Computational Insights into Multivalently Binding Polymers

by

Emiko Zumbro

S.B., Harvard University (2013)

Submitted to the Department of Materials Science and Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Materials Science and Engineering
and the Program in Polymers and Soft Matter

at the

Massachusetts Institute of Technology

September 2020

Author .....
Department of Materials Science and Engineering
August 7, 2020

Certified by .....
Alfredo Alexander-Katz
Associate Professor of Materials Science and Engineering
Thesis Supervisor

Accepted by .....
Frances M. Ross
Chair, Departmental Committee on Graduate Studies

Computational Insights into Multivalently Binding Polymers

by
Emiko Zumbro

Submitted to the Department of Materials Science and Engineering
on August 7, 2020, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Materials Science and Engineering
and the Program in Polymers and Soft Matter

Abstract

Multivalent binding is commonly used throughout biology to create strong, conformal bonds using multiple weak binding interactions simultaneously. Bonds are considered multivalent when multiple ligands on one species simultaneously bind to multiple receptors on another species. Together, this bond can be much stronger than the sum of its parts. Throughout this thesis, we use theory and coarse-grain Brownian dynamics simulations with specific reactive-binding to explore general characteristics of multivalent polymer interactions. Our simulations bridge length and timescales and can sample large polymer systems that bind targets at the sub-nanometer lengthscale. While the simulation and theory presented is very general and can be applied to many different systems of multivalent polymers, this thesis specifically explores consequences for two applications: multivalent polymers as decoys to inhibit infection and polymers as scaffolds for biocondensates.

Many pathogens use multivalent bonds to attach to our cell surfaces before entering and causing infection. Therefore, there is significant interest in preventing infection from viruses, bacteria, and toxic proteins by inhibiting this attachment step using multivalent decoys. There have been many experiments showing successful binding of long polymers or other large multivalent architectures to colloids or small proteins that pathogens use to bind to our cells. While these experiments have shown how promising multivalent inhibitors are for preventing infection, a theoretical understanding of why design parameters of multivalent polymers result in a particular binding affinity is still missing. Simulations can easily isolate a single design parameter to provide direct links between structure and function, when experiments cannot always do so. This research is intended to provide a systematic study linking structure of multivalent polymers to their binding behavior.

In the first half of this thesis, we explore design properties of polymeric binders and how degree of polymerization, solvent quality, binding site affinity patterns, backbone stiffness, and target concentration change the multivalent binding affinity. We provide simple theory to show that multivalent polymers are limited by their ability and the energetic costs of forming polymer loops. We go on to show how these results and theory have implications on the binding affinity of polymers with heterogeneous binding sites and determines the effect of polymer backbone flexibility and solvent quality on binding affinity.

Multivalent polymers are also an essential component of biocondensates, liquid-like droplets comprised of proteins and nucleic acids are found throughout cells. Although the function of these biocondensates is still an active field of study, it is clear that multivalent polymers are essential to their formation through liquid-liquid phase separation (LLPS). There is little theoretical study of biocondensates that contain binding between species of asymmetric

size and valency and the effects of multivalent polymers on the dynamics of these liquid droplets is not well understood. Studying how multivalent polymers modulate droplet dynamics is important because droplet crystallization or solidification is often associated with neurodegenerative disorders such as dementia and amyotrophic lateral sclerosis (ALS).

Therefore, in the second half of this thesis, we present research on the role multivalent polymers play in LLPS droplets and their resulting dynamics. We consider how a host of design parameters can change the phase boundary of systems with multivalent polymers binding to smaller targets including solvent quality, valency, binding affinity, specific versus non-specific binding sites, and backbone stiffness. We found that consistent with previous work on other systems, asymmetric valency systems also showed increased phase separation with increased binding affinities and valencies. We show that phase separation due to non-specific bonds is highly sensitive to changes in attraction, but that phase separation through specific-bonds is much more robust. By combining specific and non-specific multivalency, systems can precisely tune the phase separation boundary. Polymer stiffness can also modulate the phase boundary, where stiff, rod-like polymers were less able to cause phase separation than their flexible counterparts. We also elucidate how polymer-target binding affinities can be used to form micro-phase separated droplets. Lastly, we show that increasing attraction to polymers can slow target diffusion inside droplets while decreasing the density of droplets, with implications for droplet solidification.

We hope that this work will provide direction for the rational design of synthetic multivalent polymer systems such as pathogen inhibitors as well as improve understanding of native biological systems like biocondensates.

Thesis Supervisor: Alfredo Alexander-Katz
Title: Associate Professor of Materials Science and Engineering

Acknowledgments

This work could not have happened without the help of my many supporters.

First and foremost, I'd like to thank Alfredo for being the best advisor and friend a graduate student could ever ask for. Alfredo was first recommended to me as a great scientist and advisor before I became a graduate student when I was actually visiting another school, and he has definitely lived up to his reputation. I entered graduate school with no previous education on soft materials and no traditional laboratory experience, but Alfredo wasn't deterred from welcoming me into his research group. Alfredo is a brilliant scientist, who has spent countless hours discussing research problems with me, practicing for qualifying exams, and teaching me polymer physics. No research problem is too small or too big for him to be fully dedicated to solving, and he always knew just what to say to encourage me when a project wasn't going well or started to lose steam. Alfredo deeply cares about all of his students, as evidenced by him winning every award there is at MIT for his advising. He places equal emphasis on working hard and taking time for our families, our friends, and our wellbeing, all of which make up so much of the graduate student experience. His compassion and support have carried me through my graduate experience, and I could not be more grateful to him.

To the Alexander-Katz group: Shayna, Hejin, Karim, Mukarram, Shahrzad, Zhen, Josh, Yi, and Pierre, you are all amazing. I feel so lucky to have been a part of this group. When I joined the group, one of you told me that it was the best group because it is like a little family, and that couldn't be more accurate. Even though all of our projects are separate, each of you would drop anything to help me on a research or homework problem I was stuck on, practice a conference talk, study for quals, or just take a snack break. Even after graduating and moving away, you're still so generous with your time and energy. So much of my time here was shaped by having such wonderful people to share my office with, and I am very thankful to have gained such incredible lifelong friends throughout this experience.

Thank you to my thesis committee, Prof. Darrell Irvine, Prof. Katharina Ribbeck, and Prof. Jim Swan. You are all amazing scientists and have provided me with many insightful scientific suggestions over the past 5 years that have greatly improved the quality of my thesis research.

A huge part of the grad school is all of the wonderful people you meet outside of the lab, and I have met so many friends all over MIT. You have brought so much joy and fun to my graduate experience. A special thank you to Niels and the Holten-Andersen lab, the Ribbeck lab

who welcomed me with open arms and taught me everything they knew about mucus, the Polymer Graduate Student Association, TD's cohort from Nuclear Science and Engineering, and all of the wonderful friends in my materials science cohort especially Hugo, PJ, and Irina.

To my friends from before grad school, thank you for constantly checking on me and making sure I sometimes left the MIT bubble. I feel your love and support always. You all inspire me every day with all the amazing things you're accomplishing and I am excited to continuing watching you grow.

Thank you to my extended family, family friends, and in-laws. The MacDonald family treated me like one of their own since the first time I met them. They always make me smile when I need it, and I can't imagine a family I could be happier to join. To my extended family and family friends, you have helped me every step of the way with your endless love and support. From Friday night dinners to helping with grade school science projects to putting together a last minute wedding, you have all contributed so much to getting me here. Thank you to Katy and Christina for making me an Aunt by giving us Tilly and Billie, and bringing so much happiness when our family needed it most.

Thank you to my siblings, Mari and Austin, who have been there since day 1, supporting me and challenging me to be the best person I can. Thank you for always stepping up when I need you, and for your endless love. Thank you to my mom and dad, for their unconditional love and support. Thank you for always pushing us to be the best we could be, for building the stable foundation and loving family that made this all possible, and thank you for for always being my biggest cheerleaders. With a dad as a math teacher and a mom as an elevator mechanic, one of your children had to turn out to be a scientist. I love you all so much and feel incredibly lucky to have you.

Last but not least, thank you to my husband TD for being there for every day of my PhD. The best part of this experience has been meeting you. You have been my rock and best friend throughout all of this, and I could not have done it without you. I love you very much.

Contents

This Page Intentionally Left Blank

List of Figures

This Page Intentionally Left Blank

List of Tables

This Page Intentionally Left Blank

Chapter 1

Introduction

Multivalent binding interactions are used throughout biology and synthetic systems to enhance weak, monovalent binding between molecules or surfaces. Multivalent binding occurs when multiple ligands on one species interact with multiple receptors on another species simultaneously. In biology, multivalency has evolved for a variety of reasons including enhancing weak monovalent interactions, creating conformal interfaces such as those between cells or those inducing endocytosis, and increasing specificity of binding using a limited number of receptor and ligand types [1,2]. Examples of native multivalency include targeting of antibodies, endocytosis, binding of a viral or bacterial pathogen to a host cell, and cell-cell adhesion [1,3,4].

1.1 Thermodynamics and kinetics of multivalency

Multivalency is a robust strategy for increasing binding affinities of individually low affinity ligands, with the energetic benefits of multivalency explained in different ways throughout literature such as decreased loss of entropy, increased local binding site concentrations, and increased rebinding [1,5,6]. Mammen et al. provides a clear thermodynamic description of multivalent binding where binding of a target depends on contributions from enthalpy and entropy. When a ligand and receptor bind, enthalpy is gained from favorable ligand-receptor interactions, such as hydrophobicity or charge, while entropy is lost from the decrease in rotational and translational degrees of freedom. Although there can be enthalpic penalties for a multivalent species where the linker between ligands has to stretch or compress to match the receptor spacing on the target, to a first approximation, the change in enthalpy of a multivalent species with valency n is approximately equal to the enthalpy change of n individual or “monovalent" binding events. In contrast, the entropy of multivalent

versus monovalent binding can be very different. Translational and rotational entropy are only weakly dependent on particle mass and dimensions, so the translational and rotational entopy cost of binding is roughly the same for a monovalent or multivalent species [1]. Therefore, binding of n monovalent targets costs entropy:

$$\Delta S^{\text{mono}} = n(\Delta S_{\text{trans}} + \Delta S_{\text{rot}}) \tag{1.1}$$

while binding of an n-valent target costs only:

$$\Delta S^{\text{multi}} = \Delta S_{\text{trans}} + \Delta S_{\text{rot}} + \Delta S_{\text{conf}} \tag{1.2}$$

where S_{trans}, S_{rot}, and S_{conf} represent the translational, rotational, and conformational entropy respectively. In the case of a multivalent species with rigid linkers between binding sites, $\Delta S_{\text{conf}} = 0$ because there is only one molecule conformation. Therefore, the entropic cost of binding for a multivalent target is approximately the monovalent entropy cost divided by a factor of n, clearly demonstrating the increased binding free energy of multivalency. If the linker between sites on a multivalent binder are flexible, available conformations are reduced upon binding, but as long as the conformational entropy cost is less than $(n-1)(\Delta S_{\text{trans}} + \Delta S_{\text{rot}})$, the n-valent species will have enhanced avidity over n monovalent binders. This demonstrates the entropic avidity enhancement of multivalency.

Another way to think about the enhanced avidity of multivalent binders is through kinetic effects. While multivalency may not improve the k_{on} or initial rate of binding of a multivalent polymer to a receptor, once a multivalent species makes an initial bond, there is a higher change of making additional bonds. This is because multivalent species create an increased local concentration of ligands for the receptors to encounter [5]. This is not true in monovalent species where the ligands would be relatively dispersed throughout the system.

A third description of the avidity enhancement of multivalency is described by Weber et al. as an increased “rebinding effect" [6]. To imagine this effect, think of a piece of Velcro. If some of the hooks in the center of the strip get detached but the two ends are still stuck, they center hooks are still held nearby the free loops they were previously attached to. Because recently unbound hooks are trapped close to unbound loops, they are very likely to rebind once they unbind. In order to unbind a Velcro strip, the user has to pull it from one side in a peeling motion. This type of directional applied force is unlikely to happen randomly, and therefore, once a multivalent species is

bound, it has a much lower rate of unbinding, k_{off}, than monovalent equivalents (whose unbinding is uncorrelated with it's neighbors).

1.2 Functional benefits of multivalency

In addition to enhanced avidity, multivalency has many other advantages, including enhanced specificity. In native immune systems, antibodies use multivalency to target pathogens with highly specificity [1,4]. In synthetic systems, multivalent species have shown properties of "super-selectivity" [7–10]. The concentration of monovalent nanoparticles bound to a surface was shown to increase proportionally to the concentration of surface receptors, but multivalent nanoparticles demonstrated almost switchlike behavior, where, upon reaching a threshold surface receptor concentration, the concentration of bound nanoparticles increased almost exponentially [11]. Nanoparticles with multiple ligand types were also shown to superselectively target surfaces with similar concentration ratios of ligand types [12]. The specificity of multivalent binding makes it ideal for targeting tumor cells that often display higher concentrations of binding motifs [12,13].

Multivalency is also an efficient way to evolve new binding interactions; instead of creating a completely new molecule, organisms could just repeat the same ligand [2,1]. This same concept is helpful in a synthetic context where chemists may be able to use a simpler, weaker binding ligand and graft it to a polymer instead of searching for a complex and highly specific monomer that exactly matches a receptor binding site. Because mono and multivalent avidities can vary greatly, valency also provides a simple way to tune interaction strengths and signals.

Many architectures can be used to display multivalency, and have been investigated as multivalent pathogen inhibitors with benefits and drawbacks of each described by S. Bhatia *et al* [14]. Architecture possibilities include nanoparticles, dendrimers, linear random coil polymers, bottlebrush polymers, and sheets. With such a large design space available, we chose to reduce the number of variables we consider by focusing only on linear polymers. A wide variety of synthesis techniques can produce linear polymers [15], many experimental research studies have been conducted on the binding of multivalent linear polymers [16–21], and at least one study found them to be more potent than other architectures *in vivo* [22]. The flexibility of multivalency, enhanced affinity, and improved specificity all using a limited number of binding motifs have made multivalent binders prevalent in native systems and have made synthetic multivalent binders a promising class of therapeutics.

1.3 Thesis overview

Throughout this work, we focus on two examples of linear multivalent polymers from biology, mucins and the multivalent proteins and nucleic acids that control liquid-liquid phase separation in biocondensates. Mucins, the main gel-forming polymer found in mucus, display multiple sugar binding sites that act as decoys to prevent pathogens from binding to the glycocalyx on the surface of epithelial cells [23–25].

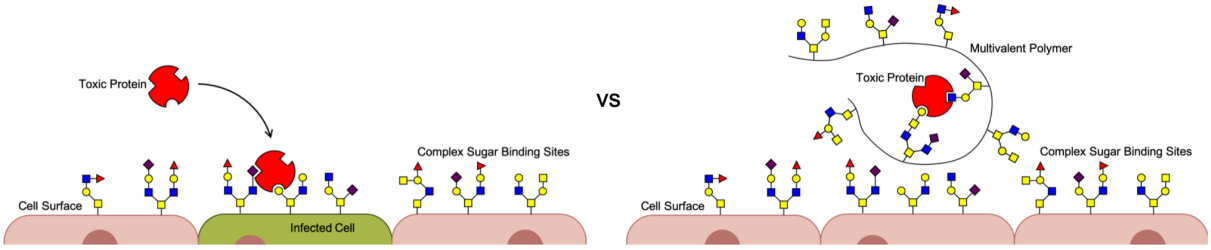


Figure 1-1: Pathogenic proteins can bind to the complex sugars on the surface of our cells as one of the first steps of infection. Using multivalent polymers as decoys is a promising avenue for blocking the binding of pathogens and the resulting infection of healthy cells. In this work we are interested in how the properties of multivalent polymers change their binding to these toxic proteins.

Often, pathogens use sugar-binding proteins called lectins, on their surface to bind to the sugar brushes on human cells in order to enter and cause harm. Many experimental and theoretical studies have been conducted to explore binding to these lectins to inhibit their attachment to the cell surface. If pathogens cannot bind to cells, this would prevent infection and allow the pathogens to wash out of the body under normal digestive flow. Therapeutics that target the binding mechanism of pathogens or their toxins instead of killing are a promising avenue for treating viral infections and for fighting bacterial infections without promoting antibiotic resistance [3,26–30]. Because inhibiting binding won't promote drug resistence, the lack of treatments for viral infections, and the versatility of multivalent platforms, linear multivalent polymers have drawn significant interest. Experimentalists have used multivalent polymers to successfully target a variety of lectins and the influenza virus [31,32,21,17,15]. While there are also many theoretical studies of multivalent binding, they have primarily focused on binding between two species of similar size [6,20,27,33–35], with relatively few theoretical or computational studies on how large polyvalent materials much larger than their targets, such as native mucins, control polyvalent interactions [1]. Amongst theoretical studies of large many-valent polymers, studies have focused on binding to larger surfaces instead of to small globular proteins such as lectins, despite the many experiments that target this size

regime [36,12,11]. Due to this lack of theoretical study, the importance of polymeric inhibitors of multivalent lectins, and the changes in polymer conformations induced by small colloids that are unique to this size regime[37,38], this work will focus on how design properties of long polymers modulate their binding to smaller targets. This study begins with the effect of polymer length on mono and multivalent binding in Chapter 3.

For targeting tumors or diseased cells, there has been an effort to make multivalent inhibitors more specific so that they bind to a single targeted species with higher affinity [36,39,10,13]. Attempts to make broad spectrum binders like mucins, which have multiple complex binding moieties [40,41] and may be high affinity to many targets is relatively unstudied. Therefore, in Chapter 4, we consider the effects of binding site affinity patterns and heterogeneity along a polymer chain. Having multiple binding site types is an important first step toward building multifunctional polymers that can bind to multiple species or include moieties for tracking as well as binding [42].

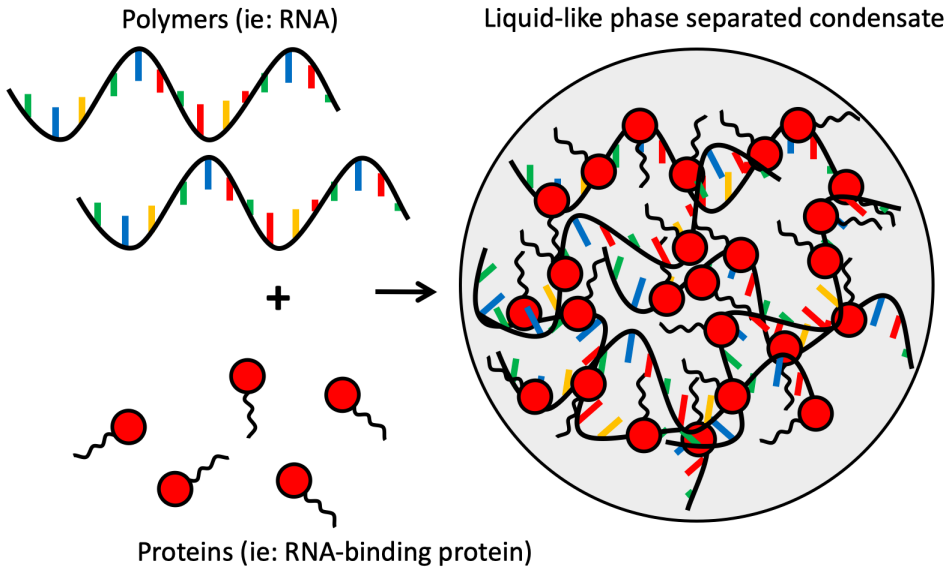


Figure 1-2: Cells contain liquid-like droplets called biocondensates that contain multivalent polymers and their cognate binding proteins. The multivalent polymers are depicted as RNA, but can also be linear multivalent proteins. We are interested in how properties of multivalent polymers can alter the phase separation that leads to these condensed droplets and the resulting dynamics of the droplet.

In addition to pathogen inhibition, we also consider multivalent polymers in the context of biocondensates, also referred to as membraneless organelles. Biocondensates are condensed polymer droplets that form through the liquid-liquid phase separation of multivalent proteins and nucleic acids inside cells. These bodies have liquid-like behavior and can contain tens to hundreds of

components [43–48]. The function of biocondensates is still an active field of study, but they are essential to for healthy cell activity. Recent research studies suggest that biocondensates can buffer concentrations of critical cell components and act as reaction crucibles by collecting associated ingredients and keeping them in close proximity [49–51].

Because biocondensates are phase separated polymer droplets, they display characteristics described by many traditional polymer physics theories. Therefore, we can use polymer physics alongside biology to understand them [52]. These polymer droplets can e viewed as phase separating following Flory-Huggins theory or can be viewed through the lens of polymer crystal nucleation and growth [53–55]. Previous research has also considered percolated networks or gels of multivalent polymers that can form either inside biocondensates or without phase separation and follow Flory-Stockmayer theory [56–58]. Biocondensates also experience microphase separation which is likely to follow concepts similar to order-disorder transitions found in block copolymers or show elements of order based on surface tension as seen in self-assembly and surfactant theory [59–62]. Therefore, we seek to use polymer physics concepts to elucidate how multivalent polymer characteristics alter the formation and dynamics of biocondensates.

Dysregulation of biocondensate formation and droplet solidification is associated with cancer and neurodegenerative diseases such as ALS, Alzheimer's Disease, and dementia [52,63–66]. Therefore, understanding the phase separation and dynamics of multivalent polymers could provide essential clues in treating and preventing these neurodegenerative disorders. Chapter 5 begins by discussing the effects of polymer stiffness on multivalent binding to pathogens before delving into how the stiffness of multivalent polymers can change their phase boundary and the resulting implications for biocondensates. In chapter 6, we further explore how multivalent polymers can regulate the formation and dynamics of biocondensates by modulating a host of design variables including comparing specific and non-specific binding interactions, binding site affinity, valency, and solvent quality.

The goal of this dissertation is to use simulation and theory to understand how design variables affect the binding of linear multivalent polymers. We explore and explain how degree of polymerization, binding site affinity patterns, backbone flexibility, and solvent quality control binding to a single small target, to many targets simultaneously, and the resulting implications for phase separation of polymers. Our coarse-grained simulations bridge length and timescales and can sample large polymer systems that bind proteins at the sub-nanometer lengthscale. This work specifically explains how our simulation results and theory provide insight into biological systems, guide the rational design of pathogen inhibitors, and control biocondensates. However, the general nature of

our simulations and simplified theory presented herein define universal design rules linking structure to function for multivalent polymers. Consequently, the results presented are applicable to many systems of similar size ratios such as polymer-colloid systems and metal-ligand coordination gels [35,67,37]. We hope this work provides helpful and clear rules for the rational design of multivalent polymeric binders.

This Page Intentionally Left Blank

Chapter 2

Methods

This chapter will explain the details of the general simulation methods used throughout this thesis. We also include a description of how the monovalent binding affinity can be estimated using the Langmuir adsorption theory and estimates of the side of a bead. Throughout this work we use a coarse-grain model of our polymers and targets to serve two purposes. First, our goal is not to recreate a specific experimental system, but to provide general structure-function relationships for multivalent polymers. Using a coarse-grain model helps us keep our results general. Second, multivalent biological polymers can be very large (megadaltons in the case of mucins) with binding targets being full folded proteins. Therefore, simulating all-atom or even close to all-atom simulations can only be done on relatively short timescales if at all, especially when we start to consider the case of multiple proteins or polymers interacting together. Typical individual receptor-ligand binding affinities in multivalent interactions are also typically low, and so long timescales are needed to provide sufficient sampling for thermodynamic and kinetic calculations. Therefore, we have chosen to use a typical coarse-grain bead spring model for our polymers to capture relevant time and lengthscales as well as keep our results as general as possible.

We would like to note that, although there are many options for pre-packaged molecular dynamics software, we made the choice to write all of our simulation code ourselves from the ground up. The primary advantage of this method is that we have complete control over and knowledge of every constant, force, protocol, and algorithm in our simulation to a detail that is difficult to achieve in a pre-packaged software. The most important part of a PhD is to build a strong scientific foundation that includes both a depth of knowledge in a particular field and general research skills that it can be applied throughout the rest of our scientific careers. Writing all of my own simulations meant that we had to make a conscious decision about which forces to include, how to include them, and

why. This provided me with a thorough and thoughtful understanding of polymer physics which was essential to my development as a graduate student in polymer science. With this overarching simulation philosophy, we have depicted our resulting general model in Figure 2-1 and detailed all of the forces and our simulation methods in the rest of this chapter. Any methods or constants specific to only a particular chapter will be detailed in a shorter methods section in the requisite chapter.

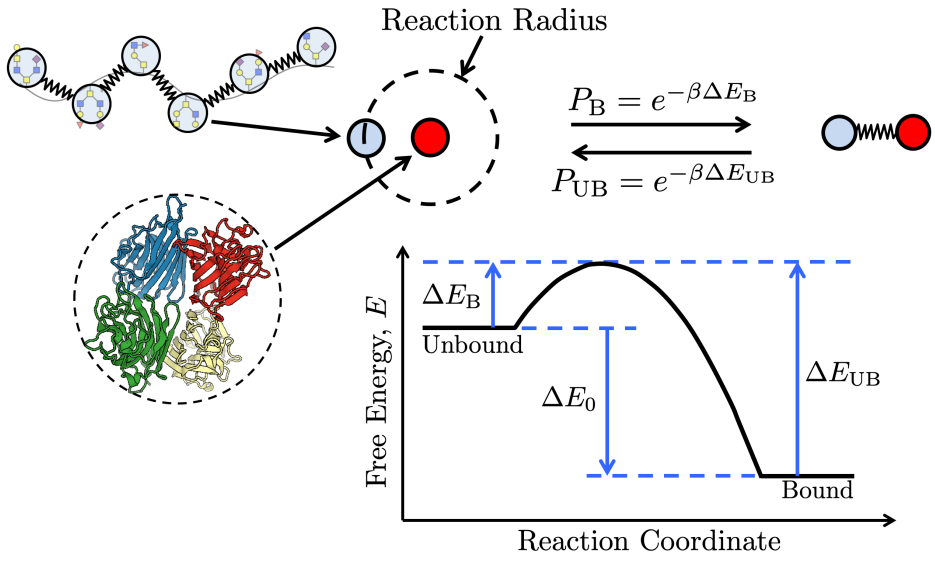


Figure 2-1: Polymers are represented by spherical beads (light blue) connected by gaussian springs. Each polymer bead has a single ligand. Targets can have multiple binding sites and are represented by a single spherical bead (red). Polymer ligands and target binding sites interact when they are within a reaction radius that is dependent on the timestep. Within this reaction radius, they have a probability of binding P_B that depends on the depicted free energy landscape. Once bound, the target and polymer beads are connected by a gaussian spring, and with some probability P_{UB} can return to being unbound and interacting solely through a Lennard-Jones potential. Rendering from the Protein Data Bank [68,69].

2.1 General simulation methods

We model our target as a single bead with M binding sites and n polymers as N_P freely jointed beads connected by harmonic springs as shown in Fig. 2-1. Each polymer bead has a single ligand and its valency is controlled by the polymer degree of polymerization N_P. Each target bead can have one or multiple binding sites. We model the chain using Brownian dynamics where the position of each polymer bead and target is governed by:

$$r(t+\Delta t) = r(t) + (-\frac{\nabla U}{\zeta})\Delta t + R\sqrt{2D\Delta t} \tag{2.1}$$

Where r is the position of the bead at time t, R is a random number drawn from a normal distribution with a mean of 0 and a standard deviation of 1, ζ is the drag coefficient, and $D = k_{\mathrm{B}}T/\zeta$ is the diffusion coefficient. The forces each bead experiences due to interactions with the surrounding polymer or target are captured in ∇U where U is a potential energy that combines contributions from connectivity, excluded volume, and binding. These are added together as $U = U_{\mathrm{sp}}+U_{\mathrm{LJ}}+U_{\mathrm{bind}}$.

The connectivity potential between adjacent polymer beads U_{sp} is modeled as a harmonic spring

$$U_{\mathrm{sp}} = \frac{\kappa}{2}k_{\mathrm{B}}T\sum_{i=1}^{N_{\mathrm{P}}-1}(r_{i+1,i} - 2a)^2 \tag{2.2}$$

where r_{ij} is the distance between polymer beads, $a = 0.5$ is the radius of a simulation bead, and κ was chosen to be $\frac{50}{a^2}$, a value sufficiently large enough to prevent the polymer from stretching apart under normal Brownian forces.

To control excluded volume and non-specific interactions, a Lennard-Jones potential U_{LJ} was applied between bead pairs as

$$U_{\mathrm{LJ}} = \epsilon k_{\mathrm{B}}T\sum_{ij}((\frac{2a}{r_{ij}})^{12} - 2(\frac{2a}{r_{ij}})^6) \tag{2.3}$$

where the value of ϵ can be adjusted to control the solvent quality [70]. Throughout this thesis, the Lennard-Jones coefficient ϵ will be denoted as ϵ_{PP} for polymer-polymer interactions, ϵ_{TT} for target-target interactions, and ϵ_{PT} for polymer-target interactions. We confirmed that modifying ϵ_{PP} for polymer-polymer interactions appropriately adjusted solvent quality by running a single polymer with a degree of polymerization $N_{\mathrm{P}} = 64$ in a box and measuring the average end to end distance at various values of ϵ_{PP}. For a theta solvent, the average end-to-end distance should be $N_{\mathrm{P}}^{1/2} = 8$, which occurs at $\epsilon_{\mathrm{PP}} = \frac{5}{12}$ (Fig. 2-2). This matches well with previous findings that $\epsilon_{\mathrm{PP}} = 0.41$ corresponds to polymer dynamics in theta solvent [70]. We also found that $\epsilon_{\mathrm{PP}} = \frac{1}{12}$ corresponds to end-to-end distances of approximately $N_{\mathrm{P}}^{3/5} = 12.1$, which is consistent with a polymer in good solvent. Therefore, throughout the following chapters, we used $\epsilon_{\mathrm{PP}} = \frac{5}{12}$ to mimic polymer configurations in a theta solvent and $\epsilon_{\mathrm{PP}} = \frac{1}{12}$ to mimic a good solvent. Any other solvent qualities will be specified explicitly in their corresponding section.

Our last type of interaction in all chapters is a reactive lock and key bond, which represents our

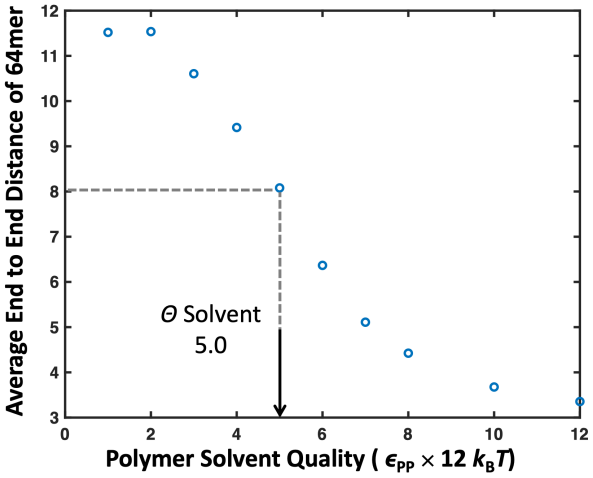


Figure 2-2: Simulated end-to-end distance of a polymer with degree of polymerization $N_{\mathrm{P}} = 64$ for various solvent qualities. The polymer has a size of $2a\sqrt{N_{\mathrm{P}}} = 8$, consistent with theta solvent conditions when $\epsilon_{\mathrm{PP}} = \frac{5}{12}$.

specific, valence-limited binding interaction. To simulate this reactive binding, harmonic springs were turned on and off between the polymer beads and the targets to dynamically represent bonded and unbonded states. This was implemented using the prefactor $\Omega(i,j)$ multiplied by a harmonic potential as follows:

$$U_{\mathrm{bind}} = \frac{\kappa}{2} k_{\mathrm{B}} T \sum_{i=1}^{M} \sum_{j=1}^{nN_{\mathrm{P}}} \Omega(i,j)(r_{ij} - 2a)^2 \tag{2.4}$$

$\Omega(i,j) = 1$ when the ith binding site on the target is bound to the jth bead of the inhibitor, and $\Omega(i,j) = 0$ when the target binding site or polymer bead is unbound. To control the probability of binding and unbinding, we use a piecewise function based on the energy barriers for the binding reaction from C. Sing and A. Alexander-Katz [71].

$$\Omega(i,j,t) = \begin{cases} \begin{cases} 1 & \Xi < e^{-\Delta E_{\mathrm{B}}} \\ 0 & \Xi > e^{-\Delta E_{\mathrm{B}}} \end{cases} & \text{if } \Omega(i,j,t-\tau_0) = 0 \cap r_{ij} < r_{\mathrm{rxn}} \\ \begin{cases} 0 & \Xi < e^{-\Delta E_{\mathrm{UB}}} \\ 1 & \Xi > e^{-\Delta E_{\mathrm{UB}}} \end{cases} & \text{if } \Omega(i,j,t-\tau_0) = 1 \end{cases} \tag{2.5}$$

Here, Ξ is a random number between 0 and 1, ΔE_{B} is the energy barrier to bind normalized by $k_{\mathrm{B}}T$, and ΔE_{UB} is the energy barrier to unbind normalized by $k_{\mathrm{B}}T$ as shown in Figure 1. Without

loss of generality, these energies are considered to be always positive, and the kinetics of binding are held constant by keeping ΔE_{B} at $\frac{1}{2}k_{\mathrm{B}}T$ so that binding is an accessibly frequent event. Increasing or decreasing the energy barrier will respectively slow or accelerate the kinetics of binding and unbinding equally, but not change the system's thermodynamics. The thermodynamic drive of binding is controlled by varying $\Delta E_0 = \Delta E_{\mathrm{B}} - \Delta E_{\mathrm{UB}}$. Binding becomes more favorable as ΔE_0 is made more and more negative. This method is based directly on Sing *et al.* as well as others [71–75]. Researchers studying vitrimers have extended this approach to include the additional effect of bond exchange [76–78], but in the case of ligand-receptor interactions in proteins, such additional possibilities do not apply. This is because the protein is much larger than the size of the binding site, which makes the binding very local and size exclusion prevents the swapping of bonds. Binding reactions are evaluated every time interval $\tau_0 = 100\Delta t$, where Δt is the length of one timestep and t is the current time. The reaction radius $r_{\mathrm{rxn}} = 1.1$ is the distance apart two beads would be if their surface was touching plus 0.1. Choosing $0.1 < (6D\tau_0)^{1/2}$ allows time for a target that unbinds to diffuse out of the polymer radius of influence in τ_0 and makes binding events independent [71]. We have applied the constraint that at any time, a polymer bead can only bind to one target binding site $\sum_j \Omega(i,j,t) \leq 1$, and a target site can only be bound to one polymer bead $\sum_i \Omega(i,j,t) \leq 1$. Competing reactions are sampled randomly. Note that we do not include the effect of forces in the breaking of the bonds, this is due to the fact that for forces on the order of $k_{\mathrm{B}}T/a$, this effect is negligible if the characteristic bond length is less than 1 nm. For reference, a discussion of the subject is given in [79].

The potentials are applied over the timestep $\Delta t = \frac{6\pi\eta a^3}{k_{\mathrm{B}}T}\Delta\tilde{t}$ where $\frac{6\pi\eta a^3}{k_{\mathrm{B}}T}$ is the characteristic monomer diffusion time or the time that it takes a bead to diffuse its radius a and the dimensionless timestep is $\Delta\tilde{t} = 10^{-4}$. These equations can all be made dimensionless by scaling energies by thermal energy $k_{\mathrm{B}}T$, lengths by bead radius a, and times by the characteristic diffusion time $\frac{6\pi\eta a^3}{k_{\mathrm{B}}T}$.

The simulation code is included in appendix B.

2.2 Monovalent binding affinity

To confirm that we used biologically relevant binding affinities for individual binding site interactions, we placed increasing concentrations of monomeric polymer beads ($N_{\mathrm{P}} = 1$) in a box with a single monovalent target. At each concentration of inhibitor beads, we measured the fraction of time the target spent bound plotted in Fig. 2-3. To estimate the dissociation constant K_{D} of binding, we

fit our simulation data to the Langmuir adsorption curve $\phi = \frac{[P]}{[P]+K_D}$ where ϕ is the fraction of time the target spends bound, $[P]$ is the concentration of polymer (in this case monomer) beads, and K_D is the dissociation constant of the binding reaction. We can then converted our unitless K_D to a concentration by assuming a target bead diameter to be 5 nm which is the approximate diameter of a peanut agglutinin lectin[80]. This method resulted in K_Ds in the mM to μM range for ΔE_0s between $-2k_BT$ and $-6k_BT$. This monovalent binding affinity is well within the weakly binding mM to μM range typical of lectins and sugars binding as well as the mM to μM affinity range of some monovalent protein-protein and RNA-protein binding found in biocondensates [81–85].

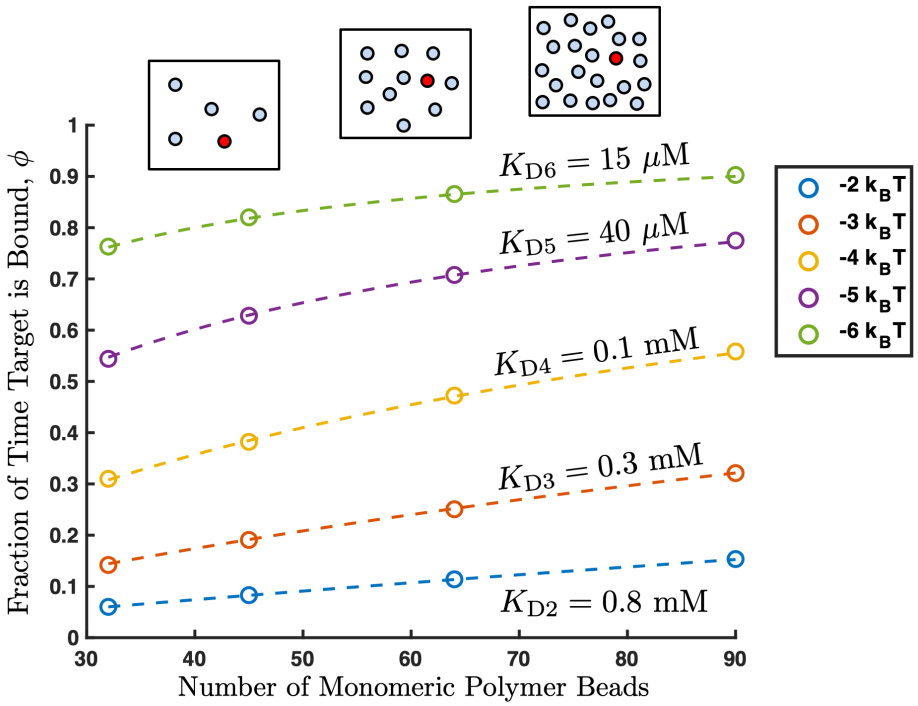


Figure 2-3: A plot of the fraction of time bound for different monomeric polymer bead concentrations interacting with a monovalent target. By fitting to the Langmuir adsorption curve $[P]/([P]+K_D)$ (dashed lines), we can find the K_D of monovalent binding corresponding to the values of ΔE_0 noted in the plot legend. Simulation results (circles) and Langmuir adsorption equation fits (dashed lines) are shown for $\Delta E_0 = -2$, -3, -4, -5, and $-6k_BT$. When results are rescaled to Molar using an estimated target diameter of 5 nm, the K_D for these binding energies is in the mM to μM range typical of monovalent glycan – lectin and some protein - nucleic acid binding interactions [81–85]. Fit values of K_D are shown above their corresponding lines.

Chapter 3

Polymer Length Dependence of Multivalent Binding Avidity

Abstract

Multivalent binding interactions are commonly found throughout biology to enhance weak monovalent binding such as between glycoligands and protein receptors. Designing multivalent polymers to bind to viruses and toxic proteins is a promising avenue for inhibiting their attachment and subsequent infection of cells. Several studies have focused on oligomeric multivalent inhibitors and how changing parameters such as ligand shape and size, and linker length and flexibility affect binding. However, experimental studies of how larger structural parameters of multivalent polymers such as degree of polymerization affect binding avidity to targets have mixed results with some finding an improvement with longer polymers and some finding no effect. Here, we use Brownian dynamics simulations to provide a theoretical understanding of how degree of polymerization affects the binding avidity of multivalent polymers. We show that longer polymers increase binding avidity to multivalent targets, but reach a limit in binding avidity at high degrees of polymerization. We also show that when interacting with multiple targets simultaneously, longer polymers are able to use inter-target interactions to promote clustering and improve binding efficiency. We expect our results to narrow the design space for optimizing the structure and effectiveness of multivalent inhibitors, as well as be useful to understand biological design strategies for multivalent binding. The work presented in this chapter is primarily sourced from Zumbro *et al.*, *Biophys. J.* **115** (2019) 892-902 [86].

Statement of Significance

Multivalent polymers show promise as inhibitors of toxins and microbial infection. Experimental studies have demonstrated that only certain traits of such polymers are useful for binding proteins. Here we demonstrate that the length of the polymer is an important parameter to consider as well as the weak protein-protein interactions of the toxins themselves. Our results provide a guide for the design of a new generation of polymeric binders that will have enhanced avidity.

3.1 Introduction

Biology uses multivalent interactions for a variety of reasons including enhancing weak monovalent interactions, creating conformal interfaces such as those between cells or those inducing endocytosis, or increasing specificity and affinity of binding using a limited number of receptor and ligand types [1]. Multivalent binding occurs when multiple ligands on one species bind to multiple receptors on another species simultaneously. While each individual binding site–ligand interaction might have weak binding affinity, when multiple sites bind simultaneously, they can produce a much stronger binding avidity than the sum of the corresponding monovalent interactions [1]. Here we use the term “avidity" as the overall binding affinity of multivalent interactions and “affinity" to refer to the binding affinity of single binding site interaction [87].

Because multivalent binding can be used to enhance low-affinity binding interactions such as those commonly found between glycoligands and sugar-binding proteins called lectins, designing synthetic multivalent polymers that target specific lectins is of great interest [88]. Binding strongly to lectins is a promising avenue for treating common diseases from diarrhea and colitis to influenza by inhibiting protein targets such as AB5 toxins including Shiga or Cholera toxin [89,27,90,26,91,92] or the hemaglutinin receptor on the influenza virus [21,20].

To narrow the design space for multivalent inhibitors, several theoretical and experimental models have looked at how spacing of binding sites and flexibilities of linkers affect binding avidity [93,94,32,18,95]. Previous studies have explored optimizing parameters on the size scale of individual binding sites. For example, Liese et al. modeled how changing the linker length and flexibility between two ligands changes their binding avidity while Papp et al. explored matching the size between ligands exactly to the target [33,20]. In contrast, the field has had relatively few studies on how large polyvalent materials and design decisions at size scales much larger than individual binding

sites control polyvalent interactions. Some experimental studies have attempted to understand the effect of degree of polymerization on a polymer's multivalent enhancement. Several of these studies found that longer polymers are more effective binders for influenza virus [17,21,96], but other groups have found that there is a limit to this binding enhancement from increased length when interacting with proteins[31,97]. A general theoretical understanding of how polymer length contributes to multivalent binding has yet to be developed.

In this article, we show that while polymers with higher degrees of polymerization bind more tightly to multivalent targets, the enhancement in binding energy tapers off with polymer length. We find that the entropic penalty from forming long loops is a likely explanation for this effect. We also demonstrate that favorable interactions between targets, such as hydrophobic attraction between proteins, can enhance the binding of the inhibiting polymer to the target, but only for higher degrees of polymerization. We use a coarse grain Brownian dynamics simulation to establish rules for how the degree of polymerization can influence the strength of multivalent binding interactions between a polymer and a globular target such as a lectin.

3.2 Methods

This chapter follows all of the simulation methods detailed in Chapter 2 and depicted in Figure 2-1.

Across the simulations we used a Lennard-Jones potential of $\epsilon_{\mathrm{PP}} = \frac{5}{12}$ to mimic polymer configurations in a theta solvent and $\epsilon_{\mathrm{PP}} = \frac{1}{12}$ to mimic a good solvent as shown in Figure 2-2. We chose to run both theta solvent and good solvent because these bound the solvent conditions for soluble polymers, putting a limit on any characteristics that depend on solvent quality. For reference, the excluded volume parameters for each of our simulation scenarios are listed in Table 3.1.

Table 3.1: ϵ values for Polymer-Polymer (PP), Polymer-Target (PT), and Target-Target (TT) bead Lennard-Jones Interactions

Case #	ϵ_{PP}	ϵ_{PT}	ϵ_{TT}	Case Description
1	5/12	1/12	N/A	Theta Solvent, Dilute Targets
2	1/12	1/12	N/A	Good Solvent, Dilute Targets
3	5/12	1/12	1/12	Theta Solvent, Many Low Attraction Targets
4	5/12	1/12	18/12	Theta Solvent, Many High Attraction Targets

3.3 Results and Discussion

To better inform the design of multivalent polymeric binders, we seek to determine how degree of polymerization changes the inhibitor's binding avidity. We examine two cases. In the first case, we observe how the binding avidity between a single monovalent or multivalent target and a polymer changes with the degree of polymerization of the polymer. In the second case, we probe how this result changes when the polymer is in the presence of multiple targets which have some favorable inter-target interactions.

3.3.1 Biologically relevant binding affinities

To establish a baseline binding affinity and ensure we are at biologically relevant binding affinities for individual binding sites, we first characterize the monovalent binding interactions of monovalent targets and monovalent free inhibitor beads. We place a single monovalent target in a box with a constant concentration of free inhibitor beads and measure the fraction of time the target is bound and unbound from an inhibitor bead. We run these simulations for 10^8 timesteps and run either 50 or 100 simulations in parallel to ensure that we capture sufficiently long timescales that are much longer than the typical time for a single binding and unbinding to have accurate averaging. The resulting fraction of time bound plotted for different inhibitor concentrations and fitted with a Langmuir adsorption curve is shown in Figure 2-3. Throughout this chapter, we use binding site affinities of $\Delta E_0 = -4k_{\mathrm{B}}T$. By assuming a target diameter of 5 nm, we can show this corresponds to a K_{D} on the order of 1×10^{-4} M. This binding affinity is similar to the monovalent binding interactions between lectins and their corresponding sugars, which typically have a K_{D} in the mM to μM range. [98,16].

3.3.2 Effect of length on binding avidity to an individual toxin

Using $\Delta E_0 = -4k_{\mathrm{B}}T$ as an experimentally relevant individual binding site affinity, we then explored how the degree of polymerization of our inhibiting polymer changed its binding avidity to a monovalent or multivalent target. We compared binding and unbinding kinetics of a single target with one or two binding sites to a constant concentration of inhibitor beads with varying connectivity. This scenario is depicted in Fig. 3-1, where a single target is placed with 64 inhibitor beads with increasing degrees of connectivity, such as 64 free inhibitor beads, 32 dimers, or a single 64mer.

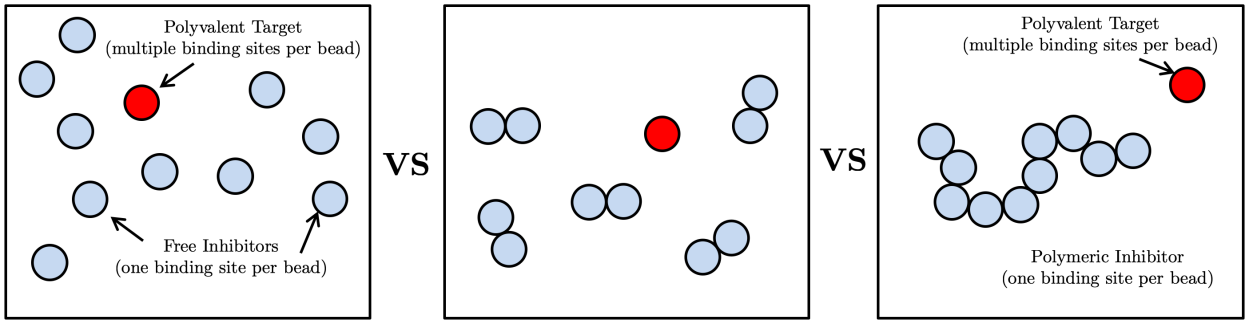


Figure 3-1: Schematic of the scenarios tested when comparing binding avidity's dependence on degree of polymerization. The volume and number of inhibitor ligands are held constant to maintain a constant concentration of ligands at 64 ligands per box. The connectivity of the inhibitor ligands was varied from monomers to 64mers in multiples of two so that degrees of polymerization, 1, 2, 4, 8, 16, 32, and 64 were all investigated. This ensured that all polymers in each simulation were monodisperse.

To compare the binding avidity of the polymeric inhibitors to the target, we counted the time that a target stayed bound to an inhibitor bead, where a target was considered bound whenever at least one of the target's binding sites was occupied. Unsurprisingly, the average time interval spent bound for monovalent targets does not depend on length (Fig. 3-2). The target's single binding site can only interact with one inhibitor ligand at a time, so interactions with neighboring ligands have no impact on the duration the target spends bound. Therefore, the polymeric structure and valency of the inhibitor do not affect the average time bound for a single or dilute monovalent target.

In contrast, for the divalent targets, switching from monomeric inhibitors to polymeric inhibitors shows a significant increase in time the target spends bound (Fig. 3-2), demonstrating the enhancement in binding avidity created by multivalent binding. This follows the multivalent binding theories of increased local concentration [5] and decreased loss of entropy over free ligands [99,1,87]. Both the constant duration of time spent bound for our monovalent target and the increase in duration of time spent bound with the lengthening of our inhibiting polymer is consistent with these previous theories.

However, these theories have not previously captured how degrees of polymerization much larger than the size scale of the target can change binding avidity. Revisiting Fig. 3-2, we can see that the time spent bound approaches a limit at high degrees of polymerization for both theta solvent (Case 1 in Table 3.1) and good solvent qualities (Case 2 in Table 3.1). To explain this phenomenon, we considered the proportion of time a target is bound in a system with a given degree of polymerization. This proportion can be transformed into a free energy of binding, which we term ΔG_{B}:

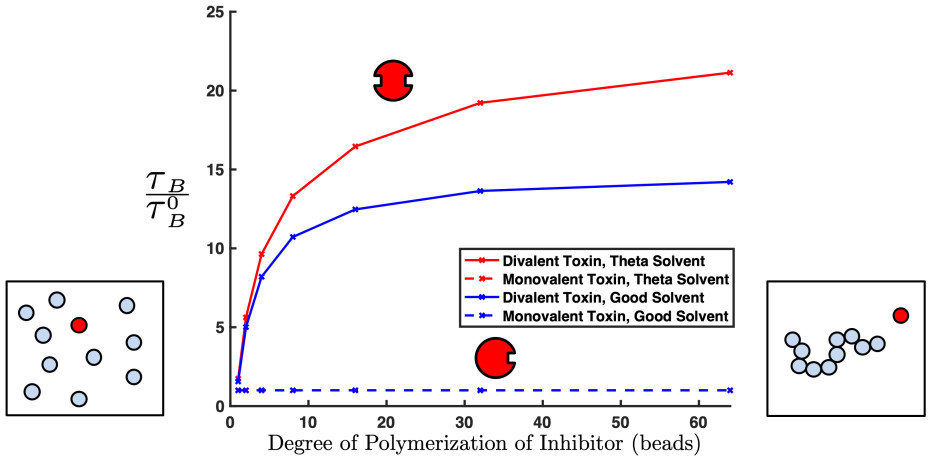


Figure 3-2: Average time interval a monovalent or divalent target are bound to polymeric inhibitors of various lengths, normalized by the average time bound for a monovalent target, τ_{B}^0. The time interval bound for a monovalent toxin does not depend on the inhibitor length (dashed blue line and dashed red line overlap). For the divalent target (solid), oligomeric inhibitors spend significantly more time bound than monomeric inhibitors, exhibiting the enhancement of multivalent binding avidities over monovalent binding. At high degrees of polymerization of the inhibiting polymer, as the length of the inhibitors is increased further, there is only a small gain in the average time interval the target is bound. The time spent bound approaches some maximum value with increasing inhibitor length. Inhibitors in theta solvent (red, Case 1 in Table 3.1) spend more time bound than good solvent (blue, Case 2 in Table 3.1). Error bars are smaller than symbol size.

$$\Delta G_{\mathrm{B}} \approx -k_{\mathrm{B}}T \ln\left(\frac{\tau_{\mathrm{B}}}{\tau_{\mathrm{UB}}}\right) \tag{3.1}$$

where τ_{B} and τ_{UB} are the average time spent bound and unbound respectively. Relative to the K_{D} of monovalent binding, we can find the dissociation constant of multivalent binding by using the difference in ΔGs. For example, assuming a bead diameter of 5 nm, we can estimate the 64mer-divalent target dissociation constant in theta solvent to be approximately 6×10^{-6} M. While τ_{B} as examined in Fig. 3-2 is difficult to treat theoretically, we found this ΔG_{B} more theoretically tractable. Note that in Eq. 3.1 we have not included second order corrections for finite size effects which will reduce the binding affinity measured in small simulations [100], but we expect this will not change qualitative results. We developed a model predicting ΔG_{B} as a function of the degree of polymerization and valency of the target as well as other factors, described in detail in Appendix A. Briefly, the model is loosely inspired by the Poland-Scheraga model of DNA denaturation, in that a polymer bound multivalently to a target can be represented as a sequence of loops alternating with sites bound to the target [101]. In the general case, the partition function of this model can only be evaluated numerically, but in the limit of high N_{P}, where N_{P} is the degree of polymerization,

there is an analytical result for ΔG_{B}. The full function, given in Eq. A.22, is complex, but the dependence on N_{P}, the number of polymers n, and volume V_{box}, is simple:

$$\Delta G_{\mathrm{B}} = C - k_{\mathrm{B}}T \ln(\frac{nN_{\mathrm{P}}}{V_{\mathrm{box}}}) \tag{3.2}$$

where C is a value not dependent on N_{P} related to the persistence length, solvent quality, ligand density, and valency of the target. Note that in our simulations, both nN_{P} and V_{box} are held constant; specifically, $nN_{\mathrm{P}} = 64$. Thus, Eq. 3.2 predicts that at high N_{P}, polyvalency no longer increases avidity. So, for example, if $N_{\mathrm{P}} = 32$ is high enough to approach the limit (a question we address shortly), ΔG_{B} should be the same for two 32-mers and one 64-mer. Our theoretical treatment successfully reproduces the qualitative behavior of ΔG_{B}: as predicted, we observe that ΔG_{B} initially decreases sharply, representing the benefits of polyvalency, before reaching a limit at higher degrees of polymerization, shown in Fig. 3-3.

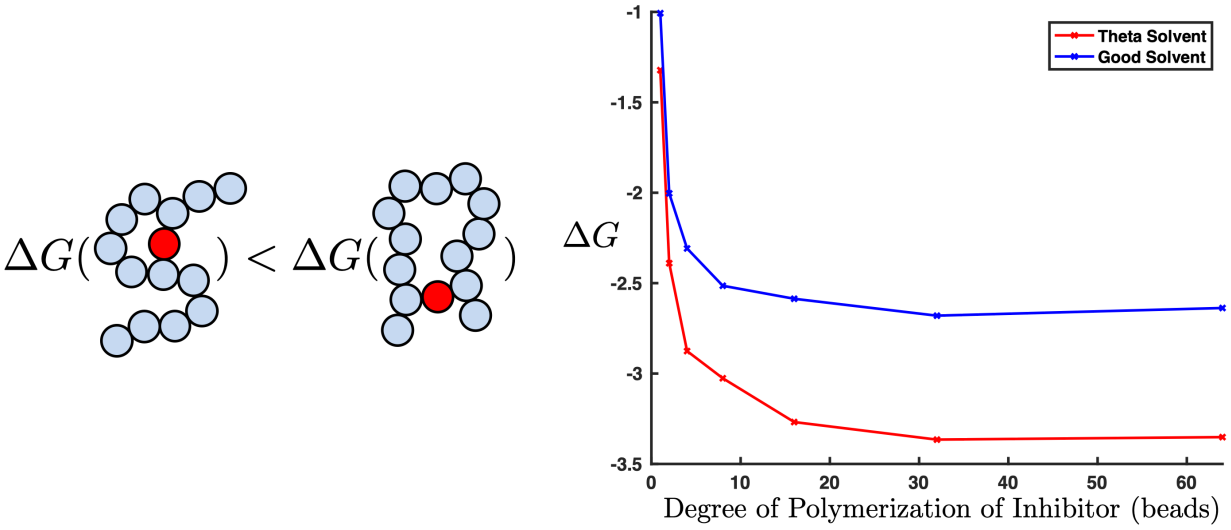


Figure 3-3: A plot of the free energy of binding for the degree of polymerization of the inhibitor. The free energy of binding is calculated using the average time interval the target spends bound to the polymer τ_{B} (meaning one or more binding sites is bound) divided by the average time interval the target spends completely unbound τ_{UB}. Longer loops are entropically unfavorable, so while they are possible in longer polymers, they are unlikely to form. We can see that this leads to a limiting minimum binding energy as degree of polymerization of our inhibitor increases. This is true in both good (blue) and theta (red) solvents.

The leveling off of ΔG_{B} in our theoretical model is due to loop entropy: when two faraway monomers each bind the target, the polymer is forced into a large loop, which restricts the conformational degrees of freedom of the polymer chain. This free energy penalty increases with the

size of the loop, and for large enough loops, the free energy penalty becomes larger than the free energy of binding. Beyond the length where binding loops are no longer thermodynamically favored (which corresponds to where the high-N_P limit begins to be reached), increasing the degree of polymerization will provide no benefit. Thus, we predicted that the flattening of the ΔG_B curve should coincide with the length at which loops stop forming. Indeed, the frequency of loops drops precipitously with loop lengths, and loops larger than a length of 9 for good solvent and 13 for theta solvent are vanishingly rare as shown in Fig. 3-4. This is in agreement with the fact that ΔG_B flattens beyond $N_P = 8$ for good solvent and $N_P = 16$ for theta solvent (Fig. 3-3). Note that if entropic costs were turned off, multivalency would continue to yield increases in avidity for longer polymers. As Kitov and Bundle show, when all possible binding sites on a multivalent ligand can bind equally to the receptor, ΔG will not plateau [102]. Our theory describes the plateau in binding behavior well for an ideal chain and the good solvent scenario also seems to follow. Thus, loop entropy is the likely culprit for the diminishing returns of increasing N_P.

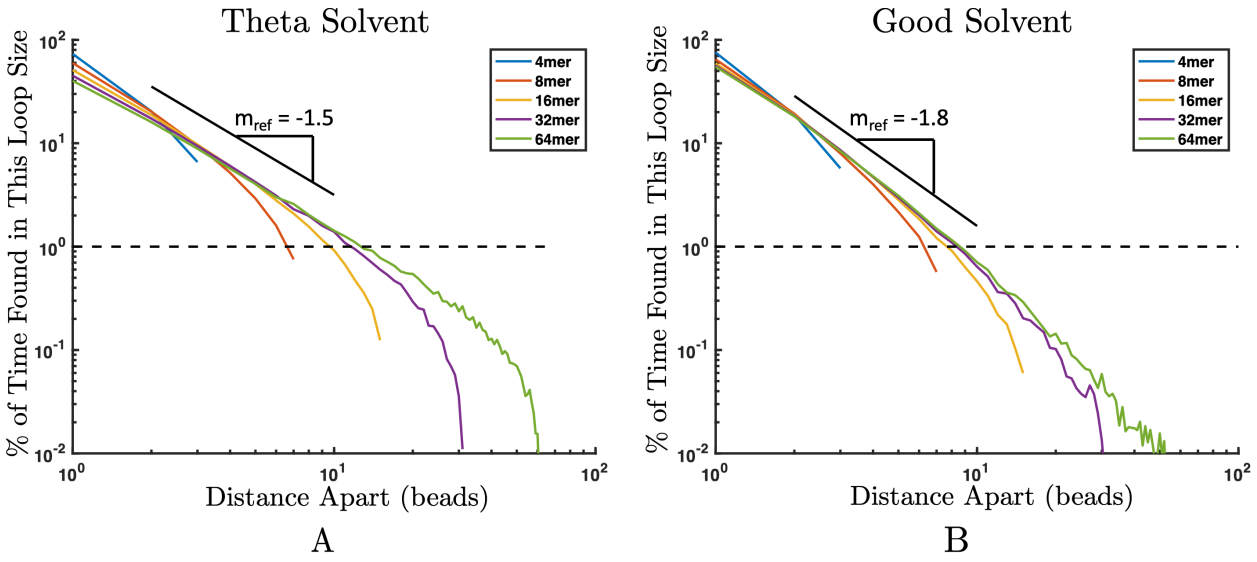


Figure 3-4: Log-log plot of the percent of time a divalent target forms various loop sizes for different length polymers. For reference, 1% frequency is shown with dashed black line. In theta solvent (A) and good solvent (B), loops larger than 13 and 9 beads, respectively, are formed less than 1% of the time. $m_{ref} = -3\nu$ is a reference slope where ν is the Flory exponent.

Ultimately, our simulation and model results match excitingly well with the experimental results that increasing polymer length leads to only a limited increase in polymer avidity to lectins [31,97].

3.3.3 Effect of length on binding avidity in the presence of multiple toxins

In vivo, environments can be crowded and multiple targets might interact with a single inhibiting polymer. If the target is a protein, hydrophobicity and charge can create target-target interactions leading to a wide range of solubility maximums from 1 mg/ml for wheat germ agglutinin to more than 50 mg/ml for serum albumin [103,104]. In this section, we examine binding between multiple targets and the inhibiting polymer and consider how target-target interactions influence binding avidity. To investigate the affect that target-target interactions have on the binding avidity of the inhibitors, we added a Lennard-Jones potential between targets and explored how changing the attraction between the targets modified their binding with the inhibiting polymer.

To examine the effect of multiple targets interacting with the inhibitor simultaneously, we placed 64 divalent targets in a box with inhibiting polymers. To compare the effect of polymer length, we again varied the connectivity of the inhibiting beads while maintaining the same total concentration of polymer binding sites, as depicted in Fig. 3-5. We modified our target-target attraction by changing ϵ in Eq. 2.3, and compared two target-target attraction scenarios: a relatively neutral condition where $\epsilon_{\mathrm{TT}} = \frac{1}{12}$ (Case 3 in Table 3.1)and a weakly attractive condition where $\epsilon_{\mathrm{TT}} = \frac{18}{12}$ (Case 4 in Table 3.1). These ϵ_{TT} values correspond to scenarios where the target-target interaction has a positive and negative second virial coefficient, respectively. To ensure we were at biologically relevant target-target interactions, we calculated the concentration of our targets by making the following assumptions. Assuming a target diameter of 5 nm and molecular weight of 70 kDa, 64 targets corresponds to a concentration of 7 mg/ml. By running 64 targets in a box without an inhibiting polymer present, we confirmed that at both $\epsilon_{\mathrm{TT}} = \frac{1}{12}$ and $\epsilon_{\mathrm{TT}} = \frac{18}{12}$ the targets do not aggregate and phase separate. This shows that both levels of target-target Lennard-Jones interactions are within the range of relevant protein solubilities.

Increased competition

Normally one does not have isolated targets, but a finite concentration of them. Thus, it is interesting to ask the following question: if one had multiple targets with a given degree of solubility binding to the same inhibitor, would that have a marked effect on the kinetics? To answer this, we examined the binding kinetics of 64 targets to our inhibiting polymers to compare to our single or dilute target case. Similarly to when interacting with single targets, the binding avidity of polymers

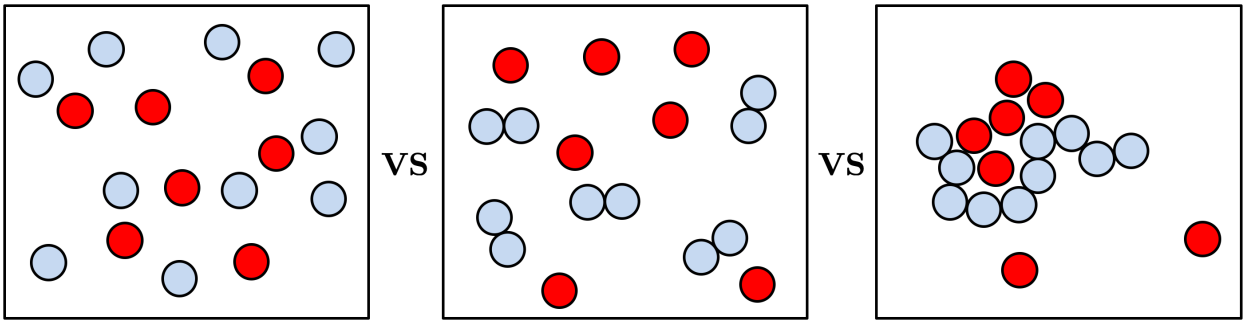


Figure 3-5: Schematic of the scenarios tested when comparing binding avidity's dependence on degree of polymerization with multiple targets present. The volume and number of inhibitor ligands are held constant to maintain a constant concentration of ligands at 64 ligands per box. The connectivity of the inhibitor ligands was varied from monomers to 64mers in multiples of two so that degrees of polymerization, 1, 2, 4, 8, 16, 32, and 64 were all investigated. This ensured that all polymers in each simulation were monodisperse. The concentration of targets was held constant in all simulations.

initially increases with increasing degree of polymerization before tapering off at high polymerization as shown in Fig. 3-6. More interestingly, in the presence of multiple targets, increased attraction between targets decreases the maximum τ_B. To investigate this phenomenon, we compared the rate of unbinding in Fig. 3-7. Here, we see two timescales at which targets unbind, a fast and a slow timescale. The fast timescale represents targets that only become singly bound before unbinding, whereas the slow timescale represents targets that transition from being doubly bound to unbound. By comparing the slope of the linear best fit line in both regions, we find that the rate of unbinding for single bonds is unchanged when there is inter-target attraction, but the rate of unbinding for doubly bound targets increases with inter-target attraction. The increased rate of unbinding for doubly bound targets leads to the decrease in average τ_B seen in Fig. 3-6.

