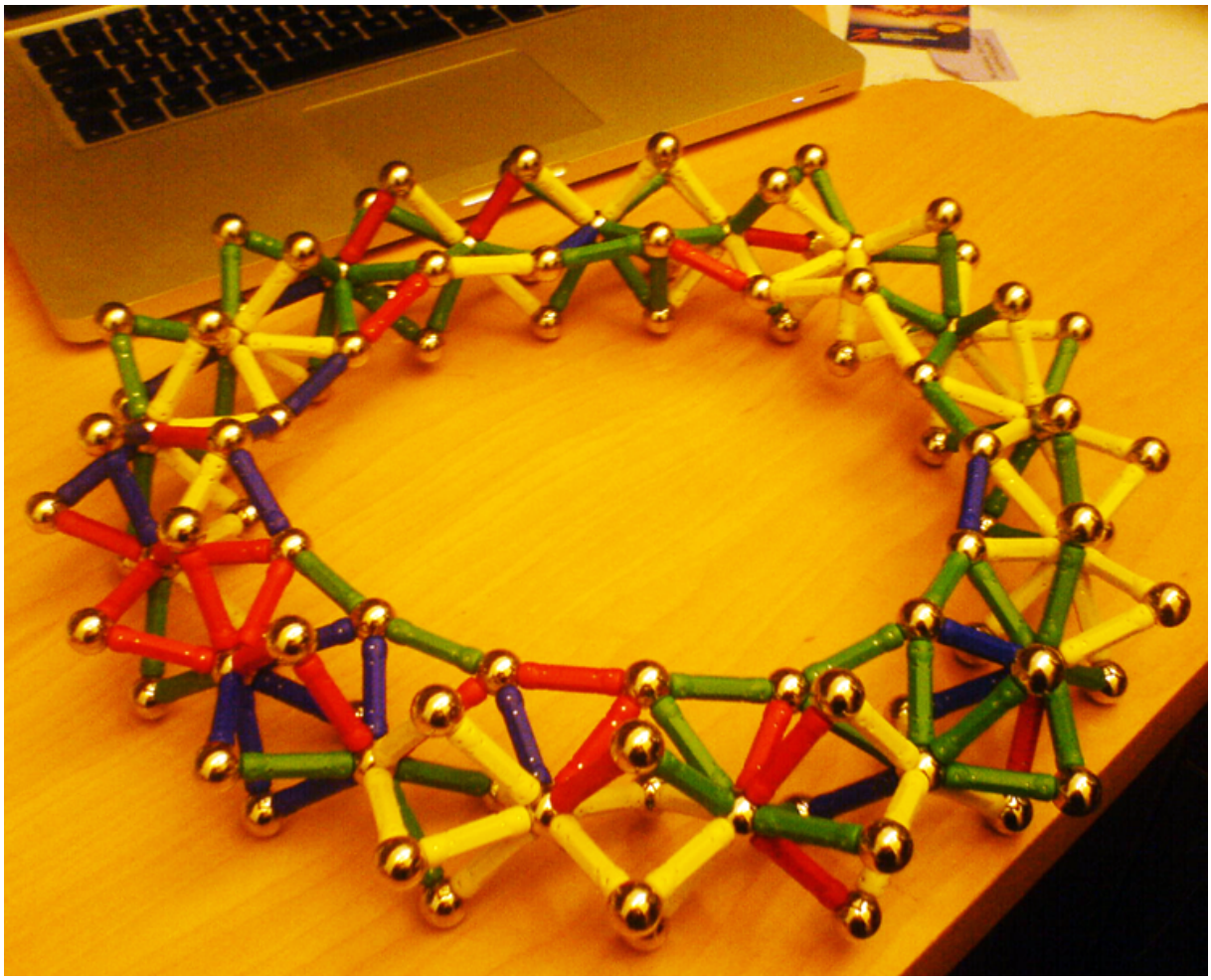


Figure A.3: The highly specific octahedron requires that 4 particles be alphabetically constrained to be in a loop, but geometrically constrained to be in a plane. The structure above, however, does not constrain any set of 4 or more particles to be both in a loop and coplanar. If specified with each particle a unique species, need it have any local minima, besides this ground state?

Appendix B

Controlling Particle and Species Numbers

When studying or building clusters of multiple species, it is important to have a method of controlling or finding a particular set of particles. For example, I set out to study 6 particles, 2 each of 3 differently colored species. I could make droplets or spread the particles into microwells, but the odds of finding a well with that particular set of colors is fairly low, about 1/50 to 1/100. Therefore, I'd prefer to have a more precise method of manipulating the particles into a small volume. I spent some time on a two main approaches to solving this problem in a scalable way. The first involved fabricating anisotropic, optically transparent and non-sticky well-plats to be used with a flow cytometer. The second involved reinjecting, synchronizing and coalescing droplets in a PDMS microfluidic device.

B.1 Anisotropic Wells Plates with Fluorescence-Assisted Sorting

I designed and fabricated plates of microwells with few mm wide tops tapering to 10s of micron or smaller in width at the bottom, compatible with both fluorecence activated cell sorting (FACS) and with optical and fluorescence microscopy. To prepare these well plates, I start with a bare silicon wafer. Then I deposit via chemical vapor low-stress silicon nitride around the surface. Then I UV crosslink an SU-8 grid pattern, i.e. squares of width 4mm, carefully alignd to the silicon crystal axes. Next, I submerge the wafer in hot KOH to etch the anisotropic well profile. Then I fill the wafer with a UV curable adhesive (Norland Optical) and cure it under a UV lamp. Then I cure PDMS around this plastic master and use the PDMS plate for the experiment. A schematic of this work flow is in figure B.1.

I perform the KOH etching process very carefully at high temperature (see Fig. B.2). The etching solution is >20% KOH in water. It is in a bath held at 80 ° C outfitted with a condenser. I cover the glass bath in tin foil so that the wafer temperature is uniform. I also coat the backside of the wafter with kryton fluorinated grease and a piece of fluorinated plastic and wrap its edges in PTFE tape to prevent any unwanted etching through any cracks or scratches in the silicon nitride coating.

After the KOH etch the silicon wafer contains many anisotropic wells (see Fig. B.3).

I fill the wells with NOA under vaccumm so remove any air bubbles. Then I cure the NOA under a UV lamp for at least 10 minutes. I carefully peel the cured optical adhesive with tweezers. This leads to a master for making the PDMS plates (see Fig. B.4).

To make the PDMS plates, I start by pouring about 10 mls of PDMS into a petri dish. Then I press the optical adhesive master into the PDMS, tips down. I place some weights on top of the master to apply pressure. Then I cure the PDMS at elevated temperature for at least about an hour, and usually longer. Then I carefully peel the PDMS from the master. Then I PEGylate the PDMS. That process involves first swelling the PDMS wells in acetone with benzophenon photointiator, deswelling, then washing excess benzophenone away with ethanol. Then I submerge the wells in a PEG-diacrylate solution and place them under UV illumination, photopolymerzing PEG onto the PDMS surface. Then I swell the PDMS-PEG wells to remove photoiniator. Finally I plasma bond the PDMS wells onto a thin (1-5 micron) spin-coated PDMS layer. Finally I plasma bond this PDMS-backed-PDMS-PEG well plate onto a large, thin glass slide (Fig. B.5).

I fill the plate with buffer and bring it to the FACS machine to fill each well with the precise amount of microspheres, such as 2 each of a given type (Fig. B.6).

B.2 Microfluidics

In a collaborative work with Manhee Lee and others, I designed and fabricated microfluidic chips and emulsions for controlling particle number. We reinjected two emulsions into a PDMS device, synchronized the rates via pressure balance, and coalesced pairs and triplets of droplets using an electric field. I wrote python code to perform image analysis on Gb-sized high-speed video to measure inter-droplet spacing as a function of applied pressure (see Fig. B.7). This work is reported in Lab on a Chip [57].

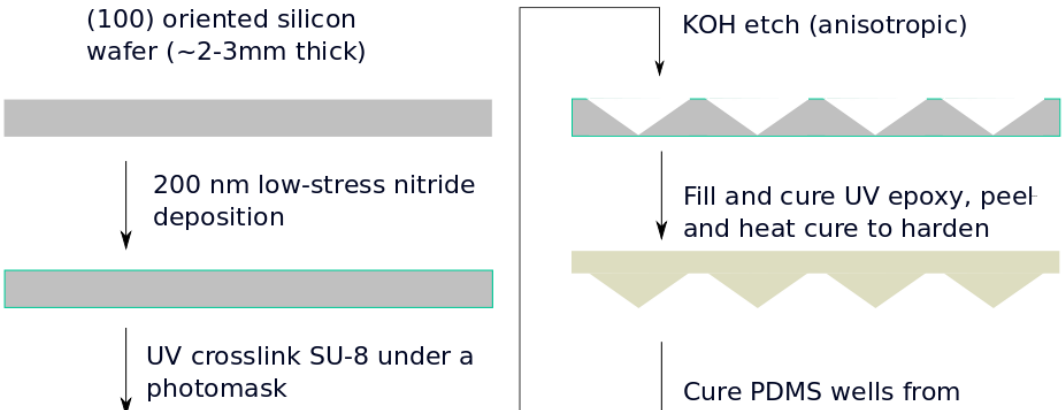


Figure B.1: Step-by-step procedure flow diagram for making anisotropic well plates.

Figure B.2: Left: experimentalist wearing a facemask after etching wafer. Right: Hot KOH wafer etching set-up.

