View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by K-State Research Exchange

Statistical mechanics models in protein association problems

by

Nelson Javier Ramallo

Lic., University of Cordoba, Argentina, 2014

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Abstract

Protein-Protein interactions can lead to disordered states such as precipitates or gels, or to ordered states such as crystals or microtubules. In order to study the different natures of protein-protein interactions we have developed statistical mechanics models in order to interpret the varied behavior of different protein systems. The main point will be to develop theoretical models that infer the time a length scales that characterize the dynamics of the systems analyzed. This approach seek to facilitate a connection to simulations and experiments, where a high resolution analysis in length and time is possible, since the theories can provide insights about the relevant time and length scales, and also about issues that can appear when studying these systems.

The first system studied is monoclonal antibodies in solution. Antibody solutions deviate from the dynamical and rheological response expected for globular proteins, especially as volume fraction is increased. Experimental evidence shows that antibodies can reversibly bind to each other via F_{ab} and F_c domains, and form larger structures (clusters) of several antibodies. Here we present a microscopic equilibrium model to account for the distribution of cluster sizes. Antibody clusters are modeled as polymers that can grow via reversible bonds either between two F_{ab} domains or between a F_{ab} and a F_c. We propose that the dynamical and rheological behavior is determined by molecular entanglements of the clusters. This entanglement does not occur at low concentrations where antibody-antibody binding contributes to the viscosity by increasing the effective size of the particles. The model explains the observed shear-thinning behavior of antibody solutions.

The second system is protein condensates inside living cells. Biomolecule condensates appear throughout the cell serving a wide variety of functions, but it is not clear how functional properties show in the concentrated network inside the condensate droplets. Here we model disordered proteins as linear polymers formed by “stickers” evenly spaced by “spacers”.

The spacing between stickers gives rise to different network toplogies inside the condensate droplet, determining distinguishing properties such us density and client binding.

The third system is protein-protein binding in a salt solutions. Biomolecular simulations are typically performed in an aqueous environment where the number of ions remains fixed for the duration of the simulation, generally with a number of salt pairs intended to match the macroscopic salt concentration. In contrast, real biomolecules experience local ion environments where the salt concentration is dynamic and may differ from bulk. We develop a statistical mechanics model to account for fluctuations of ions concentrations, and study how it affects the free energy of protein-protein binding.

Statistical mechanics models in protein association problems

by

Nelson Javier Ramallo

Lic., University of Cordoba, Argentina, 2014

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Approved by:

Major Professor
Jeremy Schmit

Copyright

© Nelson Javier Ramallo 2020.

Abstract

Protein-Protein interactions can lead to disordered states such as precipitates or gels, or to ordered states such as crystals or microtubules. In order to study the different natures of protein-protein interactions we have developed statistical mechanics models in order to interpret the varied behavior of different protein systems. The main point will be to develop theoretical models that infer the time a length scales that characterize the dynamics of the systems analyzed. This approach seek to facilitate a connection to simulations and experiments, where a high resolution analysis in length and time is possible, since the theories can provide insights about the relevant time and length scales, and also about issues that can appear when studying these systems.

The first system studied is monoclonal antibodies in solution. Antibody solutions deviate from the dynamical and rheological response expected for globular proteins, especially as volume fraction is increased. Experimental evidence shows that antibodies can reversibly bind to each other via F_{ab} and F_c domains, and form larger structures (clusters) of several antibodies. Here we present a microscopic equilibrium model to account for the distribution of cluster sizes. Antibody clusters are modeled as polymers that can grow via reversible bonds either between two F_{ab} domains or between a F_{ab} and a F_c. We propose that the dynamical and rheological behavior is determined by molecular entanglements of the clusters. This entanglement does not occur at low concentrations where antibody-antibody binding contributes to the viscosity by increasing the effective size of the particles. The model explains the observed shear-thinning behavior of antibody solutions.

The second system is protein condensates inside living cells. Biomolecule condensates appear throughout the cell serving a wide variety of functions, but it is not clear how functional properties show in the concentrated network inside the condensate droplets. Here we model disordered proteins as linear polymers formed by “stickers” evenly spaced by “spacers”.

The spacing between stickers gives rise to different network toplogies inside the condensate droplet, determining distinguishing properties such us density and client binding.

The third system is protein-protein binding in a salt solutions. Biomolecular simulations are typically performed in an aqueous environment where the number of ions remains fixed for the duration of the simulation, generally with a number of salt pairs intended to match the macroscopic salt concentration. In contrast, real biomolecules experience local ion environments where the salt concentration is dynamic and may differ from bulk. We develop a statistical mechanics model to account for fluctuations of ions concentrations, and study how it affects the free energy of protein-protein binding.

Table of Contents

List of Figures vii

List of Tables viii

Acknowledgements ix

1 Introduction 1

  1.1 Introduction 1

  1.2 Background 2

    1.2.1 Basics in Polymer Theory 2

    1.2.2 Polymer Rheology 6

  1.3 Systems of Interest 13

Bibliography 15

2 Cluster formation and entanglement in the rheology of antibody solutions 16

  2.1 Introduction 16

  2.2 Antibody cluster morphology depends on the location of binding sites 17

    2.2.1 Head-to-Head binding results in linear aggregates 17

    2.2.2 Head-to-Tail binding permits the formation of branched clusters . . . 19

    2.2.3 The average cluster size increases with binding affinity and protein concentration 21

  2.3 The viscosity of dilute solutions is determined by the volume occupied by antibody clusters 24

  2.4 Equilibrium Dynamics of antibody solutions 28

2.5 Shear thinning results from the release of entanglements 29

2.6 The binding affinity predicts both the zero shear viscosity and the onset of shear thinning 31

2.7 Summary 34

Bibliography 36

3 Static properties in biological condensates determined by topology of the dense phase 39

3.1 Introduction 39

3.2 Free Energy 41

3.3 Concentration in condensate phase 45

3.4 Client Binding 50

3.5 Conclusions 55

3.6 Methods 56

3.6.1 Excluded Volume Free Energy 56

3.6.2 Calculation of Entropy 56

3.6.3 Free energy minimization 57

3.6.4 Large Valence 58

Bibliography 61

4 Protein-Protein binding under macroscopic salt conditions 63

4.1 Introduction 63

4.2 Theory 64

4.3 Conclusions 68

Bibliography 69

List of Figures

1.1 (a) Monomers; (b) Polymer chain 3

1.2 Freely jointed chain model for a polymer. 3

1.3 The Dumbbell model. 9

1.4 (a) Entanglements points, (b) Tube model 11

2.1 Cartoon and concentration for the monomer, dimer, and trimer states in the HH model. 18

2.2 Cartoon of the monomer, dimers, and trimers described by the HT model (Eq. 2.8). As discussed in the Supporting Information, several of these structures are related by rotational symmetry and, therefore, are over-represented in the partition function q_{HT}. The corrected partition function $\tilde{q}_{\mathrm{HT}}$ removes this over-counting. 21

2.3 Average number of monomers per cluster for the HH model (blue) and the HT model (red) at two different concentrations: 100mg/mL (solid lines) and 200mg/mL (dotted lines) at the same binding energy $\epsilon_{\mathrm{HH}} = \epsilon_{\mathrm{HT}} = \epsilon$. 23

2.4 Behavior of the relative concentrations for HH model (A) and HT model (B) as a function of binding energy $\epsilon_{\mathrm{HH}} = \epsilon_{\mathrm{HT}} = \epsilon$ at different concentrations: 50mg/mL (solid lines), 150mg/mL (dotted lines). At low binding energies monomers dominate, but as the binding energy increases the monomer concentration decreases monotonically as larger structures start to form. Clusters larger than the monomer have a non-monotonic concentration because the large clusters formed at high binding energies consume the pool of smaller clusters. 24

2.5 Comparison of Eq. 2.21 to the viscosities measured in [11] (left), and [12] (right). Binding parameters are fit to the first three data points ($c \leq 20$ mg/ml). The increasing deviation at higher concentrations, especially in panel A, indicates that the antibodies are exiting the dilute regime and becoming entangled. 26

2.6 Comparison of Eq. 2.30 to the shear dependent viscosity of an antibody at A) 200 mg/ml [7], B) 150 mg/ml [7], and a different molecule at C) 120 mg/ml [28], and D) 146 mg/ml [28]. Solutions with lower viscosity deviate more strongly from the theory due to the fact smaller complexes are less polymer-like. 33

3.1 Looping entropy. 43

3.2 P_{SC} Eq. 3.12 behavior as a function of linker length for different binding energies for bivalent case. 47

3.3 Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$. 48

3.4 Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$. 49

3.5 Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different binding energies. 50

3.6 Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$. 53

3.7 Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$. 53

3.8 Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$. 54

3.9 Partition coefficient of clients P_{cl} Eq. 3.20 behavior as a function of free type "B" molecules c_{1B} for different binding energies. 55

4.1 Modeling the protein binding process as having two separate events: 1. Two boxes with charge q each represent the proteins separate by an infinite distance. 2. there is one box with charge $2q$ and one empty box, representing having the two proteins bound to each other. 66

4.2 ΔF behavior as a function of the solute charge Q for the cases with and without fluctuations. We see that as the charge gets larger the difference in energy also increases. 67

4.3 ΔF behavior as a function of box length R for the cases with and without fluctuations. We see a non monotonic behavior in both cases, and the case with fluctuations is always larger than the case without them. 68

List of Tables

2.1 Fitted free energies for the dilute antibody solutions plotted in Fig. 2.5. Also shown are the overlap concentration, at which Eq. 2.21 is expected to fail, and the average cluster size at 10 and 10 mg/ml. 27

2.2 Parameters extracted from the fits in Fig. 2.6, including the dimerization equilibrium constant and the global proportionality constant. Also shown is the average cluster size $\langle n \rangle$, which increases with the equilibrium constant. 34

Acknowledgments

First of all, I would like to acknowledge my major professor Dr. Jeremy Schmit for all the support and guidance along all these years at Kansas State University. I would like to thank Professors Chris Sorensen, Bret Flanders and Paul Smith for accepting my request to be part of my committee and providing valuable suggestions during my research.

I am very much thankful to department of Physics at Kansas State University for selecting me as a graduate student to pursue my PhD degree.

I want to thank my family for their unconditional support.

Chapter 1

Introduction

1.1 Introduction

Protein-protein interactions occur in many situations inside cells or any living organism. Different systems can behave in a variety of ways depending the type of protein and the environment in which it is immerse, and the outcomes can be ordered, such as in crystals, or disordered, such as in gels or aggregates. Protein-protein interactions handle a wide range of biological processes, including cell-to-cell interactions and metabolic and developmental control [1]. Protein-protein interactions pose a challenge, and its study is of interest to many fields in science and industry. Usually, when it comes to protein-protein interactions it is meant to be thought as a two body problem, as in enzyme-substrate problems, however we are interested in many-bodies systems, where it is not clear the connection to the two-body case. We will study systems that have properties that arise inherently from them being many-body systems, that can not be boil down to two-body problems. Our approach implies the development of theoretical models that can tell us what is the important physics within them, and seek connections to experiments and simulations, that are high-resolution in nature.

The first challenge that protein-protein interaction presents is its spatial and time scales: computer simulations offer high resolution details of the interactions between proteins, but it is limited to small systems and short time scales; on the other hand, analytic theories give low resolution models, but they can handle longer time scales and give information regarding the underneath principles governing the phenomenon.

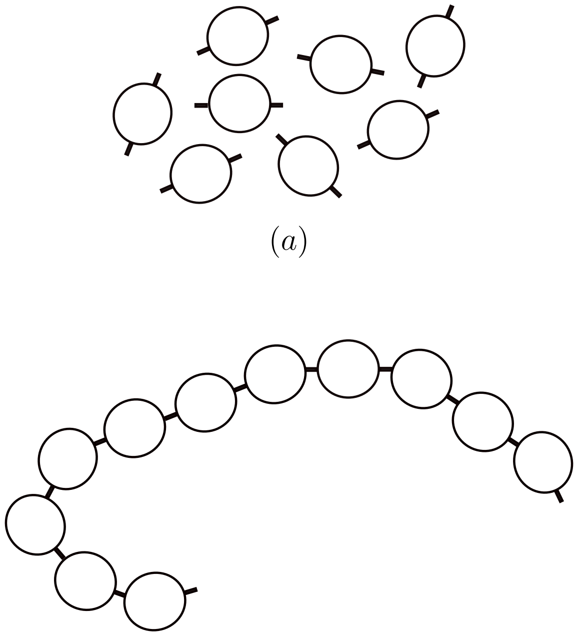
In this work we show how analytic theories can prove a tool to better target the main points of experiments and simulation for in order to understand the the interactions between proteins. Analytic models can also gives information regarding the main time scales of the process and help coarse-grain simulations to those time frames. Analytic theories can prove to be general but can also target the specifics a particular biological system, making the transition between general physical principles to the particulars a specific molecule. The overall goal is to use statistical mechanics theories to help make better choices when going to the high resolution methods of simulations.

1.2 Background

The next question is what the best representation of protein would be. We can represent proteins as colloids or as polymers. Proteins behave like string polymers given the nature of its backbone in multiple scales, so we will proceed to discuss polymer theory, and compare the rheological behavior of proteins to that of colloids.

1.2.1 Basics in Polymer Theory

Polymers are long string-like molecules made of certain chemical units called monomers. The monomers are connected sequentially as shown in Fig. 1.1. The number of monomers in a polymer is typically several thousand, and can be as large as tens of millions. Polymers are indispensable materials in modern technology; they are used as plastics, rubbers, films, and textiles.



(a)

(b)

Figure 1.1: *(a) Monomers; (b) Polymer chain*

Polymers are also the basic molecules of life. The machinery of life is realized by proteins which are natural polymers made of amino acids. The genetic information of life is inscribed in another important class of bio-polymer, DNA.

One commonly used model to represent polymers is the freely jointed chain model shown in Fig. 1.2, where a polymer molecule is represented by N segments of constant length b. The segments are connected by flexible joints and can point in any direction independently of other segments.

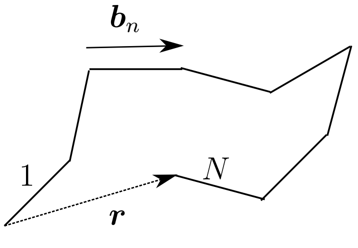


Figure 1.2: *Freely jointed chain model for a polymer.*

If there is no external forces acting on the chain ends, the distribution of the vector $\boldsymbol{r}$ is isotropic. Therefore the average $\langle \boldsymbol{r} \rangle$ is equal to zero. Let us therefore consider the average $\langle \boldsymbol{r}^2 \rangle$.

Let $\boldsymbol{b}_n$ $(n = 1, 2, \ldots, N)$ be the end-to-end vector of the n-segment. Then $\boldsymbol{r}$ is written as

$$\boldsymbol{r} = \sum_{n=1}^{N} \boldsymbol{b}_n \tag{1.1}$$

Therefore the mean square of $\boldsymbol{r}$ is calculated as

$$\langle \boldsymbol{r}^2 \rangle = \sum_{n=1}^{N} \sum_{m=1}^{N} \langle \boldsymbol{b}_n \cdot \boldsymbol{b}_m \rangle \tag{1.2}$$

For a random distribution of segments, the $\boldsymbol{b}_n$ are independent of each other. Thus

$$\langle \boldsymbol{b}_n \cdot \boldsymbol{b}_m \rangle = \delta_{nm} b^2 \tag{1.3}$$

Therefore

$$\langle \boldsymbol{r}^2 \rangle = \sum_{m=1}^{N} \langle \boldsymbol{b}_n^2 \rangle = N b^2 \tag{1.4}$$

The average size of a polymer in a force free state can be estimated by $\overline{r} = \sqrt{\langle \boldsymbol{r}^2 \rangle}$, which is equal to $\sqrt{N} b$. This is the length of an ideal chain, since we are not considering the excluded volume of each monomer. The next step is to to add this effect and compute the length of a real chain.

The conformations of a real chain are determined by the balance of the effective repulsion energy between monomers that tends to swell the chain and the entropy loss due to such deformation. One of the simplest models that captures the essence of this balance is the **Flory Theory**, which makes rough estimates of both the energetic and the entropic contributions to the free energy.

Consider again a polymer with N monomers, swollen to size $R > R_0 = \sqrt{N} b$. Flory theory assumes that monomers are uniformly distributed within the volume R^3 with no correlations between them. The probability of a second monomer being within the excluded volume v of a given monomer is the product of the excluded volume v and the number density of monomers in the pervaded volume of the chain N/R^3. The energetic cost of being excluded from this volume (the energy of excluded volume interaction) is kT per exclusion or $kTvN/R^3$ per monomer. For all N monomers in the chain, this energy is N times larger, then

$$F_{\text{int}} \sim kTv\frac{N^2}{R^3} \tag{1.5}$$

The Flory estimate of the entropic contribution to the free energy of a real chain is the energy required to stretch an ideal chain to end-to-end distance R (see [2])

$$S \sim kT\frac{R^2}{Nb^2} \tag{1.6}$$

The total free energy of a real chain in the Flory approximation is the sum of both contributions

$$F = F_{\text{int}} + S \sim kT\left(v\frac{N^2}{R^3} + \frac{R^2}{Nb^2}\right) \tag{1.7}$$

The minimum free energy of the chain (obtained by setting $\partial F/\partial R = 0$) gives the optimum size of the real chain in the Flory approximation, $R = R_F$

$$\begin{aligned} &\frac{\partial F}{\partial R} = 0 = kT\left(-3v\frac{N^2}{R_F^4} + 2\frac{R_F}{Nb^2}\right), \\ &R_F^5 \sim vb^2N^3 \\ &R_F \sim v^{1/5}b^{2/5}N^{3/5} \end{aligned} \tag{1.8}$$

The size of long real chains is much longer than that of ideal chains with the same number of monomers due to the 3/5 exponent.