The higher probability that doubly bound targets unbind can be explained by increased competition. If a lone or very dilute target becomes doubly bound and then unbinds with one binding site, this unbound site could easily rebind. In contrast, in a crowded environment with many targets, a site that unbinds has to compete with neighboring targets to rebind. This increase in competition comes from both neighboring bound targets (Fig. 3-8A) as well as unbound targets that are aggregated by a high density of bound targets (Fig. 3-8B). We will show that the increase in inter-target attraction from $\epsilon_{TT} = \frac{1}{12}$ to $\epsilon_{TT} = \frac{18}{12}$ leads to a drastically higher local concentration of targets in the polymer's radius of influence, exacerbating this competition and shortening the maximum τ_B.

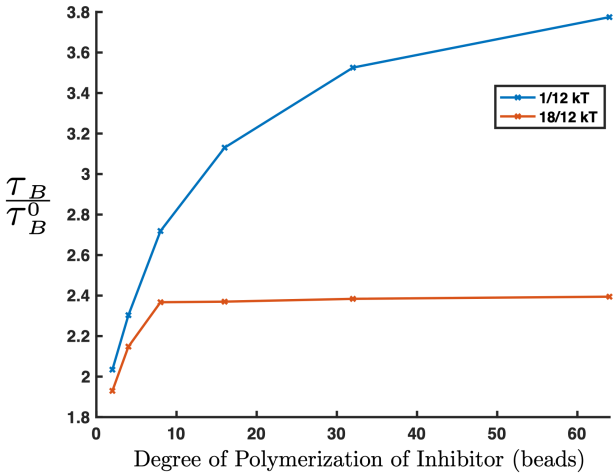


Figure 3-6: Plot of average time bound (τ_{B}) for targets when multiple targets are present. Y-axis is normalized by the average time bound for monomeric inhibitors, τ_{B}^0. Data presented is for polymer-target binding energy of $\Delta E_0 = -4\ k_{\mathrm{B}}T$ in theta solvent. Similar to with a single target, τ_{B} has a limited increase with degree of polymerization of the inhibiting polymer. More attractive inter-target potentials (orange, $\epsilon_{\mathrm{TT}} = \frac{18}{12}$) decrease the maximum τ_{B}. Error bars are smaller than symbol size.

Polymer induced phase separation

Because the kinetic changes were correlated to changes in local concentration of the target, we next considered the thermodynamics of the system, where we found an increased concentration of targets bound to the inhibitor. In Fig. 3-9A, we show that in theta solvent for ~ 0.1 mM binding affinity ($\Delta E_0 = -4k_{\mathrm{B}}T$), the average number of targets bound to the polymer increased for higher target-target attraction, for both mono and divalent targets. Therefore, although individual targets unbind more quickly, inter-toxin attraction leads to higher inhibiting polymer avidity overall. Attraction between targets causes a significant increase in the number of targets bound because it induces a collapse transition where bound targets collapse the polymer and themselves into a globule or liquid phase. When the polymer/bound target system collapses to form a globule, the enthalpic benefit of an additional target joining the globule becomes greater than the loss of entropy of binding, leading to a significant increase in the number of targets bound to the target. This leads to a target rich liquid-like phase attached to the polymer and a low concentration gas-like target phase in the supernatant. Similar data for good solvent can be seen in Supplemental Figure 3-13.

We confirmed that the marked increase in avidity was caused by a polymer collapse transition by examining the end to end distance. Fig. 3-10 shows the decrease in the average end to end distance for a 64mer polymer in theta solvent interacting with divalent (Fig. 3-10A) and monovalent (Fig.

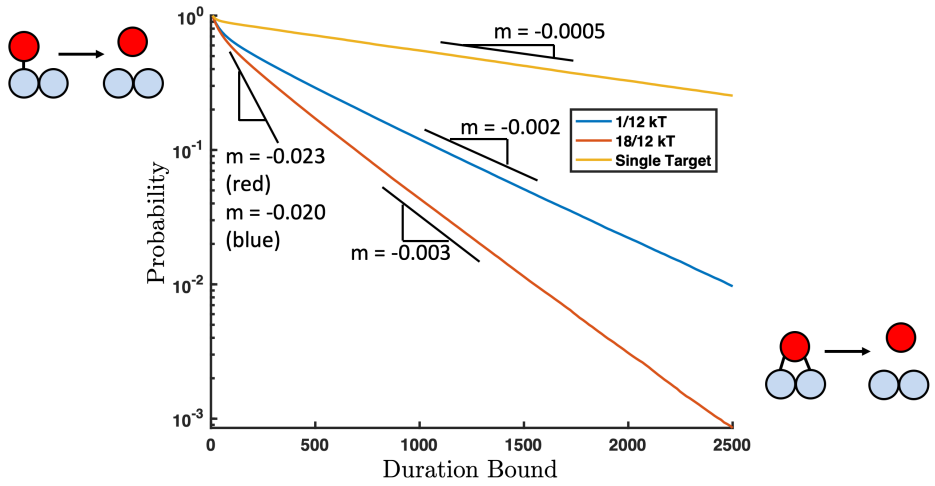


Figure 3-7: Distribution of times spent bound for single target scenarios (yellow) and scenarios with multiple targets (orange and blue). The rate of unbinding corresponds to the slope of the line in these two regions, shown in black. There is a fast and a slow timescale on which targets unbind. The former corresponds to singly bound targets unbinding and does not change with inter-target potentials. The second, longer timescale corresponds to doubly bound targets that unbind. When there are favorable target-target interactions (orange), the decay in doubly bound times is faster than if there is not an attraction between targets (blue), or if there is no competition from other targets (yellow).

3-10B) targets with $\epsilon_{TT} = \frac{1}{12}$ and $\epsilon_{TT} = \frac{18}{12}$ attractions between targets. End to end distances for polymers in good solvent interacting with multiple targets can be seen in Supplemental Figure 3-14. As expected, the theta polymer is at its normal random walk size of 8 with no targets present, but when divalent targets are added, the polymer collapses to a globule for both levels of inter-target attraction. For 64mers interacting with monovalent targets, we only see a collapse in the end to end distance when the inter-target attraction is $\epsilon_{TT} = \frac{18}{12}$, meaning that the collapse transition does not occur when $\epsilon_{TT} = \frac{1}{12}$. Instead, for monovalent targets with positive second virial coefficient, bound targets have high enough excluded volume to extend the polymer chain, causing the swelling seen in Fig. 3-10B.

This collapse transition that leads to globular polymers and higher target binding is caused by a competition between entropy and enthalpy and can be induced by increasing the polymer length. This can be seen in the large jump in targets bound for monovalent targets with $\epsilon_{TT} = \frac{18}{12}$ in Fig. 3-9A as the degree of polymerization is increased from 8 to 32 beads. Examining this case more closely, we see that as degree of polymerization is increased, the percent of inhibitor beads bound initially drops, before a sudden increase in binding after a polymerization of approximately 10 beads.

When the targets have a positive second virial coefficient, a bound target reduces the volume

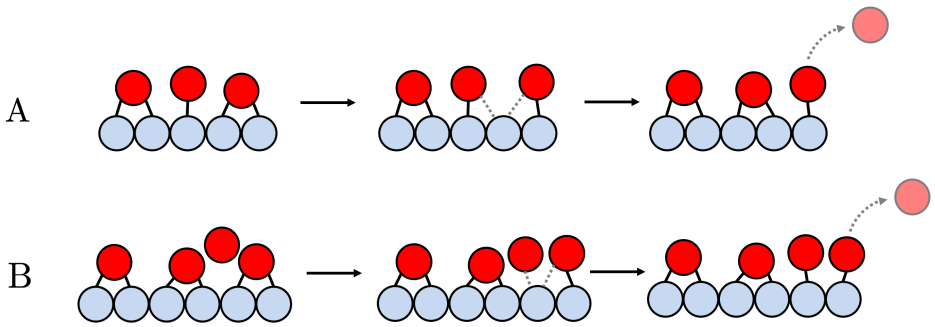


Figure 3-8: There are two types of competition bound targets experience that lead to shorter times bound for divalently bound targets. (A) shows competition from neighboring bound targets and (B) shows competition from nearby unbound targets which is increased for more favorable inter-target potentials.

that a target on a neighboring bead has. Because of this, targets prefer to bind to monomeric inhibitors or polymer ends that have more available volume around them, or targets prefer to be unbound. When the targets have high attraction, a neighboring target creates favorable interactions and smaller excluded volume so targets prefer to bind places that have more neighbors such as the center of the polymer. This can be seen in Fig. 3-11, where low inter-attraction monovalent targets prefer to bind polymer ends and monomeric inhibitors and high inter-attraction targets prefer to bind the center of the polymeric inhibitors.

Targets can also overcome the unfavorable excluded volume created by their neighbors when the polymer-target binding affinity is high enough, such as in the case where the target is divalent and $\Delta E_0 = -4k_\mathrm{B}T$. Divalent targets benefit from two factors, they get the energy benefit of binding twice to the polymer and the benefit of monopolizing two polymer beads worth of space, reducing interactions with neighboring bound targets.

In addition to increased binding of targets, the polymeric inhibitor promotes aggregation and increased local concentration of unbound targets. By measuring the minimum distance between all unbound targets and the polymer and normalizing by the volume of the shell, we compared the concentration of targets at each distance R away from the polymer as shown in Fig. 3-12 for theta solvent and Supplemental Figure 3-15 for good solvent. From these plots, it is clear that at small inter-target potentials such as $\epsilon_{\mathrm{TT}} = \frac{1}{12}$ there is a negligible increase in the concentration of unbound targets near the polymer. In contrast, with an attractive inter-target potential of $\epsilon_{\mathrm{TT}} = \frac{18}{12}$, there is a significant increase in the concentration of unbound targets near the polymer - almost 5 times the bulk concentration for theta solvent with $\Delta E_0 = -4k_\mathrm{B}T$. Overall, this means that inter-target attraction leads to significant increases in both bound targets and unbound target clustering,

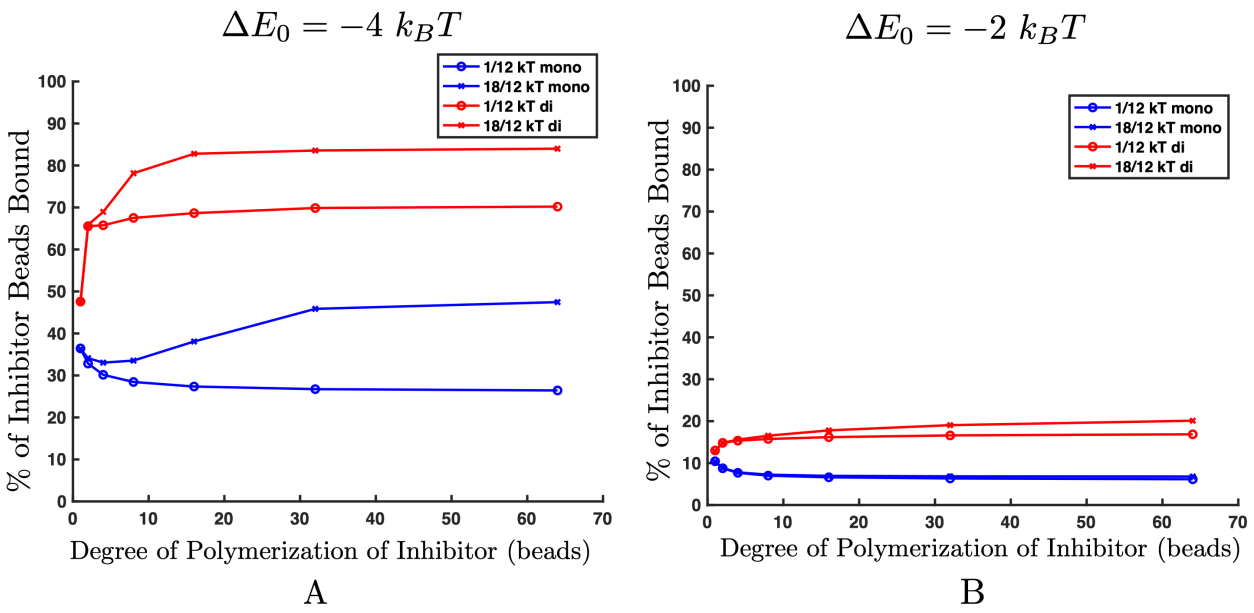


Figure 3-9: Percent of inhibitor beads bound in theta solvent when the target-polymer binding affinity is $\Delta E_0 = -4k_BT$ (A) and $-2k_BT$ (B). (A) As inhibitor length increases, a transition occurs that allows the polymer to bind a significantly higher percentage of targets when there is some inter-target attraction. This transition happens at approximately degree of polymerization of 10 for monovalent targets and degree of polymerization of around 3 for divalent targets. (B) At very low polymer-target binding affinities, such as $-2k_BT$, a critical percentage of targets never bind to the inhibiting polymer, so even at high degrees of polymerization, a transition in binding does not occur. Error bars are smaller than symbol size.

encouraging the collapse transition that makes the polymeric inhibitors more effective binders.

At a target-polymer binding affinity of $-4k_BT$, this effect is not specific to the divalent targets, and increased aggregation can also be seen for monovalent targets, although less extreme. But at lower binding affinities such as $-2k_BT$ shown in Fig. 3-9B and Fig. 3-12B, targets are unaffected by target-target attraction because they do not bind strongly enough to the polymer to create the critical concentration needed on the polymer to attract more targets. Therefore, the average number of targets bound to the polymer barely increases at higher inter-target attraction.

Above a critical length or a critical binding affinity, polymers are able to take advantage of weakly attractive inter-target interactions, increasing inhibitor binding avidity. Though competition for binding sites lowers the τ_B for individual targets, inter-target attraction allows the polymer to induce a collapse transition that clusters unbound targets, significantly increasing the binding of the polymer overall. With a sharper collapse transition and a diminished entropy of collapse and re-swelling, longer polymers should show an amplified effect.

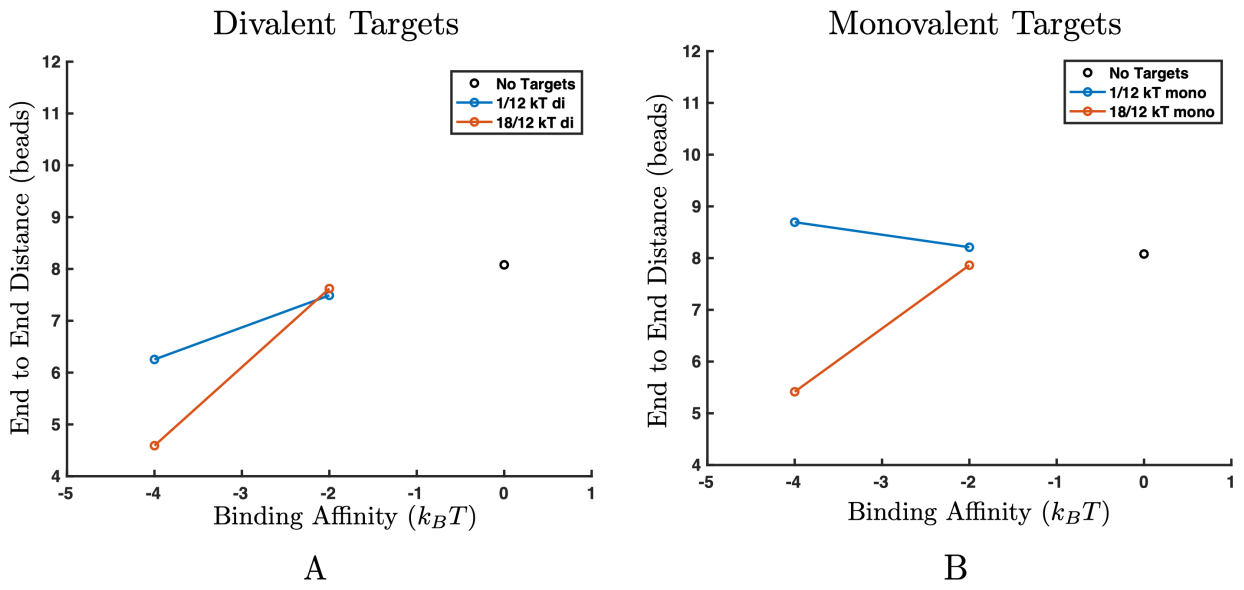


Figure 3-10: End to end distance for 64-mer polymers in theta solvent in the presence of divalent targets (A) and monovalent targets (B). (A) Increasing binding affinity between the targets and polymers induces a transition where the polymer collapses in size for both inter-target attractions. (B) Only high inter-target attraction leads to a collapse transition (orange). Low inter-target attraction (blue) does not provide enough enthalpic gain to overcome the entropic loss of phase separation. Error bars are smaller than symbol size.

3.4 Conclusion

This work has shown that increasing the degree of polymerization of a multivalent inhibitor increases the overall avidity of binding, but there is a limited increase in avidity at high degrees of polymerization. To explore the effect of multivalent polymer structure, we used a Brownian dynamics bead-spring model coupled with a reactive polymer-target binding model to investigate how degree of polymerization influences a polymeric inhibitor's avidity. First, we examined how the length of our inhibiting polymer modulates binding interactions with a single mono and divalent target model. We found that, consistent with previously reported experimental results for polymer binding to lectins, increasing the inhibitor length did increase binding avidity for multivalent targets, but interestingly, this effect was limited. We provide evidence that this limit can be explained by the entropic penalty of forming large loops; long polymers theoretically provide more possible loops when bound to a target in two places, but the entropic cost of forming long loops makes them unachievable in practice. Therefore, if the target is a globular protein, polymers longer than the maximum achievable loop length will demonstrate the maximum binding avidity. From an inhibitor design perspective, this means that the easier it is for loops to form, the greater the benefits from

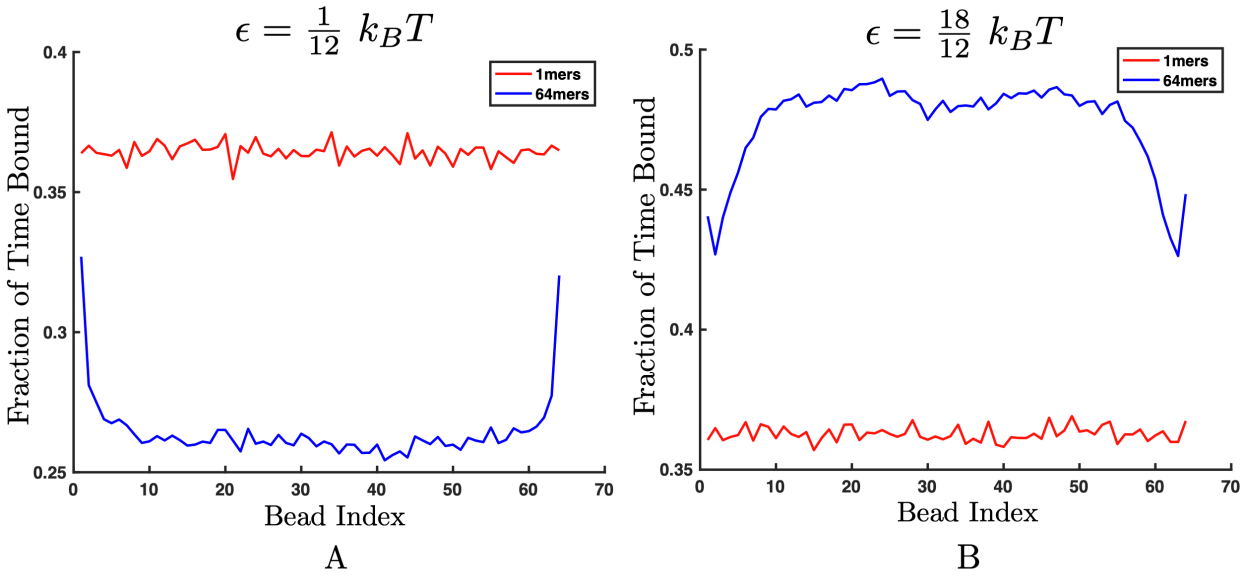


Figure 3-11: The amount of time each inhibitor bead spends bound when interacting with monovalent, $-4\ k_BT$ binding targets. Plots compare binding times for monomeric inhibitor beads (red) and beads that are part of a 64-mer polymer (blue). (A) Time bound when interacting with targets that have low ($\epsilon = 1/12\ k_BT$) target-target attraction. Monomers are each bound for a uniform amount of time, but the polymer ends are bound much more frequently than the polymer beads in the center of the chain. (B) When interacting with targets that have higher target-target attraction ($\epsilon = 18/12\ k_BT$), the polymer collapses, making the center beads bound more frequently than the chain ends. Monomeric inhibitor beads continue to experience uniform binding preference.

multivalent binding and lengthening a polymer. For example, our simulations show that increasing solvent quality discourages loop formation and causes ΔG to plateau more quickly and at a less favorable value. Likewise, our theory predicts that factors that discourage loops such as increased polymer stiffness and high amounts of swelling will reduce avidity. However, we do not address precise ligand engineering in this work; if ligands are spaced exactly to fit the receptor's binding sites, making a polymer stiffer may be an effective method to increase avidity [33].

Due to its estimation of the targets as point particles, our model works well for systems in which the binding sites are clustered in areas smaller than the distance between polymer binding sites, such as lectins or possibly clustered receptors on a surface. Our model does not address the experimental results that increasing polymer length continues to increase avidity for larger many-valent surfaces such as viruses. In this case, longer polymers may continue to show increased avidity as they are able to reach more binding sites along the surface and benefit from increased combinatorial entropy [11]. Consequently, for targeting viruses, researchers may want to continue creating polymers with higher degrees of polymerization.

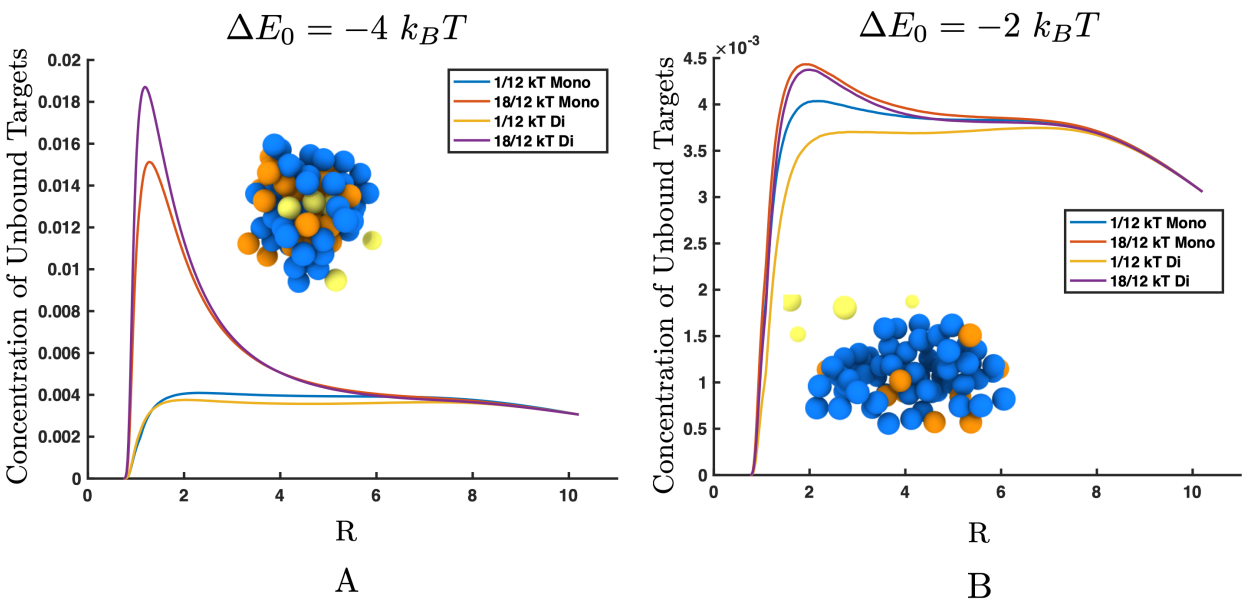


Figure 3-12: Plot of the minimum distance away from the polymer that unbound targets are found, normalized by the volume of a sphere with radius R, where R is the distance the center of the target is from the center of the nearest polymer bead. Data is shown for 64mer polymers in theta solvent with polymer-target binding affinities of (A) $-4k_{\mathrm{B}}T$ and (B) $-2k_{\mathrm{B}}T$. (A) The concentration of unbound targets is approximately the same as the bulk when there is low inter-target attraction, but the concentration of unbound target near the polymer is higher than the bulk concentration when the inter-target potential is increased. The rendering in the inset shows unbound targets (yellow) clustered inside the polymer (blue) by the bound targets (orange). (B) Fewer targets have bound to the polymer, so the polymer has not collapsed. This makes the local concentration of unbound targets near the polymer approximately the same as the bulk concentration for both high and low inter-target attractions.

In the presence of multiple targets, we found that longer polymers are able to use inter-target interactions to increase their avidity further. We show that despite decreased time bound for individual targets, longer polymers are able to bind to more targets simultaneously in the presence of favorable inter-target interactions. When inter-target attraction is present, longer polymers are able to induce a collapse transition where targets precipitate into a globule with the polymer, helping the polymer draw in a significant number of unbound targets. Increasing the concentration of unbound targets near the polymer makes the polymer better at clustering and binding targets. This could be a desirable effect in both the inhibition of targets and in other scenarios such as controlling biological signaling [105].

Our results suggest design rules for creating multivalent polymeric binders. With the understanding that increasing degree of polymerization has a limited effect on avidity in low target con-

centration environments and that inhibitor length can be used to induce phase separation in high concentration environments, future designers can focus on other variables when creating multivalent polymeric binders for proteins.

3.5 Appendix

3.5.1 Additional figures for good solvent

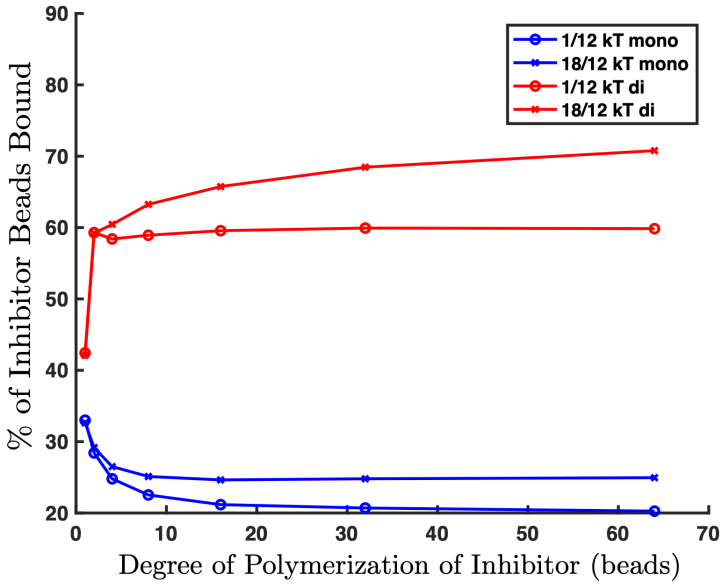


Figure 3-13: Percent of inhibitor beads bound in good solvent when the target-polymer binding affinity is $-4\ k_BT$. As inhibitor length increases, fewer monovalent targets (blue) are bound for both inter-target attractions because the enthalpic gain of targets binding does not overcome the loss of entropy. For divalent targets (red), longer polymers lead to an increase in binding avidity with higher inter-target attraction.

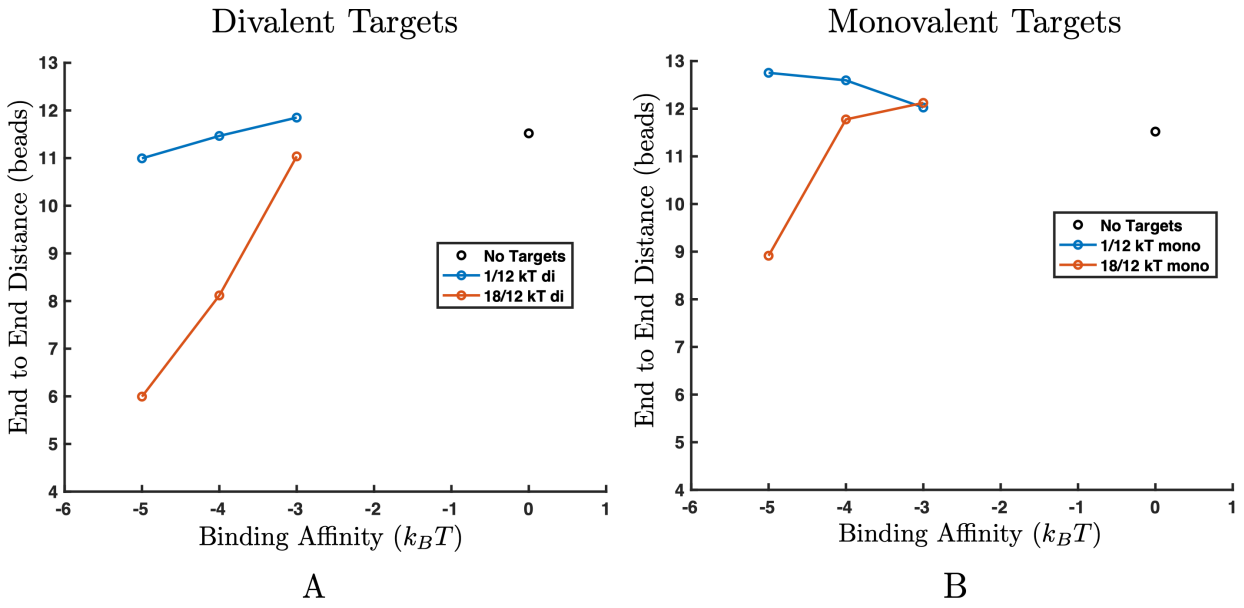


Figure 3-14: End to end distance for 64-mer polymers in good solvent in the presence of divalent targets (A) and monovalent targets (B). (A) Increasing binding affinity between the targets and polymers induces a collapse transition where the polymer distinctly collapses in size for higher inter-target attractions. This collapse in good solvent occurs at a stronger target-polymer binding affinity than in theta solvent. (B) Only high inter-target attraction leads to a transition where the polymer collapses. Low inter-target attraction does not provide enough enthalpic gain to overcome the entropic loss of phase separation.

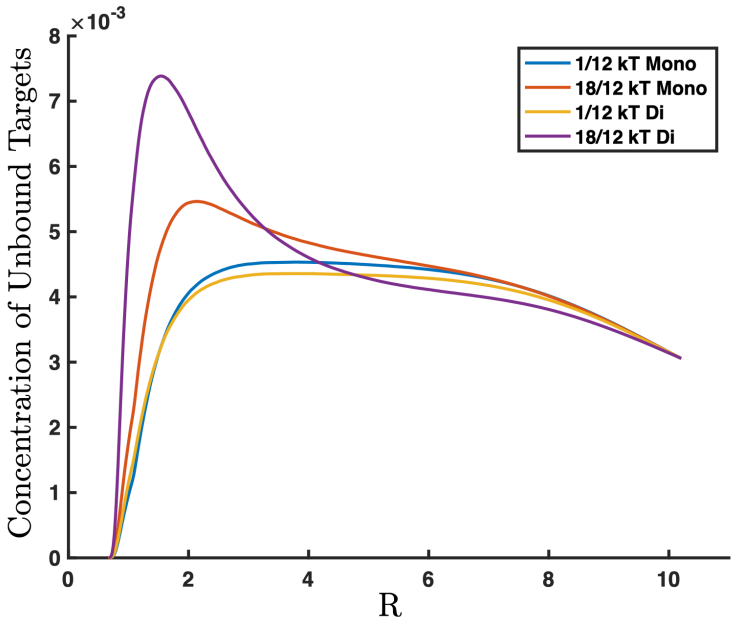
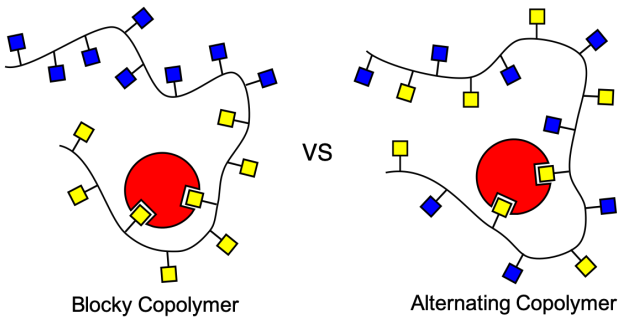


Figure 3-15: Plot of the minimum distance away from the polymer that unbound targets are found, normalized by the volume of a sphere with radius R, where R is the distance the center of the target is from the center of the nearest polymer bead. Data is shown for polymer-target binding affinity of $-4\ k_BT$ in good solvent. The concentration of unbound targets is approximately the same as the bulk when there is low inter-target attraction, but the concentration of unbound target near the polymer is higher than the bulk concentration when the inter-target potential is increased. This clustering of unbound targets is slight for monovalent targets because the polymer has not gone through a collapse transition, but unbound target clustering is significant for divalent targets because the polymer end to end distance has been greatly reduced.

Chapter 4

Influence of Binding Site Affinity Patterns on Binding of Multivalent Polymers



Abstract

Using inspiration from biology, we can leverage multivalent binding interactions to enhance weak, monovalent binding between molecules. While most previous studies have focused on multivalent binders with uniform binding sites, new synthetic polymers might find it desirable to have multiple binding moieties along the chain. Here, we probe how patterning of heterogeneous binding sites along a polymer chain controls binding affinity of a polymer using a reactive Brownian Dynamics scheme. Unlike monovalent binders which are pattern agnostic, we find that divalent binding is dependent on both polymer pattern and binding target concentration. For dilute targets, blocky polymers provide high local concentrations of high affinity sites, but at high target concentrations, competition for binding sites makes alternating polymers the strongest binders. Subsequently, we show that random

copolymers are robust to target concentration fluctuations. These results will assist in the rational design of multivalent polymer therapeutics and materials. The work presented in this chapter is primarily sourced from Zumbro and Alexander-Katz, *ACS Omega* **5** (2020) 10774–10781 [106].

4.1 Introduction

Multivalent polymers that bind to smaller targets are of interest in both biological and physical applications. In biology, multivalent interactions are used for a variety of reasons including enhancing weak monovalent binding or increasing specificity of binding using a limited number of receptor and ligand types [1]. Multivalent binding is defined as when multiple ligands on one species bind to multiple receptors on another species simultaneously. This can create a much stronger binding interaction than the sum of the corresponding monovalent single receptor/ligand interactions. In chemistry and materials science, multivalent polymers have been used to bind to multivalent crosslinkers to modulate gel characteristics [67]. Similarly, membraneless organelles also depend on the binding sequences of multivalent polymers to control gelation and liquid-liquid phase separation [58,63]. Furthermore, glycosylation of proteins in vivo often appears as a random process leading to a random arrangement of binding sites, but dysregulation of the sequence has been linked to neurodegenerative disorders [107]. Understanding the role of sequence in multimodal multivalent polymers and their influence on aggregation is thus of great interest to biology.

Synthetic multivalent polymers have also shown promise binding to sugar-binding proteins called lectins [94,92]. Sugar-protein binding sites frequently create low-affinity bonds, so multivalency can be essential to creating strong binding interactions [98,16]. Lectins are of special interest to us because viruses and bacteria use lectins to bind-to and subsequently infect cells, and microbes can release toxic lectins such as cholera or shiga toxin that cause diarrheal diseases [89,90]. Building synthetic multivalent inhibitors of lectins is a promising avenue for combating viruses, antibiotic resistant bacteria, and diarrheal diseases such as cholera [89,27,90,26,91,92,21,20] as shown in Figure 4-1.

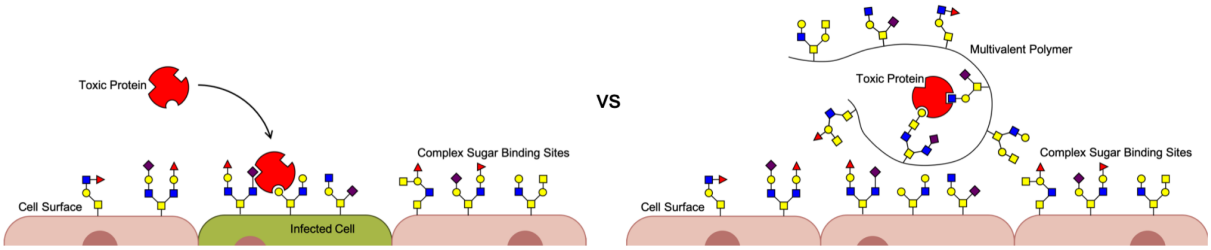


Figure 4-1: Multivalent polymers have shown promise as inhibitors for toxic lectins by preventing their attachment and subsequent infection to cells, as shown in the right panel.

Previous theoretical studies of multivalent structures with heterogeneous binding sites discussed the case of binding to a much larger flat multivalent surface such as Curk *et al.* who assumed very flexible ligands and focused on how changing overall receptor concentrations modulated binding of nanoparticles [12] and Tito *et al.* who examined the case of multivalent polymers binding to larger flat surfaces [108]. While these studies were well done, we wanted to investigate whether similar results could be found for multivalent polymers binding to much smaller targets such as folded proteins or nanoparticles. Theoretical studies have shown that interacting with small colloids can induce only a local conformational change in the polymer [37] whereas copolymers binding to a surface can create a strong conformational change leading to a stretched or even brushlike structure depending on other conditions [109,110]. This makes the scenario of binding to a much smaller target unique from binding to a surface. Experimental studies on polymers binding to multivalent proteins like lectins have focused on homopolymers with sites matched to a specific target lectin [90,32,33,97]. The ability to carefully control glycopolymer sequence was developed recently, and so comparatively few experimental studies have examined the effect of binding site sequence of heteropolymers on lectin binding [111]. Zhang *et al.* found some dependence of binding on copolymer sequence, but overall binding site concentration dominated the results, muddling the effects of sequence on binding to DC-SIGN [112].

Here, we examine polymers with multiple binding site types binding to globular protein targets such as a lectin. While holding the concentration of all binding site types constant, we explore how changing the pattern of binding sites along the chain affects binding. The study of copolymers as multivalent binders is interesting because of their potential use for binding to multiple targets for example targeting multiple lectins in the galactose-binding family. The binding specificity of

lectins to complex glycans is an active field of research. While lectins often target a particular monosaccharide or oligomeric sugar, the binding affinity can change based on the linkage or placement in a larger complex glycan ligand. For example, some galactose binding proteins can bind to both galactose and N-acetylgalactosamine, and the mannose-binding lectin concanavalin A binds to monomeric mannose, as well as mannose connected to various complex glycans with significantly different affinities [82,113]. Therefore, it is reasonable to assume that a binding site meant for one lectin might interact with another lectin or conversely that a single lectin might bind to two binding sites with different affinities. This "cross-talk" could significantly affect overall polymer binding. Unintentional heterogeneity is also important to investigate since imperfect grafting or other synthesis methods can create random binding site copolymers which could have a significant effect on target binding [42]. Additionally, in biological polymers such as mucins, the regulation and sequence of complex sugars are still not fully understood and might be heterogeneous [24].

In this chapter, we show that multivalent binding affinities are very different depending on polymer heterogeneity compared to monovalent binding. The binding affinity of monovalent targets to multivalent polymers is dependent on only the number and affinity of the highest affinity sites and not location. For multivalent targets, however, the results are more interesting. In dilute target conditions, the strength of the bond between the polymer and target is controlled by the highest affinity binding sites and the relative location between them. "Blockier" or clustered high affinity polymer binding sites create stronger binding to dilute multivalent targets. Alternatively, when many multivalent targets interact with patterned copolymers, the highest affinity polymers have alternating affinity binding sites while "blocky" copolymers have the lowest average binding to divalent targets. This results from competition between target for the same binding sites. Furthermore, we find that random copolymers are more robust to target concentration and perform mid-way between blocky and alternating copolymers in all target concentrations. We expect that these results will assist in the rational design of multivalent polymer therapeutics and materials.

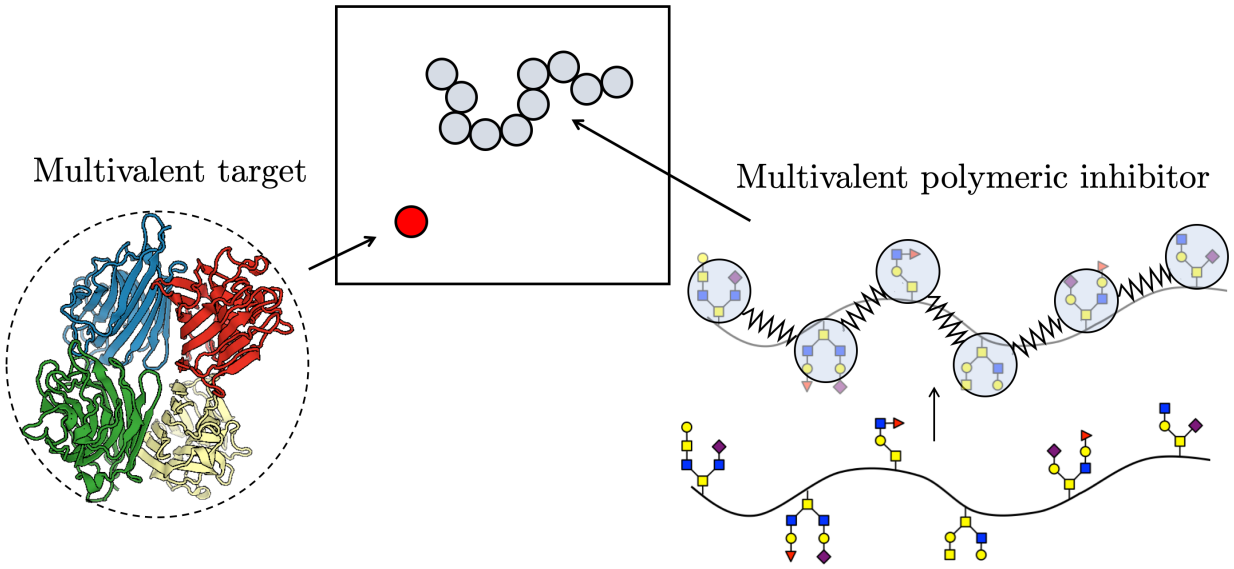


Figure 4-2: Schematic of simulation. The globular protein target is approximated as a sphere with one or more binding sites. The polymeric inhibitor is represented by a bead spring model where each bead has a single binding site and is connected to its neighbors through harmonic springs. Rendering from the Protein Data Bank [68,69]. This figure is reprinted from Zumbro *et al.* with permission from Elsevier [86].

4.2 Results and Discussion

To examine the effects of polymer binding site patterns, we placed four polymers with degree of polymerization of $N = 16$ beads in a cubic box with periodic boundaries. We chose a polymer length of $N = 16$ beads because previous work showed that increasing polymer length leads to a plateau in binding affinity after approximate lengths of $N = 13$ beads [86]. Using the same methods described in Chapter 2 with chapter-specific constants detailed in the methods section. Targets were represented by single beads of the same size as a polymer bead as shown in 4-2. Target beads were assigned one or multiple binding sites to represent monovalent or multivalent binding scenarios respectively.

Every polymer set was assigned a binding site pattern where each polymer bead was given a single binding site with a particular binding affinity ΔE_0 as shown in Figure 4-3. The binding site pattern parameter space is very large when we consider binding site energy, arrangement, and fraction of sites in the chain. Therefore, we have shrunk the parameter space to a more tractable subset where we consider polymers with 50% higher affinity binding sites and 50% lower affinity

sites. We believe that this case is still relevant to experimentalists who may only have two ligand chemistries available or who plan to target two proteins in the same family. We used polymers that had various patterns of 50% high affinity binding sites ($\Delta E_0 = -6k_BT$) and 50% low affinity binding sites ($\Delta E_0 = -2k_BT$), corresponding to monomeric binding affinities of $K_D = 0.02$ mM and $K_D = 0.8$ mM, respectively. Additional dissociation constant data for polymers with $\Delta E_0 = 0k_BT$ and $\Delta E_0 = -6k_BT$ binding sites and with $\Delta E_0 = -3k_BT$ and $\Delta E_0 = -5k_BT$ binding sites are included in the chapter appendix (4.5). In all cases we observe identical trends, and thus we only present the $(-2, -6)$ scenario. To generate randomly patterned polymers, we randomly selected half of the polymer bead indices and labeled those as high affinity sites $\Delta E_0 = -6k_BT$, the remaining half of the beads were labeled as low affinity sites. This created randomly patterned polymers while maintaining a $50 : 50$ ratio of high and low affinity sites. All of the four polymers in a simulation were assigned the same binding site pattern. For comparison, we also ran homogeneous polymers with uniform binding sites with $\Delta E_0 = -4k_BT$, corresponding to a monovalent binding site affinity of $K_D = 0.1$ mM. These binding affinities were calculated by fitting the Langmuir adsorption curve using the fraction of time bound (ϕ) of a monovalent target binding at different monomeric inhibitor concentrations. As detailed in Chapter 2, we can convert the unitless dissociation constant K_D to Molar by estimating a size of each bead in nm. These binding affinities capture relevant biological affinities of monovalent binding between sugars and proteins, commonly on the order of mM to μM [81,82].

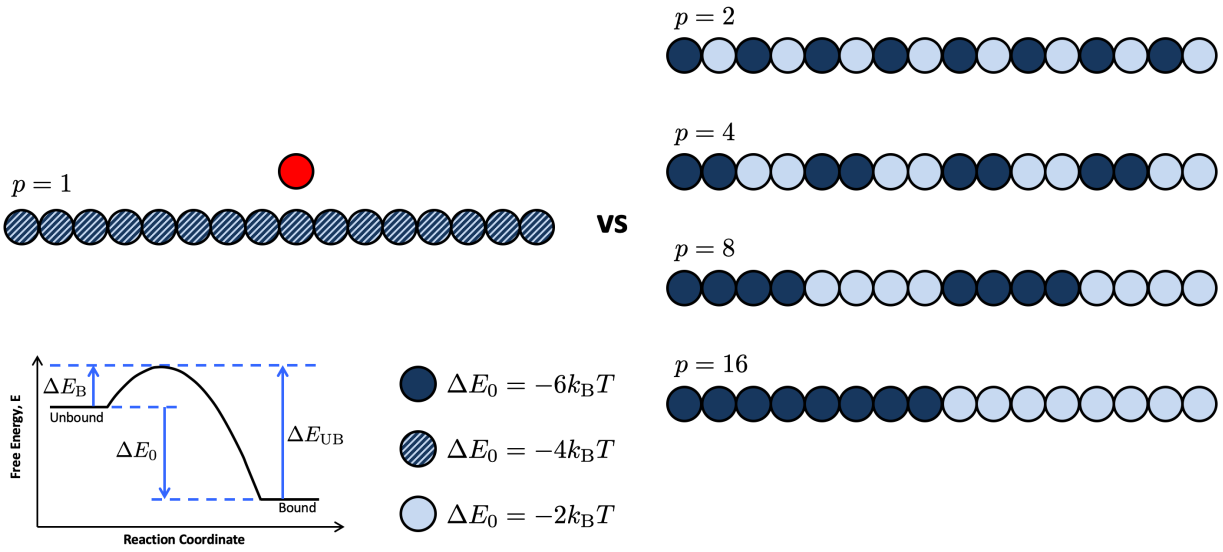


Figure 4-3: Schematic of the polymer patterns tested when exploring binding of a target (red) to homopolymers and copolymers (blues). The periodicity, p is labeled above each polymer pattern. Here, dark circles indicate high affinity binding sites with $\Delta E_0 = -6k_{\mathrm{B}}T$, light circles represent low affinity binding sites with $\Delta E_0 = -2k_{\mathrm{B}}T$, and striped circles represent a medium binding affinity used only for the homopolymer comparison with $\Delta E_0 = -4k_{\mathrm{B}}T$.

Throughout this work, we consider a target "bound" if one or more of its binding sites are bound to the polymer and "unbound" if the target has no bonds to the polymer. We analyzed the average time interval the target spent bound, τ_{B}, as we varied the binding site periodicity p while maintaining the $50:50$ high affinity and low affinity bead ratio. For example, an alternating high and low affinity polymer is considered to have a periodicity $p = 2$ and a polymer with half high affinity beads and half low affinity sites split down the center has $p = 16$ as shown in Figure 4-3. Results for four polymer periodicities $p = 2, 4, 8, 16$ with comparisons to a uniform binding site polymer and randomly patterned polymers are discussed in this work.

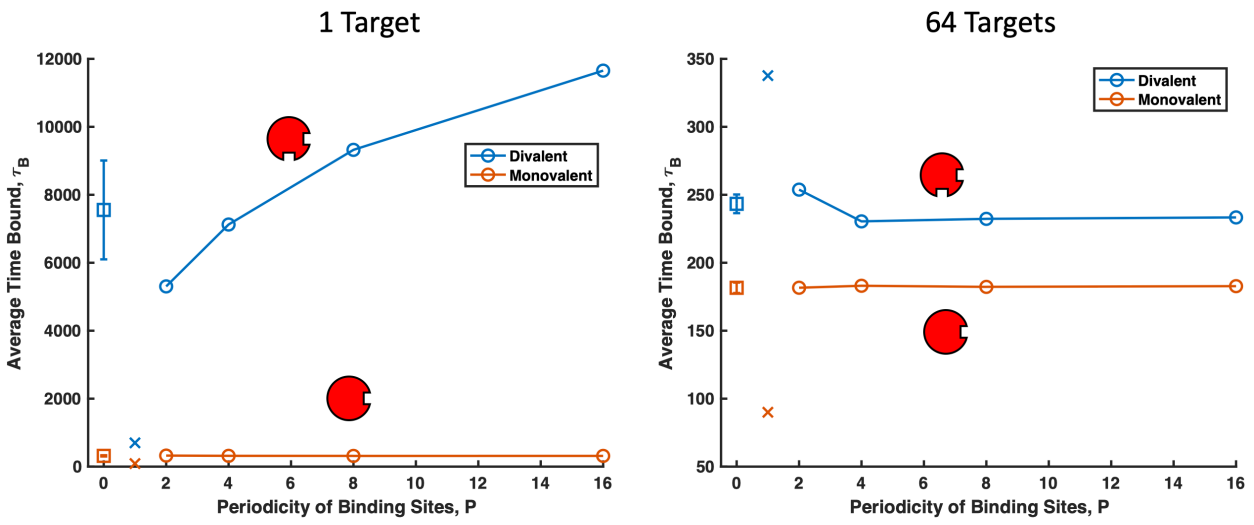


Figure 4-4: Plot of the average time bound τ_B vs the periodicity of the polymer p. The binding dependence on polymer pattern is different for divalent targets (blue) and monovalent targets (orange). Periodically patterned polymers are represented by connected circles (-o), homopolymers are represented as x's (x), and random copolymers are represented by squares (□). Because the binding of 100 co-polymer patterns were averaged, the standard deviation of the τ_B across random polymer patterns is depicted as error bars. The effect of pattern is also dependent on the concentration of targets. (A) At dilute target concentrations, target binding increases with copolymer periodicity but (B) at higher target concentrations low periodicity copolymers have higher τ_B. The sampling error for all data points is smaller than the symbol size.

4.2.1 Dilute Target Case

First, we considered a dilute target case where one target interacts in a box with four 16mer polymers. Assuming a target protein size of 5 nm, this corresponds to a target concentration of approximately 1.6μM. Results for τ_B at this dilute target concentration are shown in Figure 4-4A. For monovalent targets, τ_B is only affected by individual affinities of sites and is pattern agnostic. Shown in orange circles (-o) in Figure 4-4A, τ_B is higher for polymers with 50% $\Delta E_0 = -6k_BT$ affinity sites, than the uniform polymer (shown as an orange x in Figure 4-4A) with $\Delta E_0 = -4k_BT$ affinity sites. By plotting the fraction of time each site on the polymer chain is bound to a monovalent target in Figure S4A, we show that low affinity polymer sites are rarely bound, regardless of pattern periodicity. The design relationship for monovalent targets is straightforward: the affinity but not the relative position of sites controls the τ_B. Note that sites at the polymer ends do experience slightly higher binding than the center beads because polymer ends have less excluded volume

from neighbors, and so more available volume from which targets can bind. These end effects are relatively small contributors and are found across all polymer patterns. With only one binding site, monovalent targets can only sense the non-specific interactions of the polymer around them such as the Lennard-Jones potential, so they cannot distinguish between binding site patterns. Therefore, the binding of dilute monovalent targets is pattern agnostic and depends only on the strength and number of high affinity binding sites.

Next, we consider a single divalent target interacting with uniform and patterned polymers. Unlike monovalent targets, τ_{B} of divalent targets increases with p as shown in Figure 4-4A. A divalent target spends significantly more time bound to polymers with clustered high affinity binding sites than polymers with distributed high affinity sites. By examining which polymer beads are bound in Figure 4-5B, we find that for uniform polymers, beads in the center of the polymer are bound more often because they have the highest local concentration of binding site neighbors. Having the most binding site neighbors provides the highest chances for the target to create two simultaneous bonds.

From Figure 4-4A, we also see that on both the alternating polymer ($p = 2$) and the blocky polymer ($p = 16$), the low affinity binding sites are almost never bound (although the low affinity sites on the $p = 2$ polymer are bound slightly more often than those in $p = 16$). Comparatively, the high affinity sites on the blocky polymer are bound significantly more than the high affinity sites on the alternating polymer. This follows directly from our observation that clustered sites create increased opportunity for targets to become double bound. Blocky polymers have clustered high affinity sites, so targets can navigate to the high affinity block and will most likely become bound to two high affinity sites, creating a strong bond. In contrast, alternating high affinity sites are less occupied because for two sticky sites to be bound simultaneously, a divalent target has to form an entropically unfavorable loop. Targets prefer to bind to sites directly next to each other on the polymer to limit loop size and the corresponding polymer entropy loss as previously demonstrated in Zumbro et al [86]. A similar entropic penalty of loop formation has also been seen previously in the case of polymers binding to surfaces [108]. These loops make the alternating polymer less sticky than the blocky polymer in the case of dilute multivalent targets. While precise ligand design on the order of the target size is not considered in this work, previous research has shown that to

minimize entropic cost, binding sites should be spaced to exactly match the distance between target sites [33,34]. Therefore, when designing a polymer to bind with high affinity for a dilute target, the designer should use a blocky polymer whose binding sites are spaced the same distance apart as on the target.

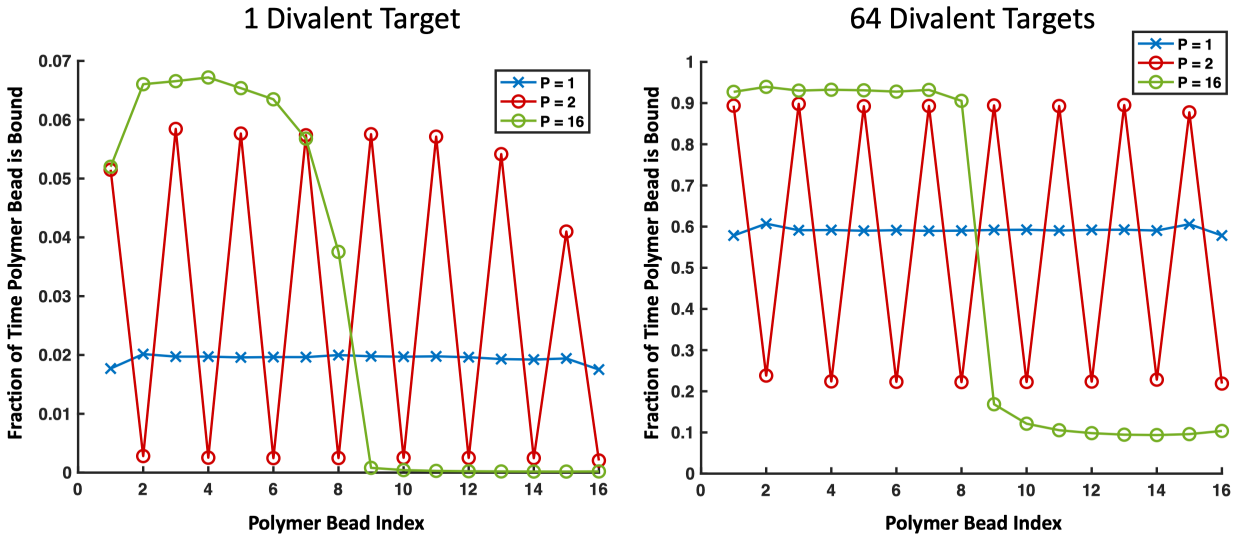


Figure 4-5: Frequency that a polymer bead is bound throughout the simulation when (A) a single divalent target and (B) 64 divalent targets are present for homopolymers (blue), alternating copolymers (red), and blocky copolymers (green). (A) For the patterned copolymers, low affinity binding sites are bound with almost the same frequency. However, the high affinity binding sites on the blocky polymer are bound much more frequently than the low affinity binding sites on the alternating polymer. (B) For the patterned copolymers, attractive binding sites are bound with almost the same frequency. However, the low affinity binding sites on the blocky polymer are bound much less frequently than the low affinity binding sites on the alternating polymer. Error bars are smaller than the symbol size.

4.2.2 High Target Concentration Case

We continued our exploration of the effect of polymer pattern by simulating the same polymer patterns shown in Figure 4-3 interacting with 64 targets to capture the case where multiple targets compete for binding sites. While previous theoretical investigation into competition of patterned polymers was between the polymers for the binding surface instead of between the targets for binding to the polymer, competition has been shown to significantly change the binding statistics [108]. Therefore again, we placed four 16mers in the box with our targets, so in this scenario, the

number of targets match the number of binding sites on the polymers. This higher concentration corresponds to approximately 100 μM assuming a 5 nm target diameter. Creating target competition for binding sites allows us to ask the question: how does pattern modulate multivalent binding when a target may not have access to the highest affinity sites? Competition for sites encourages faster turnover in bound targets because neighboring targets can steal polymer binding sites from each other. This faster turnover leads the drastically shorter τ_{B}s seen between Figure 4-4A and 4-4B. With competition, monovalent target binding was qualitatively unchanged. Monovalent targets were pattern agnostic and on average spent the highest τ_{B} on the patterned polymers with $-6k_{\mathrm{B}}T$ as shown in Figure 4-4B. For divalent targets, increased binding competition inverted τ_{B}'s dependence on polymer binding site periodicity as shown in Figure 4-4B.

When multiple targets interact with a single binding polymer, a uniform polymer with medium binding affinity sites has the highest overall avidity. The next highest τ_{B} is to the alternating high and low affinity polymer ($p = 2$), with blockier polymers $p = 4, 8, 16$ showing the shortest τ_{B}. By investigating which polymer sites are bound in Figure 4-5B, we find that the high affinity sites on the alternating polymer are now bound almost as often as the high affinity sites on the blocky polymer. In contrast, low affinity sites on the alternating polymer are significantly stickier than the low affinity sites on the blocky polymer. This is a result of restricted access to high affinity binding sites in blocky copolymers. When multiple targets are present, high affinity sites on the blocky polymer fill up, and unbound targets are forced into the low affinity region. In the low affinity half, targets are only able to bind two low affinity sites simultaneously – making relatively weak bonds. For the alternating polymer, targets forced to bind to the low affinity sites are still in close proximity to high affinity sites and can do a better job sharing sites with their target neighbors by binding to a high affinity site and a low affinity site simultaneously. This sharing makes alternating polymers the highest overall affinity of the patterned polymers for multivalent targets.

Because there is a transition in the binding as the concentration increases, there is some critical target concentration where the polymer pattern should not matter reflected as when target binding time is not dependent on the polymer periodicity. Because competition between targets for high affinity sites is causing the transition, we expect that the transition concentration should be approximately the concentration at which competition starts. Whenever there are multiple targets, there

will be some competition for sites, but we believe this competition will start to dominate when there are enough targets to bind to all high affinity polymer sites. This can be described quantitatively as when $C_{\mathrm{t}} = \frac{C_{HA}}{v_{\mathrm{t}}} = 16$ where C_{t} is the concentration of targets, $C_{\mathrm{HA}} = 32$ is the concentration of high affinity binding sites, and v_{t} is the valency of the target, in this case $v_{\mathrm{t}} = 2$. We expect the critical concentration to be slightly above this because the number of targets must exceed the available binding sites to create competition.

To investigate this critical target concentration, we plotted the dissociation constant K_{D} from simulations with C_{t} between 1 and 96 in Figure 4-6. We calculated the dissociation constant using $K_{\mathrm{D}} = \frac{\tau_{\mathrm{UB}}}{\tau_{\mathrm{B}}}$, where τ_{UB} is the average time interval spend unbound. We consider a target unbound whenever both binding sites are unbound. From this data we can see that the critical concentration occurred somewhere between $C_{\mathrm{t}} = 20$ and $C_{\mathrm{t}} = 24$. This is very close to our theoretical estimate of 16 targets as our critical concentration. The difference of 4 to 8 targets is most likely due to significant concentration occurring only when there is an additional target (above the full capacity) for each chain to compete with, but could also be explained by there being a low affinity site for each of the 4 polymer chains that has access to a high affinity site. A target that bound there could form a relatively favorable high and low affinity bond, almost creating another good binding site per chain. Either of these effects or a combination of both could increase the critical concentration slightly above $C_{\mathrm{t}} = 20$. Following these results, we expect that designers can perform our simple estimation that the alternating polymer becomes higher affinity than the blocky polymer when the target concentration exceeds $C_{\mathrm{t}} = \frac{C_{HA}}{v_{\mathrm{t}}}$.

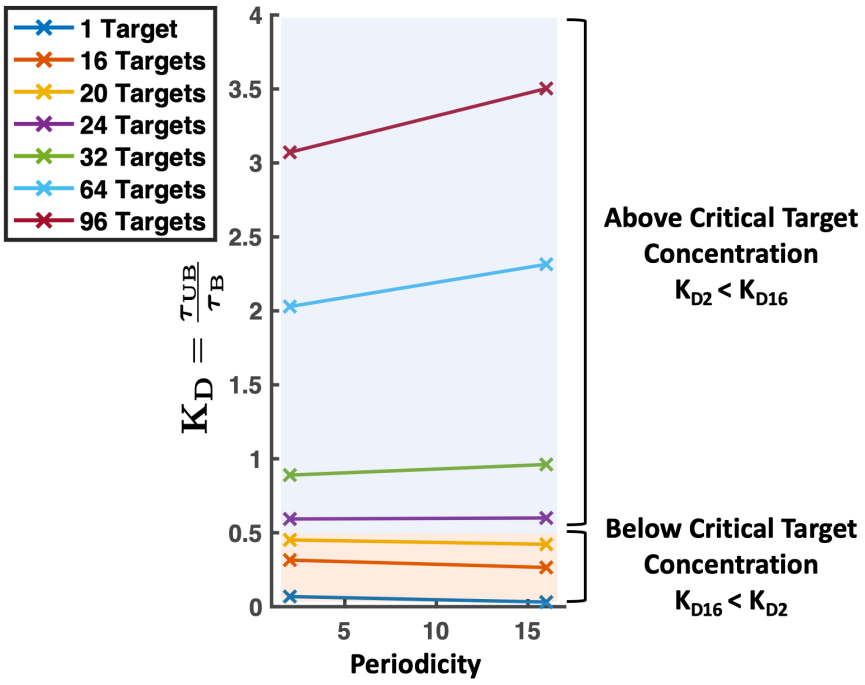


Figure 4-6: Dissociation constant K_D versus periodicity of polymer pattern for target concentrations from 1 to 96. We have marked the concentrations below the critical target concentration where the blocky polymer ($p = 16$) has a K_D less than that of an alternating polymer ($p = 2$) with an orange background. The values above the critical target concentration where the alternating polymer has a lower K_D than the blocky polymer has been labeled with a blue background.

4.2.3 Unknown Concentration

Because binding dependence on polymer pattern changes with target concentration, we subsequently explored the use of a random copolymer containing some blocky and some alternating areas. We hypothesized that polymers with both high and low periodicity binding sites would have binding behavior more robust to fluctuations in target concentration. We examined simulations with randomly patterned binding sites. To create random patterns while maintaining the 50/50 ratio of high to low affinity sites, we randomly chose 50% of the beads along the polymer chain to be high affinity ($-6k_BT$) sites, and the rest were labeled as low affinity ($-2k_BT$) sites. We averaged the performance of 100 of these different polymers, with their standard deviation of performance denoted as error bars in Figure 4-4A and 4-4B. As expected, we found that randomly patterned copolymers resulted in τ_B between that of polymers with $p = 1$ and $p = 16$ for both dilute and more concentrated divalent target scenarios as shown by the squares (□) plotted at $p = 0$ in Figure 4-4. Pattern continued

to have negligible affect on binding of monovalent targets. This suggests that in an unknown or fluctuating target concentration, a polymer with both blocky and alternating regions, such as a randomly patterned multivalent polymer may provide the broadest binding capabilities.

4.3 Conclusion

We have examined how binding site patterns along the polymer chain influence their average binding time to both monovalent and multivalent targets. In this paper, we have shown that for targets with a single binding site, the polymer is only as sticky as its highest affinity site. For targets with multiple binding sites, the effects of polymer binding site pattern are more nuanced. In dilute target conditions, polymers bind multivalent targets more tightly when high affinity sites are concentrated, so blocky copolymers are better binders than alternating copolymers. Blocky polymers also provide areas of high local concentration of high affinity sites, assisting divalent targets in to form two strong bonds. For targets to bind two sticky sites on an alternating polymer, they must form an entropically unfavorable loop with a low affinity bead, making these polymers worse binders. In crowded environments, the opposite result was found; with multiple competing targets, alternating high and low affinity binding sites were bound the longest of patterned polymers. When many targets bind to the same polymer, blocky designs with clusters of high affinity sites performed the worst because high affinity sites filled up and leftover targets were excluded from the high affinity region. Alternating polymers were able to share their high affinity sites to improve binding performance overall. Consequently, our work suggests that the pattern of multivalent polymers should be adjusted to their binding target application.