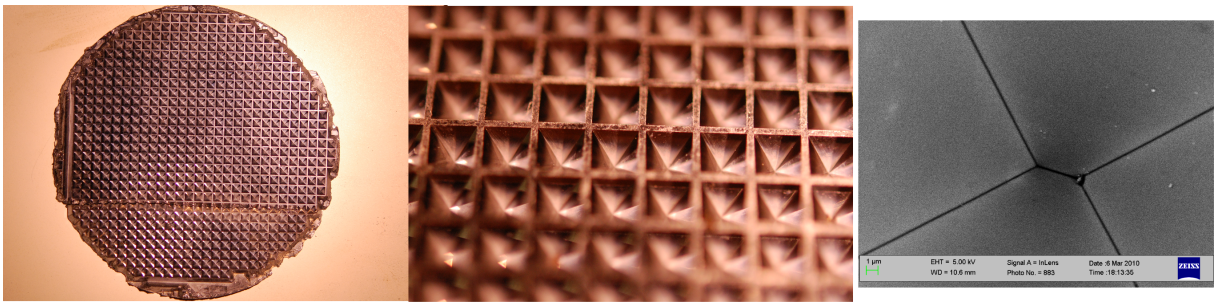


Figure B.3: ~2.7mm deep inverse pyramids in silicon wafer

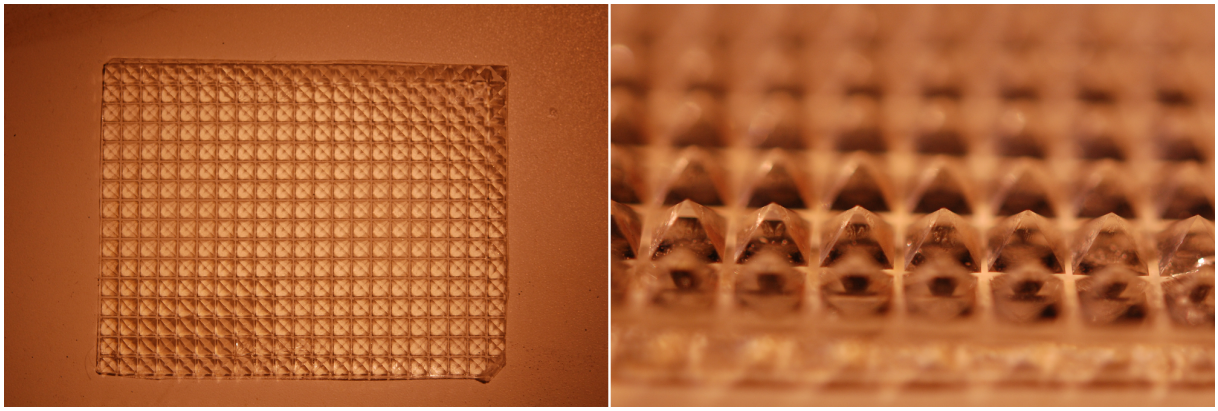


Figure B.4: Norland optical adhesive master.

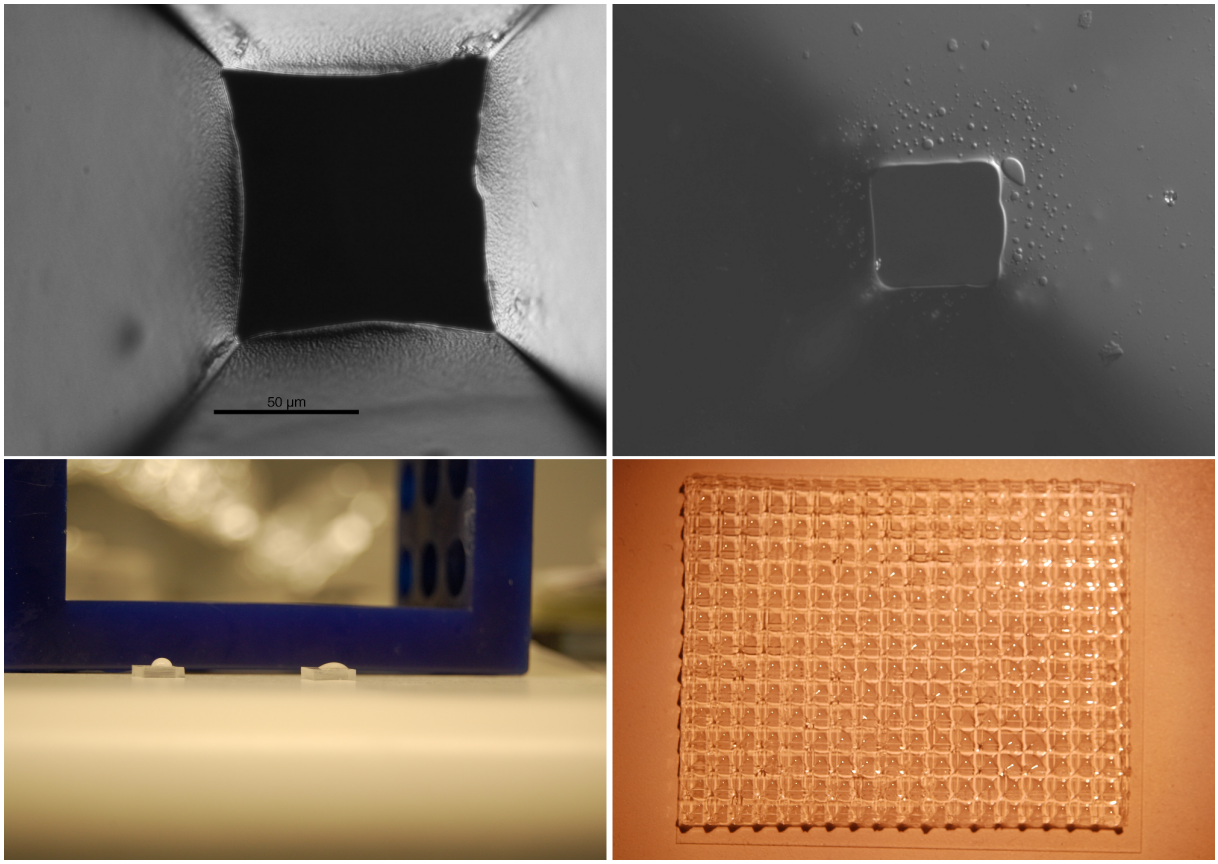


Figure B.5: PDMS wells. Upper left: well prior to bonding to surface. Upper right: well after bonding to PDMS coated glass (same scale as in upper left). Lower left: water droplet on PDMS without PEGDA treatment on left, water droplet on PEG-PDMS on right. Lower right: PEG-PDMS well plate, each well is 4mm wide.

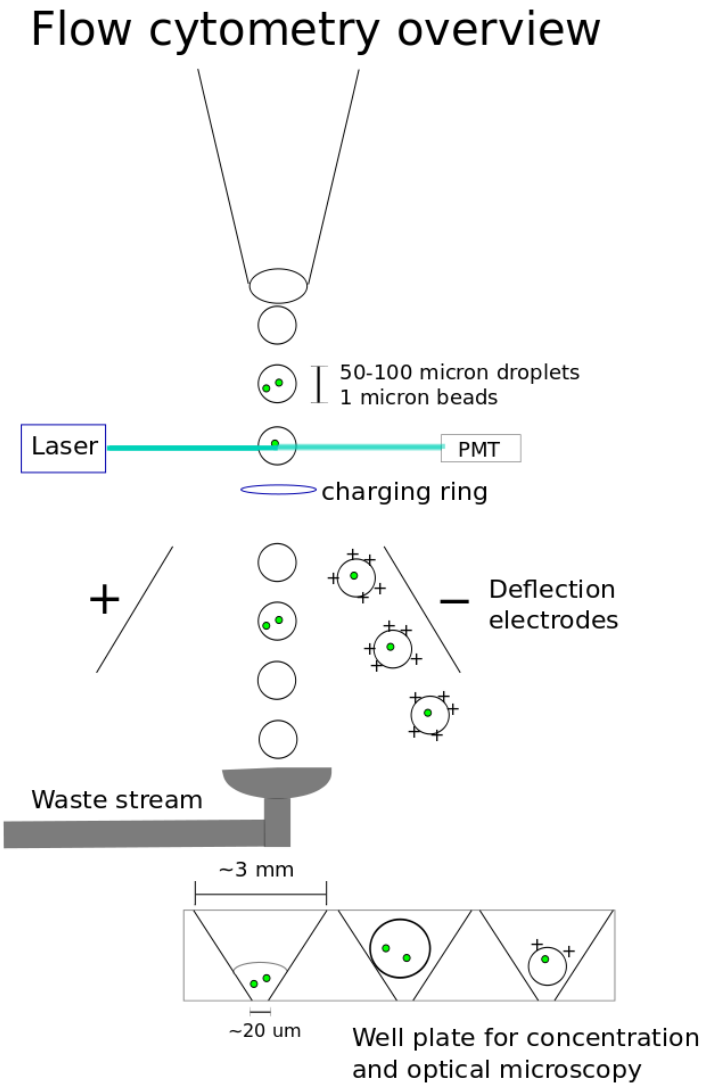


Figure B.6: Schematic of fluorescence based sorting of beads into PDMS wells.

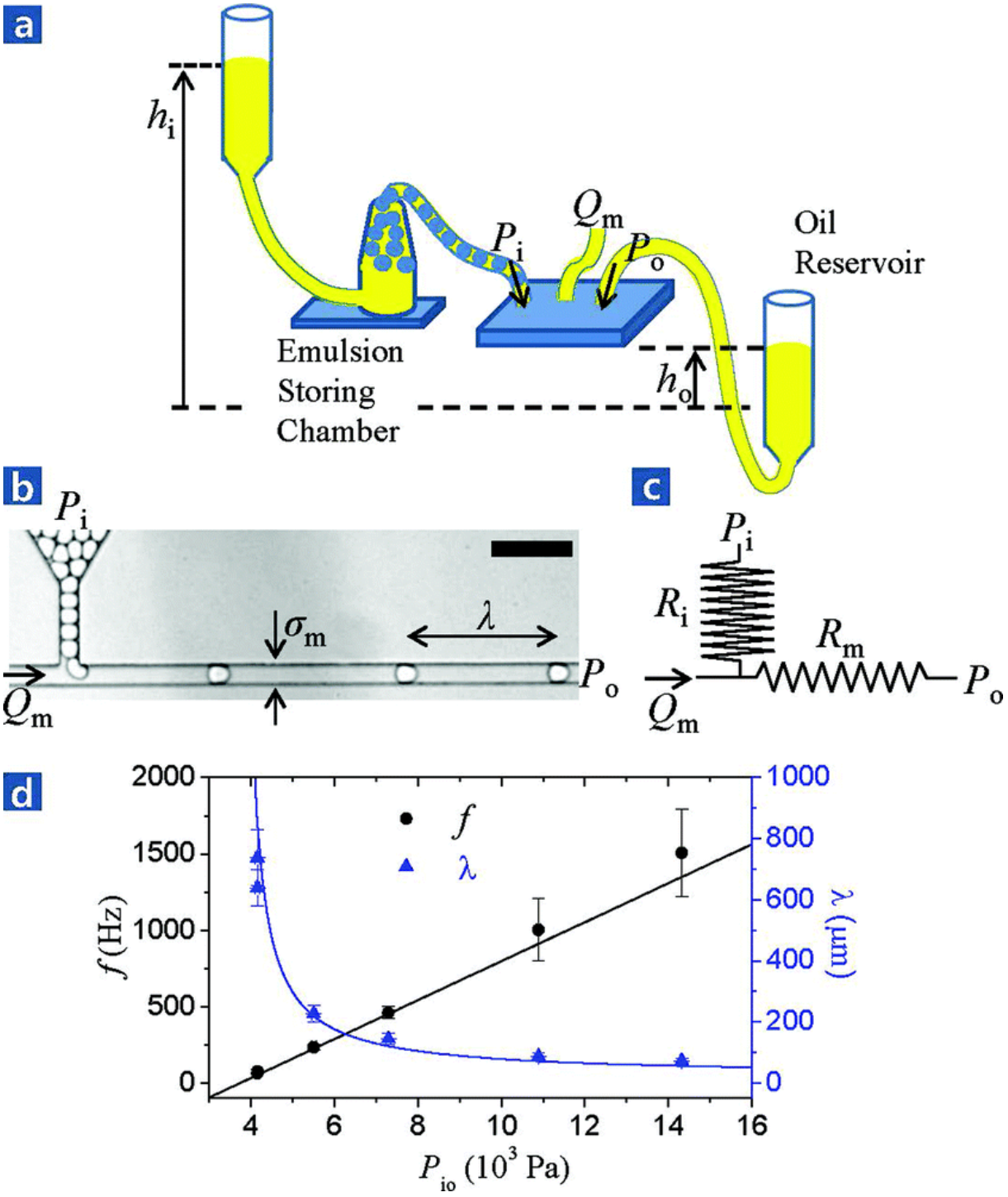


Figure B.7: a) Pressure based microfluidic emulsion reinjection schematic. b) high speed video frame showing λ, inter-droplet spacing. c) Circuit Diagram. d) Droplet frequency and spacing as a function of hydrostatic pressure. Figure courtesy of Manhee Lee.

