INSIGHTS INTO THE ROLE OF NUCLEOSOMAL DNA FOLDING ON CHROMATIN FIBER PROPERTIES

By

STEFJORD TODOLLI

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Quantitative Biomedicine

written under the direction of

Wilma K. Olson

and approved by

New Brunswick, New Jersey

January 2019

ABSTRACT OF THE DISSERTATION

Insights into the role of nucleosomal DNA folding on chromatin fiber properties

by STEFJORD TODOLLI

Dissertation Director:

Wilma K. Olson

DNA in eukaryotic cell nuclei is packaged in a highly compact, yet dynamic chromatin structure that provides a regulatory mechanism for many biological processes, such as gene expression. The basic packaging unit of chromatin is the nucleosome, which consists of ~1.7 turns of DNA wrapped around an octamer core of histone proteins (H3, H4, H2A, H2B). Chains of nucleosome-decorated DNA, which resemble beads on a string, fold into a higher-order arrangement, often referred to as the 30-nm fiber. However, the structure of this fiber remains poorly understood, despite decades of research. Many proposed models for the 3D organization of the nucleosomes and intervening DNA in chromatin vary quite significantly, and the very existence and relevance of a 30-nm structure *in vivo* has been questioned.

An analysis of the available high-resolution nucleosome structures shows subtle, yet significant differences in DNA wrapping around the histone core. Monte Carlo simulations of regular nucleosome arrays generated using a meso-scale representation of DNA suggest that these local differences can lead to large changes in global nucleosome arrangements, comparable to the effect of changes in nucleosome spacing by ~2–3 base pairs. Our results suggest that a regular nucleosome array with a 177-base-pair (bp) repeat can display a loose three-stack or a more compact two-stack arrangement, on average, depending on the DNA wrapping profile of the nucleosome. These findings imply a very dynamic chromatin fiber with a multitude of mechanisms to control its folding.

Using this meso-scale model, we have studied the role of chromatin fiber architecture and histone

tails on chromatin compaction and long-range communication in constructs containing 177-bp repeats. Our predictions for chromatin fibers with a loose three-stack nucleosome arrangement can qualitatively account for experimental data from *in vitro* assays of enhancer-promoter communication (EPC) under physiologically-relevant conditions. On the other hand, fibers that display a two-stack arrangement are in better agreement with sedimentation velocity experiments performed under a different set of ionic conditions. Removal of histone tails diminishes EPC efficiency, and our simulations predict that H3/H4 tail removal has the biggest impact, in agreement with *in vitro* experiments.

Dedication

To Kelly, my wonderful life partner

Acknowledgments

The PhD journey has been a long, challenging, yet rewarding one. I consider myself extremely lucky to have always had the support to carry onwards when the going was rough. First and foremost, I want to thank my PhD advisor and mentor Wilma K. Olson, PhD. She has been a constant source of intellectual clarity, emotional support and patience. Her availability and near-instantaneous feedback even in the midst of overwhelming work are a source of inspiration for me. I am grateful to Vasily Studitsky, PhD for the continuing collaboration and amazing support. My graduate student research record would certainly not be the same without our combined effort. I also want to thank my other committee members David Case, PhD and Sagar Khare, PhD. A special thanks goes to Willy Wriggers, PhD, who personally taught me how to work with cryoEM density maps and made the work in Chapter 2 possible. I also want to thank Kevin Range, PhD, who was my undergraduate research mentor at Lock Haven University and inspired me to apply to Rutgers to continue my academic training. It truly takes a village!

I would like to thank Guohui Zheng, PhD and Nicolas Clauvelin, PhD, who initiated the work on the meso-scale model of chromatin. A great amount of the technical knowledge that I developed during my first years of the PhD program is thanks to the mentorship from Nicolas. I also want to thank Ekaterina Nizovtseva, PhD for her experimental data, and continuing collaboration. My graduate school experience would not be the same without the daily interactions with the current and former members of the Olson lab: Pamela, Robb, Gautam, Andrew, Shuxiang, Seyed, Tahir, Nil, Ondrej, John, Tim, to name just a few. Thank you all for the support in good and bad days!

I want to thank my parents and my brother, for the continuous support of all forms: emotional, intellectual and financial, since well before I started my graduate program. I am thankful for all my amazing friends, who are many to name here, fortunately for me. A special thanks goes to Kenneth

McGuinness, PhD, my colleague, friend and housemate of four years.

Lastly and most importantly, I want to thank my wonderful life partner, Kelly. My PhD journey would not be the same without your constant support, patience, love, encouragement and intellectual stimulation. Thank you for always helping me see the light at the end of the tunnel and for making me a better person!

Acknowledgment of previous publications

Portions of the methods described in Chapters 3 and 4 have been reported in previous publications from our lab in collaboration with the research group of Vasily Studitsky [1–6].

The work presented in Chapter 3 is currently in the early stages of being prepared for publication.

Chapter 4 is being prepared for submission in collaboration with the research group of Vasily Studitsky, who is leading the *in vitro* experiments.