If target concentration is unknown, our results show that the most robust polymer pattern to bridge many target concentrations is a polymer with both blocky and alternating regions. While this could be achieved with a carefully crafted blocky and alternating copolymer, here we showed an example of this concept with a random copolymer which had τ_B's between those of alternating and blocky copolymers in both target concentrations. Therefore, for improved performance in fluctuating target concentrations, a random copolymer or other design with blocky and alternating regions may be the best choice of polymeric inhibitor. Understanding how patterns of multiple types

of binding sites on polymeric inhibitors affect the polymer's binding behavior to a single target type is an essential first step toward rational design of polymers that display multiple moieties to fulfill several simultaneous functions. The ability to tune a single polymer design to bind to multiple types of targets means that multivalent polymers could be used as "broad-spectrum" inhibitors of microbial or viral infections. Finally, our results clearly show that the effective interactions between multivalent biopolymers/proteins is sequence dependent, and modifications to such sequences can lead to clear changes in binding behavior. For example, in liquid-liquid phase separation, small changes in sequence could lead to large repercussions in the assembly and should be studied further.

4.4 Computational Methods

This chapter uses all of the same general simulations methods as described in Chapter 2. Across the simulations in this work, we have chosen $\epsilon_{\mathrm{PP}} = \frac{5}{12}$ for the Lennard-Jones parameter (Eq. 2.3) to mimic polymer configurations in a theta solvent [70]. We used polymer target potential $\epsilon_{\mathrm{PT}} = \frac{1}{12}$ and target-target potential $\epsilon_{\mathrm{TT}} = \frac{1}{12}$ to mimic a good solvent as summarized in Table 4.1. We chose theta solvent because we previously demonstrated that polymer loops are easiest to form when the polymer is in the smallest size because the entropic penalty of forming a loop is the lowest [86]. Since having a more collapsed polymer creates a higher local concentration of binding sites, a target within reach of the polymer should find more accessible binding sites on a collapsed chain as opposed to a swollen chain. Therefore, overall pattern should matter less for a collapsed chain, so we have used theta solvent as our limiting case. We expect that using a better solvent would further restrict the binding sites available to a target and magnify the effects of local pattern on binding.

Table 4.1: ϵ values for Polymer-Polymer (PP), Polymer-Target (PT), and Target-Target (TT) bead Lennard-Jones Interactions

ϵ_{PP}	ϵ_{PT}	ϵ_{TT}
5/12	1/12	1/12

4.5 Appendix

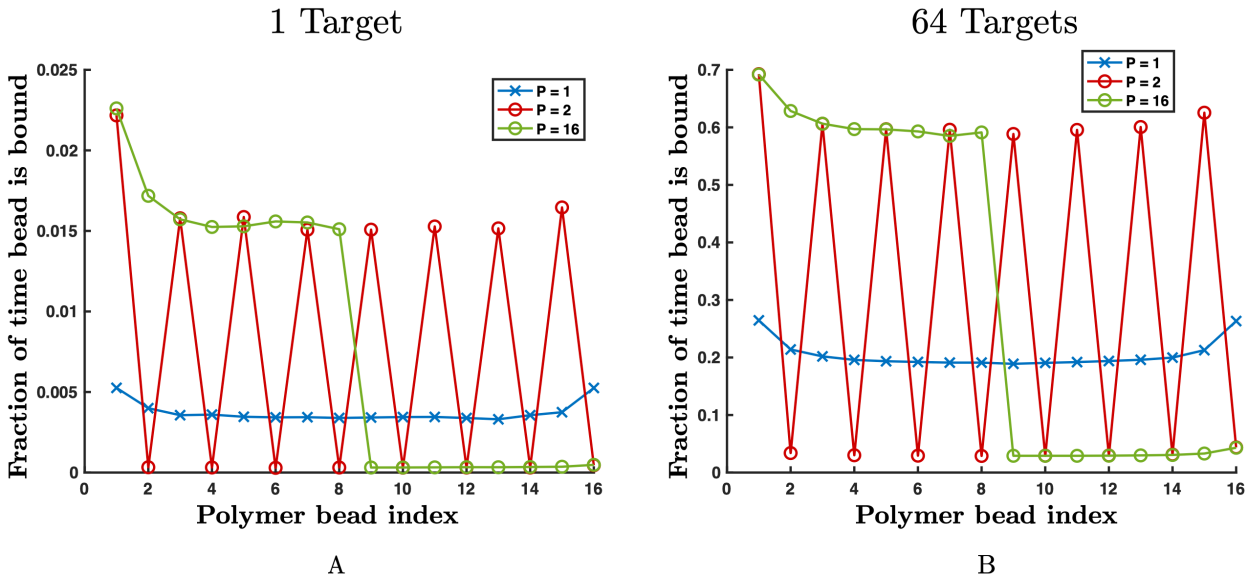


Figure 4-7: Frequency monovalent targets are bound to binding sites on homopolymers and copolymers with alternating and blocky patterns. Results are shown for (A) when a single target is placed with 4 16mer inhibiting polymers and (B) when 64 targets are placed with 4 16mer inhibiting polymers. Frequency of time bound depends on the affinity of that polymer binding site and not on polymer binding site pattern.

4.5.1 Monovalent binding frequency

In Figure 4-7, we have plotted the frequency each bead along the polymer chain is bound for monovalent targets to homopolymers $p = 1$, alternating heteropolymers $p = 2$, and blocky polymers $p = 16$. High affinity beads are bound with approximately the same frequency for all copolymer patterns. While the absolute fraction of time bound is different for low and high target concentrations, the qualitative results are the same. Monovalent targets bind to sites of the same affinity with the same frequency, regardless of copolymer pattern. The same is true of low affinity sites.

4.5.2 Other binding affinities

We also tested other binding affinities pairs both farther apart in energy $(0,-6k_{\mathrm{b}}T)$ and closer in energy $(-3,-5k_{\mathrm{b}}T)$. Without competition, dilute targets will still try to minimize the entropic cost of loop formation by binding to the two highest affinity sites, so we expect that in the dilute case

blocky polymers will always be higher affinity than alternating polymers even with different binding affinity pairs. In competition with 64 targets, we also expect the same results as the main text for different binding affinity pairs, which we confirmed with simulation shown in Figure 4-8. In the case of competition, even when the low affinity sites go to $\Delta E_0 = 0k_bT$, we still see a lower K_D for the alternating polymer most likely due to the larger available free volume around the high affinity sites. Larger spacing between the high affinity sites gives unbound targets more free volume to approach and bind to the high affinity sites, even when they are already bound by competitors.

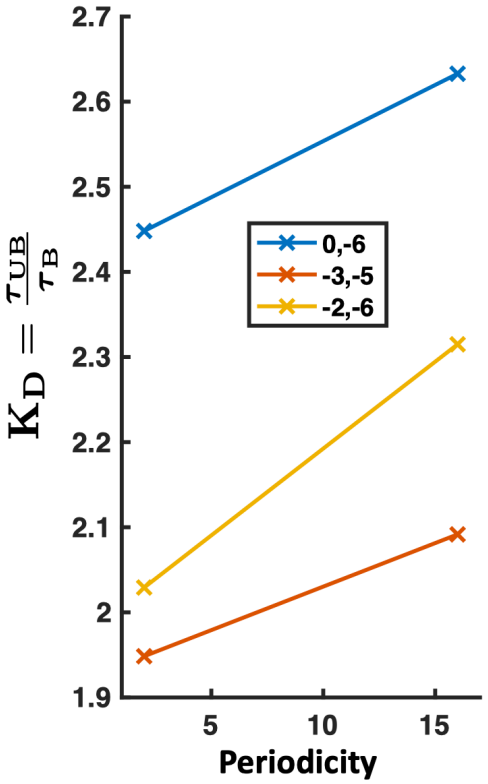


Figure 4-8: Dissociation constant for alternating ($p = 2$) and blocky ($p = 16$) polymers with ΔE_0 pairs (0,$-6k_bT$), (-2,$-6k_bT$), and (-3,$-5k_bT$). All data shown is for high competition simulations with 64 targets.

4.5.3 Effective target valency

We also thought it was interesting to examine how the target bonding changes as competition increases and with the polymer binding site pattern. In Figure 4-9, we have plotted the fraction of all time bound that the target is bound divalently, as well as the fraction of time spent in each

type of divalent bond. We show separately, the fraction of bound time the target spends with two high affinity ($-6k_{\mathrm{b}}T$) bonds, two low affinity bonds ($-2k_{\mathrm{b}}T$), and one of both low and high affinity bonds. From this plot we can see that as the blockiness or periodicity of the polymer increases, the total fraction of time spent bound divalently stays almost constant, but the types of divalent bonds change drastically. For example, in the 64 target case, two high affinity bonds account for 62% of all bonds for the blockiest copolymer, but only 41% of all bonds in the alternating polymer. Divalent bonds with both a high and low affinity bond follow the opposite trend. For the same 64 target concentration we can see that these combination bonds account for almost 28% of bonds in the alternating polymer and much less, only 7% of bonds in the blocky copolymer. These results align well with those presented in the main text showing that targets attempt to make divalent bonds with the polymer; the majority of all bonds formed are divalent. When bonding divalently, targets prefer to bind twice to high affinity beads, but also seek to decrease loop length. This can be seen by the lower number of two high affinity bonds for the alternating polymer than for the blocky polymer, and the increase in both low affinity/high affinity bonds for the alternating polymer over the blocky polymer.

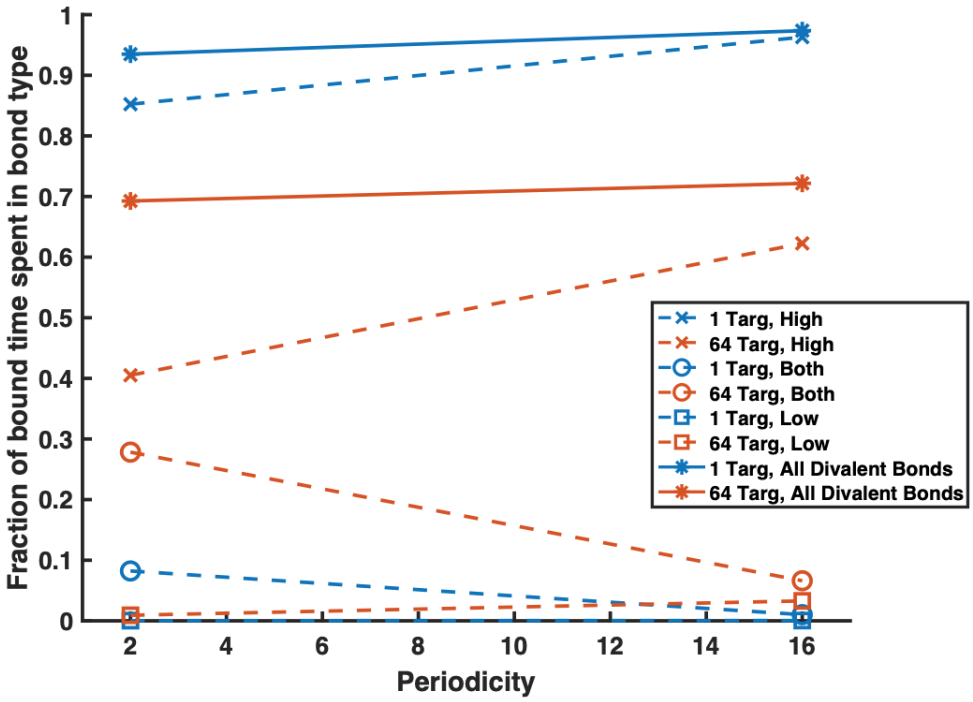


Figure 4-9: Fraction of all time spent bound that a target is bound divalently for a single target interacting with four polymers in orange (-$*$) and for 64 targets interacting with polymers in blue (-$*$). Fraction of time bound is also plotted for all three divalent bond types: two high affinity bonds (–x), two low affinity bonds (–□), and bonds with one low and one high affinity bonds, labeled as "Both" in the legend (–o). Values are shown for two polymer periodicities where ($p = 2$) is an alternating polymer and ($p = 16$) is a block copolymer.

This Page Intentionally Left Blank

Chapter 5

Polymer Stiffness Regulates Multivalent Binding and Liquid-Liquid Phase Separation

Abstract

Multivalent binding is essential to many biological processes because it builds high affinity bonds by using several weak binding interactions simultaneously. Multivalent polymers have shown promise as inhibitors of toxins and other pathogens, and they are important components in the formation of biocondensates. Explaining how structural features of these polymers change their binding and subsequent control of phase separation is critical to designing better pathogen inhibitors and also to understanding diseases associated with membraneless organelles. In this work, we will examine the binding of a multivalent polymer to a small target. This scenario could represent a polymeric inhibitor binding to a toxic protein or RNA binding to an RNA-binding protein in the case of liquid-liquid phase separation. We use simulation and theory to show that flexible random-coil polymers bind more strongly than stiff rod-like polymers and that flexible polymers nucleate condensed phases at lower binding energies than their rigid analogues. We hope these results will provide insight into the rational design of polymeric inhibitors and improve our understanding of phase separation in cells and membraneless organelles. The work presented in this chapter is primarily sourced from Zumbro and Alexander-Katz, *In Review* (2020) [114].

Statement of Significance

Multivalent polymers are essential for many biological systems, including targeting pathogens and controlling the formation of liquid-liquid phase separated biocondensates. Here, we explain how increasing polymer stiffness can reduce multivalent binding affinity to a small target such as a toxic protein and how modulating polymer stiffness can change the phase boundary for liquid-liquid phase separation. These results have implications for designing stronger pathogen inhibitors and provide insights on neurodegenerative diseases associated with abnormal biocondensate formation.

Introduction

Multivalent binding interactions are commonly found throughout biology and synthetic applications. These interactions use multiple weak binding sites to simultaneously bind to another species. Using many low-affinity binding events simultaneously enhances the overall binding affinity much more than the sum of the constituent monovalent binding interactions [1]. Multivalent binding can take on many different geometries and previous research has been done on nanoparticles, sheets, dendrites, and polymers for numerous applications [90,14]. In this work, we will focus on multivalent polymers as they pertain to toxin inhibition along with implications for nucleating liquid-liquid phase separation in biocondensates.

Synthetic multivalent polymers have shown promise at binding to and inhibiting multivalent sugar-binding proteins called lectins [94,32,95,31,97,16]. Monovalent sugar-protein binding affinities are typically weak, in the millimolar to micromolar range, so multivalency is essential to creating high binding affinities [98,16]. Binding to lectins is an exciting avenue for combatting infection because many toxins are lectins such as cholera toxin, shiga toxin, and others that cause diarrheal diseases, and because bacteria and viruses use lectins on their surface to bind to the glycocalyx on our cells [89,90]. Mucins, the megadalton weight glyprotein polymer found in mucus, are thought to use their glycan brushes as binding decoys, exploiting multivalency to bind to pathogenic lectins and prevent infection [23]. Attempts to mimic this capability with synthetic polymers have been successful, but the effects of polymer backbone flexibility and characteristic ratio C_∞ on the binding of polymers much larger than their targets has not received theoretical study. Previous studies on the flexibility

of multivalent binding have focused on species of similar size binding to each other. In these cases, small molecules or oligomers with binding sites precisely spaced to the target found rigid linkers to minimize the entropic cost of binding, resulting in the highest binding affinity [33,115,90,32,97]. When the polymer chain's end-to-end distance is on the same scale as the binding target, stiff linkers between binding sites on an antibody and rigid sections near the ligands of a divalent binder were shown to be higher affinity than their flexible counterparts [116,35]. In contrast, very few studies have considered polymers much larger than the size of their targets, even though this is the scale of native mucins and many of the previously mentioned synthetic inhibitors tested experimentally [37,86]. We anticipate that because, unlike small divalent oligomers, many-valent polymers allow for many binding site pairs with different spacings, large multivalent polymers may benefit from higher flexibility which allows them to sample more binding combinations [14,1,83]. Theoretical research on this relevant size scale has not considered the effect of stiffness of the polymer chain and how this controls binding affinity to a small multivalent target. Here, we examine how a single target binds to large many-valent polymers of increasing stiffness and provide a theoretical explanation for the difference in binding modes between a flexible random coil polymer and a stiff wormlike polymer chain.

Understanding multivalent polymers and their binding is also essential to controlling liquid-liquid phase separation in membraneless organelles [50,49,63]. Research has shown that polymeric binding characteristics such as valency and individual binding site strength can be used to control the phase separation boundary [83,58,50]. Other studies have shown that polymer properties indirectly related to binding sites such as solvation volume can determine the difference between a cross-linked gel and a phase separated system [58]. Because of this, we expect that flexibility of the polymer could also be an important factor in controlling liquid-liquid phase separation. Dysregulation of the phase separation in membraneless organelles is a common feature of neurodegenerative diseases like Alzheimer's, Parkinson's, and ALS [49–51], and so investigating how features of multivalent polymers can change the phase boundary are essential. We hope that this research will contribute understanding to how the stiffness of polymer chains can modulate nucleation of condensed phases and thus how changes in polymer stiffness could lead to aberrant condensates or disease.

In this work, we focus on how the change in polymer stiffness modifies its binding affinity to a

much smaller target. This scenario could represent a coarse gain model of a polymer binding to lectin in the case of toxin inhibition or a long section of RNA binding to a smaller RNA-binding protein in the case of biocondensates [117,118]. First, we will discuss the case of a single target binding to the polymer and provide a theoretical understanding of how polymer stiffness changes binding affinity. In the second half of the paper, we will present results for many targets simultaneously binding to the polymers for an array of target solubility limits. We explore how polymers can nucleate condensed phases and how polymer stiffness changes this phase boundary. We hope that these results can aid in the study of polymeric inhibitors as well as in the understanding of liquid-liquid phase separation of biocondensates.

5.1 Computational Methods

In this chapter we used the general simulation methods from Chapter 2 and added an additional spring force to prevent polymer bending. This adds another term to the potential energy U from Eq. 2.1 so that in this chapter U combines contributions from connectivity, bending, excluded volume, and binding as $U = U_{\text{sp}} + U_{\text{Bend}} + U_{\text{LJ}} + U_{\text{Bind}}$.

To control the flexibility of the polymer chain we use the Kratky-Porod wormlike chain model [120,121], and introduce an additional spring placed between every next nearest neighbor along the chain. This is a commonly employed scheme in some force fields such as MARTINI [119]. This spring imposes an energetic penalty for bending around the straight/stiff configuration between two bonds as shown in Figure 5-1. This added spring is implemented using the potential below:

$$U_{\text{Bend}} = \frac{\gamma}{2a} \sum_{j=2}^{N_{\text{P}}-1} \boldsymbol{u}_j \cdot \boldsymbol{u}_{j-1} \tag{5.1}$$

where γ is the bending rigidity and $2a$ is the equilibrium length between two bead centers. $\boldsymbol{u}_j$ is the unit vector pointing from bead j to bead $j+1$ [122,123]. This allows us to control the persistence length or C_∞ by modulating γ. When $\gamma = 0$ we reproduce a freely jointed chain [124], and as γ increases, the chain becomes a semiflexible polymer and a stiff, rod-like polymer at high γ. The end-to-end distance of the polymer with increasing values of γ is shown in Figure 5-2. The corresponding

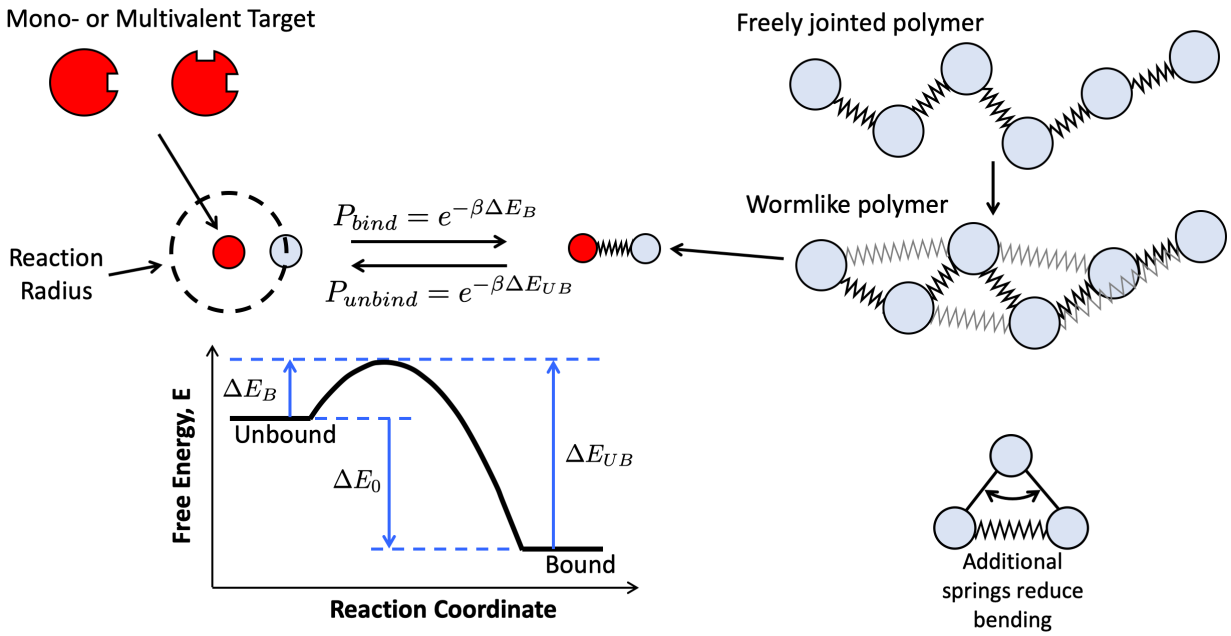


Figure 5-1: Depiction of simulation scheme. Polymers are represented by spherical beads (light blue) connected by harmonic springs. To introduce stiffness, we employ a simple scheme used also by some commonly utilized force fields (e.g. MARTINI [119]), where an additional spring is placed between every next nearest neighbor along the chain. Each polymer bead has a single ligand, meaning it can only bind monovalently, but making the polymer as a whole multivalent. Targets, on the other hand, can have multiple binding sites and are represented by a single spherical bead (red) with one or two binding sites as shown. Polymer ligands and target binding sites interact when they are within a reaction radius. Within this reaction radius, they have a probability of binding P_{B} that depends on the free-energy landscape, as depicted. Once bound, the target and polymer bead are connected by a harmonic spring, and they can unbind with probability, P_{UB}. Apart from the reactive kinetics that we include here to model the specific binding mechanisms, we use a Lennard-Jones potential to maintain the chain conformation and prevent target-target and target-polymer overlap. This figure is adapted from Zumbro *et al.* with permission from Elsevier [86].

values of persistence length p and characteristic ratio C_∞ are also displayed for reference.

Across the simulations in this work, we have chosen $\epsilon_{\mathrm{PP}} = 0.41$ to mimic polymer configurations in a theta solvent [70]. We used polymer target potential $\epsilon_{\mathrm{PT}} = 0.1$ and target-target potential $\epsilon_{\mathrm{TT}} = 0.1$ (unless otherwise stated) to mimic a good solvent as summarized in Table 3.1 Case 1. We chose a theta solvent for the polymer because this is the lower limit in end-to-end distance of a soluble polymer. We would expect a polymer in good solvent to follow similar trends as shown in our previous work, with a smaller range of possible chain end-to-end distances, although this distinction for stiff polymers becomes even more irrelevant.

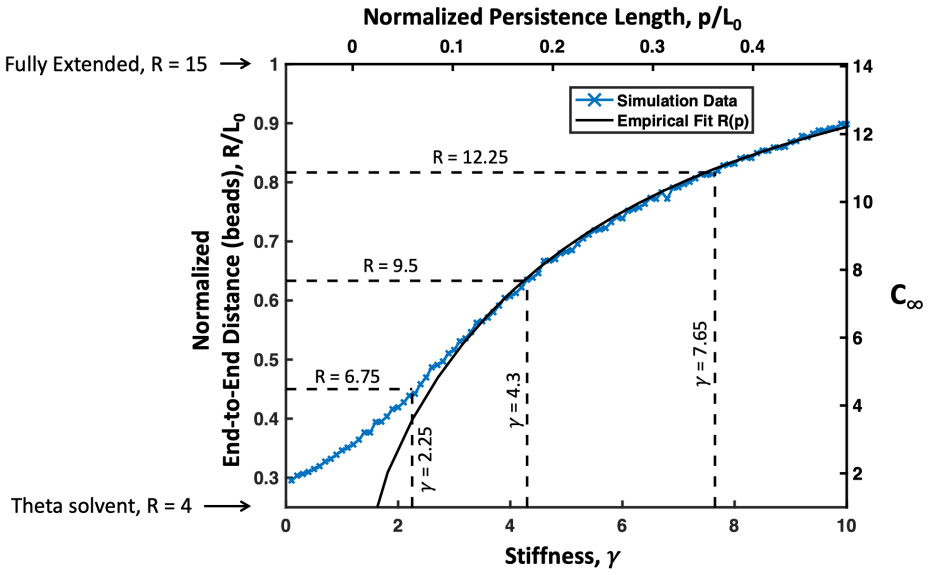


Figure 5-2: Simulated average end-to-end distance R of 16mer polymer chain normalized by the contour length L_0 plotted versus chain stiffness spring coefficient γ is shown as blue X's. Values of R and γ at which simulations were run are highlighted with dashed lines. These values of γ were chosen to explore a wide range of polymer flexibilities and represent the point where $R \approx 4$ for a perfectly flexible polymer, and 25%, 50%, and 75% of the distance between the most flexible chain $R \approx 4$ and a perfectly rigid rod where $R = L_0 = 15$. End-to-end distances were converted to C_∞ on the right axis and persistence length, p on the top axis using the empirical wormlike chain fit relating R/L_0 to p/L_0 (black solid line) [121,55].

We used our reactive binding scheme with a free energy of binding per site of $\Delta E_0 = -4k_BT$. Each polymer bead contained a single binding site, and each target bead was given $M = 1$ or $M = 2$ binding sites so that we could compare the monovalent case to the divalent case. The binding sites are isotropic to avoid unnecessary assumptions about binding site orientation and geometry in this coarse-grain model. This means that binding is attempted and can be successful whenever the centers of an unoccupied polymer bead and unoccupied target bead are within a distance of $r_{\text{rxn}} = 1.1$. Choosing a particular orientation of binding sites begins to enter the realm of precise ligand engineering, in which extensive previous theory and experiments have shown that perfectly matched receptor and ligand spacing enhances binding avidity [33,115,90,32,97].

The simulation box had periodic boundaries with side length $l_{\text{box}} = 41a$. Because the box size is dependent on the size of the beads, we can convert the target concentration from beads per box to Molar by assuming a binding protein radius. Unless otherwise specified, the polymer and target beads are the same size. Assuming the target bead radius to be $a = 2.5$ nm results in a target

Table 5.1: ϵ values for Polymer-Polymer (PP), Polymer-Target (PT), and Target-Target (TT) bead Lennard-Jones Interactions

Case #	ϵ_{PP}	ϵ_{PT}	ϵ_{TT}
1	0.41	0.1	0.1
2	0.41	0.1	1.0
3	0.41	0.1	1.25
4	0.41	0.1	1.5
5	0.41	0.1	1.7
6	0.41	0.1	2.0

concentration of approximately 1.6 μM per bead (ie: 64 targets per box is approximately 102 μM, 96 targets $\approx$ 154 μM). Similarly, by assuming the size of a target bead radius to be approximately $a = 2.5$ nm and using Langmuir adsorption theory, we can convert the $\Delta E_0 = -4k_BT$ binding energy into a dissociation constant in Molar, resulting in a monovalent binding affinity of $K_D = 0.1$ mM. This monovalent binding affinity is well within the weakly binding range typical of lectins and sugars (K_Ds between mM to μM) as well as the affinity range of some monovalent protein-protein and RNA-protein interactions found in biocondensates [81–85].

Results are averaged over the second half of the total simulation time and over at least 10 different runs, with a typical system energy profile over time shown in the chapter appendix (5.2) Figure 5-13.

Results and Discussion

For all simulations we placed $n = 4$ polymers in a box with a degree of polymerization $N_P = 16$ because this is slightly above the length at which increasing polymer length leads to only a small increase in binding affinity for a perfectly flexible polymer chain binding to a divalent target [86]. The binding dependence on length for flexible polymers is primarily influenced by the loop sizes that can favorably occur when bonded twice to a target. Thus, we expect that in wormlike chains, intra-chain loops will be even shorter, and so overall polymer length will be relatively unimportant. Simulations for both the dilute case of a single target binding to our polymers as well as higher concentration cases where many targets interact with the polymer simultaneously are discussed in

this work.

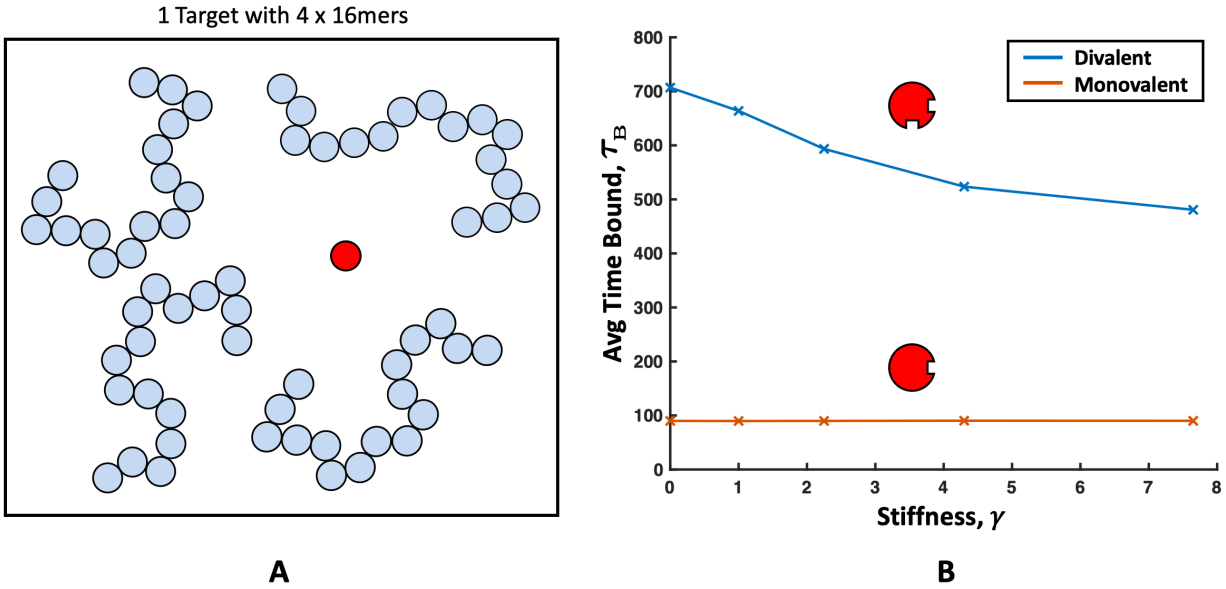


Figure 5-3: (A) Schematic of the single target simulation set up with a single mono- or divalent target shown in red and four polymers with a length of 16 beads. (B) The average time interval bound τ_B of a single divalent target (blue) and a monovalent target (orange) versus the polymer stiffness controlled by the angle-bending spring coefficient γ. Higher γ corresponds to stiffer springs and more rigid polymers. The monovalent target τ_B seems unaffected by the polymer chain stiffness while the divalent targets show a decrease in τ_B with γ. Error bars are smaller than symbol size.

Binding to a single target

It is important to consider how a single mono or multivalent target binds to a polymer without competition from other targets for available binding sites. We did this by placing a single target with either one or two binding sites in with four identical 16mer polymers as shown in Figure 5-3A. To examine the binding affinity of our polymers we calculated the average time bound τ_B for our target, where we considered our target bound whenever at least one of its binding sites was bound to the polymer. Accordingly, we consider the target unbound whenever none of its binding sites were bound to the polymer and denote the average time interval unbound as τ_{UB}.

We have plotted the τ_B for for divalent and monovalent targets in Figure 5-3B. From this plot, we can see that the τ_B for the monovalent target does not depend on polymer stiffness, but the τ_B of a divalent target decreases with increasing bending spring coefficient γ. Remember that higher γ

corresponds to a stiffer polymer. While this decrease was predicted by Zumbro *et al.*, the previous theory for random coil chains only partially applies [86]. Previous work showed that the entropic cost of forming loops limits the polymer binding affinity, but in the case of a rod-like polymer or worm-like chain, loop entropy is not the limiting factor. Instead, we predict that the major energetic factor limiting multivalent binding affinity is the enthalpic cost of bending the polymer to make two contacts with the target. If bending is the limiting factor, we can estimate the energy as:

$$E = E^0_{\text{bind}} + E_{\text{bend}} \tag{5.2}$$

where E^0_{bind} is a constant denoting the favorable energy of forming two bonds while E_{bend} is the quantity that is changing most drastically with loop length. When the target is bound twice to the same polymer chain, once to polymer bead i and simultaneously once to polymer bead j, we define the loop length by subtracting the two values $l_{\text{loop}} = |i - j|$. This results in $l_{\text{loop}} = 1$ when i and j are right next to each other, $l_{\text{loop}} = 2$ when there is a single unbound bead between them, etc. To minimize bending penalty for a particular l_{loop}, the polymer will want to minimize curvature or maximize the bending radius R_{loop}. If we consider very small loops where $l_{\text{loop}} = 1, 2, 3$, the maximum R_{loop} is constant at $2a$ where a is the radius of the beads. Because our target is very coarse-grained, we allow its binding sites to be accessible anywhere on the surface; this makes the largest R_{loop} for $l_{\text{loop}} = 1, 2, 3$ occur when the polymer beads all exactly touch the surface of the target. Since at small l_{loop}, R_{loop} is constant, we can estimate the probability of forming a loop as:

$$\ln(P_{\text{loop}}) \propto -E = -E^0_{\text{bind}} - C_0 \gamma R_{\text{loop}} l_{\text{loop}} = C_1 - C_2 \gamma l_{\text{loop}} \tag{5.3}$$

where C_0 is a constant representing the bending cross-section, and where, for small loops, we have replaced $-E^0_{\text{bind}}$ and $C_0 R_{\text{loop}}$ with generic constants C_1 and C_2 respectively. From Equation 5.3, we find that if bending energy is the most significant influence on τ_{B}, we should see the ln of the frequency of increasing loop size decay linearly with a rate proportional to the chain stiffness γ.

We can probe this theory directly by examining the length of loops formed and their frequency depicted in Figure 5-4. From this plot, we can see that the more flexible polymers ($\gamma = 0,\ 1,$ or 2.25) follow an exponential decay characteristic of entropic loop costs [86], while the stiffer chains ($\gamma =$

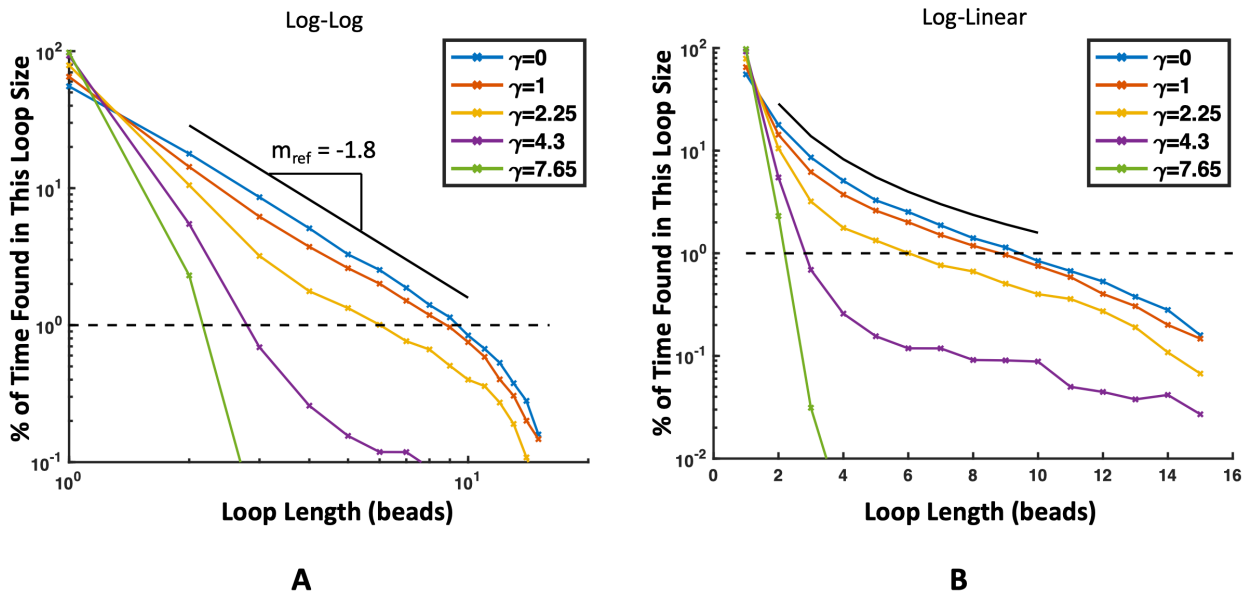


Figure 5-4: Percent of time that a polymer bound twice to a target is in a certain loop length plotted in (A) log-log scale and (B) log-linear scale. Each color represents a different polymer stiffness, the dashed black line represents 1%, and the solid black line is an example of $y = x^{-1.8}$. The frequency of long loops decreases as polymer stiffness increases. (A) More flexible chains ($\gamma = 0, 1$) have a power law decay in loop size due to the entropic cost of forming loops[86]. This manifests as a straight line in the log-log scale. (B) Stiffer chains ($\gamma = 4.3, 7.65$) have an exponential decay in loop lengths for short loops due to the energetic cost of bending. We can see this manifest in the log-linear plot as a straight line for short loop lengths ($l_{\text{loop}} = 1, 2, 3$). Lines are for aiding the eye and are not a theoretical fit.

4.3 or 7.65) follow a more linear decay in loop length characteristic of an enthalpic bending loop cost. This cost increases as C_∞ increases, leading to a drastic drop off in loop lengths with only loops of length $l_{\text{loop}} \leq 2$ being formed more than 1% of the time for $\gamma = 4.3$ and 7.65. Since only small loops are formed for highly stiff polymers, we can test our theory on them by measuring their slope.

Fitted values for C_1 and C_2 from Equation 5.3 are shown in Table 5.2 with corresponding bending theory lines shown with simulation data in Figure 5-5A. For the values shown in Table 5.2, we have chosen to show the Y-intercept at $l_{\text{loop}} = 1$ because this corresponds to the case of 0 angular springs between bound polymer beads i and j and because a loop length of 0 is nonsensical in this context. From these best fit lines, we find an excellent fit for such a simple theory for the stiffest polymer ($\gamma = 7.65$) with worsening fit as γ decreases, likely due to increasing contributions from entropic loop costs dominating flexible polymers.

Unfortunately, loops longer than 3 beads were almost never formed for stiff chains where $\gamma = 7.65$ and 4.3, and so we were unable to fit lines to more than 3 loop sizes. Therefore, to sample longer loop lengths for the stiff polymers, we also compared a case where the polymer and target beads had mismatched sizes. Note that simulation data for mismatched sizes is only used in this section on polymer loops. To do this, we doubled the target bead radius to $a_{\mathrm{T}} = 1.0$ and halved the polymer bead radius to $a_{\mathrm{P}} = 0.25$. Smaller polymer beads binding to a larger target allows for longer possible l_{loop}s while maintaining the same constant bending radius R_{loop} because there is more surface area available on the target, and each polymer bead has less excluded volume. These mismatched size simulations resulted in loops up to l_{loop} =5 beads for $\gamma = 7.65$. We also included results for a very rigid polymer at $\gamma = 12.0$ which show that even at these mismatched bead sizes, loops longer than 3 beads are not achievable in an extremely rigid polymer. Simulation results, theoretical linear fits, and fit variables are shown in Figure 5-5B and Table 5.3. Note that again, the Y-intercepts shown in Table 5.3 are also calculated at $l_{\mathrm{loop}} = 1$. Similarly to the matched-size case, in this new mismatched-size case we continue to see a good linear fit for small loops based on fitting to the shortest 5 loop frequencies. Again, stiffer chains with $\gamma = 7.65$ and 4.3 see a better goodness of fit as measured by the high R-squared value of 0.99. Using just the first five loop sizes, the more flexible polymer $\gamma = 2.25$ has a worse fit than the stiffer chains, and the fit would significantly worsen if fit on all loop sizes that occur more than 1% of the time (up to $l_{\mathrm{loop}} = 8$).

As predicted, the value of C_1, which represents the binding energy unaffected by chain stiffness is almost constant for the three highest stiffness chains. The small differences in C_1 are likely due to some dependence of E^0_{bind} on γ which our model does not capture such as increased stress on target-polymer bonds in stiffer chains, and unbound polymer beads in the center of the loop pressing in toward the target and into its excluded volume. Both of these secondary effects are energetically unfavorable and would lead to an increase in E^0_{bind} with γ, which aligns well with our calculated values of C_1.

As for the slope coefficient C_2, the behavior of the chain with $\gamma = 2.25$ is the least well-captured. This is likely because the chain is still relatively flexible and experiencing a blend of the bending costs of wormlike chains and the entropic cost of freely jointed chains. In contrast, we found good agreement for the two stiffest chains ($\gamma = 4.3, 7.65$) with our estimate that the decay rate of loop

length should be constant C_2 times the chain stiffness γ. The fit C_2 values for the two stiffest chains are within 10% of each other. Thus, with this very simple theory of bending, we capture quite well the behavior of stiff chains.

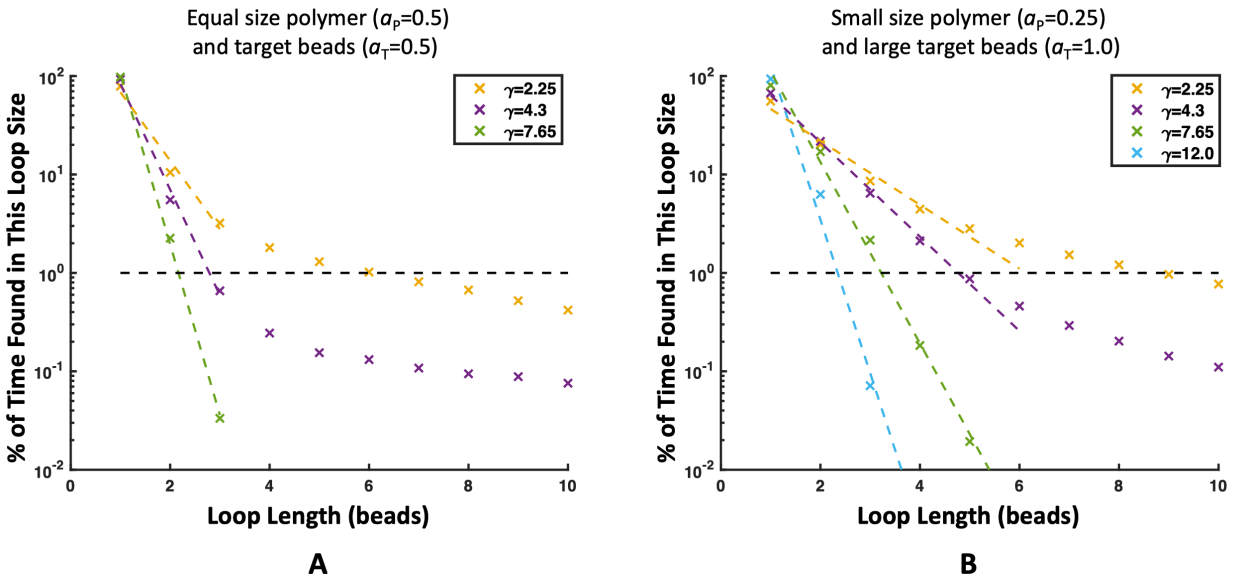


Figure 5-5: Frequency of loop sizes in Log-Linear scaling. 'x's denote simulation data and dashed (- - -) lines represent the best linear fit following equation 5.3 with values of C_1 and C_2 listed in Table 5.2 and 5.3. (A) Loop data and linear fits for matched size polymer and target beads ($a_P = 0.5$, $a_T = 0.5$). (B) Loop data and linear fits for mismatched bead sizes with smaller polymer beads and larger target bead ($a_P = 0.25$, $a_T = 1.0$) in order to sample longer loops in the stiff chains.

Table 5.2: Slopes and intercepts for lines fitted to loop lengths for Eq. 5.3 and plotted in Figure 5-5A - matched bead sizes.

γ	Y-intercept (C_1)	Slope constant (C_2)	Goodness of fit (R^2)	# of points used in fit
2.25	4.23	-0.71	0.978	3
4.30	4.41	-0.57	0.993	3
7.65	4.66	-0.52	0.999	3

Table 5.3: Slopes and intercepts for lines fitted to loop lengths for Eq. 5.3 and plotted in Figure 5-5B - mismatched bead sizes.

γ	Y-intercept (C_1)	Slope constant (C_2)	Goodness of fit (R^2)	# of points used in fit
2.25	3.83	-0.33	0.978	5
4.30	4.16	-0.26	0.997	5
7.65	4.71	-0.28	0.994	5
12.0	4.84	-0.30	0.980	3

As a result, we conclude that in the case of a much larger polymer binding to a single multivalent target, stiff polymers are limited by the enthalpic cost of bending when forming an intra-polymer loop. Because long loops cost high amounts of bending energy, they are almost impossible to form, resulting in fewer possible binding arrangements for the target and overall decreasing the τ_{B} of the rigid polymer over a flexible polymer. We can extend our results from τ_{B} to relative overall binding affinity by also measuring the average time unbound τ_{UB} and calculating the relative dissociation constant as $K_{\mathrm{D}} = \frac{\tau_{\mathrm{UB}}}{\tau_{\mathrm{B}}}$ with measured τ_{UB} and K_{D} shown in Figure 5-6A and B respectively.

The τ_{UB} is a combination of the time it takes for a target to come within reach of a free polymer binding site, which is dependent on the shape of the polymer, multiplied with the binding attempt rate and success rate, which are constant for our simulations. By approximating the flexible polymer as a sphere and the rigid polymer as a thin cylinder, we found that the diffusive flux of targets toward the sphere is smaller than a cylinder for our polymer concentration. We can think of this as the more rigid polymers having binding sites more uniformly distributed throughout the volume. This means that on average, targets take longer to find flexible polymers than rigid ones and the τ_{UB} shortens with γ. Diluting the polymer concentration can flip the relationship between the target flux toward a cylinder and sphere so that $\tau_{UB}^{\mathrm{cyl}} > \tau_{UB}^{\mathrm{sp}}$, but this will only exaggerate the effect of stiffness on K_{D} plotted in Figure 5-6B. A detailed discussion of the flux calculation are presented in the chapter appendix in Section 5.2.2, Figure 5-14.

While τ_{UB} decreases with increasing chain stiffness shown in Figure 5-6A, the decrease is not enough to overcome the decrease in τ_{B} with chain stiffness for divalent targets. This results in a

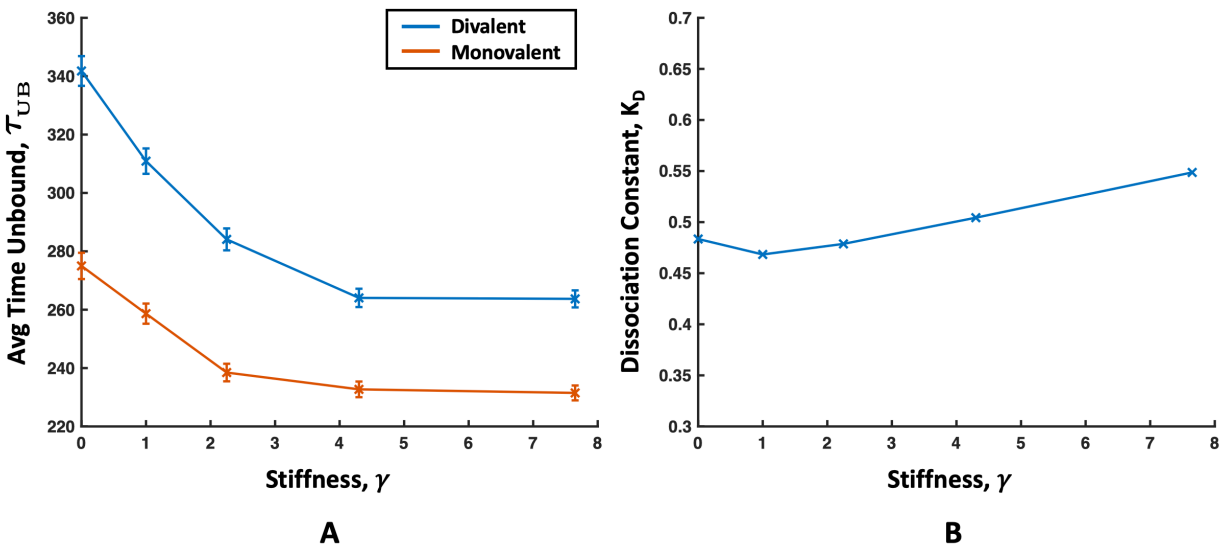


Figure 5-6: (A) The average time interval unbound τ_{UB} for a single mono- (orange) or divalent (blue) target binding to a polymer. The τ_{UB} decreases similarly for both target valencies because it is dependent on the distribution of polymer binding sites throughout the simulation volume. Standard error is denoted by error bars. (B) Dissociation constant K_{D} for a divalent target versus polymer stiffness. The longer τ_{UB} is not enough to overcome the longer τ_{B} for flexible polymers and flexible polymers show a lower K_{D} (higher affinity) than rigid ones. A plot of the K_{D} for a monovalent target is dominated by τ_{UB} and is shown in Figure 5-16.

larger dissociation constant K_{D} (lower binding affinity) for stiffer chains than flexible polymers. The effect is not monotonic, with a slight decrease in the K_{D} as γ increases from 0 to 1, which is probably due to increased free volume and therefore more accessible binding sites as seen in an experiment with polymers binding to lectins [95]. But as we move away from slightly extended chains that still follow a random walk toward highly extended chains with end-to-end distances much greater than $\sqrt{N_{\mathrm{P}}}$, there is a distinct upward trend in K_{D}. Because smaller K_{D} corresponds to a stickier polymer, we find that the affinity of the polymer generally decreases with increasing C_∞, and conclude that this result is primarily affected by the transition of the binding regime from a flexible chain, where the entropic cost of loop formation dominates, to the regime of a wormlike chain where the high enthalpic cost of bending to create divalent loops dominates and reduces binding affinity.

Here, we showed simulations for targets with one or two binding sites, but these results also have implications for globular targets of higher valencies. For example, with cholera toxin which is pentavalent or concanavalin A (conA) which is tetravalent, following directly from our results, we expect that anything that discourages loops will even more drastically decrease the binding affinity

to higher-valent targets. This is because binding to all four binding sites on conA necessarily requires three polymers loops (or more than one polymer to bind, which costs additional entropy of constraining an second polymer chain). Therefore, we would expect the dependence on stiffness to be magnified in higher-valency binding resulting in flexible polymers binding with much higher affinity for tetra and pentavalent small targets than stiff polymers. Clearly, our analytical model should serve as a starting point for future work on systems of higher valency involving stiff polymers.

Our results do not contradict previous research on precise ligand engineering where it was found that for matched sizes on the size scale of the target, stiff linkers have higher affinity [33,115,90, 32,97]. Conversely, our research is complementary to previous works; our result of large flexible polymers showing higher binding affinity than large rigid polymers suggests that the most sticky polymer might consist of flexible regions between the stiff, matched-size binding sites detailed in the aforementioned studies.

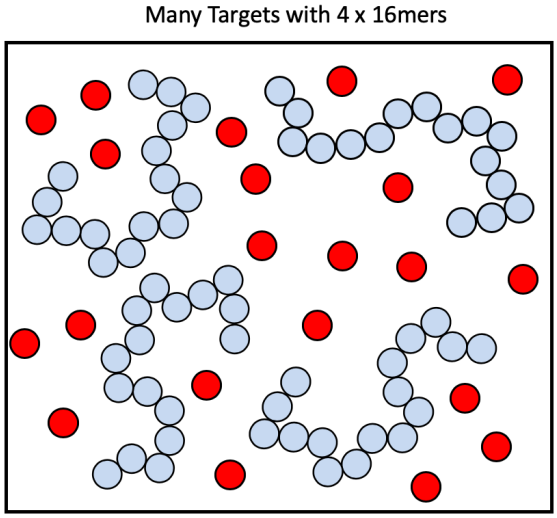


Figure 5-7: Schematic of simulations with multiple targets. In this case, 32, 64, 96, or 128 targets are placed in a box with four 16mer polymers to examine how target-target interactions and competition between targets for binding sites on the polymer can change the phase behavior of the system.

Binding to multiple targets

In this section, we consider the binding of many targets to polymers with varying backbone flexibility as shown in Figure 5-7. In this case, we place 32, 64, 96, or 128 mono or divalent targets in with our four 16mer polymers to create competition, explore the effect of varying stoichiometry, and examine

the conditions under which a condensed phase might nucleate with the help of the polymer. Past studies have shown that modulating stoichiometry and the mismatch between receptor and ligand concentrations can lead to switchlike control of the composition and function of biocondensates, so we wanted to explore how polymer stiffness can change this phase boundary [45]. At ≥ 64 target concentration, target binding sites exceed polymer binding sites for divalent targets, and there is significant competition. We can consider the concentration of targets in real units by assuming a target diameter. For example, assuming a target protein diameter of 5 nm, 64 targets corresponds to approximately 100 μM and assuming a weight of 70 kDa, approximately 7 mg/ml.

Because we have multiple targets, we now need to consider the interactions between targets. To capture these non-specific interactions that control protein solubility limits we added a generic Lennard-Jones potential between the targets themselves as described in Eq. 2.3. We used several values of potential well energies ranging from low attraction ($\epsilon_{\mathrm{TT}} = 1.0$) to moderate levels of attraction ($\epsilon_{\mathrm{TT}} = 2.0$) to explore the wide range of solubilities found in proteins. Parameters for the cases studied are summarized in Table 3.1 Cases 2-6. By themselves, up to 64 targets are soluble throughout this entire range of intra-target interaction strengths, and only condense on their own at approximately $\epsilon_{\mathrm{TT}} = 2.5 k_{\mathrm{B}}T$ Lennard-Jones attraction shown in a rough phase diagram in Figure 5-10A. 96 and 128 target concentrations can phase separate on their own at $\geq 2.0\ k_{\mathrm{B}}T$. Areas of the phase diagrams where target proteins can phase separate alone, without the addition of polymer are shaded with a purple background in Figures 5-8B and 5-9.

Previous work has shown that increasing multivalent polymer length can induce a condensed phase, so we wanted to further investigate how changing the polymer backbone stiffness changes the phase boundary [86]. In doing so, we hope to provide insights for research on liquid-liquid phase separation as well as those targeting inhibition of high concentrations of multivalent toxins. To this end, we used a combination of visual inspection, the Binder cumulant, and a collapse in the system radius of gyration R_{g} to look for a persistent condensed phase in our simulations. We first used visual inspection to look for the proportion of 10 runs in which a condensed droplet of targets and polymers persisted for the last half of the simulation time. Simulations in which a stable droplet formed more than 70% of the time are marked with a green “Y", systems that formed a droplet in 60% of runs are marked with a yellow “Y", and systems where less than 50% of runs formed

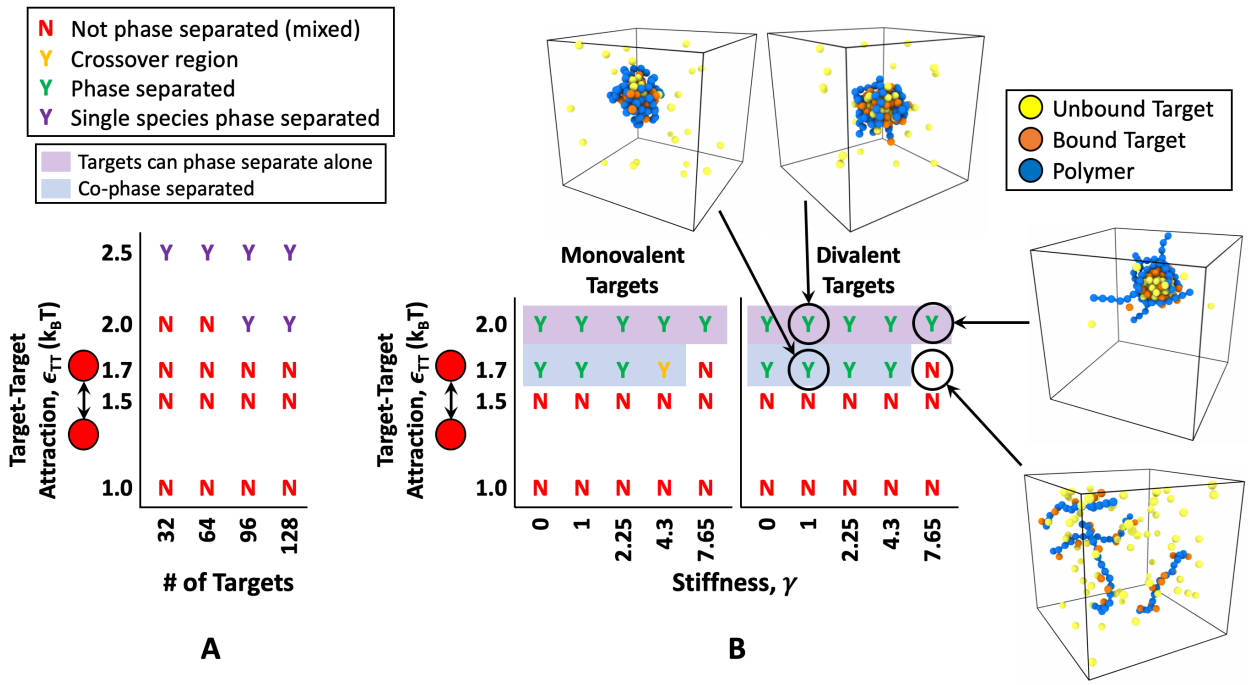


Figure 5-8: Phase diagrams of (A) targets only with increasing concentration of targets on one axis and increasing target-target Lennard-Jones attraction on the other and (B) simulations of 96 targets mixed with four 16mer polymer. To highlight the change in phase separation with stiffness, the polymer stiffness on one axis and target-target Lennard-Jones attraction on the other. Results are shown for both mono and divalent targets. Not phase separated or "mixed" systems are denoted by a red letter "N" for "no", a phase separated system where the polymer and targets are both components of the condensed phase is denoted by a green "Y" for "yes", and a purple "Y" denotes a system where the targets phase separated by themselves, in this case because no polymer was added. Regions where targets can phase separate by themselves, without the help of the polymer are shaded with a purple background. Systems where phase separation only occurs through interaction between polymers and targets we call "co-phase separation" and is shaded with a blue background.

a droplet are marked with a red "N" for no phase separation. Visual inspection is very similar to measuring density inhomogeneities used in previous work by Choi et al. because our eyes are excellent at capturing such changes [125].

We quantitatively confirmed these initial phase diagram determinations by calculating the average energy of the last half of run time and using it to compute the Binder cumulant $\frac{<E^4>}{3<E^2>^2}$. This quantity compares the ratio of the energy variance, which is equivalent to the specific heat of the system, to the average energy, and shows a maximum at the phase transition [126]. Plots of the Binder cumulant for simulated cases are provided in the chapter appendix Figures 5-19-5-27. By comparing the Binder cumulant along lines of constant target concentration, we confirmed our

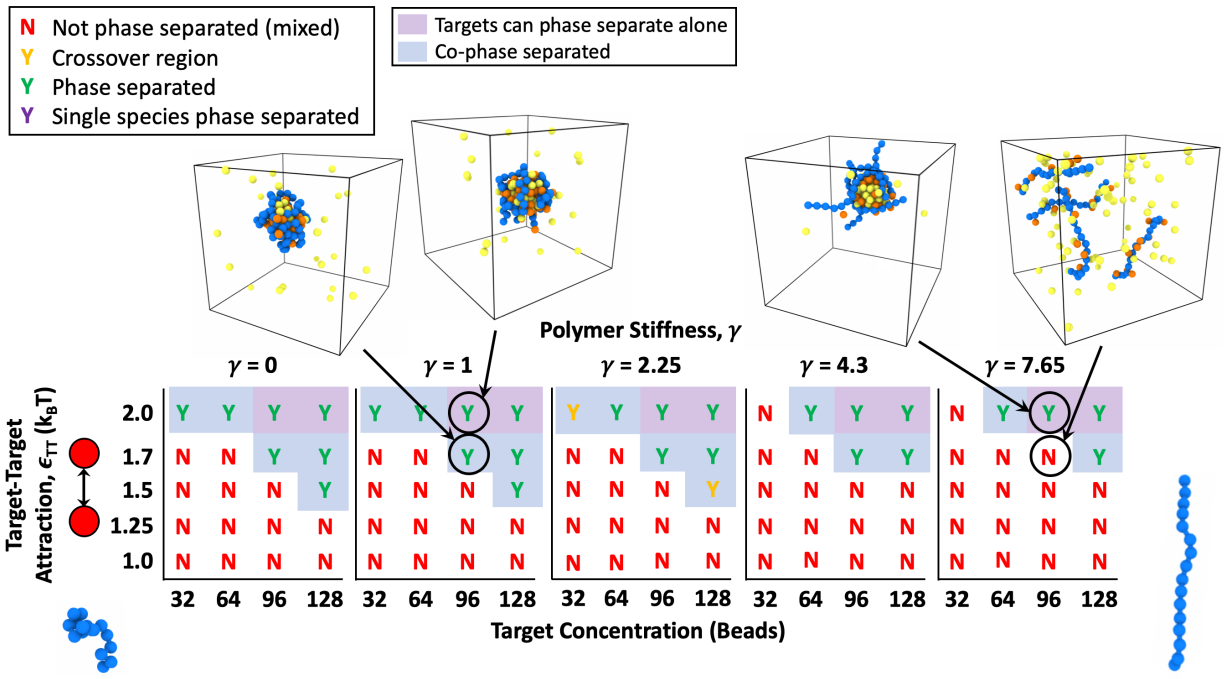


Figure 5-9: Phase diagrams of divalent targets mixed with four 16mer polymers with increasing concentration of targets on one axis and increasing target-target Lennard-Jones attraction on the other. Phase diagrams are shown for five polymer flexibilities. Phase separation occurs at lower energies and target concentrations for flexible polymers than stiff polymers. Not phase separated or "mixed" systems are denoted by a red letter "N" for "no", a phase separated system where the polymer and targets are both components of the condensed phase is denoted by a green "Y" for "yes", and a purple "Y" denotes a system where the targets phase separated by themselves, without polymers. Regions where targets can phase separate by themselves, without the help of the polymer are shaded with a purple background. Systems where phase separation only occurs through interaction between polymers and targets we call "co-phase separation" and is shaded with a blue background.

initial phase diagrams created through visual inspection.

In addition, we further corroborated our evidence of phase transitions by calculating the radius of gyration R_g of the polymers individually and the R_g of the complete polymer system to capture when the polymers showed collapse and aggregation, respectively. Methods of measuring aggregation through R_g were used in previous computational work on phase separation of biocondensates[58]. When a polymer system phase separates the system should see a collapse in the system R_g as the polymers come together, followed by a swelling of they polymer system and individual polymers as the droplet swells with targets. We considered the system phase separated when the system and individual polymer R_gs were at a minimum or showed a significant decrease before swelling. This

aligned well with our previous phase separation assessment using the Binder cumulant and visual inspection. R_gs and Binder cumulants for each system are plotted in the chapter appendix Figures 5-19-5-27. The results of this inspection were compiled into a phase diagram in Figure 5-8B and Figure 5-9 with renderings of specific cases provided for reference. Areas where the Binder cumulant and R_g predict phase separation are shaded with a blue or purple background in Figure 5-8B and 5-9.

Using our resultant phase diagrams, we find that, first, the addition of more flexible polymers ($\gamma \leq 4.3$) lowers the phase boundary for 96 targets below $\epsilon_{TT} = 1.7$ for both monovalent and divalent targets. The stiffest polymers at $\gamma = 7.65$ do not appear to change the phase boundary. In this case, there is only a slight increase in phase separation as the targets valency is increased from mono to divalent. This can be seen by the shift in the phase boundary in the $\epsilon_{TT} = 1.7$ row from right at $\gamma = 4.3$ (denoted by a yellow "Y") in the monovalent system where only some systems formed stable droplets to a fully phase separated system for divalent targets at $\epsilon_{TT} = 1.7$ (denoted by a green "Y") in Figure 5-8B. Although this shift is slight, it is consistent with previous research showing that increasing the valency of the target lowers the phase boundary in liquid-liquid phase separation of multivalent polymers [83,58,50]. We will return to the combined effects of stiffness and valency later on in this paper.

For now, looking at the renderings of simulations with divalent targets, it is clear that even though all divalent systems are phase separated at $\epsilon_{TT} = 2.0$ attraction, the resulting complexes look very different depending on the chain stiffness. For flexible polymers $\gamma = 1$, the resulting complex is spherical, characteristic of a liquid globule, and the polymers are coating the surface relatively tightly. In the case of the stiffest polymer $\gamma = 7.65$, we still see a rounded globule of targets, but now the polymer is not completely stuck to the condensed target surface. Instead, stiff polymers have peeled off the globule and are sticking out making the complex look spiky or hairy. Because these stiff chain ends are sticking away from the condensed target phase, we expect that they are less bound, making the K_D larger (lower affinity) for the targets and lowering the number of total sites occupied on the polymer. These effects lower the binding efficiency of the stiff polymers and, in the case of multivalent targets, result in the phase boundary being pushed to higher target-target attractions as the polymer stiffens. This appears as the solubility limit of our divalent

targets being pushed to higher energies as polymer stiffness increases, from below $\epsilon_{TT} = 1.7$ to $\epsilon_{TT} = 2.0$ in Figure 5-8B. We have also plotted the phase separation on a more traditional diagram for divalent targets at four different concentrations in Figure 5-9. In this figure, the blue shaded region shrinks and the phase boundary clearly moves to higher target concentrations and higher target-target attraction as the polymers stiffen. This phenomenon has also been seen in complex coacervates where it was shown that the two phase region shrinks as the stiffness of the binding polycation/polyanion species increases [127].

