THE UNIVERSITY OF CHICAGO

CONSTRUCTING CHROMATIN THROUGH MULTISCALE SIMULATIONS: A NUCLEOSOME-CENTRIC APPROACH

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

BY
JOSHUA J MOLLER

CHICAGO, ILLINOIS
AUGUST 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

There are many people without whom this work would not have been completed. I have enjoyed my tenure in Chicago and owe the world to the people who helped me along the way. Whether personal or research help, there are so many people that I would like to take the time to thank. One section is not enough to express the gratitude I have for all the help my family, friends, and coworkers have given me throughout my graduate career, but it will have to serve.

To begin, I would like to thank my family for supporting me throughout my Ph.D. My parents, Deborah and Neal, and sisters, Alyssa and Kayla, who have been with me throughout the entire process. I also have to thank my dog Milo who may be the goofiest golden retriever to exist, but that's the reason why we love him. During my tenure, my parents may have decided to uproot and move across the country, but their calls and visits pushed me forward.

I have developed many long-lasting friendships throughout my Ph.D. These friendships helped ease my mind from the tougher challenges I faced during my Ph.D. There are many members within my research group who I consider some of my greatest friends and their assistance can not be understated. I would like to express gratitude towards the following people: Abhinendra Singh, Alec Bowen, Ashley Guo, Cody Bezik, Emre Sevgen, Grant Garner, Lucas Antony, Phil Rauscher, and Yamil Colón. Outside of my research group, I would especially like to thank Elizabeth Ashley, Hung-Shen Chang, Moshe Dolejsi, Paul Jerger, Peyman Hosseinchi, Ruben Waldman, Taylor Gray, and Zack Jarin. I also thank my office mates in Giulia Galli's lab for accepting me as one of their own, especially Arin Greenwood, Viktor Rozsa, Federico Giberti, Nicholas Brawand, and Ryan McAvoy. I would also like to thank my undergraduate friends who stayed in touch and supported me throughout my time in graduate school, especially Marissa Castner, Adam Connor, Xinyu He, Sarah Kuhl, Tim McGregor, and Michael Richards. I also want to take the time to thank Viviana

Palacio-Betancur. She has been an incredible scientific inspiration, executive chef, and my best friend. I hope to one day teach her as much as she has taught me.

Next, I would like to take the time to thank my research mentors and teachers. I would personally like to thank the de Pablo group and their hospitality that led me to succeed as a researcher. As a starting graduate student, I joined the chromatin project due to the strong impression that Andrés Córdoba and Joshua Lequieu had on me. Their advice and modeling intuition developed my own understanding as a scientist, leading to my development as a researcher. I especially thank Joshua Lequieu for continuing to advise me long after he was a member of the group. Most of these chapters would be empty without his mentorship. I also thank Artem Rumyantsev, Michael Webb, and Nicholas Jackson for their polymer physics lessons. I also extend thanks to the now-bustling chromatin subgroup members of Aria Coraor, Soren Kyhl, and Walter Alvarado. This work will continue on through you. Thanks are also in order for my committee members Andrew Ferguson (UChicago), Tobin Sosnick (UChicago), and Vadim Backman (Northwestern) for their exceptional guidance.

I would also extend thanks to the PME faculty and staff who bring life to the department. In particular, I thank Novia Pagone, Rovana Popoff, Janet Boland, and Heather Crews for their incredible administrative support. Their contributions to the department allowed me to focus on my research.

Lastly, I extend my deepest thanks to my advisor, Juan José de Pablo, for his never-ending support and guidance throughout my graduate career. He guided me through every personal and professional development opportunity. Without his guidance, I would not be able to get where I am today. I consider him to be an incredible advisor and great mentor. I hope this work is sufficient to repay him for the years of his advice and guidance.

I have called the city of Chicago home for many years. Being an Illinois-lifer has not been without its struggles. The state still seems to be lacking in topography and the sports teams have for the most part been mediocre or abysmal. While my favorite team, the Chicago

Bears, have not won in my lifetime, I still hold out hope. However, I have to thank the 2016 World Series Champion Chicago Cubs who showed me that if they can take 108 years to finally win a championship, then I can finish a Ph.D.

ABSTRACT

Chromatin is a hierarchically structured fiber that regulates gene expression. Consisting of a complex network of DNA and proteins, chromatin is host to dynamic modes that facilitate genomic packaging. A comprehensive description of chromatin structure and dynamics is invaluable for a fundamental understanding of how inheritable genetic diseases originate. We now know that genetic diseases can result from epigenomic phenomena which alter the thermodynamics of the nucleosome, the fundamental unit of chromatin. Therefore, it is important to not only establish a physical description of the link between nucleosome physics and the structure of the chromatin fiber, but also the factors which alter said link. Within the last few decades, imaging and chromosome conformation capture techniques have revealed a number of structural and statistical features of the packaged chromatin fiber at a hitherto unavailable level of resolution. In this work, we deploy a comprehensive, multi-scale modeling approach to bridge the gap between chromatin structure length scales: from the nucleosome to the supramolecular chromatin fiber. We begin by quantifying the anisotropic interaction potential between nucleosomes to reconcile discrepancies in experimental endeavors. This nucleosome pair-interaction serves as the backbone for the development of a new coarse-grained model, known as the 1-Cylinder-Per-Nucleosome model (1CPN). Through free energy analysis of the 1CPN model, varying both linker DNA lengths and the linker histone, we determine that the level of description for mesoscopic structures of chromatin can be reduced to a three-nucleosome repeat unit. Through this result, we introduce a statistical framework to grow large chains of chromatin that we use to determine the correlation length, in kbp, of nucleosome-level dynamics. Lastly, we implement our findings into a machine vision workflow to determine the *in situ* structure of chromatin from ChromSTEM results. Taken together, this dissertation describes a comprehensive bottom-up modeling approach to chromatin structure that reveals unique motifs that arise from nucleosome-level physics.

CHAPTER 1

INTRODUCTION

The fundamental processes of life are of considerable interest from a physical perspective. Currently, we have a biological understanding of the processes that follow a central dogma. This dogma posits that DNA stores the code necessary for function, RNA "reads" the DNA, and proteins carry out the function provided by the RNA. While the human genome project fully decoded the genome [1], the next hurdle relies on determining how the factors surrounding the genome – known as the epigenome – influence genomic regulation. Such external, or epigenetic, factors rely on local accessibility to transcribe DNA. From this, one can draw connections between epigenomic regulation and the local DNA structure. As such, a comprehensive picture of the packaged genome structure, referred to as the 3D-genome, is integral in decoding the epigenome. A mechanistic understanding of epigenomic regulation would provide a fundamental understanding of genomic regulation, which could lead to significant advancements in the fields of medicine, agriculture, and energy.

Central to the epigenome and DNA accessibility is the chromatin fiber, a macromolecular complex consisting of DNA and proteins. While the genome encodes the processes necessary for life through the central dogma of biology, chromatin is the catalyst which renders the genome functional. The chromatin fiber enables such functionality through its hierarchical and dynamic structure, overcoming strict DNA packaging constraints to facilitate organization and epigenomic regulation across length scales. [23, 73] In particular, genome packaging poses intriguing questions that are relevant not only to biology, but also to polymer physics and chemistry. The human genome consists of billions of base pairs of DNA that are densely packed in the cell nucleus, well below the theoretical packing limit dictated by the persistence length of double-stranded DNA (~ 50 nm). Such DNA packaging occurs over multiple length scales, eventually leading to a dense nuclear environment. Despite the dense packaging, DNA must be rendered accessible at the gene-scale ($\sim$ kbp) for necessary

processes such as replication [54], transcription [80] and DNA repair [2]. In particular, epigenomic regulation influences chromatin structure through covalent modifications of the fiber such as post-translational modifications (PTMs) and DNA methylation, and substitutions of structural proteins.

At the primary level of chromatin organization, the nucleosome is considered to be the fundamental unit of the chromatin fiber. As determined by X-ray crystallography, the nucleosome is a stable complex consisting of an octamer of histones wrapped ~ 1.7 times in a left-handed superhelix by 147 bp of DNA. [87, 34] Through this structure, the nucleosome facilitates packing of DNA while also providing function to the chromatin fiber; nucleosomes in the chromatin fiber attract each other, resulting in condensed fiber motifs. While condensed fibers are required to package the genome, cellular function depends on the accessibility of the genome, which is regulated at the nucleosome scale through PTMs. Each mark induces a change in the nucleosome structure, dynamics, or binding affinity of proteins to locally tune functionality. [23, 26] For example, acetylation of the H4 tail at lysine 16 (H4K16ac) is associated with active transcription, while trimethylation of the H3 tails at lysine 9 (H3K9me3) is associated with repressed chromatin, or constitutive heterochromatin. [151, 139] The vast number of potential PTMs and their combinations comprise what is referred to as the "histone code." [64] High-throughput sequencing methods, like Chromatin-immunoprecipation sequencing (ChiP-Seq), have established clear links between PTMs to distinct chromatin compartments and motifs. Additionally, researchers have connected defective epigenetic marks to numerous genetic diseases, including cancer. [59, 118, 161] Unfortunately, researchers have yet to crack the histone code in its entirety. It is still unclear whether PTMs have a direct effect on chromatin structure, or alter it through changes to the proteome.

To characterize modifications of the chromatin structure, it is important to establish as a reference the structure and dynamics of unmodified chromatin. The structure of the nucleosome is stable, but recent studies have uncovered several spontaneous dynamic modes.

Some of these modes include DNA breathing [105, 116], sliding [22, 78, 104], and gaping [106]. Like in chromatin structure, these modes are also directly related to epigenomic regulation. [23] For example, DNA breathing facilitates the recruitment of DNA binding proteins and histone chaperones, and is altered by PTMs. [17, 152]

The dynamics of each individual nucleosome, combined with the histone code, contribute to the 3D genome. The organization of chromatin above the nucleosome scale, however, has been the source of controversy. Central to this debate are the *in vivo* existence and structure of the so-called 30-nm chromatin fiber,[70, 110, 128, 130, 147] which was first imaged *in vitro* using Electron Microscopy. Over the span of several decades, the chromatin community has questioned the equilibrium structure of the 30-nm fiber. The argument can be distilled into two candidate structures: the one-start solenoid, and the two-start zigzag. [45, 138] It has been proposed that the 30-nm fiber is an artifact of dilute experimental conditions, and that the dense conditions in the nucleus lead to liquid-like 10-nm fibers. [92, 94, 93] Recently, novel imaging techniques that probe chromatin structure in the cell nucleus, such as Chrom-EMT, have supported this claim by resolving fibers ranging from 5 to 24 nm in diameter. [112]

Important results pertaining to the 3D genome have emerged from high-throughput sequencing and chromosome conformation capture methods (3C, 5C, Hi-C). [37, 122, 123, 131] In particular, Hi-C uses high-throughput sequencing to measure the contact probability of genomic segments as a function of genomic distance. [84] These methods are helping elucidate some of the higher length-scale organization of the genome. Phenomena such as chromatin looping, topologically associating domains (TADs), and chromatin compartmentalization have been elucidated by relying on Hi-C studies. [37, 122, 135] Such 3C methods link the epigenome to structural organization through the discovery that chromatin compartments correlate with epigenetic marks. Recently, single cell Hi-C and novel sequential fluorescence *in situ* hybridization have revealed that population average features are in fact present in

single nuclei, [18] thereby providing an additional incentive to rely on modeling approaches to resolve the 3D genome.

Currently, we have snapshots of the discrete organizational modes of chromatin packaging. However, there is little information that addresses how the chromatin length scales are connected. For chromatin studies to progress, a complete toolset that addresses the following questions is needed: What physics at the nucleosome scale influence the 3D genome? What physics do not? How can they be manipulated? And is there a fundamental unit of chromatin larger than the nucleosome that governs the structure of supramolecular chromatin? These questions are not just important from a fundamental standpoint, but have implications in health as irregularities in chromatin structure are linked to inheritable genetic diseases, including Alzheimer's, Parkinson's disease, and Huntington's disease. [148] In this dissertation, we introduce and highlight a multi-scale coarse-grained modeling workflow with the intent to find a physical link between the primary level of organization in the nucleosome and supramolecular chromatin fiber structures. As a parallel effort, we connect epigenetic phenomena to changes in structure and dynamics within this multi-scale framework.

In Chapter 2, we utilize coarse-grained simulations of nucleosomes to quantify the anisotropic nucleosome pair-potential. In the last decade, multiple experiments have quantified the nucleosome interaction strength, with differing results. Through variable forms of microscopy, experiments probe the interaction resulting in a range of interactions of 3 - 15 k_BT. To find a reliable estimate, we turn to molecular dynamics simulations with free energy sampling to determine a generalized nucleosome pair-potential. We ultimately validate our results and, in the process, reconcile discrepancies between experiments. We find the interaction maximum to be of order 2.69 k_BT, and reduced to 1.70 k_BT in the presence of H4 acetylation, consistent with experiments. We also find that acetylation induces chirality in the interaction landscape, which disrupts interactions favorable for secondary organization. With a good quantitative understanding of the nucleosome interactions, we can incorporate this

work into a new coarse-grained model that focuses on the rich nucleosome physics. This chapter is adapted from Ref. [101].

In Chapter 3, we develop and extend the 1-Cylinder-Per-Nucleosome (1CPN) model to incorporate the H1 linker histone and validate the model. Aimed at feasible chromatin simulations with an emphasis on the nucleosome, the 1CPN model is an additional method to validate our anisotropic potential in the context of a chromatin fiber. Relevant experiments at the oligonucleosome scale come in the form of microscopy and sedimentation analysis of reconstituted chromatin fibers. To quantify the structure and packaging of chromatin, experimentalists evaluate the sedimentation coefficient of fibers in water, $\langle S_{20,w} \rangle$. From this work, we know that linker histone binding results in significantly condensed fibers and the amount of linker length results in non-monotonic packing behavior. [53, 55, 130] We prove here that the bottom-up approach to the 1CPN model and the nucleosome pair-potential without any modification reproduce the sedimentation coefficients from oligonucleosome fibers experiments. Such a result validates our nucleosome-centric approach to the 1CPN model and provides an opportunity to be predictive of the structure of chromatin. This chapter is adapted from Refs. [76, 100]

In Chapter 4, given the promising potential of the 1CPN model, we analyze oligonucleosome fibers to provide a candidate for a fundamental unit of larger chromatin fiber structure. Evidence for larger fundamental units of chromatin that regulate supramolecular structure is becoming prominent. Innovations in super-resolution microscopy and chromosome conformation capture techniques have proposed fundamental units of chromatin larger than the nucleosome. [109, 124, 111] This poses a revolutionary chromatin modeling approach as it promises a reduced level of description, rendering chromosome simulations feasible with little loss of information. Using the 1CPN model, we find that the three-nucleosome free-energy surface agrees with oligonucleosome behavior. We test this finding with other factors such as varying linker lengths, linker histone binding, and H4K16ac modification and find the

three-nucleosome approximation holds. With this finding, we can scale-up our bottom-up approach to focus on trinucleosome behavior without significant loss of detail. This chapter is adapted from Ref. [99]

In Chapter 5, we develop a statistical framework to generate large chains of chromatin from a trinucleosome unit. We refer to this approach as the ideal chromatin chain model (ICCM). First, we test our hypothesis of the trinucleosome by comparing and contrasting with results from a dinucleosome and tetranucleosome framework and find the trinucleosome approximation holds. To evaluate the extent to which individual nucleosomes influence supramolecular stucture, we consider the statistical segment length of the ICCM with variable linker lengths. The ICCM predicts chromatin fibers have correlation lengths to be on the order of 2 - 10 kbp. From this, we deduce that the nucleosome influences the packaging and dynamics of the chromatin fiber on the length scale of genes. We also note that correlation lengths are sensitive to the length of the linking DNA. Lastly, we find that the ICCM predicts recent results that show TADs in single cells without the presence of the cohesin molecular motor, demonstrating that TAD structure is influenced by nucleosome physics. This chapter is adapted from Ref. [103]

In Chapter 6, we use this statistical framework with a machine vision approach to ChromSTEM to determine the properties and structure of chromatin in the cell nucleus. We use the 1CPN model as a form of data augmentation to train a convolutional neural network to predict the number of nucleosomes in a simulation box. The network is tested on recent ChromSTEM data to predict the number of nucleosomes *in situ*. This prediction feeds into a workflow to determine the optimal topology of the chromatin fiber. The resulting structure is analyzed and we determine the physically-driven structure of chromatin, with an understanding of the distribution of linker lengths. This chapter is adapted from Ref. [102]

In Chapter 7, we conclude this dissertation and discuss the potential for future directions. In short, we find that the multi-scale approach outlined here demonstrates the predictive

capability of chromatin modeling and is able to provide insight into the structure of chromatin *in vivo*. We note that a majority of this work considers a first-order approximation of chromatin structure; the current approach focuses on the influence of the nucleosome on the larger structure, subject to a fraction of the PTMs that are known to alter chromatin structure. An extension of this work should incorporate a greater amount of PTMs. Additionally, there are many hierarchical organizational modes that have been left out which could improve predictive capability, such as CTCF, cohesin, and chromatin compartmentalization. Nonetheless, we find the trinculeosome approximation discovered here to be an excellent starting point to continue this multi-scale simulation approach. We believe that the work here is instrumental in pushing towards a comprehensive understanding of chromatin structure and dynamics and its relation to epigenomic regulation.

CHAPTER 2

THE FREE ENERGY LANDSCAPE OF INTER-NUCLEOSOME INTERACTIONS AND ITS RELATION TO CHROMATIN FIBER STRUCTURE

2.1 Abstract

The supramolecular chromatin fiber is governed by molecular scale energetics and interactions. Such energetics originate from the fiber's building block, the nucleosome core particle (NCP). In the past, the chromatin fiber has been examined through perturbative methods in attempts to extract the energetics of nucleosome association in the fiber. This body of work has led to different results from experiments and simulations concerning the nucleosome-nucleosome energetics. We demonstrate the capabilities of coarse-grained simulations to expand on previous experiments and evaluate the energetics inherent to nucleosomes across a variety of parameters in configurational and environmental space. In particular, we describe the NCP-NCP interactions by relying on an anisotropic energetic landscape, rather than a single potential-energy value. The attractions in that landscape arise predominantly from the highly anisotropic interactions provided by the NCP histone N-terminal domain (NTD) tails. These results are found to be in good agreement with prior nucleosome interaction experiments that suggest a maximum interaction energy of $2.69 k_B T$. Furthermore, we examine the influence of crucial epigenetic modifications, such as acetylation of the H4 tail, and how they modify the underlying landscape. These results for acetylated NCP-interactions are also in agreement with experiment. We additionally find an induced chirality in NCP-NCP interactions upon acetylation that reduces interactions that would correspond to a left-handed superhelical chromatin fiber.

2.2 Introduction

The process by which eukoryotic DNA is hierarchically packaged into the cell nucleus is epicentric to cell function and introduces steric barriers for DNA processes such as replication, transcription, and repair. At the smallest length scales, 147 base pairs of DNA are wrapped 1.7 times superhelically around a histone octamer comprised of an H3-H4 tetramer and two H2A-H2B dimers. The resulting packaging unit is known as the nucleosome core particle (NCP).[87, 125, 70] These NCPs then form a "beads-on-a-string" fiber that can self-associate into the chromatin fiber.[70] The dynamic ability of chromatin to locally condense and decondense is central to epigenomic regulation. Despite its crucial role in biology, we have a limited understanding of chromatin's condensed structure and condensation mechanism.

Available evidence on the secondary condensed chromatin structure has led to debate over the last several decades. [39, 70, 127, 151, 157, 158] Two secondary structures of chromatin have been observed *in vitro*: the one-start solenoid fiber [45], and the two-start zigzag [138]. Discussion has gradually shifted from a defined secondary structure *in vitro* towards a disordered, but dynamic network of proteins and DNA *in vivo*. [92] Such a disordered state is supported by recent results from advanced imaging techniques. [7, 38, 112]

Attempts to measure different structural and energetic features of the condensed DNA fiber have relied on approaches that capture the energetics of deformation, such as optical and magnetic tweezers. These tools probe the energetics of chromatin through extension of a single fiber. [79] More specifically, the groups of Bustamante and van Noort have extracted the average association energy of nucleosomes under varying tensions and pull rates. [33, 72] Differing experimental conditions such as salinity, fiber length, and relaxed chromatin fiber ultimately incur into discrepancies in the average nucleosome-nucleosome interaction energy (3.2 and 13.4 k_BT, respectively). As previously mentioned, the condensed fiber is not a well-defined structure, which introduces additional sources of uncertainty. As a result, we have yet to develop a comprehensive and definitive understanding of the nucleosome-nucleosome

interaction energy. An important feature that must also be taken into account is the highly anisotropic distribution of charges that comprise the nucleosome, which results in anisotropic interactions between nucleosomes. With this in mind, it is difficult to define inter-nucleosome energies by relying on an individual order parameter.

The anisotropic and dynamic distribution of charges on the NCP can be partially attributed to the flexibility and availability of the N-terminal domain (NTD) histone tails. These tails are rich in positively charged lysine and arginine residues that attract negatively-charged DNA and negatively-charged histone residues. [41] The H3 and H4 tails have been studied in the context of their positively-charged residues and positioning on the NCP; [31, 114, 162] they are grafted at the dyad axis and the sides of the nucleosome, respectively, which is of particular importance for chromatin fiber condensation. [87] Recently, these tails have been reported to be mobile in the presence of highly dense chromatin fibers, further supporting that the availability of these tails serves to stabilize condensed fibers. [51] Despite its length and flexibility, the H3 tail is believed to predominantly stabilize intra-nucleosome interactions, rather than inter-nucleosome interactions in the absence of divalent salt. [164] In contrast, the H4 tail predominantly contributes to inter-nucleosome interactions; it interacts with the H2A acidic patch at the 16th lysine residue (H4K16), which provides a strong electrostatic contribution to inter-nucleosome energetics. [31, 41, 67, 113] Removal of this interaction can be accomplished through methods such as acetylation or tail removal, which lead to a decrease of inter-nucleosome energetics and chromatin fiber unfolding. It has also been shown through chromatin array cross-linking studies that H4K16 acetylation provides the same energetic decrease as acetylation of the H4 tail at the 5th, 8th, 12th, and 16th lysines combined. [4] This is further supported by a study that demonstrates that removal and acetylation of the H4 tail leads to a significant decrease in the inter-nucleosome interaction energy. [50]

The histone tails also serve as hosts to epigenetic processes. These tails contain specific

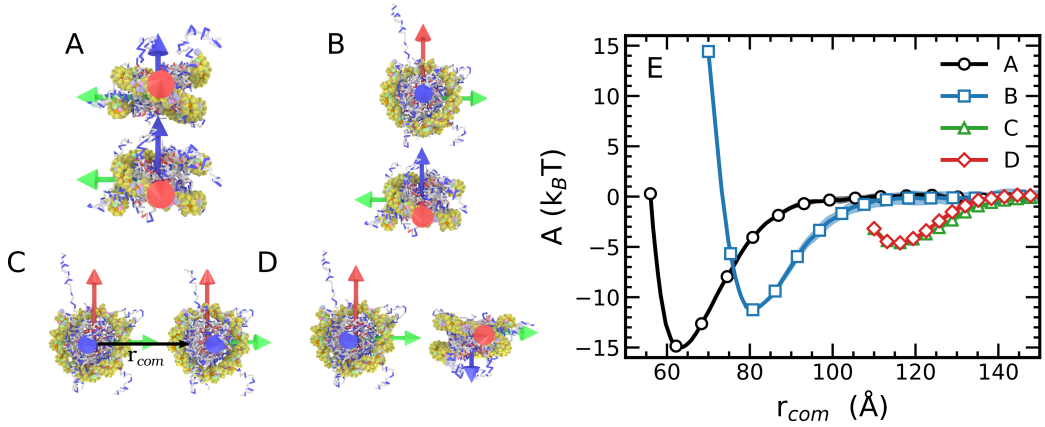

Figure 2.1: Nucleosome pair potential system orientations and resulting energetic landscape. Orientations of the nucleosomes considered here are shown on the left. The axes represent each nucleosome's coordinate system. The red axis is the nucleosome dyad vector, the blue is orthogonal to the face of the nucleosome, and the green is orthogonal to the other two. A) The "stacked" nucleosome orientation. B) The "side-side" nucleosome orientation. C) The "rotated-stack" orientation. D) The "rotated-side" orientation. Free energy landscape shown on the right in panel E. The colors correspond to the orientation shown to the left. Error bars shown are approximately the same order as the thickness of the lines.

residues that are subject to post-translational modifications (PTMs), including methylation, acetylation, and ubiquitination, [44, 6] which regulate and maintain nuclear functions such as transcription and DNA repair. Of particular interest to this work are the charged residues (eg. lysine, arginine, histidine) that lose their charge upon acetylation. Electrostatic interactions are inherently long-ranged, and play a significant role in regulating biological functionality. These charged residues can mediate nucleosome-nucleosome and nucleosome-DNA interactions, contributing to fiber condensation. Of these charged residues, lysines, especially those occurring on the H4 and H3 NTD, have been the focus of numerous epigenetic studies for their potential to be acetylated or methylated. [6, 64, 81, 113, 139, 154] It has also been proposed that acetylation of the tails reduces their flexibility, and therefore diminishes their ability to reach other nucleosomes. [31, 162] Note that irregular methylation or acetylation of lysine residues, like H3K4 and H3K27, has also been linked to carcinogenesis.[30, 117]

The innate connection between PTMs and inter-nucleosome energetics implies that epigenetic states can be linked to the structure of the chromatin fiber. Thus, an understanding of the energetics at play in the condensation of the chromatin fiber is important for studies of epigenetic states. Despite this connection, concrete links between epigenetic state and chromatin structure have yet to be drawn. The chromatin fiber is dynamic and sensitive to environmental conditions, making it difficult to disentangle the various energetic contributions to structure by relying solely on experimental deformation studies. Theoretical and computational studies could help elucidate a number of molecular-level processes that, until now, have been hidden in the experimental data.

It is important to emphasize that chromatin modeling efforts have relied extensively on available chromatin fiber experiments. [53, 58, 71, 82, 144] In particular, the Three Site Per Nucleotide (3SPN) combined with the Atomic-Interaction-based Coarse-Grained (AICG) model has been useful in studies of the nucleosome.[58, 82] The model has shown good agreement with experimental results on nucleosome energetics and dynamics, including competitive reconstitution experiments that study binding strength of DNA sequence to the histone octomer[47], force-induced nucleosome-DNA unraveling[77], and nucleosome repositioning mechanism analyses.[78]

Building on that work, in what follows we use the 3SPN and AICG nucleosome models to examine the interaction energy landscape between unmodified and modified nucleosomes, and we study the effects of several modifications on chromatin structure. In doing so, we aim to identify some of the key inter-nucleosome interactions that are relevant to chromatin condensation. Here we note that similar work at the atomistic scale has allowed researchers to extract key structural aspects of the tails when the nucleosomes are stacked. [31, 132] Building on that work, here we quantify the anisotropic inter-nucleosome free energy landscape, and provide new insights into previously reported nucleosome interaction energies by considering the roles of varying orientations, salt concentrations, and counterion condensa-

tion. We also consider the effects of lysine acetylation on the H4 tails, which lead to energetic decreases that are consistent with experimental findings. [73] Lastly, we examine how H4 lysine acetylation induces chirality of the nucleosome interaction energy landscape, away from a left-handed superhelical motif. [138, 158]

2.3 Methods

2.3.1 Nucleosome Core Particle Model

Our work is carried out with the 3SPN.2C[47] coarse-grained DNA model, parameterized using X3DNA.[86] The 3SPN.2C model represents a nucleotide with three sites: one for the sugar, one for the phosphate, and one for the base. We make use of the most recent version of the model, where there is no attractive Lennard-Jones potential added between the DNA and histones.[78] As 3SPN.2C is a sequence-dependent model, we use the strongly-binding 601 DNA sequence in view of its extensive use in experiments. [85] In future efforts, we will consider the effect of DNA sequence on the results reported here. The histone octomer is represented using the AICG protein model applied to the 1KX5 nucleosome crystal structure, generated using CafeMol.[34, 69, 82, 126] Electrostatics are treated at the level of Debye-Hückel theory. A temperature of 300K and a salt concentration of 150mM are used in all calculations unless otherwise noted, resulting in a debye length of $\lambda_d = 7.84$Å. The simulation timestep in all calculations is 20fs. Post-translational acetylations are incorporated into our model by setting the charges of those amino acids to zero. Note that evaluation of the results for different methods of acetylation are discussed in Figure 4.5.

In order to quantify the pair potential between nucleosomes, we make use of a second coarse-grained NCP realization. The second NCP is a copy of the first, and we move and rotate it into its designated orientation and location. The system is then restrained at those relative orientations, varying only the center of mass separation for our calculations. This

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

Table 2.1: Definition of Nucleosome-Nucleosome Orientations for Pair-Potential Calculations

approach is justified given the symmetry of the nucleosome core particle. Before gathering statistics, the nucleosomes are equilibrated for 20ns at their respective orientations.

2.3.2 Nucleosome Orientation and Restraint

We define 6 distinct groups of histone residues that serve to restrain the two nucleosomes at their designated orientation. These 6 groups are located at the nucleosome dyad, the nucleosome center of mass, and an edge orthogonal to the dyad axis of the nucleosome. We provide a detailed description of the specific protein sites that comprise these groupings in Table 2.4. For any calculation, a center of mass separation vector of any two groupings serves to define the orientational vectors, $(\hat{\boldsymbol{f}}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}})$, which we use to define the orientation of each NCP. For our system, $\hat{\boldsymbol{f}}$ corresponds to the vector orthogonal to the face of the nucleosome, $\hat{\boldsymbol{u}}$ corresponds to the vector through the dyad, and $\hat{\boldsymbol{v}}$ corresponds to the vector orthogonal to both $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{f}}$. Any given free energy calculation makes use of five orientational restraints between the two nucleosomes. These restraints are applied by attaching harmonic springs to a specific value of the angle between subsequent vectors in the nucleosomes. The vector combinations and values that correspond to each orientation are listed in Table 3.1.

The orientations highlighted in the snapshots in Fig. 2.1 are defined by the center of mass separation distance, r_{com}, and the orientations of the nucleosome reference unit vectors, $\hat{\boldsymbol{f}}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}$. The "stacked" interaction is reminiscent of nucleosome stacking in the 30-nm fiber proposed by Finch and Klug [45] and maximizes inter-nucleosome tail interactions. Every other orientation favors unique histone tail interactions (eg. rotated interaction highlights

the interactions of the H2A or H2B). Additionally, the nucleosome pair orientations were held away from the dyad so as to avoid DNA unwrapping events that may alter the calculations. To keep DNA from unwrapping, a small spring force was included between the ends of the DNA and the dyad. We note that this spring diminishes the effect of intra-nucleosomal positioning on these calculations, which is a parameter that will be considered in future calculations using a more coarse-grained representation of DNA.

2.3.3 Free Energy Calculations

For free energy calculations, we use umbrella sampling with the weighted histogram analysis method (WHAM). [68] Convergence was determined by calculating the free energy of the system from a subset of the time series. When each subset overlapped with the overall curve, the simulation was deemed converged. The error bars on each curve originate from an average over three independent umbrella sampling calculations.

The primary order parameter for the simulations was the center of mass distance, r_{com}, ranging from 50Å to 150Å, which was divided into 20 umbrella sampling windows. The 2D surface was generated from a 2D umbrella calculation that varied r_{com} and ϕ, the offset angle from restraint orientation A, where

$$\phi = \arccos(\hat{v}_i \cdot \hat{r}_{ij}) - 90 \tag{2.1}$$

This was calculated over 112 umbrella sampling windows, 14 for distances by 8 for angles. The rotational free energy simulation held $\boldsymbol{r_{com}}$ fixed at the calculated global nucleosome minimum of 63.3Å with the same restraints at orientation A, except θ, defined by Eq. 2.2, which was varied from $-180°$ to $180°$.

$$\theta = \pm \arccos\left(\frac{\hat{u}_i \cdot \hat{u}_j}{|\hat{u}_i||\hat{u}_j|}\right) \tag{2.2}$$

In order to ensure unique states for $\pm\theta$, the negative sign criteria was determined by $(\hat{\boldsymbol{u}}_{\boldsymbol{i}} \times \hat{\boldsymbol{u}}_{\boldsymbol{j}}) \cdot \boldsymbol{r}_{\boldsymbol{com}} \leq 0$. For the case of > 0, the positive sign was chosen.

The contact probabilities in Fig. 2.2 were calculated through analysis of the resulting umbrella trajectories. For each trajectory, the center of mass distance was calculated and binned over the collective variable distance from 50Å to 150Å. In the event that the furthest any charged histone tail residue was within one Debye length of the opposite nucleosome, that interaction was recorded. The probabilities were evaluated over a range of at least three umbrella trajectories, each of 2μs. Contacts were recorded every 50000 time steps to ensure that they corresponded to uncorrelated configurations.

2.4 Results

A schematic representation of the orientations used in this work is shown in Fig. 2.1. With the orientations chosen, the systems are subsequently restrained so that only r_{com} varies. This allows us to not only evaluate the energy of attraction between two nucleosomes at unique orientations, but also the inherent range of the interactions as well.

The strongest inter-nucleosome interaction occurs at the stacked configuration, with a potential minimum of $15.0 k_B T$ at a separation of 63.3Å (Fig. 2.1E). This result agrees with previous experimental and computational work that cites accessibility of positively charged sites on the histone tails as significant contributors to inter-nucleosome interactions.[64, 139] Here, we notice a decay to zero after ~ 2.5 Debye lengths ($\lambda_d = 7.84$) from the minimum at a separation of 83Å. The other notable minima show that the rotated and stacked orientation exhibits a well of $11.2 k_B T$ at a slightly larger separation of 80.8Å, and the side by side orientations both show a much reduced interaction minimum of $4.5 k_B T$ at a separation of 116.0Å. The rotated stacked has more histone tail contacts than either the rotated-side or the side-side orientations. We note that the energy scales obtained from this first-order calculation are in quantitative agreement with the $13.4 k_B T$ reported by Kruithof *et al* .[72]

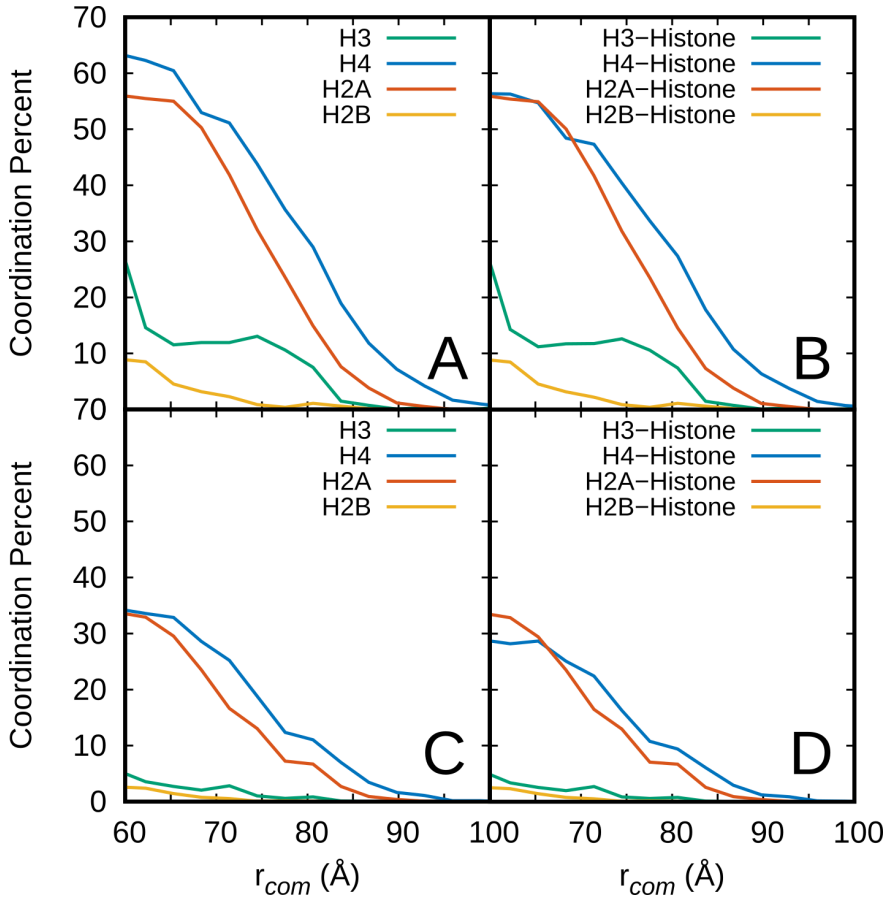

Figure 2.2: Coordination analysis of the tails with regards to the "stacked" orientation. Results were calculated based on sites within one debye length for each snapshot, λ_d. A) The percentage of interactions with the other nucleosome, including both DNA and histone contacts. B) The fraction of contact sites that were histone contacts as compared to DNA. C) Schematic of histone tails in the dinucleosome system from the front to highlight positioning of the H3 and H4 tails. D) Histone tail snapshot from the back to highlight positioning of the H2A and H2B tails. The color of the tails correspond to the graphs and the histone core is removed for ease of viewing.

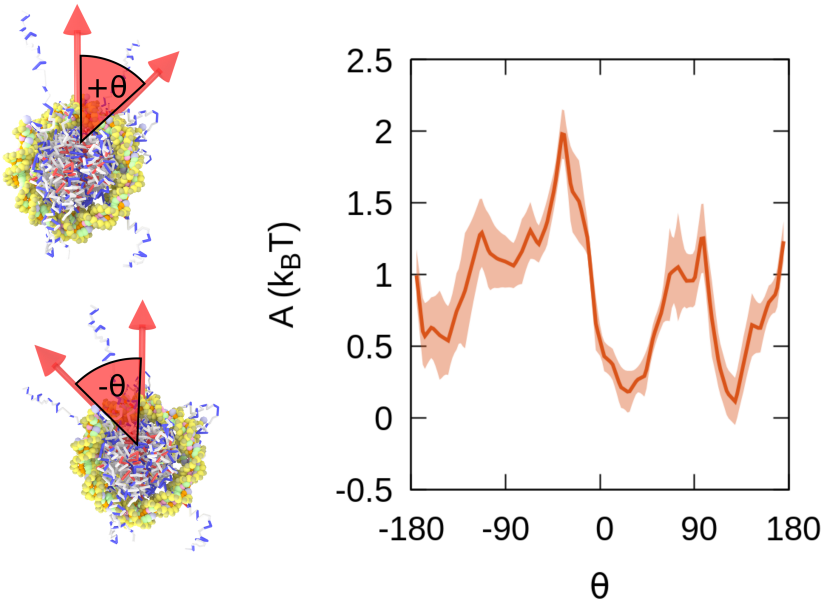

Figure 2.3: Reduction of degrees of freedom through stacked rotation of the nucleosomes. On the left is a schematic representation of the definitions of $\pm\theta$ used from simulation. On the right panel is the free energy of rotation of two stacked nucleosomes. At most the interaction is a difference of $\sim 2k_BT$. We show through this graph that separation distance and orientations are a much more dominant determinant of nucleosome interactions than relative rotation.

2.4.1 Degree of Freedom Reduction

We extend our nucleosome interaction free energy definition by also evaluating the effect of nucleosomal rotation. The results of Funke *et al* demonstrate that rotation of the nucleosomes results in little change to the pair potential. In this vein, we expect that a rotation of one nucleosome relative to the other (while keeping positional orientation unchanged) should not alter the number of histone contacts, and therefore the energetics of the system. Through this order parameter, we strengthen our results by demonstrating that rotation of the nucleosomes at their energetic minima does not significantly influence the inter-nucleosome interactions.

To accomplish this, we generate a free energy surface for rotation at the global simulation minima (stacked orientation, 63.3Å). In this orientation the top nucleosome is rotated 360°

as shown in Fig 2.3. The curve (Fig. 2.3) shows that this rotation does result in a very low energetic change. We notice that the largest change is $\sim 2k_BT$, which is a minimal change relative to the minimum of $15.0k_BT$. This demonstrates that a rotational change is not a key determinant of the interaction landscape, which motivates us to analyze the histone tails further.

2.4.2 Histone Tail Contributions

Of particular importance for this dinucleosome system is the ability to connect physical changes in the interaction landscape to modifications to the nucleosomes. The most relevant of these modifications are post-translational modifications. Through chemical modifications to the histone tails, chromatin can be regulated to become more accessible, or even further condensed. With this system, we link PTMs, namely histone H4 acetylation, to free energy landscape modifications.

We first break down the contribution of each histone tail on the free energy landscape to determine the relative importance of each tail. The role played by the histone tails in mediating inter-nucleosome energetics is analyzed here through a set of contact probability curves for each tail on the opposing nucleosome. (Fig. 2.2). These curves are calculated by assessing the probability that a residue is in contact with an opposite nucleosome. To expand upon these results, the calculations are separated into two categories: histone tail - protein inter-nucleosome contacts and total contact probability. (Fig. 2.2) Both are provided to demonstrate that a greater fraction of total inter-nucleosome interactions come from histone-histone interactions.

