UNIVERSITY OF CALIFORNIA
Los Angeles

The Physics of RNA-Protein Binding

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Chemistry

by

Zachary Gvildys

2023

ABSTRACT OF THE DISSERTATION

The Physics of RNA-Protein Binding

by

Zachary Gvildys
Doctor of Philosophy in Chemistry
University of California, Los Angeles, 2023
Professor Robijn F. Bruinsma, Chair

RNA-protein binding is important in many biological processes. These processes depend on the structural properties of the RNA, DNA and proteins involved. In this thesis we have investigated the nature of these interactions at the scale of coarse grained interactions, in atomistic detail and with analytical continuum theories. By exploring the dynamics of these biopolymers at these different scales, we hope to shed light on the important phenomena and characteristics that govern the biological processes involved.

First we show that the structural flexibility of these polymers affects he binding energy measured in MD simulations and can help understand the different packaging scenarios in in-vitro viral reconstitution. Then using analytical models we show how flexibility can aid it non-specific binding given the strength of interaction compared to thermal energy between the binding of polymers.

We finish this thesis by looking at coarse grained brownian dynamics simulations for the bonds in a chain to represent the non-linear viscoelastic behavior of proteins deforming under force. We see both equilibrium mechanical induced melting of the chain and a dynamical phase transition in the response of the chain under a varying drive amplitude and frequency of and AC applied force out of equilibrium. Using the model to understand the dynamics of a deformable chain under force, we explore how this model can provide insight into the

binding of deformable structure forming chains that mimic biopolymers like RNA, DNA and proteins.

The dissertation of Zachary Gvildys is approved.

Anastassia Alexandrova

William M. Gelbart

Justin R. Caram

Robijn F. Bruinsma, Committee Chair

University of California, Los Angeles

2023

To always remaining a student

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, professor Robijn Bruinsma, for being understanding and kind throughout my PhD journey. Our collaboration has been an entertaining one, him being from a physics background, and myself from chemistry. I have enjoyed the moments over the years where we have debated topics only for me to realize we are agreeing from different perspectives. His advisory has broadened my ability to approach problems from different perspectives and my understanding of science in general. Robijn has provided many opportunities and supported me advancing my education and career in my own way that I feel is unique in a student advisor relationship, for which I am grateful.

I am also very grateful to professor William Gelbart who has advised me on more than just science over the years. I see my relationship with Bill as instrumental in my development as a scientist, an educator and as a compassionate human being. I have had the pleasure of teaching with Bill many times over the years for which I was awarded a Distinguished teaching award, which I largely contribute to his discussion on the philosophy of teaching and pedagogy. I would also like to thank the other members of my committee professor Anastassia Alexandrova and Justin Caram for their guidance. Also, professor Alexander Levine, who is no longer with us, but whom I enjoyed teaching with very much and would bring humor to any scientific discussion, or any discussion for that matter.

I would also like to thank a lot of my colleagues over the years who have helped in many fruitful discussions. Dr. Chen Lin and I were both chemistry students in a physics group and we helped each other a lot in developing ideas and understanding topics we both were not familiar with. Dr. Jesus Iniguez for being a brother to me from undergraduates in quantum mechanics to PhD students. Both Chen and Jesus helped in taking breaks from our PhD schedules to play intramural sports to make sure our bodies struggled as much as our minds. I would also like to thank the members of the Gelbart group, Cheylene Tanimoto, Rich Sportsman, Abby Thurm, Christain Beren for being my family on campus. I enjoyed both the scientific and less than scientific arguments.

I would like to thank all of the people involved at the California Nanosystems Institute, where I had unique opportunities of scientific outreach and communication. Through them I was able to develop a computational workshop for understanding molecular self-assembly with the help of Andy Vong, David Reilley and Patricia Poths, who all did tremendous work to develop the workshop during the pandemic when in-person outreach was not possible.

Lastly, I would like to thank my parents and family for unconditional support in my unorthodox path to arrive where I am today. The immense patience they had to displace was invaluable.

VITA

2015–2017	B.S. in Physical Chemistry, UCLA, Los Angeles, CA
2017–2023	Ph.D. in Chemistry, UCLA, Los Angeles, CA

PUBLICATIONS

Theory of Diffusion-Limited Reactions between Proteins in a Confined Environment. ***in preparation.***

Chen Lin, **Zach Gvildys**, Joseph Rudnick, Robijn Bruinsma.

CHAPTER 1

Introduction

1.1 Protein-RNA Interactions

RNA-protein interactions are involved in a wide variety of cell-regulatory processes[3, 4, 5, 6]. Following synthesis, RNA transcripts are covered by RNA-binding proteins (RBPs) that regulate subsequent RNA splicing, editing, transport and localization. RBPs have different amounts of RNA sequence specificity. The zinc-finger and RNA-recognition motifs are important examples of specific RNA-protein interactions.

There is an interesting distinction between the interactions of proteins with single-stranded (ss) RNA molecules and with double-stranded (ds)DNA or ssDNA molecules[7]. For protein/ds DNA interaction there is a clear distinction between specific, sequence-sensitive interactions, involving the matching of particular protein residues with particular nucleic acid bases, and non-specific protein-dsDNA interactions between the negatively charged phosphate groups of the DNA sugar-phosphate backbone with positively charged protein residues. The non-specific protein-DNA affinity plays a distinct role in keeping DNA-associating proteins in proximity of the DNA and in nucleosome-DNA interaction. In actuality, the distinction between specific and non-specific interaction is not quite sharp. The "induced-fit mechanism" refers to the fact that the DNA sequence in the proximity of a protein binding site may code for DNA deformation in such a way that the protein can fit more easily to the DNA near a binding site through its non-specific interactions.

1.2 Protein-RNA interactions in viral self assembly

The role of non-specific interactions between ssRNA molecules and proteins is not so well understood. The complicating factor is the much greater structural flexibility of single-stranded nucleotide sequences as compared to double-stranded sequences. However, the interplay between non-specific and specific RNA/protein interactions have been studied experimentally for the case of the packaging of ssRNA viral genome molecules by capsid proteins during viral assembly. The electrostatic affinity of negatively charged ssRNA genome molecule for positively charged groups of the capsid proteins is a key thermodynamic force driving viral assembly. On the other hand, specific RNA-protein interactions between RNA hairpins ("packaging signals") and capsid protein play a key role in terms of distinguishing viral RNA molecules from cellular mRNA molecules during packaging. In a systematic series of experiments by the group of Gelbart and Knobler[8, 9, 10] on the bromoviruses it was demonstrated that purified viral capsid pro-teins of Cowpea Chlorotic Mottle Virus (CCMV) and Brome Mosaic Virus (BMV) are capable of completely packaging non-viral and non-native RNAs, sometimes even more efficiently than the viral genome! They also carried out packaging competition experiments between single-stranded polyU and polyA sequences as well as between single-stranded and double-stranded hybridized polyU and polyA sequences. The total number of nucleotides was the same while the RNA structural flexibility systematically decreased from PolyU to PolyA to dsPolyU-PolyA. Based on simple electrostatic arguments, one would expect that the binding affinity should be comparable for these three cases but in actuality single-stranded sequences out-compete double-stranded sequences while polyU outcompetes polyA. This indicates that structural flexibility of RNA nucleotide sequences indeed plays an important role during protein-RNA interactions.

With the recent SARS-CoV-2 pandemic, the need for a greater understanding of the life cycle and assembly of viruses has become even more relevant. CoVs have four structural proteins; the envelope, spike, nucleocapsid and membrane protein. We are interested in understanding the nucleocapsids protein's interactions with the viral genome because it plays an important role in the assembly and disassembly of the virus particle and potentially every

step in the viral life cycle. There are three highly conserved regions of the nucleocapsid (N) protein across different CoVs; the N-terminal domain (NTD), the C-terminal Domain (CTD) and an intrinsically disordered linker region comprised of many positive charges (LKR) connecting the other two domains. Though the disordered region has been called the RNA binding domain, all three domains have been shown to bind the genome [11]. The LKR also has a serine/arginine (SR) rich region where it is believed phosphorylation occurs, which could decrease the affinity for nucleic acid and be important for disassembly and release of the genome [11]. Within the N protein there are IDRs (intrinsically disordered regions), regions of the protein that do not have a fixed 3D structure, that are conserved between different CoVs. It has been shown that the IDR's in the N protein play an important role in increasing the binding affinity for the N protein to nucleic acids and the structural locations of the IDR's in the N protein are similar for different CoVs despite having varying amino acid sequences[12]. Though it is thought that the IDRs provide an allosteric cooperativity that enhances binding affinities, little is known about what physical properties of the IDR allow for enhanced binding to nucleic acids. There has also been evidence strongly suggesting a region of the genome, the Psi region, that first binds the disordered LKR region and may act as the initiator sequence for the organization of the genome. In general, intrinsically disordered proteins (IDPs) have evidence supporting that they are important for transcription, translation and cell-signaling [11]. We believe our knowledge of understanding the nature of binding between the CCMV capsid protein tail, an IDR, and various RNA's can also provide insight into the role of the IDR's in other IDP-RNA binding complexes.

The N protein has both specific and nonspecific interactions with genetic material and is involved in every step of the viral replication process. Understanding these interactions can lead towards the production of antivirals to help treat patients who are infected with CoVs. Using PMFs we can find small oligomers of RNA and small molecules that may bind stronger to the N protein than the viral genome which could interfere with the assembly process of the virus. We can also use markov-state modeling to find the binding mechanism of the IDR regions of the N protein. This could more broadly shed light on what about IDRs enhance

binding strengths. Unveiling the mechanism of binding could also lead to theraputics that interfere with specific conformational changes that are critical for the binding to viral RNA. We hypothesize that the IDR region could order upon binding, or move between metastable ordered states that strengthen the effective interaction while allowing an initial extension of the IDR to "grab" the binding partner.

1.3 Continuum electrostatics for polyelectrolyte binding

The theory of aqueous electrostatics in a continuum model is widely applicable to many problems in computational chemistry, problems in relation to biochemistry and materials science[13]. It provides a simplified yet insightful approach to studying electrostatic phenomena in aqueous solutions, the environment where almost all biology occurs. One of the key strengths of the continuum model is its ability to consider water as a homogeneous medium with uniform dielectric properties, significantly reducing computational complexity when compared to explicit solvent models such as molecular dynamics simulations. This enables researchers to explore large systems and complex biomolecules without compromising the overall accuracy of the calculations. It is effectively a level of "coarse graining" a simulation. Or decreasing the resolution of the system of interest to focus on the problem of interest. This is not applicable in every situation, since in some instances higher resolution is needed, as is the case in the effective attraction of polyelectrolyte binding [14, 15]. Additionally, the continuum model incorporates the concept of solvation, allowing for the inclusion of the solvent effects on molecular behavior, a crucial aspect in understanding biochemical processes and interactions.

The success of the continuum model in describing aqueous electrostatics lies in its foundation on the Poisson-Boltzmann equation.

$$\nabla^2\psi = \frac{c_o e}{\epsilon\epsilon_o}(e^{e\frac{\psi}{k_B T}} - e^{-\frac{e\psi}{k_B T}}) \tag{1.3.1}$$

Here ∇ is the differential operator, k_B is the Boltzmann constant, T is the temperature, c_o is the initial concentration of the ionic species, e is the electron charge, ψ is the electrostatic

potential, ϵ is the dielectric constant of the solvent and ϵ is the permittivity of free space. This partial differential equation characterizes the electrostatic potential in the presence of charges and ionic species, and it takes into account the distribution of ions surrounding the charged species. Usually analytical solutions are not available to this highly non-linear equation and numerical solutions are needed.

The continuum model has been widely validated against experimental data. Despite its merits, the continuum model does have limitations that need to be acknowledged. One primary concern arises from neglecting atomic details of the solvent molecules, which could be vital in certain situations where explicit solvent models or molecular dynamics simulations are better suited. The continuum model might not capture the subtle changes in solvent structure and dynamics that occur during specific interactions.

Additionally, the accuracy of the continuum model relies on the correct determination of solute charges and radii, which can be challenging, especially for large biomolecules or exotic species. This introduces uncertainties that may propagate through the calculations and lead to deviations from experimental results.

1.4 Effects of structural rigidity on biopolymer binding

One of the aims of this thesis is to explore the effect of the structural flexibility on RNA-protein interaction using all-atom molecular dynamics(MD) simulations [16]. This is done for the specific case of short polyU and polyA ssRNA sequences, as well as hybridized polyU-PolyA sequences, that interact with short (ten residue) positively charged peptide sequences of CCMV capsid proteins that are known to associate with ssRNA molecules. The sequence of calculated binding affinities agree with the observed sequence of packaging competition experiments. The variation of the binding affinity for different degrees of structural ssRNA flexibility is comparable in magnitude to the binding affinity and has been shown to account for binding affinity experimentally as well [17]. Next, we find that the *counter-ion release* mechanism – involving the negatively charged phosphate groups and Na^+ ions –

plays an important role for protein-ssRNA binding, as it does for protein-dsDNA interactions. However, we also encounter substantial variation between the spatial distribution of the small ions further away from the phosphate groups. In contrast with the case of non-specific protein-dsDNA interactions where electrostatic interactions dominate, we find that hydrogen bonding and hydrophobic interactions also make significant contribute contributions to non-specific protein-ssRNA interactions. We with a discussion of viral assembly in the context of our results.

1.5 Intrinsically Disordered Proteins

Intrinsically Disordered Proteins (IDPs), identical to the IDR regions we have mentioned above, are of significant interest due to their unique characteristics, primarily their lack of stable three-dimensional structures. These flexible proteins possess an array of binding sites and can interact with multiple partners in a dynamic way, transitioning between many metastable states. This poses a problem for traditional methods of calculating things like binding energies, where thermal equilibrium is usually required, because there are long-lived states that are not the minimum free energy structure.

The binding properties of IDPs stem from their structural plasticity. Unlike well-folded proteins with rigid tertiary structures, IDPs exhibit an ensemble of conformations that enable them to adapt their shape to that of their binding partners. This inherent flexibility facilitates both high specificity and diversity in their interactions with other biomolecules[18]. In cases such as that of the disordered N-termini of CCMV, this can be advantageous for packaging non-indogenous RNA and other charged macro molecules for the purpose of drug and gene therapy delivery methods. IDP binding regions often contain short linear motifs, which can be recognized by their partner's binding domains, such as SH3, SH2, and PDZ domains, among others[19].

IDP binding interactions play crucial roles in various cellular processes, including cellular signaling, transcriptional regulation, and post-translational modifications. By acting as

molecular switches, IDPs can rapidly transition between disordered and ordered states upon binding to different partners [20]. This ability allows IDPs to participate in diverse signaling pathways, where they mediate signal transduction and cellular responses to environmental changes. Moreover, IDPs are frequently involved in complex formation and assembly, where their flexible nature enables the formation of large macromolecular complexes with high specificity.

Dysregulation of IDP binding interactions has been linked to numerous diseases, including cancer, neurodegenerative disorders, and viral infections[21, 22]. In some cases, IDP binding may lead to uncontrolled cell proliferation, misfolding of proteins induced by the IDP binding, or impaired immune response. Understanding the binding properties of IDPs and their dysregulation in disease settings has provided new insights into potential therapeutic strategies.

Studying IDP binding presents various challenges, including the lack of well-defined structures and the transient nature of interactions. Advanced biophysical techniques, such as NMR, surface plasmon resonance (SPR), and molecular dynamics simulations, have been instrumental in investigating IDP binding kinetics and thermodynamics[23, 24, 18]. Here we hope to expand on general knowledge of general phenomena of proteins and IDPs. The ability to exist in an ensemble of structures that are not well defined posits a challenge in defining things like classical flexibility when dealing with these systems. We propose highly nonlinear models that can reproduce experimental results that may be of use in understanding binding dynamics of IDPs.

CHAPTER 2

All-Atom Molecular Dynamics Simulations of protein-RNA binding

The viral encapsidation process and RNA-protein binding systems have been investigated by many mean-field continuum theoretical models, coarse-grained molecular dynamics simulations and by in vitro experiments[25, 26]. With the large improvements of empirically derived force fields over the past twenty years, all-atom molecular dynamics(MD) simulations have greatly improved their ability to mimic real systems and provide resolution not attainable from experiment. We investigate how viral capsid protein binds to different RNAs and different RNA motifs, focusing on the case of Cowpea Chlorotic Mottle Virus (CCMV), a small single-stranded genome plant virus that is especially well-characterized experimentally. We aim to explain the different protein-RNA interactions and packaging scenarios observed experimentally for "normal-composition" RNAs with extensive duplex formation and single-stranded loops, versus single-letter RNAs like polyU, for which no intramolecular base-pairing is possible.

2.1 Simulation details

Prior to performing simulations, most of the structures used in the MD simulations were built in the Chimera[27, 28] molecular visualization tool. Due to the disordered nature of both the protein and the RNA in these simulations, crystal structures were not readily available to use. This meant the structures were in some predefined geometry that was often not of biological relevance. The structures where then paired together in the docking software

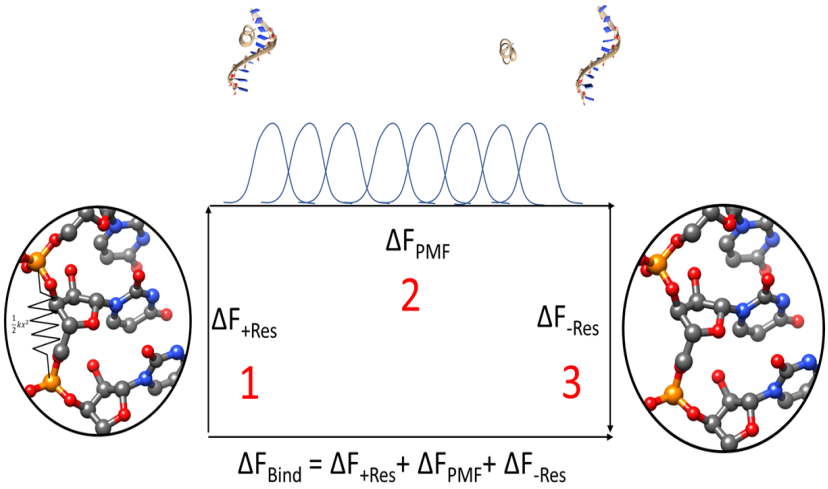


Figure 2.1.1: Schematic of thermodynamic path to estimate binding energy.

NPDock[29] that took the two structures and paired them in the best possible binding configuration in accordance with some scoring algorithm. Once the best binding structure had been assembled, that PDB file was used as the starting point for the subsequent MD simulations. All simulations were performed using the GROMACS 2018.1 software package[30]. The AMBER99SB forcefield was used and a particle mesh ewald (PME) method was used to compute electrostatics. A TIP3P representation of water molecules was used with a Noose-Hoover thermostat. simulations were carried out at 310K and a NaCl concentration of 0.1 mM (physiological conditions) with periodic boundary conditions. We have measured the binding energy of the CCMV CP N-terminus with various RNAs: polyU, polyA and an AU duplex. Each RNA was chosen to be long enough such that the number of phosphates along the RNA backbone matched the number of charged of the protein tail. The binding configurations were determined by constructing the RNA and protein sequence in CHIMERA, then using patchdock to generate a binding configuration between the two structures. The bound structures were then simulated for 100 nanoseconds to allow for relaxation to an equilibrium binbding structure. The structures after 100 nanoseconds were then used to compute binding energies.

2.2 Calculation of Binding Energy

Potential-of-mean-force (PMF) calculations is the choice method as it allows us to simplify the reaction coordinate down to a simple separation of binding partners. We estimate the binding energy to be the difference between the maximum and minimum points on the PMF curve. However, the flexible nature of the RNA and protein poses problems for a traditional PMF due to the slow orthogonal degrees of freedom. Thus, an alchemical path that includes a PMF was developed to measure binding energies. The orthogonal degrees of freedom (orientation of nucleobases with respect to the phosphate backbone, conformation of the phosphate backbone and translational and rotational motion) are restrained during the PMF calculation. Due to the large number of interactions that can change in the pulling process, it is necessary to restrain the molecule since the distributions measured in the PMF would not be equilibrium distributions in the umbrella potential. Once restrained, each umbrella window was simulated for 20ns since all slow degrees of freedom in the RNA are removed. The free energy of restraining the system was in the bound and unbound states were calculated via thermodynamic integration. Since the restrains are harmonic, a simple form of the free energy of the system is given by the following relation (details in appendix).

$$\Delta F_{res} = \int_0^{k_{umax}} \frac{1}{2} \langle (\Theta - \Theta^*)^2_{k_u} \rangle dk_u \tag{2.2.1}$$

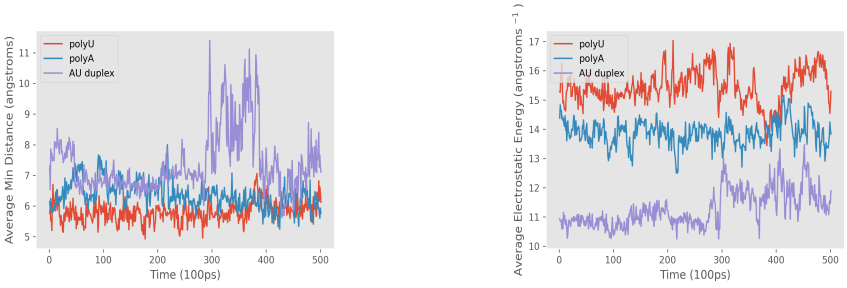
Where k_u is the force constant used for harmonic restraints, Θ is the coordinate representing the conformation of the molecule and Θ^* is the configuration of the molecules in the bound state. A schematic for the total binding energy is shown in figure 1.

2.3 Results and Discussion

RNA	Binding Energy (kT)	SASA(nm^2)	Hydrogen Bonds	Na released	Cl released
polyU	38	14.71	15.93	2.7	2.72
polyA	30	14.00	14.53	2.2	3.13
AU Duplex	23	8.95	8.17	1.8	2.66

Table 2.1: Binding energies for different RNAs along with quantities that contribute to the binding energy.

The binding energy was largest for the highly flexible polyU despite having the same charge density as polyA and half that of an AU duplex (tentative). Most likely when binding energies are redone, the order will be polyU > polyA > duplex. The new duplex being used has half the number of phosphates and is more representative of an expected duplex length in single stranded RNA. The large binding energy can be the result of a number of different molecular characteristics that are the result of polyU being a more flexible homopolymer. PolyA, and other polypurines, exhibit strong single stranded base stacking interactions that effectively increase the stiffness of the polymer [31]. The duplex has the bases involved in base paring, holding both chains together, keeping a more rigid structure than the polypyrimidines and polypurines. Do to the highly flexible backbone of polyU, the molecule can adopt a large



(a) (b)

Figure 2.3.1: (a) Averagem minimum distance between charged groups of protein and RNA.(b) Electrostatic energy for various bound RNAs

number of different conformations. Since the polyU backbone is able to bend easier than other homoRNAs, and certainly moreso than RNAs containing any secondary structure, it is better able to bring the negatively charged phosphates closer to the positively charged lysines and arginines located on the N-terminus of the protein. Table 2 and figure 2 shows that polyU brings the phosphate groups closest when compared to polyA and the duplex and has a higher average estimated electrostatic energy.

RNA	Average Minimum Distance (nm)	Average Electrostatic Energy (nm^{-1})
polyU	0.58 ± 0.01	1.55 ± 0.5
polyA	0.64 ± 0.02	1.32 ± 0.2
AU Duplex	0.7 ± 0.1	1.13 ± 0.4

Table 2.2: Average minimum distance between charged groups of protein tail and RNA and corresponding average approximate electrostatic energy ($\sum \frac{1}{|r_i - r_j|}$).

The small differences in the distances is also supported by the relative small difference in number of released counterions between the RNAs. Surprisingly, polyU is able to further stabilize these highly deformed structures as well due to non-canonical U-U transient base pairing. Another feature of this larger flexibility , polyU can rotate its bases about its backbone, and this feature allows for intramolecular hydrogen bonding similar to how single stranded hetero RNA folds. In comparison with polyA, which possesses single stranded base stacking characteristics in its coil configuration (pH > 6, the nucleobases prefer to stay aligned in helical order, due to the pi orbital interactions between the pyrimidine rings. This extra interaction stabilizes ordering of the bases and reduces the change the base will rotate about the phosphate backbone. Shown in figure 3 are two typical configurations for polyU and polyA while bound to the protein (with the protein not shown). The bases in polyA prefer to stay aligned with one another while the bases in polyU are more widely distributed around the backbone. The AU duplex is the most rigid structure, held in stable configurations by the base pairing between the A and U strands of RNA. Therefore its ability to neutralize charge is the most limited by the native structure of the RNA, fitting

through rotations and translations rather than deformations of the RNA. Table 1 shows that the polyU and polyA do better at neutralizing charge through release of counterions, maximizing the number of intermolecular hydrogen bonds and maximizing the change in solvent accessible surface area upon binding than the duplex. PolyA and polyU release on average 1 more counterion thn the duplex and average twice the number of hydrogen bonds. The also go through greater conformational changes upon binding the protein tail allowing them to maximize the difference in solvent accessible surface area between bound and unbound states, increasing the difference by around 5 nm^2.

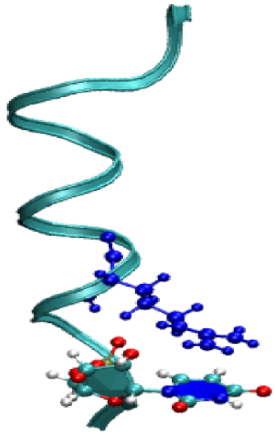


Figure 2.3.2: Illustration of a "cation-pi" interaction found in the protein-RNA complex between the N-termini and polyU. The long ribbon is the protein backbone, the multicolored structure with the blue hexagon is one residue from the polyU and the blue arginine, only one of the protein residues, is shown extending from the protein backbone.

We also found that the ability for polyU to more easily rotate the bases about the backbone allows for "cation-pi" interactions between the bases of the RNA and the positive charges on the protein. The quotations around cation-pi interactions are there because though these interactions are not explicitly modeled in molecular mechanics force fields, they can roughly be accounted for through the lennard-jones interactions. Force fields have

been well parametrized to account for base stacking interaction in nucleic acids[32], but explicit terms that would model cation-pi interactions would require knowledge of where the cation is around the nucleic acid ring (which to my knowledge has not been done yet). Both the guanidinium group on the arginine and the uracil ring on the polyU RNA in figure 3 have larger surface areas to interact to maximize the dispersion attraction term in the force field[32]. This could be the reason we see these apparent "cation-pi" interactions in simulation where the guanidine group and uracil group are lined up co-planarly as seen in figure 3. In figure 2 it is apparent how polyU can make its bases available to the protein tail and then the positive charge of the arginine finger can interact with the base of the RNA. It suffices to inform us that these types of interactions can and do happen in the case of a highly flexibly single stranded RNA. These interactions were not observed in polyA and the duplex case, though they are possible, because cation-pi interactions would have to compete with pi-pi interactions that are all ready present from the double stranded and single stranded base stacking in the duplex and RNA respectively.

(a) (b)

Figure 2.3.3: (a) Configuration of polyU in bound state and (b) configuration of polyA in bound state.

