THE UNIVERSITY OF CHICAGO

SELF ASSEMBLY WITH DNA: FROM MATERIALS DESIGN TO CHROMATIN

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE INSTITUTE FOR MOLECULAR ENGINEERING
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
JOSHUA PAUL LEQUIEU

CHICAGO, ILLINOIS
AUGUST 2017

For Amanda

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

No person has had a greater impact on this dissertation than my advisor Professor Juan de Pablo. Juan has had a profound impact on my development as a young scientist and I'm grateful for his mentorship throughout my graduate studies. For Juan, the core responsibility of a scientist is to teach, whether through courses, presentations, publications or proposals. He has pushed me to be a better scientist, a better teacher, and a better person. I will miss our regular interactions dearly.

I'm also grateful to former students and postdocs with whom I've worked closely throughout graduate school. Daniel Hinckley, Gordon Freeman, and Jonathan Whitmer deserve special thanks for serving as mentors during my first several years of graduate school, and for providing invaluable help with the nuts-and-bolts of research. In particular, I'm grateful for their encouragement to invest in skills and tools early on. I continue to reap the rewards of this advice and am grateful that they helped me get off on the proverbial "right foot". I'm also very grateful to Andres Cordoba who played a big role in teaching me about polymer dynamics and transport phenomena, and to Joshua Moller for his assistance with model development. All of these men have fought beside me in the "trenches" of research, and few pages in this dissertation do not reflect their help and advice.

It is sometimes said that, "it takes a village to raise a child" and I think the same can be said about a graduate student. I've been incredibly grateful for the many other outstanding members of the de Pablo research group who I've been privileged to work with. I've been humbled by the quality of individuals that have passed through the group, both in terms of intellectual depth and in quality of character. Our countless discussions have enriched me as a young scientist, and I will deeply miss being in close proximity to so many outstanding individuals. I would like to thank, in no particular order, Lucas Antony, Grant Garner, Yongrui Su, Aaron Fluitt, Gurdaman Khaira, Tyler Roberts, Kyle Hoffmann, Brandon Peters, Abhijeet Joshi, Sadanand Singh, Sumi Hur, Arnout Boeleans, Chi-cheng Chiu, Abelardo Ramirez-Hernandez, Marat Andreev, Rui Zhang, Hadi Ramezani-Dakhel,

Jian Qin, Julian Helfferich, Mohammad (Amin) Rahimi, Vikram Thapar, Mike Webb, Nick Jackson, Minirosadat (Sanaz) Sadati, Ivan Lyubimov, Yamil Colon, Jose (Adrian) Martinez-Gonzalez, Julio C. Armas-Pérez, Nader Taheri Qazvini, Daniel Reid, Johnny Alfaro, Weiwei Chu, Ye Zhou, Alec Bowen, Ashley Guo, Emre Sevgen, Jiyuan Li, Cody Bezik, Gustavo Andres Vazquez Montoya, Viviana Palacio Betancur, and lastly (but certainly not least) Kirk Swanson. Also thanks to my Wisconsin cohort-mates Anubhav Kushwaha, Luke Rolling and Tony Plauck.

I would also like to thank the numerous professors who have served as collaborators, mentors, or instructors throughout my graduate career. In particular, I am grateful to David Schwartz for his constant encouragement to "think big" and for graciously hosting me for a year in his laboratory. I'd also like to thank my committee members, James Skinner and Benoit Roux, for their feedback and advice throughout this dissertation. I'm also grateful to Juan Hernandez-Ortiz (UNalMed), Andrew Spakowitz (Stanford), Helmut Schiessel (Leiden), James Rawlings (UW), Mike Graham (UW), Brian Pfleger (UW), Sidney Nagel (UC), Heinrich Jaeger (UC) and Matt Tirrell (UC), for their help and guidance along the way.

The administrators at both Wisconsin and Chicago have taken care of so many things so that I could focus on doing research. Thank you to Sheri Severson (UW), Christi Balas Levenson (UW), Janet Boland (UC), Sandra Marijan (UC), Novia Pagone (UC) and Rovana Popoff (UC).

Lastly, I would like to thank my friends and family who have been a support throughout graduate school. To my parents, for always prioritizing my education and for giving me the tools and opportunities necessary to succeed. To Daniel Wilson, for our friendship, countless discussions, and for helping me see beyond the immediate demands of my work. Finally, to Amanda McMillan Lequieu, my wife and best friend, who has been the greatest supporter and cheerleader of this dissertation. She has been with me through the thick and the thin, and has truly made me a better man. SDG.

ABSTRACT

The ability to engineer the self-assembly of nano-scale objects to create highly ordered materials is of considerable scientific and practical interest. This new class of materials represents a powerful approach for engineering a next generation of devices, whose mechanical, optical, and electrical properties can be precisely tuned at the molecular scale. Though significant strides towards this goal have been achieved in recent years, the complexity achieved in engineered systems is still far surpassed by that achieved by nature. Precise self-assembly is achieved by nature through proteins and nucleic acids that fold into intricate, three-dimensional, and importantly, functional structures. A promising avenue towards improved engineered systems is to draw on discoveries from biophysics in order to inspire new approaches and paradigms for self-assembly and materials design. In this work, we explore the interplay between biophysics and engineering by exploring the self-assembly of DNA. Our discussion begins at the smallest length-scales of DNA, first by understanding the hybridization of DNA, and then at how hybridization can be used in materials to direct the self-assembly of gold nanoparticles. We report the first evidence of a tunable mechanical response in these assemblies, thereby suggesting the possibility of mechanical meta-materials constructed using DNA. Our discussion then proceeds to larger length scales, where we examine the biophysical processes that control the compaction of DNA into chromatin. Using a detailed molecular model, we explore the free energies and dynamics of smallest building block of chromatin, a protein-DNA complex called the nucleosome. Our results are in quantitative agreement with existing experimental measurements, and help to explain the molecular factors that dictate the first stages of DNA compaction into chromatin. Lastly, we present a multi-scale approach that can couple information across different length scales of chromatin in order to examine the folding of large regions DNA. By drawing on both the biophysics and engineering literature, the findings presented here suggest new approaches for materials design, and offer new paradigms for synthetic systems that seek to mimic the complexity achieved by nature.

CHAPTER 1

INTRODUCTION

The ability to engineer the self-assembly of nano-scale objects to create highly ordered materials is of considerable scientific and practical interest. By controlling the precise locations of molecular-scale building blocks, materials constructed via self-assembly promises to open up a new era of materials and devices, whose mechanical, optical, and electrical properties can be precisely controlled. Though an initial focus of this field has been to achieve materials with static properties, an ultimate goal is to design materials that are dynamic and responsive. Engineered materials with dynamic properties that can be modified by subtle optical, mechanical, or electrical cues could dramatically reshape the world in which we live, with applications ranging from medicine, to agriculture, to energy.

In many regards, the aim of these dynamic materials is to emulate the complexity that has been achieved by biology. Life itself is built upon molecular processes that interact collectively to dictate incredibly complex behavior, ranging from the function of entire organisms, down to the folding and function of a single protein. At the smallest length scales, these biological systems are truly molecular materials, whose function is achieved through mechanisms that far surpass those that that can currently be engineering in the lab. A promising avenue towards improved engineered systems is to draw on discoveries from biophysics in order to inspire new approaches and paradigms for self-assembly and materials design.

In this dissertation, we explore the interplay between biophysics and engineering by exploring the self-assembly of DNA. Though typically associated with genetics, the physics of the DNA molecule represents a rich and exciting example of biological assembly, and as we will demonstrate, has many applications to improved materials design. While much attention has been given to the so-called "protein folding" problem, a similar "DNA folding" problem exists, which aims to understand the complex processes that dictate the folding and compaction of DNA into chromatin. Chromatin is a complex, hierarchical material that spans many length scales, ranging the nano-meter dimensions of the DNA molecule itself,

to the micron dimensions of the eukaryotic nucleus. Nature has developed many processes that control the assembly of DNA, and these processes that are essential for achieving the precise, robust, and dynamic compaction of DNA characteristic of higher-order life. In this dissertation, we will use these many length scales within chromatin as an outline for our discussion, first beginning at the smallest length scales of DNA, and then progressing to larger and larger length scales of DNA assembly.

In Chapter 1, we examine the hybridization of two single strands of DNA into the well-known DNA double-helix, and examine how the underlying DNA sequence can impact these processes. We then demonstrate that this hybridization process can be used to understand a promising platform for materials design that uses DNA to dictate the self-assembly of gold nanoparticles. Though much previous research on these DNA-nanoparticle systems has focused on their phase behavior and assembly, little has been done to precisely characterize the pairwise interaction between particles. We present a detailed calculation of the association between DNA-nanoparticle conjugates using 3SPN.2, a coarse-grained model of DNA that accounts for molecular structure and base-pairing. We compare our results to those obtained experimentally using μm sized particles and analyze the free energy surfaces that characterize interparticle hybridization. Next, we study how free energies are influenced by three-body effects and their impact on particle association and melting. Lastly, we explore the observation by Park *et al.* [Nature **451**, 553 (2008)] that DNA-nanoparticle crystallization can be inhibited by the deletion of a single nucleotide. Using our model, we suggest that the role of this nucleotide is to disrupt frustration. This chapter is reproduced from Ref. [94]

In Chapter 2, we extend this work in order to examine how DNA-conjugated nanoparticles can be used to design engineered materials with predetermined, and tunable properties (reproduced from Ref. [91]). By using DNA to encode nanoscale interactions, macroscale crystals can be formed with mechanical properties that can, at least in principle, be tuned. We present *in silico* evidence that the mechanical response of these assemblies can indeed

be controlled, and that subtle modifications of the linking DNA sequences can change the Young's modulus from 97 kPa to 2.1 MPa. We rely on a detailed molecular model to quantify the energetics of DNA-nanoparticle assembly and demonstrate that the mechanical response is governed by entropic, rather than enthalpic contributions and that the response of the entire network can be estimated from the elastic properties of an individual nanoparticle. The results here provide a first step towards the mechanical characterization of DNA-nanoparticle assemblies, and suggest the possibility of mechanical metamaterials constructed using DNA.

Our discussion then proceeds to larger length scales, where we examine the biophysical processes that control the compaction of DNA into chromatin. In Chapter 3, we turn our attention to the nucleosome, a DNA-protein complex that is central to chromatin structure and function (reproduced from Ref. [92]). Nucleosomes form the basic unit of compaction within eukaryotic genomes and their locations represent an important, yet poorly understood, mechanism of genetic regulation. Quantifying the strength of interactions within the nucleosome is a central problem in biophysics and is critical to understanding how nucleosome positions influence gene expression. By comparing to single-molecule experiments, we demonstrate that a coarse-grained molecular model of the nucleosome can reproduce key aspects of nucleosome unwrapping. Using detailed simulations of DNA and histone proteins, we calculate the tension-dependent free energy surface corresponding to the unwrapping process. The model reproduces *quantitatively* the forces required to unwrap the nucleosome, and reveals the role played by electrostatic interactions during this process. We then demonstrate that histone modifications and DNA sequence can have significant effects on the energies of nucleosome formation. Most notably, we show that histone tails contribute asymmetrically to the stability of the outer and inner turn of nucleosomal DNA and that depending on which histone tails are modified, the tension-dependent response is modulated differently.

In Chapter 4, we then examine the dynamics of the nucleosome, and the factors that dictate their position along the genome (reproduced from Ref. [93]). Previous work has shown that the locations of nucleosomes across the genome are not random, but instead

depend on both the underlying DNA sequence and the dynamic action of other proteins within the nucleus. These processes are central to cellular function, and the molecular details of the interplay between DNA sequence and nucleosome dynamics remain poorly understood. In this work we investigate this interplay in detail by relying on a molecular model, which permits development of a comprehensive picture of the underlying free energy surfaces and the corresponding dynamics of nucleosome repositioning. The mechanism of nucleosome repositioning is shown to be strongly linked to DNA sequence, and directly related to the binding energy of a given DNA sequence to the histone core. It is also demonstrated that chromatin remodelers can override DNA-sequence preferences by exerting torque, and the histone H4 tail is then identified as a key component by which DNA-sequence, histone modifications and chromatin remodelers could in fact be coupled. This chapter is reproduced from Ref. [93].

In Chapter 5, we progress to yet even larger length scales, and examine the collective processes that control the interaction of many nucleosomes together, and the resulting folding of DNA into chromatin. The compaction of eukaryotic DNA in to chromatin represents an essential, yet poorly understood mechanism for controlling gene expression. Errors during compaction are associated with numerous diseases, and therefore elucidating the molecular factors that control compaction is a central goal of genetics. A significant challenge in studying chromatin compaction is the many length scales involved, typically ranging from angstroms to microns. We present a multi-scale approach that couples a detailed molecular model of the nucleosome to a coarse-grained mesoscale model of chromatin. We show that this approach can reproduce the dynamics and thermodynamics of available experimental measurements, and is computationally efficient enough to examine the self-assembly of large regions of chromatin. Notably, we show that subtle features of the chromatin fiber, such as the underlying DNA sequence, can have a significant impact on the stability and accessibility of different chromatin structures. The work presented here represents important steps toward understanding the molecular processes that dictate the dynamics and structure of chromatin.

Finally in Chapter 6 we present an overview of our findings. In sum, the main conclusion of this dissertation is that the biological processes that dictate the compaction of DNA are wonderfully complex, yet comprehensible using existing techniques. There is therefore a large opportunity to draw from both the biophysics and engineering literature, in order to gain inspiration for new approaches for materials design, and achieve new paradigms for synthetic systems that seek to mimic the complexity achieved by nature.

CHAPTER 2

A MOLECULAR VIEW OF DNA-CONJUGATED NANOPARTICLE ASSOCIATION ENERGIES

2.1 Abstract

Nanoparticles functionalized with short sequences of DNA represent a promising platform for customizable self assembly. Though much recent research has focused on the phase behavior and assembly of these structures, little has been done to precisely characterize the pairwise interaction between particles. Here we present a detailed calculation of the association between DNA-nanoparticle conjugates using 3SPN.2, a coarse-grained model of DNA that accounts for molecular structure and base-pairing. We compare our results to those obtained experimentally using μm sized particles and analyze the free energy surfaces that characterize interparticle hybridization. Next, we study how free energies are influenced by three-body effects and their impact on particle association and melting. Lastly, we explore the observation by Park *et al.* [Nature **451**, 553 (2008)] that DNA-nanoparticle crystallization can be inhibited by the deletion of a single nucleotide. Using our model, we suggest that the role of this nucleotide is to disrupt frustration.

2.2 Introduction

The ability to precisely control the self assembly of μm or nm particles represents one of the grand challenges for colloidal and nano-science. Colloids and nanoparticles provide a promising platform for highly customizable materials whose mechanical, optical and magnetic properties could be precisely tuned[161]. One powerful approach to achieve colloidal and nanoparticle assembly relies on short sequences of DNA to control the inter-particle interactions[116, 2]. By covalently binding strands of DNA to the surface of a gold nanoparticle, the sequence specific base-pairing of DNA is utilized to mediate particle-particle inter-

actions. By modifying the composition and length of the conjugated DNA sequences, both the length and energy scales of the interactions can be specified. Further, because interactions are governed by DNA (and not nanoparticle properties), this approach is amenable to tertiary interactions that are difficult to achieve using other approaches.

The flexibility of this design has led to important successes. Starting with the initial observations that DNA-nanoparticles could assemble into crystals with long-range order [135, 129], recent studies have demonstrated the formation of a wide variety of super-lattices with intricate symmetries[107]. Most recently, Auyeung *et al.* have demonstrated that with sufficiently slow cooling, DNA-nanoparticles will assemble with well defined facets and a uniform crystal habit[10] Combined with the ever increasing toolkit of particle functionalities[70, 202, 204], this discovery represents an opportunity to use DNA-nanoparticle conjugates to gain insights into crystallization processes occurring at the atomic scale.

In spite of this recent progress, several challenges remain. The development of protocols to assemble structures exhibiting long range order took nearly a decade of work, with most early systems forming disordered aggregates[134]. Several studies have now demonstrated that subtle variations in DNA sequence can disrupt ordered assembly. For example, Nykypanchuk *et al.* studied five different DNA sequences that differed only in their unreactive "linker" regions[129]. Ordered assembly was only achieved for two of these sequences. More dramatically, Park et al. [135] demonstrated that the inclusion of a single additional base pair (the so-called "spacer base") can enhance or impair ordered assembly[135]. In addition to sequence effects, experiments indicate that ordered assembly only occurs in a narrow temperature range[108]. Further, the ability to form faceted crystals is strongly dependent on the relative surface energies of the crystals. Auyeung *et al.* observed that though bcc systems form well-defined facets, fcc systems do not[10]. These nuanced effects have motivated on-going computational and theoretical efforts that seek to elucidate the physical origins of this behavior.

DNA-nanoparticle assembly is an inherently multi-scale process. The macroscopic assem-

bly of thousands of nanoparticles occurs at the μm scale, but particle-particle interactions depend on DNA-mediated contacts occurring at lengths of nanometers. Accordingly, recent work has relied on a multi-stage approach that develops models with various levels of detail depending on the property of interest. Using experimentally obtained pair-potentials of colloidal sized particles[15], Scarlett *et al.* used simulation to study the mechanism of binary superlatice formation[155]. Li *et al* demonstrated that a scale-accurate coarse-grained model could reproduce experimental phase diagrams [96] and then proceeded to determine the kinetics and DNA properties necessary for optimal assembly [97]. This model was then used to inform the pair-potentials used in a coarsened model that predicted the assembly of faceted crystals for BCC but amorphous crystals for FCC [10]. Mladek *et al.* started with a highly-detailed model of a DNA-nanoparticle to inform an increasingly coarse-grained "core-blob" model[118, 119]. Using this core-blob model, they calculated the free energy of various crystal structures and obtained results in excellent agreement with experiment. The core-blob model was then used to parameterize an even coarser pair-potential model for DNA-nanoparticle assembly.

Though these approaches have provided significant insights into DNA-nanoparticle systems, little work has been done to precisely characterize the smallest length scales of the assembly process; in particular, the pairwise assembly of individual particles. Current approaches rely on custom-built models to mimic DNA-nanoparticle assembly which have not been validated in terms of their ability to describe DNA rehybridization[178, 119]. In these models, many base-pairs are coarse-grained into one bead, thereby preventing study of subtle sequence effects such as those observed by Park et al. [135]. Further, these models typically include base pairing energy as a parameter that is hard-coded into the model, and is changed manually for different temperatures and DNA sequences[119]. Because the base-pairing energy is an input parameter, the use of these models to precisely characterize the energy scale of DNA-nanoparticle interactions should be viewed with caution. Experimentally, the pair-potential for micron sized particles has been determined [15, 150]. These results, however,

are of limited applicability to nanoscale assembly.

In this work, we present a high-precision characterization of DNA-nanoparticle pairwise interactions. We adapt a widely used coarse-grained DNA model[59] for study of DNA-nanoparticle systems to represent DNA with single base resolution. This model is used to calculate the pair-potential between particles and the details of strand-strand interactions. We recapitulate experimentally observed melting behavior and calculate the energy barriers of particle association. We compare our pair-potentials to those calculated experimentally using μm sized particles and explore the balance of entropy and enthalpy at these different length scales. Lastly, we use the sequences of Park et al. [135] to examine the impact of the so-called "spacer-base" on the inter-particle pair-potential.

2.3 Methods

2.3.1 DNA-nanoparticle Model

The model employed here is represented schematically in Fig. 2.1. For DNA, we adapt 3SPN.2[59], the latest model in the 3SPN family[82, 154], for use in conjunction with nanoparticles. In 3SPN.2, three sites are used to represent each nucleotide. These are located at the phosphate, the sugar, and the base, respectively. In addition to correctly capturing structural properties (e.g. helix width, major and minor groove widths, etc.) 3SPN.2 correctly captures several aspects of DNA physics relevant to DNA-nanoparticle assembly. Firstly, 3SPN.2 can naturally describe both single and double stranded DNA and has been used previously to study the hybridization of short oligomers[60]. The study of strand hybridization in DNA-nanoparticles (where many strands are present) represents a natural progression of those efforts. The ability to describe both forms of DNA is also important for modeling the wide range of linkers that are employed experimentally (see Fig. 2.1b). Further, 3SPN.2 was rigorously parameterized to describe experimental melting temperatures (of both hairpins and duplexes) as a function of both salt and sequence. The ability to accurately

predict the thermodynamics as a function of sequence is particularly relevant when comparing the energy scales that arise in DNA-nanoparticle systems. For these reasons, we expect 3SPN.2 to predict *quantitatively* the properties of pairwise DNA-nanoparticle assembly.

To model the nanoparticle we use an icosphere, which we create by first subdividing each triangular face of an icosahedron into four equilateral triangles. The locations of the new vertices are then mapped radially onto the surface of a sphere. Subsequent recursions on the resulting triangles can be used to increase the resolution of the sphere arbitrarily. This approach, frequently used in computer graphics, leads to an approximately constant density of points on a sphere. The effective radius of each nanoparticle site is a function of the site density and is discussed in the Appendix. In our model, each vertex is represented by a force site and is bonded to each of its nearest neighbors. Thus each site is effectively fixed at its relative location on the surface of the sphere (i.e. sites cannot diffuse). We include an additional site at the center of the sphere that is bonded to all surface sites to increase nanoparticle rigidity (Fig. 2.1c).

In order to approximate the thiol group used experimentally to covalently tether DNA to the nanoparticle surface, we use a series of force sites that bridge a nanoparticle site and the sugar site on the DNA (Fig. 2.1b). The functional form of the bonded and nonbonded interactions of nanoparticle and linkers are identical to those introduced in 3SPN.2; the parameters for these interactions are given in the Appendix. With the exception of the bonded interactions with the thiol tether, DNA sites only interacts with these added sites via excluded volume interactions.

In general, the DNA sequences used in DNA-nanoparticle assembly consist of an inert "linker" region and a reactive "sticky end" region. The "sticky ends" are different for each particle type and are chosen to be complementary in order to induce an attractive interaction between particles. The "linker" region projects the "sticky end" away from the surface of the nanoparticle and is thought to roughly control particle-particle separation. In this study, the "sticky ends" are always chosen to be identical to the experiment of interest. The "linker"

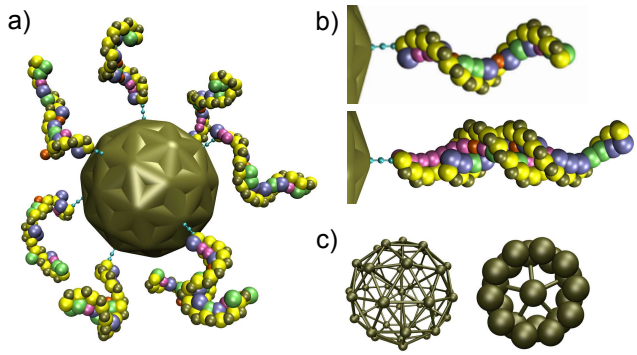

Figure 2.1: DNA-nanoparticle model. a) Representative configuration of 5 nm particle with the 15 base pair Rogers8-7 sequence. b) Both single and double-stranded DNA sequences can be represented explicitly with 3SPN.2. The top snapshot is the Rogers8-7 sequence, the bottom is Park7. DNA sequences are tethered to the nanoparticle via a short covalent linker (light blue beads). c) Nanoparticles are represented by interaction sites placed evenly along the surface of the sphere and a single site at the center.

region, however, is scaled to account for differences between experimental and simulated particle diameters. This will be discussed as necessary throughout this manuscript. The number of strands per particle was specified to be 19 pmol/cm^2 (unless otherwise noted) as shown experimentally[64]. The DNA strands were distributed uniformly across the particle surface.

All simulations were performed in the NVT ensemble using a Langevin thermostat. The Debye-Hückel approximation was used to model the interactions between phosphate sites which carry a charge of -0.6 (Ref. [59]). All simulations were performed using a 20 fs timestep and 200 mM ionic strength. At 200mM, the Debye length is $\sim 6-7\mathring{A}$ and represents a small, but not vanishing electrostatic repulsion between the negatively charged phosphates. 200mM was chosen as a intermediate value between the wide range of salt concentrations used experimentally.

2.3.2 Free Energy Calculation

Free energy surfaces were calculated using metadynamics[89]. Metadynamics is an adaptive technique where the trajectory of a molecular dynamics simulation is biased by penalizing

subsequent visits to points in state space. In this way, metadynamics forces a system away from energetic minima and efficiently samples all of phase space. The cumulative bias can then be used to estimate the free energy of the system along a given order parameter.

In our simulation of DNA-nanoparticles we bias along two order parameters simultaneously. The first, denoted by r, is the separation distance between the center site of each nanoparticle. The second, denoted by U_{bp}, is the attractive energy contribution from base-pairing interactions (i.e. hydrogen bonds between complementary DNA bases). By using only the attractive energy, we decrease the degeneracy of the U_{bp} order parameter. This resulted in faster convergence and more accurate results. The energy U_{bp} decreases linearly with increasing base-pair formation and thus, $-U_{bp}$ (i.e. "minus" U_{bp}) increases as the number of base pairs increases. The choice of U_{bp} is motivated by two considerations. One is practical: the formation and disruption of base-pairing events proceeds on a slow time scale, with the energy difference between hybridized and unhybridized states separated by many k_BT. To accelerate the convergence of our pair-potentials, it is necessary for metadynamics to bias along this slow mode of relaxation. The second consideration is scientific. By calculating the free energy along the base-pairing coordinate, we gain access to information about individual DNA strand configurations and the free energy differences between them. This information cannot be obtained experimentally and is important for understanding the details of particle-particle interactions. The details of the information gained by U_{bp} will be explored in Section 6.5.

Throughout this work, one order parameter of the 2-dimensional free energy surface is routinely "averaged over" in order to obtain a single dimensional free energy surface. To do this, we first convert the free energy into a probability via $P(r, U_{bp}) = e^{-A(r,U_{bp})/k_BT}/Q$, where Q is a normalizing factor. We then calculate an average for the order parameter of interest using

$$P(r) = \frac{1}{-U_{bp,min}} \int_{U_{bp,min}}^{0} P(r, U_{bp})\, \mathrm{d}U_{bp}, \tag{2.1}$$

where $U_{bp,min}$ is a negative number corresponding to the minimum base pairing energy

accessed by the system. Finally, the probabilities are converted back to free energies by $A(r) = \frac{1}{k_B T}\log(P(r))$. Performing the averages over the probabilities instead of the free energies reduces the statistical noise.

An analogous procedure is used to calculate $A(U_{bp})$ by integrating Eq. 2.1 over dr from 0 to r_{max}. The distance r_{max} corresponds to the maximum separation experienced by the particles. For computational efficiency, r_{max} is chosen to be slightly larger than the value where the interaction energy between particles vanishes. An artificial repulsive wall is applied at r_{max} to ensure that the system only samples configurations of interest.

The calculation of the free energy as a function of separation, $A(r)$, requires that the Jacobian corresponding to the transformation from Cartesian coordinates be included[74]. In three dimensions, this corresponds to including a factor of $2k_B T\log(r)$ in $A(r)$. Because $A(U_{bp})$ does not correspond to a coordinate change, no correction for this order parameter is necessary.

Metadynamics simulations were performed with multiple walkers [142]. Convergence was measured by comparing the evolution of the free energy surface throughout the simulation. To speed convergence, each walker was initialized to a different position in the 2-d order parameter space. The Gaussian height was 0.1 kJ/mol, the width was 0.25 $\mathring{A}$ and 20 kJ/mol. Deposition frequency was varied slightly to improve convergence (or to increase result accuracy). Typically, Gaussians were added every 5000 time steps or every 100 fs. When summing hills to obtain the entire free energy surface, averages included all deposited Gaussians. The so-called "inversion condition" was used to avoid boundary effects [27].

An analogous procedure is used to calculate the effect of three-body interactions described in Section 2.4.2. The only difference is that the order parameter r instead represents the side of the equilateral triangle containing the three particles (see Fig. 2.5a). As with the two-body case, U_{bp} represents the total attractive energy contribution from base-pairing interactions. In practice, the equilateral triangle is maintained by placing harmonic restraints on the third particle to ensure that it is equidistant from the original two particles. The equilibrium

length of these restraints is then adjusted dynamically as the distance between the two initial particles vary. This maintains the desired equilateral triangle with a variable edge length.

2.3.3 Simulated Melting Curve

Experimentally, the melting curve of a DNA-nanoparticle is calculated by monitoring the absorbance of an ensemble of DNA-nanoparticles as a function of temperature[39]. Bound and unbound DNA-nanoparticles have different absorbances, and the system exhibits a sharp step-wise transition between these two states. The melting temperature is calculated as the maximum in the derivative of this transition.

Computationally, we use the free energy surfaces described in Section 2.3.2 to calculate the probability of two particles being in the bound or unbound state. For an ensemble of DNA molecules, this represents the percentage of molecules in each state and is a close analogue to the experimentally-observed change in absorbance. We first calculate free energy surfaces of a nanoparticle system over a range of temperatures. On each free energy surface, we calculate the saddle point that separates the bound and unbound nanoparticle basins. Since, the free energy as a function of U_{bp} is known, the bound basin can be represented by values of $U_{bp} < U_{saddle}$ and the unbound basin by $U_{bp} > U_{saddle}$. This saddle point, U_{saddle}, is well defined as the first local maximum in $A(-U_{bp})$ that separates these two states and is shown pictorially by the horizontal dotted line in Fig. 5.3. The probability of hybridization, $P_{hybridized}$, can then be calculated as

$$P_{hybridized} = \frac{1}{Q} \int_{U_{bp,min}}^{U_{saddle}} \int_{0}^{r_{max}} e^{-A(r,U_{bp})/k_B T} \mathrm{d}r \, \mathrm{d}U_{bp}. \tag{2.2}$$

$P_{dehybridized}$ is calculated in an analogous manner by performing the outer integral in Eq. 2.2 from U_{saddle} to 0. In both cases, Q is a normalizing factor obtained by performing the outer integral from $U_{bp,min}$ to 0.

As in Ref. [59], we correct for fluctuations in concentration that are not captured by a

simulation with two molecules[131]. The corrected fraction of hybridized DNA, $f_{hybridized}$, is given by

$$f_{hybridized} = -\frac{1}{2\Phi} + \sqrt{\left(1 + \frac{1}{2\Phi}\right)^2 - 1}, \tag{2.3}$$

where $\Phi = \frac{P_{hybridized}}{P_{dehybridized}}$. We then define the melting temperature, T_m, as the temperature where $f_{hybridized} = 0.5$. To find T_m, we fit the simulation results to a sigmoid function, where

$$f_{hybridized}(T) = \frac{1}{1 + e^{\alpha(T - Tm)}}, \tag{2.4}$$

where α is the width of the function.

The melting temperature of bulk DNA was calculated as described in Ref. [59]. This approach is analogous to the method described here, except that free energy surfaces are calculated between two strands of ssDNA. To facilitate comparison of the melting behavior of DNA-nanoparticles to that of bulk DNA, we scale the concentration of bulk DNA to be equal to the probe density in the DNA-nanoparticle case. or a detailed description of the order parameters used and the method of scaling the concentration, interested readers are referred to Ref. [59].

2.4 Results

The results of Rogers et al. [150] represent one of the only experimental measurements of the pair-potential between DNA-coated particles[150]. The particles used in that work, however, measured 1 μm in diameter - two orders of magnitude larger than the particles used in most demonstrations of assembly[135, 129, 111]. Micron-sized particles coated with DNA differ from nano-sized particles in several regards, including DNA surface density, particle curvature, and length of DNA sequences. Since the impact of these differences is not well understood, we first seek to compare the results of Rogers et al. [150] to the pair-potential obtained here between nano-sized particles.

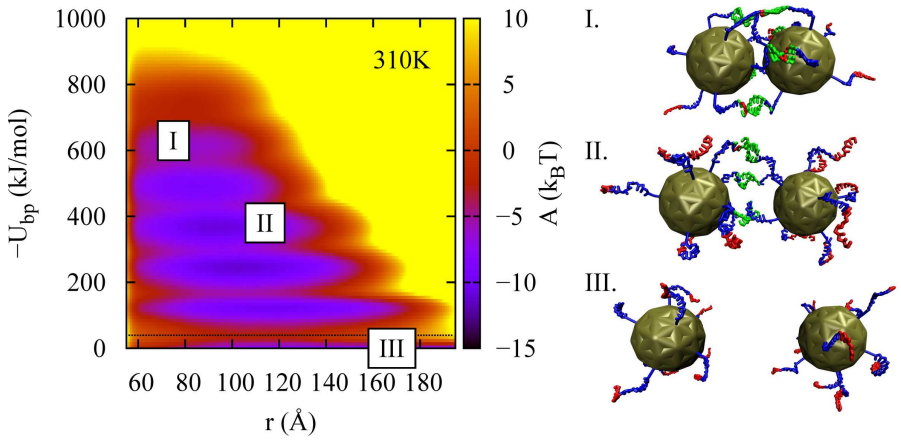

Figure 2.2: Free energy surface for Rogers8-7 at 310 K. The free energy is given as a function of both particle separation, r, and base pairing energy, U_{bp}. Lower values of base pairing energy (or increasing values of $-U_{bp}$) correlate directly with increasing hybridization between particles. Representative snapshots I, II, and III (and their position on the free energy surface) are shown on the right. The inert "linker" regions are colored blue and the reactive "sticky end" regions are colored red. Hybridized (i.e. reacted) sticky ends are colored green for clarity. The horizontal dotted line denotes the saddle point used to calculate the melting temperature.

In our simulation setup, we use sticky ends identical to Rogers et al. [150]. Since our particles are much smaller than those used experimentally, we truncate our DNA linkers by excluding the (T_{50}) spacer region. The results presented here are for DNA sequences consisting of an 8-base pair "linker" and a 7-base pair "sticky end". This sequence is denoted as "Rogers8-7" and is shown in Table 2.1. The effect of the truncation of the (T_{50}) spacer will be explored in Section 2.4.1. We use 5 nm diameter particles conjugated to nine strands of DNA, corresponding to a DNA density of 19 nmol/cm^2 as calculated experimentally[64]. The radius of gyration, R_g, for this system is 5.6 nm.

To characterize this system, we first calculate a two-dimensional free energy surface for the Rogers8-7 sequence as a function of particle separation, r, and base-pairing energy, U_{bp} (Fig. 5.3). Three configurations are shown for illustrative purposes with the "sticky end" bases colored in red and the "linker" bases colored in blue. For emphasis, "sticky end" bases are recolored green if they are successfully forming a base pair. Moving left to right (I to II to III) along the x-axis (r) corresponds to increasing inter-particle separation. Moving bottom

Table 2.1: DNA sequences used in this work. "NP" denotes nanoparticle, underlined bases represent complementary sticky ends, the bold "A" in the Park7 represents the "spacer base".

Name	Sequence
Rogers8-7	NP-TAATGCCTGTCTACC
	NP-TGAGTTGCGGTAGAC
Rogers3-7	NP-CCTGTCTACC
	NP-TGCGGTAGAC
Rogers13-7	NP-TTTTTTAATGCCTGTCTACC
	NP-TTTTTTGAGTTGCGGTAGAC
Park7	NP-AAAAAAAGACACAA
	TTCTGTGTT**A**TTTCCTT
	NP-AAAAAAAGACACAA
	TTCTGTGTT**A**AAGGAAA

to top (III to II to I) along the y-axis ($-U_{bp}$) corresponds to the formation of additional base pairs formed between the complementary sticky ends. These two order parameters are coupled; depending on the value of r, different values of U_{bp} are energetically accessible. For example, if $r = 170\,\mathring{A}$ (i.e. Configuration III), it is energetically unfavorable to form many base pairs (i.e. $-U_{bp} > 200$ kJ/mol). On the other hand if $r = 75\,\mathring{A}$ (i.e. Configuration I), the formation of many base pairs can be achieved with a small change in free energy.

One of the hallmark features of DNA-nanoparticle assembly is its temperature dependence. By changing the temperature by several degrees, the strength of interaction between particles can be strongly modulated. To examine the dependence on temperature, we calculate free energy surfaces for a range of temperatures using the Rogers8-7 sequence. Though each calculation results in a 2-D free energy surface similar to that shown in Fig. 5.3, the result is easier to visualize if averaging is performed as described in Section 2.3.2. The resulting curves curves $A(r)$ and $A(-U_{bp})$ are shown in Figs. 2.3a and b, respectively.

The temperature dependence of the pair-potential, $A(r)$, demonstrates several distinct trends (Fig. 2.3a). First, there is a significant change in the energy scale of the interaction.

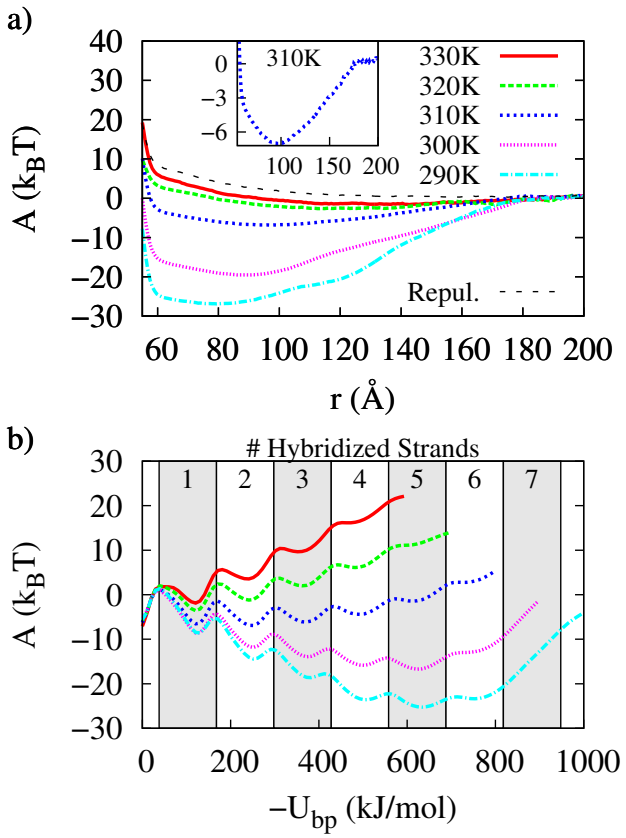

Figure 2.3: Temperature dependence of Rogers8-7 System. a) Free energy as a function of particle separation, r. Particle-particle interactions are strongly temperature dependent. The black dotted line, labeled "Repul", represents the potential for two particles containing the same DNA sequence. Since sequences are not self-complementary, this represents the repulsive contribution to the pair-potential. b) Free energy as a function of base-pairing, U_{bp}. Interactions between particles are the result of a balance between entropic and enthalpic contributions. The transition between additional hybridization events is well defined.

As temperature is changed, the strength of the pairwise interaction shifts from a $-25\,k_BT$ attractive minimum (290 K) to a net repulsion (330 K). Temperature also has a strong effect on the length scale of interaction. As temperature decreases, the energetic minima (r_{eq}) shift to lower values of r (e.g. $79\,\mathring{A}$ at 290 K, $92\,\mathring{A}$ at 300 K, $99\,\mathring{A}$ at 310 K). The qualitative features of the pair-potential also change. At higher temperatures (i.e. ≥ 310 K) the pair-potential is fairly simple: an attractive minimum and a repulsive wall that could be fit to a simple analytical potential. At low temperatures (i.e. < 310 K), significant shoulders emerge in the pair-potential.

Before we compare our simulated pair-potential with that observed by Rogers et al. [150], it is useful to examine the results obtained by averaging $A(r, U_{bp})$ over r to obtain $A(U_{bp})$ (see Fig. 2.3b). In general, the formation of base pairs (i.e. increasing $-U_{bp}$) results in an enthalpic gain and corresponds to an energetic minimum. Formation of base pairs, however, comes at an entropic cost as strand mobility is severely reduced upon hybridization. Fig. 2.3b demonstrates the balance between these two terms. At high temperatures (e.g. > 320 K), entropy dominates and the free energy increases approximately monotonically with the formation of additional base pairs. This behavior is in contrast with that observed at intermediate temperatures (e.g. ~ 310 K), where these terms nearly balance, and the free energy is approximately constant. At low temperatures (e.g. < 300 K), enthalpy dominates and free energy decreases with additional hybridization events. Note that even at low temperatures, however, the free energy only correlates with base pair formation up to a point. Since the particles exhibit curvature, additional hybridization events require DNA to stretch, resulting in an enthalpic penalty that eventually overwhelms the gains achieved by base-pair formation.

These observations are helpful for understanding the features of Fig. 2.3a. At low temperatures, additional hybridization events stabilize the interactions between particles resulting in a deeper energetic minimum. Additional base pairs require the particles to be closer together (so as to minimize DNA "stretching") and thus yield minima at lower values of r. As

temperature increases, entropy plays a larger role, the strength of interaction decreases, and the locations of the minima move to larger values of r.

2.4.1 Comparison of Nanometer and Micrometer Particles

We can now compare the simulated pair-potential in Fig. 2.3a with those measured experimentally. For sticky ends identical to those used in our simulations, Rogers et al. [150] observe a strength of attraction of $6\,k_BT$ at 303.5 K. The inset in Fig. 2.3a shows that for the nanoparticle system, a minimum of $\sim 6\,k_BT$ is obtained at 310 K, namely 6.5 K higher. At 300 K, the attraction between nano-sized particles is $\sim 19\,k_BT$. Thus, at a given temperature, the attraction between a nano-sized system is greater than that between the μm system. To explore these differences, we examine the impact of two system parameters.

One parameter that differs between nano and micron-sized systems is the number of DNA strands that form inter-particle contacts. As the curvature of the particles decrease (i.e. particle diameter increases), the number of strands that can associate between particles increases. To further examine this effect, we change the density of DNA on the 5nm particle, ρ_{DNA}, to $\sim 50\%$ ($10\,\mathrm{pmol/cm^2}$) and $\sim 200\%$ ($36\,\mathrm{pmol/cm^2}$). The results are shown in Fig. 2.4a. As ρ_{DNA} increases, the strength of interaction increases as the number of hybridization events increase. Further, the equilibrium separation, r_{eq} shifts to slightly larger values of r. Increasing ρ_{DNA} results in a denser DNA brush that leads to repulsions that favor larger particle-particle separations.

Another parameter that differs between these two systems is the length of the DNA linker regions. Though our simulations use identical sticky ends to those of Rogers et al. [150], our linker DNA regions are much shorter. Accordingly, we can vary the length of the linker bases, $n_{linker\,bases}$, and observe the effect on the pair-potential (Fig. 2.4b). These sequences, denoted Rogers13-7 and Rogers3-7 for $n_{linker\,bases} = 13$ and 3, are shown in Table 2.1.

As expected, increasing $n_{linker\,bases}$ increases r_{eq}. As the radius of gyration of the particles increases ($R_g = 51\,, 56\,, 62\,\mathring{A}$ for $n_{linker\,bases} = 3, 8, 13$, respectively) the equilibrium

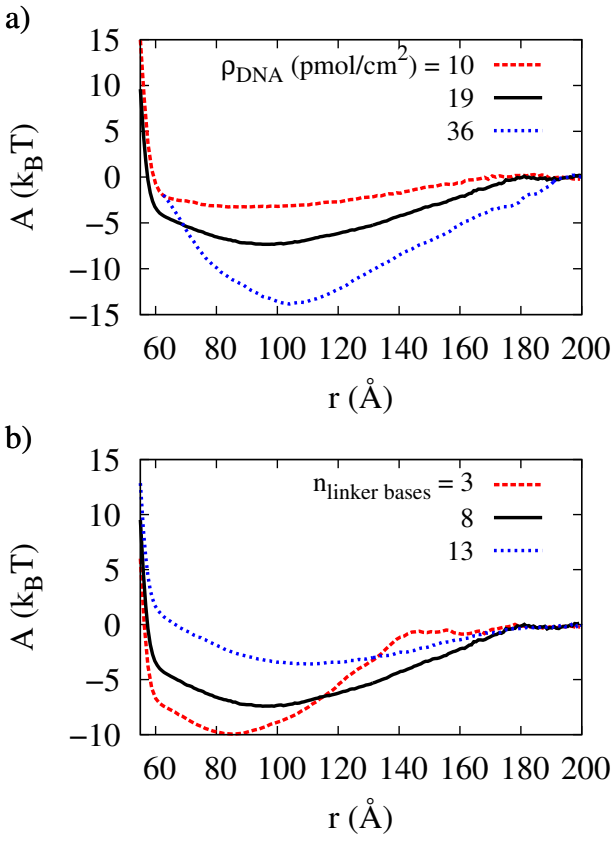

Figure 2.4: a) Effect of increasing DNA density, ρ_{DNA}, on effective pair-potential. The enthalpic gain from hybridization events results in stronger particle-particle association for higher DNA densities. b) Effect of DNA linker length, $n_{linker\,bases}$. The entropic penalty of strand hybridization increases with increasing $n_{linker\,bases}$. This results in weaker particle-particle association for longer linkers. The black lines in both panels are identical to the 310 K curve in Fig.2.3a.

separation increases. A less obvious consequence is that increasing $n_{linker\,bases}$ leads to a decrease in the minimum energy of interaction. Though longer linkers increase the "range" of each strand interaction and enable additional hybridization events, longer linkers also incur a larger entropic cost, leading to weaker interactions between particles.

Our observation that nanoparticles bind more strongly than colloidal particles can again be understood as a balance between enthalpic and entropic factors. On the one hand, the increased curvature of nano-sized particles leads to fewer DNA contacts (and thus weaker association). On the other, the shorter linkers used in nanoparticle systems minimize the entropy lost due to hybridization and result in stronger association. For the systems studied here, the entropic penalty dominates and nano-sized particles associate more strongly. Note however, that association is strongly temperature dependent — only near the melting temperature, where enthalpy and entropy are balanced, will such phenomena be apparent.