Looking at the breakdown of the most common tail interactions, we find that the H4 and H2A tails have the greatest probability of reaching the opposite nucleosome. We note that the long and flexible H3 tail provides significantly less contacts than either the H4 or H2A tails. This observation is consistent with recent evidence that the H3 tail mediates

linker DNA and intra-nucleosome interactions, rather than inter-nucleosome interactions in the absence of divalent salt. [81, 164]

A surprising feature of this analysis is the number of contacts of the H2A tail relative to the H4. To understand this result, we consider the structure of the nucleosome and the histone tails. Despite its strong coordination to the opposite nucleosome, the H2A tail contains the smallest number of positive residues. In order of lowest to highest in number of positively-charged residues, the histone tails are: H2A (5) < H4 (8) = H2B (8) < H3 (10).[87] The contributions of these tails to the free energy are a result of the number of positive residues and the accessibility of these tails to the opposing nucleosome. The positioning of the H2A and H4 tails on the nucleosome face make them highly accessible to the opposing nucleosome, as can be seen in Fig. 2.2C and D. These results suggest that accessibility of the histone tails and, to a lesser extent, the number of positively-charged residues influence the coordination probability of the tails. Here we conclude that the H4 tail contributes the most to the free energy results in Fig. 2.1, followed by the H2A tail.

While coordination and the number of positive residues is a qualitative argument for the free energy, we still lack a quantitative understanding of these tails on the free energy surface. The free energy provided by these tails is an important metric for understanding the physical basis of biological processes like transcription. From our results of tail contacts, we show that the H4 tail has the highest amount of contacts and residues. Literature suggests that post-translational modifications contribute greatly to chromatin dynamics, including fiber condensation. In particular, the acetylation of the H4 tail is highly associated with regions of transcriptionally active chromatin. Through this modification, positively-charged lysine amino acids become neutral acetyl-lysine. This modification suggests that transcriptional regulation can be linked to inter-nucleosome energetic changes. As a result, we incorporate such modifications of the nucleosome into the workflow and evaluate their affect on the potential landscape.

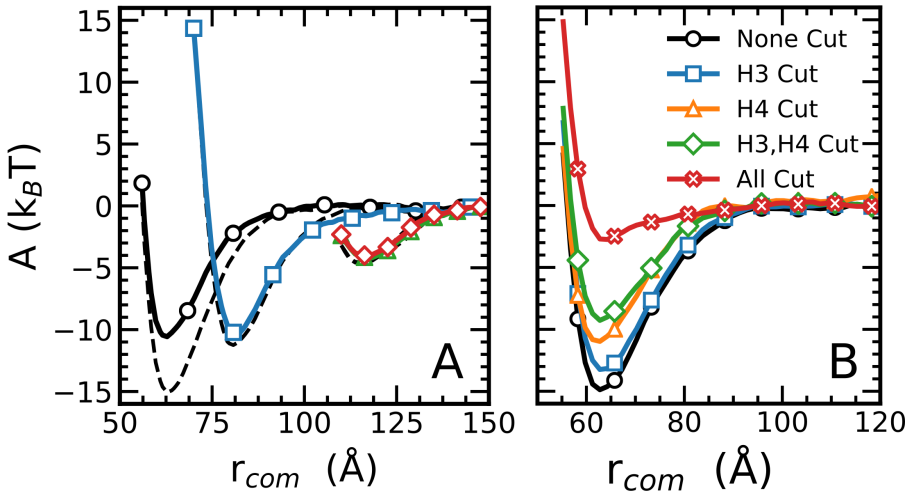

Figure 2.4: Free energy contributions of the histone tails. A) the full landscape subject to H4 histone acetylation. The unmodified landscape is shown for comparison as dashed lines. The comparison shows that acetylation predominantly affects the stacked pair potential. B) The effect of removing the H3, H4, both H3 and H4, and all the histone tails. The energy decreases with each tail removal. The H4 tail provides a larger energy of $3.8k_BT$ contribution than the H3 tail of $1.5k_BT$. Removal of all tails decreased the energy to $2.72k_BT$.

To evaluate this effect, we calculate the free energy surface for nucleosomes with acetylated tails. (Fig 2.4) It can be seem from the modified interaction landscape that the "stacked" configuration changes from $15.0k_BT$ to $10.4k_BT$. Additionally, minor reductions are calculated in the minimum of the other interactions. The "rotated-stack" in this case reduces to $10.2k_BT$, the "side-side" reduces to $4.21k_BT$, and the "rotated-side" reduces to $3.99k_BT$. This result suggests that the acetylated H4 tail predominantly affects the "stacked" orientation.

Building on this finding, we choose only the stacked configuration as the focal point for studying the effects of further modifications. We decide to highlight the energetics of the histone tails through both the removal of a small contact tail (H3) and a large contact tail (H4) and assess the resultant free energy landscape. (Fig. 2.4) By removing the H3 and H4 tails and all of the tails, we notice that the H4 tail does indeed provide a larger energetic impact on the pair potential than the H3 tail. We also note that removal and acetylation of

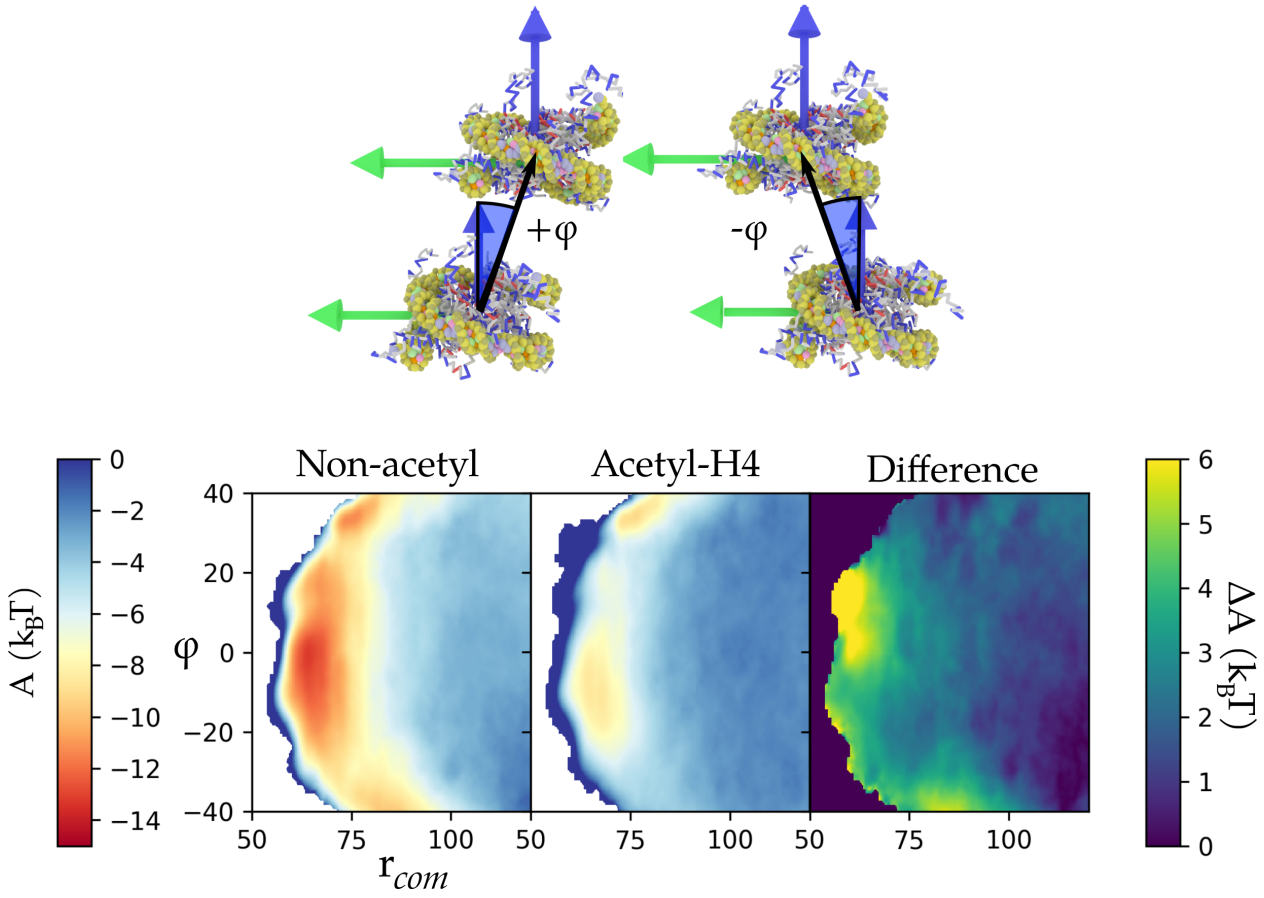

Figure 2.5: Analyzing the effect of H4 acetylation on the stacked nucleosome interaction. On top are snapshots of the different configurations for the calculations. A positive value of ϕ corresponds to a right-handed superhelical structure, and a negative ϕ corresponds to a left-handed motif. Below is the 2D internucleosome surface free energy difference calculated using two dimensional umbrella sampling with nucleosome separation distance, r, and angle, ϕ. The free energy of the standard surface with no modifications is in the left-most panel. In the middle is the free energy of the acetylated surface with reference color bar to the left for both the left and middle panels. The free energy difference between the two surfaces is shown as the right most panel with reference color bar shown to the right.

the H4 tail shows no difference in free energy. (See Supporting Information) Additionally, even when the H3 and H4 tails are removed, a significant energy well of $8k_BT$ persists. We expect a large portion of this to correspond to H2A interactions, and to a lesser extent, H2B. This is consistent with previous findings that the H2A tail provides a non-negligible interaction to the pair potential, as well as to a lesser extent the H2B. As expected, the pair potential interaction drops significantly upon removal of all tails, further proving that the energetic contribution is predominantly in the flexible histone tails. The effect of acetylations are further examined through multiple charge-removal analyses.

To understand modifications further, we analyze the potential chirality induced in the "stacked" interactions. We determined that the stacked interactions are predominantly modified, but are unsure if this is a symmetric change across the face of the NCP. Prior analysis of NCP crystal structures has shown that the chromatin fiber exhibits a preferred left-handed superhelical structure. [138, 158] We expect that decondensed fibers must have some inherent energetics that preserve this structure. To examine the potential of transcription further, we assess the H4 modified landscape through multiple continuous orientations of the two nucleosomes (Fig. 2.5). A 2D surface is constructed to highlight this area of largest attraction under both acetyl-H4 and unmodified nucleosome interactions. A comparison of the two as well as the difference between the two surfaces is shown in Fig. 2.5.

It can be seen in Fig. 2.5 that there are small lobes above and below the center, corresponding to the "stacked" orientation seen in Fig. 2.1. Upon acetylation of the H4 tail in Fig. 2.5, it can also be noticed that the bottom lobe disappears relative to the minimum, which is highlighted by the difference spike in the same area in Fig. 2.5. This suggests that the bottom lobe corresponds to H4 tail contacts providing a significant free energy reduction to the surface. Specifically, this region, what we are referring to as the "H4-contact lobe", highlights the energetics that must arise at the nucleosome for local transcription of the fiber to occur. Consistent with theory, these interactions support a left-handed superhelical struc-

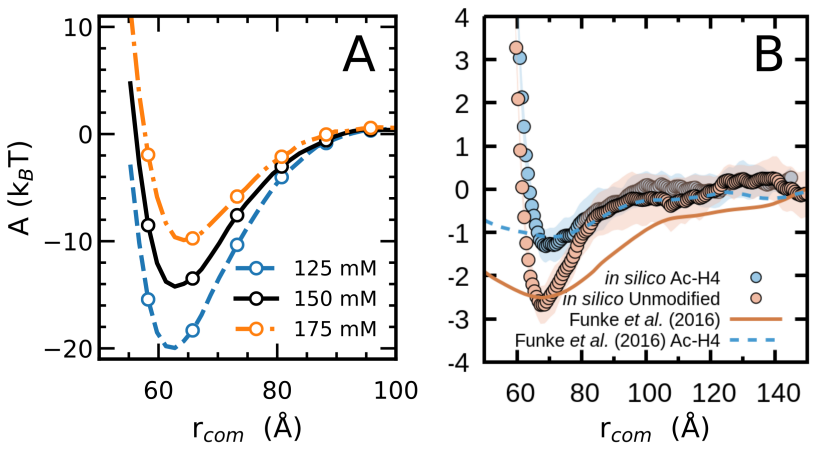

Figure 2.6: The effects of ionic condition on nucleosome interaction. A) The energetic changes to the stacked nucleosome energy with varying salt concentrations. B) A comparison of our results to experiment. This comparison shows good quantitative agreement for both the depth and location of the minima of the normal and H4-acetylated simulations.

ture and are disrupted upon acetylation of the H4 tail, demonstrating an induced repulsive chirality.[154]

2.4.3 Ionic Environment on Dinucleosome Interactions

With a better understanding of the direct inter-nucleosome energetics, we now turn our attention to environmental effects. The cell heavily regulates ionic conditions, as disruptions or stresses can result in cell death.[63] We investigate how the structure of chromatin can be altered in the event of deviations in ionic environment. For all prior calculations we used a salt concentration of $150mM$, representing physiological salt strength. Experimental work has pointed to the salt concentration playing a significant role in changing the chromatin fiber structure.[130, 55] These effects propagate from modifying local to global chromatin structure. We examine the effect of local and long-range solution effects with two approaches: changing both the environmental salt concentration and including localized ionic coordination. Keeping close to physiological concentrations, the resultant nucleosome-nucleosome interaction strength is evaluated at $150 \pm 25mM$. We show the effect of salt concentrations

changes in Fig. 2.4. Of importance, we see a minimal shift of $\sim \frac{1\mathring{A}}{25mM}$ in terms of separation distance, showing that monovalent salt slightly affects the range of these interactions, but not significantly. However, we observe large changes in the interaction depth. The minima depth for a 25 mM decrease in monovalent salt shows a $5.73k_BT$ increase in strength, while a 25 mM increase in salt concentration results in a decrease of depth by $4.25k_BT$ These results suggest that the NCP physics are highly sensitive to changes in monovalent salt concentration and implicate altered chromatin structure in cell death.

We have thus far neglected the effects of counterion condensation on the pair potential landscape. The charge distribution from exposed DNA on opposing nucleosomes satisfies the condition required by Manning counterion condensation theory. [58, 95] From Hinckley *et al* it is suggested that 3SPN.2 carry a counterion effective charge of 0.6. [58] As such, the landscape is re-evaluated with the inclusion of counterion condensation as a means to incorporate local ion effects into the system. Counterion condensation in this model consists of a reduction in potential between inter-nucleosome histone-DNA contacts shown in the SI. The results can be observed in Fig. 2.6.

It can be seen from Fig. 2.6B that the energy minima shifts significantly with counterion condensation from 63.3 to 68.8Å and includes drops of the interaction potential from 15.0 to $2.69k_BT$. This significant reduction is in quantitative agreement with the results from Funke *et al* and Cui *et al* .[33, 50] The incorporation of post-translational modifications into the landscape shows that the deepest minimum reduces even further to $1.70k_BT$ showing good agreement with experimental results. We show in Fig. 2.6B that both the depths of the wells and the location of the minimum from experiment agree with those of the counterion condensation calculations. While we note the quantitative agreement with experiment for the location and depth of the wells, we find these results to be of shorter range than in experiments. We note that Manning-condensation and Debye-Hückel electrostatics are a only a first-order approximation of the environment, more rigorous electrostatic treatments

will be considered in the future. Additionally, we acknowledge that divalent salts are present in the cell nucleus. Such salts are likely to influence the interactions examined in this work, and will also be investigated in a future study.

2.5 Conclusion

In this work we have examined the primary factors that govern the strength and shape of the interaction landscape between two nucleosomes. The underlying pair potentials are highly anisotropic, but their strength is well correlated with histone tail contacts. Predominantly, we show that the H4 and H2A NTD provide more tail contacts and a greater contribution to inter-nucleosome interactions. We also demonstrate that acetylation of the H4 tail, an epigenetic mark associated with active genes, is directly related to a free energy change in chromatin structure, which has been theorized for decades.[6, 23] We have also uncovered an induced chirality in the strongest interaction configuration upon acetylation of the H4 histone tail, suggesting that acetylation of the H4 tail disrupts the left-handed superhelical organization of the chromatin fiber. Upon consideration of local and global ion effects, one arrives at a free energy landscape that is in good agreement with available experimental reports. The results further predict a high sensitivity of the chromatin fiber structure to the ionic environment in the cell. The results reported here agree quantitatively with experiment. Taken together, the inter-nucleosome interactions studied in this work paint a clearer picture of the energies associated with the chromatin fiber, and pave the way for studies of higher length scale chromatin towards an energetic analysis of the 30-nm fiber.

2.6 Supporting Information

2.6.1 3SPN.2C Model

The 3SPN.2C model is a coarse-grained model of DNA that represents each nucleotide as three beads: one for the phosphate, one for the base, and one for the sugar. We restate the model parameters to make this work self-contained. We use a combination of bond, angle, and dihedral forces to preserve the B-DNA structure. [58] These bonded forces are

$$\begin{aligned} U_b =& U_{bond} + U_{ang} + U_{dihe} \\ =& \sum_{i}^{bonds} k_b(r_i - r_{eq,i})^2 + 100k_b(r_i - r_{eq,i})^4 \\ &+ \sum_{i}^{angles} k_\theta(\theta_i - \theta_{eq,i})^2 \\ &+ \sum_{i}^{dihedrals} -k_\phi \exp\left(\frac{-(\phi_i - \phi_{eq,i})^2}{2\sigma_{\phi,i}^2}\right), \end{aligned} \tag{2.3}$$

where k_b and $r_{eq,i}$ are the force constant and the equilibrium bond length for the i^{th} bond, k_θ and $\theta_{eq,i}$ are the force constant and equilibrium angle for the i^{th} angle, and k_ϕ, $\phi_{eq,i}$, and $\sigma_{\phi,i}$ are the well depth, equilibrium angle, and well-width for the i^{th} dihedral. In this model, dihedrals are only applied to the backbone of the system, which represents the sugars and the phosphates.

There is also an added dihedral function to the DNA backbone in the form of

$$U_{\phi,periodic} = k_{\phi,periodic}[1 + \cos(\theta - \theta_0)] \tag{2.4}$$

The non-bonded potentials can be broken down into electrostatic interactions, excluded volume interactions, and base-pairing interactions. The full non-bonded set of contributions

is:

$$U_{nonbond} = U_{exclude} + U_{elec} + U_{bp} + U_{cs} + U_{bs} \tag{2.5}$$

The excluded volume interaction, $U_{exclude}$, follows a purely repulsive Lennard-Jones interaction of the form

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

In this case, the cutoff, $r_{Cut} = \sigma_{ij}$ is the arithmetic average of the size of the interacting particles. The potential assumes interactions between all sites that are not bonded together or interacting by a base-pair non-bonded potential. The electrostatic interactions occur between the charged phosphates and residues on the proteins. We use the implicit, Debye-Hückel screened electrostatic potential of the form:

$$U_{elec} = \sum_{i<j}^{n_pairs} \frac{q_i q_j e^{-r_{ij}/\lambda_D}}{4\pi\epsilon_0 \epsilon(T,C) r_{ij}}, \tag{2.6}$$

where q_i and q_j are the charges on the i^{th} and j^{th} sites, λ_D is the Debye length, and $\epsilon(T,C)$ is the dielectric permittivity of the solution.

The base-pairing interactions can be broken into three interactions: base-pair (U_{bp}), base-stacking ($U_{base-stack}$), and cross-stacking interactions ($U_{cross-stack}$). All three rely on a Morse potential of the form

$$U_{Morse}(\epsilon_{ij}, \alpha_{ij}, r_{ij}) = \epsilon_{ij}(1 - e^{(-\alpha_{ij}(r_{ij} - r_{eq,ij}))})^2 - \epsilon_{ij} \tag{2.7}$$

which is broken down into a repulsive component,

$$U^r_{Morse}(\epsilon_{ij}, \alpha_{ij}, r_{ij}) = \begin{cases} \epsilon_{ij}(1 - e^{(-\alpha_{ij}(r_{ij} - r_{eq,ij}))})^2 & r < r_{eq,ij} \\ 0 & r \geq r_{eq,ij} \end{cases} \tag{2.8}$$

and an attractive component,

$$U^a_{Morse}(\epsilon_{ij}, \alpha_{ij}, r_{ij}) = \begin{cases} -\epsilon_{ij} & r < r_{eq,ij} \\ \epsilon_{ij}(1 - e^{(-\alpha_{ij}(r_{ij} - r_{eq,ij}))})^2 - \epsilon_{ij} & r \geq r_{eq,ij} \end{cases} \tag{2.9}$$

.

For these functions, ϵ_{ij} is the well depth of the interaction, $r_{eq,ij}$ is the equilibrium distance of the interaction, and α_{ij} is a parameter that controls the range of the attraction. We also incorporate a modulating function f to the angles of interaction in the cross-stacking, base-stacking, and base-pairing interaction to smoothly scale the interactions between non-hydrogen bonded and hydrogen bonded base pairs. The modulating function is of the form

$$f(K, \Delta\theta) = \begin{cases} 1 & \frac{-\pi}{2K} < \Delta\theta < \frac{\pi}{2K} \\ 1 - \cos^2(K\Delta\theta) & \frac{-\pi}{K} < \Delta\theta < \frac{-\pi}{2K} \text{ or } \frac{\pi}{2K} < \Delta\theta < \frac{\pi}{K} \\ 0 & \Delta\theta < \frac{-\pi}{K} \text{ or } \Delta\theta > \frac{\pi}{K} \end{cases} \tag{2.10}$$

where K is a modulating constant depending on the type of interaction. With these definitions in place, we can fully describe the force field of interaction of a base pair, which

is

$$U_{bp} = \sum^{n_{BP}} \begin{cases} U^{rep}_{Morse}(\epsilon_{ij}, \alpha_{BP}, r_{ij}) & \\ \quad +\frac{1}{2}(1+\cos(\Delta\phi_1))f(K_{BP}, \Delta\theta_{1_{ij}})f(K_{BP}, \Delta\theta_{2_{ij}})U^{a}_{Morse}(\epsilon_{ij}, \alpha_{BP}, r_{ij}) & r_{ij} < r_{eq,ij} \\ \frac{1}{2}(1+\cos(\Delta\phi_1))f(K_{BP}, \Delta\theta_{1_{ij}})f(K_{BP}, \Delta\theta_{2_{ij}})U^{a}_{Morse}(\epsilon_{ij}, \alpha_{BP}, r_{ij}) & r_{ij} \geq r_{eq,ij} \end{cases} \tag{2.11}$$

.

In this case, θ_1 is defined as the angle between the sense-strand sugar, base and the anti-sense base, while θ_2 is the same angle but on the opposite strand of DNA, ϕ_1 is the dihedral between the sugar and base on the sense and anti-sense strands of DNA. When two base pairs are hydrogen bonded, we also consider the pi-stacking interactions in the form of base-stacking and cross-stacking interactions.

The base-stacking interaction is

$$U_{bs} = \sum^{n_{BS}} \begin{cases} U^{rep}_{Morse}(\epsilon ij, \alpha_{BS}, r_{ij}) + f(K_{BS}, \Delta\theta_{BS,ij})U^{a}_{Morse}(\epsilon_{ij}, \alpha_{BS}, r_{ij}) & r_{ij} < r_{eq,ij} \\ f(K_{BS}, \Delta\theta_{BS,ij})U^{a}_{Morse}(\epsilon_{ij}, \alpha_{BS}, r_{ij}) & r_{ij} \geq r_{eq,ij} \end{cases} \tag{2.12}$$

,

where the angle, θ_{BS}, is defined by the angle between the vector connecting the 5' direction sugar and base and the vector connecting the two bases in the 3' direction.

The cross-stacking interaction is

$$U_{cs} = \sum^{n_{CS}} f(K_{BP}, \Delta\theta_{3,ij})f(K_{CS}, \Delta\theta_{CS,ij})U^{a}_{Morse}(\epsilon_{ij}, \alpha_{CS}, r_{ij}) \tag{2.13}$$

where θ_3 is the angle between the vectors connecting the sugars to the bases in a W-C base pair and θ_{CS} is the vector connecting the sugar to the base, and the vector connecting the base on the anti-sense strand in the 5' to the base in the 3' direction on the sense strand.

Table 2.2: 3SPN.2C Parameters

Parameter	Value
k_b	0.6 kJ/mol/$\mathring{A}^2$
k_θ	200 kJ/mol/rad^2
k_ϕ	6.0 kJ/mol
ϵ_r	1.0 kJ/mol
K_{BS}	6.0
α_{BS}	3.0
K_{CS}	8.0
α_{CS}	4.0
K_{BP}	12.0
α_{BP}	2.0
σ_{AT}	5.82$\mathring{A}$
σ_{GC}	5.52$\mathring{A}$
ϵ_{AT}	16.37 kJ/mol
ϵ_{GC}	20.73 kJ/mol

For completeness, we use the following parameters for the DNA model.

Table 2.3: Base-stacking and cross-stacking energies for 3SPN.2C from Freeman et al [47]. Section (a) describes base-stacking energy scales. Sections (b) and (c) describe cross-stacking energy scales. Upward-pointing arrows denote the sense strand while downward-pointing arrows denote the anti-sense strand (for cross-stacking interactions).

(a)

	Base $^{3'}\uparrow$ ϵ_{ij} (kJ/mol)			
Base $_{5'}\uparrow$	A	T	G	C
A	13.82	15.05	13.32	15.82
T	9.15	12.44	9.58	13.11
G	13.76	14.59	14.77	15.17
C	9.25	12.42	8.83	14.01

(b)

	Base $\downarrow^{5'}$ ϵ_{ij} (kJ/mol)			
Base $_{5'}\uparrow$	A	T	G	C
A	1.882	2.388	2.439	1.680
T	2.388	1.882	2.187	2.566
G	2.439	2.187	3.250	0.972
C	1.680	2.566	0.972	4.135

(c)

	Base $\uparrow^{3'}$ ϵ_{ij} (kJ/mol)			
Base $\downarrow_{3'}$	A	T	G	C
A	1.882	2.388	2.566	2.187
T	2.388	1.882	1.680	2.439
G	2.566	1.680	4.135	0.972
C	2.187	2.439	0.972	3.250

2.6.2 AICG Model

The AICG model of the histones employ a Go-like interaction force field. We define the potentials below.[82]

$$\begin{aligned}
U_{AICG} =& U_{bond} + U_{ang} + U_{dihe} + U_{natv} \\
=& \sum_{i}^{bonds} k_b(r_i - r_{eq,i})^2 \\
&+ \sum_{i}^{angles} k_\theta(\theta_i - \theta_{eq,i})^2 \\
&+ \sum_{i}^{dihedrals} k_{\phi,1}[1 - cos(\phi_i - \phi_{eq,i})] + k_{\phi,3}[1 - \cos 3(\phi_i - \phi_{eq,i})] \\
&+ \sum_{i<j-3}^{natcontact} \epsilon_{go} \left[5 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{10} \right] \\
&+ \sum_{i<j-3}^{non-native} \epsilon_{excl} \left(\frac{r_{ij,ex}}{r_{ij}} \right)^{12}, \qquad (2.14)
\end{aligned}$$

The parameters are amino-acid dependent and are calculated using the CafeMol simulation package.[69]

Description of Histone Groups

For a consistent definition of the restraints for the pair potentials, the groups are comprised of residues from the histone core. There are three unique groups of residues and one set of groups for each nucleosome. All sites displayed correspond to both copies of the histone, unless otherwise noted. (Eg. residue 63 references the 63rd residue on both H3 and H3') Here, the (') denotes the second instance of that specific histone in the octamer. These residues hold for all simulations, including tail removed/modified runs. The groups are displayed in the following table:

Table 2.4: Definition of Residue Groups for Orientation Definitions

Group Description	H3 Residues	H4 Residues	H2A Residues	H2B Residues
Nucleosome COM	63-120	30-82	26-87	33-97
Nucleosome Dyad	81-131	-	-	-
Nucleosome Orthogonal	63-120 (H3)	30-82 (H4)	26-87 (H2A')	33-97 (H2B')

Table 2.5: Definition of Nucleosome Restraint Vectors From Residue Groups

Vector	Group 1	Group 2
$\hat{\boldsymbol{u}}_i$	Nucl_i COM	Nucl_i Dyad
$\hat{\boldsymbol{v}}_i$	Nucl_i COM	Nucl_i Ortho
$\hat{\boldsymbol{u}}_j$	Nucl_j COM	Nucl_j Dyad
$\hat{\boldsymbol{v}}_j$	Nucl_j COM	Nucl_j Ortho
$\hat{\boldsymbol{r}}_{ij}$	Nucl_i COM	Nucl_j COM

From these groups, we define the vectors for the restraints. The groups were chosen to create as much of an orthogonal definition as possible. We show in the following table which groups are involved in each restraint and the values of the angle between those restraints to hold the orientations for sampling.

The restraints of the two nucleosomes are then characterized by harmonic springs placed between angles of the above vectors. For the orientations in this study, the restraint angles were calculated from the initial time step of the nucleosomes at these orientations. The restraints were then put in place on the following angles:

The values of the angles for the four orientations are defined as:

Table 2.6: Definition of Harmonic Restraint Angles

Angle Definition	Vector 1	Vector 2
θ_1	$\hat{\boldsymbol{u}}_i$	$\hat{\boldsymbol{r}}_{ij}$
θ_2	$\hat{\boldsymbol{u}}_j$	$\hat{\boldsymbol{r}}_{ij}$
θ_3	$\hat{\boldsymbol{v}}_i$	$\hat{\boldsymbol{r}}_{ij}$
θ_4	$\hat{\boldsymbol{v}}_j$	$\hat{\boldsymbol{r}}_{ij}$

Orientation	$\theta_{1,eq}$	$\theta_{2,eq}$	$\theta_{3,eq}$	$\theta_{4,eq}$
A	99.0°	99.0°	104.0°	104.0°
B	99.0°	99.0°	14.0°	14.0°
C	98.65°	8.87°	104.48°	93.01°
D	87.55°	89.18°	13.74°	14.35°

Table 2.7: Definition of Nucleosome-Nucleosome Orientations for Pair-Potential Calculations

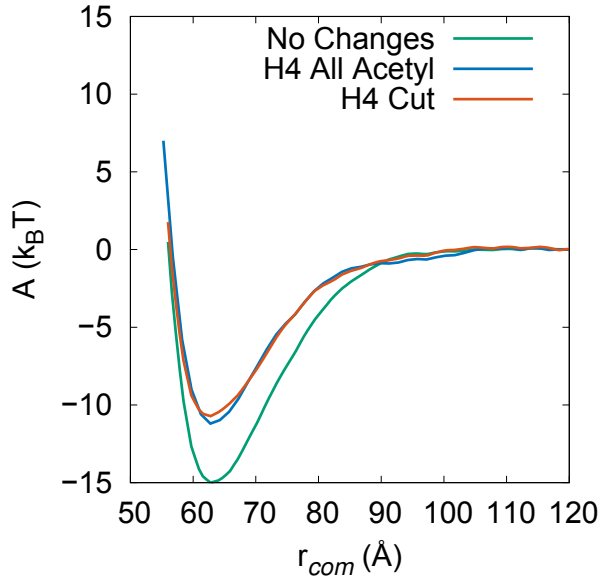

Figure 2.7: The free energy surface for different simulation methods of acetylation.

Acetylation of the H4 Tail

We considered multiple effects for modifications of the H4 Tail. The effects of the H4 tail shown in 4.8 highlight the energetic contributions.

Acetylation of the lysine residues of the H4 tail and removal of the H4 tail show no quantitative or qualitative change on the free energy surface. For this reason, we choose to use the acetylated tail to represent acetylations and cut-off tails in the manuscript. This result is consistent with the experimental observations of Funke et al. [50]

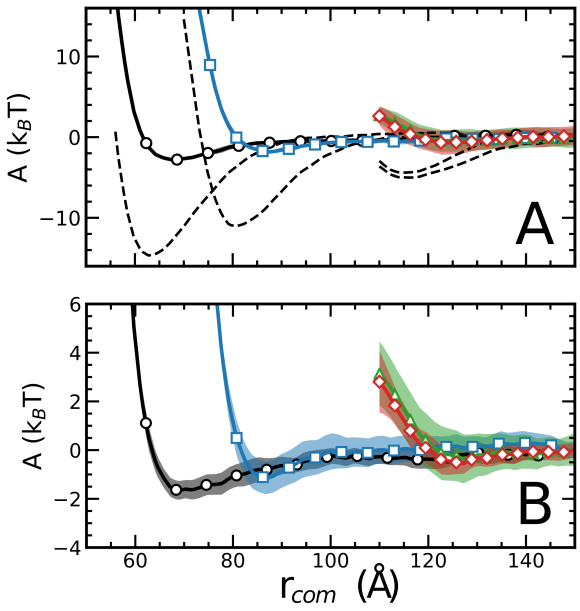

Figure 2.8: Free energy surface of nucleosome interactions with A) Manning condensation electrostatics and B) Manning condensation with an acetylated H4 tail.

2.6.3 Counterion Condensation

We use the Manning condensation electrostatic definition to quantitatively compare our results to experiment. The interactions drop proportionally for all orientations. The "rotated-stack" orientation exhibits a shift to 87.7Å with an attraction of $1.80k_BT$, the "side-side" shifts to 124.7Å with a weak interaction of $0.82k_BT$, and the "rotated-side" shifts to 123.6Å with a weak attraction of $0.28k_BT$.

Applying the same analysis for the Debye-Hückel system here, we look at the histone coordination probability of the tails. As can be seen in Fig. 2.8, the coordination probability decreases for every tail. Comparing to the Debye-Hückel result, the highest probability still corresponds to the H4 tail at 33 percent. We see no relative change of the tails, except now we note that the relative contribution of the H2A tail is approximately equivalent to that of the H4 tail. In Fig. 2.9 D, the fraction of H4 contacts with the opposite nucleosome histone is lower than that of the H2A tail. We attribute this to the DNA versus histone

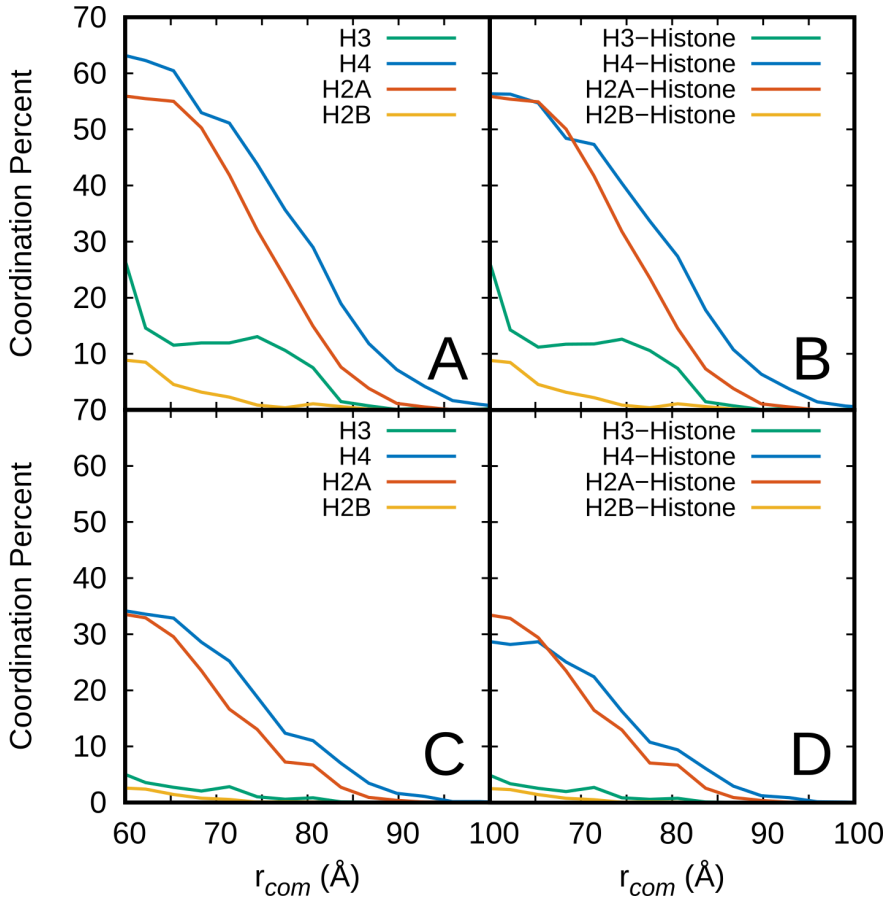

Figure 2.9: A) Tail interactions for all positive residues B) Histone fraction of tail interactions for all positive residues C) Tail interactions with counterion condensation D) Histone fraction of tail interactions with counterion condensation

contributions of the H4 tail. It can be seen that the relative coordination of the H4 is lowered when considering the fraction of nucleosome contacts, even in the Debye-Hückel case, suggesting a larger contribution of DNA contacts. As the interactions between DNA and histones are reduced in the implementation of Manning condensation, the free energy magnitude is reduced in this range.

CHAPTER 3

VALIDATION OF THE NUCLEOSOME-CENTRIC 1CPN COARSE-GRAINED CHROMATIN MODEL

3.1 Abstract

Reconstitution of chromatin *in vitro* reveals that unique structural motifs arise from subtle topological changes. Notably, the fiber structure alters when the amount of linking DNA is shifted and the linker histone is present. These structures and their compaction are quantified through sedimentation analysis of isolated oligonucleosomes. In general, the presence of the H1 or avian variant H5 linker histone leads to condensed fiber motifs and is associated with repressed chromatin. In the past, we introduced the 1-Cylinder-Per-Nucleosome (1CPN) model, designed with a bottom-up approach incorporating previous simulation pair-potential work. Here, we extend the 1CPN model to incorporate the H1 linker histone and validate the model through sedimentation coefficient analysis. We find the 1CPN model to be in excellent agreement with experiments. Additionally, through dinucleosome free-energy analysis, we demonstrate the sensitivity of fiber packing to linker length. Moving forward, we expect the 1CPN model to be a great candidate to be predictive of chromatin structure. The 1CPN model and the extensions developed here are publicly available in the LAMMPS molecular dynamics simulation package.

3.2 Introduction

The first isolated chromatin fiber was imaged through electron microscopy in 1974 by Olins *et al.* [110] Decades of studies of isolated and reconstituted chromatin fibers followed, focusing on the structure of the now-controversial "30-nm fiber." [45] In short, the 30-nm fiber is a regular, ordered fiber of diameter $\sim$ 30 nm, long thought to be the secondary organization

of the genome. There are two such candidate structures of the 30-nm fiber that have been the source of controversy throughout this history: the one-start solenoid and the two-start zigzag. The defining difference between these structures is the positioning of neighboring nucleosomes; in the one-start solenoid, sequential nucleosomes are packed next to each other with bent linker DNA, while in the two-start zigzag, sequential nucleosomes are on the opposite ends of the helix, with straight linker DNA connecting them. However, in recent years the 30-nm fiber is being challenged as an artifact of dilute experimental conditions.[92] Recently, novel ChromEMT experiments reveal that chromatin exists in fibers of radius $5 - 24$nm. [112]

Of course, there are a number of parameters known to change the local structure of the chromatin fiber. Included in these parameters are: DNA-binding proteins, such as the H1 (or H5 avian variant) linker histone, the amount of linker length connecting the nucleosomes, and the solvent composition (eg. divalent salt concentration). To address how these factors affect the structure of chromatin, experimentalists use sedimentation analysis to measure fiber compaction. In summary, the linker histone leads to much more compact fibers, divalent salt facilitates compact fibers, and there is an unknown relation between structure and linker DNA. It is thought that longer linker lengths have a propensity to bend and, therefore, lead to a solenoid structure, and shorter linker lengths lead to a zig-zag. Pioneering work by Routh *et al* demonstrates the extent to which both the linker histone and linker length influence local chromatin fiber structure *in vitro*. [130]

The rich problem of solving the chromatin structure through modeling is an enticing prospect. Currently, multiple modeling strategies at the oligonucleosome scale have been employed. In general, such models coarse-grain DNA as a few bp per bead, and the nucleosome is represented as a single entity. The DNA is treated as a wormlike-chain (WLC) model, [96] and different variants treat the nucleosome and inter-nucleosome interactions in their own way. Pioneering work by Arya *et al* introduced a coarse-grained scheme that repro-

duces the electronic field of the nucleosome with a reduced representation, and it has been successful in predicting chromatin fiber structure. [10, 11] This method, known as the discrete surface charge optimization (DISCO), reduces the nucleosome to ~ 300 pseudo-charges distributed on the nucleosome surface. Building on that work, in recent years, additional modeling approaches at this length scale have emerged. A different approach to nucleosome coarse-graining draws from experimental values of nucleosome-interaction strengths. [107] Additional efforts have sought to develop a nucleosome-resolution model with specified topological interactions to inform the assembly of nucleosomes in the chromatin fiber.[24] Another model at this scale incorporates non-histone proteins that bend DNA, resulting in heterogeneous packing in oligonucleosome fibers. [12] Recently, we have introduced the 1-cylinder-per-nucleosome (1CPN) model, built off of previous nucleosome-centric results and other, well-founded models at this length scale.[76, 155]

In this work, we discuss the development and validation of the 1CPN model. First, we highlight the bottom-up approach of the 1CPN model from previous work of nucleosome simulations. Then, we we extend the model to incorporate the linker histone model of Luque *et al* into the 1CPN model. We then validate the 1CPN model through calculation of the sedimentation coefficient and find it to be in excellent agreement with experiments. We find the model to be capable of predicting structural features of the elusive chromatin fiber. We provide insight into the non-monotonic behavior of variable linker lengths on chromatin fiber structure through free energy analysis of dinucleosomes. Through this work, we determine that chromatin packing is a result of the complex interplay between energetic penalties of bending DNA and favorable interactions that can overcome such penalties.