Figures from published literature are used with permission and include references to the original work. Where necessary, the permission is explicitly noted in the figure caption. For these referenced works, further permissions related to the material excerpted should be directed to the original publisher.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

One of the most crucial questions in cellular biology is: why are higher organisms made up of cells that look and function in wildly different manners, yet have the exact same copy of DNA? Well before the discovery of the DNA double helix [7], the scientific consensus was that DNA associates with proteins in the form of chromatin inside the eukaryotic cell nucleus and it was here that the differential gene expression was subsequently believed to be regulated. During the metaphase stage of cell division, chromatin condenses into chromosomes that are visible under the light microscope. In order to fit about two meters of human DNA into a cell nucleus that is a few microns in diameter, the DNA must undergo significant compaction. However, the structural details of chromatin organization are still an open question despite decades of research, yet have profound implications on biological processes, including transcription, DNA replication and DNA repair. The long-held textbook view has been that this compaction is hierarchical in nature (Figure 1.1) and the first level of compaction is the nucleosome, which represents the basic repeating structural unit of chromatin [8].

Nucleosomes

Work done prior to the early 1970s using biochemical analysis, X-ray diffraction, light microscopy of stained chromatin, among other techniques [10, 11] established the view that chromatin consists of interactions between DNA and histone proteins, forming spheroid chromatin particles that were

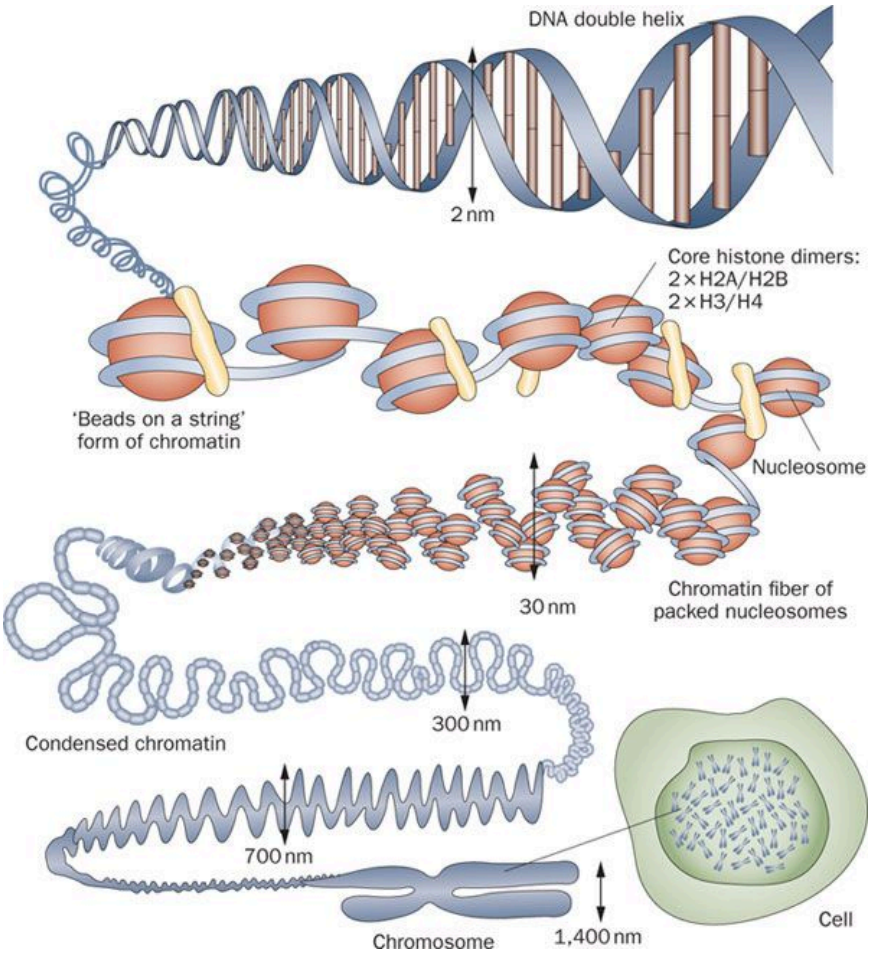


Figure 1.1: Schematic figure that illustrates how DNA in eukaryotic cell nuclei needs to condense into chromatin and chromosomes through a hierarchy of structures. The DNA double helix is about 2 nm in diameter and there is about 2 m of DNA inside each human cell nucleus, which is on average less than 10 µm in diameter. The double helix first wraps around a protein octamer made up of four types of core histone proteins to form a nucleosome, which represents the first level of compaction. Nucleosomes along a DNA chain resemble beads on a string. A fifth histone called linker histone (depicted by yellow long shapes) interacts with the intervening linker DNA entering and exiting the nucleosome core. Nucleosomes interact with one another to form a more compact fiber called the 30-nm fiber, which has been subject to recent debate about its relevance *in vivo*. This motif can then undergo further levels of compaction through self-interaction. Figure reprinted from ref. [9] with permission from Springer Nature.

initially referred to as "ν bodies". When chromatin extracted from ruptured cells was first spread on a stage plate and observed under the microscope, it was described as "particles-on-a-string" by Olins and Olins [11]. This arrangement was later referred to as "beads-on-a-string" by Kornberg, who proposed a model that chromatin consists of flexibly jointed repeating units of eight histone molecules and about 200 base pairs of DNA [8]. A subsequent study named this unit the "nucleosome" and provided support for that model using electron microscopy and biochemical studies of chromatin from ruptured cells as well as *in vitro* reconstituted nucleosomes [12].