Rigidity plays a strong role not only in the ability to align the phosphates with positive charges, but also to increase the frequency of other intermolecular forces. The protein-RNA

complex has many possible hydrogen bonding partners. Having a more flexible backbone can increase the number of hydrogen bonds between RNA and protein to further stabilize the RNA-protein complex and increase the the binding energy. The number of potential hydrogen bonding partners between the RNA and protein is roughly the same for each of the RNAs. However, a large portion of the hydrogen bonding partners in the duplex are occupied in hydrogen bonding between the two strands of RNA. That is why the duplex has an average number of hydrogen bonds around half of that of the homopolymers. This is consistent with existing theories suggesting that nonspecific binding sequences in viral assembly bind to hairpins, which contain single stranded regions of RNA. PolyU has more hydrogen bonds on average than polyA because of the stiffness of polyA in regards to rotations of the bases around the backbone and deformation of the backbone itself making hydrogen bonding partners slightly harder to find each other. Similarly to how hydrogen bonding can be optimized by deformation of the RNA, solvent accessible surface area (SASA), can also be optimized from deformation of the RNA. This was calculated by looking at the difference in SASA before and after binding, giving a rough measure of van der waals contacts between the RNA and protein. It is a rough estimate because it is not exclusively given by hydrophobic contacts, but also charge neutralization and formation of hydrogen bonds. The same trend seen in hydrogen bonds is seen with SASA.

2.4 RNA characteristics

We find that polyU decreases in size upon binding the N-terminus of CCMV capsid protein. The change is only a difference of about half a nanometer in the radius of gyration. Since polyU and polyA are linear polyelectrolytes, this could suggest that there exists some level of condensation of RNA around the highly charged protein tails. The degree of condensation could depend on the effective flexibility of the homopolymer and could affect the geometry of the fully formed capsid. On a more macroscopic scale, we know that the viral RNA (duplex forming) has a smaller radius of hydration (R_h) when compared to the homopolymers as seen in figure 4[9]. It has also been shown that unlike DNA in the presence of sufficient salt

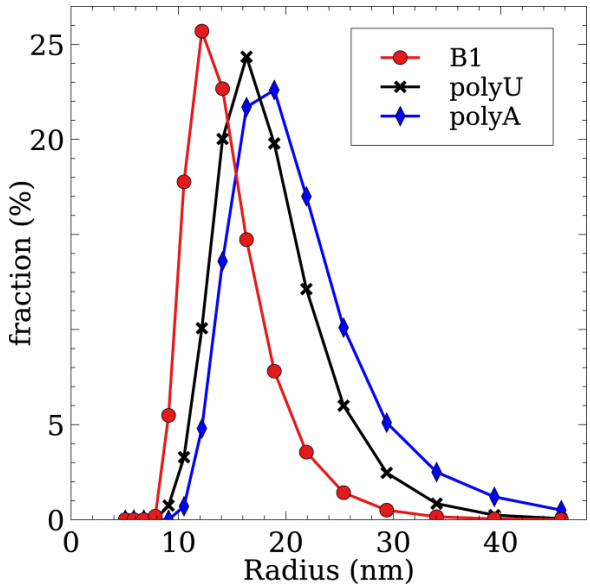


Figure 2.4.1: Size distributions of B1 RNA compared with those of polyA and polyU given by dynamic light scattering(DLS) of similar length RNAs.

concentration, RNA undergoes a continuous transition in size.

2.5 Stiffness dependent counterion release

In non-specific binding of RNA to protein in viruses, aqueous electrostatics plays a dominant role. The driving force for these binding events is the entropy release of bound counterions for the protein upon binding. In our simulations we have shown that counterion release does play a prominent role in binding and that the different RNAs exhibit different levels of counterion release that are consistent with experimental results. Figure 4 shows that upon binding of RNA to the protein tail, the distinct peaks of the coutnerion probability density for both sodium and chlorine start to diminish and approach a uniform distribution. Thus in binding the protein, the RNA neutralizes the positive charges on the protein.

We can see in figure 5 the radial distribution functions(rdf) for the different RNAs in bound vs unbound states, that polyU does the best job at both binding counterions and unbinding them upon complexing with protein. The rdfs are representative of the distance

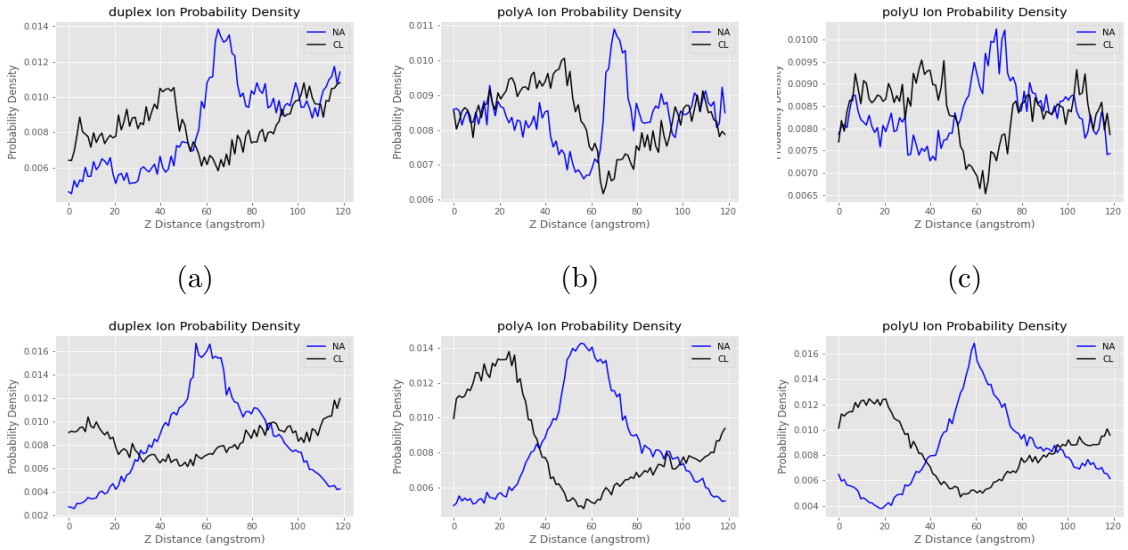


Figure 2.5.1: (a)-(c) Counterion probability density for RNAs in the bound state.(d)-(f) Counterion probability density for RNAs in the unbound state

of sodium atoms from the phosphates along the backbone of the RNA, and the distance of the center of mass of the charged groups on the protein (mostly nitrogen atoms on the lysine and arginine groups) and the chlorine atoms. The enhanced counterion release is due to the larger flexibility of polyA and polyU, which are better able to charge match the positive charges along the protein tail. Interestingly, the discrepency in bound counterions are largely due to a difference in binding the 0.7-2nm regime. These are counterions that are not tightly bound to the phosphates along the backbone of the RNA, but rather counterions that are loosely bound to the RNA. Investigating the trajectories, we find that in both the bound and unbound cases, there is usually one sodium bound closely (0.2-0.5nm) to the RNA represented by the first peak in the rdf. This sodium atom is usually transient and invariant among the different phosphates along the backbone, meaning that it can be found at any phosphate group and usually does not persist for long. In the bound case, we see that there is still a tightly bound sodium, but since the number of available negative charges is smaller due to the binding of charged groups from the protein, the frequency of observing this sodium is lower. By integrating the radial distribution function and multiplying by

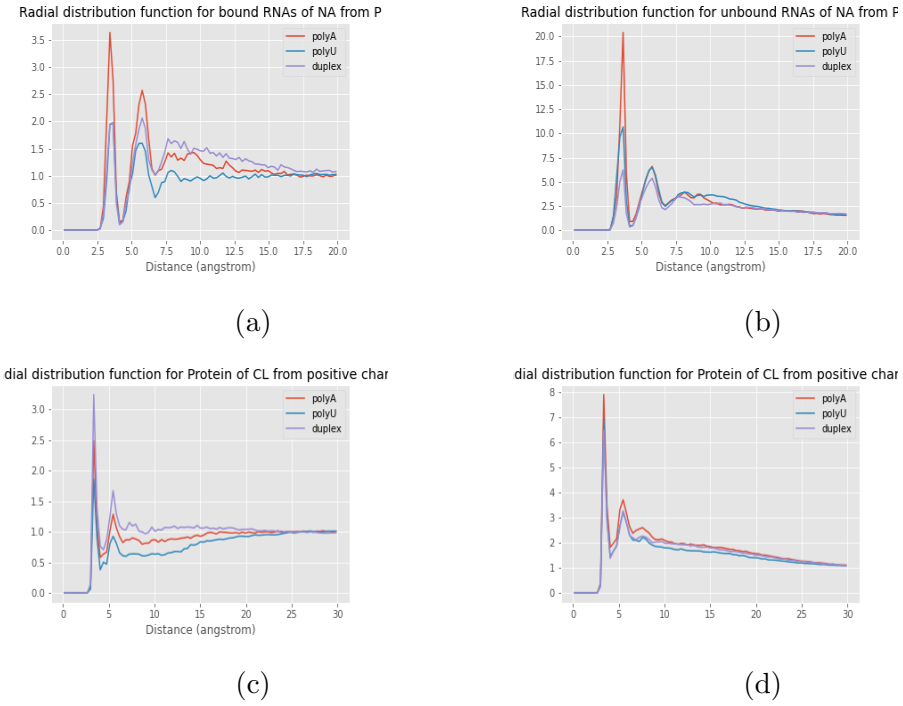


Figure 2.5.2: (a) and (b) Counterion structure around RNA.(c)-(d) Counterion structure around protein

the density(generating the cumulative distribution function(cdf)), we can see the average number of sodiums around the RNA and chlorines around the protein. By comparing the cdf's before and after binding to the RNA, we can see how many counterions are released upon binding. The results are shown in table 2.1.

2.6 Concluding remarks

We have shown that for short homonucleotide RNAs binding to an unstructured disordered protein, the RNAs with more flexibility bind more favorably. The work of Lui[26] showed that in the case of packaging of branched and linear polymers with associated stiffness, flexible polymers only when in the case of shorter chain length. Though it would be irresponsible to assert that the rational for stiffer polymers binding less favorably would lead to preferential encapsidation, it is interesting that on a molecule-molecule level of interaction the trend we

see is similar. At the very minimum, more favorable binding of the biopolymer would at least kinetically favor the preferential encapsidation of the more flexible polymer. In a multistep assembly process, like those performed in-vitro, the initial step is the binding of the RNA to the capsid protein. Then the solution conditions are changed to allow for reorganization to form the assembled capsids. It is understood that the mechanism of capsid formation can affect the nature of the formed capsid [25], to the best of our knowledge, it has not been investigated whether or not the mechanism of assembly affect which polyelectrolyte is preferentially packaged. If the reorganization occurs on a timescale larger than that of the unbinding process, the mechanism of assembly could be biased towards the preferred binding partner rather than the lower free energy of the fully formed capsid.

We have also shown that the counterion condensation is enhanced for the more flexible polyelectrolytes. It has been shown that for polyelectrolytes there is an effect on the effective persistence length of the polymer due to effective screen of intrachain repulsion and effective attractions due to correlations in the fluctuations of bound counterions [33]. In future work it would be interesting to investigate if the inverse were true; flexibility of the polymer and fluctuations in the conformation of the polymer, related to polymer stiffness, affect the ability for counterions to bind.

CHAPTER 3

Analytical Models for the Non-specific and Specific Binding of Polymers

In Section II we found that when an RNA sequence binds to a peptide sequence then the binding energy to DNA of an RNA sequence with shorter persistence length is larger than that of an RNA sequence with longer persistence length. A number of papers in the soft-matter literature have discussed the problem of the adsorption of polymers onto rigid rods[34, 35].

In general, decreasing the stiffness of the polymer increases the role of thermal fluctuations causing unbinding of the polymer from the rod. An interesting way to see this is by using the mapping between a directed polymer and a quantum particle, as first developed by de Gennes, where thermal fluctuations in three dimensions translate to the quantum fluctuations of a quantum particle in two dimensions[36].

For the present case, the binding of a polymer to a rigid rod translates to the interaction of a quantum particle moving in two dimensions with an attractive potential well at the origin. The adsorption of the polymer to the rod corresponds to appearance of a bound state. Decreasing the stiffness corresponds to decreasing the mass of the particle and hence increasing the role of fluctuations, causing unbinding. This type of fluctuation-induced unbinding is quite general. The physics of the adsorption of a polymer to a surface shows the same effect[37, 38].

This well-established body of literature seem to be in conflict with our simulation results! The more flexible ssRNA has a higher binding energy than the more rigid duplex forming RNA and the more flexible ssRNA (polyU) has a higher binding energy than polyA which 19

has increased rigidity due to single stranded base stacking. In this section we discuss simple statistical models that show that there are competing physical mechanisms in terms of the effects of flexibility when both specific and non-specific interactions are included.

3.1 Semi-flexible polymers interacting with a rigid rod via non-specific and specific interactions.

We start with a model where a semi-flexible polymer interacts with a rigid polymer, first for the case where only non-specific interactions are included

3.1.1 Semi-flexible polymer interacting with a rigid rod via non-specific interactions

Consider a semi-flexible homo-polymer of length L interacting with a long rigid rod. The system is confined to two-dimensions. The "bending energy Hamiltonian" of a semi-flexible polymer is given by

$$H_b = \frac{1}{2}\kappa \int_0^L \left(\frac{d\theta}{ds}\right)^2 ds \tag{3.1.1}$$

L represents the total length of the polymer. Next, κ is the bending modulus, which is connected to the persistence length, l_p, by the relation $\kappa = k_b T l_p$, where k_b is the Boltzmann constant and T is the temperature. As shown in figure 3.1.1, θ is the angle between the tangent vector of the polymer and the x-axis and s is the arc length coordinate that goes from 0 to L. The derivative of θ w.r.t to s, $\frac{d\theta}{ds} = \frac{1}{R(s)}$ has the meaning of the local curvature of the polymer with $R(s)$ the curvature radius, as can be seen in figure (INSERT IMAGE). The figure shows how the infinitesimal segment length of a circle of radius R is given by $Rd\theta = ds$, which leads to the relation between the curvature radius, $R(s)$, and the derivative of θ above. This Hamiltonian can be rewritten in therms of the curvature as

$$H_b = \frac{1}{2}\kappa \int_0^L \frac{1}{R(s)^2} ds \tag{3.1.2}$$

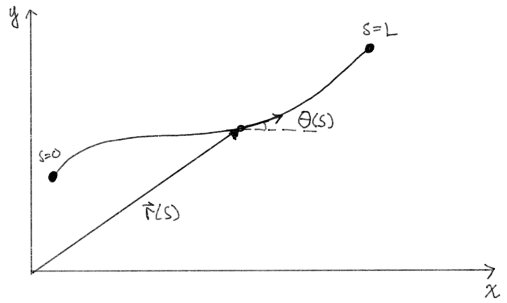


Figure 3.1.1: Diagram of system.

This expression has the same form as the elastic energy of a rod that has been bend in a particular shape. It has been extensively used as a description for the deformation energy cost of bio-polymers such as DNA and RNA on large length-scales[39].

For double-stranded DNA and RNA, the persistence length is in the range of 50 nm while for single-stranded DNA and RNA it is in the range of one nm. One can add a term to include a "preferred curvature", for bio-polymers like RNA and DNA that are heterogeneous, by subtracting a curvature function, $f(s)$, from the inverse of the radius of curvature in equation 3.

The Hamiltonian H_b must be viewed as the continuum, or long wavelength, limit of a more detailed microscopic model. It will be necessary for the calculation of partition sums and free energies to discuss this. Suppose a polymer is modeled as a sequence of rigid links – the monomers – that are joined into a chain via joints that have a limited amount of rotational degrees of freedom. Saturated hydrocarbon chains can, for example, be described in this way. Confine this chain again to two-dimensions and let θ_n be the angle of the n'th link of the chain with respect to the x-axis. A possible microscopic Hamiltonian for such a chain would be $H_{X-Y} = -(\kappa/a)\sum_{m=0}^{N_m}\cos(\theta_{m+1} - \theta_m)$ with θ_m restricted to the interval

$[0, 2\pi]$. This is a well-studied Hamiltonian, known as the one-dimensional X-Y model. Note that the angular degrees of freedom are properly periodic with period 2π. The partition sum of the model is $Z = \prod_{m=0}^{m=N} \int_0^{2\pi} d\theta_m \exp -\beta H_{X-Y}$ with $\beta = 1/k_B T$. The free energy has the form $F_{X-Y} = -k_B T ln Z = N k_B T f(\beta\kappa/a)$ with $f(x)$ a dimensionless function. The important feature here is that the microscopic length scale a – the monomer size –, which does not appear in H_b, does enter in an essential way in the free energy expression.

The next step is to specify the interaction of the polymer with the rod, which will be assumed to lie along the x-axis. Assume a non-specific interaction potential $v(x, y)$ per unit length that is independent of x and that has a minimum with respect to y at y=0. Expansion in a Taylor series gives

$$v(x, y) \simeq v_o + \frac{1}{2} k y^2 + ... \tag{3.1.3}$$

with $k = \frac{d^2 v(x, y=0)}{dy^2}$ and with v_o the binding energy per unit length for the case that the polymer also lies along the x-axis. The binding energy v_o is estimated as has $k_b T$/nm while the quantity k, which has dimensions of energy per unit volume ($[k] = \frac{[E]}{[L]^3}$), is estimated as $k_b T/nm^3$. The interaction contributes a term to the Hamiltonian given by

$$H_{int} \simeq -v_o L + \frac{1}{2} k \int_0^L y(s)^2 ds \tag{3.1.4}$$

where $y(s)$ is understood as the local displacement from the x-axis for a monomer that is located at an arc distance s along the polymer.

We must combine the bending and interaction energies. We used two notations to represent the conformation of the polymer: $\theta(s)$ and $y(s)$. The $\theta(s)$ representation is, as noted, convenient to compute the free energy of the polymer that is unbound and free in solution. On the other hand, for a nearly straight polymer located close to the x-axis, the $y(s)$ representation could be replaced by a $y(x)$ representation. This is known as the "Monge Representation" and this representation happens to be very convenient to compute thermal fluctuations for the case where the polymer is bound to the rod. However, our aim is to compute the free energy of assembly. The Monge representation is known to lead to serious errors in the computation of free energies and heat capacities.

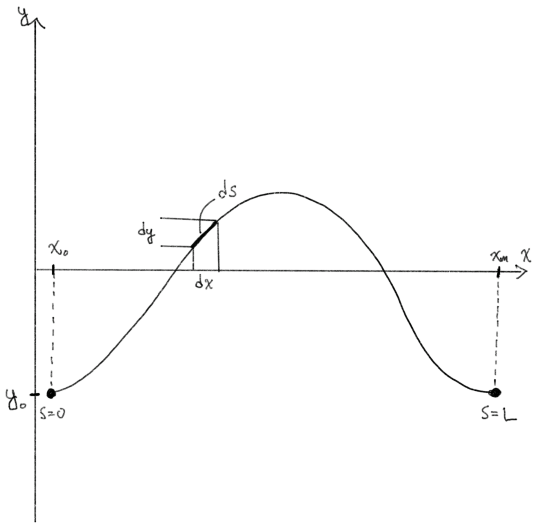


Figure 3.1.2: Diagram of system.

For this reason we must do our calculation in the $\theta(s)$ representation, which requires us to express $x(s)$ and $y(s)$ in terms of $\theta(s)$. Following the schematic of figure 3.1.2, and assuming that the potential is small so that $\theta(s)^2$ is small, we can transform our coordinates to

$$ds = (dx^2 + dy^2)^{\frac{1}{2}}$$

pulling out dx^2, we get

$$ds = dx\left(1 + \left(\frac{dy}{dx}\right)^2\right)^{\frac{1}{2}}$$

Taylor expanding and inserting $dy \approx \theta dx \approx \theta ds$

$$ds \approx dx(1 + \frac{1}{2}\theta(s)^2)$$

Solving for $x(s)$ we then get

$$x(s) = x_o + \int_0^s \frac{1}{1 + \frac{1}{2}\theta(s')^2} ds'$$

Taylor expanding the integrand and integrating we finally reach

$$x(s) \simeq x_0 + s - \frac{1}{2}\int_0^s \theta(s')^2 ds' \tag{3.1.5}$$

with x_0 the x-coordinate of the $s = 0$ end of the polymer. Doing the same with y we have

$$y(s) \simeq y_0 + \int_0^s \theta(s')ds' \tag{3.1.6}$$

with y_0 the y-coordinate of the $s = 0$ end. We will compute the free energy for fixed x_0 and y_0, which can be viewed as a potential of mean force. Plugging equation 7 into the expression for H_{int} gives to lowest order

$$H_{int} \simeq -v_o L + \frac{1}{2}k\left[\int_0^L ds\left(\int_0^s ds'\theta(s')\right)^2 + y_0\int_0^L ds\int_0^s ds'\theta(s') + y_0^2\right] \tag{3.1.7}$$

The boundary conditions are determined by the fact that there is no bending torque applied at the ends of the polymer, which means that $\kappa\frac{d\theta}{ds}|_{s=0} = \kappa\frac{d\theta}{ds}|_{s=L} = 0$. Since this expression is a quadratic functional of $\theta(s)$ it is natural to expand the interaction Hamiltonian in a mode expansion. Rewrite

$$\theta(s) = \sum_{n=1}^{N_m} A_n \cos q_n s \tag{3.1.8}$$

Where $q_n = \frac{2\pi}{\lambda_n} = \frac{n\pi}{L}$ is the wavenumber of the mode. The two modes corresponding to rigid body translation are included separately by retaining x_0 and y_0 as two separate degrees of freedom. Note that each term in the expansion obeys the zero torque boundary condition. Next, N_m is the number of monomers, equal to $\frac{L}{a}$ with a the monomer spacing of the polymer and L still the polymer length. It is easy to check that

$$\int_0^s \theta(s')ds' = \sum_{n=1}^{N_m} \frac{A_n}{q_n}\sin q_n s \tag{3.1.9}$$

Using this result and plugging it into the terms of equation 8 we have

$$\int_0^L ds\int_0^s ds'\theta(s') = \sum_{n=1}^{N_m} \frac{A_n}{q_n^2}(1 - \cos q_n L)$$

Since the expression in the parenthesis is equal to 0 when n is even and 2 when n is odd, we can simplify to

$$\int_0^L ds\int_0^s ds'\theta(s') = 2\sum_{n=1}^{N_m} \frac{A_n}{q_n^2} \tag{3.1.10}$$

Plugging equation 10 into the square term in the interaction Hamiltonian we get

$$\int_0^L ds\left(\int_0^s ds'\theta(s')\right)^2 = \int_0^L ds\left(\sum_{n,n'=1}^{N_m} \frac{A_n A_{n'}}{q_n q_{n'}}\sin q_n s \sin q_{n'} s\right)$$

The term in the parenthesis simplifies to $-\frac{1}{2}\sum_{n,n'}\frac{A_nA_{n'}}{q_nq_{n'}}[\cos(q_n+q_{n'})s-\cos(q_n-q_{n'})s]$. Integrating this with respect to s from 0 to L then gives

$$-\frac{1}{2}\sum_{n,n'}\frac{A_nA_{n'}}{q_nq_{n'}}\left(\frac{1}{q_n+q_{n'}}\sin(q_n+q_{n'})L-\frac{1}{q_n-q_{n'}}\sin(q_n-q_{n'})L\right) \tag{3.1.11}$$

The terms in the sum are equal to 0 unless $n=n'$ which then simplifies everything to

$$\int_0^L ds\left(\int_0^s ds'\theta(s')\right)^2=\frac{L}{2}\sum_{n=1}^{N_m}\frac{A_n^2}{q_n^2} \tag{3.1.12}$$

Rewriting the interaction Hamiltonian in the mode expansion gives

$$H_{int}=\frac{1}{2}k\left(y_o^2L+4y_o\sum_{n=odd}^{N_m}\frac{A_n}{q_n^2}+\frac{L}{2}\sum_{n=1}^{N_m}\frac{A_n^2}{q_n^2}\right)-v_oL \tag{3.1.13}$$

If we plug the mode expansion into equation 2 to get the bending energy of the polymer, we get

$$H_b=\frac{1}{2}\kappa\sum_{n,n'}A_nA_{n'}\int_o^L dsq_nq_{n'}\sin q_ns\sin q_{n'}s=\frac{\kappa L}{4}\sum_{n=odd}^{N_m}A_n^2q_n^2 \tag{3.1.14}$$

Thus, the total Hamiltonian for the system is

$$H=\frac{L}{4}\sum_n A_n^2[\kappa q_n^2+kq_n^{-2}]+2ky_o\sum_{n=odd}^{N_m}\frac{A_n}{q_n^2}+\frac{1}{2}ky_o^2L-v_oL \tag{3.1.15}$$

The Hamiltonian is expressed in terms of the A_n and y_0 coordinates, which are coupled because of the cross term. We can rewrite the Hamiltonian in terms of uncoupled coordinates by defining $\alpha_n=A_n-A_n^*$ where $\frac{dH}{dA_n^*}=0$. Plugging this into equation 14 and solving for A_n^* we arrive at

$$A_n^*=-\frac{4ky_o}{L}\frac{1}{k+\kappa q_n^4} \tag{3.1.16}$$

Inserting this back into equation 16 we get

$$H=-v_oL+\frac{L}{4}\sum_{n=odd}^{N_m}\left[\kappa q_n^2+\frac{k}{q_n^2}\right]\alpha_n^2+\left[\frac{kL}{2}-\frac{4k^2}{L}\sum_{n=odd}^{N_m}\frac{1}{q_n^2}\left(\frac{1}{k+\kappa q_n^4}\right)\right]y_o^2 \tag{3.1.17}$$

Where we now have the decoupled modes α_n and y_o. Define the function $g(z)$ with $z=\frac{\kappa\pi^4}{kL^4}$ where

$$g(z)=1-\frac{8}{\pi^2}\sum_{n=odd}^{N_m}\frac{1}{n^2+zn^6} \tag{3.1.18}$$

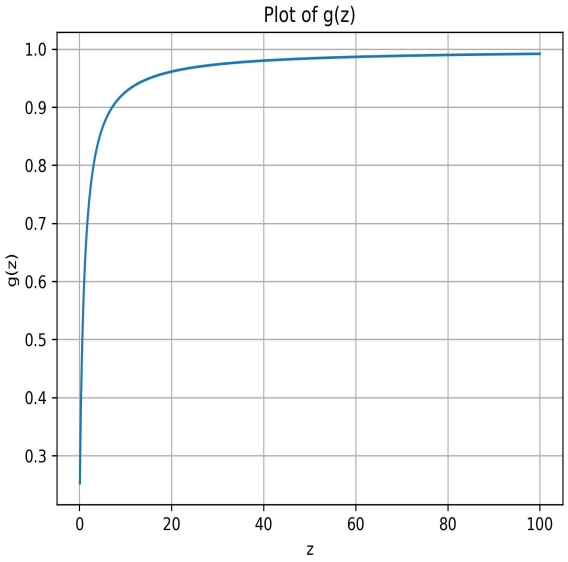


Figure 3.1.3: Plot of $g(z)$

Note that $g(0) = 0$. The limiting relations for small and large z are

$$g(z) \simeq \begin{cases} c_0 z^{\frac{1}{4}} & z \to 0 \\ 1 - \frac{c_1}{z} & z \to \infty \end{cases} \tag{3.1.19}$$

with $c_0 \simeq 0.45$ and $c_1 \simeq 0.81$. This function can be seen in figure 3.1.3 This lets us rewrite the Hamiltonian as

$$H = -v_o L + \frac{kL^3}{4\pi^2} \sum_{n=odd}^{N_m} \left(zn^2 + \frac{1}{n^4} \right) \alpha_n^2 + \frac{1}{2} kLg\left(z\right) y_o^2 \tag{3.1.20}$$

It follows from this expression that in the presence of a series of harmonic modes, the polymer continues to interact with the rod via a parabolic potential when displaced along the y direction as a rigid body. The spring constant has however changed from kL to $kLg(z)$. Since $g(z)$ is an increasing function of z and since increased flexibility decreases z, increased flexibility weakens the harmonic restoring force. It is clear from this expression that the length scale $\xi = \left(\frac{\kappa}{k}\right)^{1/4}$ plays an important role. If the polymer length L is small compared to ξ than the effective spring constant is close to kL while in the limit that L is large compared to ξ, the spring constant is "renormalized" to, roughly, $k\xi$. In the statistical physics literature, this is known as a scaling relation with L/ξ as the scaling variable. The

length ξ is similar to the Odijk deflection length [40], which is the length scale at which a polymer collides with the wall of a confining cylinder. In our case the confining cylinder is represented by the binding potential and its curvature k, while κ is related to the persistence length of the polymer. Thus the effectiveness of the binding is dependent on the statistics of a confined polymer. Note that there are now three length scales: the microscopic length scale a, the persistence length κ/k_BT, and the effective Odijk length ξ. The scaling variable z is proportional to $\left(\frac{\xi}{L}\right)^4$.