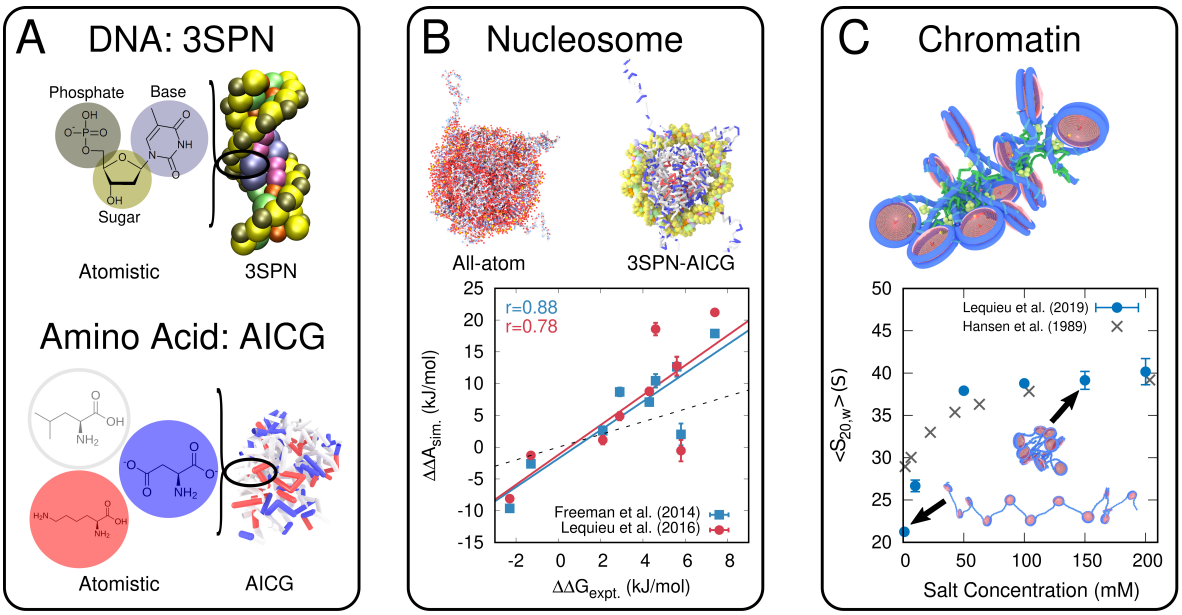

Figure 3.1: A multi-scale approach to develop the 1CPN model. A) Schematic representation of both the three site per nucleotide (3SPN) and atomic interaction-based coarse-grained (AICG) models compared to the respective atomistic structures. The colors of the AICG model represent the net charge of the residue - white for no charge, blue for +1 charge, and red for -1 charge. B) Comparison of the all-atom nucleosome to the 3SPN-AICG model. The 3SPN-AICG model is tested to reproduce the sequence-dependent binding of DNA to the nucleosome.[48, 77] The dotted-line represents a 1:1 fit for experiment and simulation. C) The 1CPN fiber takes in information from the 3SPN scale and reproduces experimental sedimentation coefficient. [55, 76]

3.3 Modeling

3.3.1 Nucleosome-based coarse-graining: The 1CPN Model

The 1CPN model was developed to study questions at the chromatin-fiber or gene scales. Here, we discuss the physics of the 3SPN-AICG model that influenced the design of the 1CPN model. For a full description of the 1CPN force-field, we refer the reader to Lequieu *et al.* [76] This model represents the nucleosomes at a single rigid-body per nucleosome level, while preserving the physics derived from the 3SPN-AICG model. In a previous study, we evaluated the anisotropic nucleosome pair-potential at the 3SPN-scale. Using relative-entropy coarse-graining, the anisotropic pair potential was mapped into a lossless

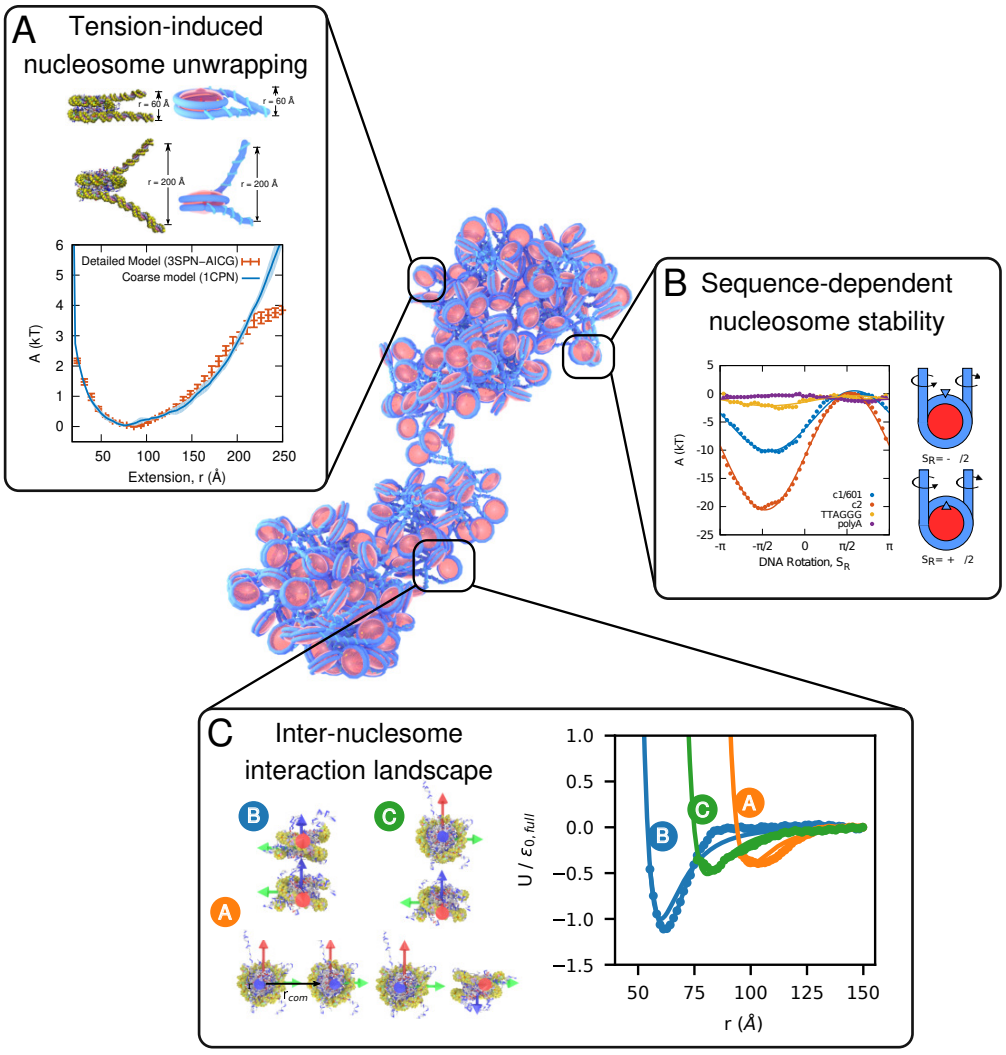

Figure 3.2: The design philosophy of the 1CPN model. The force-field for the model draws heavily from results at the 3SPN-AICG level.[76] A) Tension-induced unwrapping of nucleosomal DNA is implemented in the form of an anisotropic gaussian potential. The unwrapping of DNA matches between the 1CPN and 3SPN models up to the first transition state. [77] B) The sequence-dependent binding of DNA to the nucleosome is implemented in the form of a cosine function, with a modular amplitude depending on the nucleosomal DNA sequence.[48] C) The internucleosome-interactions are included through an anisotropic S-function expansion. The 1CPN model maps the free energy of the 3SPN and incorporates it as a potential for unmodified and acetylated H4 tail nucleosomes.[101]

and parameter-free functional form by means of an S-function expansion formalism (Figure 3.2C).[136] Additionally, a ghost site placed at the dyad of the nucleosome was introduced to mimic the effects of the H3 histone tails. This bead interacts with DNA sites through an

anisotropic Gaussian function to reproduce the frequency of DNA breathing as seen in Figure 3.2A. As the H3 tails are both flexible and positively-charged, they "hook" onto flanking DNA segments, thereby helping modulate the breathing motion. The underlying sequence-dependent positioning energetics were incorporated through an orientation dependent cosine function on the entering and exiting DNA (Figure 3.2B). Because 3SPN influences the design of 1CPN, PTMs and their impact on the structure of chromatin can be effectively taken into consideration.

3.3.2 Linker Histone Implementation

The structure of the chromatin fiber is influenced by DNA-binding proteins, with the H1 or avian variant H5 linker histone being ubiquitous examples [15, 43, 45, 56, 146, 159]. Sedimentation analyses of small chromatin fibers have demonstrated that chromatin with linker histones produces more condensed fibers than those without. [27, 53, 130, 146, 145] The density of positively-charged residues in the linker histone reduces the electrostatic repulsion and promotes binding of the linker DNA, favoring a more compact structure. Given its direct influence on chromatin structure, a model of the H1/5 linker histone by Luque *et al* [88] has been adapted to fit the 1CPN model.

The 1CPN linker histone model is at a coarse-grained level of description comparable to that of 1CPN and draws information from atomistic structures and charge distributions. Structurally, it consists of three domains: the central, rigid globular head (GH) domain consisting of $\sim$ 80 amino acid residues, the flexible, highly positively-charged C-terminal domain consisting of $\sim$ 110 residues, and the short N-terminal domain of $\sim$ 25 residues. [5] In its original model definition, the GH region was coarse-grained to be a fixed rigid-body consisting of 6 charged sites. The positions of these sites relative to the nucleosome center of mass are described in Table-3.5.

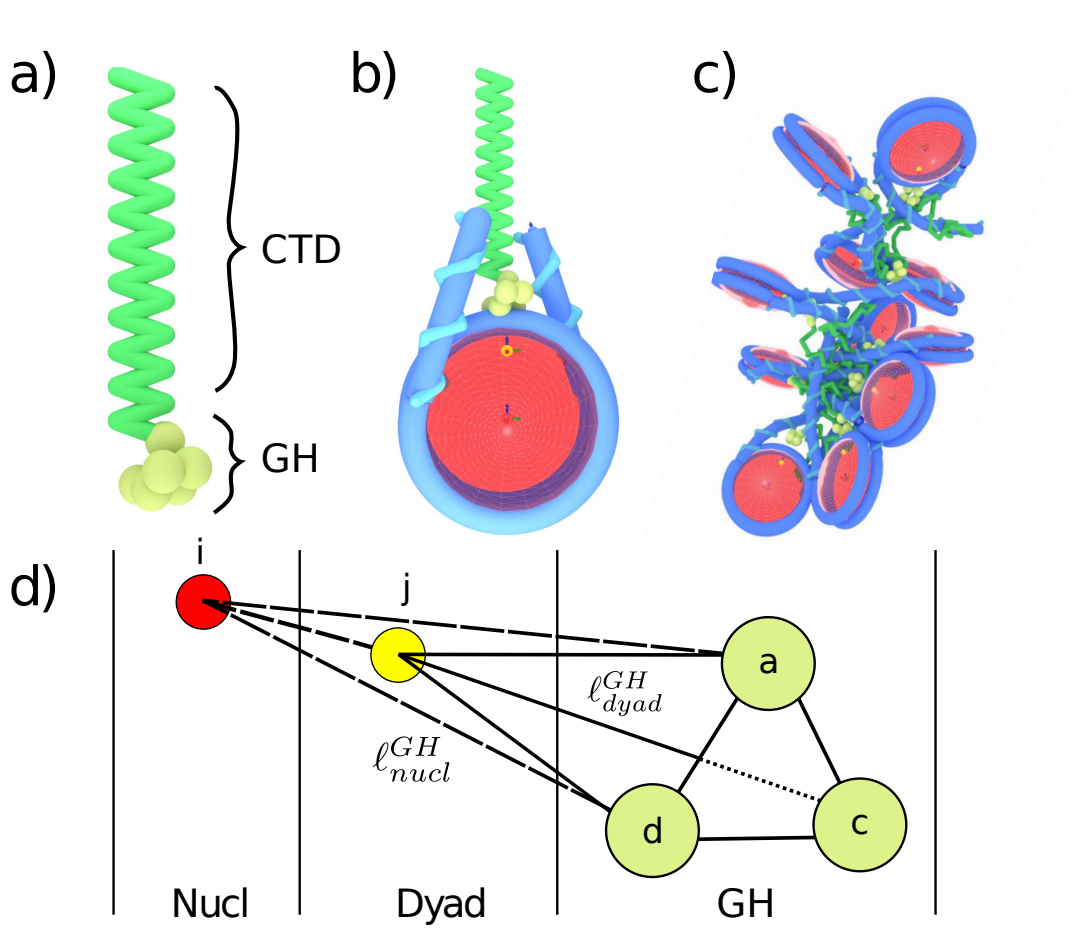

Figure 3.3: The Linker Histone in the 1CPN Model. A representation of the A) independent linker histone, B) a representation of the chromatosome particle consisting of the linker histone binding at the dyad axis and a 167 base pair nucleosome core particle, C) a representation of a condensed 12 nucleosome fiber with linker histones included, and D) a schematic representation of the bonds between the nucleosome and dyad sites with the globular head.

Bonded Interactions

The 1CPN linker histone model is adapted to work with molecular or Brownian dynamics in order to support 1CPN. To adapt the GH to the 1CPN model, stiff harmonic springs are placed to preserve a rigid structure, $U_{\mathrm{GH}}^{\mathrm{geom}}$.

$$\begin{aligned} U_{\mathrm{GH}}^{\mathrm{geom}} = & U_b^h(\{\boldsymbol{r_{GH}}\}; k_{b,GH}, \{\boldsymbol{\ell_{GH}}\}) \\ & + U_a^h(\{\boldsymbol{r_{GH}}\}; k_{a,GH}, \{\boldsymbol{\theta_{0,GH}}\}) \\ & + U_{\mathrm{d}}^{\mathrm{phi}}(\{\boldsymbol{r_{GH}}\}; k_{\phi,GH}, \{\boldsymbol{\phi_{0,GH}}\}) \end{aligned} \tag{3.1}$$

where $\{\boldsymbol{r_{GH}}\}$ is the set of all coordinates of the rigid GH sites, $\{\boldsymbol{\ell_{GH}}\}$ is the set of lengths between the positions of the rigid sites, and $\{\boldsymbol{\theta_{0,GH}}\}$ and $\{\boldsymbol{\phi_{0,GH}}\}$ are the sets of angles and dihedrals of the rigid structure, respectively. In order to ensure that the GH stays rigid, but not too constrained, a subset of the possible bonds, angles, and dihedral restraints are placed on the system and shown in Table 3.5.

Although there is discussion regarding the binding location of the linker histone to the nucleosome, we choose to bind the linker histone GH on the dyad axis. [15, 121] To accomplish this, we employ two harmonic spring potentials to three sites of the GH. We define the potential, $U_{\mathrm{GH}}^{\mathrm{bind}}$ as:

$$\begin{aligned} U_{\mathrm{GH}}^{\mathrm{bind}} = & \sum_{k=a,c,d} U_b^h(\ell_{jk}; k_{dyad}^{GH}, \ell_{dyad}^{GH}) \\ & + U_b^h(\ell_{ik}; k_{nucl}^{GH}, \ell_{nucl}^{GH}) \end{aligned} \tag{3.2}$$

Sites 1, 3, and 4 of the GH are constrained by utilizing two harmonic potentials - one to the dyad and the other to the nucleosome center of mass. The use of three constrained sites

restricts rotation of the GH and excessive lateral movement of the GH from the nucleosome dyad.

Unlike the rigid globular head, the C-terminal tail domain is long and flexible. The implementation of the C-terminal domain remains mostly unchanged from that of Luque *et al.* We describe the bonded forces of the model here for completeness. The geometric bonds of the C-terminal domain $U_{\mathrm{CTD}}^{\mathrm{geom}}$ are defined as:

$$U_{\mathrm{CTD}}^{\mathrm{geom}} = \sum_{i,i+1,i+2}^{N_{\mathrm{CTD}}} U_a^h(\boldsymbol{r}_{i,i+1}, \boldsymbol{r}_{i+1,i+2}; k_\beta^{\mathrm{LH}}, \beta^{LH}) + \sum_{i,i+1}^{N_{\mathrm{CTD}}} U_b^h(\ell_{i,i+1}; k_b^{CTD}, \ell_b^{CTD}) \tag{3.3}$$

where $\boldsymbol{r}_{i,i+1}$ is the position vector between any two subsequent CTD beads. In order to bind the CTD to the GH and the GH to the nucleosome, we place a harmonic spring between site f of the GH and the first bead of the CTD. This potential $U_{\mathrm{LH}}^{\mathrm{bind}}$ is defined as:

$$U_{\mathrm{LH}}^{\mathrm{bind}} = U_b^h(\ell_{LH}; k_b^{CTD}, \ell_b^{CTD}) \tag{3.4}$$

where ℓ_{LH} is the distance between site f and the first bead of the CTD.

Nonbonded Interactions

All linker histone sites are treated similarly for non-bonded interactions. Between linker histone sites, an excluded volume interaction is adopted in the form of a Lennard-Jones potential:

$$U_{\mathrm{LJ}}(\boldsymbol{r}_{ij}, \sigma_{ev}, \epsilon_{ev}) = 4\epsilon \left[\left(\frac{\sigma_0}{r_{ij}} \right)^{12} - \left(\frac{\sigma_0}{r_{ij}} \right)^{6} \right] \tag{3.5}$$

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

Table 3.1: Definition of Nucleosome-Nucleosome Orientations for Pair-Potential Calculations

For all spherical beads, a geometric average of σ_0 is applied based on the sizes of each species. The energy of interaction ,ϵ_0, is set low to make the interaction purely repulsive. Consistent with the DNA sites, linker histone electrostatics are treated at the level of Debye-Hückel. The electrostatic interactions are only used between linker histone and DNA.

3.4 Methods

3.4.1 Nucleosome-Nucleosome Pair Potentials

After mapping the 3SPN-AICG nucleosome model to a position and orientation, the relative orientation of the two nucleosomes i and j is given by six angles that represent all the possible combinations of angles between $\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j, \hat{\boldsymbol{r}}_{ij}$ and $\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}$. In the 1CPN model, our choice of the Zewdie potential assumes uniaxial symmetry about $\hat{\boldsymbol{f}}_i$ and $\hat{\boldsymbol{f}}_j$, and therefore the relative orientation of two nucleosomes only depends on the three vectors $\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}$. Accordingly, we avoid configurations where $\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij} \neq 0$ or $\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij} \neq 0$ which are prevented in chromatin fibers by the entering/exiting nucleosomal DNA. The precise definitions of the 3SPN-AICG orientations used in this work are listed in Table 3.1. Note that Orientations C and D gave nearly identical free energies, and therefore Orientation D is omitted from Figure 3.4 for clarity.

To compute the effective pair-potential, the 3SPN-AICG nucleosomes were constrained to these orientations with strong harmonic angle potentials, and umbrella sampling[68] was

performed along the center-of-mass separation, r, between the two nucleosomes. The parameters in the 1CPN model's Zewdie potential (see Table 3.2) were chosen to minimize the total error with the 3SPN-AICG effective pair potentials over all four orientations.

3.4.2 DNA-mediated nucleosome-nucleosome pair potential

The DNA-mediated nucleosome-nucleosome pair potential (Section 3.5.3) was obtained using umbrella sampling[68] along an order parameter defined by the center-to-center distance, r, between the two nucleosome sites. At short separations, inter-nucleosomal DNA can adopt many different configurations, each of which are separated by relatively large energy barriers. As a consequence, diffusion orthogonal to the order parameter was very slow, and obtaining accurate sampling of all possible DNA configurations was slow.

To improve sampling , we used 10 independent umbrellas at each r and choose relatively weak umbrella force constants in order to allow a given configuration to sample a wide range of r, thereby permitting the two nucleosomes to unfold and refold many times in each umbrella.

3.4.3 Sedimentation Coefficients

Sedimentation coefficients were obtained by first initializing a 12 nucleosome, 207 nucleosome repeat length (NRL) fiber in an extended configuration where the nucleosomes were not in contact. These fibers were then relaxed using Brownian dynamics until they condensed (typically 1×10^9 time steps, or 20 μs). This condensed configuration was then used as the initial configuration for a replica exchange simulation at the specified salt concentration with 24 temperatures spaced geometrically between 300K and 700K. Since the energy barriers between different fiber configurations are relatively large, this replica exchange simulation was necessary to accelerate the exploration of many possible fiber conformations, and improve the estimate of the sedimentation coefficient. The replica exchange simulations were typically

$\approx 1 \times 10^8$ steps, or 2 μs.

The sedimentation coefficient was determined, as described previously[11], according to

$$S_{20,w} = S_1 \left(1 + \frac{2R}{N'} \sum_{i}^{N'} \sum_{j>i}^{N'} \frac{1}{R_{ij}} \right) \tag{3.6}$$

where N' is the number of nucleosomes in the fiber, R is the effective radius of the nucleosomes and is assumed to be 54.6$\mathring{A}$. S_1 is the $S_{20,w}$ of a mono-nucleosome taken as equal to 11.1 Svedberg (S), and R_{ij} is the distance between two nucleosomes. The values of R and S_1 were chosen to be consistent with the values used in prior work[11]. The sedimentation coefficient was monitored throughout the replica exchange simulation (at 300K), and was determined as the average once the sedimentation coefficient converged. Error bars represent the standard deviation from four independent simulations.

3.5 Results

3.5.1 Nucleosome-Nucleosome Pair Potential

To parameterize the nucleosome-nucleosome interaction in 1CPN we first compute the anisotropic pair-potential using the 3SPN-AICG nucleosome model (Figure 3.4). The pair-potential is strongly dependent on the orientation of the nucleosomes, with the energy and length-scales of the interaction ranging from $U/\epsilon_0 \approx 1.0$ to 0.4 and 65$\mathring{A}$ to 100$\mathring{A}$, respectively. The lowest energy configuration corresponds to two nucleosomes stacked face-to-face (Orientation B).

Note that 3SPN-AICG can be used to explore the effects of epigenetic markers, such as the acetylation of the histone tails, on the nucleosome-nucleosome pair potential. As a proof-of-principle, we again calculate the anisotropic pair-potential using 3SPN-AICG, but modify the histone proteins so that the histone H4 tail is removed. The H4 tail is thought to mediate interactions between nucleosomes[128], and modifications to H4 represent an important epigenetic mechanism by which chromatin compaction is controlled. The removal

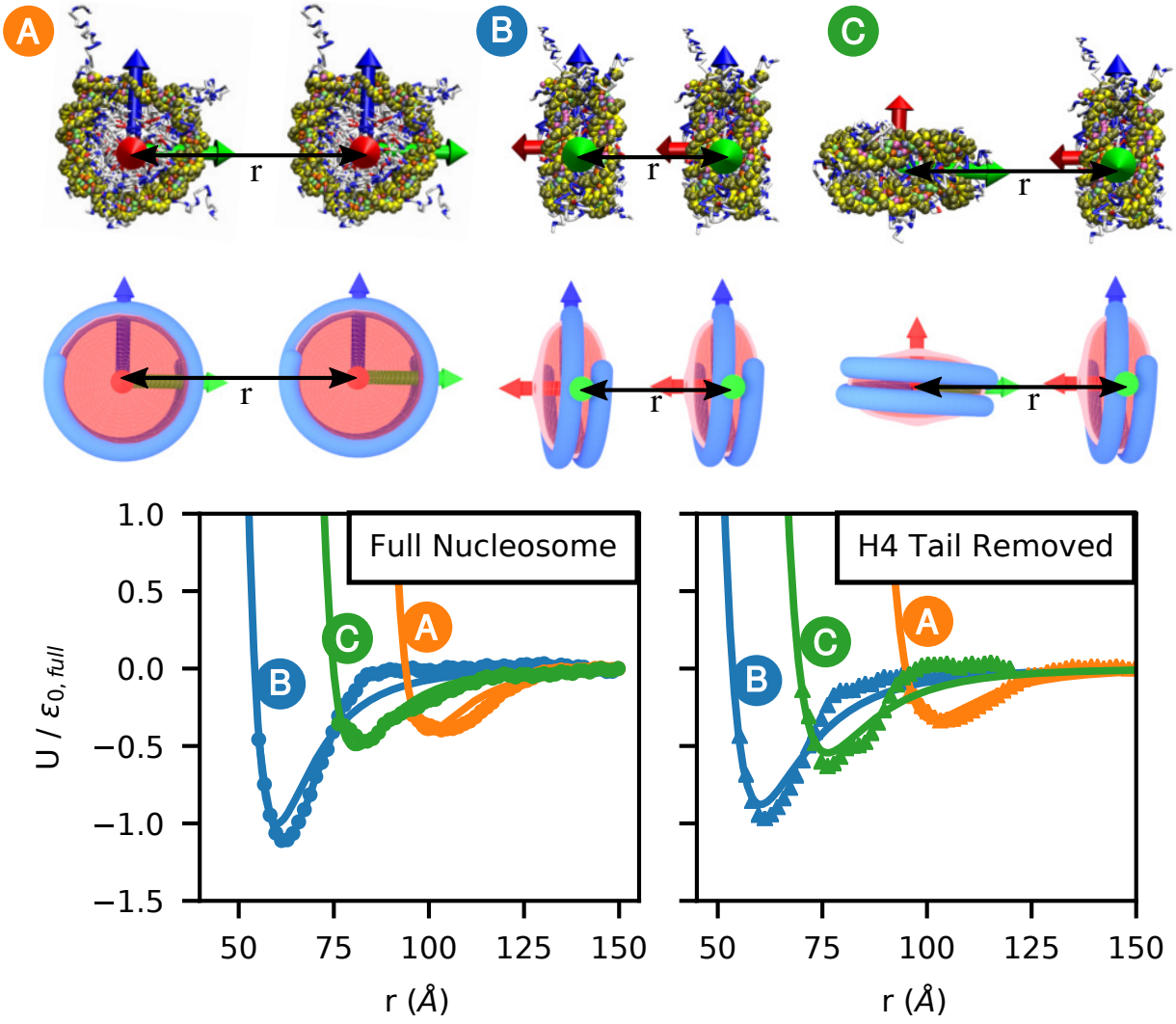

Figure 3.4: Pair-Potential between two nucleosomes for several orientations for 3SPN-AICG (points) and 1CPN models (lines). The orientations of the nucleosomes are defined by the individual coordinate systems and the displacement vector between them, r_{ij}. The coordinate systems are defined as $\hat{\boldsymbol{f}}$ is red, $\hat{\boldsymbol{u}}$ is blue, and $\hat{\boldsymbol{v}}$ is green.

of the H4 tail is observed to decrease the attraction of nucleosomes in Orientation B, and leads to a significant widening of the potential in Orientation B (Figure 3.4).

Having obtained these free-energy surfaces with the 3SPN-AICG model, we then parameterize the nucleosome-nucleosome interactions energies in 1CPN (see Methods). The resulting pair potentials obtained with 1CPN are shown by the solid lines in Figure 3.4, and the parameters are reported in Table 3.2 in Section 3.7. The agreement between the pair potentials from 3SPN-AICG (points) and those from 1CPN (lines) is good, with all length and energy scales matched between the two models. Notably, good agreement between the models is obtained for both the "full" nucleosome and for nucleosomes lacking the H4 tail. This result indicates that both the complex orientation-dependent interactions between nucleosomes and the effects of histone modifications on these interactions can be effectively coarse-grained into the 1CPN model's Zewdie potential. When computing these pair potentials in 3SPN-AICG, the maximum energy of attraction, $\epsilon_{0,full}$, was found to be 2.69k_BT, in agreement with experiment. [101, 33, 50] In the 1CPN model presented here, we choose $\epsilon_{0,full}$ as an adjustable parameter that sets the energy scale for the nucleosome-nucleosome interactions. This choice allows the 1CPN model to retain the relative energies of the different orientations and histone modifications, yet provides the flexibility to vary the strength of attraction between nucleosomes as additional experimental or simulation become available.

3.5.2 Chromatin Fiber Sedimentation Coefficients

As mentioned earlier, 1CPN is well suited to examine the structure and dynamics of chromatin fibers. To illustrate this idea, we prepared chromatin fibers consisting of 12 nucleosomes with a nucleosome repeat length of 207 bp and various salt concentrations. Following their relaxation, we computed the sedimentation coefficient (see Methods), and compared our predictions to experimental measurements for an identical system. [55] (Figure 3.5) Sedimentation coefficients have been used extensively to assess the relative compaction of different

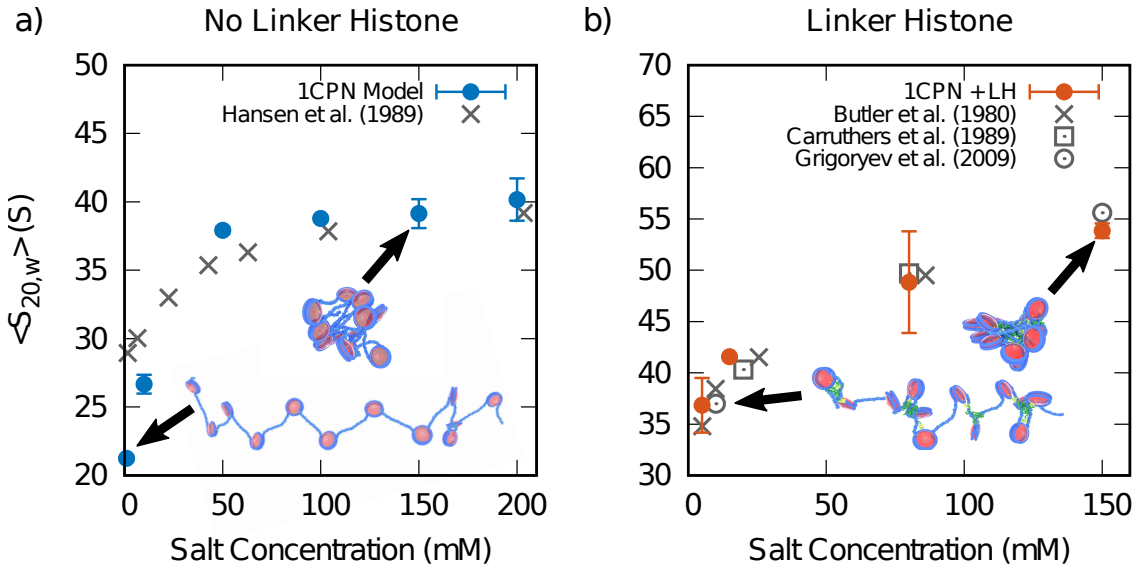

Figure 3.5: Sedimentation Coefficient, $\langle S_{20,w} \rangle$, for short chromatin fibers containing 12 nucleosome, as a function of salt concentration in the A) absence and B) presence of the Linker Histone. The nucleosome repeat length of 207 base pairs. the agreement between the 1CPN model (colored points) and available experimental measurements (grey points) is good.

chromatin fibers, and these calculations represent an important test of 1CPN [53, 130]. As can be seen in the figure, simulations and experiments agree with each other, particularly for the effects of salt on the compaction and expansion of chromatin.

3.5.3 Dinucleosome Analysis

To understand how PTMs influence the structure of chromatin, the structure of unmodified chromatin and what drives its formation must be assessed. The basis for chromatin fiber structure seemingly derives from DNA deformation penalties and the mechanical modes that limit them. Generally, these can be thought of as the energetic penalties for deforming DNA, and the interactions that are favorable, such as inter-nucleosome contacts, nucleosome positioning energy, and H3-tail induced flexibility. Of particular importance to these interactions is the amount of linker DNA connecting each nucleosome. From experiments and simulation, it is found that linker DNA lengths that are in integer amounts of DNA pitch

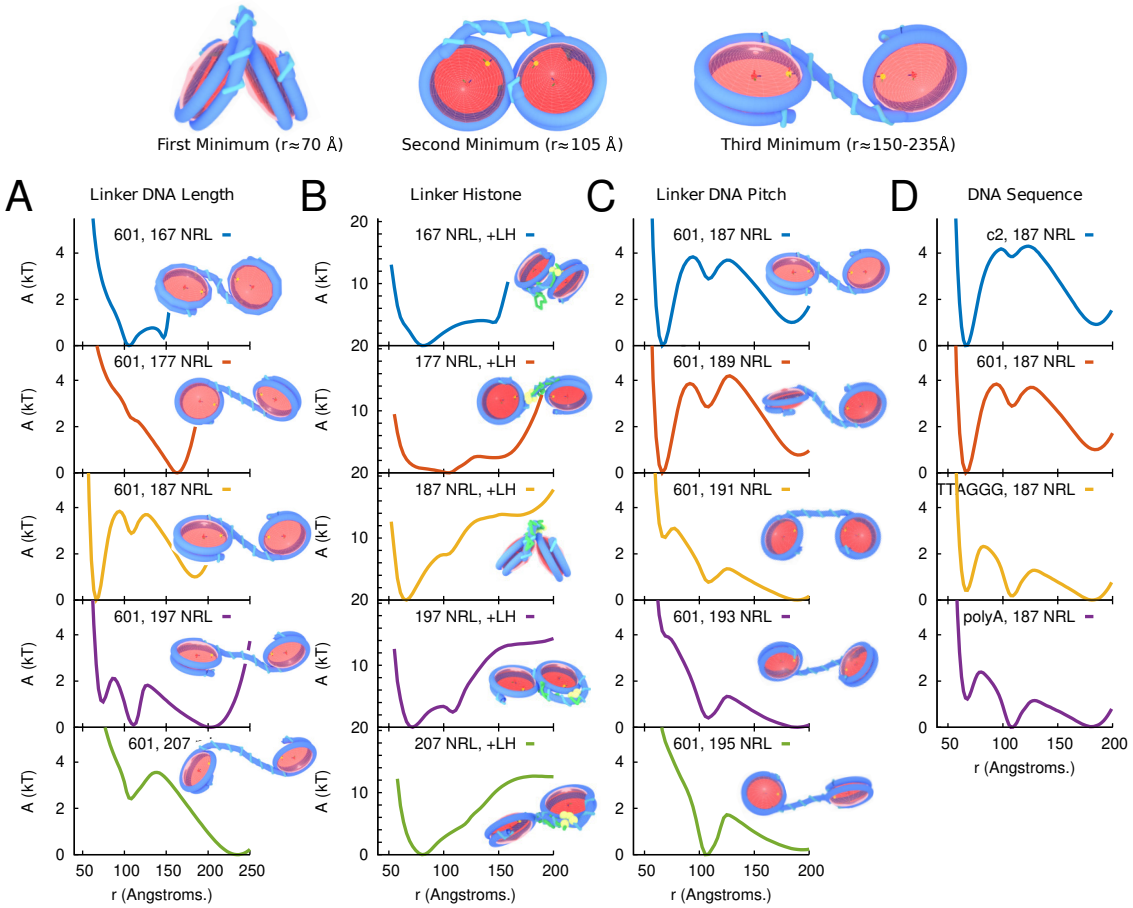

Figure 3.6: Dinucleosome pair potentials demonstrate the sensitivity of chromatin physics to variable linker lengths. All surfaces fall into the category of three potential minima. A) Dinucleosome potentials of mean force for variable linker lengths of even pitch, B) even pitch with the linker histone, C) variable pitch, D) variable DNA sequence. These results are adapted from Lequieu *et al.* [76]

(∼10 bp) lead to compact structures, whereas deviations frustrate the fiber, leading to larger structures. This property influences the eukaryotic genome, as high-throughput sequencing of linker DNA has revealed peaks of intensity at integer lengths of DNA pitch. [25, 29, 156] To highlight the pronounced effect of linker length on the structure of chromatin, we have considered the energetics of the dinucleosome system as seen in Figure 4.1. Using umbrella sampling, one can calculate a potential of mean force between the nucleosome centers of mass and analyze the resulting curves. This calculation is considered for linker lengths of 20-60 bp, variable nucleosomal DNA sequence, and linker histone presence. We note that

linker DNA length significantly alters the shape of the curves. Most notably, three structural motifs of the dinucleosome arise in these curves: (1) a stacked or condensed structure, (2) a side-side structure, and (3) a fully separated structure. Additionally, increasing the strength of the nucleosomal DNA positioning or including the linker histone shifts the curves towards condensed structures and decreases the relative free energy of the condensed structure, consistent with our prior explanations. While we note that these dinucleosomes display structural features similar to those seen in the 30-nm fiber, we find that the 1CPN model does not predict the 30-nm fiber to be a stable structure. In fact, larger fibers display motifs of fibers that are fluid in radius, much like the results of ChromEMT. [112]

Unfortunately, DNA linker length is not a static quantity. To accommodate transcription factors, nucleosomes can reposition through sliding events or dissociate from the DNA, both altering the length of linker DNA. Theoretical, simulation, and experimental work has been performed to unravel the nucleosome repositioning network. Simulations and experiments have uncovered the mechanism of sliding events and indicate that they are not only dependent on the underlying DNA sequence, but also on nucleosome modifications. For example, recent work has demonstrated that the CenpA histone induces a dramatic sliding event that is relevant for centromeric chromatin. [142] In other words, sliding events may only be a factor in variable linker lengths, but not the sole determinant. As shown at a dinucleosome scale, variable linker lengths significantly alter the chromatin physics, and increasing the number of nucleosomes exponentially increases the complexity of the resulting chromatin structure.

3.6 Conclusion

In this work, the 1CPN model has been extended to include the linker histone and validated. By relying on a multi-scale approach, 1CPN incorporates physics that occur over nano-meter length scales, such as histone modifications and DNA sequence. The model, however, is computationally efficient and permits simulation of many kilobases of chromatin.

The parameterization of 1CPN has relied both on extensive simulations with the detailed 3SPN-AICG model of the nucleosomes, as well as experimental measurements on the structure and dynamics of chromatin. 1CPN has been parameterized to reproduce the many free energies that govern the interactions within chromatin, such as the interactions between nucleosomes, the interaction between DNA and histone tails and how these interactions can be modulated by histone modifications. 1CPN has also been parameterized to reproduce the salt-dependent stiffness of DNA, the effect of DNA sequence on rotation of DNA within the nucleosome, and the dynamics of short chromatin fibers. Following this parameterization, 1CPN has been used to examine the free energies of association between two nucleosomes separated by different lengths of DNA. It is found that the length of this DNA, as well as the relative pitch between two nucleosomes, have a dramatic effect on the interactions between nucleosomes. Finally, it was demonstrated that 1CPN achieves quantitative agreement with experimental measurements of sedimentation coefficients of short chromatin fibers, both in the presence and absence of linker histone H1.

Building on this foundation, it is anticipated that 1CPN will be useful for studies of the many dynamic processes that dictate chromatin compaction. The wide variety of mechanisms that link chromatin structure to gene expression are still poorly understood, and the 1CPN model provides a tool to interrogate these relationships.

3.7 Model Parameters

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

Parameter	Value	
	Full	-H4
σ_0	55.0 $\mathring{A}$	55.0 $\mathring{A}$
ϵ_0	1.163 kcal/mol	1.303 kcal/mol
σ_0'	30.0 $\mathring{A}$	30.0 $\mathring{A}$
ϵ_0'	0.01 kcal/mol	0.01 kcal/mol
σ_{000}	1.559	1.564
σ_{cc2}	-0.756	-0.754
σ_{220}	0.143	0.151
σ_{222}	0.292	0.267
σ_{224}	0.000	0.00
ϵ_{000}	0.605	0.743
ϵ_{cc2}	0.528	0.565
ϵ_{220}	-0.029	-0.246
ϵ_{222}	-0.270	-0.580
ϵ_{224}	0.000	0.00
Source: Fig. 3.4		

Table 3.2: Nonbonded parameters for U_{zewdie}. See Fig. 3.4 for justification of parameter values

Equil. Lengths and Angles

Parameter	Value	Source
ℓ_{a-c}	9.655702 Å	Ref.[88]
ℓ_{a-d}	10.080155 Å	Ref.[88]
ℓ_{b-c}	14.904854 Å	Ref.[88]
ℓ_{b-f}	12.619831 Å	Ref.[88]
ℓ_{d-e}	16.080128 Å	Ref.[88]
ℓ_{e-f}	16.815193 Å	Ref.[88]
ℓ_{b-e}	11.760170 Å	Ref.[88]
ℓ_{c-d}	13.166473 Å	Ref.[88]
$\ell_{i,\{a,c,d\}}$	68 Å	Geometry
$\ell_{j,\{a,c,d\}}$	33 Å	Geometry
$\ell_{CTD_{eq}}$	15 Å	Ref.[88]
$\theta^{\mathrm{GH}}_{\mathrm{c-a-d}}$	83.662981°	Ref.[88]
$\theta^{\mathrm{GH}}_{\mathrm{a-d-e}}$	143.550333°	Ref.[88]
$\theta^{\mathrm{GH}}_{\mathrm{c-b-f}}$	57.281993°	Ref.[88]
$\theta^{\mathrm{GH}}_{\mathrm{b-e-f}}$	58.119728°	Ref.[88]
$\theta^{\mathrm{GH}}_{\mathrm{a-f-e}}$	81.942273°	Ref.[88]
$\theta^{\mathrm{GH}}_{\mathrm{b-c-d}}$	56.308143°	Ref.[88]
β^{LH}	110°	Ref.[88]
$\phi^{\mathrm{GH}}_{\mathrm{c-a-d-e}}$	72°	Ref.[88]
$\phi^{\mathrm{GH}}_{\mathrm{c-b-f-e}}$	123°	Ref.[88]

Force Constants

Parameter	Value	Source
$k^{\mathrm{GH}}_{\mathrm{b}}$	50 kcal/mol/Å^2	Section 3.3.2
$k^{\mathrm{GH}}_{\mathrm{a}}$	10 kcal/mol/deg^2	
$k^{\mathrm{GH}}_{\mathrm{d}}$	10 kcal/mol	
$k^{\mathrm{GH}}_{\mathrm{nucl}}$	50 kcal/mol/Å^2	
$k^{\mathrm{CTD}}_{\mathrm{b}}$	0.1 kcal/mol/Å^2	Ref.[88]
k^{LH}_{β}	1 kcal/mol/deg^2	Ref.[88]

Table 3.3: Bonded parameters for the linker histone globular head and c-terminal tail domains.