The term "nucleosome" refers to the fundamental repeating unit of chromatin. Each nucleosome consists of a nucleosome core particle (NCP) and the linker DNA that connects consecutive nucleosome cores. Often in the literature the term "nucleosome" refers to the nucleosome core [13]. Nucleosome cores are made up of about 145–147 base pairs (bp) of DNA wrapped around an octamer core of four types of histone proteins: H2A, H2B, H3 and H4 [14, 15]. With the addition of a fifth type of histone called the linker histone (H1 and H5), which is believed to interact with about 10 bp of DNA on either side of the nucleosome core, the complex with a total of about 165–167 bp of DNA is referred to as the chromatosome [13, 16, 17].

The four types of core histones share a similar overall three-dimensional fold, despite limited sequence similarity. A long α-helix flanked by two shorter helices that connect via loops, one on either side, makes up the bulk of the central and C-terminal portions of each core histone [14, 18]. This helix-turn-helix-turn-helix motif is known as a "histone fold" and has been observed in many other DNA binding proteins [18, 19]. The histone fold facilitates the formation of H3/H4 and H2A/H2B heterodimers through a hand-shake motif (Figure 1.2). Two H3/H4 heterodimers associate via four-helix bundles to form a $(H3/H4)_2$ tetramer, while an H2A/H2B heterodimer binds to either side of the $(H3/H4)_2$ tetramer, through similar four-helix bundle interactions [13, 14, 17]. The resulting octamer forms a protein "ramp" interface for about 145–147 base pairs of nucleosomal DNA to wrap around in 1.65 turns as a left-handed superhelix. A two-fold pseudo-symmetry axis that passes through a single base pair at the center of the nucleosomal DNA defines the dyad of the nucleosome (Figure 1.2) [13].

Highly conserved arginines interact with the phosphates near the minor groove at 14 positions along nucleosomal DNA, one per super-helical turn [14, 15]. While the central ~120 bp of DNA bind directly to the octamer of histone folds, the remaining DNA binds to the non-histone fold N-terminal