The free energy of a system of decoupled harmonic degrees of freedom can be calculated from the partition sum. Note that the mode amplitudes A_n are dimensionless. The resulting potential of mean force

$$F_T(y_0) = \frac{1}{2}k_bT \sum_{n=odd}^{N_m} \ln\left(n^2\pi^3 l_p/L\left(1+\frac{1}{zn^4}\right)\right) + \frac{1}{2}kLg\left(z\right)y_o^2 - v_oL \tag{3.1.21}$$

does not depend on x_0. We will define the binding energy of the system as the difference between the free energy when it is bound to the rod and when it is free in solution. In the latter case, only the bending energy contributes to the Hamiltonian, which is equal to $\frac{1}{2}k_bT\sum_{n=odd}^{N_m}\ln\left(n^2\pi^3 l_p/L\right)$ This gives, after thermal averaging, over y_0

$$\Delta F = \frac{1}{2}k_bT \sum_{n=odd}^{N_m} \ln\left(1+\frac{1}{zn^4}\right) + \frac{1}{2}k_bT\ln(\beta\pi kL\delta^2 g(z)) - v_oL \tag{3.1.22}$$

where δ is a microscopic length that does not depend on κ (hint: take the classical limit of the free energy of a quantum harmonic oscillator).

We can now examine the effects of flexibility on the binding energy in different limiting cases. Note that the flexibility only enters in ΔF through the scaling variable z.

The first regime will be the limit where $z >> 1$. This is the "stiff" limit where the Odijk length ξ is large compared to the polymer length L. The first term can be Taylor expanded for any n since z is large compared to one. This comes down to having to evaluate the sum $\sum_{n=1}^{\infty}\frac{1}{n^4} = \gamma \simeq 1.06$, which leads to $\frac{\gamma}{2}\frac{k_bT}{z}$. In this limit $g(z) \simeq 1 - \frac{c_1}{z}$. The second term in equation 24 is the sum of a term independent of z plus a negative term $-\frac{1}{2}c_1k_BT/z$ proportional to k_BT/z. The two terms are both proportional to $1/z$, which produces a net

positive contribution to the binding free energy of about $0.25k_BT/z$. Increasing the flexibility means decreasing z. This reduces the magnitude of the binding free energy ΔF, so flexibility indeed decreases the binding affinity, as well as the restoring force, in accordance with the earlier studies cited above. However, the $0.25k_BT/z$ contribution is small compared to the thermal energy scale so this correction term is in general negligible compared to the bare binding energy $-v_oL$.

In the second regime one assumes the opposite, namely $z << 1$ so where the Odijk length ξ is small compared to the length L of the polymer, but we will also assume that z still is larger than $1/N_m^4$. Physically, this means that the Odijk length is large compared to the monomer size a. In this regime, one can split the sum in equation 24 as

$$\Delta F(z) \approx \frac{1}{2}k_bT\left[\sum_{n=1}^{\left(\frac{1}{z}\right)^{\frac{1}{4}}}\ln\left(\frac{1}{n^4z}\right)+\sum_{n=\left(\frac{1}{z}\right)^{\frac{1}{4}}}^{N_m}\left(\frac{1}{n^4z}\right)\right]+\frac{1}{2}k_bT\ln\left(d_oz^{\frac{1}{4}}\right)-v_oL \qquad (3.1.23)$$

with d_0 a constant independent of z. Turning summation into integration gives the estimate

$$\Delta F(z) \approx \frac{1}{2}k_bT\left[\left(\frac{1}{z}\right)^{\frac{1}{4}}\ln\left(\frac{1}{zd_1}\right)\right]+\frac{1}{2}k_bT\ln\left(d_0z^{\frac{1}{4}}\right)-v_oL \qquad (3.1.24)$$

with d_1 a numerical constant of the order of one. The first term is proportional to the polymer length L since z is inversely proportional to $1/L^4$. The correction term no longer can be neglected with respect to the bare binding energy. The second term in the expression is not extensive and can be neglected, which gives

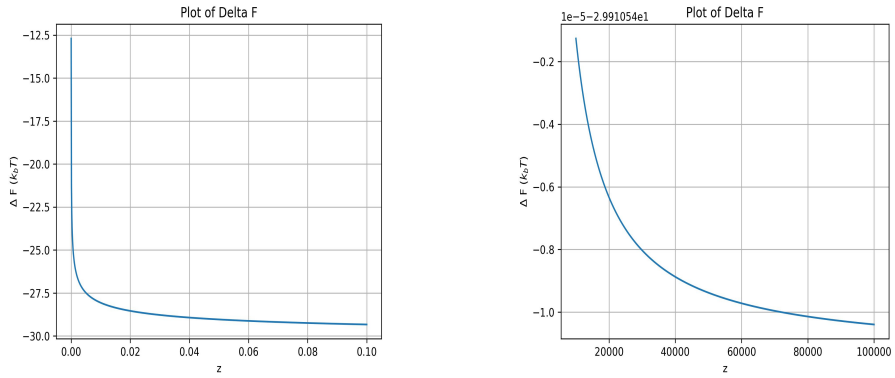
$$\Delta F(z) \approx \frac{2}{\pi}k_bT\left(\frac{L}{\xi}\right)\ln\left(\frac{L}{d_2\pi\xi}\right)-v_oL \qquad (3.1.25)$$

with d_2 a numerical constant of the order of one. Assuming the logarithm to be of the order one, leads to the estimate that thermal fluctuations produce unbinding of the polmer from the rod when k_BT/ξ is larger than the v_0.

In regime three, one looks at the case where $z << 1/N_m^4$, which is when the Odijk length is small compared to the monomer size a. Under this assumption one can approximate the binding free energy as

$$\Delta F \approx \frac{1}{2}k_bT\sum_{n=1}^{N_m}\ln\left(\frac{1}{zn^4}\right)+\frac{1}{2}k_bT\ln g(z)-v_oL \qquad (3.1.26)$$

One can again estimate $\Delta F(z)$ by treating the logarithms as quantities that are of the order of one. The first two terms are together of the order of $k_B T L/a$. Now, one expects thermal fluctuations to produce unbinding when $k_B T/a$ is larger than v_0. In each of the three regimes,



(a) ΔF as a function of z for $z << 1$. (b) ΔF as a function of z for $z- > \infty$.

the bound state becomes less favorable when the bending modulus κ is reduced as seen in the plots of figures 3.1.4a and 3.1.4b plotting (3.1.22). The polymer is bound better when the bending cost is large so it can stay in a straight configuration and best line up with the rigid rod. If a polymer has a large amount of curvature due to thermal fluctuations then this pushes the polymer farther up the potential well and further from the bound state. This is related to concept of the Odijk length ξ as the length at which the polymer collides with a confining tube after touching the wall once. Every collision decreases the entropy by an amount of the order of k_B and increases the free energy by an amount of the order of $k_B T$. As the Odijk length gets smaller, more "collisions" with the confining potential happen causing the system to lose about $k_b T$ in free energy per collision. The entropic cost of confining the polymer gets larger and a bound state gets less favorable. As the Odijk length gets larger, the more favorable binding should be. That is to say that given similar potentials, the stiffer polymer should bind more favorably with non-specific interactions due to entropy loss of confining a flexible polymer. This argument leads one to estimate that binding the polymer to the rod reduces increases the entropic free energy $-TS$ by an amount $k_B T L/\xi$ with every collision reducing the entropy by an amount k_B. Recall that that is exactly what we found in the second regime where the Odijk length was small

compared to the polymer length L but large compared to the monomer size a. We conclude by noting, first, that the concept of the Odijk length loses validity when the Odijk length exceeds the persistence length, which happens in the limit of very small k, because in that case the polymer can coil up in between the collisions and, second, that a more quantitative theory of the polymer/rod unbinding unbinding transition requires including in the partition sum the appearance of "loops" where a section of the polymer moves outside the range of the attractive potential. With these caveats in mind, the concept of the Odijk Length still provides us with an important tool for the analysis of the molecular dynamics simulations of the previous section.

3.1.2 Specific Interactions and the Quenched Average

The interaction between the polymer and the rod did not include in any way local molecular structure or specific interactions. In this section we extend the model by including these effects. This will be done in a minimalistic way by generalizing the interaction potential $v(x, y)$. Specifically, assign again Cartesian coordinates $x(s), y(s)$ to a polymer segment located an arc-distance s along the polymer. The segment now experience a potential per unit length given by

$$v(x, y) = -v_o + \frac{1}{2}k\left(y(s) - y_s(x(s))\right)^2 \tag{3.1.27}$$

with $y_s(x)$ a certain function. The first interpretation is to view $y_s(x)$ as a lateral displacement of the straight rod of the previous section in the y direction that varies along the x direction. An alternative interpretation is obtained by expanding the square:

$$v(x, y) = -v_o + \frac{1}{2}ky(s)^2 - f(x(s))y(s) + \frac{1}{2}ky_s(s)^2 \tag{3.1.28}$$

where $f(x(s)) = ky_s(x(s))$ has dimensions of force per unit length. In this case we maintain the straight rod configuration of the previous section but introduce a specific local interaction in the form of a force $f(x)$ that varies along the rod. Note that we can drop here the last term since it does not affect the interaction between the rod and the polymer. In the following we will retain this second interpretation

Assume that the polymer is in a solution that contains rod-like molecules with a wide variety of specific force distributions $f(x)$. Our aim is to compute the binding energy for a generic force function $f(x)$. The free energy will be computed for an unspecified force function $f(x)$ after which the typical free energy is obtained by an unbiased average over a range of different force functions. In the physics literature this procedure of computing statistical averages in random environment is known as a "quenched average". Performing quenched averages often leads to a considerable mathematical complication known as "replica symmetry breaking". The averaging step is done by treating the force function $f(x)$ as a random variable with zero mean and with no correlations between different locations. Specifically, the force-force correlation is $\langle f(x)f(x')\rangle = \Delta\delta(x-x')$ with Δ having dimensions of the square of force divided by a length or energy squared divided by length cubed. This quantity should be compared with the second derivative k of the non-specific interaction, which has dimensions of energy divided by length cubed. It follows that Δ/k has dimensions of energy, which suggests that the specific interactions become important when this energy scale exceeds the thermal energy scale k_BT

The specific force per per unit length introduces a new term in the Hamiltonian denoted by H_f, given by

$$H_f \simeq -y_o \int_0^L ds f(x_o+s) - \int_0^L ds f(x_o+s) \int_0^s ds' \theta(s') \tag{3.1.29}$$

Note that H_f now depends on the initial location of the polymer along the x direction, which was not the case when we had only non-specific interactions. As in the previous section, we switched from Cartesian to polar coordinates and assumed $\theta(s)$ to be small compared to one. Using the same mode expansion as before

$$\theta(s) = \sum_{n=1}^{N_m} A_n \cos q_n s$$

gives for the new term in the Hamiltonian:

$$H_f = -y_o \int_0^L ds f(x_o+s) - \sum_{n=1}^{N} \frac{A_n}{q_n} \int_0^L ds f(x_o+s) \sin q_n s \tag{3.1.30}$$

Define the sine transform of the force profile as $\bar{f}_n \equiv \int_0^L ds f(x_o + s) \sin q_n s$ and $\bar{f}_o \equiv \int_0^L ds f(x_o + s)$. Note that the f_n depend on the coordinate x_0. We can rewrite the equation as

$$H_f = -y_o \bar{f}_o(x_o) - \sum_{n=1}^{N} \frac{A_n}{q_n} \bar{f}_n(x_o) \tag{3.1.31}$$

This gives us a full Hamiltonian

$$H = \frac{L}{4} \sum_{n=odd}^{N} A_n^2 [\kappa q_n^2 + k q_n^{-2}] - \sum_{n=odd}^{N} A_n \left[\frac{\bar{f}_n}{q_n} - \frac{2ky_o}{q_n^2} \right] - y_o \bar{f}_o + (\frac{1}{2} k y_o^2 - v_o) L \tag{3.1.32}$$

Using the same method as before definition $A_n = A_n^* + \alpha_n$ such that $\frac{dH}{dA_n^*} = 0$ we define A_n^* in terms of the modes of the Hamiltonian.

$$A_n^* = \frac{2}{L} \frac{\left[\frac{\bar{f}_n}{q_n} - \frac{2ky_o}{q_n^2} \right]}{[\kappa q_n^2 + k q_n^{-2}]} \tag{3.1.33}$$

In terms of the α_n, the Hamiltonian becomes

$$H = \frac{L}{4} \sum_{n=odd}^{N} [\kappa q_n^2 + k q_n^{-2}] \alpha_n^2 - \frac{1}{L} \sum_{n=odd}^{N} \frac{\left[\frac{\bar{f}_n}{q_n} - \frac{2ky_o}{q_n^2} \right]^2}{[\kappa q_n^2 + k q_n^{-2}]} - y_o \bar{f}_o + (\frac{1}{2} k y_o^2 - v_o) L \tag{3.1.34}$$

This equation shows that the quenched average will be straightforward because Hamiltonian breaks up as a term that contains the thermal stochastic variables α_n – namely the first terms - and terms that contain the quenched stochastic variables f_n without any cross-coupling. The partition sum is the product of a term that involves the thermal stochastic variables and a term that only involves the quenched variables. There is no need to perform a quenched average of the logarithm of a partition sum that depends on the quenched variables, which in other cases leads to the mathematical complications of the replica method.

Minimize the Hamiltonian with respect to the lateral displacement y_o. The minimum energy lateral displacement is y_o^* with

$$y_o^* = \frac{1}{Lkg(z)} \left[\bar{f}_o - \frac{4}{L} \sum_{n=odd}^{N} \frac{k \bar{f}_n}{q_n} \frac{1}{\kappa q_n^4 + k} \right] \tag{3.1.35}$$

The Hamiltonian can now be expressed as

$$\begin{aligned} H =& \frac{L}{4}\sum_{n=odd}^{N}[\kappa q_n^2 + k q_n^{-2}]\alpha_n^2 - v_o L + \frac{L}{2}kg(z)(y_o - y_o^*)^2 \\ & - \frac{1}{2Lkg(z)}\left[\bar{f}_o - \frac{4}{L}\sum_{n=odd}^{N}\frac{k\bar{f}_n}{q_n}\frac{1}{\kappa q_n^4 + k}\right]^2 - \frac{1}{L}\sum_{n=odd}^{N}\frac{\bar{f}_n{}^2}{[\kappa q_n^4 + k]} \end{aligned} \tag{3.1.36}$$

We first perform the partition sum over the α_n and over y_o. This reproduces the partition sum Z_o for the case that there are no specific interactions. The free energy is the free energy of the system in the absence of specific interactions plus the last two term in the Hamiltonian that do not depend on α_n and y_o. The specific interactions introduce an added term ΔF_s in the free energy given by

$$\Delta F_s = -\frac{1}{2Lkg(z)}\left[\bar{f}_o - \frac{4}{L}\sum_{n=odd}^{N}\frac{k\bar{f}_n}{q_n}\frac{1}{\kappa q_n^4 + k}\right]^2 - \frac{1}{L}\sum_{n=odd}^{N}\frac{\bar{f}_n{}^2}{[\kappa q_n^4 + k]} \tag{3.1.37}$$

that is negative definite. This term needs to be quench-averaged. The quenched average of term linear in f_n, and hence linear in $f(x)$, are zero. The quenched average of quadratic terms

$$\begin{aligned} \langle \bar{f}_n \bar{f}_m \rangle &= \int_0^L ds \int_0^L ds' \langle f(x_o + s) f(x_o + s') \rangle \sin q_n s \sin q_n s' \\ &= \Delta \int_0^L ds \int_0^L ds' \delta(s - s') \sin q_n s \sin q_m s' \\ &= \frac{L}{2}\Delta \delta_{n,m} \end{aligned} \tag{3.1.38}$$

The exception is f_0 in which case $\langle f_o^2 \rangle = \Delta L$. This gives

$$\langle \Delta F_s \rangle = -\frac{\Delta}{2kg(z)}\left[1 - \frac{8}{\pi^2}\sum_{n=odd}^{N}\frac{1}{n^2}\frac{1}{[zn^4+1]^2}\right] - \frac{\Delta}{2k}\sum_{n=odd}^{N}\frac{1}{[zn^4+1]} \tag{3.1.39}$$

Using the fact that $\frac{8}{\pi^2}\sum_{n=odd}^{N}\frac{1}{n^2} = 1$, this reduces to a form that is again negative definite

$$\langle \Delta F_s \rangle = -\frac{2\Delta}{\pi^2 kg(z)}\left[\sum_{n=odd}^{N}\frac{zn^2(2+zn^4)}{[zn^4+1]^2}\right] - \frac{\Delta}{2k}\sum_{n=odd}^{N}\frac{1}{[zn^4+1]} \tag{3.1.40}$$

Note that $\langle \Delta F_s \rangle$ has the form of the natural scale Δ/k times a function of z. The dependence on the scaling variable z can be investigated in the same manner as before. In the regime

of small z, or large L, we replace summation by integration. First set $n = 2m - 1$ with m running over the positive integers. Define $x = z^{1/4}(2m + 1)$ and $dx = 2z^{1/4}dm$ with $dm = 1$. Then

$$\langle \Delta F_s \rangle \simeq -\frac{\Delta}{k}\left(\frac{z^{1/4}}{4\pi^2 g(z)}\int_0^\infty \frac{x^2(x^4+2)}{[x^4+1]^2}dx + \frac{1}{4z^{1/4}}\int_0^\infty \frac{1}{[x^4+1]}dx\right) \tag{3.1.41}$$

both integrals converge. In the limit of small z, $g(z)$ is proportional to $z^{1/4}$ so the first term inside the brackets goes to a constant but the second term diverges. In terms of the Odijk Length, we can express this result as

$$\langle \Delta F_s \rangle \simeq -\frac{\Delta}{k}(f_1 + f_2(L/\xi)) \tag{3.1.42}$$

with f_1 and f_2 numerical constants of the order of one. Adding the contribution to the free energy coming from the non-specific interactions, and including only the extensive terms proportional to L, gives for the quenched-averaged binding energy

$$\langle \Delta F \rangle \approx \left[\frac{2}{\pi}k_b T \ln\left(\frac{L}{d_2\pi\xi}\right) - f_2\frac{\Delta}{k}\right]\left(\frac{L}{\xi}\right) - v_o L \tag{3.1.43}$$

Increased flexibility means reducing the Odijk length ξ. If the coefficient inside the square brackets is positive, then this reduces the magnitude of the binding free energy. Note here that if the Odijk length is changed by an order of magnitude then the logarithmic changes by only a modest amount while L/ξ changes by an order of magnitude. If the coefficient is negative increasing the flexibility increases the magnitude of the binding free energy. The model leads to a simple result: if the thermal energy scale $k_B T$ is large compared to the energy scale Δ/k of the specific interactions then increasing the flexibility promotes unbinding. If $k_B T$ is small compared to the energy scale Δ/k then increasing the flexibility promotes binding. Formally, thermal unbinding wins in the large L "wins out" but this somewhat academic.

3.1.3 Calculating Odijk lengths from simulation

To crudely calculate the Odijk length from simulation we can combine experimental results for persistence length [41, 26]with the calculated PMFs to approximate the curvature of

the binding potential k. Here we will assume the binding potential is similar to a parabolic potential and use a least squares method to fit the curve and calculate k. Doing so we obtain the following results.

RNA	Odijk Length λ (nm)
polyU	0.20
polyA	0.31
AU Duplex	0.7

Table 3.1: The Odijk length calculated from experimental results for κ and results from simulation for k

We see that as the polymer stiffness increases, so does the λ. However, interpreting this in the context of the various forms of the binding energy we arrived at in equations 32, 29 and 25, it follows the opposite trend of binding energies we see in simulation. This suggests that binding energy should always favor the stiffer polymer because temperature effects destabilize the bound state for the more flexible the polymer.

CHAPTER 4

Modelling MD simulations

In the past 20 years single molecule pulling experiments have become standard for understanding the nature and function of biomolecules. Interpretation of the data of the force extension curves(FECs) and what is known and what can be determined from these experiments is still being studied[1]. Work has been done to build on previous theoretical work exploring FECs that exhibit "force rips" or regions of negative slope in the FEC, explaining the behavior through the bistable nature of different protein domains[42]. This has been attributed to the temperature and force dependent melting temperature of the protein. Once the critical temperature or force is reached, the protein will denature and give rise to these force rips[43]. Using a simple double well model, Bonilla et al were able to reproduce experimental behavior of the FECs for modular proteins and DNA hairpins. This is of particular interest to us do to the nature of RNA-protein binding involving partners that exhibit these interesting behaviors.To understand the response of proteins to mechanical force studied experimentally and theoretically[2, 44], we have attempted to connect the microscopic behavior of a small protein segment under force to the macroscopic behavior of folded proteins and duplex forming RNA/DNA in response to constant force, constrained length and an applied AC force. We are exploring the possibility that the mechanical response of the protein is predominantly determined by the hydrogen bonding dynamics within the protein. We have performed MD simulations to generate force extension curves of a small alpha helical protein (14 alanine residues) and have shown that there exist regions of the force extension curves with a negative slope as shown in the experiments to produce figure 4.2.6. In figure 4.1.3 we see the same behavior in MD simulations for a simple alpha helical protein. We chose an alpha helix as the model system because of its simple "1D" structure and its ability to

capture the important features of the force pulling dynamics of a much larger protein. We also see non-exponential relaxation of the hydrogen bonding correlations within the protein. We choose alanine for its high propensity to form alpha helices, while also allowing for non-canonical alpha helical hydrogen bonds to stabilize highly deformed structures.

A correlation function developed in the works of Chandler [45]was used to explore the hydrogen bonding dynamics. Using a model system with a potential that under certain tensions allows for bistable configurations of the monomers, we have recreated equilibrium FECs similar to those calculated from MD simulations and dynamic FECs seen in experiment. We see that the equilibrium distribution of the length of the chain is a many peaked distribution with the peaks corresponding to the stable solutions to the force balance equation for the zero temperature case, or the minimum of the Gibbs/Helmholtz potential in the force/length controlled cases. The initial response to the force and thermal exchange between the metastable states happen on different timescales and give rise to non-exponential relaxation in the correlations of the extension. Then by exploring the compliance of the response of the system under and AC force we are able to replicate the dynamical phase transition seen in the experimental works of [46].

4.1 MD simulations reveal complex hyrogen bonding dynamics of a small peptide under force

4.1.0.1 MD simulation details

The first tool we used to explore the system are all-atom molecular dynamics simulations. The forcefield parameters and simulation's set up is similar to the chapter two simulations and were also performed in the GROMACS simulation software [30]. The AMBER99SB forcefield was used and a particle mesh ewald (PME) method was used to compute electrostatics. A TIP3P representation of water molecules was used with a Noose-Hoover thermostat. simulations were carried out at 310K and a NaCl concentration of 0.1 mM (physiological conditions and enough to ensure system neutrality) with periodic boundary conditions in a

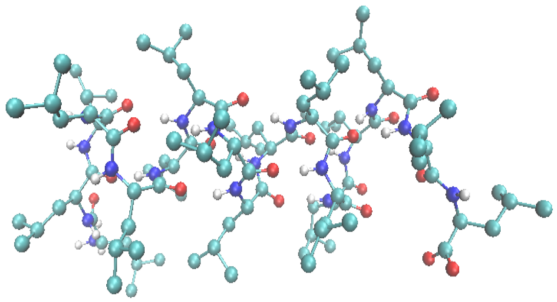
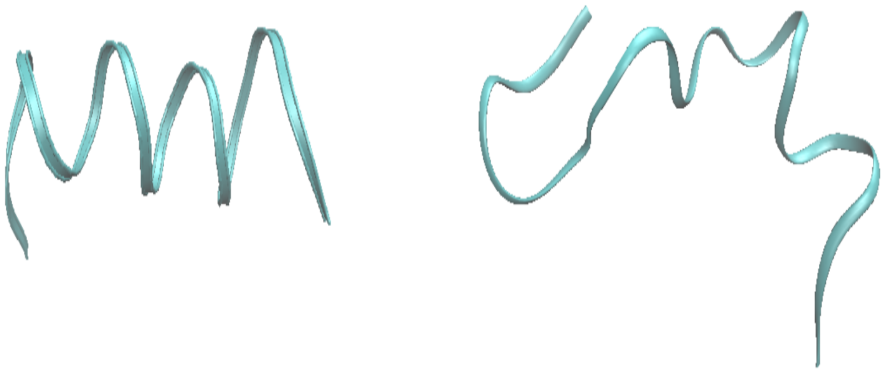


Figure 4.1.1: Image of the alanine residue chain. The teal colored sections are the alanines, the red is the oxygens and the blue the nitrogens. The pinning and pulling potentials are applied to the fourth and eleventh residues in the chain on the interior of the alpha-helix.



(a) (b)

Figure 4.1.2: Visualizations of the alpha helical state (a) and the extended state(b) after being pull in a ribbon representations. The alpha helix has been destroyed in the extended state.

rectangular box that was 6x6x12 nm. Simulation windows for both constant force and constant length simulations were run for 20 to 40ns. To conduct constant force simulations, umbrella sampling like in chapter 2 was used. A pinning potential of the form

$$V(x) = \frac{1}{2} k_{pin} \left(X_{COM}(x) - X_{pin} \right)^2 \tag{4.1.1}$$

was used. Where k_{pin} is the harmonic potential stiffness, X_{COM} is the instantaneous center-of-mass of the reside and X_{pin} is the location of the minima of the potential. It is important here not to use a potential too strong, as restricting the motion so much will lead to inaccuracies in the measurements of the force on this potential. A potential stiffness of 1200Kcal/nm^2 was sufficeint and used. A chain of 20 alanine resides was used and the protonation states set in accordance with physiological pH. Alanine was chosen due to its propensity to form an alpha helix[47]. Alpha helixes are primarily held together by intramolecular hydrogen bond interactions, making this an ideal system to investigate hydrogen bonding dynamics. The alanine chain was held at a fixed distance and a force distribution at each distance was calculated. To perform constant force simulations, one peptide at an end of the peptide chain is held in place by a strong harmonic potential on its center of mass, and the center of mass of the peptide on the opposite side of the chain is pulled at a constant force and the total end to end length of the chain is measured. For both the constant force and constant length simulations, the 4th and 11th peptides were chosen as the residues to be pinned or pulled. This was done because because in equilibrium the ends of the alanine chain were denatured due to thermal fluctuations and allowed for deformation enough of the chain for residues not under force to interact with part of the peptide that was of interest.