2.4.2 Three-body interactions

An additional difference between nano and micron-sized systems is the length of the linking DNA sequences relative to the nanoparticle radius. In micron-sized systems, the DNA linkers are short relative to the particle size and the radius of gyration of the coated-particle, R_g, is approximately equal to the radius of the uncoated particle, $R_{particle}$. In this regime, the range of DNA-mediated interactions between particles is short and many-particle interactions can be approximated as pair-wise additive. In other words, many-body effects do not play a significant role.

In nano-sized systems however, this is not the case. In these systems, the length of DNA linkers is comparable to the size of the particle (i.e. $R_g \sim 2 - 3R_{particle}$) and many-body effects can become important. This situation arises because the number of DNA strands available for hybridization is finite and different particles must compete for binding sites on complementary particles. This competition prevents interactions from being strictly pair-wise additive; any study of nano-sized DNA-particle assembly must account for these effects.

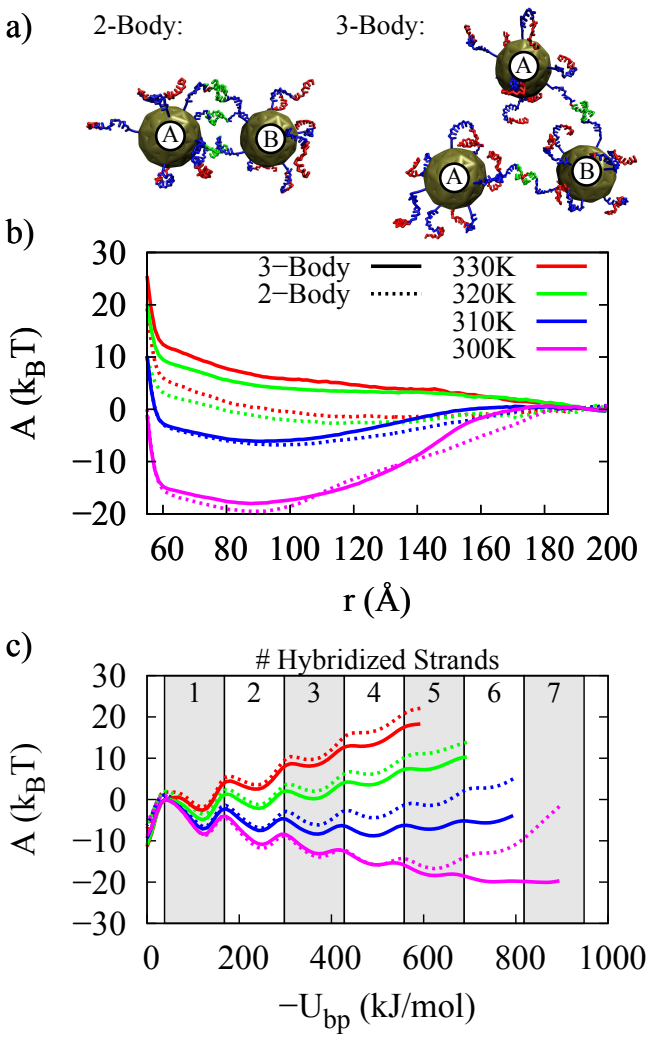

Figure 2.5: The role of three-body interactions in the Rogers8-7 system. a) Schematic of the two and three-body systems. In the three-body system, the three particles are constrained to the edges of an equilateral triangle. b) Free energy as a function of particle separation, r for two and three-body system. c) Free energy as a function of base-pairing energy, U_{bp}.

To explore the role of many-body effects on the calculated pair-potential we adopt the approach of Mladek et al. [119]. First, an equilateral triangle containing three DNA-conjugated nanoparticles is constructed (Fig. 2.5a). Two of the particles, denoted "A", contain the same linking DNA (e.g. Table 2.1, Rogers8-7 top sequence) while the third, denoted "B", contain the corresponding complementary sequence (e.g. Table 2.1, Rogers8-7 bottom sequence). That is, "A" and "B" particles can associate while two "A" particles cannot. In this configuration, the additional 3rd particle disrupts the pair-wise interactions and we can examine three-body effects. The calculation of the free-energy was performed in a manner analogous to that of the two-body case, except that r now represents the side of the equilateral triangle (instead of the distance between the two particles). Additional details of this calculation are provided in Section 2.3.2.

The calculated three-body pair-potential for the Rogers8-7 system, $A(r)$, is shown in Fig. 2.5b. For comparison, the two-body results from Fig. 2.3 are shown as dotted lines. At 320K and 330K, the three-body interactions are weaker than those observed in the two-body case. This originates from the increased repulsion that occurs when an additional particle is present. This repulsion also leads to a sharper melting transition for the three-body system (this result will be discussed in Section 2.4.3). At 300K and 310K however, the two and three-body interactions are very similar. Though the second "A" particle presents additional sticky ends to hybridize with the "B" particle, the energetic minima of the potentials remain unchanged.

The origins of this behavior can be seen in the free energy as a function of base-pairing energy, $A(U_{bp})$ (Fig. 2.5c). Despite additional sticky ends present in the three-body system, the free energy of base-pair formation for $-U_{bp} < 400$ is only marginally lower. As additional strands hybridize, however, the differences between the two systems become increasingly apparent. This is most clear at 300K: though the two- and three-body results are indistinguishable for $-U_{bp} < 500$, the results deviate from each other at large $-U_{bp}$. At high values of $-U_{bp}$ additional hybridization events in the three-body system incur a lower

entropic cost than the two-body system. Thus, the third particle only affects the free energy when many strands are hybridized between particles.

The observation that the pair-potential is essentially unchanged by the addition of a third particle deserves comment. Firstly, this demonstrates that the pair-potentials calculated here represent a good approximation to the effective potential present in DNA-nanoparticle crystals. We expect the pair-potential to be a reasonable approximation of the inter-particle spacing and (for low coordination numbers) the energy-scales within a crystal. However, we emphasize that since the interactions between multiple particles are not pair-wise additive, the pair-potentials presented here should be used in additional coarse-graining with caution.

2.4.3 Simulated Melting and Hybridization Kinetics

The results in Fig. 2.3b include additional information regarding the subtleties of interparticle interactions. Since details of the free energies of the hybridized and dehybridized states are known, a melting curve can be obtained (see Section 2.3.3). The simulated melting curves for the Rogers8-7 system are shown in Fig. 2.6a. The melting curve for the 7 base-pair "sticky end" in the bulk is shown for comparison. Note that the concentration of DNA strands in the bulk and DNA-nanoparticle simulation were chosen to be equal. Consistent with literature[39], the melting transition of the DNA-nanoparticle is much sharper than that of the bulk DNA. Additionally, the melting transition is sharper in the three-body system than in the two-body system. This result is consistent with the experimental observation that ordered assembly of many DNA-conjugated nanoparticles only occurs within a narrow temperature range. When many particles are present, the melting transition becomes increasingly sharp and the range of temperatures amenable to ordered assembly decreases.

Knowledge of $A(-U_{bp})$ provides access to kinetic information about DNA-nanoparticle assembly. A feature in Fig. 2.3b (up to this point left unmentioned) is the presence of multiple local minima and maxima in the free energy surface. These local minima correspond exactly to the hybridization of an additional DNA strand, while the local maxima correspond to

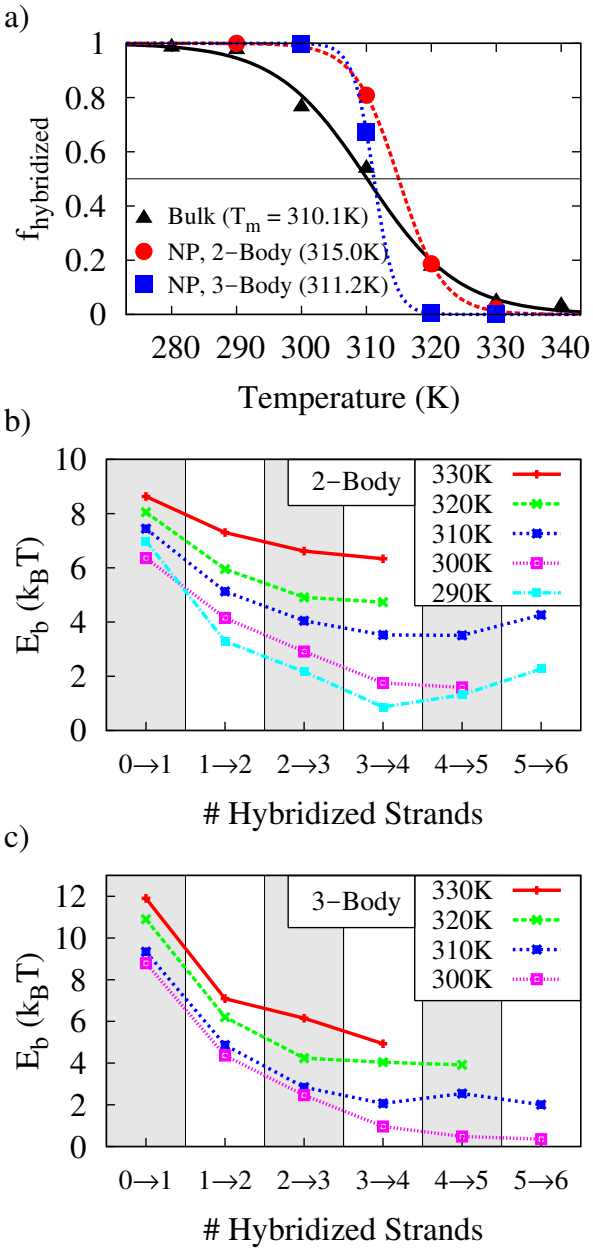

Figure 2.6: a) Simulated melting curve for Rogers8-7 DNA-nanoparticle system. The melting curve of the sticky end in the bulk is shown for comparison. As observed experimentally, DNA-nanoparticles (denoted by "NP") demonstrate a steeper melting transition than the same DNA sequence in the bulk. b) Energy barrier as a function of hybridization event for two-body interactions. c) Energy barrier as a function of hybridization event for three-body interactions. At all temperatures, the formation of the first hybridized strand is rate-limiting.

the activated state separating the transition from n to $n+1$ bound DNA strands. Since these local extrema are well defined, we can approximate each hybridization event as an independent reaction complete with its initial, final and activated energies. Here we define the energy barrier, E_b, as the difference in the free energy of the initial state, A_n, and the transition state, $A_n^\dagger$ ($E_b = A_n^\dagger - A_n$). The energy barrier as a function of number of hybridized strands is shown in Fig. 2.6b,c. For all temperatures studied, E_b is the largest for the first hybridization event. In other words, the formation of the first hybridization is rate-limiting. Further, E_b is also a strongly dependent on temperature. Thus, the kinetics of DNA-nanoparticle assembly (in addition to the thermodynamics as discussed in Fig. 2.3) are strongly affected by temperature.

The magnitude of E_b differs between the two and the three-body system in two noteworthy ways (Fig. 2.6c). First, the $0 \rightarrow 1$ barrier is $\sim 3k_BT$ larger in the three-body system. The enhanced repulsion that results from the additional particle slows the rate-limiting, initial hybridization event. Another important difference is the absence of an upward trend in E_b for the $4 \rightarrow 5$ and $5 \rightarrow 6$ transitions in the three-body configuration. Because there are additional sticky ends available and DNA stretching is reduced, E_b can decrease monotonically. These observations highlight the complicated role that many-body effects might play during assembly. On the one hand, additional particles increase the first barrier to particle association. On the other hand, the additional particles accelerate association by decreasing subsequent barriers to hybridization. Additional particles could both enhance and inhibit the kinetics of assembly.

2.4.4 Comparison to Park et al.[135]

We now use our DNA-nanoparticle model to explore a result obtained by Park et al.[135], who demonstrated assembly of DNA-nanoparticles into structures with long range order. These authors showed that a DNA-nanoparticle system that assembles into a ordered crystal and one that assembles into a disordered aggregate can differ by only a single base. Park et al.

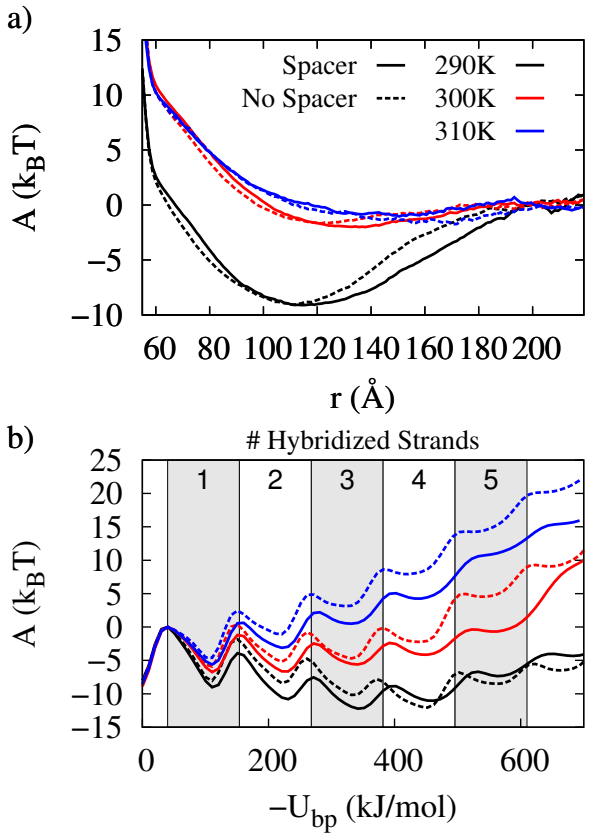

Figure 2.7: Effect of "spacer base" on free energy of Park7 Sequence. a) Pair Potential. The "spacer base" leads to the same strength of interactions but a notably wider energetic minima. b) Free energy as a function of base-pairing energy, U_{bp}. The "spacer base" stabilizes additional hybridization events.

refer to this as a "spacer base". In contrast to Rogers et al. [150] (who used ssDNA linkers), Park et al. [135] used DNA linkers consisting of both ssDNA and dsDNA regions. We model this explicitly using 3SPN.2 with no modification to the model (see Fig. 2.1b). As before, we adapt the DNA sequence used in Park et al. [135] for use in our study. For computational efficiency, we use a nanoparticle 5 nm in diameter (compared to the 10 nm particles used by Park et al.[135]). Accordingly, we scale the length of our "linker" DNA to maintain a constant ratio of particle diameter to linker length. In order to maintain a comparable energy scale, we use a 7-base pair sticky end identical to Park et al.[135]. The exact sequences used here are shown in Table 2.1. Interactions are simulated with and without the "spacer base".

The resulting pair-potential is shown in Fig. 2.7a. One consequence of the "spacer base" is the displacement of the equilibrium separation, r_{eq}, to slightly larger values of r. This slight shift is due to the longer strands that result from the "spacer base". The energy scale, however, remains unchanged. A more nuanced result is the broadening of the energetic minimum. This result is most obvious at 290 K. Whereas with the spacer, the width of the pair-potential within $3\,k_BT$ of the minimum is $\sim 70\,\mathring{A}$, the absence of the spacer reduces this width to $\sim 50\,\mathring{A}$. This result suggests that frustration could be responsible for disordered assembly. By widening the width of the minima, the "spacer base" disrupts this frustration and allows the system to reach a well-ordered, crystalline state.

The origins of this wider minimum can be seen in Fig. 2.7b. At each temperature, the free energy of the sequence containing the spacer is lower than that without the spacer. This indicates that the spacer helps stabilize additional base pairing events. By allowing additional strands to hybridize, the "spacer base" increases the degeneracy of states accessible in the vicinity of the minimum. The increased degeneracy leads to wider minima, as observed in Fig. 2.7a.

Validating this "frustration" hypothesis would require a more detailed study and is beyond the scope of this work. Nonetheless, a study of this nature would be useful towards understanding the nuances of DNA-conjugated nanoparticle assembly. This problem will be

examined in future work.

2.5 Conclusion

A coarse-grained model of DNA, whose hybridization has been studied in the bulk, was used here to provide an in-depth view of DNA-assisted nanoparticle association. We calculated the pair-potential between nanoparticles, and showed that the association of DNA-nanoparticles relies on a delicate balance of enthalpic and entropic contributions to the free energy. We also showed that this balance is particularly important for understanding the differences between nano- and colloidal-sized particles in the vicinity of their melting temperatures. Our calculations also revealed that three-body effects can alter the resulting pair-potential and the details of particle binding, particularly when the particles are small. Using information about individual base pairing events, we characterized the melting transition of DNA-nanoparticles, as well as the kinetics of subsequent strand hybridization events. Because our model can resolve the DNA backbone and individual bases, we explored the finding of Park et al. [135] that the inclusion of a single "spacer base" can result in ordered assembly, and proposed that the main role of this "spacer base" is to reduce frustration. The model presented here is highly flexible in design and can be applied to a wide variety of DNA-nanoparticle systems with different sequences, particle sizes or shapes. In particular, we expect it to be useful for studying of the thermodynamic and mechanical properties of DNA-nanoparticle crystalline assemblies.

2.6 Appendix

2.6.1 Model Parameters

The parameters used for DNA were identical to those in 3SPN.2 [59]. The interactions within the nanoparticle and thiol linker consisted of only bonded, bending and excluded volume interactions (U_{bond}, U_{bend}, and U_{exe}, respectively). The form of these interactions

were identical to 3SPN.2. That is

$$\begin{aligned} U_{bond} &= \sum_{i}^{bonds} k_{b,2}(r_i - r_{0,i})^2 + k_{b,4}(r_i - r_{0,i})^4, \\ U_{bend} &= \sum_{i}^{bonds} k_{theta}(\theta_i - \theta_{0,i})^2 \end{aligned} \tag{2.5}$$

and

$$U_{exe} = \sum_{i<j} \begin{cases} \epsilon_r \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \epsilon_r & r < r_c \\ 0 & r \geq r_c \end{cases}, \tag{2.6}$$

where r is the separation between two interaction sites.

The parameters of these interactions are given in Table 2.2. For excluded volume interactions, $\epsilon_r = 1$ kJ/mol. The radius of the thiol linker is chosen as, $\sigma_L = 1\,\mathring{A}$. The value of σ_B was varied depending on the number of sites used to define the particle and that particles diameter. σ_B was chosen as the radius of the circle that circumscribes the equilateral triangle formed by 3 neighboring sites. This created a sphere that was effectively impermeable. Mathematically this is given by

$$\sigma_B = \frac{\sqrt{3}}{3}\Delta x \tag{2.7}$$

where Δx is the distance between two adjacent sites on the surface of the nanoparticle. Because the particles are spaced evenly on the surface of the sphere, Δx is independent of the neighbor chosen.

Table 2.2: Model Parameters for bonded interactions involving particles (denoted "B") and linking thiol group (denoted "L").

Bond	$k_{b,2}$ $(kJ/mol/\mathring{A}^2)$	$k_{b,4}$ $(kJ/mol/\mathring{A}^4)$	r_0 $(\mathring{A})$
L-S	0.6	60.0	3.33
L-L	0.6	60.0	3.33
L-B	0.6	60.0	3.33
B-B	80.0	60.0	–

Bend	k_θ $(kJ/mol/rad^2)$	θ_0 $(^\circ)$
L-S-P	200.0	134.03
L-S-T	200.0	103.28
L-S-A	200.0	103.28
L-S-G	200.0	103.28
L-S-C	200.0	103.28
L-L-S	200.0	178.00
L-L-L	200.0	176.00
B-L-L	200.0	178.00
B-B-L	200.0	105.19

CHAPTER 3
MECHANICAL RESPONSE OF DNA–NANOPARTICLE CRYSTALS TO CONTROLLED DEFORMATION

3.1 Abstract

The self-assembly of DNA-conjugated nanoparticles represents a promising avenue towards the design of engineered hierarchical materials. By using DNA to encode nanoscale interactions, macroscale crystals can be formed with mechanical properties that can, at least in principle, be tuned. Here we present *in silico* evidence that the mechanical response of these assemblies can indeed be controlled, and that subtle modifications of the linking DNA sequences can change the Young's modulus from 97 kPa to 2.1 MPa. We rely on a detailed molecular model to quantify the energetics of DNA-nanoparticle assembly and demonstrate that the mechanical response is governed by entropic, rather than enthalpic contributions and that the response of the entire network can be estimated from the elastic properties of an individual nanoparticle. The results here provide a first step towards the mechanical characterization of DNA-nanoparticle assemblies, and suggest the possibility of mechanical metamaterials constructed using DNA.

3.2 Introduction

Nanoparticles functionalized with short sequences of DNA represent a highly-customizable platform for multiscale materials design. In such systems, interactions between nanoparticles are mediated by short strands of DNA, typically with lengths on the order of tens of base pairs[116, 2]. By varying the length and composition of these linking DNA sequences, the strength, range and specificity of inter-particle interactions can be precisely tuned. The ability to customize and specify DNA-mediated interactions promises to facilitate the design of hierarchical structures whose macroscopic properties could be tuned by manipulating the

corresponding nano-scale building blocks of which they are composed.

DNA-functionalized nanoparticles have now been shown to assemble into crystals with long range order [135, 129], and with tunable lattice parameters and crystal symmetries based on DNA sequence alone[107]. Through advances in nano-scale synthesis, it has also become possible to assemble nanoparticles of different shapes and properties[70, 205, 130, 104]. The resulting materials exhibit intriguing properties, such as dynamic reprogramming [206] and single crystal assembly with well-defined facets[10], and have been predicted to demonstrate re-entrant melting[4]. Recent work has demonstrated that DNA-programmed assemblies have useful, tunable plasmonic properties[174, 133, 151, 198]. A largely under-explored aspect of DNA-functionalized nanoparticle assemblies, however, is related to the tunability of their mechanical properties. Though two-dimensional films of DNA-functionalized nanoparticles have been assembled[164] and shown to have extraordinary mechanical properties[24], little work [128] has been done to characterize or tune the mechanics of DNA-nanoparticle assemblies, especially in three-dimensions.

Our interest in mechanically tunable DNA-nanoparticle assemblies builds on recent work examining the mechanical response of nanoparticles conjugated with other short organic ligands. Extremely strong, two-dimensional nanoparticle sheets have been prepared with nanoparticles functionalized with simple dodecanethiol ligands[124, 187, 153]. At the other end of the chain-length spectrum, Williams *et al.*[191] have shown that polymer-grafted nanoparticles interacting via hydrogen-bonds assemble into a fcc crystal with a Young's modulus of $26 - 82$ MPa that is capable of self-healing[191]. Importantly, these authors demonstrated that the mechanical and optical response can be tuned by varying the length of polymer-grafts, and that the optical properties can be altered by mechanical deformation. More generally, the mechanical properties of ligand-conjugated nanoparticle assemblies are of fundamental interest because they arise from non-linear combinations of their constituents; indeed, they possess characteristics of both granular (due to the nanoparticles) and viscoelastic (due to the ligands) systems[90].

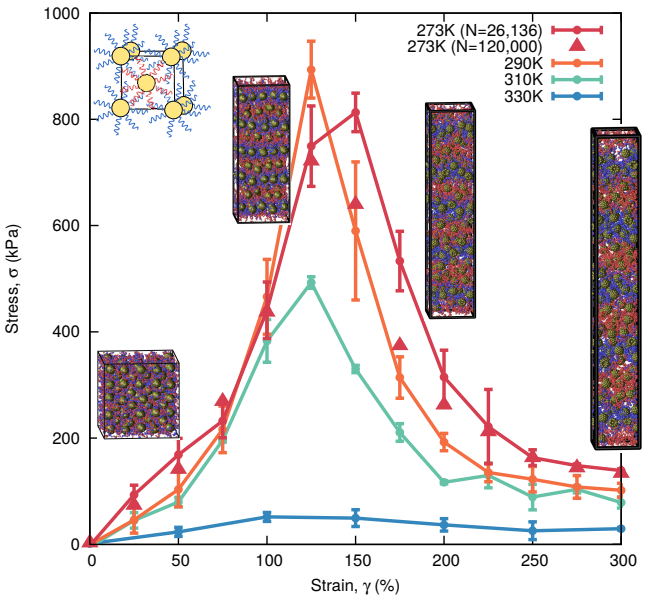

Figure 3.1: Stress-strain response of DNA-nanoparticle assembly under uniaxial extension for Sequence B (see Fig. 3.2). Simulation snapshots show the material after 0%, 100%, 200%, and 300% strain. All DNA sequences of the same type are given the same color (i.e. red or blue). Error bars denote an standard deviation over three independently initialized nanoparticle assemblies.

In this work, we present a first step towards the detailed characterization of the mechanical response of DNA-nanoparticle assemblies. To achieve this, we rely on a detailed molecular model to examine the mechanical properties of DNA-conjugated nanoparticle assemblies *in silico*. Our results demonstrate that such properties can indeed be tuned by using different DNA sequences, DNA loading densities, and temperature, thereby providing a potentially useful platform for creation of mechanical metamaterials.

3.3 Results and Discussion

3.3.1 Mechanical Response of DNA-conjugated Nanoparticle Assembly

DNA-conjugated nanoparticle lattices were assembled as described in the Methods section, and their mechanical response was measured under uniaxial extension (Fig. 3.1). The deformation was applied quasi-statically, where a nanoparticle assembly was deformed to the

specified strain and then held fixed until the stress re-equilibrated. The resulting stress-strain curve exhibits elasticity (i.e. linear response) for all temperatures up to strains of 50%. Following this elastic regime, the material stiffens until its peak stress at $\approx 125-150\%$ strain.

At strains above the peak stress, the material does not immediately yield, but instead the stress decreases slowly: even at strains of 300% there is still a small but non-zero stress. Notably, even at these large strains, the material does not exhibit necking and is characterized by constant densities throughout the sample (Fig. 3.1 snapshots, Fig. 3.7). Yet despite the lack of macroscopic defects (i.e. necking), the microscopic structure of the nanoparticle assembly is significantly perturbed under large strains. For deformations beyond the peak stress, the long-range crystalline order within the nanoparticle network is disrupted, and the assemblies become amorphous (Fig. 3.1 snapshots, Fig. 3.8,3.9). This observation suggests that under sufficiently slow deformation (i.e. quasi-static), the microstructure of the assembly can rearrange, and thereby prevent the formation of failure-prone morphologies, including necks or voids. This behavior can be thought of as a form of microscopic self-healing, where individual nanoparticles reposition themselves within the network to maintain the integrity of the material.

Temperature also plays an important role in the mechanical response (Fig. 3.1). While at low temperatures, the material exhibits a response in the kPa range, higher temperatures result in material softening and, eventually, a negligible mechanical response (e.g. 330K). The extreme softening of the material occurs for this assembly between 310K and 330K, corresponding extremely well to the 310K-315K melting temperature of this DNA sequence calculated previously[94].

In order to confirm that our measurements are representative of a bulk material and do not suffer from finite-size effects, simulations were also performed for systems consisting of 125 bcc unit cells, corresponding to 120,000 total coarse-grained sites (Fig. 3.1, red triangles). The results from this larger system are largely indistinguishable from those for the 27 unit cell

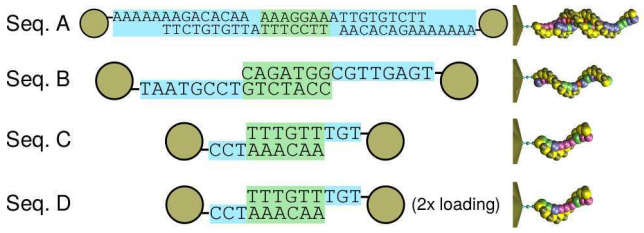

Figure 3.2: Different DNA sequences conjugated to nanoparticle surface. Green highlights the complementary "sticky end" region, while Blue highghts the non-reactive "linker" region. Snapshots corresponding to the molecular topology of each DNA linker is shown. "2x loading" denotes that Seq. D contains twice the number of DNA strands per nanoparticle. Note that Seq. A's linker consists of two DNA strands and contains a non-reactive double-stranded region, while the linkers of Seq. B, C, and D only consist of a single DNA strand.

assembly (N = 26,136), confirming that our simulations are representative of a bulk material. Accordingly, nanoparticle assemblies consisting of 27 unit cells will be used throughout the remainder of this work.

3.3.2 Tunable Sequence-Dependent Mechanical Response

A potential feature of DNA-nanoparticle assemblies is that subtle changes in the linking DNA sequences may be used to generate materials with different properties. To examine the effect of DNA sequence on the mechanical response, we repeated our analysis above using different DNA sequences (Fig. 3.2). Because the parameter space corresponding to all possible DNA sequences is prohibitively large, we focused here on a subset of DNA parameters that have been varied elsewhere in the DNA-nanoparticle literature[129, 64, 94, 107]. These include "one strand" vs "two strand" linkers (i.e. Seq. A vs Seq B.), effect of linker length (Seq. A vs Seq C) and the effect of DNA loading density (Seq. C vs Seq. D) and the effect of the complementary DNA "sticky ends" (Seq A vs Seq B vs Seq C and D). Consistent with experimental systems[135, 150, 10], the sequences chosen here contain complementary "sticky" DNA regions of 6-7 base pairs.

The mechanical response is shown to be dramatically dependent on DNA-sequence (Fig. 3.3). One sequence-dependent effect is the qualitative difference in mechanical response between

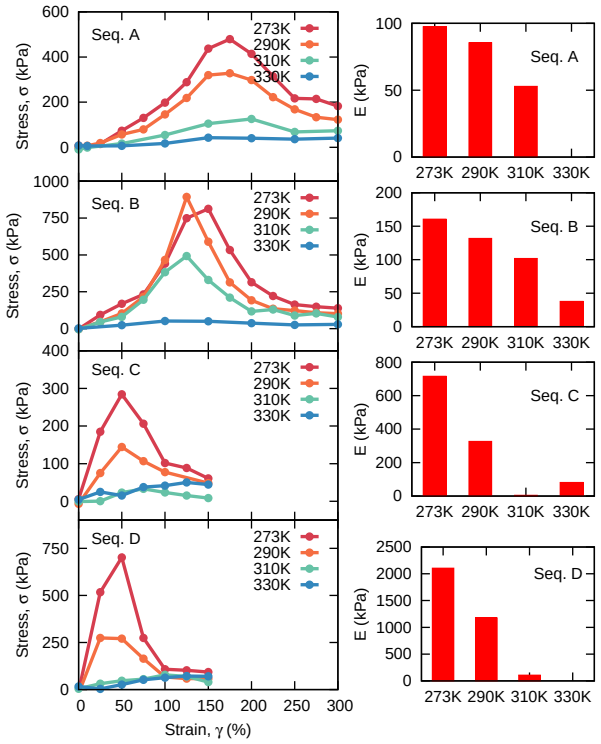

Figure 3.3: DNA sequence-dependent mechanical response. Stress-strain response of sequences A, B, C and D and corresponding Young's modulus at different temperatures. Variations in both temperature and sequence result in both qualitative and quantitative changes in the stress-strain response of DNA-nanoparticle assemblies.

sequences with short linkers (i.e. C and D) relative to sequences with long linkers (i.e. A and B). Sequences C and D demonstrate a stiff initial response ($E \approx 700-2100$ kPa) but exhibit little strain-hardening, and lose nearly all stiffness at strains greater than 100%. In contrast, Sequences A and B exhibit a softer initial response ($E \approx 100$ kPa), but a large peak strain ($\approx 500-750$ kPa). Additionally, long linkers enable strains of up to 300% without mechanical rupture. Thus, short linkers give rise to "brittle" mechanical behavior, while longer linkers result in more pliable materials. Curiously, though Sequence A and Sequence B are formed from very different linker types (i.e. double-stranded vs single-stranded), they nonetheless demonstrate a qualitatively similar mechanical response. Though double-stranded linkers are stiff, the similarities between our results for Seq. A and B suggests that the mechanical response is predominantly driven by the single-stranded linker of the assemblies constructed using Sequence A. This results suggests, that it is not the type of DNA linker but rather its single-stranded length, that dictates mechanical response in DNA-nanoparticle assemblies.

The effect of DNA coverage can be assessed by comparing Sequence C and D. Assemblies using these linkers use identical DNA sequences, and only differ in the density (i.e. loading) of DNA strands conjugated to the nanoparticle surface. That is, Seq. D has twice the number of DNA strands on the nanoparticle surface relative to Seq. C. Our results demonstrate that DNA loading density results in significant changes of the stress-strain response. For example, by simply doubling the DNA loading (i.e. nine strands for Seq. C vs 18 strands for Seq. D), the Young's modulus changes by a factor of three (≈ 700 kPa to ≈ 2100 kPa) and the peak-stress more than doubles (≈ 300 kPa to ≈ 740 kPa). Yet, despite these changes, the qualitative features of the the mechanical responses of Seq. C and D remain largely unchanged: the stress increases linearly up until a peak-strain at $\gamma \approx 50\%$ and then decreases significantly. This result suggests that DNA loading represents an important parameter for tuning mechanical behavior. Whereas the qualitative features of the mechanical response can be adjusted by changing the characteristics of the DNA sequences themselves, the magnitude of the response can be adjusted independently by varying the DNA loading on the

nanoparticle surface. In that sense, DNA coverage represents a system parameter orthogonal to the DNA sequence itself, which can be tuned to obtain the desired mechanical response. Additionally, since the number of strands increases approximately with the square of the nanoparticle radius, the larger 10-20 nm nanoparticles used experimentally[129, 135, 10] will posess even more particle-particle connections and likely a significanlty stronger mechanical response.

Interestingly, the different DNA sequences studied here exhibit different temperature-dependent mechanical properties. The mechanical response of Seq B is nearly unchanged between 273K and 290K, for example, while those of Sequences C and D over this same temperature range change dramatically. These different behaviors can be explained in terms of the melting temperature, T_m , of the DNA "sticky ends" that link complementary particles. By choosing "sticky ends" with a higher T_m (as with Seq B), the nanoparticles demonstrate higher thermal stability and are more mechanically robust. Because DNA-nanoparticles are known to exhibit extremely sharp melting curves[39], our results here suggest the possibility of extremely sensitive thermo-responsive materials whose mechanical properties could change by orders of magnitude with temperature changes of only several degrees.

3.3.3 Molecular Origin of Mechanical Response

Having considered the mechanical properties of DNA-nanoparticle assemblies, we now turn our attention to the molecular processes responsible for specific sequence-dependent mechanical responses. Specifically, we seek to provide a molecular explanation of the stress-strain curves reported in Figure 3.3. In particular, we will examine whether the mechanical response correlates more strongly with (a) the enthalpic penalty arising from the disruption of base pairs or (b) the entropic penalty arising from the anisotropic ordering of DNA strands on the nanoparticle surface.

Since nanoparticle-nanoparticle interactions are mediated by complementary DNA-DNA base pairs, deformation of the network is expected to result in base pair disruption. Disrup-

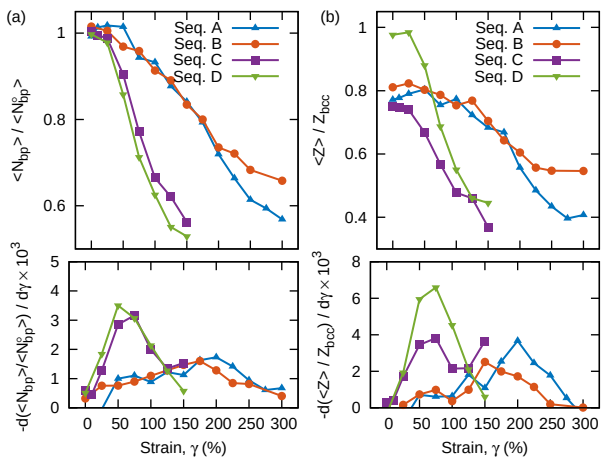

Figure 3.4: Effect of deformation on connectivity of DNA-nanoparticle network. a) Average number of base-pairs, $\langle N_{bp} \rangle$, and numerical derivative, $\frac{\mathrm{d}\langle N_{bp} \rangle}{\mathrm{d}\gamma}$, and b) Average coordination number, $\langle Z \rangle$, and numerical derivative $\frac{\mathrm{d}\langle Z \rangle}{\mathrm{d}\gamma}$ for different DNA-nanoparticle sequences at 273K. Though the disruption of base pairs influences the mechanical behavior, analysis of network connectivity alone is insufficient to fully explain the mechanical response.

tion of base pairs will incur an enthalpic penalty, and therefore we anticipate that disruption of base pairs might correlate strongly with the observed stress-strain response. Figure 3.4a shows the network connectivity (expressed as the number of base pairs in the network) for each DNA sequence as a function of strain. As expected, strain disrupts the DNA-nanoparticle network and the number of base pairs is observed to decrease. Notably, the slope of the base pair-strain curve (i.e. the derivative) differs between different DNA sequences. For short DNA linkers (i.e. Sequences C and D), the decrease in base pairs is more rapid than for DNA-nanoparticles with longer linkers (Sequences A and B). The slope of the base pair-strain curve is quantified by calculating the numerical derivative shown in the bottom panel of Fig. 3.4a. Sequences C and D demonstrate a sharply peaked derivative at 50% strain, corresponding to the maximum in their respective stress-stain response (Fig. 3.3). This correlation suggests that the stress-strain response is influenced by the rate of base pair disruption and the associated enthalpic penalty. This possibility, however, is only weakly supported by the results for Sequences A and B. For these two sequences, the derivatives are approximately independent of strain, and only show a subtle peak at 175%-200% strain, a

value higher than the peak in the stress-strain response at 125%-175%. Therefore, though the disruption of base pairs influences the mechanical behavior, our results suggest that it is not the predominant explanation for the observed sequence-dependent mechanical response.

A notable feature of Fig. 3.4a is that the number of base pairs does not decrease to zero. Even after the assembly has been mechanically disrupted, 60% to 70% of the original base pairs are still intact. This observation led us to propose that it is not the absolute number of base pairs that matter for the mechanical response, but it is instead the connectivity of the network. To test this hypothesis, we measure the connectivity of the network on the basis of nanoparticle coordination number. In this context, we define the coordination number, Z, as the number of nanoparticles that a given nanoparticle is base paired to, normalized by that expected in a bcc crystal lattice, $Z_{bcc} = 8$, (Fig. 3.4b). At rest, sequences with ordinary DNA loading densities (Sequences A,B,C) have coordination numbers that are $\approx 80\%$ of the ideal value. This result is expected since these particles contain only nine DNA strands per particle, and therefore the probability of forming DNA "bonds" between all 8 bcc neighbors is small. However for particles with a higher loading density (Seq. D), the connectivity of the network is essentially complete, with $Z/Z_{bcc} \approx 1.0$ at 0% strain. The near perfect connectivity of Sequence D helps explain its high mechanical strength: as the connectivity approaches the ideal value, mechanical properties increase considerably. Note, however, that our analysis of the coordination number is still insufficient to completely explain the sequence-dependent stress-strain response. As with the number of base pairs, N_{bp} shown in Fig. 3.4a, the coordination number in Fig. 3.4b decreases with strain and the derivative (lower panel) correlates weekly with the stress-strain curves.

Another possible explanation for the role of sequence on mechanical properties relates the deformation of individual DNA-conjugated nanoparticles themselves to the overall deformation of the network. During network deformation, it might be expected that the DNA strands conjugated to the nanoparticle surface would anisotropically orient themselves along the direction of the applied strain. Such ordering would incur an entropic penalty, and could be

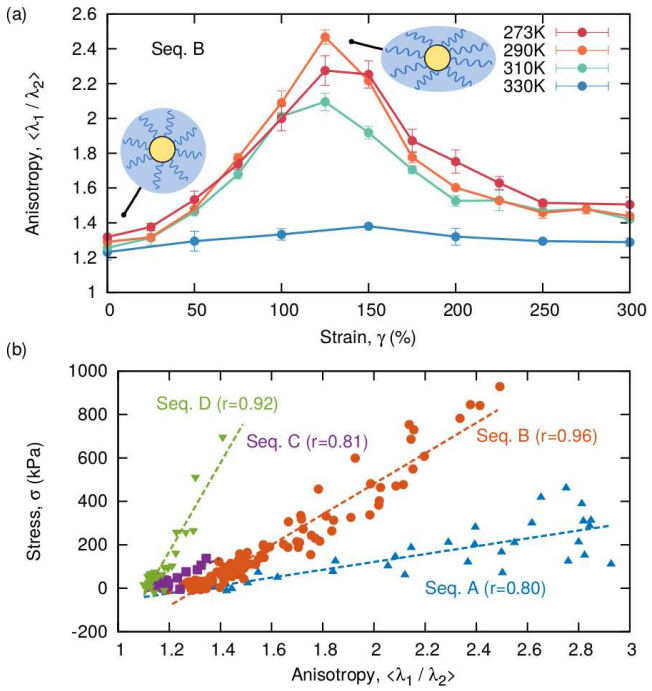

Figure 3.5: Anisotropic deformation of individual nanoparticle shape. a) Average anisotropy of DNA-nanoparticle shape during deformation. The anisotropy is qualitatively similar to the mechanical response (cf. Fig. 3.1). b) Correlation between DNA-nanoparticle anisotropy and calculated stress during deformation. For all sequences and temperatures, a strong correlation exists.

responsible for the observed sequence-dependent mechanical response. Indeed, such entropic penalties dominate the mechanical properties of polymer melts, and could be significant in DNA-nanoparticle assemblies as well.

To examine the role of this entropic ordering, the shape of individual nanoparticles was quantified by calculating the radius of gyration tensor, $S_{\alpha\beta}$, for each DNA-conjugated nanoparticle, and then determining its three eigenvalues, $\lambda_1 > \lambda_2 > \lambda_3$. The ratio of the two largest eigenvalues, λ_1/λ_2 gives the anisotropic shape of each nanoparticle with $\lambda_1/\lambda_2 \approx 1$ representing a sphere and $\lambda_1/\lambda_2 \to \infty$ an infinite rod[9]. Since the nanoparticles are very rigid, values of $\lambda_1/\lambda_2 \neq 1$ represent deformations of the conjugated DNA strands, and not the nanoparticles themselves (Fig. 3.5a, graphic). We will refer to λ_1/λ_2 as the "anisotropy" of the DNA-nanoparticles.

The average anisotropy for different values of strain and temperature is shown in Fig. 3.5a for Sequence B. At zero strain, the anisotropy is small and the individual DNA-nanoparticles are nearly spherical, as expected. As strain is increased, the average anisotropy increases up to a peak at 125% and then decays toward its initial value for strains of up to 300%. The anisotropy is also temperature dependent; higher temperatures lead to less anisotropy. The most important feature of this result, however, is the striking similarity between the anisotropy and the stress (cf. Fig. 3.1). In fact, both curves appear to be qualitatively identical, with the same maximum, curvature, and temperature dependence. This observation suggests that the shape anisotropy and the stress might be fundamentally related within the DNA-nanoparticle network.

To quantify this relationship, the correlation between anisotropy and stress was calculated for all sequences and temperatures used in this study (Fig. 3.5b). Confirming the previous observation for Sequence B, a strong correlation ($0.8 < r < 0.96$) exists between anisotropy and stress for all DNA sequences considered here. This striking result suggests that strain within the DNA-nanoparticle network is strongly related to the anisotropic deformation of the DNA strands on the particle surface, which gives rise to a free energy penalty that

manifests itself as a restoring stress when the network is deformed.

An important observation from Fig. 3.5b is that each DNA sequence exhibits a different relationship between stress and anisotropic deformation; or in other words, each DNA sequence shows a different slope. The softest assembly examined here, Sequence A, has the smallest slope (i.e. low stress per anisotropic deformation), whereas increasing assembly stiffness (i.e. Seq. D) leads to increasingly large slopes (i.e. high stress per anisotropic deformation). These results suggest that each DNA sequence possesses an inherent property, which we refer to as "shape stiffness", that corresponds to the entropic penalty incurred by causing the DNA-strands to order anisotropically. Notably, shape stiffness is sequence dependent: for Sequence A, the long DNA-linkers incur a relatively small penalty upon ordering, leading to small "shape stiffness". In contrast, the multiple short DNA linkers in Sequences C and D give rise to a larger penalty upon ordering, causing the "shape stiffness" to be large. Our results indicate that the "shape stiffness" of a single DNA-nanoparticle represents a key parameter that dictates the mechanical response of the entire DNA-nanoparticle network.

This observation has important implications for the design of DNA-nanoparticle assemblies. First, our results suggest that by simply quantifying the "shape stiffness" for a single nanoparticle the mechanical response of the network can be estimated. As such, "shape stiffness" might provide a simple metric for screening the high parameter space of different DNA-sequences in order to obtain the desired mechanical response. Second, the importance of entropic, rather than enthalpic contributions to the mechanical response has important implications for tuning the mechanical response of DNA-nanoparticle assemblies. Though previous studies have emphasized the energy scale of the complementary "sticky ends" in dictating DNA-nanoparticle assembly[96, 97], our results indicate that it is instead the properties of the unreactive linker DNA that dominate the mechanical response. This observation raises the intriguing prospect of tuning independently the mechanical response and the underlying crystal structure, leading to the possibility of creating DNA-nanoparticle assemblies with complex and precisely tunable mechanical properties.