Appendix C

DNA sequences and design

C.1 Designing DNA Sequences

I designed and/or implemented a variety of computer programs for finding DNA sequences to use to confer a specificity matrix to a set of particles.

C.1.1 Genetic Algorithm

The first method I implemented was an algorithm that accepted a base pairing matrix of an abitrary number of strands and strand lengths. The C++ program generated a random initial guess of the A-C-G-T letters to fill the matrix. Then it checked to see what the base pairing probability was for each pair in the specified matrix. It summed the probabilities. Then the program randomly changed one base at a time and repeated it probability summation. If the probability increased, it accepted the random change, and if not, it rejected it. This approach is based on work by Dirks et al.[16], and used Nupack[58, 59] to calculate thermodynamic base pair probabilities.

A problem with this method is that it did not often generate sequences that were predicted to have similar melting temperatures.

C.1.2 Brute Force Generate and Filter

I wrote an algorithm that exhaustively generated all sequences with a fixed length and fixed total number of C's and G's, then filtered these for similar melting temperatures and checked combinations of them meeting an input specificity matrix. This approach found sequences with similar predicted melting temperatures. The algorithm however took a long time to run, so I wrote it to be run on a parallel computing grid, so its calculation time of 150 computer days was reduced to 3 actual days. Finding a set of sequences corresponding to a given specificity matrix corresponds to a clique problem in graph theory and computer science. The clique decision problem is an NP-complete problem, so the main problem with this approach is the computational time and power required.

C.1.3 Random with 3-word Lookup Table

A approach used by others [25], is to generate sequences one base at a time while checking to see if adding the new base results in a word that has already been generated. Ben Rogers used this technique to

generate many sequences. First we generated many sets of 3 pairs of sequences without any repeated 3 letter words. Each of these pairs consisted of a sequence of 9 bases with a 3 letter “toehold”, then 6 bases to hybridize with another 6 bases on the other sequence in the pair, which lacked a “toehold” sequence.

C.1.4 Findings on Stacked Sequences

I explored the possibility of placing two different 6 base sequences on the same strand. The idea would be that each microsphere could be coated with a single type of DNA. So particle type 1 say would have a sequence with sticky ends A and B’. This could in principle bind to another particle coated with a sequence that had A’ and C sticky ends on it, and so on. I checked the thermodynamics of designer stacked sequences with Nupack. Although the predicted thermodynamics suggested that each of the 3 pairs of particles would have similar melting temperatures, in practice we observed vastly different melting temperatures. Turning the temperatures with these sequences to match was either not possible or reduced the surface coverage too much to have viable particle-on-particle surface diffusion.

Appendix D

DNA particle coating procedsample preparation

D.1 F108-DNA chemistry

We prepare DNA-coated colloids by reacting DNA with F108 surfactant (BASF), then grafting these molcules to the surface of polymeric microspheres, with slight modification to a protocol described by Kim et al.[60].

In a clean and dry round bottom flask I add .5 g, 2 ml DCM (Spectrum, $\geq$ 99.5%), 30 ul TEA (Sigma-Aldrich, $\geq$ 99%), and 100 mg 4-NPCF (Spectrum, $\geq$ 97%) with stirring. After 4 hours, I pour the reaction contents into a $-20\,^{\circ}\mathrm{C}$ centrifuge tube (SafeSeal, Sorbio) containing 14.4 ml pure EtOH (Koptec) and .6 ml 34-37% HCl (EMD), returning the tube to $-20\,^{\circ}\mathrm{C}$ for $\sim$ 20 minutes. I centrifuge this tube at $2\,^{\circ}\mathrm{C}$ for 5 minutes at 1000 g to pellet the activated F108, and remove the supernatant. I repeat this washing step using cold .15 ml 30% HCl in 14.75 ml EtOH three times. Finally I warm the pellet of F108 by hand to redisperse. I dry the activated F108 typically in either a rotovap with heating to $40 - 50\,^{\circ}\mathrm{C}$ over 2-3 hrs or at room temperature in a vacuum desiccator overnight.

To react amino modified DNA with NPCF-F108, I add 1 ul of 1 M pH 9.5 Carbonate buffer (0.85 M Sodium Bicarbonate, EMD 99.7-100.3%, 0.15 M Sodium Carbonate, EMD $\geq$ 99.5%) to a microcentrifuge tube, then add 15 ul of 1-3 mM DNA suspended in 10 mM Tris EDTA buffer, pH 8.0. Finally I add 4 ul of NPCF-F108 suspended at 15 mg/ml in 10 mM pH 4.0 Citric Acid buffer (1.1 mM Anhydrous Citric Acid, EMS 99.5%, 8.9 mM Sodium Citrate, Spectrum, 99.0-100.5%). Typically I mix the reaction contents for 4 hrs on a shaker (VWR) at high enough RPM such that the solution creeps up the walls of the tube but remains a single droplet. The solution slowly converts from clear to yellow over the course of an hour or so. After 4 hrs I use the F108-DNA as is or store these solutions for weeks to months at either room temperature or $4\,^{\circ}\mathrm{C}$ for future use.

D.2 Grafting DNA-F108 to Microspheres

We thoroughly wash microspheres prior to DNA-F108 grafting. Sulfate polystyrene (0.96 micron diameter, Invitrogen), carboxylate-modified polystyrene latex (1.1 micron diamater, Invitrogen), or ptBMA beads are diluted to 1.0 % v/v in pH 4.0 10 mM Citric Acid buffer. We centrifuge and wash the beads five times

in the citric acid buffer. Just prior to DNA-F108 coupling, we sonicate the beads for about 20 seconds in a bath sonicator (VWR).

We pipette 4 ul total F108-DNA solution into a non-adhesive 1.5 ml centrifuge tube (USA Scientific). To this we add 95 ul 1% v/v microspheres washed in citric acid buffer, pipetting up and down to mix. Finally we add 1 ul of toluene (EMD, 99.5%) or toluene/dye mixture to the top of the bead DNA-F108 solution. We prepare toluene dye mixtures by adding excess dye to toluene so that dye precipitant is visible, then pippetting off dye-saturated toluene and using as is or diluting to 5%-75% saturation in additional toluene. For this work, we used Nile Red (TCI) at 5% saturation and Pyrromethane 546 (Exciton) at 75% saturation. In equilibrium experiments, DCM was used for the gray particles at 100% saturation in toluene. We rotate the bead/DNA-F108/toluene solution at 30 RPM on a rotary mixer (Glas-Col) to mix the toluene into the suspension, and to keep the beads from sedimenting while the DNA-F108 adsorbs to their surfaces. After about 1 hour of rotating, we place the open centrifuge tubes in a $90\,^{\circ}\mathrm{C}$ drying oven for 10-15 minutes to remove toluene. Finally we wash the beads five times in 10 mM 1X TE buffer pH 8.0, at 2,000 g for 3 minutes per wash. After these wash steps the beads are ready for use.

D.3 Sample Prepartion for DIC or Brightfield Microscopy

To prevent particles from sticking to glass, we treat our cover slips and glass slides (VWR) with oxygen plasma for 1-2 minutes. After plasma treatment, the slips are more hydrophilic and we observe little to no adhesion of DNA-F108 coated microspheres to the glass at any salt conditions explored here. Typically we seal samples using NOA 63 (Norland) and 10-60 s of UV illumination while blocking the UV from the bulk of the sample using a small piece of aluminum foil. Alternatively we seal samples using silicone grease. In many cases we use 30-50 micron glass spacer beads (Polysciences) to control the sample thickness.

Appendix E

Specificity and Connectivity Matrices

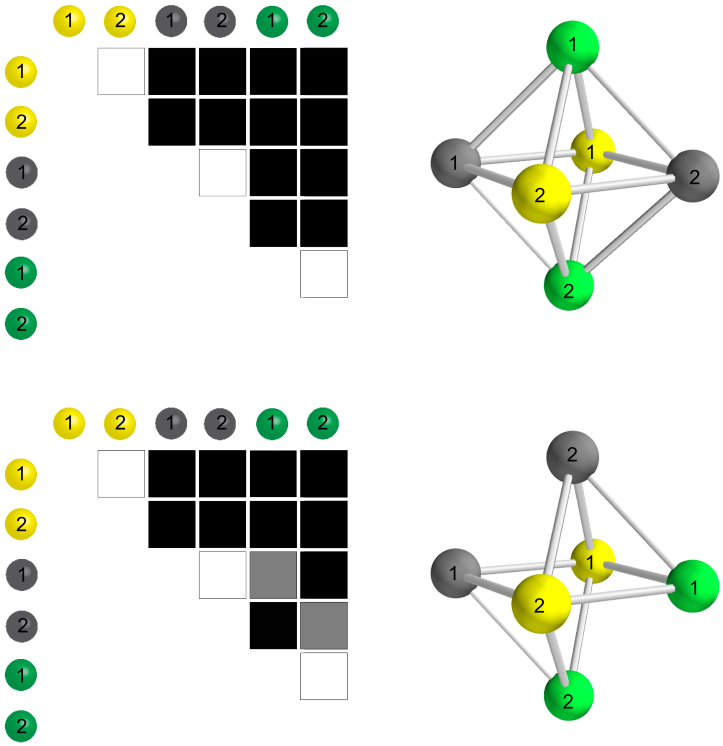


Figure E.1: N=6 Specificity and Connectivity. White squares indicate purely repulsive interaction, Gray indicates attraction, black indicates bond present in cluster to right of matrix. N=6 local minima are subgraphs of ground state.