It is important to realize that the Flory theory leads to a universal power law dependence of polymer size R on the number of monomers N:

$$R \sim N^{\nu} \tag{1.9}$$

The excluded volume only enters as a prefactor, but does not change the value of the scaling exponent ν for any $\upsilon > 0$. The Flory approximation of the scaling exponent is $\nu = 3/5$ for a swollen linear polymer. For the ideal linear chain the exponent is $\nu = 1/2$. More sophisticated theories lead to a more accurate estimate of the scaling exponent of the swollen linear polymer in three dimensions [3]:

$$\nu \cong 0.588 \tag{1.10}$$

which is very close to the Flory result $3/5 = 0.6$.

1.2.2 Polymer Rheology

The mechanical behavior of a simple elastic material (Hookean elastic material) is represented by the linear relation between strain and stress. This is represented by the linear relation between the shear strain γ and the shear stress σ:

$$\sigma = G\,\gamma \tag{1.11}$$

where the coefficient G is the shear modulus of the material.

On the other hand, the mechanical behavior of a simple fluid (Newtonian fluid) is represented by the linear relation between the shear rate $\dot{\gamma} = d\gamma(t)/dt$ and the shear stress σ

$$\sigma = \eta\,\dot{\gamma} \tag{1.12}$$

where η is the viscosity of the material.

Soft matter generally has both viscosity and elasticity, an attribute that is referred to as viscoelasticity.

Suppose now that when a time-dependent strain $\gamma_1(t)$ is applied to a material, the stress response is $\sigma_1(t)$, and that when $\gamma_2(t)$ is applied, the stress is $\sigma_2(t)$. The superposition principle states that when the superposed strain $\gamma(t) = \gamma_1(t) + \gamma_2(t)$ is applied, the stress response is $\sigma(t) = \sigma_1(t) + \sigma_2(t)$.

If the superposition principle holds, the stress response to any time-dependent strain $\gamma(t)$ can be expressed by the relaxation modulus $G(t)$. This is because any time-dependent strain $\gamma(t)$ can be regarded as a superposition of step strains $\Delta\gamma_i = \dot{\gamma}(t_i)\Delta t$ applied at time t_i , i.e.,

$$\gamma(t) = \sum_i \Delta\gamma_i \Theta(t - t_i) \tag{1.13}$$

The step strain $\Delta\gamma_1$ applied at time ti creates the stress $G(t - t_i)\Delta\gamma_i$ at a later time t. Summing up all the stresses created by the step strains applied in the past, we have the stress at time t:

$$\sigma(t) = \sum_i G(t - t_i)\Delta\gamma_i = \sum_i G(t - t_i)\dot{\gamma}(t_i)\Delta t \tag{1.14}$$

In the limit $\Delta t \to 0$, this gives

$$\sigma(t) = \int_{-\infty}^{t} dt' G(t - t')\dot{\gamma}(t') \tag{1.15}$$

Viscoelasticity which satisfies the superposition principle is called linear viscoelasticity. The linear viscoelasticity of incompressible materials is completely characterized by a single function $G(t)$. If $G(t)$ is given, the stress response for arbitrary strain can be calculated. For example, consider that a shear flow of constant shear rate $\dot{\gamma}$ is started at $t = 0$. The stress at time t is calculated as

$$\sigma = \int_0^t dt' G(t - t')\dot{\gamma} = \dot{\gamma}\int_0^t dt' G(t') \tag{1.16}$$

The shear stress increases monotonically, and approaches a constant value $\sigma(\infty)$. The steady state viscosity η_0 is defined by the ratio $\sigma(\infty)/\dot{\gamma}$. According to Eq. 1.12, the steady state viscosity is given by

$$\eta_0 = \int_0^\infty dt G(t) \tag{1.17}$$

The next step is to derive η_0 for different systems and concentrations. We start with a simple dilute colloidal system of hard spheres. Einstein [4] calculated the viscosity η_0 to be

$$\eta_0 = \eta_s \left(1 + \frac{5}{2}\phi\right) \tag{1.18}$$

where η_s is the viscosity of the solvent and ϕ is the volume fraction of spheres in the suspension.

We now continue by analyzing the case of a polymeric fluid. Polymeric fluids (polymer solutions and polymer melts) are generally viscoelastic. Suppose that a step shear is applied at time t = 0. When the shear strain is applied, polymer chains are distorted from their equilibrium conformation, which creates a stress. The stress has the same origin as that in rubbers and gels, i.e., it is due to the elasticity of the polymer chains. Unlike rubbers and gels, the polymer chains in polymeric fluids can recover their equilibrium conformation while the system is macroscopically deformed. As the polymer chains recover their equilibrium conformation, the stress decreases in time, and finally vanishes completely even though the material remains distorted. Therefore, the stress relaxation in polymeric fluids is a direct consequence of the conformational relaxation of polymer chains.

How the relaxation of polymer conformation takes place depends on whether the polymers are entangled or not. If the chains are not entangled, the problem is how a deformed chain placed in a viscous medium recovers the equilibrium conformation. If the chains are entangled, an entirely different thinking is needed.

We now discuss the viscoelasticity of polymer fluids for which the entanglement effect is not important. This is the case of dilute polymer solutions. To discuss the conformational dynamics, we consider the simple model shown in Fig. 1.3. Here a polymer molecule is represented by a dumbbell consisting of two segments connected by a spring. The segment represents half of the polymer molecule.

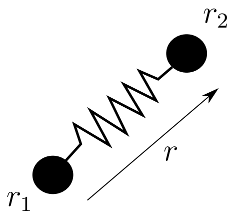


Figure 1.3: *The Dumbbell model.*

We assume that the medium surrounding the segments is a Newtonian fluid. This assumption is reasonable for dilute polymer solutions. Let $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ be the position vectors of the segments. Then the potential energy of the dumbbell is written as

$$U(\boldsymbol{r_1}, \boldsymbol{r_2}) = \frac{k}{2}(\boldsymbol{r_1} - \boldsymbol{r_2})^2 \tag{1.19}$$

where k is the spring constant of the dumbbell.

Following Doi et. al. [5] we can obtain the relaxation modulus

$$G(t) = G_0 e^{-t/\tau} \tag{1.20}$$

where the initial shear modulus G_0 is given by

$$G_0 = c\, k_B T \tag{1.21}$$

and the relaxation time is

$$\tau = \frac{\zeta}{4k} \tag{1.22}$$

where ζ is the fluid friction coefficient.

The next step is to represent the polymer by N beads. This is called the Rouse model. The chain in the Rouse model is represented as N beads connected by springs of root-mean-square size b. The beads in the Rouse model only interact through the connecting springs. Each bead is characterized by its own independent friction with friction coefficient ζ. Solvent is assumed to be draining through the chain as it moves.

The total friction coefficient of the whole Rouse chain is the sum of the contributions of each of the N beads:

$$\zeta_R = N\zeta \tag{1.23}$$

The diffusion coefficient of the Rouse chain is obtained from the Einstein relation [6]

$$D_R = \frac{k_B T}{\zeta_R} = \frac{k_B T}{N\zeta} \tag{1.24}$$

The polymer diffuses a distance of the order of its size during a characteristic time, called the Rouse time, τ_R

$$\tau_R \approx \frac{R^2}{D_R} \approx \frac{\zeta}{k_B T} N R^2 \tag{1.25}$$

On time scales shorter than the Rouse time, the chain exhibits viscoelastic modes that will not be discussed here, but on time scales longer than the Rouse time, the motion of the chain is simply diffusive.

For an ideal chain we have from Eq. 1.4 that

$$R^2 = N b^2 \tag{1.26}$$

hence

$$\tau_R \approx \frac{\zeta}{k_B T} b^2 N^2 \tag{1.27}$$

It can be seen that the relaxation modulus is given by [5]

$$G(t) = \frac{\sqrt{\pi}}{2} c k_B T \left(\frac{\tau_R}{t}\right)^{1/2} \tag{1.28}$$

Then, the steady state viscosity of the Rouse model is calculated as

$$\eta_0 = \int_0^\infty dt\, G(t) \approx c k_B T \zeta N^2 \tag{1.29}$$

The steady state viscosity is proportional to cN^2.

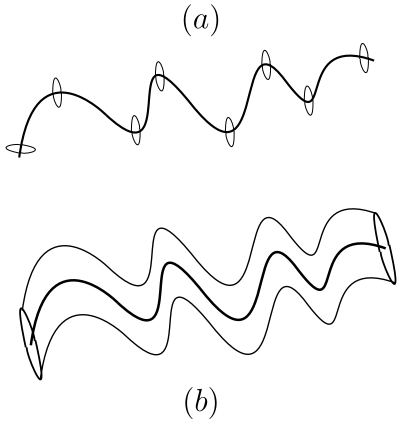


Figure 1.4: *(a) Entanglements points, (b) Tube model*

We now proceed to study the case in which entanglement points are present Fig. 1.4(a). The topological constraints imposed by neighboring chains on a given chain restrict its motion to a tube like region called the confining tube 1.4(b). The motion of the chain along the contour of the tube (the primitive path) is unhindered by topological interactions. Displacement of monomers in the direction perpendicular to the primitive path is restricted by surrounding chains to an average distance a, called the tube diameter. The number of monomers in a strand of size equal to the amplitude of transverse fluctuations (the tube diameter) is N_e, the number of monomers in an entanglement strand. For dilute solutions, excluded volume interactions are screened and the tube diameter is determined by ideal chain statistics:

$$a \approx b\sqrt{N_e} \tag{1.30}$$

The tube can be thought of as being composed of N/N_e sections of size a, with each section containing N_e monomers. The chain can be considered as either a random walk of entanglement strands (N/N_e strands of size a) or a random walk of monomers (N monomers of size b).

$$R \approx a\sqrt{\frac{N}{N_e}} \approx b\sqrt{N} \tag{1.31}$$

The average contour length $\langle L \rangle$ of the primitive path (the center of the confining tube, see Fig. 1.4(b)) is the product of the entanglement strand length a and the average number of entanglement strands per chain N/N_e

$$\langle L \rangle \approx a\frac{N}{N_e} \approx \frac{b^2 N}{a} \approx \frac{bN}{\sqrt{N_e}} \tag{1.32}$$

The average primitive path contour length $\langle L \rangle$ is shorter than the contour length of the chain bN by the factor $a/b \approx \sqrt{N_e}$.

In de Gennes' reptation model [7], an entangled chain diffuses along its confining tube in a way analogous to the motion of a snake or a worm. This motion of the chain consists of diffusion of small loops, along the contour of the primitive path. This curvilinear motion of a polymer along its tube satisfies the topological constraints imposed by surrounding chains and is characterized by the Rouse friction coefficient $N\zeta$. The curvilinear diffusion coefficient D_c that describes the motion of the chain along its tube is simply the Rouse diffusion coefficient of the chain

$$D_c = \frac{k_B T}{N\zeta} \tag{1.33}$$

The time it takes for the chain to diffuse out of the original tube of average length $\langle L \rangle$ is the reptation time

$$\tau_{\text{rep}} \approx \frac{\langle L \rangle^2}{D_c} \approx \frac{\zeta b^2}{k_B T}\frac{N^3}{N_e} = \frac{\zeta b^2}{k_B T} N_e^2 \left(\frac{N}{N_e}\right)^3 \tag{1.34}$$

Here, Eq. 1.32 was used for the average contour length of the tube. The reptation time is predicted to be proportional to the cube of the molar mass, compare this result to the Rouse time that is proportional to the square of the molar mass (Eq. 1.27).

It can be seen that the relaxation modulus is given by [5]

$$G(t) \approx G_0 e^{-t/\tau_{\mathrm{rep}}} \tag{1.35}$$

where the initial shear modulus is given by $G_0 = ck_BT$ Then, the steady state viscosity will be given by

$$\eta_0 \approx G_0 \tau_{\mathrm{rep}} \tag{1.36}$$

Since G_0 is independent of the molecular weight, the viscosity is also proportional to N^3.

Following Doi et al [5] the steady state viscosity in the semidilute state can be found to be given by

$$\eta_0 = \eta_s \left(\frac{c}{c^*}\right)^{3/(3\nu-1)} \tag{1.37}$$

where η_s is the solvent viscosity, $\nu \simeq 3/5$ is the Flory exponent, and $c^* \sim \langle L \rangle^{1-3\nu} b^{-3}$ is the overlap concentration. Then,

$$\eta \sim c^{3/(3\nu-1)} N^3 \tag{1.38}$$

1.3 Systems of Interest

In this work we will be focusing in three different systems where statistical mechanics models will prove to be a useful tool in the scales and order parameters that govern the structure and dynamics the particular system.

In Chapter 2 we will study a system of monoclonal antibodies in a semidilute solution. It will be shown that this antibodies can bind to each other and form elongated structures. We use polymer theory in order to study the solution's rheology and dynamics, and conclude that the relevant time scale is given by the reptation time of the cluster chains. This shows the time scale where experiments and simulation may be performed in order to observed the critical aspects of the system's behavior in high resolution. Experiments and simulations can next study the binding affinities between antibodies and extract the distribution of cluster sizes and shapes, and how it affects the rheology and dynamics; and it also allows us to extract the on and off rates in order to find the system's relaxation modes.

In Chapter 3 we focus in systems of proteins experiencing liquid-liquid phase separation. We apply polymer theory in order to obtain the phase diagrams and study the behavior inside the condensate phase. In this type of systems we are dealing with disorder protein where it is no clear what is the correct length scale that characterize the system; it may be mesh size of the network inside the condensate, the thermal blob size, the entanglement length. The main goal will be to extract the order parameter that governs the system's behavior. In previous works it was shown that for SPOP/DAXX [8] the order parameter is given by the ratio of N_{SPOP} to N_{DAXX}; for SUMO-SIM [9] the order parameter is the defect density in the network. The point is to choose an order parameter that describes the condensate's functionality. Once the order parameter is found, the length scale that characterizes the system's behavior emerges; this can then tell future simulations how to coarse-grain the system.

In Chapter 4 we study the protein-protein binding process, and how the addition of ions fluctuations affects it. The main point of this analytic model is to show the problems that may appear in high resolution approaches such as explicit ions simulations where the macroscopic salt concentration is fixed. This seems like a two-body problem at a first glance, however the solvent and ions give a many-body effect. The question we are trying to answer is when this pose a local problem or a general one.

Bibliography

[1] Braun, P. and Gingras, A.-C., PROTEOMICS **12** (2012) 1478.

[2] Rubinstein, M. and Colby, R., *Polymer Physics*, Oxford University Press, Oxford, U.K., 2003.

[3] McKenzie, D., Physics Reports **27** (1976) 35.

[4] Einstein, A., Annalen der Physik **324** (1906) 289.

[5] Doi, M., *The theory of polymer dynamics*, Clarendon Press, Oxford, 1986.

[6] Einstein, A., Annalen der Physik **322** (1905) 549.

[7] Gennes, *Scaling concepts in polymer physics*, Cornell University Press, Ithaca, N.Y, 1979.

[8] Schmit, J. D., Bouchard, J. J., Martin, E. W., and Mittag, T., Journal of the American Chemical Society **142** (2019) 874.

[9] Bhandari, K. and Schmit, J. D., pending of publication .

Chapter 2

Cluster formation and entanglement in the rheology of antibody solutions

2.1 Introduction

Efficient dosing of therapeutic antibodies often requires concentrations in excess of 100 or even 200 mg/ml [1–3]. However, many antibodies have a sharp rise in the viscosity that renders production and delivery prohibitive at these concentrations [2, 4–6]. Unfortunately, this problem is only apparent late in the development pipeline when it is not feasible to alter the sequence to reduce viscosity. A better approach would be to choose low viscosity target molecules early in the pipeline so that the problem can be avoided altogether. To achieve this goal it is necessary to understand how minor sequence perturbations within the complementarity determining regions (CDR) contribute to the many-body interactions responsible for the elevated viscosity.

The sharp rise in antibody viscosity has characteristics that are very different from solutions of rigid bodies [4]. First, the onset of the nonlinear regime occurs at volume fractions on the order of 5–10%, which is much less than the jamming transition for comparably shaped rigid bodies (e.g. 58% for spheres). Secondly, the viscosity of a given molecule correlates well with attractive intermolecular interactions [7–9]. While this is intuitively reasonable, the viscosity of a flocculated solution is primarily a function of the solute volume fraction. But, the volume fraction does not change upon aggregation, although entrained water cavities could account for a factor of 2 or 3 increase. To explain these discrepancies, we proposed an alternative model in which transient interactions between antigen binding domains result in long, flexible antibody complexes [10]. These complexes entangle with each other giving the solution viscosity characteristics of a semi-dilute polymer solution. In this paper we expand on this polymer model to explain shear thinning behavior, dilute solution viscosity, and show how the ensemble of complexes depends on the affinity and location of intermolecular interactions.

