

9-1-2017

# Study of the Effects of Anisotropy in Self-Assembling Systems

Jason Leigh Haaga

Lehigh University, jay.haaga@gmail.com

Follow this and additional works at: <https://preserve.lehigh.edu/etd>



Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

---

## Recommended Citation

Haaga, Jason Leigh, "Study of the Effects of Anisotropy in Self-Assembling Systems" (2017). *Theses and Dissertations*. 4283.  
<https://preserve.lehigh.edu/etd/4283>

This Dissertation is brought to you for free and open access by Lehigh Preserve. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Lehigh Preserve. For more information, please contact [preserve@lehigh.edu](mailto:preserve@lehigh.edu).

Study of the Effects of Anisotropy in  
Self-Assembling Systems

by

Jason L. Haaga

A Dissertation

Presented to the Graduate Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Physics

Lehigh University

September 2017

Copyright  
Jason L. Haaga

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Jason L. Haaga

Study of the Effects of Anisotropy in Self-Assembling Systems

---

**Date**

---

**James D. Gunton**, Dissertation Director, Chair  
Committee Members

---

**Jeffrey Rickman**

---

**Daniel Ou-Yang**

---

**Dimitrios Vavylonis**

---

**Javier Buceta**

---

**Accepted Date**

## Acknowledgments

Thank you, Megan, for somehow keeping me on track and being amazingly supportive through all of this. This would not have been remotely possible without you.

Thank you, Mom and Dad, for caring even when it probably just seemed like jibberish. I promise: it's not.

Thank you, Jim, for always giving me just enough rope to get tangled up before reeling me back in. I have learned a great deal from you both as a scientist and a great person.

Thank you to my thesis committee, for your help, for your sometimes stumping me, and for the periodic reminder to be fascinated.

Thank you to a too long to cite list of fellow grad students I now call friend. While the past four years were not always the best of times, even then I certainly got to spend them with the best of people.

Thank you to every person who ever uttered my least favorite phrase: "You're a smart kid. You'll figure it out." You were at least half correct.

And lastly, thank you, Min, for pretty much changing the entire world overnight. Like just about everything else, this is for you.

# Contents

List of Tables	vii
List of Figures	viii
Abstract	1
<b>1 Introduction</b>	<b>2</b>
1.1 Background . . . . .	2
1.2 Simulation Methods . . . . .	5
1.2.1 Molecular Dynamics . . . . .	6
1.2.2 Monte Carlo Simulation . . . . .	9
1.3 Protein Structure . . . . .	19
<b>2 Amelogenin: Effects of Coulomb Repulsion on Phase Diagram</b>	<b>24</b>
2.1 Abstract . . . . .	24
2.2 Introduction . . . . .	25
2.3 Model and Method . . . . .	27
2.4 Results . . . . .	30
2.4.1 Phase Diagrams . . . . .	30

2.4.2	Structure and Correlation Measurements . . . . .	30
2.5	Conclusion . . . . .	33
<b>3</b>	<b>Immunoglobulin: Effect of Angle Between Interaction Sites on Ki-</b>	
	<b>netics and Morphology</b>	<b>38</b>
3.1	Abstract . . . . .	38
3.2	Introduction . . . . .	39
3.3	Model and Methodology . . . . .	41
3.4	Results . . . . .	43
3.4.1	Rate of aggregation and size of aggregates . . . . .	43
3.4.2	Morphology . . . . .	46
3.5	Conclusion . . . . .	52
<b>4</b>	<b>Huntingtin: Effect of Repeat Length on Early Stage Aggregation</b>	<b>59</b>
4.1	Abstract . . . . .	59
4.2	Introduction . . . . .	60
4.3	Model and Methodology . . . . .	63
4.4	Results . . . . .	69
4.4.1	$Q_{10}$ . . . . .	69
4.4.1.1	Aggregate Mass Growth . . . . .	69
4.4.1.2	Nucleation Mechanism . . . . .	71
4.4.1.3	Oligomer Growth . . . . .	74
4.4.2	$Q_{30}$ . . . . .	77
4.4.2.1	Aggregate Growth and Nucleation Mechanism . . . . .	77
4.4.2.2	Oligomer Growth . . . . .	77

4.4.3	Nucleation Barrier . . . . .	79
4.5	Conclusion . . . . .	82
<b>5</b>	<b>Conclusion</b>	<b>86</b>
	<b>Bibliography</b>	<b>89</b>
	<b>Vita</b>	<b>104</b>



# List of Tables

4.1	Bonded interaction parameters for the PLUM coarse grained protein model. . . . .	65
4.2	General unbonded interaction parameters for the PLUM coarse grained protein model. . . . .	66
4.3	Normalized amino acid specific hydrophobicity for the PLUM coarse grained protein model. . . . .	67

# List of Figures

1.1	Examples of configurational Monte Carlo moves: translation (A), rotation (B), reptation (C), pivot (D), and crankshaft (E). Pre-move positions shown in grey. Post-move positions in black. . . . .	11
1.2	Examples of volume Monte Carlo move. Inter-particle distances are all scaled by the same ratio as the change in box length. . . . .	14
1.3	Examples of common secondary motifs: cartoon representations of an $\alpha$ -helix (left) and a three stranded $\beta$ -sheet (right) overlaid on their underlying atomistic composition. . . . .	21
2.1	Phase diagram in the $\phi_p - \phi_c$ plane with $\xi = 0.8$ and the screening length $\lambda = 0.4$ for various values of the screening strength, $\epsilon$ . $\phi_p$ and $\phi_c$ are the volume fractions of PEG and the model particles, respectively.	31
2.2	Same as Fig. 1, except that $\xi = 0.1$ and $\lambda = 0.4$ for various values of $\epsilon$ .	31
2.3	Same as Fig. 1, except that $\xi = 0.8$ and $\epsilon = 0.1$ for various $\lambda$ .	32
2.4	Radial distribution function, $g(r)$ , for state point in the low density liquid phase along $\phi_p = 0.4$ . . . . .	33
2.5	Radial distribution function, $g(r)$ , for state point in the high density liquid phase along $\phi_p = 0.4$ . . . . .	34

2.6	Radial distribution function, $g(r)$ , for state point in the solid phase along $\phi_p = 0.4$ . . . . .	35
2.7	Probability distribution of the orientational correlation for various values of $\epsilon$ with $\xi = 0.1$ and $\lambda = 2.5$ . . . . .	35
2.8	Representative snapshots for $\epsilon = 5.0$ (left), $\epsilon = 8.0$ (center), and $\epsilon = 11.0$ (right) after equilibration. Model protein particles shown as large blue spheres with the point charge represented by a small red sphere on the surface. . . . .	36
3.1	Rate of monomer loss for different angles, for the volume fraction $f = 0.0025$ . As discussed in the text, there are three different regions for the cases of $\theta = 120$ and $135$ degrees. Typical configurations for these three regions are also shown. . . . .	44
3.2	Cumulative distribution function for different angles. The equilibrium size of the aggregates increases as the angle between the patches is increased. . . . .	45
3.3	Log-log plot of the radius of gyration, $R_g$ . The aggregates form approximately spherical, compact aggregates below sixty degrees. As one increases the angle further, the significant variation in $R_g$ shown here indicates the formation of larger, non-spherical aggregates. Typical final aggregates for the angle $\theta$ equal to 60, 65, 75 and 90 degrees, respectively. For $\theta = 75$ degrees, the growth kinetic exponent shown here is close to the DLCA kinetic exponent of 0.55, indicating that the aggregates are fractals, within the accuracy of our simulation. . .	46

3.4	Plot of the anisotropy vs time for several different angles. The anisotropy is defined as the ratio of the square of the largest and smallest principle radii of gyration. From this plot we conclude that the clusters have almost identical morphologies for $\theta = 30$ and 60 degrees and remains essentially constant through the simulation. The large values of the anisotropy parameter shown for the case of $\theta = 75$ degrees indicates that the clusters are more irregular in their shape. . . . .	47
3.5	Morphology for $\theta = 30$ degrees at time $t=10,000$ . . . . .	48
3.6	Morphology for $\theta = 60$ degrees at time $t=10,000$ . . . . .	49
3.7	Morphology for $\theta = 75$ degrees, at $t=10,000$ . The transition to the rod-like behavior shown in this figure occurs around $\theta = 65$ degrees . .	50
3.8	Morphology for $\theta = 90$ degrees, at $t=10,000$ , shows the beginning of an interconnected gel-like structure. . . . .	51
3.9	Morphology for $\theta = 120$ degrees, at $t=10,000$ . . . . .	52
3.10	Morphology for $\theta = 135$ degrees. . . . .	53
3.11	Morphology for $\theta = 145$ degrees at time $t=10,000$ , which shows a significantly more compact structure. . . . .	54
3.12	Nucleation occurring after a long induction time $t=13,100$ , for $\theta = 150$ degrees. . . . .	54
3.13	Closer to equilibrium, at $t=16,250$ . . . . .	55
3.14	Equilibrium aggregate, at $t=20,000$ . . . . .	55
3.15	The correlation function $\vec{q}_6(i) \cdot \vec{q}_6(j)$ , defined in the text. . . . .	56
3.16	The radial distribution function, $g(r)$ , as a function of spatial separation $r$ . . . . .	56

3.17	log-log plot of the structure factor at early times, $t = 100$ . There appear to be no structure at large length scales. At short length scales, the aggregates are less compact and exhibit similar structures for all the angular variations. . . . .	57
3.18	log-log plot of the structure factor at late time, $t = 10,000$ . As discussed in the text, the aggregates at 75 and 90 degree angles appear to be more fractal like at large length scales, with the value of $S(q) \sim q^{-1.7}$ , which is close to the DLCA value of 1.8. . . . .	57
4.1	Schematic representation of an amino acid in the PLUM model from Reference 12. Dashed lines represent beads from neighboring residues to illustrate local geometry. . . . .	63
4.2	Schematic of the PLUM hydrogen bonding interaction from Reference 12. Dashed beads are the implicit hydrogen and oxygen atom positions, calculated from the backbone heavy atom positions. . . . .	68
4.3	Aggregate mass growth over time for two supersaturations of $Q_{10}$ . The higher value of supersaturation, $s=1.99$ , (red) demonstrates barrierless growth. Its underlying cluster distribution shows stable clusters across a range of sizes. At lower supersaturation, $s=0.78$ , (blue), a lag period is followed by the nucleation of a cluster with size above $n_c$ , and finally growth. Such systems contain a large population of transient oligomers of size less than $n_c$ with a gap in sizes before stable clusters emerge. . . . .	70

4.4	Scaling of lag phase with initial protein concentration disagrees with classical nucleation theory. The best fit slope corresponding to $\gamma = -0.57$ is shown in black. The slope corresponding to $\gamma = -0.5$ , when primary nucleation is not the dominating contribution to the lag phase, is shown in dashed red. . . . .	72
4.5	Critical nucleus size as a function of supersaturation, fit to $s^{-2}$ . Inset are critical nuclei for $S=0.47, 0.65$ , and $1.13$ , which feature 3, 2, and 1 layer $\beta$ -sheet conformations respectively, with peptides color coded to layer. . . . .	74
4.6	Average radius of gyration of clusters over time for three supersaturations of $Q_{10}$ . Higher supersaturations have smaller average size, reflecting the stability of clusters smaller than the critical nucleus size of lower supersaturations. . . . .	75
4.7	Average cluster anisotropy (ratio of largest to smallest moments of inertia) over time for three supersaturations of $Q_{10}$ . Clusters formed are highly anisotropic due to the typically extended conformation of $Q_{10}$ monomers within the cluster and the preferred direction imparted by hydrogen bonding. . . . .	76
4.8	Cluster distribution plot for $Q_{10}$ aggregates. For $S = 1.99$ (left), clusters exist across a wide range with no gaps. For the case of $S = 0.72$ (right), there exist a large number of transient small clusters, followed by a gap in the distribution, and finally a number of stable clusters above $n_c = 7$ . . . . .	76

4.9	Aggregate mass growth over time for three supersaturations of $Q_{30}$ . All show the same qualitative behavior and consistent values of $A$ and $\kappa$ . . . . .	78
4.10	Typical $Q_{30}$ tetramer forms from stacked helices in contrast to the sheet-like $Q_{10}$ clusters. $Q_{30}$ monomers adopt a more helical conformation than $Q_{10}$ , with significantly larger self-interactions. . . . .	79
4.11	Distribution of cluster sizes for $Q_{30}$ aggregates at monomer plateau. Clusters span the range of dimers to 16-mers with no gaps in the distribution. Growth at this stage is being driven by cluster combination.	80
4.12	Average radius of gyration of clusters over time for three supersaturations of $Q_{30}$ shows a quick increase during initial cluster formation followed by a period of slow growth. . . . .	81
4.13	Average cluster anisotropy (ratio of largest to smallest moments of inertia) over time for three supersaturations of $Q_{30}$ . Clusters maintain a very high degree of anisotropy over time due to their stacked helix arrangement. . . . .	82
4.14	Arrhenius plot of the nucleation rates of $Q_{10}$ and $Q_{30}$ . $Q_{10}$ shows two distinct regions, one with and without a nucleation barrier at high and lower temperature respectively. $Q_{30}$ shows good agreement to a zero-slope linear fit, indicating the lack of a barrier to nucleation. . .	83
4.15	Arrhenius plot of the nucleation rates of $Q_{15}$ , $Q_{20}$ , and $Q_{25}$ . Only $Q_{15}$ shows a non-negligible slope in the high temperature regime, and for a smaller range of temperatures than $Q_{10}$ . . . . .	84

4.16 Snapshots of system progression to long time scales of  $Q_{10}$  at high supersaturation. Initially a large number of small clusters are formed (left), which is followed by a period of consolidation into large, disordered oligomers (center). Finally these clusters reorganize into characteristic fibril forms. . . . . 85



# Study of the Effects of Anisotropy in Self-Assembling Systems

Jason L. Haaga

## Abstract

There are many things in the world that are not spheres. As a result, isotropic interaction potentials can only serve as a crude approximation to complex molecules such as proteins. In order to better understand the often harmful self-assembly phenomena that proteins can undergo, study of how anisotropic features alter collective behavior is required. Beyond the biological, these lessons can also guide rational design of new materials in the area of colloidal science. This dissertation examines the role of specific anisotropic features in coarse-grained representations of three proteins that undergo self-assembly processes. The first, amelogenin, is the primary protein involved in the formation of dental enamel; chapter 2 will explore the effect of the charged hydrophilic tail on the phase diagram of this otherwise hydrophobic protein. In chapter 3, the role of hinge angle between binding sites will be assessed in a simplified model of human antibodies; this angle has dramatic effect on aggregate morphology. Lastly, in chapter 4, the nucleation and growth mechanisms of polyglutamine tracts of different lengths will be studied, relevant to the formation of a class of neurodegenerative diseases, including Huntington's Disease. This process shows a strong dependence on repeat length, and at shorter lengths, variation due to concentration.

# Chapter 1

## Introduction

### 1.1 Background

Biological systems provide a wealth of test cases to probe many fascinating aspects of self-assembly. From the helpful, such as biomineralization, to the harmful, as in the case of diseases such as cataracts, sickle-cell anemia, or Alzheimer's, self-assembling systems are foundational to living matter. Self-assembly is ubiquitous in the realm of biochemical production and pharmaceuticals; additionally, new bioinspired materials are being created at a rapid pace. In protein science, the morphology and phase behavior of proteins is also strongly affected by anisotropic interactions. Proteins are inherently anisotropic in nature due to the differences in the exposed surface amino acid groups and their individual contributions to the overall interaction between proteins. There are thus hydrophobic, hydrophilic, charged and other regions on the protein surface, and a full understand of protein phase diagrams requires taking such effects into account.<sup>55,80,115</sup> These phase

behaviors include gas-liquid coexistence, liquid-liquid phase separation,<sup>56</sup> crystallization,<sup>36, 56, 78, 97, 135</sup> gelation,<sup>26, 112, 114, 123</sup> and aggregation.<sup>51, 55, 62, 74, 75, 137</sup> One field in which protein interactions are especially important is protein crystallization. For example, the growth of high-quality crystals is vital to the accurate determination of protein structure through x-ray diffraction. Additionally, an understanding of the kinetics of protein crystallization is important in biomedical applications. In particular, the release rates of medications such as insulin<sup>18, 76</sup> are based on the dissolution rates of protein crystals; steadier doses over extended time periods can be obtained by carefully controlling the nucleation of such crystals. Finally, numerous diseases, such as cataracts and sickle cell anemia, result from undesired protein condensation.<sup>55</sup> Study of these systems offers an opportunity to identify crucial features in the biology while simultaneously exploring the physics of anisotropy's role in self-assembly.

In colloidal science, developments in particle synthesis are beginning to provide a set of experimental techniques to create colloidal particles with different shapes and surface chemical compositions.<sup>39, 109, 136</sup> As a consequence one can create colloids with a desired number of solvophilic and solvophobic regions on the surface.<sup>52, 84</sup> This allows one to provide valence to colloids<sup>52, 53, 124</sup> and, hence, opens the new field of supra-particle colloidal chemistry. These colloidal particles have been the subject of recent experimental investigation<sup>58, 60, 100, 105</sup> as they exhibit strongly directional interactions. These advances in our ability to synthesize such materials are also permitting developments in photonic devices and biomaterials.<sup>16, 105</sup> As colloidal particles of increasing complexity can be fabricated, such as Janus particles<sup>128</sup> or functionalized DNA-coated colloids,<sup>29</sup> a more complete understanding of how surface

inhomogeneity effects phase behavior is therefore highly desirable. Long term, by building a “dictionary” of anisotropic features and their resulting effects, it could be possible to tailor solutions to specific applications with a bottom-up approach.

Computational study of the collective behaviors of complex molecules, such as proteins, is inherently limited by issues of scale. Proteins are comprised of hundreds to tens of thousands of atoms, undergoing motions on time scales many orders of magnitude smaller than the biological processes involved. To make such problems tractable, the molecular interactions are typically approximated by primarily isotropic potentials with smaller anisotropic perturbations. A number of methods have been developed to reproduce the directional nature of protein-protein interactions, such as patchy colloidal models,<sup>68</sup> coarse grained bead representations,<sup>1,10,12,32,93</sup> and polymer inspired tube approaches.<sup>6,59</sup> These greatly reduce the degrees of freedom present in the system, lessening computational requirements, and accelerating dynamics. Such models can hopefully provide insight into both the proteins of interest in addition to simpler systems they directly represent.

The work done here will utilize both Monte Carlo (MC) and Molecular Dynamics (MD) simulations to examine both equilibrium and dynamical properties of models inspired by proteins to better understand the assembly of the biological system and colloidal analogs. This chapter presents an overview of the simulation techniques used to study assembly in colloidal and protein systems. We then provide a brief review of protein structure as well as coarse grained representations thereof.

## 1.2 Simulation Methods

Computer simulation provides a powerful tool set to address problems in many-body systems. Simulation allows for a direct “measurement” of properties potentially unaccessible via physical experiment, since we know the positions and momenta of all microscopic components. Conversely *in silico* models are limited by both our knowledge of the underlying physics and the number of calculations required to evaluate the system at each simulation step. The techniques used in the simulation of molecular systems typically fall into one of two major categories.

The first, molecular dynamics (MD)[find good citation for MD], involves numerically solving the equations of motion of the system to obtain particle positions and velocities at discretized time steps. Particle trajectories are followed until reaching equilibrium, and time averaged measurements are then obtained for desired properties. In many respects, this is an *in silico* equivalent of an actual experiment. This method has the advantage of allowing us to probe dynamical quantities as well as follow the likely physical path of the system through phase space.

In the other, Monte Carlo simulation (MC),<sup>88</sup> the dynamics of the system are disregarded. System configurations are updated from one simulation step to the next by a series of random trial moves. An individual move is accepted or rejected based on an acceptance criteria that generates a valid statistical mechanical ensemble. As a result, measurements performed over a series of MC steps generate an ensemble average of the desired properties. The primary advantage of this technique is that trial moves need not be physical, and the system is allowed to find shortcuts through phase space to reach equilibrium with fewer calculations required.

In general, we will use MD when examining kinetic properties of the system and

MC when the equilibrium state is of interest. Within each category are a wealth of specific techniques tailored to a variety of different problems. We will briefly review those utilized within this thesis.

### 1.2.1 Molecular Dynamics

Molecular dynamics evolves a system by solving Newton's equations of motion in discretized time steps. Each step, the sum of the forces acting on each particle is calculated and used to update their respective trajectories. The system is stepped forward a small unit of time and the process repeated until measured quantities no longer change with time. At its simplest, this method is the evaluation of the classic  $\vec{F} = m\vec{a}$  repeated billions of times.

#### Velocity Verlet Integrator

While there are numerous techniques for numerically integrating these equations, the most common in use today is the Velocity Verlet scheme.<sup>127</sup> This method is a combination of two Taylor expansions of position as a function of time; one taking a step forward from  $t$  to  $t + \Delta t$ :

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2x(t)}{dt^2}\Delta t^2 + \frac{1}{3!}\frac{d^3x(t)}{dt^3} + O(\Delta t^4) \quad (1.1)$$

and another in the reverse direction from  $t$  to  $t - \Delta t$ :

$$x(t - \Delta t) = x(t) - \frac{dx(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2x(t)}{dt^2}\Delta t^2 - \frac{1}{3!}\frac{d^3x(t)}{dt^3} + O(\Delta t^4) \quad (1.2)$$

Combining equations 1.1 and 1.2 cancels odd terms, resulting in:

$$x(t + \Delta t) = x(t) + \frac{1}{2} \frac{d^2 x(t)}{dt^2} \Delta t^2 + O(\Delta t^4) \quad (1.3)$$

This algorithm is accurate to third order, and thus has a local error of order  $O(\Delta t^4)$ , despite only containing a second order derivative. Additionally this method does not require calculation of velocities in order to obtain changes in positions, only the acceleration which is directly obtained through calculation of the net force on a particle. As a result, the Velocity Verlet scheme provides an excellent balance of accuracy and efficiency for most problems.

### Statistical Ensemble and Solvent Effects

A typical MD simulation is done at fixed particle number (N) and system volume (V). Energy is conserved when integrating Newtonian equations of motion, to within a local error dependent on the integration method. This results in a simulation sampling states in the NVE ensemble; experiments, however, are typically done under constant temperature, rather than energy, conditions, resulting in an NVT ensemble.<sup>133</sup> Fortunately, there are multiple methods to ensure sampling of the proper ensemble. In systems involving large macromolecules or colloidal particles in a bath of much smaller solvent particles, it is often efficient to take a continuum approach and model the effects of solvent implicitly. Such methods, known as Brownian Dynamics (BD) are based on integrating the Langevin Equation.<sup>99</sup>

$$m_i \ddot{\vec{x}}_i(t) = -m_i \Gamma_i \dot{\vec{x}}_i(t) + \vec{\nabla} U_i + \vec{\xi}_i(t) \quad (1.4)$$

Which is Newtons Second Law with the addition of a stochastic white noise term,  $\vec{\xi}(t)$ , and a viscous drag term,  $\Gamma$ .  $\vec{\xi}(t)$  satisfies the following criteria:

$$\langle \xi_i(t) \rangle = 0 \tag{1.5}$$

$$\langle \vec{\xi}_i(t) \cdot \vec{\xi}_j(t) \rangle = \delta_{ij} \tag{1.6}$$

$$\langle \vec{\xi}(t) \cdot \vec{\xi}(t') \rangle = 2D\delta(t - t') \tag{1.7}$$

$\xi$  and  $\Gamma$  are related via the fluctuation-dissipation relation yielding the following combined relationship:

$$\langle \xi_i(\vec{t}) \cdot \xi_j(\vec{t}') \rangle = 6k_B T \Gamma \delta_{ij} \delta(t - t') \tag{1.8}$$

Integrating these equations motion has been shown to yield trajectories sampling from the canonical distribution.<sup>63</sup>

Alternatively the same sampling can be achieved through a stochastic coupling, known as the Anderson thermostat.<sup>2</sup> Particle velocities are randomly reassigned to satisfy a Maxwell-Boltzmann distribution at the desired temperature. This simulates the effect of collisions with the solvent bath, leading to an NVT ensemble.<sup>47</sup>

In cases where the solvent is explicitly modeled, a simple velocity rescaling thermostat may be used. Under this scheme, system temperature is periodically calculated, and all particle velocities proportionally rescaled to result in the desired temperature.

In protein systems, solvent molecules typically outnumber protein molecules by several orders of magnitude. As a result, explicit solvation requires considerably more computational investment to calculate these interactions or a corresponding



reduction in system size. Many aspects of the solvent can be successfully incorporated into implicit models, from simple screened electrostatics to mimic ion concentration<sup>31</sup> to complex hydrodynamic effects.<sup>3</sup>

## 1.2.2 Monte Carlo Simulation

From statistical mechanics, the expectation value of a quantity in a system with Hamiltonian  $H$  is given by an integral over all potential particle positions and momenta:

$$\langle A \rangle = \frac{\int A(q^N, p^N) \exp\left[-\frac{H(q^N, p^N)}{k_B T}\right] dq^N dp^N}{\int \exp\left[-\frac{H(q^N, p^N)}{k_B T}\right] dq^N dp^N} \quad (1.9)$$

For most systems, analytic solutions to this integral are impossible and the dimensionality is sufficiently high to rule out simple numeric quadrature approaches. Additionally, over most of phase space, the integrand is often zero. The Monte Carlo method is an effective alternative in many such cases.

If the quantity of interest is not dependent on the kinetic portion of the Hamiltonian, we can reduce the expression to:

$$\langle A \rangle = \frac{\int A(q^N) \exp\left[-\frac{U(q^N)}{k_B T}\right] dq^N}{\int \exp\left[-\frac{U(q^N)}{k_B T}\right] dq^N} = \int A(q^N) P(q^N) dq^N \quad (1.10)$$

where  $P(q^N)$  is a probability distribution of states corresponding to the statistical ensemble. Sampling the quantity of interest at random points according to this

distribution will converge to the desired average. In order to achieve this, a random walk through the non-trivial portions of phase space is constructed. A trial move is proposed to take the system from state  $o$  to a new state  $n$  and is accepted or rejected such that  $P(q^N)$  is reproduced.