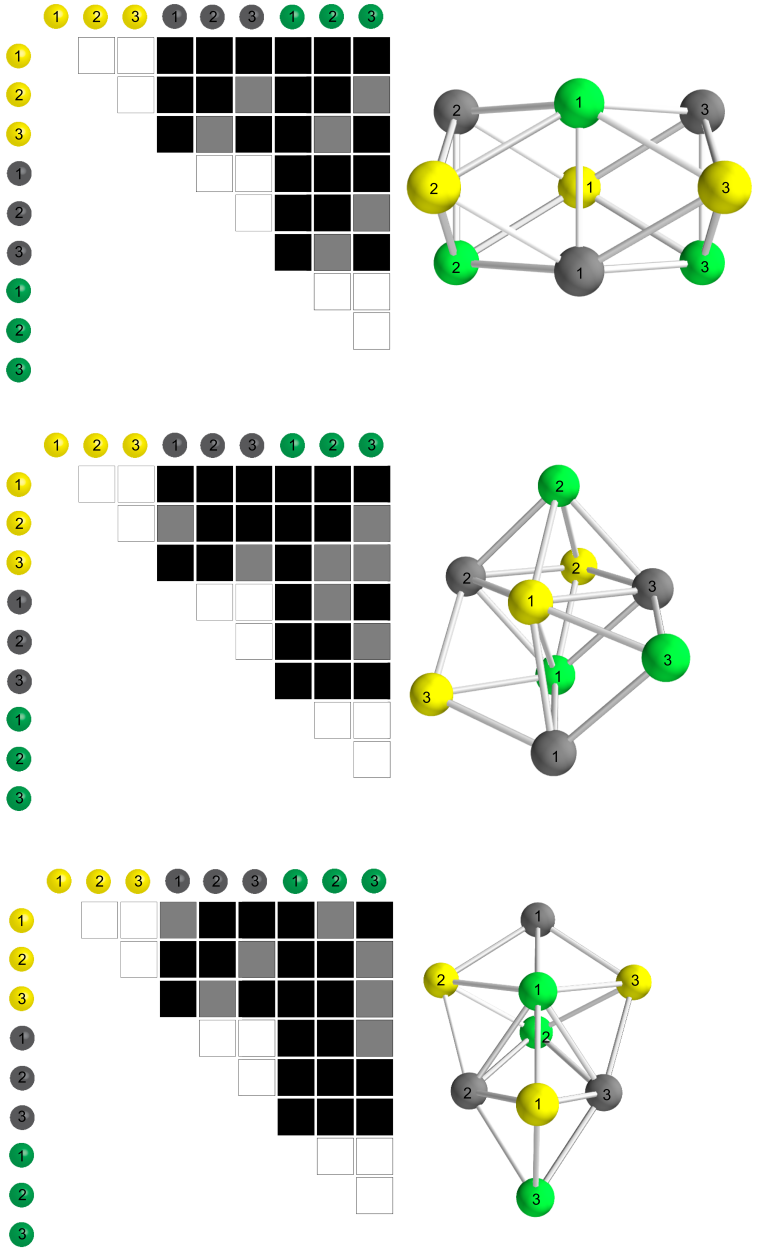


Figure E.2: N=9 Specificity and Connectivity. White squares indicate purely repulsive interaction, Gray indicates attraction, black indicates bond present in cluster to right of matrix. N=9 local minima are not subgraphs of ground state.

Appendix F

Strong Binding Limit

I performed preliminary experiments suggesting that non-equilibrium assembly of specific spheres may be, to my surprise, a viable alternative to equilibrium assembly. I do not expect sufficiently small clusters to form kinetic traps, such as a 5 particle cluster composed of two species. The first three octahedra I observed, I formed via a two-step process. First I let 5 of the 6 particles assemble into a pentagonal cap, forming 8 of the 12 bonds of the octahedron. This cap is a subunit of the true octahedron but is not a subunit of the false minimum. Then I added the 6 particle, and let the cluster undergo a conformation change strictly downhill in potential energy space, leading directly to the true minimum (see Fig. F.1). 3/3 of the clusters assembled via this procedure form the octahedron.

Before I was able to observe equilibrium clusters, I observed 12 out of 13 clusters form the true octahedron in the strong binding limit, and only once form a false octahedron. These clusters did not break specific sphere-sphere contacts, only formed them. Although the number of different arrangements of the local minimum is high compared to the ground state, the number of pathways to the ground state is higher. These experiments typically start with the cluster in a chain-like conformation. Becca Perry pointed out that by counting the number of each possible 6 particle path in each of the local minima, it may be possible to make an estimate for the likelihiood of an initial path leading to the ground state relative to the stable excited states (See Fig. F.2).

F.1 6 species experiment

I designed and observed the assembly of a pentagonal dipyramidal cluster composed of 7 particles of 6 different species. The central, gray particles were each of the same species, as each bound all 5 of the other particles. I observed this cluster assembly in the strong binding limit. I controlled the order by which I added particles to the initial chain. I made sure that the 5 non-gray particles formed a linear chain of 4 bonds before adding the gray particles. This initial state circumvented many of the possible kinetic traps (see Fig. F.3).

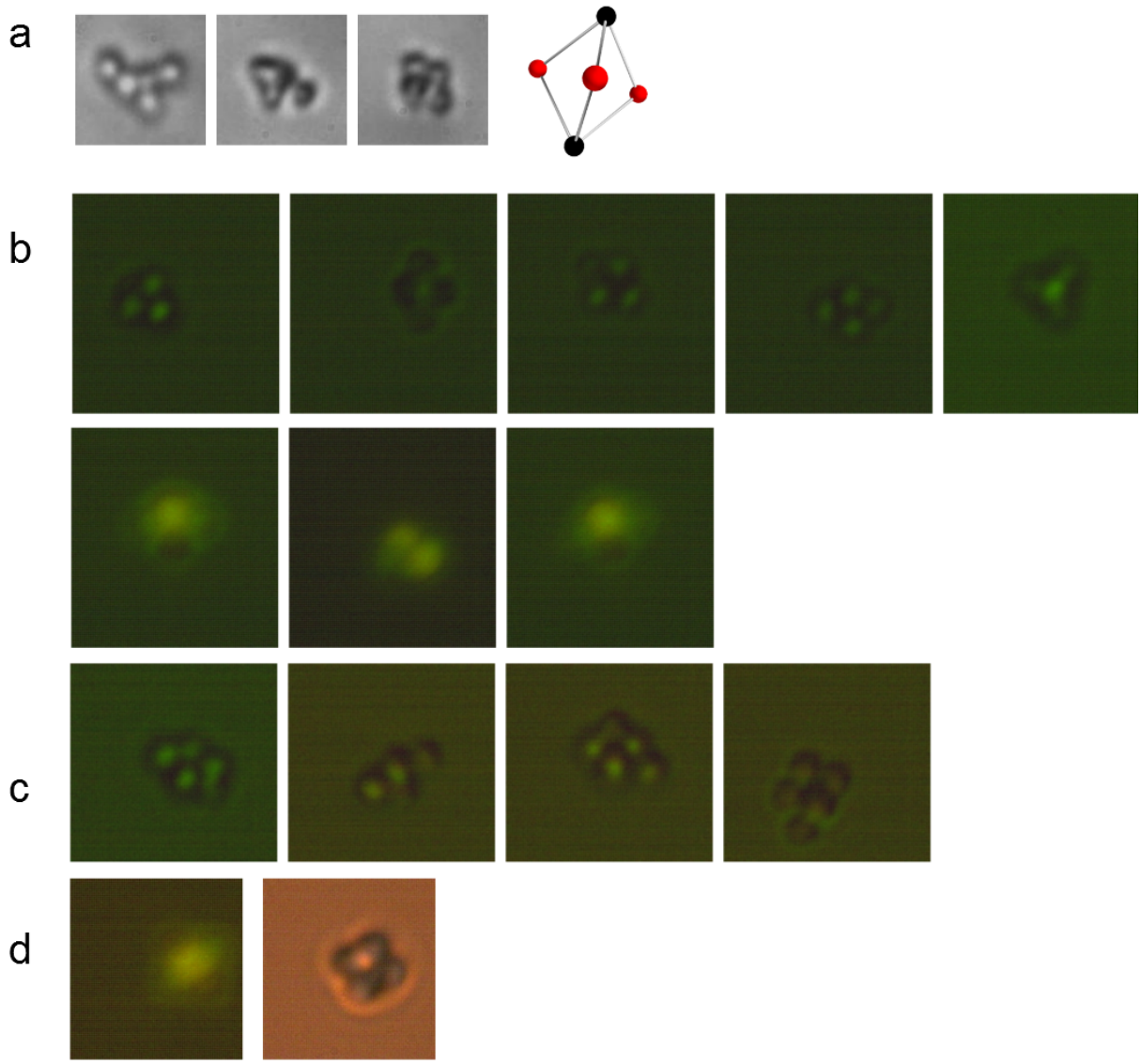


Figure F.1: Strong binding limit experiments that get to the ground state 100% of the time. a) A two-species, 5 particle system, two of one type and three of another, has only one stable state. b) 5 of the 6 particles for the octahedron assemble into a cap (only one gray particles) that is already in the basin of the true octahedron. c) Adding the 2nd gray particle, the system proceeds directly to the octahedron, shown in d)