2.2 Antibody cluster morphology depends on the location of binding sites

The large variation in the viscosity of different antibodies requires that the causative interactions involve the variable region. Experiments have shown examples of antibodies where the CDR self-associates or binds to the F_c domain [11, 12]. The type of interaction, and hence the allowed structures, will depend on the specific antibody. The presence of F_ab-F_c interactions is expected to have a significant effect on the rheological behavior, as these interactions can lead to branched structures, which cannot relax by the reptation mechanism that dominates in semi-dilute polymer solutions [13]. For now, we ignore dynamical effects and compute the equilibrium ensemble of complex structures as a function of the binding location.

2.2.1 Head-to-Head binding results in linear aggregates

To begin, we review the simplest case of F_ab-F_ab interactions, as described in [10]. We refer to this as "head to head" (HH) binding. HH binding results in the formation of linear structures, as shown schematically in Fig. 2.1

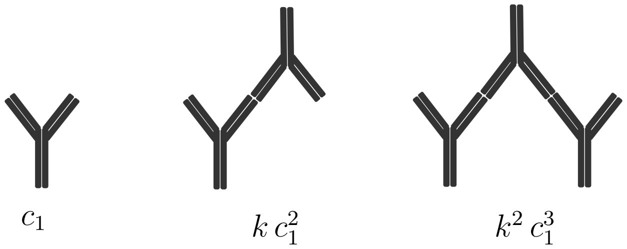


Figure 2.1: *Cartoon and concentration for the monomer, dimer, and trimer states in the HH model.*

The equilibrium constant for HH association is defined by

$$k = \frac{c_2^{\mathrm{HH}}}{c_1^2} \tag{2.1}$$

where c_2^{HH} is the concentration of dimers formed by HH binding and c_1 is the monomer concentration. The dimer equilibrium constant provides a valuable connection between dilute solution properties, which can obtained early in the development pipeline, and the viscosity of concentrated solutions. This connection can also be made using numerical methods or MD simulations [14–17]. Here we present calculations to make this connection analytically.

It follows from Eq. 2.1 that the concentrations c_i of complexes containing i molecules will be given by

$$c_i = c_1\,(k\,c_1)^{i-1} \tag{2.2}$$

where we are assuming that the equilibrium constant k is independent of the number of molecules in the i-mer.

Summing over all HH complexes, the grand partition function is given by

$$q_{\mathrm{HH}} = c_1\left(1 + (k\,c_1) + (k\,c_1)^2 + \cdots + (k\,c_1)^i + \cdots\right) \tag{2.3}$$

where the monomer concentration plays the role of the fugacity $c_1 = e^{\mu/k_{\mathrm{B}}T}$. We can rewrite the partition function as the following recursion relation

$$q_{\mathrm{HH}} = c_1\,(1 + k\,q_{\mathrm{HH}}) \tag{2.4}$$

This equation can be physically interpreted as follows: the two terms on the right hand side of Eq. 2.4 correspond to the two possible states for an antibody head, i.e., one of the F_{ab} domains. The head can be unbound, which terminates the complex and results in the factor of 1, or the head can be bound to another molecule which, in turn, can be bound to another, and so on. In the grand canonical formalism, the number of molecules in this aggregate can range from one to infinity. It follows that the sum of all possible outcomes can be replaced by the factor q_{HH} on the right hand side of Eq. 2.4. This is depicted schematically as

$$\begin{aligned} Q &= \mathsf{Y} + \mathsf{YY} + \mathsf{YYY} + \cdots \\ &= \mathsf{Y} \times (1 + \mathsf{Y} + \mathsf{YY} + \cdots) \\ &= \mathsf{Y} \times (1 + Q) \end{aligned}$$

Rearranging Eq. 2.4 we get the following expression for the partition function

$$q_{\mathrm{HH}} = \frac{c_1}{1 - k\,c_1} \tag{2.5}$$

which can also be obtained by summing the power series in Eq. 2.3.

2.2.2 Head-to-Tail binding permits the formation of branched clusters

We now consider the self association of antibodies that form bonds between an F_{ab} domain and a F_{c}, which we refer to as "head to tail"(HT) binding.

Defining s as the association constant for the HT binding, the grand partition function for complexes formed entirely by HT associations can be written as the following recursion relation

$$q_{\mathrm{HT}} = c_1 \left(1 + 2\, s\, q_{\mathrm{HT}} + s^2\, q_{\mathrm{HT}}^2\right) \tag{2.6}$$

As with Eq. 2.4, this equation has a physical interpretation: the three terms correspond to the available states for the heads of a single molecule. The first term corresponds to the state where both heads are unbound, the second term corresponds to the case where only one of the heads binds to a tail, and the third term corresponds to the state where each head binds to another molecule. Since each bound molecule can initiate a cluster of any number of molecules, the factor q_{HT} is introduced. Solving the quadratic Eq. 2.6, and keeping the root with the correct low concentration limit ($q_{\mathrm{HT}} \to c_1$), we obtain an expression for the grand partition function

$$q_{\mathrm{HT}} = \frac{1 - 2sc_1 - \sqrt{1 - 4sc_1}}{2\, s^2 c_1} \tag{2.7}$$

Fig. 2.2 shows schematically some of the complexes that can be expected from HT association. As in the HH case, we can associate each cluster to a term in the grand partition function (Eq. 2.8). This can be seen by Taylor expanding Eq. 2.7

$$q_{\mathrm{HT}} = c_1 \left(1 + 2\,(s\, c_1) + 5\,(s\, c_1)^2 + \cdots\right) \tag{2.8}$$

The coefficients in Eq. 2.8 indicate the degeneracy of the relevant i-mer states. For example, the term $5\,(s\, c_1)^2$ corresponds to the five dimer states in Fig. 2.2. Note that several of these structures are related by rotational symmetry. Therefore, these states should be considered "undistinguished" particles in the sense used in [18, 19] for classical particles. This overcounting can be corrected in an approximate way by dividing the final partition function by two

$$\tilde{q}_{\mathrm{HT}} \simeq \frac{c_1 + q_{\mathrm{HT}}}{2} \tag{2.9}$$

where $\tilde{q}_{\mathrm{HT}}$ is the corrected partition function and the extra factor of c_1 ensures that the monomer term is accurate. In the Supporting Information we present a more accurate correction and discuss the accuracy of the simple approximation made in Eq. 2.9.

Comparing HH and HT cases (Eqs. 2.3 and 2.8) we notice that in the HH model there is only one possible cluster structure for a given number of monomers, but the HT model allows for many structures for each cluster size. Therefore, for equivalent binding energies, there is an additional entropic factor promoting HT binding.

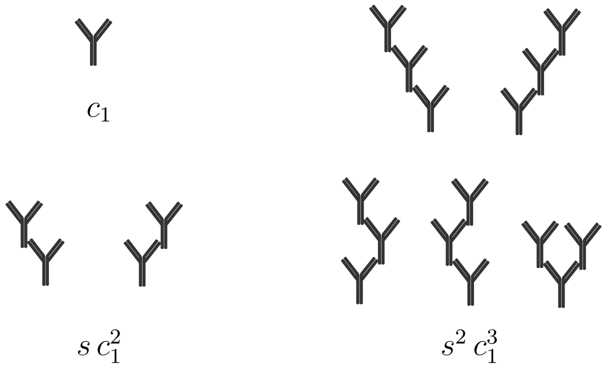


Figure 2.2: *Cartoon of the monomer, dimers, and trimers described by the HT model (Eq. 2.8). As discussed in the Supporting Information, several of these structures are related by rotational symmetry and, therefore, are over-represented in the partition function* q_{HT}. *The corrected partition function* $\tilde{q}_{HT}$ *removes this over-counting.*

2.2.3 The average cluster size increases with binding affinity and protein concentration

The cluster concentrations obtained in the previous sections are functions of the monomer concentration c_1. This quantity is less experimentally accessible than the total protein concentration c. The relationship between c_1 and c can be obtained from the mass conservation law, which requires that the total concentration c must satisfy

$$c = \sum_{i=1}^{\infty} i\, c_i = c_1 \frac{dq}{dc_1} \tag{2.10}$$

The average size of the aggregates can be written as

$$\langle n\rangle_{\mathrm{HT}} = c_1 \frac{d\,(\ln q)}{dc_1} \tag{2.11}$$

Both equations are valid for both the HH and HT models. Now, the monomer concentration c_1 can be obtained by solving Eq. 2.10 in terms of the total concentration. This result can be used in Eq. 2.11 to find the average size. For the HH model the result is [10]

$$\langle n\rangle_{\mathrm{HH}} = \frac{2kc}{\sqrt{1+4kc}-1} \tag{2.12}$$

and for the HT model we have

$$\langle n\rangle_{\mathrm{HT}} = \frac{\sqrt{2}\,s\,c}{\sqrt{1+4sc+s^2c^2-(1+sc)\sqrt{1+6sc+s^2c^2}}} \tag{2.13}$$

We focus now on computing the distribution of aggregates. To do this, we need to compute the concentration, c_i, of all clusters of size i. This can be done by Taylor expanding the partition function and selecting the term proportional to c_1^i. This procedure is formally given by

$$c_i = \frac{c_1^i}{i!}\left.\frac{d^{i-1}q}{dc_1^{i-1}}\right|_{c_1=0} \tag{2.14}$$

In order to analyze the results, we take the approach in [10] and express the binding affinities in terms of the free energy of the binding interaction

$$k = \frac{1}{m}e^{\epsilon_{HH}} \tag{2.15}$$

$$s = \frac{1}{m}e^{\epsilon_{HT}} \tag{2.16}$$

where m is the molecular mass, and ϵ_{HH} and ϵ_{HT} are the free energies for HH and HT binding association at the standard concentration of 1M, respectively, in units of $k_{\mathrm{B}}T$.

Fig. 2.3 compares the average size of cluster $\langle n\rangle$ for both HH and HT models at the same binding energy ϵ. It is clear that increasing concentration or increasing binding energy allows for a larger number of monomers per cluster. Also, at the same ϵ the HT model gives larger structures. This comes from the fact that the HT model allows a larger number of different clusters with the same number of monomers, increasing the probability of populating those states. Finally, we see that the onset of aggregation occurs for binding free energies around 5-6 $k_\mathrm{B}T$ in the range of concentrations 100-200 mg/ml.

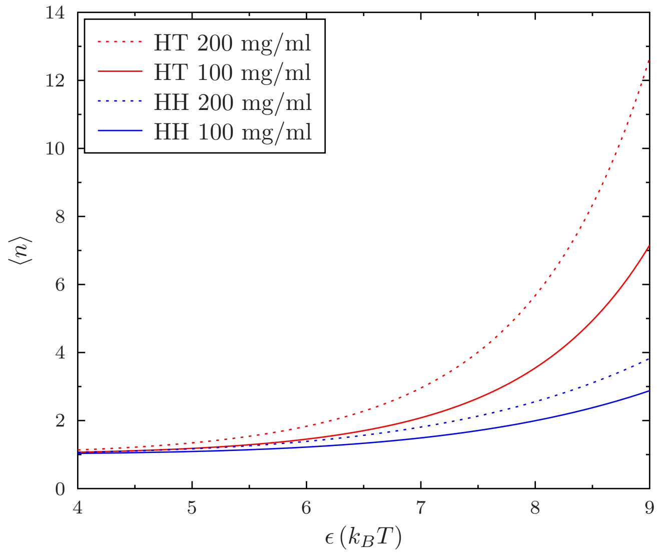


Figure 2.3: *Average number of monomers per cluster for the HH model (blue) and the HT model (red) at two different concentrations: 100mg/mL (solid lines) and 200mg/mL (dotted lines) at the same binding energy* $\epsilon_{HH} = \epsilon_{HT} = \epsilon$*.*

We now focus on the population of different cluster sizes. Fig. 2.4 shows the concentration of clusters from monomers to heptamers as a function of the binding free energy for both HH and HT models. The monomer concentration decreases monotonically as a function of binding energy as the monomer pool is depleted to form larger structures. In contrast, the concentration of larger structures shows a non-monotonic dependence. This is because initial increases in the binding energy permit the formation of larger structures, while much larger values shift the weight of the distribution to larger aggregates which deplete the pool of smaller structures.

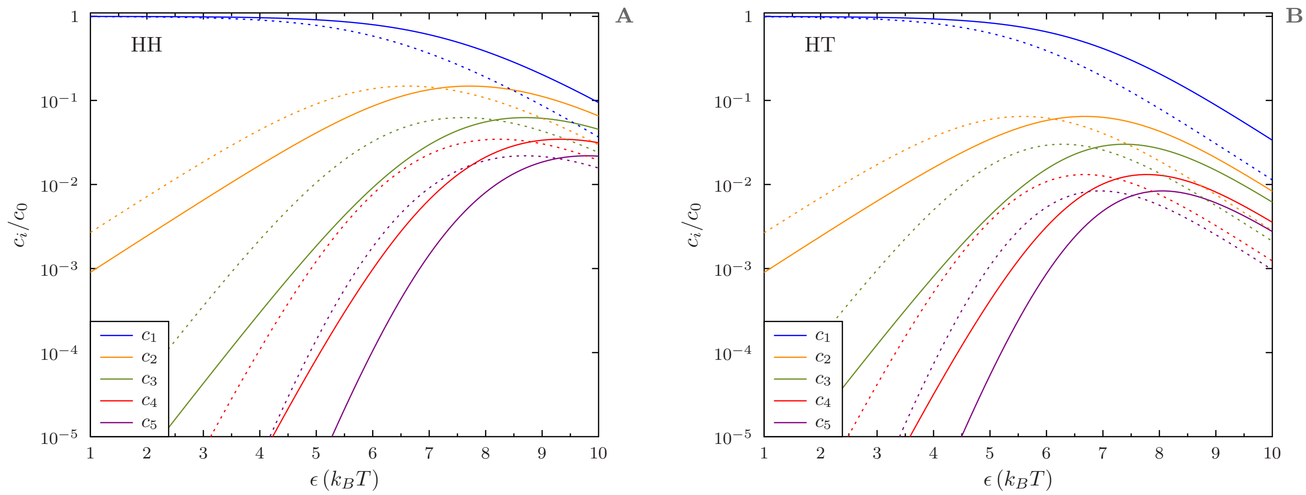


Figure 2.4: *Behavior of the relative concentrations for HH model (A) and HT model (B) as a function of binding energy* $\epsilon_{HH} = \epsilon_{HT} = \epsilon$ *at different concentrations: 50mg/mL (solid lines), 150mg/mL (dotted lines). At low binding energies monomers dominate, but as the binding energy increases the monomer concentration decreases monotonically as larger structures start to form. Clusters larger than the monomer have a non-monotonic concentration because the large clusters formed at high binding energies consume the pool of smaller clusters.*

2.3 The viscosity of dilute solutions is determined by the volume occupied by antibody clusters

The equilibrium distribution of antibody cluster sizes can be applied to obtain expressions for the viscosity at low concentrations. For a dilute solution of hard spheres immersed in a solvent, the viscosity is given by the Einstein relation [20]:

$$\frac{\eta}{\eta_0} = 1 + \frac{5}{2}\phi \tag{2.17}$$

where η_0 is the solvent viscosity and ϕ is the volume fraction occupied by the spheres. Approximating the antibody complexes as spheres of radius equivalent to radius of gyration

R_g and mass $\langle n \rangle$ times that of a monomer, we can approximate the volume fraction as

$$\phi = \frac{4}{3}\pi R_g^3 \frac{c}{\langle n \rangle m} \tag{2.18}$$

where m is the molecular mass.

The radius of gyration can now be computed by considering the antibody complex as a polymer chain of $3n$ segments of length b [21]

$$R_g = b \langle 3n \rangle^{\nu} \tag{2.19}$$

where $\langle n \rangle$ is the average number of monomers per cluster, ν is the Flory exponent, and the factor of three comes from the number of statistically independent segments (domains) per molecule. Plugging Eq. 2.19 into Eq. 2.18, we get an expression for the volume fraction that only depends on the binding energy of a particular molecule,

$$\phi = \frac{4}{3}\pi \frac{c}{m} \tilde{b}^3 \langle n \rangle^{3\nu - 1} \tag{2.20}$$

where $\tilde{b} \sim 3^{\nu} b$ is the effective molecular radius. Plugging Eq. 2.20 into Eq. 2.17, we obtain an expression for the viscosity of antibodies in dilute solutions

$$\frac{\eta}{\eta_0} = 1 + \frac{5}{2}\frac{c}{\rho} \langle n \rangle^{3\nu - 1} \tag{2.21}$$

where $\rho = \frac{3}{4\pi}\frac{m}{\tilde{b}^3}$ is the domain density. Note that even though viscosity depends linearly on volume fraction, it does not have a linear dependence on concentration, since $\langle n \rangle$ is non-linear in c.

The HH and HT binding affinities determine the average number of monomers per cluster $\langle n \rangle$. While binding location is not known for most antibodies, we can exploit two systems where Hydrogen-Deuterium exchange experiments have identified the intermolecular interactions as predominantly HH and HT, respectively [11, 12]. In making the latter comparison, we note that Eq. 2.19, which gives the radius of gyration, is valid only for linear polymers. Nevertheless, in the low concentration regime, small structures (without branching) dominate, making Eq. 2.21 an adequate approximation for antibodies with HT interactions.

We can compare this expression to the results obtained in [10]. Using the concentration dependence of $\langle n \rangle$ (Eq. 2.12) in Eq. 2.21 we see that viscosity scales like $\eta \sim c^{1.4}$ in the dilute regime, which is a weaker dependence than the semi-dilute result, $\eta \sim c^{3.75}$, found in [10].