4.1.0.2 Force extension curves for small protein

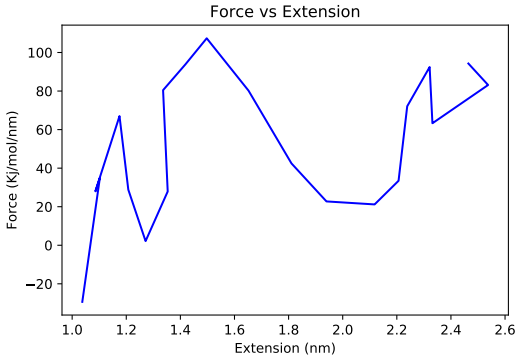


Figure 4.1.3: FEC of stretching a single 20 alanine residue alpha helix, by the 4th and 11th residues, in MD simulations and holding at constant lengths while measuring the force.

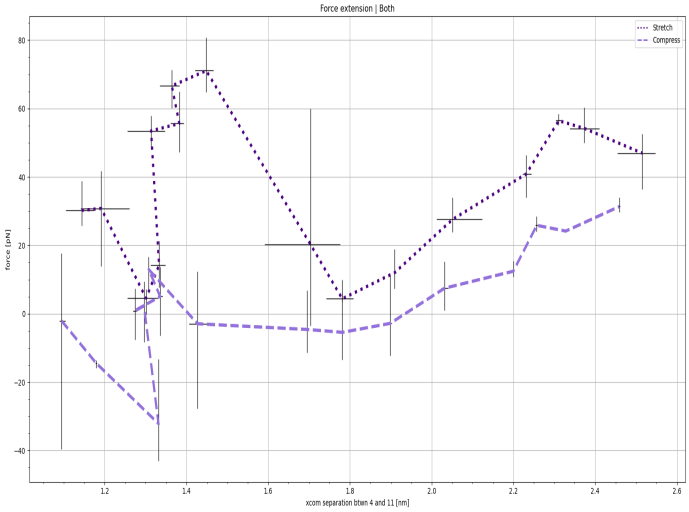


Figure 4.1.4: FEC of stretching (black dashed curve) and compressing(purple dashed curve) a single 20 alanine residue alpha helix by the 4th and 11th residues. Generated from MD simulations holding at constrained lengths while measuring the force. There is variation in force due to the properties of the protein, and variation in the length due to the harmonic potential holding the 14th residue's center of mass allowing for small fluctuations.

Turning first to the MD simulations, we see in a realistic system, in a sample as small as one short peptide, there exists regions where the protein exhibits behaviors similar to what is shown in experiment. Without looking at an ensemble measurement [44], there still exists regions of the constant length simulation FECs with regions of negative slope in figure 4.1.3. This simple system helps illustrate that the dynamical response of proteins seen in experiments are largely driven by the hydrogen bonds that hold the secondary and tertiary structures of the protein together.

Though it is often thought that proteins have a static structure, and while that may be the case for many proteins, particularly globular proteins, it is also true that these biopolymers are indeed polymers that exist in an ensemble of conformations. IDPs and proteins with less structure exhibit even more "ensemble" character since the hydrogen bonding pairs are not as well defined. Looking at the constant length FEC (or force extension plot) from the MD

simulations this becomes even more apparent. Looking at figure **??** we do see not the usual force rips or regions of negative slopes shown in the previous FEC's shown. Instead we see that varying forces can have similar lengths measured. This is counter intuitive because we would imagine that as the force is increased, the length should match, and in a purely elastic system it does. However, in a non-linearly elastic system it is possible.

Due to thermal fluctuations in the system, it is possible for a smaller force to push a particle out of a potential well, and it is also possible for a larger force to not escape the well. Going back to the work of Evans and Ritchie [48], given an infinite amount of time, a bond being pulled on by an external force will eventually escape the well. What this implies for the constant length FECs we see in MD simulations, which are simulated at a very short timescales, is that the bond breaking event is dependent on both the magnitude of the pulling force and the length of the simulations.

Another thing worth noting to further validate the results seen in the MD simulations is the reproduction of the existence of irreversible deformation. In the subsequent simulation and theoretical framework and the work done by Fogle et al [2], there is always the possibility of the system to return to elastic behavior upon lowering the applied force. In the work of Wang [44] the observe the existence of a regime where the protein loses its linear elasticity at all forces and frequency, and it is well known that upon a certain level of deformation, proteins can become kinetically trapped and unable to find their thermodynamically stable state again, leading to irreversible deformation [49]. What we have also found looking at the MD FECs, is that the viscoelastic behavior for the system, even when irreversibly deformed, can still persist. This is largely in part of non-canonical hydrogren bonds forming. Here, non-canonical hydrogen bonding refers to hydrogen bonds that occur that are not in the natural alpha helical state. Looking at figure 4.1.6, we see that the every four residue hydrogen bond normally seen in alpha helices are intact for the first fifteen nanoseconds. Then with the breaking of those bonds, irregular hydrogen bonds appear. Within the length of our simulations, it is highly unlikely that these canonical hydrogen bonds will reform. These irregular hydrogen bonds can also help in keeping the viscoelastic character of the protein.

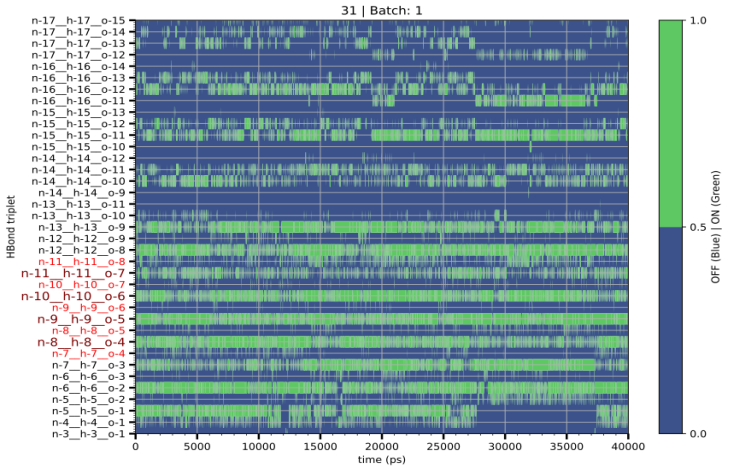
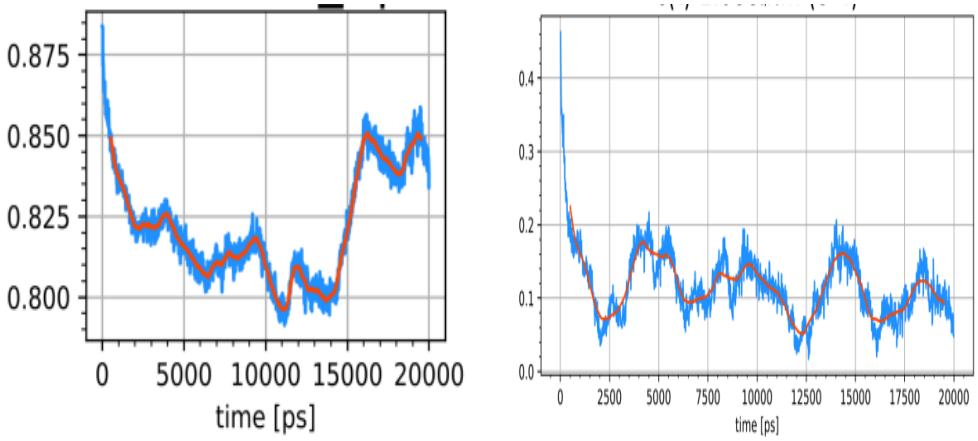


Figure 4.1.5: Contact map of hydrogen bonds within the alanine alpha helix. The numbers refer to residue number. Blue squares refer to a hydrogen bond not in-tact and green refers to an intact hydrogen bond. Shown is a simulations corresponding to one similar to figure 4.1.2(a)

Figure 4.1.6: Contact map of hydrogen bonds within the alanine alpha helix. The numbers refer to residue number. Blue squares refer to a hydrogen bond not in-tact and green refers to an intact hydrogen bond. This is a time trace of a simulation stretched to a distance. The red marked labels on the y-axis are hydrogen bonds comprised of resides that are only in the pulled part of the chain.

We see in figure 4.1.4 evidence that smaller extensions are needed to see the force rips and that smaller forces are measured throughout the different distant simulations. When compressing the peptide, it is forced to bend and twist in irregular ways that kinetically trap it in these misfolded states. However, what keeps the peptide in kinetically trapped is what also allows the peptide to retain its viscoelastic characteristic. Being forced into misfolded states brings regions of the peptide backbone that are usually separated in its native state close together and allows them to form a hydrogen bond. This would be advantageous for disordered proteins to bind in a variety of environments like those seen in cell signalling pathways. The plots in figure 4.1.7 are calculated by defining hydrogen bonds being intact or



(a) (b)

Figure 4.1.7: Plots of the correlation functions, equation 4.1.0.2,for a folded state(a) and a stretched state(a). In both cases $c(t)$ does not decay to zero and show long time correlations.

broken by a geometric criterion of donor-hydrogen-acceptor pair and and hydrogen-acceptor distance. These distances and angle are set at 2.5 angstrom and 150 degrees respectively. For the analysis of hydrogen bond correlations, a correlation function similar to Luzar's[45] was used.

$$c(t) = \langle h(0)h(t)\rangle / \langle h\rangle \tag{4.1.2}$$

Where h has a value 1 when a hydrogen bond is intact and 0 when it is not intact. In a perfectly diffuse system, these correlations would decay exponentially to zero. In figure 4.1.7 evidence of long time correlations for hydrogen bonds in both elongated and folded states

are shown. It is reasonable to suspect that the protein under force exists in metastable state and is not truly in equlibrium on the timescale of these simulations.

4.2 Brownian dynamics simulations to model MD results

Similar to the work done by [42] in the appendix to their paper, we use an asymmetric potential that allows for bistable monomers when a force is applied. Instead of simulating the postions of indidvidual monomers and the interactions between them, we simulate the bonds themselves. Using the individual bonds as collective degrees of freedom where a bond is defined as

$$x_i = y_{i+1} - y_i \tag{4.2.1}$$

where x_i represents the bond between the i^{th} and $i+1$ monomer and y_i is the position of the i^{th} monomer. For the rest of the chapter, we will use this description of simulating the bonds. We use a lennard-jones(LJ) potential coupled to a harmonic spring. The justification for the combination of these potentials is that the hydrogen bonds in a protein create strong short-range interactions, represented by the LJ potential, and that hydrophobic interactions ,which help keep the general 3D shape of the protein and holds the chain together when the covalent bonds are pulled on as well, can be modelled as a soft harmonic spring. The cooperative nature of hydrogen bonding in proteins [50] requires the inclusion of a coupling potential to mimic cooperativity. We include a coupling harmonic potential that depends on the state of the neighboring bonds. The neighboring bonds are coupled in a way that there is a restoring force that holds the two bonds at similar lengths. If the neighboring bond is in a stretched state, then it will pull on the i^{th} bond until it has reached the same stretched state. Thus, as the polymer chain becomes stretched, it will become easier to further stretch the chain. The potential for the i^{th} bond is defined for the force and length

controlled simulations

$$\begin{aligned} U\left(\{x_i\}\right) = & \sum_{i=1}^{N} 4\epsilon((\frac{\sigma}{x_i})^{12} - (\frac{\sigma}{x_i})^{6}) \\ & + \frac{1}{2}k(x_i - \sigma 2^{\frac{1}{6}})^2 + \frac{1}{2}k_s((x_i - x_{i-1})^2 + (x_i - x_{i+1})^2) \\ & + V_{FC}(\{x_i\}) \vee V_{LC}(\{x_i\}) \end{aligned} \tag{4.2.2}$$

Where $\{x_i\}$ is the set of bonds describing the system, ϵ is the depth of the LJ, σ is the zero energy position of the LJ, k is the stiffness of the harmonic restoring force and k_s is the stiffness of the coupling potential to neighboring bonds. The minima of the harmonic potential and the LJ potentials are placed on top of each other to distinguish a change in state from intact to broken hydrogen bond once the system is placed under force. In the instance of of no force on the system, the minima of both the potentials lie on top of each other. When the system is pulled or held at constant length, however, the minima split and the individual bonds can exist in an extended or compressed state. The potential is then modified according to the thermodynamic nature of the system. For force controlled simulations, or ones operating in the Gibbs potential, the extra potential is

$$V_{FC}(\{x_i\}) = \sum_{i=1}^{N} -F(t)x_i \tag{4.2.3}$$

Where the time dependence of $F(t)$ is given by $F(t) = \pm\nu_{FC}t$ and ν_{FC} is the rate at which the force is increased. For length controlled simulations, or ones operating in the Helmholtz potential, we put a harmonic constraint on the total length of the chain, so the added potential becomes

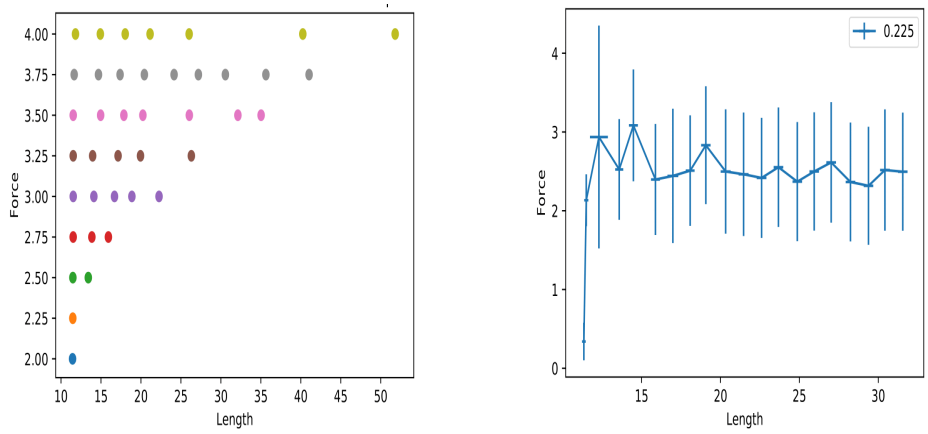
$$V_{LC}\left(\{x_i\}\right) = \frac{1}{2}K(\sum_{i} x_i - L(t))^2 \tag{4.2.4}$$

Where K is the stiffness of the constraint (or transducer in experimental settings) and $L(t)$ is the total length of the chain. Again, the time dependence for $L(t)$ is given by $L(t) = \pm\nu_{LC}t$ where ν_{LC} is the rate at which the length is increased. If the rates in either cases are equal to zero, then we observe the FEC corresponding to the equilibrium change for Gibbs/Helmholtz potentials for the force/length controlled simulations respectively. The

time dependence is only used when doing simulations of the dynamics of the polymer. For a time independent simulation we replace the potential in equation 4.2.3 with a constant force and the $L(t)$ in equation 4.2.4 with a constant length. The governing equations then for the langevin simulation, derived in the appendix, is then

$$\gamma \frac{dx_i}{dt} = -\nabla U\left(\{x_i\}\right) + \sqrt{2k_B T} f(t) \tag{4.2.5}$$

Where γ is the friction coefficient, k_B is the Boltzmann constant, T is the temperature and $f(t)$ is a random fluctuating force. In simulation, the Boltzmann constant has been set to unity along with the friction coefficient. Here we assume the system is heavily damped and thus we ignore the acceleration term. The random force is chosen such that $\langle f(0) f(t) \rangle = 2k_B T \delta(t)$ and the fluctuation dissipation theorem is obeyed. As shown in the appendix, the characteristic time and length are given by the physical constants in the langevin equation.



(a) (b)

Figure 4.2.1: FECs for the BD simulations. (a) plot of the different peaks for the length distributions for a given force. (b) Equilibrium force extension curve measured with harmonic potentials holding the chain at fixed lengths and measuring the force.

4.2.1 Equilibrium extension studies

We have used BD simulations to test the ideas of dynamic protein response in simulation beyond what we are capable of doing in MD simulations. Using the potential and formal-

ism described in the methods section we simulation a 1D polymer that is being stretched and compressed in accordance with the force. Doing equilibrium studies identical to the techniques used in the MD simulations we are able to recreate the same equilibrium measurements.

In figure 4.2.1 the similar FEC and scatter plots are seen in the BD simulations. The force extension curve in figure 4.2.1 (b) is also the same in the cases of stretching and compressing the polymer. The polymer was pulled and compressed at different temperatures as well to test the effects of temperature on the chain.

As the temperature gets larger we see a large increase in the variance of the length measurements. This behavior is similar to the gas-liquid phase coexistence regime seen in introductory thermodynamics courses[51]. In figure 4.2.2, we see the histograms of these length distributions in the constant force simulations. Each curve is pulled with the same force at a different temperature. The higher temperature distributions are many-peaked with each peak corresponding to roughly an integer multiple of the possible total lengths of the chain in accordance with the location of the minima of systems potentials. As the temperature increases, the polymer under force starts to melt and the individual monomers go into their extended states.

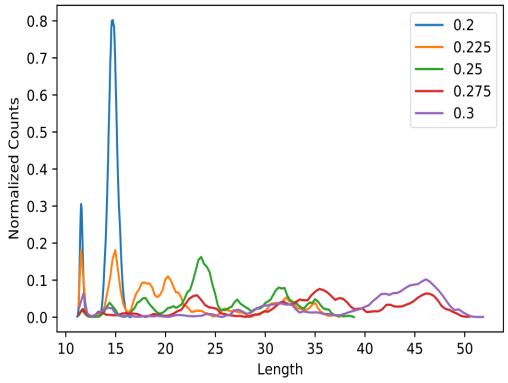


Figure 4.2.2: Histograms of the lengths for the constant force simulations and the peaks of the distribution for a given temperature. The variation of the constant force is shown on the inset legend. As the force increases distribution shifts towards larger lengths with peaks centered at integer lengths of the minima of the potential.

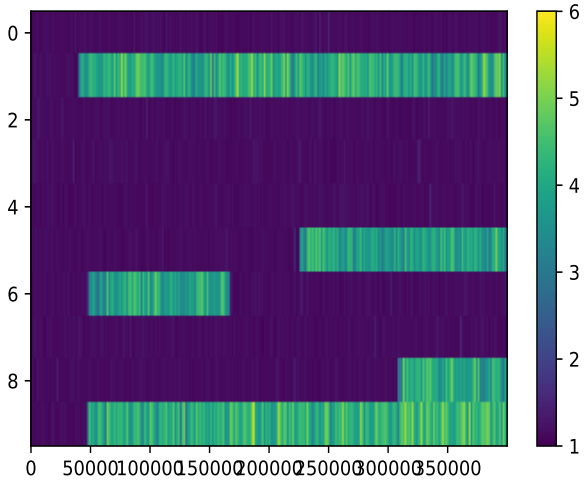


Figure 4.2.3: A time trace of all of the lengths of all monomers with spatial consistency in a constant length simulation at a fixed temperature and fixed length. The numbers on the y axis represent the monomer number and the numbers on the color bar represent the extension of that monomer. The x-axis shows the number of time steps.

Using the system, we can look at how the hydrogen bonds fluctuate and behave in individual force and length windows, similar to what was done for the MD simulations. In the MD simulations, there are expected hydrogen bonds between the nitrogen and oxygens along the peptide backbone around every four residues. In the BD simulations, the coupling is a bit simpler and is between neighboring monomers in the chain. The appearance of bonds that are not coupled to each other should appear as the "native" structure of the polymer is broken. In the BD simulation time traces, a native state will be a big chunk of blue in these plots. Looking at figure 4.2.3 we see that as the structure is force into a specific length, the appearance of single bonds breaking in isolation does occur.

These isolated bond breaks appear similarly to the non-canonical hydrogen bonds forming in the MD simulations. Due to the coupling, one would expect that if multiple bond breaks were to happen, then they would happen with neighboring monomers due to the decreased strength of the bond as its neighbor breaks. In actuality, it appears that when bonds break

they seems to be at separated but different parts on the chain. Since this is also not an exact two state system due to the ability for a monomer to extend or compress further than some well-defined lengths, is it possible to have different values of bonds breaking for a similar length. Thus, though the total length of the system is conserved, the extended state for an individual monomer may be larger. This is seen in figure 4.2.3 where throughout the time trace there are three, two then 4 extended monomers at a given total length. Though it is minimal, when the system is only at two extended monomer states, the extension of those monomers rises to their maximum value to compensate for the loss of length.

4.2.2 Dynamic force extension curves replicate experimental results

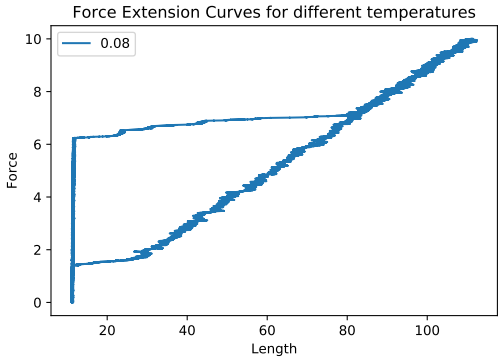


Figure 4.2.4: Force extension cycle. Here a maximum force of 10 dimensionless force units was used while pulling on the chain at a ramping force.

We first aim to recreate the experimental results with our model of the dynamic FECs seen in experiments. Completing a force cycle where the chain is dynamically stretched and then pulled back together we see force rips seen similar to those seen in figure 4.2.6. In figure 4.2.4 we pull and compress on the chain at a constant rate and measure the resultant length. Performing the simulation where instead the chain is held at a specified length that is changed at a constant rate, a more accurate representation of the traditional force extension experiments, the force rips are seen in figure 4.2.5.

For the case of a varying force on the system, we see the rupture force starts around six

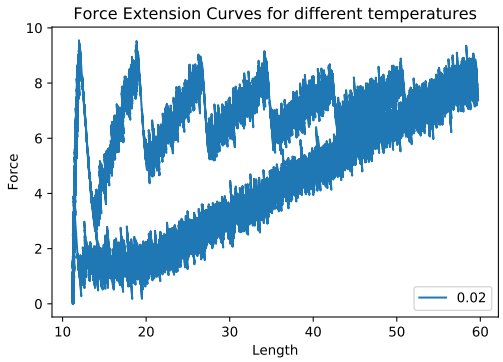


Figure 4.2.5: Force extension cycle where the length in the transducer potential is varied with time. Here a maximum length of the chain was used as the halfway point for the simulation and the force on the transducer is measured.

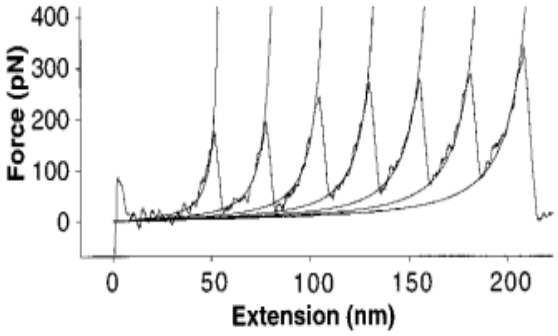


Figure 4.2.6: Force extension curves of stretching a single lg8 titin fragment with the dark black line with sawtooth pattern the experimental data and the thin black line modeled with the WLC model[1].

of the dimensionless force units. That is, once this force is reached, parts of the chain start to extend. If we increase the ramping rate of the force, the rupture force becomes larger better defined and more similar to a discontinuous transition. If the system is at a smaller force, but one where in the time spent at that force, is able to escape the bound state, it will slowly start to extend. If the force becomes larger faster, as is the case with the larger ramping rate, then before the monomers have a chance to escape the well at smaller forces, the force will become large enough to instantaneously break the bond. This is apparent in

figure 4.2.7 where when a dimensionless force of 9 is reached, the system immediately starts to extend.

In the previous figure, figure 4.2.4, we see that at a smaller force the system does start to extend, however, the system will stay at specific lengths until the monomers start to escape the bound state potential for a given period of time.

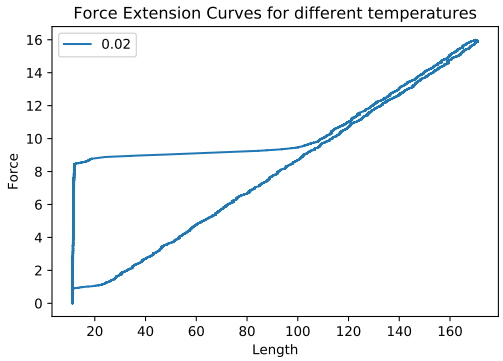


Figure 4.2.7: Force extension cycle. Here a maximum force of 17 dimensionless force units was used while pulling on the chain at a ramping force. The larger maximum force increases the ramping rate of the force increase.

4.3 Dynamics and statics of the response of protein in MD and BD simulations

In order to measure the response of the protein, similar to what was done in [2], a dynamic force was applied to the protein. That is the force was changed in accordance with

$$F(t) = F_o \sin\left(\frac{2\pi t}{\gamma}\right) \tag{4.3.1}$$

Where F_o is the maximum force that is applied to the polymer and γ is the pull rate of the transducer. However, in the case of MD simulations, the rate in which the the force is modified in experiment is not achievable in simulation due to the timescale of the period of oscillation.

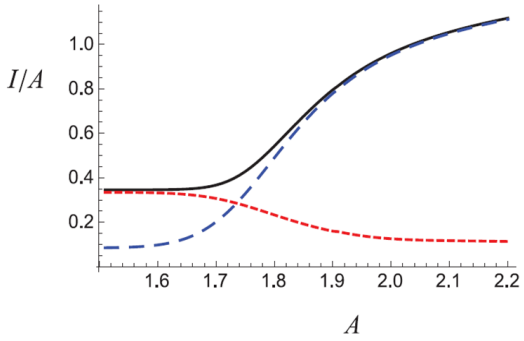


Figure 4.3.1: Example of a viscoelastic transtion looking at plotting the compliance vs force amplitude[2]. The black line is the total compliance, red short dash is elastic compliance and blue long dash is the dissipative compliance.

The experiments of Wang [44, 52, 46] show the existence of a dynamical phase transition in globular proteins. By changing the drive amplitude of the transducer in their experiments, they see a change in the dynamical response of the protein. They see that as the force in increased, the response of the protein goes from elastic to viscoelastic. Subsequent theoretical work by Fogle et al [2] showed that by reducing the system to a single collective variable representing the displacement of the protein with a simple potential could describe this viscoelastic transition as shown in figure 4.3.1. Using a potential which resulted in the restoring force of the form

$$F(x) = \begin{cases} 0 & |x| \geq x_o \\ -kx & |x| < x_o \end{cases}$$

they were able to construct a phase diagram showing how the frequency and amplitude of the oscillating force can define this transition, and how the dissipative and elastic components of the compliances can define this transition. In addition, they also showed how the presence of noise in a langevin formulation can turn the discontinuous transition in the a continuous one akin to going from a first order to second order transition in familiar thermodynamics, also shown in figure 4.3.1.

We explore using browning dynamics to find this transition using the model potential above. Without the explicit inclusion of a constant restoring force beyond some cutoff

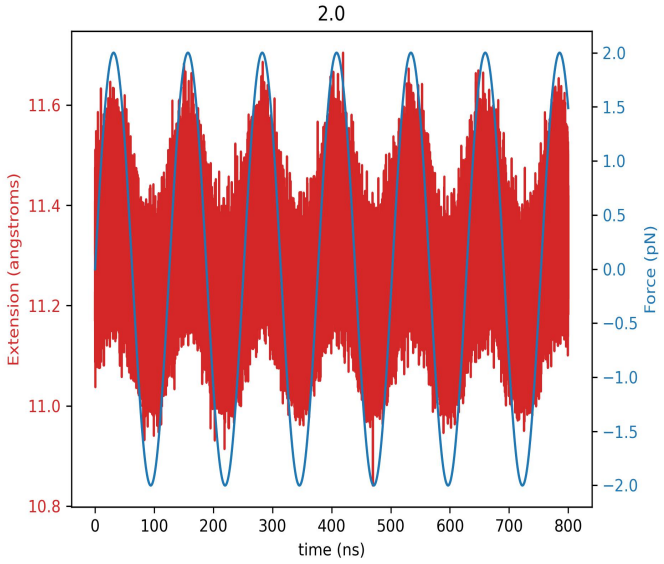


Figure 4.3.2: A time trace of a BD simulation where the title is the given forcing amplitude. Here a relatively small force amplitude is used. The blue is the drive force magnitude and the red is the displacement of the chain.