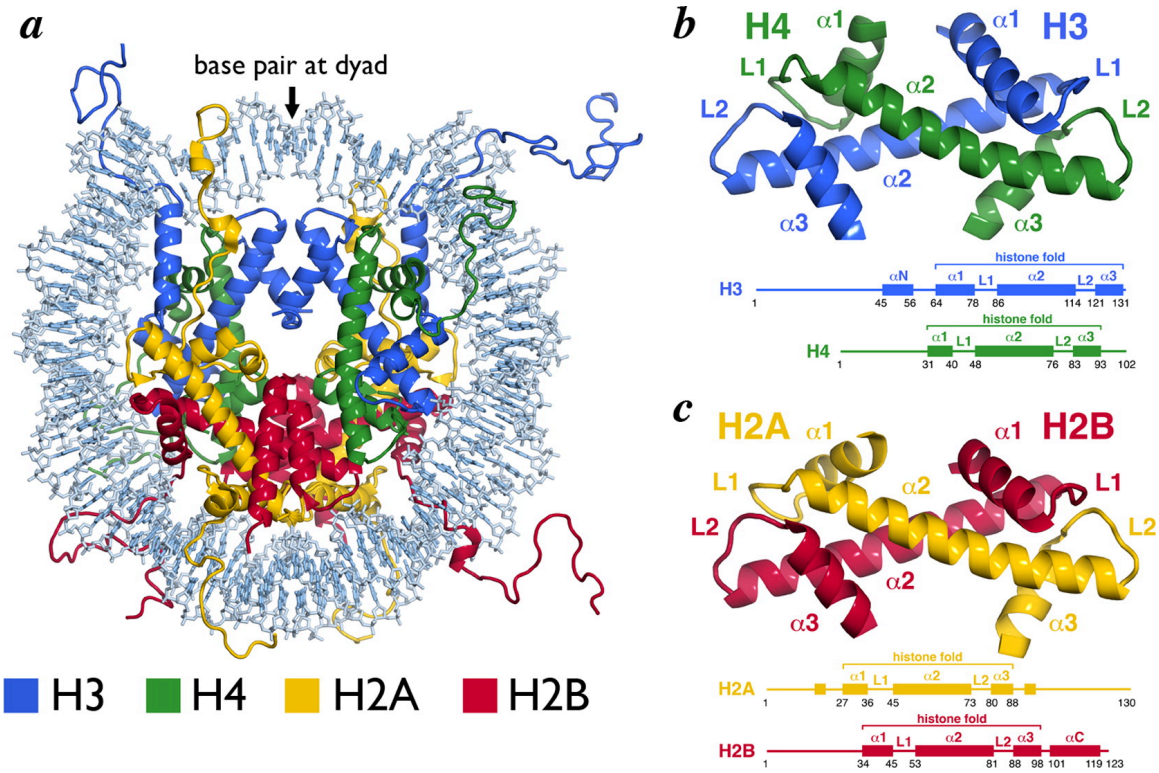


Figure 1.2: Structure of the nucleosome core particle (NCP). (a) DNA wraps in 1.65 turns around an octamer core of four types of histone proteins, with a pseudo-symmetry axis passing through a single DNA base pair (dyad). H3/H4 (b) and H2A/H2B (c) heterodimers are formed through an anti-parallel arrangement of the histone folds into a handshake motif. Although the N-terminal regions of all four histones as well as the C-terminal region of the H2A histone are mostly unstructured, a non-histone fold H3 N-terminal α-helix makes contact with the terminal regions of nucleosomal DNA, and H2B contains a relatively long C-terminal α-helix. Figure reprinted from ref. [13] with permission from ACS publications.

α-helix of H3 as well as part of the H3 N-terminal tail [14]. The N-terminal region of each histone outside of the core histone fold and the C-terminal region of H2A consist of flexible regions that are rich in cationic residues. These regions, commonly referred as tails, are mostly unstructured in solution but may adopt transient α-helical conformation [20]. The histone tails interact with both DNA and the histone core, as well as with other binding partners, such as various transcription factors or other nucleosomes [14, 21]. The tails are not required for nucleosome assembly but they play an important role in the thermal stability of the nucleosome [22]. The tail regions, usually defined by their protease sensitivity to trypsin [23], contain many targets sites for post translational modifications (PTM), which can alter their affinity for their interacting partners and consequently affect nucleosome stability. It has been shown, for example, that the presence of the N-terminal

histone tails is crucial to the role of the nucleosome as a barrier to traversal by RNA polymerase II [24, 25]. While the histone tails and the histone core are rich in positively charged residues that attract the highly negatively charged DNA, the exposed surface on the top and bottom of the NCP disk contains a cluster of negatively charged residues called the acidic patch at the H2A/H2B dimer interface [14, 15]. This feature is also involved in nucleosome-nucleosome interactions and in binding of many chromatin factors [13].

Despite the strong interactions found to hold the complex together in crystal structures, nucleosomes are not static entities whose function is simply to condense DNA. They are dynamic structures that provide a thermodynamic barrier to DNA accessibility by various chromatin factors, instead of a permanent blocking mechanism [17]. Indeed, Förster resonance energy transfer (FRET) measurements of mononucleosomes have shown that nucleosomal DNA partially unwraps from the nucleosome core in a spontaneous manner at a rate of ~4 s^{-1}. This process allows various transcription factors to bind to their target DNA and stabilize the unwrapped state, and if unbound the DNA rapidly re-wraps within 10-50 ms [26]. This partial DNA unwrapping, also called breathing, is often asymmetric and depends on nucleosomal DNA sequence, solvent conditions and a multitude of post translational modifications [27–29]. In addition to their role in local DNA accessibility, the nucleosome fluctuations can affect the arrangement of neighboring nucleosomes in an array. While structural details of nucleosome breathing have only recently been reported for single nucleosome core particles [30], the effect of nucleosome flexibility on chromatin structure remains unclear.