Here, we would like to note that while some polymer systems, such as chromatin folding, undergo Polymer-Polymer Phase Separation (PPPS), where the targets act as linkers to condense the polymers, we believe we are seeing Liquid-Liquid Phase Separation (LLPS) for several reasons. First, for PPPS to occur, the divalent binders should act as bridges that collapse the polymer without attraction between divalent bridges [128]. By measuring the bonds between two chains (inter-bonds), we found that bridges are more frequent in phase separated systems and in systems with stiffer polymers, but even then only constitute less than a third of all divalent bonds as shown in Figure 5-17. Also, stiffer polymers which encourage more divalent bridges are more difficult to phase separate, pointing to divalent bridges not being the cause of phase separation. Instead, our system only forms droplets when targets are attracted to themselves, with phase separation occurring only above $\epsilon_{TT} = 1.5$ for the most flexible polymers at the highest target concentration. Additionally, increasing the concentrations of binding proteins, where target receptors far outnumber the polymer ligands, results in droplet growth and more instances of droplet formation instead of dissolution. If the multivalent targets were acting primarily as bridges, higher protein concentrations would mean that more polymer sites are occupied monovalently, dissolving divalent bridges, and discouraging phase separation, as seen in biocondensates and synthetic polymer gels undergoing PPPS [128,129]. Therefore, we believe our system to be undergoing LLPS. If we increased the binding affinity of our reactive sites, it might be possible for the bridges to become more permanent and exhibit more elements of PPPS, which could be explored in a later study.

To further explore the interactions between targets and polymers in our system, in Figure 5-10A, we have plotted the dissociation constant for simulations of 96 divalent targets as $K_D = \frac{\tau_{UB}}{\tau_B}$ vs γ for several values of ϵ_{TT} with the values of τ_B, τ_{UB}, and monovalent K_Ds plotted in the

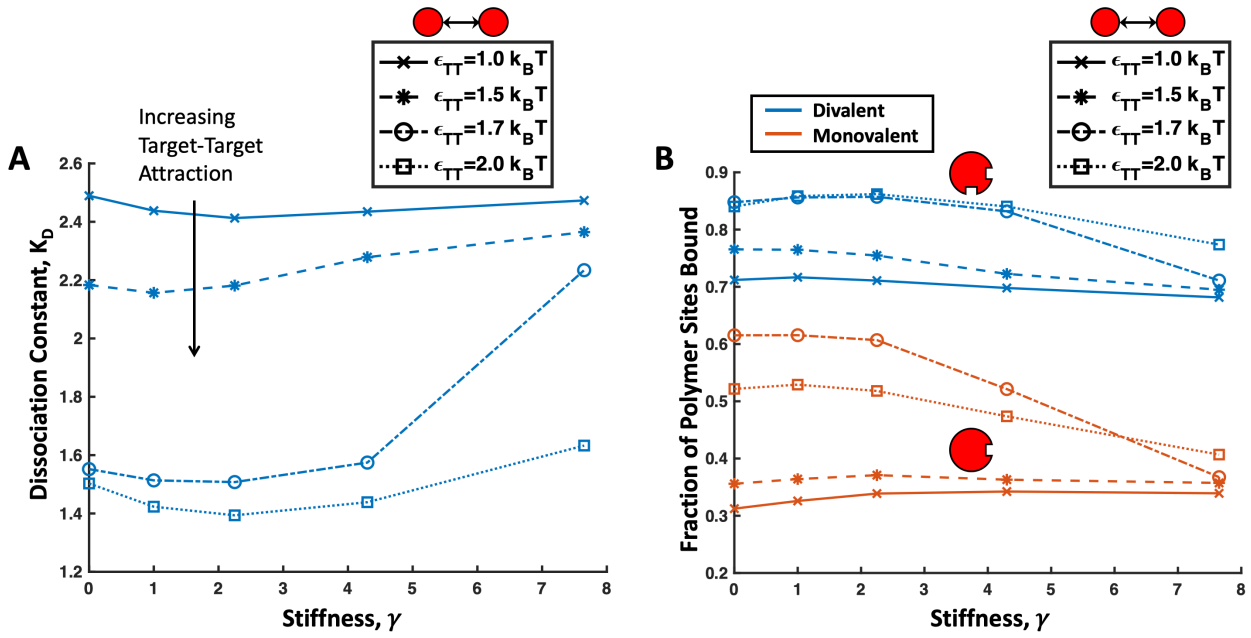


Figure 5-10: (A) K_D for 96 divalent targets binding to four 16mer polymers. As target-target attraction ϵ_{TT} increases, K_D decreases. For $\epsilon_{TT} = 1.0$, binding affinity is dominated by the increased τ_{UB} and flexible polymers are slightly lower affinity than stiff ones. At $\epsilon_{TT} \geq 1.5$, τ_B dominates and flexible polymers have higher affinity than stiff polymers. We can see a sharp increase in K_D for $\epsilon_{TT} = 1.7$ as γ increases from 4.3 to 7.65 signaling the phase boundary where flexible polymers are able to nucleate a condensed target phase but stiff polymers are not. (B) Binding efficiency of polymers calculated as the average fraction of sites on the polymer bound versus γ. This plot closely mimics the one for K_D, with a sharp decrease in binding efficiency for divalent targets at $\epsilon_{TT} = 1.7$ denoting the phase transition between $\gamma = 4.3$ and 7.65. For phase separated systems at $\epsilon_{TT} = 2.0$, there is an approximately 10% decrease in sites bound on the polymer between the $\gamma = 2.25$ and $\gamma = 7.65$ for both target valencies. This is due to rigid polymer resistance to bending and their tails sticking out away from the condensed phase as shown in Figure 5-8B. Error bars are smaller than symbol size.

chapter appendix Figure 5-15 and 5-16. As expected from the earlier single target case and our phase diagram, intra-target attraction and chain stiffness have a significant effect on the binding of divalent targets. As target-target attraction increases, the K_D decreases for all polymer stiffnesses. This is because when many targets are bound to the polymer, bound targets benefit energetically from being near other neighbors also bound to the polymer chain. This additional energy benefit makes the polymer appear stickier and is magnified when the polymer can nucleate a condensed target phase because then bound targets can gain the energy benefit of being near both bound neighbors and unbound neighbors. This can be seen in Figure 5-10A by looking at the values of K_D for intra-target attraction $\epsilon_{TT} = 1.7$. More flexible polymers $\gamma = 0, 1, 2.25, 4.3$ have a low

and almost constant K_D because the polymers nucleated a condensed target phase and so their binding affinity is benefiting greatly from the favorable energy of clustered unbound targets. As γ increases from 4.3 to 7.65 we see a sharp increase in the K_D, due to the fact that phase separation is harder to induce with a stiff polymer. Subsequently, targets that bind to a rigid polymer will have fewer target neighbors and benefit less from favorable target-target energies, making the polymer effectively lower affinity.

We can confirm this reasoning by directly examining the clustering of unbound targets near the polymers. We plot the radial distribution function (RDF) of unbound targets for simulations with 96 divalent targets in Figure 5-11. While general curve shapes for $\epsilon_{TT} = 1.0, 1.5$, and 2.0 show only small changes due to polymer stiffness, the profile of unbound targets near the phase boundary $\epsilon_{TT} = 1.7$ is strongly dependent on polymer stiffness. From the RDFs of unbound targets at $\epsilon_{TT} = 1.7$ (Figure 5-11B), we can see that the flexible polymers are able to stabilize unbound targets at this target solubility while stiff polymers at $\gamma = 7.65$ are not. In order for stiff polymers to be able to form a stable cluster of unbound targets, the target-target attraction must be increased to $\epsilon_{TT} = 2.0$. Even when phase separation occurs at $\epsilon_{TT} = 2.0$ in Figure 5-11C, the stiffer polymers stabilize fewer unbound targets shown by the lower RDF, making them lower affinity than their flexible counterparts.

We can also look at the polymer binding efficiency shown in Figure 5-10B to understand why stiff polymers phase separate at higher target-target attractions than flexible polymers. Here, we can see that for $\epsilon_{TT} = 2.0$, where all polymers phase separate, the percent of polymer sites occupied drops more than 10% as γ goes from 0 to 7.65. This is likely due to the hairy or spiky ends shown in the rendering of stiff chains in Figure 5-8B. These stiff chain ends that extend away from the condensed phase rarely get to interact with the target globule since they would have to bend in to access them, and bending for stiff chains is energetically costly. This results in the parts of the chain extending away from the globule being relatively unbound and decreasing the binding efficiency of the whole polymer. Since the rigid ends are extended out into the dilute target phase, they interact with targets more like our dilute target case discussed above, where they are already lower affinity than flexible chains due to the costs of bending associated with polymer loops. Therefore, resistance to bending makes stiff polymers bind less efficiently, making it harder to collect the critical

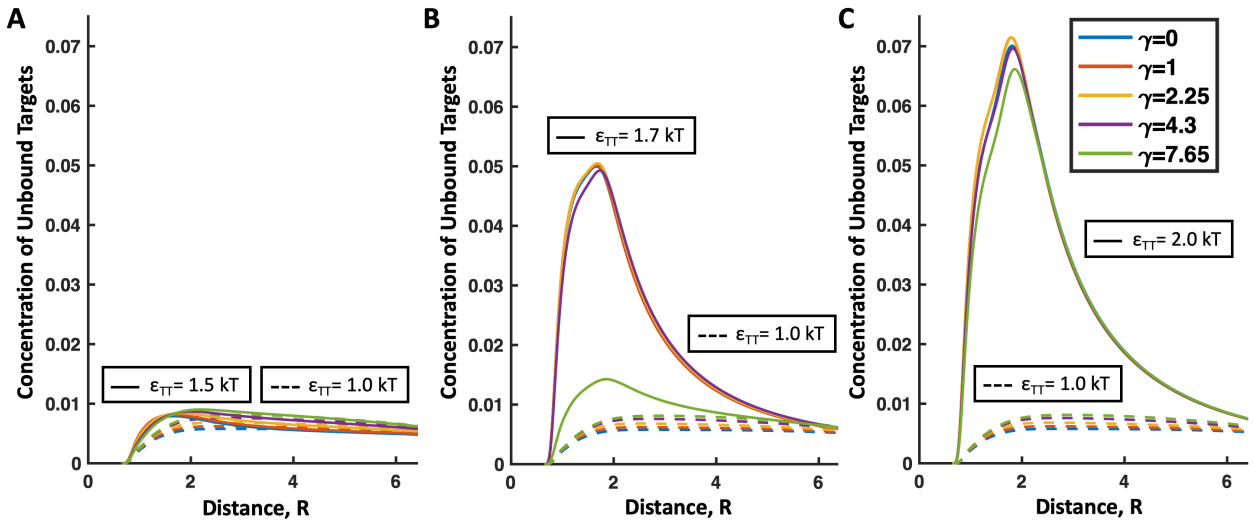


Figure 5-11: Radial distribution function (RDF) for concentration of unbound targets found near the polymer chain where the x-axis R is the distance from a target center of the closest polymer bead. Data is for 96 divalent target simulations. In all plots, the dashed line represents low target-target interaction $\epsilon_{\mathrm{TT}} = 1.0$. (A) The solid lines represent the RDF of targets for $\epsilon_{\mathrm{TT}} = 1.5$. Only small changes in the RDF occur with stiffness. (B) The solid lines represent the RDF of targets for $\epsilon_{\mathrm{TT}} = 1.7$. Note that the solid blue, red, and yellow ($\gamma = 0, 1, 2.25$) lines overlap. Here, flexible polymers show a much higher concentration of unbound targets near the chain because they are able to induce phase separation at this target-target potential. (C) The solid lines represent the RDF of targets for $\epsilon_{\mathrm{TT}} = 2.0$. Note that the blue, red, yellow, and purple lines overlap ($\gamma = 0, 1, 2.25, 4.3$). All polymers cause phase separation at this ϵ_{TT}, so all flexibilities show increased concentration of unbound targets near the polymer.

concentration of bound targets needed to stabilize and collect unbound targets. This explains the change in the phase boundary to higher ϵ_{TT} and target concentration as stiffness γ increases.

In general, we have found that increasing polymer stiffness can be thought as if one was lowering the polymer's effective valency and overall avidity because it strongly discourages polymer loops, thereby reducing the possible binding configurations of the polymer. Previous research on LLPS has shown that lower avidity and valency make it more difficult for systems to phase separate [83,58,60,130]. Thinking of increasing polymer stiffness as reducing valency or affinity ties it nicely into previous research on the LLPS of multivalent polymers.

Last, we have included an abbreviated set of phase diagrams for mono and trivalent targets to provide a better sense of how the effect of stiffness changes with target valency in Figure 5-12. Divalent targets may also be influenced by the "magic number" effect in multivalent systems. This effect is seen when the valency of one species is an interval multiple of another and makes

phase separation is more difficult [131,132]. Trivalent targets are example of a system that cannot experience this “magic number" effect. We find that trivalent targets binding to flexible polymers with $\gamma = 0$ show more phase separation than lower valencies, which is consistent with both magic number research and research showing that higher valencies lead to lower energy phase boundaries [131,132,83,58,60,130]. While it is possible that our divalent and monovalent targets are experiencing some magic number effects, we see the same dependence of phase separation on stiffness with the trivalent targets as with other valencies. It is also interesting to see that the mono, di, and trivalent targets all phase separate under the same limited set of conditions when mixed with stiff polymers $\gamma = 7.65$. This means that phase separation of these higher valency targets experience an even more drastic on dependence polymer stiffness. This matches well with our earlier prediction in the dilute target section. In order for higher valency targets to be fully bonded (all binding sites occupied), it inherently creates more than one polymer loop. Therefore, anything that prevents loops will more drastically reduce the affinity of higher-valent binding interactions. Stiffer polymers discourage loop formation, and so the effect of stiffness is more drastic for the trivalent targets than the mono or divalent targets. The phase diagrams are the same across valencies for the stiffest polymer ($\gamma = 7.65$) because at this stiffness very few loops can occur and so di and trivalent targets become effectively monovalent.

In the context of native biocondensates, the results presented in this section are relevant for asymmetrically-valent species where the binding protein has small valency. Examples include RNA binding proteins FUS, hnRNP’s, and TDP-43, which have 1, 2, and 3 RNA recognition motifs (RRM) [130,117,133]. By thinking of these RRMs as the specific binding sites in our model and RNA as our multivalent polymer, our results might aid in understanding ribonucleic protein granules. This is just one example, but we expect that there are other biocondensates in which these results would be relevant. Phase separated droplets that contain DNA also might be especially impacted by changes in stiffness, since DNA can go through drastic changes in stiffness if it transitions from single to double stranded. There is evidence that membraneless organelles that contain DNA can selectively absorb single stranded DNA, the more flexible option [134], which aligns well with our simulation results showing that more flexible polymers phase separate at lower attractions.

Overall, we have found that the phase boundary between a gas-like and condensed target/poly-

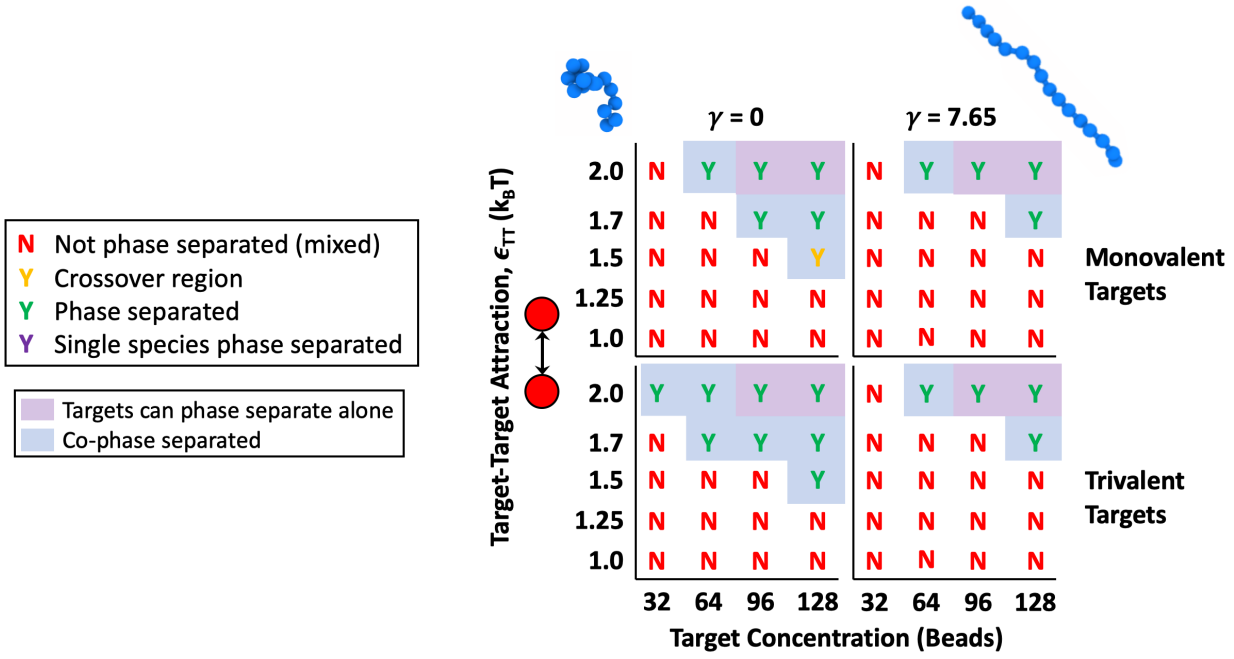


Figure 5-12: Phase diagrams of monovalent and trivalent targets mixed with four 16mer polymers with increasing concentration of targets on one axis and increasing target-target Lennard-Jones attraction on the other. Phase diagrams are shown for five polymer flexibilities. Phase separation occurs at lower energies and target concentrations for flexible polymers than stiff polymers. Not phase separated or "mixed" systems are denoted by a red letter "N" for "no", a phase separated system where the polymer and targets are both components of the condensed phase is denoted by a green "Y" for "yes", and a purple "Y" denotes a system where the targets phase separated by themselves, without polymers. Regions where targets can phase separate by themselves, without the help of the polymer are shaded with a purple background. Systems where phase separation only occurs through interaction between polymers and targets we call "co-phase separation" and is shaded with a blue background.

mer phase depends on the polymer's C_∞ or stiffness. Flexible random-coil polymers can lower the solubilities of target proteins more significantly than rigid wormlike polymers, and thus we expect that modulating polymer stiffness could play a role in controlling the phase separation of synthetic systems as well as biological liquid-liquid phase separation.

Conclusion

In this work, we have studied on how the flexibility of multivalent polymers influences binding to much smaller targets. When binding to dilute targets where there is little competition for binding sites, we have shown that there are two multivalent binding regimes. Flexible random coil polymers

fall into the first regime, where binding twice to the target is dominated by the loss of entropy of the the polymer loop as described earlier by Zumbro *et al.* [86]. We have shown here that stiff polymers fall into a second regime where binding affinity is dominated by the enthalpic cost of bending into a loop when binding divalently to a target. The high cost of bending makes stiff rod-like polymers have lower binding affinity for targets than random coil polymers. Therefore, combined with previous research showing that rigid molecules with precisely spaced binding sites have the highest affinity for targets of similar size, we expect that long polymers should ideally contain small rigid binding sections with flexible linkers connecting them into a larger chain in order to achieve the highest affinity [33,115,90,32,97].

Next, we extended our simulations to the case of many targets binding to the polymer at the same time. This adds competition between our targets as well as allowing us to consider the non-specific interactions between the targets themselves. We show that the presence of polymers can lower the solubility limit of the targets for both multivalent and monovalent binding. We showed that flexible polymers can nucleate a condensed phase at lower intra-target attractions than rigid polymers. When stiff polymer do phase separate, the shape of the resultant condensate is different for from flexible polymers; flexible polymers form a relatively smooth spherical droplet, conformally coating the condensed target surface but rigid polymer tails stick out away from the target droplet resulting in a hairy or spiky condensate. Because resistance to bending lowers the binding efficiency of the rigid polymers, it makes it more difficult for stiff polymers to nucleate a condensed phase. We also showed that stiffness has a stronger effect on higher-valent targets because they inherently require more polymer loops to be fully bonded and stiff polymer discourage loops. We believe that these results are relevant in biocondensates with asymmetrically-valent binding species such as between some proteins and RNA. Since more flexible polymers phase separate at lower energies, perhaps the drastic changes possible in DNA flexibility could also have implications on DNA-containing biocondensate formation. While more investigation is needed on this topic, a recent experimental study showed that more flexible DNA favors liquid-liquid phase separation [135].

We hope our results will aid in the design of new polymeric toxin inhibitors as well as help scientists better understand the formation of membraneless organelles and how changes in polymer stiffness can modify the phase boundary of biocondensates.

5.2 Appendix

5.2.1 System energy profile

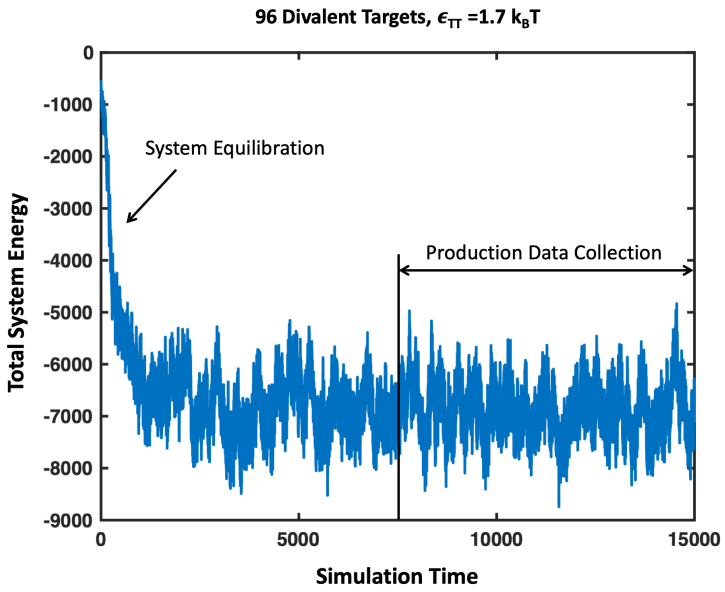


Figure 5-13: Example of a typical system energy profile over time. The total energy is shown for a system with four 16mer polymers and 96 divalent targets with a target-target attraction $\epsilon_{\mathrm{TT}} = 1.7k_{\mathrm{BT}}$. Simulation energy is shown every 10000 timesteps. There is an initial large drop in energy while the system equilibrates. Production research data is taken from the second half of the simulation, past this equilibration time period.

5.2.2 Time unbound

The average time interval unbound τ_{UB} is the inverse of the binding rate k_{on}. The τ_{UB} is controlled by the probability of binding when within reach of the polymer and the average time it takes to find a free polymer binding site. The rate of binding upon finding a free binding site is controlled by raising or lowering the energy barrier ΔE_{B}, which we fix at $0.5k_{\mathrm{B}}T$ for all simulations to get good sampling on binding events. We also fix concentration of targets and polymers in each set of simulations and the binding attempt rate, so in this case, the change in τ_{UB} with γ seen in Figure 6A is caused by changes in the polymer geometry and the resulting changes in how polymer sites are distributed throughout the volume. To understand the time it takes a target to find a binding site, we can look at the diffusive flux toward our polymers with different geometries. We can estimate

the binding surface of our flexible polymer as a sphere with a radius equal to the radius of gyration $R_g = 2a\sqrt{N_P/6}$ plus the equilibrium distance between the centers of two bound beads $2a$. We can estimate our stiff polymer as a rod with radius equal to the distance between two bound bead centers $2a$ and height $h = 2aN_P$. Where $a = 0.5$ and $N_P = 16$ from the main text. Then we apply the known steady state diffusion equations for a cylinder and a sphere from J. Crank [136], equations 5.4 and 6.7 respectively. We calculate the diffusive flux for targets toward the polymer as:

$$J = \oint D \frac{\mathrm{dC}}{\mathrm{dr}}\Big|_{r=s} \mathrm{d^2S} \tag{5.4}$$

where D is the diffusion constant, $C(r)$ is the concentration, r is the radial variable, s is the radius at which the binding surface lies, and S is the binding surface to integrate the partial flux over to get the full flux of targets toward the polymer. This results in the flux for a sphere J_{sp} and flux for a cylinder J_{cyl}:

$$J_{sp} = 4\pi D C_0 \frac{sb}{b-s} \tag{5.5}$$

$$J_{cyl} = \frac{2\pi h D C_0}{\ln(b/s)} \tag{5.6}$$

where C_0 is the concentration of targets in the bath or very far from the polymer and b is the outer radius of the system. Increasing b increases the volume of the polymer-target system, making the polymer species more dilute. Plotted in Figure 5-14 is the flux of a sphere and cylinder with the appropriate surface radii for our systems versus the system volume where volume for a sphere is $4/3\pi b^3$ and for a cylinder $\pi b^2 h$. At our simulation volume per polymer, marked with a black line, we can see that $J_{cyl} > J_{sp}$ meaning that the time for a target to diffuse to a polymer is shorter for the stiff rod-like polymer than the flexible globular polymer resulting in the behavior seen in Figure 6A, where $\tau_{UB}^{stiff} < \tau_{UB}^{flex}$. If we increase the volume per polymer the ratio of the fluxes inverts so that $\tau_{UB}^{stiff} > \tau_{UB}^{flex}$. Since system volume does not affect the τ_B, we expect that more dilute polymer conditions would magnify the trends seen in the K_D of divalent targets (Figure 6B) resulting in the binding affinity of flexible polymers becoming much higher than the binding affinity of stiff polymers.

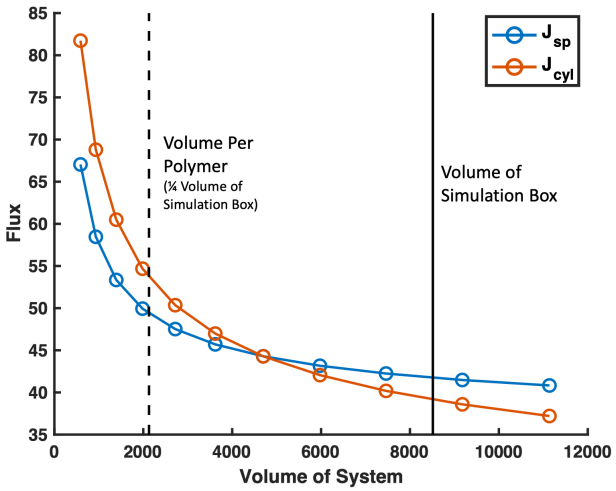


Figure 5-14: Flux of an unbound target toward a cylinder (orange) and sphere (blue) vs the system volume. The cylinder and sphere represent a rigid and flexible polymer respectively. At the simulation volume per polymer (- - -black line), the diffusive flux toward the cylinder (rigid polymer) is greater than the diffusive flux toward the sphere (flexible polymer).

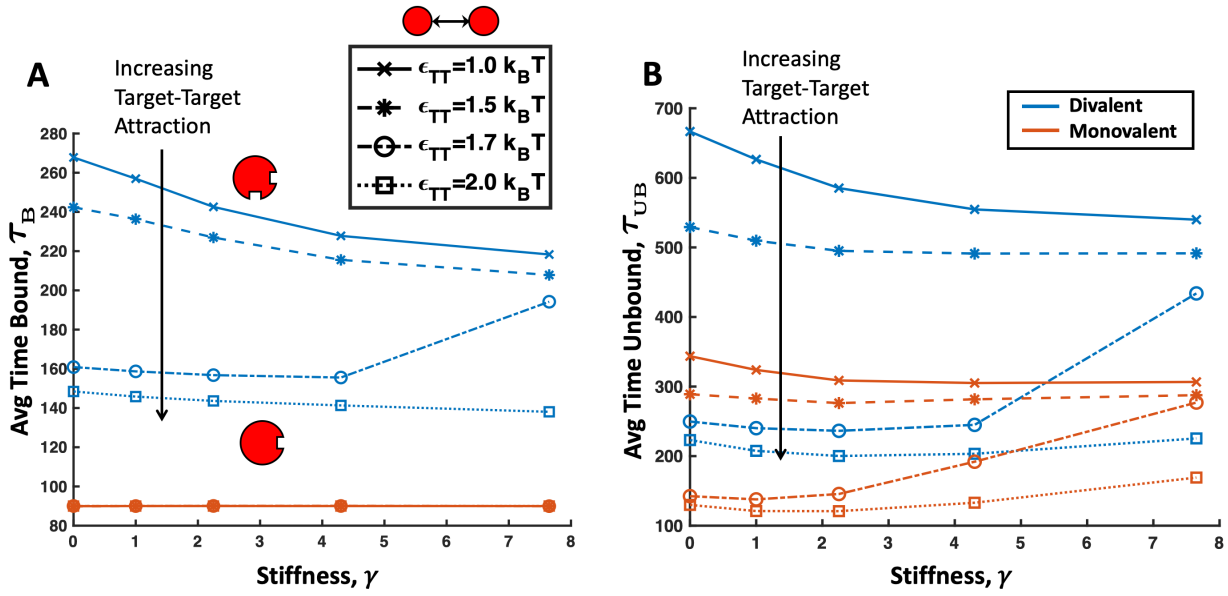


Figure 5-15: (A) Average time interval bound τ_B and (B) unbound τ_B for 96 targets. Monovalent targets are shown in orange and divalent targets are shown in blue, with different values of ϵ_{TT} denoted by different line styles and points. (A) Divalent targets see a decrease in τ_B with increasing ϵ_{TT} due to additional competition for sites between targets. Nucleation of a condensed polymer/target phase also results in increased competition, lowering the τ_B more than in the mixed/not phase separated state. Monovalent target τ_B is unaffected by stiffness or phase separation and lines for all ϵ_{TT} overlap. (B) For mixed systems, where no condensed phase is nucleated, τ_{UB} is dominated by diffusion and flexible polymers with spherical morphology experience longer τ_{UB} than rigid polymers for both divalent and monovalent targets. When systems are phase separated, flexible polymers have slightly shorter τ_{UB} than stiff polymers, likely due to a higher concentration of polymer binding sites in the condensed phase. Stiff polymers lower their concentration of binding sites in the condensed phase by extending their tails away from the targets as shown in Figure 5-8B and 5-9. Error bars are smaller than symbol size.

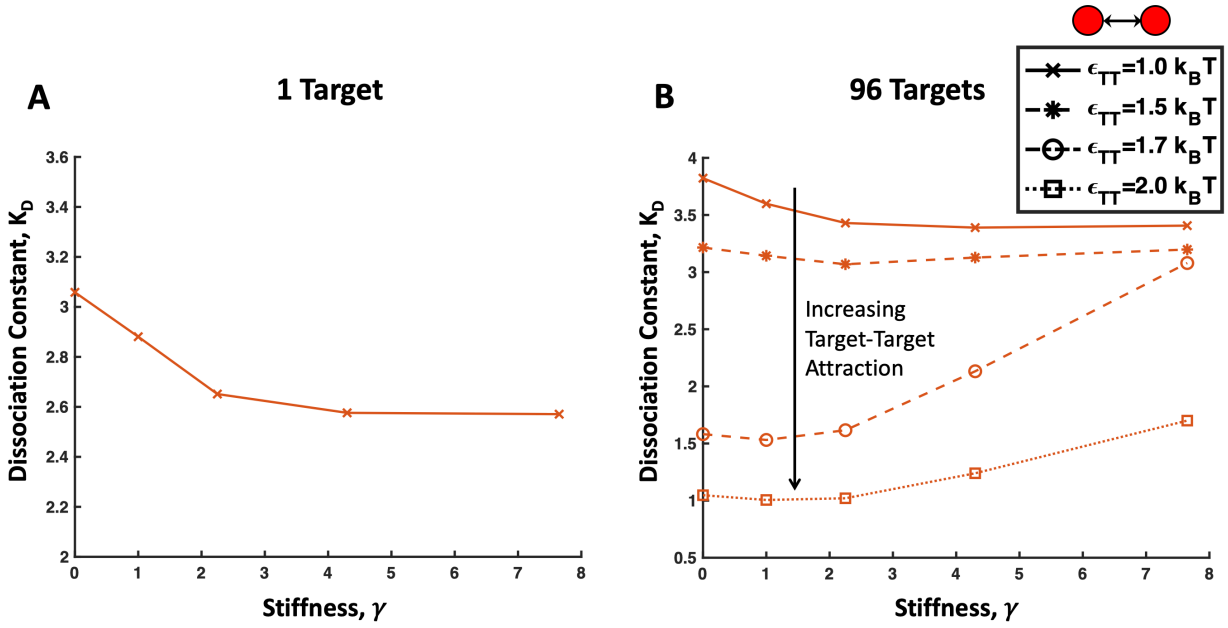


Figure 5-16: Dissociation constant K_D of monovalent targets for the single target case (A) and the 96 target case (B). (A) For one monovalent target, K_D is dominated by the time it takes the target to diffuse to a polymer. Because it takes longer for a target to diffuse to a sphere than to a rod, τ_{UB} is longer for flexible polymers than rigid polymers, so flexible polymers are lower affinity (higher K_D) for dilute monovalent targets. (B) For 96 monovalent targets, systems that don't phase separate behave similarly to the single target case; flexible polymers have lower affinity (higher K_D) than stiff polymers. When the system phase separates at $\epsilon_{TT} = 2.0 k_B T$, flexible polymers become higher affinity (lower K_D) than stiff polymers because stiff polymers extend away from the condensed target and are therefore bound less efficiently with a lower concentration of polymer binding sites in the condensed droplet. At $\epsilon_{TT} = 1.7 k_B T$, flexible polymers are significantly higher affinity than stiff polymers because they can induce phase separation at $\gamma \leq 4.3$ while stiff polymers ($\gamma = 7.65$) cannot. Error bars are smaller than symbol size.

5.2.3 Intra- and inter- divalent polymer bonds

Stiffer polymers have a slightly higher percentage of divalent bonds that are between two polymers (inter-polymer) than within a single polymer (intra-polymer). This is shown for a single target in Figure 5-17A and for many targets in Figure 5-17B. This is likely due to the high energy cost for the stiff polymers to form intra-polymer loops longer than a few beads. Since intra-polymer bonds of stiff polymers can have a high energy cost, this makes them closer to the energy of inter-polymer bonds, which are less entropically favorable because they confine the movement of two polymer instead of one. We suspect this leads to a higher percentage of bonds between two polymers as the polymers stiffen.

An upward trend as the polymer stiffness increases is seen for both target concentrations. Additionally for the 96 target concentration shown in Figure 5-17B, phase separation makes inter-target bonds more favorable. Since the polymers in a condensed phase are already held close together, the entropy cost of binding two polymers is lower than in the mixed or low concentration phase. This makes bonds across polymers much more frequent than in non-phase separated systems. Targets that form inter-polymer bonds cross-link the polymers and we can see that upon phase separation, the frequency of crosslinks goes up, likely nucleating a polymer gel in the condensed phase. In the mixed or not phase separated state, the low frequency of crosslinks suggests that there is no percolated network or gel. This phase separation with gelation is characteristic of many biocondensates [58].

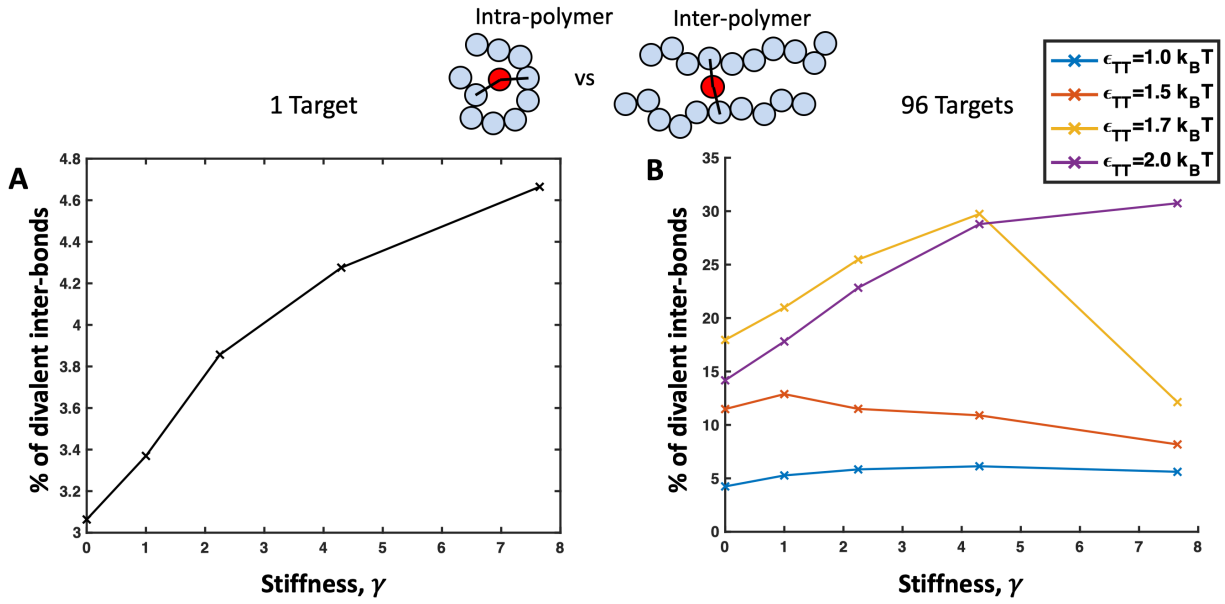


Figure 5-17: Percent of time the divalent target spends bound in loops between two polymers (inter-polymer) out of all loops formed. (A) Results for the single target case. Stiffer polymers have a higher percentage of inter-loops than flexible polymers, likely due to the energetic cost of bending for stiff polymers to form intra-polymer loops. (B) Results for 96 targets. For low inter-target attraction (blue, $\epsilon_{\mathrm{TT}} = 1.0$) and systems where all polymer stiffnesses are phase separated (purple, $\epsilon_{\mathrm{TT}} = 7.65$), behavior is similar to single target case where stiffness increases inter-polymer crosslinks. For $\epsilon_{\mathrm{TT}} = 1.7$ (yellow), flexible polymers have more crosslinks than stiff ones. In this case, more flexible polymers lead to droplets at $\epsilon_{\mathrm{TT}} = 1.7$ which brings polymer chains close together in a condensed phase and reduces the penalty for bonds across two polymers. At $\epsilon_{\mathrm{TT}} = 1.5$, flexible polymers are likely on the verge of phase separation and there are some transient small polymer-target droplets even though they don't nucleate a stable condensed phase. We suspect that crosslinks might occur less in phase separated systems with $\epsilon_{\mathrm{TT}} = 2.0$ than in $\epsilon_{\mathrm{TT}} = 1.7$ because the targets can phase separate by themselves and exclude the polymer from the droplet center through microphase separation. The effects of microphase separation will be explored in future work.

5.3 Bond types

We were also interested in considering what types of bonds were forming and whether this changed with stiffness or phase separation. The percent of time bound that a target has both of its binding sites occupied is shown in Figure 5-18. In general, stiffer polymers have a lower percentage of fully bonded divalent targets than flexible polymers, likely because the cost of bending for stiff polymers reduces the number of ways a target can bind twice to the polymer. This makes it more likely for a divalent target to only bind with a single binding site, with the other site unoccupied. As

the target-target attraction increases, the percentage of doubly bound targets decreases due to the increased competition for polymer sites from target neighbors. This can be seen very clearly for targets with $\epsilon_{\mathrm{TT}} = 1.7$, where more flexible polymers are phase separated so targets are at high concentrations and correspondingly experience high competition for binding. The stiffest polymer does not phase separate, so targets have a lower local concentration, resulting in less competition for polymer binding sites, and more frequent double bonds to the polymer.

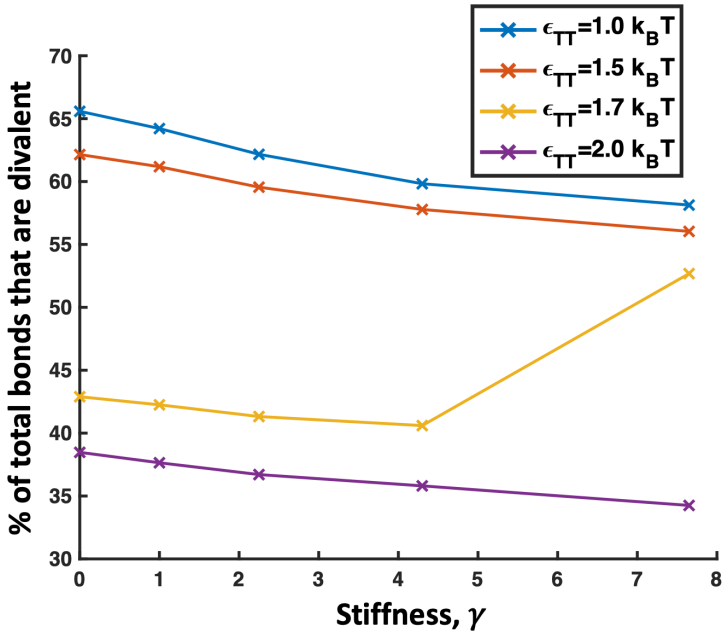


Figure 5-18: Percent of time a bound target has both binding sites bound simultaneously versus polymer stiffness. A target is considered bound if one or more of its binding sites is occupied. Data is shown for 96 divalent targets binding to four 16mer polymers. Lines represent constant target-target attraction. Error bars are smaller than symbol size.

5.3.1 Binder cumulant and radius of gyration for simulations with multiple targets

The following figures detail the average energy, Binder cumulant, system polymer radius of gyration R_{g}, and the average R_{g} for each polymer. Results are shown along lines of constant target concentration. These results were used to determine whether a system was phase separated or not. As described in the main text, the Binder cumulant compares the variance in the energy or specific heat to the average system energy and should be at a maximum when the system is on the

phase boundary. The system and individual polymer radius of gyrations should be at a minimum at the phase boundary because the polymer system collapses due to the change in effective solvent quality upon entering a target/polymer droplet. Then the system should swell as more targets and polymers aggregate into the droplet.

Divalent targets

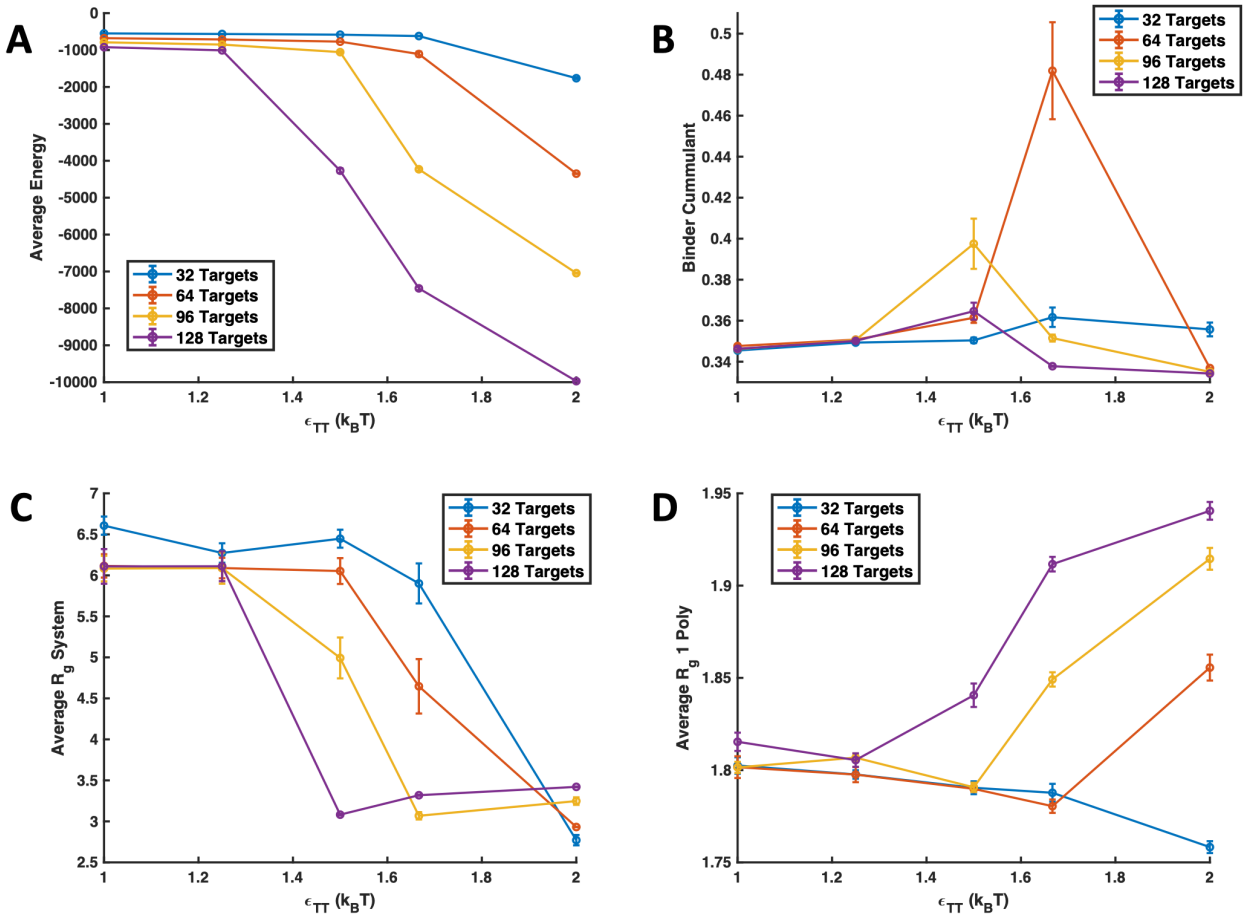


Figure 5-19: Simulation results for divalent targets with $\gamma = 0$. Data is shown for 32, 64, 96, and 128 divalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{TT} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_g of all polymer beads relative to the collective system center of mass. (D) Average R_g of each individual polymer relative to its own center of mass.

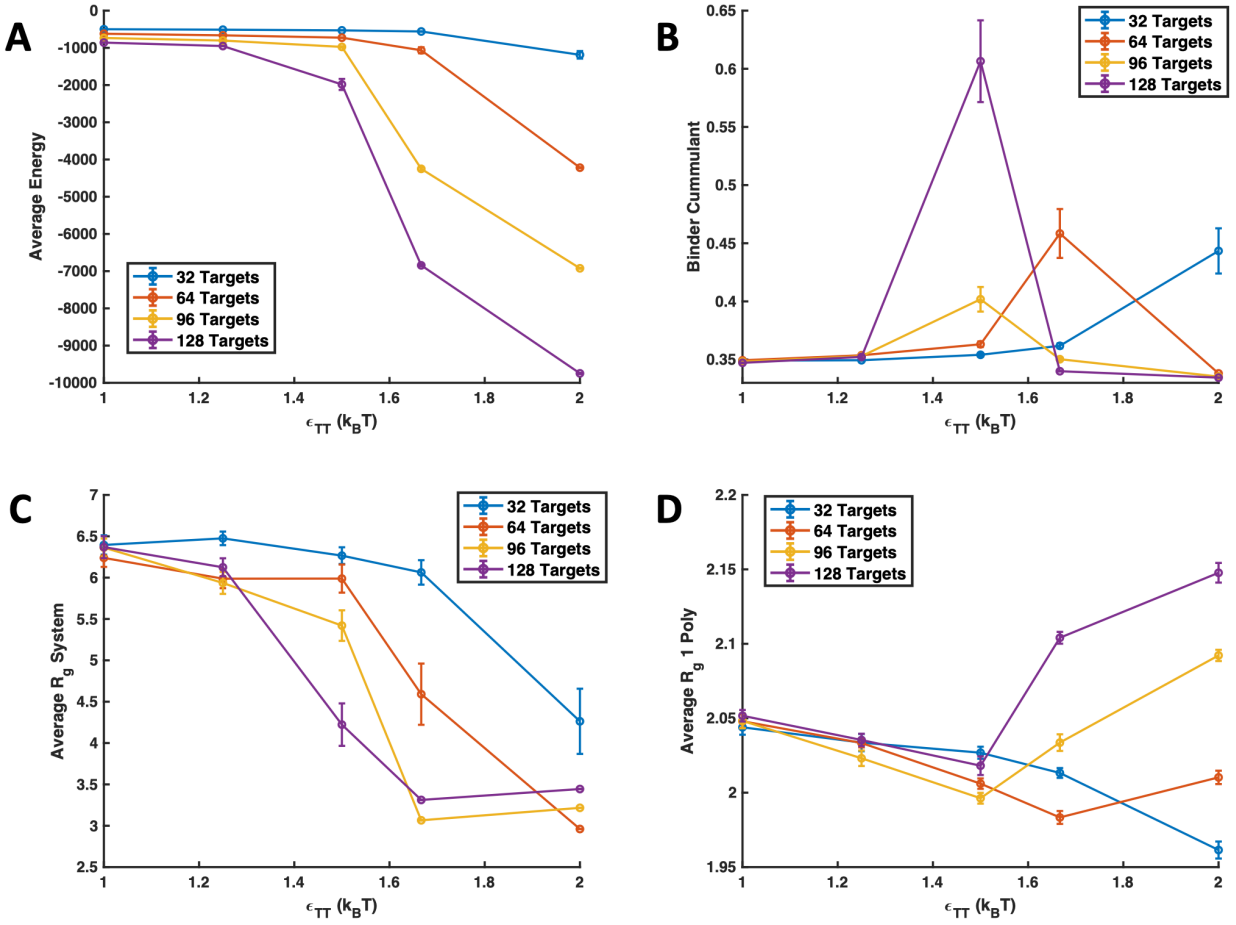


Figure 5-20: Simulation results for divalent targets with $\gamma = 1.0$. Data is shown for 32, 64, 96, and 128 divalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{TT} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_g of all polymer beads relative to the collective system center of mass. (D) Average R_g of each individual polymer relative to its own center of mass.

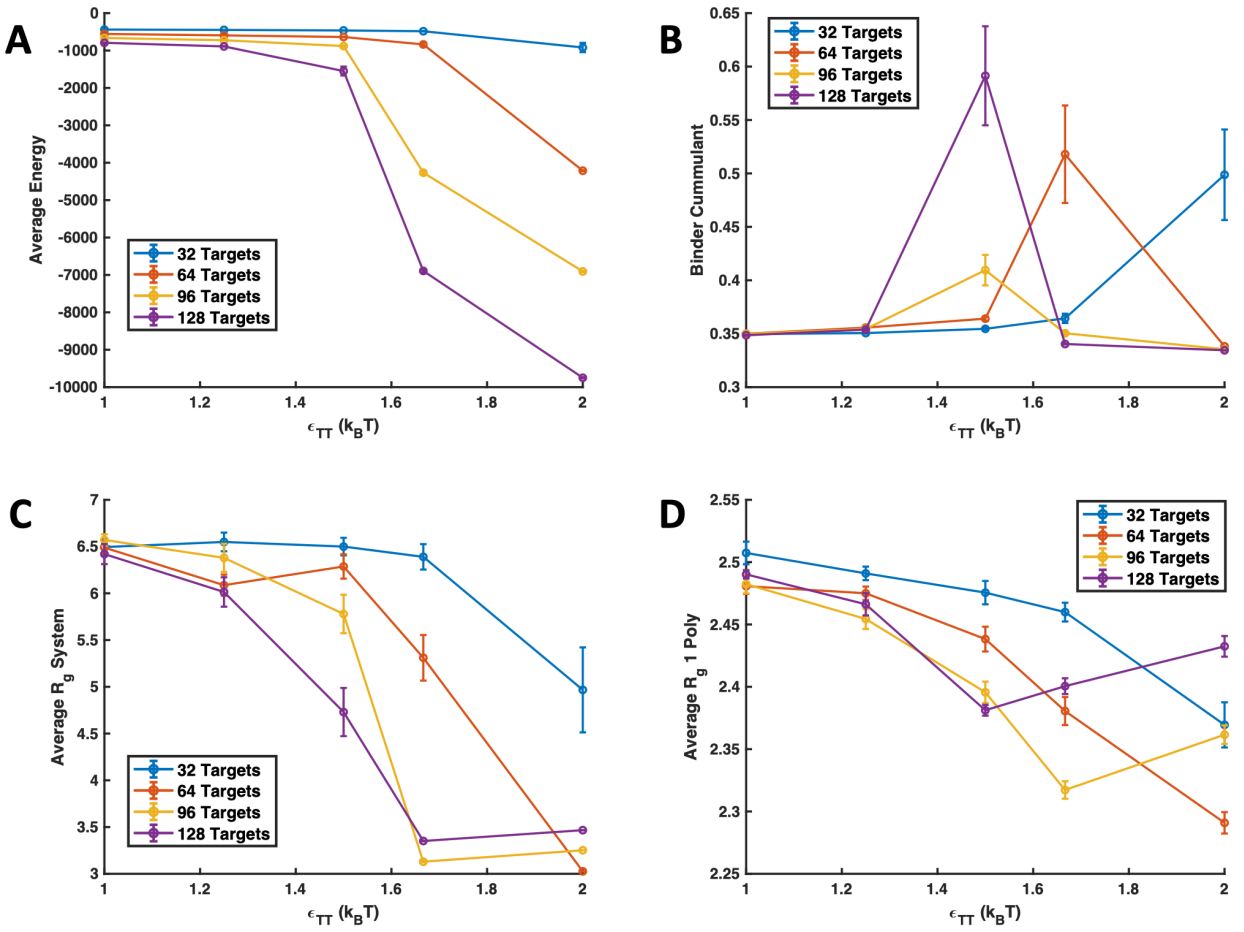


Figure 5-21: Simulation results for divalent targets with $\gamma = 2.25$. Data is shown for 32, 64, 96, and 128 divalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{TT} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_g of all polymer beads relative to the collective system center of mass. (D) Average R_g of each individual polymer relative to its own center of mass.

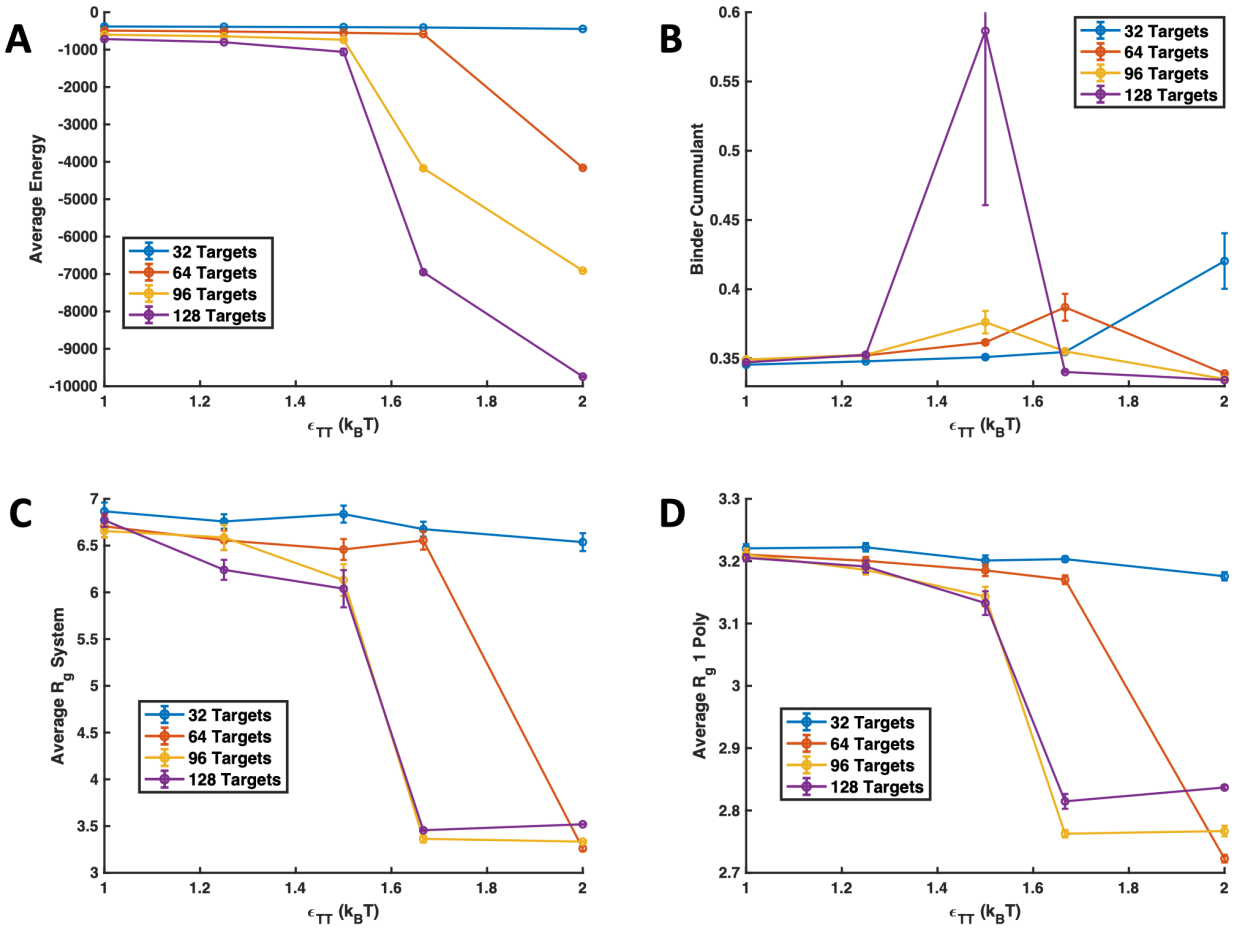


Figure 5-22: Simulation results for divalent targets with $\gamma = 4.3$. Data is shown for 32, 64, 96, and 128 divalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{TT} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_g of all polymer beads relative to the collective system center of mass. (D) Average R_g of each individual polymer relative to its own center of mass.

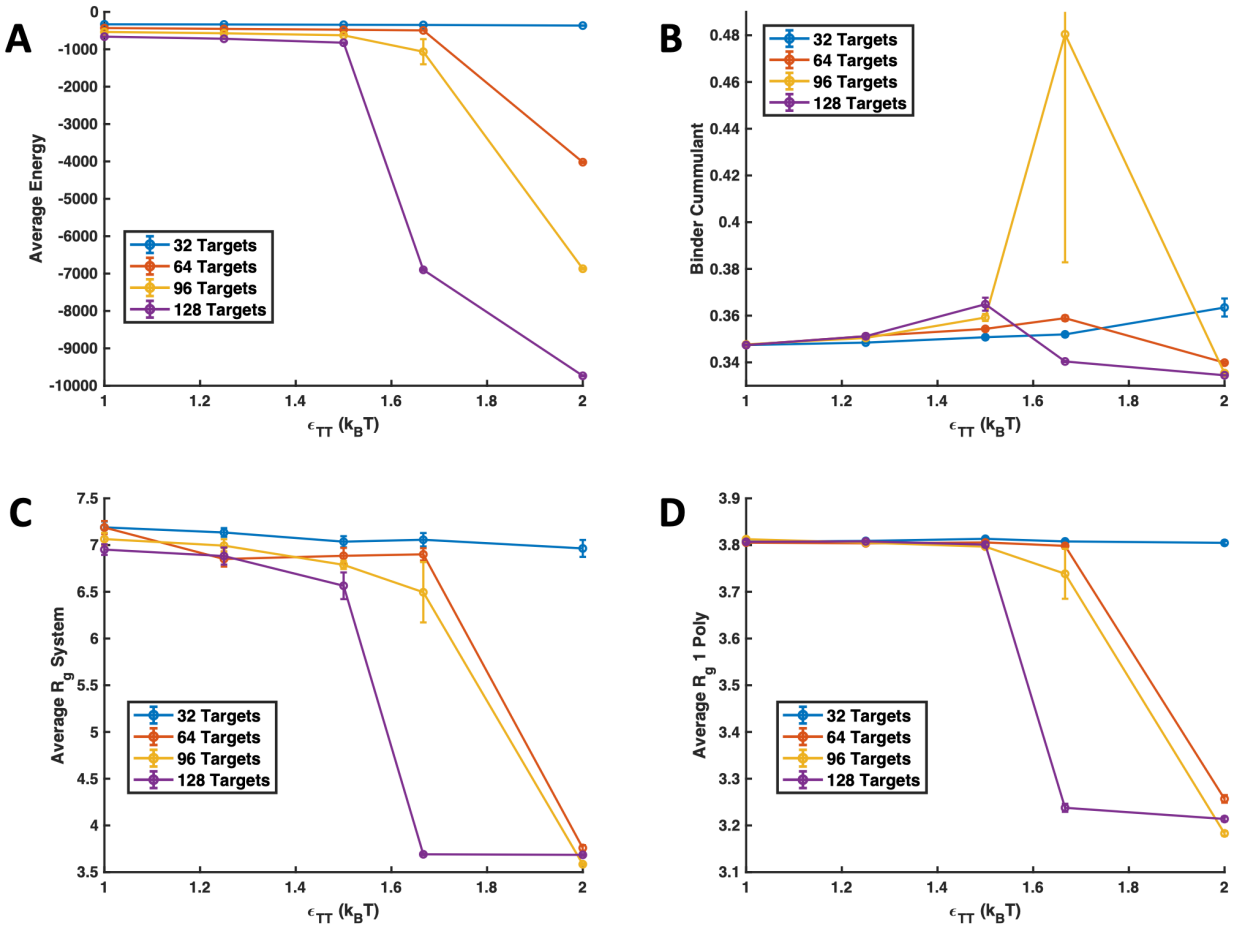


Figure 5-23: Simulation results for divalent targets with $\gamma = 7.65$. Data is shown for 32, 64, 96, and 128 divalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{TT} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_g of all polymer beads relative to the collective system center of mass. (D) Average R_g of each individual polymer relative to its own center of mass.

Monovalent targets

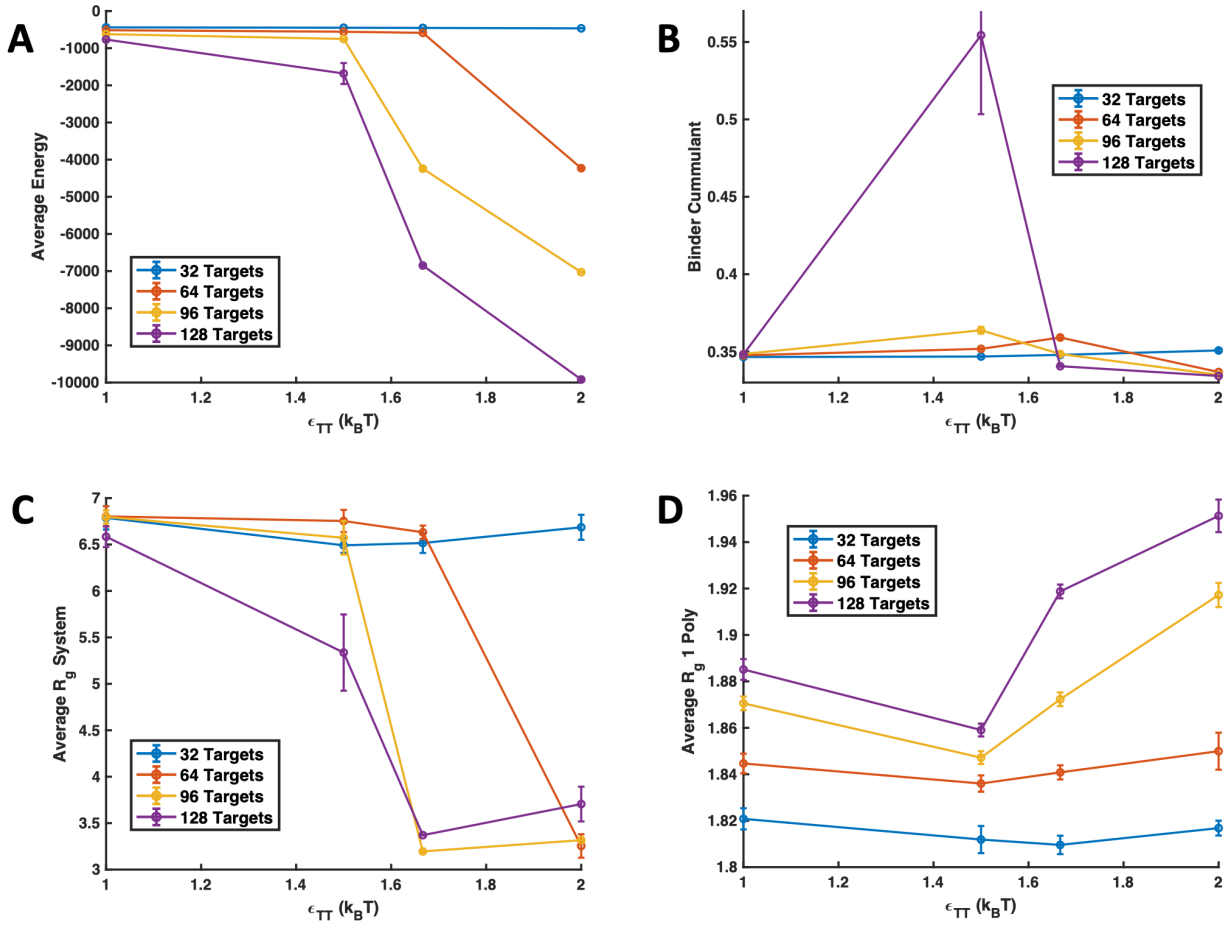


Figure 5-24: Simulation results for monovalent targets with $\gamma = 0$. Data is shown for 32, 64, 96, and 128 monovalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{TT} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_g of all polymer beads relative to the collective system center of mass. (D) Average R_g of each individual polymer relative to its own center of mass.