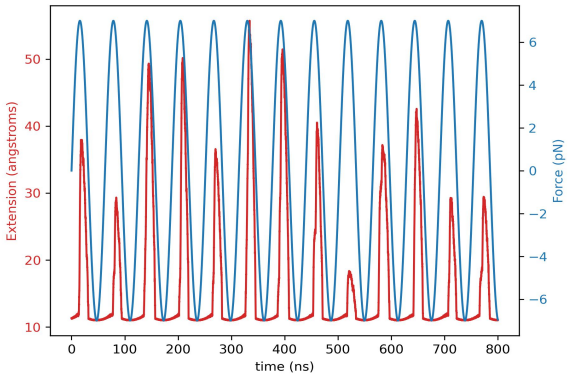


Figure 4.3.3: A time trace of a BD simulation where the title is the given forcing amplitude. A force amplitude in the transition region is used here.

distance, do you still see the emergence of viscoelastic behavior? Here, non-linearity comes from the existence of a double-well-like potential that exists from the combination of a LJ and harmonic potential that have some force applied to one end of the chain. This should resemble the system physically as there should not be infinite deformation allowed by the system. In the previous models in [44, 2], the system, once deformed, should be able to drift

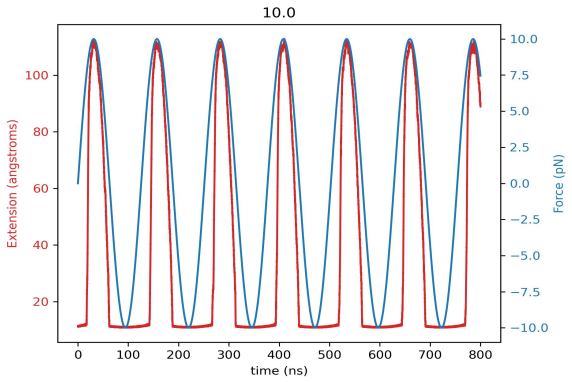


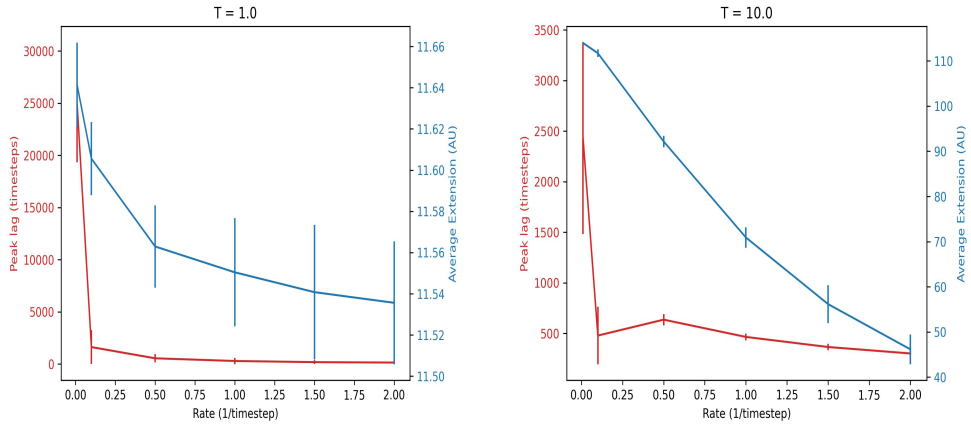
Figure 4.3.4: A time trace of a BD simulation where the title is the given forcing amplitude. A force amplitude in the large displacement region is shown here.

off to infinity. In reality, though the tertiary and secondary structure of the proteins are held together by intermolecular forces, after eventually breaking these interactions, you will pull on the strong covalent bonds that hold the amino acids together. Thus it is interesting to look at finding these viscoelastic regimes in displacement magnitudes where the system is not fully deformed.

Looking explicitly at the time traces of the driven system, it is apparent that emergent behavior occurs. In figure 4.3.2 we see that the system for a moderate forcing frequency and small force, behaviors entirely elastically and follows the drive pretty uniformly. For this small force region, the frequency is small enough that the system is able to adequately align itself with the drive and the force is small enough such that the event of switching from a bound to unbound state described in previous sections is not possible within the period of the driving force. As we increase the force, keeping the the frequency of the force constant, we see large extensions in the polymer that correspond to bond breaking. In figure 4.3.3 we see that though it only occurs once, the breaking of bonds from the driving force results in a large displacement from the minima of the LJ potentials. To look at observable quantities in this region, we look at averages over many periods. In figure 4.3.3, for example, there are slightly less than six periods of the driving force in the length of the simulation. To look at something like average extension, we would average the extension over all the periods and repeat the simulation multiple times. For slower drive frequencies, more simulations need to

be done to achieve the same number of period averaging. In figure 4.3.3, where the larger extension of the chain occurs, there is also a time lag of when the force starts to decrease and changes sign to when the bond breaks. This lag is due to the relative escape rate from hopping out of the potential well the monomer is in.

Once the drive the drive amplitude becomes large enough, we see time lag becomes smaller and breaking and forming of bonds happen twice within every period of the driving force. In terms of escape rate, the force becomes large enough that the force dependent rate also increases to where the bond breaks almost instantaneously. The systems response can be quantified in terms of the delay in the response (phase shift for a periodic one) and the maximum extension reached. Traditional signal processing techniques that look at the frequency dependence of the phase and magnitude of the response, such as bode plots [53], cannot be used here due to the very noisy nature of the signal from thermal fluctuations.



(a) (b)

Figure 4.3.5: (a) Delay in response and force ramping rate plotted with extension for a dimensionless force of 1.0. (b) Delay in response and force ramping rate plotted with extension for a dimensionless force of 10.0.

In figure 4.3.5 it is notable that as the rate is increased both the delay time and the maximum extension of the chain decrease with increasing rate. In the case of the small forcing regime, the magnitude of the delay time is also much larger than that of the larger

forcing regime. This is due to thermal fluctuations as the chain is slowly pulled by a relatively weak force. The monomers are in a relatively sharp and narrow LJ potential well and the longer time spent being pulled at one side of the well, the larger chance for a random force kick to push the monomer higher up the well. This can happen at relatively small deviations from the middle of the well and much sooner than when the force has reached its maximum value. It is also worth noting that as the frequency of the drive decreases, more simulation time is needed to create adequate sampling for those values.

In the large force regime, where the dimensionless force is greater than 8, it is also seen that the maximum extension of the chain is monotonically decreasing at higher frequencies. Looking at some of the linear force ramps such as in figures 4.2.4 and 4.2.5, once all of the monomers have been pulled out of the LJ potential, there is the familiar force-length relationship for linear response. That is, as the force is increased, the length is increased in a one to one ratio ($F = kx$). Thus, once the frequency becomes large enough where not all monomers are being pulled out of the LJ potentials in the time of half a period of the drive, then the system does not have adequate time to reach the equilibrium configuration for that given force. In other words, the drive becomes so fast that it exceeds or is comparable to the escape rate for monomers from the LJ potential.

4.3.1 Dynamical phase transition can be seen in the compliance of the response

The transition can be quantified further by looking at the compliance of the chain in response to the driving force under different conditions. For a system obeying linear response, we can separate the signal into its storage and dissipative components such that

$$x(t) = x_{el} + x_{diss} \tag{4.3.2}$$

where the subscripts notate their respective components. In the language of linear mechanical deformation, a system driven by a force $A\sin(\omega t)$, this can be written as

$$x(t) = j_1 A cos(\omega t) - j_2 A \sin(\omega t)$$

where j_1 is defined as the storage compliance, or the compliance due to elastic response, and j_2 is the loss compliance for the dissipative response. Due to the non-linear nature of their governing equations, extracting the storage and loss compliances was more challenging and also applicable to our simulations. To extract the respective components we adopt the method and notation of Fogle et al [2]. We define the total compliance for a non-linear system as

$$\frac{I_t^2}{A^2} = \frac{\omega}{2\pi} \int_0^{\frac{\omega}{2\pi}} x(t)^2 dt \tag{4.3.3}$$

Where $x(t)$ is the signal of our system and ω is the driving frequency. Integrating the signal over a period of the drive the portion of the signal that gives rise to dissipation can be extracted such that if Y_d is the dissipative part of the signal, then

$$x(t) = Y_d \cos(\omega t) + [x(t) - Y_d \cos(\omega t)] \equiv x_{diss} + x_{el} \tag{4.3.4}$$

where

$$Y_d = \frac{\omega}{\pi} \int_0^{\frac{\omega}{2\pi}} x(t) \cos(\omega t)\, dt$$

then the relationship between the total, storage and dissipative compliances are

$$I_t^2 - I_d^2 \equiv I_s^2 \tag{4.3.5}$$

where

$$I_d^2 \equiv \frac{Y_d^2}{2} \tag{4.3.6}$$

Thus showing that in terms of the coefficients of equation 4.3.1, $j_1 = \frac{I_s}{A}$ and $j_2 = \frac{I_d}{A}$ Using our simulations traces for $x(t)$ we can then extract the compliance components of our signal. They note in their paper that for when thermal noise is present $x(t)$ should be replaced by $\langle x(t) \rangle = [x(t) - x\left(t + \frac{\pi}{\omega}\right)]$, however in our situation we do this by averaging the signal over many periods.

In figure 4.3.6 the mechanical response of the system in its elastic, dissipative and total compliance. There a number of interesting features of the response of the system. First, we see the non linear response of the system after a dimensionless amplitude of six is reached. This is in the same regime as where the mechanical denaturing of the system happens in the

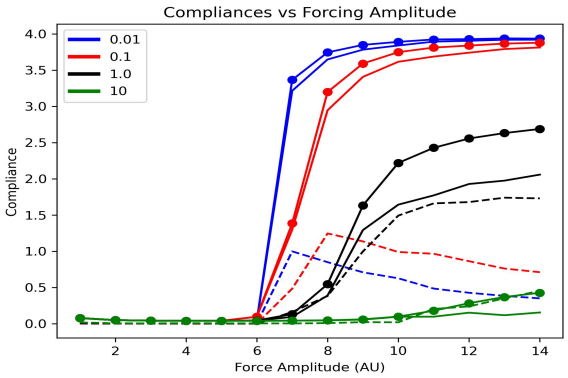


Figure 4.3.6: A plot of compliances for different simulations for varying drive amplitude A and different drive frequencies. The drive frequencies are given by the different colors and the total, dissipative and storage compliances are given for each frequency for the connected dot, straight line and dashed line respectively.

equilibrium simulations. However, in this regime we also see a dissipative component of the compliance grows as well, differing from the theoretical models shown in [2] where they were inversely correlated. There also is a decrease in the dissipative component of the compliance as the amplitude increases from the the thermodynanmic transition amplitude.

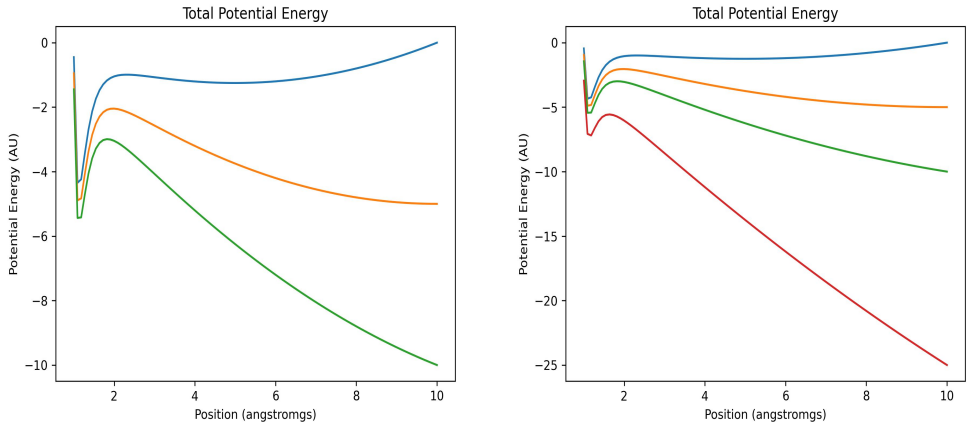
We also see that as the as the drive frequency increases the thermodynamic transition becomes shifted towards larger forcing amplitudes. As the frequency becomes large, the time spent what we will call the transition state, or pinned to a point in the LJ potential where escape is possible from a thermal fluctuation, becomes smaller. Thus, for the same forcing amplitude we see a decrease in the time spent near or at the transition state at higher frequencies. A larger force is needed to compensate and tilt the monomer potential to allow for quicker escape of the particle. As mentioned many times before, the escape rate depends on the force applied to the monomer and is shown in the dynamical response of the chain of monomers by shifting the increase of the total compliance to larger forcing amplitudes. Along with the shifting of the melting transition, the transition goes from one that is relatively abrupt to one that is more gradual. Something like a first order transition changing to a second order transition. In previous theoretical work, it was seen that adding noise to the system smoothed the transition out. Here we already have noise included with

the nature of the BD simulations so the smoothing of this transition could not be attributed to the inclusion of thermal fluctuations.

In previous works used a single collective variable to describe the response of the system. Here we have many coupled non-linear oscillators where the addition of their lengths gives the total response of the system. There is statistical variation in the number of monomers in the bound or extended state. It is possible smoothing of the transition occurs due to the variation in the number of monomers that make the transition. The longer time spent at the transition state, the more monomers that will make the transition. Given enough time near the transition state all of the monomers will make the transition whereas for shorter times only a few monomers will make that transition.

It is also noticeable that in the low to moderate frequency regime, there are both dissipative and elastic components to the total compliance. Though the major component is still from the elastic component the dissipative part is not negligible and in the moderate frequency regime equivalent to the storage compliance. We have found that this is attributed again to the nature of having many non-linear coupled oscillators. In the thermodynamic transition region and at lower frequencies, most if not all monomers make the transition to the extended state. The force ramping happens gradually enough that most of the total monomer potential is not extremely tilted towards the extended state and still has adequate time to escape the bound state. In this force regime the potential still has double well characteristics and a linear-like force regime between the two wells. The oscillating force when stretching and compressing the chain is then "dragging" the chain along this region between the two states. This drag leads to a dissipative component of the compliance. At larger forces and smaller drive frequencies this drag becomes smaller as the transition region between the two wells becomes more harmonic.

Figure 4.3.7 shows the transition region after the bound state becomes more linear as the tension on the chain gets larger. Once the critical force has been reached and the chain starts to melt, the monomers move along this transition region in response to the drive and encounter a type of drag force. This drag force then gives rise to the dissipation seen in the



(a) (b)

Figure 4.3.7: Plots of the potential energy curves for the individual monomers of the system. (a) Examples of a small pulling force on the chain. The different colors represent the surface under different tensions. (b) The same as the first plot except with a large tension curve in red. This is to illustrate how the potential energy curve becomes linear in the transition region of the potential.

system.

As the forcing frequency gets higher the dissipative component of the compliance in this transition region becomes the major contributor of the total compliance. This would be the transition from elastic to viscous material seen in the experiments and theory. In this frequency regime, if the amplitude of the drive were to increase, the total compliance is still mostly comprised of the dissipative compliance.

In figure **??** we see a phase diagram of the response of the chain. Taking the ratio of the dissipative compliance to the total compliance we describe how elastic or dissipative the system is responding. Based on figures **??** and 4.3.6, we divide the plot into 3 regions describing the dynamic response of the chain. In region I we have what we see in the low frequency low and forcing amplitude regime, and the large force low frequency region. Here the response of the chain is purely elastic. In region two we have the moderate force force amplitude and low to moderate drive frequency. In this regime the dissipative compliance of

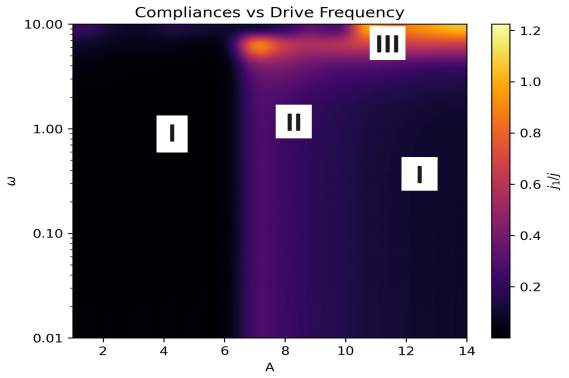


Figure 4.3.8: A phase diagram where the color indicated the fraction of the total compliance that is due to the dissipative compliance. The frequency is plotted on a logarithmic scale. Each labeled region represents a regime discussed in the text.

the chain is non-negligible and up to comparable to the elastic compliance, exhibiting both dissipative and elastic behavior. Here the forcing amplitude corresponds to a coexistence region of the stretching of the chain located in the force rips in figure 4.2.5. The chain has coexistence in the thermodynamic and dynamic phases. This could be potentially insightful into probing what conditions are suitable for molecular conformational changes in biological processes since different motions occur at different length and timescales [54]. If for a specific enzyme, for example, you could determine the thermodynamic mechanical properties of the protein, you could use that to extract information about its dynamical response to stimuli or the timescales and length scales on which its natural allosteric motions occur. The third regime in figure **??** is the viscous regime of the chain. Here the compliance almost completely dissipative and exists only in the very large frequency regime.

In the previous theoretical work [2], region II was described to be part of the smooth transition from an elastic material to a viscous one. We have chosen to label it as its own region in the phase diagram due to the possibility of re-entrance from this region to region I where the response is purely elastic. In the context of protein deformation, this region may correspond to one where there is viscous deformation of the material but it is still reversible. Then region III would correspond to irreversible deformation. Here there is no such irreversible deformation because a restoring force can always force the monomers back

into the bound state. Perhaps the addition of another dimension, though complicating the potential surfaces involved, would allow for irreversible deformation. In the following chapter we discuss a model for a structured polymer binding that could be used to show how in a similar situation, spatial ordering can lead to conformations where the native state of the chain is not recoverable upon deformation.

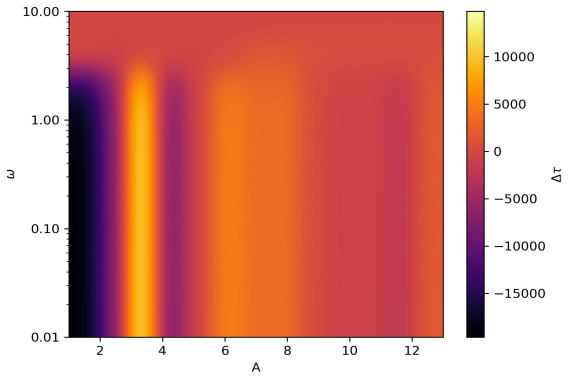
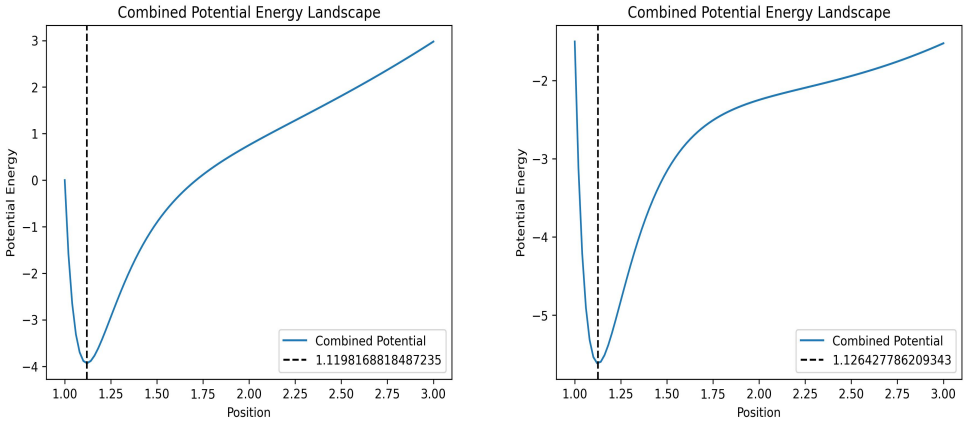


Figure 4.3.9: A phase diagram where the color indicates the size of the lagtime ($\Delta\tau$ in number of timesteps), as a function of the drive amplitude (A) and the drive frequency (ω).

Up to this point the type of response we have discussed is in the magnitude of the response and the compliance of the chain. In figure 4.3.9 we show the lagtime, or the phase shift, as a function of the different drive frequencies and amplitudes. The lagtime is calculated as the difference between the time of the maximum of the response and the time of the maximum of the drive. Though this phase diagram may seem slightly counter-intuitive and confusing, it actually reveals some of the interesting behavior of our system. At low frequencies we would expect the delay to be small and positive since the ramping rate is slow enough for the response of the material to stay relatively close to the drive. Yet in the small force and small frequency part of the phase diagram there appears to be a large negative lag time that indicates a lead in the response of the material. Incidentally, when the force and frequency are small, the system will not initially respond to the drive and thermal noise dominates the extension as the deviation from the equilibrium state is smaller than the fluctuations in the extension. However, there is still a small shift in the average of the response. To find the origin in the shift of the average for these values of drive frequency and amplitude,

we look back at figure 4.3.7(b). Though it is very small, there is a shift in the minima of the LJ potential when the tension on the system is increased. Thus, the response of the system here is not due to the response of monomers being pulled to the edge of the potential well, but, in fact, is due to the minima of the potential being shifted. In figure 4.3.10 the



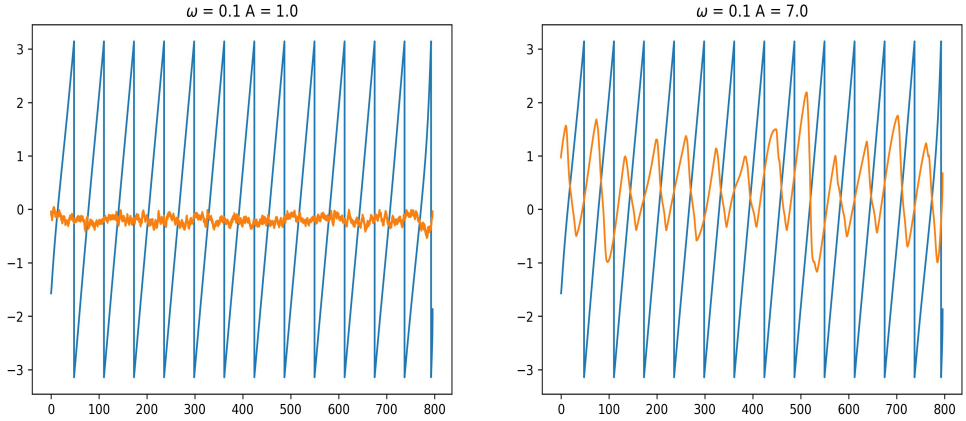
(a) (b)

Figure 4.3.10: Plots of the potential energy curves for the individual monomers of the system plotted with the minima of the LJ portion of the potentials. (a) Examples of a small pulling force on the chain with the vertical dashed line at the potential minima. (b) The same as the first plot except with a larger tension on the curve. The difference in tension is comparable to the discussion in the text.

potential energy surfaces with the relative difference in drive amplitude is plotted with the local minima of the bound state shown. The result is a shift to the right of the potential minima under force and of comparable magnitude of the change in average of the response of the chain. Deviations larger than this can therefore be attributed to direct tension on the monomer holding towards one side of the potential energy curve. There is also a notable deepening of the potential well as tension is applied to the monomer. Despite the changing of the potential energy curve, there still shows a lead in the response of the material that is still attributed to the large thermal fluctuations in position compared to the shift of the average, so the maximum extensions can be reached before the drive has reached its maximum value. This unexpected behavior of the phase of the response led us to investigate the nature of

the phase difference of the different areas of the phase diagram.

To investigate the phase behavior traces of the instantaneous phase difference were plotted for each of the time traces of the response. To get instantaneous information about the phase response of the chain a Hilbert transformation[55] was performed to plot the instantaneous phase of the drive and the response where generated. Then the difference between the instantaneous drive and response signal gives the instantaneous phase difference. A positive difference in phase indicates a lag or delay in the response and a negative phase difference indicates a lead in the response. In figure 4.3.11(b) the signal switches between lagging and



(a) (b)

Figure 4.3.11: Plots of the instantaneous phase difference (in orange) and instantaneous phase of the drive (in blue) generated from hilbert transformations of the data at the response for differing values of the drive frequency and amplitude which are shown in the title of the plots.(a) For moderate drive frequency and small driving amplitude. (b) For moderate drive frequency and moderate driving amplitude.

leading behind the drive. The early time behavior of leading the signal is coincidental as the the monomers start at slightly smaller distances than the potentials equilibrium position. Thus they move towards the equilibrium and in the direction of the drive force slightly faster than the ramping rime of the drive and gives the illusion that the response precedes the drive. In the higher frequency and force regimes the average shifts. Investigating the origin in the switching of sign of the phase response reveals that the same shifting of the potential that

leads to the lag in the small frequency and force regime is the same as the response attributed to the small force small frequency time lag. As the chain is full extended, the monomers are in the extended state where the minimum shifts farther right the larger the force is. As the force starts to be reduced, the potential shifts back towards the minima being at a smaller distance. The shift in the potential puts the monomer in an unfavorable state that then pushes it to smaller distances. If the rate of change of the potential curve is faster than the motion of the monomer, then the monomer will continue to drift towards the bound state even when the force has reached zero. Looking again at the potential curves in 4.3.10, at 0 force the potential to the right of the LJ portion is positive and restoring. Once a single monomer has entered the bound state, then its own potential and neighboring monomer's potentials will becomes stiffer due to the monomer-monomer coupling, accelerating the chain entering a compressed state. This is why the instantaneous phase difference switches sign once the force on the chain reaches zero. It is comparable to a snapping mechanism once one monomer enters the bound state.

In figure 4.3.11(a) the phase difference for the response of the material is essentially zero. Though it was not captured in the phase diagram for the lagtime, the hilbert transformations show that the response of the shifting of the potential minima follow closely to the drive at moderate frequencies. Though these plots are insightful, htey are mostly useful in these two regimes, or regimes where the elastic response dominates. The hilbert transform assumes local linearity in the response of the signal, even if the signal itself is non-linear, as is the case in the response of our chain, but in the transition regions discussed previously, and shown in the time traces of figure 4.3.3, the response can be largely non-linearly at high resolutions.

4.4 Using time varying rates alone can replicate system behavior

Maximum caliber (MaxCal) is a method of statistical inference derived from Shannon's information theory that produces distributions of dynamical trajectories near or far from equilibrium similar to how the principal of maximum entropy (MEP) produces population distributions at equilibrium[56]. The principle suggests that the path taken between two

points in space and time is such that the action, a quantity derived from the Lagrangian function, is minimized. The concept of MaxCal provides a framework for understanding the behavior of dynamic systems and systems far from equilibrium. In traditional statistical mechanics, the energy of a system is constrained and entropy maximized to understand information about the distributions of its constituent parts. In MaxCal, a dynamic quantity can be constrained while maximizing the shannon entropy [57] of the paths taken from one point in the state space to another to learn information about the distributions of trajectories of the dynamical process.