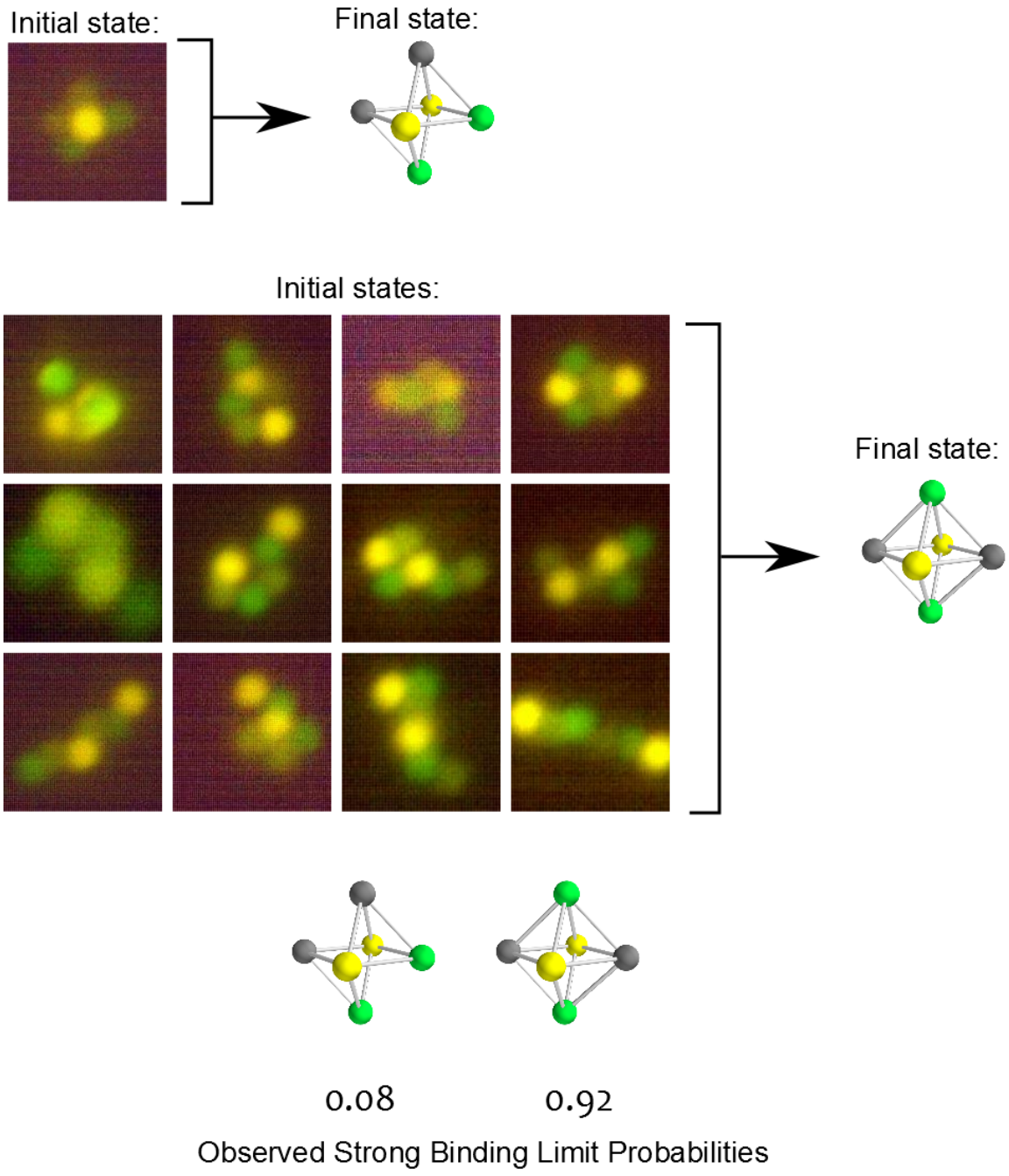


Figure F.2: Strong binding limit results on octahedral system. In these experiments, specific particle-particle contacts form but do not break. The octahedron assembles in high yield even in this limit.

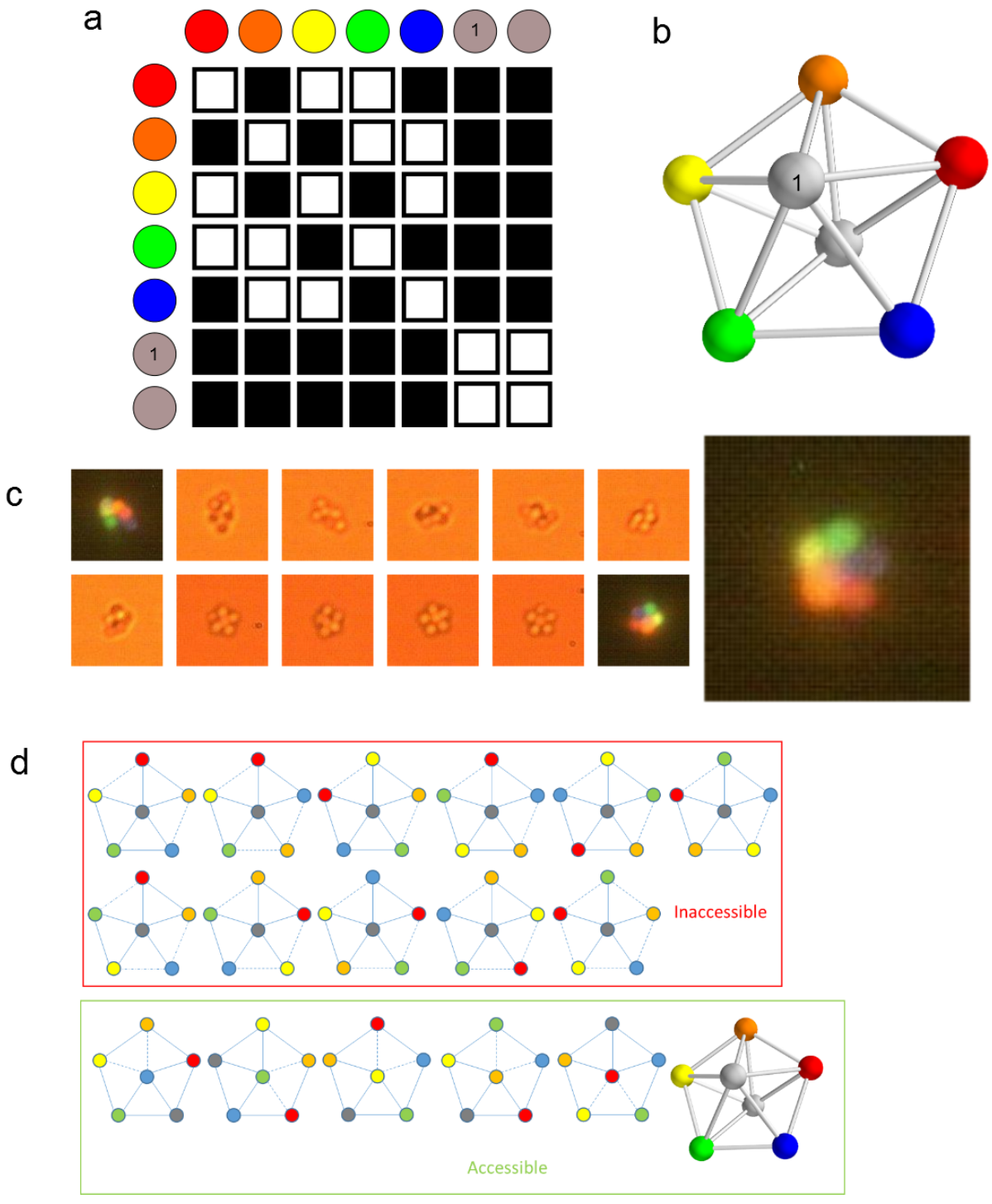


Figure F.3: 6 species, 7 particle system designed to assembly into a pentagonal dipyramidal cluster. a) specifity matrix, black squares indicate attraction, white indicates repulsion. b) rending of ground state c) time series over 10 minutes showing the assembly from a chain-like initial state (upper left) to the ground state (lower right) d) schematic of the 17 local minima, binned according to whether or not the initial state in c) was capable of organizing into the state without breaking any bonds.

Appendix G

N=7 and N=8 Experiments

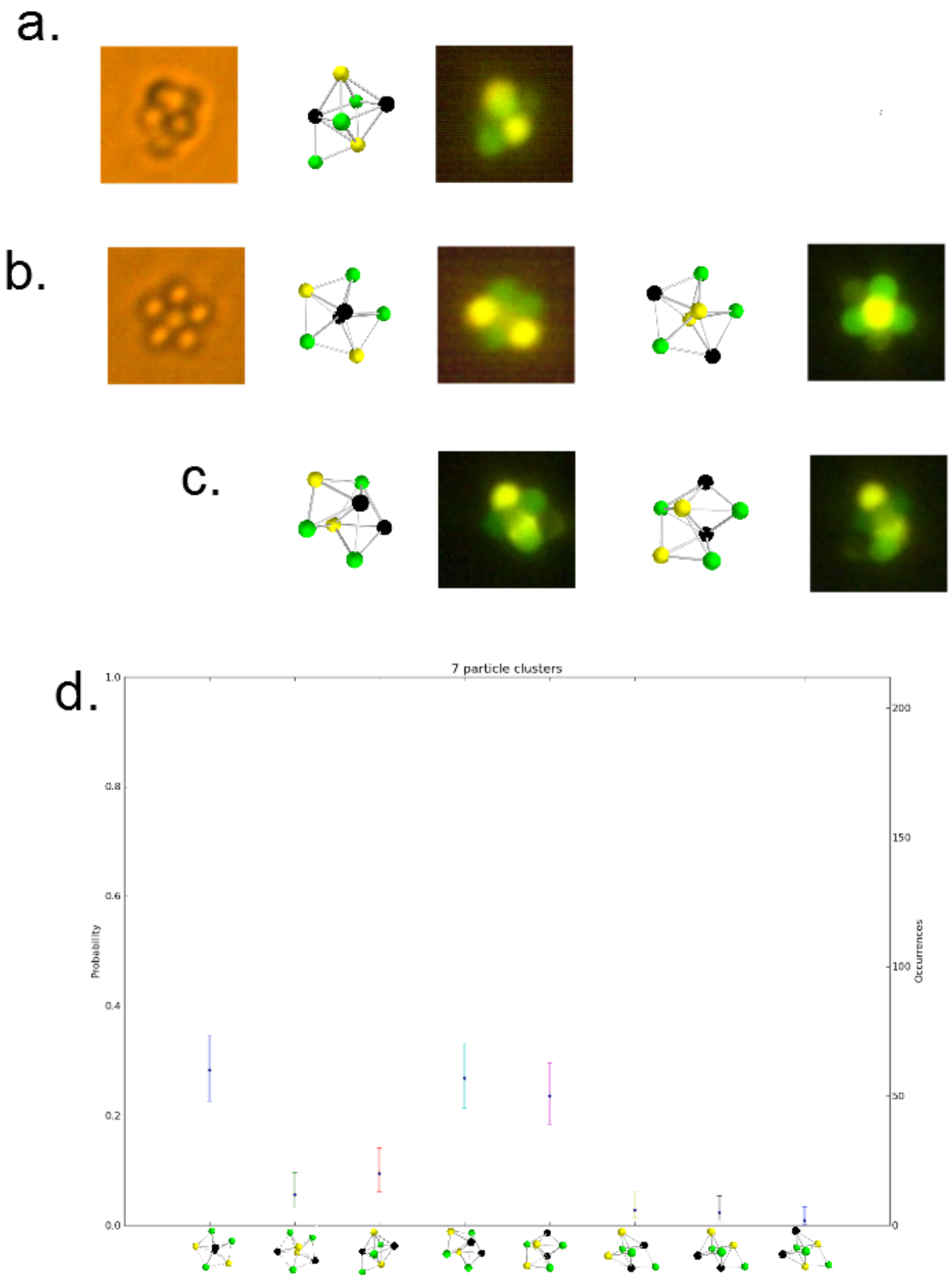


Figure G.1: N=7 results. a.) A 14 bond stable state with an octahedron subunit. From left to right bright field micrograph, rendering, and fluorescent micrograph. b.) A 14 bond stable state that can have pentagonal symmetry, and two distinguishable arrangements of it. c.) A chiral pair of 13 bond stable states. d.) Probabilities and numbers of observations of 12, 13, and 14 bond stable states.