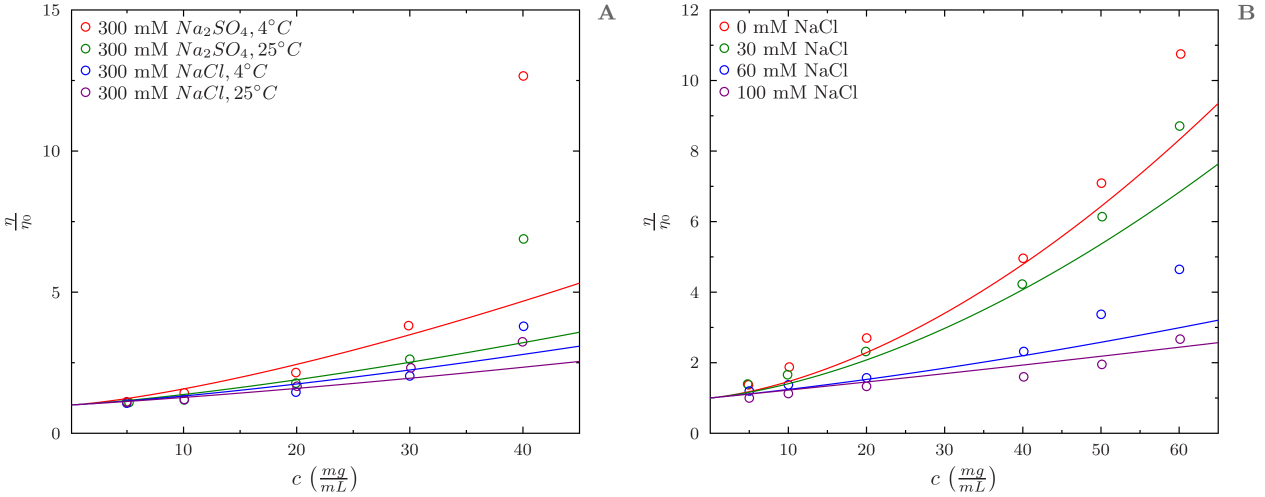


Figure 2.5: *Comparison of Eq. 2.21 to the viscosities measured in [11] (left), and [12] (right). Binding parameters are fit to the first three data points ($c \leq 20$ mg/ml). The increasing deviation at higher concentrations, especially in panel A, indicates that the antibodies are exiting the dilute regime and becoming entangled.*

The free parameters in the model are the segment length b and the association constants k and s. Since Eq. 2.21 was developed for low antibody concentrations, we limit the fitting algorithm to the first three data points for each data set (up to 20 mg/mL). We fit b by minimizing the sum of square errors for all cases, and then fit the association constants for each separate case for that value of b. Comparison between the model and the experimental results are shown in Figs. 2.5A and 2.5B. Fitting yields an effective segment length $b = 6.1$ nm, or equivalently $\tilde{b} = 11.8$ nm, which is consistent with the domain size and antibody dimension. The fitted association constants are displayed in Table 2.1.

				$\langle n \rangle$	
ref. [11]	Fig. 2.5A	$k\left(\frac{\mathrm{L}}{\mathrm{g}}\right)$	$c^*\left(\frac{\mathrm{g}}{\mathrm{L}}\right)$	$c = 10$ (g/L)	$c = 30$ (g/L)
	purple	0.021	27	1.18	1.44
	blue	0.068	22	1.46	2.01
	green	0.136	19	1.77	2.58
	red	0.609	14	3.02	4.80
ref. [12]				$\langle n \rangle$	
	Fig. 2.5B	$s\left(\frac{\mathrm{L}}{\mathrm{g}}\right)$	$c^*\left(\frac{\mathrm{g}}{\mathrm{L}}\right)$	$c = 10$ (g/L)	$c = 30$ (g/L)
	purple	0.001	34	1.02	1.06
	blue	0.007	28	1.13	1.36
	green	0.07	16	2.04	3.67
	red	0.11	14	2.53	4.96

Table 2.1: *Fitted free energies for the dilute antibody solutions plotted in Fig. 2.5. Also shown are the overlap concentration, at which Eq. 2.21 is expected to fail, and the average cluster size at 10 and 10 mg/ml.*

Eq. 2.21 gives the zero-shear viscosity of a solution in the dilute regime. This means that it is expected that the model will fail around the overlap concentration c^* that separates the dilute regime from the semi-dilute. The overlap volume fraction is reached when the molecules occupy all the space, i.e. $\phi^* = 1$. This condition is obtained from Eq. 2.20 and tabulated in Table 2.1. Inspection of Fig. 2.5 shows that c^* is predictive of when the dilute model is a good description of the solution viscosity. For concentrations below c^* the fits do an excellent job of describing the measured viscosity. However, above c^* entanglement effects lead to marked deviation. The failure of the dilute model can occur at concentrations as low as 10-20 mg/ml for antibodies with strong attraction.

Due the small complex sizes ($\langle n \rangle < 5$) formed under these conditions (Table 2.1), these systems are insensitive to the choice of HH or HT models. This is because the lack of branching and the small degeneracy factors make the HH and HT models nearly interchangeable. Because of this, and the lack of experiments that identify the binding location, only the HH model will be used when analyzing the rheological behavior of antibodies in the following sections.

2.4 Equilibrium Dynamics of antibody solutions

Above the overlap concentration antibodies have a viscous response similar to that of a semi-dilute polymer solution. This is described by the theory of Schmit *et al.* for zero-shear viscosity, which we summarize below. In their theory, each antibody is modeled as a featureless polymer with the polymer ends located at the antigen binding sites [10]. An antibody is considered to have a "polymer" length of $L = 3$ segments (with size b each) associated with the three domains of the antibody molecule. Schmit *et al.* focused on the semidilute regime, since in the range of interest, concentrations above 100 mg/mL, the center-to-center particle separation is comparable to the molecular size and the antibodies occupy a solution volume on the order of 10%. The reptation mechanism of polymers with entanglements yields a zero-shear viscosity proportional to the product of the shear modulus G and the longest relaxation time τ_{rep} [22]

$$\eta_0 \propto G\, \tau_{\text{rep}}. \tag{2.22}$$

In the semidilute regime, the reptation time is given by [23, 24]

$$\tau_{\text{rep}} \sim \frac{\eta_s L^{3\nu} b^3}{k_{\text{B}} T} \left(\frac{c}{c^*}\right)^{\frac{3-3\nu}{3\nu-1}} \tag{2.23}$$

and the shear modulus can be obtained from scaling calculations to give [22]

$$G \sim \frac{c}{L} k_{\text{B}} T \left(\frac{c}{c^*}\right)^{\frac{1}{3\nu-1}} \tag{2.24}$$

where η_s is the viscosity of the solvent, $\nu \simeq 3/5$ is the Flory exponent, c is the segment concentration, and $c^* \sim L^{1-3\nu} b^{-3}$ is the overlap concentration [22]. The reptation model for entangled polymers requires a clear separation of length scales between monomers and polymer chains. However, even after considering self-association of antibodies, the range of polymer lengths is $3 \leq L \lesssim 20$, which is still too short to satisfy the above mentioned condition. Nevertheless, the analytic model provides useful qualitative insight and reasonable agreement with measured viscosity.

Eqs. 2.22-2.24 yield a zero-shear viscosity of the form

$$\eta_0 \sim c^{\frac{3}{3\nu-1}} L^3 \tag{2.25}$$

The average length is computed from Eq. 2.12 which can be included in Eq. 2.25 to yield

$$\eta_0 = A c^{\frac{3}{3\nu-1}} \left(\frac{2kc}{\sqrt{1+4kc}-1} \right)^3 \tag{2.26}$$

where A is a constant of proportionality. Eq. 2.26 predicts that viscosity increases with concentration as $c^{3.75}$ for $\nu \simeq 0.6$. Also, we can see from Eq. 2.25 that the viscosity depends strongly on polymer length L. This increment in the zero-shear viscosity, proportional to the cube of the length scale, is found for semidilute polymer suspensions, as well as for colloidal suspensions [25].

Fitting the model to experiments, the constant was found to be $A = 5.4 \times 10^{-8}$ cP (mg/mL)$^{3.75}$ [10]. Note that the constant A absorbs the factor of 3 difference between the aggregate number and the polymer length, $3 \langle n \rangle = L$.

2.5 Shear thinning results from the release of entanglements

According to reptation theory [22, 24], the viscous response of a polymer solution under shear is the result of two relaxation mechanisms. First, the shear will stretch out the polymers by deforming the entanglements that constrain them. Then the polymer shrinks along the tube generated by the constraints to recover its equilibrium extension. This relaxation process releases entanglements and, since it is not affected by the tube, the relaxation time is given by the Rouse relaxation time τ_R. Second, the polymers undergo reptational diffusion to establish a new conformation with the original number of entanglements points. Therefore, these two processes oppose each other with the chain retraction releasing entanglements and reptation restoring them. It can be checked that Rouse time is always much less than reptation time for any entangled solution where the length of the tube is longer than its diameter, therefore, in the following we treat chain retraction as instantaneous.

Returning to the simple shear case, the shear stress immediately following a step strain is approximately $\sigma(t = 0^+) = G\,\mathrm{d}\gamma$, where G is the shear modulus. Following Milner, [26] the shear flow with a constant strain rate $\dot{\gamma}$ is approximated as a sequence of step strains $\mathrm{d}\gamma$ separated in time by some convenient interval $\mathrm{d}t$, in which the chain relaxes. After a Rouse time, the chain retracts relaxing entanglements and reducing the stress by a factor $h\left(\mathrm{d}\gamma\right)$

$$\sigma\left(t = \tau_R\right) = h\left(\mathrm{d}\gamma\right) G\,\mathrm{d}\gamma \tag{2.27}$$

where $h\left(\mathrm{d}\gamma\right)$ is the nonlinear damping function. Next, the chain relaxes by reptation, which restores the entanglements. For the steady state case, entanglements are released as fast as they are restored. Therefore, the appropriate step strain is $\mathrm{d}\gamma = \dot{\gamma}\tau_{rep}$. Since the viscosity is related to the shear stress by $\eta = \sigma/\dot{\gamma}$ we have

$$\eta\left(\dot{\gamma}\right) \approx G\tau_{rep}\, h\left(\dot{\gamma}\tau_{rep}\right) \tag{2.28}$$

where $h\left(\gamma\right)$ can be estimated by computing the root-mean-squared tube length increment immediately after the step strain [22], yielding

$$h\left(\gamma\right) = \left(1 + \gamma^2/3\right)^{-1/2} \tag{2.29}$$

The functional form of $h\left(\gamma\right)$ remains an open question and we refer the reader to [27] for a discussion of the topic. However, Eq. 2.29 is sufficient for our purposes here.

We obtain our final expression for the shear rate dependence of viscosity plugging Eq. 2.29 into Eq. 2.28 and using Eq. 2.26 for the product $G\tau_{rep}$

$$\eta = Ac^{\frac{15}{4}}\left(\frac{2kc}{\sqrt{1+4kc}-1}\right)^{3}\left(1+\frac{1}{3}\dot{\gamma}^{2}B^{2}\left(\frac{2kc}{\sqrt{1+4kc}-1}\right)^{6}\right)^{-1/2} \tag{2.30}$$

In writing Eq. 2.30 we have introduced a proportionality constant into the expression for the reptation time $\tau_{rep} = B\langle L\rangle^{3}$.

The free parameters are A, which was obtained in [10], the equilibrium constant k, which depends on both the molecule and the solution condition, and B which we obtain from a global fit to all solution conditions.

2.6 The binding affinity predicts both the zero shear viscosity and the onset of shear thinning

The model was tested against experimental data on viscosity shear-dependence for two different antibodies under a variety of conditions from Zarraga et al. [7] and Godfrin et al. [28]. Fitting Eq. 2.30 to the data sets, we obtain the prefactor B for the reptation time by minimizing the sum of square errors for both types of antibodies for all system conditions. Next, we obtain the dimerization equilibrium constant k, which can be computed by fitting the viscosity for each individual system condition. The values of B and k are shown in Table 2.2. Solid lines in Fig. 2.6 show the fits using the model in Eq. 2.30.

Eq. 2.12 shows that the average chain length is determined by k, $\langle n\rangle \propto k^{1/2}$. Both the zero-shear viscosity and the reptation time have a strong dependence on polymer length, (Eq. 2.26). We have seen that the reptation time captures the onset of the shear-thinning behavior, thus an increment in the polymer length means that η_0 will increase and the onset of shear-thinning behavior will occur for lower values of the shear rate. This correlation between η_0 and the shear thinning is observed in the experimental data and captured by the reptation model. According to the model, this correlation stems from the fact that systems with shorter chains can re-form entanglements faster than systems with longer chains, due to the shorter reptation times.

An inspection of Fig. 2.6 reveals that the drop in viscosity is noticeably steeper than the $\dot{\gamma}^{-1}$ power law predicted by Eq. 2.30. This discrepancy is also seen in polymer systems, which typically show shear thinning behavior in the range $\dot{\gamma}^{-1.2}$ to $\dot{\gamma}^{-1.4}$ [29, 30]. These exponents are consistent with shear thinning behavior shown in Fig. 2.6, although the less viscous conditions fall outside of this range, presumably because smaller antibody complexes are less polymer-like. The literature contains many refinements to the reptation model intended to improve the agreement with experiments [31, 32], however, these fall outside of the scope of the present work which is to demonstrate the utility of the entanglement model in capturing the key features of antibody rheology.

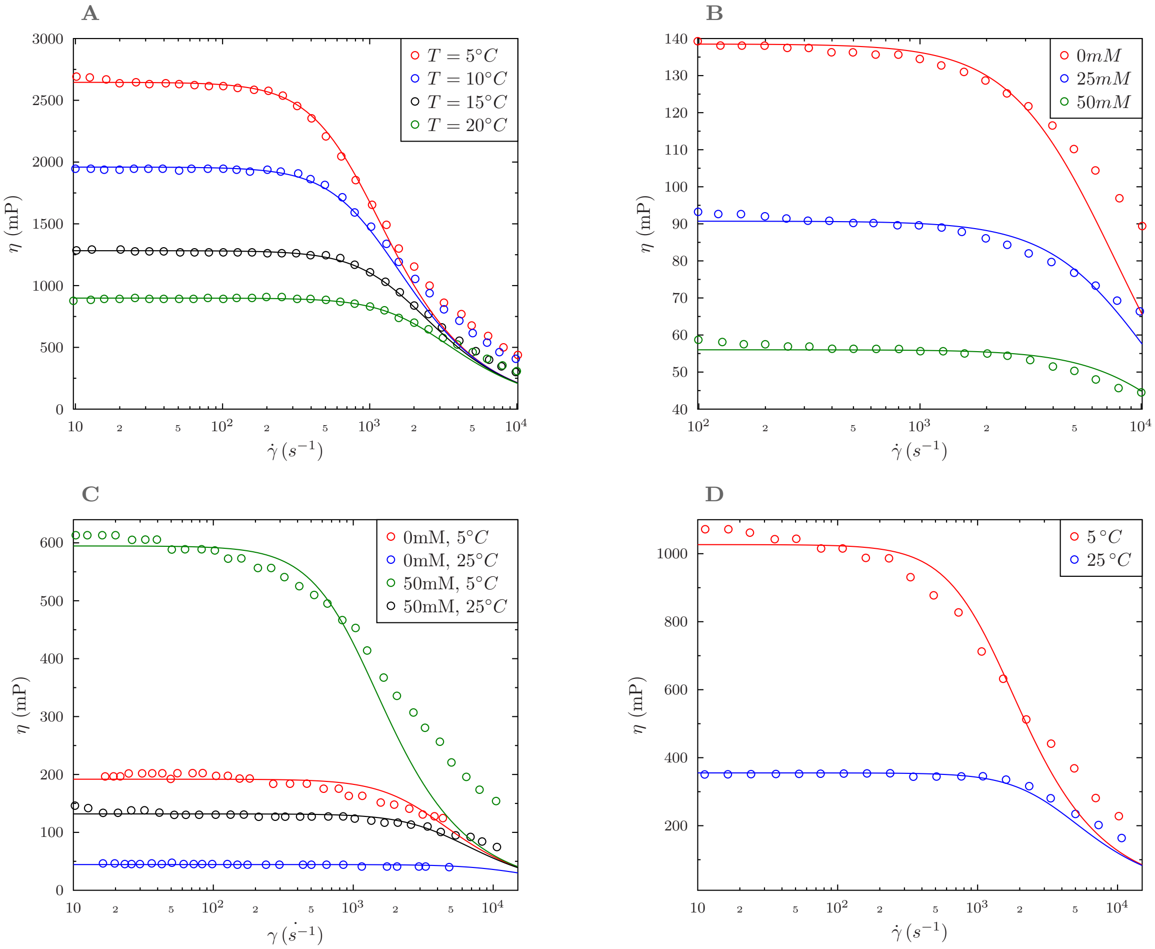


Figure 2.6: *Comparison of Eq. 2.30 to the shear dependent viscosity of an antibody at A) 200 mg/ml [7], B) 150 mg/ml [7], and a different molecule at C) 120 mg/ml [28], and D) 146 mg/ml [28]. Solutions with lower viscosity deviate more strongly from the theory due to the fact smaller complexes are less polymer-like.*

ref. [7]	Fig. 1	$k\left[\frac{\mathrm{mL}}{\mathrm{mg}}\right]$	$\langle n\rangle$	B [10^{-5} s]
	red	0.094	4.86	1.83
	black	0.074	4.40	
	blue	0.054	3.82	
	green	0.041	3.39	
	Fig. 2	$k\left[\frac{\mathrm{mL}}{\mathrm{mg}}\right]$	$\langle n\rangle$	B [10^{-5} s]
	red	0.028	2.61	1.83
	blue	0.019	2.26	
	green	0.012	1.93	
ref. [28]	Fig. 3	$k\left[\frac{\mathrm{mL}}{\mathrm{mg}}\right]$	$\langle n\rangle$	B [10^{-5} s]
	green	0.214	5.60	0.97
	red	0.091	3.84	
	black	0.067	3.39	
	blue	0.027	2.36	
	Fig. 4	$k\left[\frac{\mathrm{mL}}{\mathrm{mg}}\right]$	$\langle n\rangle$	B [10^{-5} s]
	red	0.040	2.97	0.97
	blue	0.011	1.84	

Table 2.2: *Parameters extracted from the fits in Fig. 2.6, including the dimerization equilibrium constant and the global proportionality constant. Also shown is the average cluster size* $\langle n\rangle$*, which increases with the equilibrium constant.*

2.7 Summary

We have used in this work polymer theory of rheology of polymer chains in semidilute solutions in order to obtain the viscosity in this regime and calculate the reptation time associated to entangled polymer dynamics.