At equilibrium, the net probability flow from state  $o$  to  $n$  must be equal to that of  $n$  to  $o$ . This principle of detailed balance requires:

$$P(o)p_{acc}(o \rightarrow n) = P(n)p_{acc}(n \rightarrow o) \quad (1.11)$$

$$\frac{p_{acc}(o \rightarrow n)}{p_{acc}(n \rightarrow o)} = \frac{P(n)}{P(o)} = \exp\left[\frac{U(o) - U(n)}{k_B T}\right] \quad (1.12)$$

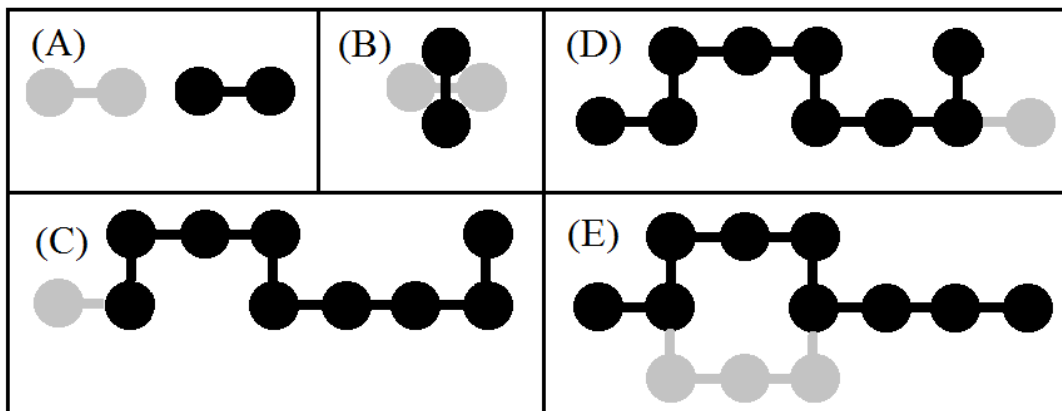
This can be satisfied by a number of different choices, but the most commonly used scheme, first proposed by Metropolis *et al*<sup>88</sup> chooses:

$$p_{acc}(o \rightarrow n) = \begin{cases} \frac{P(n)}{P(o)} & P(n) < P(o) \\ 1 & \text{else} \end{cases} \quad (1.13)$$

From this the basic MC algorithm emerges: trial moves are proposed, accepted or rejected according to this criteria, and repeated sufficiently many times to adequately sample phase space.

## Canonical Ensemble

The Canonical Ensemble (NVT) is a common choice for determining equilibrium configurations at a known density and temperature, sampling for the canonical ensemble of statistical mechanics. These variables are usually easy to control in



**Figure 1.1:** Examples of configurational Monte Carlo moves: translation (A), rotation (B), reptation (C), pivot (D), and crankshaft (E). Pre-move positions shown in grey. Post-move positions in black.

experimental study, and thus Canonical Monte Carlo provides an efficient means to validate simplified *in silico* models against their *in vitro/vivo* counterparts.

The NVT scheme is also one of the simplest of the variants of MC typically used in statistical physics. Trial moves used in this ensemble are typically only random translations and rotations. In the case of isotropic particles, only translations are required. For the case of chain-like polymer models, it is inefficient to rely only on this simple trial moves, and thus a set of multiple particle moves are typically implemented as well, such as reptation (a snake-like move forward or backward), pivot (rotating an end segment about an internal point), and crankshaft (rotating an internal section about the line connecting two pieces of the chain).

Regardless of the level of complexity, these moves only alter the configurational portion of the systems partition function, and thus all obey the same acceptance

criteria:

$$p_{acc}(o \rightarrow n) = \min[1, \exp(-\beta(U(n) - U(o)))] \quad (1.14)$$

Here  $U$  is the potential energy of a given configuration, and  $\beta$  is the standard  $1/k_B T$ . We can justify the conditions of detailed balance as follows easily. The probability of being in a specific state  $o$  is given by:

$$N(o) = \frac{\exp(-\beta U(o))}{Q} \quad (1.15)$$

With  $Q$  being the canonical partition function of the system. Our probability of generating a configuration is constant, and so the net flow for a state  $o$  to  $n$ , assuming  $n$  is higher in potential than  $o$ , is thus:

$$K(o \rightarrow n) = \frac{\exp(-\beta U(o))}{Q} \alpha \exp(-\beta(U(n) - U(o))) \quad (1.16)$$

And the reverse flow, again with  $n$  being higher in potential than  $o$ :

$$K(o \rightarrow n) = \frac{\exp(-\beta U(n))}{Q} \alpha(1) \quad (1.17)$$

Which can easily be seen to be the same expression. There is no net flow between states  $o$  and  $n$ , and thus detailed balance is enforced.

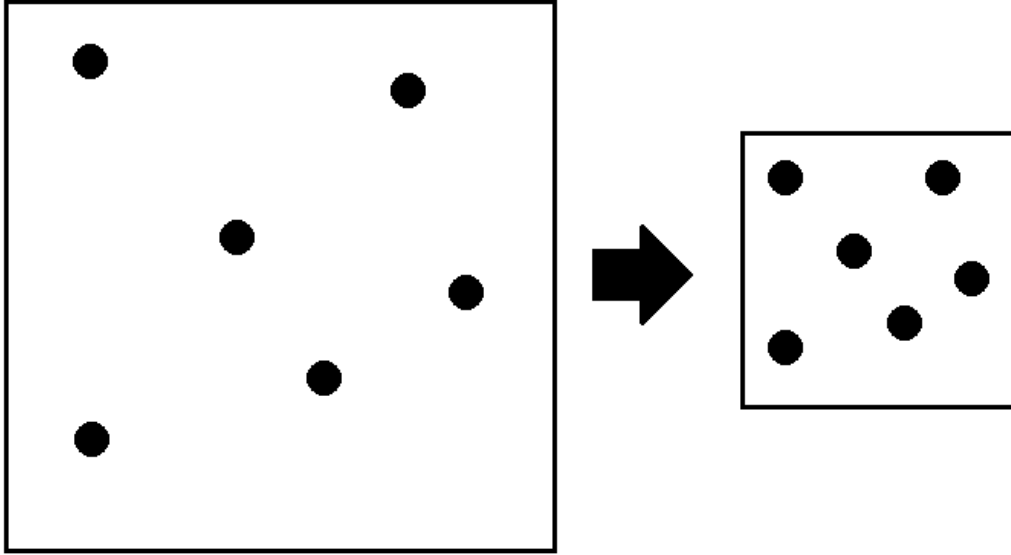
An important note is that only the portions of the system energy involving the moved particles require recalculation in order to determine move acceptance. This

results in a balance between the sophistication of the move (i.e. simple translation versus crankshaft) and the cost of calculation that should be examined when approaching a new system.

### **Isobaric-Isothermal Ensemble**

Another commonly used ensemble is the isobaric-isothermal (NPT). This ensemble has particular importance in measuring the equation of state, even for highly complex systems where analytic expressions or even reasonable expansions for the pressure may be unobtainable. Additionally this ensemble will tend towards the lowest Gibbs free energy state of the system, in contrast to NVT simulations which may attempt (and likely fail due to insufficient system size) to phase separate in the vicinity of a phase transition. This ensemble's volume altering move is also vital to equilibrating of dense phases in a reasonable time frame; periods of decreased density allow for greater success rates of the standard configurational moves seen in the NVT ensemble.

The NPT ensemble retains the configurational move set allowed by the NVT ensemble, which retains its acceptance criteria (Eq 1.14); it may also make a trial move to alter the system volume such that  $V' = V + \Delta V$ . Particle positions are uniformly scaled by a ratio  $\frac{V'}{V}$ . An important note is that in the case of molecular systems, only the center of mass coordinates should be scaled in this manner; relative positions and orientations of a molecule's components should be maintained to avoid almost guaranteed move rejection due to excessive stretching or twisting of bond interactions. The acceptance criteria for a volume move is given by:



**Figure 1.2:** Examples of volume Monte Carlo move. Inter-particle distances are all scaled by the same ratio as the change in box length.

$$p_{acc}(o \rightarrow n) = \min[1, \exp(-\beta[U(n) - U(o) + P(V' - V) - \frac{N}{\beta} \ln(\frac{V'}{V})])] \quad (1.18)$$

Alternatively the volume change can be done via a change in individual box length or in the logarithm of the volume if desired. While equally valid, such trial moves require a subtly different acceptance criteria:

$$p_{acc}(o \rightarrow n) = \min[1, \exp(-\beta[U(n) - U(o) + P(V' - V) - \frac{N+1}{\beta} \ln(\frac{V'}{V})])] \quad (1.19)$$

While configurational moves only require a recalculation of a subset of the total

internal energy, the volume move requires a complete calculation of the internal energy, with the exception of atomic systems with potentials that may be expressed purely as linear combination of powers of inter-particle distances. As a result, they should be used less frequently than configurational moves.

### Grand Canonical Ensemble

The grand canonical ensemble (GCMC, fixed  $\mu VT$ ) allows density to fluctuate by inserting and deleting particles at constant volume. This method is the gold standard in obtaining phase coexistence between gas and liquid phases. At the chemical potential corresponding to phase coexistence at a given temperature, the system will oscillate between phases, sampling each for a time before swapping; from this a coexistence curve may be constructed. For particle insertion, a random location in the simulation box is chosen, and a new particle created with randomized orientation. The acceptance criteria for this move is given by:

$$p_{acc}(N \rightarrow N + 1) = \min\left[1, \frac{V}{N + 1} \left(\frac{2\pi m}{h^2\beta}\right)^{\frac{3}{2}} \exp(\beta[\mu - U(N + 1) + U(N)])\right] \quad (1.20)$$

For deletion, a random particle is selected and destroyed. This move follow the acceptance criteria of:

$$p_{acc}(N \rightarrow N - 1) = \min\left[1, \frac{N}{V} \left(\frac{h^2\beta}{2\pi m}\right)^{\frac{3}{2}} \exp(-\beta[\mu + U(N - 1) - U(N)])\right] \quad (1.21)$$

This method encounters difficulties at high densities due to the low probability

of success of either move. As such, it is impractical for coexistence with a solid phase or highly extended molecules.

## Gibbs Ensemble

The Gibbs Ensemble (GEMC) samples from the NVT ensemble, but is of special use for simulations of phase equilibrium. Simulations of direct phase coexistence require very large system sizes, as a large portion of particles will be involved in forming the boundary between the two or more phases. This results in a significant amount of wasted computation to determine the properties of the bulk. GEMC allows one to equilibrate the pressure, temperature, and chemical potential of connected single phase boxes without the expense of forming a boundary between them. This is done by allowing these boxes to exchange particles and volume with each other, so that while the density of any individual box is allowed to fluctuate, the total number of particles and total volume are still conserved.

Within a given box, the standard set of NVT ensemble moves may be utilized with the previously mentioned acceptance criteria (Eq 1.14). Additionally, a particle may be exchanged with another box, or the volume of a box may be changed while conserving total volume. The presence of these two moves allows for equilibration of both pressure and chemical potential between the two boxes, which are both varying over the course of the simulation. This is a tremendous advantage in that we do not require having the often precise chemical potential or pressure needed for phase coexistence in advance. By requiring detailed balance in both these moves, we can derive acceptance rules without explicit knowledge of  $P$  or  $\mu$ .

Beginning with particle exchange, if we remove a particle from box 1, originally



containing  $n_1$  particles, and insert into box 2, originally containing  $n_2$  particles, the ratio of the probabilities of these configurations is:

$$\frac{N(n)}{N(o)} = \frac{n_1!n_2!V_1^{n_1-1}V_2^{n_2+1}}{n_2V_1} \exp[-\beta(U(n) - U(o))] \quad (1.22)$$

Requiring detailed balance  $K(o \rightarrow n) = K(n \rightarrow o)$  then gives:

$$p_{acc}(o \rightarrow n) = \min\left[1, \frac{n_1V_2}{n_2V_1} \exp(-\beta(U(n) - U(o)))\right] \quad (1.23)$$

In the case of a two box system,  $n_2$  is constrained to  $N - n_1$  and likewise  $V_2 = V - V_1$ . The method, however, generalizes to any arbitrary number of boxes with the above acceptance criteria.

For a volume change of size  $\Delta V$  to box 1, such that  $V_1 = V_1 + \Delta V$  and  $V_2 = V_2 - \Delta V$ , we can again set up the ratio of the statistical weights of the configurations before and after the trial move ( $o \rightarrow n$ ).

$$\frac{N(n)}{N(o)} = \frac{(V_1 + \Delta V)^{n_1}(V_2 - \Delta V)^{n_2}}{V_1^{n_1}V_2^{n_2}} \exp[-\beta(U(n) - U(o))] \quad (1.24)$$

Imposing detailed balance, we require:

$$p_{acc}(o \rightarrow n) = \min\left[1, \left(\frac{V_1 + \Delta V}{V_1}\right)^{n_1} \left(\frac{V_2 - \Delta V}{V_2}\right)^{n_2} \exp(-\beta(U(n) - U(o)))\right] \quad (1.25)$$

Again, these expressions simplify to being a function purely of  $n_1, V_1$  if only two boxes are used but can be extended to an arbitrary number.

While GEMC has obvious advantages in time compared to GCMC or NPT-MC, it notably fails in systems involving strong, short-range potentials or in the presence of very dense liquids or solids. This is largely due to the difficulty in making both effectively a particle deletion and insertion successfully in these cases. It is also often difficult to equilibrate even in the remote vicinity of a critical point; large fluctuations can cause the boxes to swap phases too quickly to allow internal equilibration.

### **Extended Ensembles**

Often in Monte Carlo simulation, a system may prove excessively time consuming to adequately sample phase space. In order to short cut around any potentially large barriers between states, extended ensemble simulations may be used. This process involves the running multiple replicas of the original configuration in one of the previous MC ensembles, each using a different value for a particular thermodynamic variable. Periodically a swap of system configurations is attempted between neighboring replicas.

Most commonly temperature is used in this role (parallel tempering). Replicas are assigned unique temperatures that range from the temperature of interest to significantly higher temperatures. The lower temperature replicas will primarily explore local minima, while the higher temperature replicas allow for more frequent potential barrier crossing. Swaps of system configurations are attempted between neighboring temperatures  $T_i$  and  $T_j$  with acceptance probability:

$$p_{acc}(i \leftrightarrow j) = \min[1, \exp((\beta_j - \beta_i)(U_i - U_j))] \quad (1.26)$$

The spacing between neighboring temperatures need not be constant for this method; these spacings and the frequency of attempting swaps should be selected such that low temperature replicas have limited but sufficient time to explore local phase space before it is likely they are swapped to a higher temperature. Parallel tempering may be employed in any ensemble to address large potential barriers between states.

Similarly one can extend in pressure or chemical potential for NPT or GCMC simulations respectively. This is extremely useful in the measurement of phase coexistence. By selecting pressure or chemical potential values that favor different density phases at their extremes, it is possible to efficiently sample both phases at a single equilibrium value in between.

### 1.3 Protein Structure

In this section, we will give a physicists review of the basics of protein structure and examine the models used to represent them. Proteins are an amazingly diverse group of biomolecules, with sizes ranging over several orders of magnitude. They are capable of a vast array of chemical interactions, large conformational changes in response to environment, and combining into complex assemblies. The degree to which different levels of the protein structure are important to a given problem shapes the models we choose to address it.

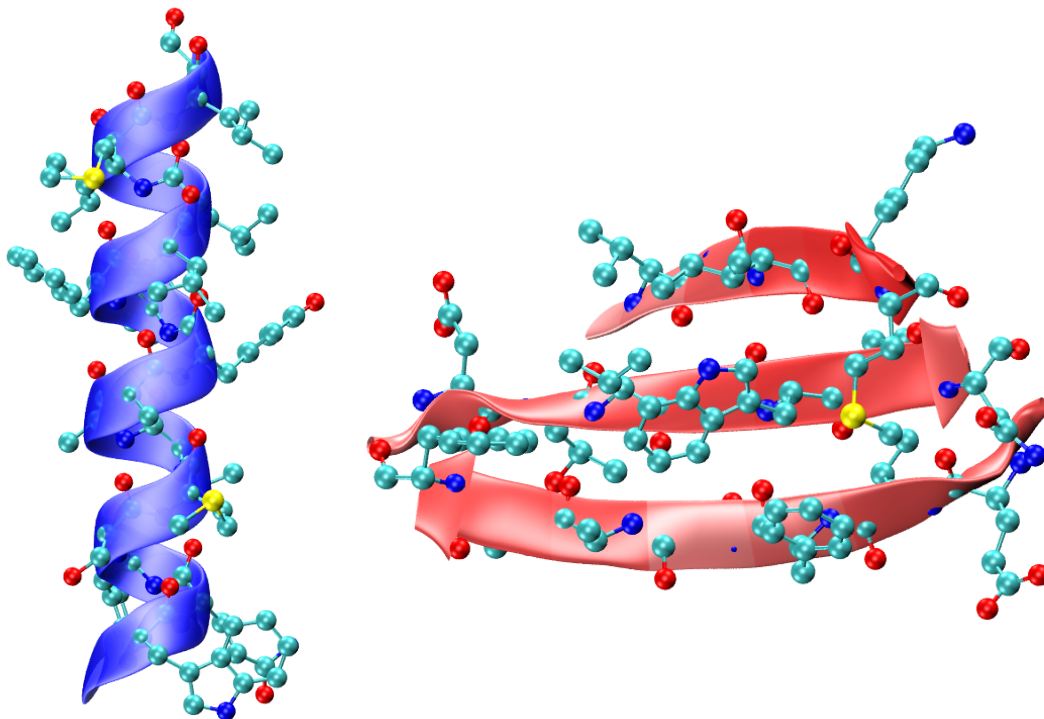
## Primary Structure

At the most basic level, all proteins are polymer chains of amino acids, also referred to as residues, of which there are 20 important to biological systems. Each amino acid consists of a central carbon atom ( $C - \alpha$ ), a carboxyl group ( $COO^-$ ), an amino group ( $NH_3^+$ ), and a sidechain (specific to each amino acid). Amino acids are linked via a peptide bond from carboxyl group of one to the amino group of the next monomer. Sidechains, or R groups, vary dramatically across the amino acids, from alanines simple four atom group to tryptophans aromatic rings. This sequence of amino acids is referred to as primary structure, and is the information encoded within DNA.

## Secondary Structure

The next tier of this hierarchy is the three-dimensional conformation of the backbone (carboxyl group, amino group, and  $C - \alpha$ ) of a protein, or its secondary structure. There are numerous effects which may contribute to the resulting structure, but it is dominated by hydrogen bonding between amino acids, which provides a powerful, and highly directional, attractive interaction. These interactions are constrained by steric restrictions, bond length and geometry, and even four body constraints arising from the delocalization of electrons in the peptide bond. Even within these constraints, there exists a world of potential conformations a protein may adopt.

One of the most common is the  $\alpha$ -helix, in which amino acids are arranged into a helical shape with 3.6 residues per turn. This is an extremely stable structure, with each residue forming two hydrogen bonds. This stability makes it relatively easy to predict from primary structure.



**Figure 1.3:** Examples of common secondary motifs: cartoon representations of an  $\alpha$ -helix (left) and a three stranded  $\beta$ -sheet (right) overlaid on their underlying atomistic composition.

Another common motif is the  $\beta$ -strand, an extended and highly linear conformation.  $\beta$ -strands can easily hydrogen bond between each other to form more complex structure as shown in Figure 1.3. This type of structure is of importance to aggregation in many neurological diseases, as the types of aggregates formed show high amounts of  $\beta$ -sheet and -helix content under dye assays.

Frequently proteins may contain significant regions that lack a well defined secondary structure under physiological conditions. These intrinsically disordered proteins may adopt a stable conformation under certain conditions (for example: in the presence of binding partners or at a particular pH or temperature), oscillate

between competing states, or simply contain large flexible sections that link ordered regions. These proteins often pose a challenge to highly simplified models, as their protein-protein interactions can be highly variable as a function of conformation.

## **Tertiary Structure**

The most crucial element in protein structure for the purposes of function, however, is the arrangement of sidechains in physical space, referred to as its tertiary structure. Hydrogen bonding, hydrophobic/hydrophilic effects, electrostatics, and van der Waals interactions all contribute, while constrained by the limitations imposed by the proteins secondary structure. As a result, predicting tertiary structure a given primary structure will fold into is a potentially very difficult task. Hydrophobicity has a powerful impact on this folding; non-polar sidechains will tend to crowd to the center of a protein, shielded by more hydrophilic or charged amino acids. Some residues, such as glutamine, are capable of forming hydrogen bonds from their sidechain in addition to those from the backbone of the molecule, which are again very strong and highly anisotropic. Net electrostatic charge or dipole-dipole interactions are longer ranged, and provide an avenue for pH response. Identifying which of these interactions are important to a given system is vital to determining the types of models to employ, as explicitly calculating all of them is typically infeasible beyond small systems without extensive computing hardware.

## **Quaternary Structure**

Lastly, many protein molecules are comprised of several polypeptide chains that have combined to form a larger macromolecule; this, when applicable, is their quaternary

structure. For example, the immunoglobulin family of proteins (antibodies) has variants that exist monomerically (ex: IgG), dimerically (ex: IgA), and even in pentamer form (ex: IgM). Typically, the subunits of such a protein are arranged symmetrically about an axis (rotations of  $360^\circ/n$ ). Subunits need not inherently be identical; for example, hemoglobin is a tetramer consisting of two pairs of two different peptide strands.

## Chapter 2

# Amelogenin: Effects of Coulomb Repulsion on Phase Diagram

### 2.1 Abstract

In this chapter we explore the role of an anisotropic charge interaction on the phase diagram of a colloidal model of the protein amelogenin. Amelogenin is the primary protein comprising the scaffolding in dental enamel upon which minerals are deposited. This globular protein is chiefly hydrophobic with a 25-residue charged C-terminus tail. Deletion of this tail has been shown to prevent the hierarchical assembly process *in vitro*.<sup>37</sup> Prior dynamics simulations have reproduced key features of this process,<sup>77</sup> representing the protein as a hydrophobic sphere with a tethered point charge. The hydrophobic interaction is given by the Asakura-Oosawa (AO) depletion potential, and the tethered charge represented by a screened Coulomb



potential. We examine the effect of altering the strength and range of these interactions on the phase diagram of the AO model in the range of values relevant to amelogenin.

## 2.2 Introduction

One approach to describing anisotropic interactions in both colloidal and protein systems is through so-called patchy models, introduced by Kern and Frenkel<sup>69</sup> and used by many groups since then, e.g. references.<sup>16,104,113</sup> These models have had reasonable success in predicting the overall features of phase diagrams. With sufficiently short-ranged attractive interactions, the phase diagrams associated with these models exhibit similar features as do those for proteins in aqueous solution: namely, a solid-liquid coexistence curve that is subtended by a metastable liquid-liquid curve. The region of metastability is of particular interest as the existence of this liquid-liquid critical point has been demonstrated to enhance crystal nucleation.<sup>122</sup>

An example of particular interest is the protein amelogenin, an important protein involved in the formation of dental enamel. Amelogenin is primarily hydrophobic with a 25-residue charged C-terminus tail. This protein undergoes a hierarchical assembly process that is crucial to mineral deposition. Experimental work has demonstrated that deletion of the C-terminus tail prevents this self-assembly.<sup>37</sup> Previous dynamics simulations<sup>77</sup> have reproduced the key features of this assembly process, utilizing a simplified model of amelogenin consisting of a hydrophobic sphere with a tethered point charge on its surface. The hydrophobic interactions were chosen as

the Asakura-Oosawa<sup>4</sup> depletion attraction interaction. The pure Asakura-Oosawa model has been of great utility in the statistical thermodynamics community, being used to model interactions between colloidal particles as well as between protein molecules, for example. In most cases the attractive interaction is produced, e.g., by the addition of non-adsorbing polymers to colloidal suspensions. The polymers are treated in the AO model as ideal soft spheres, whose fugacity controls the concentration of the polymers. A recent review of the AO model was given by Binder, Virnau and Statt.<sup>17</sup> The amelogenin model was studied as a case of a generic bipolar molecule for a variety of parameter choices in addition to those specific to amelogenin. It was found that aggregate morphology was highly sensitive to the relative strength and range of the interaction between point charges.<sup>74</sup>

In this chapter we examine the effect of these parameters on the phase diagram of this model in the range of values relevant to the previous study. This allows us to examine the impact of the point charge Coulomb repulsions between neighboring amelogenin molecules on the phase diagram of the pure AO model, studied by Evans *et al.*<sup>35</sup> and by Lekkerkerker and collaborators.<sup>72</sup> The phase diagram of the pure AO model is of considerable interest in its own right since it displays a stable liquid-liquid phase separation curve for some choices of parameter  $\xi$  (where  $\xi$  is defined as  $\xi = \sigma_c/\sigma_p$ , where  $\sigma_p$  and  $\sigma_c$  are the diameters of the polymer and colloid or protein particle, respectively), but transitions to a metastable phase diagram as the range of the attractive interaction is reduced as the parameter  $\xi$  is reduced.

## 2.3 Model and Method

As noted in the Introduction, our generic bipolar model consists of a spherical particle and a point charge located on its surface. Studies of amelogenin self-assembly are typically done in the presence of salt and a precipitant, such as polyethylene glycol (PEG). While protein-PEG interactions are complicated, we model this effect combined with hydrophobic interactions as a depletion interaction. In this chapter, we use Monte Carlo simulations to investigate the phase diagram of such a bipolar molecule. We study this model for a range of parameters, varying both the range and magnitude of each interaction.

The spherical particle is depicted as a repulsive hard core and an additional interaction, namely the well-known Asakura-Oosawa (AO) potential. The total potential between two spheres separated by a center-to-center distance  $r_{ij}^c$  is written as,  $U_c(r_{ij}^c) = U_{hc}(r_{ij}^c) + U_{AO}(r_{ij}^c)$ , where the hard core potential is given by

$$\frac{U_{hc}(r_{ij}^c)}{kT} = \begin{cases} 0 & r_{ij}^c > \sigma \\ \infty & r_{ij}^c < \sigma \end{cases} \quad (2.1)$$

and

$$\frac{U_{AO}(r_{ij}^c)}{kT} = \begin{cases} 0 & r_{ij}^c > 1 + \xi \\ \phi_p \left(\frac{1+\xi}{\xi}\right)^3 \left[\frac{3r}{2(1+\xi)} - \frac{1}{2}\left(\frac{r}{1+\xi}\right)^3 - 1\right] kT & r_{ij}^c \leq 1 + \xi \end{cases} \quad (2.2)$$

where  $\xi$  is the ratio between a PEG chain and an amelogenin particle, which controls the width of the attractive well, and  $\phi_p$  is the volume fraction of PEG, which sets

the strength of the AO interaction.<sup>4,129</sup>

The point charges on the spherical surfaces interact via a screened, due to the presence of salt ions, Coulomb potential,

$$U_p(r_{ij}^p) = \frac{\epsilon}{r_{ij}^p} \exp\left(-\frac{r_{ij}^p}{\lambda}\right), \quad (2.3)$$

in which the magnitude is controlled by  $\epsilon$ , a ratio of the strength of the Coulomb interaction to  $kT$ , and the interaction range is controlled by the Debye screening length  $\lambda$ . Parameters previously determined to be specific to amelogenin, and thus of particular interest, are given by  $\xi = 0.1$ ,  $\lambda = 0.4$  and  $\epsilon = 1.0$ .