3.4 Conclusions

In this work we have used a detailed molecular model of DNA-conjugated nanoparticles to examine the mechanical properties of DNA-nanoparticle assemblies. We demonstrate that this mechanical response is strongly dependent on temperature and suggest the possibility of thermo-sensitive materials whose mechanical properties could change by orders of magnitude upon temperature changes of only several degrees. The mechanical response is also shown to be strongly dependent on DNA sequence and that subtle changes in the linking DNA can lead to significant changes in the qualitative and quantitative features of the mechanical response. Then, moving beyond existing experiments, we interrogate our molecular model in order identify the physics that dictates the observed sequence-dependent mechanical response. By analyzing the connectivity of the network, we show that the enthalpic penalty due to base pair disruption partially explains the observed mechanical response. Most importantly however, we demonstrate that the overall mechanical response of the network is strongly correlated with the deformation of a single DNA-conjugated nanoparticle. From this observation we suggest a new sequence-dependent parameter, which we refer to as "shape stiffness", that can be used to estimate the mechanical response of a nanoparticle network from a single nanoparticle. The results presented here are the first detailed characterization of the mechanical response of DNA-nanoparticle assemblies, and are the first to demonstrate that the mechanical response can be tuned. Consequently, they represent a valuable step toward understanding the mechanical properties of DNA-nanoparticle assemblies and is useful for dictating their future directions and applications.

3.5 Methods

The molecular model of DNA-conjugated nanoparticles adopted here was introduced and validated in previous work [94]. In this model, DNA is represented by 3SPN.2[59], the latest model in the 3SPN family[82, 154]. As a coarse-grained model, 3SPN.2 represents each

nucleotide of DNA with three force sites, one at the center of mass of the sugar, phosphate and base. 3SPN.2 has been parameterized to reproduce the structural (e.g. helix width, major and minor groove widths) and mechanical (e.g. persistence length) properties of double and single-stranded DNA. Additionally, 3SPN.2 can reproduce the melting temperature of double-stranded DNA and hairpins as a function of sequence and ionic strength[59], and has been used in detailed studies of the single-stranded to double-stranded DNA transition[60]. The DNA-conjugated nanoparticles considered here are constructed by tethering 3SPN.2 to a coarse-grained nanoparticle consisting of a bonded network of sites placed on the surface of a sphere. This model results in nanoparticle sites that are fixed at their relative locations on the surface of a sphere, and therefore the positions of DNA strands are also fixed and cannot migrate along the particle surface. This model has been used to examine the precise energetics of pair-wise DNA-nanoparticle assembly and was shown to reproduce the correct energy scales and temperature dependence of DNA-nanoparticle association [94]. In this work, we use 5 nm diameter nanoparticles covered with a DNA density of 19 $\mathrm{pmol/cm^2}$ (unless otherwise noted) as shown experimentally[64]. This DNA density corresponds to nine DNA strands on a 5 nm diameter nanoparticle, distributed approximately uniformly across the particle surface. When refering to "double DNA loading", the DNA density was increased to 38 $\mathrm{pmol/cm^2}$, corresponding to 18 strands per 5 nm particle.

Several of the DNA sequences considered here (Fig. 3.2, Seq. A and B) are chosen to correspond to those used in previous experimental studies [135, 150]. As in previous work[94], the reactive "sticky end" of the DNA sequence (see Fig. 3.2 green box) is chosen to be identical to that of the experiments. The unreactive "linker" region, however, is scaled to account for differences between experimental and simulated particle diameters. The net effect of this scaling has been examined in detail previously [94].

DNA-nanoparticle latices were formed using a binary mixture of nanoparticles coated with different DNA sequences. The DNA sequences were chosen so that interactions between nanoparticles with different DNA sequences are attractive, whereas nanoparticles with the

same DNA sequence are not. As a result of these attractive and repulsive interactions, binary nanoparticle mixtures assemble experimentally into body centered cubic (bcc) crystal lattices. The protocol used for assembling the bcc lattices in this work is described in the Suppl. Information. When applying uniaxial extension, constant strain was imposed in a single dimension while constant stress was imposed the non-strained dimensions. To apply the deformation quasi-statically, the desired strain was applied and the resulting stress was measured following equilibration to its new steady state.

Simulations were performed in the NVT or NPT ensemble using a Langevin thermostat and/or a Parrinello-Rahman barostat with damping coefficients of 1ps and 20 ps, respectively, and a time step of 20 fs. The Debye-Hückel approximation was used to model the interactions between phosphate sites which carry a charge of -0.6 [59]. Simulations were performed by implementing the 3SPN.2 nanoparticle model[94] into the LAMMPS simulation package [140]. The approach described by Thompson *et al.*[179] was used to calculate the virial contribution to the pressure from 3SPN.2. The stress tensor was calculated as the sum of the kinetic and virial components,

$$\sigma_{\alpha\beta} = \frac{1}{V}\sum_{k}^{N} m_k v_{k,\alpha} v_{k,\beta} + \frac{1}{V}\sum_{k}^{N} F_{k,\alpha} r_{k,\beta} \tag{3.1}$$

Note that while the kinetic component of the stress was included, its contribution to the stress was observed to be small.

3.6 Supporting Information

3.6.1 Crystal Formation

Several approaches have been proposed in the literature to generate DNA-conjugated nanoparticle crystals via simulation. One simple, but limited, method is to impose the desired crystal structure (and lattice constant) on the nanoparticles a priori through choice of the initial

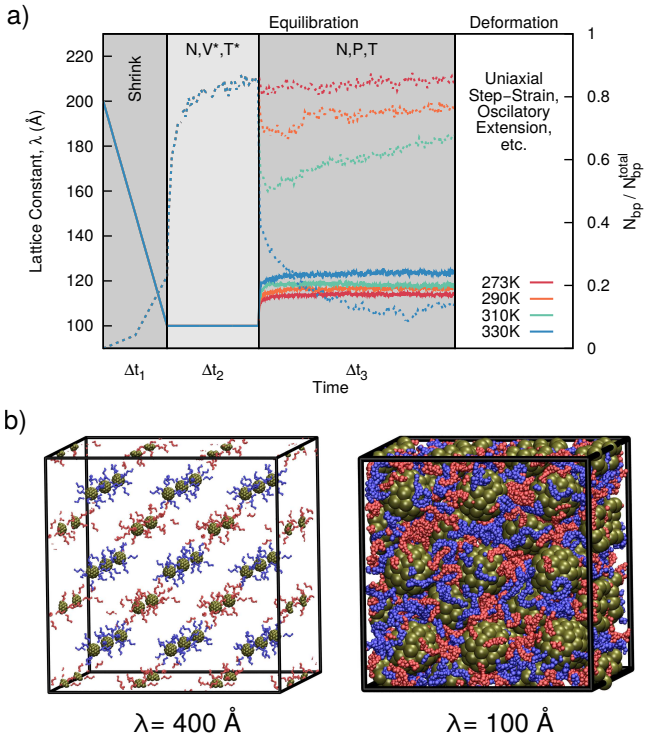

Figure 3.6: DNA-nanoparticle assembly initialization protocol. a) Lattice constant (solid lines) and number of base-pairs (dotted lines) during initialization protocol. The protocol involved initializing DNA-conjugated nanoparticles in dilute bcc lattice and then isotropically shrinking the simulation box. Next, the system was equilibrated using NV^*T^* and NPT simulations, where V^* and T^* were chosen to maximize base pair formation and P and T were the pressure and temperature of interest. b) Simulation snapshots corresponding to different lattice constants, λ.

conditions [128, 132]. This approach is computationally efficient but necessarily relies on initial assumptions that impose the desired connectivity and crystal structure of the resulting assembly. Another approach is to initialize the nanoparticle locations randomly, and then allow the crystals to assembly naturally[96, 97]. This approach has been powerful for elucidating the assembly pathways of DNA-nanoparticle crystals, but relies on simplified models of DNA in order to access the long time scales associated with assembly. In this work, we use a hybrid approach where the crystal configuration is biased towards those observed experimentally, yet without imposing DNA connectivity or crystal dimensions.

The prescription for generating the DNA-nanoparticle latices in this work is given as follows (Fig. 3.6). In the first step, DNA-nanoparticles are initialized in a dilute bcc lattice, so that the spacing between nanoparticles is much larger than their range of interaction ($\lambda = 400\mathring{A}$). The simulation box is then uniformly compressed over a timescale of Δt_1 until the nanoparticles are sufficiently close to interact. The lattice constant of this compressed state was chosen as $\lambda = 100\mathring{A}$, corresponding to an inter-particle separation of $71\mathring{A}$ and a volume of V^* (Fig. 3.6a, labeled "shrink"). In the second step, a NV^*T^* simulation of length Δt_2 is performed to allow the formation of base-pairing interactions between particles (dotted lines, Fig. 3.6). The temperature, T^* is chosen to be lower than the melting temperature of the linking DNA strands to accelerate this step, and is $T^* = 273K$ for all DNA sequences examined. Δt_2 was chosen to be sufficiently long to allow the number of base-pairs within the simulation, N_{bp}, to equilibrate. Finally in the third step, the crystals are re-equilibrated in the NPT ensemble at the temperature and pressure corresponding to the desired mechanical deformation of interest. In this study, the pressure $P = 1atm$ was chosen for all simulations; the temperature T is specified alongside each result. The time constants of each step were chosen to be $\Delta t_1 = 2$ ns, $\Delta t_2 = 600$ ns, and $\Delta t_3 = 1400$ ns. The box dimensions and normalized number of base-pairs are shown for each of these steps for four different temperatures (Fig. 3.6). Following this initialization recipe, the resulting nanoparticle assemblies were subjected to the mechanical deformation of interest.

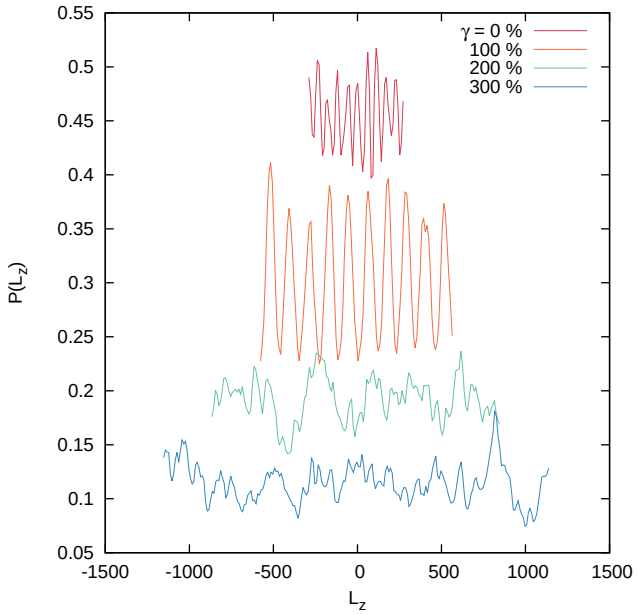

Figure 3.7: Density distribution throughout simulation box in strain direction for Sequence B at 273K. Even at high strains, no necking is observed.

3.6.2 Macroscopic and Microscopic Structure during Deformation

As observed in the main text, "At strains above the peak stress the material does not immediately yield, but instead the stress decreases slowly: even at strains of 300% there is still a small but non-zero response. Notably, even at these large strains the material does not exhibit necking and is characterized by constant densities throughout the sample (see Fig. 3.1 and Suppl. Information)". To demonstrate the absence of necking, the normalized density throughout the simulation box is shown in Fig. 3.7. At small strains $\gamma = 0\%$ and 100%, the crystal structure of the lattice leads to predictable and well defined fluctuations in the density. Upon large strains $\gamma = 200\%$ and 300% however, this crystal structure is disrupted leading to a more poorly defined density profile. Nonetheless even in the presence of this noise, the density profile is nearly uniform, without any significant minima that would correspond to necking. The nearly uniform density profile indicates that even at strains of $\gamma = 300\%$, the material does not exhibit features indicative of possible fracture.

As also noted in the main text, even "despite the lack of macroscopic defects (i.e. neck-

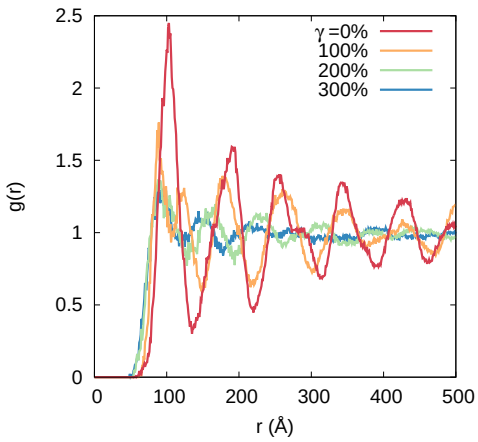

Figure 3.8: Radial distribution function as a function of strain for Sequence B at 273K. Upon deformation, the assembly looses its structure and the peaks within the radial distribution function become less pronounced.

ing), the microscopic structure of the nanoparticle is significantly perturbed under large strains". The microscopic structure is quantified using the radial distribution function, $g(r)$, during deformation. Under no deformation, $\gamma = 0\%$, the DNA-nanoparticle assembly is well ordered and exhibits a strong, long-ranged structure throughout the material. Upon increasing strain, the order within the material decreases and the peaks within the radial distribution function become less pronounced. At the maximum strain, $\gamma = 300\%$, the material has lost nearly all order and the radial distribution function only exhibits a small peak at $r \approx 95\mathring{A}$ corresponding to two particles associating together.

A similar trend is observed in the structure factor, $S(q)$, as a function of strain (Fig. 3.9). At low strains ($\gamma < 100\%$), $S(q)$ demonstrates several distinct peaks corresponding to order throughout the sample. Increasing strain progressively decreases these features until $\gamma = 200\%$, where no peaks in $S(q)$ are observed. Under strain, the location of the first peak in $S(q)$, denoted q^*, demonstrates non-monotoinc behavior: first decreasing and then increasing. This result suggests that under small strains ($\gamma \leq 50\%$), the nanoparticle network expands, whereas at larger strains, the network structure is disrupted and order becomes shorter-ranged.

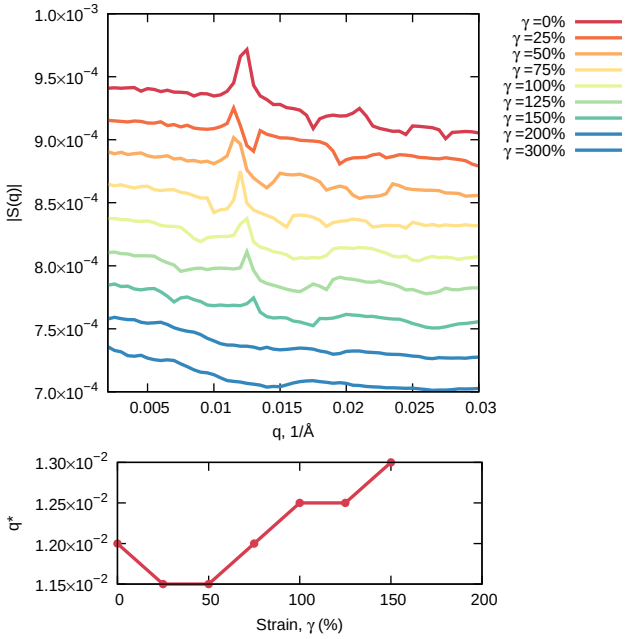

Figure 3.9: Structure factor, $S(q)$, as a function of strain for Sequence B at 273K. Upon deformation, the assembly looses its long range order and the peaks within the structure factor become less pronounced. The position of the first peak is denoted by q^*.

CHAPTER 4

TENSION-DEPENDENT FREE ENERGIES OF NUCLEOSOME UNWRAPPING

4.1 Abstract

Nucleosomes form the basic unit of compaction within eukaryotic genomes and their locations represent an important, yet poorly understood, mechanism of genetic regulation. Quantifying the strength of interactions within the nucleosome is a central problem in biophysics and is critical to understanding how nucleosome positions influence gene expression. By comparing to single-molecule experiments, we demonstrate that a coarse-grained molecular model of the nucleosome can reproduce key aspects of nucleosome unwrapping. Using detailed simulations of DNA and histone proteins, we calculate the tension-dependent free energy surface corresponding to the unwrapping process. The model reproduces *quantitatively* the forces required to unwrap the nucleosome, and reveals the role played by electrostatic interactions during this process. We then demonstrate that histone modifications and DNA sequence can have significant effects on the energies of nucleosome formation. Most notably, we show that histone tails contribute asymmetrically to the stability of the outer and inner turn of nucleosomal DNA and that depending on which histone tails are modified, the tension-dependent response is modulated differently.

4.2 Introduction

Eukaryotic genomes are packaged into a compact, yet dynamic, structure known as chromatin. The basic building block of chromatin is the nucleosome, a disk-like structure consisting of 147 base pairs of DNA wrapped into 1.7 superhelical turns around proteins known as histones[105, 29]. These histone proteins form what is known as the histone octamer, a stable protein complex consisting of two copies of histone proteins H2A, H2B, H3 and H4.

The surface of the histone octamer is highly positive, which interacts favorably with the negative backbone of DNA. As a result, at sufficiently low ionic conditions, nucleosomes are stable and spontaneously form.

The locations of nucleosomes along the genome play a central role in eukaryotic regulation. DNA segments incorporated into nucleosomes are inaccessible to other DNA binding proteins, including transcription factors and polymerases, and thus nucleosomes must be disrupted in order for the cellular machinery to access nucleosomal DNA. As such, the positions occupied by nucleosomes provide an additional, important mechanism by which eukaryotic genomes are regulated. Indeed, past work has demonstrated that deregulation of these processes is implicated in numerous diseases, including cancer[57, 14, 152]. Quantifying the strength of interactions within the nucleosome structure and the forces required to disrupt them is of fundamental importance to understanding the delicate dynamics of chromatin compaction.

Optical-trapping single-molecule techniques have been particularly effective at probing multiple interactions that underlie the nucleosome. In these experiments, chromatin fibers [21, 28, 51, 20] or single nucleosomes [115, 83, 55, 160, 109, 127] are subjected to piconewton scale forces, thereby providing the ability to precisely perturb the native nucleosome structure. By analyzing the deformations that result from these forces, one can infer the underlying strength of binding energies within the nucleosome. Following the initial work by Mihardja *et al.* [115], a consensus is emerging [83, 160, 127] in which a single nucleosome is disrupted in two stages. In the first, at 3 pN, the outer wrap of DNA is removed from the histone surface. This first wrap is removed gradually and is considered to be an equilibrium process, where spontaneous unwrapping and rewrapping events can be observed under a constant force. The second transition occurs at forces 8-9 pN and occurs rapidly via so-called "rips", where the remaining wrap of DNA is suddenly released. More recently, these transitions have been shown to depend on torque (i.e. DNA supercoiling via twist) [160], and to occur asymmetrically due to variability in the bound DNA sequence [127].

Several theoretical and computational studies have sought to help interpret these experimental results. Following the initial work of Kulić and Schiessel[86], most current treatments represent the nucleosome as an oriented spool, and the unbound DNA as a semiflexible wormlike chain [172, 120]. While earlier studies were only able to detect a single distinct unwrapping transition [86, 194], consistent with the first experiments [21], more recent work[172, 120] has been able to reproduce the two transitions observed by Mihardja *et al.* [115]. By relying on simple, primarily analytic models, these studies have provided significant insights into the fundamental physics that govern interactions within the nucleosome. Such approaches, however, have necessarily had to invoke assumptions and introduce adjustable parameters in order to describe experiments [172, 120] (e.g. the DNA-histone binding energy). This limits their ability to predict nucleosomal behavior under different conditions, such as variations in DNA sequence or ionic environment, without resorting to additional experimental data. Additionally, these models cannot explicitly account for histone modifications, which are central to nucleosome positioning and higher-order chromatin structure[68, 88, 11, 207, 20].

A complementary approach, which should in principle enable prediction of nucleosomal interactions under a wide array of situations, could rely on molecular models where the nucleosome can be assembled or disassembled explicitly. Though these approaches are particularly promising, their success has been frustrated by the inability to access the experimentally relevant time scales of stretching, typically rates of 100nm/second. Clearly, these time scales are inaccessible to atomistic simulations, yet even a highly coarse-grained spool-like model of the nucleosome only achieved stretching rates several orders of magnitude too fast[194]. There is therefore a need to develop models and methodologies to facilitate more direct comparisons between optical-trapping experiments and molecular-level calculations. If successful, such models could reveal the subtle, yet incredibly important effects of DNA-sequence and histone modifications on nucleosome stability.

In this work, we build on a recently proposed coarse-grained model of the nucleosome[47, 48, 98] to examine its response to external perturbations. A computational framework is

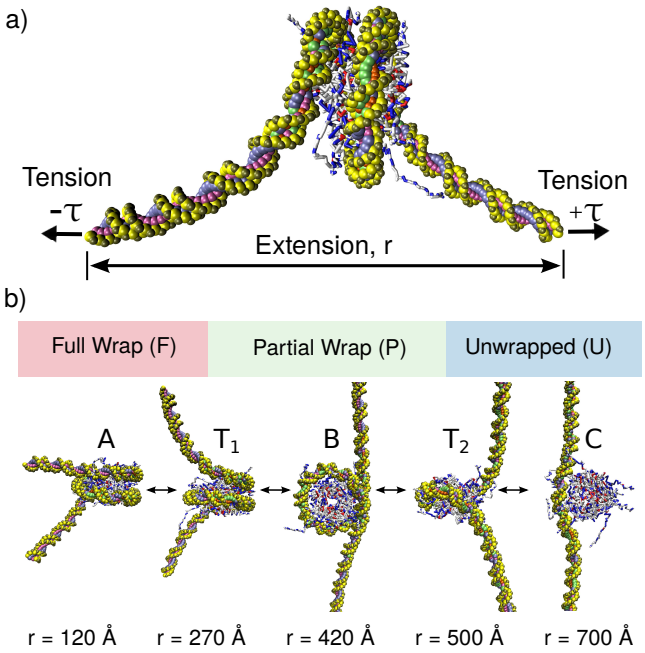

Figure 4.1: Model of Nucleosome Unwrapping. a) Coarse-grained topology of Nucleosome. DNA is represented by 3SPN2.C[47], the histone proteins by AICG[98]. Both the end-to-end extension, r, and tension, τ, are constrained during a simulation. b) Unwrapping process. During extension, the wraps of DNA around histone proteins are removed one by one. T_1 and T_2 denote the transition states separating the first ($A \leftrightarrow B$) and second ($B \leftrightarrow C$) unwrapping events. Figures were generated using VMD[63].

proposed in which the tension-dependent response of the nucleosome is examined at equilibrium, thereby providing access to the free energy of nucleosome unwrapping under tension. Our results are found to be in agreement with experimental measurements by Mihardja *et al.* [115] and Brower-Toland *et al.*[20], and serve to demonstrate that it is indeed possible to reproduce the absolute binding free energies of nucleosome formation in terms of purely molecular-level information, without resorting to additional parameters. Importantly, that model is then used to predict the impact of DNA sequence and histone modifications on the *relative* free energies of binding within the nucleosome.

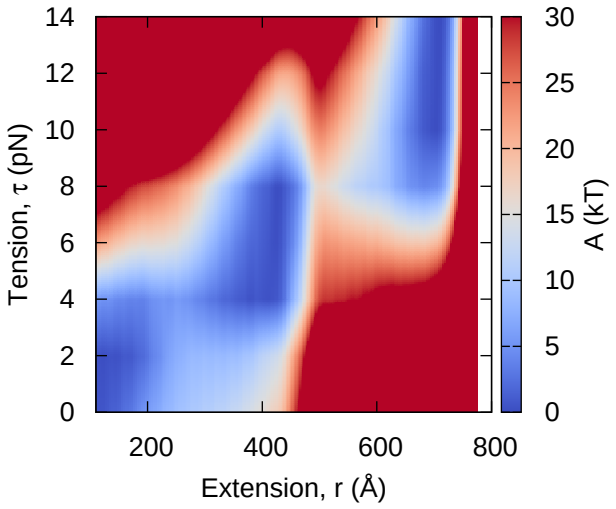

Figure 4.2: Tension-dependent free energy surface of nucleosome unwrapping for 601 positioning sequence. The free energy surface demonstrates minima at extensions of $r \approx 120\mathring{A}, \approx 420, \approx 700\mathring{A}$, depending on tension. As tension increases, the minimum-energy extension shifts to larger values. Consistent with Mihardja *et al.* [115], two transitions are observed.

4.3 Results

A schematic representation of our simulation setup is shown in Figure 4.1a. As with optical-trapping experiments, the "state" of the nucleosome is represented by two parameters: the tension (or force) exerted on the DNA molecule, τ, and the extension of the DNA ends, r. To facilitate comparison with experiments [115], the ends of the DNA are not torsionally constrained. Figure 4.1b shows instantaneous configurations of the nucleosome model for five different values of extension, r. Consistent with previous observations, the outer wrap of the nucleosome is first removed ($A \rightarrow T_1 \rightarrow B$), followed by the inner wrap ($B \rightarrow T_2 \rightarrow C$).

In order to quantify these transitions, we examine the tension-dependent free energy of nucleosome unwrapping. By calculating the tension-dependent free energy instead of a simple force-extension curve, as done previously [194, 75], we can determine the true equilibrium behavior of the unwrapping process. Additionally, by performing simulations at equilibrium, we circumvent the issue of time scales that frustrate comparisons of traditional

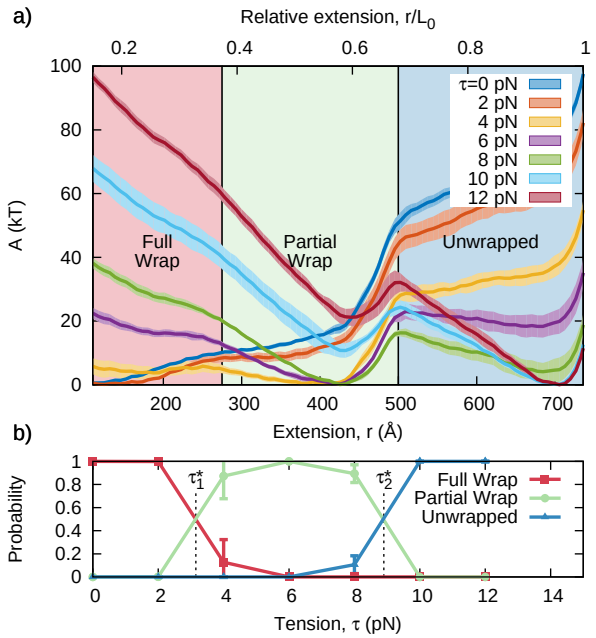

Figure 4.3: a) Free energy versus extension for different values of tension with the 601 positioning sequence. Based on the locations of the transition states, T_1 and T_2, three basins can be defined: "Fully Wrapped", "Partially Wrapped", and "Unwrapped". L_0 represents the contour length of the DNA molecule. b) Probability of observing the nucleosome in each free energy basin for different tensions. The "Fully" and "Partially" Wrapped states are at equilibrium (i.e. equal probability) when $\tau_1^* = 3.2$ pN. The "Partially" and "Unwrapped" states are at equilibrium when $\tau_2^* = 8.9$ pN. Error bars represent standard deviation across four independent simulations.

non-equilibrium molecular simulations to optical pulling experiments.

A representative two-dimensional tension-extension free energy surface for the 601 positioning sequence[103] is shown in Figure 4.2. Rather than increasing linearly with tension, the extension is quantized into three well defined vertical bands, located at $\approx 120\mathring{A}$, $420\mathring{A}$, and $700\mathring{A}$, corresponding to states "A", "B" and "C" in Figure 4.1. At low tension ($\tau < 3pN$), a low extension ($r < 200\mathring{A}$) is preferred. As tension is increased ($\tau \approx 4 - 8pN$), the minimum free energy shifts to intermediate values of extension ($r \sim 420\mathring{A}$). At higher tension ($\tau > 8pN$), the minimum free energy shifts to larger values of extension ($r = 700\mathring{A}$). The free energy penalty of low tension and high extension (e.g. $\tau = 3, r = 700$) or high tension with low extension (e.g. $\tau = 12, r = 200$) results in large energy barriers $> 40\,kT$.

The tension-dependent transition can also be visualized by plotting one-dimensional

"slices" of the free energy surface at different values of tension (Figure 4.3a). Visualizing the data in this way clearly demonstrates that there are three basins of nucleosome extension: "Fully Wrapped", "Partially Wrapped" and "Unwrapped". The basin that is favored depends on the tension applied to the DNA ends. As tension increases, the free energy minima shifts first from the "Fully Wrapped" to the "Partially Wrapped" basin, and then to the "Unwrapped" basin. The boundaries of these basins are defined by the locations of the transition states (i.e. local maximum in the free energy) that separate neighboring basins. The transition states separating the $A \rightarrow B$ transition, T_1, and the $B \rightarrow C$ transition, T_2, are shown in Figure 4.1b.

Once these three basins are defined, we can determine the precise tension at which the outer and inner DNA turns unwrap from the nucleosome. This is obtained by converting the tension-dependent free energy into probabilities, and then integrating these probabilities to determine the total probability of finding the system in each basin (see Methods). The corresponding results are shown in Figure 4.3b; it can be appreciated that the probability of finding the system in the "Fully Wrapped" or "Partially Wrapped" basin is equivalent when $\tau \approx 3.2$ pN. Thus, when $\tau \approx 3.2$ pN the outer turn of nucleosomal DNA is in equilibrium (in a statistical mechanics sense) with its unbound state. We define this tension as τ_1^*. Similarly, the probability of the nucleosome in the "Partially Wrapped" and "Unwrapped" basins is the same (i.e. the inner wrap is in equilibrium when $\tau \approx 8.9$ pN, defined as τ_2^*). These values are in quantitative agreement with those measured by Mihardja *et al.* [115], who observed that the outer and inner DNA loops were removed at 3 pN and 8-9 pN, respectively.

A complementary approach to estimate τ_1^* and τ_2^*, is to determine the tension at which the free energy barriers of the forward and reverse reactions are equal [172]. Figure 4.4 shows the corresponding tension-dependent free energy barriers of the outer ($A \leftrightarrow T_1 \leftrightarrow B$) and inner unwrapping ($B \leftrightarrow T_2 \leftrightarrow C$) events. At low tension, the energy barriers for the forward reactions, $A \rightarrow B$ and $B \rightarrow C$, dominate and the forward (i.e. unwrapping) reaction rate is low. As tension increases, the energy barriers for the forward reactions decrease, while those

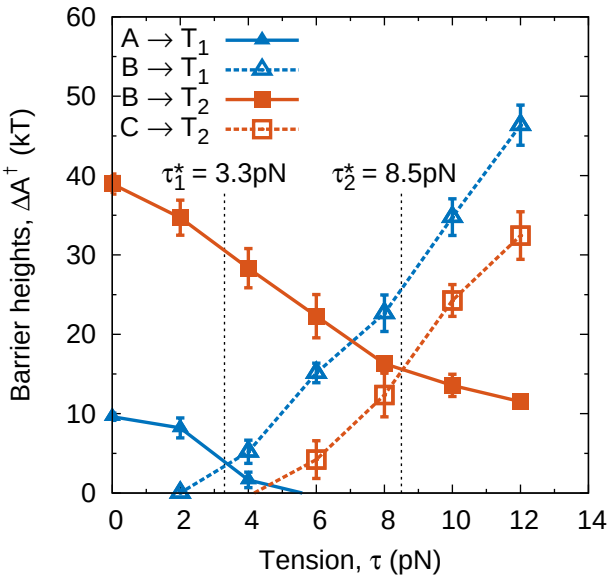

Figure 4.4: Free energy barrier heights of nucleosome unwrapping for 601 positioning sequence. Solid lines represent the unwrapping (forward) reactions, dotted lines represent wrapping (reverse) reactions. When the unwrapping and wrapping barriers are equal, the two basins are at equilibrium with one another. This is found when $\tau_1^* = 3.3$ pN for the outer wrap, and $\tau_2^* = 8.5$ pN for the inner wrap. $\Delta A^\dagger(\tau_1^*) = 4\,kT$ and $\Delta A^\dagger(\tau_2^*) = 16\,kT$. Error bars represent standard deviation across four independent simulations.

for the reverse increase, thereby causing the unwrapping reaction to proceed at a higher rate. When the energy barriers of the forward and reverse reactions are equal, the two basins are at equilibrium (in a transition state theory sense) and τ_1^* and τ_2^* can be determined. These unwrapping tensions are estimated to be 3.3 pN and 8.5 pN, in excellent agreement with the probability-based analysis of Figure 4.3b.

The magnitude of the free energy barrier also helps explain the observation by Mihardja *et al.* [115] that the outer turn of DNA can be removed reversibly, while the inner turn cannot. Since the energy barrier separating the "Fully" and "Partially" wrapped states is only $\approx 5\,kT$, the system can quickly transition between states when held at $\tau = \tau_1^*$. In contrast, the "Partial Wrap" and "Unwrapped" states are separated by an energy barrier of $\approx 18\,kT$, indicating that even at equilibrium the $P \leftrightarrow U$ transition occurs slowly. Thus, removal of the outer wrap may appear to be reversible on the time scales of a typical optical trapping experiment, while the inner wrap may not. Further, because force-extension curves are usually obtained via optical trapping by pulling a nucleosome at a fixed velocity, the experiments may not observe a $P \rightarrow U$ transition until $\tau > \tau_2^*$. This would cause the experiments to overestimate the value of τ_2^*, and lead to a sudden, irreversible "ripping" event. We also note that the barrier estimates in this work ($\Delta A_1^\dagger = 4\,kT, \Delta A_2^\dagger = 16\,kT$) are in excellent agreement to those predicted by Sudhanshu *et al.* [172] ($\Delta A_1^\dagger \approx 6\,kT, \Delta A_2^\dagger \approx 15\,kT$).

Electrostatics, Sequence Dependence, Histone Modifications

Having validated the proposed model against experimental data [115], we now examine the influence of ionic environment, DNA sequence, and histone modifications on the stability of the nucleosome. Such variations can have a significant impact on nucleosome formation, and the precise molecular origins of their impact is still poorly understood.

We first investigate the origins of the tension-dependent mechanical response by exploring the role of DNA-DNA electrostatic repulsion on the stability of the nucleosome structure.

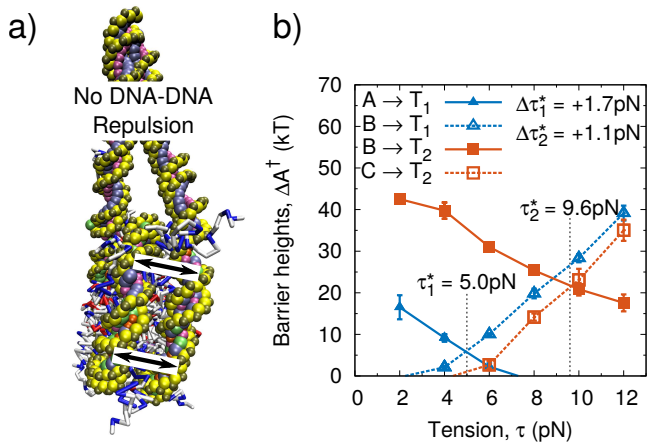

Figure 4.5: a) Schematic representation of proposed model with DNA-DNA repulsion removed. b) Resulting tension-dependent free energy barriers for 601 positioning sequence. $\Delta\tau_1^*$ and $\Delta\tau_2^*$ represent change relative to complete model. Error bars represent standard deviation across three independent simulations.

Past theoretical work [86, 120] has suggested that DNA-DNA repulsion within the nucleosome is central to its tension-dependent response. Other studies, however, have observed that DNA-DNA electrostatic repulsion is unimportant and that the correct response can be achieved by accounting for the tension-dependent orientation of the free DNA ends[172]. Since our proposed model explicitly includes both contributions, we can directly evaluate the importance of DNA-DNA repulsion on nucleosome unwrapping. To examine this effect, we disable DNA-DNA electrostatic repulsion in our model between base-pairs separated by more than 20 base pairs. Only disabling electrostatics between distant regions of DNA was necessary to avoid implicitly lowering the persistence length of DNA by neglecting Coulombic interactions between neighboring base-pairs. All electrostatics responsible for DNA-histone affinity however, remain intact.

Our results are summarized in Figure 4.5a,b. As anticipated[86], removal of DNA-DNA repulsive interactions has a greater impact on the outer DNA loop ($\Delta\tau_1^* = +1.7$pN) than on the inner DNA loop ($\Delta\tau_2^* = +1.1$pN). However in the absence of DNA-DNA repulsions, the qualitative features of the tension-dependent response remain unchanged. These results indicate that while DNA-DNA repulsions play a role in nucleosome disassembly, they are not

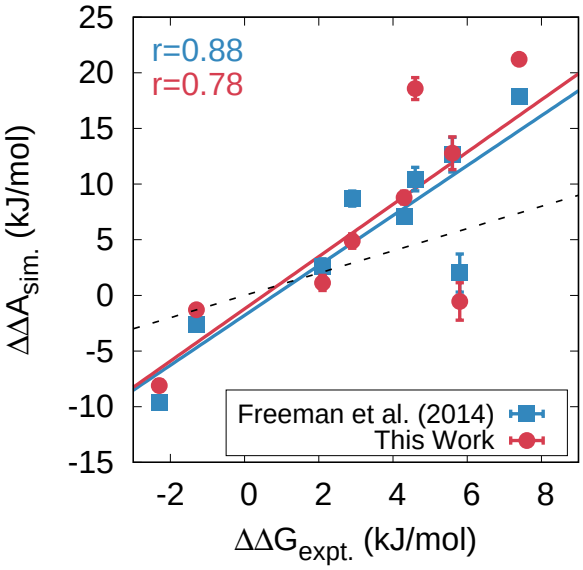

Figure 4.6: Sequence-dependent binding free energies. Squares denote model proposed by Freeman *et al.* [48] (obtained at 300K and 150mM ionic strength). Circles denote model proposed in this work, obtained at 277K and vanishing ionic strength (for consistency with Ref. [177]). Despite differing solution conditions and DNA-Protein interactions, both models reproduce the *relative* binding free energies of nucleosome formation. The DNA sequences used here are given in Ref. [48].

primarily responsible for the two unwrapping steps observed in experiments. Our results are also consistent with prior experimental measurements, where the role of DNA-DNA repulsion on the stability of the outer turn was observed to be small [122].

We next examine the impact of DNA sequence on the relative binding free energies of nucleosome formation. Optical trapping experiments could in principle be used to probe the sequence-dependent energies within the nucleosome, but recent literature studies have been limited to the 601 positioning sequence [115, 83, 160] and slight variations [127]. Instead, competitive reconstitution assays are the dominant experimental technique for characterization of sequence-dependent relative binding free energies[177, 159]. To compare model predictions to these experiments, we use the technique employed by Freeman *et al.* [48], where the relative binding free energies of different DNA sequences are assessed computationally using alchemical transformations and thermodynamic integration (see Methods). A

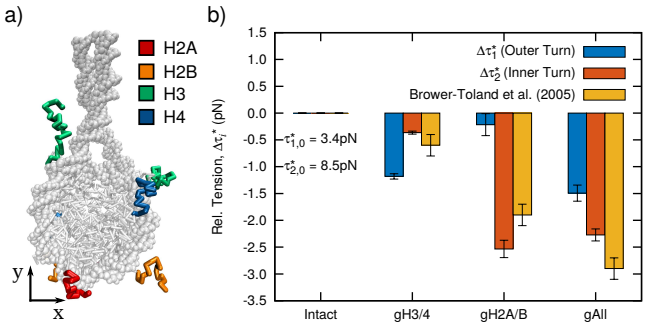

Figure 4.7: a) Molecular configuration highlighting histone tails removed by *in silico* trypsin digest. The exact residues removed are given in the original work by Brower-Toland *et al.*[20]. b) Change in equilibrium tension of outer, $\Delta\tau_1^*$, and inner DNA turn, $\Delta\tau_2^*$, resulting from removal of H3/4 tails (gH3/4), H2A/B tails (gH2A/B), and all histone tails (gAll). Tensions are reported relative to $\Delta\tau_{1,0}^*$ and $\Delta\tau_{2,0}^*$, the values reported previously in Fig. 4.4 for the 601 positioning sequence. Experimental data corresponds to removal of the inner turn of DNA (i.e. $\Delta\tau_2^*$). Error bars represent standard deviation across three independent simulations or reported in Ref. [20].

comparison of predicted and experimental free energies, shown in Figure 4.6, indicates that, as with previous work[48], the model adopted here accurately reproduces the binding free energies of many different sequences. In general, the key predictor of binding free energy is the sequence-dependent shape of the DNA molecule (i.e. minor groove widths and intrinsic curvature). Sequences that bind strongly (low $\Delta\Delta A$) posses periodic sequence motifs (e.g. TA base steps) that impart a shape that favorably "fits" underlying histone structure[159]. In contrast, weakly binding sequences (large $\Delta\Delta A$) do not posses these periodic motifs.

In addition to DNA sequence, the modification of histone tails is widely considered to be the single most important determinant of chromatin structure [68]. Methylated and acetylated histones are enriched at promoters of highly expressed genes and are thought to play a role in the strong positioning of certain nucleosomes [88, 11]. Histone tails are central to nucleosome-nucleosome interactions and their modification has important implications on chromatin's three-dimensional structure[207]. Experiments[20] have also established that removal of histone tails has a significant impact on the stability of the nucleosome.

To examine the role of histone tails on nucleosome stability at a molecular level, we return to our earlier analysis and calculate the tension-dependent free energy of nucleosome

unwrapping. Our results can be compared to the optical trapping experiments of Brower-Toland *et al.* [20], where arrays of 17 nucleosomes were disassembled for different histone tail modifications, including complete removal via trypsin digest or post-translationally via acetylation. In the model, we perform this trypsin digest *in silico* to each histone (see Figure 4.7a), and calculate the resulting tension-dependent response. Figure 4.7b shows the change in the equilibrium tension of the outer, $\Delta\tau_1^*$, and inner, $\Delta\tau_2^*$, turn of DNA due to the removal of different histone tails. The experimental measurements are also included, and correspond to the impact of histone modifications on the stability of the inner turn of DNA (i.e. $\Delta\tau_2^*$). Our results are in excellent agreement with experimental measurements and predict the effect of each histone modification to within ± 0.5pN. Yet our results provide additional insight into to these experiments, whose limited spatial resolution prevented the observation of the individual release of the outer DNA turn. Most importantly, we observe that tails of different histones contribute asymmetrically to the stability of each turn of nucleosomal DNA. The H3/4 tails dominate the stability of the outer DNA turn, $\Delta\tau_1^* = -1.2$pN, but contribute weakly to the stability of the inner turn, $\Delta\tau_2^* = -0.2$pN. In contrast, H2A/B tails have a small effect on the outer turn, $\Delta\tau_1^* = 0.4$pN, but a significant effect on the inner DNA turn, $\Delta\tau_2^* = 2.5$pN. Therefore, depending on whether histone modifications occur on H2A/B or H3/4, the stability of the nucleosome will be modulated differently. This observation suggests a potent mechanism by which each turn of nucleosomal DNA can be independently regulated, and could help to explain the importance and role of H3/4 modifications relative to those of H2A/B.

4.4 Conclusion

In this work we have demonstrated that a molecular-model of the nucleosome, composed of two coarse-grained models of DNA and proteins[48, 47, 98], can be combined parameter-free to accurately reproduce the tension-dependent response of nucleosome unwrapping. This model quantitatively reproduces the unwrapping forces observed in experiments[115, 20] and

the barrier heights predicted by prior theoretical studies [172]. We then demonstrated that this model can be used to examine, without adjustment, the role of subtle phenomena in nucleosome formation such as DNA-DNA Coulombic repulsion, DNA-sequence, and histone tail modifications.

Our proposed approach opens up a new avenue for theoretical examinations of nucleosome stability. As a first step, this model can aid the interpretation of recent optical pulling experiments where the nucleosome is subjected to torque [160], and is suitable for examining subtle features within the nucleosome such as sequence-dependent asymmetric unwrapping [127]. Further, analysis of experimental measurements can become increasingly sophisticated, because our model provides a tool for interrogating raw data, including the fluctuations, from FRET and optical pulling experiments. This combination of experiment and simulation could help to resolve nanometer-scale phenomena such as dynamic DNA-protein contacts and could effectively increase the spatial resolution of experimental measurements to the base-pair level. Yet the potential of our approach extends beyond single nucleosome experiments, and can begin to elucidate many unsolved questions within chromatin biophysics. How does the methylation of specific histone tails (and not others) enhance the positioning of certain nucleosomes? What are the free energies of different folded chromatin structures, and how do histone modifications effect this energy landscape? What is the role of DNA sequence on these processes, and do certain DNA sequences dispose chromatin to different "folds"? The approach presented in this work represents an important step towards answering these questions.

4.5 Methods

The model adopted in this work relies on a coarse-grained model of DNA [47, 48] and proteins [98], which are combined to represent the nucleosome. Both models were developed independently, but they are implemented at the same level of description, thereby facilitating their concerted use. Specifically, for DNA we use the 3SPN coarse-grained representation,

where each nucleotide is described by three force sites located at the phosphate, sugar and base[82, 154, 59, 47]. For the histone proteins, we use the "Atomistic-Interaction based Coarse-Grained model" (AICG), where the protein is represented by one site per amino acid located at the center of mass of the sidechain[98].

Interactions between the 3SPN2.C and AICG models included electrostatic and excluded volume effects. Phosphate sites with 3SPN were assigned a charge of -0.6 as described previously[59]. Each protein site was given the net charge of that residue at physiological pH (i.e. +1 for Arg, Hys and Lys; -1 for Asp and Glu, 0 for others). As with prior work[48], the effective charge of interactions between DNA and protein sites was scaled by a factor of 1.67 to bring the local charge of the phosphates back to -1. We note that DNA-Protein interactions in this work differ slightly from those employed by Freeman *et al.* [48] where, in addition to electrostatics, a small Lennard-Jones attraction was added between all DNA and protein sites. The strength of this attraction was very weak ($\epsilon_{Pro-DNA} = 0.25kJ/mol$) and was originally included to reduce fluctuations within the nucleosome structure. Here we demonstrate that this weak interaction is unnecessary; by omitting it, both the relative and absolute formation free energies of the nucleosome can be reproduced. The combined model is effectively parameter-free: both the model of DNA and Protein are included as originally proposed without any additional terms. Electrostatic forces are approximated by Debye-Hückel theory. Debye-Hückel theory invokes many assumptions about the electrostatic environment, and is not strictly valid for the highly charged association of the histone proteins and DNA. Nonetheless, Debye-Hückel theory provides a useful first-order approximation of coulombic forces and is employed here, without resorting to higher-order techniques. All simulations were performed in the canonical ensemble using a Langevin thermostat and 150mM ionic strength.