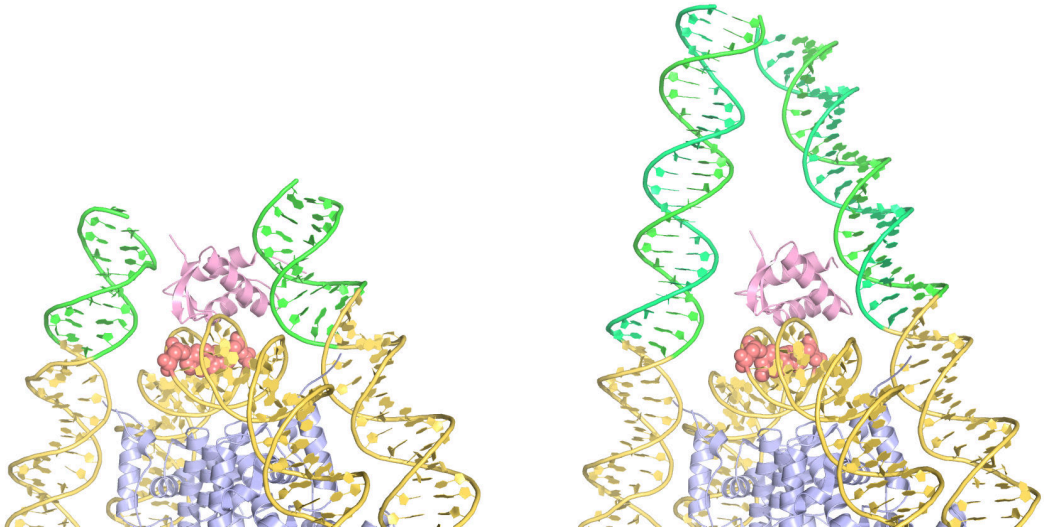
Since the solution of the first crystal structure of the nucleosome core particle to 2.8 Å resolution in 1997 [14], many more followed. This has been due in part to the work from Jonathan Widom and the engineering of a synthetic DNA sequence [31]. This sequence, called Widom 601 (from the number in a large pool of random synthetic sequences), has a much higher binding affinity for the histone octamer than natural sequences and makes the nucleosome core particles more stable. There are now many high-resolution structures available that contain valuable information on the effect that histone variants, chemical modifications and various bound proteins can have on nucleosome structure. The majority of these structures use DNA sequences that generally fall into just a few categories, including the human α-satellite repeat and the Widom 601 positioning sequences [13]. No large-scale analysis of high-resolution NCP structures has been performed to date, to quantify the differences in nucleosomal DNA wrapping as a result of these various factors. Such analysis

would enable us to understand the landscape of nucleosomal DNA wrapping, and the implications of DNA wrapping on the arrangement of nucleosome arrays. In Chapter 3, we use the available high-resolution nucleosome structures in an attempt to quantify these differences in DNA wrapping and probe the effect that they have on chromatin fiber structure based on three particular nucleosomes.

Linker histone and chromatosomes

Linker histones bind the nucleosome at the DNA entry/exit sites and interact with linker DNA to stabilize nucleosome wrapping in concert with core histone tails [32]. Linker histones are a family of lysine-rich histones that in higher organisms generally contain a tripartite structure: an unstructured ~30–40-residue N-terminal domain, an ~80-residue globular central domain and an unstructured, highly positively charged ~100-residue C-terminal domain (CTD) [33]. At least 11 different subtypes of linker histone H1 have been found in mammals, varying by cell type and developmental stage, and several more variants have been found in other organisms, such as H5 in chicken [34]. Linker histones also contain many target sites for post translational modifications. The globular domain is well conserved among different subtypes and responsible for docking onto the nucleosome core near the entry/exit site of linker DNA and protecting about 10 bp on each DNA linker [35]. An atomic structure for the globular domain of a recombinant chicken linker histone H5 was first elucidated in 1993 by Ramakrishnan et al. [36], displaying a winged helix motif that provided clues to the mode of interaction with nucleosomal and linker DNA. While the globular domain alone is sufficient for binding to the nucleosome core, the C-terminal domain is required for high affinity binding as well as formation of a DNA linker stem and higher-order chromatin compaction [35, 37]. The C-terminal domain is characterized by very low sequence conservation and high content of lysines and prolines, features typical of an intrinsically disordered protein (IDP). It has been suggested that the CTD by itself is unstructured in solution but may adopt a more compact state upon binding the nucleosome [38–40]. Many models for the interaction mode of linker histone in the chromatosome have been proposed over the years using a variety of experimental techniques and computational docking [41–45]. The proposed models generally fall into two main categories: off-dyad and on-dyad binding, but despite all this knowledge and extensive efforts, a high resolution structure of a full chromatosome complex has evaded researchers for a long time.

An atomic structure for a 165-bp chromatosome containing the globular domain of chicken linker



(a) (b)

Figure 1.3: Crystal structures of chromatosomes containing the globular domain of linker histone. The two structures are shown in the same view with respect to the nucleosome core (histone core in light blue, nucleosomal DNA in gold). The linker histone globular domains (in pink) bind on the dyad (red spheres) and interact with both DNA linkers (green) in a similar fashion. (a) Chromatosome with 165 bp DNA and the globular domain of chicken histone H5 (PDB ID 4QLC [46]). (b) Chromatosome with 197 bp DNA and globular domain of Xenopus H1.0b (PDB ID 5NL0 [47]). Figures generated with PyMol [48] from atomic coordinates.

histone H5 (gH5) was first reported in 2015 using X-ray crystallography. The structure showed an on-dyad binding mode where gH5 interacts with both DNA linkers (Figure 1.3a) [46]. An NMR study from the same group later showed that the globular domain of Drosophila H1 binds off the dyad and interacts primarily with one of the DNA linkers [49]. These results suggested that a small number of differences in sequence between the two proteins contributed to the different binding modes and led to different higher-order structures of nucleosome arrays. Using a combination of cryogenic electron microscopy (cryoEM) and X-ray crystallography, Bednar et al. [47] later reported the structure of a 197-bp chromatosome containing full length vertebrate linker histone H1. They determined that the globular domain of the linker histone (gH1) binds on the dyad and interacts with both entry/exit DNA linkers (Figure 1.3b), in a fashion similar to the interaction in the 165-bp crystal structure [46], while the C-terminal domain localizes primarily around one of the DNA linkers, although this region was not resolved to a high resolution. However, the DNA linkers contain significant differences between the two crystal structures: in the case of the 165-bp chromatosome

the linkers are more deformed, while in the 197-bp chromatosome they are mostly straight. This could be due to the poorer resolution of the latter structure, or an effect of the different crystal packings. While these structures provide valuable information on the effect of linker histone binding on the nucleosome, it still remains unclear how the interactions occurs in a chromatin environment and what the implications are for higher-order chromatin structure.