Parameter	Value
σ_{GH}	15.0 Å
σ_{CTD}	18.0 Å
ϵ_{ev}	0.001 k_BT
Source: Ref.[13]	

Table 3.4: Nonbonded parameters for linker histone interactions

Site	$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{z}$
a	-0.3413114	64.947141	10.7562
b	4.148044	57.478201	-6.935
c	7.535318	65.034297	5.458
d	-3.689099	58.350318	3.817
e	-7.547358	57.271017	-8.15
f	0.419315	71.868134	-4.953

Table 3.5: Definition of Globular Head site positions with respect to the nucleosome center of mass in Angstroms. The nucleosome is oriented that $\hat{\boldsymbol{f}} = \{1, 0, 0\}$, $\hat{\boldsymbol{u}} = \{0, 1, 0\}$, and $\hat{\boldsymbol{v}} = \{0, 0, 1\}$.

CHAPTER 4

CHROMATIN STRUCTURE BEYOND THE NUCLEOSOME: A TRINUCLEOSOME BUILDING BLOCK OF THE CHROMATIN FIBER

4.1 Abstract

Chromatin is comprised of long DNA molecules, which are hierarchically condensed from their original length (meters) into extremely small aggregates (micrometers). The underlying compaction process is partly regulated by dynamic protein-DNA interactions, which allow chromatin to organize into discrete structures. The detailed structures of the smallest features of chromatin, including DNA and the nucleosomes, are known. In contrast, the structure of chromatin beyond the "beads-on-a-string" fiber description is poorly understood. The so-called "30-nm fiber" as a chromatin construct represents a controversial concept, and recent evidence points towards a much smaller building blocks. In particular, experimental evidence points to a tetranucleosome and the nucleosome "clutch" — a cluster of 4-16 nucleosomes — as candidates for fundamental units of chromatin. In this work, we utilize the recently developed 1CPN molecular model of chromatin to interrogate the building blocks of chromatin. We use free energy sampling methods and replica exchange molecular dynamics to identify the similarities between these small units. Longer fibers are employed to identify the similarities between the structure of di and trinucleosome based constructs, and the structures that emerge in much larger fibers. We find that the trinucleosome free energy surface displays great similarities to that of fibers with homogeneous linker-DNA lengths and without the linker histone. In the presence of the linker histone, we find the fiber to favor a clutch-like motif. Laslty, we show that acetylation of the H4 tails leads to melting of the trinucleosome motifs, serving to underscore the role of H4 acetylation in controlling the structure of the chromatin fiber.

4.2 Introduction

In the cell nucleus, chromatin organizes DNA into dynamic hierarchical structures. Within these structures, DNA is regulated through sequences of epigenetic modifications known as post-translational modifications (PTMs). The complex network of PTMs is referred to as the "histone code."[64] A central target of these modifications is the nucleosome core particle (NCP) - comprised of 147 base-pairs (bp) of DNA wrapped $\sim$1.7 times in a left-handed super-helix around protein octamers called histones.[87, 34] The nucleosome has been long considered to be the building block of the chromatin fiber. At longer scales, nucleosomes form a "beads-on-a-string" structure of NCPs connected by segments of 10-80 bp linker DNA. [110] The complex transition between the "beads-on-a-string" model and compact arrays of nucleosomes is crucial to DNA replication, transcription, and repair. [54, 2, 75, 152, 80, 160, 26, 73]

Advances in chromatin analysis have revealed key structural features of the fiber at multiple length scales. New imaging techniques, such as Cryo-EM, have gradually unveiled some of chromatin's unique structural motifs. [20, 138, 133, 40] Recently, Cryo-EM images of small fibers were used to determine that the nucleosome tetramer is a potential repeat unit in the chromatin fiber, supporting previous evidence for the existence of tetranucleosome building blocks.[138] More recently, *in vivo* STORM imaging experiments have yielded evidence for nucleosome "clutches" comprising 4-24 nucleosomes.[124] The size of these clutches is positively correlated with cell pluripotency. Further experimental evidence from magnetic tweezers experiments indicates that the FACT protein regulates transcription by attenuating these clutches. [**?**] A recent study using high-resolution nucleosome orientation mapping reported two unique geometries of tetranucleosomes that combine to form the chromatin fiber. [109] Take together, this evidence suggests that a building block exists, encoded into the chromatin fiber, that is larger than the single nucleosome.

At the nuclear scale, novel imaging techniques have elucidated structural organizations

and correlations in chromatin in the nucleus. One such technique, Partial Wave Spectroscopy (PWS), uses chromatin density heterogeneity to elucidate structural fluctuations of chromatin below the diffraction limit.[143, 8] The amplitude of these fluctuations has been shown to positively correlate with human colorectal cancer cells. [7] In the last few years, ChromEMT has emerged as a powerful method that combines ChromEM staining for electron microscopy with multi-tilt tomography. Chrom-EMT has proven to be particularly useful for understanding the 3D structure of chromatin at a detailed resolution during interphase and metaphase. [112] Chrom-EMT images suggest that smaller clutches of nucleosomes exist in interphase chromatin.

The imaging techniques outlined above have revealed key aspects of chromatin's structure. However, our understanding of the underlying physics governing that structure and its hierarchical organization is very limited. A central question that we ask in this work is whether the overall structural organization of chromatin in the nucleus stems from repeating segments of short fibers, like those proposed by the "clutch" model. Past work has invoked the concept of a 30-nm fiber as a recurring motif. [45, 128, 130, 133, 39, 53] Two candidate structures have been reported: the two-start zig-zag, with straight linker DNA and next-nearest neighbor nucleosome contacts, and the one-start solenoid, with bent DNA connecting nucleosomes with nearest-neighbor contacts. Recent developments, however, have questioned its existence *in vivo*.[112, 92, 93, 94] Unfortunately, imaging techniques are limited both spatially and temporally, and additional information from different sources is needed to reconcile these different perspectives.

As an alternative solution, molecular models can provide insight into the dynamic configurations of chromatin across multiple length and time scales. [100] Recent developments in coarse-grained models have allowed for the study of the physics of dinucleosomes and trinucleosomes in considerable detail. [144] In this work, we employ the 1CPN (1-Cylinder-Per-Nucleosome) chromatin model, which builds upon these previous studies, to under-

stand the structure of chromatin at length scales between the nucleosome and larger, hierarchical, nuclear organizations.[76] The 1CPN model was developed as a coarse-grained model that incorporates key physics and dynamics from more detailed simulations of the nucleosome.[48, 58, 77] Given the rich and complex physics that arise in the nucleosome physics, it provides an avenue to study the free energy landscape of small chromatin fibers, and how such landscapes influence the structure of large fibers.

In what follows, the 1CPN model is used to determine the free energies of dinucleosomes and trinucleosomes as a function of several order parameters. Those free energies are then used to predict the structure of larger chromatin fibers. We also consider the effect of the linker histone on compaction of the chromatin fiber and quantify the resulting structural motifs through nucleosome distance distributions. From electron microscopy, it is known that linker length influences chromatin structure. [130, 53] As such, we examine the free energy changes that underlie chromatin structure with varying linker DNA length. Additionally, we explain how these free energies relate to previously proposed structures of the 30-nm fiber and estimate their relative stability. Throughout the manuscript, we provide evidence supporting that trinucleosomes serve as a building block or structural motif for homogeneous larger fibers. Lastly, we extend our analysis to illustrate that chromatin becomes more fluid in the presence of acetyl-H4 tails, stemming from weakened trinucleosome interactions. We suggest that the loss of structural order renders chromatin suceptible to external machinery such as transcription factors, resulting in a correlation between H4 acetylation and transcription.

4.3 Methods

The model adopted here is referred to as the "1-Cylinder-Per-Nucleosome" chromatin representation. It describes the nucleosome as a single anisotropic particle, and DNA as a spherical particle at a three basepair-per-bead resolution level. The nucleosome shape is dictated by calculations of nucleosome interactions at a more detailed level. [101] The nu-

cleosome interactions have also been parametrized for H4-tail acetylations. The model is designed to preserve the geometric 1KX5 structure of entering and exiting DNA.

From previous fiber reconstitution experimental studies, researchers have identified that the chromatin structure is particularly sensitive to the amount of linker DNA between nucleosomes and the inclusion of the linker histone. [130] Based on the sensitivity of the chromatin fiber to these parameters, we consider them in our analysis of chromatin building blocks. In particular, we represent the ratio of linker histones (LH) to nucleosomes as [LH]. The detailed linker histone model used here was introduced by Luque et al.[88], and it coarse-grains the 80 residue globular head (GH) domain as 6 beads of variable size and charge. These values were calculated using the Discrete Surface Charge Optimization (DISCO) method. The long and flexible 110 residue C-terminal domain (CTD) is represented by 22 spherical beads, each at a 5 residue-per-bead resolution. A more detailed description of can be found in the literature.[88, 10, 11, 13, 115]

Solvent in the 1CPN model is represented implicitly with a Langevin thermostat using a timestep of 60 fs. The electrostatics are treated at the level of Debye Hückel theory with physiological salt conditions of 150 mM. All simulations in this work were carried out using the LAMMPS molecular dynamics package. [119] The 1CPN implementation in LAMMPS is available at https://uchic.ag/1cpn.

Potentials of mean force (PMFs) were calculated using umbrella sampling with the weighted-histogram analysis method (WHAM) for the small fiber dinucleosome and trinucleosome simulations. All umbrella sampling simulations were calculated using the separation of the center of mass of the nucleosomes as the collective variable. [68] The dinucleosome free energies were run with umbrella centers spanning 50 Å to 190 - 250 Å (depending on the amount of linking DNA) and were separated by 10 Å . These centers are the same for systems with and without the linker histone. In order to avoid end effects, there was no linker DNA entering the first nucleosome or exiting the last nucleosome. Since the nucleo-

somes are anisotropic, the center of mass separation is degenerate at lower separations. To account for this degeneracy, 10 replicas of all umbrella windows were run for a length of $2x10^9$ time steps, which was found to provide sufficient sampling to construct the underlying PMF. Note that every replica corresponding to the same window was aggregated into the WHAM calculation to generate a single PMF.

A similar procedure was used to generate PMFs for trinucleosome systems. For computational reasons, the trinucleosome PMF was considered using a single NRL. The NRL of 187 was chosen for its unique property to stabilize the "stacked" configuration of the dinucleosome. We consider this system to encode for the most crystalline order in the chromatin fiber. As opposed to the dinucleosome, a 2D PMF surface was generated using the separation of the first two nucleosomes (r_α) and the first and the third nucleosome (r_β). Due to the increased computational cost of the system, only 3 replicas of the surface for the LH and LH-free systems were run.

The longer fibers consist of 24 nucleosomes with a homogeneous 187 NRL, unless stated otherwise. Relaxation of long fibers is computationally challenging. Great care is taken to ensure that fibers reach an equilibrium state that does not depend on initial conditions. First, simulations are initialized as elongated fibers where the angle between the entering and exiting DNA and the nucleosome is set at 180°. Four replicas of these simulations are run with different random number seeds until equilibration. In this case, the radius of gyration (R_g) is monitored until it reaches a steady state value. The necessary time is on the order of $\sim 100\mu$s.

After relaxation is achieved, the final configuration from the relaxation procedure is used as the initial frame for a replica-exchange molecular dynamics (REMD) simulation. For each replica-exchange simulation, we generate 24 windows with temperatures geometrically distributed between 300 and 700 K. This range is such that at the highest temperature the fiber is elongated, or "denatured." To measure convergence, we follow R_g of each fiber at

300 K until it reaches a plateau.

The REMD trajectories are analyzed to recover both contact maps and probability distributions. The contact maps are generated by calculating the frequency with which two nucleosomes in a fiber are in contact. Here, two nucleosomes are said to be in contact if they are in a "stacked" configuration. The stacked configuration corresponds to a contact distance between two nucleosomes of $63 - 77$ Å. All probability distributions are histograms generated by normalizing the separation of nucleosomes by the number of total snapshots and instances of separation within a fiber across all replicas.

4.4 Results

To understand the structure of chromatin we begin with the dinucleosome system. The linker histone (LH) is implicated in condensing the chromatin fiber, and we therefore consider dinucleosome simulations where the number of bound linker histones is varied for the same range of NRLs as before. The results for the free energy surface with one LH and two LHs are shown in Fig. 4.1B and C. The linker histone stabilizes the stacked configuration relative to the no-linker histone results. This occurs because the positively-charged and flexible linker histone C-terminal domain (CTD) compensate for the energetic penalty incurred upon bending the DNA, causing nucleosome interactions to become dominant.

Previous work established the ability of the 1CPN model to capture chromatin physics at the dinucleosome scale.[76] By this, we refer to the competing effects of the flexibility of the DNA at length scales below its persistence length and the interaction landscape of the nucleosomes. That work also examined the free energy of the dinucleosome, accounting for many relevant parameters such as nucleosome repeat length (NRL), binding strength of the nucleosomal DNA, and pitch of the DNA. There are three energetic wells corresponding to states where the nucleosomes are stacked, side-by-side, or far apart, corresponding to the state where flexibility of the DNA dominates. There are clear trends associated with

increasing the binding strength of the DNA, which lowers the free energy of the stacked configuration. Note that varying the NRL in increments of 10 bp shows another distinct trend: The 187 NRL dinucleosome is the only system where the global minimum corresponds to the stacked configuration. We reiterate these results in Fig. 4.1A.

Interestingly, we note that the 177 NRL system is the only system where the stacked configuration is not the minimum for the one LH case, but it is for two LH. Results from prior cryo-EM measurements used the LH with 177 NRL nucleosome tetramers and inferred a zig-zag like conformation with straight linker DNA. [138] Our results suggest that while chromatin with the LH trends towards bent DNA, chromatin with a repeat length of 177 bp still favors straighter linker DNA, similar to the reported images.

Based on the dinucleosome results, we next examine larger chromatin structures. We simulate a 24 nucleosome fiber with an NRL of 187 which, as indicated in Fig. 4.1A, was shown to possess the unique property of stabilizing the stacked configuration without the LH. In order to generate a statistically meaningful ensemble of structures, we simulate four replicas of equilibrated chromatin fibers with temperature replica exchange. To interpret the fiber results as consisting of a combination of many dinucleosome interactions, we calculate contact maps and histograms to highlight the probability distributions of pairwise separations of nucleosomes (Fig. 4.2). Figure 4.2C shows the frequency of contacts between any two nucleosomes in a fiber without the linker histone. It can be seen that contacts are dominated by next-nearest neighbor occurrences, which we refer to as 1-3 nucleosome interactions. In other words, every other nucleosome along the chain has a high probability of contact. In Fig. 4.2E shows that as a function of distance, the 1-2 interactions form a Gaussian distribution centered at a distance of ~ 200Å, while only 1-3 interactions show peaks at distances of 70 and 100Å. These peaks correspond to both stacked and side-by-side interactions of the nucleosomes, serving to underscore that 1-3 interactions dominate the structure of chromatin, and the flexibility of DNA controls 1-2 interactions.

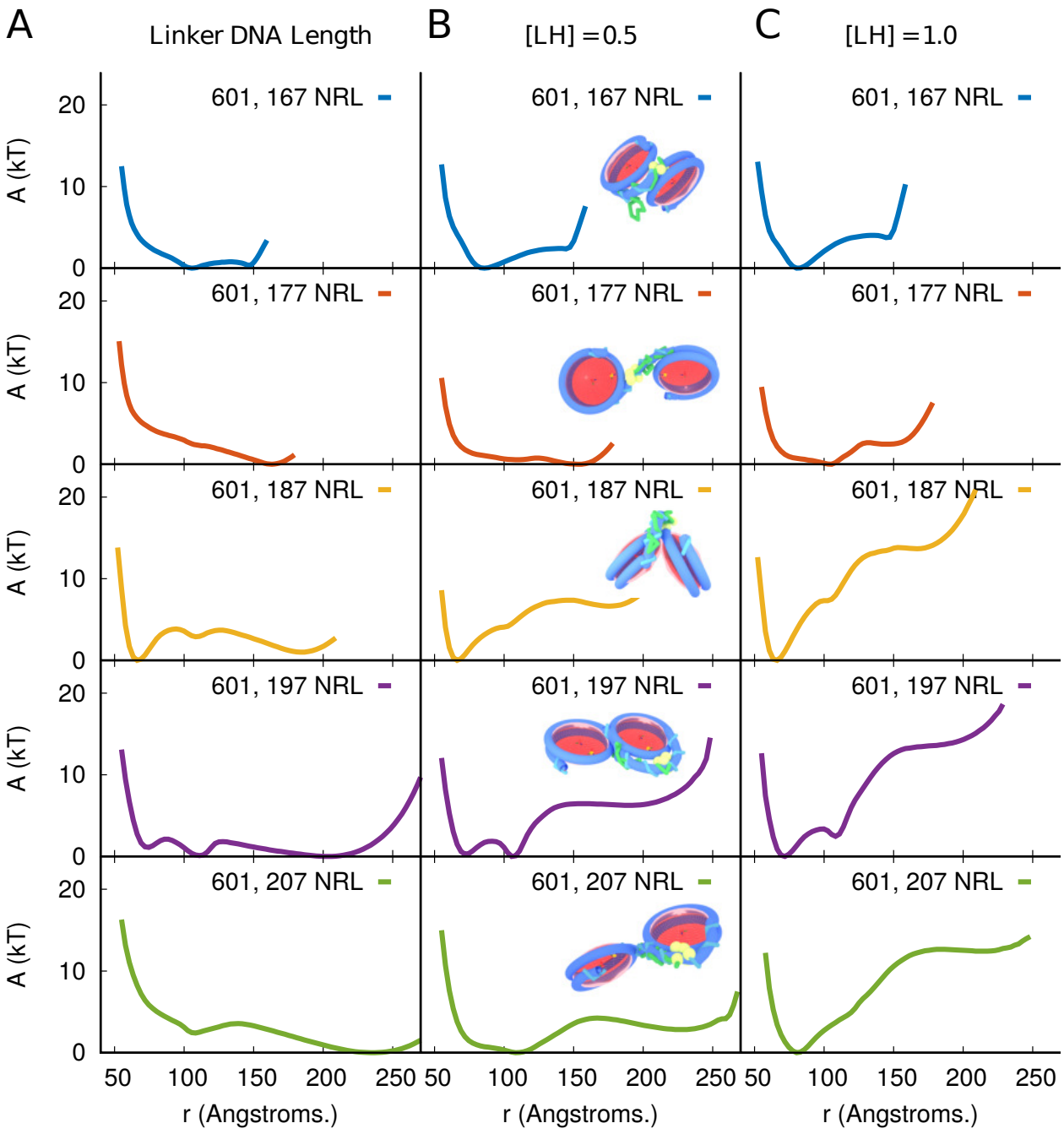

Figure 4.1: The free energy of dinucleosomes with varying concentration of bound H1 linker histones. The columns are arranged in increasing value of NRL, starting from 167 in increments of 10 to an NRL of 207. A) Free energy when [LH] = 0. Data are taken from previous work of Lequieu et al. [76] B) Free energy when [LH] = 0.5. C) Free energy when [LH] = 1.0.

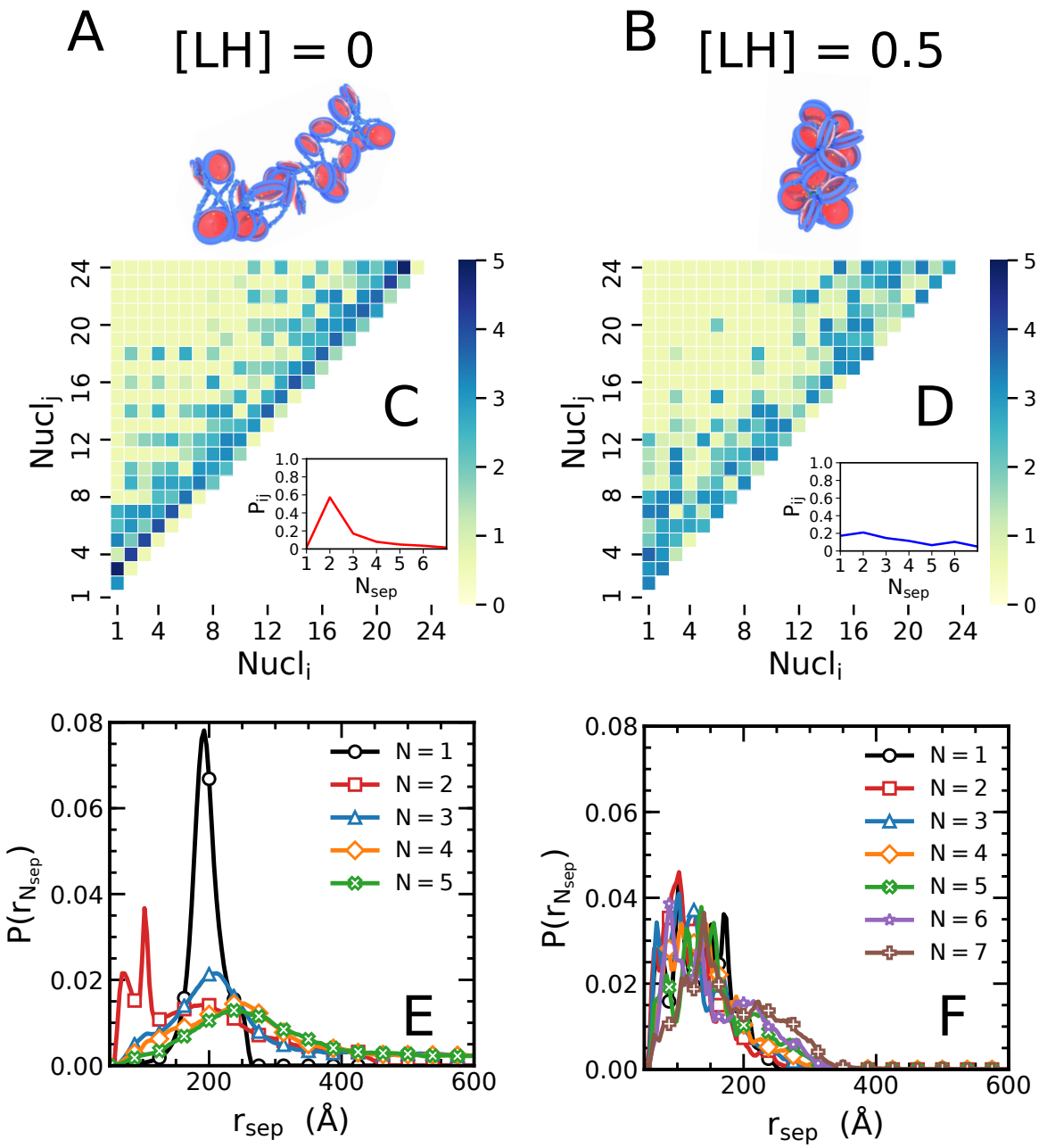

Figure 4.2: Analysis of a 24x187 chromatin fiber in terms of pairwise nucleosome separation. A) and B) Configuration of homogeneous chromatin fibers with and without the linker histone C) Contact map of nucleosomes in a fiber for a LH-free system. The color bars correspond to a log-scale of the number of contacts from the simulations. The inset highlights the probability of nucleosome contacts separated by N nucleosomes. D) Contact map of nucleosomes in a fiber with the LH present. E) Probability distribution of nucleosome contact pairs as a function of their separation distance without the LH. F) Probability distribution of nucleosome contact pairs as a function of their separation distance with the LH.

In Figure 4.2 D and F, we apply the same analysis to chromatin with a one half fraction of bound linker histones. Based on experimental data, the chromatin fiber is saturated when it has a one-half binding of LHs, which we discuss further in the Supporting Information. [130] The role of the linker histone on larger chromatin structures is reflected in Fig. 4.2D. We note that the dominance of 1-3 interactions is lost in the presence of the LH, suggesting a more condensed motif, reminiscent of the results of the pair potentials. In the inset, we also note that these contacts have a similar order of magnitude than the contacts for larger nuclesome separations as seen above. The probability curves in Fig. 4.2F show pronounced difference from the results for LH-free fibers. Up to a seven nucleosome separation, the probability distributions are the same for the separation of any two nucleosomes. These distributions are centered around 100Å, which further supports our proposition that charges on the LH compensate the energetic penalty associated with DNA bending, and nucleosome interactions dominate. In this case we see an effective "crowding" event of nucleosomes with low separation distances between any nucleosomes in the fiber. From Fig. 4.2F we see all distributions centered at ~ 100Å, which we attribute to an entropically favorable configuration of crowded nucleosomes. From both of these results, it is evident that a dinucleosome motif alone is insufficient to describe the chromatin fiber structure.

To gain a better understanding of the fiber structure, we repeat our process from above, but now considering the energetics of the three nucleosome system. In Figure 4.3, we show the free energy surface of the trinucleosome system with NRLs of 187 both with and without the LH. The free energy surface without the linker histone is shown in Figs. 4.3A and B. All panels on the left highlight the free energy curves at specific r_α (separation of the first and second nucleosome), corresponding to the wells found in the dinucleosome system. From the two-dimensional free energy surface in panel B, we come to the finding that the trinucleosome has a single global minimum, with no indication of meta-stable states when $r_\alpha = 180$Å and $r_\beta = 70$Å. Compared to the two candidate 30-nm fiber structures, our results suggest that

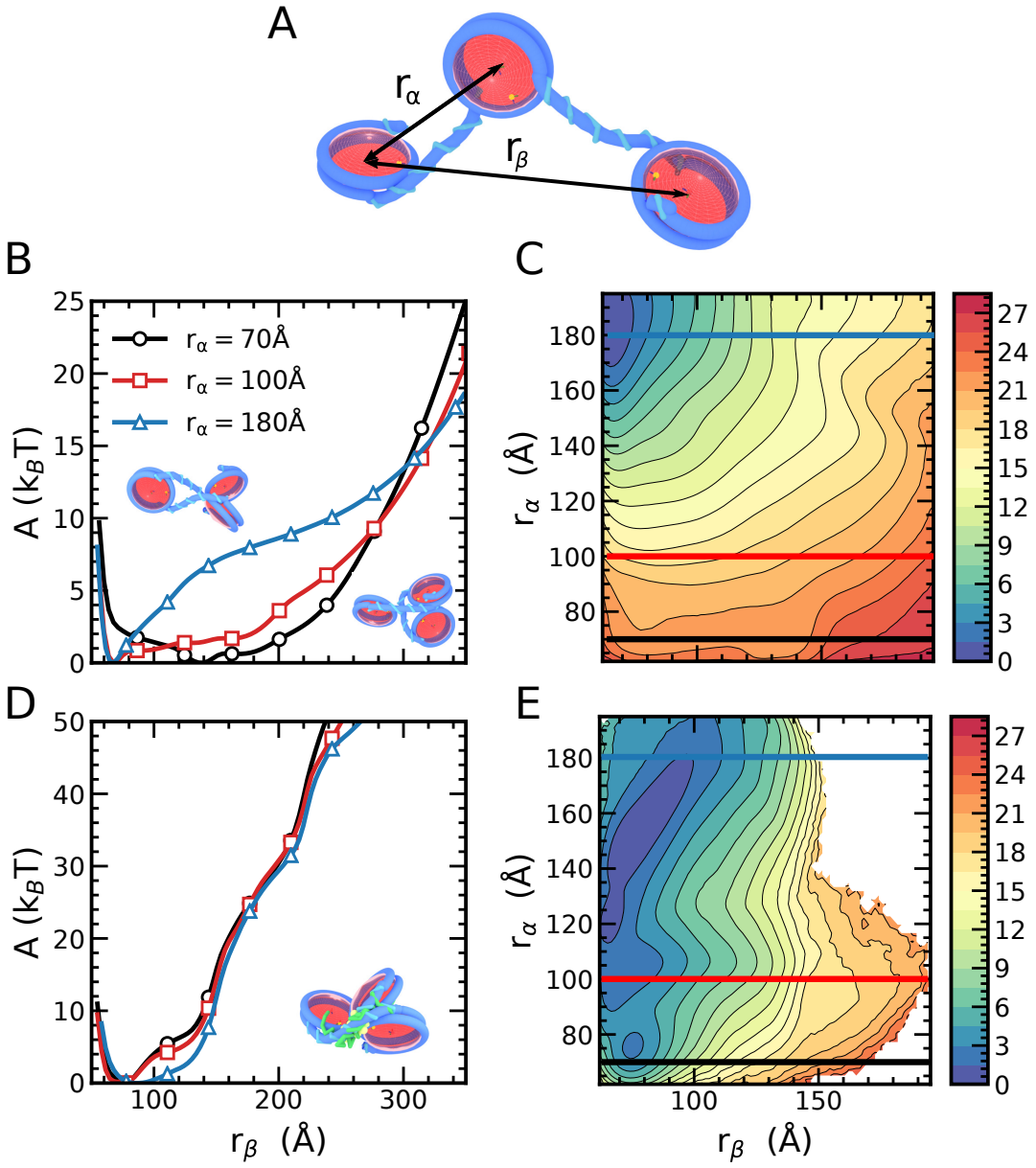

Figure 4.3: Trinucleosome free energy surface. A) Schematic of the system, where r_α is the separation between the first two nucleosomes and r_β is the separation between the first and third. The plots in the left column highlights the free energy curves when r_α is at one of the minima extracted from the dinucleosome results for B) no LH and D) with LH. The second column is the two-dimensional free energy surface in units of k_BT of the trinucleosome C) without the LH and E) with the LH.

the zig-zag has a lower free energy in the absence of the LH and with homogeneous NRL. This result is consistent with experimental evidence, which mentions the need for either the linker histone or divalent salt to form condensed and ordered solenoid-like structures.

Trinucleosomes with the LH exhibit a very different free energy surface. In contrast to the previous result, the trinucleosome free energy surface with LHs shows a wider range of separation distances over which it exhibits a minimum, consistent with our prior statement that nucleosome interactions dominate in the presence of the linker histone. We also find a metastable state associated with a condensed, solenoid-like structure at a separation distance of $r_\alpha = 75$ and $r_\beta = 75$. Comparing the trinucleosome and dinucleosome free energy surfaces, it is clear that the larger fiber should be described in terms of a combination of trinucleosome motifs and we return to the prior REMD simulations with the trinucleosome motif.

We next interpret the structure of the larger chromatin fibers in terms of the trinucleosome free energy surfaces. We construct probability distributions of all 1-3 separations (r_β) given the distance of the 1-2 separations (r_α). We show the conditional and joint probability results in Figure 4.4. The top and bottom rows correspond to data without and with LHs, respectively. We note that the free energy approximation that we generate by $-\log(p_{ij})$ has features similar to those seen in the trinucleosome results, including the global minimum at the largest extension of r_α and low extensions of r_β. The primary difference is that the global minimum occurs at an $r_\beta = 100$Å as opposed to 70Å. As the fiber consists of multiple nucleosomes, we propose that the "side-by-side" configuration is more entropically favorable. We also note that, in the free energy surface in the presence of the linker histone (Fig. 4.4D), there is an equal probability of forming a solenoid-like or zigzag-like structure. Our results here are consistent with the theoretical notion that the solenoid 30-nm fiber is formed only in the presence of the LH.

In this last section we consider the effects of acetylated H4 tails on the structure distribution of the chromatin fiber. Previous work has presented the ability of the 1CPN

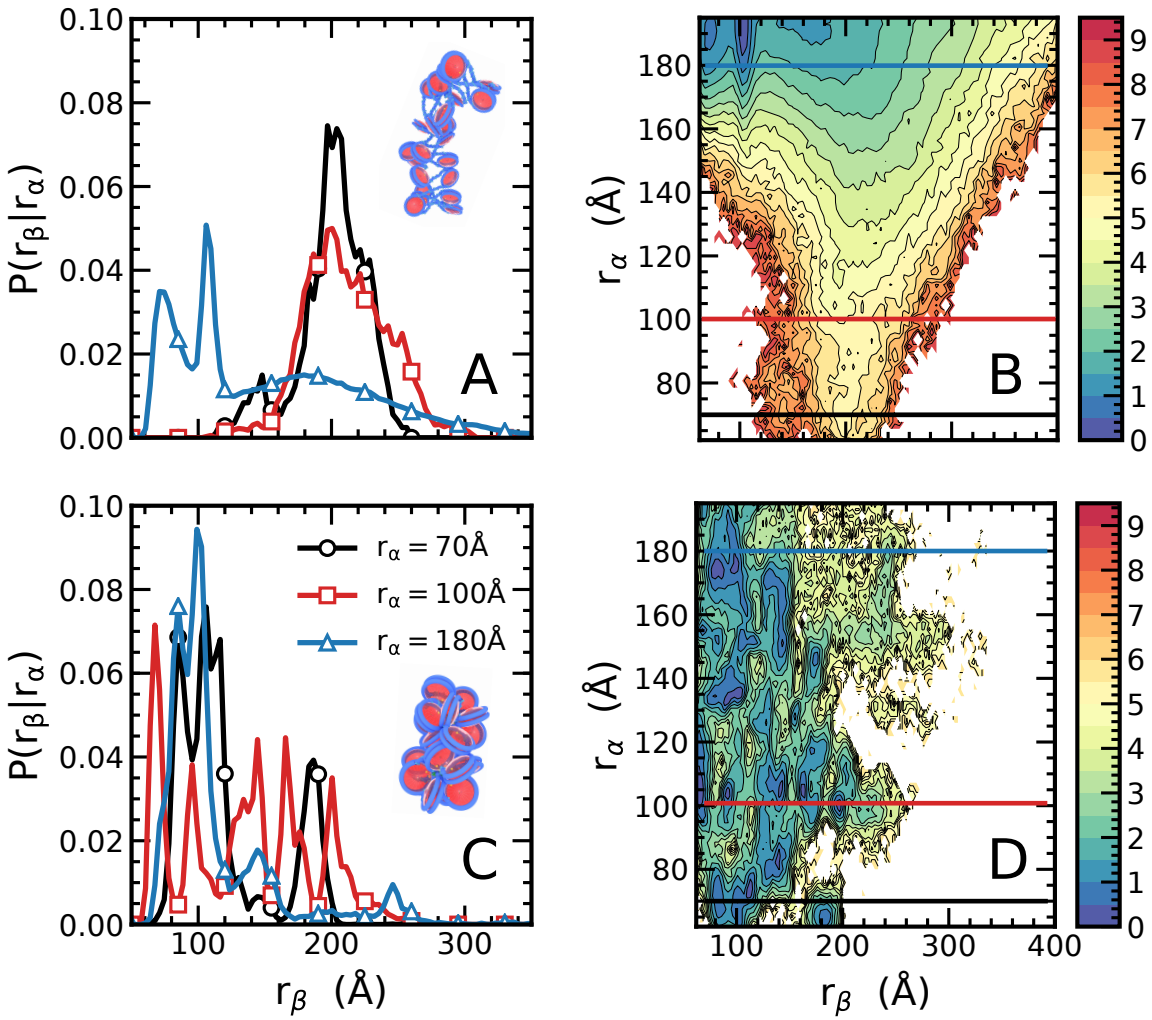

Figure 4.4: Analysis of the chromatin fiber using the joint probability distribution of r_α and r_β. A) Schematic of the system, where r_α is the separation between the first two nucleosomes and r_β is the separation between the first and third. The left column highlights the free energy curves when r_α is at one of the minima extracted from the dinucleosome results for B) no LH and D) with LH. The right column is the two-dimensional log of the histogram of states in units of k_BT of the trinucleosome C) without the LH and E) with the LH.

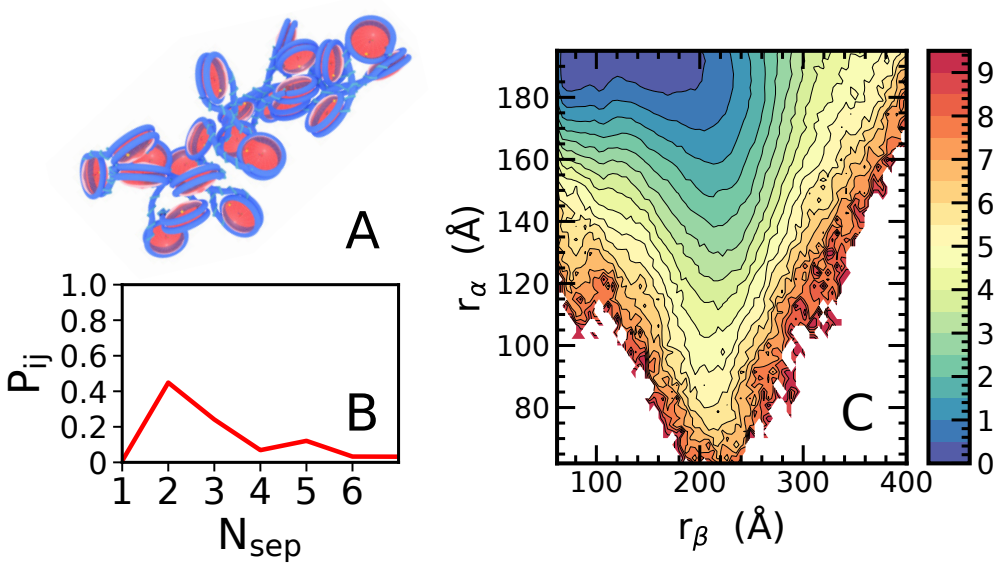

Figure 4.5: The fluid behavior of acetylated H4-tail chromatin fibers. A) A snapshot of the fiber shows a more disordered fiber. B) The probability of nucleosome contacts separated by N nucleosomes shows a decreases in the $N_{sep} = 2$ contacts. C) The free energy of an acetylated chromatin fiber.

model to capture the role of acetylation of the H4 tails in the form of a modified pairwise interaction.[76, 101] We first discuss the effect of acetyl-H4 on the trinucleosome, which we predict to be a fundamental unit of the chromatin fiber (in light of our previous results). The results included in the Supplementary Information indicate that there is no free energy difference for the trinucleosomes. In a similar manner, we run acetylated-H4 fibers with homogeneous 187 NRL and 24 nucleosomes with REMD. In contrast to the unmodified fiber, the acetylated fiber displays significant differences in the joint probability distribution when compared to its corresponding trinucleosome free energy surface (compare Fig. 4.2B with Fig. 4.5C). Instead of a small distribution of states where the fiber is stable, the region grows considerably along the r_β axis.

This result go against our prior understanding of acetylated fibers. There is a known correlation between acetylated H4-tails and transcriptionally-active chromatin.[152, 26, 75] Chromatin is theorized to extend upon H4-acetylation to facilitate transcription factor bind-

ing, serving to explain why H4-acetylated chromatin is associated with transcriptionally-active chromatin. [31, 163] Our results suggest a different mechanism. Previous work suggests that nucleosomes with acetylations to the H4 do not repel each other.[101] As there are no other factors that would introduce additional repulsive forces, we find no reason to expect the fiber to elongate. In our simulations, we find the chromatin fiber to be more fluid in the presence of acetylations. The contact probabilities of the nucleosomes in the fiber show a $\sim 20\%$ decrease in $N_{sep} = 2$ contacts (Fig. 4.5B), compared to the unmodified case (Fig. 4.2). As the previously dominant trinucleosome contacts are significantly decreased, the fiber structure becomes more compliant. We therefore expect chromatin to be malleable to external factors such as RNA polymerases or heterogeneous linker DNA sequences that are present in the cell nucleus. These external factors may play a larger role on the correlation between transcription and acetyl-H4 tails than previously theorized.

4.5 Conclusion

In this work we have identified several key structural features of the chromatin fiber using smaller fibers as a template. First, we have shown that smaller fibers are particularly sensitive to NRL and the inclusion of the linker histone. Consistent with previous experiments, we find these parameters to be equally important in dictating the larger chromatin structure. Second, we have shown that the trinucleosome system is a good predictor of the approximate free energy surface for larger chromatin fibers in the homogeneous case. Identifying a building block beyond the nucleosome has been a challenge, given the dynamic nature of chromatin and the diffraction limit of imaging equipment. Our results here support the "nucleosome clutch" model that has recently been proposed for cell pluripotency. Although the trinucleosome is smaller than the tetranucleosome repeat *in vivo*, it is consistent with evidence of a building block beyond the nucleosome, and with the notion that a building block is "encoded" from the energetics of DNA and nucleosomes.