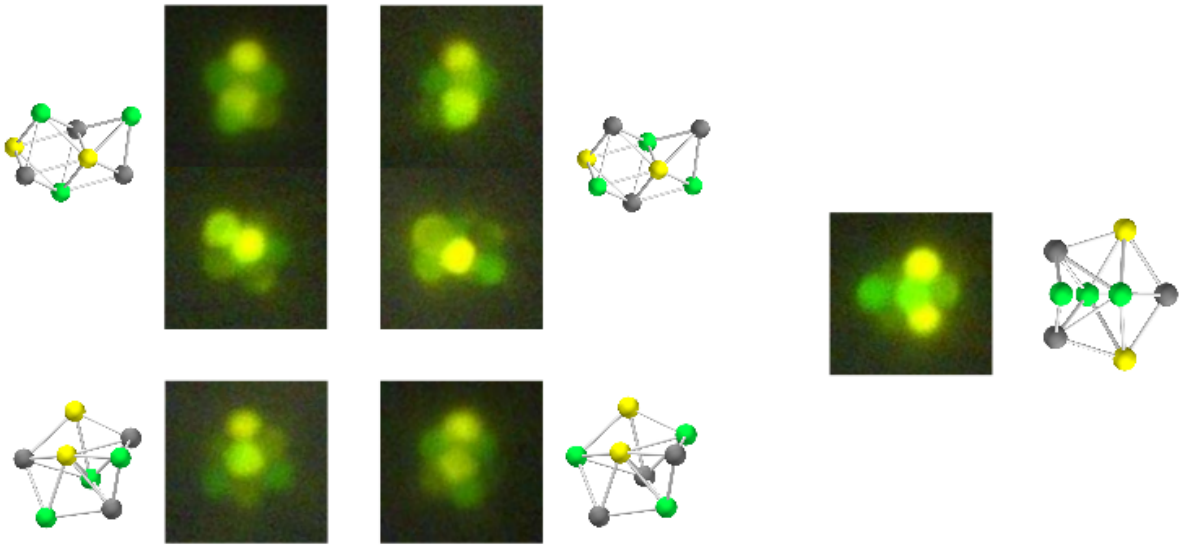


Figure G.2: N=8 clusters. Running clockwise from upper left, clusters numbers are 3,4,7,10,9.

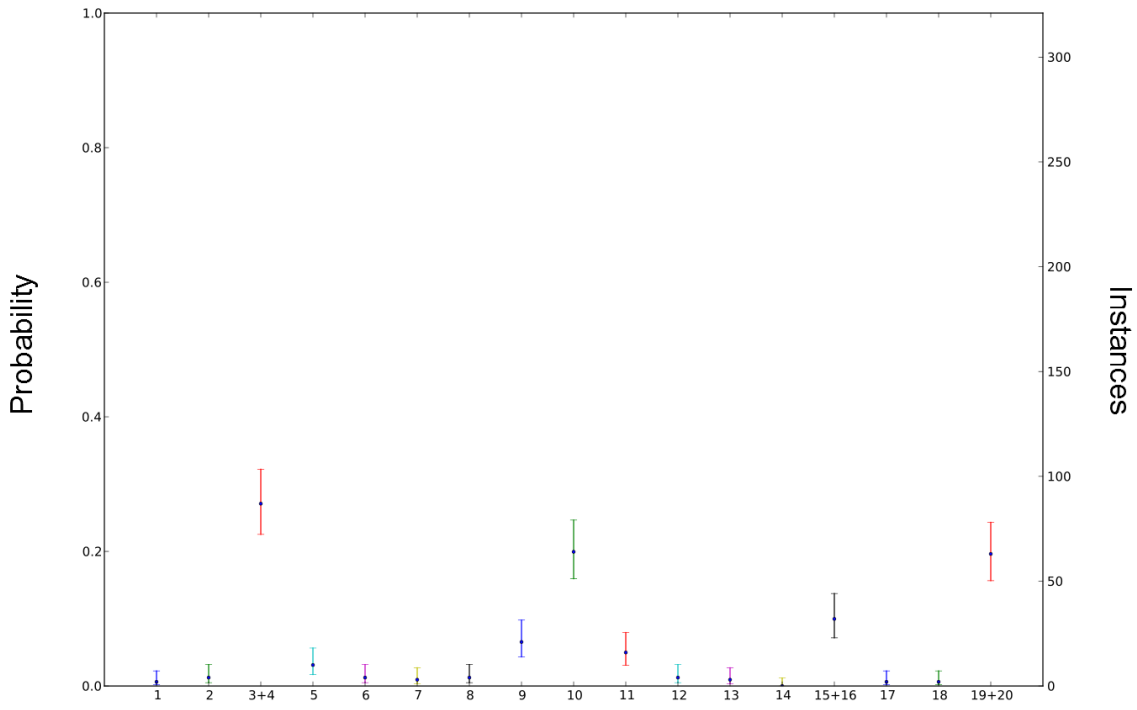
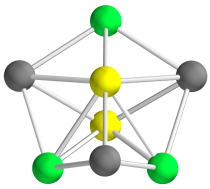
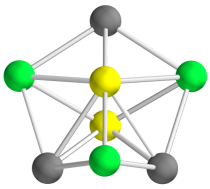


Figure G.3: N=8 probabilities.

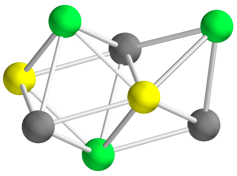
1



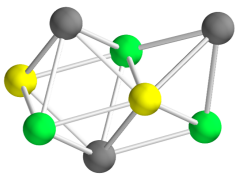
2



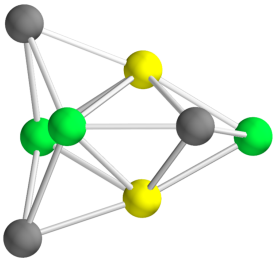
3



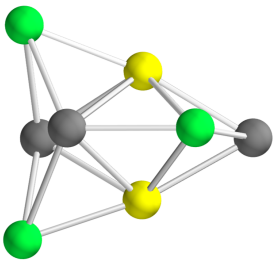
4



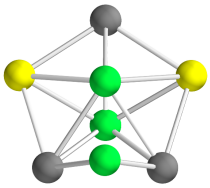
5



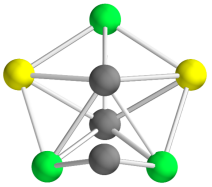
6



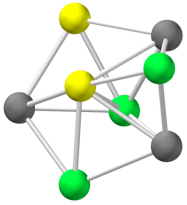
7



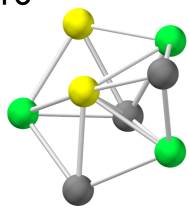
8



9



10



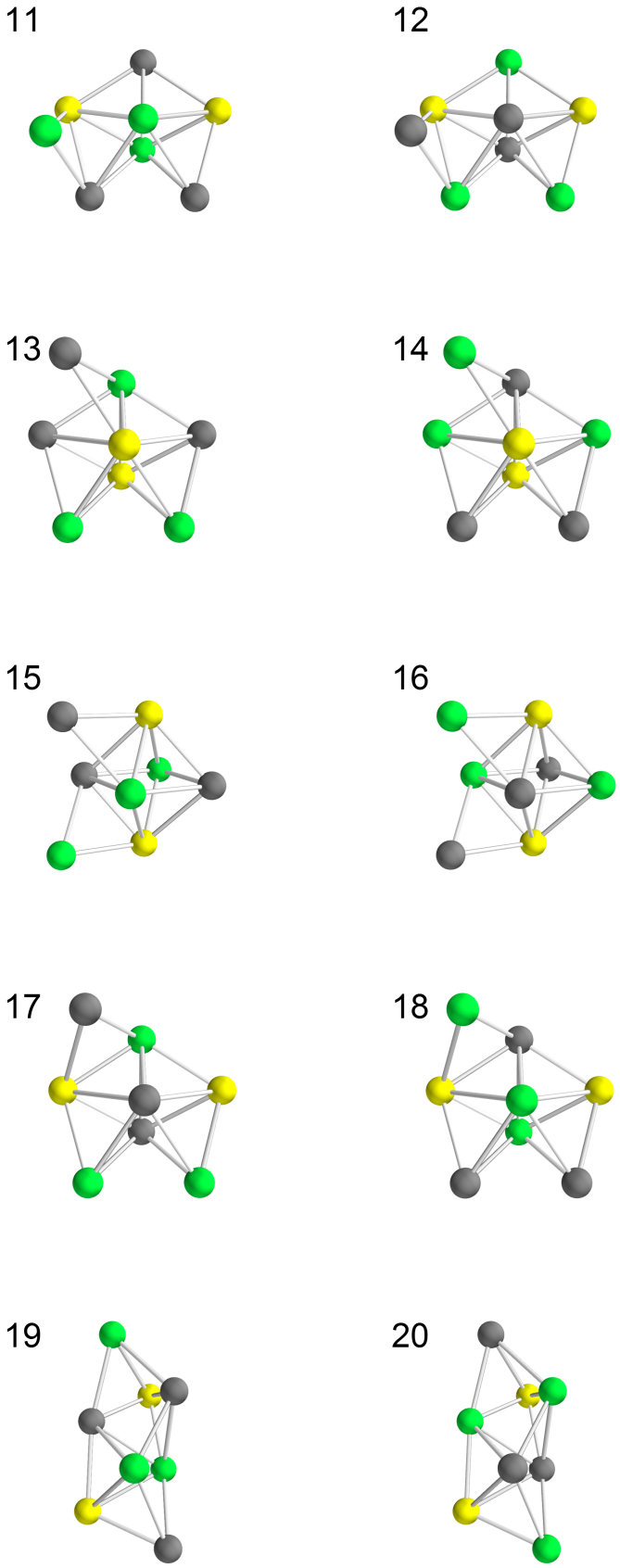


Figure G.4: N=8 renderings numerically labelled for reference in probability plot. Zorana Zeravcic provided the coordinates for these.