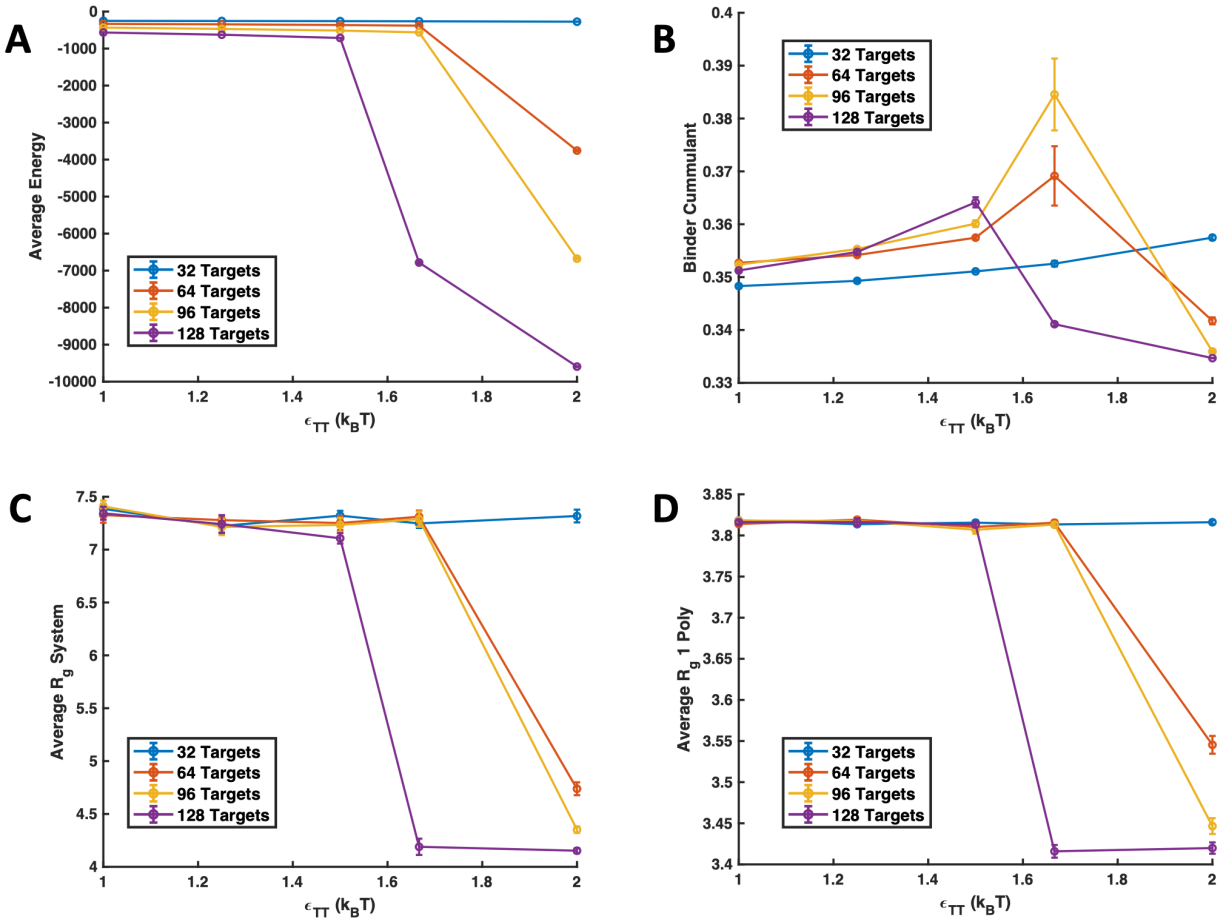


Figure 5-25: Simulation results for monovalent targets with $\gamma = 7.65$. Data is shown for 32, 64, 96, and 128 monovalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{TT} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_g of all polymer beads relative to the collective system center of mass. (D) Average R_g of each individual polymer relative to its own center of mass.

Trivalent targets

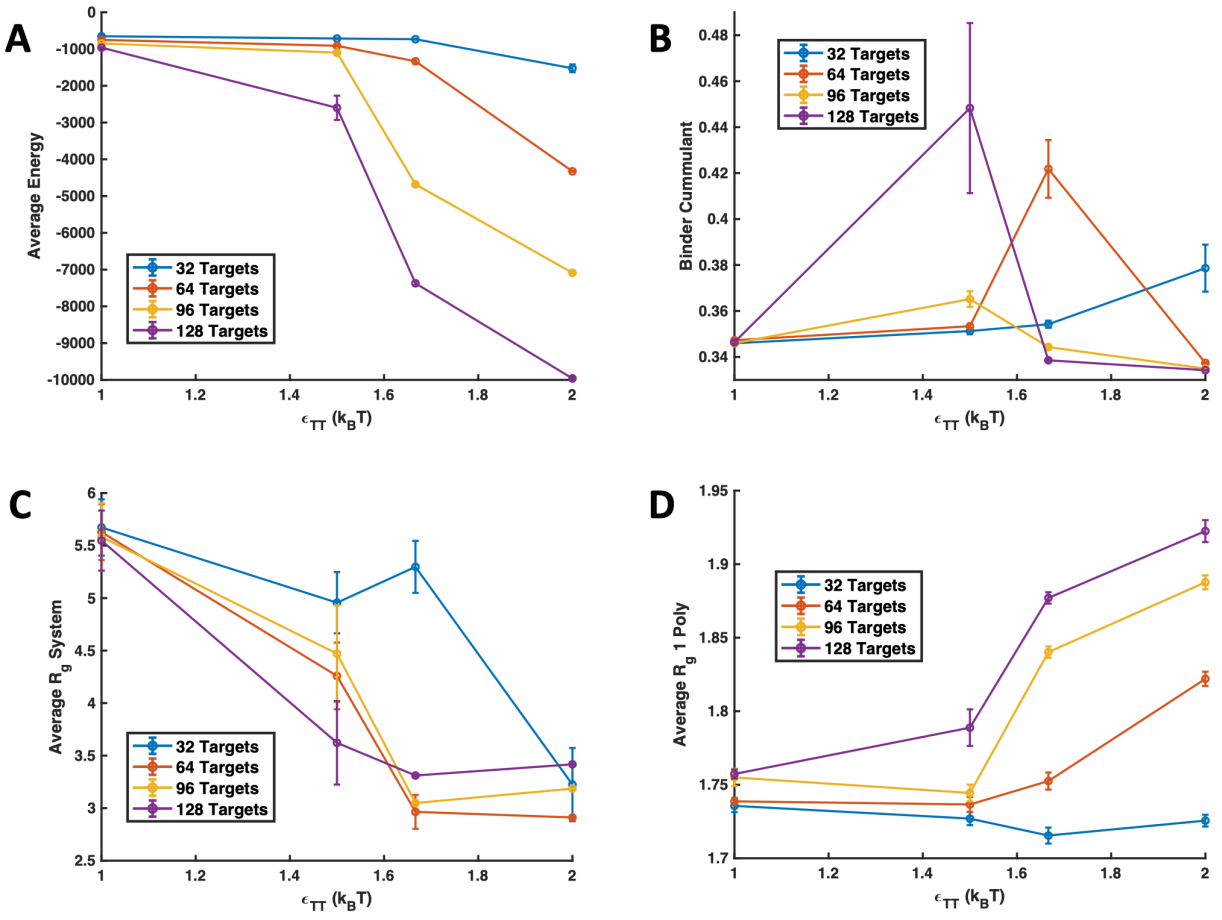


Figure 5-26: Simulation results for trivalent targets with $\gamma = 0$. Data is shown for 32, 64, 96, and 128 trivalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{TT} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_g of all polymer beads relative to the collective system center of mass. (D) Average R_g of each individual polymer relative to its own center of mass.

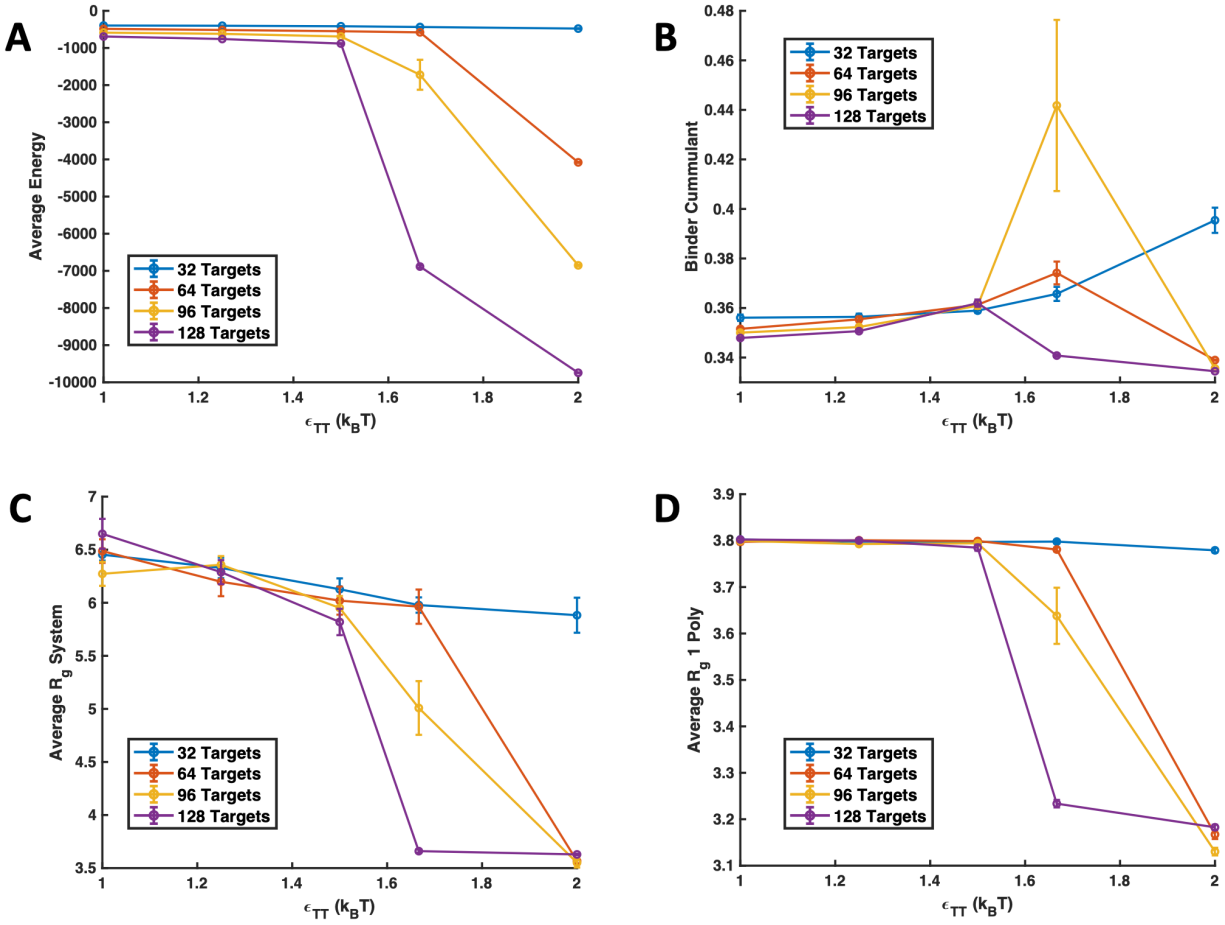


Figure 5-27: Simulation results for trivalent targets with $\gamma = 7.65$. Data is shown for 32, 64, 96, and 128 trivalent targets interacting with four 16mer polymers with target-target attractions ranging from $\epsilon_{\mathrm{TT}} = 1.0$ to 2.0. Lines connect points of constant target concentration. (A) Average energy of the system. (B) Binder cumulant. (C) Average R_{g} of all polymer beads relative to the collective system center of mass. (D) Average R_{g} of each individual polymer relative to its own center of mass.

Chapter 6

Multivalent Polymers Can Control Phase Boundary, Dynamics, and Organization of Liquid-Liquid Phase Separation

Abstract

Multivalent polymers are a key structural component of many biocondensates. When interacting with their cognate binding proteins, multivalent polymers such as RNA and modular proteins have been shown to influence the liquid-liquid phase separation (LLPS) boundary to control condensate formation and to influence condensate dynamics after phase separation. Much is still unknown about the function and formation of these condensed droplets, but changes in their dynamics or phase separation are associated with neurodegenerative diseases such as ALS and Alzheimer's Disease. Therefore, investigation into how changes in the structure of multivalent polymers relate to changes in biocondensate formation and maturation is essential to understanding and treating these diseases. Here, we use a coarse-grain, Brownian Dynamics simulation with reactive binding that mimics specific interactions in order to investigate the difference between non-specific and specific multivalent binding polymers. We show that non-specific binding interactions can lead to much larger changes in droplet formation at lower energies than their specific, valence-limited counterparts. We also demonstrate the effects of solvent conditions and polymer length on phase separation, and how modulating binding energy to the polymer can change the organization of a droplet in a

three component system of polymer, binding protein, and solvent. Finally, we compare the effects of surface tension and polymer binding on the condensed phase dynamics, where we show that both lower protein solubilities and higher attraction/affinity of the protein to the polymer result in slower droplet dynamics. We hope that the research presented in this work helps to better understand experimental systems and provides additional insight into how multivalent polymers can control LLPS. The work presented in this chapter is primarily sourced from Zumbro and Alexander-Katz, *In Preparation* (2020) [137].

6.1 Introduction

Multivalency is employed throughout biology for numerous reasons including building conformal interfaces, increasing specificity of bonds using a limited number of ligand types, and creating much stronger bonds by using many low affinity bonds simultaneously [1]. Multivalent binding is defined as when multiple binding sites on both interacting species bind simultaneously to create a much stronger bond than the sum of the constituent monovalent binding affinities [1]. Multivalent species can come in many architectures, but here, we focus on multivalent polymers and their role in biocondensates or membraneless organelles. Multivalent proteins and nucleic acids have been found in many membraneless organelles. Although these biocondensates can have tens to hundreds of components, studies have shown that multivalent polymers are key directors of the phase separation of condensates and multivalent polymers can undergo phase separation with just their target binding species in vivo, in vitro, and in simulation [50,49,51,83,58]. These studies suggests that controlling features of multivalent polymers can modulate the formation of biocondensates and the recruitment of other important components after the initial phase separation[58,138,45,47].

Because aberrant phase separation of these biocondensates is associated with neurodegenerative diseases, understanding their formation is an important area of research [64,49,65]. Since multivalent polymers control these biocondensates, exploring how multivalent polymer properties can change the kinetics and thermodynamics of liquid-liquid phase separation (LLPS) is essential. Theoretical and experimental studies of these systems have shown that increasing valency and individual binding site affinities can lower the phase separation boundary to lower species concentrations [83,138]. Another

theoretical study found that the solvation of polymeric linkers between binding sites controls whether polymers form a cross-linked gel with or without phase separation [58]. In previous work, we showed that polymer flexibility or persistence length can change the phase boundary of multivalent polymers [114]. In this paper, we further build on the understanding that multivalent polymer characteristics can significantly change the thermodynamics of biocondensates by exploring the effect of non-specific vs specific polymer binding interactions, condensed phase nucleation in smaller systems, and the dynamics of the resulting condensed polymer phase.

We use a coarse-grain Brownian Dynamics simulation to explore the phase separation of long, many-valent polymers such as RNA and smaller binding targets such as RNA-binding proteins such as those found in ribonucleicprotein (RNP) granules [118,117]. Using Brownian Dynamics allows us to capture both the thermodynamics and kinetics of phase separated polymer-target systems. Modeling the impact of polymer characteristics on the dynamics of globules could provide insight into the liquid-to-solid transition in biocondensates that is associated with disease [133]. We show that nucleating a condensed phase using non-specific interactions, such as hydrophobicity or charge, occurs at lower attraction energies than using valence limited lock-and-key type binding such as those between a ligand and a folded protein pocket. Therefore, non-specific and specific interactions can be combined to carefully adjust phase transition boundaries. By looking at the morphologies of the resulting condensates, we also explain how changing polymer interactions can control the spacial organization of the condensed phase. Last, we investigate how polymer properties can alter the kinetics inside condensed droplets.

6.2 Computational Methods

To study the condensed phase nucleation of multivalent polymers, we use coarse-grain Brownian dynamics simulations with a bead-spring polymer and a spherical binding target represented as a single bead of the same size as the polymer beads using the same methods detailed in Chapter 2. This scenario represents a general model of the protein-protein or nucleic acid-protein binding found in the formation of membraneless organelles. It most closely resembles a piece of RNA binding to an RNA binding protein such as hnRNPA1 found in stress granules, and whose solidification of

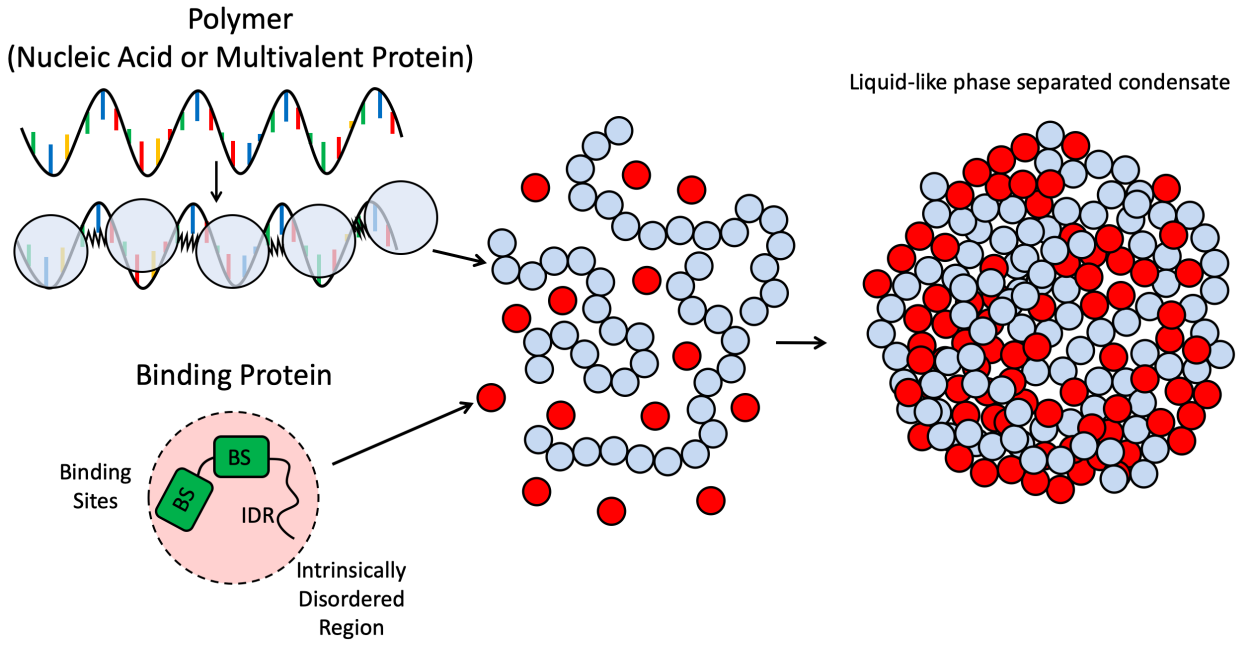


Figure 6-1: Depiction of simulation scheme. Polymers are represented by spherical beads (light blue) connected by harmonic springs. These polymers could represent either nucleic acids or long modular binding proteins found in biocondensates. Each polymer bead has a single binding ligand. Target binding proteins are represented as spherical beads (red) and can have multiple binding sites (BS) depicted as green blocks. These protein beads also encompass a intrinsically disordered region (IDR) that modulates their non-specific attraction to the polymer and between the proteins themselves. When the polymers and binding proteins are mixed together, they can undergo a phase transition into a condensed droplet.

the condensed phase is associated with amyotrophic lateral sclerosis (ALS) and fronto-temporal dementia [139,117,118].

In addition to excluded volume, we also use a Lennard-Jones potential to create non-specific attraction between target binding proteins and polymer beads as shown in Figure 6-2A. This generic potential could represent attraction due to hydrophobicity or van der Waals. Here, we could substitute a screened electrostatic potential but do not expect this to qualitatively change our results. Unless otherwise specified, we chose $\epsilon_{\mathrm{PP}} = \frac{5}{12}$ to mimic polymer configurations in a theta solvent [70] as shown in Figure 6-3A. For protein targets binding with valence-limited lock and key bonds, we used polymer target potential $\epsilon_{\mathrm{PT}} = \frac{1}{12}$ to mimic a good solvent and separate non-specific and specific binding interactions. Between the targets themselves, ϵ_{TT} was varied from $\frac{6}{12}$ to $\frac{36}{12}$ to capture a range of target protein solubilities and are specified in later requisite sections.

The last type of interaction included is a reactive lock and key bond shown in Figure 6-2B,

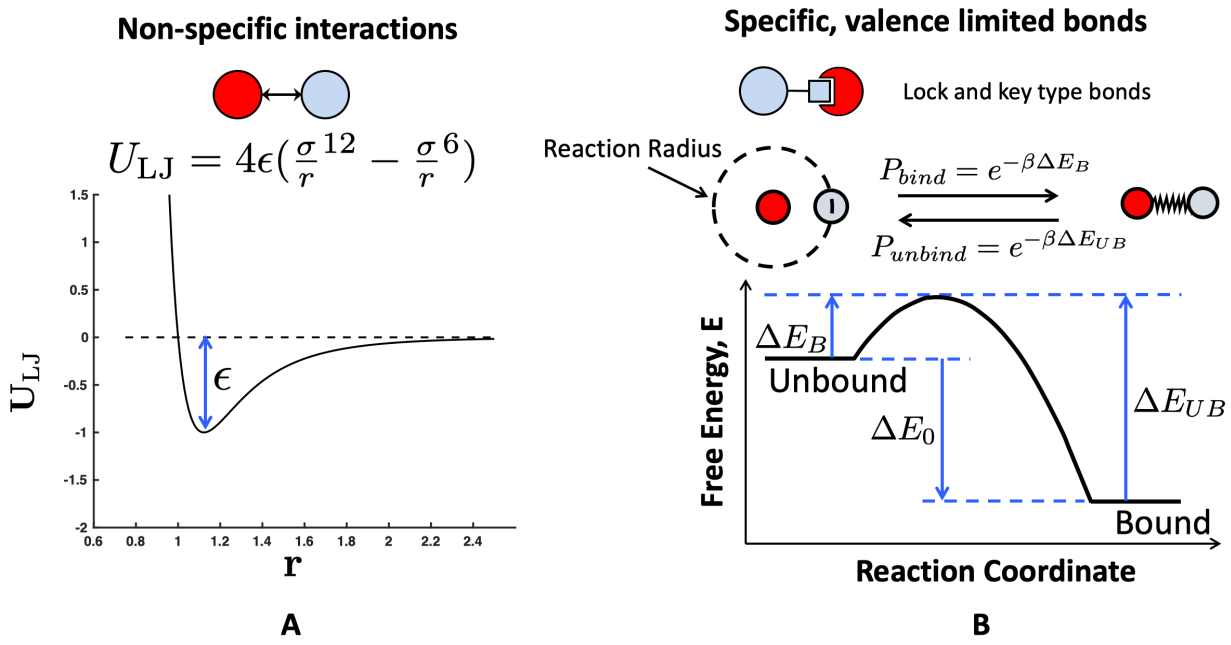


Figure 6-2: Two types types of protein-polymer interactions are explored in this work. (A) Non-specific excluded volume interactions controlled by a Lennard-Jones potential. These potentials are not valence limited and are felt by any target or polymer bead in accordance with their distance apart r. (B) Specific, valence-limited, lock-and-key type binding. Polymer ligands and target protein binding sites interact when they are within a reaction radius that is dependent on the timestep. Within this reaction radius, they have a probability of binding P_{B} that depends on the depicted free-energy landscape. Once bound, the target and polymer bead are connected by a harmonic spring, and with some probability, P_{UB}, can return to being unbound and interacting solely through a Lennard-Jones potential. This figure is adapted from Zumbro *et al.* with permission from Elsevier [86].

which represents our specific, valence-limited binding interaction between polymers and targets with details of this potential in Chapter 2. This reactive binding scheme is applied with a free energy of binding per site of $\Delta E_0 = -2, -4,$ and $-6k_{\mathrm{B}}T$. Each polymer bead contained a single binding site, and each target bead was given 1, 2, or 3 binding sites in order to capture the effects of changing binding valency. Assuming the size of a target bead to be approximately 5 nm in diameter, and using Langmuir adsorption theory, we can convert this $\Delta E_0 = -2, -4,$ and $-6k_{\mathrm{B}}T$ binding energy into a dissociation constant in Molar, resulting in a monovalent binding affinity of approximately $K_{\mathrm{D}} = 0.8, 0.1,$ and 0.02 mM, respectively. Details of this conversion are shown in Chapter 2. This monovalent binding affinity is well within the weakly binding mM to μM affinity range of some monovalent protein-protein and RNA-protein binding found in biocondensates [83–85].

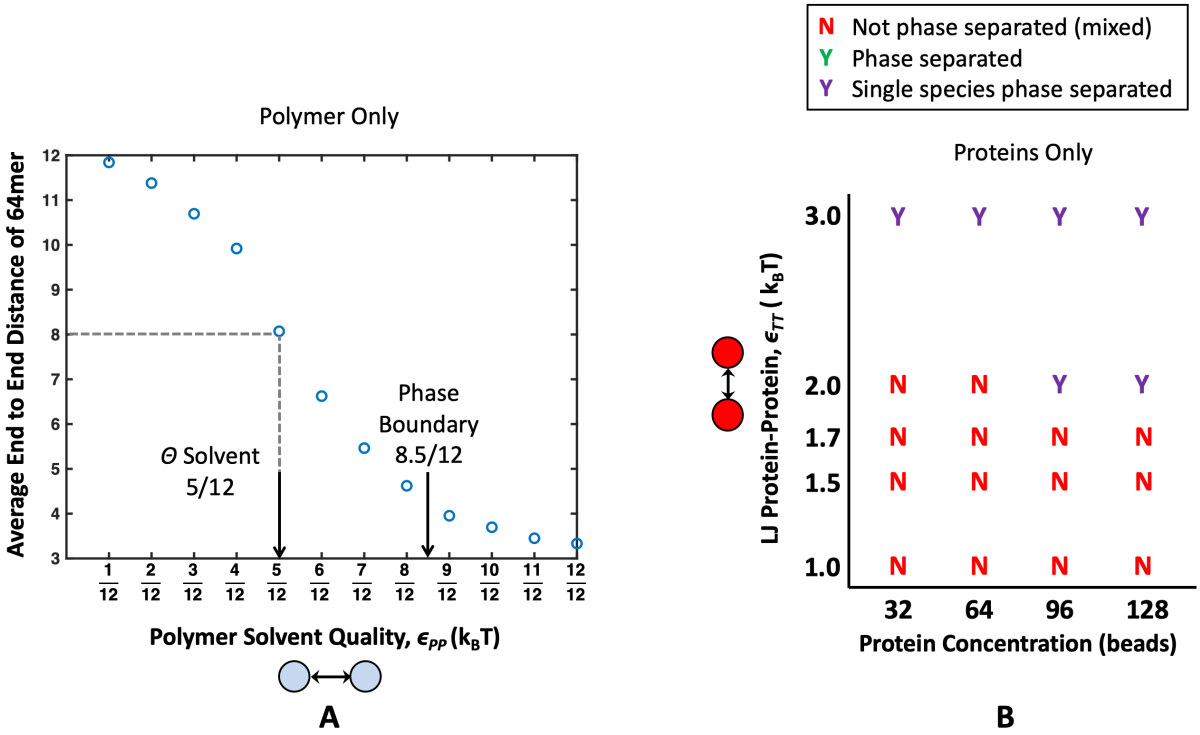


Figure 6-3: Properties of a single species alone, before mixing them together. (A) Average end-to-end distance of a 64mer polymer under various Lennard-Jones attractions ϵ_{PP}. The polymer behaves as it would in θ conditions, as a perfect random walk, when $\epsilon_{PP} = \frac{5}{12}$. $\epsilon_{PP} = \frac{8.5}{12}$ is highlighted with an arrow to denote the attraction at which four 16mer polymers aggregate into a single condensate. From this, we can see there is a region of poor solvent where polymers are collapsed but still soluble. (B) Phase diagram showing solubilities of binding proteins alone. When targets form a condensed phase without polymer, it is denoted with a purple "Y", and when they do not form a condensed phase, it is denoted with a red "N". From this chart, we see that all target concentrations tested are phase separated when $\epsilon_{TT} = 3.0$, no target concentrations nucleate a condensed phase at $\epsilon_{TT} = 1.7$, and only high target concentrations 96 and 128 targets phase separate at $\epsilon_{TT} = 2.0$. This phase diagram will serve as a control for the effects of mixing polymers and target proteins.

6.3 Results and Discussion

In biocondensates, species can often phase separate by themselves, but interactions with another species can cause phase separation at lower energies or concentrations [133,140]. As a control, we first ran simulations of a single species, only targets and only polymers, to understand their solubility before mixing at different inter-polymer and inter-target attractions. These results for polymers only and targets only are plotted in Figures 6-3A and B respectively. Polymer with a degree of polymerization of 64 beads, were shown to behave as freely-jointed random walks, characteristic of θ solvent conditions at $\epsilon_{PP} = \frac{5}{12}$, consistent with previous literature [70]. Four 16mer polymers were

seen to condense into a single droplet at $\epsilon_{PP} > \frac{8.5}{12}$. This means that, at this polymer concentration used throughout our simulations, there is a region of ϵ_{PP} where polymers can experience poor solvent quality but not phase separate on their own. Throughout this work, we will consider polymers with $\epsilon_{PP} \leq \frac{7}{12}$, meaning that in all results discussed below, the polymers do not phase separate on their own.

We also consider the potentials necessary to phase separate the targets alone at our simulation concentrations. By themselves, target concentrations of 96 and 128 forming a condensed phase at $\epsilon_{TT=2.0}$, all target concentrations nucleated a condensed droplet at $\epsilon_{TT} = 3.0$, and no targets phase separated at $\epsilon_{TT} = 2.0$. These inter-target energies bounded the parameter space for our simulations where polymers and target proteins were mixed together. In later phase diagrams, energy and concentrations regions where targets can phase separate on their own are shaded with a purple background.

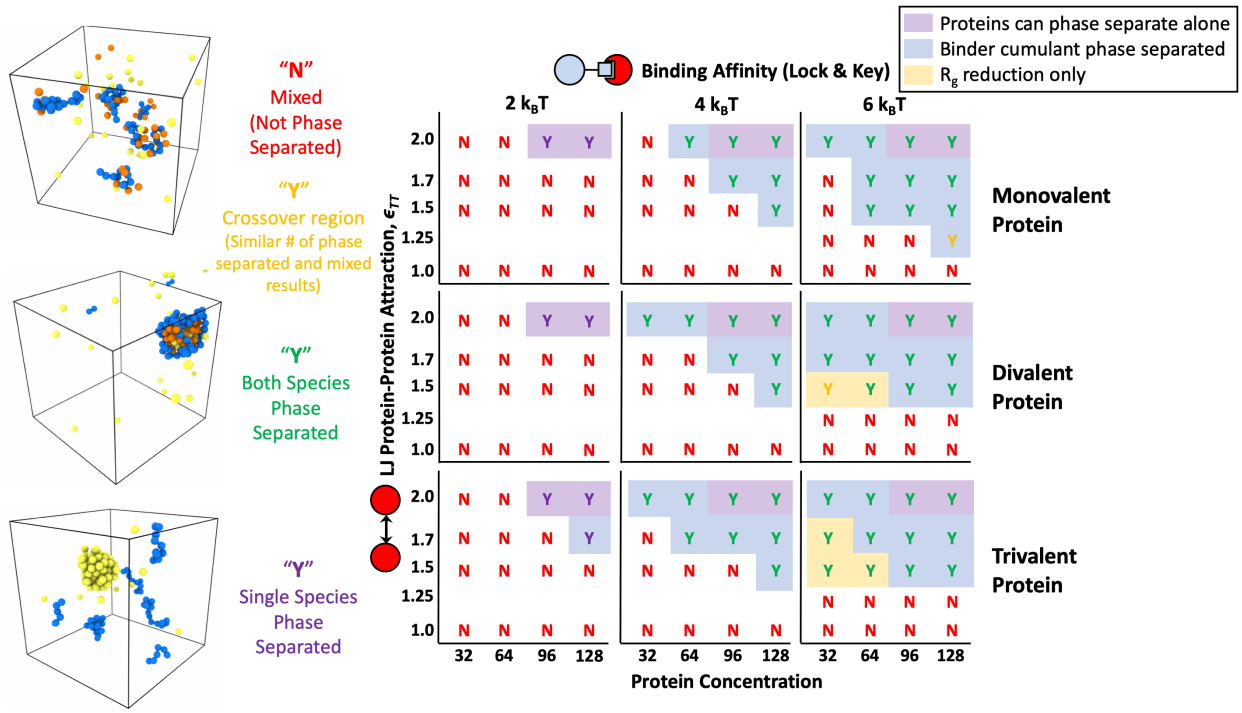


Figure 6-4: Phase diagram resulting from specific lock-and-key binding to four 16mer polymers. Results are shown for mono, di, and trivalent binding proteins with $\Delta E_0 = 2$, 4, and $6k_BT$. Letters and letter coloring were determined by visual inspection, with example renderings shown on the left of "Mixed" states labeled as a red "N" for no phase separation, fully phase separated systems with both polymers and proteins found in the condensed phase labeled with a green "Y" for yes phase separated, and purple 'Y"s denoting systems in which a single species phase separated without the other such as the proteins condensing on their own. Yellow "Y"s denote systems in a the crossover region between phase separated and mixed where 60% of simulations showed a stable condensed droplet. Purple background shading denotes regions where pure protein simulations phase separated on their own without the help of the polymer. Blue background shading denotes the regions where phase separation was also indicated by Binder cumulant of the system energy. Yellow background shading denotes that aggregation of polymers into a droplet was indicated by a significant drop in the total R_g of the polymer system accompanied by a reduction in the R_g of individual polymers. Phase separation occurs at lower protein target concentrations and lower ϵ_{TT} as valency and binding affinity are increased.

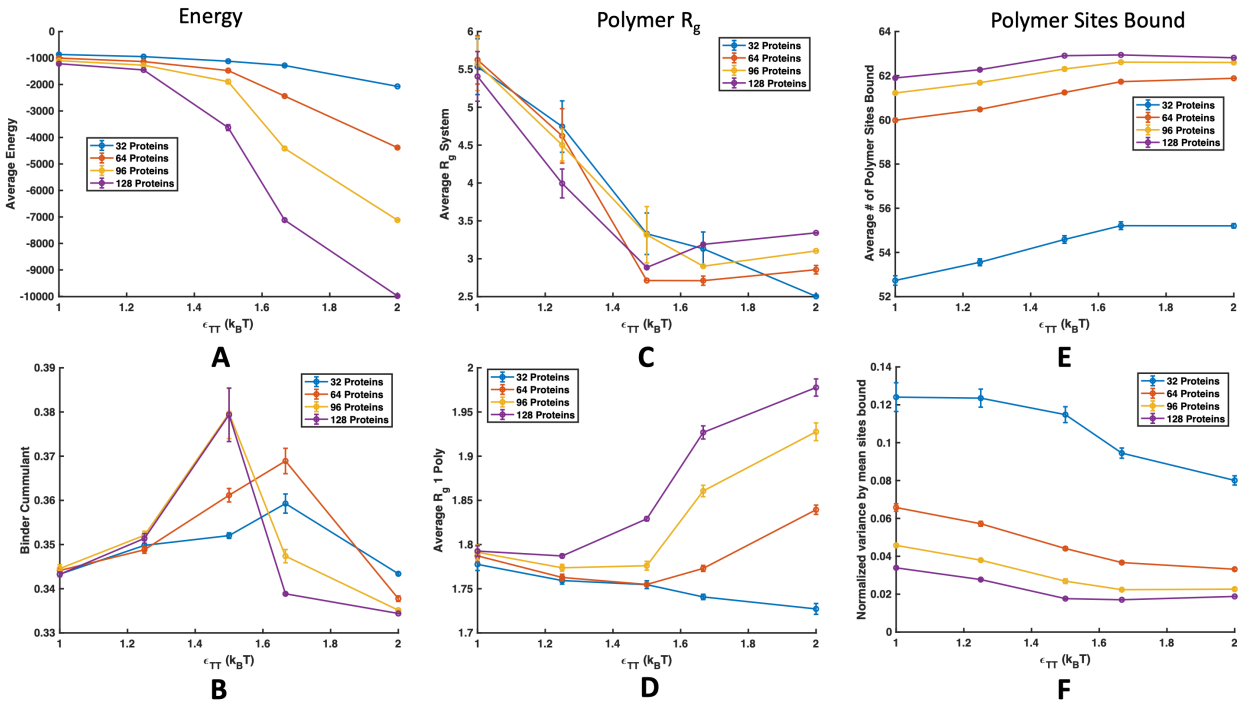


Figure 6-5: Examples of simulation properties for divalent protein targets with four 16mer polymers in theta solvent and $\Delta E_0 = 6k_BT$, shown with lines of protein concentration. (A) Total average energy of simulation (B) Binder cummulant comparing average energy fluctuations to average system energy. A maximum in the Binder cumulant corresponds to a phase boundary. (C) R_g of all polymers in the system. A large reduction in system R_g signifies that all four polymers aggregated into a single body. (D) Average R_g of individual polymers across the simulation time. A reduction in R_g signifies a change in effective solvent conditions for the polymer as a result of complexation with binding proteins. After a critical concentration of protein binding is achieved, the polymers swell if the simulation isn't protein concentration-limited. (E) Number of polymer sites bound with a maximum of 64. A plateau in sites bound occurs when a protein-polymer droplet is formed because the local concentration of protein targets reaches a maximum. (F) Variance in polymer binding sites occupied normalized by the average number of sites bound. The variance also plateaus when a condensed droplet is formed due to the smaller fluctuations in local concentration of proteins near the polymer in a liquid droplet.

6.3.1 Valency and Affinity of Specific Lock and Key Bonding

To compare our system to previous computational studies, we first investigated the effects of valency and binding site affinity on the LLPS of multivalent polymers and targets with specific, valence limited binding interactions. To do so, we placed four polymers with a degree of polymerization $N_P = 16$ beads in θ solvent with 32, 64, 96, and 128 binding protein targets. Monovalent, divalent, and trivalent targets were simulated with 3 different binding site affinities, and the resulting phase

diagrams are shown in Figure 6-4. In order to determine if a system nucleated a stable condensed phase, we initially used visual inspection to look for the proportion of 10 runs in which a condensed droplet of targets and polymers persisted for the last quarter of the simulation time. Simulations in which a stable droplet formed more than 70% of the time are marked with a green “Y", systems that formed a droplet in 60% of runs are marked with a yellow “Y", and systems where less than 50% of runs formed a droplet are marked with a red “N" for no phase separation. We further confirmed these phase separations by calculating the average energy of the last quarter of run time and using it to compute the Binder cumulant $\frac{<E^4>}{3<E^2>^2}$. This quantity compares the ratio of the energy variance, which is equivalent to the specific heat of the system, to the average energy, and shows a maximum at the phase transition [126]. Figure 6-5 shows an example of average energy and cumulant plots for a divalent protein with lock and key binding affinity of $\Delta E_0 = -6k_{\mathrm{B}}T$. By comparing the Binder cumulant along lines of constant target concentration, we confirmed our initial phase diagrams created through visual inspection. Areas where the cumulant predicts phase separation are shaded with a blue background in Figure 6-4.

We found that the cumulant did not fully predict simulations in which condensed phases were nucleated, so we also calculated the radius of gyration R_g of the polymers individually and all together to capture when the polymers themselves showed aggregation and collapse. Methods of measuring aggregation through R_g was used in previous computational work on phase separation of biocondensates[58]. When four 16mer polymers come together in θ conditions into a liquid droplet, without considering any swelling from binding proteins, they should have a similar R_g to a polymer with $N_p = 64$ which in the ideal case is $\sqrt{\frac{N_p l^2}{6}} = 3.27$ where l is the diameter of a bead $2a$. In the example shown in Figure 6-5B, all target concentrations show clear polymer aggregation with $R_g \leq 3.27$ at $\epsilon_{\mathrm{TT}} = 1.5$, and show an reduction in the individual polymer sizes around $\epsilon_{\mathrm{TT}} = 1.25$. This aligned well with our observation of droplets at energies lower than the Binder cumulant predictions, and values that showed system-wide polymer aggregation and individual polymer R_g reduction but not a phase transition using the system energy are shaded with yellow in Figure 6-4 and subsequent phase diagrams. We have also included plots of the number of polymer sites occupied by binding proteins where there is an increase in sites bound and a decrease in the variance of polymer site occupied with ϵ_{TT}. The number of sites bound and the variance in that number start to plateau

upon polymer/target condensation because the local concentration of proteins within reach of the polymer reaches a maximum in a the condensed phase.

We expect that the discrepancy of the polymer-target binding system showing aggregation and individual polymer R_g reduction at lower energies than the cumulant predicts a phase transition is the manifestation of two different transitions. As we increase the target-target attraction, bound targets create an effective interaction between polymers so that they behave as if they are in poor solvent. This leads to a first transition where the polymers aggregate and a condensed polymer droplet or small polymer gel forms. This is similar to results reported by Harmon et al. where a wider set of species concentrations resulted in gels than phase separated condensates [58]. This polymer phase separation does not manifest in the Binder cumulant because intra-target interactions are stronger than intra-polymer interactions and dominate the mean energy and specific heat of the system. As the intra-target attraction ϵ_{TT} is further increased, we see a second transition captured in the energy because the polymer clusters nucleate a condensed protein phase. This is consistent with the idea that condensed RNA can act as scaffolds for nucleating condensates [133].

By exploring our phase diagram in Figure 6-4, we show that the addition of the binding polymer leads to lower target solubilities for all target valencies studied for both $\Delta E_0 = -4k_{\mathrm{B}}T$ and $-6k_{\mathrm{B}}T$. The boundary for target phase separation in the absence of polymer is highlighted with a purple background. As the affinity of the specific binding sites increase, the phase boundary shifts down to weaker ϵ_{TT} and lower target concentration. Although less drastic, a similar shift is seen with target valency. As target valency increases, phase separation occurs at lower target attraction ϵ_{TT} and target concentration. This result matches well with previous simulations and experiments [83,58] and demonstrates that condensed droplets and similar behavior can be seen in much smaller systems than previously reported. This suggests that condensates can from droplets much smaller than can be seen through a microscope, and large condensates might grow through coalescence of these smaller droplets.

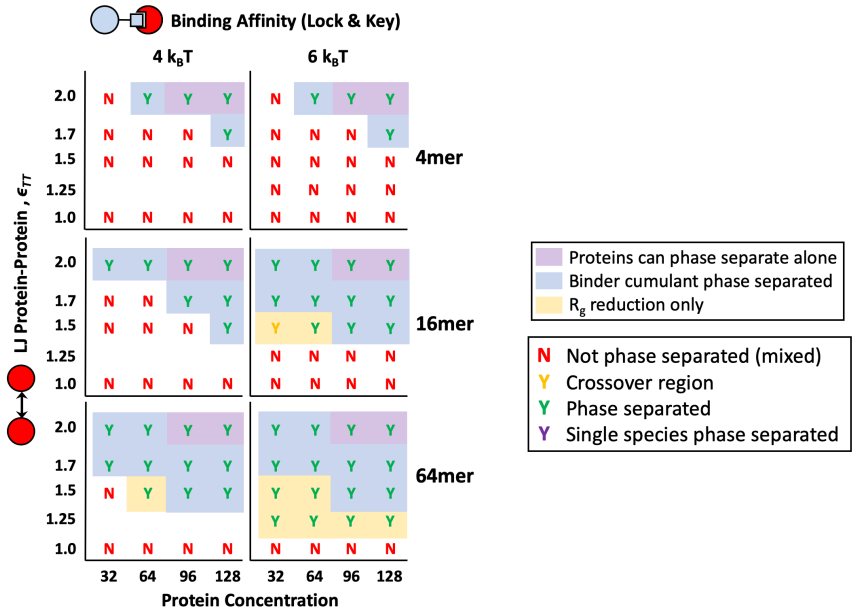


Figure 6-6: Phase diagrams for polymers of different of degrees of polymerization $N_\mathrm{p} = 4$ (Top Row), 16 (Middle Row), and 64 (Bottom Row) with reactive, specific binding affinities $\Delta E_0 = -4$ and $-6k_\mathrm{B}T$. All simulations are for divalent proteins targets and theta solvent for the polymer. Letter color coding and area shading have the same meanings as described in Figure 6-4.

We also explored phase diagrams through the lens of polymer valency or length. Figure 6-6 shows phase diagrams from divalent targets binding through reactive specific binding interactions with polymers. In these simulations, the number of polymer beads was kept at a constant concentration but the connectivity of the polymers were changed so that simulations contained sixteen 4mer polymers, four 16mer polymers, or one 64mer polymer. Consistent with previous results on increasing valency, we also saw phase separation at lower energies for polymers with higher degrees of polymerization [83,58]. The lowering of the phase separation boundary is more drastic for $\Delta E_0 = -6k_\mathrm{B}T$ when the polymer length is increased from $N_\mathrm{p} = 4$ to $N_\mathrm{p} = 16$ than when the polymer length is increased from $N_\mathrm{p} = 16$ to $N_\mathrm{p} = 64$ even though both scenarios reflect a 4X increase in length. As explored in our previous work, binding affinity of a linear multivalent polymer to a smaller protein target is significantly affected by the entropic cost of forming polymer loops when binding to the target divalently [86]. As degree of polymerization increases, longer loops are able to form, adding additional combinatorial entropy and increasing binding affinity. However, at some critical loop size, the maximum possible loop length is limited by the configurational entropy

loss of forming the loop and not limited by the polymer length. Polymers longer than the critical loop length see limited increases in avidity with longer degrees of polymerization. In these phase separation simulations, this results in a relatively small increase in binding avidity when N_p is increased from 16 to 64 and a correspondingly smaller change in the phase boundary than when N_p is increased from 4 to 16.

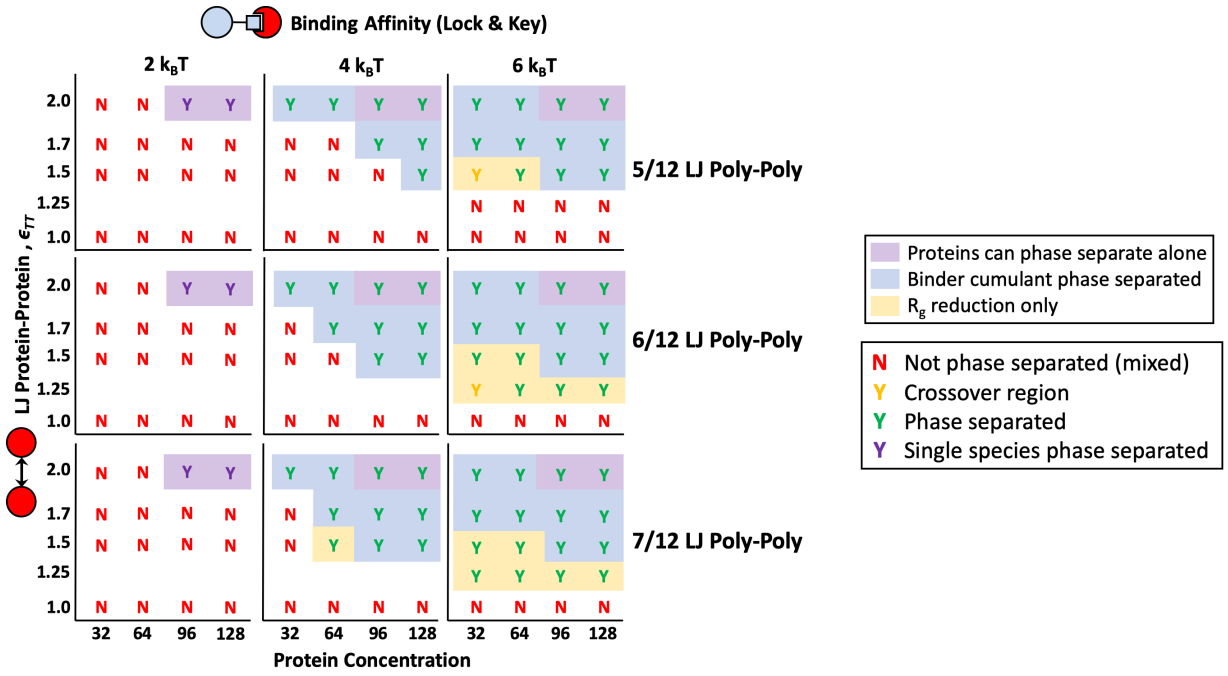


Figure 6-7: Phase diagram of simulations comparing the behavior of four 16mer polymers with lock and key binding to divalent protein targets in theta solvent ($\epsilon_{PP} = 5/12$) to two types of poor solvent ($\epsilon_{PP} = 6/12$ and $7/12$). Polymers do not phase separate on their own at any values of ϵ_{PP} tested. Lettering color codes and shading follow the same key as Figure 6-4, with "Y"s indicating "yes" phase separation occurred and "N"s representing "no" phase separation occurred. In poor solvent, phase separation occurs when polymers are mixed with binding targets at lower ϵ_{TT}s and protein concentrations than theta solvent.

6.3.2 Solvent Quality

Previous research showed that native intrinsically disordered protein (IDP) linkers can be swollen, theta condition freely random walk, or collapsed chains [58]. This same study showed that highly solvated or swollen polymers initiated gelation without phase separation and theta polymers led to phase separation with gelation [58]. Poor solvent quality is interesting because 30% of IDPs in the aforementioned study were found to have negative solvation volume, but a condensed polymer's

binding sites may be less available for target binding, or the polymer could phase separate on its own without the targets. Also, unlike good and theta solvent, polymers in poor solvent effectively have multivalent binding interactions with themselves and their binding protein targets. Using a Lennard-Jones potential and Brownian Dynamics, we show that there is a window of poor solvent conditions where polymers are soluble on their own, but can nucleate condensed droplets in the presence of binding targets.

Here, we again placed four 16mer polymers in a box with divalent targets with increasing concentration and ϵ_{TT}. If four 16mer polymers are simulated in a box by themselves, they precipitate out of solution at approximately $\epsilon_{\mathrm{PP}} = 9/12$, so we tested energies between $\epsilon_{\mathrm{PP}} = 5/12$ and $8/12$ where the 16mer polymers were collapsed but still soluble. Resultant phase diagrams for theta, and poor solvents with $\epsilon_{\mathrm{PP}} = 6/12$ and $7/12$ are shown in Figure 6-7.

In the case of poor solvent, even though collapsed polymers have less available volume for targets to bind in, polymers phase separate at lower target-target attraction and target concentrations than polymers in theta solvent. Results from our general collapsed polymer model demonstrate that a slight decrease in solvation of the polymer may trigger phase separation when the polymer is mixed with a corresponding binding target. This phenomenon can still happen when the decrease in solvent quality is not enough for the polymer to precipitate on its own. For example, changes in polymer sequence, post-translational modifications, or binding to a protein or small molecule that lowers the effective solvent quality for a multivalent polymer could cause phase separation at lower binding energies, concentrations, and target-target attractions. Our result showing how decreasing solvent quality lowers the phase boundary for systems with multiple species, builds on recent work from Martin et al. that showed that pure multivalent polymers phase separate at higher temperatures when they have more attractive self-interactions [141].

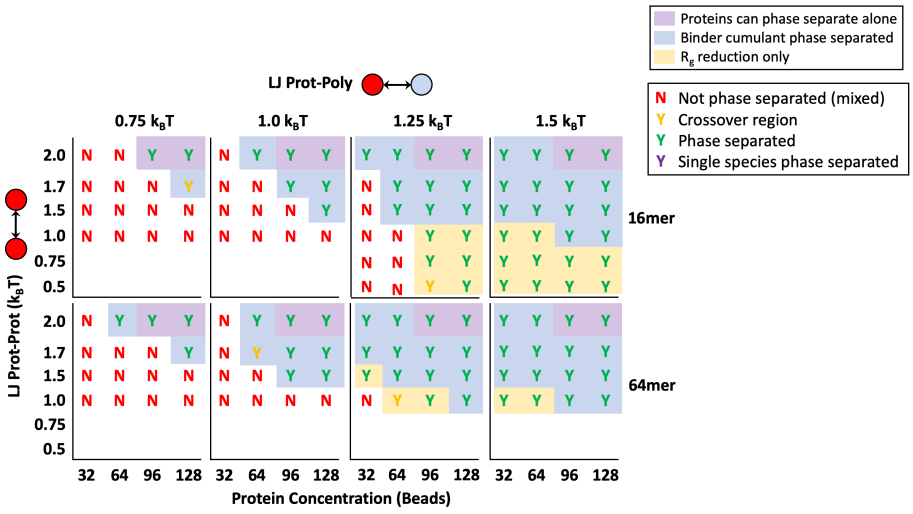


Figure 6-8: Phase diagram for proteins binding to polymers through a non-specific Lennard-Jones potential. Diagrams use the same key described in detail in Figure 6-4 where phase separated systems are marked with a “Y" for “yes" and not phase separated systems are marked with an “N" for “no". Results are shown for simulations with increasing polymer binding attractions ϵ_{TP} from left to right (columns) with four 16mer polymers (Top) and one 64mer polymer (Bottom).

6.3.3 Non-specific binding interactions

In addition to multivalency through valence-limited specific binding sites, we wanted to consider any differences in phase separation behavior associated with non-specific interactions such as charge or hydrophobicity. Non-specific interactions are commonly believed to add additional valency to lock-and-key binding through the promiscuous interactions IDRs on binding proteins such as FUS, TDP43, and hnRNPA1 [48,65]. To isolate the effects of non-specific interactions on nucleating a condensed phase, we turned off our reactive binding scheme and exclusively applied a more attractive Lennard-Jones attraction between the polymers and the targets ϵ_{TP}. We again placed four 16mer polymers in theta solvent with various target concentrations and ϵ_{TP}s. The phase behavior for $\epsilon_{\mathrm{TP}} = 0.75$, 1.0, 1.25, and $1.5k_{\mathrm{BT}}$ is shown in Figure 6-8. Results for a single longer polymer with degree of polymerization $N_{\mathrm{P}} = 64$ are also included for comparison.

With a generic Lennard-Jones potential between target proteins and the polymer, we see phase separation at much lower potential energies than with specific, reactive binding. For example, all target concentrations and solubilities as low as $\epsilon_{\mathrm{TT}} = 0.5k_{\mathrm{B}}T$showed droplet formation at $\epsilon_{\mathrm{TP}} = 1.5k_{\mathrm{BT}}$, but with the highest reactive binding energy $\Delta E_0 = -6k_{\mathrm{B}}T$, only targets with $\epsilon_{\mathrm{TT}} \geq$

$1.5k_BT$ formed droplets. This huge increase in the phase boundary energy is likely because the non-specific interaction has a much higher valency that is only limited by the maximum number of neighbors. The limited valence of lock-and-key type bonds creates competition for sites between bound and unbound protein neighbors, reducing the influence of the polymer on the targets. This reduction in binding due to competition for high affinity sites is discussed further in our previous work on multivalent binding site patterns [106]. The promiscuous nature and high effective valency of non-specific potentials reduces competition and allows polymers to interact with more targets simultaneously. This results in polymer-target phase separation occurring at much lower attractions. Therefore, non-specific interaction energies are a very sensitive dial for controlling phase separation and polymer modifications that change the non-specific interactions such as hydrophobicity or charge between the polymer and targets will have a more significant impact on the phase boundary than alterations to specific binding sites.

Still, valency and affinity of specific bonds can also change the phase boundary, although relatively large changes in these characteristics correspond to small changes in the LLPS boundary. Consequently, with a combination of non-specific and lock and key binding, the LLPS of multivalent polymers and targets can be precisely controlled. An example of using both non-specific and specific binding is shown in Figure 6-9. A small non-specific attraction $\epsilon_{TP} = 5/12k_BT$ was applied in addition to specific divalent binding with $\Delta E_0 = -2$, -4, and $-6k_BT$. The addition of the non-specific attraction provides access to phase separation at lower intra-target attractions, previously inaccessible through purely lock-and-key binding, or at such a low specific polymer attraction. The non-specific binding makes the phase separation energy barrier accessible through specific, valency limited bonds. This is helpful because the resultant phase boundary from specific bonds is less sensitive to changes in binding affinity or valence. We speculate that some reliance on the insensitivity of specific bonding could help biology to reduce the number of aberrant phase transitions caused by unintentional changes in binding sites.

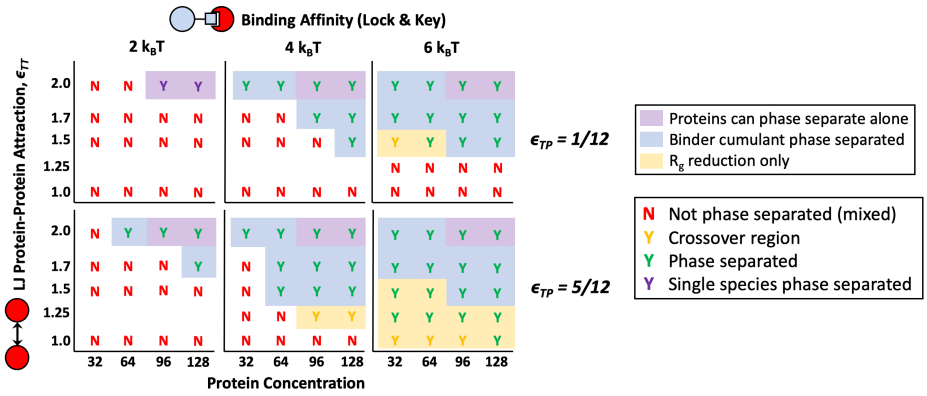


Figure 6-9: Phase diagram of targets and polymer with both nonspecific binding affinity $\epsilon_{\mathrm{TP}} = 5/12$ and specific lock and key binding (Bottom Row) compared with polymers that have lock and key binding but almost no non-specific attraction to the targets $\epsilon_{\mathrm{TP}} = 5/12$ (Top Row). Protein targets in these simulations are divalent. Diagrams use the same key described in detail in Figure 6-4 where phase separated systems are marked with a "Y" for "yes" and not phase separated systems are marked with an "N" for "no". Results are shown for simulations with specific polymer binding attractions $\Delta E_0 = -2$, -4, and $-6k_{\mathrm{BT}}$ from left to right (columns) with four 16mer polymers.

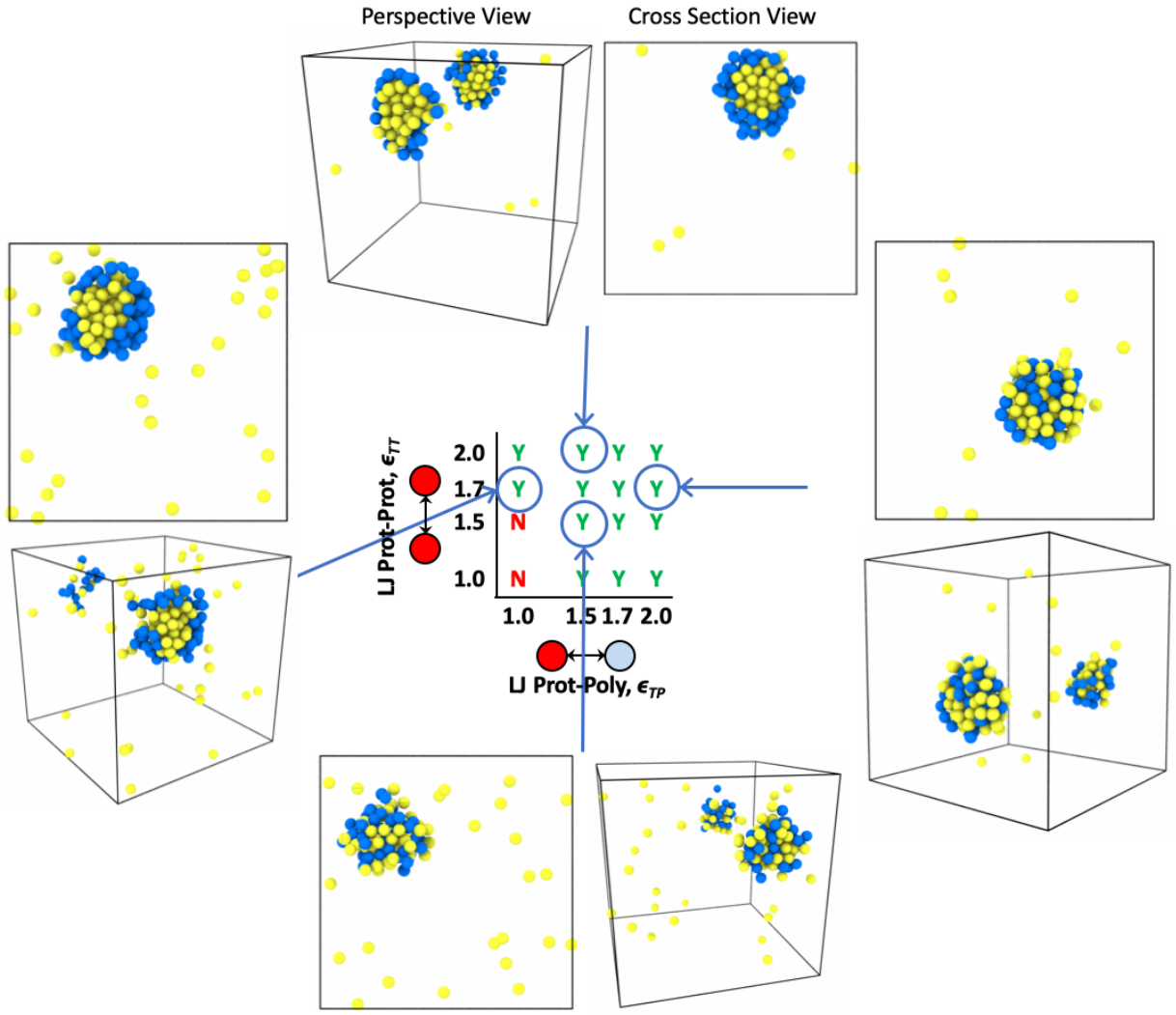


Figure 6-10: Simulation renderings depicting ordered and mixed droplets with a cross section view through the middle of the droplet and a perspective view showing the inside and outside of the droplet. Polymer beads are blue and protein target beads are yellow. Results shown are for simulations with non-specific binding to four 16mer polymers in theta solvent and 96 target binding proteins. Note that the x-axis on this phase diagram is now protein-polymer affinity ϵ_{TP} in units of k_{BT} and the y-axis is still the intra-protein attraction ϵ_{TT} seen on previous phase diagrams also in $k_{\mathrm{B}}T$. By moving vertically down the phase diagram from $\epsilon_{\mathrm{TT}} = 2.0$ to 1.5 the droplet morphology goes from ordered to mixed due to changes in surface tension of the liquid protein phase. The droplet also goes from ordered to mixed as we move from left to right across the phase diagram from $\epsilon_{\mathrm{TP}} = 1.0$ to 2.0 due to increasingly favorable protein-polymer interfacial energy χ.

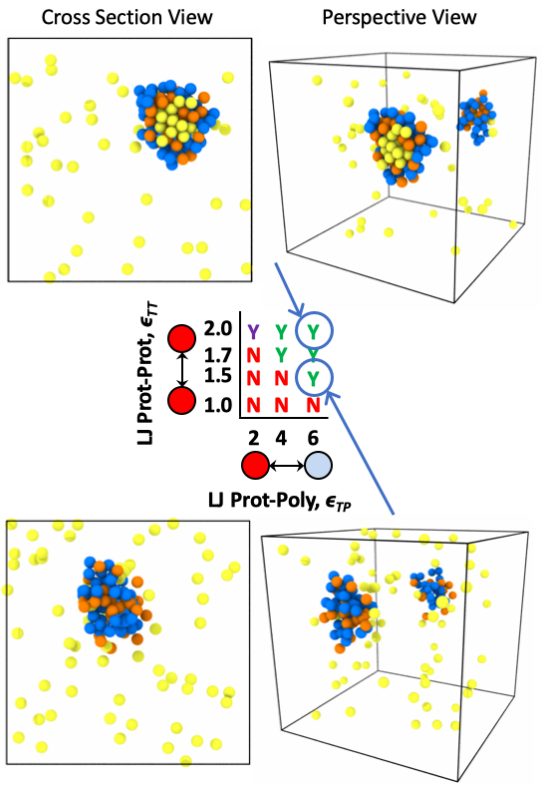


Figure 6-11: Simulation renderings depicting ordered and mixed droplets with a cross section view through the middle of the droplet and a perspective view showing the inside and outside of the droplet. Polymer beads are blue, unbound protein targets are yellow, and bound protein targets are orange. Results shown are for simulations with specific binding to four 16mer polymers in theta solvent and 96 divalent target binding proteins. Note that the x-axis on this phase diagram is now protein-polymer affinity ΔE_0 in units of $k_{\mathrm{B}}\mathrm{T}$ and the y-axis is still the intra-protein attraction ϵ_{TT} in $k_{\mathrm{B}}T$. By moving vertically down the phase diagram from $\epsilon_{\mathrm{TT}} = 2.0$ to 1.5 we also see the droplet morphology change from ordered to mixed due to changes in surface tension of the liquid protein phase.