At the high concentrations used for pharmaceutical formulation, even weak intermolecular interactions result in the formation of antibody clusters. Due to the discrete binding sites on the antibody arms, these clusters will tend to have an elongated morphology. This promotes entanglements, which have a profound effect on the dynamics of the solution. We have shown that a polymer model explains the viscosity, shear thinning behavior of antibody solutions. This model shows that viscosity can be reduced by minimizing entanglements, either by reducing intermolecular interactions or by using antibody constructs that lack the multi-valency required to make extended structures.

It is shown that the main time governing the dynamics is the reptation time. This gives information that allows simulations and experiments to look for resolutions where the reptation time is captured. Then, binding energies can be extracted to characterize the system and predict the rheological behavior.

Bibliography

[1] Shire, S. J., *Monoclonal antibodies: meeting the challenges in manufacturing, formulation, delivery and stability of final drug product*, Woodhead Publishing, 2015.

[2] Yadav, S., Shire, S. J., and Kalonia, D. S., Journal of Pharmaceutical Sciences **99** (2010) 4812.

[3] Harris, R. J., Shire, S. J., and Winter, C., Drug Development Research **61** (2004) 137.

[4] Zhang, Z. and Liu, Y., Current Opinion in Chemical Engineering **16** (2017) 48.

[5] Galush, W. J., Le, L. N., and Moore, J. M., Journal of Pharmaceutical Sciences **101** (2012) 1012.

[6] Connolly, B. D. et al., Biophysical Journal **103** (2012) 69.

[7] Zarraga, I. E. et al., Journal of Pharmaceutical Sciences **102** (2013) 2538.

[8] Li, L. et al., Pharmaceutical Research **31** (2014) 3161.

[9] Yearley, E. J. et al., Biophysical Journal **106** (2014) 1763.

[10] Schmit, J. D. et al., J. Phys. Chem. B **118** (2014) 5044.

[11] Arora, J. et al., mAbs **7** (2015) 525, PMID: 25875351.

[12] Arora, J. et al., mAbs **8** (2016) 1561, PMID: 27560842.

[13] Gennes, P.-G., *Scaling Concepts in Polymer Physics*, Cornell University Press, 1979.

[14] Kastelic, M., Dill, K. A., Kalyuzhnyi, Y. V., and Vlachy, V., Journal of Molecular Liquids **270** (2018) 234.

[15] Calero-Rubio, C., Saluja, A., and Roberts, C. J., The Journal of Physical Chemistry B **120** (2016) 6592.

[16] Barnett, G. V. et al., The Journal of Physical Chemistry B **120** (2016) 3318.

[17] Kastelic, M. and Vlachy, V., The Journal of Physical Chemistry B **122** (2018) 5400.

[18] Cates, M. E. and Manoharan, V. N., Soft Matter **11** (2015) 6538.

[19] Sethna, J. P., *Statistical Mechanics: Entropy, Order Parameters and Complexity (Oxford Master Series in Physics)*, Oxford University Press, 2006.

[20] Einstein, A., Annalen der Physik **324** (1906) 371.

[21] Teraoka, I., *Polymer solutions : an introduction to physical properties*, Wiley, New York, 2002.

[22] Doi, M. and Edwards, S. F., *The Theory of Polymer Dynamics (International Series of Monographs on Physics)*, Oxford University Press, Oxford, U.K., 1988.

[23] De Gennes, P. G., Macromolecules **9** (1976) 587.

[24] De Gennes, P. G., Macromolecules **9** (1976) 594.

[25] Mewis J. and Wagner N. J., *Collopidal Suspension Rheology*, Cambridge University Press, New York, 2012.

[26] Milner, S. T., Journal of Rheology (1978-present) **40** (1996) 303.

[27] Watanabe, H., Progress in Polymer Science **24** (1999) 1253.

[28] Godfrin, P. D. et al., J. Phys. Chem. B **120** (2016) 278.

[29] Hieber, C. A. and Chiang, H. H., Polym Eng Sci **32** (1992) 931.

[30] Fatkullin, N., Mattea, C., and Stapf, S., Polymer **52** (2011) 3522.

[31] Marrucci, G., Journal of Polymer Science: Polymer Physics Edition **23** (1985) 159.

[32] Milner, S. T. and McLeish, T. C. B., Macromolecules **31** (1998) 7479.

Chapter 3

Static properties in biological condensates determined by topology of the dense phase

3.1 Introduction

Inside living cells, in addition to vesicle-like organelles there are compartments that are not membrane-bound. These structures are referred to as biomolecular condensates or membraneless organelles, and can perform a variety of functions inside the cell, like signaling and enzymatic properties [1, 2]. These structures are often responsive to changes in environmental conditions, cell cycle, or stress [3–6].

Examples of these biomolecular condensates include RNA-protein granules such as nucleoli, Cajal Bodies and PML nuclear bodies in the nucleus [7], as well as stress granules and germ granules in the cytoplasm [8, 9]. These micron-scale structures are all defined by their ability to concentrate proteins and nucleic acids at discrete cellular sites. As these condensates lack a physical barrier to separate their internal components from the surrounding medium, it remained elusive for a long time how they concentrate molecules, maintain and regulate their structures, control their compositions and modulate internal biochemical activities.

The first membraneless compartment that was observed within the nucleus of neuronal cells was the nucleolus [10]. Since then, many such compartments have been discovered in the nucleus, cytoplasm and on membranes of essentially all eukaryotic cells. High-resolution microscopy imaging and descriptions of their molecular components have revealed similarities in their shape, dynamics and manner of assembly, despite differences in their composition, location and function. Some of the common properties are high concentration of components inside the condensate region with rapid exchange with the environment [11, 12], and coalescence of two compartments upon contact [13, 14]. These results and others led to the conclusion that the biomolecular condensates were liquid-like, and that they form by liquid-liquid phase separation from surrounding cytoplasm. Principles of phase separation can indeed explain the formation of such structures with diverse material properties as well as the complex organization of such structures (organization into subcompartments, for example). The presence of a phase boundary explains how molecules can be concentrated in one place in a cell without a surrounding membrane but still provide an environment suitable for cellular biochemistry that depends on rapid diffusion. Phase separation also provides a unifying principle that explains the formation of membraneless compartments from different types of molecules.

Biomolecular condensates are often formed by multivalent molecules that is, molecules that contain multiple elements that govern intra- or inter-molecular interactions [15, 16]. This multivalency is important because concepts in polymer theory indicate that multivalent molecules naturally assemble into large oligomers or polymers when mixed, which promotes phase separation. This idea can be applied to understand the phase separation behaviour of diverse multivalent molecules. Such molecules include proteins composed of multiple interaction domains and proteins containing disordered regions that provide multiple weakly adhesive sequence elements. RNA and DNA molecules, which can contain multiple regions that bind to other nucleic acid molecules and proteins, can also undergo phase separation.

One central problem to be address is how properties of molecules contribute to the function of these condensates. In this paper we will address the following equilibrium functional characteristics: density, which influences the diffusion and relaxation properties in the fluid; and client binding, which plays a role in the enzymes metabolism in the cell; where clients are other condensate residents concentrated within the structure by direct interactions with scaffolds, but are not required for condensate formation. It can be posed that these two quantities are related: increasing density inside the condensates diminishes the number of binding sites available to clients. Thus, it raises the question of whether this two can be decoupled, and what are the variables that control this behavior.

In considering the variables that control the system we can distinguish two categories: genetic control, i.e. variables that are fixed by evolution, and dynamic control, i.e. variables that the cell can tune and can be result in a sensitive or robust response. In the first group we can find the following variables: linker length, valence, binding affinity; and in the second group: concentration, stoichiometry and affinity via post-translational modifications.

Multivalent-proteins have been shown to achieve phase transitions at low protein concentrations when compared to globular proteins [15]. In this work we will modeling disordered proteins as linear polymers characterized by a sequence of attractive stickers "stickers" [17] and flexible, inert "spacers" [18].

3.2 Free Energy

The system under consideration will be a solution with two different types of linear polymer chains inside. Each type "A" and "B" of polymers has a number θ_A and θ_B of stickers evenly spaced, making them θ-valent sticky polymers. The first step is to compute the free energy. We notice that there are two main contributions: binding interactions and an entropy term, that includes conformational and translational entropy. It is expected that the binding interaction will promote a higher density inside the condensate and fewer sites for clients to bind. On the other hand, it is expected that the conformational and translational entropy will promote lower densities and more client sites available.

In this model we will consider two types of contributions to the system free energy. A first contribution coming from an excluded volume term and a second contribution coming from the binding interactions between the two different types of chains:

$$F = F_{excluded} + F_{bind} \tag{3.1}$$

F_{bind} promotes higher density inside the condensate and fewer client sites available. On the other hand $F_{excluded}$ promotes lower density inside the droplet and more client sites available to bind. If the mesh size of the network inside the droplet is given by the spacing between stickers R_A, it would be expected that the density inside the condensate would scale like $c_A \sim R_A^{-3}$, in other words, density should decrease as linker length increases. However, it is shown in this work that the opposite occurs and this is explained by the type of topology that the network form inside the condensate. For the case of client binding it will be shown that longer linker lengths promotes the binding of clients inside the droplet, meaning that the conformational entropy overcomes the binding energy effect.

We need to first account for the excluded volume contribution to the free energy. We consider the volume between two type A stickers, and the energy that comes from confining $l_A + l_B$ monomers (where l_A and l_B are the number of monomers between stickers of type A and B respectively) inside this volume $V_{sticker} \sim c_A^{-1}$. The excluded volume contribution comes from having real polymer chains in the solution trying to occupy all the volume available. This confinement energy per sticker will be of the order

$$f_{excluded} \sim k_B T \frac{l_a + l_B}{g} \tag{3.2}$$

where $l_A + l_B$ is the number of monomers of both types between two stickers and g is the number of monomers per blob, where a blob describes a length scale below which the polymer chain may be considered to be unperturbed by other forces or effects, in other words inside the blob the polymer exhibits ideal behavior. After some mathematical steps explained in the Methods section we obtain the following expression for the excluded volume contribution per sticker

$$\frac{f_{excluded}}{k_B T} \sim (l_A + l_B)^{\frac{3\nu}{3\nu-1}} (a^3\, c_A)^{\frac{1}{3\nu-1}} \tag{3.3}$$

We now need to include the contributions coming from pairwise interactions and the entropy of sticky chains A and B where both types of molecules have a valence θ, i.e., that each chain has θ stickers evenly distributed with a distance R between each other. We can now compute the free energy of a solution of M molecules of type A. We divide these into $\theta + 1$ groups. N_0 molecules are unbound, N_1 are bound to a molecule of type B that has $\theta - 1$ free sites, $2N_2$ are bound to a type B molecule that binds to a second type A molecule, and so for until having a type B molecule with all its sites bound to molecules of type A. Therefore $M = N_0 + N_1 + 2N_2 + 3N_3 + ... + \theta N_\theta$ and the total number of type B molecules is $N_1 + N_2 + \ldots + N_\theta$. The free energy of the system is

$$\begin{aligned} F_{bind}/k_B T = &- \left[N_1 \epsilon + N_2 \left(\epsilon + \ln(l) - \ln \binom{\theta}{2} \right) + \ldots + N_\theta \left(\epsilon + (\theta - 1) \ln(l) - \ln \binom{\theta}{\theta} \right) \right] \\ &- \mu_B \left(N_1 + N_2 + \ldots + N_\theta \right) - \ln \Omega \end{aligned} \tag{3.4}$$

where $\epsilon > 0$ is the free energy of a type A molecule binding to a type B molecule site and μ_B is the chemical potential of type B chains (both of them scaled by $k_B T$ to make them dimensionless); l is the contour length between two stickers, and the term $(i-1)\ln(l)$ correspond to the entropic penalty of looping the chain to form a bond with the same chain that already has a bond (neglecting variable length, since we are considering only the logarithm of such a length); and the term $\ln \binom{\theta}{i}$ corresponds to the number of choices that i stickers can choose among θ available.

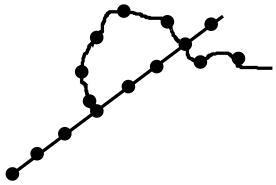


Figure 3.1: *Looping entropy.*

For simplicity of notation we write

$$\epsilon_j = \epsilon + (j-1)\ln(l) - \ln\binom{\theta}{j} + \mu \tag{3.5}$$

Then Eq. 3.4 can be rewritten as

$$F_{bind}/k_BT = -\sum_{j=1}^{\theta} N_j\epsilon_j - \ln\Omega \tag{3.6}$$

We now need to compute Ω, which is the number of ways to arrange type B molecules on the binding sites. We will be following the Flory-Huggins theory to compute it [19, 20] (see Methods).

Minimizing the free energy with respect to the sites occupancies (see Methods) yields

$$f_{bind} = \ln(n_0) + n_2 + 2n_3 + \ldots + (\theta-1)\,n_\theta \tag{3.7}$$

and the sites occupancies

$$\begin{aligned} n_1 &= n_0\, e^{\epsilon_1+\mu_B} \\ n_2 &= n_0^2\, e^{2\epsilon_2+\mu_B+\ln z} \\ \ldots \\ n_\theta &= n_0^\theta\, e^{\theta\epsilon_\theta+\mu_B+(\theta-1)\ln z} \end{aligned} \tag{3.8}$$

where $n_i = N_i/M$ and n_0 is the only real and positive solution of the polynomial equation:

$$1 = (1 + c_{1B}\exp^{\epsilon_1})\,n_0 + \left(2c_{1B}z\exp^{2\epsilon_2}\right)n_0^2 + \ldots + \left(\theta c_{1B}z^{\theta-1}\exp^{\theta\epsilon_\theta}\right)n_0^\theta \tag{3.9}$$

Eq. 3.9 can be solved analytically for low values of θ, but it must be solve numerically for higher orders. We have developed an approximation to compute n_0 in the Methods section 3.6 to obtain an approximate analytical solution to Eq. 3.9.

Summing both contribution we have the total free energy per sticker is given by

$$
\begin{aligned}
f &= f_{excluded} + f_{\theta} \\
&\sim \left(l_A + l_B\right)^{\frac{3\nu}{3\nu-1}} \left(a^3\, c_A\right)^{\frac{1}{3\nu-1}} + \ln\left(n_0\right) + n_2 + 2n_3 + \ldots + \left(\theta - 1\right) n_{\theta}
\end{aligned}
\tag{3.10}
$$

3.3 Concentration in condensate phase

We can now ask how the cell can tune the functional properties. At first glance, density and client binding seem tightly connected. However, a comparison of SPOP/DAXX [21] and SUMO/SIM [22] systems shows that network topology can decouple them. We now proceed to compare the expected topologies that the system can present. One type of topology is the zipper-like structure, characterized by long one dimensional structures, as in the SUMO/SIM system [23]. We expect that this topology promote low density and intolerance of asymmetric stoichiometry of species A and B. On the other hand, there is the random network present in systems like SPOP/DAXX that is three dimensional. We expect that this topology promote higher densities and to be tolerant of asymmetric stoichiometry.

The way topology is handled in our model comes from the z parameter that accounts for the interaction among chains: z describes the number of type A molecules within reach of a type B molecule bound with one site. This parameter has two types of contribution, one coming from same chain interactions, meaning interaction with the same chain already bound (z_{SC}); and a second coming from a mean field type of interaction, coming from interactions with other free chains (z_{MF}):

$$
\begin{aligned}
z &= z_{SC} + z_{MF} \\
&= \frac{R_A}{R_B} + c_A R_A^3
\end{aligned}
\tag{3.11}
$$

where R_A and R_B are the linker lengths of molecules of type A and type B respectively. We can see that the first term stems from the possibility of binding to the free site of the already bound chain, and the second term means a search in the volume within reach of a sticker and depends on the concentration of type A molecules.