To determine liquid-liquid coexistence, we first attempted Gibbs Ensemble Monte Carlo simulations. Validating this method for the case of zero charge against previous determinations of the phase diagram of the AO model<sup>35</sup> showed good agreement in the high-density liquid phase but a consistent over-estimation of the low-density phase. This discrepancy grew more pronounced with decreasing density of the sparser phase, which indicates the AO potential, even at moderate interaction lengths of  $\xi = 0.8$  is too “sticky” for this method. Small clusters in the low-density phase are unable to break apart, and equilibrium is not achieved between the two phases.

In order to alleviate this, we instead employ grand-canonical Monte Carlo simulations. Additionally, we use parallel-tempering<sup>46,49</sup> to accelerate equilibration and improve sampling, running 16 replicas simultaneously across a range of chemical potential,  $\mu$ . Within each replica, standard insertion/deletion, translation, and rotation moves were attempted using standard MC acceptance criteria. Periodically,

a replica was selected to attempt to swap with a neighbor with an acceptance probability

$$p_{acc}(x_i \leftrightarrow x_{i+1}) = \min[1, \exp(-\beta\Delta\mu\Delta N)] \quad (2.4)$$

Periodic boundary conditions were used for a cubic cell of  $L = 10 \sigma$ . Simulations were equilibrated for  $10^6$  MC steps and then sampled over an additional  $10^7$  MC step production run.

For liquid-solid coexistence, a similar process was employed with parallel-tempered, isobaric-isothermal (NPT) Monte Carlo simulations with 64 simultaneous replicas. In each simulation the local structure was initialized in an FCC lattice with  $N = 256$  particles. Standard MC translation, rotation, and volume change moves were attempted, in addition to a periodic swap with neighboring replicas with a probability given by:

$$p_{acc}(x_i \leftrightarrow x_{i+1}) = \min[1, \exp(-\beta\Delta P\Delta V)], \quad (2.5)$$

Simulations were equilibrated over  $5 \cdot 10^6$  MC steps and then sampled over a  $10^7$  step production run. In order to examine the effect of the Coulomb charge on the orientational ordering within the solid, a much larger range of parameters for  $\epsilon$  and  $\lambda$  than originally examined was required. Individual NPT-MC simulations were begun from a close-packed FCC lattice of  $N=108$  particles. Simulations were allowed to equilibrate over  $10^6$  MC steps, with a  $10^7$  step production run.

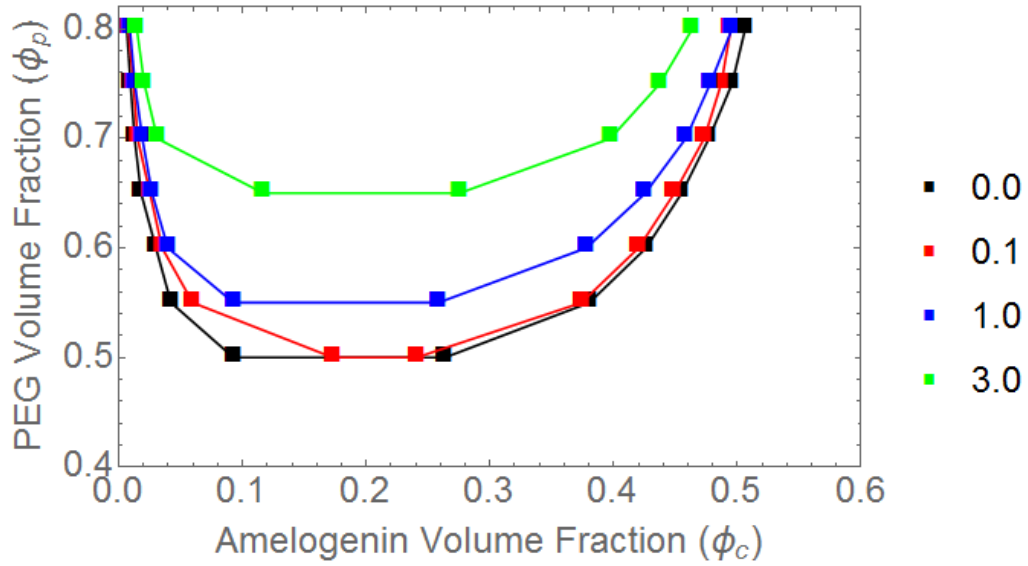
## 2.4 Results

### 2.4.1 Phase Diagrams

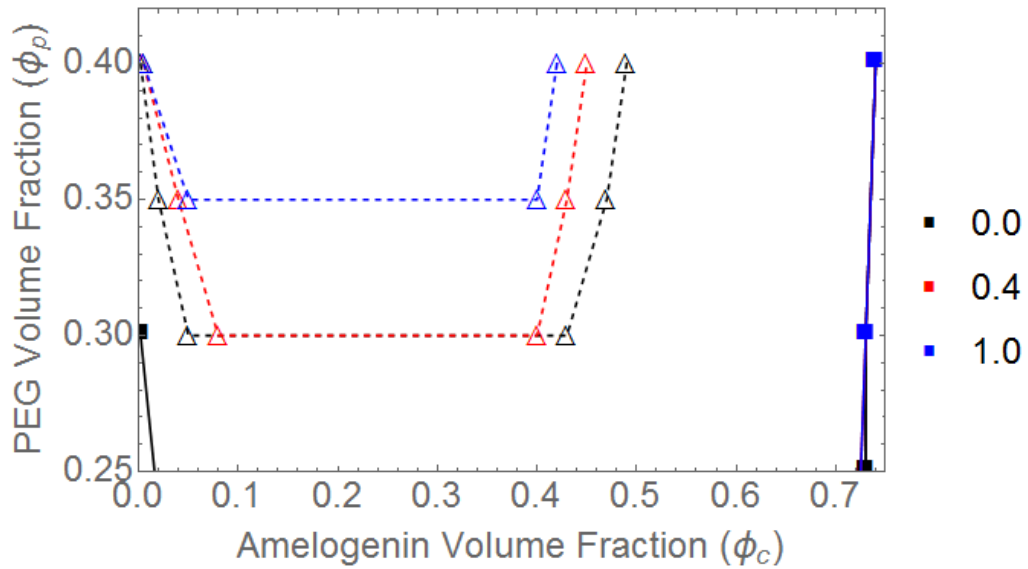
We first calculated the phase diagram for  $\xi = 0.8$  for  $\lambda = 0.4$  and varying  $\epsilon$  using the methodology outlined above, as shown in figure 2.1. We found that the stable liquid-liquid coexistence region shifts upward in  $\phi_p$  and narrows in volume fraction of our model particle,  $\phi_c$ , as repulsion is increased. A similar trend is seen in the behavior of the metastable liquid-liquid region for  $\xi = 0.1$ , as shown in figure 2.2. For this shorter AO interaction range, the effect of a small charge is significantly more pronounced than in the case of  $\xi = 0.8$ . Additionally, this behavior is also observed for changing  $\lambda$  at fixed  $\epsilon$  (see figure 2.3). In a previous study of this model,<sup>77</sup> it was shown that variations in  $\lambda$  or  $\epsilon$  produced differing aggregation morphologies. We find that the phase behavior is largely insensitive to the difference between increasing range versus strength, depending instead on their combined effect to control the overall amount of repulsion present in the system. This shift in the coexistence region is analogous to a lowering of  $T_c$  as overall attraction between particles is reduced.

### 2.4.2 Structure and Correlation Measurements

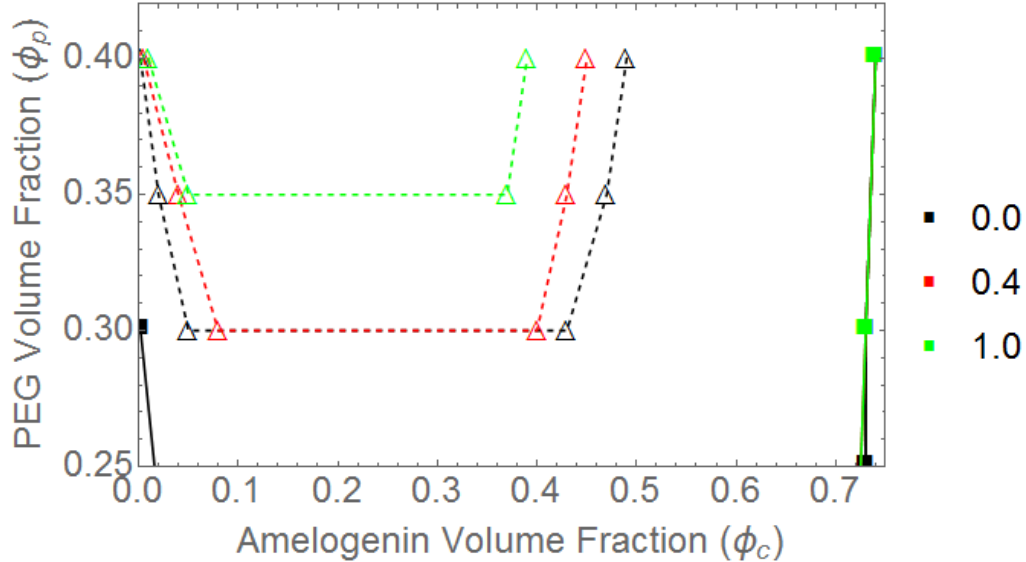
Next, we calculated the radial distribution functions,  $g(r)$ , for state points representative of the two metastable liquids and a solid phase for  $\xi = 0.1$  and  $\phi_p = 0.4$ , as shown in figures 2.4-2.6. The low-density liquid phase exhibits ideal gas-like behavior, due to the minimal amount of interaction at this concentration, whereas the high-density liquid exhibits some minimal structure typical of an attractive liquid,



**Figure 2.1:** Phase diagram in the  $\phi_p - \phi_c$  plane with  $\xi = 0.8$  and the screening length  $\lambda = 0.4$  for various values of the screening strength,  $\epsilon$ .  $\phi_p$  and  $\phi_c$  are the volume fractions of PEG and the model particles, respectively.



**Figure 2.2:** Same as Fig. 1, except that  $\xi = 0.1$  and  $\lambda = 0.4$  for various values of  $\epsilon$ .

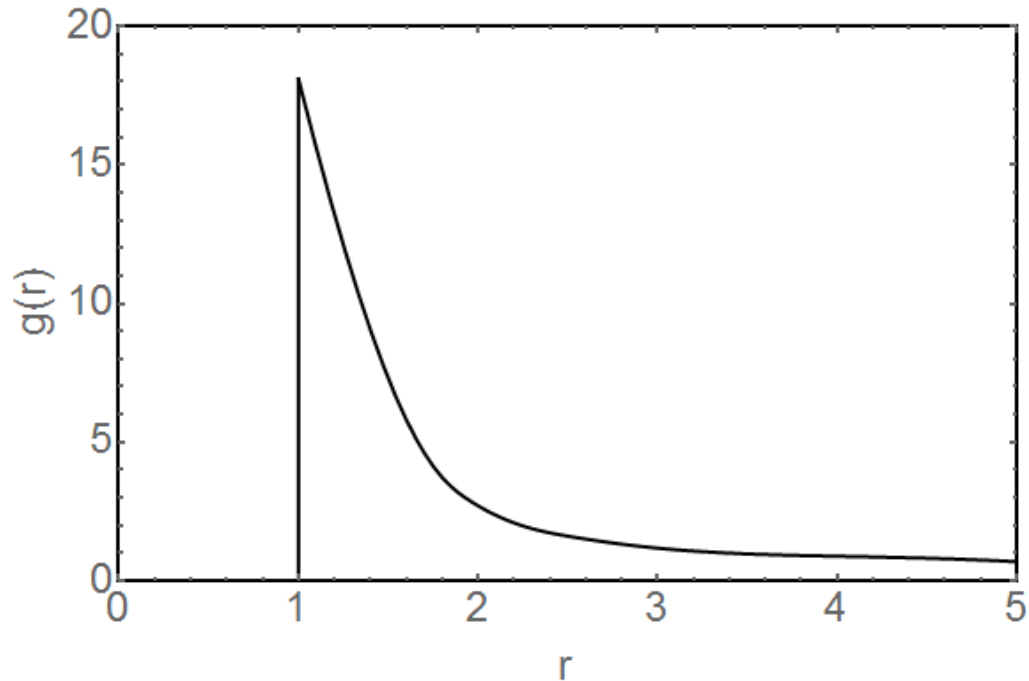


**Figure 2.3:** Same as Fig. 1, except that  $\xi = 0.8$  and  $\epsilon = 0.1$  for various  $\lambda$

with a small peak in  $g(r)$  near  $r = 2$ . The radial distribution function for the solid phase shows clear crystalline peaks at values indicative of a close-packed FCC lattice:  $r = 1, \sqrt{2}, \sqrt{3}, 2$ , etc. To verify the nature of the crystal structure, we also calculated the bond-orientational  $q_6$  order parameter.<sup>8</sup> We found average values of  $q_6 = 0.5745$ , consistent with a FCC structure. These findings are consistent across varying values of the parameters  $\epsilon$  and  $\lambda$ .

While the liquid-liquid coexistence curve shows a dramatic influence from the Coulomb interaction, the solidus line sees little effect at high  $\phi_p$  and short  $\xi$ . For small  $\lambda$  and  $\epsilon$ , the charge interaction represents a negligible fraction of the system energy for the solid or low-density liquid phases. Thermal fluctuations easily destroy any orientational ordering in the solid phase. At much larger values of these parameters, point charges align to minimize the repulsive interaction more readily. This transition can be observed in the probability distribution of orientational correlation



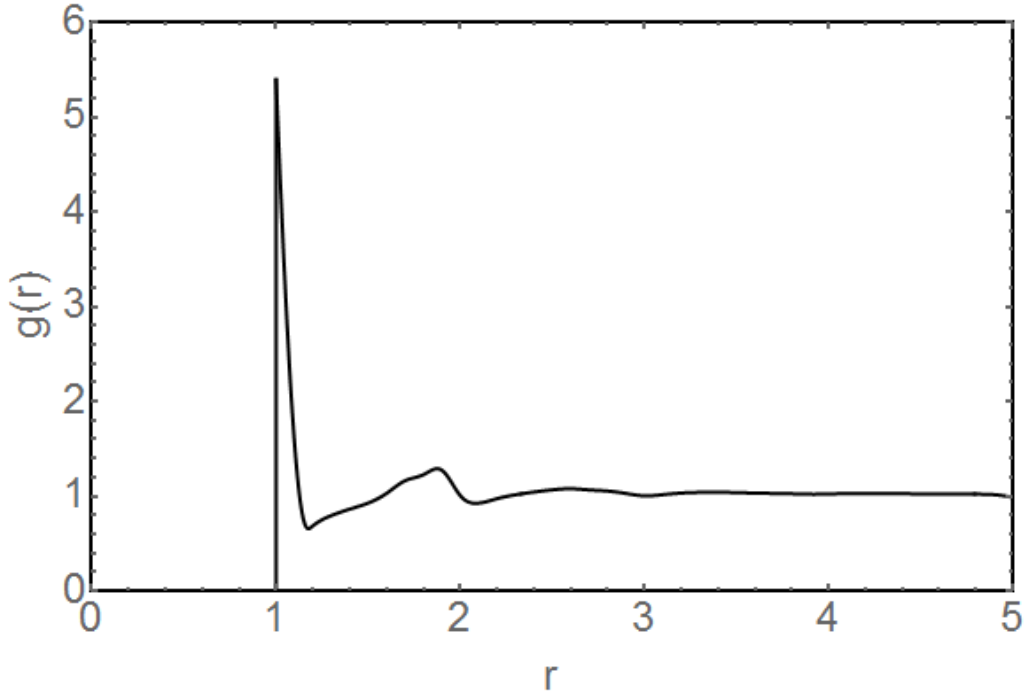


**Figure 2.4:** Radial distribution function,  $g(r)$ , for state point in the low density liquid phase along  $\phi_p = 0.4$

shown in figure 2.7. For  $\epsilon = 5.0$ , the orientation of point charges is uniformly distributed through the system. As the strength of the charge is increased, we observe increased ordering across the solid. Representative snapshots of the solid phase for these values of  $\epsilon$  are shown in figure 2.8.

## 2.5 Conclusion

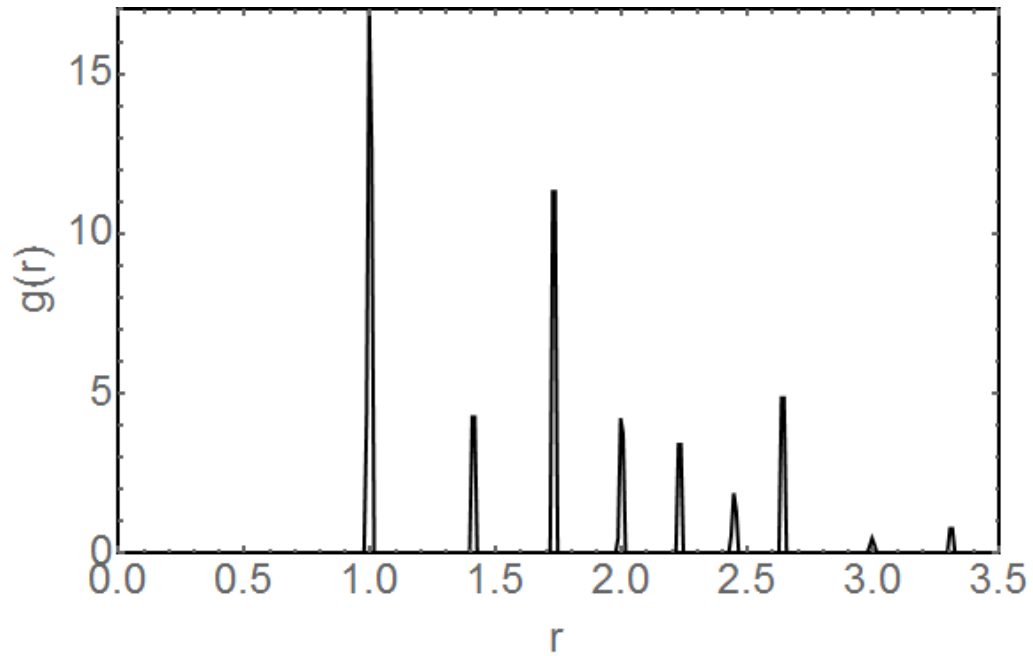
The phase diagram for our model bipolar molecule has been determined via Monte Carlo simulation with parallel tempering. Our results for the  $\epsilon = 0$  (no Coulomb interaction) case are in good agreement with a past study of the phase diagram of the



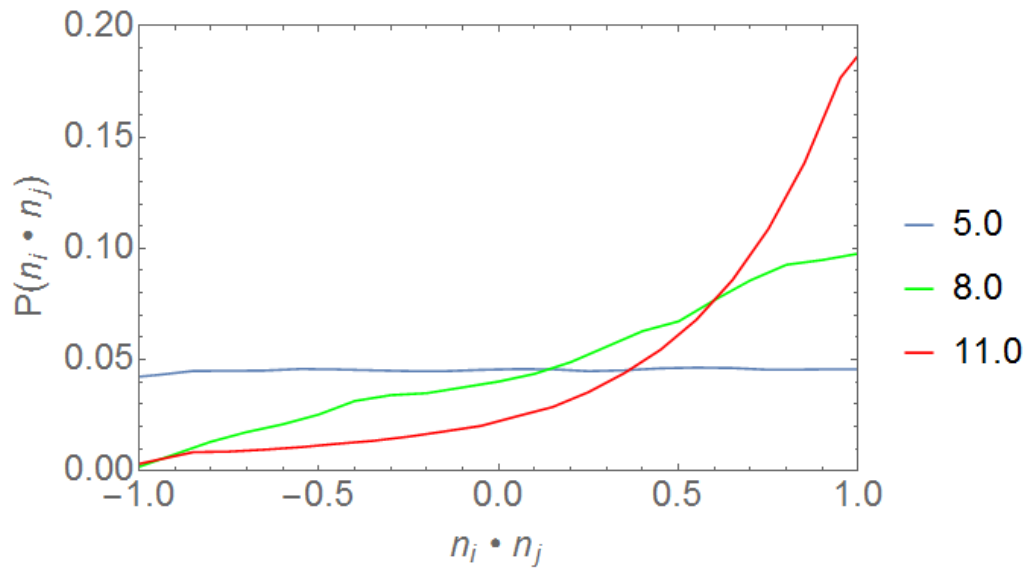
**Figure 2.5:** Radial distribution function,  $g(r)$ , for state point in the high density liquid phase along  $\phi_p = 0.4$

AO potential.<sup>35</sup> We find that the addition of a Coulomb charge to a particle interacting with an Asakura-Oosawa potential strongly affects liquid-liquid coexistence. The repulsive interaction reduces the density of the high-density phase and shifts the coexistence region to higher values of  $\phi_p$  (volume fraction of PEG). The same qualitative trend occurs in a stable liquid-liquid coexistence ( $\xi = 0.8$ ) and a metastable region ( $\xi = 0.1$ ). In the low-density liquid and solid phases, the Coulomb repulsion represents a negligible portion of the overall interaction, and therefore results in little effect.

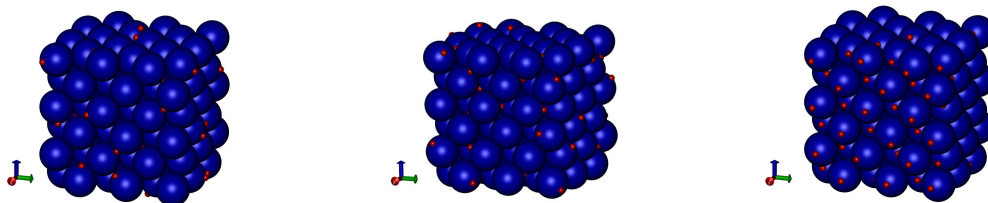
For the parameter case specific to amelogenin, we find that the previous dynamics study and typical experimental conditions place the system near the low-density edge



**Figure 2.6:** Radial distribution function,  $g(r)$ , for state point in the solid phase along  $\phi_p = 0.4$



**Figure 2.7:** Probability distribution of the orientational correlation for various values of  $\epsilon$  with  $\xi = 0.1$  and  $\lambda = 2.5$ .



**Figure 2.8:** Representative snapshots for  $\epsilon = 5.0$  (left),  $\epsilon = 8.0$  (center), and  $\epsilon = 11.0$  (right) after equilibration. Model protein particles shown as large blue spheres with the point charge represented by a small red sphere on the surface.

of the metastable region. In this area of high  $\phi_p$  and low  $\phi_c$ , the phase behavior is largely insensitive to changes in the Coulomb interaction.

An important topic for future research is to study the dynamics and morphology of self-assembly of amelogenin *in vivo*. The aforementioned experiment on amelogenin self-assembly<sup>37</sup> was carried out *in vitro*, and it is known that the process of self-assembly *in vivo* differs from that in *in vitro*.<sup>34</sup> At the very least one would have to take into account the effects of crowding due to the confinement of the molecules in the cell.

More generally, cryoelectron microscopy studies of amelogenin aggregation have shown that this system undergoes stepwise, hierarchical self-assembly.<sup>43</sup> Moreover, it has been demonstrated that interactions between hydrophilic, C-terminus tails are critical for the formation of, for example, oligomers. It would clearly be desirable, then, to model the temporal dependence of the aggregation process to characterize the morphologies of the aggregates and their distribution of sizes. In particular, at late times in the aggregation process, it has been found that large nanospheroids predominate; however, the particle morphology (i.e., spherical versus oblate) is still an open question. It is expected that simulation will be useful in elucidating stable

particle morphologies and in highlighting their internal structure.

## **Acknowledgements**

This work was supported by a grant from the G. Harold and Leila Y. Mathers Foundation. One of us (E. P.) was supported by NSF grants PHY-0849416 and PHY-1359195.

# Chapter 3

## Immunoglobulin: Effect of Angle Between Interaction Sites on Kinetics and Morphology

### 3.1 Abstract

We study the kinetics of aggregation of a two bead model of interacting spherical molecules. A given bead on one molecule can interact with one or more beads on other neighboring molecules. The beads represent the result of a simple coarse graining of a putative amino acid residue (or residues). We study the kinetics and equilibrium morphology for a fixed angle between the two beads, and then study the dependence of this process and morphology on this angle, which we study for several angles between 30 and 150 degrees. In the model the beads interact via an attractive Asakura-Oosawa potential and the molecules have the usual hard sphere

repulsion interaction. We find a transition from a micelle like morphology at small angles to a rod-like morphology at intermediate angles, to a gel like structures at values of the angle greater than about ninety degrees. However, at 150 degrees, after a long induction time during which there is no aggregation, we observe a nucleation and growth process that leads to a final spherical-like aggregate.

## 3.2 Introduction

Immunoglobulin (Ig, better known as an antibody) is a large, Y-shaped protein produced by the immune system to identify pathogens for removal. Ig binds to specific surface motifs (antigens) on the foreign body and then subsequently to receptors on specialized cells to kill or remove the bacteria or virus. The protein is comprised of four peptide chains (two heavy, two light) bound together via disulfide bonds. There are five classes of human Ig, determined by the composition of the heavy chains, of which IgG is the most common. Each chain contains a constant region, conserved among antibodies of the same class, and a variable region that defines its antigen specificity.

IgG is comprised of three structural domains: two Fab (Fragment antigen binding) and an Fc (Fragment crystallizable, which binds to receptors on immune cells), joined by a flexible hinge. These antibodies are typically monomeric in the body. High concentration IgG solutions are used in the treatment of various illnesses, typically autoimmune or chronic inflammatory diseases. Such solutions are harvested from blood plasma of thousands of individual donors. Unfortunately the likelihood of an Fab region from one donor targeting that of another donor increases with

number of donors used.<sup>107,121</sup> This causes an increase in dimerization, and a corresponding decrease in effectiveness of the treatment. At high concentrations, IgG solutions show signs of gelation and even the formation of insoluble aggregates, which can provoke a harmful immune response.<sup>25,30,57,79</sup>

In addition to the Fab-Fab interactions, the Fc regions and the hinge itself have been shown to effect aggregation.<sup>95</sup> Exposed hydrophobic residues are more concentrated in these regions than in the Fab domains. These areas have been experimentally mutated to be more hydrophilic, resulting in less aggregation prone IgG variants.<sup>23,24</sup>

In this chapter we study a greatly simplified model of Ig G, consisting of spherical molecules with two interacting sites (“beads”) on each spherical molecule. A given bead is meant to represent the result of a simple coarse graining of amino acid residues. It can interact with other beads on neighboring molecules. In addition to this bead-bead interaction, the molecules have the usual hard-core repulsive interaction. We study the situation in which we vary the hinge angle, the angle between the two beads subtended by lines drawn from each site to the center of the molecule. Our motivation in part is the case of immunoglobulin , which are flexible, so that the “hinge” angle between the two arms of the molecule is variable.