The trinucleosome building block uncovered in this work represents a stepping stone towards understanding larger chromatin structures. To arrive at a comprehensive picture, however, the influence of additional epigenomic phenomena on chromatin structure must also be considered. Recent theoretical work provides evidence that heterogeneity of linker lengths greatly influences the long-ranged looping potential of chromatin. [16] Our own results, and prior microscopy, have also shown that the amount of linking DNA significantly alters chromatin structure; [130] here, we only considered homogeneous linker-lengths, and within that constraint we were able to identify a well-defined correlation between the trinucleosome building block free energy and chromatin fiber structure. In nuclear chromatin, nucleosomes are dynamic and diffuse along the DNA in a "sliding" fashion, resulting in highly heterogeneous or broad linker length distributions. [78, 89, 98] The proposed "nucleosome clutch" and "tetranucleosome" models may be a result of longer range correlations through the joint effects of heterogeneous linker-lengths and linker histone binding. While considering every combination of linker lengths is computationally demanding with the 1CPN model, the trinucleosome building block introduced here could provide a tractable alternative. Experiments, such as Micro-C and MNase-seq, currently seek to uncover the distribution of linker lengths *in vivo*. [29, 61, 60] A combination of the trinucleosome physics with experimentally determined linker length distributions would reveal with more clarity the extent to which the trinucleosome building block influences chromatin structure. We leave such a study of heterogeneous linker lengths on chromatin structure to future work.

4.6 Supporting Information

4.6.1 Linker Histone Saturation

The work of Routh et al demonstrates that when there are effectively 1 linker histone for every 2 nucleosomes, the fiber is saturated and at its peak of compaction. [130] The same ratio is

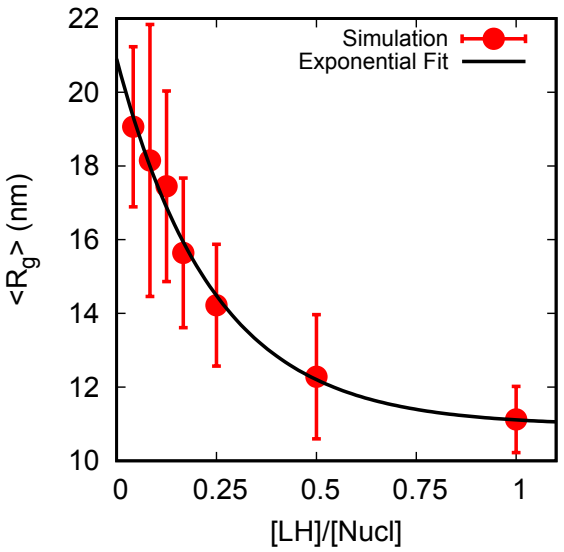

Figure 4.6: The chromatin fiber is further compacted with increasing relative amounts of linker histones. The fiber becomes saturated around a ratio of one linker histone for every two nucleosomes, consistent with experimental claims.

used in this work for the full fiber with linker histone free energy surface. To prove that this ratio is at the saturation point, we run long replica-exchange molecular dynamics simulations of 24 nucleosomes in the 1CPN model with a 187 repeat length for varying amounts of linker histones. We display the results of the average radius of gyration, $\langle R_g \rangle$, as a function of the amount of linker histones in Figure 4.6. From the figure, one can see that at a ratio of 0.5, $\langle R_g \rangle$ is at the bottom of the exponential decay, consistent with experimental claim. For this reason, the assumption of one linker histone per two nucleosomes is appropriate. Recently, results show that ratios above one linker histone per nucleosome result in unique compactions and structures of chromatin, which is a future direction of this work.

4.6.2 Comparison of multiple NRLs

While we primarily focus on the 187 NRL for this manuscript, simulations of variable linker lengths are reported for the 24 nucleosome simulations. We focus on the 187 NRL as it displays unique physical behavior in the dinucleosome case. This result is pushed towards

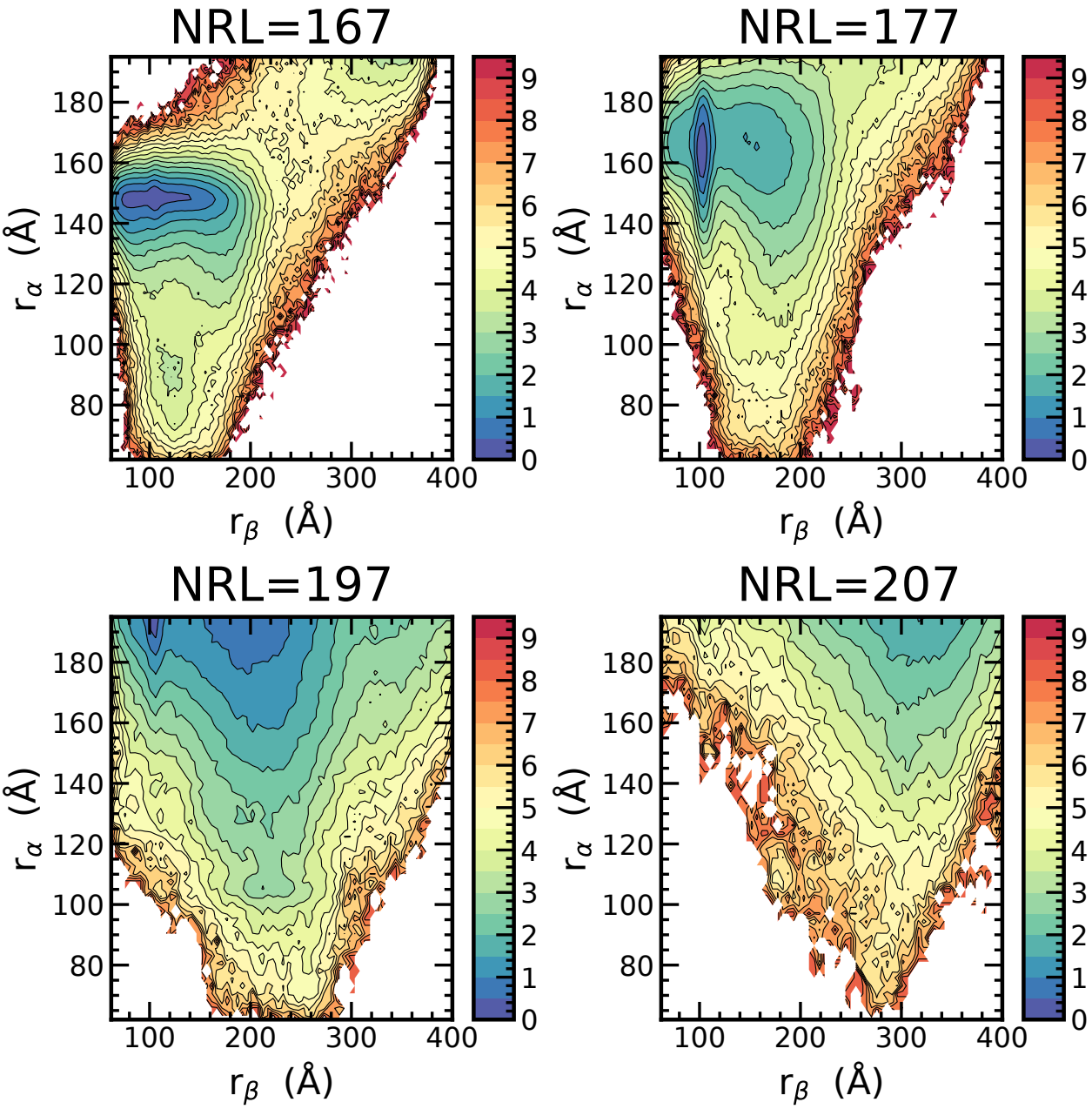

Figure 4.7: The histogram of states for fibers with varying repeat lengths.

larger systems in an attempt to reconcile the observed differences of the 30-nm fiber. As *in vitro* chromatin experiments reconstitute chromatin using even pitches of DNA, $\sim$ 10 bps, we consider fibers with similar repeat lengths. The data for the histogram of states is shown in Figure 4.7, with the NRLs displayed above each surface. These figures are generated in the same fashion as the 187 NRL fiber in the manuscript. Interestingly, we note that each surface displays unique minimum, similar to the results of the dinucleosome free energy surfaces. We claim that each repeat length influences the thermodynamics and structure of the chromatin fiber at this scale, which is of considerable interest. In future work, we plan to investigate and characterize this phenomena further.

4.6.3 Acetylation of the H4 Tail

In this work, we show that the fiber is fluid in the presence of H4K16ac, which allows for external factors to more easily disrupt the fiber structure. Additionally, this work reveals that the trinucleosome is a viable fundamental unit that describes the thermodynamics of chromatin fibers. For completeness, we show that a similar claim can be be drawn for the trinucleosome free energy surface of H4K16ac. Using the same procedure of umbrella sampling with the weighted histogram analysis method, we compute the trinucleosome free energy surface with H4K16ac and show the results in Figure 4.8. Both the acetylated and unmodified, or "wild-type" trinucleosome surfaces are shown for comparison. Unsurprisingly, we see very little change between the two free energy surfaces, consistent with what is claimed in the work.

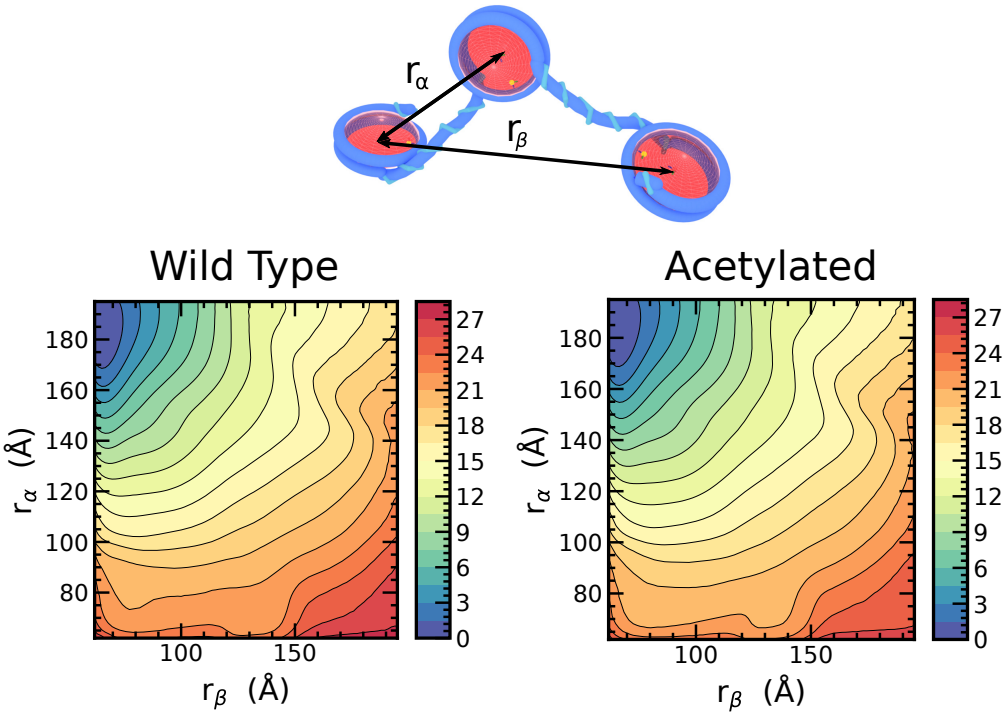

Figure 4.8: Comparison of the trinucleosome free energy surface for unmodified chromatin and H4K16ac chromatin.

CHAPTER 5

EQUILIBRIUM FEATURES OF THE CHROMATIN FIBER PREDICT TAD-LIKE STRUCTURES

5.1 Abstract

Chromatin structure uniquely facilitates the packaging of DNA through a combination of DNA-protein interactions and epigenetic modifications. As the structure of chromatin is inherently linked to the transcriptome and other essential biological processes, it is necessary to fundamentally understand the structure of chromatin and the factors that alter it. Chromatin is often treated as a model polymer wherein nucleosome-monomers are bonded via stiff DNA segments. While such models can provide a coarse treatment of the overall chromatin structure, they potentially sacrifice key physics at the nucleosome-level. A key question that arises is what is the fundamental unit for the structure of chromatin? In this work, we introduce the ideal chromatin chain model (ICCM), which provides a computationally efficient model of chromatin structure with high resolution. Using this model, which is derived from data from coarse-grained molecular dynamics simulations, we provide evidence that a trinucleosome repeat unit serves as a robust fundamental description of chromatin and that variable linker lengths have unique statistical segment lengths that span 2-10 kbps. We further show that the ICCM predicts topologically associating domain structures on the order of 1 Mbps, but when averaged over the population, such structures are not detectable and agree with recent experimental results. Through the ICCM, we provide a comprehensive statistical framework that captures local chromatin structure and can be extended to predict its effect on global chromatin organization.

5.2 Introduction

The genome is compacted in the cell nucleus through a complex and dynamic network of protein-DNA interactions. Host to this network is the chromatin fiber, which succeeds in packaging the genome through its hierarchical structure. For reference, a packaged genome requires folding 3 billion bp of DNA that is 3 meters in length into chromosomes that are $\sim 1\mu$m in size. In addition to the demanding structural constraints, the chromatin fiber is also central to biological processes such as transcription, DNA replication, and repair, which require local unwinding of chromatin. [54, 2, 29, 149, 134] The structural basis of the chromatin fiber is the nucleosome, a stable compound of DNA and proteins. From X-ray crystallography, the structure of the nucleosome was determined to be 147 bp of DNA wrapped $\sim$1.7 times in a left-handed superhelix about a histone octomer core. [87, 34] Structural motifs of the fiber beyond the nucleosome are fluid and, therefore, harder to discern using conventional methods. [93]

Despite its dynamic nature, numerous experimental breakthroughs have illuminated structural features of the chromatin fiber. In particular, electron microscopy techniques such as Cryo-EM provide significant evidence for repeating structural motifs of chromatin.[138] Recent advancements have combined multi-tilt scanning-transmission electron microscopy with a high contrast dye, known as ChromEM, to elucidate structural features of chromatin *in situ*. [112, 83] Less intrusive methods using super-resolution microscopy techniques, such as stochastic optical reconstruction microscopy (STORM), have also probed chromatin structure and organization. For example, recent STORM experiments found evidence of a smaller building block of chromatin, known as nucleosome clutches, whose size is related to cell pluripotency.[124]

With the advent of chromosome conformation capture (3C) techniques, key structural features of chromatin have been revealed.[84] In particular, the work of Dixon *et al* and Rao *et al* demonstrated the prominence of regions of high contact that are conserved across

cells referred to as topologically associating domains (TADs).[37, 123] TAD stability has been further implicated in oncogenesis.[150] Another feature, chromatin loops, form between points of high-contact between gene enhancers and promoters. Upon removal of the molecular motor cohesin, Hi-C maps show the disappearance of TADs and loops leading to an increase in chromatin compartmentalization,[135, 122] which is evidence of the epigenome influencing chromatin structure.

Of course, chromatin structure is known to be altered through a myriad of epigenomic phenomena.[139, 64] For example, Hi-C and chromatin immunoprecipitation sequencing (ChiP-Seq) reveal that chromatin elements with similar epigenetic marks aggregate. [42, 65] Recent advances in 3C techniques have demonstrated that population-based results of Hi-C can be determined from the single-cell level. [18] Another advance through a novel combination of Hi-C with orientation mapping reveals tetranucleosome structural motifs are prevalent in the genome,[109] which suggests a fundamental unit of chromatin larger than the nucleosome.

While experiments paint broad strokes of the chromatin structure and its influence by epigenetic phenomena, modeling has the potential to depict a much higher resolution image. For example, Fudenberg and Nuebler recently proposed a model by which chromatin loops form through an active mechanism of chromatin extrusion, increasing short ranged contact probability. [49, 108] Studies of how epigenetic signaling influences chromatin structure are also now emerging. [97] Notably, di Pierro *et al* incorporated epigenetic signaling via maximum entropy minimization to reproduce contact maps with high fidelity.[36, 35] Other modeling efforts by MacPherson *et al* have employed a mean-field representation of chromatin as a block-copolymer to understand compartmentalization as a result of histone tail H3 methylation and its effect on reproducing Hi-C contact maps.[90, 91]

The application of polymer physics ideals to chromatin[21, 32] is an appealing approach that would enable facile investigation of factors influencing the packaging and connectivity

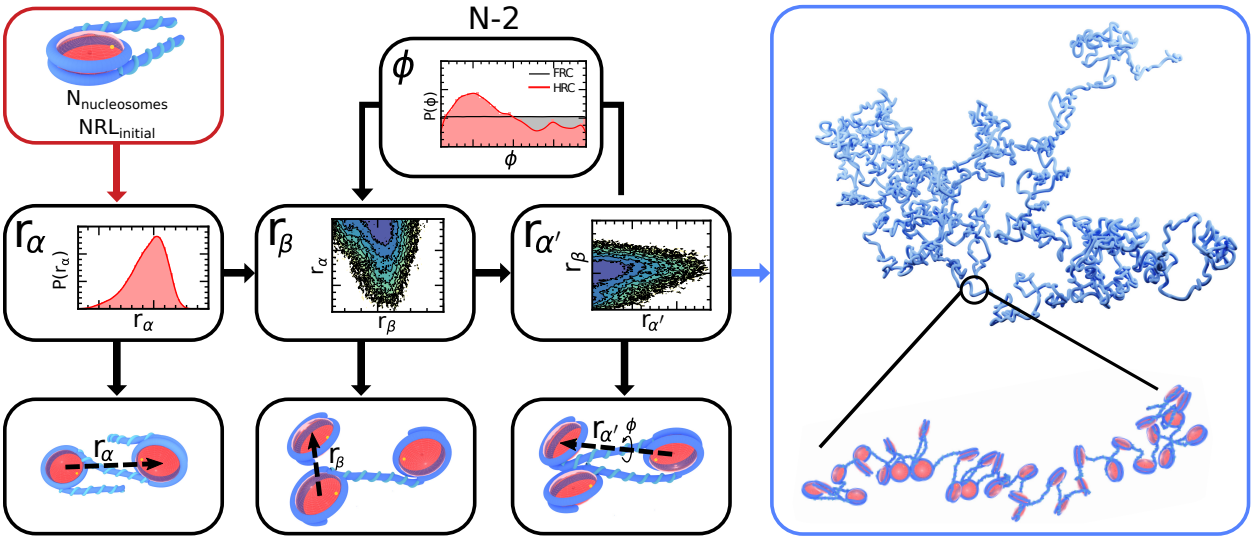

Figure 5.1: Flow diagram of the ideal chromatin chain model. In the top left, a single monomer is chosen with a specific linker length. That linker length is then used to draw a value of r_α to grow the first link, which is shown in the bottom left. The value of r_α is now used to draw from the conditional distribution to get a value of r_β. To geometrically complete the trimer a value of r'_α is drawn from a second conditional distribution. To grow the chain out, trimers are repeatedly grown for N-2 steps, where N is the desired number of monomers in the final chain. If a tetranucleosome motif is chosen, the dihedral distribution of chromatin is drawn from, otherwise we assume a uniformly distributed dihedral distribution. The output chain is shown to the right.

of the genome.[9, 129] From a polymer physics perspective, understanding the behavior of a fundamental physical unit and the interactions of those units is often sufficient to determine global-chain properties. For example, the rotational isomeric state model of Paul Flory[46] utilizes local, discretized dihedral distributions to fully describe the conformational characteristics of the entire chain; the main idea being that propagation of local phenomena on the scale of fundamental units determines properties on the scale of the entire chain. Similarly, it would be useful to find a building block of chromatin and propagate the physics of that building block and its interactions to determine a comprehensive, thermodynamic picture of chromatin structure and functionality.

In this work, we introduce the Ideal Chromatin Chain Model (ICCM) to describe global chromatin structure. Previously, we have employed mesocsale models, such as the one-cylinder per nucleosome model or 1CPN, to characterize the structural behavior of chromatin[76] and ultimately reveal that mesoscale chromatin structure heavily features a trinucleosome motif;[100] additional simulations have also characterized the trinucleosome unit. [28] Building upon this work, we develop ICCM and use it to elucidate the effect of varying linker length on the equilibrium chromatin structure for fundamental units of two, three, and four nucleosomes. This investigation supports previous hypotheses on the importance of the trinucleosome motif to chromatin structure. Moreover, the model predicts that variable linker length has a noticeable effect on the efficiency of packing DNA at the Mbp scale. Additionally, we find that distributions of linker lengths have a unique signature on the packaging of chromatin. Lastly, we find good agreement with recent single-cell Hi-C experiments that show that TADs exist on the single-cell but still show TAD-like features upon removal of cohesin. Based on these results, we predict that TADs are an equilibrium property of the chromatin fiber and that molecular motors and epigenetics influence the stability of their genomic address.

5.3 Methods

The work in this manuscript relies on ideal chain growth software written in-house. To generate appropriate distributions of polymers, chains were grown according to the partition function, Z.

$$Z = P(r_\alpha; \ell_{NRL}) * \prod_{i=1}^{N} \Big[P(\ell_{NRL}) P(r_\beta | r_\alpha; \ell_{NRL}) P(r_\alpha^{'} | r_\beta; \ell_{NRL}) P(\phi; \ell_{NRL}) \Big] \quad (5.1)$$

5.3.1 Model Parameters

Here, the trimer is decomposed into a definition of $\mathrm{r}_\alpha, \mathrm{r}_\beta, \mathrm{and} \mathrm{r}_\alpha^{'}$, where r_α is the separation distance between the first two neighboring nucleosomes in a trimer, r_β is the separation distance between the first and third nucleosomes in a trimer, and $\mathrm{r}_\alpha^{'}$ is the separating distance between the second and the third. In previous work, it was established that such a decomposition was an appropriate estimate for large fibers. [99] We extend upon this previous work by allowing for a definition of the trimer decomposition for a range of linker DNA lengths in the operating range of 10-60 bps, comparable with the range of lengths in human chromatin. Such a decomposition is analagous to the freely-rotating-chain (FRC) model of ideal polymers, however here we allow for unique functionals of each amount of linking DNA. To demonstrate the unique signature of each repeat length, we provide the free energy surface of each trimer decomposition in Section 5.6.1

To generate a robust framework which accounts for more than just trinucleosome decomposition, the additional term of $\mathrm{P}(\phi)$ is included. The functional form of dihedrals for each fiber is provided in Section 5.6.1. By doing so, we add in an energy of dihedrals along the fiber. This term can be considered as introducing information of a tetranucleosome fundamental unit, akin to the hindered-rotation chain model (HRC) from polymer physics. The trinucleosome approximation from Moller *et al* , is a convenient explanation, but perhaps

higher order organazation of chromatin requires a larger fundamental unit. In this manner, we are not only interested in finding the fundamental unit that best describes large-scale chromatin, but also the unit which provides the most marginal information per level of description.

Simulations of the coarse-grained molecular dynamics 1-cylinder-per-nucleosome (1CPN) model were used in order to calculate both the joint probability distribution, $\mathrm{P}(\mathrm{r}_\alpha, \mathrm{r}_\beta, \mathrm{r}'_\alpha; \ell_{\mathrm{NRL}})$ and $\mathrm{P}(\phi; \ell_{\mathrm{NRL}})$ for each linker length. The 1CPN model was developed with the rich nucleosome physics in mind. We refer the reader to the original paper for a full description of the development and validation of the 1CPN model. [76] The 1CPN model feasibly captures chromatin fiber physics of appreciable length (order of kbps). While kbp simulations of the model are feasible, in the characterization of the 1CPN model, it was discovered that the free energy surface of short fibers reduced into the r_β and r_α order parameters displayed similar features to a trinucleosome free energy surface. In order to extend this work to the operating range of linker lengths of multiple organisms' genomes, multiple simulations of larger fibers were performed. To get sufficient statistics for chromatin fibers, for each linker length, 3 replicas simulations of 24 nucleosomes of constant repeat length are run with replica exchange molecular dynamics for 10^9 timesteps. At a timestep of 60 fs, this results in simulations of 60 μs each. These results are then broken down into histograms of states of r_β, r_α, and r'_α. Similarily, the distribution of dihedrals, ϕ, are calculated from each set of simulations. As this is an extensive dataset, we believe that there is a significant amount of information encoded in the simulations that we have not yet uncovered. The data is available upon request.

For the purpose of this work, the linker length distribution $\mathrm{P}(\ell_{\mathrm{NRL}})$, is specified according to multiple theoretical and experimental considerations. As each linker length simulation results in unique trimer free energy surfaces, linker length should have a considerable impact on the resulting fiber. In that manner, distributions of linker length should also result

in unique fiber properties. Since nucleosome repositioning through nucleosome sliding and dissociation/association results in heterogeneous links, distributions of linker lengths are important.[78, 29, 141] If not specified, $P(\ell_{NRL})$, is a delta function, $\delta_{\ell=NRL}$ around the specified repeat length. Otherwise, we use distributions from statistical work of Jonathan Widom, experimental work of Voong *et al* , and a distribution from recent theoretical work. [156, 153, 16] The probability distribution functions for repeat lengths used in this work are plotted in Figure 5.4A.

5.3.2 Chain Growth

Chain growth steps are proposed according to the partition function in Equation 5.1. Figure 5.1 highlights the chain growth process graphically. A step is initially proposed by sampling a linker length from P(NRL). For uniform linker length growths, the probability distribution, P(NRL), is defined as the delta function centered around the chosen repeat length, otherwise the distribution is defined from the desired distribution function. After the initial growth step, the first trimer step is chosen. This involves sampling a distance from the conditional distribution, $P(r_\beta|r_\alpha)$, where r_β is the distance between the first and the third monomer. To solve the trimer geometrically, the third side of the triangle, $r_\alpha^{'}$ is chosen from the conditional distribution, $P(r'_\alpha|r_\beta)$. As a result of this process, we are left with a "ring" of degeneracy of possible chain placements; the overlap of the surface of two spheres of radius r'_α and r_β is a ring (if they do overlap). To choose the position along this ring is the difference between the FRC model and the HRC model. In the FRC case, we use the "uniform ring approximation," implying that a point is uniformally chosen on the ring. In the HRC case, the orientation on the ring is selected from the appropriate dihedral distribution, $P(\phi;\ell_{NRL})$. After this step, the process is repeated for N-2 trimer growth steps, where each r'_α becomes r_α of the next trimer. It is important to note that this leads to an inherent bias in growth steps when distributions of repeat lengths are chosen. As a result, we consider the initial repeat length

distributions to be a prior for the resulting distribution from multiple chain steps.

As an added parameter, we also investigate the influence of the linker histone. In previous work, the linker histone introduced significant compaction of the chromatin fiber.[88, 76] For this work, similar distributions are calculated from long 24 nucleosome replica exchange molecular dynamics simulations. In order to simplify the probability distribution, we assume that all sets of $\{\mathrm{r}_\alpha, \mathrm{r}_\beta\}$ are equally probable below a linker-length specific threshold.

To sample physical chains, not all proposed chain growth steps are accepted. The process of rejecting chain moves falls into one of the following categories: two proposed points are within the excluded volume range,

$$U_{EV} = \begin{cases} \infty & r \leq 60 \\ 0 & r > 60 \end{cases} \tag{5.2}$$

where r is the separation between any two nucleosomes, the probability that an r_β conditioned on r_α is low, or the spheres of radius r_β and r'_α do not overlap. In this event, another chain step growth is attempted. If 100 proposals fail, the chain is attempted to regrow from the point where the majority of chain steps are failing. Here we note that ultimately we are interested in a physical solution to the partition function and as such, we are looking for fibers that are solutions and are not growing explicitly through Monte Carlo chain growth steps. Additionally, there are no long-range interactions included in the model beyond the excluded volume. Since we are primarily interested in how linker lengths influence the long-range interactions and packaging of chromatin, we exclude specific long-range interactions as they are unnecessary for the objectives of this work.

5.3.3 Self-avoiding walk generation

To compare the information provided by the trimer information to a less-informed model, we generate self-avoiding walk data with distances between monomers sampled from $\mathrm{P}(\mathrm{r}_\alpha; \ell_{\mathrm{NRL}})$,

akin to a freely-jointed chain model. An initial chain is generated in a straight line as an initial condition. For this version of the model, Monte-Carlo pivot moves are proposed at a random monomer. Again, the only long-range interaction is excluded volume as defined above. The chains are run until equilibration according to the radius of gyration R_g. These results are then compared against the trimer to demonstrate the marginal information between a dimer and trimer-based model.

5.4 Results

To begin, we briefly describe key aspects of the ideal chromatin chain model (ICCM). The ICCM assumes a trinucleosome fundamental unit, represented by two variables: the separation between nearest neighbor nucleosomes, r_α, and next-nearest neighbors, r_β; this decomposition is motivated by previous results demonstrating that the thermodynamics of chromatin fibers projected onto these order parameters reproduces the free energy surface of a trinucleosome with good qualitative agreement.[99] We also note the definition of a convenient variable, $r_\alpha{}' = r_\beta - r_\alpha$. The growth of a polymer chain comprised of multiple units then proceeds by sampling from probability distributions related to r_α and r_β as shown in Figure 5.1. Data for these distributions are obtained from long replica-exchange molecular dynamics simulations of 24-nucleosome fibers with DNA repeat lengths in the range of 157-207 base pairs; a full description of the molecular dynamics simulations and results obtained are provided in the Supporting Information. Alternatives to the trinucleosome as a fundamental unit are explored by elimination of r_β and inclusion of dihedral angles as an order parameter.

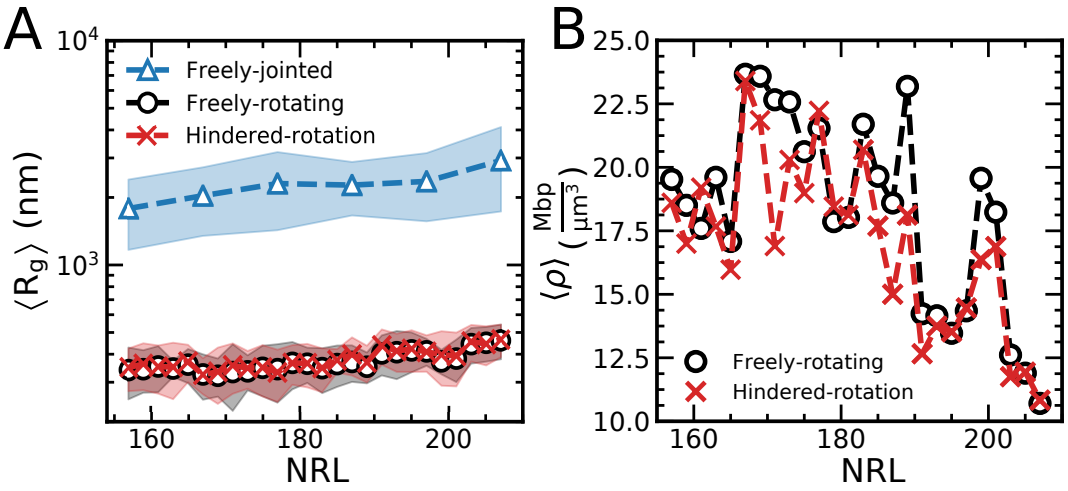

Figure 5.2: Packaging of the genome as a function of homogeneous linker-length fibers (denoted as NRL). A) The size of the average radius of gyration of the fibers, $\langle R_g^2 \rangle^{1/2}$, as a function of the fiber repeat length. The blue line is the freely-jointed chain model (FJC), equivalent to a dinucleosome fundamental unit, the black line is the freely-rotating chain model (FRC), equivalent to a trinucleosome approximation, and the red line corresponds to the hindered-rotation chain (HRC), equivalent to a tetranucleosome approximation. The shaded regions indicate statistical errors obtained from five-way block-averaging. B) The DNA density of the chains within a spherical volume with radius R_g.

5.4.1 Chromatin packaging is influenced by fundamental unit assumption and linker length

Important to chromatin is its facilitation of DNA packaging under tight constraints. Typically, approximately 100 Mbp must fit within ∼1 μm territories. In order to examine the effect that different linker lengths have on its long-range packaging, we calculate the average radius of gyration, $\langle R_g^2 \rangle^{1/2}$, as a function of repeat length. To gather appropriate statistics, 1000 replicas of 10,000 nucleosome chain growths are used for each repeat length. The results of the radius of gyration are shown in Figure 5.2A for three models: a freely-jointed chain model (FJC), which uses a dinucleosome fundamental unit (blue), a freely-rotating chain model (FRC), which uses a trinucleosome fundamental unit (black) , and a hindered-rotation chain model (HRC), which uses a tetranucleosome fundamental unit (red). Although the differences between the trinucleosome and the tetranucleosome models are small, the differ-

ences between the trinucleosome and dinucleosome models are significant. As a result, we propose the trinucleosome unit as the smallest fundamental unit with complete structural information. Consequently, presented results will be for the FRC model unless otherwise stated. It is also interesting to note that $\langle R_g^2 \rangle^{1/2}$ is not strictly monotonic with NRL, which could suggest that each repeat length imparts unique physics affecting the chromatin structure; however, this observation is not statistically significant. This result shows that the size of chains on the order of 10,000 nucleosomes is roughly independent of linker length. However, as larger repeat lengths chains contain more DNA, it also important to look at the efficiency of packaging DNA. We display the results of the DNA density, ρ, in a spherical volume element, $\hat{\mathrm{V}}$, with radius, $\langle R_g^2 \rangle^{1/2}$ in Figure 5.2B. We see overall that packing is also non-monotonic, however repeat lengths in the range of 167-187 more efficiently pack DNA than other repeat lengths. This demonstrates that variable linker lengths impact genome packaging.

5.4.2 Statistical segment length of chromatin is on the order of the length of genes

With the prevalence of 3C methods, the frequency of contact between distal chromatin loci can be calculated and quantified as the looping probability. In chromatin, the probability of looping exhibits two power-law scaling regimes: for short separations ($\leq$ 5Mbp) the scaling is ~ -0.75, whereas the scaling at larger separations is -1.8, or a polymer in good solvent.

We quantified the looping probability of the ICCM by analyzing 1000 replicas of 10,000 nucleosome chains; the results of that analysis are shown in Figure 5.3. Figure 5.3A schematically depicts how looping/contacts are identified from generated configurations; in particular, any two nucleosomes are in "contact" if their separation distance is smaller than 9 nm, which is the distance that minimizes the free energy of interaction between two nucleosomes.[101] Figure 5.3B investigates the looping probability as a function of genomic distance N at vari-

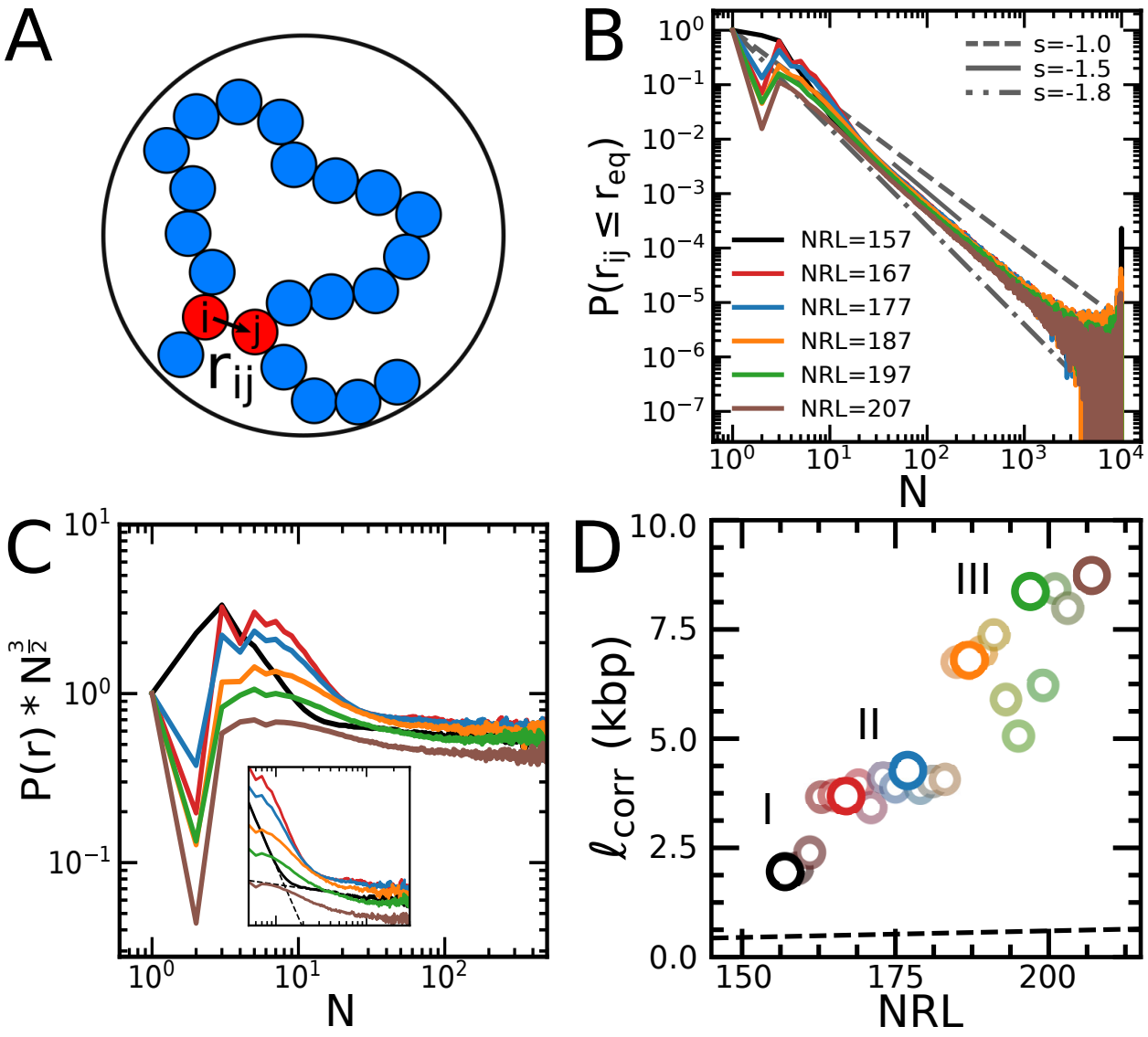

Figure 5.3: Looping probability of the ideal chromatin chain model for variable repeat lengths. A) We show a schematic for how the looping probability is evaluated. For any two monomers, i and j, separated by some distance N, we calculate the probability that they are within some distance R_{eq}, which is 9 nm for this calculation. B) The looping probability as a function of the separation distance N. The results for NRLs of even pitch of DNA are shown from 157 to 207. For reference, the scaling behavior of real chromatin, s = −1.0, a theta solvent polymer, s = −1.5, and a good-solvent polymer, s = −1.8, are shown. C) The scaled looping probability by a factor of $N^{1.5}$. The inset shows a zoomed in region to better show the inflection point and the fitting procedure to determine the statistical segment length of the chains shown in D).

ous NRL lengths, specifically for even periods of linker DNA pitch of 10 bps, which frequently occur in the genome. [156, 60, 29, 151] By comparison to various polymer physics models, one can see that the ICCM demonstrates similar scaling as that of a polymer in theta-solvent, which is expected as there is no long-range potential in the model. Of interest in this study is the separation distance between monomers required for the model to display theta-solvent behavior, which would correspond to the statistical segment length of the polymer. To more easily identify the transition in power-law scaling, we multiply the looping probability by $N^{1.5}$ and identify the emergence of any plateau-like regimes, which would correspond to the statistical correlation length, ℓ_{corr}; the results of this transformation are shown in Figure 5.3C. To quantitatively identify the onset of a plateau, we fit linear segments before and after the first inflection point following the function maximum and assign the intersection of those segments as ℓ_{corr} (see inset of Figure 5.3C). The ℓ_{corr} determined in this manner for all probed NRLs are shown in Figure 5.3D. While the precise values of ℓ_{corr} may fluctuate based on methodology, the data suggest that ℓ_{corr} generally increases with NRL (perhaps in three regimes as indicated), which is reminiscent of previous work demonstrating how the relative free energy of interaction for a dinucleosome changes with NRL.[76] Depending on NRL, we find that ℓ_{corr} ranges from $2 - 10$kbs, which is on the order of the typical length of genes. From this, we conclude that linker lengths can affect the packing and structural behavior of chromatin up to the gene scale.