As an initial condition, we combine the 1KX5 crystal structure [29] of the nucleosome core particle with a proposed configuration of exiting DNA [65, 114] to form a 223 base-pair structure, with 147 base-pairs bound to the histone proteins and 38 flanking bases on each

side. When using the 601 positioning sequence[103], the flanking bases were chosen as polyA. This configuration was only used as the initial configuration, and no information from either structure was directly encoded into the nucleosome model.

To extract the tension-dependent free energy surface, two constraints were applied to the nucleosomal model. First, a constant force (i.e. tension) was applied to each end of DNA in order to mimic the experimental setup of optical-trapping experiments. Then, harmonic constrains were applied to the end-to-end extension of the DNA molecule, and umbrella sampling was performed to determine the free-energy as a function of DNA extension. In umbrella sampling, many independent simulations are performed at specified locations in phase space and molecular fluctuations are used to estimate the local free energy surface at that location. These many local estimates are then systematically combined to obtain the total free energy surface[87, 74]. Because tension is held at a constant value during a simulation, the resulting free energy "surfaces" are not truly a continuous function of tension. They are instead a compilation of two-dimensional "curves" that are plotted co-currently to construct the "surface" presented in Figure 4.2.

The relative free energy of binding for different DNA sequences ($\Delta\Delta A$) was calculated as described in detail previously [48]. Briefly, a thermodynamic cycle was defined that represents the relative sequence-dependent free energy of nucleosome formation, $\Delta\Delta A$, as the difference between the free energy difference of two DNA sequences in the bulk, ΔA_{bulk}, and bound to the histone proteins, ΔA_{bound} (i.e. $\Delta\Delta A = \Delta A_{bulk} - \Delta A_{bound}$); ΔA_{bulk} and ΔA_{bound} are determined by thermodynamic integration. The DNA sequences analyzed are given explicitly in the original paper.

CHAPTER 5

DIRECT OBSERVATION OF SEQUENCE-DEPENDENT NUCLEOSOME SLIDING

5.1 Abstract

Nucleosomes represent the basic building-block of chromatin, and provide an important mechanism by which cellular processes are controlled. The locations of nucleosomes across the genome are not random, but instead depend on both the underlying DNA sequence and the dynamic action of other proteins within the nucleus. These processes are central to cellular function, and the molecular details of the interplay between DNA sequence and nucleosome dynamics remain poorly understood. In this work we investigate this interplay in detail by relying on a molecular model, which permits development of a comprehensive picture of the underlying free energy surfaces and the corresponding dynamics of nucleosome repositioning. The mechanism of nucleosome repositioning is shown to be strongly linked to DNA sequence, and directly related to the binding energy of a given DNA sequence to the histone core. It is also demonstrated that chromatin remodelers can override DNA-sequence preferences by exerting torque, and the histone H4 tail is then identified as a key component by which DNA-sequence, histone modifications and chromatin remodelers could in fact be coupled.

5.2 Introduction

The basic building block of eukaryotic chromatin is the nucleosome, a DNA-protein complex containing 147 base pairs of DNA wrapped around a disk-like protein complex known as the histone octamer [105]. Since nucleosomal DNA is inaccessible to other DNA-binding proteins, such as transcription factors and polymerases [81, 102, 5], the locations of nucleosomes represent an important mechanism by which cellular processes are controlled. Notably,

nucleosome positions are dynamic, with changes in transcription levels, cellular state, and environmental factors resulting in different packagings of chromatin [183, 158]. Proper packaging of genomic DNA is critical to cellular function, and a wide range of human diseases have been associated with defects in chromatin structure [57, 14]. Understanding the molecular factors that control the locations of nucleosomes, and how they are dynamically modulated, therefore represents a central goal of molecular biology and biophysics.

It is now appreciated that the DNA sequence itself represents a key factor that governs the locations of nucleosomes. Different DNA sequences exhibit different affinities for the histone proteins and, as such, they form nucleosomes with probabilities that can differ by several orders of magnitude [103, 176]. The dependence of nucleosome locations on DNA sequence originates from subtle differences in the intrinsic shape and flexibility of a specific DNA sequence, which lead to favorable electrostatic interactions between the DNA backbone and residues on the histone surface [171]. In fact, this pronounced dependence on DNA sequence has led several authors to propose that a genetic code exists [159, 73] where the positions of 50% of nucleosomes *in vivo* are dictated by DNA-sequence alone. Such a view, however, is not without controversy [203, 167], and a better understanding of the underlying processes must be developed.

Given that DNA sequence is largely constant throughout the life of a cell, other mechanisms must also be at play to achieve the dynamic nucleosome repositioning necessary for cellular function. One part of this dynamic regulation is accomplished by chromatin remodelers, ATP-dependent proteins within the nucleus that actively reposition nucleosomes along the genome [125]. Chromatin remodelers can facilitate many different modifications to chromatin by positioning [193] or removing [43, 189] nucleosomes from promoters or by evenly spacing nucleosomes across sections of DNA [67, 196, 141]. They are central to genetic compaction because, in their absence, nucleosome diffusion is extremely slow, typically on the timescale of hours [137, 113, 46]. Chromatin remodelers accelerate the packaging of DNA by moving nucleosomes away from their sequence-directed binding locations, into

metastable positions. The mechanism by which these metastable nucleosomes relax back to their equilibrium locations represents an important process through which DNA sequence and chromatin dynamics are coupled.

Several studies have sought to elucidate the mechanism by which DNA repositions around the nucleosome. One group of studies has led researchers to propose a "loop propagation" model, where DNA loops are introduced into one side of the nucleosome, and then move along the histone core in an inchworm-like manner [157, 85, 101, 143, 170, 136]. Another group of studies has led authors to propose a "twist diffusion" model, in which a twist defect in the natural helicity of nucleosomal DNA is first introduced, and then diffuses around the nucleosome in a corkscrew-like motion [84, 145, 173, 52]. There is also evidence for a roll-and-slide mechanism, whereby the nucleosome repositions by a sequence of lateral slide displacements [181]. Several single-molecule measurements suggest a repositioning mechanism that cannot be classified into any of these models [16, 32]. There is considerable evidence in support of each model, and it is likely that these repositioning mechanisms are not mutually exclusive but arise depending on other factors, including the DNA sequence [41]. Because chromatin is inherently dynamic, elucidating the mechanisms by which nucleosomes are rearranged, both in the presence and absence of chromatin remodelers, is of considerable scientific interest.

Both sequence and dynamics do matter, but the relationship between these factors has been difficult to elucidate by relying exclusively on experiments. A major obstacle to developing a comprehensive picture of nucleosome repositioning is the lack of sequence-dependent studies on the mechanism of chromatin repositioning. Some of the effects of DNA sequence on nucleosome mobility have been explored [139, 196, 16]. However, the vast majority of studies on nucleosome repositioning [55, 66, 126, 32, 80, 141, 66, 168] have only considered the 601 positioning sequence [103], which exhibits a particularly strong affinity for histones. Such studies have led to valuable insights, but it remains unclear if or to what extent observations pertaining to the 601 sequence can be generalized to other, naturally-occurring,

non-synthetic, DNA sequences. Note that the lack of sequence-dependent studies even extends to high-resolution crystal structures of the nucleosome, which are only available for several DNA sequences [175]. Several models now assume that a bias exists towards nucleosome affinity depending on sequence, but little work [42, 41] has been done to understand the effects of sequence on the dynamics of nucleosome positions, and how sequence might impact the underlying mechanisms.

In this work, we investigate the interplay of DNA sequence and nucleosome repositioning dynamics in detail, by using a detailed molecular model of the nucleosome. By relying on a variety of advanced simulation techniques, we characterize the effect of DNA sequence on both the free energies of nucleosome arrangement, as well as the time scales over which they occur. Our results indicate that different DNA sequences do indeed rely on different mechanisms to reposition, through pathways reminiscent of both the "loop propagation" and "twist diffusion" models described above. However, our results demonstrate that the original formulation of these previously proposed mechanisms is incomplete, and identify several molecular details of the histone surface that play crucial roles in repositioning. Lastly, we examine the effect of applied forces on nucleosomal dynamics, and suggest that nucleosome remodelers can override certain sequence-based positioning preferences by applying torque to nucleosomal DNA. Taken together, our results serve to provide a more comprehensive picture of the effect of DNA sequence on nucleosome repositioning dynamics than was previously available, and will help develop an understanding of the dynamic molecular processes that occur within chromatin.

5.3 Results

We rely on the coarse-grained 3SPN-AICG model of the nucleosome [49, 92]. By combining detailed, fully validated models of DNA [49] and proteins [98], the 3SPN-AICG combination has been demonstrated to accurately reproduce experimental measurements of both the tension-dependent and sequence-dependent binding free energies of nucleosome formation

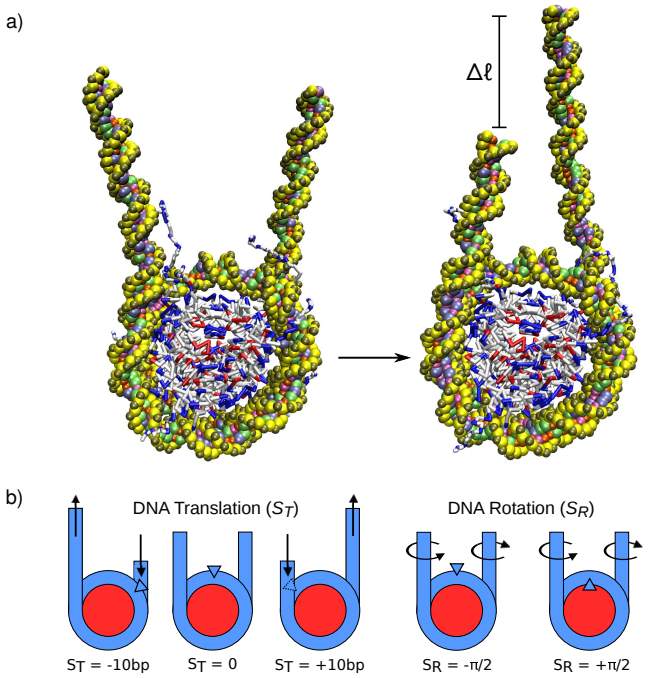

Figure 5.1: a) Molecular configurations of nucleosome repositioning. The chief aim of this study is to characterize the molecular mechanism by which this repositioning occurs. b) Order parameters used to characterize DNA translation, S_T, and rotation, S_R, relative to the histone proteins.

without introducing adjustable parameters [92]. Importantly, no information from the nucleosome crystal structure [29] or locations of DNA-histone contacts is encoded directly into our model, with all energies and dynamics arising naturally from the underlying molecular interactions between the DNA and histone proteins. As such, the 3SPN-AICG model can be viewed as truly predictive, and is an appropriate choice for a molecular-level investigation of DNA repositioning.

Two representative 3SPN-AICG molecular configurations of relevance to nucleosome sliding are shown in Figure 5.1a. They contain 223 base pairs of DNA, of which 147 base pairs are initially incorporated into the nucleosome, with 38 flanking base pairs on each side. Since we are primarily concerned with the mechanism of small rearrangements of DNA (< 20 base pairs), 38 flanking bases were used in order to minimize boundary effects that might arise from the free DNA ends. In order to characterize the degree of nucleosome sliding, an order parameter is defined, S_T, which represents the *translational* position of the central base pair

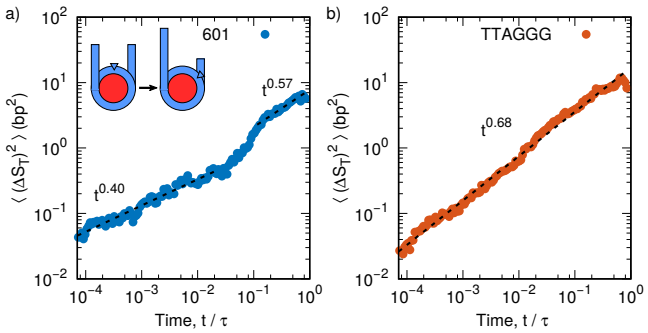

Figure 5.2: Mean-squared displacement of nucleosomal DNA around histone proteins for (a) 601 positioning sequence and (b) TTAGGG repeat. Both DNA sequences exhibit anomalous diffusion, but the 601 positioning sequence contains two dynamic regimes while the TTAGGG repeat only contains one. This difference suggests that DNA with different sequences might translate using different mechanisms. Averaging was performed over 100 independent molecular trajectories.

relative to the histone dyad (Figure 5.1b and Methods). Here S_T is given in units of base pairs (bp), with $S_T \approx 0$ bp corresponding to the native binding position of a DNA sequence, and $S_T = \pm 10$ corresponding to a translation of the DNA by one helical turn forward (+) or backward (-) relative to the histone dyad.

To characterize the timescales at play during nucleosome repositioning, we first compute the mean squared displacement (over short to intermediate times), $\langle (\Delta S_T)^2 \rangle$, of DNA around the histone core for two different DNA sequences (Fig. 5.2). In this analysis, time is reported in units of τ, the characteristic timescale of DNA unwrapping (see Methods). The first sequence, denoted "601", contains the strongly-positioning 601 sequence discovered by Lowary and Widom [103]. In contrast, the second sequence is a "TTAGGG" repeat, a sequence found in the telomeres of human chromosomes [123, 121] that positions nucleosomes poorly [44, 139]. Both sequences are characterized by anomalous sub-diffusion, with scaling exponents ranging from 0.4 to 0.69 for all time scales below τ. Sub-diffusion here was expected; entire nucleosomes are known to reposition slowly on experimental time scales [137, 113, 46] and would not be expected to exhibit the diffusive regime ($\langle (\Delta S_T)^2 \rangle \sim t$) on the time and length scales considered here.

What was not expected however, was the qualitative differences between the dynamics

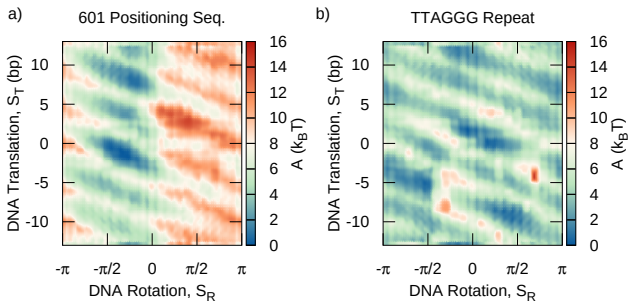

Figure 5.3: Free energy surface of DNA repositioning for (a) 601 positioning sequence and (b) TTAGGG repeat. The 601 sequence demonstrates strong rotational, S_R, and translational, S_T, positioning preferences, whereas the TTAGGG repeat does not.

of the two different sequences. Whereas the TTAGGG repeat is characterized by a single dynamic regime (i.e. $\sim t^{0.69}$ for all times), the 601 sequence is observed to demonstrate two dynamic regimes: a strongly sub-diffusive regime ($\sim t^{0.40}$) for short time scales $< 10^{-2}\tau$, and a weakly sub-diffusive regime ($\sim t^{0.57}$) at longer times $> 10^{-1}\tau$. These results suggest that the 601 sequence might reposition via a slow mode at short times, characterized by little DNA motion, followed by a faster mode at longer times, where DNA repositions more rapidly. More generally, the fact that motion of the TTAGGG repeat exhibits one time scale, whereas the 601 sequence exhibits two, leads us to hypothesize that these two DNA sequences reposition via different mechanisms.

5.3.1 Sequence-dependent Nucleosome Sliding

To explore this possibility, we focus on the free energy surface for DNA repositioning. An additional order parameter is introduced, S_R, which corresponds to the *rotational* orientation of DNA relative to the histone proteins. S_R measures what side of the DNA double helix is facing the histone core, with $S_R = \pm\pi/2$ corresponding to the major groove (+) or the minor groove (-) facing the protein core (see Figure 5.1b and Methods). Combined with S_T, which quantified the translational position of DNA, these two order parameters provide a relatively complete description of the position of DNA wrapped around the histone core.

Figure 5.3 shows the free energy surfaces as a function of S_T and S_R for the 601 sequence

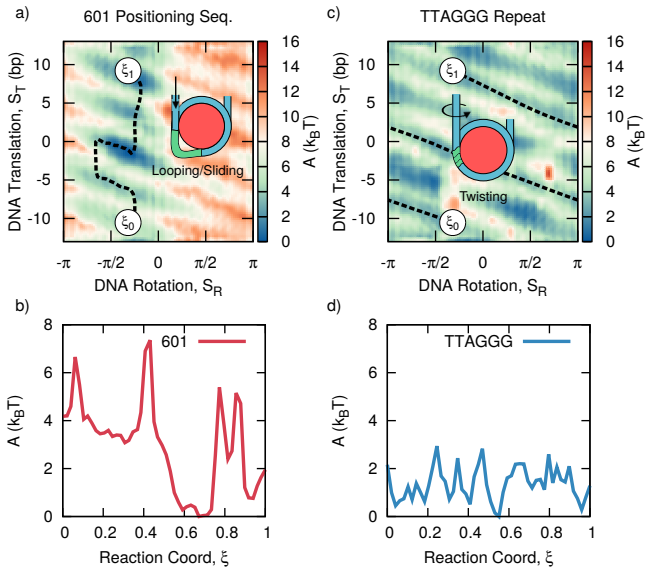

Figure 5.4: Minimum Free Energy Path corresponding to 20 bp of DNA translation for (a-b) 601 positioning sequence and (c-d) TTAGGG repeat. (a) The 601 sequence exhibits a minimum energy path similar to "loop propagation" whereas (b) the TTAGGG repeat exhibits a path characteristic of "twist diffusion". (c) The corresponding energy barriers along this path are large for the 601 positioning sequence, and (d) relatively small for the TTAGGG repeat.

and the TTAGGG repeat. As with the mean-squared displacement measurements, the two free energy surfaces are considerably different. The free energy surface of the 601 sequence reveals a strong tendency for both translational and rotational positioning. More specifically, the free energy minimum corresponds to translational DNA positions with the 601 sequence centered on the nucleosome ($S_T \approx 0$), and with the minor groove facing the histone core ($-\pi/2 > S_R > 0$). DNA movement away from this strongly-bound configurations leads to large free energy penalties. The location of this minimum corresponds very well with the experimentally established minima for the 601 sequence and serves as further validation that the 3SPN-AICG can accurately capture DNA sequence effects within the nucleosome. The free energy surface of the TTAGGG repeat lacks the pronounced translational and rotational positioning preferences of the 601 sequence, and is characterized by a diffuse and relatively flat free energy surface.

Given these differences, it is reasonable to anticipate that these two DNA sequences

would reposition through different mechanisms. In order to infer such mechanisms from the free energy surfaces, we rely on the "String Method", as implemented in the SSAGES package (see Methods), to calculate the minimum energy path corresponding to a 20 base pair rearrangement from ξ_0 to ξ_1 (Fig. 5.4). Since this minimum-energy path corresponds to the most probable transition from ξ_0 to ξ_1, the resulting string can be used to infer the most likely mechanism of DNA rearrangement.

For the 601 sequence (Figure 5.4a), we observe that the mechanism of repositioning of DNA is almost independent of rotational orientation. The minimum energy path is characterized by two distinct modes: one is characterized by increasing S_T at constant S_R, and the other is characterized by a changing S_R at near-constant S_T. Notably, these two modes alternate semi-periodically, with the first mode always followed by the second (and vice versa). An analysis of Figure 5.4b indicates that the regions corresponding to this first mode are associated with large free energy barriers (e.g. $\xi = 0.05, 0.4, 0.8$), whereas regions corresponding to the second mode have a flat free energy surface (e.g. $\xi = 0.2, 0.6$). This mechanism is consistent with the "loop propagation" model described above, where DNA translation is achieved independently of DNA rotation.

In contrast, the minimum energy path for the TTAGGG repeat exhibits a strong coupling between DNA translation, S_T, and DNA rotation, S_R. In fact, the relationship between S_T and S_R observed here corresponds exactly to the 10 base pair pitch of DNA: one complete rotation of DNA, leads to a translation of 10 base pairs. This translation-rotation coupling is similar to the "twist diffusion" model, in which DNA repositions via a cork-screw like motion, where DNA translation is accompanied by DNA rotation in order to maintain minor-groove contacts with the histone protein. The energy surface corresponding to this mechanism is rough, with barriers $\approx 2k_BT$, but largely uniform, with no global translational positioning preferences across the 20 base pair region considered here (Figure 5.4).

In previous work, Schiessel *et al.* have estimated that the energy barriers of DNA loop formation should be $\approx 20k_BT$[157], whereas other work by Kulíc *et al.* estimate the energy

barriers of twist defects to be a strong function of DNA sequence and range from several k_BT to $8 - 10k_BT$ [85]. Other work by Tolstorukov *et al.* has estimated barriers that are much higher at $\approx 50 - 100k_BT$ [181]. Though the energy barriers reported in our work are somewhat lower than these previously reported values, our approach differs significantly from this previous work by making no assumptions about the location of contacts between the DNA and the histone. Because our model permits non-canonical DNA-histone binding, its expected that the energy barriers observed in this work are somewhat lower than those reported elsewhere. Additionally, no other models of nucleosome repositioning have also been simultaneously applied to estimate the free energies of nucleosome unwrapping and nucleosome formation as with the model here [92]. We observe that these measurements are strongly coupled, and that by increasing the energy barriers to sliding, the agreement with other experimental measurements is lost (Figure 5.10). Though it is tempting to fit a model to one experimental metric at the expense of others, a comprehensive view of the nucleosome requires that a wide range of experimental measurements are matched and explained simultaneously by a single model, as in our approach here.

We should also note that the minimum energy paths observed in Figure 5.4 are not strictly equivalent to the "loop diffusion" and "twist defect" mechanisms described above. Since the order parameters, S_R and S_T were defined to give the global translational and rotational position of nuclosomal DNA, they the do not resolve the detailed energies of corresponding to the motions of a single twist or loop defect. Nonetheless, even from this global perspective, the sequence-dependent mechanisms we observe here contain striking similarities to the local mechanisms proposed previously.

The minimum free energy paths observed in Figure 5.4 can also be used to explain the mean-squared displacement measurements discussed previously in Figure 5.2. The single dynamic regime observed for the TTAGGG repeat (Figure 5.2b) is found to correspond to the uniform cork screw-like motion of repositioning, which, from the minimum energy path, appears to be the dominant repositioning mode during DNA motion. The two dynamic

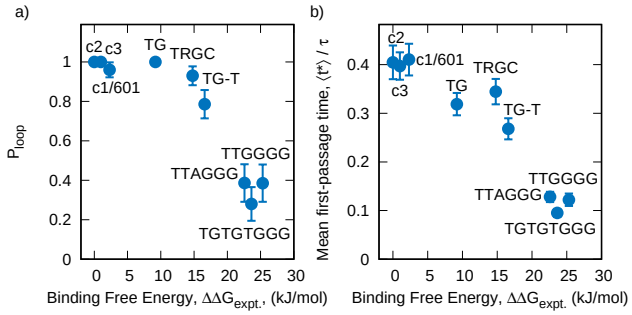

Figure 5.5: (a) Probability of repositioning through a loop-like mechanism, P_{loop}, for different DNA sequences. Stronger binding sequences (low $\Delta\Delta G$) exhibit looping, whereas weaker binding sequences exhibit a twisting mechanism. (b) The mean first-passage time, $\langle \tau^* \rangle$, of a 10 bp DNA translation for a different DNA sequences. Despite the different repositioning mechanisms exhibited by these different sequences, $\langle \tau^* \rangle$ displays a simple, near-linear decay with $\Delta\Delta G$.

regimes observed for the 601 sequence (Figure 5.2a), however, correspond to the two modes of repositioning present in the minimum energy path. We can now interpret the slow repositioning mode ($\sim t^{0.57}$), as corresponding to the formation and propagation of DNA loops, where DNA is translated quickly. The fast repositioning mode ($\sim t^{0.40}$) then corresponds to the small events that proceed the formation of the subsequent DNA loop.

Taken together, the dynamic and thermodynamic (i.e. free energy) evidence indicates that the 601 sequence and the TTAGGG repeat do indeed reposition through different processes, which are reminiscent of the "loop propagation" or "twist defect" mechanism, respectively. Note, however, that the 601 sequence and the TTAGGG repeat represent the two extremes of the DNA sequence affinity for the nucleosome; the free energy difference between the two corresponding nucleosomes is ≈ 20 kJ/mol [103, 44]. It is therefore unclear whether this result can be generalized to other sequences, especially naturally occurring genomic sequences with intermediate affinities for the nucleosome.

To investigate whether a truly sequence-dependent mechanism of repositioning indeed exists, we generated a small library of sequences across the range of binding free energies for the histone core. Binding free energies are given by $\Delta\Delta G$, with smaller values of $\Delta\Delta G$ corresponding to stronger DNA-histone binding. These sequences range from strongly bound

sequences based on the 601 positioning sequence [103, 159] (c1/601, c2, c3), to intermediately bound sequences [162] (TG, TRGC, TG-T), to weakly bound sequences based on other telomeric repeat sequences (TTAGGG in mamals, TGTGTGGG in S. cerevisiae, TTGGGG in Tetrahymena) [22, 44]. For each of these sequences, we simulated 100 independent and unbiased realizations of nucleosome repositioning by first initializing the system at $S_T \approx 0$, and then performing molecular dynamics until $S_T = \pm 10$. The mechanism of repositioning (i.e. looping or twisting), as well as the repositioning time, were examined for each realization. By performing this analysis over an ensemble of trajectories, we were able to generate a probabilistic picture of the different repositioning mechanisms and timescales that are dominant for different DNA sequences.

The mechanism of repositioning for our library of DNA sequences is illustrated in Figure 5.5a. Consistent with our earlier results, it is strongly dependent on the binding free energy, $\Delta\Delta G$ of the DNA sequence. Strongly-bound sequences (low $\Delta\Delta G$) are dominated by the loop propagation repositioning mechanism, with a looping probability, $P_{\text{loop}} \approx 1$. In contrast, weakly-bound sequences, associated with high values of $\Delta\Delta G$, are dominated by DNA twisting. Yet, even though twisting dominates in these sequences, $P_{\text{loop}} = 20 - 40\%$, indicating that DNA looping still occurs. Perhaps the most notable feature of Figure 5.5a is that the transition between these two repositioning mechanisms is smooth, with intermediately binding sequences exhibiting both the looping and twisting mechanism. Importantly, this transition occurs over the range $\Delta\Delta G = 10 - 20$ kJ/mol, which corresponds to almost all naturally occurring DNA sequences. In naturally occurring chromatin, our results therefore indicate that both twisting and looping mechanisms are present simultaneously.

We are not aware of other evidence suggesting that both mechanisms exist simultaneously for naturally occurring DNA sequences and if correct, this result has important implications for the study of chromatin. Experimentally, it suggests that single molecule experiments that rely on the 601 sequences may indeed be missing important physics that characterizes naturally occurring DNA sequences. More broadly, however, the simultaneous presence

of both looping and twisting could represent an important mechanism for the coupling of DNA sequence and chromatin dynamics, with different combinations influencing the three-dimensional structure of chromatin in both complementary and competing ways.

The timescale of repositioning as a function of DNA sequence can also be quantified from this ensemble of trajectories(Figure 5.5b) by the mean first-passage time,$\langle\tau^*\rangle$, corresponding to a transition from $S_T \approx 0$ to $S_T = \pm 10$. That timescale is found to be approximately linear with respect to $\Delta\Delta G$, even though the repositioning mechanism changes significantly over this range; $\langle\tau^*\rangle$ appears to be largely independent of mechanism, and is instead a simple function of the binding strength of the DNA to the histone surface.

A comment on the sheer magnitude of information contained within this ensemble of trajectories is in order. It consists of one hundred independent simulations for 9 different DNA sequences, each of which encompasses at least 10 base pairs of DNA translation. These trajectories consist of 350τ (≈ 4.5 ms) of cumulative simulation time, whose spatial, near-Angstrom level resolution, approaches that of fully atomistic representations. These trajectories are provided through our website (see Methods), and their analysis might yield additional insights into the structure and dynamics of the nucleosome with a spatial and temporal resolution that complements that of single-molecule experiments. In the section that follows, we further analyze this ensemble of trajectories to highlight additional details of pertaining to the loop propagation and twist diffusion repositioning events.

5.3.2 DNA loops are distributed unevenly on the histone surface

Most theories suggest that, during loop propagation, DNA loops are first introduced in the outer wrap of the nucleosome, where the DNA is known to transiently disassociate from the histone core [95]. Such theories suggest that these DNA loops then propagate deeper into the nucleosome, through the dyad, and then exit the nucleosome on the other side, thereby leading to a net DNA translation [157]. In this view, one might expect DNA loops to be, on average, evenly distributed across the histone surface, particularly since loops located at

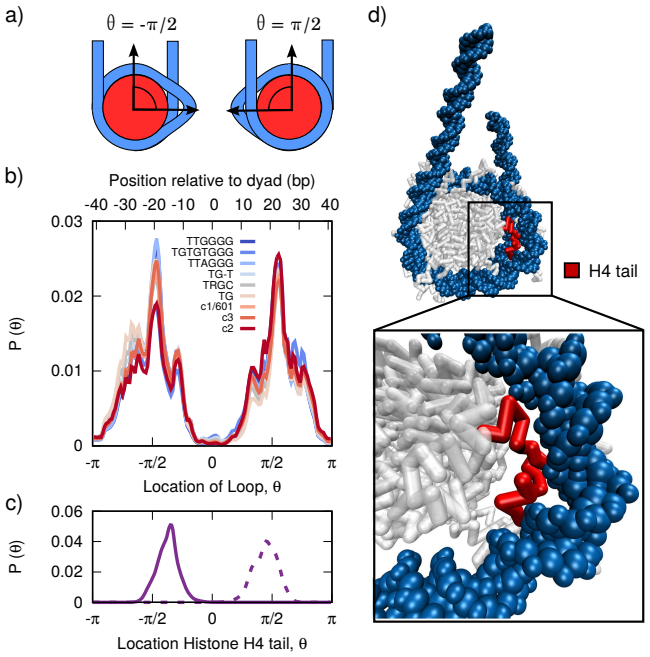

Figure 5.6: (a) Diagram describing the relationship between θ and loops on the histone surface. (b) Distribution of loop positions for different DNA sequences. Loops are distributed unevenly on the histone surface, in a manner largely independent of DNA sequence. (c) Location of histone H4 tail. The H4 tail is co-localized with the location of DNA loops. (d) Molecular snapshot highlighting the role of the H4 tail in stabilizing DNA loops at $\theta \approx \pm\pi/2$, a position ± 20 bp from the histone dyad.

different regions in the nucleosome have comparable energies. However, by analyzing our ensemble of trajectories, we demonstrate that this conventional view of loop propogation is incomplete and that DNA loops are distributed unevenly on the histone surface.

To quantify the position of loops in these trajectories, we first define a variable, θ, that measures the location of loops on the histone surface (Figure 5.6a, Methods). We then construct a histogram of the loop locations extracted from our trajectories, as shown in Figure 5.6b. In contrast to the predictions of the loop diffusion model, we observe that loops are distributed in a highly uneven manner along the histone surface, with different regions on the histone either enhancing or inhibiting the formation of loops. Loops are rarely found at the histone dyad ($\theta = 0$), and are strongly favored at $\theta \approx \pm\pi/2$, a location ± 20 bases from the dyad frequently referred to as SHL$\pm$2. These loops are observed to be ≈ 10 bp in size, and their locations are found to be insensitive to DNA sequence, with each of our nine DNA sequences exhibiting a similar histogram. This result was unexpected, since these different DNA sequences were observed to reposition by dramatically different mechanisms (c.f. Figure 5.5a). This lead us to propose that the location of loops is dictated by a feature of the histone, and not the DNA sequence.

One explanation for the observed distribution of loops could be the variable strength of different DNA-histone binding sites to DNA. Single-molecule experiments [55] have shown that certain regions of nucleosomal DNA are more strongly bound to the histone proteins than others [55]. For the 601 sequence, these experiments identified three regions of strong DNA-histone binding, with the strongest binding observed at the histone dyad, and two other regions at ± 40 bp from the dyad. These three regions correspond exactly with regions in our calculations where looping is absent (Figure 5.6b), implying that regions of strong DNA-histone binding suppress the formation of DNA loops. Additionally, the lack of sequence-dependence in our results suggests that the strong binding regions observed in [55] might be a general feature of nucleosome structure, and not simply a specific feature of the 601 sequence used in the experiments.

The observed distribution of loops could also be related to the ability of histone tails to enhance the formation of loops at certain regions in the nucleosome. The location in the nucleosome where we observe the highest probability of loop formation ($\theta = \pm\pi/2$, SHL$\pm$2) is also associated with the protrusion of the tail of histone H4 from the histone surface [105, 29]. Our molecular trajectories show that the tail of histone H4 is strongly co-localized with the DNA loops, with the highest probability of being located at $\theta \approx \pm\pi/2$ (Figure 5.6c). Beyond mere co-localization however, our simulations suggest that the H4 tails are in fact critical for stabilizing the DNA loops present at $\theta = \pm\pi/2$. When molecular configurations that exhibit loops are visualized, the H4 tail is found to be bent backwards toward the histone core, and to be in contact with the in-facing side of the looped DNA (Figure 5.6d). By adopting that orientation, the H4 tail stabilizes the formation of a loop by restoring the DNA-histone contacts that were originally disrupted. Through this process, the energy of loop formation at SHL$\pm$2 is greatly reduced (and hence the high probability of finding loops at that position).

The H4 histone tail, and the DNA region at SHL$\pm$2, is rapidly becoming a nexus for nucleosome repositioning phenomena. For many years, SHL$\pm$2 has been known to exhibit increased structural variability in crystal structures [105, 38, 18], and is now associated with weak DNA-histone contacts [55]. Furthermore, SHL$\pm$2 is the site of DNA translocation inside the nucleosome [208], and is a known site where SWI/SNF and ISWI remodelers associate [31]. Additionally, the H4 tail is now considered an integral component of chromatin remodeling, with H4 both facilitating remodeling at SHL$\pm$2 [126], and playing a role in the mechanism by which ACF senses linker DNA length [66]. Our results add to this chorus of results by identifying SHL$\pm$2 as a location of prominent DNA looping, through a mechanism that is dependent on the H4 tail. Building on these previous studies, our results suggest a potential for crosstalk between the mechanism of repositioning, the H4 tail, and chromatin remodeler action.

The uneven distribution of loops also has important implications for the mechanism of

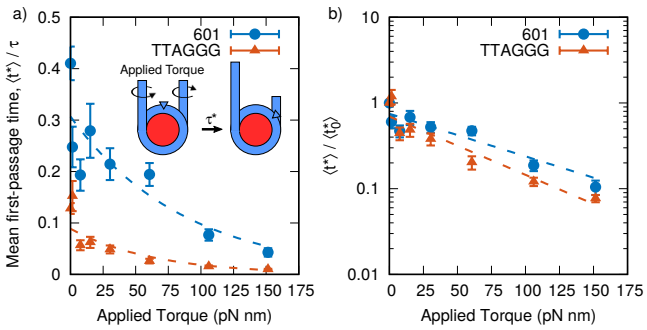

Figure 5.7: (a) Effect of applied torque on mean first-passage times, $\langle \tau^* \rangle$, for 601 sequence and TTAGGG repeat. Small amounts of torque dramatically decrease $\langle \tau^* \rangle$ with a dependence well fit by an exponential decay (dotted lines). (b) Upon renormalizing by the zero-torque value $\langle \tau_0^* \rangle$, both sequences exhibit a similar dependence on applied torque. This suggests a mechanism by which chromatin remodelers, override sequence-positioning preferences.

DNA repositioning. Our observation that loops are infrequently found at certain locations along histone (i.e. $\theta = 0, \pm\pi$), suggests that DNA loops do not slowly propagate across the histone surface as in the loop propagation model, but instead rapidly localize to SHL$\pm$2. In this view, the H4 tail at SHL$\pm$2 could function as a step-wise molecular ratchet, where a DNA loop is (1) pulled into the nucleosome and stored at SHL$\pm$2, (2) pulled across the dyad by the H4 tail on the other side at SHL$\mp$2), and finally (3) released from the nucleosome, thus leading to net nucleosome motion. To our knowledge, our simulations present the first evidence of this type of H4 tail-mediated repositioning event.

5.3.3 Nucleosome mobility can be enhanced through an applied torque

We now turn our attention to twist-diffusion repositioning. Since the results above indicate DNA translation and rotation are coupled in a corkscrew-like motion, chromatin remodelers could be expected to reposition nucleosomes through a "drill"-like mechanism. That is, a chromatin remodeler that applies a purely rotational torque to the DNA molecule could in fact facilitate translational nucleosomal motion. It is also of interest to determine how DNA sequence, especially a sequence's preference for looping or twisting, might influence this motion.

To examine this possibility, we applied a rotational torque to the DNA molecule (see Methods), and observed that a DNA can indeed be repositioned through a drill-like mechanism (SI and Supplemental Movie 5.8). This effect was then quantified by applying a range of rotational torques to the DNA molecule, and measuring the effect on repositioning dynamics. The results are shown in Figure 5.7. For both the TTAGGG repeat and the 601 sequences, repositioning times (as measured by $\langle \tau^* \rangle$) are significantly reduced when torque is applied. Small torques, of ≈ 8 pN nm, are sufficient to reduce the repositioning rate by more than 50%. For larger torques, the decay times follow a simple exponential form (dotted lines). It is of interest to note that when these results are normalized by the value corresponding to zero torque, τ_0^*, both sequences collapse onto a single line (Figure 5.7b). Since the TTAGGG repeat repositions through twisting, we had originally expected that its motion would be significantly enhanced when torque is applied. However, since the 601 sequence repositions through looping, an applied torque was not expected to have a large effect on the rate of repositioning. The fact that both sequences are affected equally by applied torque, suggests that chromatin remodelers might in fact be able to overlook or bypass many sequence effects. By simply applying a torque, remodelers can override the sequence-dependent repositioning mechanisms (i.e. looping or twisting), and force DNA to reposition via a cork-screw like motion. Further, even if the torques applied by these remodelers are small, the dynamics of nucleosomal DNA can be altered considerably.

5.4 Conclusion

The work presented here provides a first step towards the direct observation of both the loop propagation and twist diffusion mechanisms within the nucleosome. Our results indicate that such mechanisms depend on DNA sequence, and that the binding free energy of a given DNA sequence is an excellent predictor of which is the dominating repositioning mechanism. The free energy is also correlated with the characteristic timescale corresponding to a particular repositioning event. It is found that for most naturally occurring DNA sequences, which

exhibit a moderate binding energy, both looping and twisting repositioning mechanisms co-exist. Importantly, a number of previously unknown subtle features have been identified within the nucleosome, including an asymmetric distribution of loops, a strong influence of the H4 tail in their formation, and the dominating effect of torques in the mobilization of nucleosomes.

One of the central findings of this work is that DNA sequence can lead to a wide range of mechanisms and dynamics of nucleosome positioning. Yet, we have also shown that certain features are not influenced by DNA sequence, including the locations of DNA loops, and the effect of torque on nucleosome mobility. Building on these findings, it will be important to pursue experimental studies that go beyond the 601 positioning sequence, and towards naturally occurring, moderately binding DNA sequences. For cases where it is not possible to perform single molecule experiments with genomic DNA sequences, the 3SPN-AICG model presented here can serve as a complementary tool to predict situations where sequence dependence might be important.

5.5 Methods

The model of the nucleosome is identical that that used previously [92], where DNA is represented by the 3SPN.2C model [47] and the histone proteins by the AICG model [98]. The 3SPN.2C model is the latest version of the 3SPN model [82, 154, 59], where DNA is represented by three force sites, located at the center of mass of the phosphate, sugar and base. The 3SPN.2 model has been parameterized to reproduce the melting of double-stranded to single-stranded DNA and correctly predicts the effects of DNA sequence and salt. Additionally 3SPN.2C can reproduce the sequence-dependent curvature and sequence-dependent flexibility of DNA [47]. The "Atomistic-Interaction based Coarse-Grained model" (AICG) used for the histone proteins, represents each amino acid by a single site located at the center of mass of the side chain [98]. Interactions between 3SPN.2C and AICG models are represented only by electrostatic and excluded volume effects as described previously

[92]. Electrostatic forces are introduced at the level of Debye-Hückel theory. All simulations were performed in the canonical ensemble using a Langevin thermostat and 150mM ionic strength. The coarse-grained topology of the 3SPN-AICG model permits the simulation of very long time scales characteristic of nucleosome repositioning.

A Langevin thermostat represents the solvent implicitly, and the absolute magnitudes of the timescales predicted by this model are less informative than the relative ones. As a consequence, it is conventional to normalize the times in a implicit solvent simulation relative to some other timescale of interest, τ. In this work we define τ as the characteristic time scale of spontaneous nucleosomal DNA unwrapping as meausred by experimentally [95, 180]. The details of the definition of τ and how it was computed is given in Figure 5.9.

We also note that the use of Debye-Hückel theory to represent the DNA-histone interactions in the 3SPN-AICG model neglects multi-valent ions and water-mediated hydrogen bonds that have been suggested to mediate DNA-histone contacts [29]. For this reason, the 3SPN-AICG model can only capture this complex electrostatic environment to first order, and may smooth out certain aspects of the nucleosome repositioning energy surface. Higher order approximations are available for the 3SPN model [58], but significantly limit the length and time scales that can be examined.

As an initial configuration to examine nucleosome repositioning, we use the 223 base pair configuration employed previously [92] based on the 1KX5 crystal structure [29] and a proposed structure of exiting nucleosomal DNA [65]. We note that this configuration was only used as the initial configuration, and no information from either structure was directly encoded into the nucleosome model. For the c1/601, c2, c3, TG, TRGC and TG-T sequences used, the 223 base pair sequence was generated by taking the defined 147 bp sequence and periodically appending 38 bp on each side. Specifically, if X_i denotes the identity of the ith base pair of a 147 bp DNA sequence, then the resulting 223 bp sequence would be $X_{109}, X_{110}, \ldots, X_{147}, X_1, \ldots, X_{147}, X_1, \ldots, X_{38}$.

5.5.1 Order Parameters

To characterize the movement of DNA around the histone proteins, we explicitly track the rotational and translational orientation of the DNA relative to the protein dyad. To define the translational orientation of the DNA molecule, we define an order parameter, S_{T}, defined as

$$S_{\mathrm{T}} = \frac{1}{\lambda}\langle \pm \arccos\left(\frac{\boldsymbol{P}\cdot\boldsymbol{P}_0}{\|\boldsymbol{P}\|\|\boldsymbol{P}_0\|}\right)\rangle,$$

where $\mathbf{P}$ is a vector from the center of a base step to the center of the protein, $\mathbf{P_0}$ is the value of $\mathbf{P}$ from the 1KX5 nucleosomal crystal structure [29], and the angle brackets denote an average over base steps at the -15, -5, +5 and +15 positions relative to the dyad. The positive sign is chosen if $(\boldsymbol{P}\times\boldsymbol{P}_0)\cdot\hat{\boldsymbol{f}} \leq 0$ (negative if > 0), where $\hat{\boldsymbol{f}}$ is a vector that points along the center of the nucleosomal DNA superhelix. Therefore positive S_{T}, corresponds to "forward" translation of DNA (towards 5' end), whereas negative S_{T} corresponds to "backwards" motion (towards 3' end) as shown in Figure 5.1b. Lastly, λ represents a conversion factor from radians to base pairs of DNA translation and is defined as $\lambda = 0.08$ rad / bp. λ is calculated by dividing the circumference of the histone proteins [105], $r \approx 42\mathring{A}$, by 2π and the distance between adjacent base pairs, $3.3\mathring{A}$.

The rotational orientation of DNA is defined by an order parameter, S_{R}, given by

$$S_{\mathrm{R}} = \langle \pm \arccos\left(\frac{\boldsymbol{P}\cdot\boldsymbol{B}}{\|\boldsymbol{P}\|\|\boldsymbol{B}\|}\right)\rangle,$$

where $\mathbf{B}$ is a vector from the center of a given base step on the sense strand to its complementary base step on the anti-sense strand, and all other terms are the same as defined for S_{T}. The positive sign is chosen if $(\boldsymbol{P}\times\boldsymbol{B})\cdot\boldsymbol{D} \leq 0$ (negative if > 0), where $\boldsymbol{D}$ is a vector in the 5' to 3' direction along the sense strand. Notably, when $S_{\mathrm{R}} = -\pi/2$, the minor groove is oriented toward the protein core, and when $S_{\mathrm{R}} = \pi/2$, it is oriented away from it (see Figure 5.1b). The order parameter S_{R} has been used previously to quantify the orientational preferences of different DNA sequences to the histone proteins [49].

In order to apply a torque to the DNA molecule, we simply applied a constant force along the S_R order parameter. Since S_R is defined as the average over four different DNA base steps, this applied torque is divided evenly among the base steps at -15, -5, +5 and +15 positions relative to the central base pair. By definition this force results in rotation of the DNA molecule around the histone proteins, and can be converted to a torque using the diameter of the DNA.