In a previous model defined by Li et al,<sup>73</sup> the authors studied of polymer-induced phase separation and crystallization in immunoglobulin G, but kept the hinge region fixed. Their model yielded results that are similar to the experimental results obtained subsequently by Wang et al.<sup>131</sup> However, our model is not meant to be a realistic model of such a protein, but rather a simple way to simulate qualitatively the effect of varying the angle on the kinetics and morphology of the aggregation



of the molecules. In our model the site-site interactions are taken to be of the Asakura-Oosawa depletion form, which provides an attractive interaction between the molecules. Our main result is that for case I the hinge angle plays a significant role in the kinetics and morphology of aggregation. For small hinge angles of thirty to sixty degrees, the aggregates are more micelle like (although becoming anisotropic as the angle is increased). Then there is a transition window between sixty-five and seventy degrees in which the morphology becomes more rod-like. As one increases the angle even more, the aggregates form a more gel-like structure. Finally at one hundred and fifty degrees we see no aggregation until after a long induction time we observe a nucleation and growth aggregation process. One thing to note is that each value of the hinge angle has its own particular phase diagram, so that in the case of 150 degrees we have apparently quenched the system into a metastable region of the diagram.

### 3.3 Model and Methodology

We assume in our models that there are two interaction sites per spherical molecule, whose angular separation is varied in the study. There is the usual hard-core repulsion between the spheres ( $U_{hc}$ ). There is also an additional the two-body Asakura-Oosawa<sup>5</sup> attractive depletion potential ( $U_{AO}$ ) between the interaction sites on the particles. Hence the overall interaction potential is  $U = U_{AO-sites} + U_{hc-spheres}$ . The Asakura-Oosawa depletion potential is given by

$$U_{AO} = \phi_p((1 + \delta)/\delta)^3 [3r/(2(1 + \delta)) - 1/2(r/(1 + \delta))^3 - 1]kT, \quad r < 1 + \delta, \quad (3.1)$$

and the hard core repulsion is taken to be

$$U_{hc} = r^{-\alpha}kT, \quad (3.2)$$

with  $\alpha = 36$ . For the depletion potential the  $\delta$  value is set at 0.5 so that the range of interaction is  $1.5\sigma$ , where  $\sigma$  is the diameter of the sphere. The value of  $\phi_p$  is selected such that maximum depth of interaction is set at  $6kT$ .

The interactions between the sites exert torques on the molecules and hence produce a rotational motion of the molecules that is included in our Brownian equations of motion. These are given as

$$m\ddot{\vec{r}}_i = -\vec{\nabla}(U_i^c + U_i^p) - \Gamma_t\dot{\vec{r}}_i + \vec{W}_i(t), \quad (3.3)$$

$$I\ddot{\vec{\omega}}_i = \vec{\tau}_i - \Gamma_r\vec{\omega}_i + \vec{W}'_i(t), \quad (3.4)$$

where  $m$ ,  $I$ ,  $\vec{r}_i$ ,  $\vec{\omega}_i$ ,  $\vec{\tau}_i$  are the mass, moment of inertia, position vector, angular velocity, and torque, respectively, of the  $i^{th}$  molecule. The masses of the point sites are ignored in this model.  $\Gamma_t = 6\pi\eta r$  ( $\Gamma_r = 8\pi\eta r^3$ ) is the translational (rotational) friction coefficient, where  $r$  is the radius of the monomer (molecule) and  $\eta$  is the dynamic viscosity.  $\vec{W}_i$  and  $\vec{W}'_i$  are the random forces and torques acting on the  $i^{th}$  particle respectively, which satisfy fluctuation-dissipation relations  $\langle \vec{W}_i(t) \cdot \vec{W}_j(t') \rangle = 6kT\Gamma_t\delta_{ij}\delta(t-t')$  and  $\langle \vec{W}'_i(t) \cdot \vec{W}'_j(t') \rangle = 6kT\Gamma_r\delta_{ij}\delta(t-t')$ .<sup>125</sup>

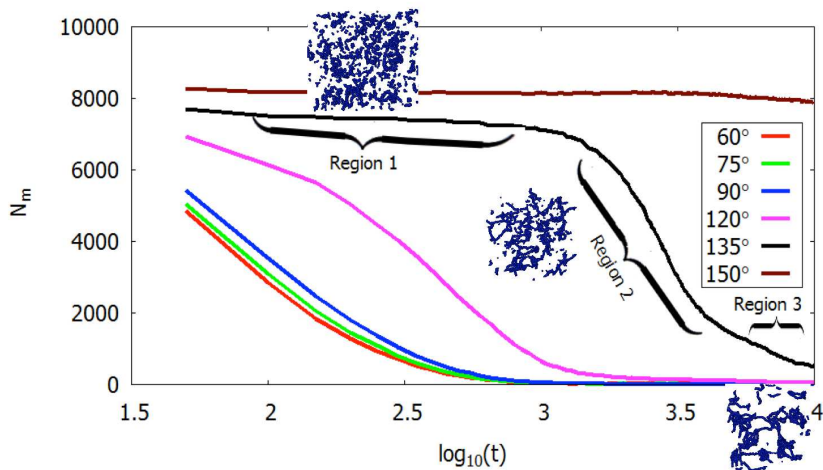
In our Brownian dynamics simulations, all length scales are measured in units of

the monomer diameter  $\sigma$ . We choose  $\Gamma_t = 0.5$ ,  $\Gamma_r = 0.167$ , and the time step  $\Delta t = 0.005$  in reduced time units of  $\sigma(m/kT)^{1/2}$  with  $m = 1$ . For this choice of  $\Gamma_t = 0.5$ , particle motion is purely diffusive for  $t \gg 1/\Gamma_t$ , i.e.,  $t \gg 2$  in our units. In order to study the anisotropic effect brought by the sites on the self-assembly process, we focus on a relatively dilute system, as described in the section below. Periodic boundary conditions are enforced to minimize wall effects. All simulations start from a random initial monomer configuration and the results for the kinetics are averaged over several (5-10) runs.

## 3.4 Results

### 3.4.1 Rate of aggregation and size of aggregates

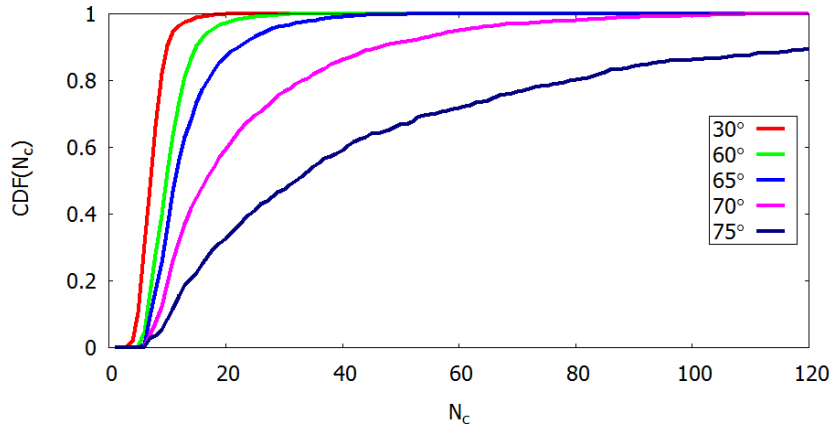
We studied the aggregation of the particles at two different volume fractions of  $f = 0.0025$  and  $f = 0.01$ . The rate of monomer loss is studied at the lower volume fraction of  $f = 0.0025$ . The comparison of monomer loss is made at six different angles between the interaction sites on the spherical particles. Our results show that the rate of aggregation slows down as the angle between the interaction sites is increased, as shown in fig.3.1. The monomer loss for the case of  $\theta = 120$  and  $\theta = 135$  degrees can be divided into three regions, as shown in fig.3.1. In the first region, the rate of monomer loss is slow due to the large degree of dissociation of the aggregates. In region 2, the dissociation of aggregates is negligible and the clusters grow with the addition of monomers. In region 3, the cluster size growth is dominated by the cluster-cluster aggregation and hence there is a slow addition of monomers to the aggregates. Typical configurations for these three regions are shown in fig.3.1.



**Figure 3.1:** Rate of monomer loss for different angles, for the volume fraction  $f = 0.0025$ . As discussed in the text, there are three different regions for the cases of  $\theta = 120$  and  $135$  degrees. Typical configurations for these three regions are also shown.

At this concentration, no aggregation takes place for the angular separation of 150 degrees (i.e.  $\theta = 150$ ) on the time scale shown in this diagram. However, in three other long runs (out of a total of ten), we observed a long induction time followed by nucleation and growth into a final aggregate. We discuss this in more detail in our discussion of the morphology of the aggregates.

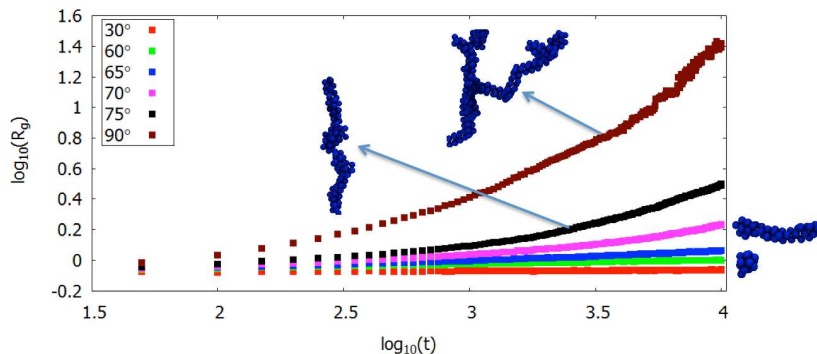
Varying the angle between the interaction sites also affects the size of the aggregates. We studied the dependence of the size of aggregates on the angle by two methods, first by calculating the cumulative distribution function of the aggregate size and second by studying the radius of gyration of the aggregates. In the first method we compared the equilibrium size of the aggregates for five different values of  $\theta$ , as



**Figure 3.2:** Cumulative distribution function for different angles. The equilibrium size of the aggregates increases as the angle between the patches is increased.

shown in fig.3.2. It can be seen that as the angle between the patches is increased, the equilibrium size of the aggregates also increases.

We next analyzed the time dependence of the radius of gyration, which gives us both a measure of the size as well as of the dynamics of aggregation for different angles between interaction sites. Fig.3.3, shows the log-log plot of the average radius of gyration  $R_g$  of the clusters as a function of time for  $f = 0.01$ . It can be seen that for angles below 60 degrees, there is a very slow variation in the size of the aggregates, which indicates the formation of approximately spherical, compact aggregates. As the  $\theta$ -value is increased further, there is a significant variation in the radius of gyration of the aggregates, indicating the formation of non-spherical larger size aggregates. It can be seen that for  $\theta = 75$  degrees, the growth kinetic exponent compares well with the Diffusion Limited Cluster Growth (DLCA) kinetic

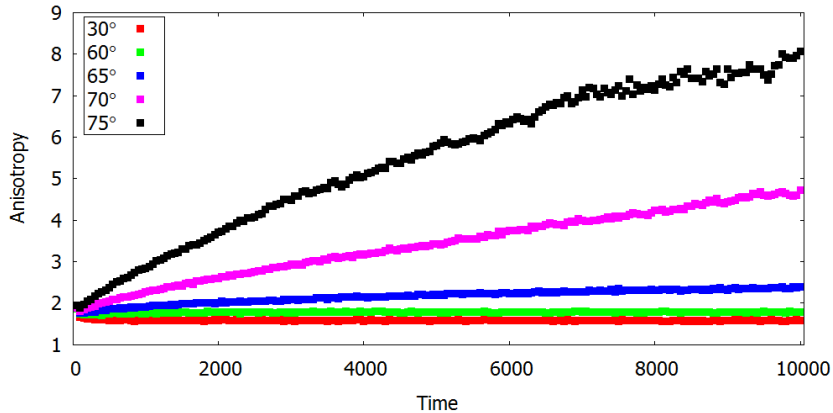


**Figure 3.3:** Log-log plot of the radius of gyration,  $R_g$ . The aggregates form approximately spherical, compact aggregates below sixty degrees. As one increases the angle further, the significant variation in  $R_g$  shown here indicates the formation of larger, non-spherical aggregates. Typical final aggregates for the angle  $\theta$  equal to 60, 65, 75 and 90 degrees, respectively. For  $\theta = 75$  degrees, the growth kinetic exponent shown here is close to the DLCA kinetic exponent of 0.55, indicating that the aggregates are fractals, within the accuracy of our simulation.

exponent of 0.55 indicating that the aggregates are fractal-like for  $\theta = 75$ . Fig.3.3 also shows typical final aggregates for the angle  $\theta$  equal to 60, 65, 75 and 90 degrees, respectively.

### 3.4.2 Morphology

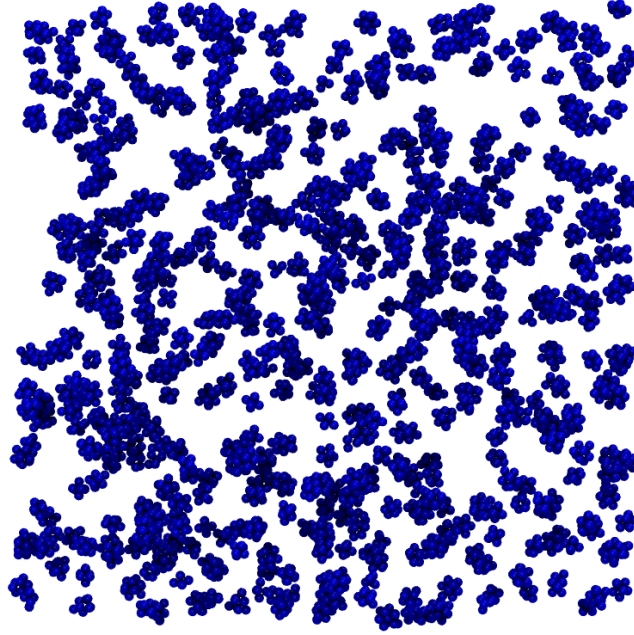
We now focus on the shape and morphology of the aggregates. First, we compare and characterize the shape of the aggregates based on the cluster shape anisotropy parameter. The cluster shape parameter ( $A_{13}$ ) is defined as the ratio of the squares of the largest and smallest principle radii of gyration. From the cluster shape anisotropy calculations, shown in fig.3.4, we can conclude that for the case of  $\theta = 30$  and 60 degrees, the clusters have an almost identical morphology and that this morphology stays the same throughout the simulation time. A transition to a more



**Figure 3.4:** Plot of the anisotropy vs time for several different angles. The anisotropy is defined as the ratio of the square of the largest and smallest principle radii of gyration. From this plot we conclude that the clusters have almost identical morphologies for  $\theta = 30$  and  $60$  degrees and remains essentially constant through the simulation. The large values of the anisotropy parameter shown for the case of  $\theta = 75$  degrees indicates that the clusters are more irregular in their shape.

rod-like morphology occurs around  $\theta = 65$  degrees. For  $\theta = 75$  degrees, the large values of the anisotropy parameter indicate that more irregular shaped clusters are formed. For  $\theta = 90$  degrees and larger, the aggregates are gel-like, with an interconnected structure. The typical equilibrium morphologies are shown in Figure 3.5 through Figure 3.11. (The morphology for  $\theta = 30$  degrees is of a similar micelle like shape as in fig.3.6.) The transition to rod-like behavior is evident in Figure 3.7, while the gel-like behavior is shown in the figures with  $\theta$  greater than or equal to 90 degrees ending at about 140 degrees. Beyond  $\theta = 140$  degrees, this extended network collapses into distinct, compact clusters.

The case of  $\theta = 150$  degrees deserves special attention, because on the time scale shown in figure 1, there is no aggregation. However, in three of ten long runs we observed nucleation and growth events (at times longer than shown in figure 1).



**Figure 3.5:** Morphology for  $\theta = 30$  degrees at time  $t=10,000$ .

Figures 3.12, 3.13 and 3.14 shows one such event. We also ran for very long times at 180 degrees but were unable to see any aggregation, including any late stage nucleation event.

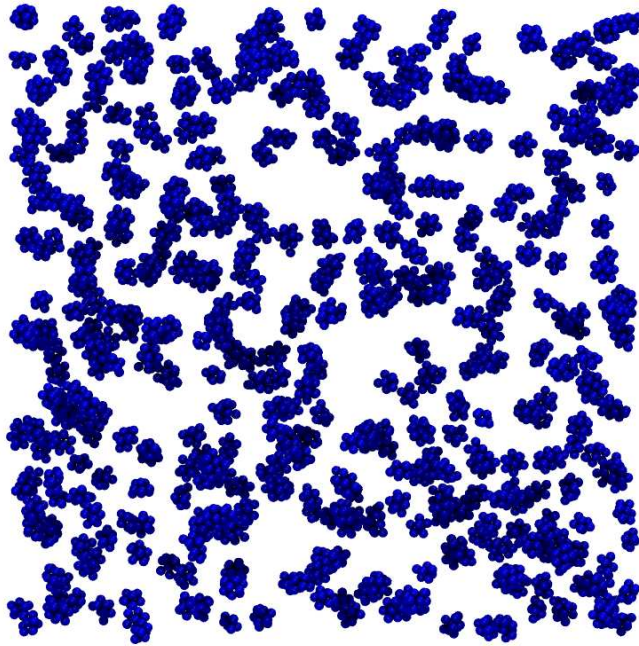
In order to determine the nature of the nucleated droplet, we performed a statistical analysis of the quantity  $q_6$  for the cluster.<sup>7</sup> We first define

$$q_{lm}(i) = \frac{1}{N_b(i)} \sum Y_{lm}(\hat{r}_{ij}) \quad (3.5)$$

where the  $Y_{lm}(\hat{r}_{ij})$  are the spherical harmonics evaluated for the normalized direction vector  $\hat{r}_{ij}$  between the neighbors  $i$  and  $j$ .  $q_6$  is defined as

$$q_6(i) = \frac{4\pi}{2l+1} \sum_{m=-l}^{m=l} |q_{6m}(i)|^2)^{1/2} \quad (3.6)$$





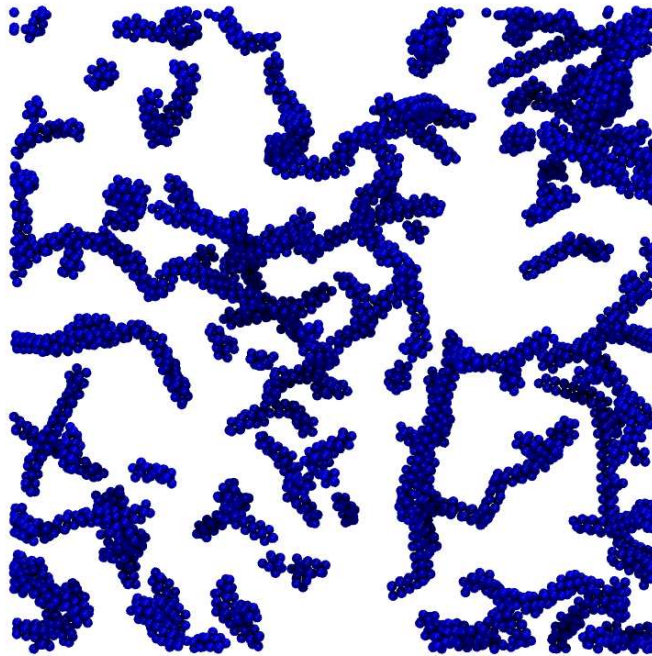
**Figure 3.6:** Morphology for  $\theta = 60$  degrees at time  $t=10,000$ .

We found a rather broad distribution function for  $q_6$ , which had its maximum around  $q = 0.45$  with a width from 0.2 to 0.7. To obtain a more definitive test of the nature of the droplet, we calculated the correlation function  $\vec{q}_6(i) \cdot \vec{q}_6(j)$ , which is defined as

$$\vec{q}_6(i) \cdot \vec{q}_6(j) = \sum_{m=-6}^{m=6} q_{6m}(i) q_{6m}^*(j) \quad (3.7)$$

where the asterisk denotes complex conjugate. This correlation function is shown in fig.3.15. The fact that it is strongly peaked in the neighborhood of zero is strong evidence that the nucleated droplet is liquid-like.<sup>7</sup>

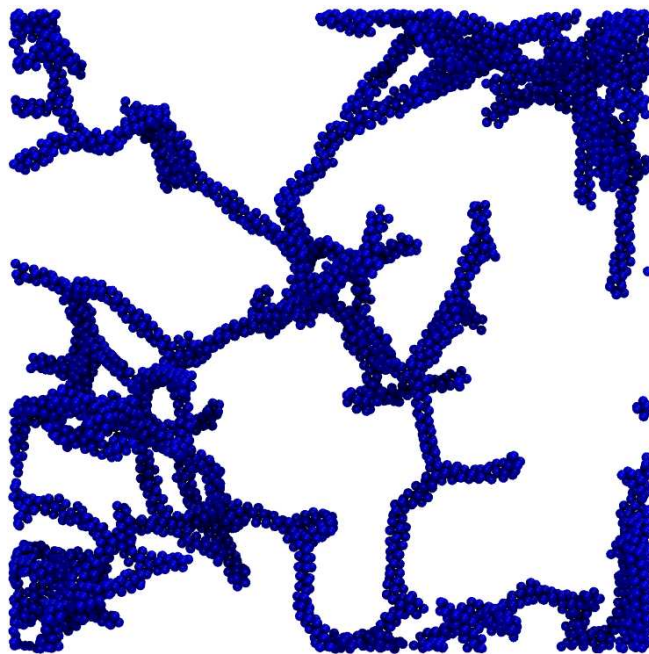
We further probed the structure of the aggregates by studying the radial distribution function of the aggregates for different  $\theta$  values, The distribution of the



**Figure 3.7:** Morphology for  $\theta = 75$  degrees, at  $t=10,000$ . The transition to the rod-like behavior shown in this figure occurs around  $\theta = 65$  degrees .

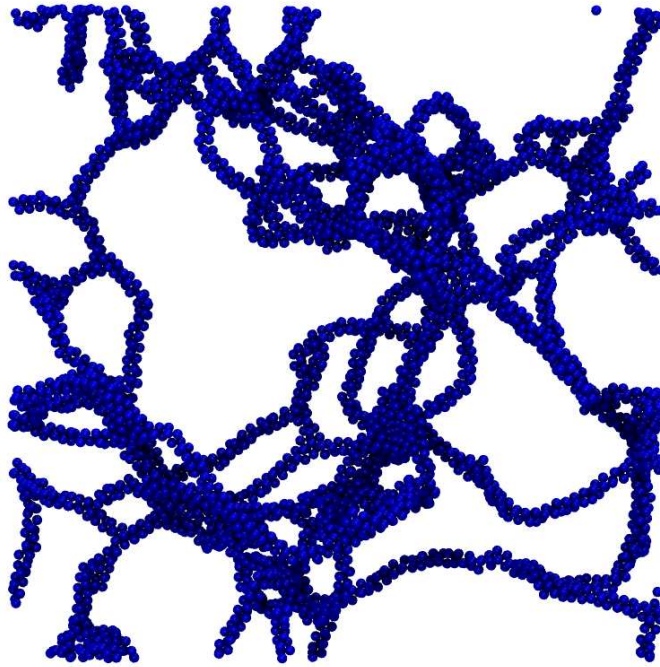
nearest neighbors appear to be very similar for the different  $\theta$  values, as indicated by the distribution shown in fig.3.16. For the next neighbors (indicated by the 2nd and 3rd maxima values), a variation in the structure can be observed. There is very little difference between the radial distribution shown at  $\theta = 120$  degrees and those for 135 and 150 degrees, so we do not display these latter two in this figure. A similar remark holds for the structure function, which we discuss below.

Finally, we examine the structure factor,  $S(q,t)$  where  $q$  is the wavenumber;  $S(q,t)$  is the Fourier transform of the pair correlation function. In fig.3.17 and fig.3.18, we show the log-log plot of  $S(q,t)$  versus  $q$  at early and late times, respectively. A fit to the data shows some interesting results. Fig.3.17 shows the structure factor results



**Figure 3.8:** Morphology for  $\theta = 90$  degrees, at  $t=10,000$ , shows the beginning of an interconnected gel-like structure.

at early times for various hinge angles. At early times there seems to be no structure at large length scales whereas at short length scales, the aggregates are less compact (the slope is approximately minus 3) and have very similar structure for all angular variations. The slope of minus 3 indicates that the aggregates are semi-compact. For spherical aggregates, a slope of minus 4 is required (Porod's law). The structure factor calculations of the late time aggregates for various hinge angles are shown in fig.3.18. For  $\theta = 75$  and 90 degrees, inspection of the results at different length scales appear to show two different structures of the aggregates. The late time analysis shows that at small length scales, the aggregates seem to be compact whereas at intermediate and large length scales, the aggregates appear to be more fractal-like i.e.  $S(q) \sim q^{-1.7}$ , approaching to the DLCA value of 1.8. For small angle variations

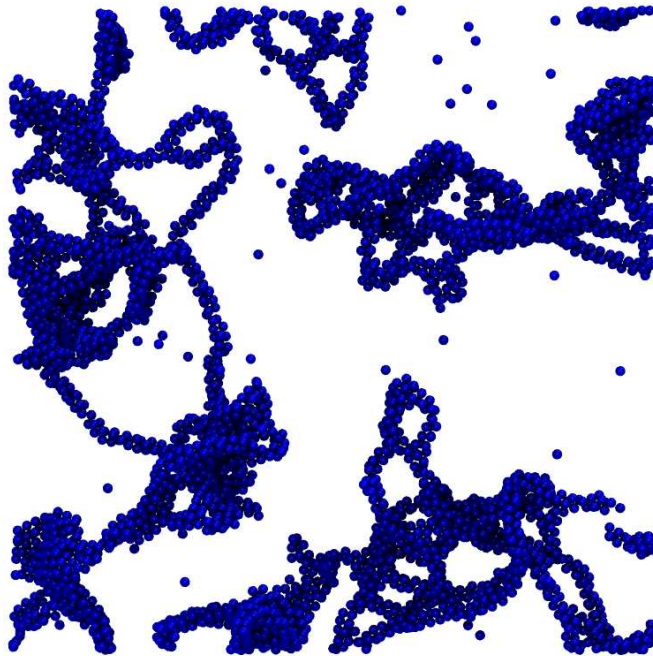


**Figure 3.9:** Morphology for  $\theta = 120$  degrees, at  $t=10,000$ .

(i.e.  $\theta = 30$  and  $60$  degrees), the aggregates are semi-compact. The results obtained from analyzing the behavior of the structure factor are complementary to the results we have obtained by other methods in this section.