5.4.3 Linker length distributions and the linker histone demonstrate chromatin density heterogeneity

Our analysis with the ICCM to this point has utilized a single NRL; however, actual chromatin is heterogeneous in linker length due to chromatin remodelers, histone chaperones, and nucleosome dissociation/association. [134, 149] To incorporate this heterogeneity, chain-generation via the ICCM is modified by drawing a linker length, growing a trimer, and

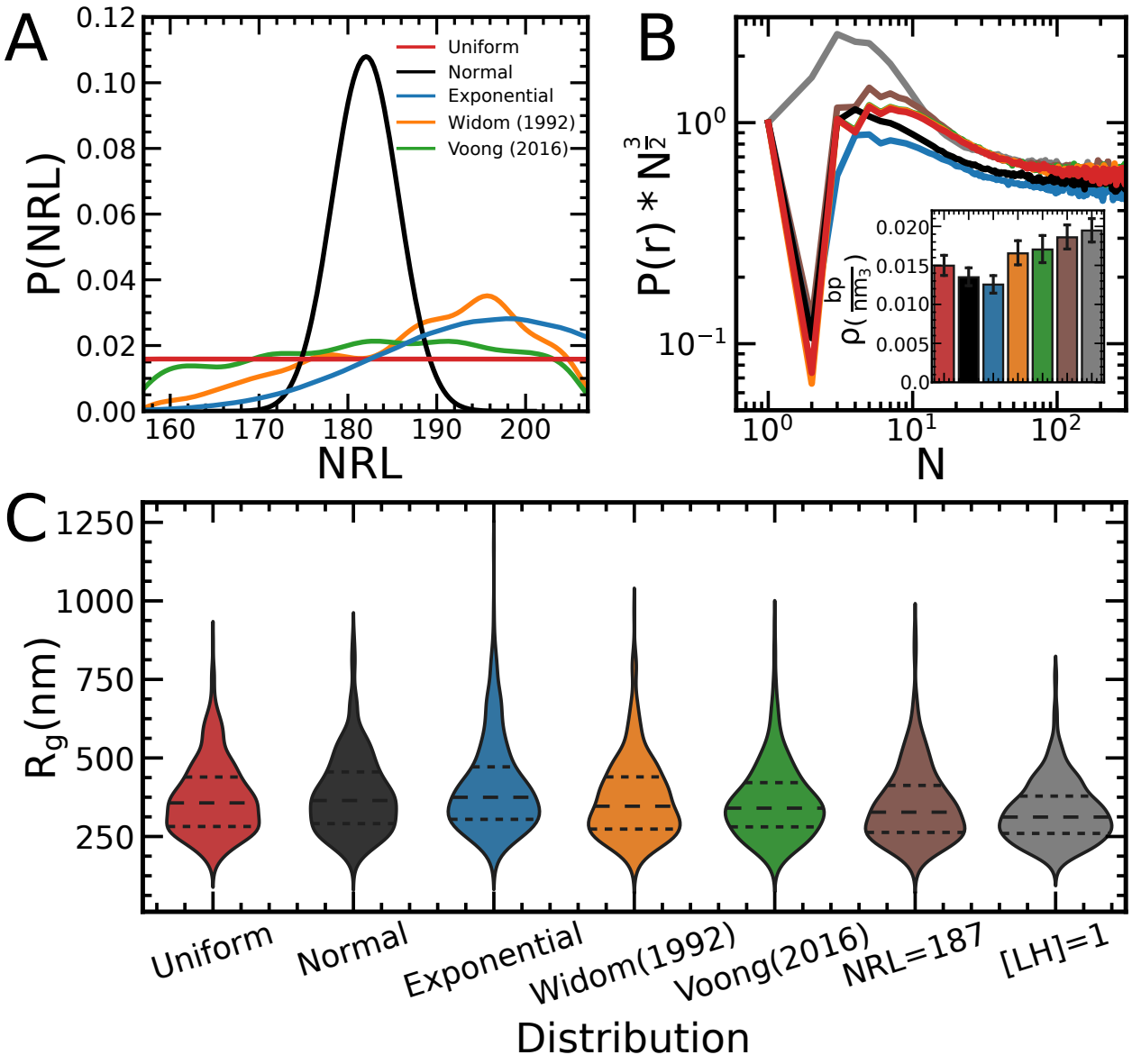

Figure 5.4: Packaging of heterogeneous linker lengths. A) The distributions of linker lengths chosen to grow chains in the ideal chain model. B) The scaling probability of the different distributions, compared against the scaling of a single repeat of 187 bp and the linker histone. The inset shows the DNA density of the different distributions. C) The radius of gyration distributions of the different heterogeneous linker length distributions.

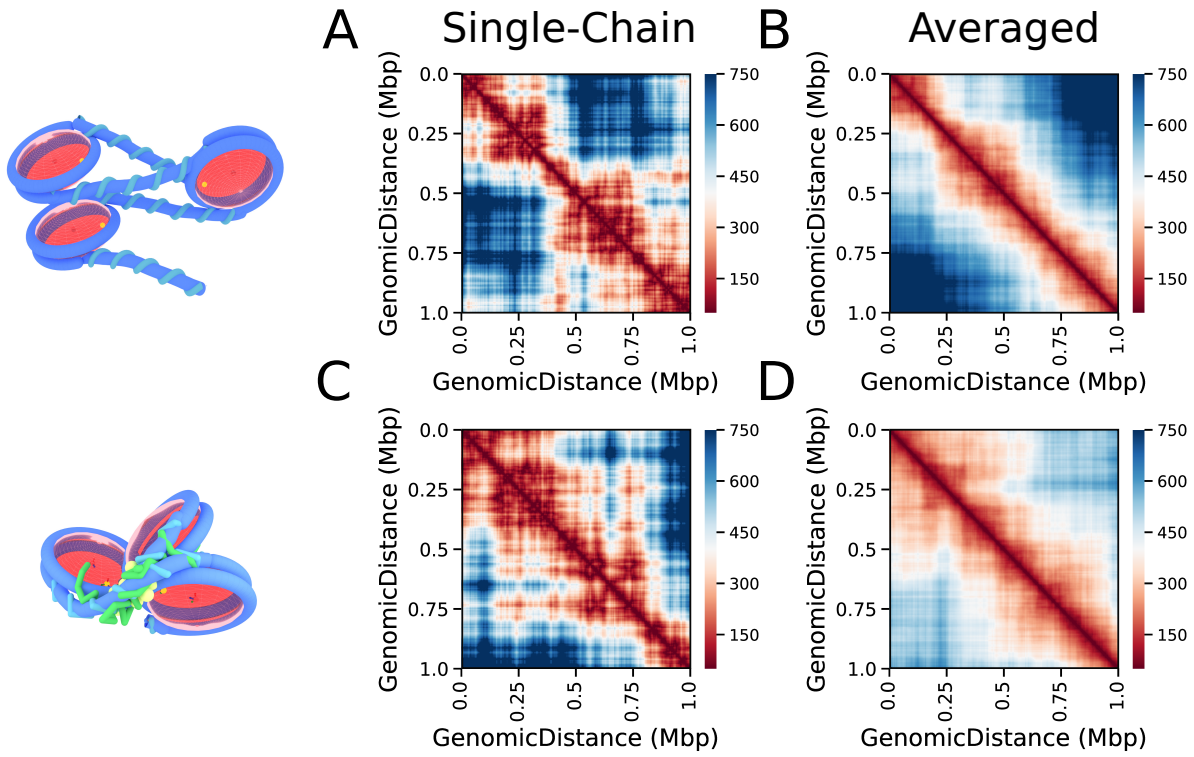

Figure 5.5: Single cell Hi-C predictions of the model in distance space. On the left, the representative fundamental units are shown to be a trinucleosome and a trinucleosome with the linker histone. A) The single chain map for a trinucleosome assumption with averaged result in B). C) The single chain map for a trinucleosome with the linker histone assumption with averaged result in D).

selecting a new linker length based on the r_α; the linker histone can be included in a similar manner. Generating chains in this manner requires knowledge of the distribution of NRLs, and we explore several possibilities ranging from basic analytic distributions, like uniform and normal distributions, to those empirically derived from experimental or statistical approaches [156, 153, 16]; the distributions employed here are shown in Figure 5.4A. It is worth noting that the NRL distribution realized via the ICCM deviates slightly from that shown in Figure 5.4A due the conditioning of the second NRL in a trimer on the trimer configuration; this small bias is demonstrated in the Supporting Information. With this modification, we examine the power-law scaling of chormatin chains with heterogeneous NRL distributions in Figure 5.4B. Interestingly, the ICCM model with the uniform distribution (red) almost perfectly tracks with the models based on empirical distributions. Meanwhile, the ICCM coupled with the normal and exponential distributions do not follow the same trends, and they are the only two distributions to not sample the full spectrum of NRLs during chain-growth. Consequently, we conclude that the relative availability, or the distribution, of NRLs influences genomic packaging. Further evidence of this is demonstrated by the inset of Figure 5.4B, which explores the packing density. While the uniform and empirical distributions yield somewhat comparable packing densities, the normal and exponential distributions result in overall less dense structures. Data for chains grown with a single, homogeneous NRL = 187 along with those with a linker histone are also shown; these latter two result in the most dense structures overall. We also show the radius of gyration distributions in the violin plots of Figure 5.4C. Overall, we see similar behavior to that from B; however, the data corresponding to the exponential distribution exhibit much greater variance. This behavior likely results from constructing chains with longer linker lengths, which are more associated with heterochromatin[149]. Relevant partial-wave spectroscopy results reveal that heterochromatin exhibits a higher density heterogeneity, consistent with what the ICCM predicts for the exponential distribution. [8]

TAD-like domains on the order of Mbp are present in equilibrium

Finally, we investigate the ability of the ICCM to recapitulate experimental observations from 3C methods. To do so, we first generate chromatin chains comprised of 5,000 nucleosomes using the ICCM for two cases: (i) a heterogeneous NRL distribution based on the Widom statistical distribution and (ii) a system of saturated linker histones, [LH] = 1.0; contact maps for additional cases are provided in the Supporting information. The resulting configurations are then analyzed to construct single cell Hi-C contact maps as shown in Figure 5.5; in the figure, the distance matrix is computed amongst loci with positions based on the average of 50 nucleosomes, which is a resolution appropriate to illustrate the effect of TAD-like domains. Contact maps for a single chain are shown in Figures 5.5A and C. At this level, we observe regions of significant contact on the order of 100s of kbps, which is similar to the size of TADs. In addition, the chain with the linker histone results in the largest TAD-like domain, which is consistent with our earlier observation on looping probability/packing density. In Figures 5.5B and D, we explore the *ensemble* behavior by averaging over multiple chromatin chains. We see that some of the long-range contacts are suppressed in the equilibrium population of chains. Nevertheless, whether linker histones are included results in a notable difference in the extent of genomic contacts.

We compare these single-chain distance maps to recent studies of TAD structure through single-cell and multi-cell Hi-C. Multi-cell work, such as the Hi-C work of Schwarzer *et al* investigates the role of the molecular motor cohesin in TAD formation they reveal that when Rad-21, a key component of cohesin, is knocked out, TADs are depleted. [135] To compare, the single-cell work of Bintu *et al* reveal similar structures to the Hi-C work of Schwarzer *et al* .[18] From this, one can conclude that the Hi-C structure of TADs exist in single-cells, similar to what is shown in Figure 5.5 A and C. Unlike Hi-C, single-cell distance maps show that TAD-like domains still remain in the absence of Rad-21. We explain this discrepancy through our distance maps: TAD-like structures on the order of Mbp are a consequence of

nucleosome-level physics, while cohesin and CTCF stabilize the size and genomic address of TADs. The ICCM predicts this phenomena solely with local information, implying that nucleosome-level physics influence the formation of TADs.

5.5 Discussion

In this work, we introduce the ideal chromatin chain model (ICCM) which demonstrates that a combination of linker lengths and linker histones influence the packaging and regulation of the genome at the gene scale. We find that heterogeneous distributions primarily influence genome packaging depending on the available subset of linker lengths. The model also predicts recent results demonstrating the irregularities of TADs at the single-cell versus the population level. We find that a trinucleosome motif is the smallest fundamental unit that preserves this information.

The simplicity of the ICCM leaves room for considerable improvement. While local structural predictions of the model are promising, more physics should be incorporated to make the model truly predictive of global chromatin structure. In particular, the lack of a long-range interaction keeps the model from directly reproducing Hi-C contact maps with high fidelity. As mentioned previously, compartmentalization and chromatin looping are known structural phenomenon that are a result of epigenomic regulation. Perhaps an incorporation of the recent bottom-up block-copolymer model with an inclusion of the ideal chain model as its homopolymer may provide exciting results. [90] Similarly, a recent model proposed long-range chromatin contacts and density heterogeneity through a simple self-returning random walk model. [62] Additionally, there are further epigenetic marks like H4K16ac which are associated with transcription and do alter the local structure of chromatin.[152] Recent mesoscale modeling approaches are beginning to investigate structural effects like the HP1 protein and a combination of different histone marks and proteins to understand their combined effects. [14, 12] Incorporating approaches such as these appear to be natural next

steps for the ICCM.

5.6 Supporting Information

5.6.1 Model Parameters

Freely-rotating Chain Parameters

As stated in the text, there is a significant amount of data that is used to generate the larger chains. Here, we include visual representations for the variables that define the ideal chain chromatin model. First, we display the joint probability distribution of r_α and r_β for all repeat lengths in the range of 157-206 in Figure 5.6. These figures are generated from long replica-exchange molecular dynamics simulations of 24 nucleosomes each of the 1CPN chromatin model. [76] After an equilibration of $\sim$ 2 million time steps, the snapshots of the fiber are broken down into values of r_α and r_β. From this data, a PMF-like quantity is generated by taking the negative log of the probability distribution and shifted up by the global minima. From the figure, it can be seen that seemingly each NRL displays a unique distribution of states which lends to the unique behavior of the ideal chain chromatin model. Additionally, we note that repeat lengths ≥ 187 begin to display similar behavior, which is similar to the boundary of the "third region" from the statistical length analysis of this work.

Hindered-rotation Chain Parameters

From the same simulations, the dihedral distributions, $P(\phi)$, for each repeat length are also evaluated. The data is displayed in Figure 5.7. We note that the longer repeat length fibers display a more uniform distribution of dihedrals, while shorter lengths display a bimodal distribution. The bimodal distribution is most likely a result of short linker lengths; the

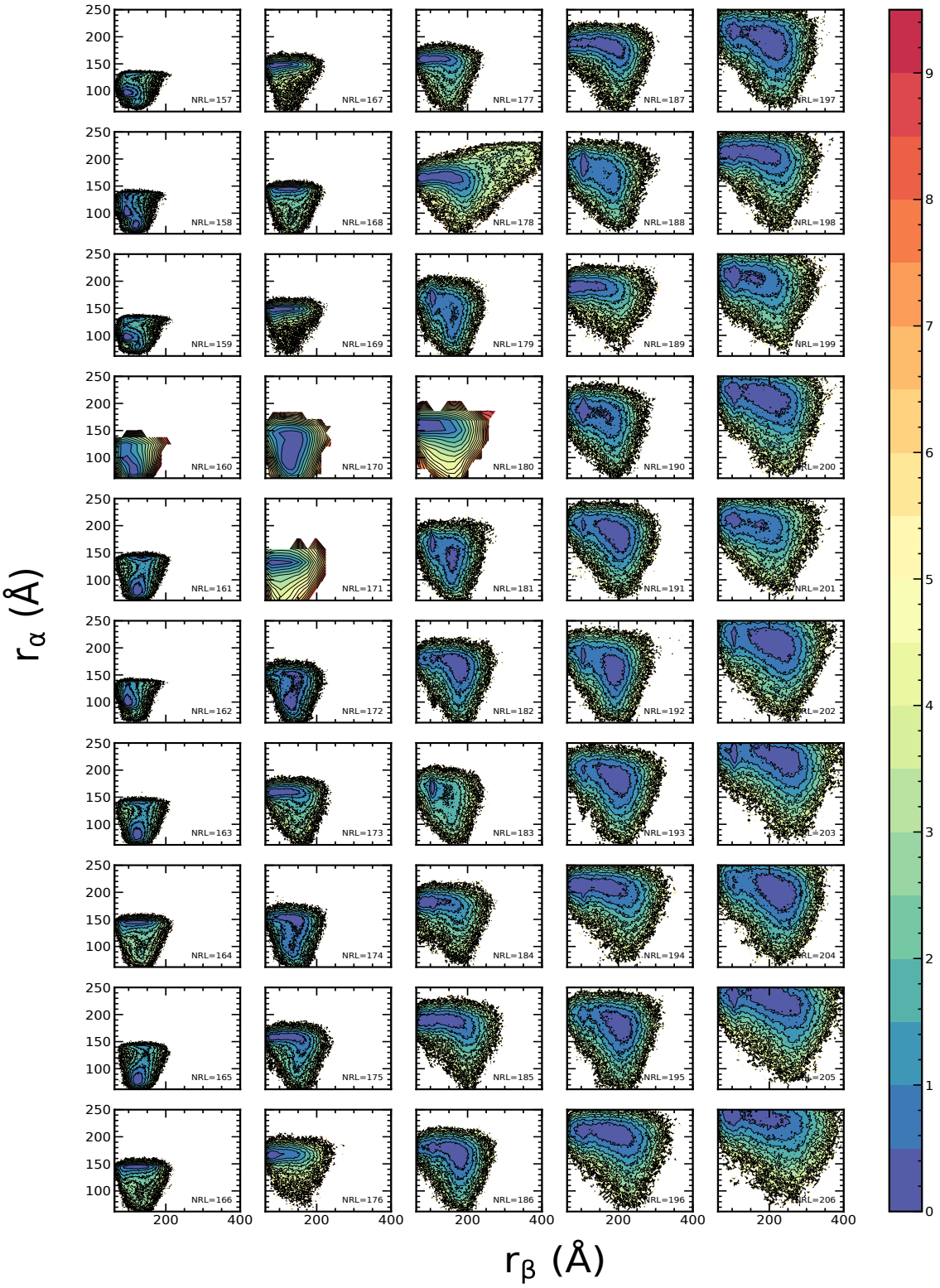

Figure 5.6: Joint probability distributions for the freely-rotating chain version of the ideal chain chromatin model. Distributions are shown for every repeat length from 157-206.

lengths are short enough that nucleosome orientation is governed by the excluded volume interaction between subsequent nucleosomes.

5.6.2 Comparison Between Model Versions

As mentioned in this work, there are slight differences between the hindered-rotation chain (HRC) and freely-rotating chain (FRC) versions of the model. Here, we demonstrate the low differences between the models and also comment on other properties of the model that aren't included in the main text. First, we plot the resulting distribution of r_α as a function of the homogeneous repeat length simulations in Figure 5.8. The results for the FRC model are shown in A and the HRC in B. These figures are generated from the 1000 replicas of the 10,000 monomer chain growth for each homogeneous repeat-length fiber. The distributions are then calculated using the kernel-density estimation (KDE) method with a bandwidth of 2. Interestingly, we note that the distributions are tighter at NRLs of ~ 170 and 187, which correspond well to the boundaries of the regimes of the statistical segment lengths. We also note that, again, there is little qualitative change between the FRC and HRC models. The same distributions were also evaluated for the heterogeneous models in Figure 5.9. The top row are the resulting distributions for the FRC model, and the bottom row are the HRC analogues. Here, we highlight a few observations from the figure. First, we see slight qualitative differences between the two models, except for the Widom distribution, which shows a much larger operating range for the HRC model. Overall, we see that narrower distributions such as the exponential and normal distribution show similarly small distributions in order parameter space. An interesting observation is that the uniform, Widom, and Voong distributions are similar, despite different input NRL distributions. As we mention in this work, these distributions are inputs and can be thought of as a "prior" in a Bayesian optimzation sense. Therefore, the resulting distributions of NRLs can be thought of as the "posterior" distributions. We posit that while this is an inherent bias in

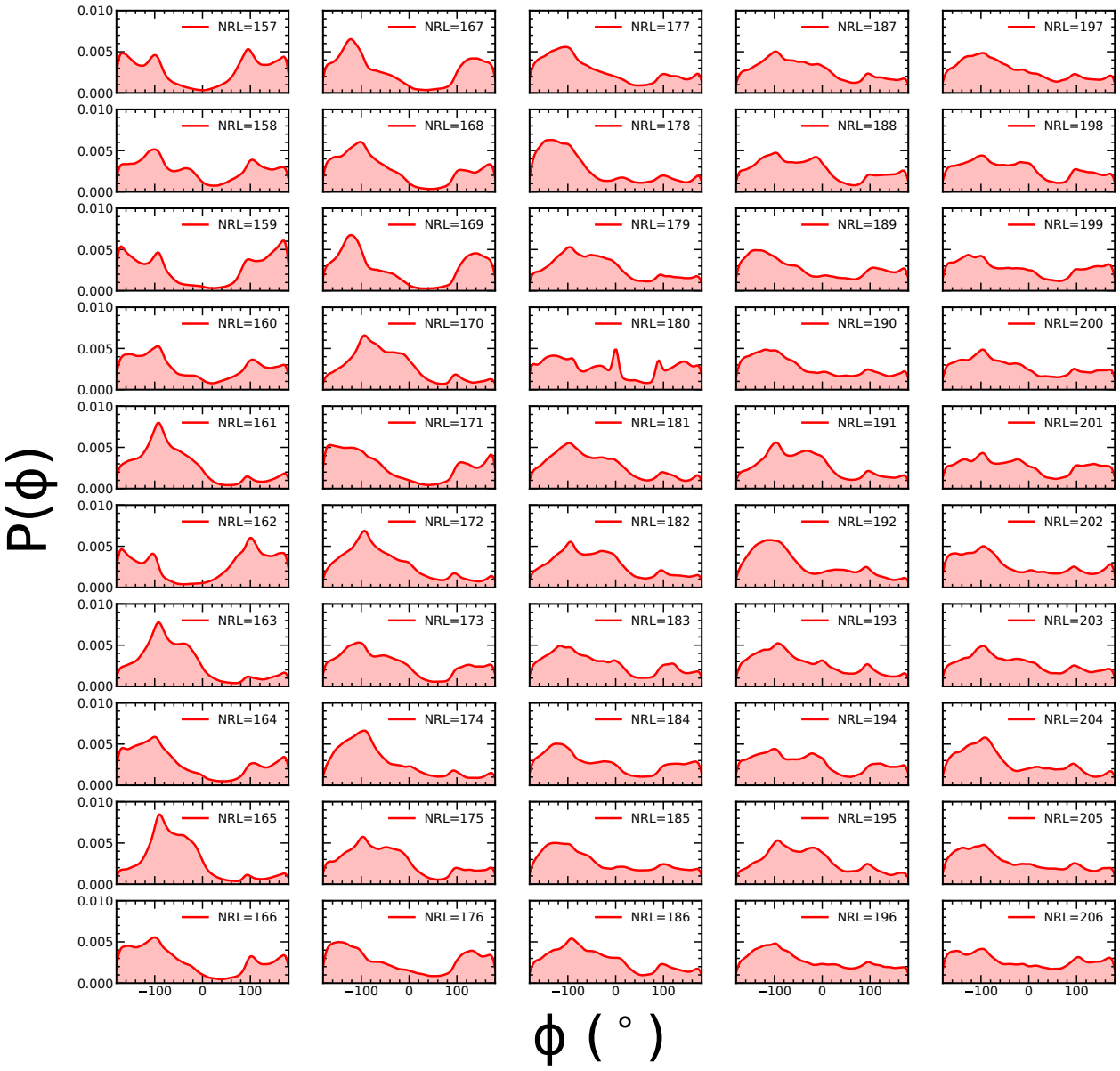

Figure 5.7: Probability distributions for the hindered-rotation chain version of the ideal chain chromatin model. Distributions are shown for every repeat length from 157-206.

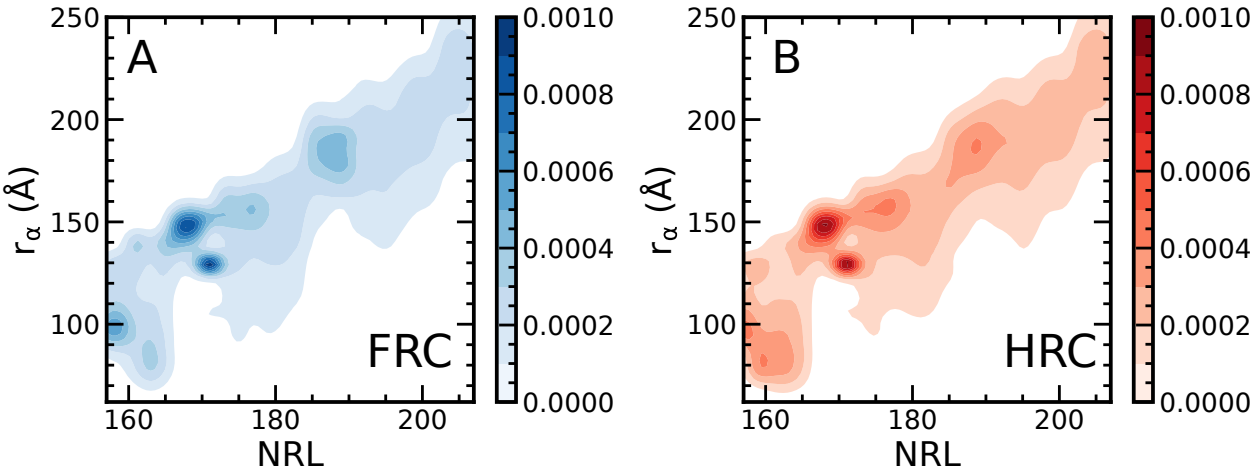

Figure 5.8: Joint distribution of NRL and r_α for the homogeneous-NRL chain growths of both the A) freely-rotating chain model and B) the hindered-rotation chain model variants.

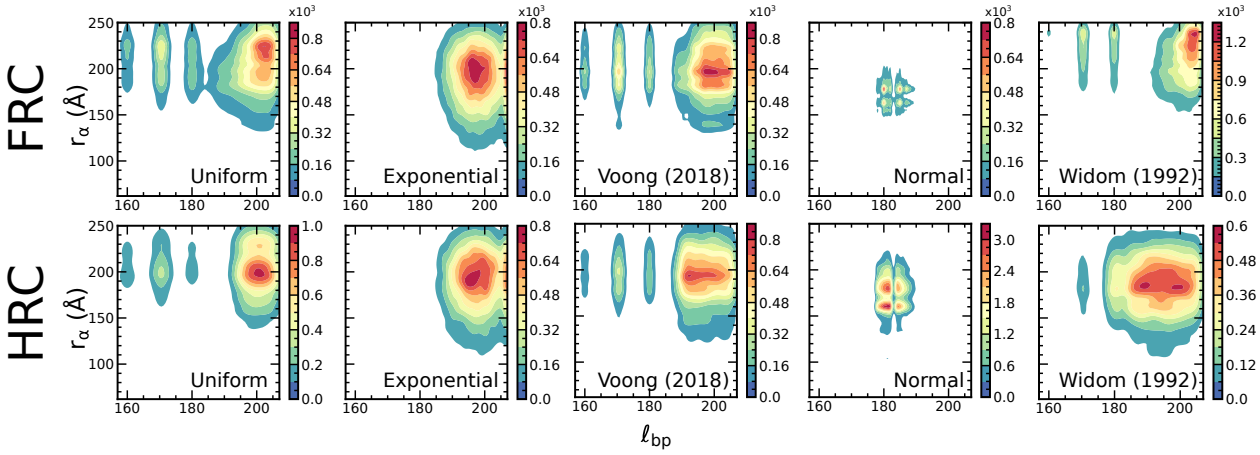

Figure 5.9: Joint distribution of NRL and r_α for the heterogeneous-NRL chain growths of both the freely-rotating chain model and the hindered-rotation chain model variant. The top row figures for the different distributions of the FRC model and the bottom row is the HRC model.

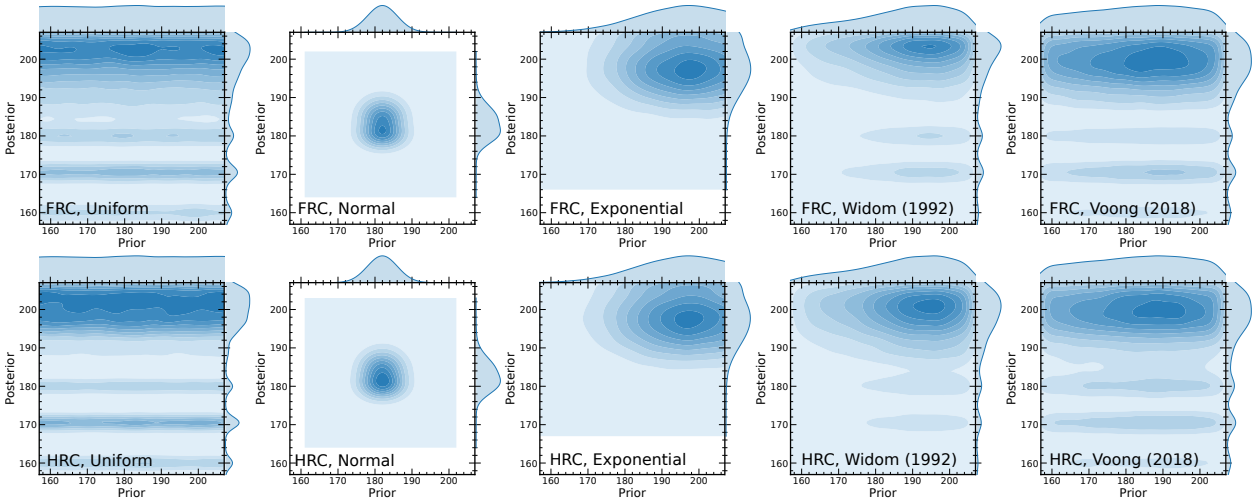

Figure 5.10: Joint plots of the prior and posterior NRL distributions for the heterogeneous chain simulations. The distribution on top is the prior and the distribution on the right is the posterior distribution. The center is a joint plot of both to qualitatively highlight similarities and differences between the prior and the posterior.

the model, we can characterize it and draw information from this. As a result, we calculate the joint probability distributions of the "prior" and "posterior" NRL distributions to better understand this phenomena. These results are displayed in Figure 5.10. For each figure, the prior distribution is displayed on top of the plot and the posterior on the right side. The joint distribution between both the prior and posterior are shown inside each plot. We note that the normal and exponential distributions both show similar behavior, which is reflected in the joint distribution. Perhaps the most interesting result is that of all the distributions that sample all repeat lengths, the Widom distribution displays similar behavior for its prior and posterior. As there are qualitative similarities between the exponential, Widom, and Voong distributions, we claim that perhaps the equilibrium distribution of NRLs follows similar behavior. As longer NRLs are associated with regions of heterochromatin, [149] it makes sense that shorter repeat lengths are less likely in the genome. Such a result bolsters the efficacy of the predictive capabilities of the ideal chain chromatin model.

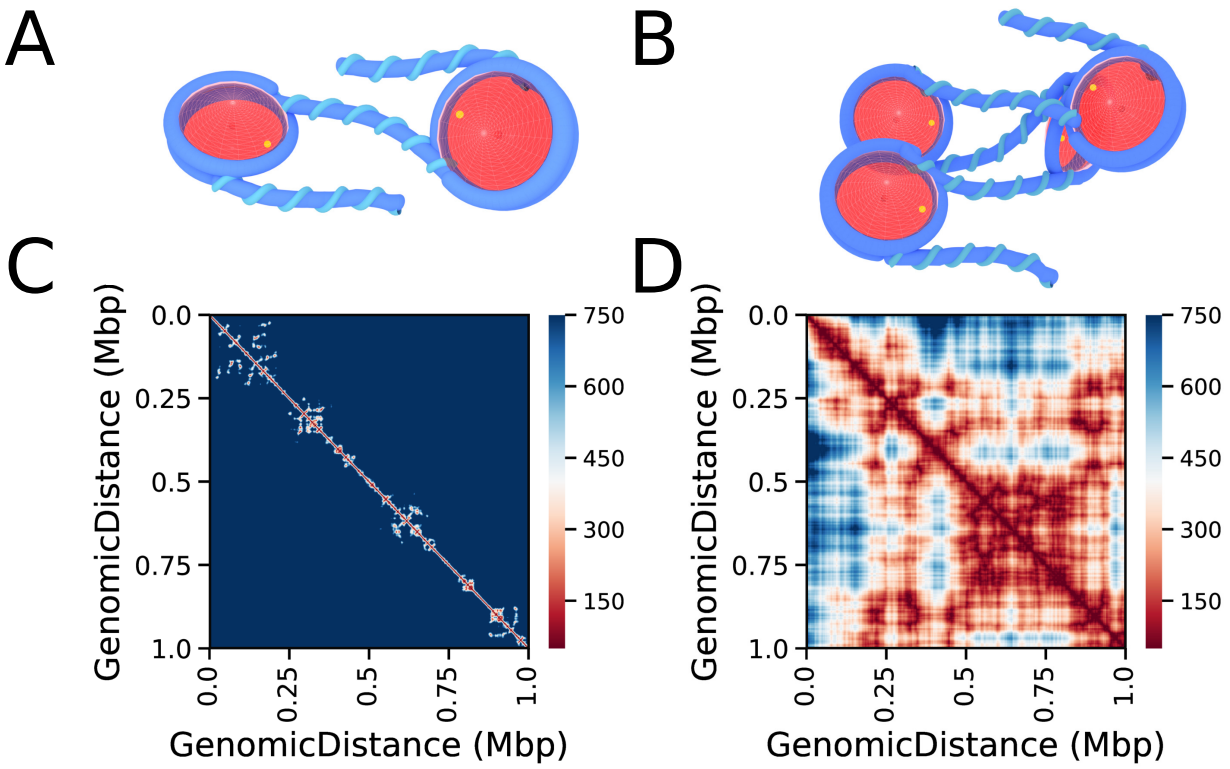

Figure 5.11: Single chain distance maps for varying levels of representation. A) A graphical representation of the dinucleosome, with corresponding distance map in C). B) A tetranucleosome with resulting distance map in D).

5.6.3 Distance Maps

In the manuscript, we highlight the predictive capabilities of our chain growth method by comparison to recent single cell Hi-C experiments. To show the unique behavior of the trinucleosome assumption, we also show the results for a freely-jointed chain (FJC) model (or dinucleosome) and the HRC model. The results are displayed for the Widom distribution in Figure 5.11

CHAPTER 6

PHYSICALLY-INFORMED DEEP LEARNING RECONSTRUCTION OF CHROMATIN *IN SITU*

6.1 Abstract

Advances in microscopy are pushing the boundary for high resolution structural chromatin information. In particular, the recent advances of ChromEMT and ChromSTEM are driving our understanding of chromatin structure *in situ*. Parallel to such advancements, the development of machine vision methods to correctly identify objects have advanced medical imaging and autonomous automobiles. Here, we combine molecular dynamics simulations and deep learning to determine the structure of chromatin from ChromSTEM data. Our approach is end-to-end, taking in raw ChromSTEM data and producing a physical structure of chromatin. To begin, we discuss modeling decisions and the training of the machine learning models. Notably, we highlight our unique data augmentation methods through molecular dynamics simulations that enhance the capability of machine learning models. Lastly, we evaluate the feasibility of the resulting structures through distributions of linker lengths and spatial nucleosome positioning. We find the approach outlined here of great promise and at the forefront of chromatin structure study.

6.2 Introduction

Chromatin structure regulates necessary biological processes. Both hierarchical and dynamic, the structure of chromatin facilitates regulation of fundamental processes such as transcription, replication, and DNA repair. Due to the dense nuclear environment, chromatin is heterogeneously packed. Two regions of packing, heterochromatin and euchromatin, correspond to regions of highly dense packaging and less dense packaging, respectively. Re-

cent experiments have proposed that euchromatin and heterochromatin form a liquid-liquid phase separation, resulting in spatially distinct departments. [74, 140] Long thought to organize into ordered structures at the kbp scale, theory has shifted towards a hetergeneous and dynamic, but still ordered structure of chromatin.

Innovations in our understanding of chromatin structure come from significant advances in imaging and chromosome conformation capture methods. Recent advances can be attributed to the development of ChromEM-based methods. [112, 83] ChromEM is an Osmium-based high-contrast dye that selectively binds to DNA. By fixing cells and introducing ChromEM, electron microscopy with multi-tilt tomography (ChromEMT) can resolve detailed features of chromatin up to ~ 1.9 nm. Through this method, the previously-established 30-nm fiber was dispelled as an *in vivo* concept. Instead, ChromEMT reveals that chromatin exists in seemingly disordered fibers of radius 5-24 nm. Further advances of ChromEM based methods comes in the form of ChromSTEM, which utilizes scanning-transmission electron microscopy with multi-tilt tomography. [83]

In contrast to such highly resolved *in situ* methods, label-free, *in vivo* methods such as partial wave spectroscopy (PWS) detect packaging heterogeneity in the nucleus. [8, 38] With sensitivity up to 20nm, PWS captures structural changes in chromatin. So far, PWS has revealed significant changes to chromatin structure in cancerous cells, linking the local chromatin packaging to cancer. [38] Such a technique is poised to be powerful in determining preliminary structural changes that can be further resolved through methods such as ChromSTEM.

While optics and microscopy are probing detailed structural features, simulation-based approaches are able to finely resolve chromatin structure. Additionally, recent advancements in technology involves the advent of machine learning in processing patterns in images, language, and data. Included in the AI and machine learning revolution is the development of technology for object detection. Standard in the self-driving car literature is the use of

object detection, specifically from the benchmark KITTI dataset. [52] Unfortunately, object detection is in the class of supervised techniques, which rely on a significant amount of data to train.

Conventionally, molecular modeling of chromatin has been able to predict features of chromatin structure at varying hierarchical length scales. [101] At the nucleosome scale, atomistic simulations characterize fluctuations in nucleosome structure and connect them to higher length scale phenomena and their associated biological implications. [19] At the oligonucleosome scale, coarse-grained modeling connects the structure of the 30-nm fiber and its changes in the presence of epigenetic changes.[10, 76, 101, 107] At the largest scale, polymer models are informed by the results of 3C methods, such as Hi-C.[84]

Despite such cutting-edge advances in imaging, a resolved picture of chromatin, consisting of nucleosomes and linker DNA has yet to be revealed. While nucleosomes can be imaged and determined through X-ray crystallography, a complete structure of the chromatin, including the linking DNA *in situ* or *in vivo* is a difficult prospect. Due to the limited amount of ChromSTEM data available and difficulty in labeling images, supervised techniques combined with ChromSTEM are simply intractable. Through recent advancements in modeling and imaging, it should now be possible to resolve the local chromatin structure. To this end, we foresee a fruitful combination of molecular modeling and deep learning in successful interpretation of ChromSTEM data.

In this work, we combine molecular simulations with deep learning to introduce a workflow designed to find the structure of chromatin from ChromSTEM. We rely on preprocessing through unsupervised learning to separate different clusters of chromatin assumed to be separate strands. We then run these clusters through a trained convolutional neural network (CNN) to predict the number of nucleosomes within each cluster. The network is trained on hundreds of frames of simulations that are reduced to the same representation of ChromSTEM data. With the number of nucleosomes, the data is then passed into a Gaussian

Mixture Model routine to determine their positions and orientations. To determine the amount of linking, the nucleosomes are treated as centers to be optimized by simulated annealing Monte-Carlo. At the end of the workflow, we demonstrate the capabilities of data-driven techniques in combinations with molecular simulations to provide the first-of-its kind resolved structure of chromatin.

6.3 Methods

We highlight our workflow below in Figure. 6.1 The data for this work comes from our collaborators. [83] ChromSTEM with multi-tilt tomography results in a rectangular prism of data, sliced from the nucleus of size 100 nm x 1200 nm x 1200 nm. ChromSTEM data in its raw form are voxels of varying intensity of size 2.9 nm x 2.9 nm x 2.9 nm, where the image intensity is directly proportional to the amount of DNA. The ChromSTEM data for this work comes from A549 adenocarcinoma cells. To render the data useable for our workflow, we create a mesh through the open-source Paraview software and export the vertices to generate a point cloud of data. [3]

6.3.1 Data preprocessing: separate chains

Due to the constraints of the method, chromatin may not be in a contiguous chain and can leave and re-enter the box. To account for this, we use a density-based hierarchical clustering method known as HDBSCAN to separate regions into multiple chains and filter out probable noise. This method allows us to feasibly split the dataset so that we can train the resulting workflow on the individual chains and not the entire volume all at once. The resulting sub-clusters are sent to a convolutional neural network (CNN) to predict the number of nucleosomes.

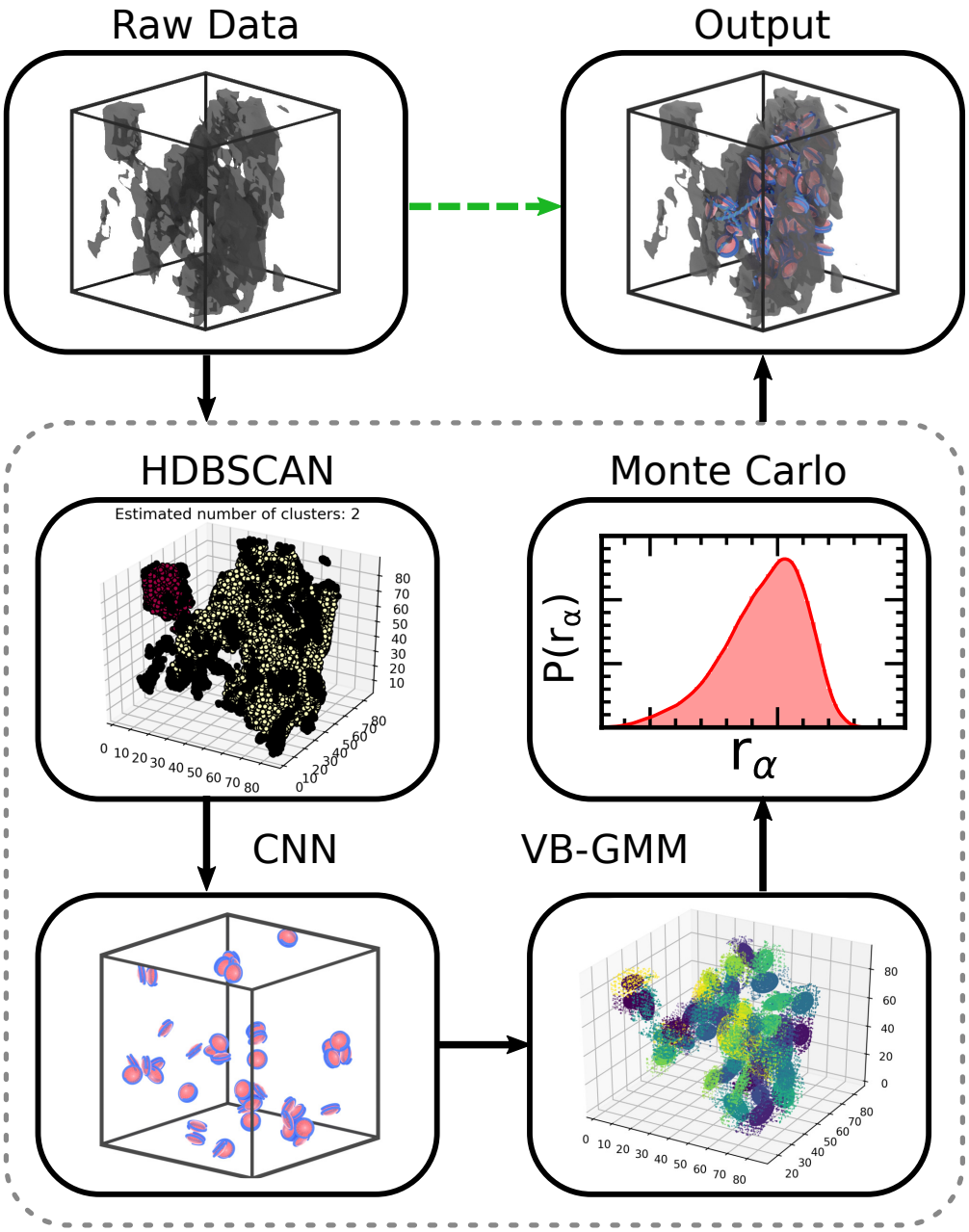

Figure 6.1: End-to-end determination of chromatin structure from ChromSTEM data. The input data is sent to an HDBSCAN routine to separate individual chains of chromatin. The number of nucleosomes in these chains are determined with a trained neural network. Once the number of nucleosomes are learned, a variational Gaussian Mixture Model determines the locations and orientations of the nucleosomes. The nucleosome positions are used as centers in a simulated annealing Monte Carlo routine to determine an optimal topology. The resulting topology is simulated to create a physically-informed structure of chromatin.