Appendix H

2D Patchy Substrate

This system consists of N_S specific spheres of radius r diffusing in 2D above a substrate with N_S square patches of diagonal length d, each seperated by a minimum patch-to-patch spacing L. Again, the patches are the $i = 0$ specie, and the sphere i binds to $i + 1$ and $i - 1$, now with $b_{2N} = 1$, favoring the closed loop configuration. Incorporating the geometric constraints $d < r$, prevents more than one particle from binding a patch, and $l > rN_S$, prevents a chain of the spheres from binding more than one patch at a time. I take the bond energy, U, to be equal for all attractive interactions.

There ground state (gs) is an N_S particle loop with any one of its particles bound to any one of the N_S patches, for a total bond number of $N_S + 1$ and multiplicity $\Omega_{gs} \propto N^2$. There are N different types minima, each of which has $N!$ multiplicity and N_S bonds. The N different types arise because of the different ways to break up the closed loop into smaller chains. Let's look at the local minima in which there is a one-to-one correspondence of all N_S particles to all N_S patches. The number of ways to arrange this local minimum configuration is $N!$, so $\Delta S \propto \frac{\Omega_{lm}}{\Omega_{gs}} = k \ln N! - k \ln N^2 \approx kN \ln N - 2k \ln N$ and

$$\lim_{N \to \infty} \Delta S/N \propto k \ln N.$$

The entropy difference between excited states and ground states is superextensive.

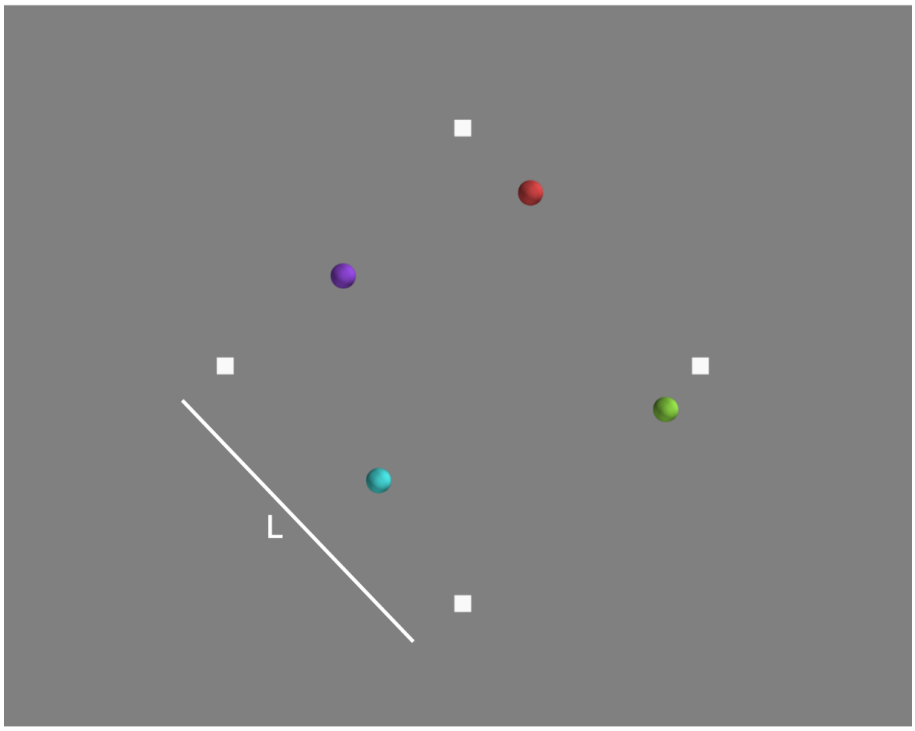


Figure H.1: High temperature state of the toy model 2D permutation glass. 4 particles diffuse above a surface. The 4 white squares represent patches of DNA capable of binding each and every one of the 4 particles. The minimum distance between patches, L, is defined to be $L > (N-1)D$ where D is the diameter of the spheres. This constraint ensures that their is a relatively small number of ground states ($\sim N^2$) compared to a large number of local minima ($\approx N * N!$), in the case that the all particle-particle and particle-patch bonds are of equal strength. When considering particle-particle and particle-patch bonds, I coarse-grain over the individual DNA strands, as in comparable laboratory systems of 1 micron spheres and 65-base pair (bp) DNA sequences.

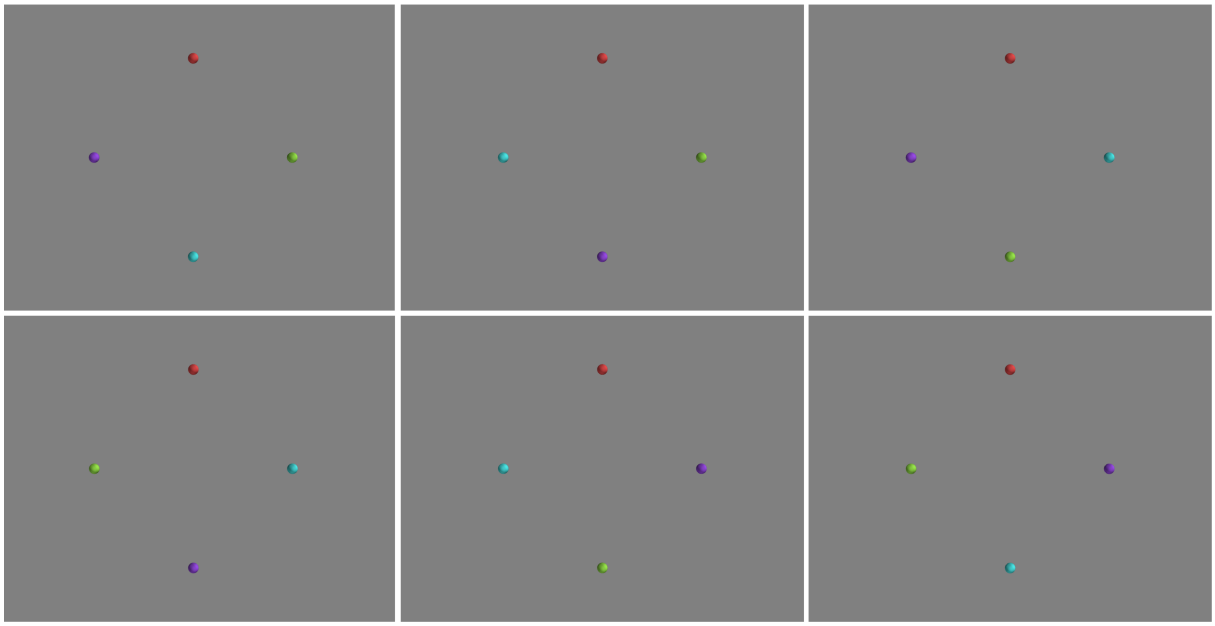


Figure H.2: Displayed here are 6 out of the 4! local minima in which each particle is bound to each patch. The other 18 local minima of this kind can be obtained by changing the position of the red particle and then placing the remaining 3 again in all 3! remaining arrangements. Although these appear to be equivalent to rotating the images in increments of 90 degrees, I know they all still contribute to the partition function because the patches are distinguishable/trackable.

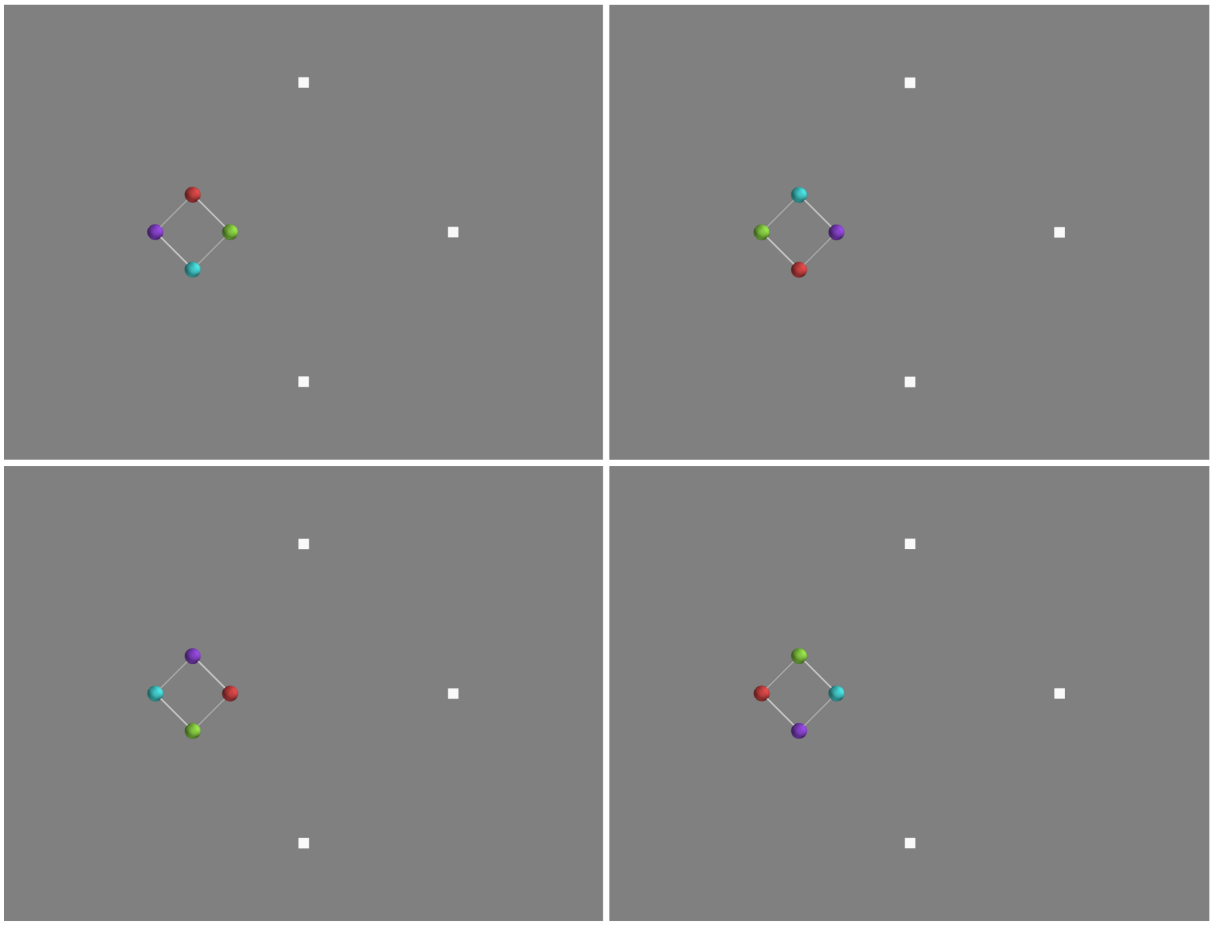


Figure H.3: 4 of the 16 ground states of the this 4 bead 2D model system. The beads are specified to bind in a loop as shown. There are N^2 ground states of this type: N different ways (particles) to bind the chain to each patch and N different patches to bind the chain. If all bonds are of equal strength, these are ground states because they have N+1 total bonds, whereas all other local minima have fewer than N+1 total bonds. In addition to the local minima in Fig. H.2, any subset of the chain could form but be impeded from ground state formation by 2 or more components of the chain binding to 2 or more patches. Were $L \leq (N-1)D$, the ground states would include open chains bound to two (or more if possible) patches.