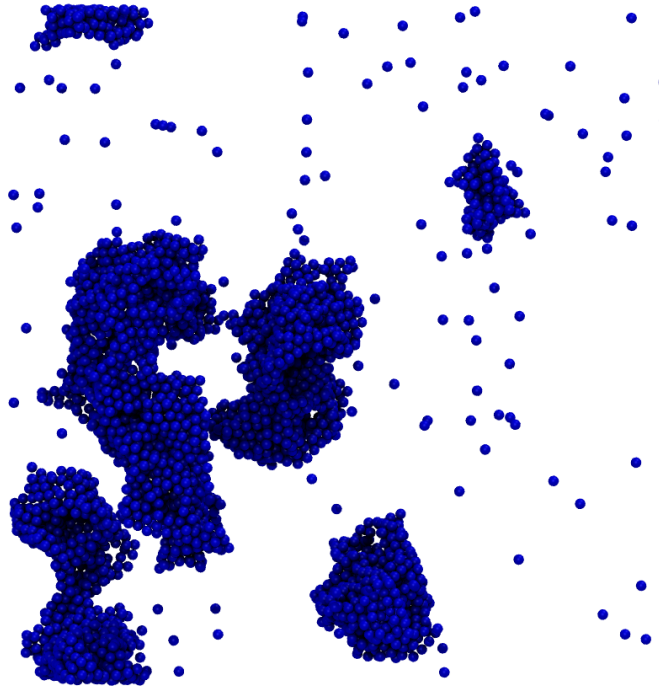
### 3.5 Conclusion

In this chapter we have examined a particular type of anisotropic interaction between molecules, resulting from two sites being located asymmetrically on the sphere representing the molecule. We studied in considerable detail the kinetics of aggregation and the dependence of the morphology of the resulting aggregate on the angular

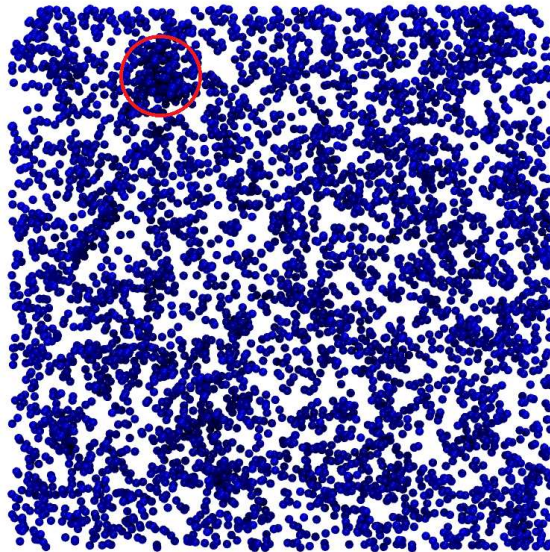


**Figure 3.10:** Morphology for  $\theta = 135$  degrees.

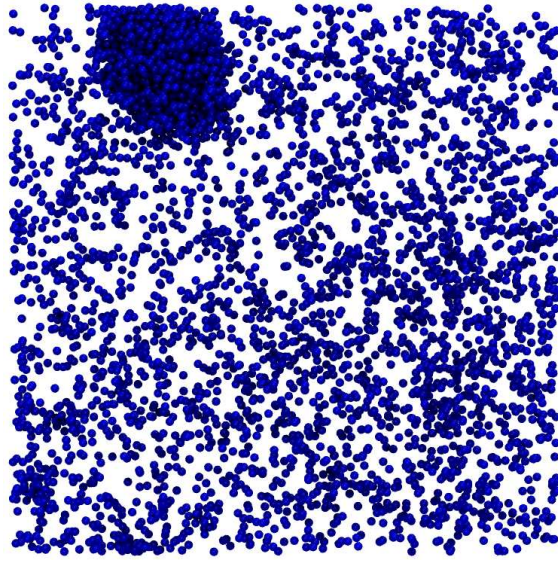
separation (“hinge angle”) between the two sites. We have found an interesting dependence on the hinge angle on both the morphology and kinetics of the aggregation. Although the rate of aggregation is different for different hinge angles, the initial morphology is found to be very similar in all cases (i.e. a quasi-spherical, micelle-like structure). As time progresses, we found that there was a strong dependence of the morphology on the hinge angle and observed a transition from a micelle-like aggregation for small angles to a more rod-like aggregations that occurs around  $\theta = 65$  degrees. As one increases the angle even further, one eventually finds a gel-like aggregate. In one particular case ( $\theta = 75$  degrees), we found that the growth kinetic exponent is close to the DLCA kinetic exponent of 0.55, indicating that the aggregates are fractals, within the accuracy of our simulation. We did not find this to be



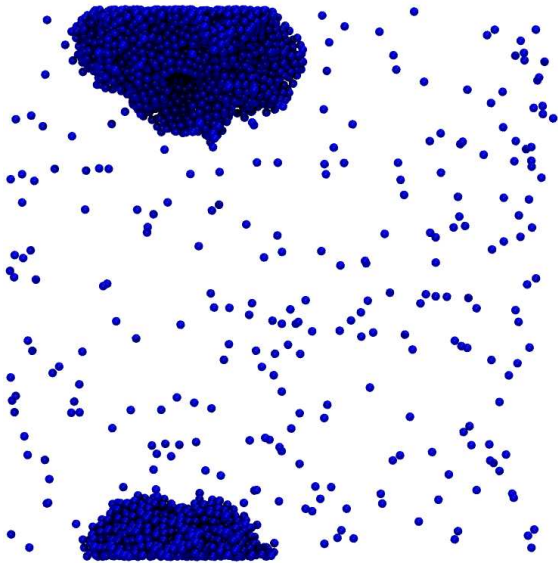
**Figure 3.11:** Morphology for  $\theta = 145$  degrees at time  $t=10,000$ , which shows a significantly more compact structure.



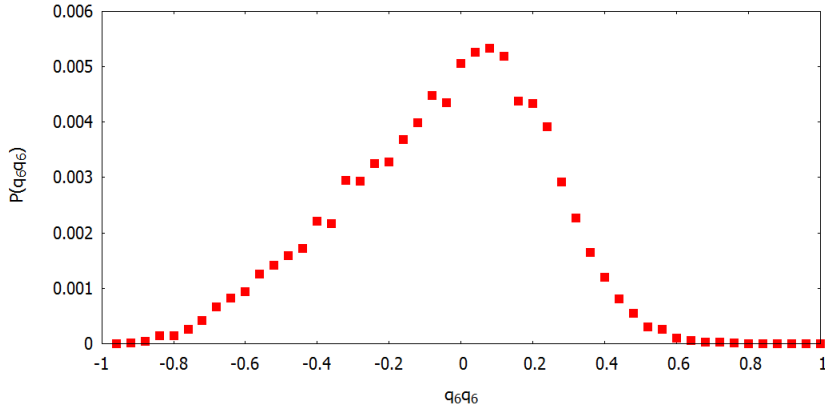
**Figure 3.12:** Nucleation occurring after a long induction time  $t=13,100$ , for  $\theta = 150$  degrees.



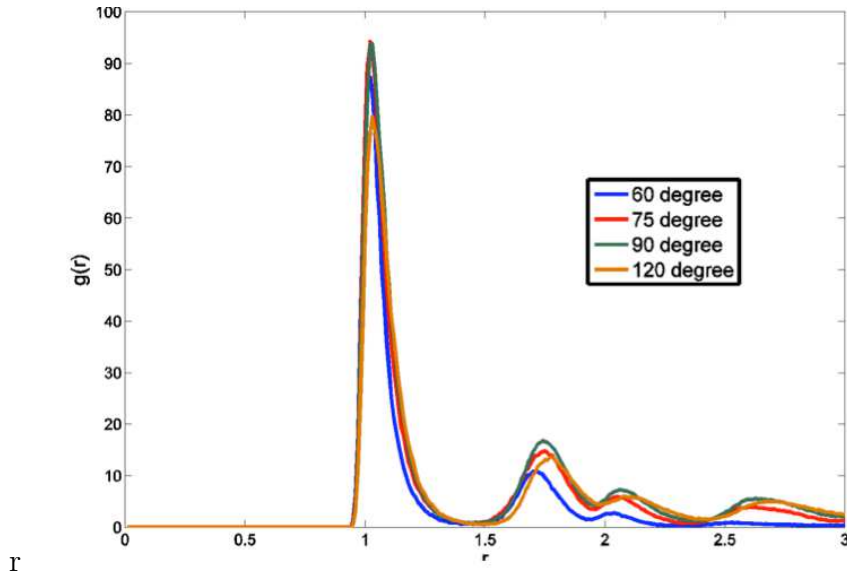
**Figure 3.13:** Closer to equilibrium, at  $t=16,250$ .



**Figure 3.14:** Equilibrium aggregate, at  $t=20,000$ .



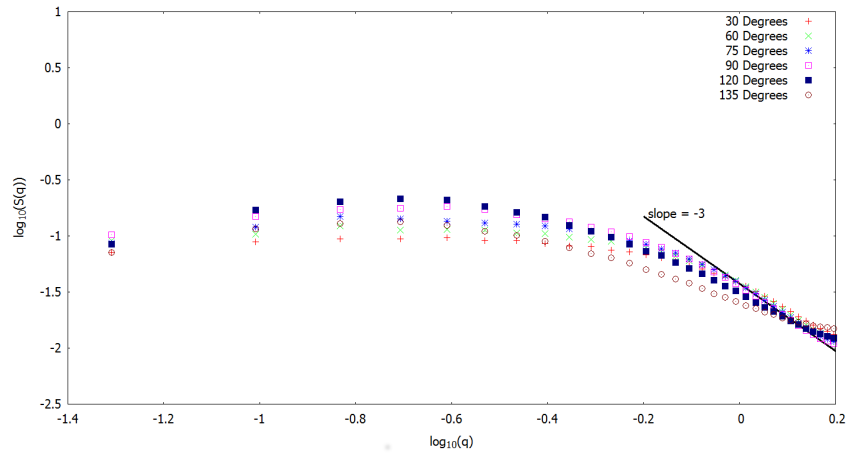
**Figure 3.15:** The correlation function  $\vec{q}_6(i) \cdot \vec{q}_6(j)$ , defined in the text.



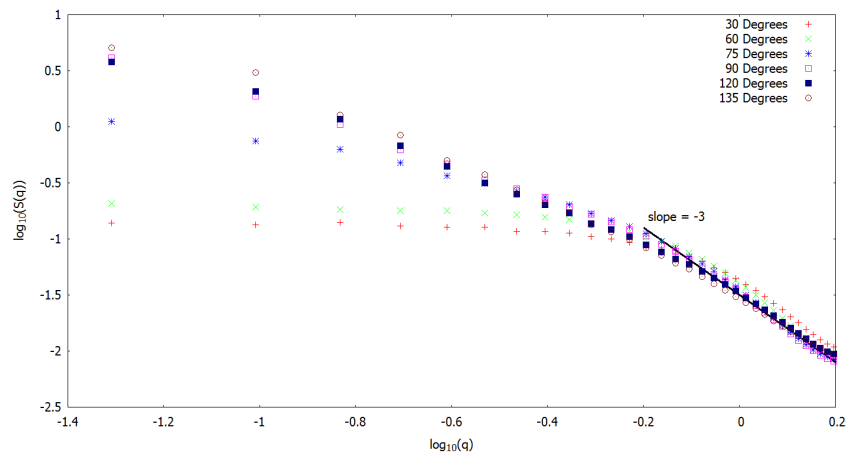
**Figure 3.16:** The radial distribution function,  $g(r)$ , as a function of spatial separation  $r$ .

the case for the other angles studied. In addition, we characterized the nature of the nucleating droplet that we observed in three long runs for the angular separation of 135 degrees by calculating the quantities  $q_6$  and the correlation function  $\vec{q}_6(i) \cdot \vec{q}_6(j)$ . In this latter case we provided evidence that the nucleating droplet is liquid-like





**Figure 3.17:** log-log plot of the structure factor at early times,  $t = 100$ . There appear to be no structure at large length scales. At short length scales, the aggregates are less compact and exhibit similar structures for all the angular variations.



**Figure 3.18:** log-log plot of the structure factor at late time,  $t = 10,000$ . As discussed in the text, the aggregates at 75 and 90 degree angles appear to be more fractal like at large length scales, with the value of  $S(q) \sim q^{-1.7}$ , which is close to the DLCA value of 1.8.

even at the stable equilibrium state. This observation is in contrast to the case of isotropic interactions, where the initial nucleating liquid-like structure leads to the formation of more order structure at large, stable sizes.

There are several simulations that one could do next, including starting this model at lower temperatures and at higher concentrations. In addition, it would be of interest to calculate the phase diagram for this model, at least for a few characteristic angles. Finally, it is possible that our results might be relevant to extending current studies of immunoglobulin, for example. In that case it is known that for IgG there is considerable variation in the flexible hinge angle of these Y shaped molecules. One could perhaps study this effect this by modifying the model defined by Li et al<sup>73</sup> in their study of polymer-induced phase separation and crystallization in immunoglobulin G to include a variable hinge angle. It would also be of interest to see whether this modification led to any significant changes in their predicted phase diagrams, in addition to extending their work to study the kinetics of aggregation of the modified model.

## Acknowledgements

This work was supported by a grant from the G. Harold and Leila Y. Mathers Foundation.

# Chapter 4

## Huntingtin: Effect of Repeat

## Length on Early Stage

## Aggregation

### 4.1 Abstract

In this chapter we study the early stages of aggregation of a model of polyglutamine (polyQ) for different repeat lengths (number of glutamine amino acid groups in the chain). We use the LAMMPS molecular dynamics simulator to study a generic coarse grained model proposed by Bereau and Deserno. We focus on the primary nucleation mechanism involved and find that our results for the initial self assembly process are consistent with the two dimensional classical nucleation theory of Kashchiev and Auer. In particular, we find that with decreasing supersaturation, the oligomer fibril (protofibril) transforms from a one-dimensional  $\beta$  sheet to two-

three- and higher layer  $\beta$  sheets, as the critical nucleus size increases. We also show that the results are consistent with several predictions of their theory, including the dependence of the critical nucleus size on the supersaturation. Our results for the time dependence of the mass aggregation are in reasonable agreement with an approximate analytical solution of the filament theory by Knowles and collaborators that corresponds to an additional secondary nucleation arising from filament fragmentation. Finally we study the dependence of the critical nucleus size on the repeat length of polyQ. We find that for the larger length polyglutamine chain that we study, the critical nucleus is a monomer, in agreement with experiment and in contrast to the case for the smaller chain, for which the smallest critical nucleus size is concentration dependent.

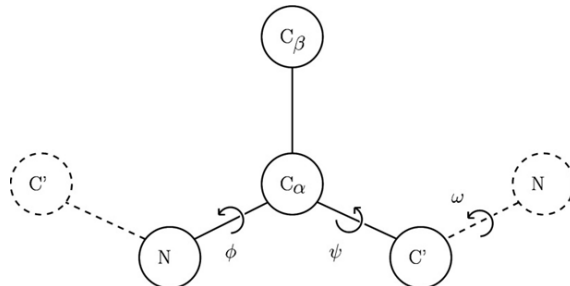
## 4.2 Introduction

It is widely accepted that a large number of proteins can change from normally functioning, soluble species into insoluble aggregates that can potentially lead to disease. The typical driving force in protein aggregation is believed to be a destabilization of the native protein that yields a group of partially folded intermediates which have a greater propensity for increased aggregation.<sup>50,106,110,116,134</sup> In particular polyglutamine (polyQ) aggregate formation has been implicated as playing an important role in a class of neurodegenerative diseases, such as Huntington's disease.<sup>11,15,22,42,119</sup> These are known as polyglutamine expansion diseases, since they all seem to be related to the aggregation of proteins whose only common sequence feature is the presence of long polyglutamine tracts, which result from the genetic

expansion of of a disease protein's polyQ sequence beyond repeat length thresholds in the range<sup>65</sup> of 35-50. Although these polyQ sequences can be of variable length, there exists a fairly sharp threshold length beyond which the disease emerges.<sup>65</sup> As the length of the polyQ increases past the threshold length for a given disease protein, the disease age-of-onset decreases and the disease severity at onset increases as well. These neurodegenerative diseases in humans are pathologically noteworthy for amyloid-like neuronal polyQ aggregates, Thus gaining an understanding of the biophysics of polyQ aggregation would seem of utmost importance.

It is generally accepted that amyloid fibrils (such as caused by polyQ) form by a nucleation and growth mechanism (see, e.g.,<sup>66,67,71,71,81,90</sup>).A reasonable amount is already known about polyQ aggregation, in particular, through experimental studies. Wetzel and collaborators have shown that relatively long polyQ peptides aggregate by nucleated growth polymerization, with a monomeric critical nucleus.<sup>15,22,42,119</sup> They subsequently showed that the size of the critical nucleus for aggregation increases from monomeric to dimeric to tetrameric<sup>65</sup> over a short repeat length range from Q26 to Q23 (for  $K_2Q_{26}$  to  $K_2Q_{23}K_2$ ). They also carried out detailed kinetics studies of the aggregation of these polyQ containing sequences and showed that the spontaneous aggregation of several disease-related, complex polyQ sequences is controlled by monomeric nucleus formation,<sup>65</sup> similar to simple polyQ peptides. This suggests that results from polyQ model sequences are relevant to at least some polyQ pathology. Most importantly, the modest differences in critical nucleus size found in these studies by Wetzel et al<sup>65</sup> project huge differences in the kinetics of aggregation predicted for different repeat lengths at physiological concentrations. A

more recent small-angle neutron scattering study<sup>101</sup> (also see<sup>120</sup>) showed that the aggregation pathway of  $NtQ_{22}P_{10}$  is very different from that of  $NtQ_{42}P_{10}$ . The initial steps of aggregation for the former peptide involve a monomer to 7-mer transition, whereas the latter involves a transition from a monomer to dimer (i.e. critical nucleus size of 7 and 1, respectively). Wetzel et al<sup>65</sup> also note that although the repeat length threshold found in their studies is 24-25, the thresholds for disease risk tend to be in the 35-45 range. Their work provides new hope for better understanding of the connections between the repeat length dependence of polyQ physical properties and the repeat length dependence of disease risk, but this remains a major challenge. It is also important to note that the polyQ flanking sequences play an important role in the nucleation and growth processes. Pappu and collaborators<sup>28</sup> showed that for one important polyQ disease protein, the Huntingtin protein, the N- and C-terminal flanking sequence modules from exon 1 of huntingtin act as gatekeepers, whereby the N-terminal flanking sequence accelerates fibril formation while destabilizing nonfibrillar species, whereas the C-terminal flanking sequence reduces the overall driving force for aggregation. Although it was initially thought that in these amyloid proteins the fibril network was responsible for the disease, it is now believed that it might result from toxic oligomers<sup>40,130</sup> formed during the nucleation and growth process. Given these experimental advances, it is important to understand theoretically the early stages of aggregation.



**Figure 4.1:** Schematic representation of an amino acid in the PLUM model from Reference 12. Dashed lines represent beads from neighboring residues to illustrate local geometry.

### 4.3 Model and Methodology

In this chapter we study the early stages of aggregation of a coarse grained model of polyQ for different repeat lengths, focusing primarily on the nucleation mechanism, using a generic coarse grained model proposed by Berau and Deserno.<sup>12</sup> This model provides an intermediate level of resolution with four beads per amino acid and an implicit solvent. It offers full sequence specificity and includes the important backbone hydrogen bonds. The Berau-Deserno model (or PLUM model) has been used by Luiken and Bolhuis<sup>82,83</sup> to study the nucleation kinetics of three peptides.

This model is not biased towards a particular secondary structure, nor does it include additional terms to the potential in order to force known structures. Both solvent and electrostatics are treated implicitly in PLUM.

In order to accurately model local geometry necessary for secondary structure, two- (length), three- (angle), and four-body (dihedral) bonded interactions are included. The two- and three-body interactions both take the form of very stiff, harmonic potentials, given by:

$$U_{bond}(r) = \frac{1}{2}k_{bond}(r - r_0)^2 \quad (4.1)$$

$$U_{angle}(\theta) = \frac{1}{2}k_{angle}(\theta - \theta_0)^2 \quad (4.2)$$

Dihedrals are written as a Fourier series in their respective rotation angle:

$$U_{dih}(\varphi) = k_n[1 - \cos(n\varphi - \varphi_{n,0})] \quad (4.3)$$

For amino acids immediately preceding a proline residue, up to the  $n = 2$  term is used, yielding two minima. For all others, only the  $n = 1$  term is kept, which has a single minimum. Additionally, dipole-dipole interactions between peptide bonded carbonyl and amide groups are folded into this potential by parameter tuning. Parameters for these interactions can be found in Table 4.1.

PLUM uses three unbonded interaction potentials. The first is a simple volume exclusion given by Weeks-Chandler-Andersen potential (the repulsive portion of the standard Lennard-Jones potential):

$$U_{bb}(r) = \begin{cases} 4\epsilon_{bb}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^6 + \frac{1}{4}\right] & r < 2^{\frac{1}{6}}\sigma_{ij} \\ 0 & r > 2^{\frac{1}{6}}\sigma_{ij} \end{cases} \quad (4.4)$$

where  $\sigma_{ij}$  is the arithmetic mean of the two involved beads' diameters. This potential applies to all backbone-backbone and backbone-side chain interactions. Bead specific parameters can be found in Table 4.2.

For side chain-side chain interactions, a piece-wise Lennard-Jones function is



**Table 4.1:** Bonded interaction parameters for the PLUM coarse grained protein model.

Bond Lengths					
	$NC_\alpha$	$C_\alpha C'$	$C'N$	$C_\alpha C_\beta$	
$r_0(\text{\AA})$	1.455	1.510	1.325	1.530	
$k_{bond}(\epsilon/\text{\AA}^2)$	300	300	300	300	
Bond Angles					
	$NC_\alpha C_\beta$	$C_\beta C_\alpha C'$	$NC_\alpha C'$	$C_\alpha C'N$	$C'NC_\alpha$
$\theta_0(deg)$	108	113	111	116	122
$k_{angle}(\epsilon/deg^2)$	300	300	300	300	300
Dihedrals					
	$\phi$	$\psi$	$\omega$	$\omega_{pro}$	Improper
$k(\epsilon)$	-0.3	-0.3	67.0	3.0	17.0
$n$	1	1	1	2	1
$\varphi_0(deg)$	0	0	180	0	$\mp 120$

used to phenomenologically model hydrophobic, side chain hydrogen bonding, and electrostatic interactions in one potential:

$$U_{hp}(r) = \begin{cases} 4\epsilon_{hp}[(\frac{\sigma_{C_\beta}}{r})^{12} - (\frac{\sigma_{C_\beta}}{r})^6] + (\epsilon_{hp} - \epsilon_{ij}) & r < 2^{\frac{1}{6}}\sigma_{C_\beta} \\ 4\epsilon_{hp}\epsilon_{ij}[(\frac{\sigma_{C_\beta}}{r})^{12} - (\frac{\sigma_{C_\beta}}{r})^6] & 2^{\frac{1}{6}}\sigma_{C_\beta} < r < r_{hp,cut} \\ 0 & r > r_{hp,cut} \end{cases} \quad (4.5)$$

All side chain beads are taken to be the same size with the same interaction range and cutoff (values given in Table 4.2). Amino acid specificity comes in through the

**Table 4.2:** General unbonded interaction parameters for the PLUM coarse grained protein model.

Backbone Excluded Volume Interaction			
$\sigma_N(\text{\AA})$	$\sigma_{C_\alpha}(\text{\AA})$	$\sigma_{C'}(\text{\AA})$	$\epsilon_{bb}(\varepsilon)$
2.9	3.7	3.5	0.02
Hydrophobic Interaction			
$\sigma_{C_\beta}(\text{\AA})$	$\epsilon_{hp}(\varepsilon)$	$r_{hp,cut}(\text{\AA})$	
5.0	4.5	10.0	
Hydrogen Bonding Interaction			
$\sigma_{hb}(\text{\AA})$	$\epsilon_{hb}(\varepsilon)$	$r_{hb,cut}(\text{\AA})$	
4.11	6.0	8.0	

term  $\epsilon_{ij}$ , which is the geometric mean of the individual residue parameters given in Table 4.3, governing the strength of the side chain-side chain attractive interaction.

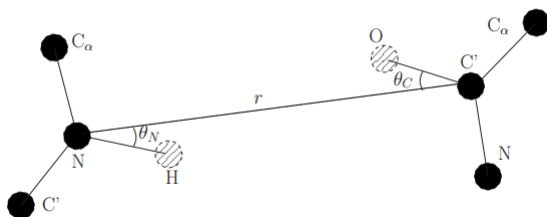
Lastly the hydrogen bonding potential is given by 12-10 Lennard-Jones combined with an angular term:

$$U_{hb}(r, \theta_N, \theta_C) = \epsilon_{hb} \left[ 5 \left( \frac{\sigma_{hb}}{r} \right)^{12} - 6 \left( \frac{\sigma_{hb}}{r} \right)^{10} \right] \times \begin{cases} \cos^2 \theta_N \cos^2 \theta_C & r < r_{hb,cut}, |\theta_N| < 90^\circ, \text{ and } |\theta_C| < 90^\circ \\ 0 & \text{else} \end{cases} \quad (4.6)$$

This form uses the inferred location of light atoms involved in the hydrogen bond based on the local geometry of the backbone, requiring the positions of six beads, as shown in Figure 4.2. This interaction exists between any pair of  $N$  and  $C'$  beads, unless the  $N$  belongs to a proline residue.

**Table 4.3:** Normalized amino acid specific hydrophobicity for the PLUM coarse grained protein model.

Amino Acid	$\epsilon_i(\varepsilon)$
Lys / K	0.00
Glu / E	0.05
Asp / D	0.06
Asn / N	0.10
Ser / S	0.11
Arg / R	0.13
Gln / Q	0.13
Pro / P	0.14
Thr / T	0.16
Gly / G	0.17
His / H	0.25
Ala / A	0.26
Tyr / Y	0.49
Cys / C	0.54
Trp / W	0.64
Val / V	0.65
Met / M	0.67
Ile / I	0.84
Phe / F	0.97
Leu / L	1.00



**Figure 4.2:** Schematic of the PLUM hydrogen bonding interaction from Reference 12. Dashed beads are the implicit hydrogen and oxygen atom positions, calculated from the backbone heavy atom positions.

We use the LAMMPS molecular dynamics simulator<sup>103</sup> and study a system of 1000 “monomers” (each representing a polyQ chain) primarily for two different repeat lengths,  $Q_{10}$  and  $Q_{30}$ . Initial peptide conformations are selected from low energy configurations obtained via dynamics simulation of a single chain of the desired length, and placed with randomized position and orientation. Simulations are run until a plateau in monomer concentration is reached. We take this value to be an estimate of the equilibrium concentration and use this as a reference to define our supersaturation value,  $s$ . However, the uncertainty in this equilibrium value could lead to a systematic error in our determination of the supersaturation. We average over 10 different starting configurations for each supersaturation examined, in order to minimize finite size effects. We find that additional runs have a negligible effect on quantities of interest. We study the nucleation mechanism involved in the initial stages of growth namely, the early kinetic stages of oligomer aggregation. Finally, we address the issue raised by the above referenced experimental studies: the dependence of the critical nucleus size on repeat length.