Here, we use MaxCal to constrain the transition rates between two states that represent the elongated states and the bound states of the individual bonds. The dynamical partition function of a single bond is related to the dynamical partition function for a series of coupled or uncoupled bonds similarly to how the grand canonical ensemble are related in classical statistical mechanics.

$$Q_1^M(t) = Q_M(t) \tag{4.4.1}$$

Where $Q_1(t)$ is the microcanonical dynamical partition function of a single bond and $Q_M(t)$ is the dynamical partition function for the entire system. This is then related to the probabilities of being in one state, say A or B, as

$$Q_M(t) = [P_A(t) + P_B(t)]^M \tag{4.4.2}$$

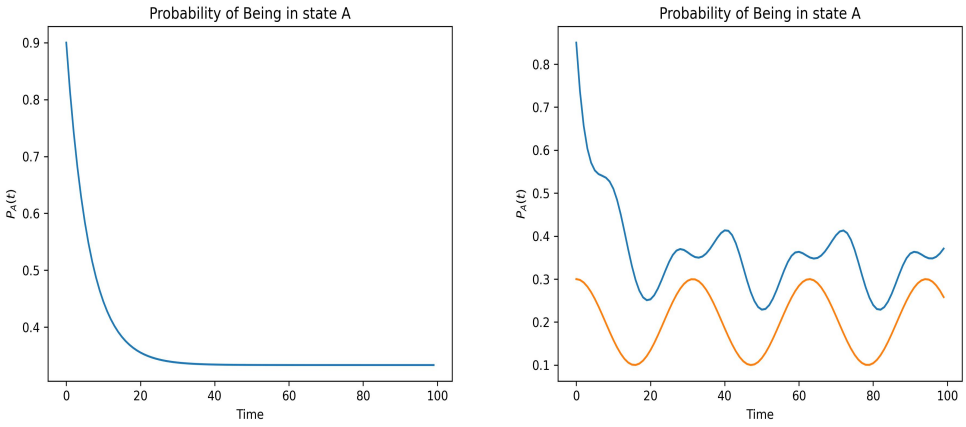
By taking an associated rate matrix as an operator to a probability vector, where the elements of the of the vector represent the probability of being in state A or B, we can accurately propagate the system forward in time such that the time constraints on the rates. If we have a rate matrix K such that

$$\begin{bmatrix} K_{AA} & K_{AB}(t) \\ K_{BA}(t) & K_{BB} \end{bmatrix}$$

where the transfer rates between states time dependence is given by

$$g_{ab}\left(1 + \left(\frac{1}{2}\right)\cos 4g_{ab}t\right) \tag{4.4.3}$$

with g_{ij} is average transfer rates. Doing this for all of the individual bonds in the system and summing them will give a distribution of the total length of the system.



(a) (b)

Figure 4.4.1: The probability of a bond being in a given state as a function of time for (a) a system with constant rates of exchange between the state and (b) for an oscillating rate. In (b) the orange line is the oscillations of the rate.

We see that the inclusion of a time varying rate, like those that are expected under dynamic loads on proteins, induces a response that is similar to the time traces we see in the previous section and the solutions to the equations of motion described in Fogle et al [2]. The same was done for a system of M particles and we can track the probability of n particles being in state A, for example, for a given time. which is shown in figure 4.4.3.

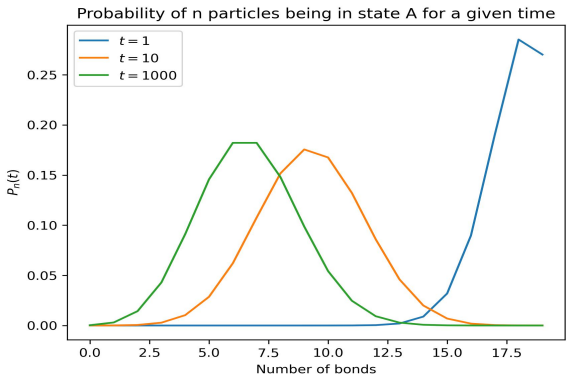


Figure 4.4.2: Plot of the probability of being in state A for three times, 1, 10 and 1000 time steps, for a constant rate matrix. The probability decreases to the steady state value as expected.

Applying this same procedure with the time dependent rates in equation 4.4, and looking at the probability of the total extension, we see the similar traces to those of the noisy system in Fogle et all [2]. We see the same time of time-delay for the extension that is in the model system as well.

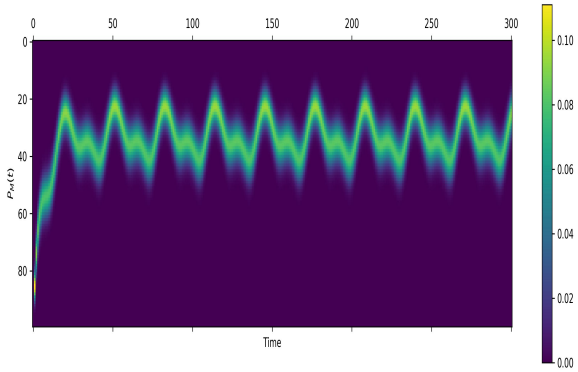


Figure 4.4.3: Time trace of the probability of the number of monomers being in state A as a function of time. The sum of the probability of being in state A and state B will give the length probability distribution of the entire chain.

4.5 Concluding remarks

Here we have shown using MD simulations a simple model for a series of collective variables that represent the bond lengths of hydrogen bonds in a short peptide can recreate both equilibrium and out of equilibrium results of deformable biopolymer dynamics. There is a mechanical induced melting transition from equilibrium simulations that is show in the force rips and variation of force measurements of the force extension curves. It is indicated that it is the complex hydrogen bonding dynamics seen in the MD simulations inside the chain that govern the mechanical response of the protein. The model system also shows long time correlations in these dynamics which supports its use.

The model system is then tested to reproduce out of equilibrium measurements where and AC force is applied to the chain to look for a dynamical phase transition represented as the change in compliance of the response of the chain. The model shows a transition at high

drive frequencies that corresponds to a transition from an elastic to a viscous solid. There is a region where the response carries comparable dissipative and elastic components to the compliance that we attribute to the region discussed in the work of Wang [46] of reversible deformation of the protein. We find that the force regime of the dynamic phase transition of region two of 4.3.8 happens at drive amplitudes that correspond to the equilibrium mechanical melting transition seen in the force rips of figure 4.2.5.

Lastly we briefly discuss if using a simple two state system with time varying transfer rates can be enough to capture the rich behavior of the system with no physical variables in the language of MaxCal. We have shown that this extremely simple model can capture the steady-state response of Fogle [2]. Further validation of the results need to be analyzed to determine if from this formalism more interesting characteristics of the system can be extracted from the results of simulation or experiments. It would be interesting to see if physical properties of biopolymers could be extracted from the results of dynamic biopolymer stretching experiments using a statistical inference method like MaxCal.

CHAPTER 5

Deformable polymers in relation to biomolecular binding

The binding of highly felxible and disordered proteins is important in many cellular signaling pathways and the nature of this flexibility and disorder is advantageous for their respective functions [58]. Insight into the underlying physics can be helpful in better understanding the driving factors for these signaling pathways and aid in the treatment of certain prion protein diseases such as Alzheimer's disease, mad cow disease and many more[21, 59, 60]. Understanding the nature of binding driven structural transitions is largely important in prion diseases where misfolding of a single protein can then induce the same conformational change in other proteins. In the previous chapter we used a non-linear model to simulate the behavior of deformable biopolymers under static and dynamic loads, and in chapter two we discussed MD simulations that showed how flexibility for short chain RNAs can help in forming stronger bonds when binding to a disordered protein. Here, we will investigate how the flexibility paired with the ability to form structure can affect the ability for biopolymers to bind. The term structure is used for interactions that model the secondary and tertiary structures of proteins and nucleic acids usually comprised of electrostatic, hydrogen bonding and hydrophobic interactions. Generally, like the work done in chapter three, flexibility is treated as the energy required to bend a polymer to the intrinsic stiffness of the chain. However, in RNAs and proteins, this classification of rigidity may not be sufficient to understand the interplay of flexibility and binding energy and mechanisms [26, 20].

The aim here will be to introduce a stiffness to the polymer in a highly non-linear way to accurately model the behavior seen in the experiments and simulations of previous chapters

and help understand the general type of binding seen in flexible and disordered systems. Previous work has been done in classifying the different binding modes for disordered and ordered systems. The work done by Fuxrieter [20] classifies these binding modes in disordered proteins as disorder-to-order, disordered and "fuzzy" binding. Where upon binding the IDP becomes ordered, stays disordered or a mixture of both respectively. The work here is especially relevant where both binding partners may be disordered. The emphasis on an interaction of an ensemble with variation in the dominant structures present provides a useful framework for how these structures interact, and will be used to describe the type of binding we see in the BD simulations. Since the simulations of chapter four show that they can accurately reproduce experimental data of the dynamics of proteins, it is a good starting point to build a model for looking at the binding between flexible partners. We showed in the previous chapter that the dynamical response of the system was closely related to the thermodynamic phase change seen in equilibrium simulations and use this as justification for applying the model to represent potentially highly out of equilibrium states seen in binding transitions.

5.1 Methods of simulation and model

Simple simulations were used to help understand what physical properties play a role in RNA-protein binding. The model system can be a bridge between the analytical theory and MD simulations in understanding the fundamental phenomena which are present in RNA-protein binding and the dependence of flexibility. The potentials used are coarse grained and designed to mimic only the most important parts of the real system. The model is largely adopted from chapter three with the addition of an extra dimension and here the individual monomers of the chain are simulated and we are not dealing with a set of collective variables. To model a deformable polymer we need at least two dimensions for them to interact. Instead of an explicit double well potential that we have in our one dimensional (and bonila et al's) simulations, we use a LJ potential for next nearest neighbors to create a structure of our choosing. We focus on a zig-zag like structure that is similar to a flattened alpha helix. On the

axis along the direction of the primary axis of the polymer there is a lennard-jones for short distances between next nearest neighbors and two harmonic springs between them, providing two "wells" along this dimension. The LJ interactions stable the secondary structure of the system and harmonic wells act as the linker between different units. Using this system we can recreate the same behavior we saw in the one dimensional system. For the interaction between the two chains we have used both a peacewise harmonic potential and a gaussian potential. The difference between both potentials in negligible so we use the gaussian for easier comparison to analytical theories. The potential for the entire system is the following;

$$\begin{aligned} U(\{r_i\}) = \sum_{i=1}^{N} 4\epsilon \left(\left(\frac{\sigma}{r_{i,i+1}} \right)^{12} - \left(\frac{\sigma}{r_{i,i+1}} \right)^{6} \right) + 4\epsilon \left(\left(\frac{\sigma}{r_{i,i-1}} \right)^{12} - \left(\frac{\sigma}{r_{i,i-1}} \right)^{6} \right) \\ + \frac{1}{2} k_h \left((r_i - r_{i-1})^2 + (r_i - r_{i+1})^2 \right) + U_{interaction}(\{r_i\}) \end{aligned} \tag{5.1.1}$$

Where $\{r_i\}$ is the set of distances, ϵ is the potential depth of the monomer-monomer interactions, σ is the distance of the minimum of the LJ potential, k_h is the stiffness of the neighboring harmonic bonds. Here the value of sigma is around one third of the length of the harmonic minima between monomers to keep the ratio of distances between the nearest neighbor interaction and the size of a single nucleic acid or amino acid realistic. $U_{interaction}(\{r_i\})$ is the potential term for the interaction between polymers which can be one of the following:

$$U_{hp}(\{y_i\}) = \sum_{i=1}^{N} \begin{cases} 0 & y_i > y_o \\ \frac{1}{2} k_b \left(y_i - y_o \right)^2 & y_i \leq y_o \end{cases}$$

Where r_o is the cutoff distance for the piecewise harmonic potential, y_o is the location of the potential axis and k_b is the stiffness of the interaction potential. We have the following for the gaussian potential simulations;

$$U_g(\{x_i\}) = \sum_{i=1}^{N} -D_e e^{\frac{-(y_i - y_o)^2}{2\sigma_g^2}}$$

Where D_e is the potential depth of the gaussian, y_o is the location of the minima axis and σ_g is the width of the gaussian potential. In these systems the width of the potentials give a physical size of the interacting rod. The minima here is the x-axis of the system and the

gaussian potential runs along its entirety to avoid edge effects from the stationary rod. The governing equation and update rule for the system is then

$$\gamma \frac{df}{dx_i} = -\nabla U(x_i, x_{i+1}, x_{i+1}) + \sqrt{2k_B T} f(t) \tag{5.1.2}$$

the same as in the previous chapter where γ is the friction coefficient, T is the temperature and k_B is the Boltzmann constant. We use dimensionless units following the procedure in appendix B. The temperature is set to around 310K in accordance with physiological conditions.

5.2 Characterization of the system

To monitor the change in state of the system we track the eccentricity, size, number of unbound monomers and spin number which will be discussed below. A combination of four observable best characterizes the system because different states may be "degenerate" in one but distinguishable in another. For example, a state may be have a small radius of gyration like a highly structured state, but instead is a disordered compact state. Therefore, to distinguish metastable states, multiple observables are needed. The eccentricity of the chain is used as a measure to determine how elongated the structure is. Both highly structured and completely denatured states can have an eccentricity close to one and a perfect random coil state will have an eccentricity of zero. The eccentricity is calculated by finding the eigenvalues of the moment of inertia tensor $\begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix}$ where the elements are the summed squared values of the x (X_i) and y (Y_i) components of the positions of the monomers in the chain, for example $I_{xy} = \sum_{i=1}^{N} X_i \times Y_i$ The eigen vectors of this matrix, λ_1 and λ_2, give the principle axises of rotation weighted by the mass of the monomers. Here all the monomers masses are equal so its an average of the square positions. The magnitude of these eigen values tell us the which axis dominates. The eccentricity is then related to the eigenvalues by

$$\xi = \sqrt{1 - \left(\frac{\lambda_2}{\lambda_1}\right)^2} \tag{5.2.1}$$

Where for monomers of equal mass the first eigenvale is equal to one and the second is bounded by the first. Therefore the values of eccentricity go from 0 to 1, 0 being minimally eccentric and 1 being maximally eccentric.

To measure the radius of gyration of the monomers we take the sum of the squared differences of the x and y coordinates from the mean

$$R_g^2 = \sum_{i=1}^{N} \left(X_i - X_{COM}\right)^2 + \left(Y_i - Y_{COM}\right)^2 \tag{5.2.2}$$

The number of unbound monomers is a measure of what fraction of the fluctuating polymer lies within the bound state of the potential. This is determined by the cutoff for the harmonic piecewise potential and close to the σ parameter in the gaussian one. For the gaussian potential the cutoff is chose such that $y_{cutoff} - > -D_e e^{\frac{-\left(y_{cutoff} - y_o\right)^2}{2\sigma_g^2}} \leq 0.1 k_B T$, or that the depth of the potential well is only a fraction of the thermal energy. To measure the spin quantity is given by

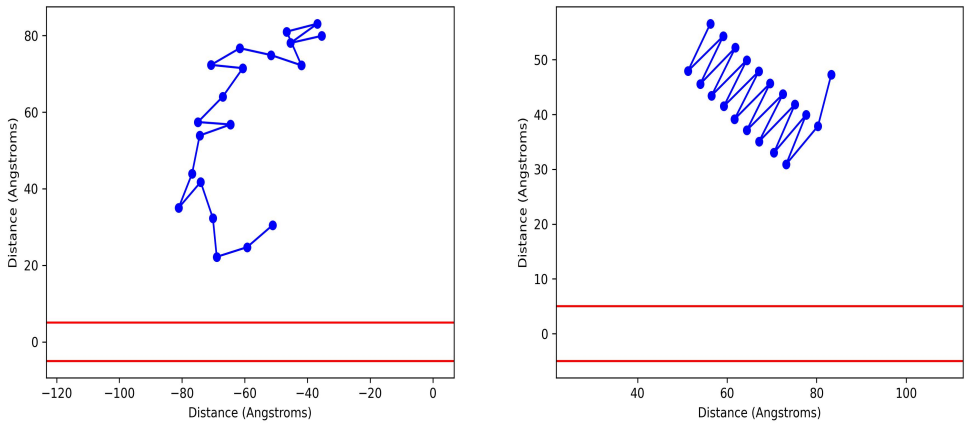
$$S = \sum_{i=1}^{N} N\left(X_i, Y_i\right) \tag{5.2.3}$$

Where

$$N\left(X_i, Y_i\right) = \begin{cases} 0 & r > r_{cutoff} \\ 1 & r < r_{cutoff} \end{cases}$$

where $r = \sqrt{X_i^2 + Y_i^2}$. A spin that is equal to the $N - 1$ is having all spins of the system intact.

In figure 5.2.1 we see examples of the simulated system in structured and non-structured cases. The unstructured example has examples of an intact next neighbor bond and the fully structured system has almost every bond intact and is in the zig-zag conformation mentioned earlier. There is no inherent stiffness between neighboring units in the form of an potential on the angle between monomers, so stiffness is introduced in the strength of the LJ interaction. Thus the interesting parameters to observe the behavior of the system would be varying D_e, the strength of the binding potential, and ϵ the strength of the nearest neighbor interactions.



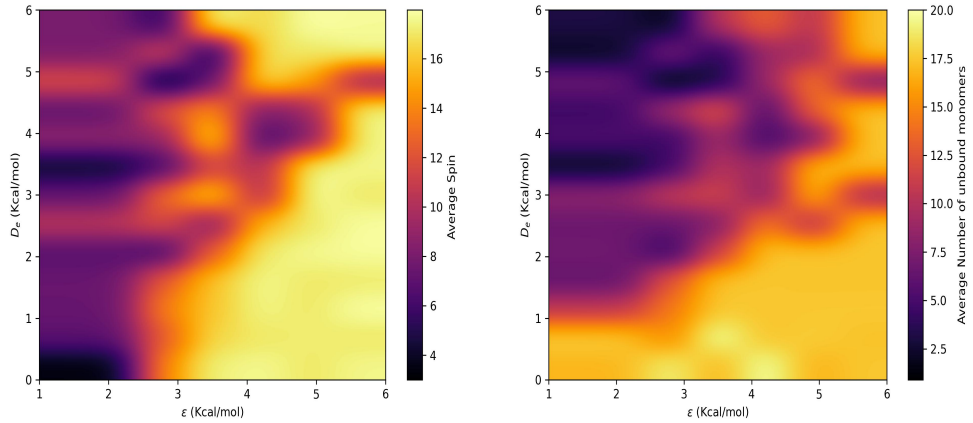
(a) (b)

Figure 5.2.1: Example configuration of the in the potential well. The blue connected dots is the positions of the monomers and the red line is the edge of the piecewise harmonic potential or the σ parameter in the guassian potential. The x and y axis are both dimensionless positions.(a) Example of a low structure state. (b) Example of a almost fully structure polymer

5.3 Simulations reveal existence of many metastable states

Using the above descriptions for observables, we vary σ and D_e for simulations of one million time steps and a value of σ of about two, which roughly correlates to a width of the potential about as large as as the distance between two monomers. Looking at figure 5.3.1 we see that as the LJ gets stronger the average number of intact bonds (spins) grows as ϵ which follows intuition. The average number of unbound monomers, for the most part, follows that as D_e gets lower, more monomers leave the potential well.

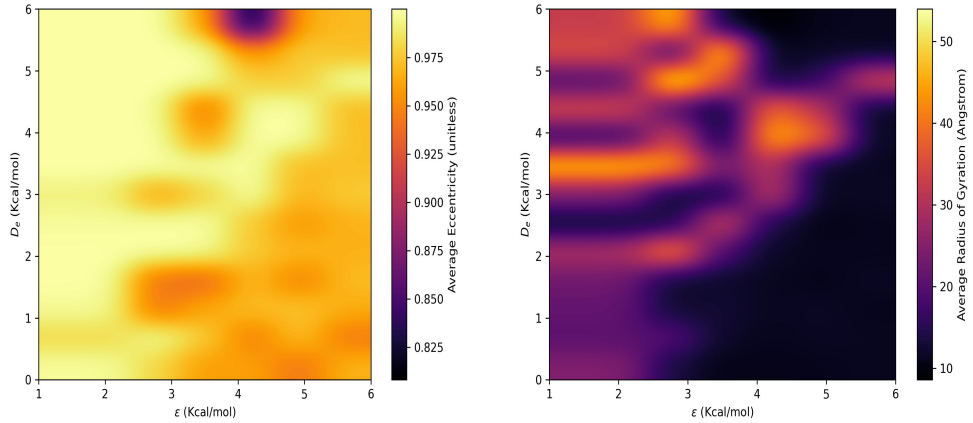
There is also a portion of the heat maps that show for $\epsilon \approx 4$ and $D_e \approx 2 - > 5$, that the number of spins goes down as ϵ increases, and that as D_e increases, the number of unbound monomers also increase. In figure5.3.2, plots of the eccentricity and radius of gyration are shown. There is a similar pattern in these plots. Generally the radius of gyration decreases as ϵ gets smaller and the eccentricity follows. It is worth noting that the eccentricity values are all very close together, indicating that even in highly structured cases they might be



(a) (b)

Figure 5.3.1: Plots of the average spin (a) and the average number of unbound monomers (b) for the $\sigma = 2$ simulations.

hard to distinguish from elongated states.



(a) (b)

Figure 5.3.2: Plots of the average eccentricity (a) and the average radius of gyration (b) for the $\sigma = 2$ simulations.

To investigate the nature of these states, we show snapshots in figure 5.3.3 with their associated simulation. States were chosen that in the various regions of the above heat maps. First to note is the the states with the large radius of gyration in the center region of the plots. They are very elongated structures with maybe a few intact bonds that are always

in the direction of the force according to the binding potential, meaning they are spanning the width of the potential and stabilized by it. The smaller radius of gyration states are the more ordered states and a moderate radius of gyration corresponds to the random coil like states or a "broken" chain. Where a good fraction of the bonds are intact, but the overall size is still small. The largest states are ones that correspond to states that have had almost all of their spins destroyed and every monomer sits inside the binding potential.

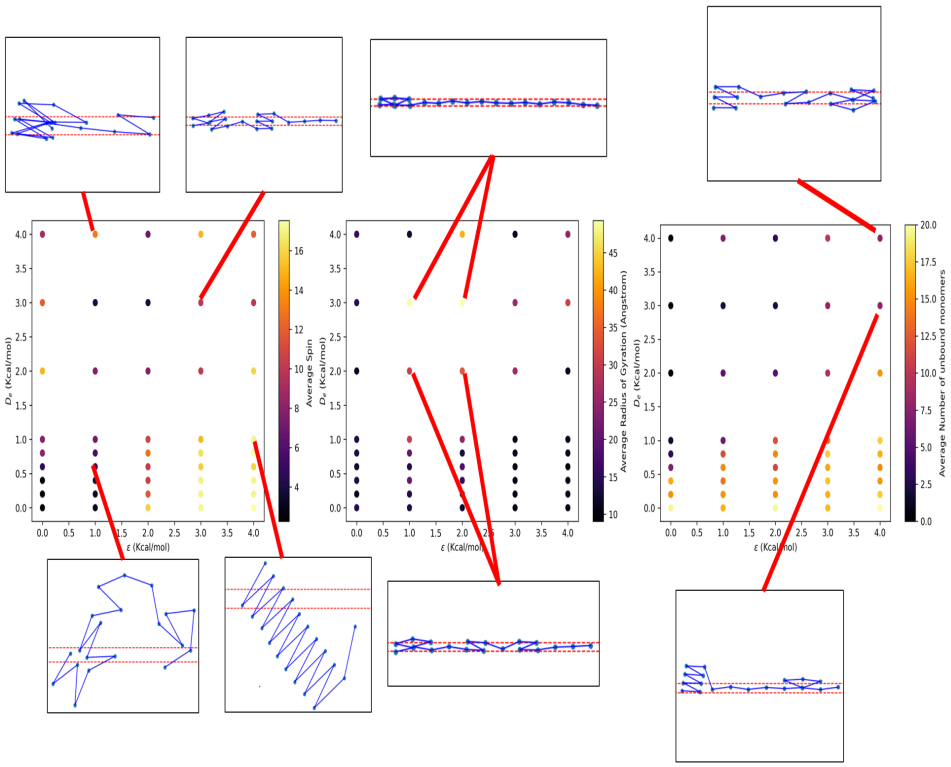


Figure 5.3.3: Data points from the above heatmaps with corresponding states shown in the insets connected to the data point with red lines.

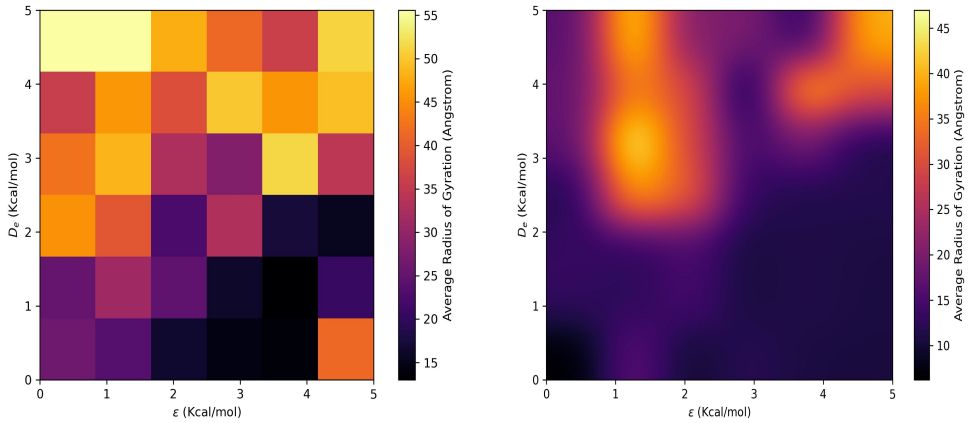
As for the number of unbound monomers, we see that for low D_e, the majority of those states are lying across or only partially in the potential. For the disordered states this would be expected because as you see an unbinding transition, the size and number of loops, the number of consecutive monomers that leave and then enter the cutoff region, increases as the temperature rises or the binding potential depth goes down. Surprisingly, from previous theory, we would expect that as the chain starts to stiffen, the binding would

be less frustrated as the conformational entropy starts to decrease and the system moves to the more enthalpically favorable bound state. We instead see that as the bonds stay intact, the primary axis of the polymer rotates about the bonding axis and appears to favor only a few monomers in the well. We will discuss this in more detail later. The two structures on the top and bottom of the plot on the far right in figure 5.3.3 we also see what the states that are on the edge of the region mentioned in the first paragraph of this section look like. Unlike the elongated states, where only one bond is inside the potential well, here we see up to 4 different bonds intact perpendicular to the binding axis. As the chain stiffens,and ϵ and D_e become comparable in strength, it becomes even more difficult to "denature" the chain. Once the bonds have oriented themselves in this way, thermal fluctuations would need to be relatively large for them to be able to remove themselves from this state. This is the first evidence we see that these state can be very long lived kinetic traps.

In the large D_e for low values of ϵ the chain appears to collapse on itself. In the plots of 5.3.1, there appeared to be values in this region of moderate spin count, but in the case of a collapsed state with distance criterion being the only thing that determines a spin, it would appear that this is an artificial increase in the spin number. For the case of $\epsilon = 0$ there will be not spins intact because that potential no longer exists. For low values of ϵ and D_e, the states are mostly random coil states that have very short lived transient bonds that form.

5.4 Metastable states gives rise to slight hysteresis

Figure 5.4.1 shows two plots of different starting configurations to look for evidence of hysteresis or dependence of observables on the initial condition of the system. The simulations show the chains are similar in behavior for the most part, indicating that the model used can help with determining binding mechanisms from out of equilibrium states as suspected from the discussions of chapter 4. There does appear to be some slight hysteresis in the regions of large binding potential depth and monomer-monomer interactions. When in the large force regimes , the system becomes more depend on on initial conditions when thermal fluctuations are not able to "jostle" the system out of a metastable state. For example, for



(a) (b)

Figure 5.4.1: Plots of the radius of gyration given two different starting configurations.(a) An elongated configuration where the chain is laid along the x-axis.(b) A zig-zag starting configuration.