5.5.2 Free energy methods

Two dimensional free energy surfaces (Figure 5.3) were obtained along S_{T} and S_{R} using two-dimensional umbrella sampling and WHAM [74, 87]. Error in these free energies were assessed by reconstructing three independent free energy surfaces and computing the standard deviation. The majority of the free energy is found to be accurate to within $\pm 1 k_B T$. The largest errors ($3 k_B T$) correspond to large values of the free energy ($12 k_B T$), leading to an acceptable relative error of $\approx 25\%$ (Figure 5.11).

To determine probable reaction paths along these free energy surface (Figure 5.4), we employ the string method [37, 184], as implemented in the SSAGES package [1]. Rather than randomly guessing initial paths, we used a hybrid approach where the finite-temperature string method [184] was used at a high temperature to generate an ensemble of possible paths, which were then energy minimized using the zero-temperature string method [37]. This hybrid approach was found to yield a much wider range of possible paths than the zero-temperature or finite-temperature string methods alone. Since the free energy surfaces used in the string-method were obtained from a complex molecular system, they are expectedly rough, and many probable paths contained similar energies. Accordingly, the paths reported in Figure 5.4, represent a representative minimum energy path chosen from an ensemble of possible paths predicted using the finite-temperature string.

5.5.3 Quantification of DNA loop locations

To quantify the locations of DNA loops on the histone surface, it was necessary to first define an orthonormal basis to represent the rotational position of the histone proteins. This orthonormal basis consisted of three vectors, $\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}$ where $\hat{\boldsymbol{u}}$ points from the center of mass of the histone to the dyad, $\hat{\boldsymbol{f}}$ (as defined above) points along the nucleosomal DNA superhelix, and $\hat{\boldsymbol{u}} \times \hat{\boldsymbol{f}} = \hat{\boldsymbol{v}}$. After this basis was defined, the center of the DNA helical axis was calculated for each base pair using the Kahn method [71] as described previously [59]. A vector, was then constructed from the center of mass of the protein to the helical axis site of the ith base pair with orientation $\hat{\boldsymbol{w}}_i'$ and magnitude $\ell_i = |\boldsymbol{w}_i'|$. The vector rejection of $\hat{\boldsymbol{w}}_i'$ and $\hat{\boldsymbol{f}}$, $\hat{\boldsymbol{w}}_i = \hat{\boldsymbol{w}}_i' - \hat{\boldsymbol{w}}_i' \cdot \hat{\boldsymbol{f}}$ is computed, where $\hat{\boldsymbol{w}}_i$ is the projection of $\hat{\boldsymbol{w}}_i'$ onto the plane containing $\hat{\boldsymbol{v}}$ and $\hat{\boldsymbol{u}}$, perpendicular to $\hat{\boldsymbol{f}}$. The location of each base pair is then defined by the angle θ_i where $\theta_i = \arccos(\hat{\boldsymbol{w}}_i \cdot \hat{\boldsymbol{u}})$ (see Figure 5.6a). This process characterizes each base pair of DNA by two values, ℓ_i the distance of that base pair to the histone center of mass, and θ_i the location of the base pair relative to the dyad. Note that since DNA is dynamically repositioning, θ_i and ℓ_i are not constant during a simulation. We therefore found it more useful to compute the average distances as a function of theta, $\langle \ell(\theta) \rangle$.

Next to define whether a loop was present, we first construct a normalization curve, $\langle \bar{\ell}(\theta) \rangle$ which characterizes the average distance of base pair at a given θ in the absence of loops. To ensure that loops did not form in this calculation, $\langle \bar{\ell}(\theta) \rangle$ was calculated at a very low salt for a strongly bound DNA sequence. $\langle \bar{\ell}(\theta) \rangle$, is then used to normalize $\langle \ell(\theta) \rangle$ to obtain $\Delta\ell(\theta)$, where $\Delta\ell(\theta) = \langle \ell(\theta) \rangle - \langle \bar{\ell}(\theta) \rangle$. Using this metric, in the absence of DNA loops, when DNA is in close contact with the histone surface $\Delta\ell \approx 0$ for all θ. However when DNA loops form and DNA-histone contacts are disrupted $\Delta\ell > 0$. We define loops as base pairs where $\Delta\ell(\theta) > \delta$, where δ is some predefined threshold. We choose $\delta = 8\mathring{A}$, corresponding to the Debye length at 150mM, the length at which the DNA-histone attraction has significantly decayed. Small variations in the value of δ had little effect on our results.

All simulation codes and results presented in this work are freely available from our

website.

5.6 Timescale of Nucleosome Unwrapping

As stated in the main text, the 3SPN-AICG model uses Langevin dynamics, where the solvent is represented implicitly by a continuum. In Langevin dynamics, the time scale of a simulation is governed exclusively by ζ, the drag coefficient of the solvent. By changing ζ, the dynamics of a system can be scaled almost arbitrarily, and thus the *relative* time scales are more important to consider than the *absolute* ones. For this reason, it is conventional to normalize all times Langevin dynamics calculations relative to a time scale of interest, τ, in order to emphasize that relative time scales are important. In this work, we define τ as the characteristic timescale of spontaneous unwrapping of DNA from the histone core as measured experimentally[95, 180].

In order to calculate τ, we first define a unit vector $\hat{\boldsymbol{u}}$ which points from the center 30th DNA base pair within the nucleosome, to the center of the 1st base pair entering it (see Figure 5.9 snapshots). Because DNA is stiff over the 30 bp length scales, $\hat{\boldsymbol{u}}$ gives the orientation of the exiting 30 bp of nucleosomal DNA. We then calculate the autocorrelation $\langle \hat{\boldsymbol{u}}(0)\hat{\boldsymbol{u}}(t)\rangle$ in order to characterize the timescale of DNA unwrapping (Figure 5.9). We find that the autocorrelation function is fit well to a two-exponential decay, $Ae^{t/\tau_a} + (1-A)e$ where τ_a is a fast timescale corresponding to local fluctuations in $\hat{\boldsymbol{u}}$, and τ is the timescale of interest. Upon fitting, $\tau = 1.4\times10^4$ ns and is used to non-dimensionalize the timescales in the main text.. What emerges from this scaling is that the 10 bp repositioning times observed in our model are comparable to the timescale of unwrapping. This finding is consistent with the hypothesis by Widom and co-workers, that spontaneous DNA unwrapping could be associated with nucleosome looping and remodeling [95].

Note that these calculations were performed using 147 bp of DNA, not the 223 bp of DNA employed in the main text. We observed that the timescales of unwrapping in the 223 bp system to be much slower than and as a result were more challenging to estimate.

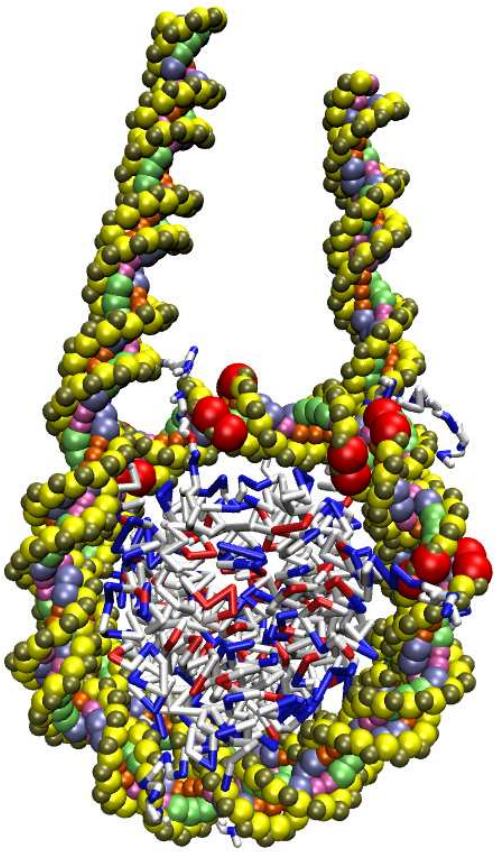

Figure 5.8: [Supplementary Movie S1 Online] Response of DNA to applied torque. A 150 pN nm constant torque was applied to the DNA molecule and the DNA was observed to rapidly reposition through a corkscrew-like motion. The constant torque was applied along the S_R order parameter described in the main text. The definition of S_R includes an average over four base steps at the -15,-5,+5 and +15 positions relative to the central base pair, and therefore the torque was subdivided evenly between these four sites. In order to highlight these sites in the movie, the radii of these four base steps are enlarged and are colored in red. S_R is defined in terms of the histone center of mass and therefore the rotational torque is applied relative to the orientation of the histone.

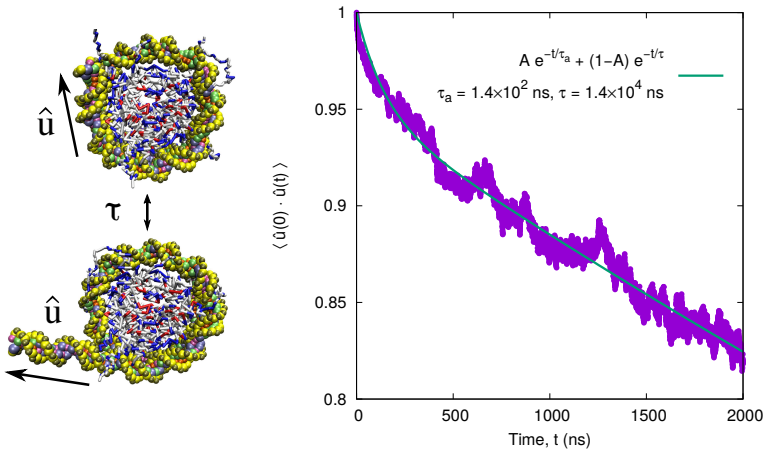

Figure 5.9: Timescale of Nucleosome Unwrapping. The timescale of nucleosome unwrapping is characterized by the autocorrelation function, $\langle \hat{\boldsymbol{u}}(0)\hat{\boldsymbol{u}}(t)\rangle$, where $\hat{\boldsymbol{u}}$ corresponds to the orientation of the 30 bp of entering nucleosomal DNA. The autocorrelation function is fit well to a two-exponential decay, $Ae^{t/\tau_a} + (1-A)e$ where τ_a is a short timescale, and τ is the timescale of interest.

From our preliminary calculations $\tau \approx 4.0 \times 10^5$ ns for this 223 bp system. We also note that the histone proteins were prevented from rotating is these calculations so that the autocorrelation function of $\hat{\boldsymbol{u}}$, was a true measure of DNA unwrapping, and not nucleosome rotational diffusion.

5.7 The energies of nucleosome sliding are not independent from other biophysical measurements

As stated in the main text, the energy barriers of nucleosome repositioning are not an independent quantity, and are strongly related to other biophysical processes within the nucleosome, such as tension-induced nucleosome unwrapping. To demonstrate this relationship, we have performed additional calculations with a slight variant of our model that had a stronger attractions between the DNA and histone proteins, and therefore was expected to have higher energy barriers to repositioning (Fig 5.10). As expected, this modified model had higher repositioning energy barriers of 11 kT, compared to 7 kT in our original model (Fig. 5.10a), and correctly predicts the sequence-dependent formations energies (Fig. 5.10c).

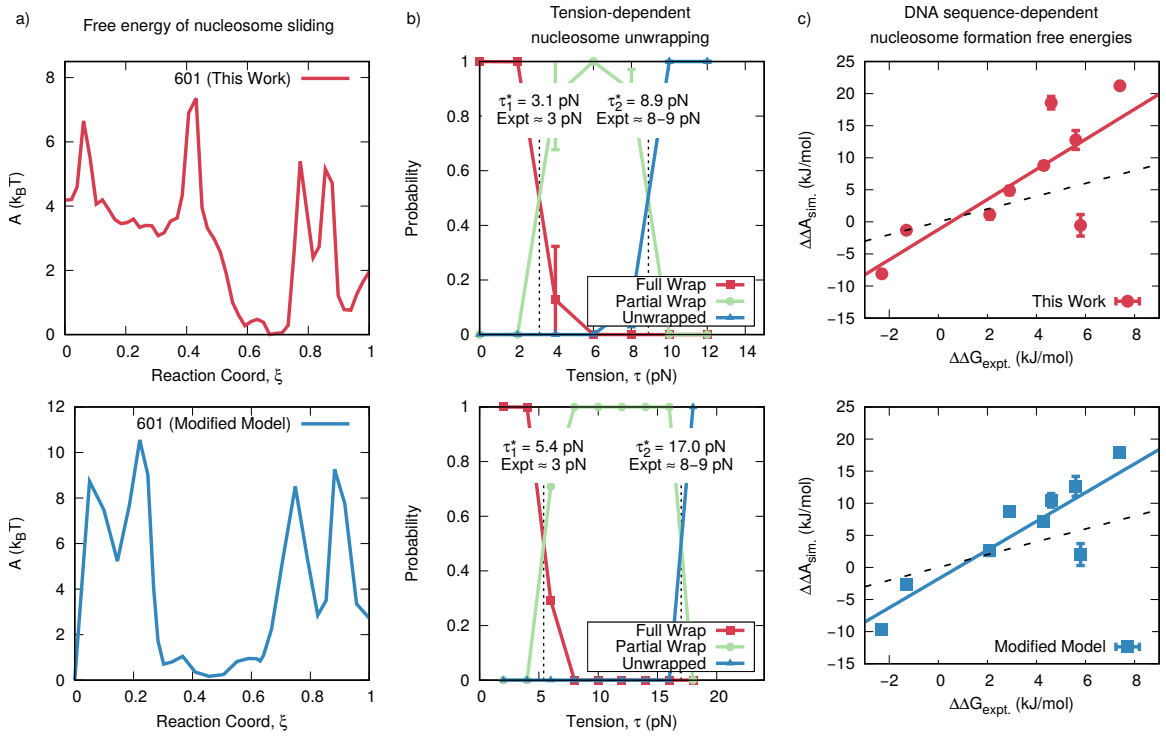

Figure 5.10: Comparison of Model presented in this work (top row), and a modified model with higher repositioning barriers (bottom row). a) Free energies of nucleosome sliding. Top panel is reproduced from main text Fig. 5.4B, bottom panel is calculated using same methods using the modified model. b) Tension-dependent nucleosome unwrapping. Top panel corresponds to Fig. 3B of Ref.[92], bottom panel corresponds to the same calculation using modified model. τ_1^* and τ_2^* correspond to the unwrapping tensions of the outer and inner DNA turns, respectively. Experimental data is from Miharadja et al.[115] . c) Sequence dependent nucleosome formation energies for model used in this work (top), and modified model (bottom). Detailed methods descriptions from b-c can be found in Ref. [92]. The modified model was designed to capture the effect of water-mediated DNA-histone contacts and consisted of a additional weak Lennard-Jones attraction between DNA and histone sites as in Freeman *et al.*[49].

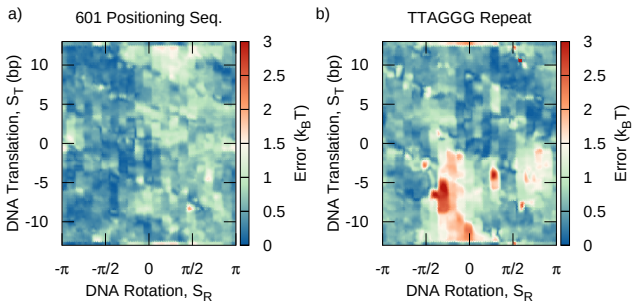

Figure 5.11: Error corresponding to free energy surfaces reported in Figure 5.3. (a) 601 positioning sequence and (b) TTAGGG repeat. Errors were obtained from the standard deviation of three independent free energy surfaces obtained by umbrella sampling, as described in the main text.

An unintended consequence of this modified model however, was a dramatic increase in the tension-dependent unwrapping of the nucleosome (Fig. 5.10b). Whereas our original model demonstrates quantitative agreement with experimentally measured unwrapping forces (top row), this modified model predicts unwrapping forces that are twice as large (bottom row). Though it is tempting to fit one experimental metric at the expense of others, a comprehensive view of the nucleosome requires that a wide range of experimental measurements are matched and explained simultaneously by a single model. The approach presented in this work is, to our knowledge, the only such treatment where this is achieved, and therefore represents important progress in our understanding of the nucleosome.

CHAPTER 6

1CPN: A COARSE-GRAINED MULTI-SCALE MODEL OF CHROMATIN

6.1 Abstract

The compaction of eukaryotic DNA in to chromatin represents an essential, yet poorly understood mechanism for controlling gene expression. Errors during compaction are associated with numerous diseases, and therefore elucidating the molecular factors that control compaction is a central goal of genetics. A significant challenge in studying chromatin compaction is the many length scales involved, typically ranging from angstroms to microns. In this work, we present a multi-scale approach that couples a detailed molecular model of the nucleosome to a coarse-grained mesoscale model of chromatin. We show that this approach can reproduce the dynamics and thermodynamics of available experimental measurements, and is computationally efficient enough to examine the self-assembly of large regions of chromatin. Notably, we show that subtle features of the chromatin fiber, such as the underlying DNA sequence, can have a significant impact on the stability and accessibility of different chromatin structures. The work presented here represents important steps toward understanding the molecular processes that dictate the dynamics and structure of chromatin.

6.2 Introduction

The assembly and compaction of eukaryotic DNA into chromatin represents a critical component of cellular function. Modifications to the processes that dictate chromatin folding, such as histone modifications, represent an additional mechanism by which cellular processes like transcription and replication are regulated. These modifications, known collectively as the epigenome, are central to human health, and there is a rapidly growing body of literature linking these processes with diseases such as cancer[197, 61]. Developing a mechanistic

understanding the processes that dictate DNA compaction, and how these processes can be manipulated, is a central question of molecular biology and biophysics.

In recent years, genome-wide maps of histone modifications have provided an unprecedented view of these modifications, and have presented researchers with molecular-level information about the processes that dictate chromatin compaction. One key finding that has emerged from these studies is that many histone modifications do not act in isolation, but instead interact together collectively to dictate cellular function[199]. By analyzing these genome-wide data sets, various groups have identified strong correlations between different patterns in histone modifications and the resulting function of chromatin [100, 40, 78]. What these studies lack however, is the ability to translate these correlations into a mechanistic understanding of chromatin that explains how these modifications collectively dictate chromatin's three-dimensional structure, and ultimately gene expression.

One approach towards such a mechanistic understanding is so-called chromatin conformation capture techniques which can map the three-dimensional contacts between different parts of the genome[33]. These measurements have provided several surprising observations, such as the fractal structure of chromatin[99, 117], and the existence of topologically associated domains[34]. Though initial experiments could only resolve genome contacts at 1000 kb resolution, recent techniques have achieved resolutions of only one to tens of nucleosomes [144, 62]. Super-resolution microscopy has also been used to directly visualize the chromatin compaction process, shedding light on multiple, previously unknown subtle features of chromatin compaction[186, 13, 17, 185, 35]. Though both chromatin capture and microscopic techniques have significantly advanced our understanding of large scale structures within the genome, they do possess sufficient resolution to resolve the interactions between individual nucleosomes. Resolving these interactions is critical, because it is these length scales on which DNA-protein binding occurs and where histone modifications likely play the largest role[147].

The challenges associated with understanding these length scales of chromatin, at which

many nucleosomes interact together, represents a continuation of a large body of literature trying to understand the so-called chromatin fiber[149, 182]. There is significant experimental evidence for both a "solenoid" model (one-start helix) [45, 190, 112, 148], where linker DNA between neighboring nucleosomes is bent, and a "zig-zag" model (two-start helix) [195, 192, 36, 156, 163], where DNA passes through the center of the fiber. Other work has called the very existence of this fiber *in vivo* into question, suggesting that at physiologically relevant concentrations, a 30 nm fiber may not even exist[110, 50]. Though progress continues to be made in this field[54, 146], understanding chromatin at these length-scales and how these processes can be modified by epigenetic modification, is a significant challenge.

Computational models provide an opportunity to bridge the gap, by providing a link between the small length scales and those observable in experiments. One of the most successful approaches has been developed by Arya, Schlick and co-workers where coarse-grained model of the chromatin fiber has been used to interpret a wide range of experimental measurements[53, 26, 54, 12]. In this model, the nucleosome core is represented by a single rigid body covered with optimized pseudo charges, and linker DNA by a twistable worm-like chain[8] Notably, this model contains flexible histone-tails[8] and the linker histone[7, 106] which mediate inter-nucleosome interactions, as well interactions between nucleosomes and linker DNA. Another successful approach, instead represents the nucleosome with an anisotropic potential, and the conformation of the exiting DNA by a set of angles [77, 165]. This model has been used to interpret the experimentally measured force-extension curves of the chromatin fiber [76], and the energies of different fiber configurations [166]. In order to be computationally tractable, both of these models necessarily make assumptions about the nucleosome, namely that with the exception of the histone tails, the nucleosome is well represented by a fully rigid body. These previous models also rely on Monte-Carlo sampling which gives a good picture of the equilibrium structures, but not the dynamics pathways that connect them.

An alternative approach is to use detailed molecular models, which make no assumptions

about the structure or energies within the nucleosome. By employing detailed coarse-grained models of DNA[47] and proteins[98, 30], these models make very few assumptions about the structure of the nucleosome, and do not not rely on nucleosome crystal structures to bias the configurations sampled by the model. This absence of assumptions makes this approach very powerful, and these models have been used to examine a wide range of phenomena associated with the nucleosome such as nucleosome unwrapping[92, 201], sequence-dependent formation energies[49], and nucleosome repositioning [93]. This lack of assumptions makes these models for the study of higher-order structures of chromatin, as these models naturally incorporate both interactions both within a single nucleosome, as well as interactions between different nucleosomes. Perhaps most importantly, the high resolution of these models can naturally incorporate histone modifications at amino-acid resolution[23], and can be used to examine the wide range of histone modifications known to impact chromatin function[199]. However despite the promise of this approach, computational limitations have limited to these models to small systems consisting of several nucleosomes[23]. A computational approach that can simultaneously capture the physics present in these detailed models, whereas permitting access to large length and time-scales could be very powerful.

In this work here, we present a new model of chromatin that relies on a multi-scale approach to map between different length scales of chromatin. This new model, which we refer to as 1CPN (1 **C**ylinder **p**er **n**ucleosome), represents each nucleosome by a single anisotropic site, and is computationally efficient enough to simulate kilobases of DNA, involving hundreds to thousands of nucleosomes. The interactions parameters in 1CPN have been rigorously determined from detailed free-energy measurements obtained using the 3SPN-AICG model of the nucleosome[49, 92]. The 1CPN model is demonstrated to accurately reproduce inter-nucleosome pair-potentials, the free energies of nucleosome unwrapping, and the sequence-dependent repositioning of nucleosomal DNA. Additionally, 1CPN reproduces the dynamics of chromatin, and can be employed to examine the dynamic folding and rearrangements of the chromatin fiber.

The remainder of this manuscript is structured as follows. In Section 6.3, we discuss the details of the 1CPN model topology and force field We next discuss the methods used to validate the 1CPN model, including the mapping between 1CPN and the 3SPN-AICG model, and the free energy techniques used (Section 6.4). Finally, in Section 6.5 we present the validation of the 1CPN model. Lastly we note that the 1CPN model is implemented in the LAMMPS simulation package[140], and is freely available at our website.

6.3 Model

While formulating the 1CPN model, we began by specifying several key "design goals" or features of the model that would guide our model development, and that we would prioritize during model validation and parameterization. Though each of these features will be discussed in detail through this manuscript, we choose to discuss them in general terms here.

1. **Robust validation and mapping between length scales**. Chromatin is a complex hierarchical material, and no single model can be used to examine all aspects of chromatin biophysics. For this reason, we expect that multi-scale approaches will be increasingly important to examining chromatin structure and dynamics. When designing 1CPN, we designed a topology that could be robustly validated using the detailed 3SPN-AICG model. We also designed a 1CPN topology that could smoothly map between 3SPN-AICG configurations (and vice versa), thereby allowing a multi stage simulation where part of the simulation is run with one length scale of interest, and then continued seamlessly using a different model.

2. **Inter-nucleosome Dynamics.** Chromatin is an inherently dynamic structure where regions of the genome of constantly being expanded or compressed depending on cellular state, the or levels of protein expression. For this reason, we desired a model that could examine the dynamics of chromatin and the various pathways by which chromatin can fold.

3. **Intra-nucleosome Dynamics and nucleosome flexibility.** The dynamics of chromatin go beyond the folding of the chromatin fiber, and extend to the motion of nucleosomes themselves. These dynamics can manifest themselves through the fluctuations of nucleosomal DNA [95, 180], the incorporation of loops and twist [], repositioning of the nucleosomes themselves[125], or dynamic modification of histone tails[199]. We desired a model that could incorporate these features, to move beyond a view where the nucleosome is completely rigid, to one where intra-nucleosome dynamics are considered.

4. **Sequence Dependence.** DNA sequence plays an important role in the position and stability of nucleosomes, and very little work has been performed examining the effect of DNA sequence on the three-dimensional structure of chromatin. When designing 1CPN, we sought a "first-order" approach to this sequence dependence where the effect of different DNA sequences could begin to be examined.

5. **Parsimony.** In order to permit a model that could access the large length- and time-scales of chromatin assembly, we sought a model that was as simple as possible but no simpler. When seeking to reproduce the other design goals, we thought extensively about how to reproduce the desired physics, with the least possible computational expense.

In the remainder of this section, we present the details of the 1CPN model, including the model topology, model force-field, and integration scheme.

6.3.1 Model Topology

All sites in 1CPN are anisotropic, whereby in addition to their positional degrees of freedom, each site also has rotational degrees of freedom as well. As a consequence each site has 12 degrees of freedom, six corresponding to the site's position and velocity, and six corresponding to the site's orientation and angular momentum. The orientation of the ith site is given

uniquely by the quaternion $\boldsymbol{q}_i = (q_0, q_1, q_2, q_3)$, which corresponds to the rotation quaternion that converts from the body-frame to the lab-frame. When discussing the model in the remainder of this section, the orientation of a site is more easily represented by an orthonormal basis $(\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$. This basis is obtained by first defining a reference basis corresponding to the body-frame $(\hat{\boldsymbol{f}}_0, \hat{\boldsymbol{v}}_0, \hat{\boldsymbol{u}}_0)$, and then using

$$\hat{\boldsymbol{f}}_i = \boldsymbol{q}_i \hat{\boldsymbol{f}}_0 \boldsymbol{q}_i^*$$

$$\hat{\boldsymbol{v}}_i = \boldsymbol{q}_i \hat{\boldsymbol{v}}_0 \boldsymbol{q}_i^*$$

$$\hat{\boldsymbol{u}}_i = \boldsymbol{q}_i \hat{\boldsymbol{u}}_0 \boldsymbol{q}_i^*$$

where $\boldsymbol{q}^*$ is the conjugate of $\boldsymbol{q}$, and $(\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$ is the orientation in the lab-frame. The reference basis can be chosen arbitrarily and so in this work, we use $\hat{\boldsymbol{f}}_0 = (1, 0, 0)$, $\hat{\boldsymbol{v}}_0 = (0, 1, 0)$, $\hat{\boldsymbol{u}}_0 = (0, 0, 1)$.

A graphical representation of the 1CPN model is shown in Figure 6.1. In the 1CPN model, three different site types are employed corresponding to nucleosomes, DNA, and the dyad. Graphically, nucleosome sites are represented by a red ellipsoid surrounded by a blue cylinder and centered on a red sphere. DNA sites are represented by blue spheres which connect adjacent nucleosomes, and the dyad sites are represented by yellow spheres within the red ellipsoid. A discretization of one DNA site per three base pairs is chosen and will be discussed in subsequent sections. The orientation of all sites are represented using a local coordinate frame centered at the site's location. The red, green and blue arrows correspond to the $\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}$ vectors, respectively. Note that for the DNA, $\hat{\boldsymbol{v}}$ is omitted for clarity, and can be obtained by $\hat{\boldsymbol{v}} = \hat{\boldsymbol{u}} \times \hat{\boldsymbol{f}}$

A particularly useful visualization of the 1CPN model involves overlaying it on top of the 1KX5 nucleosome crystal structure [29] (Figure 6.1d). From this representation, can be seen that the geometrical position of the DNA sites are located well within the nucleosome core particle. This topology was chosen to permit the partial unwrapping of DNA from the

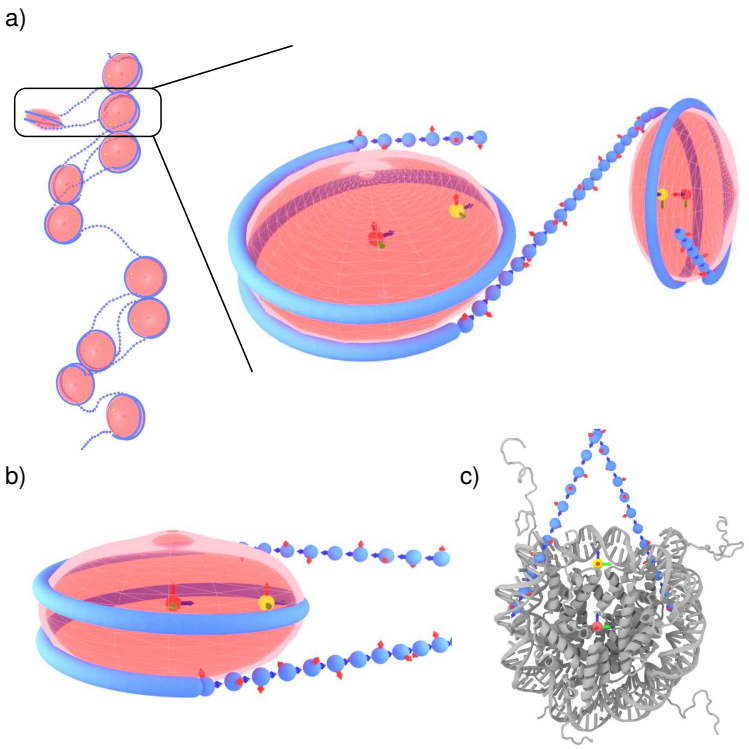

Figure 6.1: Topology of the 1CPN model. Nucleosomes are represented by a single anisotropic site (red ellipsoid with blue tube). DNA is represented as a twistable worm like chain (blue spheres), and the Dyad is included within the excluded volume of the nucleosome in order to stabilize the entering/exiting nucleosomal DNA (yellow sphere). The $\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}$ vectors that denote the orientation of each site are given by red, green and blue arrows, respectively. a) Chromatin fiber and two nucleosomes, b) Single 1CPN nucleosome, c) 1CPN nucleosome overlaid on 1KX5 crystal structure.

histone core and will be discussed in detail in Section 6.3.4. It can also be seed that the dyad site is located approximately at the histone dyad. The dyad site is used to incorporate the effect of the H3 tails on the DNA strands entering and exiting the nucleosome and will be discussed in Section 6.3.2.

We now proceed to discuss the 1CPN force field, and the interactions that govern the potential energy of our model In the following discussion, we divide the total potential energy of the 1CPN model, $U_{\mathrm{1CPN}} = U_{\mathrm{b}} + U_{\mathrm{nb}}$ into the bonded, U_{b} and non-bonded, U_{nb}, interactions. The details of these interactions will now be discussed.

6.3.2 Non-bonded Interactions

Non-bonded interactions are represented by pair-wise interactions where the total non-bonded energy is the sum over all pairs of sites,

$$U_{\mathrm{nb}} = \sum_{i}^{N} \sum_{j \neq i}^{N} U_{ij}. \tag{6.1}$$

The interaction potential between two sites, U_{ij} depends on the site type that i and j represent. The different interaction types for each site pair is given by

$$U_{ij} = \begin{cases} U_{\mathrm{Zewdie}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j\,;\, \sigma_0, \epsilon_0) & \text{if } i \in \text{Nucl.}, j \in \text{Nucl.} \\ U_{\mathrm{Zewdie-LJ}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_i;\, \sigma_0', \epsilon_0') & \text{if } i \in \text{Nucl.}, j \in \text{DNA} \\ U_{\mathrm{Zewdie-LJ}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_j;\, \sigma_0', \epsilon_0') & \text{if } i \in \text{DNA}, j \in \text{Nucl.} \\ U_{\mathrm{elec}}(\boldsymbol{r}_{ij}) & \text{if } i \in \text{DNA}, j \in \text{DNA} \\ U_{\mathrm{gauss,aniso}}(\boldsymbol{r}_{ij};\, \hat{\boldsymbol{f}}_i, \hat{\boldsymbol{u}}_i) & \text{if } i \in \text{Dyad}, j \in \text{DNA} \\ U_{\mathrm{gauss,aniso}}(\boldsymbol{r}_{ji};\, \hat{\boldsymbol{f}}_j, \hat{\boldsymbol{u}}_j) & \text{if } i \in \text{DNA}, j \in \text{Dyad} \\ 0 & \text{else} \end{cases} \tag{6.2}$$

where $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$, and $\hat{\boldsymbol{f}}_i$, $\hat{\boldsymbol{u}}_i$, $\hat{\boldsymbol{f}}_j$ and $\hat{\boldsymbol{u}}_j$ are the orientation vectors of sites i and j

respectively.

The different pair potentials, U_{Zewdie}, $U_{\text{Zewdie-LJ}}$, U_{elec}, and $U_{\text{gauss,aniso}}$ are defined as follows.

Zewdie Potential

Interactions with nucleosome sites are given by the Zewdie potential [200], an extension of the widely used Gay-Bernne potential, that has been shown previously to be a good representation of the nucleosome [77, 165]. The Zewdie potential between sites i and j is given by

$$U_{\text{Zewdie}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j\,;\,\sigma_0, \epsilon_0) = 4\epsilon\left[\left(\frac{\sigma_0}{r_{ij}-\sigma+\sigma_0}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}-\sigma+\sigma_0)}\right)^{6}\right] \tag{6.3}$$

where $r_{ij} = |\boldsymbol{r}_{ij}|$, σ_0 and ϵ_0 are parameters, and $\sigma = \sigma(\hat{\boldsymbol{r}}_{\boldsymbol{ij}}, \hat{\boldsymbol{f}}_{\boldsymbol{i}}, \hat{\boldsymbol{f}}_{\boldsymbol{j}})$ and $\epsilon = \epsilon(\hat{\boldsymbol{r}}_{\boldsymbol{ij}}, \hat{\boldsymbol{f}}_{\boldsymbol{i}}, \hat{\boldsymbol{f}}_{\boldsymbol{j}})$ incorporate the orientation dependence of the potential. In the Zewdie potential, σ and ϵ are represented by the first six terms of a S-function expansion[169] and are defined as

$$\sigma = \sigma_0(\sigma_{000}S_{000} + \sigma_{cc2}(S_{022}+S_{202}) + \sigma_{220}S_{220} + \sigma_{222}S_{222} + \sigma_{224}S_{224}) \tag{6.4}$$

$$\epsilon = \epsilon_0(\epsilon_{000}S_{000} + \epsilon_{cc2}(S_{022}+S_{202}) + \epsilon_{220}S_{220} + \epsilon_{222}S_{222} + \epsilon_{224}S_{224}) \tag{6.5}$$

.

In this formulation, the orientation dependence of the potential is embedded in the terms S_{000}, S_{022}, S_{220}, S_{222}, and S_{224} and are defined in Table 6.1. The range and strength of the Zewdie parameter is controlled by the parameters σ_0 and ϵ_0, respectively, while the anisotropic shape of the potential is given by parameters σ_x and ϵ_x for $x = \{000, cc2, 220, 222, 224\}$. A key advantage of this S-function expansion is that these differ-

$$a_0 = \hat{\boldsymbol{f}}_i \cdot \hat{\boldsymbol{f}}_j,\ a_1 = \hat{\boldsymbol{f}}_i \cdot \hat{\boldsymbol{r}}_{ij},\ a_2 = \hat{\boldsymbol{f}}_j \cdot \hat{\boldsymbol{r}}_{ij}$$
$$S_{000} = 1,\ S_{202} = (3a_1^2 - 1)/2\sqrt{5},\ S_{022} = (3a_2^2 - 1)/2\sqrt{5},\ S_{220} = (3a_0^2 - 1)/2\sqrt{5},$$
$$S_{222} = \frac{1}{\sqrt{70}}(2 - 3a_1^2 - 3a_2^2 - 3a_0^2 + 9a_0a_1a_2),$$
$$S_{224} = \frac{1}{4\sqrt{70}}(1 + 2a_0^2 - 5a_1^2 - 5a_2^2 - 20a_0a_1a_2 + 35a_1^2a_2^2)$$

Table 6.1: Definition of Orientation Dependent Terms in the Zewdie Potential. $\hat{\boldsymbol{f}}_i$ and $\hat{\boldsymbol{f}}_j$ represent the orientation of the ith and jth sites. $\hat{\boldsymbol{r}}_{ij}$ is the normalized vector pointing from site i to site j.

ent ϵ and σ parameters set the energy scales of different directions of approach between particles[200]. For example, the second and third terms S_{022} and S_{202} increase as particle orientations $(\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j)$, align with the inter-particle vector $\boldsymbol{r}_{ij}$. As a result, the value of ϵ_{cc2} controls preference of particles to stack column-wise. Similarly, S_{220} is only a function of particle orientation (e.g. $\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j$), and so ϵ_{220} controls the preference of particles to form a nematic phase. Higher order effects such as quadrupoles, are captured by the terms S_{222} and S_{224}.

The flexibility the expansion makes the Zewdie potential well-suited to represent a complex disk-like bio-molecule like the nucleosome. The 12 parameters in the Zewdie potential are fit to anisotropic pair-potentials obtained using our detailed 3SPN-AICG model of the nucleosome and will be discussed in detail in Section 6.5.1.

Interactions between DNA and nucleosome sites are also represented using the Zewdie potential. However the since non-bonded interactions between DNA sites are isotropic (i.e. DNA sites are spheres), and a slight modification is required and is denoted by $U_{\text{Zewdie}-\text{LJ}}$ In $U_{\text{Zewdie}-\text{LJ}}$, torques are only applied to the anisotropic nucleosomes and therefore are only a function of nucleosome orientation. In practice, this is achieved by setting $\hat{\boldsymbol{f}}_{DNA} = \hat{\boldsymbol{r}}_{ij}$ and $a_0 = a_2$ and solving then solving U_{Zewdie} as usual. Additionally, due to the different radii of the nucleosome and DNA sites, different values of σ_0 and ϵ_0 must be chosen. Here we use an arithmetic mixing rule to obtain $\sigma_0' = (\sigma_{DNA} + \sigma_0)/2$. The value of ϵ_0 is chosen to be small so that Nucleosome-DNA interactions are negligibly attractive, and instead interact only via an excluded volume.

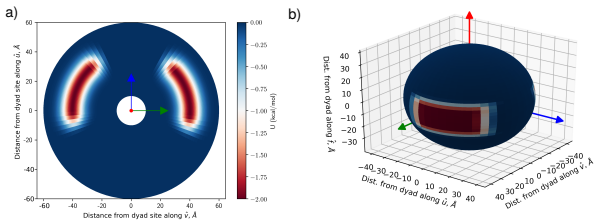

Figure 6.2: Pair-wise Dyad Potential. Potential energy of $U_{\text{gauss,aniso}}$ when a) projected onto the $\hat{\boldsymbol{u}}$-$\hat{\boldsymbol{v}}$ plane, b) plotted as a 3-d surface for $r = r_0$. In both panels, the dyad site is located at the origin and its orientation $(\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$ given by the red, green and blue arrows, respectively.

Past work using the Zewdie potential [200, 77, 165] has been limited to Monte Carlo simulations which, by construction, did not require computation of the forces and torques. In order to simulate the Zewdie potential using molecular dynamics, calculation of the forces and torques was necessary, and are given in the Appendix 6.7.1.

Non-bonded Dyad Interactions

Though the Zewdie potential provides a good representation the interactions between nucleosomes and DNA in general, the Zewdie potential alone neglects additional, and important interactions that occur within chromatin. The most important of these interactions is the role of the histone H3 tail at stabilizing DNA entering and exiting the nucleosome. Previous experimental measurements[20] and simulations[92] indicate that histone tails H3 and H4 stabilize the outer wrap of nucleosomal DNA and that the histone H3 tail is important for screening repulsive interactions between entering and exiting DNA[6].

In order to incorporate these interactions into our model, we introduce an additional pair-wise interaction between the dyad sites and DNA sites. Coarse-graining these complex interactions between DNA and histone tails into a single pair potential was challenging, especially because these interactions are highly localized near the entering/exiting DNA. In order to derive a suitable potential, we took inspiration from the base-pairing interactions in the 3SPN.2 model, where hydrogen bonds between complementary bases were given by

an anisotropic morse potential[59]. By adopting a similar approach for 1CPN, we define the dyad potential, U_{dyad}, as an anisotropic gaussian potential (Figure 6.2) that is given by the functional form

$$\begin{aligned} U_{\text{gauss,aniso}} &= f(K_\theta, \Delta\theta) f(K_\phi, \Delta\phi)\, U_{\text{gauss}} \\ &= f(K_\theta, \Delta\theta) f(K_\phi, \Delta\phi) \left(-d_0 e^{-(r-r_0)^2/2\sigma^2} \right) \end{aligned} \tag{6.6}$$

where r is the distance between the DNA and dyad sites, r_0 is the equilibrium distance between the sites, d_0 and σ control the depth and width of the gaussian well, and $f(K_\theta, \Delta\theta)$ and $f(K_\phi, \Delta\phi)$ are modulating functions that scales from zero to unity, depending on the relative orientation of the DNA and dyad sites. As with the 3SPN.2 model, the modulating function is chosen to be

$$f(K_\theta, \Delta\theta) = \begin{cases} 1 & -\frac{\pi}{2K_\theta} < \Delta\theta < \frac{\pi}{2K_\theta} \\ 1 - \cos^2(K_\theta \Delta\theta) & -\frac{\pi}{K_\theta} < \Delta\theta < -\frac{\pi}{2K_\theta} \text{ or } \frac{\pi}{2K_\theta} < \Delta\theta < -\frac{\pi}{K_\theta} \\ 0 & \Delta\theta < -\frac{\pi}{K_\theta} \text{ or } \Delta\theta > \frac{\pi}{K_\theta} \end{cases} \tag{6.7}$$

and similarly for $f(K_\phi, \Delta\phi)$, where $\Delta\theta = \theta - \theta_0$ and $\Delta\phi = \phi - \phi_0$, and θ_0 and ϕ_0 are reference angles. Here $\cos\theta = \hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{r}}$ is the polar angle between the normalized bond vector $\hat{\boldsymbol{r}}$ and the dyad orientation vector $\hat{\boldsymbol{u}}$ and $\cos\phi = \hat{\boldsymbol{f}} \cdot \hat{\boldsymbol{r}}$ is the azimuthal angle between $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{f}}$. Though the choice of the other $U_{\text{gauss,aniso}}$ parameters will be given in Section 6.5.3, the approximate symmetry of the exiting and entering DNA yields $\phi_0 = 90°$. To aid in interpreting this potential, different projections of $U_{\text{gauss,aniso}}$ are given in Figure 6.2. A comparison of the locations of the energetic minima in Figure 6.2 to the 1CPN-1KX5 overlay in Figure 6.1, gives a qualitative justification for the functional form of $U_{\text{gauss,aniso}}$; this potential it seeks to account for the role of the H3 tail in stabilizing certain configurations of entering/exiting DNA from the nucleosome. Before proceeding we note that the orientation of the DNA site

does not enter the potential, and $U_{\text{gauss,aniso}}$ is only a function of the vector connecting the two sites r, and the orientation of the dyad site.

Non-bonded DNA Interactions

Non-bonded interactions between DNA sites are represented by Debye-Huckel theory using a screened electrostatic potential, U_{elec}, and is given by,

$$U_{\text{elec}} = \frac{q_i q_j e^{-r_{ij}/\lambda_D}}{\epsilon r_{ij}} \tag{6.8}$$

where q_i and q_j are the charge of sites i and j, r_{ij} is the inter-site separation, λ_D is the Debye length, and ϵ is the solution dielectric constant. Since each DNA site represents three base-pairs, a charge of -3 is given to each DNA bead. The Debye Length is defined as

$$\lambda_D = \sqrt{\frac{k_B T \epsilon_0 \epsilon(T,C)}{2 N_A e_c^2 I}} \tag{6.9}$$

where N_A is Avogadro's number, and I is the ionic strength of the solution. The solution dielectric constant is a function of monovalent salt concentration, C, and temperature, T, such that $\epsilon(T,C)$. The salt and temperature dependence can be assumed to be independent such that

$$\epsilon(T,C) = \epsilon(T)a(C), \tag{6.10}$$

where

$$\epsilon(T) = 249.4 - 0.778T/K + 7.20 \times 10^{-4}(T/K)^2 \tag{6.11}$$

and

$$a(C) = 1.000 - 2.551C/M + 5.151 \times 10^{-2}(C/M)^2 - 6.889 \times 10^{-3}(C/M)^3 \tag{6.12}$$

as in Ref [59]. The details and justification for this treatment are given in a previous

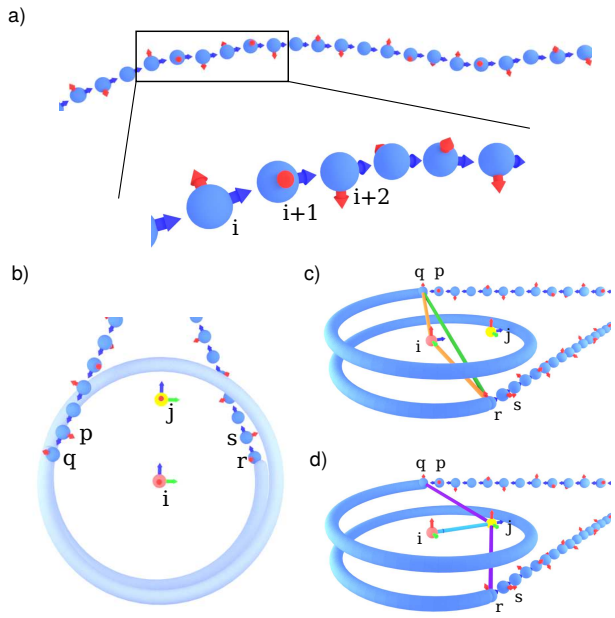

Figure 6.3: Bonded Interactions in the 1CPN Model. a) DNA is represented by a twistable worm like chain as defined in Equation 6.14. The $\hat{\boldsymbol{f}}$ and $\hat{\boldsymbol{u}}$ vectors of each DNA bead are given by the red and blue arrows. b) Bonded interactions within the nucleosome. Indicies i, j denote the nucleosome and dyad sites, while p, q, r, s denote the DNA sites involved in the bonded nucleosome interactions. These indices will be used throughout Section 6.3.3 while discussing the various aspects of the 1CPN forcefield. c) Bonded interactions within U_{NDNA} as defined in Equation 6.19. d) Bonded interactions within U_{dyad} as defined in Equation 6.24.

publication [59].