## 4.4 Results

### 4.4.1 $Q_{10}$

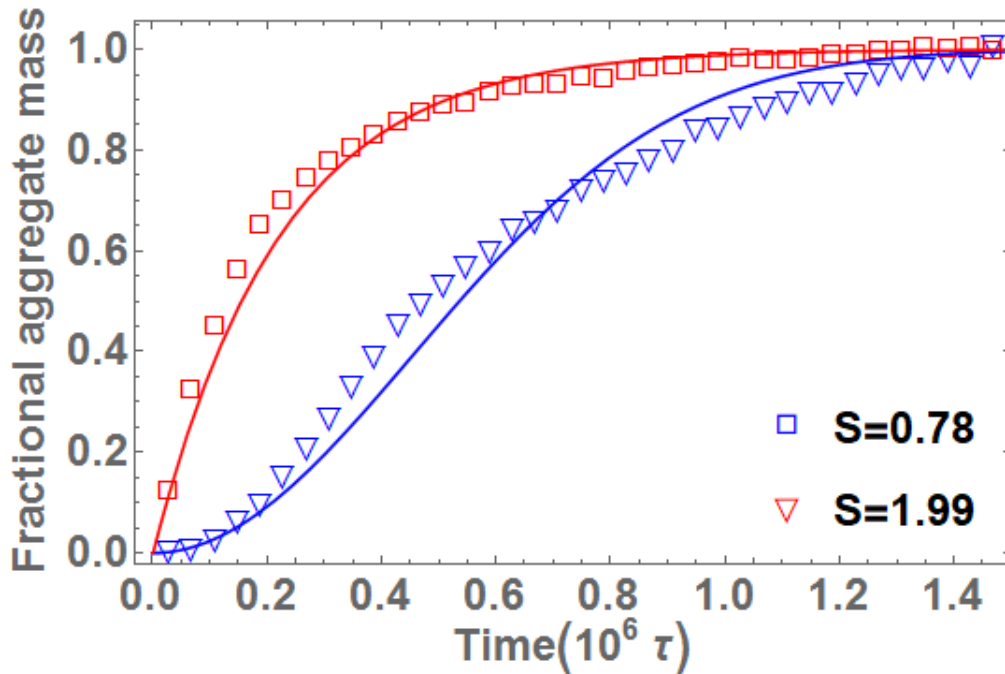
#### 4.4.1.1 Aggregate Mass Growth

We first examine the aggregation of  $Q_{10}$  at  $T = 300K$ . Figure 4.3 shows plots of the fraction of mass aggregation as a function of time<sup>1</sup>, which demonstrate two different pathways to aggregation. The first occurs at a relatively high supersaturation for which there is no induction or lag time, i.e. there is no free energy barrier to growth and hence no nucleation process is involved. This process has been termed “metanucleation”.<sup>21</sup>The second plot indicates the pathway for aggregation that occurs at smaller supersaturation, in which one finds a standard nucleation and growth curve. To determine the nucleation mechanism involved in this process, we have visually analyzed the critical nuclei at different supersaturations. We then estimate the critical nucleus size  $n_c$ , by directly examining the relative rates of monomer addition and dissociation for a cluster of size  $n$  peptides over time, averaging over all clusters of size  $n$  in the simulation and across individual runs. We define any two peptide chains as connected when their total pairwise interaction exceeds a minimum of  $2\varepsilon$  binding energy, where  $\varepsilon = kT_r = k \times 300K$  is the characteristic energy scale of the PLUM model. Clusters are then built iteratively by tracing these connections.  $n_c$  is then given by the smallest cluster size that yields a net forward rate.

We have also tried to analyze the behavior of the mass aggregation curve shown in Figure 4.3, whose behavior is a combination of nucleation and growth. There

---

<sup>1</sup>Time units are given in units of  $\tau$ , the natural time unit of the PLUM model. For more details, see reference.<sup>12</sup>



**Figure 4.3:** Aggregate mass growth over time for two supersaturations of  $Q_{10}$ . The higher value of supersaturation,  $s=1.99$ , (red) demonstrates barrierless growth. Its underlying cluster distribution shows stable clusters across a range of sizes. At lower supersaturation,  $s=0.78$ , (blue), a lag period is followed by the nucleation of a cluster with size above  $n_c$ , and finally growth. Such systems contain a large population of transient oligomers of size less than  $n_c$  with a gap in sizes before stable clusters emerge.

have been recent, significant developments in the theory of filament assembly<sup>71,89</sup> that describes the type of sigmoidal nucleation and growth behavior we see in Figure 4.3, for which analytical solutions in certain cases have been obtained. The master equation describing the filament growth includes the possibility of secondary nucleation in addition to the primary nucleation we study here. (For a review, see<sup>90</sup>). We have tried preliminary fits of the mass aggregation curve shown in Figure 4.3 using the recent theory.<sup>71,90</sup> Our best fit to the data using the approximate analytical solution obtained by Knowles et al<sup>71</sup> is shown in Figure 4.3. The solution we use

for the mass aggregation curve is

$$\frac{M(t)}{m_{tot}} = 1 - \exp[-A(\cosh \kappa t - 1)] \quad (4.7)$$

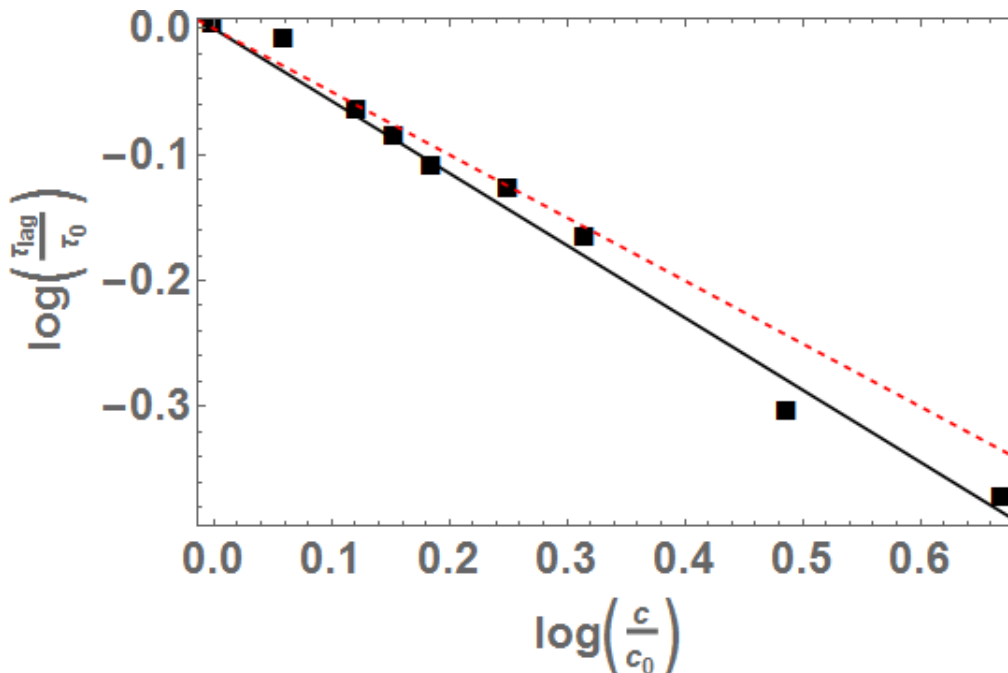
with  $A = \frac{k_n m_{tot}^{n_c-1}}{k_-}$  and  $\kappa = \sqrt{2k_+ k_- m_{tot}}$ . We also find that the induction time scales like  $\tau_{lag} \sim C_1^\gamma$ , with  $\gamma = -0.57$ , where  $C_1$  is the monomer concentration, shown in Figure 4.4. Our result is consistent with the recent theoretical prediction<sup>71</sup> that  $\gamma = -0.5$ , and disagrees with the classical prediction  $\gamma = -(n_c + 1)/2$ , in which primary nucleation is a dominant mechanism determining the lag time. Our result for the lag time scaling behavior is consistent with secondary nucleation, as might arise from fragmentation of aggregates. Finally, we note that if we use a result from filament theory that the coefficient of the early time  $t^2$  behavior of the mass aggregation curve yields the critical nucleus size,<sup>44</sup> we find values that are consistently somewhat smaller than those we report here and have a more irregular dependence on supersaturation than found in our determination. However, the values obtained this way depend on the window of time that we use to estimate the quadratic in time behavior.

#### 4.4.1.2 Nucleation Mechanism

We now focus on the nucleation mechanism involved in this self-assembly process, using an existing theory<sup>2</sup> of self assembly proposed by Kashchiev and Auer.<sup>66</sup> In

---

<sup>2</sup>We first note that for peptide chains such as considered here, the combination of a non-directional interaction, due to hydrophobic forces, and directional interactions such as hydrogen bonding, in principle can lead to a complex non-standard nucleation behavior, depending on the relative degree of anisotropy.<sup>21,67</sup> In the non-standard case there is no well defined concept of a critical nucleus, as the work of formation does not scale in a definite way with the nucleus size. In such a case the size of the fibril nucleus does not have a unique value and nanofibrils of different size and shape can act as nuclei. However, we find that, at least to a first approximation, the



**Figure 4.4:** Scaling of lag phase with initial protein concentration disagrees with classical nucleation theory. The best fit slope corresponding to  $\gamma = -0.57$  is shown in black. The slope corresponding to  $\gamma = -0.5$ , when primary nucleation is not the dominating contribution to the lag phase, is shown in dashed red.

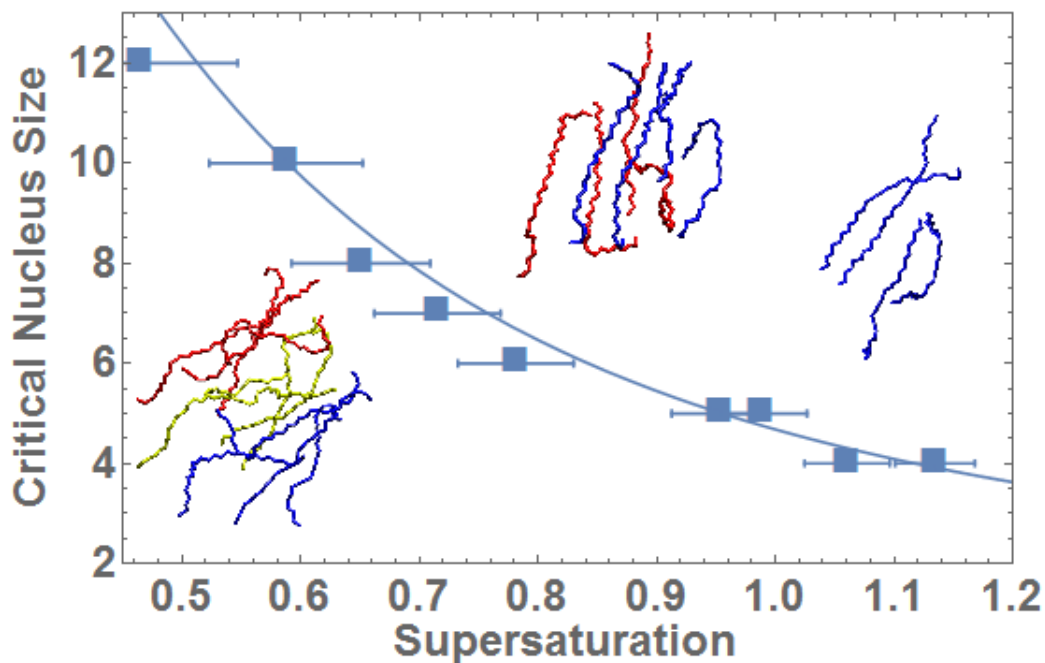
this theory nucleation is viewed as a two-dimensional process analogous to two dimensional crystallization. The amyloid fibrils (protofilaments) are assumed to be built up of successively layered  $\beta$ -sheets with fixed width and thickness. Each sheet has fully extended rodlike peptides arranged parallel to each other and perpendicular to the fibril elongation axis. The interaction between nearest neighbor peptide chains in a  $\beta$ -sheet is dominated by hydrogen bonds, which lead to a specific surface tension energy  $E$ . Interactions between nearest neighbor peptide chains in adjacent  $\beta$ -sheets involve much weaker bonds, such as hydrophobic forces, that lead to a specific surface

---

anisotropy in our model is not sufficiently large as to produce this non-standard behavior and as a consequence we attempt to analyze our results in terms of a classical theory described in the text.



tension energy  $E_h$ . There are several relevant predictions of the classical nucleation theory. The first is that fibrils begin as single  $\beta$ -sheets and develop a second  $\beta$ -sheet only after reaching a supersaturation independent transition size  $n_t$  given by  $n_t = E/E_h$ . We find that  $n_t = 4$  which fixes the ratio of the two surface tension energies and that  $n_c = 4$ ; we find that as one decreases the supersaturation, additional  $\beta$ -sheets are formed as predicted by the theory and that  $n_t$  is indeed independent of supersaturation. Figure 4.5 shows three different critical nuclei observed in our study, ranging from the minimal size  $n_c = 4$ , consisting of a single filled  $\beta$  sheet of four peptide chains up to and including a critical nucleus  $n_c = 12$ , consisting of three filled  $\beta$  sheets. The latter was obtained using a seed to nucleate the system, as the induction time was otherwise too long at that supersaturation. A second prediction of the theory is that a filled single  $\beta$  sheet occurs at the largest supersaturation,  $s_1 = 2E_h/kT$ , before one passes into the barrierless regime. We estimate that  $s_1 = 1.13$  which leads to the approximate value  $E_h = 0.6kT$  and from the relation  $n_t = E/E_h$  the estimate that  $E = 2.4kT$ . Finally, Kashchiev and Auer predict that<sup>66</sup>  $n_c$  increases with decreasing supersaturation, with  $n_c - 1 = A/s^2$  and  $A = 4EE_h/(kT)^2$ . Our data is consistent with this prediction for all  $s$  in the range studied (within our estimated error bars for  $s$ ), as shown in Figure 4.5, with  $A \approx 6.28$  as obtained from the best fit. If we substitute the values of  $E_h = 0.6kT$  and  $E = 2.4kT$  into their theoretical prediction for  $A$ , we find  $A = 5.8$  so that our results seem consistent with their predictions, within the accuracy of our data. We note that one obvious limitation of the Kashchiev-Auer model, however, is that it does not consider the full range of possible conformational changes in the peptides which occur in proteins and in our model peptide chains. As a consequence, clusters

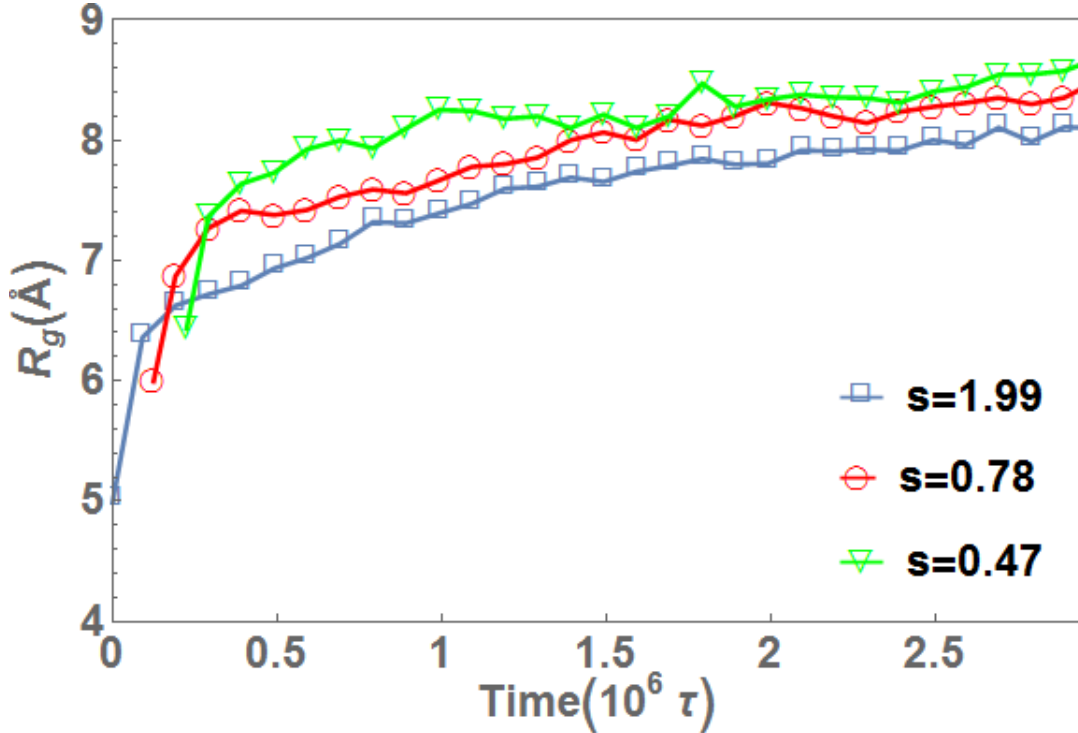


**Figure 4.5:** Critical nucleus size as a function of supersaturation, fit to  $s^{-2}$ . Inset are critical nuclei for  $S=0.47$ ,  $0.65$ , and  $1.13$ , which feature 3, 2, and 1 layer  $\beta$ -sheet conformations respectively, with peptides color coded to layer.

frequently form around a more disordered, larger than critical size, nucleus. These clusters feature a larger number of hydrophobic contacts and fewer hydrogen bonds.

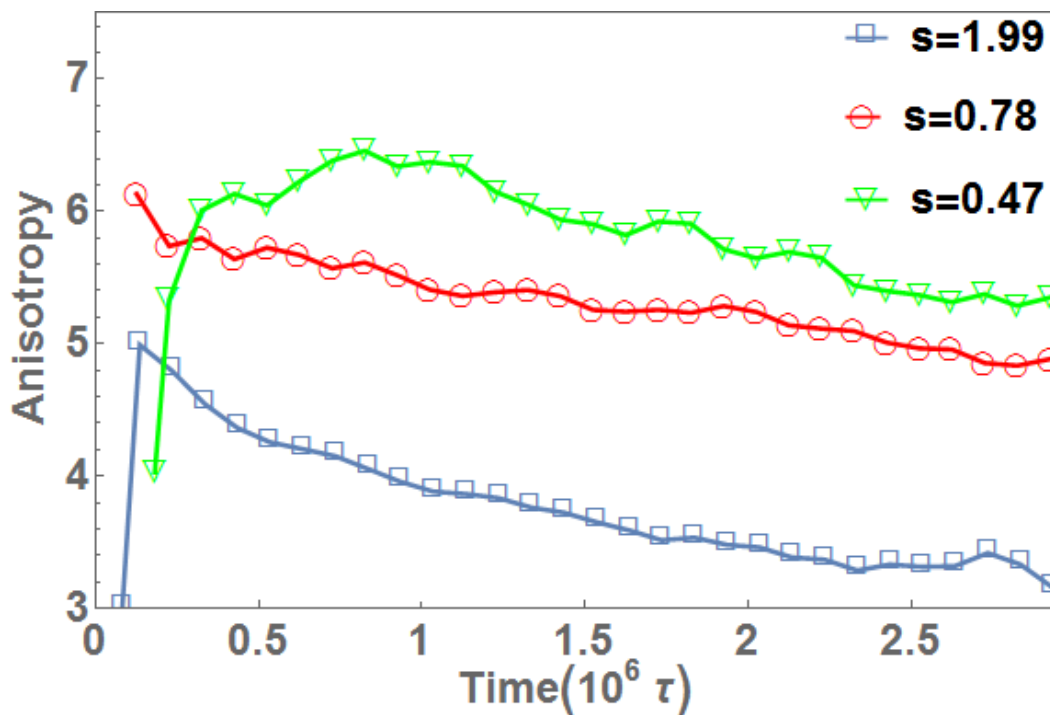
#### 4.4.1.3 Oligomer Growth

Next, we briefly provide a more quantitative description of the time evolution of the oligomers. This can be seen in Figure 4.6, which is a plot of the average radius of gyration,  $R_g$  vs time, showing its reaching a plateau value after a rapid initial growth that is due to the initial monomer state. Larger supersaturations have smaller average size (smaller  $R_g$ 's) reflecting the stability of clusters smaller than the critical nucleus size of smaller supersaturation values, reflected in their respective

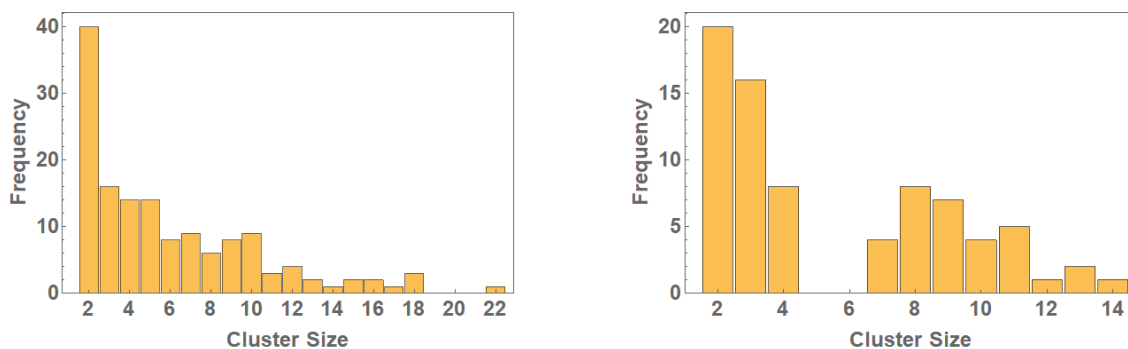


**Figure 4.6:** Average radius of gyration of clusters over time for three supersaturations of  $Q_{10}$ . Higher supersaturations have smaller average size, reflecting the stability of clusters smaller than the critical nucleus size of lower supersaturations.

cluster distribution functions (Figure 4.8). A better description of the anisotropic morphology can be obtained from a plot of the ratio of  $\lambda_1/\lambda_3$  where  $\lambda_1$  and  $\lambda_3$  are the smallest and largest eigenvalues of the mass inertia tensor, respectively. This plot is shown in Figure 4.7, for a few supersaturations, in which this ratio slowly decreases with time after an initial rapid growth, which is also due to the initial starting point being a collection of monomers. We note from the plot of the behavior of the mass inertia tensor that the growth is not of the one dimensional nature found in the case of extreme anisotropy.



**Figure 4.7:** Average cluster anisotropy (ratio of largest to smallest moments of inertia) over time for three supersaturations of  $Q_{10}$ . Clusters formed are highly anisotropic due to the typically extended conformation of  $Q_{10}$  monomers within the cluster and the preferred direction imparted by hydrogen bonding.



**Figure 4.8:** Cluster distribution plot for  $Q_{10}$  aggregates. For  $S = 1.99$  (left), clusters exist across a wide range with no gaps. For the case of  $S = 0.72$  (right), there exist a large number of transient small clusters, followed by a gap in the distribution, and finally a number of stable clusters above  $n_c = 7$ .

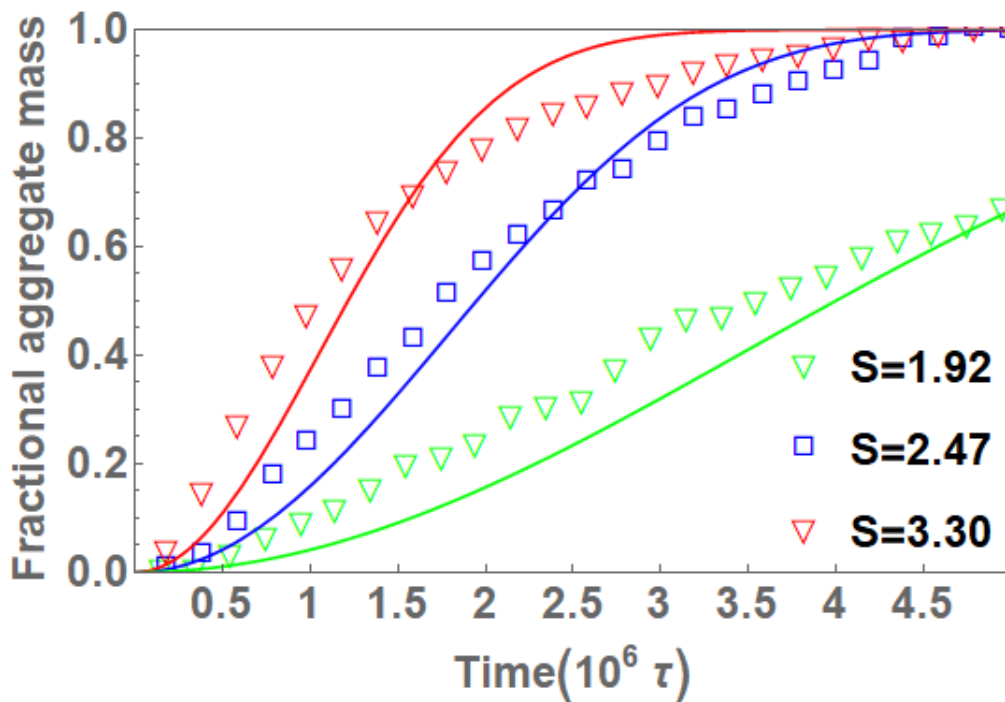
## 4.4.2 $Q_{30}$

### 4.4.2.1 Aggregate Growth and Nucleation Mechanism

Next we examine the growth of  $Q_{30}$  aggregates at  $T = 300K$ . Plotting the fraction of mass aggregation over time (Figure 4.9) consistently shows the presence of a lag time. The duration of this lag phase scales well with the square of the average separation distance between monomers at that concentration; this suggests that the aggregation process is diffusion limited, rather than limited by a nucleation event. This is further supported by examining the relative rates of monomer association and disassociation for all clusters of a given size  $n$ . We find that there is net growth for clusters of all sizes across the range of supersaturations examined ( $s = 0.78$  to  $s = 3.46$ ). Fitting to the solution given by Equation 4.7, we find the parameters  $A$  and  $\kappa$  to be consistent across all supersaturations. Due to the dependence of these values on critical nucleus size, we can further confirm there is no change in  $n_c$  as a function of supersaturation and that  $n_c = 1$ . This is in agreement with experimental observations<sup>65,101</sup> that the critical nucleus is repeat length dependent and that as one increases the length, the critical nucleus becomes a monomer. In particular, it agrees with the experimental result<sup>42</sup> for  $Q_{30}$  that the critical nucleus is a monomer.