6.3.4 Condensed phase organization

Biocondensates often show microphase separation within the condensed phase [49,50,142,143]. This disorder to order transition is a well known phenomenon in polymer physics with block copolymers where self-assembly can be controlled by the interaction (χ) between the two polymer block types. We expected the same to be true for polymer-target assemblies in LLPS. When attraction between binding proteins is higher than attraction to the polymer, condensates should undergo microphase separation where polymers surround a condensed target phase. When attraction between targets is similar to target-polymer attraction, the condensed phase will remain mixed. If the targets are

highly attracted to the polymer, and not attracted to themselves, we might also see the case where the polymer is condensed in the center of the droplet with targets decorating the outside of the condensed phase.

First, we can explore inducing droplet order by changing the self-interactions of a single species such as the binding proteins. Using self-interactions or solvation volume can be thought of as changing the surface tension of the liquid target phase. Changes in droplet organization due to surface tension or interactions of polymers with solvent were previously explored in a 4 component system with 3 types of equal size binding polymers and solvent. They found that swollen multivalent polymers could lower the surface tension of similar less-solvated polymers and induce a shell-core structure seen in some RNP bodies [142]. In Figure 6-10 we show that a similar ordering can be induced through surface tension in our asymmetric valency/size 3 component system, where the 3 components in our system are multivalent polymer, smaller binding proteins, and implicit solvent. If the target-polymer attraction is held constant, the condensates go through a demixing transition as the target-target attraction increases from $\epsilon_{TP} = 1.0 k_{\mathrm{B}}T$ to $2.0 k_{\mathrm{B}}T$, moving in the vertical direction up the phase diagram in Figure 6-10. Similar changes in droplet organization controlled by the excluded volume of the binding proteins is also seen for the specific lock-and-key binding polymers in Figure 6-11.

In addition to ordering due to surface tension, we also see demixing caused by changes in non-specific binding affinity between the polymers and targets, more akin to inducing order/disorder through modulating the χ parameter. Looking at droplet order in Figure 6-10, while moving across the phase diagram from left to right, it is clear in the simulation renderings that if target-target attraction is held constant, the system goes through and order-to-disorder transition as the target-polymer attraction increases from $\epsilon_{TP} = 1.0 k_{\mathrm{B}}T$ to $2.0 k_{\mathrm{B}}T$. In this case, when the bonding energy or attraction between the targets and polymer are very strong, the targets prefer to associate with the polymer equal to or more than themselves and subsequently decorate the polymer as much as possible. This results in the polymer becoming fully mixed with the binding targets. In contrast, when the target attraction to the polymer is less than the amount they like to associate with themselves, the targets prefer to surround themselves with only target neighbors and the polymer is pushed to the outside of the droplet. This means that modifications made to IDRs that impact

attraction to RNA and not just excluded solvent volume can change droplet ordering.

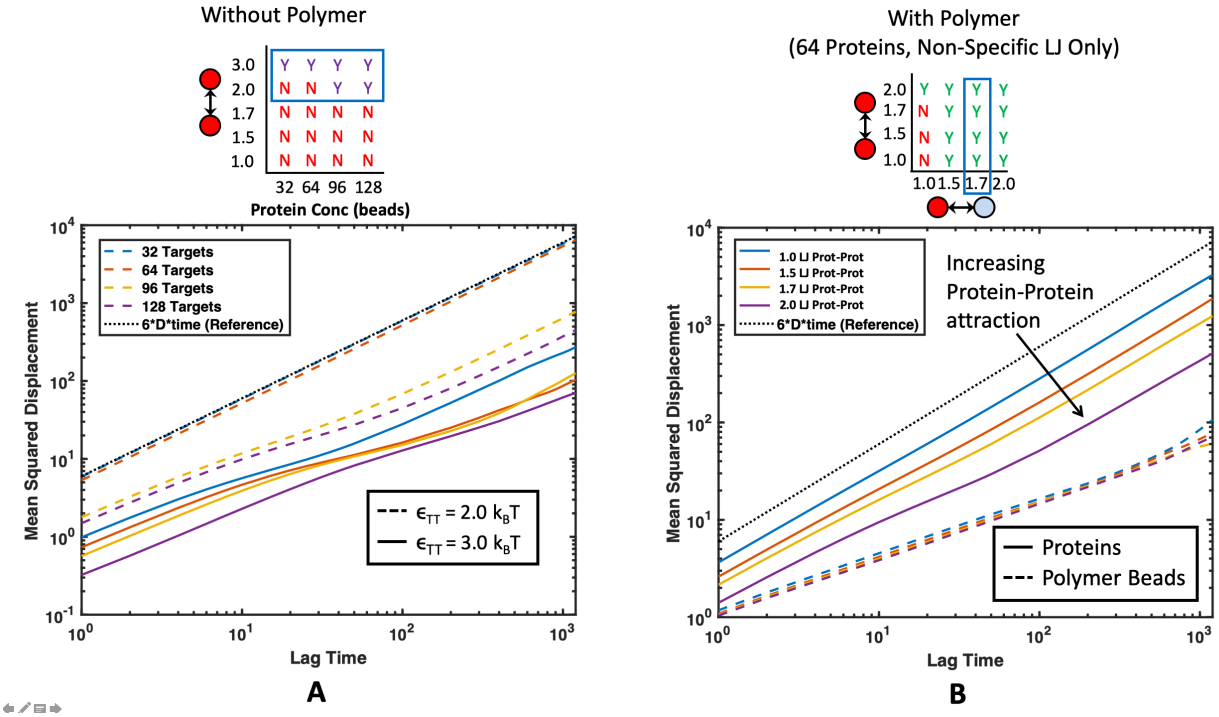


Figure 6-12: Average mean squared displacement (MSD) of all proteins for several different protein-protein affinities. Corresponding regions of the phase diagrams are highlighted with a blue rectangle above each plot. The black dotted line (·····) represents normal 3-D Brownian diffusion. (A) MSD of pure proteins without polymers present. Colors represent different protein concentrations and line pattern represents intra-protein affinity. Not phase separated proteins diffuse with normal Brownian motion whereas phase separated proteins see much slower diffusion rates. Higher ϵ_{TT} leads to lower MSD and slower protein diffusion. (B) MSD for 64 proteins interacting with 16mer polymers through non-specific attraction at $\epsilon_{TP} = 1.7$. Color corresponds to ϵ_{TT}. Average MSD for all proteins is shown with a solid line (——) and average MSD for all polymer beads is shown with a dashed line (- - -). In the presence of polymers, higher intra-protein attraction still leads to slower protein diffusion times.

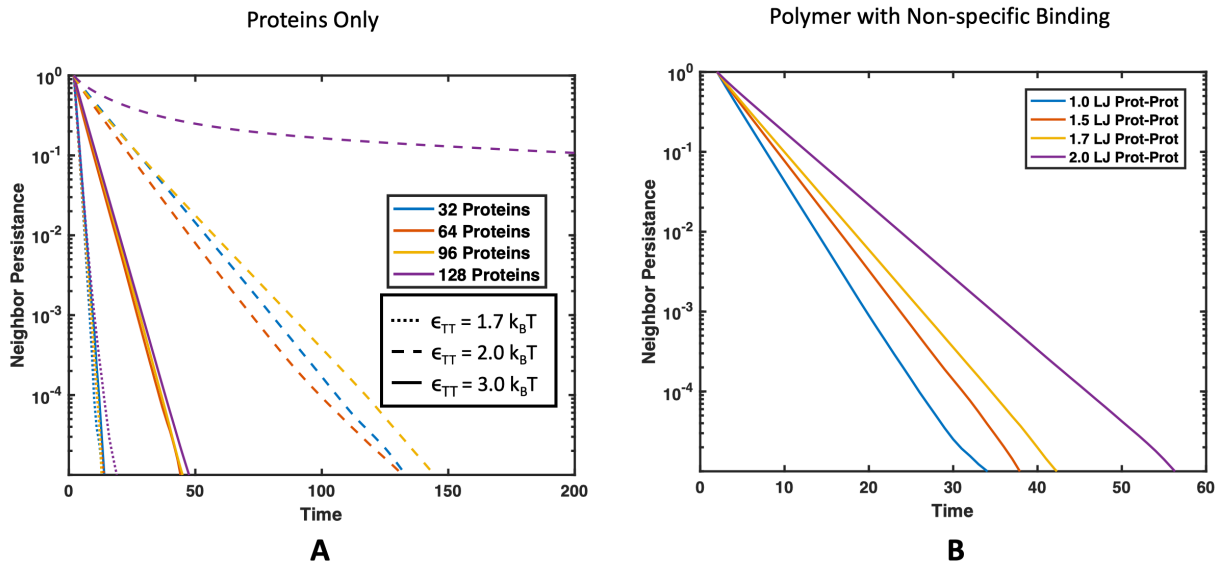


Figure 6-13: Neighbor persistence of a binding protein (A) without a polymer present and (B) with a polymer present that interacts through non-specific interactions. In both cases the time proteins spent with the same neighbors is lengthened as the intra-target attraction ϵ_{TT} increases. (A) Line color corresponds to protein concentration and line pattern denotes ϵ_{TT}. (B) Line color denotes ϵ_{TT} in simulations with 64 proteins interacting with four 16mer polymers in theta solvent with $\epsilon_{\mathrm{TP}} = 1.7 k_{\mathrm{B}}T$.

6.3.5 Kinetics inside the droplet

Our simulation methods also allow us to study the diffusion and dynamics of condensed phase species which are important to understand biocondensate function and diseases associated with altered droplet dynamics [118,66,144]. To examine the mobility of binding proteins and polymers in droplets we looked at both the mean squared displacement (MSD) and the neighbor persistence. We calculated neighbor persistence as the average number of neighbors that remained the same over a time interval, where beads were considered neighbors when their centers were within $2.5a$. The rate at which the neighbor persistence goes to zero provides a measure of how quickly the targets are exchanging with the dilute supernatant phase. A high neighbor persistence or slow decay rate means that a target in the droplet maintains many of its neighbors over a long period of time, suggesting a solid-like phase. A fast decay in neighbors to zero signifies that, in a short amount of time, the target became surrounded by an entirely new set of neighbors or completely left the droplet, signifying a more liquid-like droplet with a fast exchange rate with the outside environment.

We compared the MSD and neighbor persistence of the target binding proteins across increasing intra-protein interactions ϵ_{TT} in Figure 6-12 and Figure 6-13, respectively. Not phase separated systems show normal three dimensional diffusion, while systems that formed condensed droplets have decreased MSD consistent with transformation from a gas to liquid phase. More interestingly, we see a slow down in the diffusion and longer neighbor persistence time as ϵ_{TT} increases from 2.0 to $3.0k_BT$. At $\epsilon_{TT} = 3.0k_BT$ the system with 128 targets appears to be almost solid-like with some neighbors maintained longer than the evaluated time interval in Figure 6-13A. The same decrease in diffusion and increase in neighbor persistence with higher target-target attraction is also seen in the presence of binding polymers as shown in Figure 6-12B and 6-13B. Although only non-specific binding polymers are shown in Figure 6-12B, trends are similar for lock-and-key binding polymers. These results match well with experimental evidence that when RNA-binding proteins more attracted to themselves this result in slower protein diffusion times and more solid-like droplets [140]. Therefore, any modifications that make these RNA-binding proteins more attractive to themselves such as additional hydrophobic residues in the IDRs could lead to solidification of droplets. The increase in neighbor persistence could also correspond to lower diffusion limited reaction rates because neighboring proteins exchange more slowly.

Li *et al.* showed experimentally that valency and binding affinity of molecules inside droplets inversely correlate with FRAP recovery kinetics, which is exactly what we see in simulation [83]. Here, we provide further evidence of these results through simulation where increasing the affinity of the targets to the polymer decreased the MSD of targets and increased their neighbor persistence. This effect is demonstrated with both non-specific attraction to the polymer and specific valence-limited binding in a slow down in the MSD in Figure 6-14 and longer neighbor persistence in Figure 6-15. The slow down in dynamics caused by higher affinity binding to the polymer is concentration dependent. Unsurprisingly, a higher ratio of binding proteins to polymer results in the polymer having less influence over the droplet dynamics. Comparing the diffusion time and neighbor persistence of droplets with and without polymers in Figure 6-16, we find that our simulations also match previous experiments that showed that dynamics of pure protein droplets are slowed upon the addition of a long RNA polymer [140]. While we did not see any increase in protein diffusion upon the addition of binding polymers as seen in Maharana et al. this may be

because we were not deep enough into the energy regime where targets phase separate alone such as proteins that formed a solid-like structure such as at $\epsilon_{TT} = 3.0 k_B T$ or because the polymers we simulated were too long [133,140]. Exploring the conditions under which polymers can speed the diffusion of proteins in condensed droplets would be an interesting avenue for future work. Here, we consider beads of equal sizes, but if we used different sizes to disrupt the packing, we might see more significant changes in diffusion.

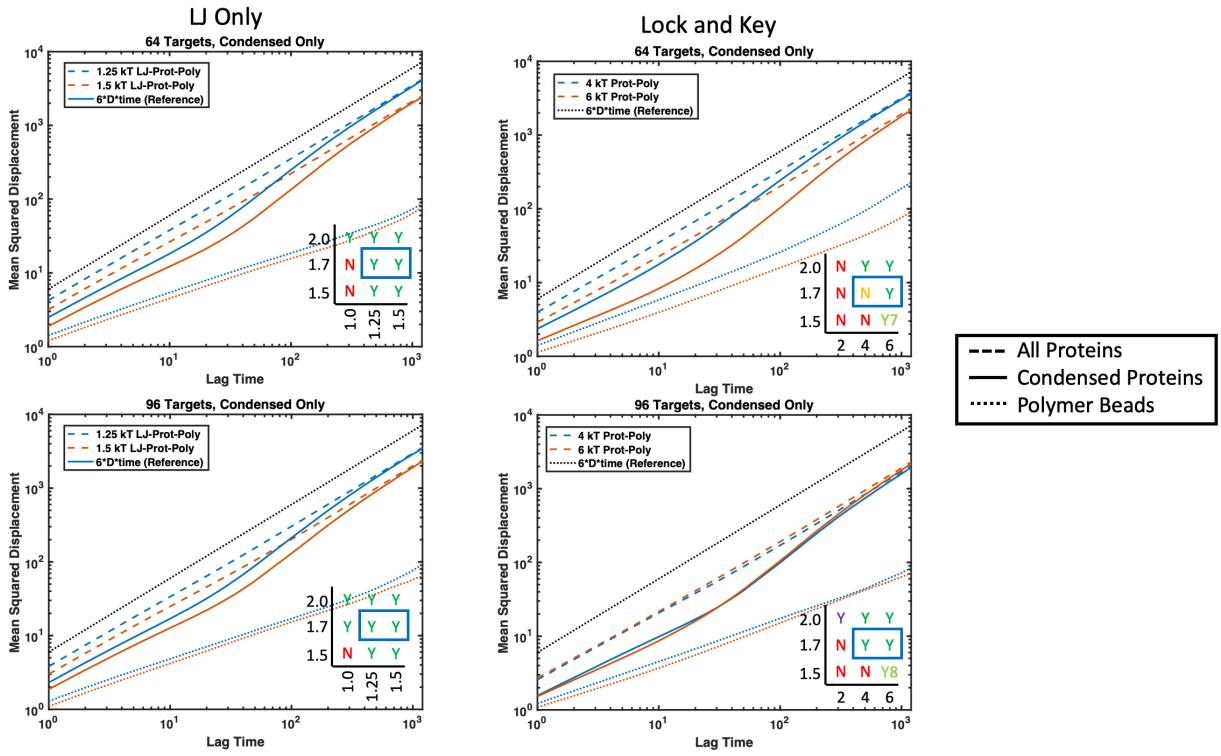


Figure 6-14: Average MSD for proteins with $\epsilon_{TT} = 1.7 k_B T$ interacting with four 16mer polymers in theta solvent. Dotted black line (·····) represents normal Brownian diffusion, dashed lines (- - -) represent the average MSD over all proteins in the simulation, solid lines (——) represent the average MSD over all proteins that started with at least one neighbor at the beginning of the time interval, and the colored dotted (·····) lines represent the average MSD over all polymer beads in the simulation. Colors represent two attraction energies between protein targets and polymers with blue denoting lower affinity than orange. Each plot contains the corresponding phase diagrams with the plotted regions highlighted with a blue rectangle. Cases plotted include (A) non-specific binding polymer with 64 targets and $\epsilon_{TP} = 1.25$ and $1.5 k_B T$, (B) non-specific binding polymer with 96 targets and $\epsilon_{TP} = 1.25$ and $1.5 k_B T$, (C) specific binding polymer with 64 divalent targets and $\Delta E_0 = -4$ and $-6 k_B T$, and (D) specific binding polymer with 96 divalent targets and $\Delta E_0 = -4$ and $-6 k_B T$. Protein diffusion slows with increasing protein-polymer attraction, but the polymer has less influence on droplet dynamics when the ratio of proteins to polymer is high.

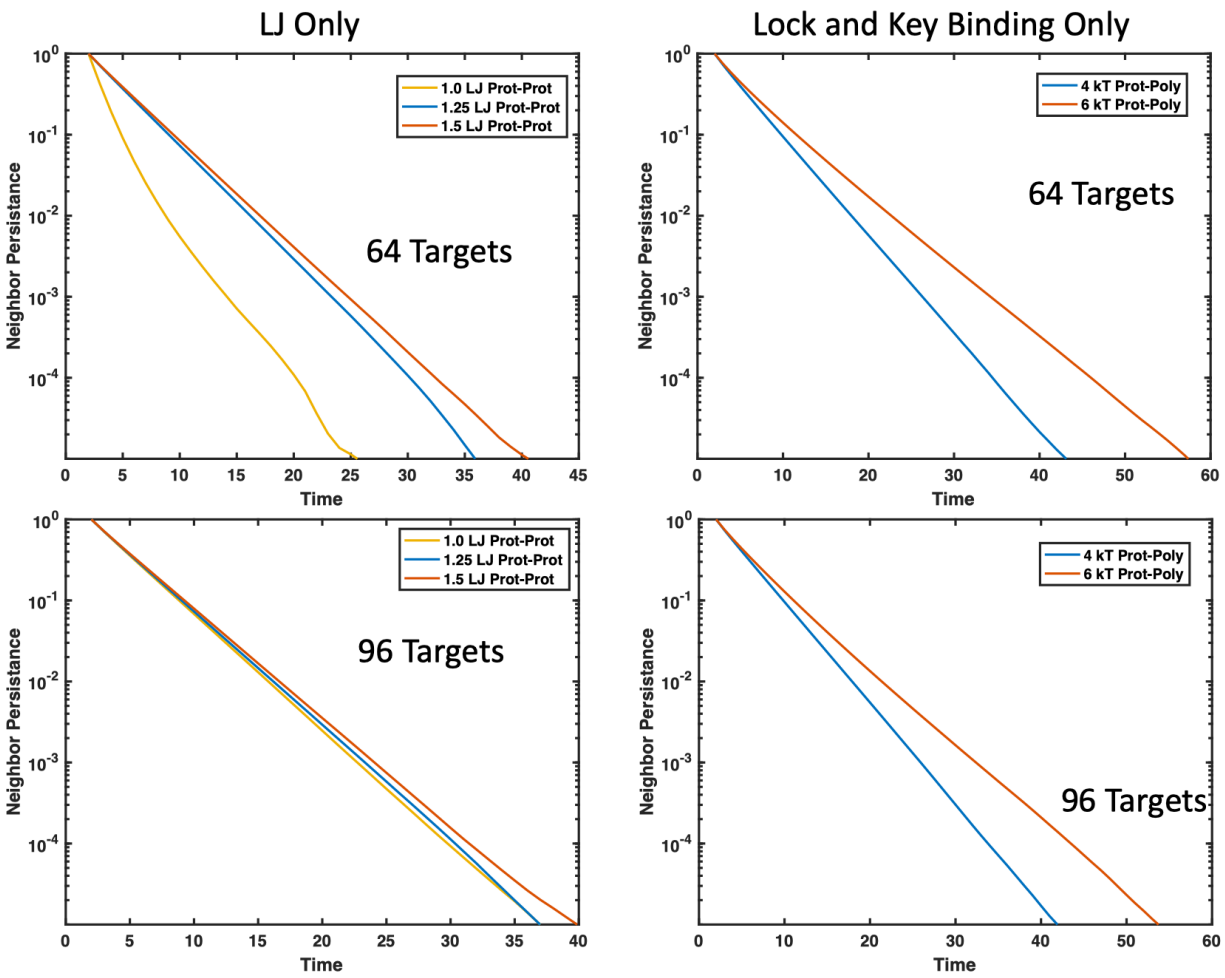


Figure 6-15: Average time proteins spend with the same neighbors normalized by the average number of initial neighbors. Faster decays to zero indicate a more liquid-like droplet where proteins can move through or exit the droplet freely. Increasing binding affinity to the polymer results in longer protein neighbor persistence. Results are shown for the same cases as Figure 6-14. (A) Non-specific binding polymer with 64 targets and $\epsilon_{\mathrm{TP}} = 1.25$ and $1.5 k_{\mathrm{B}}T$, (B) non-specific binding polymer with 96 targets and $\epsilon_{\mathrm{TP}} = 1.25$ and $1.5 k_{\mathrm{B}}T$, (C) specific binding polymer with 64 divalent targets and $\Delta E_0 = -4$ and $-6 k_{\mathrm{B}}T$, and (D) specific binding polymer with 96 divalent targets and $\Delta E_0 = -4$ and $-6 k_{\mathrm{B}}T$.

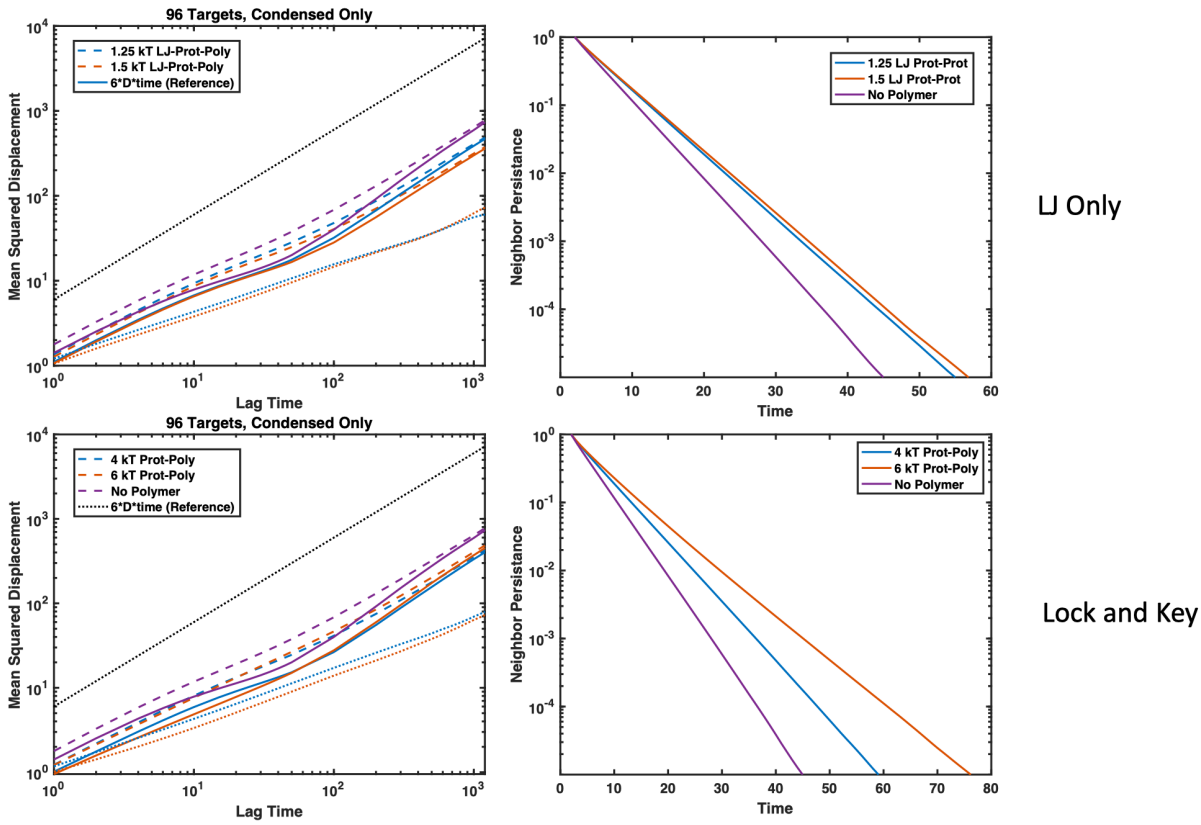


Figure 6-16: Average MSD and neighbor persistence of 96 proteins with $\epsilon_{TT} = 2.0 k_B T$ compared with the dynamics of a pure protein droplet (purple). (Top Row) Proteins experiencing non-specific attraction with the polymers $\epsilon_{TP} = 1.25$ (blue) and $1.5 k_B T$ (orange). (Bottom Row) Divalent proteins experiencing specific attraction with the polymers $\Delta E_0 = -4$ (blue) and $-6 k_B T$ (orange). In MSD plots (Left Column), dotted black line (·····) is reference for normal Brownian diffusion, dashed lines (- - -) represent the average MSD over all proteins in the simulation, solid lines (——) are the average MSD over all proteins that started with at least one neighbor at the beginning of the time interval, and the colored dotted (·····) lines represent the average MSD over all polymer beads in the simulation.

6.4 Conclusion

Understanding how the body can alter and control the phase transitions of biocondensates and their dynamics is crucial to understanding pathological aggregation. Here, we present simulations and explore the aggregation and diffusion kinetics of smaller species that bind multivalently to longer polymers. This system directly resembles a coarse-grain model of smaller proteins binding to linear polymers such as nucleic acids or other proteins with multiple repetitive binding sites such as RNP bodies.

Despite the different geometries, these results seem to align well with systems of two similarly size linear multivalent polymers and their phase transitions [83,58]. We similarly found that the addition of a multivalently binding polymers can lower the phase boundary for a protein, and that increasing protein valency and binding affinity also lower the phase boundary. To add on previous investigations, we consider the differences in non-specific binding that might come from IDRs versus the specific valency limited binding that comes with RNA recognition motifs. We show that changes in the affinity of non-specific interactions can cause more drastic changes in the phase boundary than valence-limited lock-and-key type bonds. Together, they can be used to carefully tune the phase boundary.

Next, we showed that both surface tension and binding affinity could be used to tune droplet order in a system of only three components. When proteins had higher attraction to multivalent polymers, droplets remained mixed with proteins and polymers distributed throughout. When proteins had higher attraction to themselves than to the multivalent polymer, we were able to recreate systems of concentric droplets. Pure proteins formed a central core, while the polymers were pushed into an outer shell. This could have implications for understanding how changes in polymer-protein binding can impact biocondensate function.

Last, we found that increasing attraction between targets themselves and between targets and polymers can slow the diffusion of targets within condensates and make them more solid-like, consistent with previous experimental results [140,83]. The attraction to the polymer has a greater effect on target dynamics in droplets with lower target concentrations, but after nucleation and growth of a condensed target phase where targets outnumber polymer binding sites, target-target attractions dominate the droplet dynamics. This suggests that changes in the non-specific attraction between binding proteins themselves can induce droplets to be more liquid or more solid-like in addition to polymer-binding interactions. Changes in dynamics could have big implications for the reversibility of condensate formation and for reaction rates inside condensates, leading to clear implications for diseases related to dysregulation of liquid-liquid phase separation such as ALS.

While more research needs to be done on specific systems and systems with more than two components we hope that the results presented contribute to the understanding and control of biocondensates and their associated diseases.

This Page Intentionally Left Blank

Chapter 7

Summary and Outlook

7.1 Summary of thesis

This thesis explored the rational design of linear multivalent polymers as both inhibitors of pathogens and architects of biocondensates. We sought to relate structural design variables of these polymers to their resultant function using simulation and theory. We hope that this work will narrow the design space for experimentalists and help to understand native biological systems. In Chapter 2, we developed a Brownian Dynamics bead-spring model for our multivalent polymer-target binding system that included excluded volume effects and reactive binding. We used this general model to explore the effect of degree of polymerization on multivalent binding affinity to a small globular protein target or colloid in Chapter 3. While holding the concentration of binding sites constant, we compared the binding of polymers of different connectivity (ie: 64 monomers, 32 dimers, 4 16mers) and showed that while multivalency drastically increases binding affinity to dilute targets, the improvement plateaus with increasing length. We showed that this plateau in binding affinity is likely due to the entropic loss of forming large polymer loops; when a polymer binds twice to a target, it has to form an entropically unfavorable loop.

Using a Poland-Scheraga type model for polymer-loop entropy we show that the binding free energy (and correspondingly the binding avidity) of multivalent polymers follows Equation 3.2 and plateaus at long degrees of polymerization because there is a maximum critical loop size. Long polymers can theoretically add additional binding configurations, but if the resultant polymer loops

cost more entropy than they gain in binding enthalpy, these loops won't form. Therefore, above the critical loop length, longer polymers do not add accessible longer loop configurations and will see very minimal increases in binding affinity with increasing length. Anything that makes polymer loops more energetically costly to form such as improved solvent quality or stiffer chains will lower the binding affinity to a small target. While this effect is true of dilute targets, longer polymers may have additional benefits when interacting with many targets simultaneously. In high target concentrations, long polymers can more easily induce phase separation than short ones because condensing multiple small polymers into a droplet costs additional translational entropy than the configurational entropy cost of compacting a single long polymer chain. Therefore, if the goal is to bind dilute targets, using polymers above a the critical loop length is unnecessary, but if the goal is to phase separate many target proteins, longer polymers can shift the phase boundary to lower concentrations and protein solubilities.

In Chapter 4, we consider the possibility of multifunctional polymers and how patterns of different binding site affinities along the polymer chain can control binding to a target. We show that again, behavior is dependent on binding target concentrations and the cost of forming loops within the multivalent polymer. At low target concentrations, when there isn't competition for binding sites, block copolymers bind with higher affinity than alternating copolymers. This is because binding targets want to minimize loops by binding to two polymer beads next to each other, and block copolymers provide the highest concentration of high-affinity binding sites. In contrast, when there is high competition for polymer binding sites, the high affinity sites are excluded by already bound targets, and free targets are pushed to interact with low affinity sites. This makes alternating polymers higher affinity because they are better at sharing their high-affinity sites along the polymer chain. We show that this critical competition occurs approximately when the target binding sites equal the number of polymer binding sites. Therefore, binding site arrangement for polymer designs with heterogeneous sites should be tuned to expected target concentrations, or designs should be considered with a mixture of both blocky and alternating regions such as a random copolymer.

Next, we considered wormlike chains — polymers that are stiff and extended, instead of random coils. Chains could be more rod-like due to their dense sidechains, high charge density, or complexation with another chain such as in the case of double stranded DNA. For stiff chains, binding

multiple times to a small target involves an additional enthalpic cost of bending. Because the cost of bending is very high for stiff chains, only very small polymer loops can form. This lowers the effective valency of rod-like polymers and makes them significantly lower binding affinity for small targets. This effect is not dependent on target concentration, and at high target concentrations, results in phase separation only at higher concentrations and lower solvent quality than for flexible random coil polymers. We suspect that this could affect biocondensates where changes in stiffness could induce or prevent phase separation. We also suggest that large differences in multivalent polymer stiffness may play a role in experimental systems where biocondensates were shown to selectively enrich themselves in single stranded DNA and exclude double stranded DNA [134].

Finally, in Chapter 6 we look directly at the phase boundary and dynamics of phase separation and compare binding through non-specific and specific binding sites. This work is relevant to membraneless organelles whose binding proteins often have both specific binding sites and intrinsically disordered protein regions that add non-specific binding sites. We show that non-specific binding can cause phase separation at lower energies, likely because non-specific binding sites can interact with many targets simultaneously, effectively creating a much higher valency. Therefore, these two types of interactions give the body two energetic knobs for controlling the phase boundary. Non-specific binding is highly sensitive to changes in energy, and so small changes in non-specific affinity can be used to set the overall system energy close to the phase boundary. Specific-binding is much more robust to changes in energy with relatively large changes in energy resulting in small changes to the phase boundary. Specific bonds can be applied on top of the non-specific attraction to make small changes in the phase boundary, giving both native and synthetic polymers precise control of phase separation. We suspect this also means that biological systems are less robust to changes in non-specific binding affinities than changes in specific, valence-limited binding sites, and therefore, subtle changes in intrinsically disordered protein sequences could result in large changes in phase separation, aberrant biocondensates, and disease.

Also in Chapter 6, we explore how increased polymer length, binding affinity, and valence can lower the phase boundary in our asymmetrically sized binding system, matching results previously reported in systems of two binding polymer species. Furthermore, we explore the effects of poor solvent and how even in systems of only three components (polymer, binding protein, and implicit

solvent) we see microphase separation in the droplet. Furthermore, both polymer-target binding affinity and surface tension with the solvent can control order within small droplets. We go on to investigate changes in droplet dynamics associated with changing polymer interactions. Our results show that increasing attraction between targets (lower protein solubilities) and increasing binding affinity between targets and polymers resulted in slower diffusion within the droplet and slower neighbor exchange. Although there isn't a direct correlation between diffusion and droplet crystallization because faster diffusion corresponds to both faster crystal growth and slower crystal nucleation rates, we were able to show that phase separated protein droplets had lower density when mixed with polymers. We suspect that since liquids are lower density than solids, small decreases in the density of droplets due to multivalent polymers could explain why some RNA-binding proteins droplets are more liquid-like and crystallization is suppressed when RNA is present [133]. If we ran simulations where polymer and target beads were not the same size, we might also be able to see that polymers disrupt droplet crystal structure.

Our general coarse-grained simulation platform has allowed us to apply our results to several different scenarios of multivalent polymer binding. We hope that the results presented in this thesis research will help guide the design of a next generation of polymeric pathogen inhibitors and aid in understanding causes and treatments for neurodegenerative disease associated with biocondensates.

7.2 Open questions and future work

As in any research, the scope of this thesis was limited by time, but we have several additional projects in mind for future work.

7.2.1 Different polymer geometries

As mentioned previously, there are many multivalent geometries that could have been considered — too many to explore in a single thesis! If we had more time we would be interested in either adding more detailed geometry to our target, perhaps by making the binding sites in specific patches on its surface instead of isotropic. It would also be interesting to consider the differences in using linear versus brush-like polymers to inhibit pathogens because mucins have a bottlebrush architecture

[24,145]. Other geometries such as flexible multivalent sheets may provide additional steric shielding to prevent target binding to cells [14,30]. Sheets also could have lower configurational entropy loss upon binding than polymers because they are restricted to movement in only two dimensions. This could improve their binding affinity and it would be interesting how design parameters alter their binding affinity.

7.2.2 Machine learning on patterns

In the realm of heterogeneous polymers or considering the effect of intrinsically disordered protein sequence, it could be helpful to use machine learning tools to understand the nuances of how sequence is tied to binding affinity. The sequence of proteins that phase separate in biocondensates are of particular interest, and there are several recent studies that discuss linking amino acid sequences to phase behavior [146,147]. If neural networks could accurately predict the binding affinity of polymers based on their sequence, this would speed up the time to evaluate a polymer sequence using simulation from hours or days to seconds. Comparing thousands of sequences simultaneously might clarify details of sequence-dependent phase separation. For multivalent polymeric inhibitors, sequences could be optimized for multiple objectives such as binding to two different target species but not a third or binding to a target that has multiple binding site types. During my thesis research, we dabbled in machine learning by using binding affinity sequences as input to output a binding affinity or average time bound as shown in Figure 7-1. By learning on hundreds of 8mer random polymer sequences, we were able to get relatively good predictions for overall polymer binding affinity with a fully connected feed forward, back propagation neural network. Unfortunately, the network appeared to base its predictions primarily on average site affinity and couldn't identify subtle changes in pattern. Therefore, in future work, we would suggest starting with a convolutional neural net or another method that can intrinsically capture sequence or patterns.

7.2.3 Polymer-polymer binding

It would also be interesting to consider the binding of two multivalent polymers because some biocondensates contain two linear multivalent species such as those involved in coordinating the

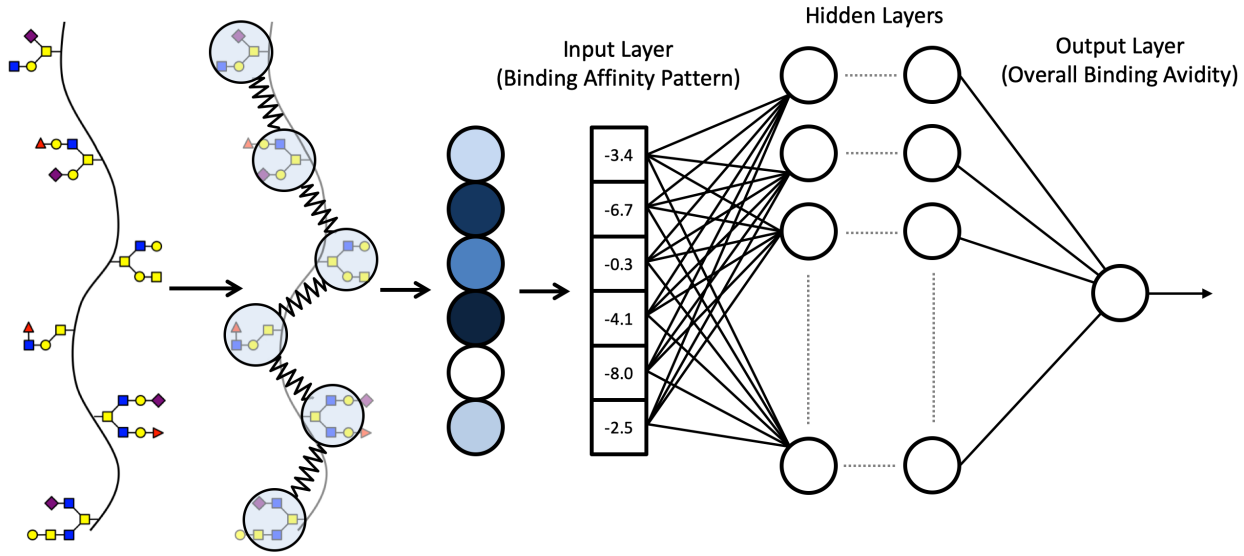


Figure 7-1: Binding site affinity patterns appear to lend themselves well to neural networks which could take the binding site affinity sequence as input and give the overall polymer binding affinity as output.

construction of actin filaments such as Nephrin, Nck and N-WASP[83,50,148]. Repeat expansion disorders that cause Huntington disease, muscular distrophy, and ALS can also be modeled as two multivalent polymers which only phase separate after a critical repeat number or valency is reached [64]. Therefore, it would be interesting to consider how the binding of two polymers with similar valency is different than the binding of a polymer to a small target, as considered in this thesis. For example, in a two polymer system, how does the ratio of the two polymer stiffnesses change binding or phase separation? In this case, we suspect that two rod-like polymers would have stronger binding if their sites are perfectly matched, but what about if there is slight mismatch? In Chapter 6, we were unable to see hemispherical phase separation where one species is enriched in the left hemisphere and one species is enriched in the right hemisphere as shown in Figure 7-2. Maybe we could capture this morphology, seen in Cajal bodies, with two polymers [49]. To test this, we placed ten 16mer polymers that bound themselves through four evenly spaced reactive binding sites in a box with ten 16mer polymers with no specific binding sites. All of the polymers were in the same quality of poor solvent so that they would phase separate without the specific binding forces. Although these results are only preliminary, we did see behavior of small hemispherically segregated droplets forming initially before aggregating into one large ordered droplet as shown in Figure 7-2.

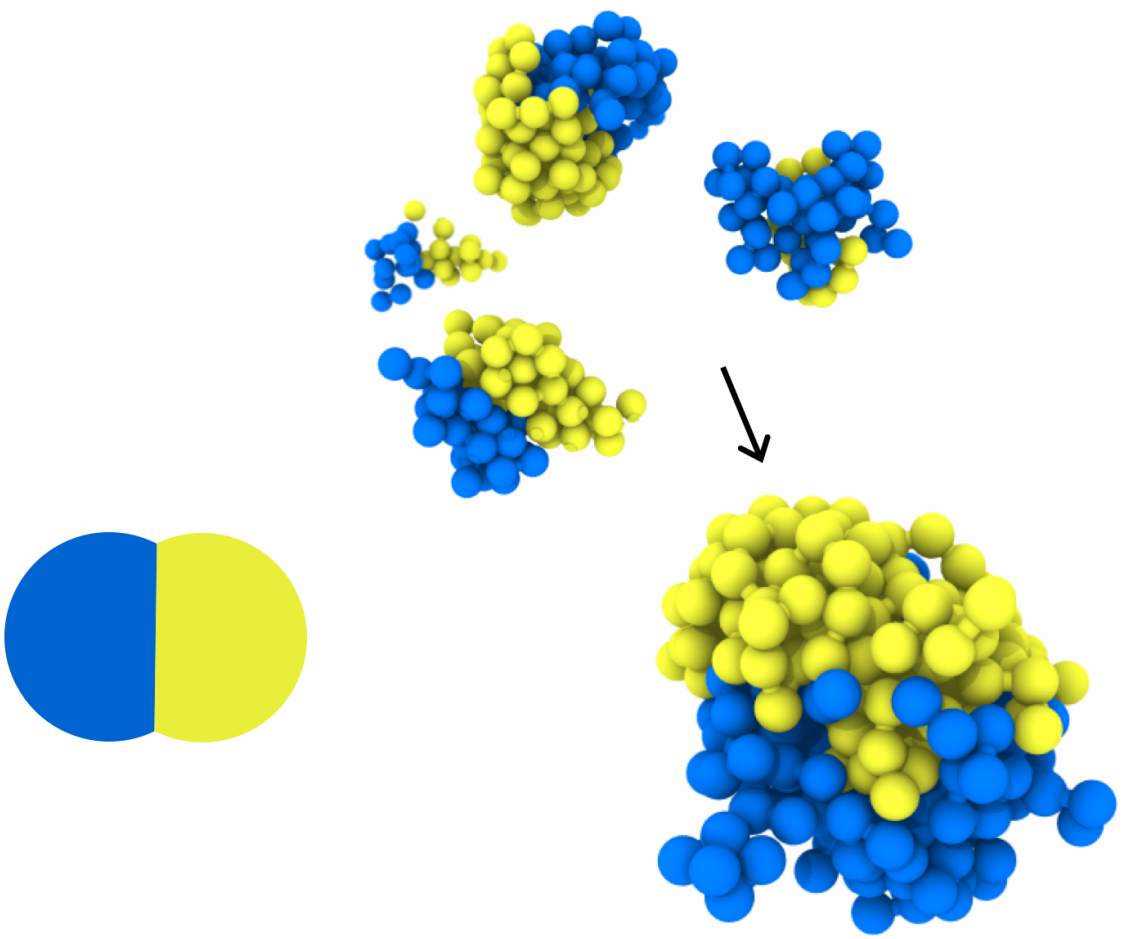


Figure 7-2: Systems of two multivalent polymers can phase separate into hemispherically segregated droplets. A schematic is shown on the bottom left, with simulation renderings on center and right. Initially, small ordered droplets formed that coalesced into one large ordered droplet at long timescales. Polymers with specific binding sites are shown in yellow and polymers without specific binding sites are shown in blue.

7.2.4 Third species

Another option for future work is considering a third or fourth species interacting with multivalent polymers. In the case of inhibitors, this could represent another pathogenic lectin the polymer is trying to inhibit, or a healthy protein the polymer wants to avoid binding to. For multivalent polymers to be effective broad-spectrum therapeutics, they must bind broadly enough to have high affinity for multiple pathogens, but be specific enough to have low affinity for important proteins, healthy cell surfaces, and beneficial microbes. Exploring how multivalent polymers interact with a

third lectin target or a lectin and a surface in the same system could help clarify these tradeoffs. Biocondensates can contain tens to hundreds of components, but phase separation appears to be controlled by only a few of these components termed “scaffolds" with the rest of the ingredients recruited post-formation and categorized as “guests". How these phase separated droplets specifically recruit guests is still an active field of study [45,47,46,149]. Simulations that contain additional species could further elucidate how multivalent polymers can control the composition of droplets. We would be especially interested in understanding how recruitment of additional species occurs in micro-phase separated or ordered droplets.

Chapter 8

Funding Sources

This work was supported by the Department of Defense (DoD) through the National Science and Engineering Graduate Fellowship (NDSEG) Program. This research was also generously supported by the Ida M. Green Fellowship and the Collamore-Rogers Fellowship through the MIT Office of the Dean of Graduate Education. Computational resources were provided in part by the MIT Supercloud [150].

This Page Intentionally Left Blank

Appendix A

Statistical Mechanical Model of Polymer Loops

The purpose of this section is to develop a statistical mechanical treatment for binding of polymers presenting many ligands to a multivalent receptor. In section A.1, we present the model for polyvalent binding. In section A.2, we calculate the partition function and provide formulas for the free energy of receptor-ligand binding, showing that numerical methods are required for precise calculations. In section A.4, we perform a large-N approximation of polyvalent binding (where N is the number of ligands present on the polymer) and find analytically that increasing N eventually stops adding any advantage for target binding. Finally, in section A.5, we present the limitations of the current model but also the interesting predictions it makes that are in agreement with the simulations performed in the main text. This section is adapted from Zumbro *et al.*, *Biophys. J.* **115** (2019) 892-902 [86].

A.1 Evaluation of the partition function for the canonical ensemble

We start with a simple model of polyvalent binding: we have an infinitely dilute random coil polymer with N ligands (N corresponds to N_P in the main text), evenly spaced with spacers of contour length l between each ligand, and an infinitely dilute receptor with M binding sites (Fig. A-1A). In an individual binding conformation in which k sites are bound, we can partition the polymer into $k+1$ fragments, with fragment lengths $y_1 l, \ldots, y_{k+1} l$ (so, $\sum_{j=1}^{k+1} y_j = N - 1$; Fig. A-1B. Note that this

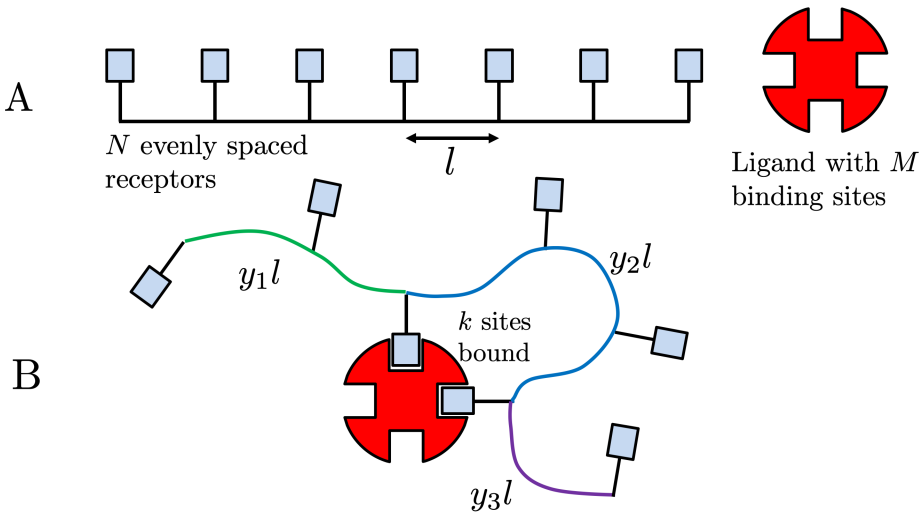


Figure A-1: Variables for polyvalent binding. A) Polymer contains N ligands, spaced by contour length l, and may bind to a receptor with M binding sites. B) In a particular binding conformation, there will be k sites bound (so $1 \leq k \leq M$), which will effectively split the polymer into $k+1$ fragments, in which all but the first and $(k+1)$th fragments must be looped. The lengths of the fragments are $y_1 l, \ldots, y_{k+1} l$

implies all of the fragments will be forced into a loop except the first and last fragment, which are free on the edges.

Assuming that the polymer is a random coil with Flory exponent ν, we have an average end-to-end distance as a function of contour length L:

$$\sqrt{\langle R^2 \rangle (L)} = \lambda \left(\frac{L}{\lambda} \right)^{\nu} = \lambda^{1-\nu} L^{\nu}$$

where λ is a constant related to the persistence length of the polymer. Then letting $\lambda = \alpha l$ for some constant α (note that α now carries the information about persistence length), let us consider the end-to-end distance as a function of y (so the contour length L here is $L = yl$):

$$\sqrt{\langle R^2 \rangle (y)} = (\alpha l)^{1-\nu} (yl)^{\nu} = \alpha^{1-\nu} l y^{\nu} \tag{A.1}$$

Therefore, the potential volume of space explored by a fragment with length y is:

$$V \sim \left(\sqrt{\langle R^2 \rangle (y)} \right)^3 = \alpha^{3-3\nu} l^3 y^{3\nu} \tag{A.2}$$

In order to form a loop, one particular point out of that volume must be chosen. Thus, the probability of a loop forming $P_{loop}(y)$ is:

$$P_{loop}(y) = \gamma\alpha^{3\nu-3}l^{-3}y^{-3\nu} \tag{A.3}$$

where γ is some constant. This probability also doubles as the Boltzmann weight given a loop. Also, let g^{Nl} be the polymer's configurational entropy, for some g, and let $w = e^{-\beta E_{bind}}$, where E_{bind} is the energy for one receptor-ligand interaction. Finally, let $P(M,k)$ be the number of permutations possible for k receptor-ligand interactions given a *particular choice* of k ligands on the polymer to bind the receptor, or to use another term, it is the bound state degeneracy; we will discuss this value later.

A.2 Statistical mechanical analysis of binding free energy

We are now in a position to write the partition function given some N and k, $Q(N,k)$:

$$\begin{aligned} Q(N,k) &= P(M,k)g^{Nl}w^k \sum_{y_1,\ldots,y_{k+1}}^{y_1+\cdots+y_{k+1}=N,*} \left(\gamma\alpha^{3\nu-3}l^{-3}y_2^{-3\nu}\right)\cdots\left(\gamma\alpha^{3\nu-3}l^{-3}y_k^{-3\nu}\right) \\ &= P(M,k)g^{Nl}w^k\gamma^{k-1}\alpha^{(3\nu-3)(k-1)}l^{-3(k-1)} \sum_{y_1,\ldots,y_{k+1}}^{y_1+\cdots+y_{k+1}=N,*} y_2^{-3\nu}\cdots y_k^{-3\nu} \end{aligned}$$

The * represents the constraint that $y_1,\ldots,y_{k+1}$ are all strictly positive because it is impossible to have a loop of length 0. Strictly speaking, it should be possible to have $y_1 = 0$ or $y_{k+1} = 0$ because the first and last fragments are not looped; it doesn't make a difference really whether we include them in the constraint so we include it for consistency. Also, for the sake of legibility, we collect our terms together the following way: let $\Phi = \gamma^{-1}\alpha^{-(3\nu-3)}l^3$ and $\Psi = w\gamma\alpha^{3\nu-3}l^{-3}$, giving us:

$$Q(N,k) = P(M,k)g^{Nl}\Phi\Psi^k \sum_{y_1,\ldots,y_{k+1}}^{y_1+\cdots+y_{k+1}=N,*} y_2^{-3\nu}\cdots y_k^{-3\nu} \tag{A.4}$$

We can now write the full partition function $Q(N)$, summing over all k:

$$\begin{aligned} Q\left(N\right) &= \sum_{k=1}^{M} Q(N,k) \\ &= \sum_{k=1}^{M} P(M,k) g^{Nl} \Phi \Psi^k \sum_{y_1,\ldots,y_{k+1}}^{y_1+\cdots+y_{k+1}=N,*} y_2^{-3\nu} \cdots y_k^{-3\nu} \end{aligned}$$

This is as far as we can get with the canonical ensemble, because the constraint $y_1 + \cdots + y_{k+1} = N$ makes calculations infeasible.

A.3 Tranformation to grand canonical ensemble

We therefore move to the grand canonical ensemble $Z(z)$ (essentially a constant chemical potential for adding new monomers to the polymer, with $z = e^{\beta\mu}$), and get some helpful simplifications:

$$\begin{aligned} Z\left(z\right) &= \sum_{N=1}^{\infty} z^N Q(N) \\ &= \sum_{N=1}^{\infty} z^N \sum_{k=1}^{M} P(M,k) g^{Nl} \Phi \Psi^k \sum_{y_1,\ldots,y_{k+1}}^{y_1+\cdots+y_{k+1}=N,*} y_2^{-3\nu} \cdots y_k^{-3\nu} \\ &= \sum_{k=1}^{M} P(M,k) \Phi \Psi^k \sum_{N=1}^{\infty} \left(zg^l\right)^N \sum_{y_1,\ldots,y_{k+1}}^{y_1+\cdots+y_{k+1}=N,*} y_2^{-3\nu} \cdots y_k^{-3\nu} \\ &= \sum_{k=1}^{M} P(M,k) \Phi \Psi^k \sum_{N=1}^{\infty} \sum_{y_1,\ldots,y_{k+1}}^{y_1+\cdots+y_{k+1}=N,*} \left(zg^l\right)^{y_1} \left[\left(zg^l\right)^{y_2} y_2^{-3\nu} \cdots \left(zg^l\right)^{k} y_k^{-3\nu}\right] \left(zg^l\right)^{y_{k+1}} \\ &= \sum_{k=1}^{M} P(M,k) \Phi \Psi^k \left(\sum_{y=1}^{\infty} \left(zg^l\right)^y\right)^2 \left(\sum_{y=1}^{\infty} \frac{\left(zg^l\right)^y}{y^{3\nu}}\right)^{k-1} \\ &= \sum_{k=1}^{M} P(M,k) \Phi \Psi^k \left(\frac{zg^l}{1-zg^l}\right)^2 \left(f_{3\nu}^{+}(zg^l)\right)^{k-1} \end{aligned}$$

where $f_{3\nu}^{+}$ is the polylogarithm or polylog function. Rearranging terms, we have:

$$Z(z) = \Phi \left(\frac{zg^l}{1-zg^l}\right)^2 \left(f_{3\nu}^{+}(zg^l)\right)^{-1} \sum_{k=1}^{M} P(M,k) \left(\Psi f_{3\nu}^{+}(zg^l)\right)^k$$

Now, we calculate our permutation term: assuming no restrictions on binding topology, we have $P(M,k) = \frac{M!}{(M-k)!}$. Using this calculation, we get a somewhat analytically tractable sum [151]:

$$\begin{aligned} Z(z) &= \Phi \left(\frac{zg^l}{1-zg^l} \right)^2 \left(f_{3\nu}^+(zg^l) \right)^{-1} \sum_{k=1}^{M} \frac{M!}{(M-k)!} \left(\Psi f_{3\nu}^+(zg^l) \right)^k \\ &= \Phi \left(\frac{zg^l}{1-zg^l} \right)^2 \left(f_{3\nu}^+(zg^l) \right)^{-1} \left[e^{\frac{1}{\Psi f_{3\nu}^+(zg^l)}} \left(\Psi f_{3\nu}^+(zg^l) \right)^M \Gamma \left(M+1, \frac{1}{\Psi f_{3\nu}^+(zg^l)} \right) - 1 \right] \end{aligned}$$

where $\Gamma(a,x)$ is the upper incomplete gamma function:

$$\Gamma(a,x) = \int_x^\infty t^{a-1} e^{-t}\, dt \tag{A.5}$$

Now, let's consider the value $\ln(Z)$, first dropping the '-1' on the assumption that it is not significant because it represents no binding. That is, it corresponds to the $k = 0$ state in which the receptor and ligand are spatially co-localized (meaning a loss of translational entropy) without physically interacting. This can either be considered a sort of 'binding' because the ligand will be localized to the receptor, or alternatively not binding. Regardless, the $k = 0$ state is is clearly unfavorable compared to binding which justifies our dropping the -1:

$$\begin{aligned} Z(z) &\approx \Phi \left(\frac{zg^l}{1-zg^l} \right)^2 \left(f_{3\nu}^+(zg^l) \right)^{-1} e^{\frac{1}{\Psi f_{3\nu}^+(zg^l)}} \left(\Psi f_{3\nu}^+(zg^l) \right)^M \Gamma \left(M+1, \frac{1}{\Psi f_{3\nu}^+(zg^l)} \right) \\ \ln(Z) &= \ln(\Phi) + 2\ln\left(zg^l\right) - 2\ln\left(1-zg^l\right) + \frac{1}{\Psi f_{3\nu}^+(zg^l)} + M\ln\Psi \\ &\quad + (M-1)\ln\left(f_{3\nu}^+(zg^l)\right) + \ln\left(\Gamma\left(M+1, \frac{1}{\Psi f_{3\nu}^+(zg^l)}\right)\right) \end{aligned} \tag{A.6}$$

Eventually we will have to relate z to the actual N which we are given. We do this by noting that $< N >= z\frac{\partial \ln Z(z)}{\partial z}$; note that one of the properties of the polylog is that $\frac{d}{dx} f_{3\nu}^+(x) = \frac{f_{3\nu-1}^+(x)}{x}$ (we

ignore fluctuations in N here):

$$\begin{aligned}
\frac{<N>}{z} &= \frac{2}{z} - \frac{2g^l}{zg^l - 1} - \frac{f^+_{3\nu-1}(zg^l)}{\Psi z\left(f^+_{3\nu}(zg^l)\right)^2} \\
&\quad + (M-1)\frac{f^+_{3\nu-1}(zg^l)}{z\left(f^+_{3\nu}(zg^l)\right)} + \frac{-\left(\Psi f^+_{3\nu}(zg^l)\right)^{-M} e^{-\left(\Psi f^+_{3\nu}(zg^l)\right)}\left(-\frac{f^+_{3\nu-1}(zg^l)}{\Psi z\left(f^+_{3\nu}(zg^l)\right)^2}\right)}{\Gamma\left(M+1, \frac{1}{\Psi f^+_{3\nu}(zg^l)}\right)} \\
<N> &= \frac{2}{1-zg^l} - \frac{f^+_{3\nu-1}(zg^l)}{\Psi\left(f^+_{3\nu}(zg^l)\right)^2} + (M-1)\frac{f^+_{3\nu-1}(zg^l)}{f^+_{3\nu}(zg^l)} \\
&\quad + \frac{f^+_{3\nu-1}(zg^l)e^{-\left(\Psi f^+_{3\nu}(zg^l)\right)}}{\Psi^{-(M+1)}\left(f^+_{3\nu}(zg^l)\right)^{-(M+2)}\Gamma\left(M+1, \frac{1}{\Psi f^+_{3\nu}(zg^l)}\right)}
\end{aligned} \tag{A.7}$$

We can also calculate thermodynamic quantities, which is what we're actually looking for. Namely, we have:

$$-k_BT\ln Z = <E> - TS - \mu <N> = A - \mu <N> \tag{A.8}$$

where A is the Helmholtz free energy, and the chemical potential is given by $\mu = \frac{\ln(z)}{\beta} = k_BT\ln(z)$. Since A is what we're after (we have constant N, after all, and assuming no substantial pressure changes A will be equivalent to G), the numerical determination of free energy would proceed as follows: we are given some length of polymer N^0. Since we have derived the function $<N>(z)$, we invert this relation (numerically, or, as we describe below, analytically in the high-N limit) to find $z(N^0)$. We then calculate the free energy of binding using:

$$A = -k_BT\ln Z\left(z(N^0)\right) + k_BT\ln\left(z(N^0)\right)N^0 \tag{A.9}$$

A.3.1 Free energy of free, unbound polymer

Let us now consider the energy of *not* binding A_{NB}, because ultimately what we want is the change in free energy of binding, $\Delta A = A - A_{NB}$. Our canonical partition function $Q_{NB}(N)$ is:

$$Q_{NB}(N) = g^{Nl}V/h \tag{A.10}$$

where V is a volume, representing the increased translation entropy that comes from not binding to the receptor, and h is a constant to make the partition function unitless. Transforming to the grand canonical ensemble for consistency, and calculating the other quantities we need:

$$\begin{aligned}
Z_{NB}(z) &= \sum_{N=1}^{\infty} z^N Q_{NB}(N) \\
&= \sum_{N=1}^{\infty} z^N g^{Nl} V/h \\
&= V/h \sum_{N=1}^{\infty} (zg^l)^N \\
&= \frac{V z g^l}{h(1 - zg^l)} \\
\ln(Z_{NB}) &= \ln(V/h) + \ln(zg^l) - \ln(1 - zg^l) \\
< N > &= z \frac{\partial \ln Z(z)}{\partial z} \\
&= 1 - \frac{zg^l}{zg^l - 1} \\
zg^l &= \frac{N-1}{N} = 1 - \frac{1}{N}
\end{aligned}$$

We can now calculate thermodynamic quantities:

$$\begin{aligned}
k_B T N \ln z &= k_B T N \ln \left(\frac{1}{g^l} \left(1 - \frac{1}{N} \right) \right) \\
&= -k_B T N \ln(g^l) + k_B T N \ln(1 - \frac{1}{N})
\end{aligned}$$

And:

$$\ln Z_{NB} = \ln(V/h) + \ln(1 - \frac{1}{N}) - \ln(1/N)$$

With the total Helmholtz free energy:

$$\beta A_{NB} = -\ln(V/h) - \ln(1 - \frac{1}{N}) - \ln(N) - N \ln(g^l) + k_B T N \ln(1 - \frac{1}{N}) \tag{A.11}$$

So to summarize, our prodecure for finding $\Delta A = A - A_{NB}$ given some N^0 is as follows. We first invert Equation A.7 to find $z(N^0)$. We then find A using Equation A.9, with $\ln Z$ given by Equation A.6. A_{NB} is given by Equation A.11. All this, particularly inverting Equation A.7, must be performed numerically; we used Mathematica [152] in our analysis in section A.5 below. (Also note that while we dropped the '-1' to get to a more legible Equation A.6, in numerical Mathematica analysis we need not drop this term. However, in the following section, we *do* drop the '-1'. We have observed that this does not make a significant difference, as expected.)

A.4 Large-N approximations for binding energy

Alternatively, it would be desirable to acquire asymptotic, large-N results. First, we consider inverting the function $< N > (z)$ given by Equation A.7. In order to make N large, we can send $1 - zg^l$ to 0, by sending zg^l to 1.

The limiting behavior of the other terms depends on the exact value of 3ν. If $1 < 3\nu < 2$, which is the case for a polymer in a theta solvent or a good solvent, then $f^+_{3\nu-1}(zg^l)$ also diverges for $zg^l \to 1$, while $f^+_{3\nu}(zg^l)$ is perfectly well-behaved. In particular, $f^+_{3\nu}(1) = \zeta(3\nu)$, where $\zeta(\cdot)$ is the Riemann zeta function. So the way to get to a high N is to send zg^l to 1. If $3\nu > 2$, as is required for the Poland-Scheraga DNA denaturation model to be a first order rather than second order phase transition [101], then $f^+_{3\nu-1}(zg^l)$ is well-behaved, converging to $\zeta(3\nu - 1)$. We will show that the $\frac{2}{1-zg^l}$ term is the only significant one in the large-N limit of Equation A.7, and will therefore assume that $1 < 3\nu < 2$; if $3\nu > 2$ then this is trivial.