6.3.3 Bonded interactions

The bonded potential, U_b, includes interactions between sites, like bonds, angles and dihedrals, that maintain the topology of the model. Since all sites in 1CPN have both a position and orientation, the bonded interactions in 1CPN include interactions that constrain both the relative position and relative orientation of bonded sites. In our discussion here, the bonded interactions will be discussed in two parts, first those interactions within the DNA molecule, U_{DNA}, and secondly the interactions within the nucleosome, U_{Nucl}. The total bonded energy is the sum of these two contributions,

$$U_{\rm b} = U_{\rm DNA} + U_{\rm Nucl}. \tag{6.13}$$

DNA bonded interactions

Interactions within the DNA molecule are represented by a twistable worm-like chain[25]. This potential represents DNA by a series of bonded sites, $U_{\rm bond}$ that interact through angle potentials that constrain the bending rigidity, $U_{\rm angle}$, and the torsional rigidity, $U_{\rm twist}$, of the DNA polymer (Figure 6.3). For a DNA chain consisting of N beads, the potential, $U_{\rm DNA}$, is given by

$$\begin{aligned} U_{\rm DNA} &= U_{\rm bond} + U_{\rm angle} + U_{\rm twist} + U_{\rm align} \\ &= \frac{k_{\rm b}^{\rm D}}{2\ell_{\rm b,0}} \sum_{i=1}^{N-1} (\ell_i - \ell_{\rm b,0})^2 + \frac{k_\beta}{\ell_{\rm b,0}} \sum_{i=1}^{N-2} (1 - \cos(\beta_i)) \\ &\quad + \frac{k_\gamma}{\ell_{\rm b,0}} \sum_{i=1}^{N-1} (1 - \cos(\omega_i - \omega_0)) + \frac{k_\psi}{\ell_{\rm b,0}} \sum_{i=1}^{N-1} (1 - \cos(\psi_i)) \,, \end{aligned} \tag{6.14}$$

where $\ell_i = ||\boldsymbol{r}_{i+1} - \boldsymbol{r}_i||$ is the distance between adjacent sites, $k_{\rm b}^{\rm D}$ is the force constant, and $\ell_{\rm b,0}$ is the equilibrium bond length; $\cos\beta_i = \hat{\boldsymbol{r}}_{i,i+1} \cdot \hat{\boldsymbol{r}}_{i+1,i+2}$ is the angle between adjacent bonds with force constant k_β; $\cos\gamma_i = \hat{\boldsymbol{f}}_i \cdot \hat{\boldsymbol{f}}_{i+1}$ is the twist around the polymer backbone between adjacent sites, k_γ is the corresponding force constant and γ_0 is the equilibrium twist. The last term, $U_{\rm align}$ constrains $\hat{\boldsymbol{u}}_i$ of each site to be aligned with the bond vector, $\boldsymbol{r}_{ij}$, where $\cos\psi_i = \hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{i,i+1}$, and k_ψ is the corresponding force constant. As mentioned previously, our model discretization is chosen to represent three base-pairs by a single DNA site. In our implementation, we use a modified form of the twistable worm-like chain implementation [19], that can account for a non-zero equilibrium twist (i.e. $\omega_0 \neq 0$). The corresponding forces and torques for this modified potential is given in Appendix 6.7.3.

Nucleosome-DNA Bonded interactions

Bonded interactions within the nucleosome, U_{Nucl}, are responsible for constraining the location and orientation of both the DNA entering/exiting sites and the dyad site relative to the nucleosome. As a consequence, it is natural to divide U_{Nucl} into two terms

$$U_{\text{Nucl}} = U_{\text{NDNA}} + U_{\text{dyad}} \tag{6.15}$$

where U_{NDNA} represents the bonded interactions corresponding to the implicit nucleosomal DNA, and U_{dyad} are bonded interactions that control the position and orientation of the dyad.

In 1CPN many of the bonded potentials have the same functional form, and only differ in their choice of parameters. As a consequence, to simplify the notation we first provide the definition of harmonic bonded potentials, U_{b}^{h}, the harmonic and cosine angle/orientation potentials, U_{a}^{h} and $U_{\text{a}}^{\cos}$, as

$$U_{\text{b}}^{\text{h}}(\ell_{ij}\,;\,k,\ell_0) = \frac{k}{2}(\ell_{ij}-\ell_0)^2 \tag{6.16}$$

$$U_{\text{a}}^{\text{h}}(\boldsymbol{x},\boldsymbol{y};k,\theta_0) = \frac{k}{2}\,(\theta-\theta_0)^2 \tag{6.17}$$

$$U_{\text{a}}^{\cos}(\boldsymbol{x},\boldsymbol{y};k,\theta_0) = k\,(1-\cos\,(\theta-\theta_0)) \tag{6.18}$$

where $\ell_{ij} = |\boldsymbol{r}_{ij}|$, $\cos\theta = \hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}$ with $\hat{\boldsymbol{x}} = \boldsymbol{x}/|\boldsymbol{x}|$ and k, ℓ_0 and θ_0 are parameters corresponding to the force constant, equilibrium length and equilibrium angles, respectively. To further simplify our notation, we label all sites involved in U_{Nucl} interactions as i, j, p, q, r, s as shown in Figure 6.3b-d. Throughout the remainder of this section Equations 6.16-6.5.3, and the indexing convention from Figure 6.3b-d will be referenced extensively.

The nucleosomal DNA contribution can be subdivided into four terms,

$$U_{\text{NDNA}} = U_{\text{NDNA}}^{\text{geom}} + U_{\text{NDNA}}^{\text{orient}} + U_{\text{NDNA}}^{\text{rotation}} + U_{\text{NDNA}}^{\text{twist}}.. \tag{6.19}$$

The first term, $U_{\text{NDNA}}^{\text{geom}}$ ensures that the locations of the DNA leaving the nucleosome correspond to their location in their crystal structure. This is shown graphically by the orange and green bonds in Figure 6.3c and is defined as

$$\begin{aligned} U_{\text{NDNA}}^{\text{geom}} = U_{\text{b}}^{\text{h}}(\ell_{iq}; k_{\text{b}}^{\text{N}}, \ell_{0,c}) + U_{\text{b}}^{\text{h}}(\ell_{ir}; k_{\text{b}}^{\text{N}}, \ell_{0,c}) + U_{\text{b}}^{\text{h}}(\ell_{qr}; k_{\text{b}}^{\text{N}}, \ell_{0,e}) \\ + U_{\text{a}}^{\text{h}}(\boldsymbol{r}_{qi}, \hat{\boldsymbol{f}}_i; k_\zeta, \zeta) + U_{\text{a}}^{\text{h}}(\boldsymbol{r}_{ir}, \hat{\boldsymbol{f}}_i; k_\zeta, \zeta). \end{aligned} \tag{6.20}$$

The second term, $U_{\text{NDNA}}^{\text{orient}}$ constrains the angle at which these DNA strands leave the nucleosome, and is given by

$$\begin{aligned} U_{\text{NDNA}}^{\text{orient}} = U_a^h(\boldsymbol{r}_{ij}, \boldsymbol{r}_{jk}; k_\alpha, \alpha) + U_a^h(\boldsymbol{r}_{km}, \boldsymbol{r}_{mn}; k_\alpha, \alpha) \\ + U_a^h(\hat{\boldsymbol{u}}_j, \hat{\boldsymbol{u}}_k; k_\eta, \eta) + U_a^h(\hat{\boldsymbol{u}}_m, \hat{\boldsymbol{u}}_k; k_\mu, \mu). \end{aligned} \tag{6.21}$$

The third term, $U_{\text{NDNA}}^{\text{rotation}}$ incorporates the ability of nucleosomal DNA to rotate around the histone proteins,

$$U_{\text{NDNA}}^{\text{rotation}} = U_a^{cos}(\hat{\boldsymbol{f}}_q, \hat{\boldsymbol{f}}_i; k_\theta, \theta) + U_a^{cos}(\hat{\boldsymbol{f}}_r, \hat{\boldsymbol{f}}_i; k_\theta, \theta). \tag{6.22}$$

Lastly, the fourth term $U_{\text{NDNA}}^{\text{twist}}$, corresponds to the ability of DNA twist to be stored *inside* of the nucleosome. The storage of twist within the nucleosome is a common feature of nucleosomal crystal structures and is a hypothesized mechanism by which nucleosomes reposition along a long stretch of DNA. This term is defined as

$$U_{\text{NDNA}}^{\text{twist}} = U_a^h(\hat{\boldsymbol{f}}_q, \hat{\boldsymbol{f}}_r; k_{\text{t}}^{\text{N}}, 0). \tag{6.23}$$

The position of the dyad site is constrained by two contributions to the potential energy, such that

$$U_{\mathrm{dyad}} = U_{\mathrm{dyad}}^{\mathrm{geom}} + U_{\mathrm{dyad}}^{\mathrm{orient}}. \tag{6.24}$$

A graphical representation of the interactions that contribute to U_{dyad} are given in Figure 6.3d. The first term, $U_{\mathrm{dyad}}^{\mathrm{geom}}$, constrains the position of the dyad relative to the nucleosome and the entering/exiting DNA sites and is given by

$$U_{\mathrm{dyad}}^{\mathrm{geom}} = U_h(\ell_{ji}; \ell_{0,d}) + U_h(\ell_{jq}; \ell_{0,e}) + U_h(\ell_{jr}; \ell_{0,e}), \tag{6.25}$$

and the second term, $U_{\mathrm{dyad}}^{\mathrm{orient}}$, constrains the alignment of the dyad to be equivalent to the nucleosome,

$$\begin{aligned} U_{\mathrm{dyad}}^{\mathrm{orient}} = {} & U_a^h(\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j; k_\psi, 90^\circ) + U_a^h(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_j; k_\psi, 90^\circ) + U_a^h(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_i; k_\psi, 90^\circ) \\ & + U_a^h(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j; k_\psi, 0^\circ) + U_a^h(\boldsymbol{r}_{ij}, \hat{\boldsymbol{u}}_j; k_\psi, 0^\circ) + U_a^h(\boldsymbol{r}_{ij}, \hat{\boldsymbol{u}}_i; k_\psi, 0^\circ), \end{aligned} \tag{6.26}$$

where the first three terms constrain $\hat{\boldsymbol{f}}$ and the final three terms constrain $\hat{\boldsymbol{u}}$.

6.3.4 Nucleosome Unwrapping and Variable Nucleosome Repeat Length

As mentioned previously, a key design goal of the 1CPN model was to capture intra-nucleosome nucleosome dynamics and flexibility. One aspect of nucleosome flexibility that was to permit DNA in our model to partially unwrap from the histone core. In order to achieve this, the entire 147 of nucleosomal DNA was not assumed to be rigid, but instead a certain number of DNA base pairs, $n_{bp,unwrap}$, were permitted to disassociate from the histone core. We chose $n_{bp,unwrap} \approx 10$ base-pairs, permitted 1CPN to reproduce the free energies of nucleosome unwrapping obtained using the detailed 3SPN-AICG model (to be discussed in detail in Section 6.5.3).

An additional consequence of introducing $n_{bp,unwrap}$ into the 1CPN model, is that any

arbitrary nucleosome repeat length (NRL) could be achieved without changing our DNA discretization. The NRL corresponds to the distance in base-pairs that separate two adjacent nucleosomes and gives a measure of how densely nucleosomes are assembled along a stretch of DNA. The distributions of NRL within cells has been measured to vary continuously, and can range from 160 to 220 base-pairs, depending on the cell type of interest, expression levels and cellular state[138]. Given that 147 base-pairs are incorporated into the nucleosome itself, these NRLs correspond to $\approx 10 - 70$ base pairs of linker DNA separating neighboring nucleosomes. In previous models of chromatin, the NRLs employed was limited to those where the linking DNA was an even multiple of their DNA discretization. And therefore comparing two different NRLs, different variations of their model would be introduced.

To permit 1CPN to simulate any arbitrary NRL without changing DNA discretization, we allow $n_{bp,unwrap}$ to vary by ± 1 base-pair, such that the specified NRL is an even For example, assume that two nucleosomes were connected by a NRL of 187 base-pairs, corresponding to a linker length of 40. We then choose $n_{bp,unwrap} = 10$ so that the number of 1CPN base pairs connecting these two nucleosomes is an even multiple of our DNA discretization (e.g. $(40 + 2n_{bp,unwrap})/(3\text{bp per bead}) = 60/3 = 20$ beads). However if a NRL of 188 is specified, corresponding to a linker length of 41, then $n_{bp,unwrap} = 11$ (e.g. $(41 + 2n_{bp,unwrap})/(3\text{bp per bead}) = 63/3 = 21$ beads). In effect permitting $n_{bp,unwrap} = 10 \pm 1$ bp involves calculating three different sets of geometric parameters (e.g. $\ell_{a,0}, \ell_{c,0}, \ell_{d,0}$) from the 1KX5 crystal structure depending on the exact value of $n_{bp,unwrap}$. In the course of model development, we found this was the only change necessary and that variations in $n_{bp,unwrap}$ to vary by a single base pair $n_{bp,unwrap} = 10 \pm 1$ base-pair had negligible effects on model results.

6.3.5 Brownian Dynamics

Dynamics in our model are described by Brownian dynamics where the translational motion of a particle is governed by

$$m_i d\boldsymbol{v}_i(t) = -\boldsymbol{\zeta}_i \cdot \boldsymbol{v}_i(t)dt + \boldsymbol{B}_i \cdot d\boldsymbol{W}_i(t), \tag{6.27}$$

$$\tag{6.28}$$

where $\boldsymbol{v}_i(t) = d\boldsymbol{r}_i/dt$ is the translational velocity of the ith particle, $\boldsymbol{r}_i(t)$ is its position, m_i is its mass, $\boldsymbol{\zeta}_i$ is the translational friction tensor, $\boldsymbol{B}_i$ is a matrix which satisfies $[\boldsymbol{B}_i] \cdot [\boldsymbol{B}_i]^\top = 2k_\mathrm{B}T\boldsymbol{\zeta}_i$ and $d\boldsymbol{W}_i(t)$ is a Wiener increment, which has white noise statistics, $\langle d\boldsymbol{W}_i(t) d\boldsymbol{W}_j(t')\rangle_\mathrm{eq} = \boldsymbol{\delta}\delta_{i,j}\delta(t-t')$.

Similarly, the rotational motion of a particle in solution can be written as [72],

$$[\boldsymbol{R}_i(t)]^{-1} \cdot \boldsymbol{I}_i \cdot d\boldsymbol{\omega}_i(t) = -\boldsymbol{\zeta}_i^\mathrm{r} \cdot \boldsymbol{\omega}_i(t)dt + \boldsymbol{B}_i^\mathrm{r} \cdot d\boldsymbol{W}_i(t), \tag{6.29}$$

$$\frac{d\hat{\boldsymbol{f}}_i(t)}{dt} = \boldsymbol{\omega}_i(t) \times \hat{\boldsymbol{f}}_i(t), \quad \frac{d\hat{\boldsymbol{u}}_i(t)}{dt} = \boldsymbol{\omega}_i(t) \times \hat{\boldsymbol{u}}_i(t), \quad \frac{d\hat{\boldsymbol{v}}_i(t)}{dt} = \boldsymbol{\omega}_i(t) \times \hat{\boldsymbol{v}}_i(t), \tag{6.30}$$

where ω_i is the angular velocity, $\boldsymbol{\zeta}^r$ is the rotational friction tensor, $\boldsymbol{B}_i^\mathrm{r}$ is a matrix which satisfies $[\boldsymbol{B}_i^\mathrm{r}] \cdot [\boldsymbol{B}_i^\mathrm{r}]^\top = 2k_\mathrm{B}T\boldsymbol{\zeta}_i^\mathrm{r}$ and $\boldsymbol{R}_i(t) \equiv \left\{\hat{\boldsymbol{f}}_i(t), \hat{\boldsymbol{u}}_i(t), \hat{\boldsymbol{v}}_i(t)\right\}^\top$ is a rotation matrix in which each row is one of the vectors in the triad of orientation vectors of the particle; $\boldsymbol{I}_i$ is the moment of inertia tensor of the particle.

Here we use freely-draining Brownian dynamics and the friction tensor is assumed to be diagonal such that $\boldsymbol{\zeta}_i = \frac{m_i}{\lambda}\boldsymbol{\delta}$, and $\boldsymbol{\zeta}^r = \frac{m_i}{\lambda^r}\boldsymbol{\delta}$ where λ and λ^r are the so-called scalar translational and rotational damping parameters. λ and λ^r are in given in units of time and are chosen to be $\lambda = 5.1$ps and $\lambda^r = 0.3\lambda$. The choice of these values are given in Appendix 6.7.5.

6.3.6 Code Availability

The 1CPN model is implemented into LAMMPS and is freely available at our website.

6.4 Validation Methods

Validation of a coarse-grained model can generally be classified by either a "top-down" or a "bottom-up" strategy. In a top-down strategy, a set of experimental measurements are chosen as a "target", and coarse-grained parameters are varied until these target measurements are reproduced. In contrast, a bottom-up strategy begins with a more detailed molecular model, frequently an atomistic one, and then applies a heuristic to approximate this detailed model into a coarse-grained one.

When validating the 1CPN model, we adopt a hybrid approach that uses elements of both the top-down and bottom-up strategies. A top-down approach was used to parameterize dynamical quantities like the diffusion constant and sedimentation coefficients. A bottom-up approach was used to inform the energetic terms within the model, and to maximize correspondence between 1CPN and the detailed 3SPN-AICG model.

6.4.1 3SPN-AICG Nucleosome Model

Though the work presented here is primarily concerned with the new 1CPN model, many of the parameters in 1CPN were obtained by mapping 1CPN results to those obtained using the 3SPN-AICG model. For this reason, we provide an overview of the 3SPN-AICG model here, and refer and interested reader to the detailed descriptions found in prior publications [49, 92]. In the 3SPN-AICG model, DNA is represented by the 3SPN.2C model of DNA, the latest version of the 3SPN model[82, 154, 59] that has been extended to incorporates the effects of DNA sequence on both the shape, local flexibility, and global flexibility of the DNA molecule [47]. In the 3SPN.2C model, each base pair is represented by three sites, located at the phosphate, sugar, and the base. The histone proteins are represented by the Atomic Interaction-based Coarse Grained (AICG) model, a Go-based protein model with sequence-specific interactions parameterized from all-atom energies obtained from AMBER[98]. In AICG each amino acid is represented by a single site, located at the center of mass of

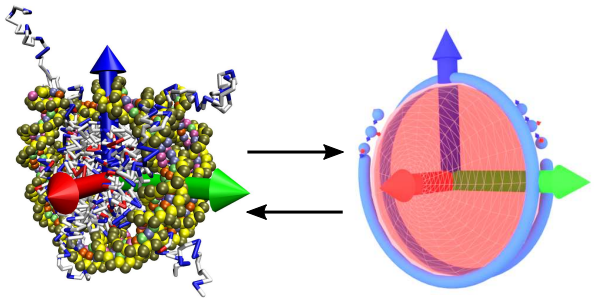

Figure 6.4: Mapping between detailed 3SPN-AICG model and coarse-grained 1CPN model.

the amino acid sidechain. We note that the AICG and 3SPN.2C models were developed independently to be accurate models of proteins and DNA, and consequently, no information about the nucleosome is encoded directly into the model.

Interactions between the 3SPN and AICG models are represented only through electrostatic interactions between the negatively charged phosphate DNA sites, and the charged amino acids on the histone protein. As a result, the configurations sampled in the 3SPN-AICG arise naturally as a balance between the elastic energies with the DNA molecule and the electrostatic interactions between the DNA and histones. This has permitted the 3SPN-AICG model to quantitatively match experimental measurements of the tension-dependent unwrapping of the nucleosome, without any additional fit parameters [92]. in 3SPN-AICG, electrostatic interactions are treated at the level of Debye-Hückel theory, and the solvent implicitly by a Langevin thermostat. All calculations using 3SPN-AICG were performed at 300K at 150mM monovalent salt unless otherwise noted.

6.4.2 Mapping between 3SPN-AICG and 1CPN Models

In order to match the nucleosome-nucleosome interactions energies in the 1CPN and 3SPN-AICG nucleosome models, it was necessary to develop a mapping function between the two models. Specifically, we desired a projection of a high-dimensional 3SPN-AICG configuration into lower-dimension 1CPN configuration. This reduction in dimension is most pronounced for a nucleosome where a $\mathcal{R}^{3N}$ where $N \approx 1800$ 3SPN-AICG configuration is mapped to

Orientation	A	B	C
$\hat{\boldsymbol{f}}_i \cdot \hat{\boldsymbol{f}}_j$	1	1	0
$\hat{\boldsymbol{f}}_j \cdot \hat{\boldsymbol{r}}_{ij}$	0	1	0
$\hat{\boldsymbol{f}}_i \cdot \hat{\boldsymbol{r}}_{ij}$	0	1	1
$\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_j$	1	1	0
$\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij}$	0	0	0
$\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij}$	0	0	0

Table 6.2: Definition of Nucleosome-Nucleosome Orientations for Pair-Potential Calculations

a 1CPN configuration in $\mathcal{R}^6$ defined by a single position, $\boldsymbol{r}$, and orientation $\{\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}\}$. In general, such a high reduction in dimension is prone to high loss of information, but we found that the rigid-structure of the histone core permitted a robust mapping from $\mathcal{R}^{3N}$ to $\mathcal{R}^6$.

To map from the 3SPN-AICG nucleosome to the 1CPN nucleosome site, the position, $\boldsymbol{r}_{\text{1CPN}}$, is simply the center of mass of the 3SPN-AICG nucleosome. The orientation vectors $\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}$ were obtained by defining geometric relationships within the nucleosome, such that $\hat{\boldsymbol{u}}$ points from the center of mass of the histone to the dyad, $\hat{\boldsymbol{f}}$ points along the nucleosomal DNA superhelix, and $\hat{\boldsymbol{u}} \times \hat{\boldsymbol{f}} = \hat{\boldsymbol{v}}$. Once the position and orientation of the nucleosome site is obtained, the dyad site is positioned $\ell_{d,0}$ along $\hat{\boldsymbol{u}}$. DNA sites in 1CPN are positioned along the helical axis of 3SPN DNA corresponding to the central base pair of a three base pair segment. The orientation of these DNA sites are chosen such that $\hat{\boldsymbol{f}}$ points toward the 3SPN minor groove, and $\hat{\boldsymbol{u}}$ points along the helical axis. Additional details of this mapping procedure are described in Appendix 6.7.7.

6.4.3 Nucleosome-Nucleosome Pair Potentials

In the 3SPN-AICG nucleosome model, the relative orientation of the two nucleosomes is given by six angles which represent the all combinations of angles between $\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j, \hat{\boldsymbol{r}}_{ij}$ and $\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}$. Thus to fully capture all possible orientations of two nucleosomes in the 3SPN-AICG model, an biaxial anisotropic potential would be necessary. However, the Zewdie potential is a uniaxial potential and only depends on a three angles $\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j, \hat{\boldsymbol{r}}_{ij}$. Thus when choosing the values of $\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}$ for the 3SPN-AICG pair potentials in order to parameterize 1CPN, we

sought ones that would best characterize the orientations that occur within a chromatin fiber. In chromatin, entering/exiting DNA from nucleosomes prevents nucleosome-nucleosomes interactions involving the dyad. Accordingly we avoid configurations where $\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij} \neq 0$ or $\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij} \neq 0$. The precise definitions of Orientations A,B,C in Figure 6.5 and listed in Table 6.4.3.

To compute the pair-potential, the 3SPN-AICG nucleosomes were constrained to these orientations using strong harmonic angle potentials, and then umbrella sampling[74] was performed along the center-of-mass separation, r, between the two nucleosomes.

6.4.4 Nucleosomal DNA Rotation

In order to calculate the free energies of DNA rotation around the nucleosome, we first define an order parameter, S_{R}, that quantifies the rotational position of the nucleosomal DNA. This order parameter has been used in previous work with the nucleosome [49, 93], and is described again here. The definition of S_{R} is given by

$$S_{\mathrm{R}} = \langle \pm \arccos\left(\frac{\boldsymbol{P}\cdot\boldsymbol{B}}{\|\boldsymbol{P}\|\|\boldsymbol{B}\|}\right)\rangle,$$

where $\mathbf{B}$ is a vector from the center of a given base step on the sense strand to its complementary base step on the anti-sense strand, $\mathbf{P}$ is a vector from the center of a base step to the center of the protein, and the angle brackets denote an average over base steps at the -15, -5, +5 and +15 positions relative to the dyad. The positive sign is chosen if $(\boldsymbol{P}\times\boldsymbol{B})\cdot\boldsymbol{D} \leq 0$ (negative if > 0), where $\boldsymbol{D}$ is a vector in the 5' to 3' direction along the sense strand. Notably, when $S_{\mathrm{R}} = -\pi/2$, the minor groove is oriented toward the protein core, and when $S_{\mathrm{R}} = \pi/2$, it is oriented away from it. The free energy along S_{R} was obtained using Metadynamics, an advanced sampling technique where a molecular dynamics trajectory is continuously biased toward unexplored regions of phase-space, thereby ensuring that all of phase space is sampled[89].

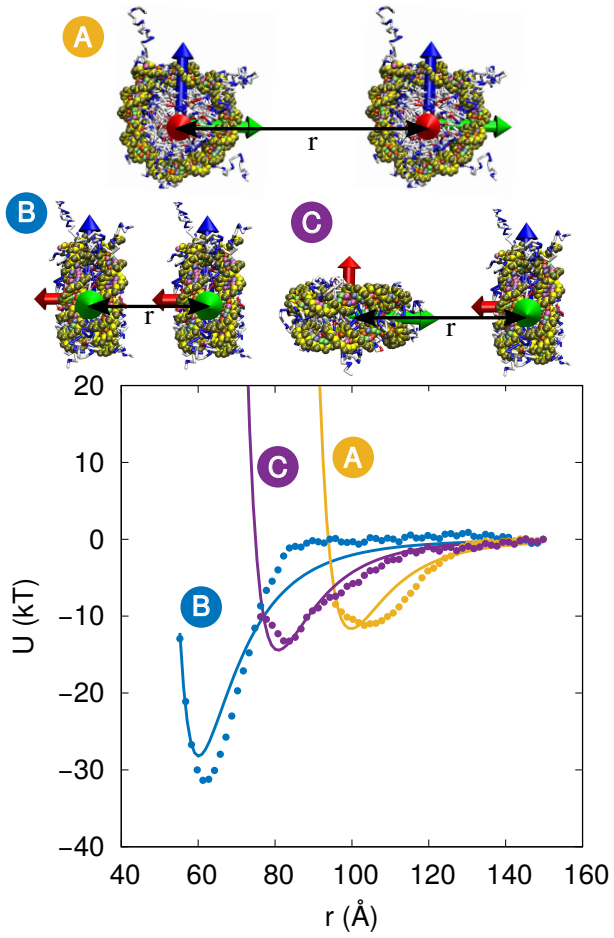

Figure 6.5: Pair-Potential between two nucleosomes for several orientations for 3SPN-AICG (points) and 1CPN models (lines).

6.5 Results

6.5.1 Nucleosome-Nucleosome Pair Potential

The pair-potential for the 3SPN-AICG nucleosome model was calculated for three different nucleosome orientations (dots, Figure 6.5). These orientations were achieved by calculating the orientation vectors $\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}$ of each nucleosome , and then constraining the relative angle between the nucleosomes. Once the orientation of the nucleosomes were constrained, the pair-wise free energy as a function of separation, r, was calculated using umbrella sampling (see Methods).

From these free energy surfaces the lowest energy configuration corresponds to when

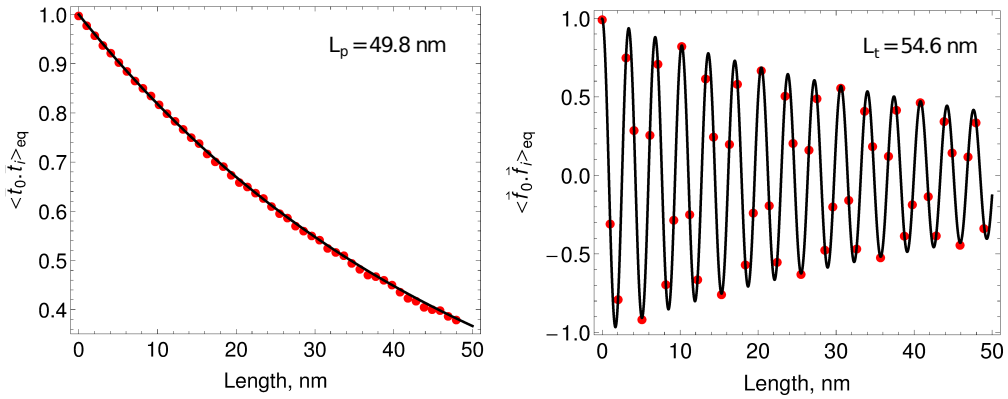

Figure 6.6: Bend and Twist persistence length of the 1CPN DNA model.

the two nucleosomes are when they are face-to-face (Orientation B), consistent with many models of the chromatin fiber. As nucleosomes adopt alternate orientations (e.g. Orientation A,C) the energy scales of interactions decrease, and the length scales increase. The length scales range from $\approx 50 - 95\mathring{A}$ consistent with the geometry of the nucleosome core particle, whereas the energy scales vary from $\approx 10 - 30 k_B T$ depending on the orientation of approach.

Having obtained these free-energy surfaces, we then parameterize nucleosome-nucleosome interactions energies in 1CPN in order to maximize agreement with the 3SPN-AICG results. Specifically, the parameters for the Zewdie potential, U_{Zewdie}, are chosen as to minimize the error between the 3SPN-AICG and 1CPN model distributions for the three different nucleosome orientations (see Methods). The resulting parameters are reported in Table 6.6 and the resulting pair potentials are shown by solid lines in Figure 6.5. The correspondence between the 3SPN-AICG (points) and the 1CPN (lines) pair potentials are fairly good, indicating that the complex orientation-dependent interactions between two nucleosomes are well approximated by the Zewdie potential.

6.5.2 DNA Persistence Length

In addition to the pair-wise interactions between nucleosomes, inter-nucleosome interactions within chromatin are also mediated by the elasticity of the DNA that connects neighboring nucleosomes. To ensure that these elastic contributions are represented faithfully in our

model, we calculate the bend and twist persistence lengths of the twistable worm-like chain used in 1CPN (Figure 6.6)..

The symbols are obtained by averaging results from a Brownian dynamics simulation of a chain with $N = 50$ beads with the torques and forces derived above and using the constants given by Wedemann [188]. The line on the left panel is a fit of the equation, $\langle \hat{\boldsymbol{t}}_0 \cdot \hat{\boldsymbol{t}}_j \rangle_{\text{eq}} = ke^{-\ell_0(j-1)/\ell_p}$, to the data using ℓ_p and k as fitting parameters and $j = 1, 2, ..., N$ where N is the number of beads. Form this a value of the persistence length, ℓ_p, of 49.8 nm is obtained. The line on the right panel is a fit of the equation, $\langle \hat{\boldsymbol{f}}_0 \cdot \hat{\boldsymbol{f}}_j \rangle_{\text{eq}} = k_1 e^{-\ell_0(j-1)/\ell_t} \cos[k_2 \ell_0 (j-1)]$, to the data using ℓ_t, k_1 and k_2 as fitting parameters. Form this a value of the twist persistence length, ℓ_t, of 54.6 nm is obtained. Both values of the bending persistence length, ℓ_p, and the twist persistence length, ℓ_t are in good agreement with experimental values[19].

6.5.3 DNA-Nucleosome Coupling

Having validated the predictions of 1CPN for nucleosome-nucleosome interactions and DNA elasticity, we now turn our attention to validating the interactions that couple Nucleosome-DNA interactions together. These interactions can be thought of as the "boundary conditions" between the DNA and nucleosome models described above, and represents an important component of our model. More generally, capturing this coupling accurately is a central to our goal of a chromatin model that captures intra-nucleosome dynamics.

Entering/Exiting Nucleosomal DNA

The first aspect of DNA-nucleosome coupling we choose to address is the angle at which DNA exits the nucleosome. In previous models of chromatin, this angle is fixed at a constant value, which is either extracted from the nucleosomal crystal structure[8, 7] or is chosen as an input parameter in the model [165, 77]. In both cases, the resulting configurations of the chromatin fiber is shown to be very sensitive to the choice of these angles, with different choices of angles leading to a wide range of different configurations. Additionally, an implicit assumption in

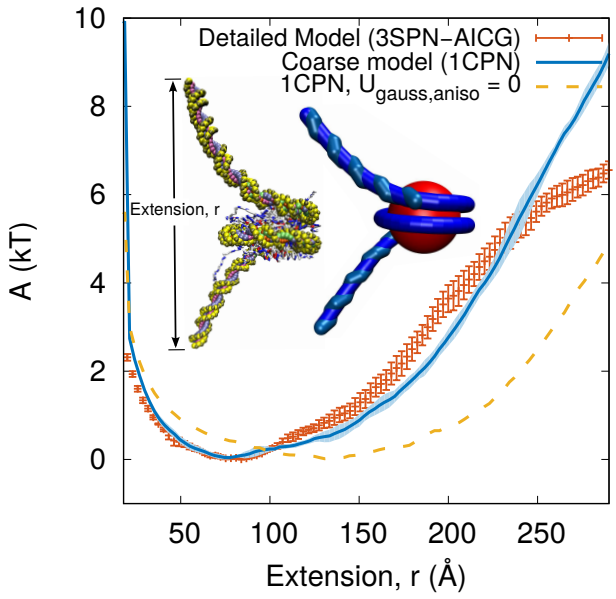

Figure 6.7: Free energy of partial nucleosome unwrapping for 3SPN-AICG model (points) and 1CPN model (lines).

choosing a fixed angle is that the DNA within the nucleosome is rigid, and that partial DNA unwrapping cannot occur.

Instead of assuming a fixed entry/exit DNA angle of nucleosomal DNA, in our approach here we parameterize the 1CPN model to reproduce the distribution of possible angles as observed using the detailed 3SPN-AICG model. To achieve this, we rely on recent results using the 3SPN-AICG model where the free energies of nucleosome unwrapping were obtained [92]. These free energies, and the forces associated with them were found to be in in excellent agreement with the experimental measurements of [115]. To compare the predictions of these different models, repeat the procedure employed by [92] using the 1CPN model. Specifically, this entailed generating a 1CPN configuration consisting of a single nucleosome containing 223 bp of DNA, and then using umbrella sampling to obtain the free energy as a function to the end-to-end extension, r.

The free energies predicted by 1CPN and those obtained using 3SPN-AICG are shown in Figure 6.7, and are found to be in excellent agreement over a large range of extensions. Both the DNA-DNA repulsion at low extension and the free energy penalty as the extension is

increased are well reproduced by the 1CPN model. The largest disagreement between the two models occurs at large separation $r > 225$, where DNA begins to unwrap from the histone surface in the 3SPN-AICG model in order to decrease the elastic penalty of bending the DNA (see snapshot). Though the 1CPN model permits the unwrapping of $n_{bp,unwrap} \approx 10$ bp of nucleosomal DNA (see Figure 6.7 1CPN snapshot), at extensions larger than $225\mathring{A}$ no additional DNA is permitted to unwrap, and the free energy of 1CPN continues to increase, whereas the free energy of 3SPN-AICG begins to plateau. Nonetheless, the good agreement between the coarse-grained 1CPN model with 3SPN-AICG indicates that 1CPN accurately reproduces the free energies of possible DNA configurations exiting the nucleosome.

This result also indicates the importance of the anisotropic gaussian potential, $U_{\text{aniso,gauss}}$, in governing the exiting configurations of DNA (as discussed in Section 6.3.2). Whereas inclusion of $U_{\text{aniso}-\text{gauss}}$ is sufficient to reproduce the 3SPN-AICG free energies (Figure 6.7, blue line), removal of this potential results in free energies that are significantly too low (Figure 6.7).

Sequence-Dependent Rotation of Nucleosomal DNA

An additional form of DNA-Nucleosome coupling occurs through the rotation of DNA around the histone proteins. This motion can be thought of as a coupling between the orientation of the DNA on of the nucleosome with the twist on the other. This twisting can manifest itself in two ways, either as twist *stored within* the nucleosome, or as twist *transferred through* the nucleosome. We will discuss each of these terms here.

The storage of DNA twist within the nucleosome is a well known phenomena that dates back to the observation in early nucleosome crystal structures that DNA within the nucleosome was over-twisted. Many crystal structures of the nucleosome are observed to have over-twisted DNA [105, 38, 175], and the incorporation of twist into the nucleosome has long been hypothesized as a mechanism by which the nucleosome could reposition[84]. This storage of twist within the 1CPN model is dictated by the value of k_{t}^{N}, which enters in the

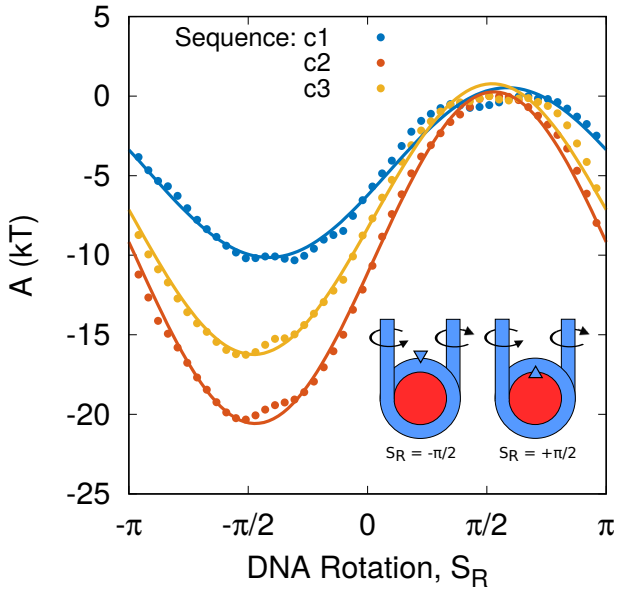

Figure 6.8: Free Energy of Nucleosomal DNA rotation for 3SPN-AICG model (points) and 1CPN model (lines).

expression for $U^{\text{twist}}_{\text{NDNA}}$ in Equation 6.23, and couples the orientation of the DNA entering and exiting the nucleosome. k^{N}_{t} dictates the strength of coupling between the entering and exiting DNA and is inversely related to how much twist can be stored within the nucleosome. Since most analyses suggest that a single twist defect can be stored in the nucleosome (corresponding to $\approx 36^{\circ}$ excess twist), we choose $k^{\text{N}}_{\text{t}} = 0.5$kcal/mol so that the energy cost of a single twist defect is $\approx 5k_BT$.

The transfer of twist through the nucleosome is derived from the observation that different DNA sequences exhibit positional phasing. It is well established that different DNA sequences bind to the histone proteins with different affinities, with different sequences exhibiting mechanical properties or equilibrium shapes that make nucleosome formation more or less favorable[171]. Certain DNA sequences bind stronger than others because these sequences contain periodic base-step motifs that energetically prefer to bind to the histone proteins. When all of these favorable base-steps occur on the side of the DNA facing the histone protein (with the 10 bp pitch of DNA), very strong binding can occur. A consequence of these periodic strong-binding base-steps is that the DNA molecule has a strong rotational

preference for the histone core. DNA Rotation disrupts the favorable DNA-histone contacts, and results in an energetic penalty.

In recent work using the 3SPN-AICG model, we have quantified the free energies corresponding to DNA rotation within the nucleosome for three different DNA sequences (points, Figure 6.8) [49]. The orientation of the DNA molecule is given by the order parameter, S_R, that describes which side of the DNA molecule is facing the histone core (see Methods). Notably, a value of $S_{\mathrm{R}} = -\pi/2$ corresponds to the minor groove facing the protein, whereas when $S_{\mathrm{R}} = \pi/2$ the minor groove is facing away from it. As expected, these DNA sequences have a strong rotational preferences. All three sequences exhibit free energy minima at $\approx -\pi/2$, where the favorable DNA binding motifs are facing the histone protein and maxima at $\approx \pi/2$ these these motifs are facing away from the histone protein. The magnitude of the free energy penalty of this rotation is strongly dependent on sequence, ranging from $\approx 20 k_B T$ for the c2 sequence, to $\approx 10 k_B T$ for the c1 sequence.

This sequence-dependent rotation is incorporated into the 1CPN model through the term $U_{\mathrm{NDNA}}^{\mathrm{rotation}}$, defined in Equation . $U_{\mathrm{NDNA}}^{\mathrm{rotation}}$ is a cosine angle potential (see Equation that contains two parameters θ_0, the equilibrium angle between the nucleosome and entering/exiting orientation vector $\hat{\boldsymbol{f}}$, and k_θ the spring constant constant corresponding to rotation. By fitting θ_0 and k_θ in Equation to the different sequences, we obtain the solid lines in Figure 6.8. For all sequences, the functional form of Equation can fit the 3SPN-AICG data very well. Therefore, sequence-dependent rotation of DNA around the histone proteins can be incorporated into the 1CPN model through the choice of θ_0 and k_θ.

6.6 Conclusion

In conclusion, we have shown that the 1CPN coarse-grained model of chromatin can accurately reproduce a wide array of biophysical processes within the nucleosome. The 1CPN model has been rigorously parameterized using free energies measured using the detailed 3SPN-AICG nucleosome model, and can achieve exceptional agreement with these free

energies at a fraction of the computational cost. 1CPN can capture the free energies of nucleosome-nucleosome attraction, nucleosome unwrapping, and sequence-dependent rotation of nucleosomal DNA. Additionally 1CPN captures the correct dynamics of chromatin, as well as the mechanical properties of the linking DNA. We expect the 1CPN model to be useful for studying the dynamic processes that dictate chromatin compaction and can shed light on the processes by which genome compaction and gene expression are linked. These are important and unsolved problems, and the 1CPN represents important progress towards addressing them.

6.7 Appendix

6.7.1 Zewdie Potential Forces and Torques

At usual, the forces are given by,

$$\boldsymbol{F}_{ij} = -\frac{\partial U_{ij}}{\partial \boldsymbol{r}_{ij}} \tag{6.31}$$

Though the derivative $\frac{\partial U_{ij}}{\partial \boldsymbol{r}_{ij}}$ can be computed by brute force, [3] provide a simpler solution by wisely applying the chain rule . Following this procedure gives for force, F_{ij} as

$$\boldsymbol{F}_{ij} = -\left(\frac{\partial U_{ij}}{\partial r_{ij}}\right)\frac{\partial r_{ij}}{\partial \boldsymbol{r}_{ij}} - \left(\frac{\partial U_{ij}}{\partial a_0}\right)\frac{\partial a_0}{\partial \boldsymbol{r}_{ij}} - \left(\frac{\partial U_{ij}}{\partial a_1}\right)\frac{\partial a_1}{\partial \boldsymbol{r}_{ij}} - \left(\frac{\partial U_{ij}}{\partial a_2}\right)\frac{\partial a_2}{\partial \boldsymbol{r}_{ij}}. \tag{6.32}$$

The advantage of this form is that the vector derivatives (e.g. $\partial r_{ij}/\partial \boldsymbol{r}_{ij}$) are completely separated from the scalalar derivatives (e.g. $\partial U_{ij}/\partial r_{ij}$). This greatly simplifies the calculation by allowing the vector derivatives to be neglected until the final step.