We now introduce the parameter P_{SC} to account for the portion of same chain interaction in the system

$$\begin{aligned} P_{SC} &= P_{SC}\left(\epsilon, c_A, c_B, R_A, R_B\right) \\ &= \frac{z_{SC}}{z_{SC} + z_{MF}} \end{aligned} \tag{3.12}$$

When $P_{SC} \approx 1$ same chain interactions dominate and the topology present is 1D zipper-like structures. On the other hand, when $P_{SC} << 1$, mean field interactions dominate and we are in presence of a random 3D network.

Before continuing we need to get the concentration of type "A" molecules in the condensate phase, and its dependency on the concentration of free type "B" molecules and on the linkers length. Given the fact that representative systems have usually low concentration in the dilute phase, as a first approximation we will assume that the phase transition condition is given by

$$\left.\frac{\partial f}{\partial c_A}\right|_{c_{A\_dense}} = 0 \tag{3.13}$$

This results in a equation for the concentration of type "A" molecule inside the dense phase of a droplet $c_A = c_A\left(\epsilon, c_{1B}, R_A, R_B\right)$. Solving Eq. 3.13 numerically we can see the concentration inside the droplet as a function of type "B" molecule concentration. The free parameters that can be varied are the binding energy, and the linker lengths.

We now proceed to look at Eq. 3.12 and see the dependency of topology on the linker length and binding energy. In Fig. 3.2 we see that linker length ($R = R_A = R_B$) determines the type of topology and that the network is robust to changes in binding energy. We notice that there is a region $3nm < R < 5nm$ where a mixed topology is present. The next step would be to analyze the density of the condensates for the different types of topologies.

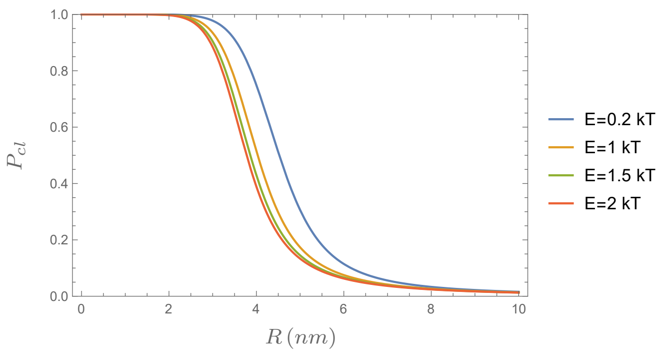


Figure 3.2: *P_{SC} Eq. 3.12 behavior as a function of linker length for different binding energies for bivalent case.*

In Fig. 3.3 we see the behavior of the concentration of type A molecules as a function of concentration of free type B molecules, for different linker lengths for the bivalent case. We have already seen in Fig. 3.2 that linker length determines the type of topology that will form inside the condensate, now we see how the density inside the droplet is affected by the type of topology that the network adopts. We show three different cases of topologies: zipper-like ($R = 2nm$), mixed topology, and 3D random network ($R = 6nm$). We can see that there is a clear distinction between the behavior of each topology. The 3D random network has a higher density, around two order of magnitude, than both the zipper-like and the mixed network. This result shows the expected trends, since the zipper-like network promotes lower densities due to its tendency to form long 1D structures disfavoring cross-links; whereas the random network presents higher densities due to its propensity to form cross-links.

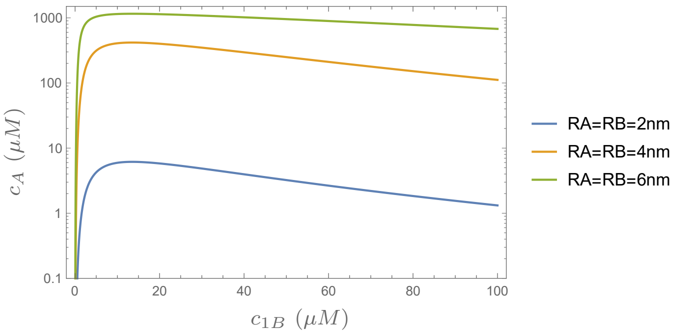


Figure 3.3: *Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$.*

In Fig. 3.4 we can see a comparison between different valences in the density behavior inside the droplet of $R_A = R_B = 6nm$ and $\epsilon = 2k_BT$. We can see that in all cases, concentrations are of the same order but the n-valent case shows a slightly higher value. Another point is that as valence goes higher the peak in concentration occurs earlier, this can be interpreted as the higher the valence the sooner it saturates the available binding sites. One third point to look at is the behavior for large concentration, the higher the valence the more random the network is, taking higher concentrations to dissolve the droplet.

Another point is the almost flat behavior of the n-valent case: this can be interpreted as the excluded volume effect canceling the compression effect that the binding interaction is exerting.

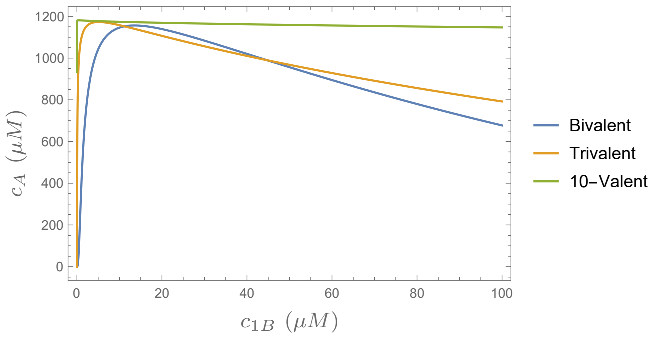


Figure 3.4: *Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths* $R_A = R_B$.

We are now interested in how the system responds to changes in binding energy between molecules. In Fig. 3.5 we can see the concentration of type A molecules behavior inside the droplet as a function of concentration of free type B molecules for different binding energies in the case we have symmetric linker lengths ($R_A = R_B = 2$ nm), same trends are present for larger linker lengths. We can notice that the bivalent case behaves differently compared to the other two cases: the peak in concentration in the bivalent case is the same for all energies. This can be understood as the network formation is constrained by the fact that once a chain forms a bond it is limited to a fixed distance for the next bond. In other words, the network mesh size is given by the linker length. On the other hand, for the other two cases higher energies show higher peaks: now the network mesh size is not limited by one linker length and the chain can bind twice or more times with one chain and still form a cross-link with a different one. In other words, higher energies favor a more condensed network.

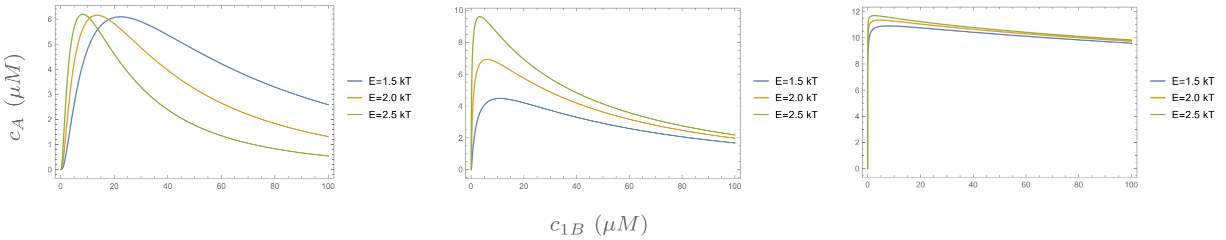


Figure 3.5: *Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different binding energies.*

3.4 Client Binding

We now shift attention to the effect of adding clients to type "A" molecules into the solution. We are going to be interested in computing the ratio of bound clients to that of free client molecules. As a first measure we need to write an entropy term that accounts for the presence of free type "A" molecules bound to clients. We will call N_{client} the number of type "A" molecules bound to a client molecule and N_{00} the number of free type "A" molecules. Therefore, the number of possibilities for having an already free type "A" molecule binding to a client molecule is given by

$$\Omega_{new} = \frac{N_0!}{N_{client}!N_{00}!} \tag{3.14}$$

We can now write the entropy term as follows

$$\begin{aligned} \Omega_{client} &= \Omega \times \Omega_{new} \\ &= \left(\frac{z}{M}\right)^{N_2+2N_3+(\theta-1)N_\theta} \frac{M!}{N_{client}!N_{00}!N_1!\ldots N_\theta!} \end{aligned} \tag{3.15}$$

where now $M = N_{client} + N_{00} + N_1 + 2N_2 + \ldots + \theta N_\theta$.

The free energy of the system is

$$\begin{aligned} F_M/k_BT = &-(N_1\epsilon_1 + N_2\epsilon_2 + \ldots + N_\theta\epsilon_\theta) - \mu_B(N_1 + N_2 + \ldots + N_\theta) \\ &- \epsilon_{client}N_{client} - \mu_{client}N_{client} - \ln\Omega_{client} \end{aligned} \tag{3.16}$$

The same method as before can be followed to obtain each occupation number n_i. Minimizing the free energy with respect to the occupation numbers it gives the following occupation numbers

$$\begin{aligned} n_{client} &= n_0\, e^{\epsilon_{cl}+\mu_{cl}} \\ n_1 &= n_0\, e^{\epsilon_1+\mu_B} \\ n_2 &= n_0^2\, e^{2\epsilon_2+\mu_B+\ln z} \\ &\dots \\ n_\theta &= n_0^\theta\, e^{\theta\epsilon_\theta+\mu_B+(\theta-1)\ln z} \end{aligned} \tag{3.17}$$

where n_0 can be obtained from the polynomial equation

$$1 = \left(1 + c_{1B}\, e^{\epsilon_1} + c_{qcl}\, e^{\epsilon_{cl}}\right) n_0 + \left(2c_{1B} z\, e^{2\epsilon_2}\right) n_0^2 + \ldots + \left(\theta c_{1B} z^{\theta-1}\, e^{\theta\epsilon_\theta}\right) n_0^\theta \tag{3.18}$$

Giving the same expression for the free energy as before

$$f = \ln\left(n_0\right) + n_2 + 2n_3 + \ldots + \left(\theta - 1\right) n_\theta \tag{3.19}$$

We now want to compute the ratio of clients outside and inside the droplet as a function of concentration of free type ”B” molecules for different linker lengths.

$$\begin{aligned} P_{cl} &= \frac{c_{clients\,inside}}{c_{clients\,outside}} \\ &= \frac{c_{bound\,cl} + c_{free\,cl}}{c_{free\,cl} + k_{cl} c_{free\,cl} c_{scaffold}} \end{aligned} \tag{3.20}$$

where $k_{cl} = e^{\epsilon_{cl}}$. We will be using the same binding energy used for the binding interactions between type ”A” and type ”B” for simplicity. We can obtain $c_{bound\,cl}$ using Eq. 3.17 as

$$c_{bound\,cl} = c_A \times n_{cl} \tag{3.21}$$

The concentration of scaffolds $c_{scaffold}$ is the concentration of type "A" molecules outside the droplet, which can be obtained from the equilibrium condition

$$\begin{aligned} \mu_{A\,out} &= \mu_{A\,in} \\ \ln\left(c_{scaffold}\right) + N\, f_M(z_{intra}) &\approx \ln(c_{A\,in}) + N\, f_M(z_{intra} + z_{MF}) + f_{blob} \end{aligned} \tag{3.22}$$

where N is the number of stickers in a type "A" chain, and $f_M(z_{intra})$ is the free energy without considering the Mean Field interaction. Notice that we are using the same number of stickers for both type "A" and type "B" molecules. plugging Eq. 3.22 in Eq. 3.20 we obtain the partition coefficient that gives the ratio of clients inside to clients inside the droplet.

Again we are interested in comparing the behavior of the different topologies, now regarding the partition coefficient behavior as a function of concentration of type B molecules. In Fig. 3.6 we can see the partition coefficient for the following topologies: zipper-like network ($R = 2nm$), mixed network ($R = 4nm$), and random network ($R = 6nm$) for the bivalent case (similar behavior is shown for other valences). We notice that in all cases the random network presents a higher partition coefficient, there are two plausible explanations: first, the random network is more disorder and presents more space for the clients to bind; and second, the zipper-like network inherently disfavors asymmetric stoichiometry, making that the scaffolds compete for clients outside the droplet. It can be concluded that if a high concentration of clients inside the droplet is what it is looked for, then a random network is preferable. However, if a fine control of the clients in the condensate is required, a zipper-like network should be considered.

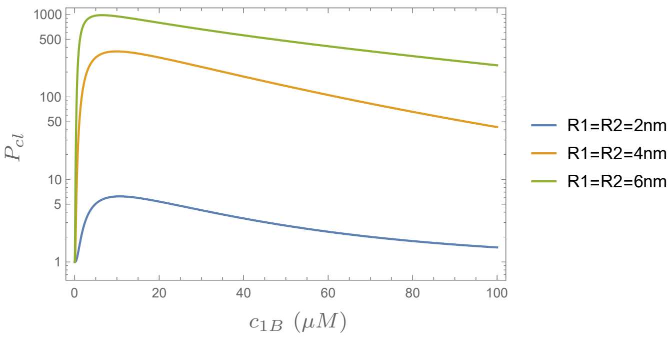


Figure 3.6: *Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$.*

In Fig. 3.7 we can see the behavior of the partition coefficient P_{cl} as a function of concentration of free type B molecules for two different client binding energies ϵ_{cl} at conditions $\epsilon = 2k_BT$, $c_{cl} = 50\mu M$ and $R_A = R_B = 6nm$, for the trivalent case (similar trends are present for all other cases) . In all cases we see that increasing the client binding energy results in a higher partition coefficients, meaning that a stronger affinity between client and the scaffold promotes a higher concentration inside the droplet.

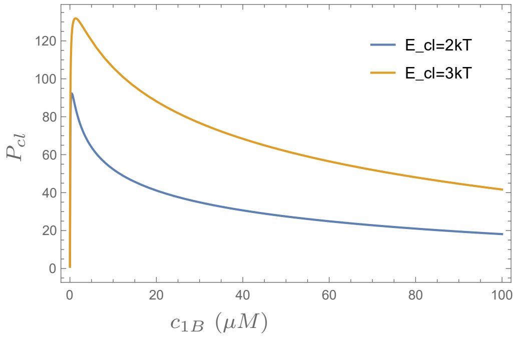


Figure 3.7: *Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$.*

In Fig. 3.8 we can see the behavior of the partition coefficient P_{cl} as a function of concentration of free type B molecules for two different client concentrations c_{cl} at conditions

$\epsilon = 2k_BT$, $\epsilon_{cl} = 2k_BT$ and $R_A = R_B = 2nm$, for the bivalent case (same behavior for all other cases). It is worth mentioning that from Eq. 3.20 it may seem that c_{cl} is factor out, but from Eq. 3.18 n_0 depends on c_{cl}, hence $c_{boundcl}$ depends on c_{cl} too. We can see that a higher client concentration results in larger partition coefficients. The higher concentration of scaffold inside the droplet means more spaces available for clients, hence higher client concentrations will result in higher partition coefficients.

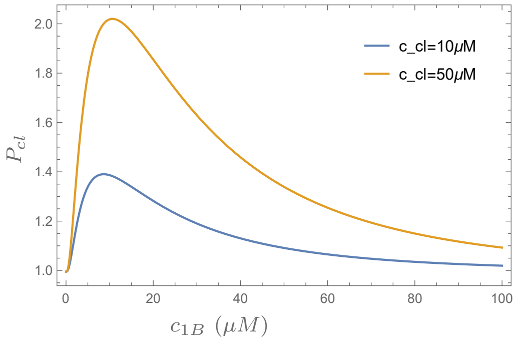


Figure 3.8: *Concentration of type "A" molecules c_A behavior inside the droplet as a function of concentration of free type "B" molecules c_{1B} for different symmetric linker lengths $R_A = R_B$.*

From Figs. 3.7 and 3.8 we can conclude that both the client binding energy ϵ_{cl} and the client concentration c_{cl} can be used as parameters to control the concentration of clients inside the droplet, since increasing or decreasing one has a similar effect on the partition coefficient.

Now we will focus on varying the network parameters. In Fig. 3.9 we can see the behavior of the partition coefficient P_{cl} as a function of concentration of free type B molecules for two different network binding energies ϵ at conditions $\epsilon_{cl} = 2k_BT$, $c_{cl} = 50\mu M$ and $R_A = R_B = 2nm$, for the bivalent, trivalent and n-valent cases. We see that increasing binding energy lowers the partition coefficient. Stronger affinities favors a tighter network, meaning less space for clients to bind to. Here we have a parameter that presents an interesting behavior, we have seen in Fig. 3.5 that for the bivalent case, varying the binding energy does not change the peak in concentration inside the droplet, but it does have an effect in the partition coefficient, so we have a network that is robust to changes in binding energy but it is responsive to client binding.

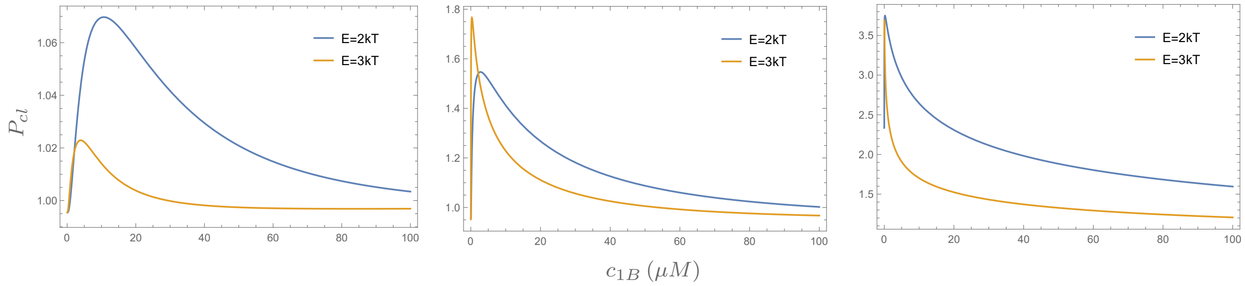


Figure 3.9: *Partition coefficient of clients P_{cl} Eq. 3.20 behavior as a function of free type "B" molecules c_{1B} for different binding energies.*

3.5 Conclusions

We have used in this work polymer theory of real chains in order to obtain the excluded volume interactions among polymers and developed a model for the binding interaction between sticky polymers.