Chromatin structure and function

Shortly after the description of the basic chromatin structure as a string of nucleosome repeats and the visualization of the beads on a string structure, Finch and Klug proposed the first model of how nucleosomes associate into higher-order structures [50]. Using electron microscopy they observed that in the presence of Mg^{2+} ions and linker histones, chromatin extracted from rat liver nuclei condenses into a thick fiber with a diameter of about 30 nm. They proposed a model where nucleosomes arrange in a solenoid with pitch of 11 nm, corresponding to the diameter of a nucleosome. In this model, the nucleosomes stack side-to-side in the direction of the fiber axis, while consecutive nucleosomes pack radially into a linear filament connected by a highly bent DNA linker [50, 51]. This solenoid model has persisted in literature for a long time and the observed filaments became known as the "30-nm fiber". The proposed structure was considered to be the next level of DNA compaction beyond the nucleosome, and one among several levels in a hierarchical model of chromatin compaction (Figure 1.1).

Once the first details about the general organization of the nucleosome became available, many more chromatin fiber models were proposed over the following years [52]. A helical ribbon model, where the repeating unit consists of two nucleosomes twisted around the fiber axis, was proposed in sharp contrast to the solenoid model [53]. This arrangement implied that the intervening DNA was mostly straight as it crossed back and forth between consecutive nucleosomes in the ribbon. A later study proposed that the chromatin fiber consisted of two stacks of non-consecutive nucleosomes twisted onto each other around the fiber axis in a left-handed manner, while maintaining the earlier observation that the DNA linker is mostly straight [54]. In another study, the chromatin fiber was described as a triple helix model, with a loose three-dimensional zigzag arrangement of nucleosomes in low salt that form three stacks when compacted under high salt [55]. The models proposed from these early studies were built from observations of chromatin extracted from nuclei, and this "native"

chromatin is generally heterogeneous in terms of DNA sequence, nucleosome spacing and various histone compositions. The development of a nucleosome array model system reconstituted from tandem repeat 5S rRNA sequences [56] and later of the synthetic 601 strong nucleosome positioning sequence [31] allowed for more precise control over the spacing between the nucleosomes and led to many *in vitro* studies of regular nucleosome arrays.

Using nucleosome arrays reconstituted *in vitro*, Rhodes and colleagues conducted EM measurements of fiber dimensions, and proposed an inter-digitated nucleosome array model [57]. Another study in the Grigoryev research group used analytical centrifugation to characterize the effects of nucleosome repeat length (NRL) and various ionic conditions on the compaction of regular nucleosome arrays, and [58]. The data from these *in vitro* studies (and others) of regular nucleosome arrays, has led to many analytical and computational models that capture the dependence of chromatin structure on NRL [59–61]. In the absence of detailed chromatin fiber structures observed experimentally, however, these models were all supported by one set of data or another and often contained contradictory features [52].

A crystal structure of a tetra-nucleosome by the Richmond group revealed a stacked di-nucleosome structure with crossed linkers [62]. The structure provided strong support for a two-start model of the 30-nm fiber and suggested that nucleosomes within a stack interact via acidic patch–H4 tail interactions. This importance of this interaction for chromatin fiber folding was previously demonstrated using sedimentation coefficient experiments [63]. A recent study by the same group, however, has revealed two tetra-nucleosome arrangements with shorter linker DNA and different nucleosome-nucleosome interactions that challenge the necessity for the previously proposed H4 tail–acidic patch interaction [64]. A lower resolution cryoEM density map of a reconstituted 12mer nucleosome array also displayed two different sets of interactions between nucleosomes [65]. The differences among these few nucleosome array structures highlight the inherent flexibility of the chromatin fiber, and reflect the challenges in the effort to elucidate chromatin structure. Chromatin fiber heterogeneity is likely a feature of chromatin that allows for a regulatory mechanism and contributes to its function in biology.