Equation 6.32 can be further simplified since a_0 is not a function of r_{ij} and so the second

term vanished. Additionally using $a_1 = \hat{\boldsymbol{f}}_i \cdot \boldsymbol{r}_{ij}/r_{ij}$, and applying the quotient rule gives,

$$\frac{\partial a_1}{\partial \boldsymbol{r}_{ij}} = \frac{1}{r_{ij}} \frac{\partial}{\partial \boldsymbol{r}_{ij}} \left(\hat{\boldsymbol{f}}_i \cdot \boldsymbol{r}_{ij} \right) + \hat{\boldsymbol{f}}_i \cdot \boldsymbol{r}_{ij} \frac{\partial}{\partial \boldsymbol{r}_{ij}} \left(\frac{1}{r_{ij}} \right) \tag{6.33}$$

$$= \hat{\boldsymbol{f}}_i / r_{ij} - a_1 \boldsymbol{r}_{ij} / r_{ij}^2, \tag{6.34}$$

with a similar result for $\partial a_2 / \partial \boldsymbol{r}_{ij}$. Taken together, the force is now

$$\boldsymbol{F}_{ij} = -\left(\frac{\partial U_{ij}}{\partial r_{ij}} \right) \frac{\boldsymbol{r}_{ij}}{r_{ij}} - \left(\frac{\partial U_{ij}}{\partial a_1} \right) \left(\frac{\hat{\boldsymbol{f}}_i}{r_{ij}} - \frac{a_1 \boldsymbol{r}_{ij}}{r_{ij}^2} \right) - \left(\frac{\partial U_{ij}}{\partial a_2} \right) \left(\frac{\hat{\boldsymbol{f}}_j}{r_{ij}} - \frac{a_2 \boldsymbol{r}_{ij}}{r_{ij}^2} \right). \tag{6.35}$$

A similar approach is used to compute the torques. The torques of a uniaxial potential are given by,

$$\boldsymbol{\tau}_{ij} = -\hat{\boldsymbol{f}}_{\boldsymbol{i}} \times \frac{\partial U_{ij}}{\partial \hat{\boldsymbol{f}}_i} \tag{6.36}$$

where

$$\frac{\partial U_{ij}}{\partial \hat{\boldsymbol{f}}_{\boldsymbol{i}}} = \left(\frac{\partial U_{ij}}{\partial r_{ij}} \right) \frac{\partial r_{ij}}{\partial \hat{\boldsymbol{f}}_i} + \left(\frac{\partial U_{ij}}{\partial a_0} \right) \frac{\partial a_0}{\partial \hat{\boldsymbol{f}}_i} + \left(\frac{\partial U_{ij}}{\partial a_1} \right) \frac{\partial a_1}{\partial \hat{\boldsymbol{f}}_i} + \left(\frac{\partial U_{ij}}{\partial a_2} \right) \frac{\partial a_2}{\partial \hat{\boldsymbol{f}}_i} \tag{6.37}$$

based on the chain rule. This equation can again be simplified since $\frac{\partial r_{ij}}{\partial \hat{\boldsymbol{f}}_i} = \frac{\partial a_2}{\partial \hat{\boldsymbol{f}}_i} = 0$ and the first and the fourth terms disappear. Further, since $\frac{\partial a_0}{\partial \hat{\boldsymbol{f}}_i} = \hat{\boldsymbol{f}}_{\boldsymbol{j}}$ and $\frac{\partial a_1}{\partial \hat{\boldsymbol{f}}_i} = \frac{\boldsymbol{r}_{ij}}{r_{ij}}$, Equation 6.36 reduces to

$$\boldsymbol{\tau}_{ij} = -\hat{\boldsymbol{f}}_i \times \left[\left(\frac{\partial U_{ij}}{\partial a_0} \right) \hat{\boldsymbol{f}}_j + \left(\frac{\partial U_{ij}}{\partial a_1} \right) \frac{\boldsymbol{r}_{ij}}{r_{ij}} \right] \tag{6.38}$$

It should be noted that $\tau_{ij} \neq -\tau_{ji}$ and that τ_{ji} should be calculated by reversing the labels of i and j and then using Equation 6.38 accordingly.

The next step is then to calculate the derivatives of the potential, U_{ij} with respect to r_{ij} , a_0, a_1, and a_2 and then substitude these values into Equations 6.35 and 6.38 to obtain the forces and torques, respectively.

The derivative with respect to r_{ij} is the simplest and is given by

$$\frac{\partial U_{ij}}{\partial r_{ij}} = 4\epsilon \left[\frac{-12\sigma_0^{12}}{(r_{ij} - \sigma + \sigma_0)^{13}} + \frac{6\sigma_0^6}{(r_{ij} - \sigma + \sigma_0)^7}) \right]. \tag{6.39}$$

The derivative with respect to a_0, a_1 and a_2 is more complicated, and again we benefit from a judiciously applied chain-rule. The derivitive with respect to a_0 is then

$$\begin{aligned} \frac{\partial U}{\partial a_0} = 4\epsilon & \left[\left(\frac{\sigma_0}{r_{ij} - \sigma + \sigma_0} \right)^{12} - \left(\frac{\sigma_0}{r_{ij} - \sigma + \sigma_0} \right)^6 \right] \times \\ & \left[\left(\frac{12}{r_{ij} - \sigma + \sigma_0} - \frac{6}{r_{ij} - \sigma + \sigma_0} \right) \frac{\partial \sigma}{\partial a_0} + \frac{1}{\epsilon} \frac{\partial \epsilon}{\partial a_0} \right] \end{aligned} \tag{6.40}$$

where $\partial\sigma/\partial a_0$ is given by

$$\frac{\partial \sigma}{\partial a_0} = \sigma_0 \left[\sigma_{000} \frac{\partial S_{000}}{\partial a_0} + \sigma_{cc2} \left(\frac{\partial S_{022}}{\partial a_0} + \frac{\partial S_{202}}{\partial a_0} \right) + \sigma_{220} \frac{\partial S_{220}}{\partial a_0} + \sigma_{222} \frac{\partial S_{222}}{\partial a_0} + \sigma_{224} \frac{\partial S_{224}}{\partial a_0} \right] \tag{6.41}$$

and $\partial\epsilon/\partial\sigma$ is given by

$$\frac{\partial \epsilon}{\partial a_0} = \epsilon_0 \left[\epsilon_{000} \frac{\partial S_{000}}{\partial a_0} + \epsilon_{cc2} \left(\frac{\partial S_{022}}{\partial a_0} + \epsilon_{cc2} \frac{\partial S_{202}}{\partial a_0} \right) + \epsilon_{220} \frac{\partial S_{220}}{\partial a_0} + \epsilon_{222} \frac{\partial S_{222}}{\partial a_0} + \epsilon_{224} \frac{\partial S_{224}}{\partial a_0} \right], \tag{6.42}$$

where the partial derivatives of the S-functions are given in Table 6.3. A similar approach is used to obtain $\partial U/\partial a_1$ and $\partial U/\partial a_2$. Finally Equations 6.39, 6.40, as well as the corresponding equations for $\partial U/\partial a_1$ and $\partial U/\partial a_2$ can be substituted into equation 6.35 and 6.38 to determine the forces and torques.

	$\partial S/\partial a_0$	$\partial S/\partial a_1$	$\partial S/\partial a_2$
S_{000}	0	0	0
S_{202}	0	$3a_1/\sqrt{5}$	0
S_{022}	0	0	$3a_2/\sqrt{5}$
S_{220}	$3a_0/\sqrt{5}$	0	0
S_{222}	$\frac{(-6a_0+9a_1a_2)}{\sqrt{70}}$	$\frac{(-6a_1+9a_0a_2)}{\sqrt{70}}$	$\frac{(9a_0a_1-6a_2)}{\sqrt{70}}$
S_{224}	$\frac{(4a_0-20a_1a_2)}{4\sqrt{70}}$	$\frac{(-10a_1-20a_0a_2+70a_1a_2^2)}{4\sqrt{70}}$	$\frac{(-20a_0a_1-10a_2+70a_1a_1^2)}{4\sqrt{70}}$

Table 6.3: Partial Derivatives of S functions with respect to a_0, a_1 and a_2

6.7.2 Angle Orient Potential

In order to constrain the f vectors of the entering and exiting DNA with the nucleosome, we needed to introduce an additional potential. This potential has the form,

$$U_{ij} = \frac{1}{2}k\left[(\theta_1 - \theta_{1,0})^2 + (\theta_2 - \theta_{2,0})^2 + (\phi - \phi_0)^2\right] \tag{6.43}$$

where $\theta_1 = \text{acos}(a_1)$, $\theta_2 = \text{acos}(a_2)$, and $\phi = \text{acos}(a_0)$ and a_0, a_1, and a_2 were defined in Table 6.1.

To calculate the forces, we can begin from Equation 6.35 from our derrivation of the Zewdie potential. Now, the $\partial U_{ij}/\partial r_{ij}$ term is zero, and the forces reduce to

$$\boldsymbol{F}_{ij} = -\left(\frac{\partial U_{ij}}{\partial a_1}\right)\left(\frac{\hat{\boldsymbol{f}}_i}{r_{ij}} - \frac{a_1\boldsymbol{r}_{ij}}{r_{ij}^2}\right) - \left(\frac{\partial U_{ij}}{\partial a_2}\right)\left(\frac{\hat{\boldsymbol{f}}_j}{r_{ij}} - \frac{a_2\boldsymbol{r}_{ij}}{r_{ij}^2}\right). \tag{6.44}$$

Our derivation of the torques proceeds similarly, and is identical to Equation 6.38. With Equation 6.35 and 6.38, the remaining step is to calculate the partial derivatives with respect to a_0, a_1, and a_2 and are equal to

$$\frac{\partial U_{ij}}{\partial a_0} = -k(\phi - \phi_0)^2/\sqrt{1-a_0^2} \tag{6.45}$$
$$\frac{\partial U_{ij}}{\partial a_1} = -k(\theta_1 - \theta_{1,0})^2/\sqrt{1-a_1^2} \tag{6.46}$$
$$\frac{\partial U_{ij}}{\partial a_2} = -k(\theta_2 - \theta_{2,0})^2/\sqrt{1-a_2^2}. \tag{6.47}$$

The forces and torques are found by substituting Equation 6.47 into Equations 6.35 and 6.38, respectively.

6.7.3 tWLC Forces and Torques

In this Section the twist and alignment torques and forces for the twistable wormlike chain implementation proposed by Brackley et al. [19] are derived. The only difference from the torques and forces derived by Brackley et al.[19] is that a non-zero twist angle, w_0, is considered here.

In the twistable wormlike chain implementation proposed by Brackley et al. [19] the free energy penalties for twist and alignment are assumed to be decoupled and are simply added to the bend and stretch free energies,

$$U = U_{\text{twist}} + U_{\text{align}} + U_{\text{stretch}} + U_{\text{bend}} \tag{6.48}$$

Both the alignment and twist energies are taken to be harmonic in the alignment, ψ_i, and twist w angles, respectively,

$$U_{\text{align}} = \frac{\kappa_a}{\ell_0}\sum_{i=1}^{N+1}\psi_i^2, \quad U_{\text{twist}} = \frac{\kappa_t}{\ell_0}\sum_{i=1}^{N+1}(w_i - w_0)^2, \tag{6.49}$$
$$U_{\text{stretch}} = \frac{\kappa_s}{2\ell_0^3}\sum_{i=1}^{N}(\ell_i - \ell_0)^2, \quad U_{\text{bend}} = \frac{\kappa_b}{\ell_0}\sum_{i=1}^{N+1}\beta_i^2.$$

Parameter	Value
κ_s	1.10×10^{-18} J nm
κ_b	2.06×10^{-19} J nm
κ_t	2.67×10^{-19} J nm

Table 6.4: Twistable wormlike chain parameters for DNA given by Widemann [188] and used thought this document.

Where N is the number of sites in the linker strand, $\boldsymbol{t}_i = \boldsymbol{r}_{i+1} - \boldsymbol{r}_i$, $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, $\ell_i = |\boldsymbol{t}_i|$, $r_{ij} = |\boldsymbol{r}_{ij}|$. $\alpha_i, \beta_i, \gamma_i \sim$ Euler angles, rotation from triad $\{\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{f}}_i, \hat{\boldsymbol{v}}_i\}$ to to triad $\{\hat{\boldsymbol{u}}_{i+1}, \hat{\boldsymbol{f}}_{i+1}, \hat{\boldsymbol{v}}_{i+1}\}$. The twist angle is defined as $\omega_i = \alpha_i + \gamma_i$, while the alignment angle is given by, $\cos\psi_i = \hat{\boldsymbol{u}}_i \cdot \boldsymbol{t}_i/\ell_0$.

The parameters $\kappa_s, \kappa_p, \kappa_t, \kappa_a$ are stretch, bend, twist and alignment moduli, respectively; and w_0 is the equilibrium twist angle per bead. Here we use a coarse-graining resolution of three base pairs per bead, such that $\ell_0 \approx 0.99$ nm. On the other hand, for DNA the equilibrium twist angle per base pair is 36°, therefore the equilibrium twist angle per bead used here is 108°.

By means of Taylor series expansions it can be shown that up to second order in ψ_i, β_i and ω_i the angle dependent potentials in eq.(6.49) can be conveniently written in terms of cosines,

$$U_{\text{align}} = \frac{\kappa_a}{\ell_0} \sum_{i=1}^{N+1} [1 - \cos\psi_i], \quad U_{\text{twist}} = \frac{\kappa_t}{\ell_0} \sum_{i=1}^{N+1} [1 - \cos(\omega_i - \omega_0)] \tag{6.50}$$

$$U_{\text{bend}} = \frac{\kappa_b}{\ell_0} \sum_{i=1}^{N+1} [1 - \cos\beta_i].$$

To obtain the torques and forces from eq.(6.50) consider that in a time δt, the position and orientation of the ith bead change by $\delta\boldsymbol{r}_i$ and $\delta\phi_i$, respectively, where the latter is a a rotation about and axis $\hat{\boldsymbol{p}}_i$. Assuming that $\{\delta\boldsymbol{r}_i, \delta\phi_i\}$ are independent variables, then via the principle of virtual work the force and torque on each bead due to the free energy U is

given by,

$$\delta U = -\sum_i \boldsymbol{F}_i \cdot \delta \boldsymbol{r}_i - \sum_i \boldsymbol{\tau}_i \cdot \hat{\boldsymbol{p}}_i \delta\phi_i. \tag{6.51}$$

Where $\boldsymbol{F}_i$ and $\boldsymbol{\tau}_i$, are the force and torque, respectively.

To proceed is useful to introduce the following three relations,

$$\cos(w_i - w_0) = \cos w_i \cos w_0 + \sin w_i \sin w_i, \tag{6.52a}$$

$$\cos w_i = \frac{\hat{\boldsymbol{f}}_{i+1} \cdot \hat{\boldsymbol{f}}_i + \hat{\boldsymbol{v}}_{i+1} \cdot \hat{\boldsymbol{v}}_i}{(1 + \hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_{i+1})}, \qquad \sin w_i = \frac{\hat{\boldsymbol{v}}_{i+1} \cdot \hat{\boldsymbol{f}}_i - \hat{\boldsymbol{f}}_{i+1} \cdot \hat{\boldsymbol{v}}_i}{(1 + \hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_{i+1})}. \tag{6.52b}$$

Using the relations in eq.(6.52) one can write,

$$\begin{aligned} &[1 + \hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_{i+1}] \cos(w_i - w_0) = \\ &\left[\hat{\boldsymbol{f}}_{i+1} \cdot \hat{\boldsymbol{f}}_i + \hat{\boldsymbol{v}}_{i+1} \cdot \hat{\boldsymbol{v}}_i\right] \cos w_0 + \left[\hat{\boldsymbol{v}}_{i+1} \cdot \hat{\boldsymbol{f}}_i - \hat{\boldsymbol{f}}_{i+1} \cdot \hat{\boldsymbol{v}}_i\right] \sin w_0. \end{aligned} \tag{6.53}$$

Other useful relations between the local triad vectors and the bend, β_i, the angle, θ_i, between consecutive tangent vectors, $\boldsymbol{t}_i$ and $\boldsymbol{t}_{i+1}$, and the alignment angle ψ_i are,

$$\begin{aligned} \hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_{i+1} = \cos\beta_i, \qquad & \cos\theta_i = \frac{\boldsymbol{t}_i \cdot \boldsymbol{t}_{i+1}}{b_i b_{i+1}}, \\ b_i \cos\psi_i \quad = \quad & \hat{\boldsymbol{u}}_i \cdot \boldsymbol{t}_i. \end{aligned} \tag{6.54}$$

Now on can proceed to take the derivative of eq.(6.48) using eq.(6.50) for the alignment and twist energies,

$$\delta U = \frac{\kappa_t}{\ell_0} \sum_i \sin(w_i + w_0)\delta(w_i) + \frac{\kappa_a}{\ell_0} \sum_i \sin\psi_i \delta\psi_i. \tag{6.55}$$

To rewrite eq.(6.55) in terms of the local triad vector one needs to make use of eq.(6.53).

Taking the derivative of eq.(6.53) one obtains,

$$\begin{aligned}
&\cos(w_i - w_0)\left[\delta\hat{\boldsymbol{u}}_{i+1}\cdot\hat{\boldsymbol{u}}_i + \hat{\boldsymbol{u}}_{i+1}\cdot\delta\hat{\boldsymbol{u}}_i\right] - \\
&\sin(w_i - w_0)\left[\hat{\boldsymbol{u}}_i\cdot\hat{\boldsymbol{u}}_{i+1}\right]\delta(w_i) = \\
&\left[\delta\hat{\boldsymbol{f}}_{i+1}\cdot\hat{\boldsymbol{f}}_i + \hat{\boldsymbol{f}}_{i+1}\cdot\delta\hat{\boldsymbol{f}}_i + \delta\hat{\boldsymbol{v}}_{i+1}\cdot\hat{\boldsymbol{v}}_i + \hat{\boldsymbol{v}}_{i+1}\cdot\delta\hat{\boldsymbol{v}}_i\right]\cos w_0 + \\
&\left[\delta\hat{\boldsymbol{f}}_{i+1}\cdot\hat{\boldsymbol{v}}_i + \hat{\boldsymbol{f}}_{i+1}\cdot\delta\hat{\boldsymbol{v}}_i - \delta\hat{\boldsymbol{v}}_{i+1}\cdot\hat{\boldsymbol{f}}_i + \hat{\boldsymbol{v}}_{i+1}\cdot\delta\hat{\boldsymbol{f}}_i\right]\sin w_0.
\end{aligned} \tag{6.56}$$

The infinitesimal change in the axis vector $\hat{\boldsymbol{u}}_i$ due to a rotation of $\delta\phi_i$ about a vector $\hat{\boldsymbol{p}}_i$ is given by $\delta\hat{\boldsymbol{u}}_i = \hat{\boldsymbol{p}}_i \times \hat{\boldsymbol{u}}_i\delta\phi_i$ and similar for $\hat{\boldsymbol{f}}_i$ and $\hat{\boldsymbol{v}}_i$,

$$\begin{aligned}
-\left[\delta\hat{\boldsymbol{u}}_{i+1}\cdot\hat{\boldsymbol{u}}_i + \hat{\boldsymbol{u}}_{i+1}\cdot\delta\hat{\boldsymbol{u}}_i\right] &= \sin\beta_i\delta(\beta_i) \\
&= -\left[\hat{\boldsymbol{u}}_{i+1}\times\hat{\boldsymbol{u}}_i\right]\cdot\left(\hat{\boldsymbol{p}}_{i+1}\delta\phi_{i+1} - \hat{\boldsymbol{p}}_i\delta\phi_i\right).
\end{aligned} \tag{6.57}$$

With this, one can show that the first term on the right hand side of eq.(6.55) can also be written as,

$$\begin{aligned}
&\left[1 + \hat{\boldsymbol{u}}_{i+1}\cdot\hat{\boldsymbol{u}}_i\right]\sin(w_i + w_0)\delta(w_i) = \\
&\Bigg[\left(\hat{\boldsymbol{u}}_{i+1}\times\hat{\boldsymbol{u}}_i\right)\cos(w_i - w_0) - \left(\hat{\boldsymbol{f}}_{i+1}\times\hat{\boldsymbol{f}}_i + \hat{\boldsymbol{v}}_{i+1}\times\hat{\boldsymbol{v}}_i\right)\cos w_0 - \\
&\left(\hat{\boldsymbol{f}}_{i+1}\times\hat{\boldsymbol{v}}_i - \hat{\boldsymbol{v}}_{i+1}\times\hat{\boldsymbol{f}}_i\right)\sin w_0\Bigg]\cdot\left(\hat{\boldsymbol{p}}_{i+1}\delta\phi_{i+1} - \hat{\boldsymbol{p}}_i\delta\phi_i\right).
\end{aligned} \tag{6.58}$$

On the other hand, the second term on the right hand side of eq.(6.55) can be written as,

$$\sin\psi_i\delta\psi_i = \frac{1}{b_i}\left(\cos\psi_i\hat{\boldsymbol{t}}_i - \hat{\boldsymbol{u}}_i\right)\cdot\left(\delta\boldsymbol{r}_{i+1} - \delta\boldsymbol{r}_i\right) - \frac{1}{b_i}\left(\hat{\boldsymbol{u}}_i\times\hat{\boldsymbol{t}}_i\right)\cdot\hat{\boldsymbol{p}}_i\delta\phi_i. \tag{6.59}$$

Inserting eqs. 6.58 and 6.59 into eq. 6.55, and matching terms in eq. 6.51 gives expressions

for the force and torque,

$$\begin{aligned} \boldsymbol{F}_i &= \frac{\kappa_a}{\ell_0}\left(\boldsymbol{G}_i - \boldsymbol{G}_{i-1}\right), \\ \boldsymbol{\tau}_i &= \frac{\kappa_a}{\ell_0}\hat{\boldsymbol{u}}_i \times \hat{\boldsymbol{t}}_i + \frac{\kappa_t}{\ell_0}\left(\boldsymbol{H}_i - \boldsymbol{H}_{i-1}\right). \end{aligned} \tag{6.60}$$

Where

$$\begin{aligned} \boldsymbol{H}_i = \Bigg[& (\hat{\boldsymbol{u}}_{i+1} \times \hat{\boldsymbol{u}}_i)\cos(w_i - w_0) - \left(\hat{\boldsymbol{f}}_{i+1} \times \hat{\boldsymbol{f}}_i + \hat{\boldsymbol{v}}_{i+1} \times \hat{\boldsymbol{v}}_i\right)\cos w_0 \\ & - \left(\hat{\boldsymbol{f}}_{i+1} \times \hat{\boldsymbol{v}}_i - \hat{\boldsymbol{v}}_{i+1} \times \hat{\boldsymbol{f}}_i\right)\sin w_0 \Bigg] \Big/ \left(1 + \hat{\boldsymbol{u}}_{i+1} \cdot \hat{\boldsymbol{u}}_i\right), \\ \boldsymbol{G}_i = & \frac{1}{b_i}\left(\cos\psi_i \hat{\boldsymbol{t}}_i - \hat{\boldsymbol{u}}_i\right). \end{aligned} \tag{6.61}$$

6.7.4 Langevin Dynamics

6.7.5 Calculating Damping Parameter

The purpose of this subsection is to estimate the damping parameter to be used with the Langevin thermostat in LAMMPS to perform MD simulations of the 1CPN model in the canonical ensemble. The Langevin thermostat in LAMMPS only allows for the specification of a scalar friction coefficient. More precisely, in LAMMPS the friction coefficient of a site is specified by means of a scalar damping parameter $\lambda \equiv m/\zeta$. Where ζ is the friction coefficient of a site and m is the mass of that site. However in the level of description of 1CPN a nucleosome is represented as an ellipsoidal particle. Ellipsoidal particles have tensorial friction coefficients [56, 79]. Therefore to find a good scalar approximation of the scalar damping parameter to be used in LAMMPS we first study the diffusion of a single nucleosome in a dilute solution accounting for the tensorial nature of the friction coefficient.

The Langevin equation for the translational motion of a microscopic particle in solution

can be written as,

$$m_i d\boldsymbol{v}_i(t) = -\boldsymbol{\zeta}_i \cdot \boldsymbol{v}_i(t)dt + \boldsymbol{B}_i \cdot d\boldsymbol{W}_i(t),$$
$$\frac{d\boldsymbol{r}_i(t)}{dt} = \boldsymbol{v}_i(t).$$

Where $\boldsymbol{v}_i(t)$ is the translational velocity of the the particle, $\boldsymbol{r}_i(t)$ is its position and m is its mass, $\boldsymbol{\zeta}_i$ is the translational friction tensor, $\boldsymbol{B}_i$ is a matrix which satisfies $[\boldsymbol{B}_i] \cdot [\boldsymbol{B}_i]^\top = 2k_{\mathrm{B}}T\boldsymbol{\zeta}_i$ and $d\boldsymbol{W}_i(t)$ is a Wiener increment, which has white noise statistics, $\langle d\boldsymbol{W}_i(t) d\boldsymbol{W}_j(t')\rangle_{\mathrm{eq}} = \boldsymbol{\delta}\delta_{i,j}\delta(t-t')$. The mass of a nucleosome core particle is $M_N = 196.666$ Kg/mol.

The inertial time scale of the nucleosome motion using $M_N = 196.666$ Kg/mol is,

$$\lambda_1 = \left(\frac{\frac{1}{4}k_{\mathrm{B}}T}{ma^2}\right)^{1/2} = 3.1 \times 10^{-9}\text{seconds}. \tag{6.62}$$

where $m_N \equiv M_N/N_A$ and N_A is Avogadro's constant. Where a is the equatorial (largest) radius of the nucleosome.

For an ellipsoidal particle the translational friction tensor, $\boldsymbol{\zeta}$, can be written as,

$$\boldsymbol{\zeta} = 16\pi\mu a e \left[X^A \hat{\boldsymbol{f}}\hat{\boldsymbol{f}} + Y^A(\boldsymbol{\delta} - \hat{\boldsymbol{f}}\hat{\boldsymbol{f}})\right] \tag{6.63}$$

where $\mu \sim 8.90 \times 10^{-4}$ Kg/(s m) is the viscosity of the solvent, mostly water in this case. The eccentricity of the oblate ellipsoid is $e \equiv \sqrt{1 - \frac{c^2}{a^2}}$ and for oblate ellipsoids the equatorial radius a is always greater than the polar radius c. The constants X^A and Y^A are given by [79],

$$X^A \equiv \frac{8}{3}e^3 \left[2(2e^2-1)\cot^{-1}\left(\frac{\sqrt{1-e^2}}{e}\right) + 2e\sqrt{1-e^2}\right]^{-1}, \tag{6.64a}$$

$$Y^A \equiv \frac{8}{3}e^3 \left[2(2e^2+1)\cot^{-1}\left(\frac{\sqrt{1-e^2}}{e}\right) - e\sqrt{1-e^2}\right]^{-1}. \tag{6.64b}$$

For a spherical particle, which is the case of DNA beads, the translational friction tensor reduces to $\boldsymbol{\zeta} = 6\pi R\mu\boldsymbol{\delta}$, where R is the radius of the spherical bead.

The Langevin equation for the rotational motion of a microscopic particle in solution can be written as [72],

$$
\begin{aligned}
&[\boldsymbol{R}_i(t)]^{-1} \cdot \boldsymbol{I}_i \cdot d\boldsymbol{\omega}_i(t) = -\boldsymbol{\zeta}_i^{\mathrm{r}} \cdot \boldsymbol{\omega}_i(t)dt + \boldsymbol{B}_i^{\mathrm{r}} \cdot d\boldsymbol{W}_i(t), \\
&\frac{d\hat{\boldsymbol{f}}_i(t)}{dt} = \boldsymbol{\omega}_i(t) \times \hat{\boldsymbol{f}}_i(t), \quad \frac{d\hat{\boldsymbol{u}}_i(t)}{dt} = \boldsymbol{\omega}_i(t) \times \hat{\boldsymbol{u}}_i(t), \quad \frac{d\hat{\boldsymbol{v}}_i(t)}{dt} = \boldsymbol{\omega}_i(t) \times \hat{\boldsymbol{v}}_i(t).
\end{aligned}
$$

Where ω_i is the angular velocity, $\boldsymbol{\zeta}^r$ is the rotational friction tensor, $\boldsymbol{B}_i^{\mathrm{r}}$ is a matrix which satisfies $[\boldsymbol{B}_i^{\mathrm{r}}] \cdot [\boldsymbol{B}_i^{\mathrm{r}}]^\top = 2k_{\mathrm{B}}T\boldsymbol{\zeta}_i^{\mathrm{r}}$ and $\boldsymbol{R}_i(t) \equiv \left\{\hat{\boldsymbol{f}}_i(t), \hat{\boldsymbol{u}}_i(t), \hat{\boldsymbol{v}}_i(t)\right\}^\top$ is a rotation matrix in which each row is one of the vectors in the triad of orientation vectors of the particle, $\boldsymbol{I}_i$ is the moment of inertia tensor of the ellipsoidal particle. In the coordinate system defined by $\left\{\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{u}}_i, \hat{\boldsymbol{v}}_i\right\}$ the moment of inertia tensor can be wrritten as,

$$
\boldsymbol{I}_i = \begin{pmatrix} \frac{2}{5}m_i a^2 & 0 & 0 \\ 0 & \frac{1}{5}m_i\left(a^2+c^2\right) & 0 \\ 0 & 0 & \frac{1}{5}m_i\left(a^2+c^2\right) \end{pmatrix}. \tag{6.65}
$$

The rotational friction tensor for an ellipsoidal particle can be written as [79],

$$
\boldsymbol{\zeta}^{\mathrm{r}} = \frac{32}{3}\pi\mu a^3 e^3 \left[X^C \hat{\boldsymbol{f}}\hat{\boldsymbol{f}} + Y^C(\boldsymbol{\delta} - \hat{\boldsymbol{f}}\hat{\boldsymbol{f}})\right].
$$

For an oblate ellipsoid the constants X^C and Y^C are given by [79],

$$
X^C = \frac{2}{3}e^3 \left[\cot^{-1}\left(\frac{\sqrt{1-e^2}}{e}\right) - e\sqrt{1-e^2}\right]^{-1}, \tag{6.66a}
$$

$$
Y^C = \frac{2}{3}e^3(2-e^2) \left[e\sqrt{1-2e^2} - (1-2e^2)\cot^{-1}\left(\frac{\sqrt{1-e^2}}{e}\right)\right]^{-1}. \tag{6.66b}
$$

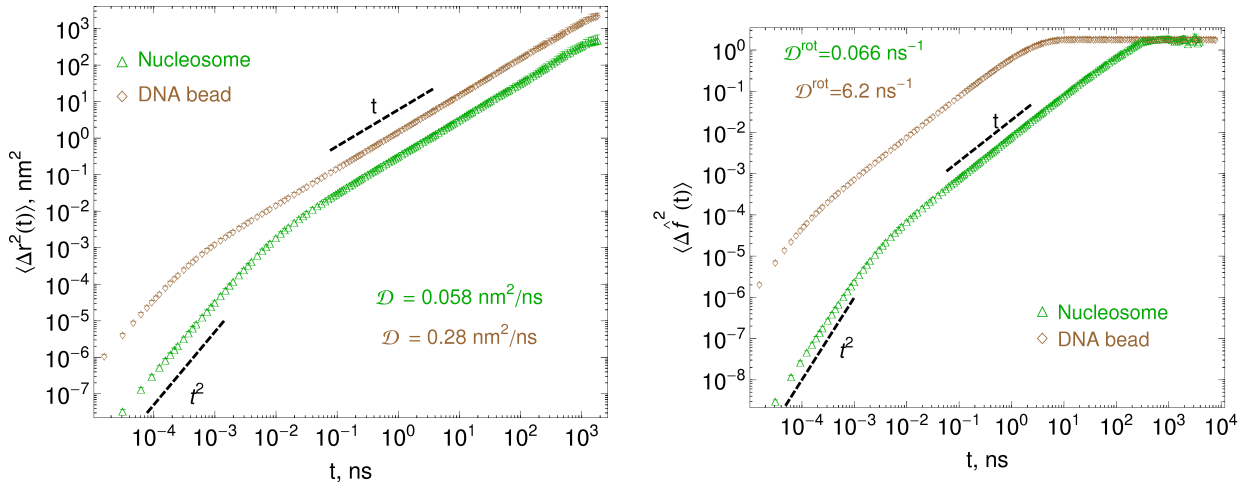

Figure 6.9: Rotational and translational dynamics of nucleosome and DNA sites. $\langle \Delta \boldsymbol{r}^2(t) \rangle \equiv \langle [\boldsymbol{r}(t) - \boldsymbol{r}(0)]^2 \rangle_{\text{eq}}$, $\langle \Delta \hat{\boldsymbol{f}}^2(t) \rangle \equiv \langle [\hat{\boldsymbol{f}}(t) - \hat{\boldsymbol{f}}(0)]^2 \rangle_{\text{eq}}$. A scaling of t^2 indiates ballistic motion, a scaling of t^1 indicates diffusive motion.

For a spherical particle, which is the case of DNA beads, the translational friction tensor reduces to $\boldsymbol{\zeta}^{\text{r}} = 8\pi R^3 \mu \boldsymbol{\delta}$, where R is the radius of the spherical bead.

Fluorescence fluctuation spectroscopy (FFS) mesurements of the difussion coefficient of nucleosome particles in dilute solution [69] have yield results in the range of $\mathcal{D}_{\text{nucleosome}} \sim 0.015 - 0.018$ nm^2/ns. These are transaltional diffusion coefficients, rotational difusion coefficients of nucleosomes do not appear to have been reported in the literature. The hydrodynamic calculations presented above are in good qualitative agreement with the available experimental measurements.

Parameter	Value
m_{DNA}	3.24×10^{-24} Kg
m_N	3.26×10^{-22} Kg
ζ_{DNA}	1.7×10^{-11} N s/m
ζ_N^{eff}	7.15×10^{-11} N s/m
m_{DNA}/ζ_{DNA}	0.195×10^{-12} s
m_N/ζ_N^{eff}	4.563×10^{-12} s

Table 6.5: Dynamic parameters for the 1CPN model.

To calculate the damping parameter for the Langevin thermostat to be used in LAMMPS when simulating 1CPN we calculate an effective friction coefficient of the nucleosome using

the Einstein relation, $\zeta^{\text{eff}} \equiv \dfrac{k_{\text{B}}T}{\mathcal{D}} \sim 7.9{\times}10^{-11}$ N s /m. With this effective friction a damping parameter is calculated $\lambda \equiv m_N/\zeta^{\text{eff}} \sim 4.6$ ps. This damping parameter is larger than the one that is obtained if one uses the friction coefficient and the mass of a 1CPN DNA bead (see Table 6.5). The damping parameter determines how rapidly the temperature is relaxed in an MD simulation in the canonical ensemble. The damping parameter of the nucleosome which produces a slower relaxation of the temperature is choosen as the damping coefficient to be used when running 1CPN in LAMMPS. Additionally, in LAMMPS the rotational friction is also specified as scalar. Therefore we use the scalar relation between translational and rotational friction that can be derived for a sphere, e.g. $\lambda^{\text{r}} \sim 0.3\lambda_m$.

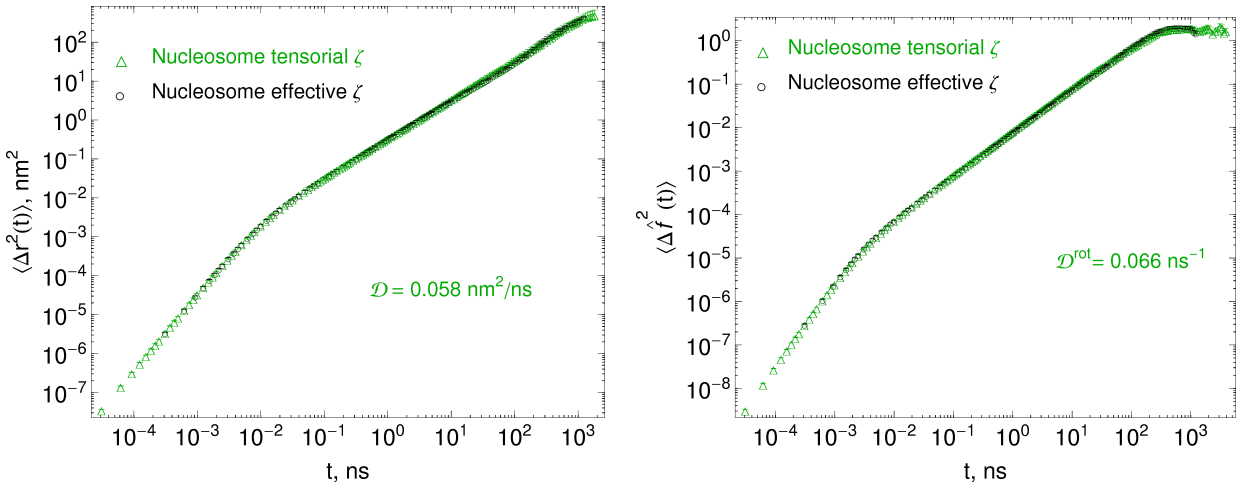

Figure 6.10: Rotational and translational dynamics of the 1CPN nucleosome sites.

6.7.6 Model Parameters

The parameters used in the model are described in the following tables.

6.7.7 Definition of Orientation Vectors in 3SPN-AICG

The three vectors, $\hat{\boldsymbol{f}}$, $\hat{\boldsymbol{v}}$, and $\hat{\boldsymbol{u}}$, are defined using several structural features on the 3SPN-AICG representation of the histone octamer. The vector $\hat{\boldsymbol{u}}$ is directed from the center of mass of the protein through the dyad on the nucleosome. The location of the dyad was

U_{zewdie}				U_{elect}			
σ_0	48.5	σ_0'	30.0	ϵ	Eq. 6.10	λ_D	Eq. 6.9
ϵ_0	3.2	ϵ_0'	0.1	$U_{\text{gauss,aniso}}$			
σ_{000}	1.7	ϵ_{000}	2.5	d_0	2.0	r_0	45
σ_{cc2}	-0.60	ϵ_{cc2}	2.3	σ	2.0		
σ_{220}	-0.02	ϵ_{220}	1.5	K_θ	4	θ_0	70°
σ_{222}	0.0	ϵ_{222}	0.0	K_ϕ	8	ϕ_0	90°
σ_{224}	0.0	ϵ_{224}	0.0				
Source: Fig. 6.5				Source: Fig. 6.7			

Table 6.6: Nonbonded parameters

parameter	explanation	value
k_{b}^{D}	kbond dna	163.3 kcal/mol/$\mathring{A}^2$
k_β	kbend dna	30 kcal/mol
k_ω	ktwist dna	38.9 kcal/mol
k_ψ	kalign dna	388.9 kcal/mol
$\ell_{\text{b},0}$	eq length dna	9.9 $\mathring{A}$
ω_0	eq twist dna	46°

Table 6.7: DNA Bonded parameters

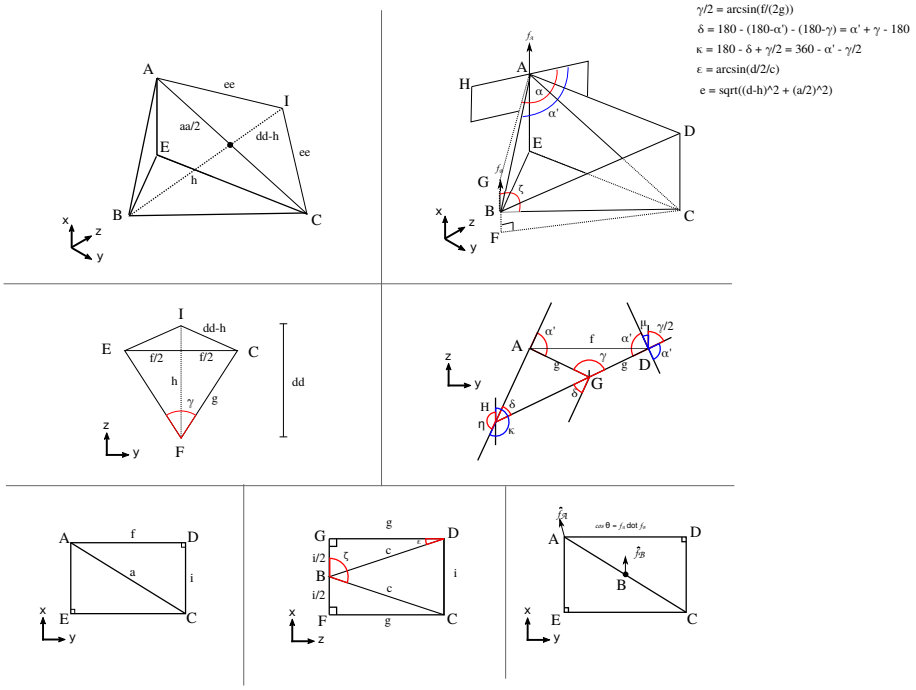

Figure 6.11: Diagram showing the definition of different lengths used within 1CPN Model.

Equil. Lengths and Angles

parameter	value	source
$\ell_{a,0}$	(9) 90.11, (10) 91.87, (11) 91.85	Ref. [29]
$\ell_{c,0}$	(9) 48.17 , (10) 48.02 , (11) 47.18	Ref. [29]
$\ell_{d,0}$	input parameter (dyad nucl length)	Ref. [29]
$\ell_{e,0}$	$\ell_{e,0} = \sqrt{(\ell_d - \ell_h)^2 + (\ell_a/2)^2}$	Geometry
α	90°	Fig. 6.7
ζ	$\zeta = \frac{\pi}{2} + \arcsin\left(\frac{i/2}{c}\right)$	Geometry
θ	$\theta = (36^\circ n_{bp,unwrap}) \mod 360^\circ$	Geometry
η	$\eta = 2\pi - \kappa$	Geometry
μ	$\mu = \pi - \alpha' - \gamma/2$	Geometry
κ	$\kappa = (2\pi - \alpha' - \gamma) + (\gamma/2)$	Geometry
γ	$\gamma = 2\arcsin(0.5\ell_f/\ell_g)$	Geometry
$\ell_{f,0}$	$\ell_f = \sqrt{\ell_{a,0}^2 - \ell_{i,0}^2}$	Geometry
$\ell_{g,0}$	$\ell_g = \sqrt{\ell_{c,0}^2 - (\ell_{i,0}/2)^2}$	Geometry
$\ell_{i,0}$	(9) 36.93, (10) 35.36, (11) 32.37	Ref. [29]
$\ell_{h,0}$	$\ell_h = \sqrt{\ell_{g,0}^2 - (\ell_{f,0}/2)^2}$	Geometry

Force Constants

parameter	value	source
k_α	10 kcal/mol	Fig. 6.7
k_ζ	100 kcal/mol	Geometry
k_θ	1.6 kcal/mol	Fig. 6.8
$k_{\mathrm{t}}^{\mathrm{N}}$	10 kcal/mol	Sec. 6.5.3
k_η	100 kcal/mol	
k_μ	—	Same as k_η

Table 6.8: Bonded parameters entering U_{Nucl}. When three values are given, these correspond to the value when $n_{bp,unwrap} = 9, 10, 11$ bp, respectively.

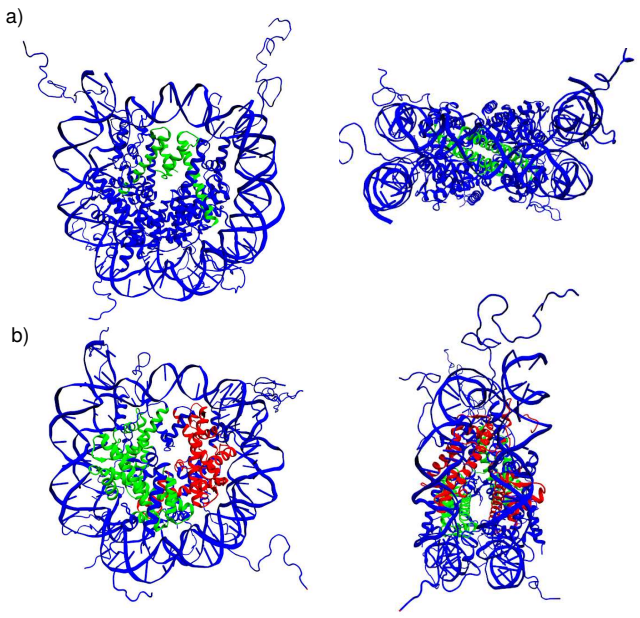

Figure 6.12: Protein sites used to define fvu vectors in 3SPN-AICG

defined as the point between the center of mass of the histone core and the center of mass of the sites consisting of the $\alpha 1$ through $\alpha 3$ sites of the H3 and H3' histones.

The vector $\hat{\boldsymbol{f}}$ is the vector normal to the flat face of the nucleosome and roughly points along the nucleosomal DNA superhelix. In order to calculate $\hat{\boldsymbol{f}}$, we first defined two clusters of sites on the histone proteins (Figure 6.12 red and green), and define a vector $\hat{\boldsymbol{w}}$ that points between these two clusters. $\hat{\boldsymbol{f}}$ is then defined as $\hat{\boldsymbol{f}} \equiv \hat{\boldsymbol{u}} \times \hat{\boldsymbol{w}}$. Once vector $\hat{\boldsymbol{f}}$ was determined, the last vector was determined by taking the cross product between $\hat{u}$ and $\hat{f}$.

CHAPTER 7

CONCLUSIONS

In conclusion, it has been demonstrated in this dissertation that the DNA molecule is a versatile tool for achieving self-assembly across a wide range of length scales. As the smallest length scales, we have shown that the hybridization of two DNA molecules is strongly dependent on DNA sequence, and that this sequence-dependence can be exploited for materials applications. Specifically, we have shown that that DNA sequence can be used to engineer the energy scales and mechanical properties of DNA-conjugated nanoparticles, thereby suggesting the possibility of mechanical metamaterials constructed using DNA. At larger length scales, we then turned to the biophysical processes of chromatin compaction, with the goal of understanding the mechanisms employed by nature to achieve complex and dynamic assembly. We initially focused on the smallest unit of chromatin, a DNA-protein complex called the nucleosome, and examined its energetics and dynamics using a detailed molecular model. Even at these larger length scales, we show that DNA sequence continues to play a significant role, and that the energies and dynamics of nucleosome formation are strongly dictated by the DNA sequence incorporated into the nucleosome. DNA sequence manifests itself in nucleosome repositioning through different mechanisms, suggesting that even the higher-order folds of chromatin may be dictated, to some degree, by the DNA sequence itself. Finally, we have presented a multi-scale approach that seeks to understand even larger structures within chromatin, and the properties that emerge from chromatin when many nucleosomes interact together.