### 4.4.2.2 Oligomer Growth

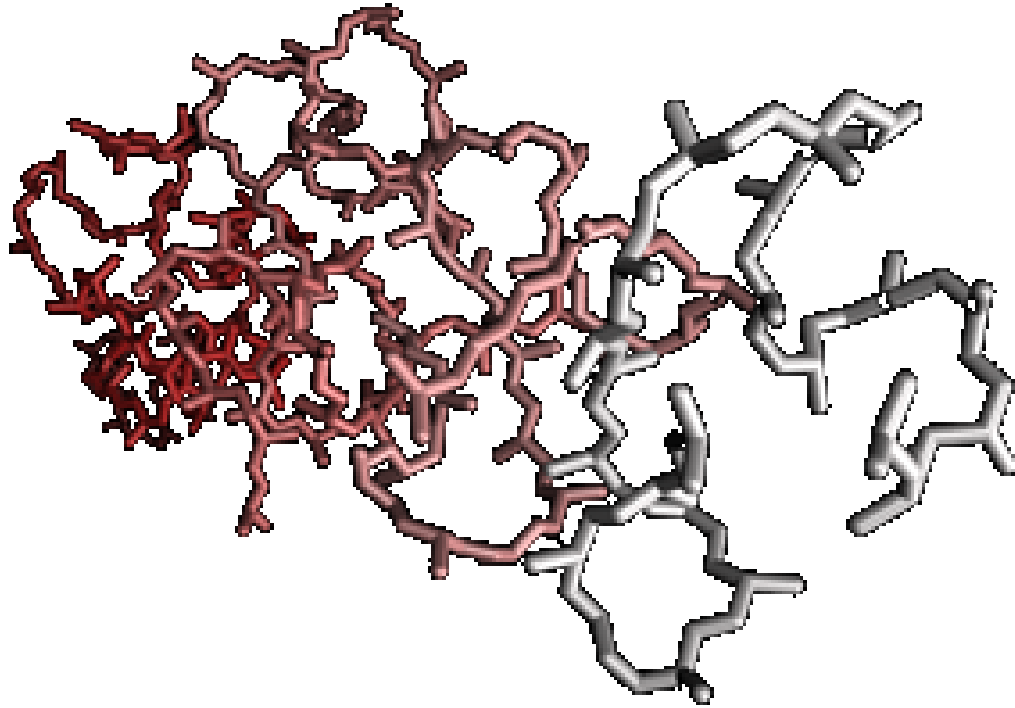
We show in Figure 4.10 a typical small oligomer for the longer repeat length case. Individual monomers in this arrangement have a much larger number of internal hydrogen bonds than in the case of  $Q_{10}$ , and tend towards a stacked helix arrangement. This is again similar to experimental observations<sup>65,101</sup> and computational



**Figure 4.9:** Aggregate mass growth over time for three supersaturations of  $Q_{30}$ . All show the same qualitative behavior and consistent values of  $A$  and  $\kappa$ .

study.<sup>70, 87, 94, 98</sup>

The average radius of gyration as a function of time is very similar for different supersaturations, as shown in Figure 4.12, with the exception of the length of the induction period. Unlike the case of  $Q_{10}$ , all cluster sizes are viable at a given supersaturation (Figure 4.11), resulting in each supersaturation having a similar underlying distribution. As such, the typical size and anisotropy of clusters are comparable. Figure 4.13 quantitatively shows the average anisotropy of clusters as a function of time.  $Q_{30}$  clusters are considerably more anisotropic than  $Q_{10}$ , reflecting the tendency to arrange in proto-fibrillar configurations more readily. Also in contrast to  $Q_{10}$ , over time these clusters are not tending towards more spherical

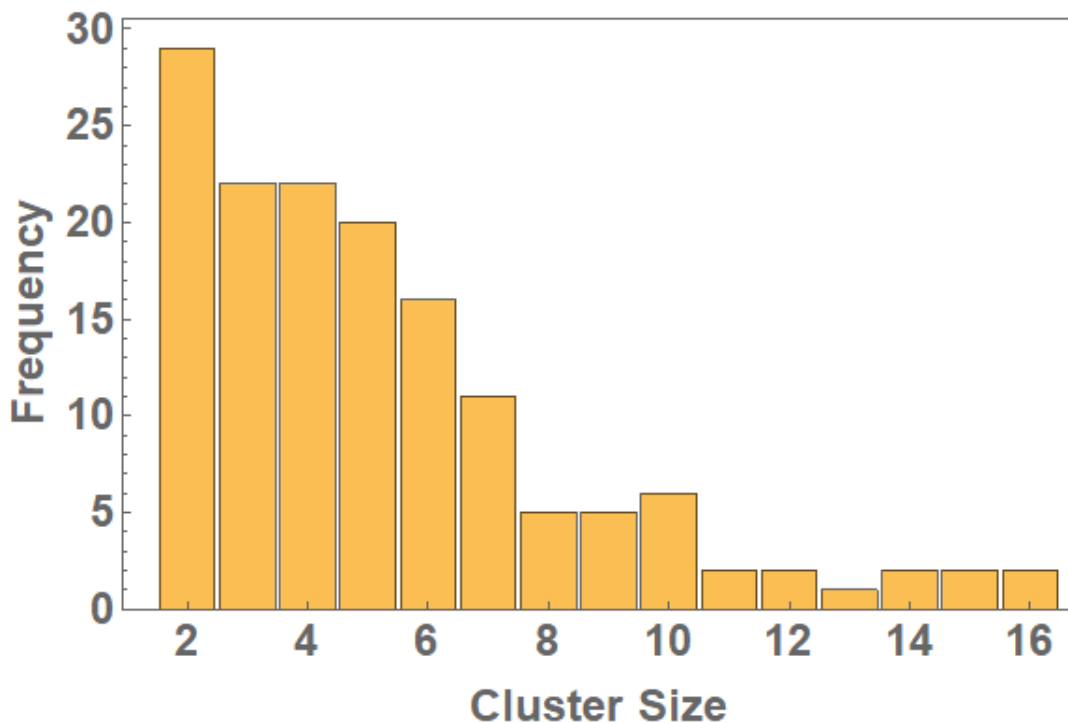


**Figure 4.10:** Typical  $Q_{30}$  tetramer forms from stacked helices in contrast to the sheet-like  $Q_{10}$  clusters.  $Q_{30}$  monomers adopt a more helical conformation than  $Q_{10}$ , with significantly larger self-interactions.

oligomers, instead maintaining very linear structures.

### 4.4.3 Nucleation Barrier

Lastly we examine the presence of a barrier to nucleation as a function of repeat length. We performed simulations across the range of temperatures  $275 - 300K$  at fixed concentration for polyQ constructs ranging from  $Q_{10}$  to  $Q_{30}$  in increments of five repeats. The concentration for all lengths was chosen such that  $Q_{10}$  had a critical nucleus size of 4 at  $T = 300K$ . We then repeat previous analysis to obtain estimates of the nucleation rate,  $k_n$ . Figure 4.14 shows a normalized plot of the

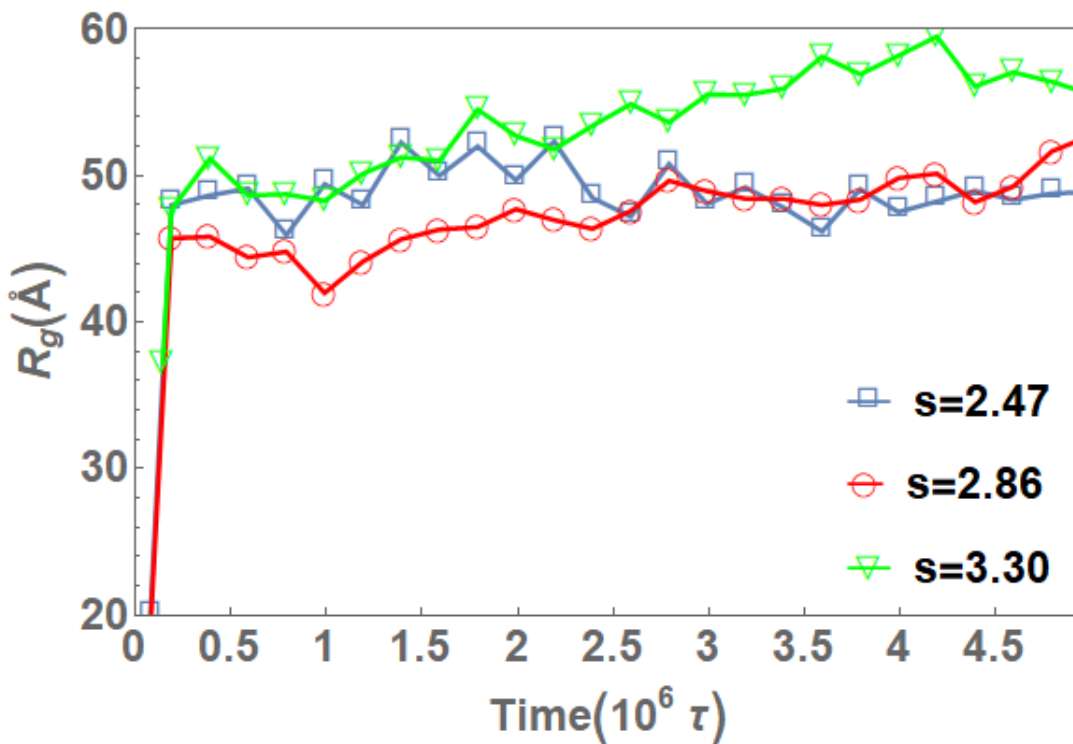


**Figure 4.11:** Distribution of cluster sizes for  $Q_{30}$  aggregates at monomer plateau. Clusters span the range of dimers to 16-mers with no gaps in the distribution. Growth at this stage is being driven by cluster combination.

log of the nucleation rate versus inverse temperature for  $Q_{10}$  and  $Q_{30}$ . Assuming the nucleation rate, at least locally, scales like an Arrhenius process, the slope of this curve gives the negative of the height of the nucleation barrier. In the case of  $Q_{30}$ , this slope is negligible, indicating the absence of a barrier to nucleation. In contrast,  $Q_{10}$  shows a barrierless regime at low temperature, and clear evidence of a substantial nucleation barrier at high temperatures.

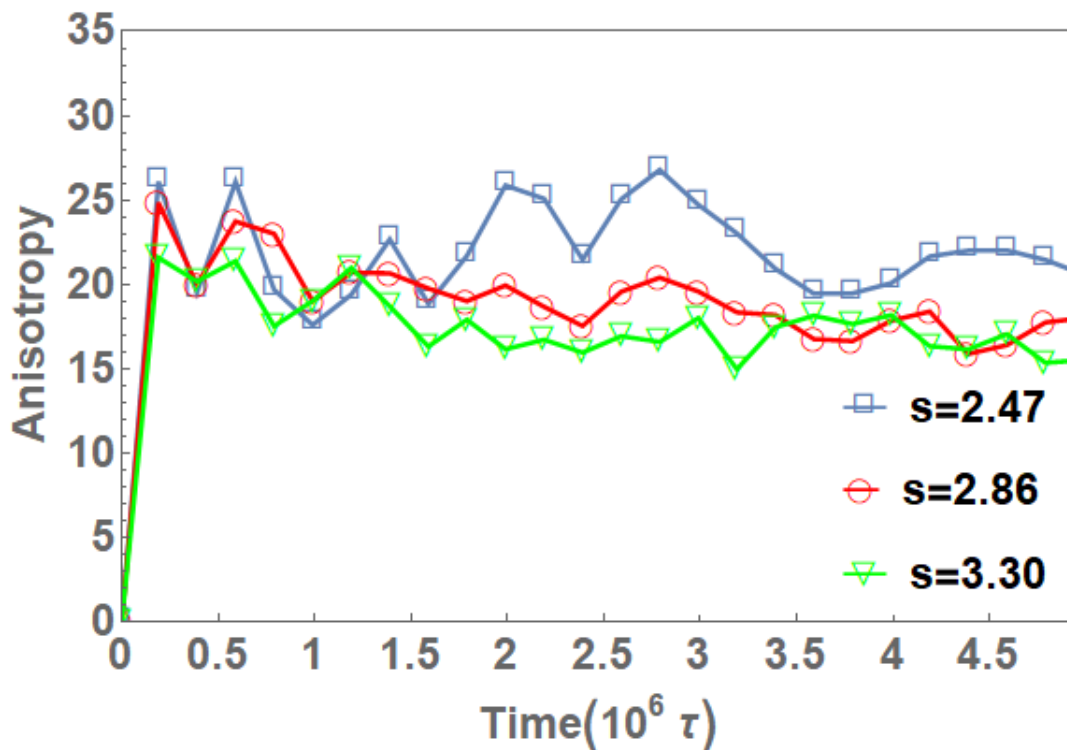
Repeating this analysis for the intermediate cases (Figure 4.15), we find that this model of polyQ rapidly shifts to barrierless behavior as repeat length increases. Both  $Q_{20}$  and  $Q_{25}$  exhibit near zero slopes across the range of temperatures examined.





**Figure 4.12:** Average radius of gyration of clusters over time for three supersaturations of  $Q_{30}$  shows a quick increase during initial cluster formation followed by a period of slow growth.

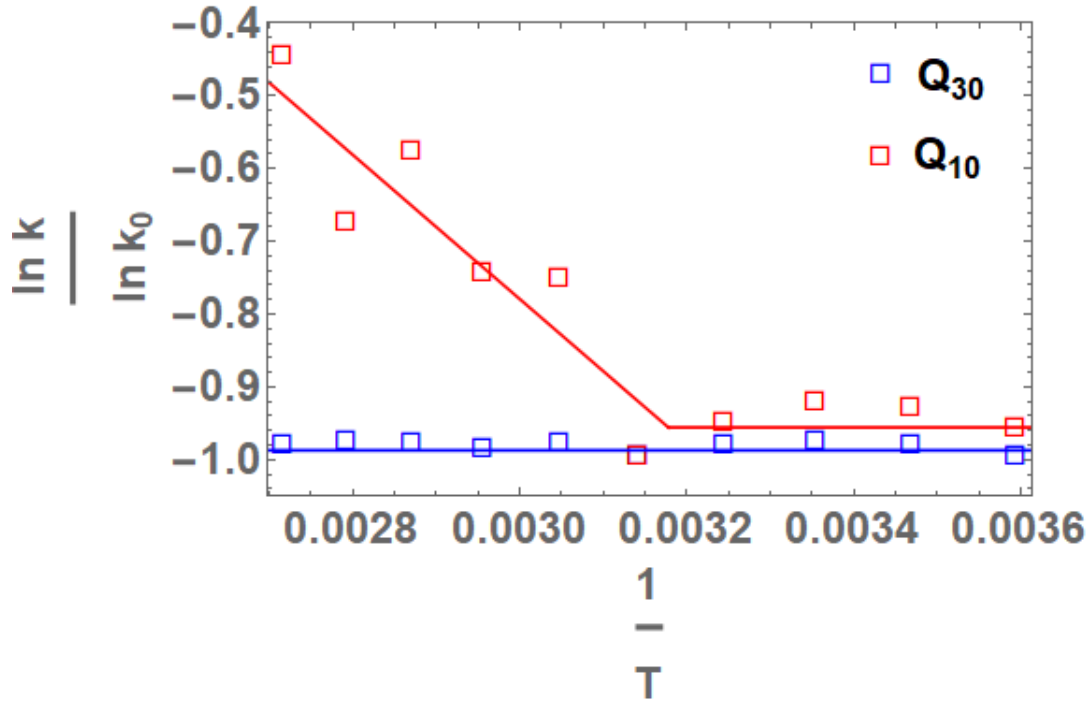
$Q_{15}$  shows a small nucleation barrier only at the highest end of this temperature range. This suggests there is a somewhat continuous, albeit rapid, transition from the nucleation process seen in  $Q_{10}$  at smaller supersaturations to the ever present metanucleation of  $Q_{30}$ . This crossover from nucleation to metanucleation is at a significantly smaller length than experimental work,<sup>65</sup> however. The difference between a pure polyglutamine construct and the more complicated peptides may be substantial in this area though, with flanking sequence playing a critical role.



**Figure 4.13:** Average cluster anisotropy (ratio of largest to smallest moments of inertia) over time for three supersaturations of  $Q_{30}$ . Clusters maintain a very high degree of anisotropy over time due to their stacked helix arrangement.

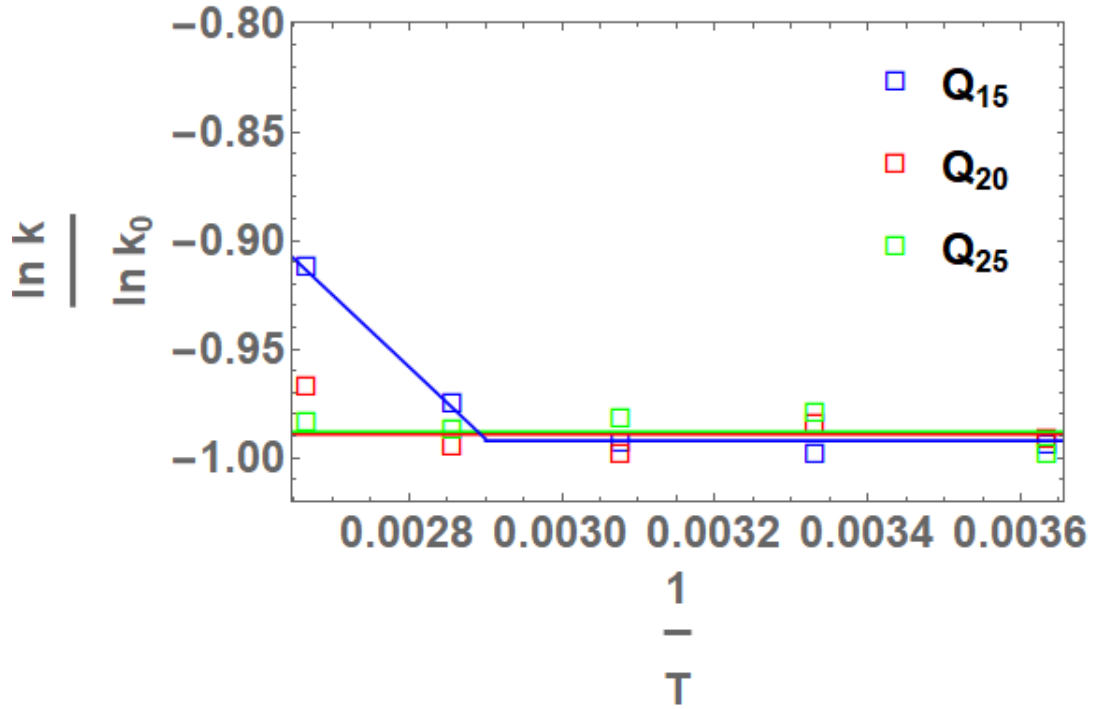
## 4.5 Conclusion

In conclusion, we have found that the nucleation mechanism for the PLUM model polyglutamine is in reasonable agreement with the theory of Kashchiev and Auer<sup>66</sup> for the case of  $Q_{10}$ . In particular, we have found that the maximum number of peptide chains in the one dimensional  $\beta$  sheet is independent of supersaturation and in the case of  $Q_{10}$  is equal to 4 at  $T = 300K$ . We have also found that the critical nucleus size is inversely proportional to the square of the supersaturation, as predicted by Kashchiev and Auer, subject to the caveat about the uncertainty



**Figure 4.14:** Arrhenius plot of the nucleation rates of  $Q_{10}$  and  $Q_{30}$ .  $Q_{10}$  shows two distinct regions, one with and without a nucleation barrier at high and lower temperature respectively.  $Q_{30}$  shows good agreement to a zero-slope linear fit, indicating the lack of a barrier to nucleation.

in our estimate of  $C_e$ . In addition, we have found that the critical nucleus is repeat length dependent,<sup>65</sup> equal to one peptide chain for our model for the case  $Q_{30}$ , in agreement with experiment,<sup>42</sup> though we find this transition occurs at lower repeat length than expected. Additionally, we see a change in oligomer composition as a function of repeat length, with short constructs forming  $\beta$  sheets and  $Q_{30}$  adopting  $\beta$  helix structures, in agreement with experimental observations.<sup>101</sup> Finally, although our focus in this chapter has been on the early stages of self-assembly, we have done a few long runs and have seen the oligomers first form a few large clusters which eventually form fibrils, shown in Figure 4.16, to verify the long time trajectory of

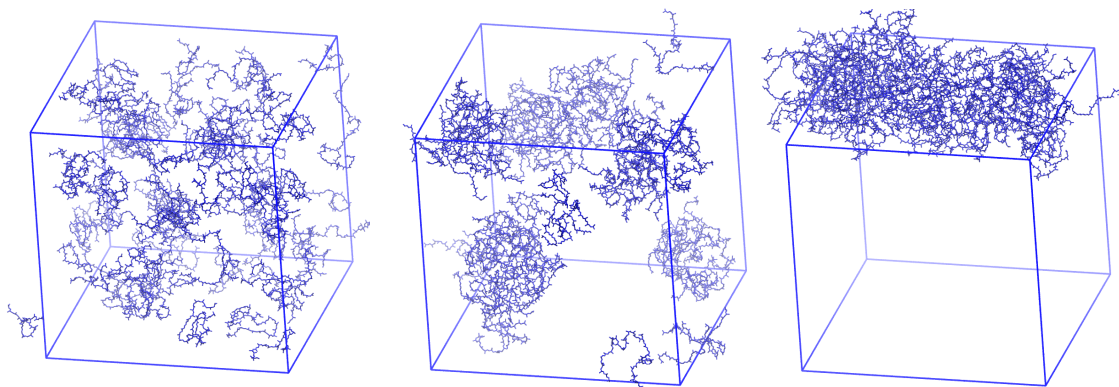


**Figure 4.15:** Arrhenius plot of the nucleation rates of  $Q_{15}$ ,  $Q_{20}$ , and  $Q_{25}$ . Only  $Q_{15}$  shows a non-negligible slope in the high temperature regime, and for a smaller range of temperatures than  $Q_{10}$ .

this system.

## Acknowledgments

This work was supported by a grant from the G. Harold and Leila Y. Mathers Foundation. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. One of us (N. B) was supported by NSF grant PHY-1359195.



**Figure 4.16:** Snapshots of system progression to long time scales of  $Q_{10}$  at high supersaturation. Initially a large number of small clusters are formed (left), which is followed by a period of consolidation into large, disordered oligomers (center). Finally these clusters reorganize into characteristic fibril forms.

# Chapter 5

## Conclusion

This chapter briefly examines "What's next?", the near term direction of work within the field of coarse grained protein simulation and colloidal science.

Coarse grained protein models come with a powerful advantage in their ability to reach time scales unobtainable by atomistic models. This, however, comes at the price of internal degrees of freedom, and as a result loss of realistic interactions with smaller components of the environment (i.e. complex solvent interactions such as retrograde solubility<sup>118</sup> or the effects of macromolecular crowding<sup>14,91,92,138</sup>). These shortcomings must be addressed to more faithfully model the cellular environment,<sup>41</sup> in which the concentration of the peptide of interest is typically vastly lower than *in silico*, yet the environment is more crowded than the usual coarse grained simulation.

Additionally, while models exist that combine both peptides and lipids,<sup>85</sup> this is typically done at a fairly high to moderate resolution. The role of membranes on protein aggregation, particularly in the area of neurodegenerative diseases appears to be of considerable importance.<sup>9,20,45,48,86,140</sup> Many of these proteins are membrane

bound or cleaved, and lipid membranes have been shown to induce aggregation in previously stable solutions and dramatically alter the free energy landscape of the aggregation process.<sup>54,111</sup> Combining single particle colloidal models of proteins with highly coarse grained lipid models<sup>27,33,126</sup> or network models of membranes<sup>19,96</sup> has great potential to provide insight into the aggregation process in these diseases as it progresses *in vivo*.

On top of better incorporation of these more complex interactions, is the ever continuing work of improving existing tools. More rigorous coarse graining techniques,<sup>108</sup> as well as refinements in the atomistic force fields to better capture intrinsically disordered proteins,<sup>61</sup> represent important steps forward in this area. In hardware, the relatively low cost of modern accelerators is pushing down the cost of supercomputing, expanding access and increasing the number of ongoing projects. At the other extreme, specialized hardware<sup>117,132</sup> is pushing orders of magnitude beyond what was possible even a short time ago, allowing simulations of tens of millions of particles on the microsecond scale.<sup>102</sup>

Colloidal and bioinspired materials with anisotropic features have an astounding number of potential uses as photonic materials, catalysts, electronic paper, and biomedical probes and delivery mechanisms.<sup>64,139</sup> Preparation of these particles on a scale needed for application development, however, is still lacking. Existing techniques are too time consuming per unit produced or too unreliable in quality. Designer peptides that can reproduce specific features of more complex proteins and be grown in bulk are likely to have more success in the immediate term.<sup>13,38</sup>

The next phase of research in these areas will hopefully see computational biology continue to better reflect experiment and experimental colloidal work able to produce

on a scale to sift through the great many predictions from simulation and theory waiting to be tested.



# Bibliography

- <sup>1</sup> S. Yadav T. W. Patapoff S. J. Shire A. Chaudri, I. E. Zarraga and G. A. Voth. *J. Phys. Chem. B*, 117:1269, 2013.
- <sup>2</sup> Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.
- <sup>3</sup> Clment Arnarez, Jaakko J. Uusitalo, Marcelo F. Masman, Helgi I. Inglfsson, Djurre H. de Jong, Manuel N. Melo, Xavier Periole, Alex H. de Vries, and Siewert J. Marrink. Dry martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *Journal of Chemical Theory and Computation*, 11(1):260–275, 2015.
- <sup>4</sup> S. Asakura and F. Oosawa. On Interaction between Two Bodies Immersed in a Solution of Macromolecules. *J. Chem. Phys.*, 22:1255, 1954.
- <sup>5</sup> S. Asakura and F. J. Oosawa. *J. Chem. Phys.*, 22:1255, 1954.
- <sup>6</sup> S. Auer. *J. Chem. Phys.*, 135, 2011.
- <sup>7</sup> S. Auer and D. Frenkel. *Adv. Polym. Sci.*, 207:14, 2005.