6.3.2 Predicting the number of nucleosomes

As a proof of concept, we use a CNN to predict the number of nucleosomes in a box. Traditionally used in image processing, CNNs are very capable of learning features from images and predicting what the features represent. For example, the standard ResNet is a deep network consisting of convolutional and fully-connected layers trained to accurately predict 1000 object classes. [57] For simplicity, we take the original voxelized representation of ChromSTEM to train the network given the efficiency of 3D CNNs to predict objects from voxelized data. To reduce the size of the data, we break down the original ChromSTEM data into boxes of dimensions 100 nm x 100 nm x 100 nm. We note that due to all data being the same volume, the network will only be able to predict nucleosomes accurately for a fixed-volume. A significantly larger amount of data is required to train a network to learn nucleosome signatures independent of volume.

Data generation and augmentation

ChromSTEM data is often expensive and time-consuming to produce, not to mention impossible to label nucleosome coordinates/positions. As a result, we choose to "augment" the ChromSTEM dataset by running simulations of the established 1-Cylinder-Per-Nucleosome (1CPN) model and represent the simulations in the exact same manner as ChromSTEM data. [76] The 1CPN model has been integral in predicting chromatin structure at the kbp length scale, primarily due to the comprehensive definition of the nucleosome pair-interactions. [101] Not only are we capable of augmenting the ChromSTEM dataset through 1CPN, but we also have knowledge of the positions and orientations of all nucleosomes in a box, converting the problem from unsupervised to supervised learning. The resulting dataset consists of 600,000 frames of simulations of 5 to 100 nucleosomes, sufficient for training a CNN. The data is split into training, validation, and testing sets according to the standard 70-15-15 split. The network was trained on the 70% of the training data for each epoch and

the epoch with the best validation accuracy was selected for the final network parameters. We display the results of the network training in Figure 6.4. The loss for both the training and validation phases is in Figure 6.4B and the accuracies in A.

Data manipulation

To prevent overfitting of the network to the data, each image is filtered with random Gaussian noise and Poisson noise. The Gaussian noise is defined as:

$$\mathcal{N}_{Gauss} \sim \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\vec{\rho}-\mu)^2}{2\sigma^2}} \tag{6.1}$$

where σ is the defined standard deviation of the distribution, μ is the mean amount of noise, and $\vec{\rho}$ is the chromstem density in basepairs/voxel. The Poisson noise is applied after Gaussian noise is applied in the form:

$$\begin{aligned} \mathcal{N}_{Poisson} &\sim e^{-\lambda}\frac{\lambda^k}{k!} \\ \lambda &= \vec{\rho} * 2^{\lceil \log_2(\{\rho_i\}_{i\in i,\ldots,n})\rceil} \end{aligned} \tag{6.2}$$

where $\{\rho_i\}_{i\in i,\ldots,n}$ represents the set of all unique values of DNA density in the chromstem dataset and k is the number of events occuring within the interval λ. We display the effects of noise on a vertical slice of the input data in Figure 6.2.

Network Parameters

The final neural network consists of five layers that we show in Figure 6.3. The first two layers correspond to the featurization layers, each of 3D convolutional layers with max pooling. The last three layers are fully-connected layers. The middle layer consists of a dropout layer, which randomly disconnects 30% of the connections, resulting in a more robust network.

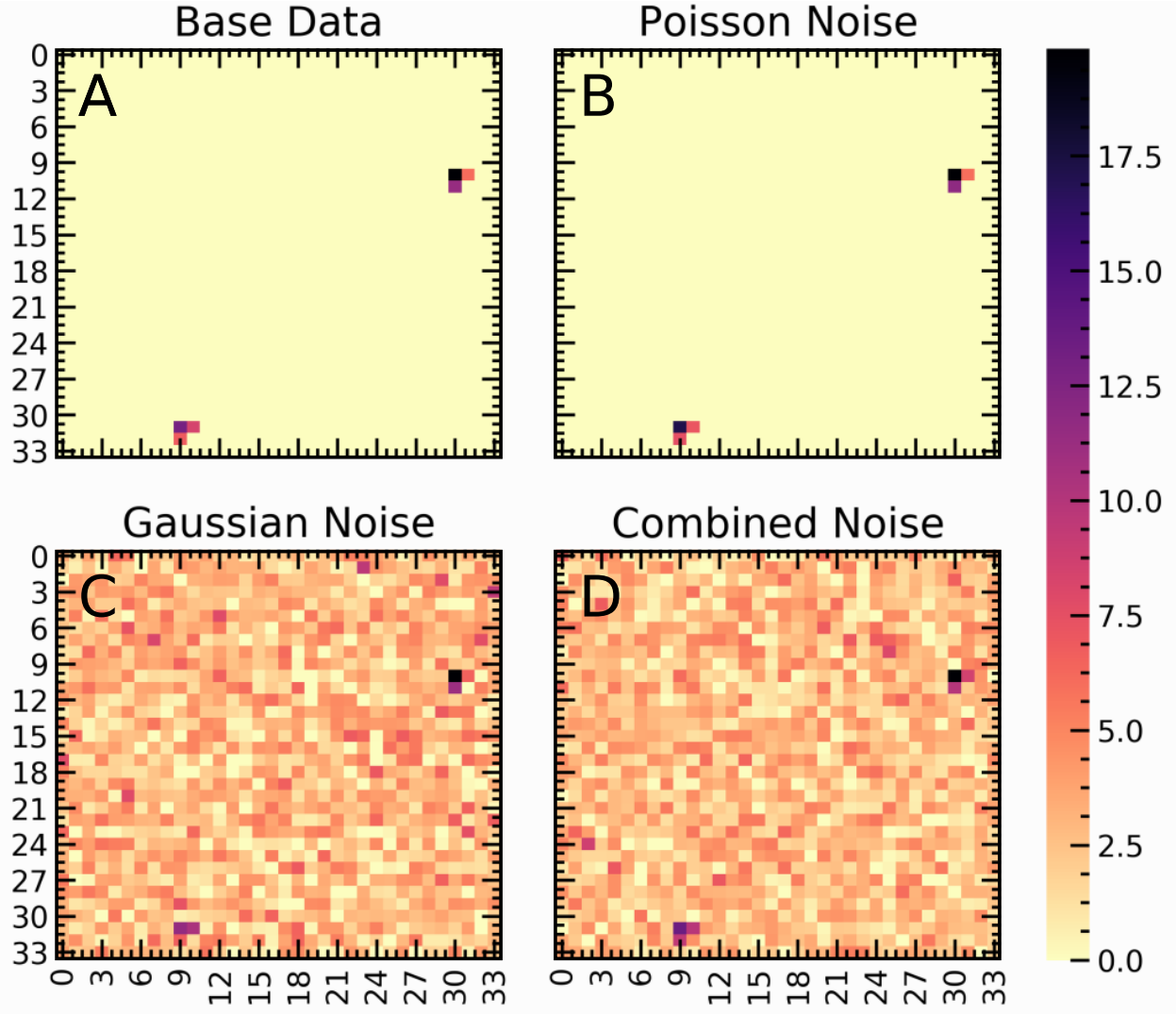

Figure 6.2: Noise incorporated into the dataset to make the network more robust to the actual data. A) Vertical slice of simulation snapshot voxelized into DNA density without modifications B) Poisson noise filtered data C) Gaussian noise filtered data D) Gaussian + Poisson noise

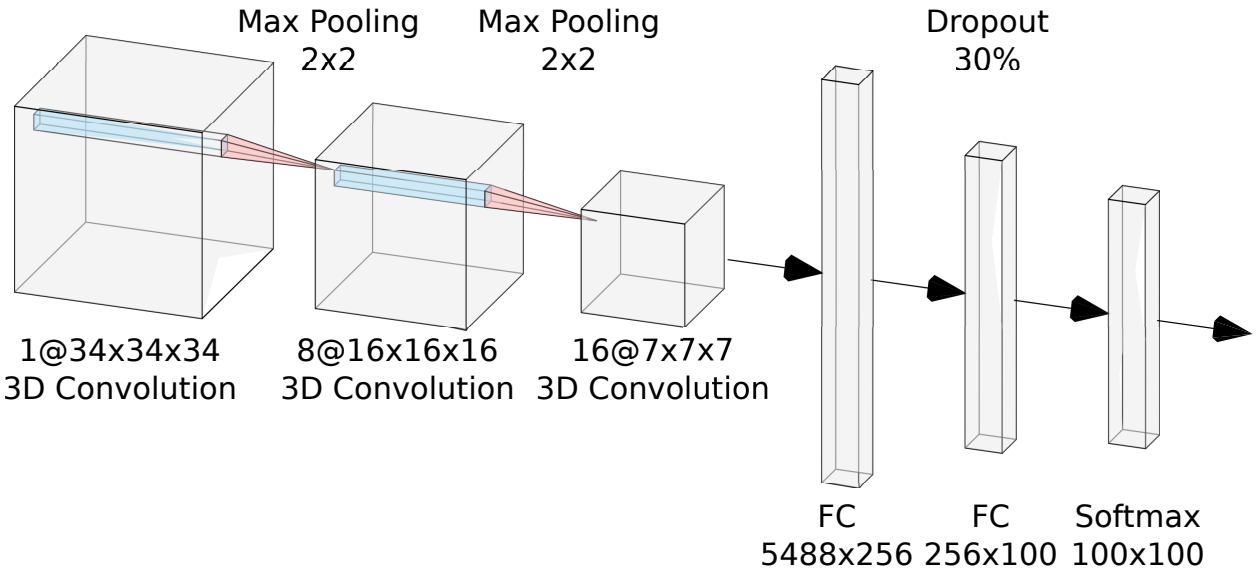

Figure 6.3: Deep network architecture. The network is a deep network consisting of a featurization layer with successive convolutional layers and a prediction layer with three fully-connected layers. After each convolutional layer, a max-pooling layer is evaluated. After the second fully-connected layer, there is a dropout layer of 30%. The output loss function utilizes a softmax layer to predict the most-probable amount of nucleosomes in the snapshot.

The output layer is a softmax layer that results in the most probable number of nucleosomes for the snapshot. The parameters for all of the layers are available in Table 6.1.

Layer	Features	Height	Width	Depth	Filter
Conv3D	1	34	34	34	(2x2x2)
MaxPool3D	8	16	16	16	2
Conv3D	8	16	16	16	(2x2x2)
MaxPool3D	16	7	7	7	2
FC	1	5488	256	1	-
FC	1	256	100	1	-
Softmax	1	100	100	1	-

Table 6.1: Neural network layer parameters

6.3.3 Predicting nucleosome positions

With a trained network, we have the number of predicted nucleosomes for each volume subset. We pass this number into a variational Bayesian Gaussian mixture model (VBGMM, or

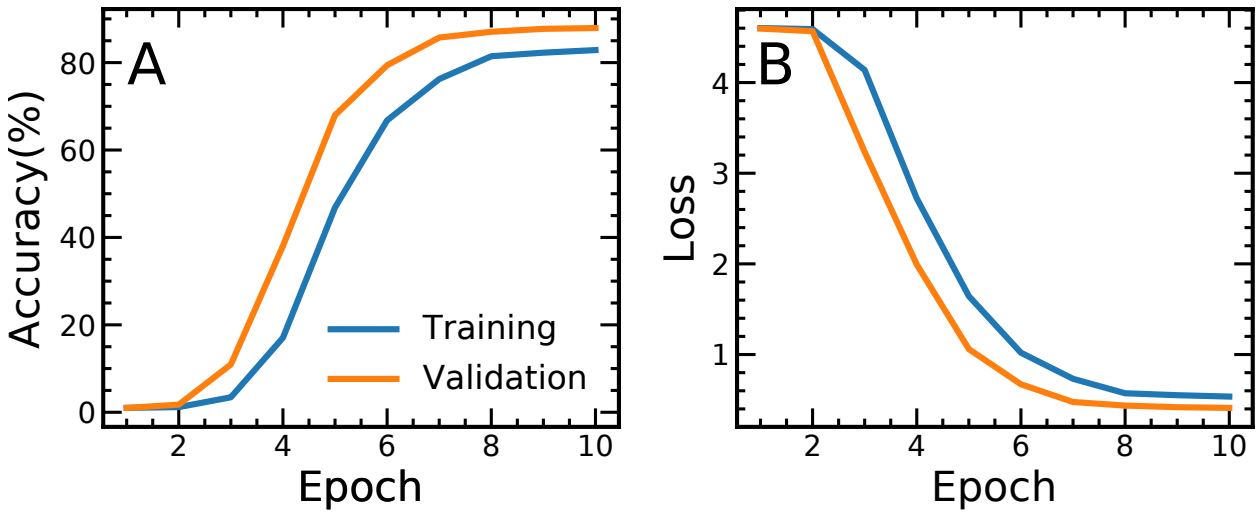

Figure 6.4: Network training. A) The accuracy of both the validation and training phases over the training length of the network. Final training accuracy is 83% and final validation accuracy is 89%. The accuracy of the validation phase is higher due to the use of a dropout layer in training. B) The loss function of the network during training. The network is trained with stochastic gradient-descent (SGD) with momentum and regularization with parameter $\lambda = 0.01$.

GMM for short) to predict the locations of the nucleosomes. Gaussian mixture models are a form of unsupervised learning and work by fitting multivariate Gaussian distributions to datasets. In the case of the ChromSTEM data, we use the aforementioned point cloud representation and fit Gaussian distributions through the VBGMM routine. The mean of these distributions corresponds to the location of a nucleosome within the volume. Additionally, we perform an eigenvector decomposition of the covariance matrix of the distribution to get the orientation of each nucleosome. We provide a representation in Figure 6.1.

6.3.4 Predicting the chromatin topology

With the nucleosome positions and orientations in place, we next tackle the question of linking the nucleosomes. The problem of finding the optimal linking is reminiscent of the "postman's dilemma, " a non-polynomial hard algorithm. Therefore, we consider the approximate solution through simulated-annealing Monte-Carlo (SA-MC). With previous work,

we generated a significant amount of data to determine the distance distribution between sequential nucleosomes along a chromatin fiber as a function of the amount of linking DNA. This data was generated by running fibers of 24 nucleosomes each of homogeneous linker DNA with replica exchange for 100 μs each. Using this data and the reference ChromSTEM data, we generate the following Hamiltonian, $\mathcal{H}$, for SA-MC.

$$\mathcal{H} = \sum_{i=0}^{N_{voxel}} \left(\rho_{sim,i} - \rho_{ChromSTEM,i}\right)^2 + \sum_{i=0}^{N_{links}} f_{NRL}(|r|_i) \tag{6.3}$$

We start with a dimensionless temperature T = 100 and reduce by one until a local minima is found. Our MC moves consist of swapping links and swapping functional forms of the linking DNA, representing changes in the amount of linking DNA. At the end of the routine, we have the resulting topology of the chromatin fiber(s).

6.4 Results

Here, we discuss the resulting structure of chromatin through the workflow highlighted above. We show how the structure overlaps with the reference dataset in Figure 6.5A. For this region, the neural network determined that there are 62 nucleosomes, 59 of which are shown. The gray represents a mesh of the raw ChromSTEM data, created in Paraview. The nucleosomes are represented by pink oblate ellipsoids and the linker DNA is shown by the blue connections between them. The figure shows qualitative agreement with the reference data, with a few notable deviations. For one, we note that a large amount in the top left corner of the data does not show any nucleosomes. The HDBSCAN routine detected this as a separate chain, so it was removed from the image for interpretability.

We analyze the structure by calculating the displacement between nearest neighbor nucleosomes and next-nearest neighbor nucleosomes, the orientation of the nucleosomes relative to the z-axis, and the linker length distribution. Following previous work, we evaluate

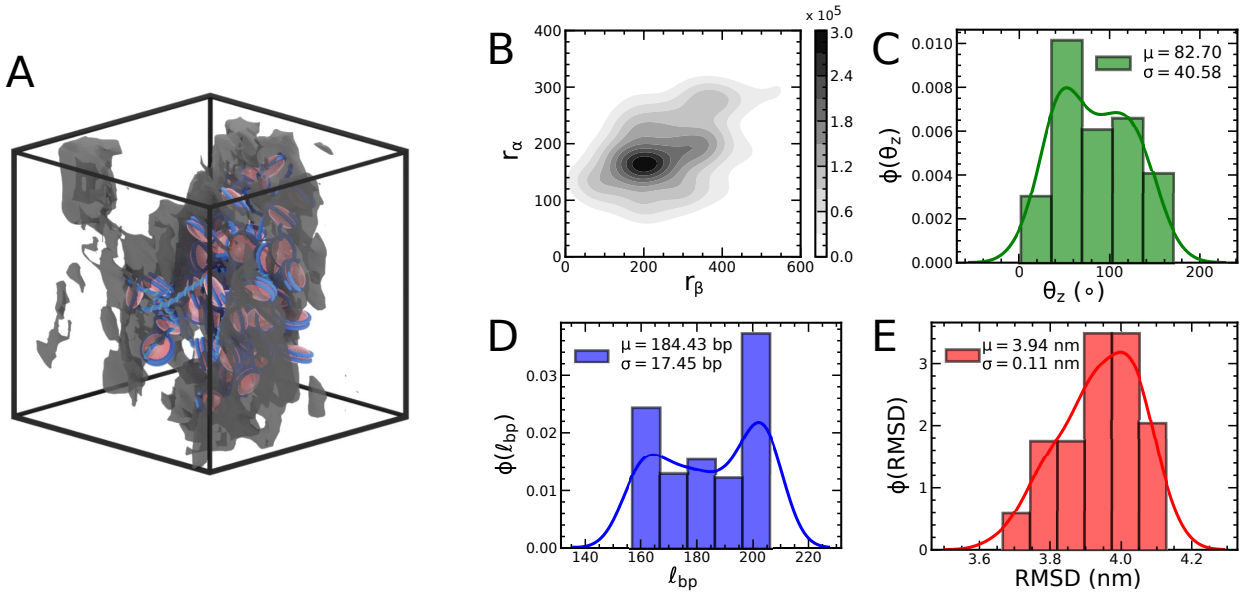

Figure 6.5: The physically-informed deep learning structure of chromatin. A) The simulation result of the workflow (blue and pink) with the reference data overlay (gray). The data was meshed using the Paraview software and was imported into Blender 2.82a. The nucleosomes are in pink and the linker DNA is in blue with helix orientation in light blue. B) The local nucleosome distance distribution where r_α is the nearest-neighbor nucleosome distance and r_β is the next-nearest neighbor distance. C) The orientations of the nucleosomes with respect to the z-axis, θ_z. D) The linker length distribution of the fiber. E) The root-mean-squared distance distribution of 10 copies of the nucleosome positions of the GMM routine. Each unique run was translationally and rotationally aligned using the Kabsch algorithm.

the joint probability distribution between nearest-neighbor (r_α) and next-nearest neighbor distances (r_β), seen in Figure 6.5B. We notice the majority of nucleosomes are at a distance of $r_\alpha = 180$Å and $r_\beta = 210$Å. The maximum probability for the spatial distribution corresponds to a less condensed fiber locally than previous work dictates. Such a result demonstrates that locally, the fiber is not compact, however long-range interactions seemingly dominate, resulting in the condensed fiber. We also show the relative orientations of the nucleosomes relative to the z-axis, θ_z, in Figure 6.5C. The orientations are calculated from the angle between the face vector of the nucleosomes, $\hat{f}$. We find the nucleosomes to be oriented towards the fiber, however that may be an artifact of the dilute simulation conditions. As we are interested in resolving the linker length distribution, we calculate the distribution of NRLs that result from the final fiber topology, which we display in Figure 6.5D. We note that there is a bimodality in the distribution at short and long linker lengths. As a result of potential variance in the GMM, nucleosomes can be connected at unfavorable distances to minimize the reference dataset energy. Apart from peaks associated with these phenomena, we also notice a small peak at 177-187 bp. This range of linker lengths is associated with euchromatin, suggesting that the region of chromatin in this box is more associated with euchromatin than heterochromatin.

Additionally, there are regions where the reference data and resulting structure deviate, namely with the linker DNA. These deviations arise from a few sources of error that we discuss in depth here. The workflow relies heavily on the assumption of each individual method minimizing the bias-variance trade-off. As such, any method which introduces errors will result in error propagation throughout the workflow. Perhaps the greatest sources of error come from the GMM in the workflow. There is an inherent variance in the method when running it multiple times. We quantify said variance by calculating the root-mean-squared distance (RMSD) distribution of the nucleosome positions with respect to the average positioning. We calculate this quantity by using the Kabsch algorithm for rotational and

translational invariance between two reference structures. [66] The resulting RMSD distribution is displayed in Figure 6.5E. We see an average deviation of $\sim$ 4nm, which while low, could significantly propagate in the SA-MC routine.

6.5 Discussion

In this work, we develop a workflow which combines molecular simulations and recent microscopy advances to determine the structure of chromatin *in situ*. With this work, we are able to provide an estimate for the chromatin structure. We find that long-range interactions seemingly dominate in chromatin compaction. Additionally, we find the distribution of linker lengths in the structure we found to be similar to that of euchromatin. Despite the workflow defined here, there is still propagation of error which can be reduced in the future. We find that this work is just the beginning and can be extended significantly. Advances are being made every day with object detection by deep learning. In particular, it would be of great benefit to extend the deep learning portion of this work to reduce the variance in the final solution. Namely, we deem it feasible to integrate recent advances in 3D object detection through networks such as PointNet++ [120], PV-RCNN [137], or VoxelNet [165] to predict the positions of the nucleosomes directly, without the need for a GMM approach. Additionally, it could be possible that a deep network approach could also learn topology of the fiber and provide a one-shot solution.

CHAPTER 7

CONCLUSION

In summation, we have shown the power of computational tools to provide insight into the current field of chromatin structure. Through the pair-potential work, we demonstrate a comprehensive picture of the nucleosome interaction landscape and evaluate the role of each histone tail. This work directly led to the development of the 1CPN chromatin model which is emerging as the premiere model for mesoscopic chromatin simulations. Here, we extend the model to incorporate the linker histone to enhance the model's capabilities. The model was validated here with and without the linker histone through sedimentation coefficient scaling with salt concentration. We find the 1CPN model to be in excellent agreement with experiments. The 1CPN model is first utilized to determine structural correlations in the chromatin fiber. Through this, we find chromatin structure has a trinucleosome structural basis. We also conclude that acetylation of the nucleosome tails results in liquid-like behavior of the chromatin, rendering it malleable by external factors. The trinucleosome assumption is tested further by the introduction of the ICCM. We find the assumption to hold and provide insight into the role of the nucleosome on supramolecular chromatin structure. In particular, we propose the nucleosome to influence TAD structure formation and individual linker lengths to influence the structure and dynamics on length-scales relevant to transcription. We lastly test the developed frameworks by proposing an end-to-end deep learning and molecular simulation workflow to determine the structure of chromatin *in situ*.

We believe each chapter of this work has the potential to be expanded upon, especially in predicting the relation between thermodynamics and epigenetics of the nucleosome and chromatin fiber. We believe that the trinucleosome functional shown here is a rich physical definition of chromatin structure with low overhead. We expect that incorporating the trinucleosome into a mean-field model with the potential to predict chromatin structure influenced by CTCF, cohesin, and chromatin compartmentalization to be an exciting prospect.

Such a model would have nucleosome-level information at the chromosomal scale. Lastly, we find an extension of the machine learning workflow to predict nucleosomes directly using cutting-edge neural networks to enhance the resolution of the model and lower variances.

In a broader sense, chromatin structure plays a part in the formation of genetic diseases. However, researchers have yet to decouple the contributions from proteomic effects, epigenomic phenomena, and the structure of chromatin. In this work, we provide a computational workflow that can be predictive of chromatin structure in the absence of the other phenomena to understand the extent of contributions from the chromatin fiber. We find that this dissertation provides a blank-canvas upon which we can paint a comprehensive picture of the relation between chromatin structure and epigenomic regulation. Future work will build upon this workflow and understand how both the proteome and epigenetics work in concert with chromatin structure. Such a comprehensive picture would provide experimentalists with a fundamental understanding of the biology at play and would pave the way for targeted design to counter genetic diseases.

REFERENCES

[1] Zahra Abdellah, Alireza Ahmadi, Shahana Ahmed, Matthew Aimable, Rachael Ainscough, Jeff Almeida, Claire Almond, Andrew Ambler, Karen Ambrose, Kerrie Ambrose, Robert Andrew, Daniel Andrews, Neil Andrews, Dan Andrews, Eva Apweiler, Hazel Arbery, Beth Archer, Gareth Ash, Kevin Ashcroft, Jennifer Ashurst, Robert Ashwell, Deborah Atkin, Andrea Atkinson, Barry Atkinson, John Attwood, Keith Aubin, Katherine Auger, Terry Avis, Anne Babbage, Sarah Babbage, Joanne Bacon, Claire Bagguley, Jonathan Bailey, Andrew Baker, Ruby Banerjee, Simon Bardill, Darren Barker, Gary Barker, Daniel Barker, Karen Barlow, Laurent Baron, Anika Barrett, Rebecca Bartlett, David Basham, Victoria Basham, Alex Bateman, Karen Bates, Caroline Baynes, Lisa Beard, Susan Beard, David Beare, Alastair Beasley, Helen Beasley, Oliver Beasley, Stephan Beck, Emma Bell, Damian Bellerby, Tristram Bellerby, Richard Bemrose, James Bennett, David Bentley, Andrew Bentley, Mary Berks, Michael Berks, Graeme Bethel, Christine Bird, Ewan Birney, Helen Bissell, Suzanne Blackburne-Maze, Sarah Blakey, Christel Bolton, James Bonfield, Ralph Bonnett, Richard Border, Amanda Bradley, Nicola Brady, Jason Bray, Sarah Bray-Allen, Anne Bridgeman, Jonathan Brook, Shane Brooking, Andrew Brown, Clive Brown, Jacqui Brown, Margaret Brown, Mary Brown, Richard Bruskiewich, Jackie Bryant, David Buck, Veronica Buckle, Claire Budd, Sarah Buller, Jill Burberry, Deborah Burford, Joanne Burgess, Wayne Burrill, Christine Burrows, John Burton, Christine Burton, Phil Butcher, Adam Butler, Murray Cairns, Nick Camm, Christopher Campbell, Bruno Canning, Carol Carder, Paul Carder, Nigel Carter, Tamara Cavanna, Shani Chalk, Ka Chan, Joanna Chapman, Rachel Charles, Neil Chillingworth, Tom Chothia, Connie Chui, Rob Clack, Michele Clamp, Anthea Clark, Graham Clark, Kevin Clark, Sarah Clark, Sue Clark, Richard Clark, Betty Clarke, Eddie Clarke, Kay Clarke, Adrian Clarke, Laura Clarke, Chris Clee, Sheila Clegg, Karen Clifford, Julia Coates, Victoria Cobley, Alison Coffey, Penelope Coggill, Lotte Cole, Rachael Collier, Simon Collings, John Collins, Philip Collins, Louise Colman, Anthony Connolly, Richard Connor, Jennie Conquer, Donald Conroy, Doug Constance, Leanna Cook, Jonathan Cooper, Rachel Cooper, Robert Cooper, Maria Coppola, Teresa Copsey, Nicole Corby, Linda Cornell, Ruth Cornell, Christine Cornell, Amanda Cottage, Alan Coulson, Gez Coville, Anthony Cox, Tony Cox, Robert Coxhill, Matthew Craig, Tom Crane, Matt Crawley, Victor Crew, James Cuff, Karl Culley, Auli Cummings, Kirsti Cummings, Paul Cummings, Adam Curran, Valery Curwen, Jeffrey Cutts, Rachael Daniels, Lucy Davidson, Jonathon Davies, Joy Davies, Nicholas Davies, Robert Davies, John Davis, Jayne Davis, Matthew Davis, Elisabeth Dawson, Rebecca Deadman, Peter Dean, Simon Dear, Frances Dearden, Marcos Delgado, Panos Deloukas, Janet Dennis, Pawandeep Dhami, Catherine Dibling, Ruth Dobbs, Richard Dobson, Catherine Dockree, Daniel Doddington, Steven Dodsworth, Norman Doggett, Thomas Down, Andrew Dunham, Ian Dunham, Anne Dunn, Matthew Dunn, Richard Durbin, Jillian Durham, Ireena Dutta, Ruth Dwyer, Lauren Dyer, Mark Earthrowl, Timothy Eastham, Emma Eastham, Carol Edwards, Karen Edwards, Andrew Ellington, David Elliott, Matthew Ellwood, Becky Emberson, Helen Errington, Gareth Evans, John Evans,

Katie Evans, Richard Evans, Eduardo Eyras, Louisa Faulkner, Charlotte Fellingham, Theresa Feltwell, Stephen Fennell, Robert Finn, Tina Flack, Claire Felming, Kerry Fleming, Jonathan Flint, Mark Flint, Yvonne Floyd, Simon Footman, John Fowler, Deborah Frame, Matthew Francis, Stephen Francis, Adam Frankish, John Frankland, Audrey Fraser, David Fraser, Lisa French, David Fricker, Daniel Frost, Jackie Frost, Lorna Frost, Carole Frost, Liam Fuller, Kathryn Fullerton, Alison Gardner, Patrick Garner, Jane Garnett, Leigh Gatland, Lindsay Gatland, Jilur Ghori, Ben Gibbs, Diane Gibson, Elizabeth Gibson, James Gilbert, Lisa Gilby, Christopher Gillson, Rebecca Glithero, April Gooderham, Matthew Gorton, Darren Grafham, Michael Grant, Susan Grant, Iain Gray, Emma Gray, Lisa Green, James Greenhalgh, Joe Greenhill, Sam Griffiths-Jones, Philippa Gregg, Simon Gregory, Susan Gribble, Coline Griffiths, Ed Griffiths, Mark Griffiths, Russell Grocock, Ian Guthrie, Rhian Gwilliam, Rebekah Hall, Karen Halls, Gretta Hall-Tamly, John Hamlett, Sian Hammond, Julie Hancock, Adam Harding, Joanne Harley, David Harper, Georgina Harper, Patrik Harper, Grant Harradence, Charlene Lou Harrison, Elliott Harrison, Ruth Harrison, Elizabeth Hart, Daniel Hassan, Natalie Hawkins, Kellie Hawley, Kerry Hayes, Paul Heath, Rosemary Heathcott, Cathy Hembry, Carl Henderson, Tim Herd, Stephen Hewitt, Douglas Higgs, Guy Hillyard, Russell Hinkins, Sara Jane Ho, David Hodgson, Michael Hoffs, Jane Holden, Janet Holdgate, Ele Holloway, Ian Holmes, Sarah Holmes, Simon Holroyd, Alison Hooper, Lucy Hopewell, Ben Hopkins, Gary Hornett, Geoff Hornsby, Tony Hornsby, Sharon Horsley, Roger Horton, Philip Howard, Philip Howden, Kevin Howe, Gareth Howell, Timothy Hubbard, Elizabeth Huckle, Jaime Hughes, Jennifer Hughes, Louisa Hull, Holger Hummeric, Sean Humphray, Matthew Humphries, Adrienne Hunt, Paul Hunt, Sarah Hunt, Giselle Hunter, David Hyde, Michael Ince, Judith Isherwood, Vivek Iyer, Janet Izatt, Monica Izmajlowicz, Niclas Jareborg, Bijay Jassal, Grant Jeffery, Kim Jeffery, Colin Jeffrey, Kerstin Jekosch, Lee Jenkins, Tina Johansen, Cheryl Johnson, Christopher Johnson, David Johnson, Keith Jolley, Abigail Jones, Claire Jones, Juliet Jones, Matthew Jones, Michael Jones, Steven Jones, Shirin Joseph, Ann Joy, Linsey Joy, Victoria Joy, Gillian Joyce, Mark Jubb, Kanchi Karunaratne, Michael Kay, Danielle Kaye, Lyndal Kearney, Stephen Keenan, Simon Kelley, Joanna Kershaw, Ross Kettleborough, Cathy Kidd, Peter Kierstan, Andrew Kimberley, Andrew King, Simon Kingsley, Colin Kingswood, Gillian Klingle, Andrew Knights, Ian Korf, Anders Krogh, Heena Lad, Philip Laidlaw, Michael Laing, Gavin Laird, Christine Lambart, Ralph Lamble, Cordelia Langford, Ben Larke, Timun Lau, Stephanie Lawlor, Sampsa Leather, Minna Lehvaslaiho, Johannes Lemke, Steven Leonard, Daniel Leongamornlert, Margaret Leversha, Julia Lightning, Sarah Lindsay, Matthew Line, Sally Linsdell, Peter Little, Christine Lloyd, David Lloyd, Victoria Lock, William Lock, Anne Lodziak, Ian Longden, Howard Loraine, Rachel Lord, Jane Loveland, Jamie Lovell, Georgina Lye, Neil Marriott, Anna Marrone, Paul Marsden, Victoria Marsh, Matthew Martin, Sancha Martin, Gareth Maslen, Debbie Mason, Lucy Matthews, Paul Matthews, Nick Matthews, Madalynne Maynard, Owen McCann, Joseph McClay, Craig McCollum, Louise McConnachie, Bill McDonald, Louise McDonald, Jennifer McDowall, Sarah McGuire, Carole McKeown, Stuart

McLaren, Kirsten McLay, James McLean, John McMurdo, Amanda McMurray, Des McMurray, Natalie McWilliams, Nalini Mehta, Patrick Meidl, Noel Menuge, Simon Mercer, Asab Miah, Gos Micklem, Simon Miles, Sarah Milne, Dippica Mistry, Shailesh Mistry, Jake Mitchell, Jeff Mitchell, Maryam Mohammadi, Christophe Molina, Paul Mooney, Madeline Moore, Andrea Moreland, Beverley Mortimore, Richard Mott, Ian Mullenger, Jim Mullikin, Brian Munday, Elaine Munday, Andy Mungall, Clare Murnane, Kerry Murrell, Alison Myers, David Negus, Bee Ng, David Niblett, Jonathan Nicholson, Tim Nickerson, Sukhjit Nijjar, Zemin Ning, James Nisbet, Karen Novik, Christopher Odell, Daniel O'donovan, Francess Ogbighele, Tom Oinn, Hayley Oliver, Karen Oliver, Helena Orbell, Anthony Osborn, Joan Osborne, Emma Overton-Larty, Sophie Palmer, Richard Pandian, Adrian Parker, Christopher Parkin, Kim Parkin, Ginny Parry-Brown, Dina Patel, Ritesh Patel, Alexandra Pearce, Danita Pearson, Anna Peck, Richard Peck, John Peden, Sarah Pelan, Chantal Percy, Andrew Perito, Isabelle Perrault, Anna Peters, Roger Pettett, Ben Phillimore, Kim Phillips, Samantha Phillips, Darren Platt, Emma Playford, Bob Plumb, Matthew Pocock, Keith Porter, Tarryn Porter, Christopher Potter, Simon Potter, Don Powell, Radhika Prathalingham, Elena Prigmore, Michael Quail, Hanna Quarrie, Chris Quince, Matloob Qureshi, Helen Ramsay, Yvonne Ramsey, Sally Ranby, Richard Rance, Vikki Rand, Joanne Ratford, Lewis Ratford, Daniel Read, Donald Redhead, Richard Redon, Christine Rees, Mary Reid, Astrid Reinhardt, Alex Rice, Catherine Rice, Peter Rice, Suzzanne Richard, Susan Richardson, Kerry Ridler, Lyn Riethoven, Rachel Rigby, Melanie Robinson, Rebecca Rochford, Jane Rogers, Lisa Rogers, Hugh Ross, Mark Ross, Angela Rule, James Rule, Ben Russell, Jayne Rutter, Kamal Safdar, Natalie Salter, Javier Santoyo-Lopez, David Saunders, Carol Scott, Deborah Scott, Ian Scott, Fiona Seager, Margaret Searle, Paul Searle, Stephen Searle, Harminder Sehra, Joe Shakespeare, Jason Shardelow, Greg Sharp, Teresa Shaw, Charles Shaw-Smith, Jennifer Shearing, Karen Sheppard, Richard Sheppard, Elizabeth Sheridan, Ratna Shownkeen, Richard Silk, Matthew Sims, Sarah Sims, Shanthi Sivadasan, Carl Skuce, Luc Smink, Andrew Smith, Laura Smith, Lorraine Smith, Michelle Smith, Russell Smith, Stephanie Smith, James Smith, Hannah Sneath, Cari Soderlund, Victor Solovyev, Erik Sonnhammer, Elizabeth Sotheran, William Spooner, Lee Spraggon, Janet Squares, Suzanna Squares, Michael Stables, James Stalker, Steve Stamford, Melanie Stammers, Helen Steingruber, Yvonne Stephens, Charles Steward, Aengus Stewart, Michael Stewart, Ian Still, Mo Stock, Lisa Stoppard, Philip Storey, Roy Storey, Sarah Stowe, Carol Strachan, Greg Strachan, Claire Stribling, John Sturdy, John Sulston, Chris Swainson, Mark Swann, David Swarbreck, Neil Sycamore, Matthew Tagney, Steven Tan, Elizabeth Tarling, Amy Taylor, Gillian Taylor, Kate Taylor, Ruth Taylor, Sam Taylor, Susan Taylor, Louise Tee, Julieanne Tester, Andrew Theaker, David Thexton, Craig Thomas, Daniel Thomas, Karen Thomas, Ruth Thomas, Roselin Thommai, Andrea Thorpe, Karen Thorpe, Glen Threadgold, Scott Thurston, Emma Tinsley, Alan Tracey, Jonathan Travers, Anthony Tromans, Ben Tubby, Cristina Tufarelli, Kathryn Turney, Darren Upson, Zoe Van Helmond, Mark Vaudin, Ramya Viknaraja, Wendy Vine, Paul Voak, David Vollenhoven, Sarah Walker, Melanie Wall, Justine Wallis, Michelle Wallis, Graham Warren,