We have seen in this work that the network formed by the sticky polymer can adopt different topologies inside the droplet depending on the polymer linker length, and a transition from one to the other is found. These different topologies have different behavior when it comes to density inside the droplet and client biding. We see in all cases a reentrant behavior, marking the conditions in which the condensed phase is possible, for each topology. The same reentrant behavior is seen for client binding.

It was expected that the mesh size of the network given by the linker length will determine a behavior for the concentration inside the droplet of inverse proportionality, but the opposite was found, and this was explained for the topologies that the network adopts. It was also found that longer linker lengths promotes the binding of clients inside the droplet.

It is worth mentioning that we have obtained an order parameter, the length between stickers, to determine the topology of the network and the values for this transition. This allows experiments and simulations to coarse-grain their search and analyze the systems for each one of these topologies.

3.6 Methods

3.6.1 Excluded Volume Free Energy

The number of monomers per blob is given by

$$\frac{g}{a^3 g^{3\nu}} = \frac{l_a + l_B}{V_{sticker}} \tag{3.23}$$

which result in

$$g = \frac{1}{\left(a^3\, c_A (l_a + l_B)\right)^{\frac{1}{3\nu-1}}} \tag{3.24}$$

Plugging this into Eq. 3.2 we get

$$\begin{aligned} \frac{f_{excluded}}{k_B T} &\sim \frac{l_A + l_B}{g} \\ &\sim (l_A + l_B)\left(a^3\, c_A (l_a + l_B)\right)^{\frac{1}{3\nu-1}} \\ &\sim (l_A + l_B)^{\frac{3\nu}{3\nu-1}} (a\, c_A)^{\frac{1}{3\nu-1}} \end{aligned} \tag{3.25}$$

and this is the expression in Eq. 3.3

3.6.2 Calculation of Entropy

The first step is to place the first site for each of the θ-bound type B molecules. The number of ways to do this is $M!/\left(N_\theta!\,(M - N_\theta)!\right)$. Next, we place the second binding site for each of these molecules. These sites are more restricted due to the fact they are constrained by the first bond. The number of nearby sites, z, depends on whether the type A molecule is in dilute phase or in a dense phase. However, we need to account for the probability that these sites are already occupied, For the placement of the first double attachment, N_θ out of M sites are already occupied, so the probability that each of the z sites is available is $\left(M - N_\theta\right)/M$. For the placement of the second free tail the probability the neighbouring sites are free is $\left(M - (N_\theta + 1)\right)/M$. Therefore, the number of ways of placing the second binding sites is

$$\begin{aligned}\left(z\frac{M - N_\theta}{M}\right) \times \left(z\frac{M - (N_\theta + 1)}{M}\right) \times \ldots \times \left(z\frac{M - (2N_\theta - 1)}{M}\right) \\ = \left(\frac{z}{M}\right)^{N_\theta} \frac{(M - N_\theta)!}{(M - 2N_\theta)!}\end{aligned} \tag{3.26}$$

Now for the next bond there are $(M - 2N_\theta)/M$ sites available to attach to a type B molecule. We can follow the exact same argument as before until we only have left the N_1 single bound type B molecules among the remaining $M - \theta N_\theta - \ldots - 2N_2$ sites leaving N_0 sites unbound. The number of ways to do this is $(M - \theta N_\theta - \ldots - 2N_2)!/(N_0!N_1!)$. So the total number of ways to arrange the type B molecules is

$$\Omega = \left(\frac{z}{M}\right)^{N_2 + 2N_3 + (\theta - 1)N_\theta} \frac{M!}{N_0!N_1!\ldots N_\theta!} \tag{3.27}$$

3.6.3 Free energy minimization

We now insert Ω into Eq. 3.4 and apply Stirling's approximation on the factorials. The free energy becomes

$$\begin{aligned}F_M/k_BT = & -\epsilon\left(N_1 + 2N_2 + \ldots + \theta N_\theta\right) - \mu_B\left(N_1 + N_2 + \ldots + N_\theta\right) \\ & + \sum_{i=2}^{\theta} N_i \ln\left(\frac{N_i}{M}\frac{1}{z^{i-1}}\right) + \sum_{i=2}^{\theta}(i-1)N_i + N_1 \ln\frac{N_1}{M} \\ & + N_0 \ln\frac{N_0}{M} + \lambda\left(M - N_0 - N_1 - 2N_2 - \ldots - \theta N_\theta\right)\end{aligned} \tag{3.28}$$

where λ is a Lagrange multiplier that will be used to constrain the total number of sites. For convenience this can be re-written as

$$f = F_M/K_BT = (1 + \lambda) + \sum_{i=0}^{\theta} n_i(\ln n_i - 1 + A_i) \tag{3.29}$$

where $n_i = N_i/M$ and

$$\begin{aligned}
A_0 &= -\lambda \\
A_1 &= -\epsilon - \mu_B - \lambda \\
A_2 &= -2\epsilon - \mu_B - 2\lambda - \ln z \\
&\dots \\
A_\theta &= -\theta\epsilon - \mu_B - \theta\lambda - (\theta - 2)\ln z
\end{aligned} \tag{3.30}$$

Minimizing the free energy with respect to the site occupancies yields $n_i = e^{-A_i}$, *or*

$$\begin{aligned}
n_0 &= e^{\lambda} \\
n_1 &= e^{\epsilon+\mu+\lambda} \\
n_2 &= e^{2\epsilon+\mu_B+2\lambda+\ln z} \\
&\dots \\
n_\theta &= e^{\theta\epsilon+\mu_B+\theta\lambda+(\theta-2)\ln z}
\end{aligned} \tag{3.31}$$

which can be expressed in terms of the concentration of unbound type B molecules using the relation $\mu_B = \ln c_{1B}$.

Inserting Eqs. 3.31 into Eq. 3.29 yields Eq.3.7. Using the condition $1 = n_0 + n_1 + 2n_2 + \dots + \theta n_\theta$ and the expressions in Eq. 3.31, gives the polynomial equation 3.9 for n_0. We choose the root that gives $n_0 \to 1$ as $c_{1B} \to 0$.

3.6.4 Large Valence

We are now interested in the case of polymers with large valence. Our starting point is going to be Eq. 3.9, we will use the approximation in Eq. 3.5 $f_j = \epsilon + \mu$

$$
\begin{aligned}
1 &= \left(1 + c_{1B}\exp^{\epsilon}\right) n_0 + \left(2c_{1B}z\exp^{2\epsilon}\right) n_0^2 + \ldots + \left(\theta c_{1B}z^{\theta-1}\exp^{\theta\epsilon}\right) n_0^{\theta} \\
&= n_0 + \frac{c_{1B}}{z}\sum_{i=1}^{\theta} i\,\left(z\,e^{\epsilon}n_0\right)^i \\
&= n_0 + \frac{c_{1B}}{z}\sum_{i=1}^{\theta} i\,x^i
\end{aligned}
\tag{3.32}
$$

where $x = z\,e^{\epsilon}n_0$. For $\theta \to \infty$ the last term in Eq. 3.32 looks like a power series that can be solved as

$$
\sum_{i=1}^{\theta} i\,x^i = \frac{x}{(x-1)^2}
\tag{3.33}
$$

for $x < 1$. Plugging Eq. 3.33 into Eq. 3.32 we obtain

$$
1 = n_0 + c_{1B}\frac{e^{\epsilon}n_0}{(z\,e^{\epsilon}n_0 - 1)^2}
\tag{3.34}
$$

This is a cubic equation in n_0 that can be re-written as

$$
0 = z^2\,e^{2\epsilon}n_0^3 - (z^2\,e^{2\epsilon} + 2z\,e^{\epsilon})n_0^2 + (2z\,e^{\epsilon} + c_{1B}e^{\epsilon} + 1)n_0 - 1
\tag{3.35}
$$

which can be solved analytically.

Next step is to re-write Eq. 3.7 for the case of large θ making use of Eq. 3.8:

$$
\begin{aligned}
f_{\theta} &= \ln\left(n_0\right) + n_2 + 2n_3 + \ldots + \left(\theta - 1\right) n_{\theta} \\
&= \ln(n_0) + \sum_{i=2}^{\theta}(i-1)n_i \\
&= \ln(n_0) + \frac{c_{1B}}{z}\sum_{i=1}^{\theta}(i-1)(z\,e^{\epsilon}n_0)^i \\
&= \ln(n_0) + \frac{c_{1B}}{z}\sum_{i=1}^{\theta}(i-1)\,x^i
\end{aligned}
\tag{3.36}
$$

where $x = z\,e^{\epsilon}n_0$. We need now to write an expression for the last sum for the case of large θ but finite.

$$\begin{aligned}\sum_{i=1}^{\theta}(i-1)\,x^i &= \sum_{i=1}^{\theta} i\,x^i - \sum_{i=1}^{\theta}\,x^i \\ &= 1 + x\left(\frac{1-(\theta+1)x^\theta+\theta x^{\theta+1}}{(x-1)^2}\right) + \frac{1-x^{\theta+1}}{x-1} \\ &= x\frac{x-\theta x^\theta+(\theta+1)x^{\theta+1}}{(x-1)^2}\end{aligned} \tag{3.37}$$

We can now plug this expression into Eq. 3.36 to obtain

$$f_\theta = \ln(n_0) + c_{1B}\,e^\epsilon n_0 \frac{z\,e^\epsilon n_0 - \theta(z\,e^\epsilon n_0)^\theta + (\theta+1)(z\,e^\epsilon n_0)^{\theta+1}}{(z\,e^\epsilon n_0-1)^2} \tag{3.38}$$

and for $\theta \to \infty$ we will have

$$f_{\theta\to\infty} = \ln(n_0) + c_{1B}\,z\,\frac{(\,e^\epsilon n_0)^2}{(z\,e^\epsilon n_0-1)^2} \tag{3.39}$$

We can now numerically solve Eq. 3.13 to obtain the concentration of type "A" molecules inside the condensed phase.

Bibliography

[1] Shin, Y. and Brangwynne, C. P., Science **357** (2017) eaaf4382.

[2] Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K., Nature Reviews Molecular Cell Biology **18** (2017) 285.

[3] Franzmann, T. M. et al., Science **359** (2018) eaao5654.

[4] Rai, A. K., Chen, J.-X., Selbach, M., and Pelkmans, L., Nature **559** (2018) 211.

[5] Nott, T. J. et al., Molecular Cell **57** (2015) 936.

[6] Riback, J. A. et al., Cell **168** (2017) 1028.

[7] Mao, Y. S., Zhang, B., and Spector, D. L., Trends in Genetics **27** (2011) 295.

[8] Decker, C. J. and Parker, R., Cold Spring Harbor Perspectives in Biology **4** (2012) a012286.

[9] Wu, H., Cell **153** (2013) 287.

[10] Pederson, T., Cold Spring Harbor Perspectives in Biology **3** (2010) a000638.

[11] Dundr, M. et al., Journal of Cell Biology **164** (2004) 831.

[12] Phair, R. D. and Misteli, T., Nature **404** (2000) 604.

[13] Platani, M., Goldberg, I., Swedlow, J. R., and Lamond, A. I., Journal of Cell Biology **151** (2000) 1561.

[14] Fu, L. et al., Molecular Biology of the Cell **16** (2005) 4905.

[15] Li, P. et al., Nature **483** (2012) 336.

[16] King, O. D., Gitler, A. D., and Shorter, J., Brain Research **1462** (2012) 61.

[17] Wang, J. et al., Cell **174** (2018) 688.

[18] Harmon, T. S., Holehouse, A. S., Rosen, M. K., and Pappu, R. V., eLife **6** (2017).

[19] Flory, P. J., *Principles of Polymer Chemistry (The George Fisher Baker Non-Resident Lectureship in Chemistry at Cornell University)*, NCROL, 1953.

[20] Dill, K. A. and Bromberg, S., *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience, 2nd Edition*, Garland Science, 2010.

[21] Bouchard, J. J. et al., Molecular Cell **72** (2018) 19.

[22] Banani, S. F. et al., Cell **166** (2016) 651.

[23] Bhandari, K. and Schmit, J. D., pending of publication .

Chapter 4

Protein-Protein binding under macroscopic salt conditions

4.1 Introduction

When studying biological macromolecules, the conformations, dynamics, function and binding process can be very sensitive to the salt concentration and composition of the local environment. In the living cell, ion composition and concentrations can differ between inter/intracellular environments [1] and organelles [2]. The local ionic concentration in the environment around real biological macromolecules, however, can significantly deviate from macroscopic concentrations.

Molecular Dynamics simulations are powerful tools in studying the structure, dynamics and functions of biomolecular systems in the atomic scale. Current state-of-the-art approaches simulate a small volume around the biomolecule using explicit atomistic solvent to model the local environment [3–6]. To emulate electrostatic screening effects in the local solvent environment, explicit ions are added, to achieve net neutrality. Nevertheless, statistical fluctuations in the total number of ions in the region around the biomolecule may result in significant variance in the local salt concentration, where relative concentration fluctuations diminish slowly with increasing simulation volume.

In this work we develop a model to account for the statistical fluctuations in ions concentrations inside the volume where the biomolecule is immersed. We will study particularly the phenomenon of protein binding and analyze how adding this statistical fluctuations in ions concentration affects the free energy of binding.

4.2 Theory

We will consider a solute of charge Q immersed in a box with volume V. We assume the box can exchange salt ions with a large solvent reservoir with a concentration c_0 of symmetric, monovalent salt. The grand canonical partition function for the system can be written as

$$\mathcal{Z} = \sum_{N_+,N_-}^{\infty} e^{\frac{N_+\mu}{k_BT}} e^{\frac{N_-\mu}{k_BT}} \frac{V^{N_+}}{N_+!} \frac{V^{N_-}}{N_-!} e^{\frac{E(Q,N_+,N_-)}{k_BT}} \tag{4.1}$$

where N_+ and N_- are the number of positive and negative ions inside the box respectively, and μ is the chemical potential for the exchange of ions. Since the box is in contact with a bath of ions $\mu = k_BT\ \ln(c_0)$.

We now need to write the energy term, in first approximation we can assume that the charge is spread inside the box like a background charge. In this way, there will be an electrostatic contribution to the energy like a Bohr term, in other words, we compute the energy of bringing charges to the surface of a sphere of radius R:

$$E(Q, N_+, N_-) = \frac{1}{8\pi\epsilon_0}\frac{1}{R}e^2\left(q + N_+ - N_-\right)^2 \tag{4.2}$$

where $R^3 \sim V$, $Q = e\,q$ and e is the electron charge.

plugging this result into Eq. 4.1 we obtain

$$\mathcal{Z} = \sum_{N_+,N_-}^{\infty} e^{\frac{N_+\mu}{k_BT}} e^{\frac{N_-\mu}{k_BT}} \frac{V^{N_+}}{N_+!} \frac{V^{N_-}}{N_-!} e^{-\frac{l_B}{2R}(q+N_++N_-)^2} \tag{4.3}$$

where $l_B = e^2/(4\pi\epsilon_0 K_BT)$ is the Bjerrum length.

We can now write the grand canonical partition function as a function of the concentration of ions in the bath c_0:

$$\begin{aligned} \mathcal{Z} &= \sum_{N_+,N_-}^{\infty} c_0^{N_+} c_0^{N_-} \frac{V^{N_+}}{N_+!} \frac{V^{N_-}}{N_-!} e^{-\frac{l_B}{2R}(q+N_++N_-)^2} \\ &= \sum_{N_+,N_-}^{\infty} \frac{(Vc_0)^{N_++N_-}}{N_+!N_-!} e^{-\frac{l_B}{2R}(q+N_+-N_-)^2} \end{aligned} \tag{4.4}$$

From this expression we can obtain the free energy

$$F_{\mathcal{Z}} = -k_BT \ln(\mathcal{Z}) \tag{4.5}$$

We want to compare this result to the case in which there is a fixed number of ions inside the box and no fluctuations in its number. In order to compute the number of ions in the box we follow Schmit et al [7] and compute the average number of ions inside a simulation box in contact with a reservoir of ions of concentration c_0.