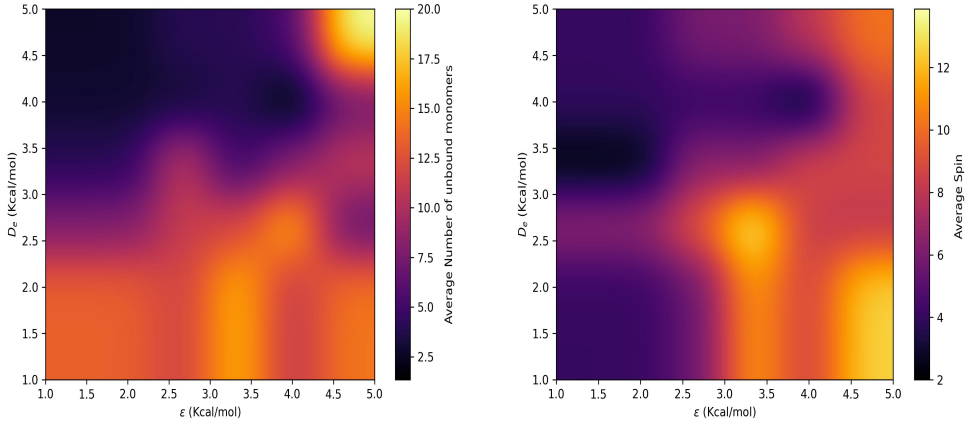
large D_e, if the starting configuration is elongated at the base of the potential well, then very large fluctuations would be needed to push the particles out of the well so they can become available to form bonds with neighboring monomers. This is most notable when ϵ and D_e are comparable in magnitude to each other.

5.4.0.1 Excluded volume aids in prevented trapped metastable states

The evidence of collapsed states where monomers may lie on top of each other suggests that an excluded volume interaction may be necessary. Inclusion of excluded volume interactions shifts the configuration space the chain may occupy by restricting the positions of monomers to certain regions dependent on the positions of other monomers. To include excluded volume interactions, a soft repulsive gaussian potential was added to the center of the position of each monomer. The potential has the form

$$U_e(\{r_{ij}\}) = \sum_{i=1}^{N}\sum_{j=1}^{N} A_e e^{\frac{-\left(r_{ij}\right)^2}{2\sigma_e^2}} \tag{5.4.1}$$

where $\{r_{ij}\}$ is the set of monomer-monomer distances , A_e is strength of the excluded volume interaction and σ_e represents the physical size of the monomer.It is notable that this interaction is summed over all potential pairs at all distances and thus greatly increases the computational cost.



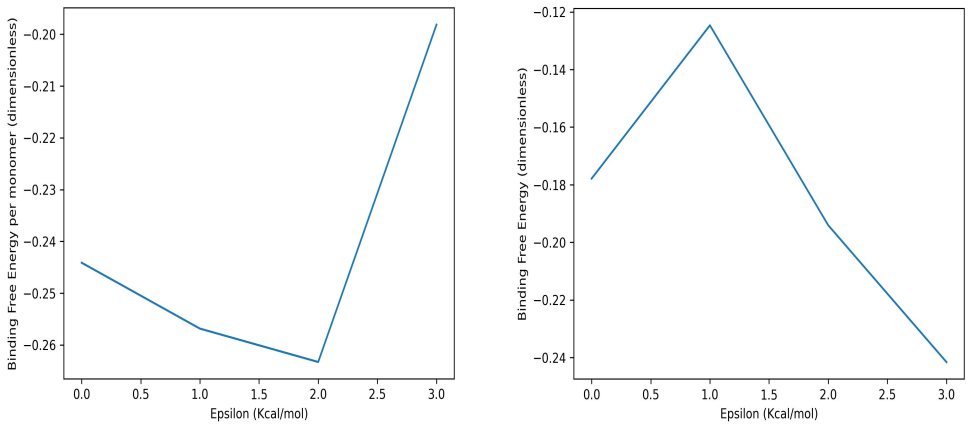
(a) (b)

Figure 5.4.2: Plots of the average number of unbound monomers(a) and the average sping (b) for the system with excluded volume.

We see in figure 5.4.2 that the trends for included excluded volume interactions and without excluded volume in 5.3.1 are similar. However, in the large D_e and large ϵ regime, it appears that there is less variation in the observables between different values of these parameters. It is shown that the inclusion of excluded volume effects are helpful in helping the chain escape from metastable states. It also suggests that metastable states may be avoided if you can restrict the configuration space of the chain.

5.5 Evidence of stiffening induced unbinding

Figure 5.5.1 shows that for small changes the binding energy per monomer, calculated as detailed in appendix A, decreases as as a function of ϵ. The binding energy per monomer increases as a function of ϵ for a chain of 100 monomers. Looking back to figure 5.3.3, in this regime the structure is kept intact but the majority of monomers rotate outside of the well.



(a) (b)

Figure 5.5.1: Plots of the free energy of binding per monomer for (a) a chain of 20 monomers and (b) a chain of 100 monomers.

Thus rotational motion must be larger for short chains than it is for long chains and a finite size effect. In the work of Li et al [61], they show for packaging high flexibly polymer chains in an aqueous environment, the free energy of the packaged polymer is lower than that of a branched system for highly flexibly chains only when they are short enough. It is possible that these two phenomena are related.

Including the stiffness of the chain in this nonlinear provides an alternate mechanism for how flexibility induced binding could occur. In classical elasticity theory, there is a bending energy penalty for deforming the chain and then a favorable change in configurational entropy. Here the trade-off is the energy cost of the loss of a rigid for a favorable change in configurational entropy. These "kinks" can form and allow the two new sections of the chain that have been created to bend without energy cost. For short chains, the gain in configurational entropy is not enough to compensate for the loss of rotational entropy of the short chain and the change in energy for breaking monomer-monomer bonds. Figure 5.5.2 displays a configuration of the long $N = 100$ chain where there is an example of this happening.

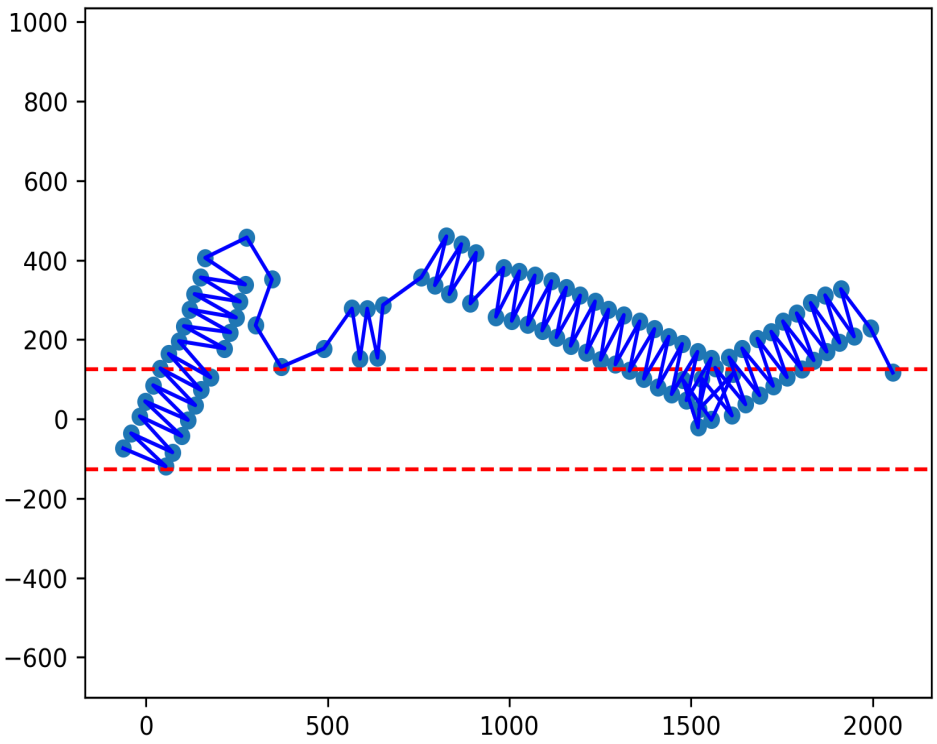


Figure 5.5.2: Configuration of a large ϵ state in a moderate D_e potential well. The appearance of a kink in the chain forms to maximize binding energy at the cost of losing a monomer-monomer bond.

5.6 The inclusion of position dependence of the binding potential highlights how flexibility can aid in binding to a variety of environments

Up to this point the interactions between the monomers and the binding potential have been invariant along the x-axis, or the binding potential axis. Since real proteins, and specifically IDPs, can bind to a wide variety of targets, it is of interest to investigate how the ability for a free chain to form rigid structure can aid or hinder the chains ability to bind. To include a position depends of the binding potential we add it in two ways: (1) a varied gaussian that is modulated by a sine wave to provide regions of repulsion and attraction and (2) a varied gaussian that is modified by a squared sine wave to provide neutral regions and attractive

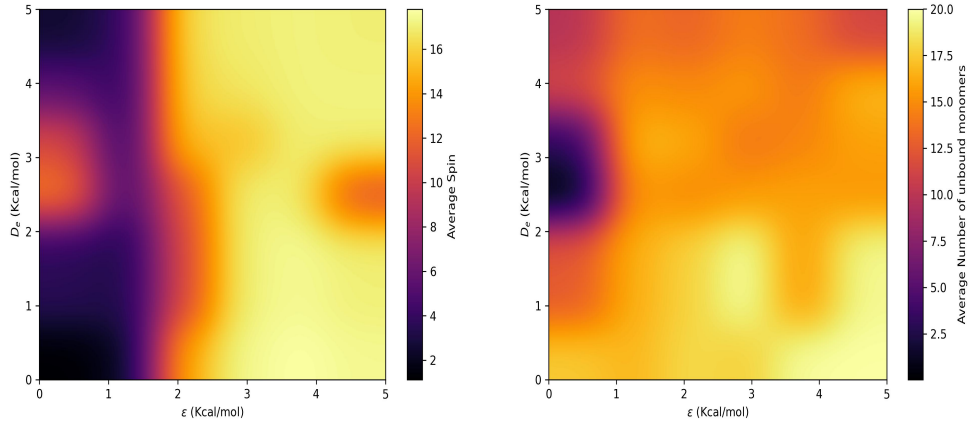
regions. All of the simulations are have exluded volume effects from the previous section included.

5.6.1 Position dependence of alternating repulsive and attractive regions

Using a binding of the potential of the form

$$U_{bp}(\{x_i, y_i\}) = \sum_{i=1}^{N} -D_e \sin\left(\frac{2\pi x_i}{\lambda}\right) e^{\frac{-(y_i - y_o)^2}{2\sigma_g^2}} \tag{5.6.1}$$

Where $\{x_i, y_i\}$ are the set of x and y positions of the monomers and λ is the wavelength of the sine wave. This parameter effective controls the spacing of the attractive parts of the potential along the x-axis. The other parameters are the same as defined in the original binding potential.



(a) (b)

Figure 5.6.1: Plots of the average unbound monomers(a) and spin(b) for a wavelength of $\lambda = 1$ for an alternating attractive and repulsive potential. The majority of monomers in these cases are largely unbound and the spins are intact for moderate to large values of ϵ.

In figure 5.6.1 we see that the monomers are largely unbound for these given parameters and the monomer-monomer bonds are intact for moderate to large values of the ϵ. In figure 5.6.2 an example structure from these simulations is plotted. The x and y axis are the coordinates of the positions of the monomer and the contour on the plot indicates the values of the potential at that point.

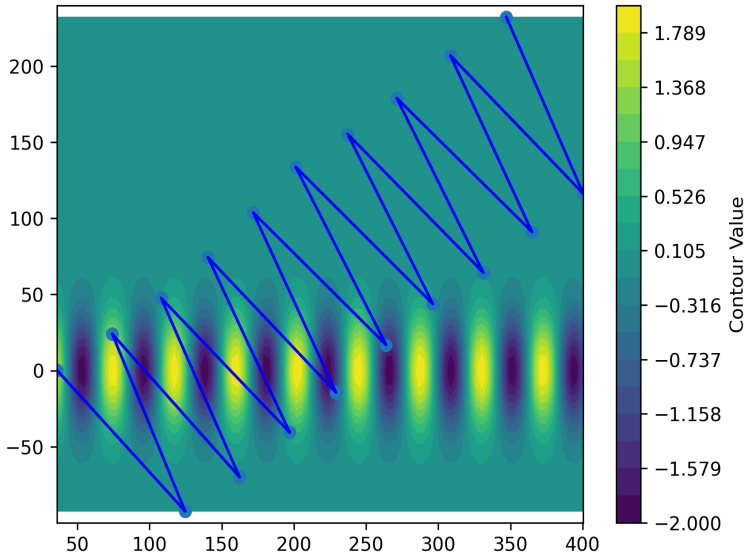
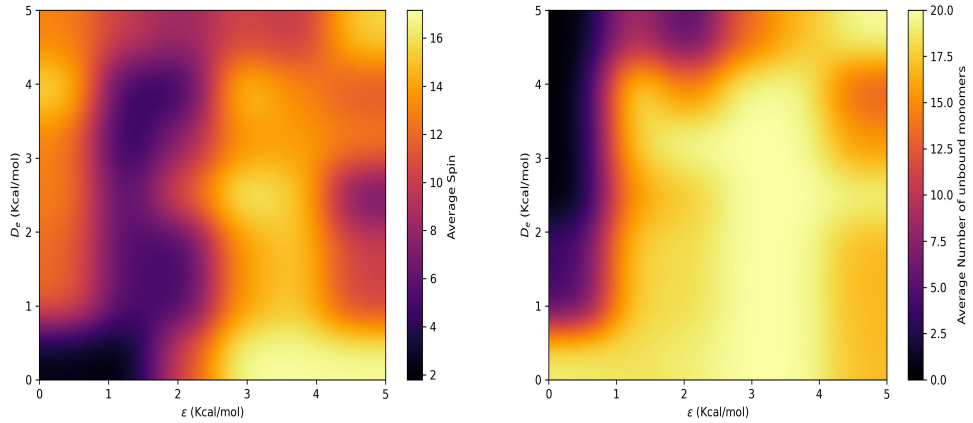


Figure 5.6.2: Configuration of an example structure of $\lambda = 1$ for large a ϵ. The color indicated the value of the binding potential at that point indicated by the color bar on the plot.

The regions of repulsion force the monomer to rotated about the axis to try and maximize the number of monomers that can be in attractive parts of the potential. The wavelength is not an integer multiple of the next nearest neighbor bond distance, so the chain is not perfectly perpendicular to the binding axis. Adding a position dependence smaller than the bond distance severely restricts the number of configurations that the chain can adopt while in a bound state.

In figure 5.6.3 the same plots are shown for a much larger wavelength than the monomer-monomer bond distance. There are now even a larger fraction of the unbound monomers at most values of ϵ. Only in the case of $\epsilon = 0$ do we see any significant number of monomers bound. The origin of this behavior is seen in figure 5.6.4. The chain has dissociated entirely from the binding potential. The spacing between attractive and repulsive regions has become so large that for larger values of ϵ the chain cannot deform in any way to adjust to the binding potential.

For a smaller value of epsilon, $\epsilon = 2$, we see in figure 5.6.5 that the bonds are able to break and that the spacing between attractive regions of the potential are roughly around



(a) (b)

Figure 5.6.3: Plots of the average unbound monomers(a) and spin(b) for a wavelength of $\lambda = 4$ for an alternating attractive and repulsive potential. The majority of monomers in these cases are largely unbound and the spins are intact for moderate to large values of ϵ, the same as in the $\lambda = 1$ case.

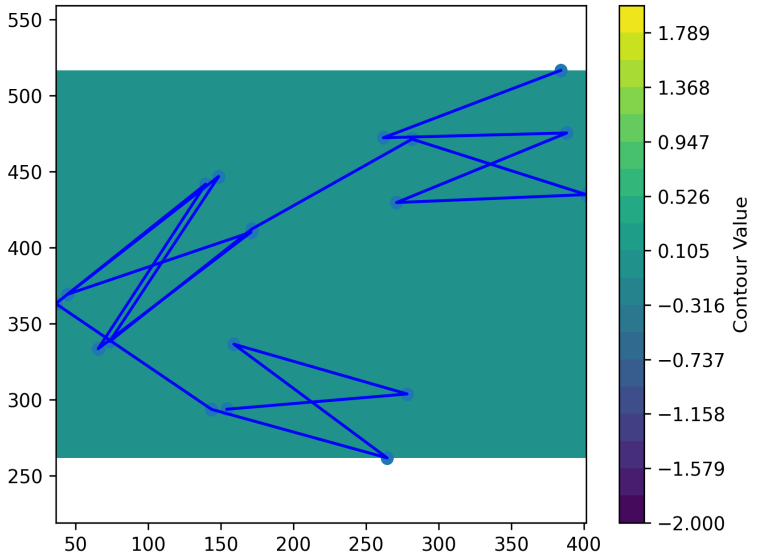


Figure 5.6.4: Configuration of an example structure of $\lambda = 4$ for large a ϵ. The color indicated the value of the binding potential at that point indicated by the color bar on the plot. The resultant structure is completely unbound from the potential.

the same length as the equilibrium distance between neighboring monomers. The chains is then able to deform to accommodate the potential. The next nearest neighbor bonds are also able to form within these attractive regions for certain configurations, as seen in the $-200 > y > -100$ region. Three monomers are able to bind to each other in the attractive well, which further stabilize the state when compared to a chain where no structure formation is possible. For a purely elastic chain as well, this type of deformation would be too large an energy cost, proving it to be advantageous to have this type of nonlinearity in the elasticity of the chain.

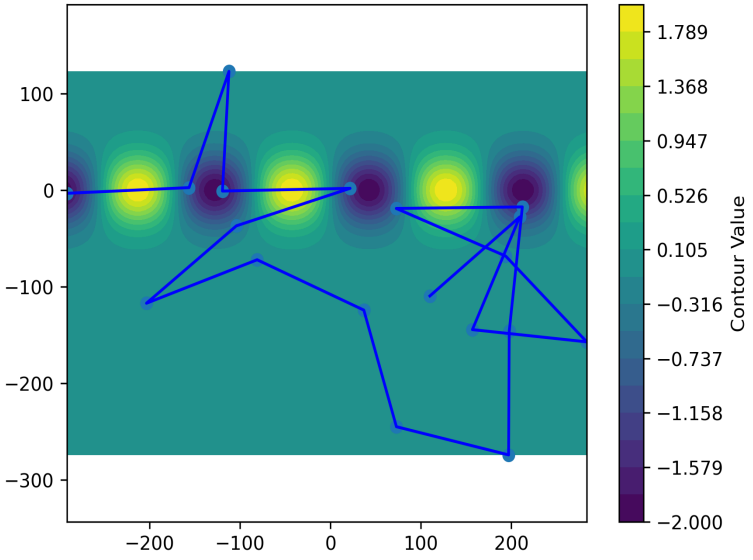


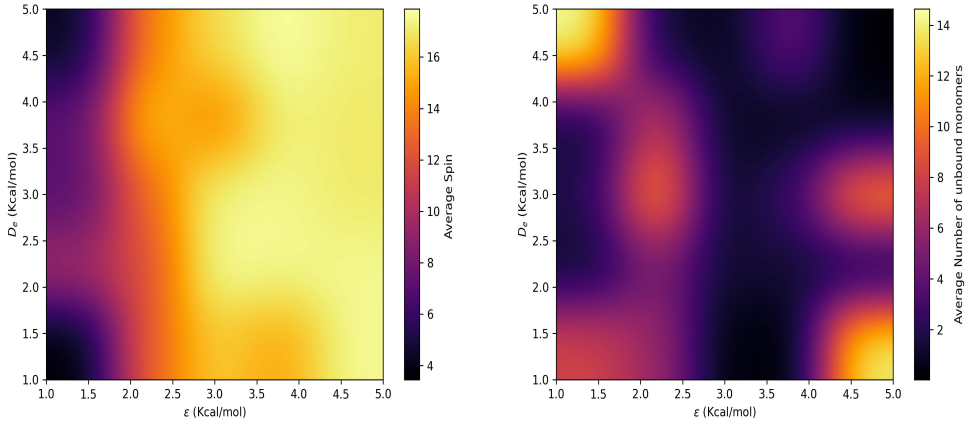
Figure 5.6.5: Configuration of an example structure of $\lambda = 4$ for large a small ϵ. The color indicated the value of the binding potential at that point indicated by the color bar on the plot. The resultant structure is highly deformed but still bound to the binding axis.

5.6.2 Position dependence of strictly attractive potential

A system without repulsive regions along the x-axis is also chosen to simulation binding situations where the interaction between alternation wells is neutral. Using a binding potential of the form

$$U_{bp}(\{x_i, y_i\}) = \sum_{i=1}^{N} -D_e \sin^2\left(\frac{2\pi x_i}{\lambda}\right) e^{\frac{-(y_i - y_o)^2}{2\sigma_g^2}} \tag{5.6.2}$$

Where $\{x_i, y_i\}$ are the set of x and y positions of the monomers and λ is the wavelength of the sine wave. This parameter effective controls the spacing of the attractive parts of the potential along the x-axis. The other parameters are the same as defined in the original binding potential and the alternating repulsive potential. Here the $\sin^2$ allows for all the the wells to be attractive and the regions between them neutral.



(a) (b)

Figure 5.6.6: Plots of the average unbound monomers(a) and spin(b) for a wavelength of $\lambda = 2$ for a purely attractive potential. The majority of monomers in these cases are largely unbound and the spins are intact for moderate to large values of ϵ, the same as in the $\lambda = 1$ case.

Unlike the previous potential, where the observables for these parameters showed that the chain was mostly unbound, here the monomers are largely in the bound state for larger values of ϵ. Figure 5.6.7 shows that for this spacing, which is equal to the bond distance between next nearest neighbors, the chain can retain its structure and fit nicely into the potential wells of the binding potential.

Here a large effective stiffness of the chain is inconsequential to the binding. Looking at figure 5.6.8 We see that for small epsilon the structure is partially broken, but the chain remains bound.

Seen from the spin plots in figure 5.6.6, even at small values of epsilon the spins remain

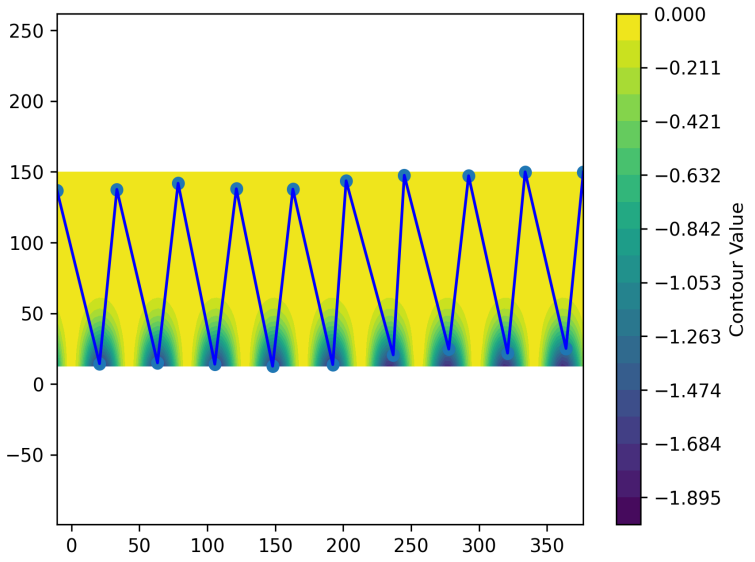


Figure 5.6.7: Configuration of an example structure of $\lambda = 2$ for large ϵ. The color indicated the value of the binding potential at that point indicated by the color bar on the plot. The resultant structure a zig-zag structure that has fit half of its monomers into the potential well.

largely intact. The potential helps stabilize these spin interactions in two ways. In figure 5.6.7 the bonds are stabilized by the matching of the potential well separation and the equilibrium bond distance. However, this forces half o the monomers out of the minima in the y direction of the potential. In figure 5.6.8, the bonds are stabilized by the being perpendicular to the binding axis. In the y direction the potential is still gaussian and the spacing between potential wells such that neighboring monomers can extend over 4 potential wells such that in the direction of the harmonic force is parallel to the binding axis. Orienting parts of the chain this way allows for the gaussian potential to push next nearest neighbors together, stabilizing the bonds between those monomers.

In figure 5.6.9 the spin of the larger wavelength potential is relatively similar to that of the smaller wavelength. However, for large epsilon we see that unbinding occurs when the stiffness of the chain increases with ϵ. For all values of D_e there is stiffening induced unbinding.

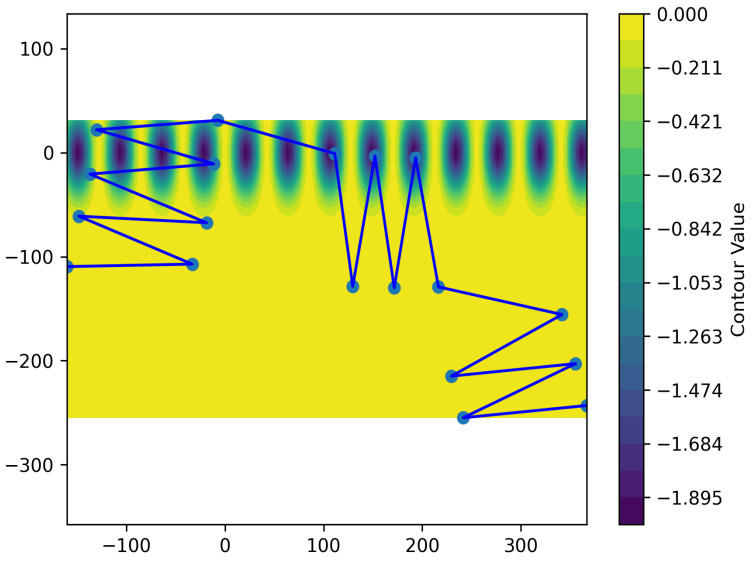
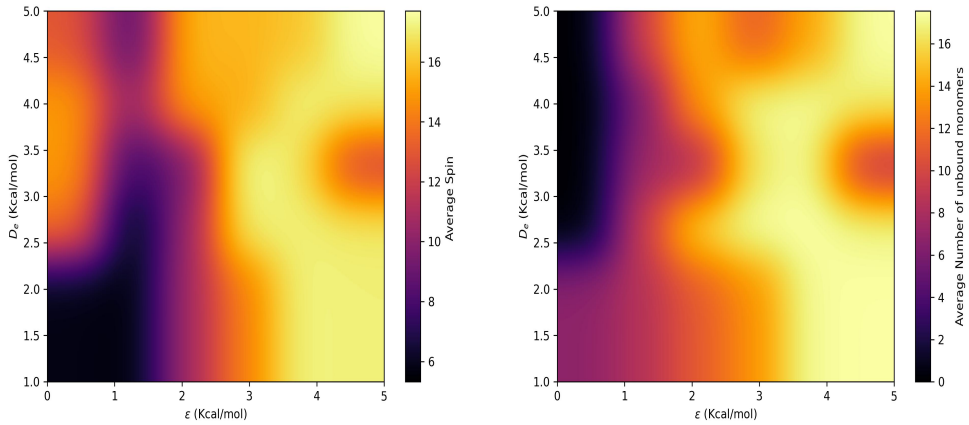


Figure 5.6.8: Configuration of an example structure of $\lambda = 2$ for small ϵ. The color indicated the value of the binding potential at that point indicated by the color bar on the plot. The resultant structure is highly deformed and not bound as well as the larger ϵ case.