Georgina Warry, Andy Watson, Nicola Watt, Anthony Webb, Jeannette Webb, John Weir, Alan Wells, Sarah Wells, Robert Welton, Paul West, Tony West, Angela Wheatley, Carl Wheatley, Gideon Wheeler, Hayley Whitaker, Adam White, Amelia White, Brian White, Johnathon White, Simon White, Sally Whitehead, Matthew Whiteley, Pamela Whittaker, Adam Whittaker, Sara Widaa, Anna Wild, Jane Wilkinson, Paul Wilkinson, David Willey, Andy Williams, Bill Williams, Leanne Williams, Helen Williamson, Tamsin Wilmer, Laurens Wilming, Brian Wilson, Gareth Wilson, Margaret Wilson, Nyree Wilson, Siobhan Wilson, Wendy Wilson, Piers Wilson, Philip Window, Jenny Winster, Claire Winzar, James Witt, Fred Wobus, Emma Wood, Joe Wood, Sharon Woodeson, Kathryn Woodfine, Rebecca Woodhouse, Rebecca Woodmansey, Richard Wooster, Matthew Wray, Paul Wray, Charmain Wright, Kathrine Wright, Debbie Wright, Julia Wyatt, Jane Xie, Louise Young, Sheila Young, Ruth Younger, Shenru Zhao, Amanda Abbott-Ozersky, Amber Isak, Amy Berghoff, Amy D. Reily, Andrea Holmes, Andrew Levy, Andrew Van Brunt, Aniko Sabo, Anthony Harris, Anu Desai, Asif Chinwalla, Aye M. Tin-Wollam, Betty Lamar, Brian Mailey, C. Richard Harkins, Caryn Wagner-Mcpherson, Catherine Marquis-Homeyer, Catrina Fronick, Chad Tomlinson, Charlene Pearman, Christine Nguyen, Chunyan Wang, Colin Kremitzki, Craig Pohl, Cynthia Strong, Dan Layman, Dan Bentley, Darin Blasiar, David Dooling, Delali Buatsi, Douglas Johnson, Edward A. Belter, Edward Paulson, Elaine R. Mardis, Elizabeth Boatright, Ernest Goyea, Feiyu Du, George Hu, Glendoria Elliott, Holly Bradshaw-Cordum, Hui Du, Hui Sun, James Randolph, Jason Carter, Jason E. Waligorski, Jason Maas, Jeffrey Woessner, Jennifer Edwards, Jennifer Murray, Jennifer Randall-Maher, Joanne Nelson, Joelle Kalicki-Veizer, Johar Ali, John McPherson, John Spieth, John Wallace, Jon Armstrong, Joshua Heyen, Kelly Carpenter, Kelie Kang, Kelly Mead, Kelsi Scott, Kimberley Delhaunty, Kyriena L. Schatzkamer, Kris Wylie, Krista Haglund, Kym Hallsworth-Pepin, Kyung Kim, Lachlan Oddy, Ladeana Hillier, Laura P. Courtney, Lauren Caruso, Lee Trani, Li Ding, Lucinda A. Fulton, Maria Cedroni, Marco Marra, Mark Johnson, Martin Yoakum, Matt Cordes, Maxim Radionenko, Michael Becker, Michael D. McLellan, Mike Nhan, Mike Wendl, Mundeep Sekhon, Nancy Miller, Neenu Grewal, Neha Shah, Nicolas Berkowicz, Patrick J. Minx, Patty Wohldmann, Phil Latreille, Philip Ozersky, Prashant Sinha, Rachel Maupin, Rekha Meyer, Rick Meyer, Richard K. Wilson, Robert S. Fulton, Robert H. Waterston, Sandra W. Clifton, Sara Kohlberg, Sara Jaeger, Shiaw Pyng Yang, Scott Abbott, Scott Martinka, Scott S. Kruchowski, Sharhonda Swearengen-Shahid, Shawn Leonard, Shunfang Hou, Stacie Gattung, Stephanie Andrews, Stephanie Chissoe, Susan M. Rock, Tamberlyn Bieri, Theresa Rohlfing, Tina A. Graves, Tony Gaige, Tracie Miner, William Nash, Yoram Shotland, Eric S. Lander, Chad Nusbaum, Bruce Birren, Kerstin Lindblad-Toh, Jill P. Mesirov, Robert Nicol, Michael C. Zody, Jean L. Chang, Christina A. Cuomo, Ken Dewar, Mike Fitzgerald, David B. Jaffe, Xiaoping Yang, Amr Abouelleil, Nicole R. Allen, Harindra Arachchi, Jennifer Baldwin, Toby Bloom, Mark Borowsky, Boris Boukhgalter, Jon Butler, Bruno Chazaro, Mieke Citroen, April Cook, Ben Corum, Kurt Dearellano, Kathy Dooley, Lester Dorris, Matthew Endrizzi, Jeffery Erickson, Gary Gearin, Sante Gnerre, Andreas Gnirke, Nabil Hafez, Daniel

Hagopian, Jennifer Hall, Catherine Hosage-Norman, Charlien Jones, Michael Kamal, Asha Kamat, Chinnappa Kodira, Patrick Kuharic, Kurt Labutti, Teresa Lai, Jessica Lehoczky, Rosie Levine, Xiaohong Liu, Tashi Lokyitsang, Annie Lui, Richard Mabbitt, Pendexter Macdonald, John Major, Jonathan Manning, Lisa Marinelli, Charles D. Matthews, Evan Mauceli, Atanas Mihalev, Glen Munson, Jerome Naylor, Cindy Nguyen, Paula O'donnell, Sinead O'leary, Keith O'neill, Stephen Parker, Bruno Piqani, Anthony Rachupka, Umadevi Ramasami, Christina Raymond, Joseph Rodriguez, Rebecca Schupbach, Christopher Seaman, Ted Sharpe, Andrew Sheridan, Cherylyn Smith, Carrie Sougnez, Sabrina Stone, Matthew Stubbs, Jessica Talamas, Pema Tenzin, Kerri Topham, Sarah Towey, Vijay Venkataranan, Charles Whittaker, Andrew Zimmer, Emmanuel Adekoya, Mostafa Ait-Zahra, Alla Ali, Thaddeus Allen, Mechele Anderson, Scott Anderson, Jeff Armbruster, Pasang Bachatsang, Tashi Bayul, Berta Blitshsteyn, Jason Blye, Leonid Boguslavskiy, Jody Camarata, Kevin Campo, Yama Cheshatsang, Al Collymore, Tony Considine, Patrick Cooke, Robert David, Tenzin Dawoe, Stewart Degray, Chemey Dhongsar, Passang Dorje, Kunsang Dorjee, Noah Duffy, Alan Dupes, Abderrahim Farina, Susan Faro, Pat Ferriera, Heather Fischer, Sheila Fisher, Karen Foley, Diane Gage, Steph Gardyna, Seth Gordon, Audra Goyette, Joe Graham, Edward Grandbois, Kunsang Gyaltsen, Birhane Hagos, Beah Hatcher, Andrew Heaford, Andrew Heller, Tracy Honan, Nathan Houde, Lee Hughes, Bill Hulme, Ilian Iliev, Cristyn Kells, Alix Kieu, Peter Kisner, Dawa Lama, Tom Landers, Jean Pierre Leger, Doreen Lewis, Tammy Lewis, Yeshi Lokyitsang, Christine Maclean, Dick Marabella, Kebede Maru, Megan McCarthy, Tina McGhee, James Meldrim, Louis Meneus, Tanya Mihova, Val Mlenga, Leon Mulrain, Christian Newes, Na Nguyen, Thu Nguyen, Chou Dolma Norbu, Nyima Norbu, Ose Okoawo, Jennifer O'loughlin, Bamidele Omotosho, Sahal Osman, Tenzin Phulchung, Pema Phunkang, Nadia Pierre, Rayale Rameau, Vernada Ray, Cecil Rise, Peter Rogov, Steven Seaman, Ngawang Sherpa, Brian Spencer, John Stalker, Nicole Stange-Thomann, Sharon Stavropoulos, Kevin Stetson, Casey Stone, Pierre Tchuinga, Senait Tesfaye, Joumathe Theodore, Yama Thoulutsang, Tsamla Tsamla, Nawang Tsomo, Duane Valle, Helen Vassiliev, Rose Veil, Andy Vo, Tsering Wangchuck, Tsering Wangdi, Dudley Wyman, Shailendra Yadav, Shane Yeager, Rahel Retta Yeshitela, Geneva Young, Joanne Zainoun, Lisa Zembeck, Anne Abrajano, Catherine Adam, Aaron Adamson, Andrea Aerts, Arun Aggarwal, Dana Alcivare, Michelle Alegria, Michelle Allegria-Hartmann, Jennifer Alleman, Susan Allen, Michael R. Altherr, Chris Amemiya, Gina Amico-Keller, Janice Andora, Carol Andredesz, Tim Andriese, Sinoula Apostolou, Terisita Arcaina, Marlon Arcaina, Andre Arellano, Perry Arellano-Jones, Linda K. Ashworth, Cara Aslanidis, Lawreen Asuncion, Cristina Attix, Aaron Avila, Julie Avila, Phil Bach, Hummy Badri, Eva Bajorek, M. Baker, Richard Baker, Dan Baker, Michele Bakis, Michael Banda, Suba Basu, Mark Batzer, Jason Baumohl, Keith Beall, Leslie Beaucham, George I. Bell, J. Benke, William Benner, Douda Bensasson, John Bercovitz, Anne Bergmann, Tony J. Beugelsdijk, Rita Bhakta, Aleksandr Bituin, Stacey Black, Robert Blazej, Heather Blumer, John Boehm, Juanan Boen, Marnel Bondoc, Eric Bowen, Brigitte F. Brandriff, Wade Brannon, Elbert Branscomb, Thomas Brettin, Michael Bridgers, Peter

Brokstein, Nancy C. Brown, Robert Bruce, David C. Bruce, William J. Bruno, Jennifer E. Bryant, Judith M. Buckingham, Kerem Bulbul, Nathan Bunker, Matt Burgin, Karolyn Burkhart-Schultz, Paul Butler, Sean Caenepeel, David F. Callen, Connie S. Campbell, Evelyn W. Campbell, Mary L. Campbell, Chenier Caoile, Lolo Cardenas, Faviola Cardenas, Anthony V. Carrano, Jason Carriere, Mario Cepeda, Patrick Chain, Jean F. Challacombe, Yee Man Chan, Jarrod Chapman, Jeffrey Chapple, Leslie A. Chasteen, Simeon Chavarria, Jin Chen, Ray Chen, Chira Chen, Jan Fang Cheng, Olga Chertkov, Han C. Chi, Sylvia Chin, Corey Chinn, Sally Chiu, Mari Christensen, A. Chung, Michael J. Cinkowsky, Lynn Clark, Lynn M. Clark, David Cleveland, Jackie Cofield, Judith D. Cohn, Rick Colayco, Jessie Combs, Robin Comstock, Karen Connolly, Alex Copeland, Rebecca Cordray, Earl Cornell, Sara Cotton, Olivier Couronne, Terrance Critchlow, Paul Critz, Stephanie Cummings, Eileen Dalin, Linda Danganan, Christopher Daum, Stuart Davidson, Cheryl Davis, Larry L. Deaven, Kerry Deere, David Degusta, Maria Deguzman, Paramvir Dehal, Pieter De Jong, Mirian Denys, John C. Detter, Laurie Devlin, Jennifer Dias, Victoria Dias, Genevieve Dibartolo, Mark Dickson, Jeanne Dietz-Band, Richard Digennaro, Karen Dilts, Mira Dimitrijevic-Bussod, Kami Dixon, Long Do, Norman A. Doggett, Kimberly Dong, Victor Dorsett, Suzanne Duarte, Inna Dubchak, Erin Dunwell, Evan Eichler, Chris Elkin, Jeffrey Elliott, Ger van den Engh, David Engle, Anne Marie Erler, Julio Escobar, Gerald Eveleth, Joseph J. Fawcett, Alicia Ferguson, Anne Fertitta, Jim Fey, Patrick Fidel, Elizabeth Fields, Marie Fink, Laurice Fischer, Pat Fitch, J. Flanagan, Dave Flowers, Peg Folta, Jonathan Fong, Amanda F. Ford, Dea Fotopulos, Matthew Fourcade, Jessica Fox, Ken Frankel, Marvin Frazier, Jane Fridlyand, Sharin Fuller, Celsa Gallegos, Michael Galvez, Stuart Gammon, Emilio Garcia, Carmen Garcia, Antonio Garcia-Martinez, Consuelo Garduno, Jeffrey A. Garnes, Joseph M. Gatewood, Binyam Gebreyesus, Cynthia Gelein, Maartin Gelpke, Anca Georgescu, Amy Geotina, Mary Gifford, Isaias Gil, Paul Gilna, Jeff Gingrich, Kira Giovanelli, Tijana Glavina, Darren Gold, Kristen Golinveaux, Maria Gomez, Eidelyn Gonzales, David Goodstein, Lynne A. Goodwin, Laurie A. Gordon, Catherine Gordon, Ann Gorvad, Christine Gould, Deborah L. Grady, Jennifer Grant, Bruce Gray, Jeffrey K. Griffith, Igor Grigoriev, Jane Grimwood, Matthew Groza, Hannibal Guarin, Metzalli Guel, Kate Gunning, Jane Guo, Chi Ha, Allen Haim, Patrick Hajek, Matt Hamilton, Nancy Hammon, Sha Hammond, Khalid Hamza, Shunsheng Han, Trevor Hawkins, Lauren Haydu, Keven Helfenbein, Uffe Hellsten, Nina Henderson, David Hendrix, Karla Henning, Victor Hepa, Charles Herman, Carl E. Hildebrand, Issac Ho, Susan M.G. Hoffman, Ann Holtz, Caleb Holtzer, Roya Hosseini, Katherine Huang, Wayne Huang, Zhengping Huang, Heather Hue, Hillary Hughes-Hull, Dav. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, 2004.

[2] Nicholas L. Adkins, Hengyao Niu, Patrick Sung, and Craig L. Peterson. Nucleosome dynamics regulates DNA processing. *Nature Structural & Molecular Biology*, 20(7):836–842, 2013.

[3] James Ahrens. ParaView: An End-User Tool for Large Data Visualization. Technical

report, 2005.

[4] Abdollah Allahverdi, Renliang Yang, Nikolay Korolev, Yanping Fan, Curt A. Davey, Chuan Fa Liu, and Lars Nordenskiöld. The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Research*, 39(5):1680–1691, 2011.

[5] J. Allan, P. G. Hartman, C. Crane-robinson, and F. X. Aviles. The structure of histone H1 and its location in chromatin. *Nature*, 288(5792):675–679, 1980.

[6] V. G. Allfrey, R. Faulkner, and A. E. Mirsky. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proceedings of the National Academy of Sciences*, 51(5):786–794, 1964.

[7] L. M. Almassalha, A. Tiwari, P. T. Ruhoff, Y. Stypula-Cyrus, L. Cherkezyan, H. Matsuda, M. A. Dela Cruz, J. E. Chandler, C. White, C. Maneval, H. Subramanian, I. Szleifer, H. K. Roy, and V. Backman. The global relationship between chromatin physical topology, fractal structure, and gene expression. *Scientific Reports*, 7(1):41061, 2017.

[8] Luay M. Almassalha, Greta M. Bauer, John E. Chandler, Scott Gladstein, Lusik Cherkezyan, Yolanda Stypula-Cyrus, Samuel Weinberg, Di Zhang, Peder Thusgaard Ruhoff, Hemant K. Roy, Hariharan Subramanian, Navdeep S. Chandel, Igal Szleifer, and Vadim Backman. Label-free imaging of the native, living cellular nanoarchitecture using partial-wave spectroscopic microscopy. *Proceedings of the National Academy of Sciences*, page 201608198, 2016.

[9] A. Amitai and D. Holcman. Polymer physics of nuclear organization and function, 2017.

[10] Gaurav Arya and Tamar Schlick. A Tale of Tails: How Histone Tails Mediate Chromatin Compaction in Different Salt and Linker Histone Environments †. *The Journal of Physical Chemistry A*, 113(16):4045–4059, 2009.

[11] Gaurav Arya, Qing Zhang, and Tamar Schlick. Flexible Histone Tails in a New Mesoscopic Oligonucleosome Model. *Biophysical Journal*, 91(1):133–150, 2006.

[12] Gaurav Bajpai and Ranjith Padinhateeri. Irregular Chromatin: Packing Density, Fiber Width, and Occurrence of Heterogeneous Clusters. *Biophysical Journal*, 118(1):207–218, 2020.

[13] Gavin D Bascom, Taejin Kim, and Tamar Schlick. Kilobase Pair Chromatin Fiber Contacts Promoted by Living-System-Like DNA Linker Length Distributions and Nucleosome Depletion. *Journal of Physical Chemistry B*, 121(15):3882–3894, 2017.

[14] Gavin D. Bascom, Christopher G. Myers, and Tamar Schlick. Mesoscale modeling reveals formation of an epigenetically driven HOXC gene hub. *Proceedings of the National Academy of Sciences*, 116(11):4955–4962, 2019.

[15] Jan Bednar, Isabel Garcia-Saez, Ramachandran Boopathi, Amber R. Cutter, Gabor Papai, Anna Reymer, Sajad H. Syed, Imtiaz Nisar Lone, Ognyan Tonchev, Corinne Crucifix, Hervé Menoni, Christophe Papin, Dimitrios A. Skoufias, Hitoshi Kurumizaka, Richard Lavery, Ali Hamiche, Jeffrey J. Hayes, Patrick Schultz, Dimitar Angelov, Carlo Petosa, and Stefan Dimitrov. Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Molecular Cell*, 66(3):384–397.e8, 2017.

[16] Bruno Beltran, Deepti Kannan, Quinn Macpherson, and Andrew J. Spakowitz. Geometrical Heterogeneity Dominates Thermal Fluctuations in Facilitating Chromatin Contacts. *Physical Review Letters*, 123(20):208103, 2019.

[17] Morgan Bernier, Yi Luo, Kingsley C Nwokelo, Michelle Goodwin, Sarah J Dreher, Pei Zhang, Mark R Parthun, Yvonne Fondufe-Mittendorf, Jennifer J Ottesen, and Michael G Poirier. Linker histone H1 and H3K56 acetylation are antagonistic regulators of nucleosome dynamics. *Nature Communications*, 6:10152, 2015.

[18] Bogdan Bintu, Leslie J. Mateo, Jun-Han Su, Nicholas A. Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N. Boettiger, and Xiaowei Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, 2018.

[19] Mithun Biswas, Jörg Langowski, and Thomas C. Bishop. Atomistic simulations of nucleosomes. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(4):378–392, 2013.

[20] Melissa J. Blacketer, Sarah J. Feely, and Michael A. Shogren-Knaak. Nucleosome interactions and stability in an ordered nucleosome array model system. *Journal of Biological Chemistry*, 285(45):34597–34607, 2010.

[21] Th Boveri. Die blastomerenkerne von ascaris megalocephala. *Archiv für Zellforschung*, 3:181, 1909.

[22] Gregory D Bowman. Mechanisms of ATP-dependent nucleosome sliding. *Current Opinion in Structural Biology*, 20(1):73–81, 2010.

[23] Gregory D Bowman and Michael G Poirier. Post-translational modifications of histones that influence nucleosome dynamics, 2015.

[24] C. A. Brackley, J. Johnson, D. Michieletto, A. N. Morozov, M. Nicodemi, P. R. Cook, and D. Marenduzzo. Nonequilibrium Chromosome Looping via Molecular Slip Links. *Physical Review Letters*, 119(13):138101, 2017.

[25] Kristin Brogaard, Liqun Xi, Ji Ping Wang, and Jonathan Widom. A map of nucleosome positions in yeast at base-pair resolution. *Nature*, 486(7404):496–501, 2012.

[26] James E. Brownell, Jianxin Zhou, Tamara Ranalli, Ryuji Kobayashi, Diane G. Edmondson, Sharon Y. Roth, and C.David Allis. Tetrahymena Histone Acetyltransferase

A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation. *Cell*, 84(6):843–851, 1996.

[27] P. J.G. Butler and Jean O. Thomas. Changes in chromatin folding in solution. *Journal of Molecular Biology*, 140(4):505–529, 1980.

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

[29] Răzvan V. Chereji, Srinivas Ramachandran, Terri D. Bryson, and Steven Henikoff. Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biology*, 19(1):19, 2018.

[30] Tomasz Cierpicki, Laurie E Risner, Jolanta Grembecka, Stephen M Lukasik, Relja Popovic, Monika Omonkowska, David D Shultis, Nancy J Zeleznik-Le, and John H Bushweller. Structure of the MLL CXXC domain–DNA complex and its functional role in MLL-AF9 leukemia. *Nature Structural &38; Molecular Biology*, 17(1):62–68, 2009.

[31] Rosana Collepardo-Guevara, Guillem Portella, Michele Vendruscolo, Daan Frenkel, Tamar Schlick, and Modesto Orozco. Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: A multiscale computational study. *Journal of the American Chemical Society*, 137(32):10205–10215, 2015.

[32] Thomas Cremer, Marion Cremer, Steffen Dietzel, Stefan Müller, Irina Solovei, and Stanislav Fakan. Chromosome territories - a functional nuclear landscape, 2006.

[33] Y. Cui and C. Bustamante. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proceedings of the National Academy of Sciences*, 97(1):127–132, 2000.

[34] 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 Å resolution. *Journal of Molecular Biology*, 319(5):1097–1113, 2002.

[35] Michele Di Pierro, Ryan R. Cheng, Erez Lieberman Aiden, Peter G. Wolynes, and José N. Onuchic. De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proceedings of the National Academy of Sciences*, 114(46):12126–12131, 2017.

[36] Michele Di Pierro, Bin Zhang, Erez Lieberman Aiden, Peter G. Wolynes, and José N. Onuchic. Transferable model for chromosome architecture. *Proceedings of the National Academy of Sciences*, 113(43):12168–12173, 2016.

[37] 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.

[38] 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. *Proceedings of the National Academy of Sciences*, page 201602202, 2016.

[39] Benedetta Dorigo. Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. *Science*, 306(5701):1571–1573, 2004.

[40] Babatunde Ekundayo, Timothy J. Richmond, and Thomas Schalch. Capturing Structural Heterogeneity in Chromatin Fibers. *Journal of Molecular Biology*, 429(20):3031–3042, 2017.

[41] Jochen Erler, Ruihan Zhang, Loukas Petridis, Xiaolin Cheng, Jeremy C. Smith, and Jörg Langowski. The Role of Histone Tails in the Nucleosome: A Computational Study. *Biophysical Journal*, 107(12):2911–2922, 2014.

[42] 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):43–49, 2011.

[43] He Fang, Sijie Wei, Tae Hee Lee, and Jeffrey J Hayes. Chromatin structure-dependent conformations of the H1 CTD. *Nucleic Acids Research*, 44(19):9131–9141, 2016.

[44] Gary Felsenfeld and Mark Groudine. Controlling the double helix. *Nature*, 421(6921):448–453, 2003.

[45] J T Finch and A Klug. Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences*, 73(6):1897–1901, 1976.

[46] Paul. J. Flory and M. Volkenstein. *Statistical mechanics of chain molecules*, volume 8. Wiley, 1969.

[47] Gordon S. Freeman, Daniel M. Hinckley, Joshua P. Lequieu, Jonathan K. Whitmer, and Juan J. De Pablo. Coarse-grained modeling of DNA curvature. *Journal of Chemical Physics*, 141(16), 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), 2014.

[49] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, 15(9):2038–2049, 2016.

[50] Jonas J Funke, Philip Ketterer, Corinna Lieleg, Sarah Schunter, Philipp Korber, and Hendrik Dietz. Uncovering the forces between nucleosomes using DNA origami. *Science Advances*, 2(11):e1600974–e1600974, 2016.

[51] Min Gao, Philippe S. Nadaud, Morgan W. Bernier, Justin A. North, P. Chris Hammel, Michael G. Poirier, and Christopher P. Jaroniec. Histone H3 and H4 N-terminal tails in nucleosome arrays at cellular concentrations probed by magic angle spinning NMR spectroscopy. *Journal of the American Chemical Society*, 135(41):15278–15281, 2013.

[52] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. Technical report, 2012.

[53] Sergei A Grigoryev, Gaurav Arya, Sarah Correll, Christopher L Woodcock, and Tamar Schlick. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proceedings of the National Academy of Sciences*, 106(32):13317–13322, 2009.

[54] Anja Groth, Walter Rocha, Alain Verreault, and Geneviève Almouzni. Chromatin challenges during DNA replication and repair. *Cell*, 128(4):721–33, 2007.

[55] Jeffrey C. Hansen, Juan Ausio, Valerie H. Stanik, and K. E. Van Holde. Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1. *Biochemistry*, 28(23):9129–9136, 1989.

[56] Sean W Harshman, Nicolas L Young, Mark R Parthun, and Michael A Freitas. H1 histones: Current perspectives and challenges, 2013.

[57] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. Technical report, 2016.

[58] 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. *Journal of Chemical Physics*, 139(14), 2013.

[59] Denes Hnisz, Abraham S. Weintrau, 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.

[60] Tsung-Han S Hsieh, Geoffrey Fudenberg, Anton Goloborodko, and Oliver J. Rando. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nature Methods*, 13(12):1009–1011, 2016.

[61] 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.

[62] Kai Huang, Vadim Backman, and Igal Szleifer. Interphase chromatin as a self-returning random walk: Can DNA fold into liquid trees? *bioRxiv*, page 413872, 2018.

[63] Gyung Hye Huh, Barbara Damsz, Tracie K. Matsumoto, Muppala P. Reddy, Ana M. Rus, José I. Ibeas, Meena L. Narasimhan, Ray A. Bressan, and Paul M. Hasegawa. Salt causes ion disequilibrium-induced programmed cell death in yeast and plants. *Plant Journal*, 29(5):649–659, 2002.

[64] Thomas Jenuwein. Translating the Histone Code. *Science*, 293(5532):1074–1080, 2001.

[65] Daniel Jost, Pascal Carrivain, Giacomo Cavalli, and Cédric Vaillant. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*, 42(15):9553–9561, 2014.

[66] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976.

[67] P.-Y. Kan, T. L. Caterino, and J. J. Hayes. The H4 Tail Domain Participates in Intra- and Internucleosome Interactions with Protein and DNA during Folding and Oligomerization of Nucleosome Arrays. *Molecular and Cellular Biology*, 29(2):538–546, 2009.

[68] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

[69] Hiroo Kenzaki, Nobuyasu Koga, Naoto Hori, Ryo Kanada, Wenfei Li, Kei-ichi Okazaki, Xin-Qiu Yao, and Shoji Takada. CafeMol: A Coarse-Grained Biomolecular Simulator for Simulating Proteins at Work. *Journal of Chemical Theory and Computation*, 7(6):1979–1989, 2011.

[70] Roger D. Kornberg. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science*, 184(4139):868–871, 1974.

[71] Elena F. Koslover, Colin J. Fuller, Aaron F. Straight, and Andrew J. Spakowitz. Local Geometry and Elasticity in Compact Chromatin Structure. *Biophysical Journal*, 99(12):3941–3950, 2010.

[72] Maarten Kruithof, Fan-Tso Chien, Andrew Routh, Colin Logie, Daniela Rhodes, and John van Noort. Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nature Structural & Molecular Biology*, 16(5):534–540, 2009.

[73] Siavash K. Kurdistani and Michael Grunstein. Histone acetylation and deacetylation in yeast. *Nature Reviews Molecular Cell Biology*, 4(4):276–284, 2003.

[74] Adam G. Larson, Daniel Elnatan, Madeline M. Keenen, Michael J. Trnka, Jonathan B. Johnston, Alma L. Burlingame, David A. Agard, Sy Redding, and Geeta J. Narlikar. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, 547(7662):236–240, 2017.

[75] Daniel Y. Lee, Jeffrey J. Hayes, Dmitry Pruss, and Alan P. Wolffe. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*, 72(1):73–84, 1993.

[76] Joshua Lequieu, Andrés Córdoba, Joshua Moller, and Juan J. De Pablo. 1CPN: A coarse-grained multi-scale model of chromatin. *Journal of Chemical Physics*, 150(21):215102, 2019.

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

[78] Joshua Lequieu, David C. Schwartz, and Juan J. de Pablo. In silico evidence for sequence-dependent nucleosome sliding. *Proceedings of the National Academy of Sciences*, page 201705685, 2017.

[79] Sanford H Leuba, Mikhail A Karymov, Yanzhang Liu, Stuart M Lindsay, and Jordanka Zlatanova. Mechanically stretching single chromatin fibers. *Gene Therapy and Molecular Biology Gene Ther Mol Biol*, 4(4):297–301, 1999.

[80] Bing Li, Michael Carey, and Jerry L. Workman. The Role of Chromatin during Transcription. *Cell*, 128(4):707–719, 2007.

[81] Wei Li, Ping Chen, Juan Yu, Liping Dong, Dan Liang, Jianxun Feng, Jie Yan, Peng Ye Wang, Qing Li, Zhiguo Zhang, Ming Li, and Guohong Li. FACT Remodels the Tetranucleosomal Unit of Chromatin Fibers for Gene Transcription. *Molecular Cell*, 64(1):120–133, 2016.

[82] 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*, 108(9):3504–3509, 2011.

[83] Yue Li, Eric Roth, Vasundhara Agrawal, Adam Eshein, Jane Fredrick, Luay Almassalha, Anne Shim, Reiner Bleher, Vinayak P Dravid, and Vadim Backman. Quantifying Three-dimensional Chromatin Organization Utilizing Scanning Transmission Electron Microscopy: ChromSTEM. *bioRxiv*, c:636209, 2019.

[84] 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, Leonid A Mirny, Eric S Lander, and Job Dekker. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326(5950):289–293, 2009.

[85] 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, 1998.

[86] Xiang J. Lu and Wilma K. Olson. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research*, 31(17):5108–5121, 2003.

[87] Karolin Luger, Armin W Mäder, Robin K. Richmond, David F Sargent, and Timothy J Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648):251–260, 1997.

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

[89] Yuri L. Lyubchenko. Nanoscale nucleosome dynamics assessed with time-lapse AFM, 2014.

[90] Quinn MacPherson, Bruno Beltran, and Andrew J. Spakowitz. Bottom–up modeling of chromatin segregation due to epigenetic modifications. *Proceedings of the National Academy of Sciences*, 115(50):12739–12744, 2018.

[91] Quinn MacPherson, Bruno Beltran, and Andrew J. Spakowitz. Chromatin Compaction Leads to a Preference for Peripheral Heterochromatin. *Biophysical Journal*, 118(6):1479–1488, 2020.

[92] Kazuhiro Maeshima, Saera Hihara, and Mikhail Eltsov. Chromatin structure: does the 30-nm fibre exist in vivo? *Current Opinion in Cell Biology*, 22(3):291–297, 2010.

[93] Kazuhiro Maeshima, Satoru Ide, Kayo Hibino, and Masaki Sasai. Liquid-like behavior of chromatin. *Current Opinion in Genetics and Development*, 37:36–45, 2016.

[94] Kazuhiro Maeshima, Ryosuke Imai, Sachiko Tamura, and Tadasu Nozaki. Chromatin as dynamic 10-nm fibers. *Chromosoma*, 123(3):225–237, 2014.

[95] Gerald S. Manning. Limiting laws and counterion condensation in polyelectrolyte solutions. 7. Electrophoretic mobility and conductance. *Journal of Physical Chemistry*, 85(11):1506–1515, 1981.

[96] John F. Marko and Eric D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, 1995.

[97] D. Michieletto, E. Orlandini, and D. Marenduzzo. Polymer model with epigenetic recoloring reveals a pathway for the de novo establishment and 3D organization of chromatin domains. *Physical Review X*, 6(4), 2016.

[98] Atsushi Miyagi, Toshio Ando, and Yuri L Lyubchenko. Dynamics of nucleosomes assessed with time-lapse high-speed atomic force microscopy. *Biochemistry*, 50(37):7901–7908, 2011.

[99] Joshua Moller, Walter Alvarado, Yue Li, Vadim Backman, and Juan J. de Pablo. Physically-informed deep learning reconstruction of chromatin.

[100] Joshua Moller and Juan J. de Pablo. Bottom-Up Meets Top-Down: The Crossroads of Multiscale Chromatin Modeling, 2020.

[101] Joshua Moller, Joshua Lequieu, and Juan J. De Pablo. The Free Energy Landscape of Internucleosome Interactions and Its Relation to Chromatin Fiber Structure. *ACS Central Science*, 5(2):341–348, 2019.

[102] Joshua Moller, Joshua P. Lequieu, and Juan J. de Pablo. Chromatin Structure Beyond the Nucleosome: A Trinucleosome Building Block of the Chromatin Fiber.

[103] Joshua Moller, Michael A. Webb, and Juan J. de Pablo. Equilibrium Features of the Chromatin Fiber Predict TAD-like Structures.

[104] Felix Mueller-Planitz, Henrike Klinker, and Peter B Becker. Nucleosome sliding mechanisms: new twists in a looped history. *Nature Structural & Molecular Biology*, 20(9):1026–1032, 2013.

[105] 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.

[106] Thuy T.M. Ngo and Taekjip Ha. Nucleosomes undergo slow spontaneous gaping. *Nucleic Acids Research*, 43(8):3964–3971, 2015.

[107] Davood Norouzi and Victor B Zhurkin. Topological polymorphism of the two-start chromatin fiber. *Biophysical Journal*, 108(10):2591–2600, 2015.

[108] Johannes Nuebler, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid Mirny. Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation. *bioRxiv*, page 196261, 2017.

[109] Masae Ohno, Tadashi Ando, David G. Priest, Vipin Kumar, Yamato Yoshida, and Yuichi Taniguchi. Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs. *Cell*, 176(3):520–534.e25, 2019.

[110] A. L. Olins and D. E. Olins. Spheroid Chromatin Units (ngr Bodies). *Science*, 183(4122):330–332, 1974.

[111] Jason Otterstrom, Alvaro Castells-Garcia, Chiara Vicario, Pablo A Gomez-Garcia, Maria Pia Cosma, and Melike Lakadamyali. Super-resolution microscopy reveals how histone tail acetylation affects DNA compaction within nucleosomes in vivo. *Nucleic Acids Research*, (1), 2019.

[112] Horng D. Ou, Sébastien Phan, Thomas J. Deerinck, Andrea Thor, Mark H. Ellisman, and Clodagh C. O'Shea. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349):eaag0025, 2017.

[113] Sharon Pepenella, Kevin J Murphy, and Jeffrey J Hayes. A distinct switch in interactions of the histone H4 tail domain upon salt-dependent folding of nucleosome arrays. *Journal of Biological Chemistry*, 289(39):27342–27351, 2014.

[114] Sharon Pepenella, Kevin J Murphy, and Jeffrey J Hayes. Intra- and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure, 2014.

[115] Ognjen Perišić, Stephanie Portillo-Ledesma, and Tamar Schlick. Sensitive effect of linker histone binding mode and subtype on chromatin condensation. *Nucleic Acids Research*, 2019.

[116] C. Phelps, W. Lee, D. Jose, P. H. von Hippel, and A. H. Marcus. Single-molecule FRET and linear dichroism studies of DNA breathing and helicase binding at replication fork junctions. *Proceedings of the National Academy of Sciences*, 110(43):17320–17325, 2013.

[117] Andrea Piunti, Rintaro Hashizume, Marc A Morgan, Elizabeth T Bartom, Craig M Horbinski, Stacy A Marshall, Emily J Rendleman, Quanhong Ma, Yoh-hei Takahashi, Ashley R Woodfin, Alexander V Misharin, Nebiyu A Abshiru, Rishi R Lulla, Amanda M Saratsis, Neil L Kelleher, C David James, and Ali Shilatifard. Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. *Nature Medicine*, 23(4):493–500, 2017.

[118] Andrea Piunti and Ali Shilatifard. Epigenetic balance of gene expression by polycomb and compass families. *Science*, 352(6290):aad9780, 2016.

[119] S J Plimpton. Fast parallel algorithms for short-range molecular dynamics, 1995.

[120] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, volume 2017-Janua, pages 77–85. Institute of Electrical and Electronics Engineers Inc., 2017.

[121] V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, 362(6417):219–223, 1993.

[122] Suhas S.P. Rao, Su-Chen Huang, Brian Glenn St Hilaire, Jesse M Engreitz, Elizabeth M Perez, Kyong-Rim Kieffer-Kwon, Adrian L Sanborn, Sarah E Johnstone, Gavin D Bascom, Ivan D Bochkov, Xingfan Huang, Muhammad S Shamim, Jaeweon Shin, Douglass Turner, Ziyi Ye, Arina D Omer, James T Robinson, Tamar Schlick, Bradley E. Bernstein, Rafael Casellas, Eric S Lander, and Erez Lieberman Aiden. Cohesin Loss Eliminates All Loop Domains. *Cell*, 171(2):305–320.e24, 2017.

[123] 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, 2014.

[124] Maria Aurelia Ricci, Carlo Manzo, María Filomena García-Parajo, Melike Lakadamyali, and Maria Pia Cosma. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, 160(6):1145–1158, 2015.

[125] T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes, and A. Klug. Structure of the nucleosome core particle at 7 resolution. *Nature*, 311(5986):532–537, 1984.

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

[127] Philip J.J. Robinson, Woojin An, Andrew Routh, Fabrizio Martino, Lynda Chapman, Robert G. Roeder, and Daniela Rhodes. 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. *Journal of Molecular Biology*, 381(4):816–825, 2008.

[128] Philip JJ Robinson and Daniela Rhodes. Structure of the '30nm' chromatin fibre: A key role for the linker histone. *Current Opinion in Structural Biology*, 16(3):336–343, 2006.

[129] Angelo Rosa and Ralf Everaers. Structure and Dynamics of Interphase Chromosomes. *PLoS Computational Biology*, 4(8):e1000153, 2008.

[130] Andrew Routh, Sara Sandin, and Daniela Rhodes. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proceedings of the National Academy of Sciences*, 105(26):8872–8877, 2008.

[131] Adrian L. Sanborn, Suhas S. P. Rao, Su-Chen Huang, Neva C. Durand, Miriam H. Huntley, Andrew I. Jewett, Ivan D. Bochkov, Dharmaraj Chinnappan, Ashok Cutkosky, Jian Li, Kristopher P. Geeting, Andreas Gnirke, Alexandre Melnikov, Doug McKenna, Elena K. Stamenova, Eric S. Lander, and Erez Lieberman Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, 112(47):E6456–E6465, 2015.

[132] Suman Saurabh, Matthew A. Glaser, Yves Lansac, and Prabal K. Maiti. Atomistic Simulation of Stacked Nucleosome Core Particles: Tail Bridging, the H4 Tail, and Effect of Hydrophobic Forces. *Journal of Physical Chemistry B*, 120(12):3048–3060, 2016.

[133] 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):138–141, 2005.

[134] 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(5):887–898, 2008.

[135] Wibke Schwarzer, Nezar Abdennur, Anton Goloborodko, Aleksandra Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A Fonseca, Wolfgang Huber, Christian H. Haering, Leonid Mirny, and Francois Spitz. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551(7678):51–56, 2017.

[136] M. Scott Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics*, 129(14):144108, 2008.

[137] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection. 2019.

[138] 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, 2014.

[139] Brian D Strahl and C. David Allis. The language of covalent histone modifications. *Nature*, 403(6765):41–45, 2000.

[140] Amy R. Strom, Alexander V. Emelyanov, Mustafa Mir, Dmitry V. Fyodorov, Xavier Darzacq, and Gary H. Karpen. Phase separation drives heterochromatin domain formation. *Nature*, 547(7662):241–245, 2017.

[141] Kevin Struhl and Eran Segal. Determinants of nucleosome positioning. *Nature Structural and Molecular Biology*, 20(3):267–273, 2013.

[142] Micah P Stumme-Diers, Siddhartha Banerjee, Mohtadin Hashemi, Zhiqiang Sun, and Yuri L Lyubchenko. Nanoscale dynamics of centromere nucleosomes and the critical roles of CENP-A. *Nucleic Acids Research*, 46(1):94–103, 2018.

[143] Hariharan Subramanian, Prabhakar Pradhan, Yang Liu, Ilker R. Capoglu, Jeremy D. Rogers, Hemant K. Roy, Randall E. Brand, and Vadim Backman. Partial-wave microscopic spectroscopy detects subwavelength refractive index fluctuations: an application to cancer diagnosis. Technical Report 4, 2009.

[144] Shoji Takada, Ryo Kanada, Cheng Tan, Tsuyoshi Terakawa, Wenfei Li, and Hiroo Kenzaki. Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Accounts of Chemical Research*, 48(12):3026–3035, 2015.

[145] F. Thoma and Th Koller. Influence of histone H1 on chromatin structure. *Cell*, 12(1):101–107, 1977.

[146] F. Thoma, Th Koller, A. Klug, Koller Th., and A. Klug. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *Journal of Cell Biology*, 83(2 I):403–427, 1979.

[147] David J. Tremethick. Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*, 128(4):651–654, 2007.

[148] Rocio G. Urdinguio, Jose V. Sanchez-Mut, and Manel Esteller. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies, 2009.

[149] 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.

[150] Anne Laure Valton and Job Dekker. TAD disruption as oncogenic driver, 2016.

[151] Kensal E. van Holde. *Chromatin*, volume 59 of *Springer Series in Molecular Biology*. Springer New York, New York, NY, 1989.

[152] M. Vettese-Dadey, P. A. Grant, T. R. Hebbes, C. Crane- Robinson, C. D. Allis, and J. L. Workman. Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *The EMBO Journal*, 15(10):2508–2518, 1996.

[153] Lilien N. Voong, Liqun Xi, Amy C. Sebeson, Bin Xiong, Ji Ping Wang, and Xiaozhong Wang. Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell*, 167(6):1555–1570.e15, 2016.

[154] Masatoshi Wakamori, Yoshifumi Fujii, Noriyuki Suka, Mikako Shirouzu, Kensaku Sakamoto, Takashi Umehara, and Shigeyuki Yokoyama. Intra- and inter-nucleosomal interactions of the histone H4 tail revealed with a human nucleosome core particle with genetically-incorporated H4 tetra-acetylation. *Scientific Reports*, 5(1):17204, 2015.

[155] Gero Wedemann and Jörg Langowski. Computer Simulation of the 30-Nanometer Chromatin Fiber. *Biophysical Journal*, 82(6):2847–2859, 2002.

[156] J. Widom. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proceedings of the National Academy of Sciences*, 89(3):1095–1099, 1992.

[157] 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. *Biophysical Journal*, 49(1):233–248, 1986.

[158] C. L F Woodcock, L. L Y Frado, and J B Rattner. The higher-order structure of chromatin: Evidence for a helical ribbon arrangement. *Journal of Cell Biology*, 99(1 I):42–52, 1984.

[159] Christopher L Woodcock, Arthur I Skoultchi, and Yuhong Fan. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Research*, 14(1):17–25, 2006.

[160] J. L. Workman and R. E. Kingston. Alteration of Nucleosome Structure As a Mechanism of Transcriptional Regulation. *Annual Review of Biochemistry*, 67(1):545–579, 2002.

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

[162] Bin Zhang, Weihua Zheng, Garegin A. Papoian, and Peter G Wolynes. Exploring the Free Energy Landscape of Nucleosomes. *Journal of the American Chemical Society*, page jacs.6b02893, 2016.

[163] Ruihan Zhang, Jochen Erler, and Jörg Langowski. Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction. *Biophysical Journal*, 112(3):450–459, 2017.

[164] Chunyang Zheng, Xu Lu, Jeffrey C. Hansen, and Jeffrey J. Hayes. Salt-dependent intra- and internucleosomal interactions of the H3 tail domain in a model oligonucleosomal array. *Journal of Biological Chemistry*, 280(39):33552–33557, 2005.

[165] Yin Zhou and Oncel Tuzel. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 4490–4499, 2018.