The solute of charge Q and volume v_p is immersed in a box with volume $v_t = v_p + v_w$, where v_w is the volume occupied by solvent. Inside the box, the ion concentrations are perturbed by interactions with the solute charges. The perturbed concentrations are related to the electrostatic potential, ϕ, by

$$\begin{aligned} c_+(\boldsymbol{x}) &= c0e^{-e\phi(\boldsymbol{x})/kT} \\ c_-(\boldsymbol{x}) &= c0e^{e\phi(\boldsymbol{x})/kT} \end{aligned} \tag{4.6}$$

The total number of ions in the box can be obtained by integrating over the solvent volume:

$$N_\pm = \int_{v_w} c_+(\boldsymbol{x}) = c0e^{\mp e\phi(\boldsymbol{x})/kT} \mathrm{d}^3\boldsymbol{x} \tag{4.7}$$

where e is the electron charge, kB is the Boltzmann constant, T is the absolute temperature, and the integral is evaluated over the ion accessible volume. Next step is to replace the spatially dependent potential $\phi(\boldsymbol{x})$ with an average potential $\overline{\phi}$ that is uniform across the simulation box. With this approximation, Eq. 4.7 becomes $N_\pm = v_w c_0 e^{\mp e\overline{\phi}/kT}$. The value of the average potential can be determined from the charge neutrality condition $Q/e = N_+ + N_-$ which yields

$$\frac{e\overline{\phi}}{kT} = \mathrm{ArcSinh}\left(\frac{Q}{2ev_wc_0}\right) \tag{4.8}$$

The required ion numbers are then given by

$$N_\pm = \nu_w c_0 \exp\left[\mp \mathrm{ArcSinh}\left(\frac{Q}{2e\nu_w c_0}\right)\right] \tag{4.9}$$

where ν_w is the volume occupied by the solvent. In our case $\nu_w = V$ since the solute is only background charge.

We can now write the canonical partition function for this system

$$\mathcal{Q} = \frac{V^{N_+}}{N_+!}\frac{V^{N_-}}{N_-!}e^{-\frac{l_B}{2R}(q+N_+-N_-)^2} \tag{4.10}$$

We want to compute the difference between allowing fluctuations and not allowing fluctuations in the event of protein binding. We will model the protein binding event as having two boxes of solute charge q each (separate proteins) and having the total charge $2q$ in one box and the other box empty (bound proteins), as shown in the picture:

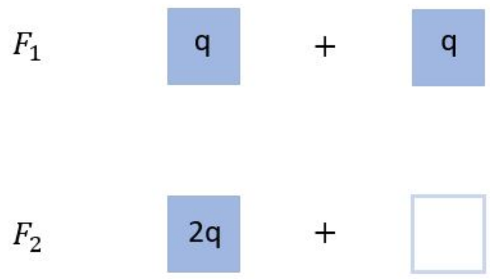


Figure 4.1: *Modeling the protein binding process as having two separate events: 1. Two boxes with charge* q *each represent the proteins separate by an infinite distance. 2. there is one box with charge* $2q$ *and one empty box, representing having the two proteins bound to each other.*

We will compute the free energy of each one of the events F_1 and F_2 and get the difference in free energy of the process $\Delta F = F_2 - F_1$. We will do this for both cases, with and without fluctuations.

In Fig. 4.2 we see the behavior of ΔF as a function of the solute charge Q for a fixed box length $R = 7nm$ and a bulk salt concentration $c_0 = 200mM$, for both cases with and without fluctuations. We first notice that in both cases the difference in free energy increases monotonically as the solute charge increases: the barrier for binding gets larger as the charge increases. We next notice that the difference between allowing and not allowing ions fluctuations gets larger as we increase the solute charge.

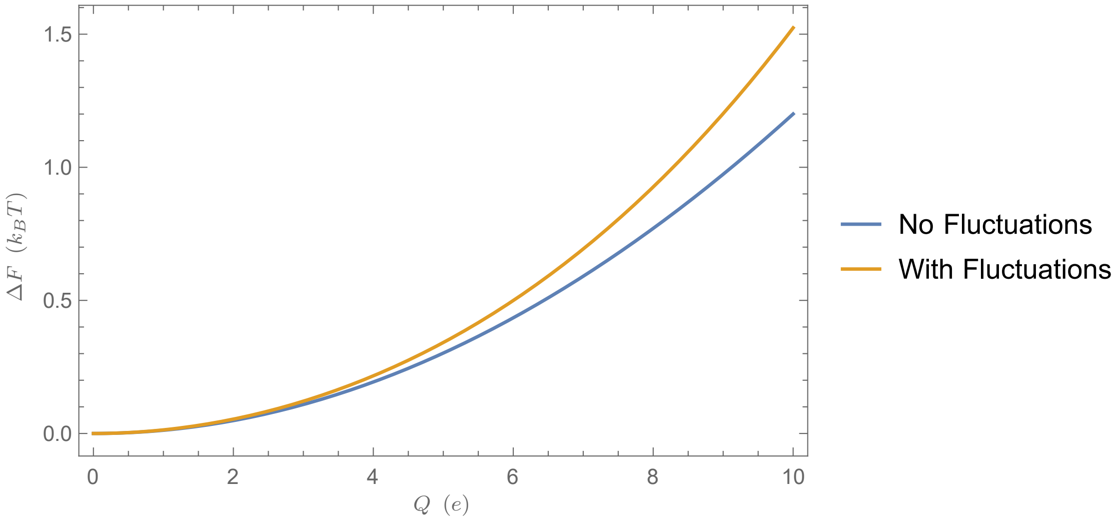


Figure 4.2: ΔF *behavior as a function of the solute charge* Q *for the cases with and without fluctuations. We see that as the charge gets larger the difference in energy also increases.*

In Fig. 4.3 we see the behavior of ΔF as a function of the box length R for a fixed solute charge $Q = 6e$ and a bulk salt concentration $c_0 = 200mM$, for both cases with and without salt fluctuations. We first notice that in both cases ΔF present a non monotonic behavior. Next we notice that there is region where there is no much difference between allowing and not allowing salt fluctuations.

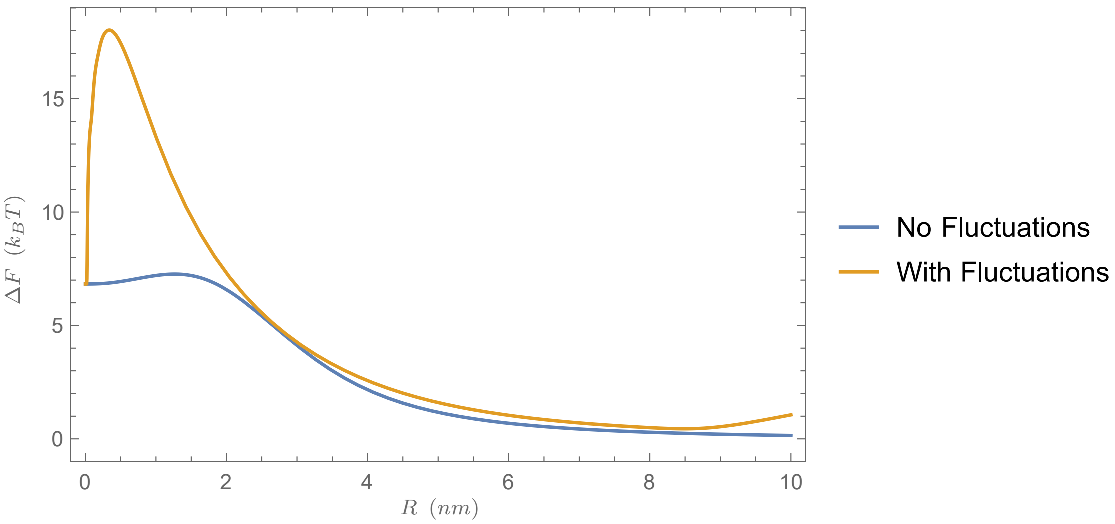


Figure 4.3: *ΔF behavior as a function of box length R for the cases with and without fluctuations. We see a non monotonic behavior in both cases, and the case with fluctuations is always larger than the case without them.*

4.3 Conclusions

We have used in this work Statistical Mechanics tools, such as, canonical partition functions and grand canonical partition functions in order to obtain the free energy of the protein binding event.

We have shown in this work a calculation for the Free Energy of an event of two protein binding, where fluctuations of ions was allowed and compare the result to the Free Energy obtained when salt concentration is kept constant. We have pointed the regions where there is a difference in Free Energy, which allows future simulations to consider whether adding ions fluctuations may affect the the results.

Bibliography

[1] Phillips, *Physical Biology Of The Cell, 2/E (Pb-2013)*, Garland Science, 2019.

[2] Moore, R. and Morrill, G., Biophysical Journal **16** (1976) 527.

[3] Abraham, M. J. et al., SoftwareX **1-2** (2015) 19.

[4] Jo, S., Kim, T., Iyer, V. G., and Im, W., Journal of Computational Chemistry **29** (2008) 1859.

[5] Eastman, P. et al., Journal of Chemical Theory and Computation **9** (2012) 461.

[6] Eastman, P. et al., PLOS Computational Biology **13** (2017) e1005659.

[7] Schmit, J. D., Kariyawasam, N. L., Needham, V., and Smith, P. E., Journal of Chemical Theory and Computation **14** (2018) 1823.

Appendix A

Corrections to the Partition Function

As mentioned in the main text, the partition function for the HT model overestimates the number of states due to over-counting of cluster degeneracies. Here we correct the over-counting of states and compare it to the approximate form previously obtained. Following similar works, we neglect the formation of closed loops [1].

An inspection of the Taylor expansion of the partition function for the HT model shows that starting at the dimer level (c_1^2) there is already an over-counting of states:

$$q_{\text{HT}} = c_1 \left(1 + 2\,(s\,c_1) + 5\,(s\,c_1)^2 + \cdots\right). \tag{A.1}$$

The partition function shows two possible states, whereas it is clear that those states are related by a rotation around the symmetry axis of the first molecule (see fig. A.1), and should only be counted once in the partition function. While the partition function q_{HT} over-counts the number of clusters (fig 9A), the recursion formula does correctly count the number of states branching off the arms of the base molecule. This is because the presence of the starting molecule breaks the rotational symmetry (fig 9B).

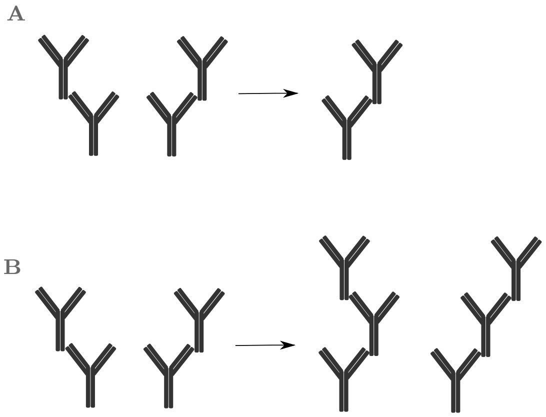


Figure A.1: *(A) The partition function in the main text gives two possible structures for dimers, but those are related to each other by a rotation around a symmetry axis. (B) The same degenerate dimers, considered as part of a trimer, give two different trimer structures. This idea will exploited to count the total number of states.*

The first attempt is to assume that, neglecting the monomer term, it would be enough to divide by two in order to get rid of the spurious rotation states (Eq. 2.9). Unfortunately this approach does not solve the over-counting problem. A closer look reveals that some states are not repeated twice in the partition function. Those states arise when considering structures with the same number of antibodies in each arm (Fig. A.2 shows an example with $N = 5$). Fig. A.2A shows two structures that are related by a rotation around the symmetry axis, and this cluster is counted twice in the partition function. On the other hand figure A.2B shows a structure that is only counted once.

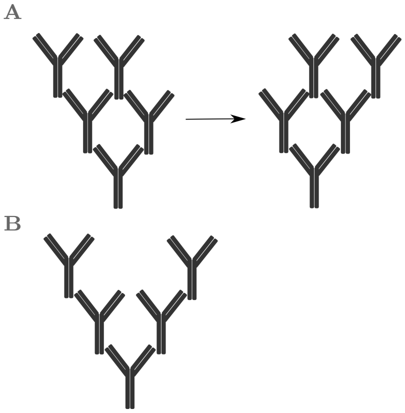


Figure A.2: *(A) Structure with same amount of mAbs in each arm is counted twice in the partition function, since a rotation around the symmetry axis is also included in the sum. (B) A a structure with rotational symmetry.*

The states that are single counted in the partition function are those that already posses rotational symmetry. Therefore it is possible to compute the partition function by artificially adding these symmetric states once more to the sum and then dividing q_{HT} by two in order to get the partition function corrected for symmetry over-counting.

$$\tilde{q}_{\text{HT}} = \frac{q_{\text{HT}} + q_{\text{sym}}}{2} \tag{A.2}$$

where the correction factor is given by

$$q_{\text{sym}} = c_1 + c_1 \sum_{i=1}^{\infty} \frac{1}{(i-1)!} \left. \frac{d^{i-1}q}{dc_1^{i-1}} \right|_{c_1=0} (sc_1)^{2i} \tag{A.3}$$

To derive this expression we note that each term in the uncorrected partition function q_{HT} can be used to generate a symmetric structure. This is done by taking each structure of size i, duplicating it, and add each copy to the heads of a base molecule to generate a structure of size $2i+1$. The statistical weight of these clusters is $c_1(sc_1)^{2i}$ times the degeneracy factor given by Eq. 2.14. This product is summed over cluster sizes to give Eq. A.3. By Taylor expanding this expression we can see the number of rotationally symmetric clusters

$$q_{\mathrm{sym}} = c_1 + 1\, s^2 c_1^3 + 2\, s^4 c_1^5 + 5\, s^6 c_1^7 + 14\, s^8 c_1^9 + \cdots \,. \tag{A.4}$$

The next step is to add Eq. A.3 to the partition function computed in the main text Eq. 2.8 and divide the resulting expression by two (neglecting the monomer state). The first few terms of the corrected partition function for the HT model are given by

$$\tilde{q}_{\mathrm{HT}} = c_1 + s\, c_1^2 + 3s^2\, c_1^3 + 7s^3\, c_1^4 + 22s^4\, c_1^5 + \cdots \,. \tag{A.5}$$

The clusters represented by these terms are shown in Fig. A.3.

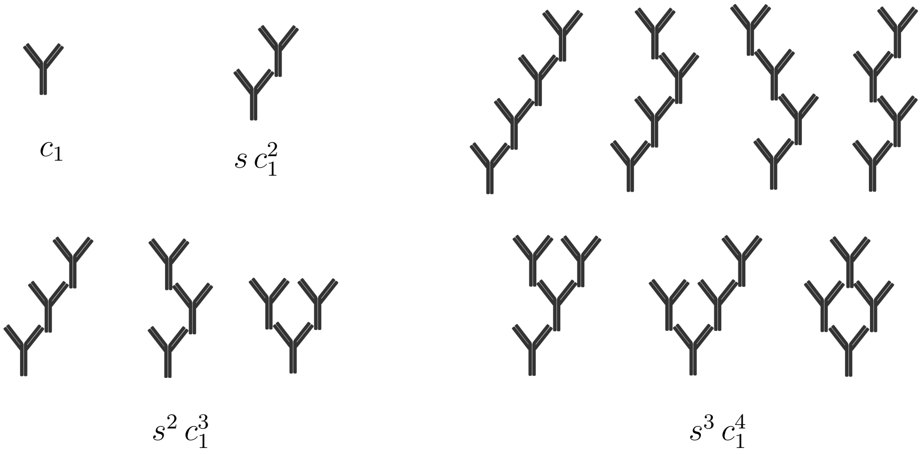


Figure A.3: *sketches of allowed structures' types in the HT model from monomer to trimers.*

Now that we have the corrected for symmetry over-counting solution for the partition function, we evaluate the quality of the approximate partition function derived from the recursion relationships by comparing it to the corrected value. Given the fact most of the physical quantities obtained in the text are functions of the number of monomers units in the cluster, it is natural to take $\langle n \rangle$ as the quantity to compare to.

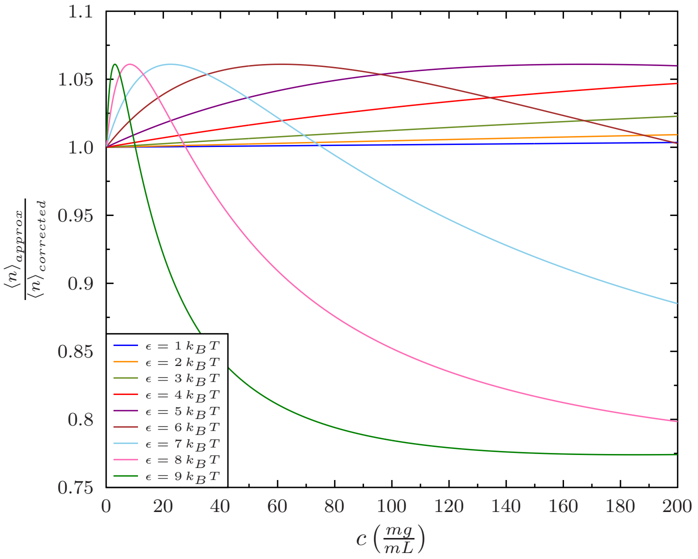


Figure A.4: *Comparison of the average cluster sizes for approximate and corrected solution for different binding energies:* $1 - 9\, k_B T$ *as a function of concentration. As the interaction energy gets stronger the difference is more noticeable, but even for the largest case it is around 20%.*

Figure A.4 shows the comparison between the average size obtained with the approximation in the main text ($\langle n \rangle_{\text{approx}}$) and the one corrected for rotational symmetry ($\langle n \rangle_{\text{corrected}}$), for different values of the binding energy ϵ_{HT}, where $s = c^{-1} e^{\epsilon_{\text{HT}}/k_B T}$. It can be seen that the main text $\langle n \rangle$ gives a good approximation to the expected values. For stronger binding energies the deviation is close to 20% for high concentrations.

Bibliography

[1] Rubinstein, M. and Colby, R., *Polymer Physics*, Oxford University Press, Oxford, U.K., 2003.