Due to its role in DNA compaction and the reduced accessibility of the nucleosomal DNA, chromatin has been generally considered to act as a transcription repression mechanism. The more actively transcribed chromatin regions (euchromatin) are defined by a lower density of nucleosomes

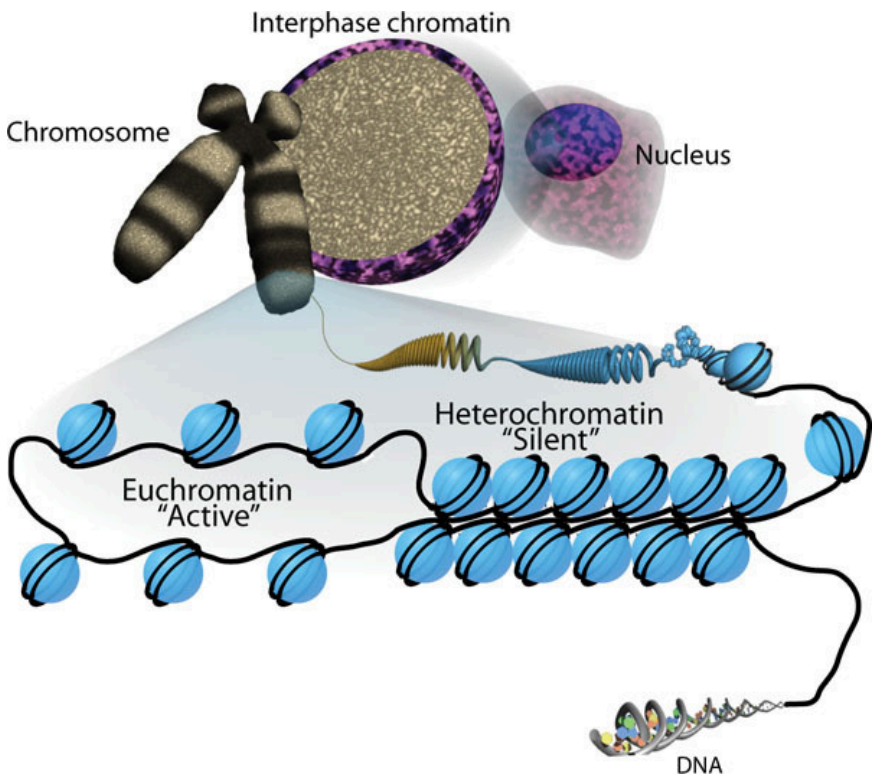


Figure 1.4: Euchromatin is characterized by looser packing of the nucleosomes and generally associated with higher levels of transcription, whereas heterochromatin contains a much denser packing of the nucleosomes that prevents DNA access to many of the chromatin factors necessary for the transcription machinery to work. Figure reprinted from [66].

and a looser fiber arrangement, while the more silent regions (heterochromatin) are defined by higher nucleosome density (Figure 1.4). However, these two states are not necessarily static and are subject to regulatory mechanisms. Recent advances in genetics have led to the analysis of the spatial arrangements of nucleosomes in whole genomes with very high precision and the identification of topologically associated domains (TAD), which are large regions of chromatin with high frequency of local interactions and are found in many species [67]. The barriers between these domains, characterized by low interactions, were shown to be associated with the spread of heterochromatin regions and their disruption can lead to misregulation [67, 68]. Furthermore, complex biological processes often require distant elements in the DNA sequence to come into close proximity. Such is the role of enhancers that activate transcription at distant promoter sites, often thousands of bases away in the sequence [6, 69]. How an enhancer identifies its target promoter in three-dimensional nuclear space over such long distances still remains largely unknown, from a mechanistic perspective. It has been shown from *in vitro* studies and simulations of precisely positioned nucleosome arrays

that chromatin enhances the efficiency of enhancer-promoter communication (EPC) compared to naked DNA [1, 2]. While the N-terminal tails increase this enhancement effect, the presence of nucleosome gaps can either strengthen or weaken the EPC enhancement, depending on the length of the system [2, 5]. In this work we use computational modeling and simulations in an effort to understand how chromatin structure can contribute to the level of EPC enhancement and what the role of specific N-terminal tails is in this enhancement. We compare our results with *in vitro* studies led by the Vasily Studitsky lab to understand the implications of this data on chromatin fiber architecture.

It has been suggested that the 30-nm chromatin fiber may not exist *in vivo* [70]. The view of chromatin that has emerged in the last several years supports an irregular chromatin fiber that exists in a disordered state [71]. An *in situ* study using electron microscopy tomography in combination with advanced DNA staining (ChromEMT) revealed a disordered chromatin chain, 5–24 nm in diameter, that comprises a heterogeneous mix of arrangement motifs [72]. Recent *in vivo* work found that chromatin forms compact domains organized by inter-nucleosome interactions that display a liquid-like behavior and move coherently [73, 74]. These domains appear to share some properties with TADs, although they are products of different physical processes. On the other hand, a super-resolution study revealed that topologically associated domains are found in single cells and while regulatory proteins are necessary to maintain TAD boundaries at preferred sites, TAD-like structures are found in the absence of these proteins [75]. These recent advances are shifting the paradigm for chromatin structure and necessitate new computational tools that can capture the heterogeneity observed from experiments. Our work represents an effort to develop such methods to aid in our understanding of chromatin behavior.