Where we let zg^l approach 1, substituting in $f^+_{3\nu}(zg^l) = \zeta(3\nu)$, we get :

$$< N > = \frac{2}{1 - zg^l} - \frac{f^+_{3\nu-1}(zg^l)}{\Psi\left(\zeta(3\nu)\right)^2} + (M-1)\frac{f^+_{3\nu-1}(zg^l)}{\zeta(3\nu)} + \frac{f^+_{3\nu-1}(zg^l)e^{-(\Psi\zeta(3\nu))}}{\Psi^{-(M+1)}\left(\zeta(3\nu)\right)^{-(M+2)}\Gamma\left(M+1, \frac{1}{\Psi\zeta(3\nu)}\right)} \tag{A.12}$$

Note that the last three terms are now independent of z except for a $f^+_{3\nu-1}(zg^l)$ term in each. We will factor that out and collect all the other terms into a constant κ (which we will shortly be able

to ignore):

$$< N >= \frac{2}{1 - zg^l} - \kappa f^+_{3\nu-1}(zg^l) \tag{A.13}$$

This is still not invertible, so let us find out which term blows up *faster*, $\frac{2}{1-zg^l}$ or $\kappa f^+_{3\nu-1}(zg^l)$. First, let $x = 1 - zg^l$. Then we have:

$$< N >= \frac{2}{x} - \kappa f^+_{3\nu-1}(1 - x) \tag{A.14}$$

The expansion about $x = 0$ of $f^+_{3\nu-1}(1 - x)$ is:

$$f^+_{3\nu-1}(1 - x) \approx \zeta(3\nu - 1) + \frac{\Gamma(2 - 3\nu)}{x^{2-3\nu}} + O(x^{3\nu} - 1) \tag{A.15}$$

which means that this term goes to infinity only with $2 - 3\nu$th power of $\frac{1}{x}$, and $2 - 3\nu < 1$ [153]. Thus, the first term, $\frac{2}{x}$, dominates as $zg^l \to 1$, so we drop all other terms. This at long last leaves us with the invertible approximation:

$$< N >= \frac{2}{1 - zg^l} \tag{A.16}$$

$$zg^l = \frac{N - 2}{N} = 1 - \frac{2}{N} \tag{A.17}$$

So, we now immediately have the extra term in the Helmholtz free energy:

$$\begin{aligned} k_B T N \ln z &= k_B T N \ln\left(\frac{1}{g^l}\left(1 - \frac{2}{N}\right)\right) \\ &= k_B T N \left(-\ln(g^l) + \ln(1 - \frac{2}{N})\right) \\ &\approx -k_B T N \ln(g^l) - k_B T N \left(\frac{2}{N}\right) \\ &= -k_B T N \ln(g^l) - 2k_B T \end{aligned} \tag{A.18}$$

The approximation $\ln(1 - 2/N) \approx -2/N$ allowed the equation to take a more elegant form. We can also now substitute $1 - \frac{2}{N}$ for zg^l in our equation for $\ln Z$. Again, we will make first-order

approximations because we are working with high N:

$$\begin{aligned}\ln(Z) &= \ln(\Phi) + 2\ln\left(1-\frac{2}{N}\right) - 2\ln\left(1-(1-\frac{2}{N})\right) + \frac{1}{\Psi f_{3\nu}^{+}(1-\frac{2}{N})} + M\ln\Psi \\ &+ (M-1)\ln\left(f_{3\nu}^{+}(1-\frac{2}{N})\right) + \ln\left(\Gamma\left(M+1, \frac{1}{\Psi f_{3\nu}^{+}(1-\frac{2}{N})}\right)\right) \\ &= \ln(\Phi) - \frac{4}{N} - 2\ln(2) + 2\ln(N) + \frac{1}{\Psi f_{3\nu}^{+}(1-\frac{2}{N})} + M\ln\Psi \\ &+ (M-1)\ln\left(f_{3\nu}^{+}(1-\frac{2}{N})\right) + \ln\left(\Gamma\left(M+1, \frac{1}{\Psi f_{3\nu}^{+}(1-\frac{2}{N})}\right)\right) \end{aligned} \tag{A.19}$$

Also, in the large N limit, A_{NB} reduces to:

$$\beta A_{NB} = -\ln(V/h) - \frac{1}{N} - \ln(N) - N\ln(g^l) - 1 \tag{A.20}$$

So, the total free energy of binding is:

$$\begin{aligned}\beta\Delta A = &-\ln(\Phi) + \frac{4}{N} + 2\ln(2) - \ln(N) - \frac{1}{\Psi f_{3\nu}^{+}(1-\frac{2}{N})} - M\ln\Psi + \ln(V/h) \\ &- (M-1)\ln\left(f_{3\nu}^{+}(1-\frac{2}{N})\right) - \ln\left(\Gamma\left(M+1, \frac{1}{\Psi f_{3\nu}^{+}(1-\frac{2}{N})}\right)\right) - 1\end{aligned}$$

Note that all contributions of g^l have canceled out, which was what must happen because the polymer has the same flexibility regardless of binding. In the high-N limit, we can continue to simplify:

$$\begin{aligned}\beta\Delta A = &-\ln(\Phi) + 2\ln(2) - \ln(N) - \frac{1}{\Psi\zeta(3\nu)} - M\ln\Psi + \ln(V/h) \\ &- (M-1)\ln(\zeta(3\nu)) - \ln\left(\Gamma\left(M+1, \frac{1}{\Psi\zeta(3\nu)}\right)\right) - 1\end{aligned} \tag{A.21}$$

Re-expanding out the terms Φ and Ψ, we get: $\Phi = \gamma^{-1}\alpha^{-(3\nu-3)}l^3$ and $\Psi = w\gamma\alpha^{3\nu-3}l^{-3}$, giving us:

$$\begin{aligned}\beta\Delta A = &-\ln(\gamma^{-1}\alpha^{-(3\nu-3)}l^3) + 2\ln(2) - \ln(N) - \frac{l^3\alpha^{3-3\nu}}{w\gamma\zeta(3\nu)} - M\ln\left(w\gamma\alpha^{3\nu-3}l^{-3}\right) \\ &- (M-1)\ln(\zeta(3\nu)) - \ln\left(\Gamma\left(M+1, \frac{1}{w\gamma\alpha^{3\nu-3}l^{-3}\zeta(3\nu)}\right)\right) - 1 + \ln(V/h)\end{aligned} \tag{A.22}$$

Equation A.22 is what we've been after all along, an analytical form for free energy of binding; we can now start to draw some conclusions based on it. First, the energy of binding grows only logarithmically with N. This makes sense in retrospect, because for large enough loops, the loop entropy will eventually dominate the binding energy because the loop entropy can increase indefinitely but the binding energy is fixed. Thus, all polyvalent binding is necessarily "local," with the definition of local being dependent on the various parameters. The energy grows logarithmically because the number of binding locations grows linearly with N (indeed, *is* N), which gives an extra binding entropy on the order of $\ln N$. Analysis of the effects of other variables is unfortunately dependent on the incomplete gamma function term, which makes analytical statements difficult.

Lastly, let us bring β over to the right side of the equation, group together all the terms that include neither N nor V into a constant C, and assume changes in pressure are negligible (so $\Delta A = \Delta G$). We then have:

$$\Delta G = C - k_B T \ln(N/V) \tag{A.23}$$

In other words, ΔG depends solely on $d = N/V$, where d the density of individual receptors. In the simulation described in the main text, we have $d = \frac{nN_P}{V_{box}}$ Substituting this into Equation A.23 gives us:

$$\Delta G = C - k_B T \ln\left(\frac{nN_P}{V_{box}}\right) \tag{A.24}$$

This is Equation 7 in the main text, as desired.

A.5 Conclusions and discussion

There are some clear caveats to this model that limit its potential efficacy:

1. There may be kinetic constraints on various types of binding conformations, or increased steric hindrance of binding when there are many ligands bound to the receptor.

2. We are ignoring fluctuations in $< N >$ in the grand canonical ensemble, which is fine in the large-N limit but this could be problematic for medium N values.

3. We are assuming that the polymer forms a random coil in between each binding site, which would not apply to systems without flexible linkers.

4. We are assuming infinitely dilute polymer and receptor. Thus, this model is not directly applicable to, for example, the situation with multiple targets, or gel formation mediated by a multivalent cross-linker.

5. We are assuming that the receptor is a point mass, or at least small with respect to the polymer. This assumption does not hold for viruses and other large targets for polyvalent inhibitors.

We note that points 1, 2 ,and 3 do not affect the asymptotic result of diminishing returns for increasing the degree of polymerization, though they will affect the avidities for specific polymer constructs. Points 4 and 5 do somewhat circumscribe the applicability of the model. We do not apply this theoretical model to the multiple-target situation in the main text due to point 4, and we do not expect that the model will provide accurate results (or even necessarily accurate scaling) for viruses due to point 5.

Despite these caveats, a number of conclusions and predictions may be drawn from the model. Most prominently, we predict diminishing returns of increased polymer length, as observed in the main text. This perhaps suggests that more compact polymers, such as star-shaped highly branched, or even looped polymers, may be more effective at polyvalent binding because they will not have as much of an unfavorable free energy contribution from loop entropy.

This model also predicts that diminishing returns will be reached more quickly in good solvent, and furthermore that good solvent will reduce the maximum achievable avidity. To show this, we will first assign plausible values for various polymer constants constants. Assuming a flexible polymer, let "Persistence length" $\lambda = 2$ nm. For l, let's start with $l = 4$ nm, so $\alpha = \lambda/l = 0.5$. To set a value for γ, note that it has units of volume, so is essentially the size of the binding site. We should be fine, order of magnitude wise, letting $\gamma = 1$ nm^3. We ignore V/h; those only shift things by a constant, so we'll let $V = h = 1$. Thus, we should not necessarily expect that absolute ΔG values to be accurate, though the relative ΔG values should be. For our binding strength, let $w = e^3$. We also let $M = 2$ to match the simulations. For good solvent, we use a Flory exponent of $\nu = 0.588$, and for theta solvent, we use $\nu = 0.5$.

Using these values, Fig. A-2 shows the ΔG given these values in good and theta solvent, using the numerical calculations rather than high-N approximation. In this figure, we correct the ΔG for constant receptor per volume density, to match the simulations performed. As described in the main

text and observed in our simulations, ΔG initially decreases due to increased avidity but reaches a limit at high N_P as predicted from Equation A.24. Furthermore, ΔG is more negative in the theta solvent case, as observed in the simulations. The plot for the theta solvent also appears to be reaching its asymptote slightly more slowly than the good solvent case, which is likewise consistent with the simulations described in the main text.

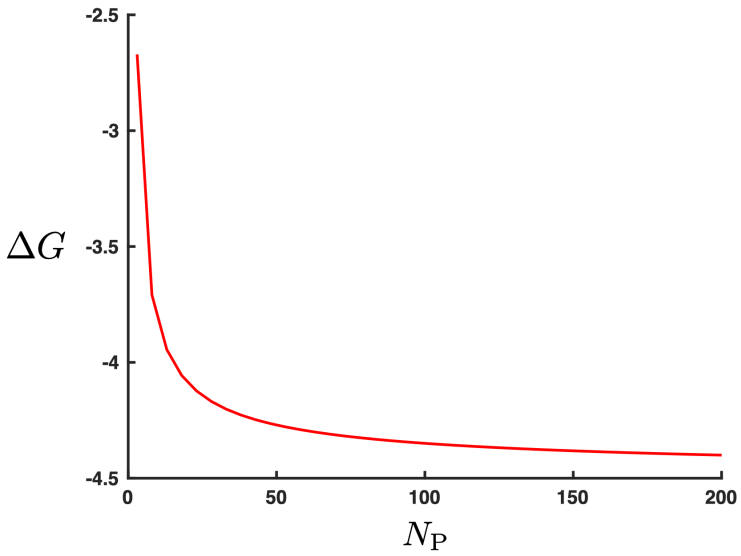


Figure A-2: Free energy of binding ΔG as a function of degree of polymerization N_P for good and theta solvent. Note that we are holding the total number of receptors in the system constant, only changing the connectivity.

To conclude, our theoretical approach makes qualitatively accurate predictions about two interesting phenomena: the diminishing returns on avidity for extending polymer length after a certain point, and the effect of good versus theta solvent on avidity.

This Page Intentionally Left Blank

Appendix B

Simulation Code

This Appendix includes the general simulation code for all of the simulations used in this thesis. It includes four Fortran files:

1. toxinSolubilityNVT.f95, primary simulation file
2. routinesMultTox.f95, routines for calculating forces, neighbor lists, etc.
3. functionsEmi.f95, small generic functions to round numbers and search arrays
4. parameters.f95, global simulation variables and constants

B.1 toxinSolubilityNVT.f95

```
! To Compile:
! gfortran -c parameters.f95 routinesMultTox.f95 functionsEmi.f95
! gfortran -O2 toxinSolubilityNVT.f95 parameters.o routinesMultTox.o
   functionsEmi.o -o toxinSolubility.exe

program toxinSolubilityNVT

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! This program runs the simulations of multivalent polymers and small
   targets
```

```
! used for my thesis research. It calls on variables and constants from parameters.f95,
! and routines and functions from routinesMultTox.f95 and functionsEmi.f95.
!
! Emiko Zumbro Jul 2020
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

USE parameters
USE routinesMultTox
USE functionsEmi

IMPLICIT NONE

! Initiate all of the variables
INTEGER :: i,j,n,totSteps,iStart

! positions
REAL(DP), DIMENSION(:), ALLOCATABLE :: rx,ry,rz
REAL(DP), DIMENSION(:,:), ALLOCATABLE :: posTox ! dim x maxTox
REAL(DP), DIMENSION(:,:), ALLOCATABLE :: dRTox ! beads x maxTox

! forces
REAL(DP), DIMENSION(:), ALLOCATABLE :: fljx,fljy,fljz,fljxTox,fljyTox,fljzTox ! beads
REAL(DP), DIMENSION(:), ALLOCATABLE :: fljxToxTox,fljyToxTox,fljzToxTox ! maxTox
REAL(DP), DIMENSION(:), ALLOCATABLE :: fljxToxTot,fljyToxTot,fljzToxTot ! beads
REAL(DP), DIMENSION(:), ALLOCATABLE :: fspringx,fspringy,fspringz,fspringxTox,fspringyTox,fspringzTox
REAL(DP), DIMENSION(:), ALLOCATABLE :: FbindMag,fbindx,fbindy,fbindz
REAL(DP), DIMENSION(:), ALLOCATABLE :: fwx,fwy,fwz ! polymer beads
```

```
36
37 ! delta positions and cummulative potentials
38 REAL(DP), DIMENSION(:), ALLOCATABLE :: delXToxUnit,delYToxUnit,delZToxUnit
39 REAL(DP), DIMENSION(:), ALLOCATABLE :: dUdx,dUdy,dUdz,dUdxTox,dUdyTox,
       dUdzTox
40
41 ! random multipliers
42 REAL(DP), DIMENSION(:,:), ALLOCATABLE :: deltaWt,deltaWtTox
43 INTEGER :: nRun, nSeed
44 INTEGER*8, DIMENSION(:) :: readSeed(seedSize)
45 INTEGER*4, DIMENSION(:) :: seed(seedSize)
46
47 ! Gaussian random number holders
48 REAL(DP) :: gausNum = 0.0_DP
49 INTEGER :: rowG,colG
50
51 ! binding information
52 INTEGER, DIMENSION(:,:,:), ALLOCATABLE :: prevOmegaTot
53 INTEGER, DIMENSION(:,:), ALLOCATABLE :: prevOmega, newOmega
54 INTEGER, DIMENSION(:), ALLOCATABLE :: beadsBound, timeBoundUnbound,
       timeTypeBond
55 INTEGER, DIMENSION(:), ALLOCATABLE :: idxToxBnd, bdBndIdx
56 REAL(DP), DIMENSION(:), ALLOCATABLE :: delE_0
57 REAL(DP), DIMENSION(:), ALLOCATABLE :: delE_UB
58
59 ! Variables for keeping toxin concentration constant
60 REAL(DP), DIMENSION(:,:), ALLOCATABLE :: dRToxTox, existTox
61
62 ! Variables for assigning random indexes
63 INTEGER, DIMENSION(:), ALLOCATABLE :: old_idxs, rand_idxs
64 INTEGER :: nIdx = 0
65 REAL(DP), DIMENSION(:,:), ALLOCATABLE :: delE0Tot
66
```

```
67 ! Variables for neighbor lists
68 INTEGER, DIMENSION(:), ALLOCATABLE :: nNghbrsPP, nNghbrsToxTox,
      nNghbrsPolyTox
69 INTEGER, DIMENSION(:,:), ALLOCATABLE :: nghbrListPP, nghbrListToxTox,
      nghbrListPolyTox
70
71 ! restart info
72 INTEGER :: restart
73
74 CHARACTER(80) :: restart_conf
75 CHARACTER(80) :: restart_rand
76 CHARACTER(80) :: configuration
77
78 restart_conf="dataOut/restart_conf.res"
79 restart_rand="dataOut/restart_rand.res"
80
81 OPEN (UNIT=20,FILE="dataOut/posToxX.dat",STATUS="REPLACE")
82 CLOSE (20)
83 OPEN (UNIT=20,FILE="dataOut/posToxY.dat",STATUS="REPLACE")
84 CLOSE (20)
85 OPEN (UNIT=20,FILE="dataOut/posToxZ.dat",STATUS="REPLACE")
86 CLOSE (20)
87 OPEN (UNIT=20,FILE="dataOut/confX.dat",STATUS="REPLACE")
88 CLOSE (20)
89 OPEN (UNIT=20,FILE="dataOut/confY.dat",STATUS="REPLACE")
90 CLOSE (20)
91 OPEN (UNIT=20,FILE="dataOut/confZ.dat",STATUS="REPLACE")
92 CLOSE (20)
93 OPEN (UNIT=20,FILE="dataOut/idxToxBnd.dat",STATUS="REPLACE")
94 CLOSE (20)
95 OPEN (UNIT=20,FILE="dataOut/numBnd.dat",STATUS="REPLACE")
96 CLOSE (20)
97
```

```
print *, "about to input.par"

OPEN(10,file="input.par",status="unknown")
REWIND (10)
READ (10,FMT=*)
READ (10,FMT=*) restart
READ (10,FMT=*)
READ (10,FMT=*) Nmer
READ (10,FMT=*)
READ (10,FMT=*) beads
READ (10,FMT=*)
READ (10,FMT=*) delE_0_center
READ (10,FMT=*)
READ (10,FMT=*) bindingSites
READ (10,FMT=*)
READ (10,FMT=*) tot_t
READ (10,FMT=*)
READ (10,FMT=*) isPolymer
READ (10,FMT=*)
READ (10,FMT=*) randNewToxin
READ (10,FMT=*)
READ (10,FMT=*) stdDevPolyAff
READ (10,FMT=*)
READ (10,FMT=*) nRun
READ (10,FMT=*)
READ (10,FMT=*) nTox
READ (10,FMT=*)
READ (10,FMT=*) epsToxTox
READ (10,FMT=*)
READ (10,FMT=*) gammaw
READ (10,FMT=*)
READ (10,FMT=*) epsilon
READ (10,FMT=*)
```

```
READ (10,FMT=*) epsToxPoly
CLOSE (10)

totSteps = tot_t/deltaT
print *, totSteps
!This size was calculated to have the same concentration of beads as I had
    in all the data I ran with only 10 beads 1.1*beads;
sizeBox = ((11.0_DP**3.0_DP)*(64.0_DP/10.0_DP))**(1.0/3.0_DP) !((11.0_DP
   **3.0_DP)*(beads/10.0_DP))**(1.0/3.0_DP) ! 200
print *, "sizeBox = ", sizeBox

! This is for NVT (we aren't going to change the number of tox, so don't
   need
! to allocate extra space in advance)
maxTox = nTox

! Get out your random seed and seed your function
!print *, "about to rSeed.par"
OPEN(10,file="rSeed.par",status="unknown")
REWIND (10)
DO nSeed = 1,seedSize
READ (10,FMT=*) readSeed(nSeed)
END DO

CLOSE (10)

seed = INT(readSeed,4)

! Initiate your random seed
CALL RANDOM_SEED(PUT=seed)
!print *, "size of seed = ", SIZEOF(seed)

print *, "seeded"
```

```
161
162 ALLOCATE (rx(beads),ry(beads),rz(beads),posTox(dim,maxTox),dRTox(beads,
      maxTox))
163 ALLOCATE (fljx(beads),fljy(beads),fljz(beads),fljxTox(beads),fljyTox(beads
      ),fljzTox(beads))
164 ALLOCATE (fljxToxTox(maxTox),fljyToxTox(maxTox),fljzToxTox(maxTox))
165 ALLOCATE (fljxToxTot(beads),fljyToxTot(beads),fljzToxTot(beads))
166
167 ALLOCATE (fspringx(beads),fspringy(beads),fspringz(beads))
168 ALLOCATE (fspringxTox(beads),fspringyTox(beads),fspringzTox(beads))
169 ALLOCATE (fwx(beads),fwy(beads),fwz(beads))
170
171 ALLOCATE (fbindx(beads),fbindy(beads),fbindz(beads),FbindMag(beads))
172 ALLOCATE (prevOmegaTot(beads,bindingSites,maxTox),prevOmega(beads,
      bindingSites), newOmega(beads,bindingSites))
173 ALLOCATE (timeBoundUnbound(maxTox),timeTypeBond(maxTox))
174 ALLOCATE (delXToxUnit(beads), delYToxUnit(beads), delZToxUnit(beads))
175 ALLOCATE (deltaWt(dim,beads), deltaWtTox(dim,maxTox),beadsBound(beads))
176 ALLOCATE (dUdx(beads),dUdy(beads),dUdz(beads),dUdxTox(maxTox),dUdyTox(
      maxTox),dUdzTox(maxTox))
177 ALLOCATE (delE_0(beads),delE_UB(beads))
178 ALLOCATE (dRToxTox(maxTox,maxTox))
179 ! ALLOCATE (nToxAggTime(timeAvgInt),dRNewTox(maxTox),closeToxIdx(maxTox))
180 ALLOCATE (idxToxBnd(beads),bdBndIdx(bindingSites))
181 ALLOCATE (old_idxs(Nmer),rand_idxs(Nmer/2))
182 ALLOCATE (delE0Tot(50,beads))
183 ALLOCATE (nNghbrsPP(beads),nghbrListPP(beads,beads),nNghbrsToxTox(maxTox),
      nghbrListToxTox(maxTox,maxTox))
184 ALLOCATE (nNghbrsPolyTox(maxTox),nghbrListPolyTox(maxTox,beads))
185
186 delE_0 = delE_0_center
187
188 ! Alternate binding site pattern assignments are commented out below
```

```
189 ! ! Assign a random number for every bead in the polymer
190 ! DO colG=1,beads
191 !   CALL gasdev_s(gausNum)
192 !   delE_0(colG) = gausNum
193 ! END DO
194
195 ! delE_0 = stdDevPolyAff*delE_0 + delE_0_center
196
197 ! ! ! Making half and half polymers
198 ! DO colG=1,beads/Nmer
199 !   rowG = Nmer*(colG-1)+1
200 !   delE_0(rowG:rowG-1+Nmer/2) = delE_0_center-stdDevPolyAff
201 !   delE_0(rowG+Nmer/2:rowG-1+Nmer) = delE_0_center+stdDevPolyAff
202 ! END DO
203
204 ! ! Making alternating polymers
205 ! delE_0 = delE_0_center-stdDevPolyAff
206 ! DO colG=1,beads/2
207 !   delE_0(colG*2) = delE_0_center+stdDevPolyAff
208 ! END DO
209
210 ! Making alternating every 2 polymers
211 ! delE_0 = delE_0_center-stdDevPolyAff
212 ! DO colG=4,beads,4
213 !   delE_0(colG) = delE_0_center+stdDevPolyAff
214 !   delE_0(colG-1) = delE_0_center+stdDevPolyAff
215 ! END DO
216
217 ! ! Making alternating every 4 polymers
218 ! delE_0 = delE_0_center-stdDevPolyAff
219 ! DO colG=8,beads,8
220 !   delE_0(colG) = delE_0_center+stdDevPolyAff
221 !   delE_0(colG-1) = delE_0_center+stdDevPolyAff
```

```
!   delE_0(colG-2) = delE_0_center+stdDevPolyAff
!   delE_0(colG-3) = delE_0_center+stdDevPolyAff
! END DO

! Making polymers with two binding affinities that are randomly located
   along the chain

! delE_0 = delE_0_center+stdDevPolyAff
! DO colG=1,beads/Nmer
!   DO nIdx = 1,Nmer
!     old_idxs(nIdx) = nIdx+(Nmer*(colG-1))
!   END DO
!   print *, old_idxs

!   nIdx = Nmer/2
!   ! pick nIdx random indexes from old_idxs and return them in rand_idxs
!   CALL pick_random(old_idxs, rand_idxs, nIdx)

!   print *, "rand_idxs = ", rand_idxs, "\n"

!   delE_0(rand_idxs) = delE_0_center-stdDevPolyAff
! END DO

! ! Get polymer affinities from a previously written matrix
! OPEN(10,file="delE0Tot.txt",status="unknown")
! REWIND (10)
! DO nSeed = 1,50
! READ (10,FMT=*) delE0Tot(nSeed,:)
! END DO
! CLOSE (10)


delE_UB = -delE_0 + 0.5_DP
```

```
254 posTox = 0.0
255
256 print *,"allocated"
257
258 !initialize the configuration
259 IF (restart == 0) THEN
260
261   iStart = 0
262   ry=0.0
263   rz=0.0
264
265   ! Only works for polymers with 64 beads or less
266   CALL setInitialPos(rx,ry,rz)
267   CALL setInitialPosTox(posTox(1,1:nTox),posTox(2,1:nTox),posTox(3,1:nTox)
      ,rx,ry,rz)
268
269 ELSE
270   ! Restart from previous configuration - Not Implemented
271   iStart = 0
272 END IF
273
274 print *, "iStart = ",iStart
275
276 OPEN(10,file="dataOut/Output.dat",status="REPLACE")
277 ! WRITE (10,*)
278 ! WRITE (10,*) "This is the file that contains the parameters"
279 ! WRITE (10,*) "used in the simulation."
280 ! WRITE (10,*)
281 WRITE (10,*) "startingStep",iStart
282 WRITE (10,*) "restart",restart
283 WRITE (10,*) "beads",beads
284 WRITE (10,*) "Nmer",Nmer
285 WRITE (10,*) "bindingSites",bindingSites
```

```
286 WRITE (10,*) "maxTox",maxTox
287 WRITE (10,*) "nTox0",nTox
288 WRITE (10,*) "tot_t",tot_t
289 !WRITE (10,*) "isPolymer",isPolymer
290 WRITE (10,*) "delE_0_center",delE_0_center
291 !WRITE (10,*) "delE_UB",delE_UB
292 WRITE (10,*) "deltaT",deltaT
293 WRITE (10,*) "tot_t",tot_t
294 WRITE (10,*) "checkBindingInterval",checkBindingInterval
295 WRITE (10,*) "D",D
296 WRITE (10,*) "DTox",DTox
297 WRITE (10,*) "dim",dim
298 WRITE (10,*) "sizeBox",sizeBox
299 WRITE (10,*) "d_wrt",d_wrt
300 WRITE (10,*) "nRun",nRun
301 WRITE (10,*) "epsilonLJPP",epsilon
302 WRITE (10,*) "epsLJToxTox",epsToxTox
303 WRITE (10,*) "epsLJToxPoly",epsToxPoly
304 WRITE (10,*) "gammaw",gammaw
305 CLOSE (10)
306
307 OPEN (UNIT=10,FILE="dataOut/iPos.dat", STATUS="REPLACE")
308   DO j=1,beads
309      WRITE (UNIT=10,FMT=*) rx(j),ry(j),rz(j)
310   END DO
311 CLOSE (10)
312
313 OPEN (UNIT=10,FILE="dataOut/iPosTox.dat", STATUS="REPLACE")
314   DO j=1,maxTox
315      WRITE (UNIT=10,FMT=*) posTox(1,j),posTox(2,j),posTox(3,j)
316   END DO
317 CLOSE (10)
318
```

```
319 OPEN (UNIT=10,FILE="dataOut/randSeed.dat", STATUS="REPLACE")
320    WRITE (UNIT=10,FMT=*) seed
321 CLOSE (10)
322
323 OPEN (UNIT=10,FILE="dataOut/delE0.dat", STATUS="REPLACE")
324   DO j=1,beads
325      WRITE (UNIT=10,FMT=*) delE_0(j)
326   END DO
327 CLOSE (10)
328
329 OPEN (UNIT=10,FILE="dataOut/delEUB.dat", STATUS="REPLACE")
330   DO j=1,beads
331      WRITE (UNIT=10,FMT=*) delE_UB(j)
332   END DO
333 CLOSE (10)
334
335 ! Initiate everything to start unbound and unaggregated
336 IF (restart==0) THEN
337   prevOmegaTot = 0
338   prevOmega = 0
339   timeBoundUnbound = 0
340   idxToxBnd = 0
341   timeTypeBond = 0
342 ELSE
343   prevOmega = 0
344   timeBoundUnbound = 0
345   timeTypeBond = 0
346 ENDIF
347
348 print *, "made it to function"
349
350 DO i=iStart+1,totSteps
351
```

```
352 DO colG=1,dim
353   DO rowG=1,beads
354     CALL gasdev_s(gausNum)
355     deltaWt(colG,rowG) = gausNum
356   END DO
357 END DO
358
359 deltaWt = deltaWt * SQRT(deltaT)
360
361 DO colG=1,dim
362   DO rowG=1,nTox
363     CALL gasdev_s(gausNum)
364     deltaWtTox(colG,rowG) = gausNum
365   END DO
366 END DO
367
368 deltaWtTox = deltaWtTox * SQRT(deltaT)
369
370 ! Calculate the spring forces due to connectivity
371 IF (isPolymer) THEN
372   CALL springForces(fspringx,fspringy,fspringz,rx,ry,rz)
373   IF (gammaw > 0.0) THEN
374     ! Calculate the wormlike chain forces to control bending
375     CALL worm_like(fwx,fwy,fwz,rx,ry,rz)
376   ELSE
377     fwx = 0.0
378     fwy = 0.0
379     fwz = 0.0
380   END IF
381 ELSE
382   ! there is no connectivity and all the spring forces are 0
383   fspringx = 0.0
384   fspringy = 0.0
```

```
385     fspringz = 0.0
386
387     fwx = 0.0
388     fwy = 0.0
389     fwz = 0.0
390   END IF
391
392   ! print *, "Calculated spring forces"
393   IF (beads > 1) THEN
394
395     ! Update the neighbor list between inhibitor beads
396     IF ((i==(iStart+1)).OR.(MODULO(i,checkBindingInterval/10)==0)) THEN
397       CALL updateNeighborListPP(rx,ry,rz,nNghbrsPP,nghbrListPP)
398     END IF
399
400     ! Calculate the lennard-jones excluded volume forces between inhibitor
      beads
401     CALL rljmodNLBC(fljx,fljy,fljz,rx,ry,rz,nNghbrsPP,nghbrListPP)
402
403   END IF
404
405   fbindx = 0.0
406   fbindy = 0.0
407   fbindz = 0.0
408
409   fljxTox=0.0
410     fljyTox=0.0
411     fljzTox=0.0
412
413     fljxToxTot=0.0
414     fljyToxTot=0.0
415     fljzToxTot=0.0
416
```

```
417     dUdxTox=0.0
418     dUdyTox=0.0
419     dUdzTox=0.0
420
421   idxToxBnd = 0
422
423     ! Find LJ excluded volume between all the toxins and themselves if
       there are multiple toxins
424   IF (nTox > 1) THEN
425
426
427     ! Update the neighbor list
428     IF ((i==(iStart+1)).OR.(MODULO(i,checkBindingInterval/10)==0)) THEN
429       CALL updateNeighborListToxTox(posTox(1,1:nTox),posTox(2,1:nTox),
       posTox(3,1:nTox),nNghbrsToxTox,nghbrListToxTox)
430     END IF
431
432     ! Calculate the lennard-jones excluded volume forces between toxins
433     ! Neighbor List incompatible with changing toxin concentration
       dynamically
434     CALL rljToxToxNLBC(fljxToxTox,fljyToxTox,fljzToxTox,posTox(1,1:nTox),
       posTox(2,1:nTox),posTox(3,1:nTox), &
435       & nNghbrsToxTox,nghbrListToxTox)
436
437   END IF
438
439   IF (beads > 0) THEN
440
441       ! Update the neighbor list between toxins and inhibitor polymers
442       IF ((i==(iStart+1)).OR.(MODULO(i,checkBindingInterval/10)==0)) THEN
443         CALL updateNeighborListPolyTox(rx,ry,rz,posTox(1,1:nTox),posTox
       (2,1:nTox),posTox(3,1:nTox),&
444           & nNghbrsPolyTox,nghbrListPolyTox)
```

```
445       END IF
446
447      DO j = 1,nTox
448         ! Calculate Lennard-Jones excluded volume forces between toxins and
    inhibitor beads
449         CALL rljToxNLBC(fljxTox,fljyTox,fljzTox,dRTox(:,j),rx,ry,rz,posTox
    (:,j),delXToxUnit,delYToxUnit,&
450           & delZToxUnit,nNghbrsPolyTox(j),nghbrListPolyTox(j,:))
451
452         prevOmega = prevOmegaTot(:,:,j)
453
454         ! If it is a binding interval, check to see if a binding event
    happens
455         IF ((MODULO(i,checkBindingInterval)==0).AND.(delE_0_center<0)) THEN
456           CALL bound(prevOmegaTot, prevOmega, dRTox(:,j), FbindMag, newOmega
    , delE_0, delE_UB)
457         ELSE
458           newOmega = prevOmega
459           beadsBound = SUM(newOmega,2) ! by this one toxin
460           FbindMag = -k_bind*beadsBound*(dRTox(:,j)-l_bind)
461         END IF
462
463         ! Calculate the bound forces with their direction
464         DO n = 1,beads
465                 ! del_ToxUnit points from the toxin to the inhibitor
466                 ! I think this makes it correct for the forces on the bead,
    but the toxin will have to have a negative of the fbind_
467                 ! In order to have fbindx be dimension n, add the previous
    fbindx as you iterate through the toxins, j
468                 fbindx(n) = fbindx(n) + FbindMag(n)*delXToxUnit(n)
469                 fbindy(n) = fbindy(n) + FbindMag(n)*delYToxUnit(n)
470                 fbindz(n) = fbindz(n) + FbindMag(n)*delZToxUnit(n)
471           END DO
```

```
472
473       ! Update the Omega
474       prevOmegaTot(:,:,j) = newOmega
475
476       bdBndIdx = 0
477       ! Update the tracker for whether my toxins is overall bound or not
478       IF (ANY(SUM(newOmega,2)==1)) THEN !newOmega(beads,bindingSites)
479               ! sum along binding sites to get a 1 for every bead that is
    bound and a 0 for every polymer bead that is unbound and then count the
     bead as bound if any of its binding sites is bound (ie, any of the
    bead slots = 1)
480               timeBoundUnbound(j) = 1
481               timeTypeBond(j) = SUM(SUM(newOmega,2),1)
482               bdBndIdx = FIND(SUM(newOmega,2)==1);
483               DO n = 1,bindingSites
484                 IF (bdBndIdx(n)>0) THEN
485                   idxToxBnd(bdBndIdx(n)) = j
486                 END IF
487               END DO
488
489           ELSE
490               timeBoundUnbound(j) = 0
491               timeTypeBond(j) = 0
492           END IF
493
494           ! Add together the forces on the toxins
495           dUdxTox(j) = SUM(FbindMag*delXToxUnit,1) + SUM(fljxTox,1) -
    fljxToxTox(j) !(maxTox)
496           dUdyTox(j) = SUM(FbindMag*delYToxUnit,1) + SUM(fljyTox,1) -
    fljyToxTox(j)
497           dUdzTox(j) = SUM(FbindMag*delZToxUnit,1) + SUM(fljzTox,1) -
    fljzToxTox(j)
498
```

```
499         ! Add together all of the lj forces from each toxin on the
     inhibitor beads
500         fljxToxTot = fljxToxTot+fljxTox
501         fljyToxTot = fljyToxTot+fljyTox
502         fljzToxTot = fljzToxTot+fljzTox
503       END DO
504   ELSE
505     DO j = 1,nTox
506       dUdxTox(j) = - fljxToxTox(j) !(maxTox)
507         dUdyTox(j) = - fljyToxTox(j)
508         dUdzTox(j) = - fljzToxTox(j)
509       END DO
510   END IF
511
512   IF (beads > 0) THEN
513     ! Add together the forces on the polymer
514     dUdx = -(fspringx + fwx + fljx + fbindx + fljxToxTot)
515     dUdy = -(fspringy + fwy + fljy + fbindy + fljyToxTot)
516     dUdz = -(fspringz + fwz + fljz + fbindz + fljzToxTot)
517
518     ! Calculate the next position of the polymer
519     rx = rx + (Vf-dUdx/zeta)*deltaT + SQRT(2.0_DP*D)*deltaWt(1,:)
520     ry = ry + (Vf-dUdy/zeta)*deltaT + SQRT(2.0_DP*D)*deltaWt(2,:)
521     rz = rz + (Vf-dUdz/zeta)*deltaT + SQRT(2.0_DP*D)*deltaWt(3,:)
522   END IF
523
524   ! Calculate the next position of the toxin
525   posTox(1,:) = posTox(1,:) + (Vf-dUdxTox/zeta)*deltaT + SQRT(2.0_DP*DTox)
       *deltaWtTox(1,:)
526   posTox(2,:) = posTox(2,:) + (Vf-dUdyTox/zeta)*deltaT + SQRT(2.0_DP*DTox)
       *deltaWtTox(2,:)
527   posTox(3,:) = posTox(3,:) + (Vf-dUdzTox/zeta)*deltaT + SQRT(2.0_DP*DTox)
       *deltaWtTox(3,:)
```

```
528
529 ! Reestablish the boundary conditions for the polymer
530   DO j = 1,beads
531
532       IF (rx(j) > 0.5*sizeBox) THEN
533           rx(j) = rx(j) - sizeBox
534       ELSE IF (rx(j) < -0.5*sizeBox) THEN
535           rx(j) = rx(j) + sizeBox
536       END IF
537
538       IF (ry(j) > 0.5*sizeBox) THEN
539           ry(j) = ry(j) - sizeBox
540       ELSE IF (ry(j) < -0.5*sizeBox) THEN
541           ry(j) = ry(j) + sizeBox
542       END IF
543
544       IF (rz(j) > 0.5*sizeBox) THEN
545           rz(j) = rz(j) - sizeBox
546       ELSE IF (rz(j) < -0.5*sizeBox) THEN
547           rz(j) = rz(j) + sizeBox
548       END IF
549
550   END DO
551
552   ! Reestablish the boundary conditions for the toxin
553   DO j = 1,nTox
554
555       IF (posTox(1,j) > 0.5*sizeBox) THEN
556           posTox(1,j) = posTox(1,j) - sizeBox
557       ELSE IF (posTox(1,j) < -0.5*sizeBox) THEN
558           posTox(1,j) = posTox(1,j) + sizeBox
559       END IF
560
```

```
561         IF (posTox(2,j) > 0.5*sizeBox) THEN
562             posTox(2,j) = posTox(2,j) - sizeBox
563         ELSE IF (posTox(2,j) < -0.5*sizeBox) THEN
564             posTox(2,j) = posTox(2,j) + sizeBox
565         END IF
566
567         IF (posTox(3,j) > 0.5*sizeBox) THEN
568             posTox(3,j) = posTox(3,j) - sizeBox
569         ELSE IF (posTox(3,j) < -0.5*sizeBox) THEN
570             posTox(3,j) = posTox(3,j) + sizeBox
571         END IF
572
573     END DO
574
575     ! Save things to files at specified intervals
576   IF (mod(i,checkBindingInterval) == 0) THEN
577
578         OPEN (UNIT=10,FILE="dataOut/idxToxBnd.dat"&
579              &,STATUS="unknown",POSITION="append")
580         WRITE (UNIT=10,FMT=*) idxToxBnd
581         CLOSE (10)
582
583         OPEN (UNIT=10,FILE="dataOut/numBnd.dat"&
584              &,STATUS="unknown",POSITION="append")
585         WRITE (UNIT=10,FMT=*) SUM(timeBoundUnbound)
586         CLOSE (10)
587
588     END IF
589
590   IF (mod(i,d_wrt) == 0) THEN
591
592     ! Write out the toxin position
593     OPEN (UNIT=20,FILE="dataOut/posToxX.dat"&
```

```
594             &,STATUS="unknown",POSITION="append")
595         WRITE (UNIT=20,FMT=*) posTox(1,:)
596         CLOSE (20)
597
598         OPEN (UNIT=20,FILE="dataOut/posToxY.dat"&
599             &,STATUS="unknown",POSITION="append")
600         WRITE (UNIT=20,FMT=*) posTox(2,:)
601         CLOSE (20)
602
603         OPEN (UNIT=20,FILE="dataOut/posToxZ.dat"&
604             &,STATUS="unknown",POSITION="append")
605         WRITE (UNIT=20,FMT=*) posTox(3,:)
606         CLOSE (20)
607
608         ! Write out the polymer position
609         OPEN (UNIT=20,FILE="dataOut/confX.dat"&
610             &,STATUS="unknown",POSITION="append")
611         WRITE (UNIT=20,FMT=*) rx
612         CLOSE (20)
613
614         OPEN (UNIT=20,FILE="dataOut/confY.dat"&
615             &,STATUS="unknown",POSITION="append")
616         WRITE (UNIT=20,FMT=*) ry
617         CLOSE (20)
618
619         OPEN (UNIT=20,FILE="dataOut/confZ.dat"&
620             &,STATUS="unknown",POSITION="append")
621         WRITE (UNIT=20,FMT=*) rz
622         CLOSE (20)
623
624   END IF
625
626     IF (mod(i,100000) == 0) THEN
```

```
627
628         OPEN (UNIT=20,FILE="dataOut/counter.dat",STATUS="unknown")
629         REWIND (20)
630         WRITE (20,FMT='(I8)') i
631         CLOSE (20)
632
633         OPEN (UNIT=20,FILE=restart_conf,STATUS="unknown")
634         REWIND (UNIT=20)
635         DO j=1,beads
636            WRITE (UNIT=20,FMT=*) rx(j),ry(j),rz(j)
637         END DO
638         CLOSE (20)
639
640     END IF
641
642   IF (mod(i,1000000) == 0) THEN
643     print *, "Step ",i,"of ",totSteps
644   END IF
645
646   END DO
647
648
649 ! Deallocate all of our variables
650 DEALLOCATE (rx,ry,rz,posTox)
651 DEALLOCATE (fljx,fljy,fljz,fljxTox,fljyTox,fljzTox)
652 DEALLOCATE (fljxToxTot,fljyToxTot,fljzToxTot,fljxToxTox,fljyToxTox,
      fljzToxTox)
653 DEALLOCATE (fspringx,fspringy,fspringz)
654 DEALLOCATE (fspringxTox,fspringyTox,fspringzTox)
655 DEALLOCATE (fbindx,fbindy,fbindz,FbindMag)
656 DEALLOCATE (prevOmegaTot,prevOmega,newOmega)
657 DEALLOCATE (timeBoundUnbound)
658 DEALLOCATE (dRTox,delXToxUnit,delYToxUnit,delZToxUnit)
```

```
659 DEALLOCATE (deltaWt,deltaWtTox,beadsBound)
660 DEALLOCATE (dUdx,dUdy,dUdz,dUdxTox,dUdyTox,dUdzTox)
661 DEALLOCATE (nNghbrsPP,nghbrListPP)
662 DEALLOCATE (fwx,fwy,fwz)
663
664 print *, "Done!"
665
666 END program toxinSolubilityNVT
```

B.2 routinesMultTox.f95

```
1  MODULE routinesMultTox
2
3    IMPLICIT NONE
4
5  CONTAINS
6
7    SUBROUTINE worm_like(fwx,fwy,fwz,rx,ry,rz)
8
9    !This subroutine uses a DOT product interaction between adjacent bond
       vectors
10   !to implement the WLC model
11
12   USE parameters
13
14   IMPLICIT NONE
15
16   INTEGER :: i,n
17
18   REAL(DP) :: delta,constw
19   REAL(DP) :: X_kj,Y_kj,Z_kj,X_ij,Y_ij,Z_ij
20   REAL(DP) :: X_kj_Adj,Y_kj_Adj,Z_kj_Adj,X_ij_Adj,Y_ij_Adj,Z_ij_Adj
21   REAL(DP), DIMENSION(:), INTENT(IN):: rx,ry,rz
```

```
22  REAL(DP), DIMENSION(:), INTENT(OUT) :: fwx,fwy,fwz
23
24  fwx=0.0
25  fwy=0.0
26  fwz=0.0
27
28  constw=gammaw/(sigma**2)
29
30  IF (Nmer == 1) THEN
31    print *, fwx
32    print *, "forgot to change isPolymer to 0"
33
34  ELSE IF (Nmer == beads) THEN
35
36    DO i=2,SIZE(rx)-1
37
38    X_kj = rx(i+1)-rx(i)
39    Y_kj = ry(i+1)-ry(i)
40    Z_kj = rz(i+1)-rz(i)
41
42    X_ij = rx(i-1)-rx(i)
43    Y_ij = ry(i-1)-ry(i)
44    Z_ij = rz(i-1)-rz(i)
45
46    ! Correct for periodic boundary conditions
47    X_kj_Adj = X_kj - sizeBox*NINT(X_kj/sizeBox)
48    Y_kj_Adj = Y_kj - sizeBox*NINT(Y_kj/sizeBox)
49    Z_kj_Adj = Z_kj - sizeBox*NINT(Z_kj/sizeBox)
50
51    X_ij_Adj = X_ij - sizeBox*NINT(X_ij/sizeBox)
52    Y_ij_Adj = Y_ij - sizeBox*NINT(Y_ij/sizeBox)
53    Z_ij_Adj = Z_ij - sizeBox*NINT(Z_ij/sizeBox)
54
```

```
55    fwx(i) = fwx(i)+ X_kj_Adj + X_ij_Adj!x32-x12
56    fwy(i) = fwy(i)+ Y_kj_Adj + Y_ij_Adj
57    fwz(i) = fwz(i)+ Z_kj_Adj + Z_ij_Adj
58
59    fwx(i-1) = fwx(i-1)-X_kj_Adj ! -x32
60    fwy(i-1) = fwy(i-1)-Y_kj_Adj
61    fwz(i-1) = fwz(i-1)-Z_kj_Adj
62
63    fwx(i+1) = -X_ij_Adj ! -x12
64    fwy(i+1) = -Y_ij_Adj
65    fwz(i+1) = -Z_ij_Adj
66
67    END DO
68
69  ELSE IF (Nmer < beads) THEN
70
71    DO n = 1,beads/Nmer
72      DO i=Nmer*(n-1)+2,Nmer*n-1
73
74        X_kj = rx(i+1)-rx(i)
75        Y_kj = ry(i+1)-ry(i)
76        Z_kj = rz(i+1)-rz(i)
77
78        X_ij = rx(i-1)-rx(i)
79        Y_ij = ry(i-1)-ry(i)
80        Z_ij = rz(i-1)-rz(i)
81
82        ! Correct for periodic boundary conditions
83        X_kj_Adj = X_kj - sizeBox*NINT(X_kj/sizeBox)
84        Y_kj_Adj = Y_kj - sizeBox*NINT(Y_kj/sizeBox)
85        Z_kj_Adj = Z_kj - sizeBox*NINT(Z_kj/sizeBox)
86
87        X_ij_Adj = X_ij - sizeBox*NINT(X_ij/sizeBox)
```

```
88        Y_ij_Adj = Y_ij - sizeBox*NINT(Y_ij/sizeBox)
89        Z_ij_Adj = Z_ij - sizeBox*NINT(Z_ij/sizeBox)
90
91        fwx(i) = fwx(i)+ X_kj_Adj + X_ij_Adj !rx(i+1)+rx(i-1)-2.0*rx(i) !
     x32-x12
92        fwy(i) = fwy(i)+ Y_kj_Adj + Y_ij_Adj !ry(i+1)+ry(i-1)-2.0*ry(i) !
     (derivative of r32.r12 wrt r2?)
93        fwz(i) = fwz(i)+ Z_kj_Adj + Z_ij_Adj !rz(i+1)+rz(i-1)-2.0*rz(i)
94
95        fwx(i-1) = fwx(i-1)-X_kj_Adj ! -x32
96        fwy(i-1) = fwy(i-1)-Y_kj_Adj
97        fwz(i-1) = fwz(i-1)-Z_kj_Adj
98
99        fwx(i+1) = -X_ij_Adj ! -x12
100       fwy(i+1) = -Y_ij_Adj
101       fwz(i+1) = -Z_ij_Adj
102
103     END DO
104
105    END DO
106
107 END IF
108
109
110 fwx=constw*fwx
111 fwy=constw*fwy
112 fwz=constw*fwz
113
114
115 END SUBROUTINE worm_like
116
117 SUBROUTINE updateNeighborListPP(rx,ry,rz,nNghbrs,nghbrList)
118
```

```
119     USE parameters
120
121     IMPLICIT NONE
122
123     REAL(DP), DIMENSION(:), INTENT(IN) :: rx,ry,rz
124     INTEGER, DIMENSION(:), INTENT(OUT) :: nNghbrs(beads)
125     INTEGER, DIMENSION(:,:), INTENT(OUT) :: nghbrList(beads,beads)
126
127     REAL(DP) :: deltaX, deltaY, deltaZ, deltaXAdj, deltaYAdj, deltaZAdj
128     REAL(DP) :: delta, cutoff
129     INTEGER :: i,j
130
131     nNghbrs = 0
132     nghbrList = 0
133     cutoff = (LJcutoff+1.0_DP)*sigma
134
135     DO i=1,SIZE(rx)
136        DO j=i+1,SIZE(rx)
137
138           deltaX = rx(j)-rx(i)
139           deltaY = ry(j)-ry(i)
140           deltaZ = rz(j)-rz(i)
141
142           ! Correct for periodic boundary conditions
143           deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
144           deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
145           deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
146
147           delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
148
149           ! if you are within the cutoff distance you are a neighbor
150           IF (delta < cutoff) THEN
151           ! add one to the number of neighbors beads i and j have
```

```
152             nNghbrs(i) = nNghbrs(i)+1
153             nNghbrs(j) = nNghbrs(j)+1
154           ! then add bead i to j's neighbor list and vice versa
155             nghbrList(i,nNghbrs(i)) = j
156             nghbrList(j,nNghbrs(j)) = i
157           END IF
158
159       END DO
160     END DO
161
162   END SUBROUTINE updateNeighborListPP
163
164   SUBROUTINE updateNeighborListToxTox(rx,ry,rz,nNghbrs,nghbrList)
165
166     USE parameters
167
168     IMPLICIT NONE
169
170     REAL(DP), DIMENSION(:), INTENT(IN) :: rx,ry,rz
171     INTEGER, DIMENSION(:), INTENT(OUT) :: nNghbrs(maxTox)
172     INTEGER, DIMENSION(:,:), INTENT(OUT) :: nghbrList(maxTox,maxTox)
173
174     REAL(DP) :: deltaX, deltaY, deltaZ, deltaXAdj, deltaYAdj, deltaZAdj
175     REAL(DP) :: delta, cutoff
176     INTEGER :: i,j
177
178     nNghbrs = 0
179     nghbrList = 0
180     cutoff = (LJcutoff+1.0_DP)*sigmaToxTox
181
182     DO i=1,SIZE(rx)
183        DO j=i+1,SIZE(rx)
184
```

```
185         deltaX = rx(j)-rx(i)
186         deltaY = ry(j)-ry(i)
187         deltaZ = rz(j)-rz(i)
188
189         ! Correct for periodic boundary conditions
190         deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
191         deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
192         deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
193
194         delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
195
196         ! if you are within the cutoff distance you are a neighbor
197         IF (delta < cutoff) THEN
198         ! add one to the number of neighbors beads i and j have
199           nNghbrs(i) = nNghbrs(i)+1
200           nNghbrs(j) = nNghbrs(j)+1
201         ! then add bead i to j's neighbor list and vice versa
202           nghbrList(i,nNghbrs(i)) = j
203           nghbrList(j,nNghbrs(j)) = i
204         END IF
205
206     END DO
207   END DO
208
209 END SUBROUTINE updateNeighborListToxTox
210
211 SUBROUTINE updateNeighborListPolyTox(rx,ry,rz,tx,ty,tz,nNghbrs,nghbrList
     )
212
213 USE parameters
214
215   IMPLICIT NONE
216
```

```
217     REAL(DP), DIMENSION(:), INTENT(IN) :: rx,ry,rz,tx,ty,tz
218     INTEGER, DIMENSION(:), INTENT(OUT) :: nNghbrs(maxTox)
219     INTEGER, DIMENSION(:,:), INTENT(OUT) :: nghbrList(maxTox,beads)
220
221     REAL(DP) :: deltaX, deltaY, deltaZ, deltaXAdj, deltaYAdj, deltaZAdj
222     REAL(DP) :: delta, cutoff
223     INTEGER :: i,j
224
225     nNghbrs = 0
226     nghbrList = 0
227     cutoff = (LJcutoff+1.0_DP)*sigmaToxPoly
228
229     DO i=1,SIZE(tx)
230        DO j=1,SIZE(rx)
231
232           deltaX = rx(j)-tx(i)
233           deltaY = ry(j)-ty(i)
234           deltaZ = rz(j)-tz(i)
235
236           ! Correct for periodic boundary conditions
237           deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
238           deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
239           deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
240
241           delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
242
243           ! if you are within the cutoff distance you are a neighbor
244           IF (delta < cutoff) THEN
245           ! add one to the number of neighbors toxin i has
246             nNghbrs(i) = nNghbrs(i)+1
247           ! then add bead j to i's neighbor list
248             nghbrList(i,nNghbrs(i)) = j
249           END IF
```

```
250
251       END DO
252     END DO
253
254   END SUBROUTINE updateNeighborListPolyTox
255
256   SUBROUTINE rljmodNLBC(fljx,fljy,fljz,rx,ry,rz,nNghbrs,nghbrList)
257     ! Takes into account boundary conditions and neighbor list
258
259     USE parameters
260
261     IMPLICIT NONE
262
263     ! Declare constants
264     INTEGER :: i,j
265
266     REAL(DP) :: delta,const
267     REAL(DP) :: deltaX, deltaY, deltaZ, deltaXAdj, deltaYAdj, deltaZAdj
268     REAL(DP), DIMENSION(:), INTENT(IN) :: rx,ry,rz
269     INTEGER, DIMENSION(:), INTENT(IN) :: nNghbrs
270     INTEGER, DIMENSION(:,:), INTENT(IN) :: nghbrList
271     REAL(DP), DIMENSION(:), INTENT(OUT) :: fljx,fljy,fljz
272     REAL(DP) :: fx,fy,fz
273
274     fx=0.0
275     fy=0.0
276     fz=0.0
277
278     fljx=0.0
279     fljy=0.0
280     fljz=0.0
281
282     ! Loop through all the bead pairs and see what the LJ potential is
```

```
  between them
283   DO i=1,SIZE(nNghbrs)
284      DO j=1,nNghbrs(i)
285
286         deltaX = rx(nghbrList(i,j))-rx(i)
287         deltaY = ry(nghbrList(i,j))-ry(i)
288         deltaZ = rz(nghbrList(i,j))-rz(i)
289
290         ! Correct for periodic boundary conditions
291         deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
292         deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
293         deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
294
295         delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
296
297         IF (delta < LJcutoff*sigma) THEN
298
299           const = ((sigma/delta)**12-(sigma/delta)**6)&
300                &*(epsilon/delta)/delta
301
302           fx = - const*deltaXAdj
303           fy = - const*deltaYAdj
304           fz = - const*deltaZAdj
305
306         ELSE
307
308            fx=0.0
309            fy=0.0
310            fz=0.0
311
312         END IF
313
314         fljx(i) = fljx(i)+fx
```

```
315           fljy(i) = fljy(i)+fy
316           fljz(i) = fljz(i)+fz
317
318         END DO
319       END DO
320
321   END SUBROUTINE rljmodNLBC
322
323   SUBROUTINE rljToxToxNLBC(fljx,fljy,fljz,rx,ry,rz,nNghbrs,nghbrList)
324     ! Takes into account boundary conditions and neighbor list
325     ! Does not work with dynamically adjusting toxin concentration
326
327     USE parameters
328
329     IMPLICIT NONE
330
331     ! Declare constants
332     INTEGER :: i,j
333
334     REAL(DP) :: delta,const
335     REAL(DP) :: deltaX, deltaY, deltaZ, deltaXAdj, deltaYAdj, deltaZAdj
336     REAL(DP), DIMENSION(:), INTENT(IN) :: rx,ry,rz
337     INTEGER, DIMENSION(:), INTENT(IN) :: nNghbrs
338     INTEGER, DIMENSION(:,:), INTENT(IN) :: nghbrList
339     REAL(DP), DIMENSION(:), INTENT(OUT) :: fljx,fljy,fljz
340     REAL(DP) :: fx,fy,fz
341
342     fx=0.0
343     fy=0.0
344     fz=0.0
345
346     fljx=0.0
347     fljy=0.0
```

```
348    fljz=0.0
349
350    ! Loop through all the bead pairs and see what the LJ potential is
       between them
351    DO i=1,SIZE(nNghbrs)
352       DO j=1,nNghbrs(i)
353
354          deltaX = rx(nghbrList(i,j))-rx(i)
355          deltaY = ry(nghbrList(i,j))-ry(i)
356          deltaZ = rz(nghbrList(i,j))-rz(i)
357
358          ! Correct for periodic boundary conditions
359          deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
360          deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
361          deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
362
363          delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
364
365          IF (delta < LJcutoff*sigmaToxTox) THEN
366
367            const = ((sigmaToxTox/delta)**12-(sigmaToxTox/delta)**6)&
368                 &*(epsToxTox/delta)/delta
369
370            fx = - const*deltaXAdj
371            fy = - const*deltaYAdj
372            fz = - const*deltaZAdj
373
374          ELSE
375
376             fx=0.0
377             fy=0.0
378             fz=0.0
379
```

```
380           END IF
381
382           fljx(i) = fljx(i)+fx
383           fljy(i) = fljy(i)+fy
384           fljz(i) = fljz(i)+fz
385
386        END DO
387     END DO
388
389   END SUBROUTINE rljToxToxNLBC
390
391   SUBROUTINE rljToxNLBC(fljx,fljy,fljz,deltaTox,rx,ry,rz,posTox,
       delXToxUnit,delYToxUnit,delZToxUnit,nNghbrs,nghbrList)
392
393     USE parameters
394
395     IMPLICIT NONE
396
397     ! Declare constants
398     INTEGER :: i,j
399     INTEGER, INTENT(IN) :: nNghbrs
400     INTEGER, DIMENSION(:), INTENT(IN) :: nghbrList
401
402     REAL(DP) :: delta,const
403     REAL(DP) :: deltaX, deltaY, deltaZ, deltaXAdj, deltaYAdj, deltaZAdj
404     REAL(DP), DIMENSION(:), INTENT(IN) :: rx,ry,rz,posTox
405     REAL(DP), DIMENSION(:), INTENT(OUT) :: fljx,fljy,fljz,deltaTox(beads)
406     REAL(DP), DIMENSION(:), INTENT(OUT) :: delXToxUnit(beads), delYToxUnit
       (beads), delZToxUnit(beads)
407     REAL(DP) :: fx,fy,fz
408
409     fx=0.0
410     fy=0.0
```

```
411    fz=0.0
412
413    deltaTox=0.0
414
415    delXToxUnit=0.0
416    delYToxUnit=0.0
417    delZToxUnit=0.0
418
419    fljx=0.0
420    fljy=0.0
421    fljz=0.0
422
423    ! Loop through all the bead pairs and see what the LJ potential is
   between them
424    DO i=1,nNghbrs
425
426      deltaX = rx(nghbrList(i))-posTox(1)
427      deltaY = ry(nghbrList(i))-posTox(2)
428      deltaZ = rz(nghbrList(i))-posTox(3)
429
430      ! Correct for periodic boundary conditions
431      deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
432      deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
433      deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
434
435      delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
436      deltaTox(nghbrList(i)) = delta
437
438      ! Make unit vectors for later use in bigger function
439      delXToxUnit(nghbrList(i)) = deltaXAdj/delta
440      delYToxUnit(nghbrList(i)) = deltaYAdj/delta
441      delZToxUnit(nghbrList(i)) = deltaZAdj/delta
442
```

```
443      IF (delta < LJcutoff*sigmaToxPoly) THEN
444
445         const = ((sigmaToxPoly/delta)**12-(sigmaToxPoly/delta)**6)&
446              &*(epsToxPoly/delta)/delta
447
448         ! check to make sure the negative here is right!
449         fx = - const*deltaXAdj
450         fy = - const*deltaYAdj
451         fz = - const*deltaZAdj
452
453      ELSE
454
455         fx=0.0
456         fy=0.0
457         fz=0.0
458
459      END IF
460      ! This part still needs work
461      fljx(nghbrList(i)) = fljx(nghbrList(i))-fx
462      fljy(nghbrList(i)) = fljy(nghbrList(i))-fy
463      fljz(nghbrList(i)) = fljz(nghbrList(i))-fz
464
465    END DO
466  END SUBROUTINE rljToxNLBC
467
468  SUBROUTINE springForces(fspringx,fspringy,fspringz,rx,ry,rz)
469
470    USE parameters
471
472    IMPLICIT NONE
473
474    INTEGER :: i
475
```

```
476     REAL(DP) :: delta,const
477     REAL(DP) :: deltaX, deltaY, deltaZ, deltaXAdj, deltaYAdj, deltaZAdj
478     REAL(DP), DIMENSION(:), INTENT(IN) :: rx,ry,rz
479     REAL(DP), DIMENSION(:), INTENT(OUT) :: fspringx,fspringy,fspringz
480     REAL(DP) :: fsxMid (beads), fsyMid (beads), fszMid (beads)
481     REAL(DP) :: fsx (beads), fsy (beads), fsz (beads)
482
483     fspringx = 0.0
484     fspringy = 0.0
485     fspringz = 0.0
486
487     ! Compute spring forces from being connected to neighboring beads
488     ! Compute End Forces (because only connected to one neighbor)
489     IF (Nmer == 1) THEN
490       print *, fspringx
491       print *, "forgot to change isPolymer to 0"
492
493     ELSE IF (Nmer == beads) THEN
494
495       ! Do first bead
496       i=1
497
498       deltaX = rx(i+1)-rx(i)
499       deltaY = ry(i+1)-ry(i)
500       deltaZ = rz(i+1)-rz(i)
501
502       ! Correct for periodic boundary conditions
503       deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
504       deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
505       deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
506
507       delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
508       const = k_sp*(delta-l_sp)/delta
```

```
509
510       fspringx(i) = const*(deltaXAdj)
511       fspringy(i) = const*(deltaYAdj)
512       fspringz(i) = const*(deltaZAdj)
513
514       ! Do force on the last bead
515       i=beads-1
516
517       deltaX = rx(i+1)-rx(i)
518       deltaY = ry(i+1)-ry(i)
519       deltaZ = rz(i+1)-rz(i)
520
521       ! Correct for periodic boundary conditions
522       deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
523       deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
524       deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
525
526       delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
527       const = k_sp*(delta-l_sp)/(delta)
528
529       fspringx(i+1) = - const*(deltaXAdj)
530       fspringy(i+1) = - const*(deltaYAdj)
531       fspringz(i+1) = - const*(deltaZAdj)
532
533    ELSE IF (Nmer < beads) THEN
534
535       ! Calculate bonding of first beads
536       DO i=1,beads-1,Nmer
537
538         deltaX = rx(i+1)-rx(i)
539         deltaY = ry(i+1)-ry(i)
540         deltaZ = rz(i+1)-rz(i)
541
```

```
542         ! Correct for periodic boundary conditions
543         deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
544         deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
545         deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
546
547         delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
548         const = k_sp*(delta-l_sp)/(delta)
549
550         fspringx(i) = const*(deltaXAdj)
551         fspringy(i) = const*(deltaYAdj)
552         fspringz(i) = const*(deltaZAdj)
553
554       END DO
555
556       ! Calculate bonding of last beads
557       DO i=(Nmer-1),(beads-1),Nmer
558
559         deltaX = rx(i+1)-rx(i)
560         deltaY = ry(i+1)-ry(i)
561         deltaZ = rz(i+1)-rz(i)
562
563         ! Correct for periodic boundary conditions
564         deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
565         deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
566         deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
567
568         delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
569         const = k_sp*(delta-l_sp)/(delta)
570
571         fspringx(i+1) = - const*(deltaXAdj)
572         fspringy(i+1) = - const*(deltaYAdj)
573         fspringz(i+1) = - const*(deltaZAdj)
574
```

```
575       END DO
576
577     END IF
578
579     ! Compute middle bead connectivity forces
580     IF (Nmer > 2) THEN
581       DO i=2,beads-1
582
583         deltaX = rx(i+1)-rx(i)
584         deltaY = ry(i+1)-ry(i)
585         deltaZ = rz(i+1)-rz(i)
586
587         ! Correct for periodic boundary conditions
588         deltaXAdj = deltaX - sizeBox*NINT(deltaX/sizeBox)
589         deltaYAdj = deltaY - sizeBox*NINT(deltaY/sizeBox)
590         deltaZAdj = deltaZ - sizeBox*NINT(deltaZ/sizeBox)
591
592         delta = (deltaXAdj**2+deltaYAdj**2+deltaZAdj**2)**0.5
593         const = k_sp*(delta-l_sp)/(delta)
594
595         ! If you are the end of an Nmer, make the force to the right 0
596         ! Otherwise proceed normally
597         IF (MOD(i,Nmer) /= 0) THEN
598           fsxMid(i) = const*(deltaXAdj)
599           fsyMid(i) = const*(deltaYAdj)
600           fszMid(i) = const*(deltaZAdj)
601         ELSE
602           fsxMid(i) = 0.0
603           fsyMid(i) = 0.0
604           fszMid(i) = 0.0
605         END IF
606
607       END DO
```

```
608
609       ! Add in the forces pulling back to the left
610       ! This is 0 force pulling back to the left for left most end beads
611
612         fsxMid(1) = fspringx(1)
613         fsyMid(1) = fspringy(1)
614         fszMid(1) = fspringz(1)
615
616         fsxMid(beads) = fspringx(beads)
617         fsyMid(beads) = fspringy(beads)
618         fszMid(beads) = fspringz(beads)
619
620         ! Reassign the middle forces to a temporary variable
621         fsx = fsxMid
622         fsy = fsyMid
623         fsz = fszMid
624
625       DO i=2,beads-1
626         ! Without the temp variable, when I did -fsxMid(i-1) errors start
      to accummulate as you move right along the chain because you modify i-1
       and then you use the modified value instead of the original value
627        fsxMid(i) = fsxMid(i) - fsx(i-1)
628        fsyMid(i) = fsyMid(i) - fsy(i-1)
629        fszMid(i) = fszMid(i) - fsz(i-1)
630
631       END DO
632
633       DO i=1,beads,Nmer
634
635         fsxMid(i) = fspringx(i)
636         fsyMid(i) = fspringy(i)
637         fszMid(i) = fspringz(i)
638
```

```
639       END DO
640
641       DO i=Nmer,beads,Nmer
642
643         fsxMid(i) = fspringx(i)
644         fsyMid(i) = fspringy(i)
645         fszMid(i) = fspringz(i)
646
647       END DO
648
649       fspringx = fsxMid
650       fspringy = fsyMid
651       fspringz = fszMid
652
653     END IF
654
655   END SUBROUTINE springForces
656
657 ! dR is size beads by 1 (cycle through this funciton for each toxin)
658 ! prevOmegaTot is (beads,bindingSites,maxTox)
659 ! prevOmega = (beads,bindingSites,existingToxins(k))
660 ! newOmega = (beads,bindingSites)
661   SUBROUTINE bound(prevOmegaTot, prevOmega, dRTox, FbindMag, newOmega,
      delE_0,delE_UB)
662
663     USE parameters
664     USE functionsEmi
665
666     IMPLICIT NONE
667
668     INTEGER :: i,j,k,cnt,beadToBind,beadBound,randIdx
669     REAL(DP) :: deltaX, deltaY, deltaZ, deltaXAdj, deltaYAdj, deltaZAdj
670
```

```
671     REAL(DP), DIMENSION(:), INTENT(IN) :: dRTox(beads),delE_0(beads),
    delE_UB(beads)
672     INTEGER, DIMENSION(:,:), INTENT(IN) :: prevOmega(beads,bindingSites) !
     how this toxin was previously bound
673     INTEGER, DIMENSION(:,:,:), INTENT(IN) :: prevOmegaTot(beads,
    bindingSites,maxTox) ! how all the other toxins are bound (to prevent
    binding to same bead)
674
675     REAL(DP), DIMENSION(:), INTENT(OUT) :: FbindMag(beads)
676     INTEGER, DIMENSION(:,:), INTENT(OUT) :: newOmega(beads,bindingSites)
677
678     REAL(DP), DIMENSION(:,:) :: oper(beads,bindingSites), alreadyBoundTemp
    (beads,bindingSites)
679     INTEGER, DIMENSION(:) :: alreadyBound(beads), beadsBound(beads)
680     INTEGER, DIMENSION(:), ALLOCATABLE :: alreadyBoundIdx, closeBds,
    possibleBinds
681     REAL(DP), DIMENSION(:), ALLOCATABLE :: randPerm
682     REAL(DP), DIMENSION(:) :: dR_modifier(beads), dRAvailableBeads(beads),
     delE_B(beads)
683     LOGICAL :: mask(beads)
684
685     delE_B = delE_0 + delE_UB
686     newOmega = 0
687     oper = 0.0
688     dR_modifier = 0.0
689
690     ! Generate uniform random numbers so that you can use them later to
    determine binding
691     CALL RANDOM_NUMBER(oper)
692
693     ! Keep track of the beads you've already bound so that you don't get
    more than one tox binding site bound to it
694     alreadyBoundTemp = SUM(prevOmegaTot, 3)
```

```
695   alreadyBound = SUM(alreadyBoundTemp, 2) ! This isn't affected by
      changing the storage of toxins because unbound and nonexistent toxins
      are both 0
696
697   ! Check to make sure nothing has bound to the same bead twice
698   mask = alreadyBound > 1
699   IF (ANY(mask)) THEN
700     print *, "Error: bound to the same bead twice"
701   END IF
702
703   ! Get the indexes of the beads that are bound so that you can keep
      track of them
704   mask = alreadyBound==1
705   alreadyBoundIdx = FIND(mask)
706
707   ! Make the already bound beads too far to be picked up by the close
      beads search (so that they won't be available for new bonds to form)
708   dR_modifier(alreadyBoundIdx) = l_bind*(1.0_DP+reach)*2.0_DP; ! 2 is
      just to guarantee bead is well outside reach of binding site
709   dRAvailableBeads = dRTox + dR_modifier;
710
711   ! Iterate through each binding site and assess whether you bind/unbind
      stay unbound to it
712   DO j=1,bindingSites
713     ! If this binding site is not already bound to something
714     mask = prevOmega(:,j) == 0
715     IF (ALL(mask)) THEN
716       ! If your site is within binding range of any of the beads
717       ! This will have a problem if there is no excluded volume between
      toxins and polymers
718       mask = ((dRAvailableBeads<=(reach)).AND.(dRAvailableBeads>0))
719       IF (ANY(mask)) THEN
720         ! Find which beads are close
```

```
721         closeBds = FIND(mask)
722         ! Check if any of the close beads have a high enough energy to
    jump the binding energy barrier
723         DO i = 1, SIZE(closeBds)
724           IF (oper(i,j) < EXP(-delE_B(closeBds(i)))) THEN
725             newOmega(closeBds(i),j) = 1;
726             ! Make sure bead you bound to is out of reach
727             dRAvailableBeads(closeBds(i)) = dRTox(closeBds(i)) + l_bind
    *(1.0_DP+reach)*2.0_DP;
728           END IF
729         END DO
730         ! Check if you can bind to more than one bead
731         IF (SUM(newOmega(:,j)) > 1) THEN
732             ! Only pick one to bind to randomly
733             !randomize the index of beads you could successfully bind to
734             ALLOCATE (randPerm(SUM(newOmega(:,j))))
735             CALL RANDOM_NUMBER(randPerm)
736
737             randIdx = MAXLOC(randPerm,1)
738             mask = newOmega(:,j) == 1
739             possibleBinds = FIND(mask)
740
741             beadToBind = possibleBinds(randIdx)
742             newOmega(:,j) = 0
743             newOmega(beadToBind,j) = 1
744             ! Make sure the bead you just bound to is out of reach of
    other binding sites
745             dRAvailableBeads(closeBds) = dRTox(closeBds)
746             dRAvailableBeads(beadToBind) = dRTox(beadToBind) + l_bind
    *(1.0_DP+reach)*2.0_DP
747
748
749             ! Deallocate randperm in case it is used again for another
```

```
   binding site
750              DEALLOCATE (randPerm)
751            END IF
752          ! Otherwise, you're not within reach of the beads and everything
   should be unbound
753          ELSE
754            newOmega(:,j) = 0
755          END IF
756
757        ! If you're already bound to something
758        ELSE
759          beadBound = MAXLOC(prevOmega(:,j),1)
760          ! If you don't have enough energy to unbind then stay bound to the
    same bead
761          IF (oper(beadBound,j) > exp(-delE_UB(beadBound))) THEN
762            newOmega(:,j) = prevOmega(:,j)
763          ELSE
764            newOmega(:,j) = 0
765          END IF
766        END IF
767
768      END DO
769
770      ! Calculate the new force
771      beadsBound = SUM(newOmega,2)
772      FbindMag = -k_bind*beadsBound*(dRTox-l_bind)
773
774    END SUBROUTINE bound
775
776  ! This isn't very elegant, creates inhibitor starting positions in s or
    ladder shape, only goes up to 92 beads
777    SUBROUTINE setInitialPos(rx,ry,rz)
778      USE parameters
```

```
779     USE functionsEmi
780
781     IMPLICIT NONE
782
783     REAL(DP), DIMENSION(:), INTENT(OUT) :: rx(beads),ry(beads),rz(beads)
784     REAL(DP) :: rightEdge,leftEdge
785     INTEGER :: lengthRow,i
786     REAL(DP), DIMENSION(:) :: iPosTmp(100)
787     INTEGER :: maxBeads = 64
788
789     rightEdge = (floor(sizeBox/2.0)-1.0)
790     leftEdge = -(floor(sizeBox/2.0)-1.0)
791     lengthRow = rightEdge-leftEdge+1
792
793     ! Set up an array of x positions for your inhibitor to start from
794     DO i=1,lengthRow
795       iPosTmp(i) = leftEdge+l_sp*(i-1.0)
796     END DO
797     DO i=lengthRow+1,lengthRow+4
798       iPosTmp(i) = rightEdge
799     END DO
800     DO i=1,lengthRow
801       iPosTmp(i+lengthRow+4) = rightEdge-l_sp*(i-1.0)
802     END DO
803     DO i=2*lengthRow+4,2*lengthRow+8
804       iPosTmp(i) = leftEdge
805     END DO
806     DO i=1,lengthRow
807       iPosTmp(i+2*lengthRow+8) = leftEdge+l_sp*(i-1.0)
808     END DO
809     DO i=3*lengthRow+8,3*lengthRow+12
810       iPosTmp(i) = rightEdge
811     END DO
```

```
812     DO i=1,lengthRow
813       iPosTmp(i+3*lengthRow+12) = rightEdge-l_sp*(i-1.0)
814     END DO
815     DO i=4*lengthRow+12,4*lengthRow+16
816       iPosTmp(i) = leftEdge
817     END DO
818
819     rx = iPosTmp(1:beads)
820
821     ! Set y positions
822     iPosTmp = 0.0
823     DO i=1,lengthRow
824       iPosTmp(i) = leftEdge
825     END DO
826     DO i=lengthRow+1,lengthRow+4
827       iPosTmp(i) = leftEdge+(i-lengthRow)
828     END DO
829     DO i=1,lengthRow
830       iPosTmp(i+lengthRow+4) = leftEdge+5
831     END DO
832     DO i=2*lengthRow+4,2*lengthRow+8
833       iPosTmp(i) = leftEdge+(i+1-2*lengthRow)
834     END DO
835     DO i=1,lengthRow
836       iPosTmp(i+2*lengthRow+8) = leftEdge+10
837     END DO
838     DO i=3*lengthRow+8,3*lengthRow+12
839       iPosTmp(i) = leftEdge+(i+2-3*lengthRow)
840     END DO
841     DO i=1,lengthRow
842       iPosTmp(i+3*lengthRow+12) = leftEdge+15
843     END DO
844     DO i=4*lengthRow+12,4*lengthRow+15
```

```
845       iPosTmp(i) = leftEdge+(i+3-4*lengthRow)
846     END DO
847
848     ry = iPosTmp(1:beads)
849
850     ! Set z positions
851     rz = -4.0 !0.0
852
853     IF (beads > maxBeads) THEN
854      DO i = 1,(beads-maxBeads)
855         rx(i+maxBeads) = rx(i)
856         ry(i+maxBeads) = ry(i)
857         rz(i+maxBeads) = rz(i)-2.0
858      END DO
859     END IF
860
861   END SUBROUTINE setInitialPos
862
863   SUBROUTINE setInitialPosTox(rx,ry,rz,px,py,pz)
864
865     USE parameters
866     USE functionsEmi
867
868     IMPLICIT NONE
869
870     INTEGER :: iTox, otherTox, ibds
871     REAL(DP) :: dX,dY,dZ,dXAdj,dYAdj,dZAdj,dToxTmp
872     REAL(DP), DIMENSION(:), INTENT(OUT) :: rx(nTox),ry(nTox),rz(nTox)
873     REAL(DP), DIMENSION(:), INTENT(IN) :: px(beads),py(beads),pz(beads)
874
875
876     DO iTox=1,nTox
877
```

```
878 100   CALL RANDOM_NUMBER(rx(iTox))
879       CALL RANDOM_NUMBER(ry(iTox))
880       CALL RANDOM_NUMBER(rz(iTox))
881
882     rx(iTox) = rx(iTox)*sizeBox-sizeBox/2.0_DP
883     ry(iTox) = ry(iTox)*sizeBox-sizeBox/2.0_DP
884     rz(iTox) = rz(iTox)*sizeBox-sizeBox/2.0_DP
885
886     ! Check to make sure you didn't just put your new toxin on top of
     other ones
887     DO otherTox=1,iTox-1
888
889           dX = rx(otherTox)-rx(iTox)
890           dY = ry(otherTox)-ry(iTox)
891           dZ = rz(otherTox)-rz(iTox)
892
893           ! Correct for periodic boundary conditions
894           dXAdj = dX - sizeBox*NINT(dX/sizeBox)
895           dYAdj = dY - sizeBox*NINT(dY/sizeBox)
896           dZAdj = dZ - sizeBox*NINT(dZ/sizeBox)
897
898           dToxTmp = SQRT(dXAdj**2.0_DP+dYAdj**2.0_DP+dZAdj**2.0_DP)
899
900           IF (dToxTmp < 1.5_DP*reach) THEN
901             GOTO 100
902           END IF
903     END DO
904     DO ibds=1,beads
905
906           dX = px(ibds)-rx(iTox)
907           dY = py(ibds)-ry(iTox)
908           dZ = pz(ibds)-rz(iTox)
909
```

```
910             ! Correct for periodic boundary conditions
911             dXAdj = dX - sizeBox*NINT(dX/sizeBox)
912             dYAdj = dY - sizeBox*NINT(dY/sizeBox)
913             dZAdj = dZ - sizeBox*NINT(dZ/sizeBox)
914
915             dToxTmp = (dXAdj**2+dYAdj**2+dZAdj**2)**0.5
916
917             IF (dToxTmp < 1.5_DP*reach) THEN
918               GOTO 100
919             END IF
920
921         END DO
922     END DO
923
924   END SUBROUTINE setInitialPosTox
925
926   SUBROUTINE init_random_seed(nRun,seed)
927
928       INTEGER, INTENT(IN) :: nRun
929       INTEGER :: i, n, clock
930       INTEGER, DIMENSION(:), ALLOCATABLE, INTENT(OUT) :: seed
931
932       CALL RANDOM_SEED(size = n)
933       ALLOCATE(seed(n))
934
935       CALL SYSTEM_CLOCK(COUNT=clock)
936
937       seed = clock + 37 * (/ (i - 1, i = 1, n) /) + nRun
938       print *, "seed = ", seed
939       CALL RANDOM_SEED(PUT = seed)
940
941   END SUBROUTINE
942
```

```
SUBROUTINE findAggToxins(dRToxTox,timeBoundUnbound,nToxAgg,closeToxIdx)

  USE parameters

  IMPLICIT NONE

  INTEGER :: i,j
  INTEGER, DIMENSION(:,:) :: closeMaskINT(nTox,nTox)
  INTEGER, DIMENSION(:), INTENT(IN) :: timeBoundUnbound(nTox)
  REAL(DP), DIMENSION(:,:), INTENT(IN) :: dRToxTox
  INTEGER, DIMENSION(:), INTENT(OUT) :: closeToxIdx(nTox)
  INTEGER, INTENT(OUT) :: nToxAgg

  closeMaskINT = 0
  nToxAgg = 0
  closeToxIdx = 0

  DO i=1,nTox
    DO j = i+1,nTox
      IF (AND(dRToxTox(i,j) > 0.0, dRToxTox(i,j) < reach)) THEN
        closeMaskINT(i,j) = 1
        closeMaskINT(j,i) = 1
      END IF
    END DO
  END DO

  closeToxIdx = SUM(closeMaskINT,2)
  closeToxIdx = closeToxIdx + timeBoundUnbound

  DO i=1,nTox
    IF (closeToxIdx(i)>0) THEN
      nToxAgg = nToxAgg+1
      closeToxIdx(i)=1
```

```
976       END IF
977     END DO
978
979   END SUBROUTINE findAggToxins
980
981   SUBROUTINE adjNTox(nToxAgg,toxConc0,nToxNew)
982
983     USE parameters
984
985     IMPLICIT NONE
986
987     REAL(DP), INTENT(IN) :: nToxAgg, toxConc0
988     INTEGER, INTENT(OUT) :: nToxNew
989     REAL(DP) :: toxConc, v0
990     LOGICAL :: didntAdd = .true.
991
992     v0 = sizeBox**3
993     nToxNew = nTox
994     ! Find the current free toxin concentration
995     toxConc = (nTox-nToxAgg)/(v0-nToxAgg*(4.0/3.0)*pi*(l_sp/2.0)**3)
996
997     ! Change the number of free toxins
998       DO WHILE (toxConc < toxConc0)
999         nToxNew = nToxNew+1
1000        toxConc = (nToxNew-nToxAgg)/(v0-nToxAgg*(4.0/3.0)*pi*(l_sp/2.0)
     **3)
1001        didntAdd = .false.
1002      END DO
1003
1004
1005      DO WHILE (AND(toxConc > toxConc0,didntAdd))
1006        nToxNew = nToxNew-1
1007        toxConc = (nToxNew-nToxAgg)/(v0-nToxAgg*(4.0/3.0)*pi*(l_sp/2.0)
```

```
   **3)
1008       END DO
1009
1010       ! Make sure youre always slightly below the original concentration
   instead of oscillating between below and above because of adding and
   subtracting methods
1011       ! so if you added until higher than conc0, subtract one
1012       IF (.NOT.didntAdd) THEN
1013         nToxNew = nToxNew-1
1014       END IF
1015
1016       !print *, "nToxNew", nToxNew
1017     END SUBROUTINE adjNTox
1018
1019 SUBROUTINE squeeze_to_string_long(prefix,inum1,middle,inum2,suffix,ss)
1020 !
    ************************************************************************

1021 ! Needs:
1022 !    subroutine integer_to_string(,)
1023 !
    ************************************************************************

1024     IMPLICIT NONE
1025
1026     INTEGER inum1,inum2
1027     CHARACTER(*) :: prefix, middle, suffix
1028     CHARACTER(60) :: strnum1, strnum2
1029     CHARACTER(*), INTENT(out) :: ss
1030
1031     CALL integer_to_string(inum1,strnum1)
1032     CALL integer_to_string(inum2,strnum2)
1033
```

```
    ss = prefix // TRIM(strnum1) // middle // TRIM(strnum2) // suffix

    RETURN
  END SUBROUTINE squeeze_to_string_long


  SUBROUTINE squeeze_to_string(prefix,inum,suffix,ss)
!
   ************************************************************************

! Needs:
!    subroutine integer_to_string(,)
!
   ************************************************************************

    IMPLICIT NONE

    INTEGER inum
    CHARACTER(*) :: prefix, suffix
    CHARACTER(60) :: strnum
    CHARACTER(*), INTENT(out) :: ss

    CALL integer_to_string(inum,strnum)

    ss = prefix // TRIM(strnum) // suffix

    RETURN
  END SUBROUTINE squeeze_to_string


  SUBROUTINE integer_to_string(jj,ss)
! ***************************************************************
! 11/02/00
```

```
! Converts an integer to a character variable of same value
! Character variable has left over space so it must be trimmed.
! ****************************************************************
    IMPLICIT NONE

    INTEGER :: ii, jj
    INTEGER :: maxdigits, idig, icount
    DOUBLE PRECISION :: mod1, div1, frame
    CHARACTER(*), INTENT(out) :: ss

    frame=DBLE(jj)
    maxdigits=INT(LOG10(frame))+1
    ss=' '
    mod1=10.0
    div1=1.0

    DO ii=1,maxdigits
       idig=INT(MOD(frame,mod1)/div1)
       ss = ACHAR(idig+48) // ss
       div1=mod1
       mod1=10.0*mod1
    END DO

    RETURN
  END SUBROUTINE integer_to_string


SUBROUTINE ARRAY_COPY(src,dest,n_copied,n_not_copied)

  USE parameters

  REAL(DP), DIMENSION (:), INTENT(IN) :: src
  REAL(DP), DIMENSION (:), INTENT(OUT) :: dest
```

```
1096   INTEGER, INTENT(OUT) :: n_copied,n_not_copied
1097
1098   n_copied = MIN(SIZE(src),SIZE(dest))
1099   n_not_copied = SIZE(src)-n_copied
1100   dest(1:n_copied) = src(1:n_copied)
1101
1102
1103 END SUBROUTINE ARRAY_COPY
1104
1105
1106
1107 ! ***************************************************************
1108 ! From Numerical Recipes in Fortran 90 2nd Edition
1109 ! 01/18/18
1110 ! creates a vector of gaussian distributed numbers with zero mean
1111 ! and variance of 1, using RANDOM_NUMBER as the source of uniform
1112 ! random numbers
1113 ! ***************************************************************
1114 ! This is not working, not sure why - implemented one that does it one
       number at a time
1115 SUBROUTINE gasdev_v(harvest)
1116
1117   USE parameters
1118
1119   IMPLICIT NONE
1120
1121   REAL(DP), DIMENSION(:), INTENT(OUT) :: harvest
1122   REAL(DP), DIMENSION(SIZE(harvest)) :: rsq,v1,v2
1123   REAL(DP), ALLOCATABLE, DIMENSION(:), SAVE :: g
1124   INTEGER :: n,ng,nn,m
1125   INTEGER, SAVE :: last_allocated=0
1126   LOGICAL, SAVE :: gaus_stored=.false.
1127   LOGICAL, DIMENSION(SIZE(harvest)) :: mask
```

```
1128
1129  n=SIZE(harvest)
1130
1131  print *, "ENTERING FUNCTION...."
1132  print *, "last_allocated = ", last_allocated
1133  print *, "gaus_stored = ", gaus_stored
1134
1135  IF (n /= last_allocated) THEN
1136    IF (last_allocated /= 0) THEN
1137      DEALLOCATE(g)
1138    END IF
1139    ALLOCATE(g(n))
1140    last_allocated = n
1141    gaus_stored=.false.
1142  END IF
1143
1144  IF (gaus_stored) THEN
1145    harvest=g
1146    gaus_stored=.false.
1147  ELSE
1148    ng = 1
1149
1150    DO
1151
1152      IF (ng > n) THEN
1153        EXIT
1154      END IF
1155      print *, "ng = ", ng
1156
1157      CALL RANDOM_NUMBER(v1(ng:n))
1158      CALL RANDOM_NUMBER(v2(ng:n))
1159      ! Make your uniform random numbers from -1 to 1
1160      v1(ng:n) = 2.0_DP*v1(ng:n)-1.0_DP
```

```
      v2(ng:n) = 2.0_DP*v2(ng:n)-1.0_DP
      rsq(ng:n) = v1(ng:n)**2+v2(ng:n)**2
      print *, "rsq = ", rsq
      print *, "v1 before Array copy = ", v1
      print *, "v2 before array copy = ", v2

      mask(ng:n) = (rsq(ng:n)>0.0 .and. rsq(ng:n)<1.0)
      print *, "mask = ", mask
      print *, "v1(ng:) = ", v1(ng:)
      CALL ARRAY_COPY(PACK(v1(ng:n),mask(ng:n)),v1(ng:),nn,m)
      print *, "result of v1 PACK = ", PACK(v1(ng:n),mask(ng:n))
      print *, "v1 after array copy = ", v1
      v2(ng:ng+nn-1) = PACK(rsq(ng:n),mask(ng:n))
      print *, "v2 after pack = ", v2
      print *, "n copied = ", nn
      print *, "not copied = ", m
      ng=ng+nn

    END DO
    ! Now make the Box-Muller transformation to get two normal deviates
    ! Return the amount needed and save the rest for next time
    rsq=sqrt(-2.0_DP*log(rsq)/rsq)
    harvest=v1*rsq
    g=v2*rsq
    gaus_stored=.true.

  END IF

  print *, "LEAVING FUNCTION...."
  print *, "last_allocated = ", last_allocated
  print *, "gaus_stored = ", gaus_stored
  print *, "LEFT FUNCTION."