Appendix I

Python Code for Quasi-1D Permutation Glass Energy Landscapes

```
import itertools
import networkx as nx

class LandscapeQ1D():
    '''An energy landscape has minima and transitions between minima.

    There are different ways to represent and energy landscape or
        surface.

    A 3D surface is one.  Minima are valleys separated hy hills, or
        energy barriers.

    A 2D graph is another.  Vertices represent energy minimizing
        states.  Edges represent transition states.

    This class generates the energy landscape for the quasi-1D
        permutation glass.  The landscape includes all the local
        minima.  Edges representing the one-bond breaking transitions
        between local minima are included.  No two-bond or more
        breaking transitions are included.  This graph is directed;
        some transitions require one bond to break to go in one
        direction, but more bonds to break to go backwards.

    Ex.
    E=LandscapeQ1D()
```

```
E.GS='ABCD'
E.generate_landscape()
E.draw_showing_gs()

'''
def __init__(self,GS=None):
    self.N=None
    self.GS=GS
    self.minima=[]
    self.transitory=[]
    self.Graph=nx.DiGraph()
    self.gs_nn=None
    return
def remove_dihedral(self):
    '''We choose not to distinguish backwards and forwards words
       from each other.'''
    for i in self.minima:
        k=''.join([j for j in reversed(i)])
        if i<k:
            self.minima.remove(k)
def remove_dihedral_2(self,another_list):
    temp=[]
    for i in another_list:
        k=''.join([j for j in reversed(i)])
        if i<k:
            temp.append(i)
        else:
            temp.append(k)
    return temp
def make_minima(self):
    '''Take string in self.GS, permute, and place each
       permutation in self.minima'''
    l=[''.join(i) for i in itertools.permutations(self.GS)]
    self.minima=l
def add_nodes(self):
    self.Graph.add_nodes_from(self.minima)
def build_graph(self):
    '''Really this just sets up all the nodes, without any edges
       '''
    if self.GS:
        self.make_minima()
        self.remove_dihedral()
```

```
        self.add_nodes()
def make_all(self):
    self.build_graph()
    self.find_gs_nearest_neighbors()
    self.find_bound_pairs()
    self.singlets()
    self.strictly_dimers()
def nearest_neighbors(self,minimum):
    '''Get a list of nearest_neighbors in minima, a string, i.e
       element of self.minima.'''
    return [minimum[i:i+2] for i in range(0,len(minimum)-1)]
def find_gs_nearest_neighbors(self):
    gs_nn=self.nearest_neighbors(self.GS)
    l=[''.join(reversed(i)) for i in gs_nn]
    gs_nn.extend(l)#this list contains all the bound pairs (
       forward and reverse)
    self.gs_nn=gs_nn
def count_k(self):
    '''Get a count of the number of bonds in each node'''
    gs_nn=self.gs_nn
    for n in self.Graph.nodes():
        k=len([i for i in self.nearest_neighbors(n) if i in gs_nn
           ])
        self.Graph[n]['k']=k
def find_bound_pairs(self):
    '''Find and store the bound pairs for each node'''
    for n in self.Graph.nodes():
        self.Graph[n]['bound_pairs']=[i for i in self.
           nearest_neighbors(n) if i in self.gs_nn]
def singlets(self):
    '''Find the Letters in each string that are not bound to any
       other letter '''
    s=''
    for n in self.Graph.nodes():
        for i in self.Graph[n]['bound_pairs']:
            s=s+i
        self.Graph[n]['singlets']=[j for j in self.GS if j not in
            s]
        s=''
def strictly_dimers(self):
    '''Find the pairs of letters in each string that are bound to
       each other and no other letters'''
```

```
        s=''
        for n in self.Graph.nodes():
            s=set(self.Graph[n]['bound_pairs'])
            self.Graph[n]['dimers']=[]
            for i in s:
                s1=''
                for j in s.difference(set([i])):
                    s1=s1+j
                if set(i).isdisjoint(set(s1)):
                    self.Graph[n]['dimers'].append(i)
    def moves_singlet(self,downhill_only=True):
        '''Find the edges that take each node to another energy state
            by moving a singlet to a different position'''
        for n in self.Graph.nodes():
            self.Graph[n]['s_moves']=[]
            self.Graph[n]['bpi']=[n.find(i) for i in self.Graph[n]['
               bound_pairs']]
            for i in self.Graph[n]['singlets']:
                l=[n[:j].replace(i,'')+i+n[j:].replace(i,'') for j in
                    range(0,len(self.GS)+1) if n[j-1:j+1] not in self
                   .Graph[n]['bound_pairs']]
                self.Graph[n]['s_moves'].extend(l)
            self.Graph[n]['s_moves']=self.remove_dihedral_2(list(set(
               self.Graph[n]['s_moves'])))
    def moves_dimer(self,downhill_only=True):
        '''Find the edges that take each node to a lower energy state
            by reversing a dimer'''
        for n in self.Graph.nodes():
            l=[n.replace(i,''.join(reversed(i))) for i in self.Graph[
               n]['dimers']]
            self.Graph[n]['d_moves']=l
    def add_edges(self,singlet_moves=True,dimer_moves=True):
        '''Add the edges corresponding to single bond breaking moves,
            i.e. moving singlets or reversing dimers.'''
        for n in self.Graph.nodes():
            for s in self.Graph[n]['s_moves']:
                self.Graph.add_edge(n,s)
            for d in self.Graph[n]['d_moves']:
                self.Graph.add_edge(n,d)
    def bond_number_dictionary(self):
        d={}
        for i in self.Graph.nodes():
```

```
            d[i]=len(self.Graph[i]['bound_pairs'])
            self.Graph[i]['weight']=d[i]
        return d
    def generate_landscape(self,clean_up=True):
        '''Requiring only the GS, build the energy landscape, i.e.
           the graph of all the minima and their connections via
           single-bond breaking paths.
        If clean_up=True, also reload the Graph without the extra
           data used to build it'''
        self.make_all()
        self.moves_singlet()
        self.moves_dimer()
        self.add_edges()
        self.bond_num=self.bond_number_dictionary()
        if clean_up:
            edge_list=self.Graph.edges()
            self.Graph=nx.DiGraph()
            self.Graph.add_edges_from([i for i in edge_list if i[1].
               isupper()])
    def spectral_positions(self):
        '''Return the x,y positions of the nodes from the spectral
           layout'''
        self.spectral_pos=nx.drawing.layout.spectral_layout(self.
           Graph)
    def shells(self):
        '''Turn self.bond_num and nodes into shells for shell_layout
           '''
        r=[i for i in range(min(self.bond_num.values()),max(self.
           bond_num.values())+1)]
        d={}
        for i in r:
            d[i]=[]
        for i in self.bond_num:
            d[self.bond_num[i]].append(i)
        self.sh=[i for i in reversed(d.values())]
        self.sl=nx.shell_layout(self.Graph,self.sh)
    def draw_spectral_showing_gs(self,**kwargs):
        '''Draw the energy landscape/graph in a way that shows the
           single ground state among the many local minima '''
        nx.draw_spectral(self.Graph,nodelist=[i for i in self.Graph.
           nodes() if i != self.GS],node_color='r',node_size=30)
        nx.draw_spectral(self.Graph,nodelist=[self.GS],node_color='b
```

```
        ',node_size=300)
def get_nice_coords(self):
    '''Nice coords have one coordinate from the spectral layout
       and the other from the bond number'''
    d={}
    for i in self.sl:
        d[i]=array([self.sl[i][1],-self.bond_num[i]])
    self.nice_coords=d
def collapse_shell(self):
    '''This takes the shell layout coords and turns shells into
       crescents and rotates each crescent such that no edges
       complete overlap in the graph drawing.  It is similar to
       the nice_coords layout except that no edges overlap.  '''
    self.sl
def draw(self,**kwargs):
    nx.draw(self.Graph,**kwargs)
def paths_to_gs(self):
    '''returns a list of all the shortest path lists from each
       local minimum to the target.  if no (single-bond-breaking)
        path is possible, it returns an empty list'''
    l=[]
    for i in self.Graph.nodes():
        try:
            a=nx.algorithms.shortest_path(self.Graph,i,self.GS)
            l.append(a)
        except nx.NetworkXNoPath:
            l.append([])
    return l
def path_lengths_to_gs(self):
    '''list of lengths of shortest paths from each local minimum
       to the ground state '''
    l=self.paths_to_gs()
    l2=[len(i) for i in l]
    return l2
def shortest_path_statistics(self):
    '''returns the mean of the shortest path lengths from each
       node and the number of nodes that do not have any path
       leading to the ground state.  these nodes in other words
       require more than a single (white stripe to colored sphere
       ) bond breaking event.  they may require in addition one
       or possibly two colored-colored sphere bond breaking
       events at the same time.'''
```

```
        l=self.path_lengths_to_gs()
        m=np.mean(l)
        zs=len([i for i in l if i==0])
        return m,zs
    def ground_states(self,N=26):
        s=""
        for i in range(0+65,65+N):
            s=s+str(chr(i))
        return s
    def calc_stats_to_n(self,n):
        '''Runs shortest_statistics on graphs from 4 to n,

example output
(4, (2.0833333333333335, 0))
(5, (2.5333333333333332, 4))
(6, (3.0055555555555555, 39))
(7, (3.4813492063492064, 347))
(8, (4.0076884920634921, 3150))

runtime blows up at n=9

    '''
        stats=[]
        for i in range(4,n,1):
            s=self.ground_states(i)
            self.GS=s
            self.Graph=nx.DiGraph()
            self.generate_landscape()
            cur_stats=self.shortest_path_statistics()
            print (i,cur_stats)
            stats.append((i,cur_stats))
        return stats
```