As the complexity of engineered systems continues to advance, the driving research question is no longer what materials *can* be made, but rather what material *should* be made in order to achieve materials with a desired performance and properties. This so-called "inverse-design" problem represents a significant challenge for the research community, and is a major barrier for progress in engineered systems. Though the complexity of these materials is critical to achieve their function, this complexity also frustrates attempts to design

materials from the bottom up, because the parameter space of all possible designs is prohibitively large. Nature has discovered innovative, and frequently unexpected solutions, to take advantage of this complexity, and can offer new approaches for materials design. There is therefore a large opportunity to draw from both the biophysics and engineering literature, in order to design synthetic systems that seek to mimic the complexity achieved by nature.

REFERENCES

[1] SSAGES: An open-source, engine agnostic, C++11 based advanced sampling package (https://github.com/MICCoM/SSAGES-public). 2017.

[2] A. Paul Alivisatos, Kai P Johnsson, Xiaogang Peng, Troy E Wilson, Colin J Loweth, Marcel P Bruchez Jr, and Peter G Schultz. Organization of 'nanocrystal molecules' using DNA. *Nature*, 382:609–611, 1996.

[3] M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*. 1987.

[4] Stefano Angioletti-Uberti, Bortolo M. Mognetti, and Daan Frenkel. Re-entrant melting as a design principle for DNA-coated colloids. *Nature Materials*, 11(6):1–5, April 2012.

[5] T K Archer, M G Cordingley, R G Wolford, and G L Hager. Transcription factor access is mediated by accurately positioned nucleosomes on the mouse mammary tumor virus promoter. *Mol. Cell Biol.*, 11(2):688–698, 1991.

[6] G. Arya and T. Schlick. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Nat. Acad. Sci. USA.*, 103(44):16236–16241, 2006.

[7] Gaurav Arya and Tamar Schlick. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J. Phys. Chem. A*, 113(16):4045–4059, 2009.

[8] Gaurav Arya, Qing Zhang, and Tamar Schlick. Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophysical journal*, 91(1):133–150, 2006.

[9] Debra J. Audus, Francis W. Starr, and Jack F. Douglas. Coupling of isotropic and directional interactions and its effect on phase separation and self-assembly. *J. Chem. Phys.*, 144(7):074901, 2016.

[10] Evelyn Auyeung, Ting I N G Li, Andrew J Senesi, Abrin L Schmucker, Bridget C Pals, Monica Olvera de la Cruz, and Chad A Mirkin. DNA-mediated nanoparticle crystallization into Wulff polyhedra. *Nature*, 505(7481):73–7, January 2014.

[11] Artem Barski, Suresh Cuddapah, Kairong Cui, Tae Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129(4):823–837, 2007.

[12] Gavin D. Bascom, Karissa Y. Sanbonmatsu, and Tamar Schlick. Mesoscale Modeling Reveals Hierarchical Looping of Chromatin Fibers Near Gene Regulatory Elements. *J. Phys. Chem. B*, 120(33):8642–8653, 2016.

[13] Brian J. Beliveau, Alistair N. Boettiger, Maier S. Avendaño, Ralf Jungmann, Ruth B. McCole, Eric F. Joyce, Caroline Kim-Kiselak, Frédéric Bantignies, Chamith Y. Fonseka, Jelena Erceg, Mohammed A. Hannan, Hien G. Hoang, David Colognori, Jeannie T. Lee, William M. Shih, Peng Yin, Xiaowei Zhuang, and Chao-ting Wu. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.*, 6(May):7147, 2015.

[14] Sukesh R Bhaumik, Edwin Smith, and Ali Shilatifard. Covalent modifications of histones during development and disease pathogenesis. *Nat. Struct. Mol. Biol.*, 14(11):1008–1016, 2007.

[15] Paul Biancaniello, Anthony Kim, and John Crocker. Colloidal Interactions and Self-Assembly Using DNA Hybridization. *Phys. Rev. Lett.*, 94(5):058302, February 2005.

[16] Timothy R Blosser, Janet G Yang, Michael D Stone, Geeta J Narlikar, and Xiaowei Zhuang. Dynamics of nucleosome remodelling by individual ACF complexes. *Nature*, 462(7276):1022–1027, 2009.

[17] Alistair N Boettiger, Bogdan Bintu, Jeffrey R Moffitt, Siyuan Wang, Brian J Beliveau, Geoffrey Fudenberg, Maxim Imakaev, Leonid A Mirny, Chao-ting Wu, and Xiaowei Zhuang. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586):418–422, 2016.

[18] Gregory D. Bowman. Mechanisms of ATP-dependent nucleosome sliding. *Curr. Opin. Struc. Biol.*, 20(1):73–81, 2010.

[19] C A Brackley, A N Morozov, and D Marenduzzo. Models for twistable elastic polymers in Brownian dynamics, and their implementation for LAMMPS. *J. Chem. Phys.*, 140(13):135103, apr 2014.

[20] Brent Brower-Toland, David A. Wacker, Robert M. Fulbright, John T. Lis, W. Lee Kraus, and Michelle D. Wang. Specific Contributions of Histone Tails and their Acetylation to the Mechanical Stability of Nucleosomes. *J. Mol. Biol.*, 346(1):135–146, 2005.

[21] Brent D Brower-Toland, Corey L Smith, Richard C Yeh, John T Lis, Craig L Peterson, and Michelle D Wang. {M}echanical disruption of individual nucleosomes reveals a reversible multistage release of {DNA}. *Proc. Nat. Acad. Sci. USA.*, 99(4):1960–1965, 2002.

[22] Stefano Cacchione, Maria Antonietta Cerone, and Maria Savino. In vitro low propensity to form nucleosomes of four telomeric sequences. *FEBS Lett.*, 400(1):37–41, 1997.

[23] Le Chang and Shoji Takada. Histone acetylation dependent energy landscapes in tri-nucleosome revealed by residue-resolved molecular simulations. *Scientific Reports*, 6(March):34441, 2016.

[24] Wenlong Cheng, Michael J Campolongo, Judy J Cha, Shawn J Tan, Christopher C Umbach, David A Muller, and Dan Luo. Free-standing nanoparticle superlattice sheets controlled by DNA. *Nat. Mater.*, 8(6):519–525, 2009.

[25] Giuseppe Chirico and Jörg Langowski. Kinetics of DNA supercoiling studied by Brownian dynamics simulation. *Biopolymers*, 34:415–433, 1994.

[26] Rosana Collepardo-Guevara and Tamar Schlick. The effect of linker histones nucleosome binding affinity on chromatin unfolding mechanisms. *Biophys. J.*, 101(7):1670–1680, 2011.

[27] Yanier Crespo, Fabrizio Marinelli, Fabio Pietrucci, and Alessandro Laio. Metadynamics convergence law in a multidimensional system. *Phys. Rev. E*, 81(5):055701, May 2010.

[28] Y Cui and C Bustamante. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proc. Nat. Acad. Sci. USA.*, 97(1):127–132, 2000.

[29] Curt a Davey, David F Sargent, Karolin Luger, Armin W Maeder, and Timothy J Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *Journal of molecular biology*, 319(5):1097–113, June 2002.

[30] Aram Davtyan, Nicholas P Schafer, Weihua Zheng, Cecilia Clementi, Peter G Wolynes, and Garegin a Papoian. AWSEM-MD: Protein Structure Prediction Using Coarse-Gained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B*, 116:8494–8503, 2012.

[31] Mekonnen Lemma Dechassa, Swetansu K. Hota, Payel Sen, Nilanjana Chatterjee, Punit Prasad, and Blaine Bartholomew. Disparity in the DNA translocase domains of SWI/SNF and ISW2. *Nucleic Acids Res.*, 40(10):4412–4421, 2012.

[32] Sebastian Deindl, William L Hwang, Swetansu K Hota, Timothy R Blosser, Punit Prasad, Blaine Bartholomew, and Xiaowei Zhuang. ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell*, 152(3):442–52, January 2013.

[33] Job Dekker, Marc A Marti-Renom, and Leonid A Mirny. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.*, 14(6):390–403, jun 2013.

[34] Jesse R. Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.

[35] Biqin Dong, Luay M. Almassalha, Yolanda Stypula-Cyrus, Ben E. Urban, John E. Chandler, The-Quyen Nguyen, Cheng Sun, Hao F. Zhang, and Vadim Backman. Superresolution intrinsic fluorescence imaging of chromatin utilizing native, unmodified nucleic acids for contrast. *Proc. Nat. Acad. Sci. USA.*, 113:201602202, 2016.

[36] Benedetta Dorigo, Thomas Schalch, Alexandra Kulangara, Sylwia Duda, Rasmus R Schroeder, and Timothy J Richmond. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, 306(5701):1571–1573, 2004.

[37] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.*, 126(16):164103, apr 2007.

[38] Rajeswari S. Edayathumangalam, Philipp Weyermann, Peter B. Dervan, Joel M. Gottesfeld, and Karolin Luger. Nucleosomes in solution exist as a mixture of twist-defect states. *J. Mol. Biol.*, 345(1):103–114, 2005.

[39] R. Elghanian, James J. Storhoff, Robert C. Mucic, Robert L. Letsinger, and Chad A Mirkin. Selective Colorimetric Detection of Polynucleotides Based on the Distance-Dependent Optical Properties of Gold Nanoparticles. *Science*, 277(5329):1078–1081, August 1997.

[40] Jason Ernst, Pouya Kheradpour, Tarjei S Mikkelsen, Noam Shoresh, Lucas D Ward, Charles B Epstein, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael Coyne, Manching Ku, Timothy Durham, Manolis Kellis, and Bradley E Bernstein. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–9, 2011.

[41] Behrouz Eslami-Mossallam, Helmut Schiessel, and John van Noort. Nucleosome dynamics: Sequence matters. *Adv. Colloid Interface Sci.*, 232:101–113, 2015.

[42] Arman Fathizadeh, Azim Berdy Besya, Mohammad Reza Ejtehadi, and Helmut Schiessel. Rigid-body molecular dynamics of DNA inside a nucleosome. *The European physical journal. E, Soft matter*, 36(3):21, March 2013.

[43] Thomas G. Fazzio and Toshio Tsukiyama. Chromatin remodeling in vivo: Evidence for a nucleosome sliding mechanism. *Mol. Cell*, 12(5):1333–1340, 2003.

[44] Ilaria Filesi, Stefano Cacchione, Pasquale De Santis, Luigi Rossetti, and Maria Savino. The main role of the sequence-dependent DNA elasticity in determining the free energy of nucleosome formation on telomeric DNAs. *Biophys. Chem.*, 83(3):223–237, 2000.

[45] J T Finch and A Klug. Solenoidal model for superstructure in chromatin. *Proc. Nat. Acad. Sci. USA.*, 73(6):1897–1901, 1976.

[46] Andrew Flaus and T J Richmond. Positioning and stability of nucleosomes on MMTV 3'LTR sequences. *J. Mol. Biol.*, 275(3):427–441, 1998.

[47] Gordon S. Freeman, Daniel M. Hinckley, Joshua P. Lequieu, Jonathan K. Whitmer, and Juan J. de Pablo. Coarse-grained modeling of DNA curvature. *J. Chem. Phys.*, 141(16):165103, oct 2014.

[48] Gordon S. Freeman, Joshua P. Lequieu, Daniel M. Hinckley, Jonathan K. Whitmer, and Juan J. de Pablo. DNA Shape Dominates Sequence Affinity in Nucleosome Formation. *Physical Review Letters*, 113(16):168101, October 2014.

[49] Gordon S. Freeman, Joshua P. Lequieu, Daniel M. Hinckley, Jonathan K. Whitmer, and Juan J. de Pablo. DNA Shape Dominates Sequence Affinity in Nucleosome Formation. *Phys. Rev. Lett.*, 113(16):168101, oct 2014.

[50] Eden Fussner, Reagan W. Ching, and David P. Bazett-Jones. Living without 30nm chromatin fibers. *Trends Biochem. Sci.*, 36(1):1–6, 2011.

[51] Gregory J Gemmen, Ronald Sim, Karl a Haushalter, Pu Chun Ke, James T Kadonaga, and Douglas E Smith. Forced unraveling of nucleosomes assembled on heterogeneous DNA using core histones, NAP-1, and ACF. *J. Mol. Biol.*, 351(1):89–99, 2005.

[52] Joel M. Gottesfeld, Jason M. Belitsky, Christian Melander, Peter B. Dervan, and Karolin Luger. Blocking Transcription Through a Nucleosome with Synthetic DNA Ligands. *J. Mol. Biol.*, 321(2):249–263, aug 2002.

[53] Sergei A. Grigoryev, Gaurav Arya, Sarah Correll, Christopher L. Woodcock, and Tamar Schlick. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc. Nat. Acad. Sci. USA.*, 106(32):13317–13322, 2009.

[54] Sergei A Grigoryev, Gavin Bascom, Jenna M Buckwalter, Michael B Schubert, Christopher L Woodcock, and Tamar Schlick. Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. *Proc. Nat. Acad. Sci. USA.*, 113(5):1238–43, 2016.

[55] Michael A Hall, Alla Shundrovsky, Lu Bai, Robert M Fulbright, John T Lis, and Michelle D Wang. High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat. Struct. Mol. Biol.*, 16(2):124–129, feb 2009.

[56] John Happel and Howard Brenner. *Low Reynolds number hydrodynamics: with special applications to particulate media*, volume 1. Springer Science & Business Media, 2012.

[57] B Hendrich and W Bickmore. Human diseases with underlying defects in chromatin structure and modification. *Hum. Mol. Genet.*, 10(20):2233–2242, 2001.

[58] Daniel M. Hinckley and Juan J. De Pablo. Coarse-grained ions for nucleic acid modeling. *J. Chem. Theory Comput.*, 11(11):5436–5446, 2015.

[59] Daniel M Hinckley, Gordon S Freeman, Jonathan K Whitmer, and Juan J de Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.*, 139(14):144903, October 2013.

[60] Daniel M Hinckley, Joshua P Lequieu, and Juan J de Pablo. Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.*, 141(3):035102, July 2014.

[61] Denes Hnisz, Abraham S Weintraub, Daniel S Day, Anne-laure Valton, Rasmus O Bak, Charles H Li, Johanna Goldmann, Bryan R Lajoie, Zi Peng Fan, Alla A Sigova, Jessica Reddy, Diego Borges-rivera, Tong Ihn Lee, Rudolf Jaenisch, Matthew H Porteus, Job Dekker, and Richard A Young. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351(6280):1454–1458, 2016.

[62] Tsung Han S Hsieh, Assaf Weiner, Bryan Lajoie, Job Dekker, Nir Friedman, and Oliver J. Rando. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*, 162(1):108–119, 2015.

[63] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual Molecular Dynamics. *J. Mol. Graphics*, 14(February):33–38, 1996.

[64] Sarah J Hurst, Abigail K R Lytton-Jean, and Chad A Mirkin. Maximizing DNA loading on a range of gold nanoparticle sizes. *Anal. Chem.*, 78(24):8313–8, December 2006.

[65] Sajad Hussain, Damien Goutte-gattat, Nils Becker, Sam Meyer, and Manu Shubhdarshan. interactions and 3D organization of the nucleosome. pages 1–6, 2010.

[66] William L. Hwang, Sebastian Deindl, Bryan T. Harada, and Xiaowei Zhuang. Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. *Nature*, 512(7513):213–217, 2014.

[67] Takashi Ito, Michael Bulger, Michael J. Pazin, Ryuji Kobayashi, and James T. Kadonaga. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell*, 90(1):145–155, 1997.

[68] T Jenuwein and C D Allis. Translating the histone code. *Science (New York, N.Y.)*, 293(5532):1074–80, August 2001.

[69] Isabel Jimenez-Useche, Nathan P Nurse, Yuqing Tian, Bhargav S Kansara, Daphne Shim, and Chongli Yuan. Dna methylation effects on tetra-nucleosome compaction and aggregation. *Biophysical journal*, 107(7):1629–1636, 2014.

[70] Matthew R Jones, Robert J Macfarlane, Byeongdu Lee, Jian Zhang, Kaylie L Young, Andrew J Senesi, and Chad A Mirkin. DNA-nanoparticle superlattices formed from anisotropic building blocks. *Nat. Mater.*, 9(11):913–7, November 2010.

[71] Peter C. Kahn. DEFINING THE AXIS OF A HELIX. *Computers & Chemistry*, 13(3):185–189, 1989.

[72] Yu P. Kalmykov. Rotational Brownian motion in an external potential: The Langevin equation approach. *J. Mol. Liq.*, 69(9 SPEC. ISS.):117–131, 1996.

[73] Noam Kaplan, Irene K Moore, Yvonne Fondufe-Mittendorf, Andrea J Gossett, Desiree Tillo, Yair Field, Emily M LeProust, Timothy R Hughes, Jason D Lieb, Jonathan Widom, and Eran Segal. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236):362–6, mar 2009.

[74] Johannes Kästner. Umbrella sampling. *Wiley Interdiscip Rev Comput Mol Sci*, 1(6):932–942, November 2011.

[75] Hiroo Kenzaki and Shoji Takada. Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLoS Comput. Biol.*, 11(8):e1004443, 2015.

[76] Nick Kepper, Ramona Ettig, Rene Stehr, Sven Marnach, Gero Wedemann, and Karsten Rippe. Force spectroscopy of chromatin fibers: extracting energetics and structural information from Monte Carlo simulations. *Biopolymers*, 95(7):435–47, jul 2011.

[77] Nick Kepper, Dietrich Foethke, Rene Stehr, Gero Wedemann, and Karsten Rippe. Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation. *Biophysical journal*, 95(8):3692–705, October 2008.

[78] Peter V Kharchenko, Artyom A Alekseyenko, Yuri B Schwartz, Aki Minoda, Nicole C Riddle, Jason Ernst, Peter J Sabo, Erica Larschan, Andrey A Gorchakov, Tingting Gu, Daniela Linder-Basso, Annette Plachetka, Gregory Shanower, Michael Y Tolstorukov, Lovelace J Luquette, Ruibin Xi, Youngsook L Jung, Richard W Park, Eric P Bishop, Theresa K Canfield, Richard Sandstrom, Robert E Thurman, David M MacAlpine, John A Stamatoyannopoulos, Manolis Kellis, Sarah C R Elgin, Mitzi I Kuroda, Vincenzo Pirrotta, Gary H Karpen, and Peter J Park. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature*, 471(7339):480–485, 2011.

[79] Sangtae Kim. Singularity solutions for ellipsoids in low-reynolds-number flows: with applications to the calculation of hydrodynamic interactions in suspensions of ellipsoids. *Int. J. Multiph. Flow*, 12(3):469–491, 1986.

[80] Henrike Klinker, Felix Mueller-Planitz, Renliang Yang, Ignasi Forne, Chuan Fa Liu, Lars Nordenskiold, and Peter B. Becker. ISWI remodelling of physiological chromatin fibres acetylated at lysine 16 of histone H4. *PLoS ONE*, 9(2):e88411, 2014.

[81] Joseph A. Knezetic and Donal S. Luse. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell*, 45(1):95–104, 1986.

[82] Thomas A Knotts, Nitin Rathore, David C Schwartz, and Juan J de Pablo. A coarse grain model for DNA. *J. Chem. Phys.*, 126(8):084901, March 2007.

[83] M. Kruithof and J. van Noort. Hidden Markov Analysis of Nucleosome Unwrapping Under Force. *Biophys. J.*, 96(9):3708–3715, 2009.

[84] I. Kulić and H. Schiessel. Chromatin Dynamics: Nucleosomes go Mobile through Twist Defects. *Phys. Rev. Lett.*, 91(14):148103, oct 2003.

[85] I M Kulić and H Schiessel. Nucleosome repositioning via loop formation. *Biophys. J.*, 84(5):3197–3211, 2003.

[86] I. M. Kulić and H. Schiessel. DNA Spools under Tension. *Phys. Rev. Lett.*, 92(22):228101, jun 2004.

[87] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, 16(11):1339–1350, 1995.

[88] Siavash K Kurdistani, Saeed Tavazoie, and Michael Grunstrin. Mapping Global Histone Acetylation Patterns to Gene Expression. *Cell*, 117:721–733, 2004.

[89] Alessandro Laio and Francesco L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep Prog Phys*, 71(12):126601, December 2008.

[90] Dongyun Lee, Shengguo Jia, Sarbajit Banerjee, Joze Bevk, Irving P. Herman, and Jeffrey W. Kysar. Viscoplastic and granular behavior in films of colloidal nanocrystals. *Phys. Rev. Lett.*, 98:026103, 2007.

[91] Joshua Lequieu, Andrés Córdoba, Daniel Hinckley, and Juan J. de Pablo. Mechanical Response of DNANanoparticle Crystals to Controlled Deformation. *ACS Central Science*, page acscentsci.6b00170, 2016.

[92] Joshua Lequieu, Andrés Córdoba, David C. Schwartz, and Juan J. de Pablo. Tension-Dependent Free Energies of Nucleosome Unwrapping. *ACS Cent. Sci.*, 2(9):660–666, 2016.

[93] Joshua Lequieu, David C. Schwartz, and Juan J. de Pablo. Direct Observation of Sequence-Dependent Nucleosome Sliding. *To Be Published*, 2017.

[94] Joshua P Lequieu, M Hinckley, and Juan J de Pablo. A molecular view of DNA-conjugated nanoparticle association energies. *Soft Matter*, 11:1919–1929, 2015.

[95] Gu Li, Marcia Levitus, Carlos Bustamante, and Jonathan Widom. Rapid spontaneous accessibility of nucleosomal DNA. *Nature structural & molecular biology*, 12(1):46–53, January 2005.

[96] Ting Li, Rastko Sknepnek, Robert J Macfarlane, Chad A Mirkin, and Monica Olvera de la Cruz. Modeling the crystallization of spherical nucleic acid nanoparticle conjugates with molecular dynamics simulations. *Nano Lett.*, 12(5):2509–14, May 2012.

[97] Ting Li, Rastko Sknepnek, and Monica Olvera de la Cruz. Thermally active hybridization drives the crystallization of DNA-functionalized nanoparticles. *J. Am. Chem. Soc.*, 135(23):8535–41, June 2013.

[98] Wenfei Li, Peter G Wolynes, and Shoji Takada. Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 108(9):3504–9, March 2011.

[99] Erez Lieberman-aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, and Leonid A Mirny. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326(9 October 2009):289–293, 2009.

[100] Chih L. Liu, Tommy Kaplan, Minkyu Kim, Stephen Buratowski, Stuart L. Schreiber, Nir Friedman, and Oliver J. Rando. Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biology*, 3(10), 2005.

[101] Yahli Lorch, Barbara Davis, and Roger D Kornberg. Chromatin remodeling by DNA bending, not twisting. *Proc. Nat. Acad. Sci. USA.*, 102(5):1329–1332, 2005.

[102] Yahli Lorch, Janice W. LaPointe, and Roger D. Kornberg. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, 49(2):203–210, 1987.

[103] P T Lowary and J Widom. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of molecular biology*, 276(1):19–42, February 1998.

[104] Fang Lu, Kevin G Yager, Yugang Zhang, Huolin Xin, and Oleg Gang. Superlattices assembled through shape-induced directional binding. *Nat. Commun.*, 6:6912, 2015.

[105] Karolin Luger, Armin W Mader, Robin K Richmond, David F Sargent, and Timothy J Richmond. Crystal structure of the nucleosome resolution core particle at 2.8 A. *Nature*, 389:251–260, 1997.

[106] Antoni Luque, Rosana Collepardo-Guevara, Sergei Grigoryev, and Tamar Schlick. Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. *Nucleic Acids Res.*, 42(12):7553–7560, 2014.

[107] Robert J Macfarlane, Byeongdu Lee, Matthew R Jones, Nadine Harris, George C Schatz, and Chad A Mirkin. Nanoparticle superlattice engineering with DNA. *Science*, 334(6053):204–8, October 2011.

[108] Robert J Macfarlane, Ryan V Thaner, Keith A Brown, Jian Zhang, Byeongdu Lee, Son-Binh T Nguyen, and Chad A Mirkin. Importance of the DNA 'bond' in programmable nanoparticle crystallization. *Proc. Natl. Acad. Sci. U.S.A.*, 111(42):14995–15000, October 2014.

[109] Andrew H. Mack, Daniel J. Schlingman, Robielyn P. Ilagan, Lynne Regan, and Simon G.J. Mochrie. Kinetics and Thermodynamics of Phenotype: Unwinding and Rewinding the Nucleosome. *J. Mol. Biol.*, 423(5):687–701, 2012.

[110] Kazuhiro Maeshima, Saera Hihara, and Mikhail Eltsov. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr. Opin. Cell Biol.*, 22(3):291–7, 2010.

[111] Mathew M Maye, Mudalige Thilak Kumara, Dmytro Nykypanchuk, William B Sherman, and Oleg Gang. Switching binary states of nanoparticle superlattices and dimer clusters by DNA strands. *Nat. Nanotechnol.*, 5(2):116–20, February 2010.

[112] J D McGhee, J M Nickol, G Felsenfeld, and D C Rau. Higher order structure of chromatin: orientation of nucleosomes within the 30 nm chromatin solenoid is independent of species and spacer length. *Cell*, 33(3):831–841, 1983.

[113] G Meersseman, S Pennings, and E M Bradbury. Mobile nucleosomes–a general behavior. *EMBO J.*, 11(8):2951–9, 1992.

[114] Sam Meyer, Nils B Becker, Sajad Hussain Syed, Damien Goutte-Gattat, Manu Shubhdarshan Shukla, Jeffrey J Hayes, Dimitar Angelov, Jan Bednar, Stefan Dimitrov,

and Ralf Everaers. From crystal and NMR structures, footprints and cryo-electron-micrographs to large and soft structures: nanoscale modeling of the nucleosomal stem. *Nucleic Acids Res.*, 39(21):9139–9154, 2011.

[115] Shirley Mihardja, Andrew J Spakowitz, Yongli Zhang, and Carlos Bustamante. Effect of force on mononucleosomal dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(43):15871–6, October 2006.

[116] Chad A Mirkin, Robert L Lestsinger, Robert C. Mucic, and James J Storhoff. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature*, 382:607, 1996.

[117] Leonid A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Research*, 19(1):37–51, 2011.

[118] Bianca Mladek, Julia Fornleitner, Francisco Martinez-Veracoechea, Alexandre Dawid, and Daan Frenkel. Quantitative Prediction of the Phase Diagram of DNA-Functionalized Nanosized Colloids. *Phys. Rev. Lett.*, 108(26):1–5, June 2012.

[119] Bianca M. Mladek, Julia Fornleitner, Francisco J. Martinez-Veracoechea, Alexandre Dawid, and Daan Frenkel. Procedure to construct a multi-scale coarse-grained model of DNA-coated colloids from experimental data. *Soft Matter*, 9(30):7342, 2013.

[120] Laleh Mollazadeh-Beidokhti, Farshid Mohammad-Rafiee, and Helmut Schiessel. Nucleosome dynamics between tension-induced states. *Biophys. J.*, 102(10):2235–2240, 2012.

[121] G B Morin. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell*, 59(3):521–529, 1989.

[122] Georgette Moyle-Heyrman, Hannah S Tims, and Jonathan Widom. Structural Constraints in Collaborative Competition of Transcription Factors Against the Nucleosome. *J. Mol. Biol.*, 412(4):634–646, 2011.

[123] R K Moyzis, J M Buckingham, L S Cram, M Dani, L L Deaven, M D Jones, J Meyne, R L Ratliff, and J R Wu. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Nat. Acad. Sci. USA.*, 85(18):6622–6626, 1988.

[124] Klara E. Mueggenburg, Xiao-Min Lin, Rodney H. Goldsmith, and Heinrich M. Jaeger. Elastic membranes of close-packed nanoparticle arrays. *Nat. Mater.*, 6(9):656–660, 2007.

[125] Felix Mueller-Planitz, Henrike Klinker, and Peter B Becker. Nucleosome sliding mechanisms: new twists in a looped history. *Nat. Struct. Mol. Biol.*, 20(9):1026–32, sep 2013.

[126] Felix Mueller-Planitz, Henrike Klinker, Johanna Ludwigsen, and Peter B Becker. The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. *Nat. Struct. Mol. Biol.*, 20(1):82–9, 2013.

[127] Thuy T M Ngo, Qiucen Zhang, Ruobo Zhou, Jaya G Yodh, and Taekjip Ha. Asymmetric Unwrapping of Nucleosomes under Tension Directed by DNA Local Flexibility. *Cell*, 160(6):1135–1144, 2015.

[128] Van a. Ngo, Rajiv K. Kalia, Aiichiro Nakano, and Priya Vashishta. Supercrystals of DNA-Functionalized Gold Nanoparticles: A Million-Atom Molecular Dynamics Simulation Study. *The Journal of Physical Chemistry C*, 116(36):19579–19585, September 2012.

[129] Dmytro Nykypanchuk, Mathew M Maye, Daniel van der Lelie, and Oleg Gang. DNA-guided crystallization of colloidal nanoparticles. *Nature*, 451(7178):549–52, January 2008.

[130] Matthew N O'Brien, Matthew R Jones, Byeongdu Lee, and Chad A Mirkin. Anisotropic nanoparticle complementarity in DNA-mediated co-crystallization. *Nat. Mater.*, 14(August):833–840, 2015.

[131] Thomas E Ouldridge, Ard A Louis, and Jonathan P K Doye. Extracting bulk properties of self-assembling systems from small simulations. *J. Phys. Condens. Matter*, 22(10):104102, March 2010.

[132] Olivia Padovan-Merhar, Fernando Vargas Lara, and Francis W Starr. Stability of DNA-linked nanoparticle crystals: effect of number of strands, core size, and rigidity of strand attachment. *The Journal of chemical physics*, 134(24):244701, June 2011.

[133] Daniel J Park, Chuan Zhang, Jessie C Ku, Yu Zhou, George C Schatz, and Chad A Mirkin. Plasmonic photonic crystals realized through DNA-programmable assembly. *Proc. Natl. Acad. Sci. U.S.A.*, 112(4):977–981, 2015.

[134] SJ Park, AA Lazarides, James J. Storhoff, Lorenzo Pesce, and Chad A. Mirkin. The structural characterization of oligonucleotide-modified gold nanoparticle networks formed by DNA hybridization. *J. Phys. Chem. B*, 108:12375–12380, 2004.

[135] Sung Yong Park, Abigail K R Lytton-Jean, Byeongdu Lee, Steven Weigand, George C Schatz, and Chad A. Mirkin. DNA-programmable nanoparticle crystallization. *Nature*, 451(7178):553–6, January 2008.

[136] Marco Pasi and Richard Lavery. Structure and dynamics of DNA loops on nucleosomes studied with atomistic, microsecond-scale molecular dynamics. *Nucleic Acids Res.*, 44(11):5450–5456, 2016.

[137] S. Pennings, G. Meersseman, and E. M. Bradbury. Mobility of positioned nucleosomes on 5 S rDNA. *J. Mol. Biol.*, 220(1):101–110, 1991.

[138] Ognjen Perišić, Rosana Collepardo-Guevara, and Tamar Schlick. Modeling studies of chromatin fiber structure as a function of DNA linker length. *J. Mol. Biol.*, 403(5):777–802, 2010.

[139] Sabrina Pisano, Enrico Marchioni, Alessandra Galati, Rosella Mechelli, Maria Savino, and Stefano Cacchione. Telomeric Nucleosomes Are Intrinsically Mobile. *J. Mol. Biol.*, 369(5):1153–1162, 2007.

[140] Steve Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics*, 117:1–19, 1995.

[141] Lisa R Racki, Janet G Yang, Nariman Naber, Peretz D Partensky, Ashley Acevedo, Thomas J Purcell, Roger Cooke, Yifan Cheng, and Geeta J Narlikar. The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nature*, 462(7276):1016–21, 2009.

[142] Paolo Raiteri, Alessandro Laio, Francesco Luigi Gervasio, Cristian Micheletti, and Michele Parrinello. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B*, 110(8):3533–9, March 2006.

[143] Padinhateeri Ranjith, Jie Yan, and John F Marko. Nucleosome hopping and sliding kinetics determined from dynamics of single chromatin fibers in Xenopus egg extracts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34):13649–54, August 2007.

[144] Suhas S P Rao, Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, Ido Machol, Arina D. Omer, Eric S. Lander, and Erez Lieberman Aiden. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, dec 2014.

[145] Timothy J Richmond and Curt A Davey. The structure of DNA in the nucleosome core. *Nature*, 423(6936):145–50, 2003.

[146] Viviana I. Risca, Sarah K. Denny, Aaron F. Straight, and William J. Greenleaf. Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. *Nature*, 541(7636):237–241, 2017.

[147] Viviana I. Risca and William J. Greenleaf. Unraveling the 3D genome: Genomics tools for multiscale exploration. *Trends in Genetics*, 31(7):357–372, 2015.

[148] Philip J J Robinson, Louise Fairall, Van a T Huynh, and Daniela Rhodes. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. U.S.A.*, 103(17):6506–6511, 2006.

[149] Philip J J Robinson and Daniela Rhodes. Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Current opinion in structural biology*, 16(3):336–43, June 2006.

[150] W Benjamin Rogers and John C Crocker. Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling. *Proc. Natl. Acad. Sci. U.S.A.*, 108(38):15687–92, September 2011.

[151] Michael B Ross, Jessie C Ku, Victoria M Vaccarezza, George C Schatz, and Chad A Mirkin. Nanoscale form dictates mesoscale function in plasmonic DNA-nanoparticle superlattices. *Nat. Nanotechnol.*, 10(5):453–458, apr 2015.

[152] G. Sadri-Vakili, B. Bouzou, C. L. Benn, M.-O. Kim, P. Chawla, R. P. Overland, K. E. Glajch, E. Xia, Z. Qiu, S. M. Hersch, T. W. Clark, G. J. Yohrling, and J.-H. J. Cha. Histones associated with downregulated genes are hypo-acetylated in Huntington's disease models. *Hum. Mol. Genet.*, 16(11):1293–1306, 2007.

[153] K. Michael Salerno, Dan S. Bolintineanu, J. Matthew D Lane, and Gary S. Grest. High strength, molecularly thin nanoparticle membranes. *Phys. Rev. Lett.*, 113(25):258301, 2014.

[154] E J Sambriski, D C Schwartz, and J J de Pablo. A mesoscale model of DNA and its renaturation. *Biophys. J.*, 96(5):1675–90, March 2009.

[155] Raynaldo T. Scarlett, Marie T. Ung, John C. Crocker, and Talid Sinno. A mechanistic view of binary colloidal superlattice formation using DNA-directed interactions. *Soft Matter*, 7(5):1912, 2011.

[156] Thomas Schalch, Sylwia Duda, David F Sargent, and Timothy J Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047):138–41, jul 2005.

[157] H. Schiessel, J. Widom, R. Bruinsma, and W. Gelbart. Polymer Reptation and Nucleosome Repositioning. *Physical Review Letters*, 86(19):4414–4417, May 2001.

[158] Dustin E Schones, Kairong Cui, Suresh Cuddapah, Tae-Young Roh, Artem Barski, Zhibin Wang, Gang Wei, and Keji Zhao. Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, 132:887–898, 2008.

[159] Eran Segal, Yvonne Fondufe-Mittendorf, Lingyi Chen, AnnChristine Thåström, Yair Field, Irene K Moore, Ji-Ping Z Wang, and Jonathan Widom. A genomic code for nucleosome positioning. *Nature*, 442(7104):772–778, aug 2006.

[160] Maxim Y Sheinin, Ming Li, Mohammad Soltani, Karolin Luger, and Michelle D Wang. Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. *Nat. Commun.*, 4:2579, jan 2013.

[161] Elena V Shevchenko, Dmitri V Talapin, Nicholas A Kotov, Stephen O'Brien, and Christopher B Murray. Structural diversity in binary nanoparticle superlattices. *Nature*, 439(7072):55–9, January 2006.

[162] T E Shrader and D M Crothers. Artificial nucleosome positioning sequences. *Proc. Nat. Acad. Sci. USA.*, 86:7481–7482, 1989.

[163] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, and G. Li. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science*, 344(6182):376–380, apr 2014.

[164] Sunita Srivastava, Dmytro Nykypanchuk, Masafumi Fukuto, Jonathan D Halverson, Alexei V Tkachenko, Kevin G Yager, and Oleg Gang. Two-dimensional DNA-programmable assembly of nanoparticles at liquid interfaces. *Journal of the American Chemical Society*, 136(23):8323–32, June 2014.

[165] René Stehr, Nick Kepper, Karsten Rippe, and Gero Wedemann. The effect of internucleosomal interaction on folding of the chromatin fiber. *Biophys. J.*, 95(8):3677–91, oct 2008.

[166] Rene Stehr, Robert Schopflin, Ramona Ettlg, Nick Kepper, Karsten Rippe, and Gero Wedemann. Exploring the conformational space of chromatin fibers and their stability by numerical dynamic phase diagrams. *Biophys. J.*, 98(6):1028–1037, 2010.

[167] Arnold Stein, Taichi E. Takasuka, and Clayton K. Collings. Are nucleosome positions in vivo primarily determined by histone-DNA sequence preferences? *Nucleic Acids Res.*, 38(3):709–719, 2010.

[168] Chris Stockdale, Andrew Flaus, Helder Ferreira, and Tom Owen-Hughes. Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J. Biol. Chem.*, 281(24):16279–16288, 2006.

[169] A.J. Stone. The description of bimolecular potentials, forces and torques: the S and V function expansions. *Mol. Phys.*, 36(1):241–256, 1978.

[170] Ralf Strohner, Malte Wachsmuth, Karoline Dachauer, Jacek Mazurkiewicz, Julia Hochstatter, Karsten Rippe, and Gernot Längst. A 'loop recapture' mechanism for ACF-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.*, 12(8):683–690, 2005.

[171] Kevin Struhl and Eran Segal. Determinants of nucleosome positioning. *Nature structural & molecular biology*, 20(3):267–73, March 2013.

[172] B Sudhanshu, S Mihardja, E F Koslover, S Mehraeen, C Bustamante, and a J Spakowitz. Tension-dependent structural deformation alters single-molecule transition kinetics. *Proceedings of the National Academy of Sciences of the United States of America*, 108(5):1885–90, February 2011.

[173] Robert K. Suto, Rajeswari S. Edayathumangalam, Cindy L. White, Christian Melander, Joel M. Gottesfeld, Peter B. Dervan, and Karolin Luger. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.*, 326(2):371–380, 2003.

[174] Shawn J Tan, Michael J Campolongo, Dan Luo, and Wenlong Cheng. Building plasmonic nanostructures with DNA. *Nat. Nanotechnol.*, 6(5):268–276, 2011.

[175] Song Tan and Curt A. Davey. Nucleosome structural studies. *Curr. Opin. Struc. Biol.*, 21(1):128–136, 2011.

[176] A Thastrom, P T Lowary, H R Widlund, H Cao, M Kubista, and J Widom. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.*, 288(2):213–229, 1999.

[177] A Thåström, P T Lowary, and J Widom. Measurement of histone-DNA interaction free energy in nucleosomes. *Methods*, 33(1):33–44, may 2004.

[178] Panagiotis E Theodorakis, Christoph Dellago, and Gerhard Kahl. A coarse-grained model for DNA-functionalized spherical colloids, revisited: Effective pair potential from parallel replica simulations. *J. Chem. Phys.*, 138(2):025101, January 2013.

[179] Aidan P Thompson, Steven J Plimpton, and William Mattson. General formulation of pressure and stress tensor for arbitrary many-body interaction potentials under periodic boundary conditions. *J. Chem. Phys.*, 131(15):154107, oct 2009.

[180] Hannah S Tims, Kaushik Gurunathan, Marcia Levitus, and Jonathan Widom. Dynamics of nucleosome invasion by DNA binding proteins. *J. Mol. Biol.*, 411(2):430–48, aug 2011.

[181] Michael Y. Tolstorukov, Andrew V. Colasanti, David M. McCandlish, Wilma K. Olson, and Victor B. Zhurkin. A Novel Roll-and-Slide Mechanism of DNA Folding in Chromatin: Implications for Nucleosome Positioning. *J. Mol. Biol.*, 371(3):725–738, 2007.

[182] David J Tremethick. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell*, 128(4):651–4, February 2007.

[183] Anton Valouev, Steven M Johnson, Scott D Boyd, Cheryl L Smith, Andrew Z Fire, and Arend Sidow. Determinants of nucleosome organization in primary human cells. *Nature*, 474(7352):516–520, 2011.

[184] Eric Vanden-Eijnden and Maddalena Venturoli. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.*, 130(19):194103, may 2009.

[185] S. Wang, J.-H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C.-t. Wu, and X. Zhuang. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, 353(6299):598–602, 2016.

[186] W Wang, G W Li, C Chen, X S Xie, and X Zhuang. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*, 333(6048):1445–1449, 2011.

[187] Yifan Wang, Jianhui Liao, Sean P. Mcbride, Efi Efrati, Xiao Min Lin, and Heinrich M. Jaeger. Strong Resistance to Bending Observed for Nanoparticle Membranes. *Nano Lett.*, 15(10):6732–6737, 2015.

[188] Gero Wedemann and Jörg Langowski. Computer simulation of the 30-nanometer chromatin fiber. *Biophys. J.*, 82(6):2847–59, jun 2002.

[189] Iestyn Whitehouse, Oliver J Rando, Jeff Delrow, and Toshio Tsukiyama. Chromatin remodelling at promoters suppresses antisense transcription. *Nature*, 450(7172):1031–1035, 2007.

[190] J Widom and a Klug. Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. *Cell*, 43(1):207–213, 1985.

[191] Gregory A. Williams, Ryohei Ishige, Olivia R. Cromwell, Jaeyoon Chung, Atsushi Takahara, and Zhibin Guan. Mechanically Robust and Self-Healable Superlattice Nanocomposites by Self-Assembly of Single-Component "Sticky" Polymer-Grafted Nanoparticles. *Adv. Mater.*, 27(26):3934–3941, 2015.

[192] S P Williams, B D Athey, L J Muglia, R S Schappe, a H Gough, and J P Langmore. Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophys. J.*, 49(1):233–248, 1986.

[193] Christian J Wippo, Lars Israel, Shinya Watanabe, Andreas Hochheimer, Craig L Peterson, and Philipp Korber. The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. *EMBO J.*, 30(7):1277–88, 2011.

[194] T. Wocjan, K. Klenin, and J. Langowski. Brownian dynamics simulation of DNA unrolling from the nucleosome. *J. Phys. Chem. B*, 113(9):2639–2646, 2009.

[195] C L Woodcock, S a Grigoryev, R a Horowitz, and N Whitaker. A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc. Nat. Acad. Sci. USA.*, 90(19):9021–9025, 1993.

[196] Janet G Yang, Tina Shahian Madrid, Elena Sevastopoulos, and Geeta J Narlikar. The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing. *Nat. Struct. Mol. Biol.*, 13(12):1078–1083, 2006.

[197] Jueng Soo You and Peter A. Jones. Cancer Genetics and Epigenetics: Two Sides of the Same Coin? *Cancer Cell*, 22(1):9–20, 2012.

[198] Kaylie L. Young, Michael B. Ross, Martin G. Blaber, Matthew Rycenga, Matthew R. Jones, Chuan Zhang, Andrew J. Senesi, Byeongdu Lee, George C. Schatz, and Chad A. Mirkin. Using DNA to design plasmonic metamaterials with tunable optical properties. *Adv. Mater.*, 26(4):653–659, 2014.

[199] Gabriel E Zentner and Steven Henikoff. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol*, 20(3):259–66, 2013.

[200] Habtamu Zewdie. Computer-simulation studies of diskotic liquid crystals. *Phys. Rev. E*, 57(2):1793–1805, 1998.

[201] Bin Zhang, Weihua Zheng, Garegin A. Papoian, and Peter G Wolynes. Exploring the Free Energy Landscape of Nucleosomes. *J. Am. Chem. Soc.*, 138(26):8126–8133, 2016.

[202] Chuan Zhang, Robert J Macfarlane, Kaylie L Young, Chung Hang J Choi, Liangliang Hao, Evelyn Auyeung, Guoliang Liu, Xiaozhu Zhou, and Chad A Mirkin. A general approach to DNA-programmable atom equivalents. *Nat. Mater.*, 12(8):741–6, August 2013.

[203] Yong Zhang, Zarmik Moqtaderi, Barbara P Rattner, Ghia Euskirchen, Michael Snyder, James T Kadonaga, X Shirley Liu, and Kevin Struhl. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat. Struct. Mol. Biol.*, 16(8):847–52, aug 2009.

[204] Yugang Zhang, Fang Lu, Kevin G Yager, Daniel van der Lelie, and Oleg Gang. A general strategy for the DNA-mediated self-assembly of functional nanoparticles into heterogeneous systems. *Nat. Nanotechnol.*, 8(11):865–72, November 2013.

[205] Yugang Zhang, Fang Lu, Kevin G Yager, Daniel van der Lelie, and Oleg Gang. A general strategy for the DNA-mediated self-assembly of functional nanoparticles into heterogeneous systems. *Nat. Nanotechnol.*, 8(11):865–872, nov 2013.

[206] Yugang Zhang, Suchetan Pal, Babji Srinivasan, Thi Vo, Sanat Kumar, and Oleg Gang. Selective transformations between nanoparticle superlattices via the reprogramming of DNA-mediated interactions. *Nat. Mater.*, 14(8):840–847, 2015.

[207] Vicky W Zhou, Alon Goren, and Bradley E Bernstein. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.*, 12(1):7–18, 2011.

[208] Martin Zofall, Jim Persinger, Stefan R Kassabov, and Blaine Bartholomew. Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome. *Nat. Struct. Mol. Biol.*, 13(4):339–346, 2006.