(a) (b)

Figure 5.6.9: Plots of the average unbound monomers(a) and spin(b) for a wavelength of $\lambda = 6$ for a purely attractive potential. The majority of monomers in these cases are largely unbound and the spins are intact for moderate to large values of ϵ, the same as in the $\lambda = 1$ case.

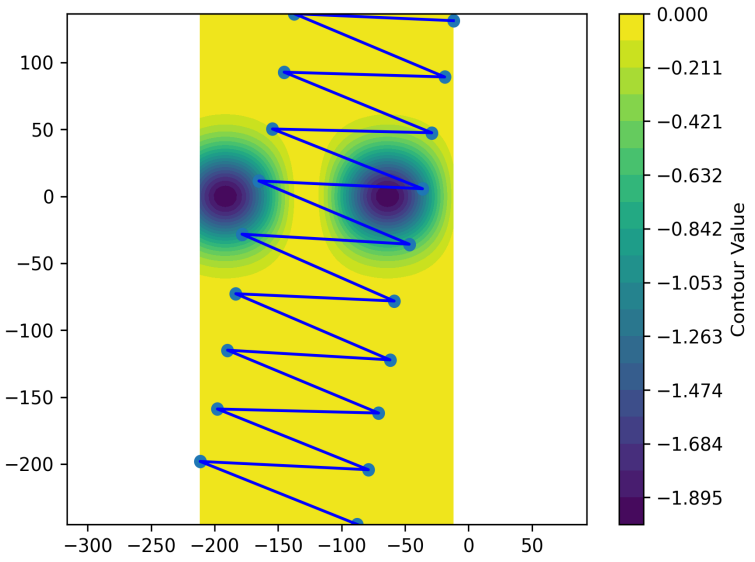


Figure 5.6.10: Configuration of an example structure of $\lambda = 6$ for $\epsilon = 4$. The color indicated the value of the binding potential at that point indicated by the color bar on the plot. The chain is highly structured and spacing between potential wells allows for rotation of the chain perpendicular to the binding axis.

Figure 5.6.10 reveals that the structure of the chain remains intact and the equilibrium configuration is one that is perpendicular to the binding axis resulting in a relatively frustrated structure as most of the monomers sit outside the potential well. As the chain becomes softer and deforms, the monomers can extend parallel to the binding axis and find a lower energy structure. Figure 5.6.11 shows an equilibrium configuration for the chain.

5.7 Concluding remarks

This section has shown that using the nonlinear elasticity model of chapter four helps in understanding the binding scenarios seen in chapter 2. By introducing the elasticity in this way, we are able to recreate a stiffening induced reduction in binding that is contradictory to traditional theory of binding of deformable chains [36]. For small chains that can form structure, like the RNA molecules in chapter two, an increase in the rigidity of this structure induces unbinding and lowers the free energy of binding. When the length of the

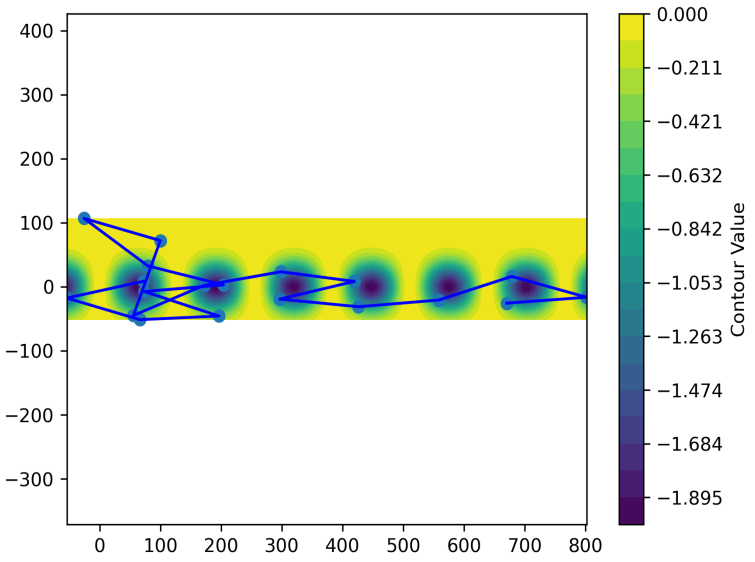


Figure 5.6.11: Configuration of an example structure of $\lambda = 6$ for $\epsilon = 1$. The chain keeps a moderate amount of its spins intact while deforming to sit along the binding axis.

chain is increased, rotational entropy becomes negligent to the overall free energy of binding and stiffening of the chain increases the binding energy, which is predicted by the classical theories.

Adding postion dependence of the binding potential along the axis of binding has shown that, for short chains, moderate elasticity aids in the ability to bind. In the case of alternating repulsive and attractive sections of the potential, binding becomes extremely hindered for potential wavelengths that are not integer multiples of the bond distance or neighboring monomer extensions. In this case softer chains will always bind more favorably to easier deform to match pattern of the potential. In the case of only attractive wells stiffening of the chain can help in binding when the bond distance is comparable to the wavelength of the well separation. This resembles a lock and key mechanism for binding. As the wavelength deviates from this value, the more elastic states are favored in binding up to a point. Moderate stiffness introduced in this way can help in binding by further stabilizing monomers in the bound state.

Further work would look into the effects of binding of longer chains to position dependent

potentials and more quantitative ways of measuring the binding free energy. Introducing system complexity such as the binding of two deformable polymers could reveal mechanisms in cooperative binding mechanism that more accurately represent real systems.

APPENDIX A

Thermodynamic integration for MD and BD simulations

To get to the expression for equation 2.2.1 we start with the generalized expression for the free energy of our system.

$$F = -k_B T \ln Z \tag{A.0.1}$$

Where k_B is the Boltzmann constant, T is the temperature and Z is the canonical partition function given by

$$Z = \int e^{-\beta H(x_1, x_2, \ldots, x_N)} dx_1 dx_2 \ldots dx_N$$

$\beta = \frac{1}{k_B T}$, x_i being the generalized coordinates of the system and H being the Hamiltonian of the system that depends on these coordinates. For any part of the Hamiltonian, we can parameterize it by the mathematical parameter λ that goes from 0 to 1, such that the partition function then becomes

$$Z = \int e^{-\beta H(x_1, x_2, \ldots, x_N, \lambda)} dx_1 dx_2 \ldots dx_N$$

To then see how the free energy of the system then depends on this parameter, we can take the derivative of the partition function w.r.t λ.

$$\frac{dF}{d\lambda} = -k_B T \frac{\int \frac{dH(x_1, x_2, \ldots, x_N, \lambda)}{d\lambda} e^{-\beta H(x_1, x_2, \ldots, x_N, \lambda)} dx_1 dx_2 \ldots dx_N}{\int e^{-\beta H(x_1, x_2, \ldots, x_N)} dx_1 dx_2 \ldots dx_N} \tag{A.0.2}$$

We recognize the quantity to the right of $k_B T$ as the thermodynamic average of the quantity in the integrand in the numerator. Thus the derivative is given by

$$\frac{dF}{d\lambda} = -k_B T \langle \frac{dH(x_1, x_2, \ldots, x_N, \lambda)}{d\lambda} \rangle \tag{A.0.3}$$

Where the brackets $\langle\cdot\rangle$ represent the thermodynamic average. Then by integrating both sides of equation A.0.3 w.r.t λ we have

$$\Delta F = -k_b T \int_0^1 \langle \frac{dH(x_1, x_2, \ldots, x_N, \lambda)}{d\lambda} \rangle_\lambda d\lambda \tag{A.0.4}$$

Where the subscript λ after the brackets refers to averages w.r.t to different values of λ. A.0.4 is the desired general result that tells us the free energy difference of the system with and without the parameter λ. In the case of the MD simulations described in chapter 2, we look to unbias the binding energy that has been altered by the restraints put on the system. The restraints of the system are in the form of harmonic restraints and the parameter λ tunes the strength of these restraints. Thus, the parameterized Hamiltonian for the system is

$$H(x_1, x_2, \ldots, x_N, \lambda) = T(\dot{x_1}, \dot{x_2}, \ldots, \dot{x_N}) + \lambda \frac{1}{2} k(\Theta(x_1, x_2, \ldots, x_N) - \Theta^*)^2 + \sum U_i(x_1, x_2, \ldots, x_N)$$

Where $T(\dot{x_1}, \dot{x_2}, \ldots, \dot{x_N})$ is the kinetic energy, the summation on the r.h.s of the equation is the potential energy terms that do not depend on λ, k is the stiffness of the restraining potential, $\Theta(x_1, x_2, \ldots, x_N)$ is a function that represents the conformation of the RNA molecule and Θ^* is a reference conformation. Inserting this into equation A.0.4 we get

$$\Delta F = -\int_0^1 \langle k(\Theta(x_1, x_2, \ldots, x_N) - \Theta^*) \rangle_\lambda d\lambda \tag{A.0.5}$$

k is directly modified by λ, so the thermodynamic averages correspond to averages w.r.t different spring constants for the restraining potentials, giving us equation 2.1.1.

In the Case of the BD simulations, the idea is the same. The binding energy is given by integrating the thermodynamic average with λ modifying the binding potential so the Hamiltonian becomes

$$H(x_1, x_2, \ldots, x_N, \lambda) = T(\dot{x_1}, \dot{x_2}, \ldots, \dot{x_N}) + \sum_i \lambda D_e e^{\frac{y_i^2}{2\sigma^2}} + \sum_i U_i(x_1, x_2, \ldots, x_N)$$

Plugging this into equation A.0.4 we get

$$\Delta F = -\int_0^1 \langle \sum_i D_e e^{\frac{y_i^2}{2\sigma^2}} \rangle_\lambda d\lambda \tag{A.0.6}$$

Thus by measuring the average quantity inside the brackets in 1.0.6 we should be able to find the binding energy of the system.

APPENDIX B

Numerical integration of the Langevin equation in dimensionless units

We start from a general one-dimensional Langevin equation:

$$m\ddot{x} + \gamma\dot{x} = F(x) + \eta(t) \tag{B.0.1}$$

Here η is the white noise with zero mean and with an autocorrelation function $< \eta(t)\eta(t') >= 2k_BT\delta(t-t')$. We next discretize this equation by integrating over a time interval Δt with

$$t_n = n\Delta t; x_n = x(t_n); v_n = v(t_n) \tag{B.0.2}$$

This gives:

$$\int_{t_n}^{t_n+\Delta t} \frac{d^2x}{dt^2}dt + \int_{t_n}^{t_n+\Delta t} \gamma\frac{dx}{dt}dt = \int_{t_n}^{t_n+\Delta t} F(x)dt + \int_{t_n}^{t_n+\Delta t} \eta(t)dt \tag{B.0.3}$$

This is approximated as

$$m(v_{n+1} - v_n) + \gamma(x_{n+1} - x_n) = \int_{t_n}^{t_n+\Delta t} F(x)dt + \int_{t_n}^{t_n+\Delta t} \eta(t)dt \tag{B.0.4}$$

at time t_{n+1} the force is expanded in a Taylor expansion, $F(x_n)+F'(x_n)v_n(t-t_n)+...\mathcal{O}(\Delta t^2)$. Inserting this into the integral we get:

$$\int_{t_n}^{t_n+\Delta t} Fdt = F(x_n)\Delta t + \frac{1}{2}F'(x_n)v_n\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{B.0.5}$$

Since $\eta(t)$ & $\eta(t')$ are uncorrelated random variables, the integral of the white noise must be a Gaussian random variable with zero expectation. Its variance can be calculated as:

$$\sigma_n^2 = \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t_{n+1}} dt'' < \eta(t')\eta(t'') >= 2k_B T\gamma\Delta t \tag{B.0.6}$$

thus the magnitude of the integral of the noise is $2k_BT\gamma\Delta t$, we here name it as Δn. This expression indicates that the random force will dominant when time step is small enough. With preparing all integral up to a range within time step, now the Langevin equation can be translated to a update rule:

$$x_{n+1} = x_n + v_n\Delta t + \mathcal{O}(\Delta t^2); v_{n+1} = v_n - \gamma\frac{v_n\Delta t}{m} - \frac{F(x_n)\Delta t}{m} + \frac{\Delta n}{m} + \mathcal{O}(\Delta t^2) \tag{B.0.7}$$

We now convert all quantities into a dimensionless manner. The important step is define a characteristic time: $\tau_0 = \frac{m}{\tau}$. This selection guaranteed when $\Delta t < \tau_0$, the inertial effect would dominant, on the contrary the viscous effect will play a main role. Then the dimensionless time step should be $\bar{\Delta t} = \frac{\Delta t}{\tau_0}$.

Then we define the characteristic energy by equipartition theorem: $\frac{1}{2}m < v^2 >= \frac{1}{2}k_BT$. Thus the characteristic velocity and displacement can be derived as:

$$v_0 = \sqrt{\frac{k_BT}{m}}; x_0 = v_0\tau_0 \tag{B.0.8}$$

similarly the dimensionless quantities can be obtained by dividing the characteristic one from raw value. This makes the updating rule in B.0.7 can be rewrite as follows:

$$\bar{x_{n+1}} = \bar{x_n} + \frac{v_n\Delta t}{x_0} = \bar{x_n} + \frac{v_n\Delta t}{v_0\tau_0} = \bar{x_n} + \bar{v_n}\Delta t \tag{B.0.9}$$

$$\bar{v_{n+1}} = \bar{v_n} - \gamma\bar{v_n}\frac{\Delta t}{m}\tau_0 - \frac{F(x_n)\bar{\Delta t}\tau_0}{mv_0} + \frac{\Delta n}{mv_0} = \bar{v_n} - \bar{v_n}\bar{\Delta t} - \frac{F(x_n)\bar{\Delta t}\tau_0}{mv_0} + \frac{\Delta n}{mv_0} \tag{B.0.10}$$

the reorganization of the second term is hold by $\frac{\gamma\tau_0}{m} = 1$ which is the definition of τ_0. Then comparing the corresponding terms, we can further define characteristic force and random noise:

$$F_0 = \frac{mv_0}{\tau_0} = \frac{mv_0^2}{\tau_0 v_0} = \frac{k_B T}{x_0}; \bar{\Delta n} = \frac{\Delta n}{mv_0} \tag{B.0.11}$$

the variance, that is the magnitude of the noise can be estimated as:

$$\bar{\sigma^2} = \frac{2k_B T\gamma\Delta t}{m^2 v_0^2} = \frac{2k_B T\Delta t}{mv_0^2\tau_0} = 2\bar{\Delta} t \tag{B.0.12}$$

REFERENCES

[1] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science (New York, N.Y.)*, 276(5315):1109–1112, May 1997.

[2] Craig Fogle, Joseph Rudnick, and David Jasnow. Protein viscoelastic dynamics: a model system. *Physical Review E*, 92(3):032719, September 2015. arXiv: 1502.00343.

[3] Gene W. Yeo, Nicole G. Coufal, Tiffany Y. Liang, Grace E. Peng, Xiang-Dong Fu, and Fred H. Gage. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nature Structural & Molecular Biology*, 16(2):130–137, February 2009.

[4] Julian König, Kathi Zarnack, Nicholas M. Luscombe, and Jernej Ule. Protein-RNA interactions: new genomic technologies and perspectives. *Nature Reviews. Genetics*, 13(2):77–83, January 2012.

[5] Stefanie Gerstberger, Markus Hafner, and Thomas Tuschl. A census of human RNA-binding proteins. *Nature Reviews Genetics*, 15(12):829–845, December 2014. Number: 12 Publisher: Nature Publishing Group.

[6] Ibrahim Avsar Ilik, Jeffrey J. Quinn, Plamen Georgiev, Filipe Tavares-Cadete, Daniel Maticzka, Sarah Toscano, Yue Wan, Robert C. Spitale, Nicholas Luscombe, Rolf Backofen, Howard Y. Chang, and Asifa Akhtar. Tandem Stem Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. *Molecular cell*, 51(2):156–173, July 2013.

[7] Arumay Pal and Yaakov Levy. Structure, stability and specificity of the binding of ssDNA and ssRNA with proteins. *PLoS Computational Biology*, 15(4):e1006768, April 2019.

[8] Ruben D. Cadena-Nava, Mauricio Comas-Garcia, Rees F. Garmann, A. L. N. Rao, Charles M. Knobler, and William M. Gelbart. Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio. *Journal of Virology*, 86(6):3318–3326, March 2012. Publisher: American Society for Microbiology Journals Section: Structure and Assembly.

[9] Abby R. Thurm, Christian Beren, Ana Luisa Duran-Meza, Charles M. Knobler, and William M. Gelbart. RNA Homopolymers Form Higher-Curvature Virus-like Particles Than Do Normal-Composition RNAs. *Biophysical Journal*, 117(7):1331–1341, October 2019.

[10] Christian Beren, Lisa L. Dreesens, Katherine N. Liu, Charles M. Knobler, and William M. Gelbart. The Effect of RNA Secondary Structure on the Self-Assembly of Viral Capsids. *Biophysical Journal*, 113(2):339–347, July 2017.

[11] Ruth McBride, Marjorie Van Zyl, and Burtram C. Fielding. The Coronavirus Nucleocapsid Is a Multifunctional Protein. *Viruses*, 6(8):2991–3018, August 2014. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[12] Chung-ke Chang, Ming-Hon Hou, Chi-Fon Chang, Chwan-Deng Hsiao, and Tai-huang Huang. The SARS coronavirus nucleocapsid protein – Forms and functions. *Antiviral Research*, 103:39–50, March 2014.

[13] Paweł Grochowski and Joanna Trylska. Continuum molecular electrostatics, salt effects, and counterion binding–a review of the Poisson-Boltzmann theory and its modifications. *Biopolymers*, 89(2):93–113, February 2008.

[14] Thomas E. Angelini, Hongjun Liang, Willy Wriggers, and Gerard C. L. Wong. Like-charge attraction between polyelectrolytes induced by counterion charge density waves. *Proceedings of the National Academy of Sciences*, 100(15):8634–8637, July 2003. Publisher: Proceedings of the National Academy of Sciences.

[15] Itamar Borukhov, Robijn F. Bruinsma, William M. Gelbart, and Andrea J. Liu. Elastically Driven Linker Aggregation between Two Semiflexible Polyelectrolytes. *Physical Review Letters*, 86(10):2182–2185, March 2001. Publisher: American Physical Society.

[16] Peter L. Freddolino, Anton S. Arkhipov, Steven B. Larson, Alexander McPherson, and Klaus Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure (London, England: 1993)*, 14(3):437–449, March 2006.

[17] Chun Chen, Xinliang Zhao, Ryszard Kierzek, and Yi-Tao Yu. A Flexible RNA Backbone within the Polypyrimidine Tract Is Required for U2AF65 Binding and Pre-mRNA Splicing InVivo. *Molecular and Cellular Biology*, 30(17):4108–4119, September 2010. Publisher: American Society for Microbiology.

[18] Peter Tompa and Monika Fuxreiter. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*, 33(1):2–8, January 2008.

[19] Szabolcs Sipeki, Kitti Koprivanacz, Tamás Takács, Anita Kurilla, Loretta László, Virag Vas, and László Buday. Novel Roles of SH2 and SH3 Domains in Lipid Binding. *Cells*, 10(5):1191, May 2021.

[20] Monika Fuxreiter. Classifying the Binding Modes of Disordered Proteins. *International Journal of Molecular Sciences*, 21(22), November 2020. Publisher: Multidisciplinary Digital Publishing Institute (MDPI).

[21] Carlo Condello, William F. DeGrado, and Stanley B. Prusiner. Prion biology: implications for Alzheimer's disease therapeutics. *The Lancet Neurology*, 19(10):802–803, October 2020. Publisher: Elsevier.

[22] Patricia Santofimia-Castaño, Bruno Rizzuti, Yi Xia, Olga Abian, Ling Peng, Adrián Velázquez-Campoy, José L. Neira, and Juan Iovanna. Targeting intrinsically disordered proteins involved in cancer. *Cellular and Molecular Life Sciences*, 77(9):1695–1707, 2020.

[23] Tanja Mittag and Julie D. Forman-Kay. Atomic-level characterization of disordered protein ensembles. *Current Opinion in Structural Biology*, 17(1):3–14, February 2007.

[24] Veronika Csizmok, Isabella C. Felli, Peter Tompa, Lucia Banci, and Ivano Bertini. Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy. *Journal of the American Chemical Society*, 130(50):16873–16879, December 2008.

[25] Jason D. Perlmutter and Michael F. Hagan. Mechanisms of virus assembly. *Annual Review of Physical Chemistry*, 66:217–239, April 2015.

[26] Ju-Hui Liu, Kun Xi, Xi Zhang, Lei Bao, Xinghua Zhang, and Zhi-Jie Tan. Structural Flexibility of DNA-RNA Hybrid Duplex: Stretching and Twist-Stretch Coupling. *Biophysical Journal*, 117(1):74–86, July 2019.

[27] Gregory S. Couch, Donna K. Hendrix, and Thomas E. Ferrin. Nucleic acid visualization with UCSF Chimera. *Nucleic Acids Research*, 34(4):e29, February 2006.

[28] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera–a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004.

[29] Irina Tuszynska, Marcin Magnus, Katarzyna Jonak, Wayne Dawson, and Janusz M. Bujnicki. NPDock: a web server for protein-nucleic acid docking. *Nucleic Acids Research*, 43(W1):W425–430, July 2015.

[30] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J. C. Berendsen. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20291.

[31] Alex Plumridge, Steve P. Meisburger, Kurt Andresen, and Lois Pollack. The impact of base stacking on the conformations and electrostatics of single-stranded DNA. *Nucleic Acids Research*, 45(7):3932–3943, April 2017.

[32] Jiří Šponer, Judit E. Šponer, Arnošt Mládek, Petr Jurečka, Pavel Banáš, and Michal Otyepka. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers*, 99(12):978–988, 2013. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.22322.

[33] Andrey V. Dobrynin. Effect of Counterion Condensation on Rigidity of Semiflexible Polyelectrolytes. *Macromolecules*, 39(26):9519–9527, December 2006. Publisher: American Chemical Society.

[34] Lennart Piculell, Christer Viebke, and Per Linse. Adsorption of a Flexible Polymer onto a Rigid Rod. A Model Study. *The Journal of Physical Chemistry*, 99(48):17423–17430, November 1995. Publisher: American Chemical Society.

[35] J. Kierfeld and R. Lipowsky. Unbundling and desorption of semiflexible polymers. *Europhysics Letters*, 62(2):285, April 2003. Publisher: IOP Publishing.

[36] P.-G. de Gennes. Some conformation problems for long macromolecules. *Reports on Progress in Physics*, 32(1):187, January 1969.

[37] E. Eisenriegler, K. Kremer, and K. Binder. Adsorption of polymer chains at surfaces: Scaling and Monte Carlo analyses. *The Journal of Chemical Physics*, 77(12):6296–6320, December 1982.

[38] A. C. Maggs, D. A. Huse, and S. Leibler. Unbinding Transitions of Semi-flexible Polymers. *Europhysics Letters*, 8(7):615, April 1989.

[39] Jeffrey Skolnick and Marshall Fixman. Electrostatic Persistence Length of a Wormlike Polyelectrolyte. *Macromolecules*, 10(5):944–948, September 1977. Publisher: American Chemical Society.

[40] Theo Odijk. The statistics and dynamics of confined or entangled stiff polymers. *Macromolecules*, 16(8):1340–1344, August 1983.

[41] Margarita I. Zarudnaya, Iryna M. Kolomiets, Andriy L. Potyahaylo, and Dmytro M. Hovorun. Structural transitions in poly(A), poly(C), poly(U), and poly(G) and their possible biological roles. *Journal of Biomolecular Structure and Dynamics*, 37(11):2837–2866, July 2019.

[42] L. L. Bonilla, A. Carpio, and A. Prados. Theory of force-extension curves for modular proteins and DNA hairpins. *Physical Review E*, 91(5):052712, May 2015. Publisher: American Physical Society.

[43] Bharat Jagannathan and Susan Marqusee. Protein Folding and Unfolding Under Force. *Biopolymers*, 99(11):860–869, November 2013.

[44] Yong Wang and Giovanni Zocchi. Elasticity of Globular Proteins Measured from the ac Susceptibility. *Physical Review Letters*, 105(23):238104, December 2010. Publisher: American Physical Society.

[45] Alenka Luzar and David Chandler. Hydrogen-bond kinetics in liquid water. *Nature*, 379(6560):55–57, January 1996. Number: 6560 Publisher: Nature Publishing Group.

[46] Yong Wang and Giovanni Zocchi. Viscoelastic Transition and Yield Strain of the Folded Protein. *PLOS ONE*, 6(12):e28097, December 2011. Publisher: Public Library of Science.

[47] Carol A. Rohl, Wayne Fiori, and Robert L. Baldwin. Alanine is helix-stabilizing in both template-nucleated and standard peptide helices. *Proceedings of the National Academy of Sciences*, 96(7):3682–3687, March 1999. Publisher: Proceedings of the National Academy of Sciences.

[48] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophysical Journal*, 72(4):1541–1555, April 1997.

[49] D.b. Wetlaufer. Reversible and irreversible denaturation of proteins in chromatographic systems. *Makromolekulare Chemie. Macromolecular Symposia*, 17(1):17–28, 1988. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/masy.19880170104.

[50] F. R. Salemme. Cooperative motion and hydrogen exchange stability in protein -sheets. *Nature*, 299(5885):754–756, October 1982. Number: 5885 Publisher: Nature Publishing Group.

[51] Donald A. McQuarrie. *Statistical Mechanics*. University Science Books, June 2000. Google-Books-ID: itcpPnDnJM0C.

[52] Yong Wang and Giovanni Zocchi. The folded protein as a viscoelastic solid. *Europhysics Letters*, 96(1):18003, September 2011.

[53] Norman S. Nise. *Control Systems Engineering*. John Wiley & Sons, Inc., USA, 3rd edition, 2000.

[54] Gilad Haran and Hisham Mazal. How fast are the motions of tertiary-structure elements in proteins? *The Journal of Chemical Physics*, 153(13):130902, October 2020.

[55] Alan V. Oppenheim, Ronald W. Schafer, and Roland W. Schafer. *Discrete-time signal processing*. Always learning. Pearson, Harlow, third edition, pearson new international edition edition, 2014.

[56] Steve Pressé, Kingshuk Ghosh, Julian Lee, and Ken A. Dill. Principles of maximum entropy and maximum caliber in statistical physics. *Reviews of Modern Physics*, 85(3):1115–1141, July 2013. Publisher: American Physical Society.

[57] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, July 1948.

[58] Peter E. Wright and H. Jane Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*, 16(1):18–29, January 2015. Number: 1 Publisher: Nature Publishing Group.

[59] D. C. Gajdusek, C. J. Gibbs, and M. Alpers. Experimental transmission of a Kuru-like syndrome to chimpanzees. *Nature*, 209(5025):794–796, February 1966.

[60] G. C. Telling, P. Parchi, S. J. DeArmond, P. Cortelli, P. Montagna, R. Gabizon, J. Mastrianni, E. Lugaresi, P. Gambetti, and S. B. Prusiner. Evidence for the conformation of the pathologic isoform of the prion protein enciphering and propagating prion diversity. *Science (New York, N.Y.)*, 274(5295):2079–2082, December 1996.

[61] Siyu Li, Gonca Erdemci-Tandogan, Paul van der Schoot, and Roya Zandi. The effect of RNA stiffness on the self-assembly of virus particles. *Journal of Physics: Condensed Matter*, 30(4):044002, December 2017. Publisher: IOP Publishing.