```

```
1194 END SUBROUTINE gasdev_v
1195
1196 SUBROUTINE gasdev_s(harvest)
1197
1198   USE parameters
1199
1200   IMPLICIT NONE
1201
1202   REAL(DP), INTENT(OUT) :: harvest
1203   REAL(DP) :: rsq,v1,v2
1204   REAL(DP), SAVE :: g
1205   LOGICAL, SAVE :: gaus_stored=.false.
1206
1207   IF (gaus_stored) THEN
1208     harvest=g
1209     gaus_stored=.false.
1210   ELSE
1211     DO
1212       CALL RANDOM_NUMBER(v1)
1213       CALL RANDOM_NUMBER(v2)
1214
1215       v1 = 2.0_DP*v1-1.0_DP
1216       v2 = 2.0_DP*v2-1.0_DP
1217       rsq = v1**2+v2**2
1218       IF (rsq > 0.0 .and. rsq < 1.0) THEN
1219         EXIT
1220       END IF
1221     END DO
1222     ! Now make the Box-Muller transformation to get two normal deviates
1223     ! Return the amount needed and save the rest for next time
1224     rsq = sqrt(-2.0_DP*log(rsq)/rsq)
1225     harvest = v1*rsq
1226     g = v2*rsq
```

```
    gaus_stored=.true.
  END IF

END SUBROUTINE gasdev_s

SUBROUTINE pick_random(old_idxs, rand_idxs, nIdx)
  ! pick a random index
      IMPLICIT NONE

    INTEGER, DIMENSION(:), INTENT(IN) :: old_idxs
    INTEGER, DIMENSION(:), INTENT(OUT) :: rand_idxs
    INTEGER, DIMENSION(:), ALLOCATABLE :: new_idxs
    INTEGER, INTENT(IN) :: nIdx
    REAL :: r ! random number from 0 to 1
    INTEGER :: r_int,i ! random index

    ALLOCATE (new_idxs(SIZE(old_idxs)))
    new_idxs = old_idxs
    DO i = 1,nIdx

      ! get a random number between 0 and 1 and turn it into an index
      call random_number(r)
      r_int = int(r*(size(new_idxs)-i+1)) + 1

      ! get the number associated with that index and return it (this is
   your actual random index you'll use in the your program)
      rand_idxs(i) = new_idxs(r_int)

      ! return a new matrix without the index you chose
      new_idxs(1:r_int-1) = new_idxs(1:r_int-1)
      new_idxs(r_int:SIZE(new_idxs)-1) = new_idxs((r_int+1):SIZE(new_idxs)
   )

```

```
1258     END DO
1259
1260   END SUBROUTINE pick_random
1261
1262 END MODULE routinesMultTox
```

B.3 functionsEmi.f95

```
1  MODULE functionsEmi
2
3  IMPLICIT NONE
4
5  CONTAINS
6
7    REAL(DP) FUNCTION round(n)
8
9      USE parameters
10
11     IMPLICIT NONE
12
13       REAL(DP), INTENT(IN) :: n
14
15       round = NINT(n)
16
17   END FUNCTION round
18
19   FUNCTION FIND(mask)
20
21     USE parameters
22
23     IMPLICIT NONE
24
25     INTEGER, DIMENSION(:), ALLOCATABLE :: find
```

```
    LOGICAL, DIMENSION(:), INTENT(IN) :: mask
    INTEGER, DIMENSION(:), ALLOCATABLE :: indxs
    INTEGER :: i, cnt

    cnt = COUNT(mask)
      ALLOCATE (indxs(cnt))
    indxs=0
      cnt=1
    DO i=1,SIZE(mask,1)
      IF (mask(i)) THEN
        indxs(cnt)=i
        cnt = cnt+1
      END IF
    END DO

    find = indxs

  END FUNCTION FIND

END MODULE functionsEmi
```

B.4 parameters.f95

```
MODULE parameters

  IMPLICIT NONE
  !
  ! THIS MODULE HAS THE DECLARATION OF MOST OF THE VARIALBES USED
  ! IN THE SIMULATION. THE ONLY VARIABLES THAT ARE IN ANOTHER MODULE
  ! ARE THE RANDOM NUMBER GENERATOR VARIABLES.
  !
  ! This statement is simply to select the precision
  INTEGER, PARAMETER :: DP=SELECTED_REAL_KIND(14)
```

```
11
12  ! Scientific constants
13  REAL(DP), PARAMETER :: pi=3.14159265358979323846264338327950288419 7_DP
14  REAL(DP), PARAMETER :: N_A = 6.0221409E23
15
16  ! Number of dimensions of space (ie: x,y,z)
17  INTEGER, PARAMETER :: dim = 3
18
19  ! Equilibrium length/starting length of the springs
20  REAL(DP), PARAMETER :: l_sp = 1.0
21  REAL(DP), PARAMETER :: kbT = 1.0 ! thermal energy
22  REAL(DP), PARAMETER :: k_sp = 200.0 ! spring constant for potential
     force connecting polymer beads
23  REAL(DP), PARAMETER :: k_bind = 10.0 !k_sp ! spring constant of bond
     between polymer and toxin
24  REAL(DP), PARAMETER :: d_tox = l_sp
25  REAL(DP), PARAMETER :: l_bind = l_sp/2.0_DP + d_tox/2.0_DP! equilibrium
     length of spring after binding
26  REAL(DP), PARAMETER :: d_bead = l_sp ! diameter of a bead
27
28  ! How often should you check if a binding event happens
29  INTEGER, PARAMETER :: checkBindingInterval = 100 !1000
30  REAL(DP), PARAMETER :: D = 1.0 ! diffusion coefficient of Polymer
31  REAL(DP), PARAMETER :: DTox = D !10; % diffusion coefficient of Toxin
32  REAL(DP), PARAMETER :: Vf = 0.0 ! velocity of the fluid
33  REAL(DP), PARAMETER :: deltaT = .0001 ! time step
34  REAL(DP), PARAMETER :: zeta = 1.0 ! Not needed for now, so set to
     arbitrary number (for hydrodynamic forces/friction of fluid)
35  INTEGER :: maxTox
36  INTEGER, PARAMETER :: timeAvgInt = 100
37  INTEGER :: nTox
38
39  ! Schmolochowsky limit for how far the toxin can see the inhibitor beads
```

```
40  REAL(DP), PARAMETER :: reach = l_bind+SQRT(D*deltaT*100_DP)
41
42  ! Species features
43  INTEGER :: Nmer
44  INTEGER :: bindingSites
45  REAL(DP) :: tot_t
46
47  ! The number of beads
48  INTEGER :: beads
49
50  ! size of side of box for periodic boundary conditions (origin is at
     center of box)
51  REAL(DP) :: sizeBox
52
53  ! Energy barriers
54  REAL(DP) :: stdDevPolyAff
55  REAL(DP) :: delE_0_center
56  INTEGER :: randNumSeedMultiplier
57  INTEGER, PARAMETER :: seedSize = 33
58
59  ! The effective step size
60  REAL(DP) :: muo
61
62  ! The strength of the flow
63  REAL(DP) :: wi
64
65  ! stretching constant for wormlike chains
66  REAL(DP), PARAMETER :: gamma=100.0
67
68  ! bk is the kuhn length in units of a
69  ! dmax is the maximum bond length
70  REAL(DP), PARAMETER :: bk=2.0,dmax=2.2
71
```

```
72   ! sigma is the space constant in the lennard-jones force
73   REAL(DP), PARAMETER :: sigma=l_sp
74   REAL(DP), PARAMETER :: sigmaToxTox=d_tox
75   REAL(DP), PARAMETER :: sigmaToxPoly=l_bind
76   REAL(DP), PARAMETER :: LJcutoff=3.0_DP
77
78   ! epsilon is the strength of the lennard-jones potential
79   REAL(DP) :: epsilon
80   REAL(DP) :: epsToxTox
81   REAL(DP) :: epsToxPoly
82
83   ! constants to control the stiffness of the chain - wormlike chain model
84   REAL(DP) :: gammaw
85
86   ! Options
87   ! Decides whether the inhibitor is a polymer or just monomers floating
       around, true = is a polymer and experiences connectivity forces, 0 =
       free monomers in solution, don't experience connectivity forces
88   LOGICAL :: isPolymer
89
90   ! This controls whether when a toxin unbinds, if it is replaced with a
       new toxin starting in a random place or if it just continues on
91   LOGICAL :: randNewToxin
92
93   ! Cycles and writing blocks
94   INTEGER :: block,d_wrt= 10000
95
96 END MODULE parameters
```

This Page Intentionally Left Blank

References

[1] Mathai Mammen, Seok-Ki Choi, and George M Whitesides. Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors. *Angewandte Chemie, International Edition*, 37:2755–2794, 1998.

[2] Ang Gao, Krishna Shrinivas, Paul Lepeudry, Hiroshi I Suzuki, Phillip A Sharp, and Arup K Chakraborty. Evolution of weak cooperative interactions for biological specificity. *Proceedings of the National Academy of Sciences*, 115(47):E11053–E11060, nov 2018.

[3] Dimiter S. Dimitrov. Virus entry: molecular mechanisms and biomedical applications. *Nature Reviews Microbiology*, 2(2):109–122, feb 2004.

[4] Ángel M. Cuesta, Noelia Sainz-Pastor, Jaume Bonet, Baldomero Oliva, and Luis Álvarez-Vallina. Multivalent antibodies: when design surpasses evolution. *Trends in Biotechnology*, 28(7):355–362, jul 2010.

[5] D J Diestler and E W Knapp. Statistical Mechanics of the Stability of Multivalent Ligand-Receptor Complexes †. *The Journal of Physical Chemistry C*, 114(12):5287–5304, apr 2010.

[6] Marcus Weber, Alexander Bujotzek, and Rainer Haag. Quantifying the rebinding effect in multivalent chemical ligand-receptor systems. *Journal of Chemical Physics*, 137(5):1–11, 2012.

[7] Galina V Dubacheva, Tine Curk, Bortolo M Mognetti, Rachel Auzély-Velty, Daan Frenkel, and Ralf P Richter. Superselective Targeting Using Multivalent Polymers. *Journal of the American Chemical Society*, 136(5):1722–1725, feb 2014.

[8] Lorenzo Albertazzi, Francisco J Martinez-Veracoechea, Christianus M A Leenders, Ilja K Voets, Daan Frenkel, and E W Meijer. Spatiotemporal control and superselectivity in supramolecular polymers using multivalency. *Proceedings of the National Academy of Sciences*, 110(30):12203–12208, jul 2013.

[9] Bradley M. Lunde, Claire Moore, and Gabriele Varani. RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology*, 8(6):479–490, jun 2007.

[10] Shihu Wang and Elena E Dormidontova. Selectivity of Ligand-Receptor Interactions between Nanoparticle and Cell Surfaces. *Physical Review Letters*, 109(23):238102, dec 2012.

[11] Francisco J Martinez-Veracoechea and Daan Frenkel. Designing super selectivity in multivalent nano-particle binding. *Proceedings of the National Academy of Sciences*, 108(27):10963–10968, 2011.

[12] Tine Curk, Jure Dobnikar, and Daan Frenkel. Optimal multivalent targeting of membranes with many distinct receptors. *Proceedings of the National Academy of Sciences*, 114(28):7210–7215, jul 2017.

[13] Coby B. Carlson, Patricia Mowery, Robert M. Owen, Emily C. Dykhuizen, and Laura L. Kiessling. Selective Tumor Cell Targeting Using Low-Affinity, Multivalent Interactions. *ACS Chemical Biology*, 2(2):119–127, feb 2007.

[14] Sumati Bhatia, Luis Cuellar Camacho, and Rainer Haag. Pathogen Inhibition by Multivalent Ligand Architectures. *Journal of the American Chemical Society*, 138(28):8654–8666, jul 2016.

[15] S. R. Simon Ting, Gaojian Chen, and Martina H. Stenzel. Synthesis of glycopolymers and their multivalent recognitions with lectins. *Polymer Chemistry*, 1(9):1392, nov 2010.

[16] René Roy. Syntheses and some applications of chemically defined multivalent glycoconjugates. *Current Opinion in Structural Biology*, 6(5):692–702, oct 1996.

[17] George B Sigal, Mathai Mammen, Georg Dahmann, and George M Whitesides. Polyacrylamides Bearing Pendant α-Sialoside Groups Strongly Inhibit Agglutination of Erythrocytes by Influenza Virus: The Strong Inhibition Reflects Enhanced Binding through Cooperative Polyvalent Interactions. *Journal of the American Chemical Society*, 118(16):3789–3800, jan 1996.

[18] Brian D Polizzotti and Kristi L Kiick. Effects of Polymer Structure on the Inhibition of Cholera Toxin by Linear Polypeptide-Based Glycopolymers. *Biomacromolecules*, 7(2):483–490, feb 2006.

[19] Jason E Gestwicki, Christopher W Cairo, Laura E Strong, Karolyn A Oetjen, and Laura L Kiessling. Influencing Receptor-Ligand Binding Mechanisms with Multivalent Ligand Architecture. *Journal of the American Chemical Society*, 124(50):14922–14933, dec 2002.

[20] Ilona Papp, Christian Sieben, Adam L. Sisson, Johanna Kostka, Christoph Böttcher, Kai Ludwig, Andreas Herrmann, and Rainer Haag. Inhibition of Influenza Virus Activity by Multivalent Glycoarchitectures with Matched Sizes. *ChemBioChem*, 12(6):887–895, 2011.

[21] Shengchang Tang, Wendy B. Puryear, Brian M. Seifried, Xuehui Dong, Jonathan A. Runstadler, Katharina Ribbeck, and Bradley D. Olsen. Antiviral Agents from Multivalent Presentation of Sialyl Oligosaccharides on Brush Polymers. *ACS Macro Letters*, 5(3):413–418, mar 2016.

[22] Sumati Bhatia, Daniel Lauster, Markus Bardua, Kai Ludwig, Stefano Angioletti-Uberti, Nicole Popp, Ute Hoffmann, Florian Paulus, Matthias Budt, Marlena Stadtmüller, Thorsten Wolff, Alf Hamann, Christoph Böttcher, Andreas Herrmann, and Rainer Haag. Linear polysialoside outperforms dendritic analogs for inhibition of influenza virus infection in vitro and in vivo. *Biomaterials*, 138:22–34, sep 2017.

[23] A Varki. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, 3(2):97–130, apr 1993.

[24] Caroline Werlang, Gerardo Cárcarmo-Oyarce, and Katharina Ribbeck. Engineering mucus to study and influence the microbiome. *Nature Reviews Materials*, 4(2):134–145, feb 2019.

[25] D. R. Strombeck and D. Harrold. Binding of cholera toxin to mucins and inhibition by gastric mucin. *Infection and Immunity*, 10(6):1266–1272, 1974.

[26] David A. Rasko and Vanessa Sperandio. Anti-virulence strategies to combat bacteria-mediated disease. *Nature Reviews Drug Discovery*, 9(2):117–128, 2010.

[27] Sumati Bhatia, Mathias Dimde, and Rainer Haag. Multivalent glycoconjugates as vaccines and potential drug candidates. *Med. Chem. Commun.*, 5(7):862–878, may 2014.

[28] Maria Bergström, Shuang Liu, Kristi L. Kiick, and Sten Ohlson. Cholera Toxin Inhibitors Studied with High-Performance Liquid Affinity Chromatography: A Robust Method to Evaluate Receptor-Ligand Interactions. *Chemical Biology & Drug Design*, 73(1):132–141, jan 2009.

[29] Susanne Liese and Roland R Netz. Quantitative Prediction of Multivalent Ligand–Receptor Binding Affinities for Influenza, Cholera, and Anthrax Inhibition. *ACS Nano*, 12(5):4140–4147, may 2018.

[30] Jonathan Vonnemann, Susanne Liese, Christian Kuehne, Kai Ludwig, Jens Dernedde, Christoph Böttcher, Roland R. Netz, and Rainer Haag. Size Dependence of Steric Shielding and Multivalency Effects for Globular Binding Inhibitors. *Journal of the American Chemical Society*, 137(7):2572–2579, feb 2015.

[31] Motomu Kanai, Kathleen H Mortell, and Laura L Kiessling. Varying the Size of Multivalent Ligands: The Dependence of Concanavalin A Binding on Neoglycopolymer Length. *Journal of the American Chemical Society*, 119(41):9931–9932, oct 1997.

[32] Miho Watanabe, Koji Matsuoka, Eiji Kita, Katsura Igai, Nobutaka Higashi, Atsushi Miyagawa, Toshiyuki Watanabe, Ryohei Yanoshita, Yuji Samejima, Daiyo Terunuma, Yasuhiro Natori, and Kiyotaka Nishikawa. Oral Therapeutic Agents with Highly Clustered Globotriose for Treatment of Shiga Toxigenic Escherichia coli Infections. *The Journal of Infectious Diseases*, 189(3):360–368, 2004.

[33] Susanne Liese and Roland R. Netz. Influence of length and flexibility of spacers on the binding affinity of divalent ligands. *Beilstein Journal of Organic Chemistry*, 11:804–816, 2015.

[34] Sara Fortuna, Federico Fogolari, and Giacinto Scoles. Chelating effect in short polymers for the design of bidentate binders of increased affinity and selectivity. *Scientific Reports*, 5(1):15633, dec 2015.

[35] Miguel A. Soler and Sara Fortuna. Influence of Linker Flexibility on the Binding Affinity of Bidentate Binders. *The Journal of Physical Chemistry B*, 121(16):3918–3924, apr 2017.

[36] Galina V Dubacheva, Tine Curk, Rachel Auzély-Velty, Daan Frenkel, and Ralf P Richter. Designing multivalent probes for tunable superselective targeting. *Proceedings of the National Academy of Sciences*, 112(18):5579–5584, may 2015.

[37] Waldemar Nowicki. Structure and Entropy of a Long Polymer Chain in the Presence of Nanoparticles. *Macromolecules*, 35(4):1424–1436, feb 2002.

[38] Rama Bansil and Bradley S Turner. Mucin structure, aggregation, physiological functions and biomedical applications. *Current Opinion in Colloid & Interface Science*, 11(2-3):164–170, jun 2006.

[39] Tine Curk, Jure Dobnikar, and Daan Frenkel. Rational design of molecularly imprinted polymers. *Soft Matter*, 12(1):35–44, 2016.

[40] Oliver Lieleg, Corinna Lieleg, Jesse Bloom, Christopher B. Buck, and Katharina Ribbeck. Mucin Biopolymers As Broad-Spectrum Antiviral Agents. *Biomacromolecules*, 13(6):1724–1732, jun 2012.

[41] Inka Brockhausen, Harry Schachter, and Pamela Stanley. *O-GalNAc Glycans*. Cold Spring Harbor Laboratory Press, 2009.

[42] Mallory A. van Dongen, Casey A Dougherty, and Mark M. Banaszak Holl. Multivalent Polymers for Drug Delivery and Imaging: The Challenges of Conjugation. *Biomacromolecules*, 15(9):3215–3234, sep 2014.

[43] Clifford P. Brangwynne, Christian R. Eckmann, David S. Courson, Agata Rybarska, Carsten Hoege, Jöbin Gharakhani, F. Julicher, and Anthony A. Hyman. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science*, 324(5935):1729–1732, jun 2009.

[44] Clifford P. Brangwynne, Timothy J. Mitchison, and Anthony A. Hyman. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences*, 108(11):4334–4339, mar 2011.

[45] Salman F. Banani, Allyson M. Rice, William B. Peeples, Yuan Lin, Saumya Jain, Roy Parker, and Michael K. Rosen. Compositional Control of Phase-Separated Cellular Bodies. *Cell*, 166(3):651–663, jul 2016.

[46] William M. Jacobs and Daan Frenkel. Phase Transitions in Biological Systems with Many Components. *Biophysical Journal*, 112(4):683–691, feb 2017.

[47] Jonathon A. Ditlev, Lindsay B. Case, and Michael K. Rosen. Who's In and Who's Out—Compositional Control of Biomolecular Condensates. *Journal of Molecular Biology*, 430(23):4666–4684, nov 2018.

[48] Vladimir N. Uversky. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current Opinion in Structural Biology*, 44:18–30, jun 2017.

[49] Yongdae Shin and Clifford P. Brangwynne. Liquid phase condensation in cell physiology and disease. *Science*, 357(6357):eaaf4382, sep 2017.

[50] Salman F. Banani, Hyun O. Lee, Anthony A. Hyman, and Michael K. Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, 18(5):285–298, may 2017.

[51] Edward Gomes and James Shorter. The molecular language of membraneless organelles. *Journal of Biological Chemistry*, 294(18):7115–7127, may 2019.

[52] Clifford P. Brangwynne, Peter Tompa, and Rohit V. Pappu. Polymer physics of intracellular phase transitions. *Nature Physics*, 11(11):899–904, nov 2015.

[53] Paul J. Flory. Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics*, 10(1):51–61, jan 1942.

[54] Maurice L. Huggins. Theory of Solutions of High Polymers. *Journal of the American Chemical Society*, 64(7):1712–1719, jul 1942.

[55] Paul Hiemenz and Timothy Lodge. *Polymer chemistry*. CRC Press, 2 edition, 2007.

[56] Walter H Stockmayer. Analysis of Linear Viscoelasticity of a Crosslinking Polymer at the Gel Point Journal of Rheology. *Thermodynamics of High Polymer Solutions The Journal of Chemical Physics*, 12(4):51, 1944.

[57] Paul J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, Ithaca, NY, 1953.

[58] Tyler S. Harmon, Alex S. Holehouse, Michael K. Rosen, and Rohit V. Pappu. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife*, 6, nov 2017.

[59] Marina Feric, Nilesh Vaidya, Tyler S. Harmon, Diana M. Mitrea, Lian Zhu, Tiffany M. Richardson, Richard W. Kriwacki, Rohit V. Pappu, and Clifford P. Brangwynne. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, 165(7):1686–1697, jun 2016.

[60] David W. Sanders, Nancy Kedersha, Daniel S.W. Lee, Amy R. Strom, Victoria Drake, Joshua A. Riback, Dan Bracha, Jorine M. Eeftens, Allana Iwanicki, Alicia Wang, Ming Tzo Wei, Gena Whitney, Shawn M. Lyons, Paul Anderson, William M. Jacobs, Pavel Ivanov, and Clifford P. Brangwynne. Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell*, 181(2):306–324.e28, apr 2020.

[61] Frank S. Bates and Glenn H. Fredrickson. Block Copolymers—Designer Soft Materials. *Physics Today*, 52(2):32–38, feb 1999.

[62] Jacob Israelachvili. *Intermolecular and Surface Forces*. Elsevier, 2011.

[63] Steven Boeynaems, Simon Alberti, Nicolas L Fawzi, Tanja Mittag, Magdalini Polymenidou, Frederic Rousseau, Joost Schymkowitz, James Shorter, Benjamin Wolozin, Ludo Van Den Bosch, Peter Tompa, and Monika Fuxreiter. Protein Phase Separation: A New Phase in Cell Biology. *Trends in Cell Biology*, 28(6):420–435, jun 2018.

[64] Ankur Jain and Ronald D. Vale. RNA phase transitions in repeat expansion disorders. *Nature*, 546(7657):243–247, jun 2017.

[65] Adriano Aguzzi and Matthias Altmeyer. Phase Separation: Linking Cellular Compartmentalization to Disease. *Trends in Cell Biology*, 26(7):547–558, jul 2016.

[66] Avinash Patel, Hyun O. Lee, Louise Jawerth, Shovamayee Maharana, Marcus Jahnel, Marco Y. Hein, Stoyno Stoynov, Julia Mahamid, Shambaditya Saha, Titus M. Franzmann, Andrej Pozniakovski, Ina Poser, Nicola Maghelli, Loic A. Royer, Martin Weigert, Eugene W. Myers, Stephan Grill, David Drechsel, Anthony A. Hyman, and Simon Alberti. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*, 162(5):1066–1077, aug 2015.

[67] Scott C Grindy, Martin Lenz, and Niels Holten-Andersen. Engineering Elasticity and Relaxation Time in Metal-Coordinate Cross-Linked Hydrogels. *Macromolecules*, 49(21):8306–8312, nov 2016.

[68] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, T N Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, jan 2000.

[69] R. Ravishankar, C.J. Thomas, K. Suguna, A. Surolia, and M. Vijayan. Crystal structures of the peanut lectin-lactose complex at acidic pH: retention of unusual quaternary structure, empty and carbohydrate bound combining sites, molecular mimicry and crystal packing directed by interactions at the combining site. *Proteins*, 43:260–270, 2001.

[70] Alfredo Alexander-Katz and Roland R Netz. Dynamics and Instabilities of Collapsed Polymers in Shear Flow. *Macromolecules*, 41(9):3363–3374, may 2008.

[71] Charles E Sing and Alfredo Alexander-Katz. Equilibrium Structure and Dynamics of Self-Associating Single Polymers. *Macromolecules*, 44(17):6962–6971, sep 2011.

[72] G. Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, may 1978.

[73] M. R. King and D. A. Hammer. Multiparticle adhesive dynamics: Hydrodynamic recruitment of rolling leukocytes. *Proceedings of the National Academy of Sciences*, 98(26):14919–14924, dec 2001.

[74] German V. Kolmakov, Krzysztof Matyjaszewski, and Anna C. Balazs. Harnessing Labile Bonds between Nanogel Particles to Create Self-Healing Materials. *ACS Nano*, 3(4):885–892, apr 2009.

[75] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics. *PLOS Computational Biology*, 15(2):e1006830, feb 2019.

[76] Simone Ciarella, Francesco Sciortino, and Wouter G Ellenbroek. Dynamics of Vitrimers: Defects as a Highway to Stress Relaxation. *Physical Review Letters*, 121(5):058003, jul 2018.

[77] Bernardo Oyarzún and Bortolo Matteo Mognetti. Programming configurational changes in systems of functionalised polymers using reversible intramolecular linkages. *Molecular Physics*, 116(21-22):2927–2941, nov 2018.

[78] Francesco Sciortino. Three-body potential for simulating bond swaps in molecular dynamics. *The European Physical Journal E*, 40(1):3, jan 2017.

[79] Charles E Sing and Alfredo Alexander-Katz. Giant Nonmonotonic Stretching Response of a Self-Associating Polymer in Shear Flow. *Physical Review Letters*, 107(19):198302, nov 2011.

[80] Patricia Targon Campana, Leandro Ramos, Souza Barbosa, and Rosangela Itri. Conformational stability of peanut agglutinin using small angle X-ray scattering. *International Journal of Biological Macromolecules*, 48:398–402, 2010.

[81] Yoann M. Chabre, Denis Giguère, Bertrand Blanchard, Jacques Rodrigue, Sylvain Rocheleau, Mathieu Neault, Subhash Rauthu, Alex Papadopoulos, Alexandre A. Arnold, Anne Imberty, and René Roy. Combining glycomimetic and multivalent strategies toward designing potent bacterial lectin inhibitors. *Chemistry - A European Journal*, 17(23):6545–6562, 2011.

[82] Halina Lis and Nathan Sharon. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chemical Reviews*, 98(2):637–674, apr 1998.

[83] Pilong Li, Sudeep Banjade, Hui-Chun Cheng, Soyeon Kim, Baoyu Chen, Liang Guo, Marc Llaguno, Javoris V. Hollingsworth, David S. King, Salman F. Banani, Paul S. Russo, Qiu-Xing Jiang, B. Tracy Nixon, and Michael K. Rosen. Phase transitions in the assembly of multivalent signalling proteins. *Nature*, 483(7389):336–340, mar 2012.

[84] Yousif Shamoo, Norzehan Abdul-Manan, Ann M Patten, Janet K Crawford, Matthew C Pellegrini, and Kenneth R Williams. Both RNA-Binding Domains in Heterogeneous Nuclear Ribonucleoprotein A1 Contribute Toward Single-Stranded-RNA Binding. *Biochemistry*, 33(27):8272–8281, jul 1994.

[85] Yousif Shamoo, Norzehan Abdul-Manan, and Kenneth R Williams. Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Research*, 23(5):725–728, 1995.

[86] Emiko Zumbro, Jacob Witten, and Alfredo Alexander-Katz. Computational Insights into Avidity of Polymeric Multivalent Binders. *Biophysical Journal*, 117(5):892–902, sep 2019.

[87] Vijay M Krishnamurthy, Lara A. Estroff, and George M Whitesides. Multivalency in Ligand Design. In Raimund Mannhold, Hugo Kubinyi, Gerd Folkers, Wolfgang Jahnke, and Daniel A. Erlanson, editors, *Fragment-based Approaches in Drug Discovery*, chapter 2, pages 11–53. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, jul 2006.

[88] Laura L. Kiessling, Travis Young, and Kathleen H. Mortell. Multivalency in Protein-Carbohydrate Recognition. In Bertram O. Fraser-Reid, Kuniaki Tatsuta, and Joachim Thiem, editors, *Glycoscience: Chemistry and Chemical Biology I-III*, pages 1817–1861. Springer, Berlin, Heidelberg, 2001.

[89] Jeffrey D Esko and Nathan Sharon. Microbial Lectins: Hemagglutinins, Adhesins, and Toxins. In A Varki, R D Cummings, J D Esko, H H Freeze, P Stanley, C R Bertozzi, G W Hart, and M E Etzler, editors, *Essentials of Glycobiology*, chapter 34, pages 489–500. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2 edition, 2009.

[90] Thomas R Branson and W Bruce Turnbull. Bacterial toxin inhibitors based on multivalent scaffolds. *Chem. Soc. Rev.*, 42(11):4613–4622, 2013.

[91] Shuang Liu and Kristi L Kiick. Architecture Effects on the Binding of Cholera Toxin by Helical Glycopolypeptides. *Macromolecules*, 41(3):764–772, feb 2008.

[92] David Deniaud, Karine Julienne, and Sébastien G Gouin. Insights in the rational design of synthetic multivalent glycoconjugates as lectin ligands. *Org. Biomol. Chem.*, 9(4):966–979, 2011.

[93] Joseph J Lundquist, Sheryl D Debenham, and Eric J Toone. Multivalency Effects in Protein-Carbohydrate Interaction: The Binding of the Shiga-like Toxin 1 Binding Subunit to Multivalent C- Linked Glycopeptides. *The Journal of Organic Chemistry*, 65(24):8245–8250, dec 2000.

[94] Xiaoxiong Zeng, Takeomi Murata, Hirokazu Kawagishi, Taichi Usui, and Kazukiyo Kobayashi. Synthesis of Artificial N -Glycopolypeptides Carrying N -Acetyllactosamine and Related Compounds and Their Specific Interactions with Lectins. *Bioscience, Biotechnology, and Biochemistry*, 62(6):1171–1178, jan 1998.

[95] Brian D Polizzotti, Ronak Maheshwari, Jan Vinkenborg, and Kristi L Kiick. Effects of Saccharide Spacing and Chain Extension on Toxin Inhibition by Glycopolypeptides of Well-Defined Architecture. *Macromolecules*, 40(20):7103–7110, oct 2007.

[96] Masanori Nagao, Yurina Fujiwara, Teruhiko Matsubara, Yu Hoshino, Toshinori Sato, and Yoshiko Miura. Design of Glycopolymers Carrying Sialyl Oligosaccharides for Controlling the Interaction with the Influenza Virus. *Biomacromolecules*, 18(12):4385–4392, dec 2017.

[97] Sarah-Jane Richards, Mathew W. Jones, Mark Hunaban, David M. Haddleton, and Matthew I. Gibson. Probing Bacterial-Toxin Inhibition with Synthetic Glycopolymers Prepared by Tandem Post-Polymerization Modification: Role of Linker Length and Carbohydrate Density. *Angewandte Chemie International Edition*, 51(31):7812–7816, jul 2012.

[98] Moira Ambrosi, Neil R Cameron, Benjamin G Davis, and Snjezana Stolnik. Investigation of the interaction between peanut agglutinin and synthetic glycopolymeric multivalent ligands. *Organic & Biomolecular Chemistry*, 3(8):1476, 2005.

[99] William P Jencks. On the attribution and additivity of binding energies (proteins/ligands/entropy/enzymes). *Biochemistry*, 78(7):4046–4050, 1981.

[100] Thomas E Ouldridge, Ard A Louis, and Jonathan P K Doye. Extracting bulk properties of self-assembling systems from small simulations. *Journal of Physics: Condensed Matter*, 22(10):104102, mar 2010.

[101] Douglas Poland and Harold A Scheraga. Phase Transitions in One Dimension and the Helix—Coil Transition in Polyamino Acids. *The Journal of Chemical Physics*, 45(5):1456–1463, sep 1966.

[102] Pavel I Kitov and David R Bundle. On the Nature of the Multivalency Effect: A Thermodynamic Model. *Journal of the American Chemical Society*, 125(52):16271–16284, dec 2003.

[103] Dreania Levine, Michael J Kaplan, and Peter J Greenaway. The Purification and Characterization of Wheat-Germ Agglutinin. *Biochem. J*, 129:847–856, 1972.

[104] 5 Human Albumin. *Transfusion Medicine and Hemotherapy*, 36(6):399–407, 2009.

[105] Christopher W Cairo, Jason E Gestwicki, Motomu Kanai, and Laura L Kiessling. Control of Multivalent Interactions by Binding Epitope Density. *Journal of the American Chemical Society*, 124(8):1615–1619, feb 2002.

[106] Emiko Zumbro and Alfredo Alexander-Katz. Influence of Binding Site Affinity Patterns on Binding of Multivalent Polymers. *ACS Omega*, 5(19):10774–10781, may 2020.

[107] Veronika Csizmok, Ariele Viacava Follis, Richard W. Kriwacki, and Julie D. Forman-Kay. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. *Chemical Reviews*, 116(11):6424–6462, jun 2016.

[108] Nicholas B. Tito and Daan Frenkel. Optimizing the Selectivity of Surface-Adsorbing Multivalent Polymers. *Macromolecules*, 47(21):7496–7509, nov 2014.

[109] Doros N Theodorou. Microscopic structure and thermodynamic properties of bulk copolymers and surface-active polymers at interfaces. 2. Results for some representative chain architectures. *Macromolecules*, 21(5):1422–1436, sep 1988.

[110] C. M. Marques and J. F. Joanny. Block copolymer adsorption in a nonselective solvent. *Macromolecules*, 22(3):1454–1458, may 1989.

[111] Jason E. Hudak and Carolyn R. Bertozzi. Glycotherapy: New Advances Inspire a Reemergence of Glycans in Medicine. *Chemistry & Biology*, 21(1):16–37, jan 2014.

[112] Qiang Zhang, Jennifer Collins, Athina Anastasaki, Russell Wallis, Daniel A. Mitchell, C. Remzi Becer, and David M. Haddleton. Sequence-Controlled Multi-Block Glycopolymers to Inhibit DC-SIGN-gp120 Binding. *Angewandte Chemie International Edition*, 52(16):4435–4439, apr 2013.

[113] Dipti Gupta, Tarun K. Dam, Stefan Oscarson, and C. Fred Brewer. Thermodynamics of Lectin-Carbohydrate Interactions. *Journal of Biological Chemistry*, 272(10):6388–6392, mar 1997.

[114] Emiko Zumbro and Alfredo Alexander-Katz. Polymer Stiffness Regulates Multivalent Binding and Liquid-Liquid Phase Separation. *In Review*, 2020.

[115] Vijay M. Krishnamurthy, Vincent Semetey, Paul J. Bracher, Nan Shen, and George M. Whitesides. Dependence of Effective Molarity on Linker Length for an Intramolecular Protein-Ligand System. *Journal of the American Chemical Society*, 129(5):1312–1320, feb 2007.

[116] Maximilian Klement, Chengcheng Liu, Bernard Liat Wen Loo, Andre Boon-Hwa Choo, Dave Siak-Wei Ow, and Dong-Yup Lee. Effect of linker flexibility and length on the functionality of a cytotoxic engineered antibody fragment. *Journal of Biotechnology*, 199:90–97, apr 2015.

[117] Yuan Lin, David S.W. Protter, Michael K. Rosen, and Roy Parker. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular Cell*, 60(2):208–219, oct 2015.

[118] Amandine Molliex, Jamshid Temirov, Jihun Lee, Maura Coughlin, Anderson P. Kanagaraj, Hong Joo Kim, Tanja Mittag, and J. Paul Taylor. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell*, 163(1):123–133, sep 2015.

[119] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, jul 2007.

[120] Emeritus Hiromi Yamakawa. *Helical Wormlike Chains in Polymer Solutions*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.

[121] O. Kratky and G. Porod. Röntgenuntersuchung gelöster Fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas*, 68(12):1106–1122, sep 1949.

[122] Patrick S. Doyle and Patrick T. Underhill. Brownian Dynamics Simulations of Polymers and Soft Matter. In *Handbook of Materials Modeling*, pages 2619–2630. Springer Netherlands, Dordrecht, 2005.

[123] Alberto Montesi, Matteo Pasquali, and F. C. MacKintosh. Collapse of a semiflexible polymer in poor solvent. *Physical Review E*, 69(2):021916, feb 2004.

[124] Paul. J. Flory. *Statistical mechanics of chain molecules*. Wiley-Interscience, New York, 1969.

[125] Jeong-Mo Choi, Furqan Dar, and Rohit V. Pappu. LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLOS Computational Biology*, 15(10):e1007028, oct 2019.

[126] Kurt Binder and Dieter W. Heermann. *Monte Carlo Simulation in Statistical Physics*. Graduate Texts in Physics. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[127] Tyler K. Lytle and Charles E. Sing. Tuning chain interaction entropy in complex coacervation using polymer stiffness, architecture, and salt valency. *Molecular Systems Design & Engineering*, 3(1):183–196, feb 2018.

[128] Fabian Erdel and Karsten Rippe. Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal*, 114(10):2262–2270, may 2018.

[129] Seth Allen Cazzell and Niels Holten-Andersen. Expanding the stoichiometric window for metal cross-linked gel assembly using competition. *Proceedings of the National Academy of Sciences*, 116(43):21369–21374, oct 2019.

[130] Jie Wang, Jeong-Mo Choi, Alex S. Holehouse, Hyun O. Lee, Xiaojie Zhang, Marcus Jahnel, Shovamayee Maharana, Régis Lemaitre, Andrei Pozniakovsky, David Drechsel, Ina Poser, Rohit V. Pappu, Simon Alberti, and Anthony A. Hyman. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*, 174(3):688–699.e16, jul 2018.

[131] Bin Xu, Guanhua He, Benjamin G. Weiner, Pierre Ronceray, Yigal Meir, Martin C. Jonikas, and Ned S. Wingreen. Rigidity enhances a magic-number effect in polymer phase separation. *Nature Communications*, 11(1):1561, dec 2020.

[132] Elizabeth S. Freeman Rosenzweig, Bin Xu, Luis Kuhn Cuellar, Antonio Martinez-Sanchez, Miroslava Schaffer, Mike Strauss, Heather N. Cartwright, Pierre Ronceray, Jürgen M. Plitzko, Friedrich Förster, Ned S. Wingreen, Benjamin D. Engel, Luke C.M. Mackinder, and Martin C. Jonikas. The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell*, 171(1):148–162.e19, sep 2017.

[133] Shovamayee Maharana, Jie Wang, Dimitrios K. Papadopoulos, Doris Richter, Andrey Pozniakovsky, Ina Poser, Marc Bickle, Sandra Rizk, Jordina Guillén-Boixet, Titus M. Franzmann, Marcus Jahnel, Lara Marrone, Young-Tae Chang, Jared Sterneckert, Pavel Tomancak, Anthony A. Hyman, and Simon Alberti. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science*, 360(6391):918–921, may 2018.

[134] Timothy J. Nott, Evangelia Petsalaki, Patrick Farber, Dylan Jervis, Eden Fussner, Anne Plochowietz, Timothy D. Craggs, David P. Bazett-Jones, Tony Pawson, Julie D. Forman-Kay, and Andrew J. Baldwin. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell*, 57(5):936–947, mar 2015.

[135] Anisha Shakya and John T. King. DNA Local-Flexibility-Dependent Assembly of Phase-Separated Liquid Droplets. *Biophysical Journal*, 115(10):1840–1847, nov 2018.

[136] John Crank. *The Mathematics of Diffusion*. Clarendon Press, Oxford, second edi edition, 1975.

[137] Emiko Zumbro and Alfredo Alexander-Katz. Multivalent Polymers Can Control Phase Boundary, Dynamics, and Organization of Liquid-Liquid Phase Separation. *In Preparation*, 2020.

[138] Saroj Kumar Nandi, Meta Heidenreich, Emmanuel D. Levy, and Samuel A. Safran. Interacting multivalent molecules: affinity and valence impact the extent and symmetry of phase separation. *arXiv*, oct 2019.

[139] Hong Joo Kim, Nam Chul Kim, Yong-Dong Wang, Emily A. Scarborough, Jennifer Moore, Zamia Diaz, Kyle S. MacLea, Brian Freibaum, Songqing Li, Amandine Molliex, Anderson P. Kanagaraj, Robert Carter, Kevin B. Boylan, Aleksandra M. Wojtas, Rosa Rademakers, Jack L. Pinkus, Steven A. Greenberg, John Q. Trojanowski, Bryan J. Traynor, Bradley N. Smith, Simon Topp, Athina-Soragia Gkazi, Jack Miller, Christopher E. Shaw, Michael Kottlors, Janbernd Kirschner, Alan Pestronk, Yun R. Li, Alice Flynn Ford, Aaron D. Gitler, Michael Benatar, Oliver D. King, Virginia E. Kimonis, Eric D. Ross, Conrad C. Weihl, James Shorter, and J. Paul Taylor. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature*, 495(7442):467–473, mar 2013.

[140] Ming-Tzo Wei, Shana Elbaum-Garfinkle, Alex S. Holehouse, Carlos Chih-Hsiung Chen, Marina Feric, Craig B. Arnold, Rodney D. Priestley, Rohit V. Pappu, and Clifford P. Brangwynne. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nature Chemistry*, 9(11):1118–1125, nov 2017.

[141] Erik W. Martin, Alex S. Holehouse, Ivan Peran, Mina Farag, J. Jeremias Incicco, Anne Bremer, Christy R. Grace, Andrea Soranno, Rohit V. Pappu, and Tanja Mittag. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science*, 367(6478):694–699, feb 2020.

[142] Tyler S. Harmon, Alex S. Holehouse, and Rohit V. Pappu. Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins. *New Journal of Physics*, 20(4):045002, apr 2018.

[143] David W. Sanders, Nancy Kedersha, Daniel S.W. Lee, Amy R. Strom, Victoria Drake, Joshua A. Riback, Dan Bracha, Jorine M. Eeftens, Allana Iwanicki, Alicia Wang, Ming-Tzo Wei, Gena Whitney, Shawn M. Lyons, Paul Anderson, William M. Jacobs, Pavel Ivanov, and Clifford P. Brangwynne. Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell*, 181(2):306–324.e28, apr 2020.

[144] B. Tüű-Szabó, G. Hoffka, N. Duro, and M. Fuxreiter. Altered dynamics may drift pathological fibrillization in membraneless organelles. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1867(10):988–998, oct 2019.

[145] Rama Bansil and Bradley S. Turner. The biology of mucus: Composition, synthesis and organization. *Advanced Drug Delivery Reviews*, 124:3–15, jan 2018.

[146] Felipe García Quiroz and Ashutosh Chilkoti. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nature Materials*, 14(11):1164–1171, nov 2015.

[147] Yi-Hsuan Lin, Julie D Forman-Kay, and Hue Sun Chan. Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. *Biochemistry*, 57(17):2499–2508, may 2018.

[148] Sudeep Banjade and Michael K. Rosen. Phase transitions of multivalent proteins can promote clustering of membrane receptors. *eLife*, 3, oct 2014.

[149] William M. Jacobs and Daan Frenkel. Predicting phase behavior in multicomponent mixtures. *The Journal of Chemical Physics*, 139(2):024108, jul 2013.

[150] Albert Reuther, Jeremy Kepner, William Arcand, David Bestor, Bill Bergeron, Chansup Byun, Matthew Hubbell, Peter Michaleas, Julie Mullen, Andrew Prout, and Antonio Rosa. LLSuperCloud: Sharing HPC systems for diverse rapid prototyping. In *2013 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–6. IEEE, sep 2013.

[151] Wolfram|Alpha. Wolfram Alpha LLC. https://www.wolframalpha.com/input/?i=sum+M!%2F(M-k)!+x%5Ek+from+k%3D1+to+M, Accessed: January 9, 2019.

[152] Wolfram Research, Inc. Mathematica, Version 11.3. Champaign, IL, 2018.

[153] Wolfram|Alpha. Wolfram Alpha LLC. https://www.wolframalpha.com/input/?i=expand+polylog(0.8,1-x)+about+0, Accessed: January 9, 2019.