- <sup>8</sup> S. Auer and D. Frenkel. Numerical simulation of crystal nucleation in colloids. *Adv. Polym. Sci.*, 173:149, 2005.
- <sup>9</sup> P. K. Auluck, G. Caraveo, and S. Lindquist.  $\alpha$ -synuclein: Membrane interactions and toxicity in parkinson's disease. *Annu. Rev. of Cell and Dev. Bio.*, 26, 2010.
- <sup>10</sup> M. Baaden and S. J. Marrink. *Curr. Opin. in Structural Biology*, 23:878, 2013.
- <sup>11</sup> G P Bates and C. Benn. *Huntington's disease*. ed G. P. Bates, P. S. Harper and L. Jones, Oxford University Press, Oxford, UK, 2002.
- <sup>12</sup> T. Bereau and M. Desiarno. *J Chem Phys*, 130:235106, 2009.
- <sup>13</sup> Bryan Berger, Zhou Yang, Leah Spangler, Victoria Berard, Qian He, Li Lu, Robert Dunleavy, Christopher Kiely, and Steven McIntosh. Scalable biosynthesis of quantum dots: evolution of size selectivity, solubility and extracellular production. In *PROTEIN SCIENCE*, volume 24, pages 1–1. WILEY-BLACKWELL 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2015.
- <sup>14</sup> A. Bhattacharya, Y. C. Kim, and J. Mittal. Protein-protein interactions in a crowded environment. *Biophys. Rev.*, 5:99–108, 2013.
- <sup>15</sup> A. M. Bhattacharyya, A. K. Thakur, and R. Wetzal. Polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction. *Proc Nat Acad Sci*, 102:15400–5, 2005.
- <sup>16</sup> E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli, and F. Sciortino. . *Phys. Rev. Lett.*, 97:168301, 2006.

- <sup>17</sup> K Binder, P. Virnau, and A. Statt. Perspective: The Asakura Oosawa Model: A colloid prototype for bulk and interfacial phase behavior. *J. Chem. Phys.*, 141:140901, 2014.
- <sup>18</sup> L. Bromberg, J. Rashba-Step, and T. Scott. Insulin particle formation in supersaturated aqueous solutions of polyethylene glycol. *Biophys. Jour.*, 89:3424, 2005.
- <sup>19</sup> Frank L. H. Brown. Continuum simulations of biomembrane dynamics and the importance of hydrodynamic effects. *Quarterly Reviews of Biophysics*, 44(4):391432, 2011.
- <sup>20</sup> Robijn Bruinsma and Philip Pincus. Protein aggregation in membranes. *Current Opinion in Solid State and Materials Science*, 1(3):401 – 406, 1996.
- <sup>21</sup> R. Cabriolu, D. Kashchiev, and S. Auer. *J Chem Phys*, 137:204903, 2012.
- <sup>22</sup> S. Chen, F. Ferrone, and R. Wetzel. Huntington’s disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci USA*, 99:11884–11889, 2002.
- <sup>23</sup> Naresh Chennamsetty, Bernhard Helk, Vladimir Voynov, Veysel Kayser, and Bernhardt L. Trout. Aggregation-prone motifs in human immunoglobulin g. *Journal of Molecular Biology*, 391:404–413, 2009.
- <sup>24</sup> Naresh Chennamsetty, Vladimir Voynov, Veysel Kayser, Bernhard Helk, and Bernhardt L. Trout. Design of therapeutic proteins with enhanced stability. *PNAS*, 106:11937–11942, 2009.

- <sup>25</sup> Arthur J. Chirino, Marie L. Ary, and Shannon A. Marshall. Minimizing the immunogenicity of protein therapeutics. *Drug Discovery Today*, 9:82–90, 2004.
- <sup>26</sup> L. Cipelletti and L. Ramos. *J. Phys.: Condens. Matter*, year = 2005, volume = 17, pages = 253,.
- <sup>27</sup> Ira R. Cooke, Kurt Kremer, and Markus Deserno. Tunable generic model for fluid bilayer membranes. *Phys. Rev. E*, 72:011506, Jul 2005.
- <sup>28</sup> Scott L. Crick, Kiersten M. Ruff, Kanchan Garai, Carl Frieden, and Rohit V. Pappu. Unmasking the roles of n- and c-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *PNAS*, 110(50):20075–20080, 2013.
- <sup>29</sup> W. Dai, C. W. Hsu, F. Sciortino, and F. W. Starr. . *Langmuir*, 26:3601, 2010.
- <sup>30</sup> Bhas Dani, Robert Platz, and Stelios T. Tzannis. Pharmaceuticals, preformulation and drug delivery. high concentration formulation feasibility of human immunoglobulin g for subcutaneous administration. *Journal of Pharmaceutical Sciences*, 96:1504–1517, 2007.
- <sup>31</sup> P. Debye and E. Huckel. The theory of electrolytes. i. lowering of freezing point and related phenomena. *Phy. Z*, 24, 1923.
- <sup>32</sup> D. H. deJong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, and S. J. Marrink. *J. Chem. Theory and Computation*, 9:687, 2013.

- <sup>33</sup> Markus Deserno. Mesoscopic membrane physics: Concepts, simulations, and selected applications. *Macromolecular Rapid Communications*, 30(9-10):752–771, 2009.
- <sup>34</sup> T. Diekwisch. Private communication.
- <sup>35</sup> M. Dijkstra, J. M. Brader, and R. Evans. . *J. Phys. Condens. Matter*, 11:10079, 1999.
- <sup>36</sup> J. P. K. Doye, A.A. Louis, I-C. Lin, L.R. Allen, E.G. Noya, A.W. Wilber, H.C. Kok, and R. Lyus. Controlling crystallization and its absence: Proteins, colloids and patchy models. *Phys. Chem. Chem. Phys.*, 9:2197, 2007.
- <sup>37</sup> Chang Du, Giuseppe Falini, Simona Fermani, Christopher Abbott, and Janet Moradian-Oldak. Supramolecular Assembly of Amelogenin Nanospheres into Birefringent Microribbons. *Science*, 307:1450, 2005.
- <sup>38</sup> Robert Dunleavy, Li Lu, Christopher J. Kiely, Steven McIntosh, and Bryan W. Berger. Single-enzyme biomineralization of cadmium sulfide nanocrystals with controlled optical properties. *Proceedings of the National Academy of Sciences*, 113(19):5275–5280, 2016.
- <sup>39</sup> A. P. R. Eberle, R. Castaneda-Priego, J. M. Kim, and N. J. Wagner. . *Langmuir*, 28:1866, 2012.
- <sup>40</sup> D. Eisenberg and M. Jucker. *Cell*, 148:1188, 2012.
- <sup>41</sup> R. J. Ellis and A. P. Minton. Cell biology: join the crowd. *Nature*, 425:27–28, 2003.

- <sup>42</sup> A. K. Thakur et al. *Nat Struct Mol Biol*, 2009.
- <sup>43</sup> P. A. Fang, J. F. Conway, H. C. Margolis, J. P. Simmer, and E. Beniash. Hierarchical self assembly of amelogenin and the regulation of biomineralization at the nanoscale. *Proc. Nat. Acad. Sci. USA*, 108:14097–14102, 2011.
- <sup>44</sup> F. Ferrone. Analysis of protein aggregation kinetics. *Methods in Enzymol.*, 309:256–274, 1999.
- <sup>45</sup> Anthony L. Fink. The aggregation and fibrillation of  $\alpha$ -synuclein. *Accounts of Chemical Research*, 39(9):628–634, 2006. PMID: 16981679.
- <sup>46</sup> D. Frenkel and B. Smit. *Understanding molecular simulation. From algorithms to applications. Second edition.* Academic Press, Boston, 2002.
- <sup>47</sup> Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications.* Elsevier, 2002.
- <sup>48</sup> C. Galvagnion, A. K. Buell, Georg Meisl, T. C. T. Michaels, M. Vendruscolo, T. P. J. Knowles, and C. M Dobson. Lipid vesicles trigger  $\alpha$ -synuclein aggregation by stimulating primary nucleation. *Nature Chemical Biology*, 11:229–234, 2015.
- <sup>49</sup> C. J. Geyer and E. A. Thompson. Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *J. Am. Stat. Assoc.*, 90:909–920, 1995.
- <sup>50</sup> J E Gillam and C E MacPhee. Modelling amyloid fibril formation kinetics: mechanisms of nucleation and growth. *J Phys: Condens Matter*, 25:373101, 2013.
- <sup>51</sup> S. C. Glotzer and M. J. Solomon. *Nat. Mater.*, 6:557, 2007.

- <sup>52</sup> S. C. Glotzer and M. J. Solomon. . *Nat. Mater.*, 6:557, 2007.
- <sup>53</sup> S. C. Glotzer, M. J. Solomon, and N. A. Kotov. . *AIChE Journal*, 50:2978, 2004.
- <sup>54</sup> Mark Goulian. Inclusions in membranes. *Current Opinion in Colloid and Interface Science*, 1(3):358 – 361, 1996.
- <sup>55</sup> J. D. Gunton, A. Shiryayev, and D. L. Pagan. *Protein Condensation: Kinetic Pathways to Crystallization and Disease*. Cambridge University Press, 2007.
- <sup>56</sup> James D. Gunton, A. Shiryayev, and D. Pagan. *Protein Condensation: Kinetic Pathways to Crystallization and Disease*. Cambridge University Press: Cambridge, 2007.
- <sup>57</sup> N. Harn, C. Allan, C. Oliver, and C.R. Middaugh. Highly concentrated monoclonal antibody solutions: Direct analysis of physical structure and thermal stability. *Journal of Pharmaceutical Sciences*, 96:532–546, 2006.
- <sup>58</sup> S. Hernandez-Navarro, J. Iñes-Mullol, F. Sagues, and P. Tierno. . *Langmuir*, 28:5981, 2012.
- <sup>59</sup> T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, and A. Maritan. *Proc. Natl. Acad. Sci*, 101:7960, 2004.
- <sup>60</sup> L. Hong, A. Cacciuto, E. Luijten, and S. Granick. . *Langmuir*, 24:621, 2008.
- <sup>61</sup> J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell Jr. Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14:71–73, 2017.

- <sup>62</sup> A. H. Huisman, P. G. Bolhuis, and A. Fasolino. Phase transitions to bundles of flexible supramolecular polymers. *Phys. Rev. Lett.*, 100:188301, 2008.
- <sup>63</sup> Philippe H. Hunenberger. *Thermostat Algorithms for Molecular Dynamics Simulations*, pages 105–149. Springer Berlin Heidelberg, 2005.
- <sup>64</sup> Chariya Kaewsaneha, Pramuan Tangboriboonrat, Duangporn Polpanich, Mohamed Eissa, and Abdelhamid Elaissari. Janus colloidal particles: Preparation, properties, and biomedical applications. *ACS Applied Materials & Interfaces*, 5(6):1857–1869, 2013.
- <sup>65</sup> K. Kar, M. Jayaraman, B. Sahoo, R. Kodali, and R. Wetzel. *Nat Struct Mol Biol*, pages 328–336, 2011.
- <sup>66</sup> D. Kashchiev and S. Auer. Nucleation of amyloid fibrils. *J. Chem. Phys.*, 132:215101, 2010.
- <sup>67</sup> D. Kashchiev, R. Cabriolu, and S. Auer. Confounding the paradigm: peculiarities of amyloid fibril nucleation. *J. Am. Chem. Soc.*, 135:1531, 2013.
- <sup>68</sup> N. Kern and D. Frenkel. *J. Chem. Phys.*, 118:9882, 2003.
- <sup>69</sup> N. Kern and D. Frenkel. . *J. Chem. Phys.*, 118:9882, 2003.
- <sup>70</sup> Sagar D. Khare, Feng Ding, Kenneth N. Gwanmesia, and Nikolay V. Dokholyan. Molecular origin of polyglutamine aggregation in neurodegenerative diseases. *PLoS Computational Biology*, 1(3):231–235, 2005.



- <sup>71</sup> T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson. *Science*, 326:1533, 2009.
- <sup>72</sup> H. N. W. Lekkerkerker, W. C. Poon, P. N. Pusey, A. Stroobants, and P. B. Warren. . *Europhys. Lett.*, 20:559, 1992.
- <sup>73</sup> J. Li, R. Rajagopalan, and J. Jiang. Polymer-induced phase separation and crystallization in immunoglobulin g solutions. *J. Chem. Phys.*, 128:205105, 2008.
- <sup>74</sup> W. Li, Chakrabarti, and J. D. Gunton. *Langmuir*, 29:4470, 2013.
- <sup>75</sup> W. Li, Y. Liu, T. Perez, J. D. Gunton, C. M. Sorensen, and A. Chakrabarti. *Biophys. J.*, 101:2501, 2011.
- <sup>76</sup> Wei Li, J. D. Gunton, Siddique J. Khan, J. K. Schoelz, and A. Chakrabarti. Brownian dynamics simulation of insulin microsphere formation from break-up of a fractal network. *J. Chem. Phys.*, 134:024902, 2011.
- <sup>77</sup> Wei Li, Ya Liu, Toni Perez, James D. Gunton, C. M. Sorensen, and Amit Chakrabarti. Kinetics of Nanochain Formation in a Simplified Model of Amelogenin Biomacromolecules. *Biophysical Journal*, 101:2502, 2011.
- <sup>78</sup> X. Li, J. D. Gunton, and A. Chakrabarti. *J. Chem. Phys.*, 131:115101, 2009.
- <sup>79</sup> Jun Liu, Mary D.H. Nguyen, James D. Andya, and Steven J. Shire. Reversible self-association increases the viscosity of a concentrated monoclonal antibody in aqueous solution. *Journal of Pharmaceutical Sciences*, 94:1928–1940, 2005.

- <sup>80</sup> A. Lomakin, N. Asherie, and G. B. Benedek. Aeolotopic interactions of globular proteins. *Proc. Natl. Acad. Sci. USA*, 96:9465–9468, 1999.
- <sup>81</sup> A. D. Lomakin, S. Chung, G. B. Benedek, D. A. Kirschner, and D.P. Teplow. *Proc. Natl. Acad. Sci. USA*, 93:1125, 1996.
- <sup>82</sup> J.A. Luiken and P.G. Bolhuis. *J Phys Chem B*, 119:12568, 2015.
- <sup>83</sup> J.A. Luiken and P.G. Bolhuis. *Phys Chem Chem Phys*, 17:10556, 2105.
- <sup>84</sup> V. N. Manoharan, M. T. Elsesser, and D. J. Pine. . *Science*, 301:483, 2003.
- <sup>85</sup> S.-J. Marrink and S. Marcelja. Potential of mean force computations of ions approaching a surface. *Langmuir*, 17:7929–7934, 2001.
- <sup>86</sup> Zak Martinez, Min Zhu, Shubo Han, and Anthony L. Fink. Gm1 specifically interacts with -synuclein and inhibits fibrillation. *Biochemistry*, 46(7):1868–1877, 2007.
- <sup>87</sup> Antonello Merlino, Luciana Esposito, and Luigi Vitagliano. Polyglutamine repeats and the beta-helix structure: Molecular dynamics study. *Proteins Structure Function and Bioinformatics*, 63(4):918–927, 2006.
- <sup>88</sup> N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. N. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.
- <sup>89</sup> T. C. T. Michaels, S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles. *Phys. Rev. Lett.*, 116:038101, 2016.

- <sup>90</sup> T. C. T. Michaels, L. X. Liu, G. Meisl, and T. P. J. Knowles. Physical principles of filamentous protein self-assembly kinetics. *J. Phys.: Condens. Mat.*, 29:153002, 2017.
- <sup>91</sup> Cayla M. Miller, Young C. Kim, and Jeetain Mittal. Protein composition determines the effect of crowding on the properties of disordered proteins. *Biophysical Journal*, 111(1):28 – 37, 2016.
- <sup>92</sup> A. P. Minton. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.*, 276:10577–10580, 2001.
- <sup>93</sup> A. Morriss-Andrews and Joan-Emma Shea. *J. Phys. Chem. Lett.*, 5:1899, 2014.
- <sup>94</sup> Miki Nakano, Hirofumi Watanabe, E.B. Starikov, Stuart M. Rothstein, and Shigenori Tanaka. Mutation effects on structural stability of polyglutamine peptides by molecular dynamics simulation. *Interdisciplinary Sciences: Computational Life Sciences*, 1(1):21–29, 2009.
- <sup>95</sup> Roald Nezlin. Interactions between immunoglobulin g molecules. *Immunology Letters*, 132:1–5, 2010.
- <sup>96</sup> Hiroshi Noguchi. Membrane simulation models from nanometer to micrometer scale. *Journal of the Physical Society of Japan*, 78(4):041007, 2009.
- <sup>97</sup> E. G. Noya, C. Vega, J. P. K. Doye, and A. A. Louis. *J. Chem. Phys.*, 127:054501, 2007.

- <sup>98</sup> Hajime Ogawa, Miki Nakano, Hirofumi Watanabe, E.B. Starikov, Stuart M. Rothstein, and Shigenori Tanaka. Molecular dynamics simulation study on the structural stabilities of polyglutamine peptides. *Computational Biology and Chemistry*, 32(2):102 – 110, 2008.
- <sup>99</sup> R. K. Pathria. *Statistical Mechanics*. Elsevier, 2005.
- <sup>100</sup> A. B. Pawar and I. Kretzschmar. . *Macromol. Rapid Commun.*, 31:150, 2010.
- <sup>101</sup> T. Perevozchikova, C.B. Stanley, H.P. McWilliams-Koeppen, E.L Rowe, and V. Berthelie. *Biophys J*, 107:411–421, 2014.
- <sup>102</sup> J. R. Perilla and K. Schulten. Physical properties of the hiv-1 capsid from all-atom molecular dynamics simulations. *Annu. Rev. of Cell and Dev. Bio.*, 8, 2017.
- <sup>103</sup> S. Plimpton. *J Comp Phys*, 117:1–19, 1995.
- <sup>104</sup> Z. Preisler, T. Vissers, G. Munao’, F. Smalenburg, and F. Sciortino. . *Soft Matter*, 10:512, 2014.
- <sup>105</sup> S. C. Bae Q. Chen and S. Granick. . *Nature*, 469:361, 2011.
- <sup>106</sup> C J Roberts. *J Phys Chem B*, 107:1194, 2003.
- <sup>107</sup> Kenneth H. Roux and Donald L. Tankersley. A view of the human idiotypic repertoire. electron microscopic and immunologic analyses of spontaneous idiotype-anti-idiotype dimers in pooled human igg. *Journal of Immunology*, 144:1387–1395, 1990.

- <sup>108</sup> Kiersten M. Ruff, Tyler S. Harmon, and Rohit V. Pappu. Camelot: A machine learning approach for coarse-grained simulations of aggregation of block-copolymeric protein sequences. *The Journal of Chemical Physics*, 143(24):243123, 2015.
- <sup>109</sup> S. Sacanna and D. J. Pine. . *Curr. Opin. Colloid Interface Sci.*, 16:96, 2011.
- <sup>110</sup> J M Sanchez-Ruiz. *Biophys Chem*, 148:1–15, 2010.
- <sup>111</sup> A. Saric and A. Cacciuto. Self-assembly of nanoparticles adsorbed on fluid and elastic membranes. 9:667–6695, 2013.
- <sup>112</sup> F. Sciortino, S. Buldyrev, C. De Michele, N. Ghofraniha, E. La Nave, A. Moreno, S. Mossa, P. Tartaglia, and E. Zaccarelli. *Comput. Phys. Commun.*, 169:166, 2005.
- <sup>113</sup> F. Sciortino, A. Giacometti, and G. Pastore. . *Phys. Chem. Chem. Phys.*, 12:11869, 2010.
- <sup>114</sup> F. Sciortino and P. Tartaglia. *Adv. Phys.*, 54:471, 2005.
- <sup>115</sup> R. P. Sear. Phase behavior of a simple model of globular proteins. *J. Chem. Phys.*, 111:4800–4806, 1999.
- <sup>116</sup> L C Serpell, M Sunda, and C C C F Blake. *Cell. Mol. Life Sci.*, 53:8761–887, 1997.
- <sup>117</sup> David E. Shaw, J. P. Grossman, Joseph A. Bank, Brannon Batson, J. Adam Butts, Jack C. Chao, Martin M. Deneroff, Ron O. Dror, Amos Even, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian

Greskamp, C. Richard Ho, Douglas J. Ierardi, Lev Iserovich, Jeffrey S. Kuskin, Richard H. Larson, Timothy Layman, Li-Siang Lee, Adam K. Lerer, Chester Li, Daniel Killebrew, Kenneth M. Mackenzie, Shark Yeuk-Hai Mok, Mark A. Moraes, Rolf Mueller, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot, John K. Salmon, Daniele P. Scarpazza, U. Ben Schafer, Naseer Siddique, Christopher W. Snyder, Jochen Spengler, Ping Tak Peter Tang, Michael Theobald, Horia Toma, Brian Towles, Benjamin Vitale, Stanley C. Wang, and Cliff Young. Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '14, pages 41–53, Piscataway, NJ, USA, 2014. IEEE Press.

- <sup>118</sup> A. Shirayayev, D. L. Pagan, J. D. Gunton, D. S. Rhen, A. Saxena, and T. Lookman. Role of solvent for globular proteins in solution. *Virtual Jour. Biol. Phys.*, 2005.
- <sup>119</sup> N. Slepko, A. M. Bhattacharyya, G. R. Jackson, J. S. Steffan, J. L. Marsh, L. M. Thompson, and R. Wetzel. Normal-repeat-length polyglutamine peptides accelerate aggregation nucleation and cytotoxicity of expanded polyglutamine proteins. *Proc. Natl. Acad. Sci. USA*, 103:14367–14372, 2006.
- <sup>120</sup> C.B. Stanley, T. Perevozchikova, and V. Berthelie. *Biophys J*, 100:2504–2512, 2011.
- <sup>121</sup> Donald L. Tankersley, M. Sue Preston, and J.S. Finlayson. Immunoglobulin g dimer: An idiotype-anti-idiotype complex. *Molecular Immunology*, 25:41–48, 1988.

- <sup>122</sup> P. R. tenWolde and D. Frenkel. . *Science*, 277:1975, 1997.
- <sup>123</sup> V. Trappe and P. Sankhuhleri. *Curr. Top. Colloid Interface Sci*, 8:494, 2004.
- <sup>124</sup> A. van Blaaderen. . *A. van Blaaderen*, 439:545, 2006.
- <sup>125</sup> W. F. van Gunsteren and H. J. Berendsen. *Mol. Phys.*, 45:637, 1982.
- <sup>126</sup> Maddalena Venturoli, Maria Maddalena Sperotto, Marieke Kranenburg, and Berend Smit. Mesoscopic models of biological membranes. *Physics Reports*, 437(1):1 – 54, 2006.
- <sup>127</sup> L. Verlet. Computer experiments on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, 159:98, 1967.
- <sup>128</sup> J. Teun Vissers, Zdenek Preisler, Frank Smalenburg, Marjolein Dijkstra, and Francesco Sciortino. . *J. Chem. Phys.*, 138:164505, 2013.
- <sup>129</sup> A. Vrij. Polymers at Interfaces and the Interactions in Colloidal Dispersions. *Pure Appl. Chem.*, 48:471, 1976.
- <sup>130</sup> D. M. Walsh and D. J. Selkoe. *J. Neurochem*, 101:1172, 2007.
- <sup>131</sup> Y. Wang, A. Lomakin, R. F. Latypov, J. P. Laubach, T. Hideshima, P. G. Richardson, N. C. Munshi, K. C. Anderson, and G. B. ; Benedek. Phase transitions in human igg solutions. *J. Chem. Phys.*, 139:121904, 2013.
- <sup>132</sup> Y. Wang and J. A. McCammon. *The Blue Waters Super-System for Super-Science*. Sitka Publications, 2012.

- <sup>133</sup> Y. Wang and J. A. McCammon. *Introduction to molecular dynamics: theory and applications in biomolecular modeling*. Springer-Verlag, 2012.
- <sup>134</sup> R. Wetzol. *Cell*, 186:699–702, 1996.
- <sup>135</sup> A. W. Wilber, J. P. K. Doye, A. A. Lewis, E. G. Nova, M. A. Miller, and P. Wong. *J. Chem. Phys.*, 127:085106, 2007.
- <sup>136</sup> T. H. Zhang, J. Klok, R. H. Tromp, J. Groenewold, and W. K. Kegel. *Soft Matter*, 8:667, 2012.
- <sup>137</sup> Z. Zhang and S. C. Glotzer. *Nano Lett.*, 4:1407, 2004.
- <sup>138</sup> H. X. Zhou, G. Rivas, and A. P. Minton. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.*, 37:375–397, 2008.
- <sup>139</sup> Jian Zhu and Mark C. Hersam. Assembly and electronic applications of colloidal nanomaterials. *Advanced Materials*, 29(4):1603895–n/a, 2017. 1603895.
- <sup>140</sup> Min Zhu and Anthony L. Fink. Lipid binding inhibits  $\alpha$ -synuclein fibril formation. *Journal of Biological Chemistry*, 278(19):16873–16877, 2003.



# Curriculum Vitae

## Jason L. Haaga

Jason Haaga was born in Hartford, CT and is a dual citizen of the United States and Finland. He has random walked around the northeastern US and settled for now in rural Berks county.

### Education

PhD in Physics, Lehigh University	2017
Advisor: James D. Gunton	
MS in Physics, Lehigh University	2015
BS in Physics, Kutztown University of Pennsylvania	2013
<i>Summa Cum Laude</i>	

### Research Experience

Lehigh University - Bethlehem, PA, USA	Jan 2014 - Aug 2017
Drexel University - Philadelphia, PA, USA	Mar 2005 - Sep 2007

## Teaching Experience

Recitation Instructor PHY 021 (Intro. Physics 2)	Spring 2017
Recitation Instructor PHY 011 (Intro. Physics 1)	Falls 2014, 2015
Laboratory Instructor PHY 022 (Intro. Physics 2)	Fall 2013 - Spring 2014

## Service

Graduate Student Senate	
Physics Department Representative	Fall 2016 - Spring 2017
Graduate Research Committee	
Graduate Student Representative	Fall 2016 - Spring 2017

## Organizations

Physics Honors Society: Sigma Pi Sigma	2012
Adult Student Honors Society: Alpha Sigma Lambda	2012
American Physical Society	2014
Biophysical Society	2014

## Publications

1. J. Haaga, C. N. Buckles, and J. D. Gunton, “Early stage aggregation of a coarse-grained model of polyglutamine”, in review (2017)
2. J. Haaga, E. Pemberton, J.D. Gunton, and J.M. Rickman, “Phase diagram of a model of the protein amelogenin”, *J. Chem. Phys.*, 145(8):085105 (2016)

3. S. Khan, J. Haaga, and J.D. Gunton, “Kinetics of aggregation of an anisotropic model of self-assembling molecules”, *J. Chem. Phys.*, 143(2):024906 (2015)
4. A. Leonard, D. M. Goldberg, J. L. Haaga, and R. Massey, “Gravitational Shear, Flexion and Strong Lensing in Abell 1689”, *Astrophys. J.* 666:51-63 (2007)

## **Presentations**

1. “Molecular dynamics study of early stage kinetics of polyglutamine aggregation”
  - BPS Annual Meeting 2017. New Orleans, LA, USA
  - BPS Mid-Atlantic Meeting 2016. Bethlehem, PA, USA
2. “Effects of Coulomb repulsion on the phase diagram of the Asakura-Oosawa model”
  - APS March Meeting 2016. Baltimore, MD, USA
3. “Simulation of the Effects of Flanking Sequences on Polyglutamine Aggregation”
  - APS Mid-Atlantic Meeting 2014. State College, PA, USA