1.2 Dissertation overview

Chapter 2 describes our approach to construct an approximate atomic-level model of a chromatin fiber based on available low resolution cryoEM density maps. We aim to understand the structural details of the fiber and pathways of inter-nucleosomal DNA and present an efficient method to model the DNA linkers. Our results hint at heterogeneity among the linkers in the cryoEM density maps and point to potential problems with the densitites.

In Chapter 3 we discuss the results of meso-scale simulations of regular nucleosome arrays and the effects of nucleosomal DNA wrapping on the global properties of the resulting fibers. We focus on three specific high-resolution nucleosome structures, including one that contains a linker histone globular domain. We attempt to characterize how the specific DNA wrapping in these nucleosome structures affects the nucleosome-spacing dependence of chromatin fiber architecture. Our results may help to understand some of the discrepancies among previously proposed chromatin fiber models and experimental results from sedimentation velocity studies. Our meso-scale model allows us to treat chromatin as a dynamic ensemble of structures and to calculate ensemble properties, a view that has recently gained support as a more accurate description of chromatin structure. We present the details of the meso-scale model and the methods we use to characterize the simulated arrays as dynamic ensembles.

In Chapter 4 we use our meso-scale model of chromatin in simulations of enhancer-promoter communication (EPC) experiments of nucleosome arrays to study the effect of array length and N-terminal histone tails on these long-range interactions. We compare our results with experimentally observed patterns of distant communication from *in vitro* studies from the research group of Vasily Studitsky, as a way to validate the utility of our model. Our EPC results support a loose, approximately three-stack arrangement of the nucleosomes in fully saturated arrays of 177-bp repeat systems, in physiologically-relevant ionic conditions.

Chapter 5 contains concluding remarks. We discuss the significance of our work and the potential shortcomings, as well as the many questions that are raised for future studies.

Chapter 2

Chromatin fiber organization: insights from modeling based on cryoEM data

2.1 Introduction

The traditional description of the chromatin fiber as a hierarchical structure with various compaction levels still persists in textbooks as an illustrative way to think about how DNA is packed inside the eukaryotic cell nuclei. Since the early observations of nucleosome-decorated DNA as beads on a string condensing into a 30 nm fiber in the presence of linker histone H1 or increasing ionic strength [32, 50], there have been many modeling attempts to understand the nature of this compaction at this scale. However, the proposed models for the 3D organization of the nucleosomes and intervening DNA vary quite significantly and the very existence and relevance of this structure *in vivo* has been called into question [70]. While the structure of the nucleosome core particle (NCP) has been solved to a very high-resolution [15], structural information about the nucleosome-decorated DNA chains from experiments has been scarce and inconsistent, mostly due to the size and flexibility of the system.

A crystal structure of a tetra-nucleosome at 9 Å published in 2005 supported a two-start helix where mostly straight DNA zigzags back and forth between nucleosomes arranged in two stacks [62]. Recently the same group published two higher resolution tetra-nucleosome crystal structures at 6.7 and 5.8 Å, which are in general agreement with the two-start model of the prior work, but contain

different nucleosome-nucleosome interactions within each stack [64]. These tetra-nucleosomes represent the highest resolution to date of any structure that resembles the 30 nm fiber and highlight the challenges of having a detailed account of the chromatin structure. The structural heterogeneity of the nucleosome arrangements at this level can have large effects on higher-order structures at the genomic scale. However, these crystal structures only contain four nucleosomes in the absence of the linker histone protein, which is assumed to interact with DNA linkers near the entry/exit sites, and do not necessarily reflect what may occur in a long chromatin fiber.

A single-particle cryogenic electron microscopy (cryoEM) study by Song et al. produced low-resolution (11 Å) density maps of two chromatin constructs with 12 nucleosome repeats of the Widom 601 positioning DNA sequence in the presence of human linker histone H1.4 [65]. One construct has a 177-base-pair (bp) repeat (12×177 bp) and the other has a 187-bp repeat (12×187 bp). The study reveals a two-start crossed linker architecture, similar to the arrangement suggested by the first tetra-nucleosome crystal structures [62]. The structure is organized in three tetra-nucleosome unit repeats, with the nucleosome stacking between each unit stabilized through H4 tail–acidic patch interactions, and adjacent H2A/H2B dimers interfaced within each unit. While no detailed account for the modeled DNA linkers is presented, it is suggested that the linker histone globular domain interacts with both DNA linkers off the dyad of the nucleosome.

In this study we revisit the density data from the cryoEM study with the goal of building an atomic-level model and getting a better understanding of the arrangement between nucleosomes and intervening DNA. There is an apparent heterogeneity present in the density map, where the nucleosomes are arranged in three tetra-nucleosome units [65] and we want to understand potential differences among the DNA linker arrangements that define each position within a unit. While the linker in this chromatin fiber arrangement are generally considered as straight, our work suggests that it may not be the case and that there may be differences between the different DNA linkers in the fiber density. These different DNA arrangements could have implications for the positioning of the linker histone. We present an efficient method for generating approximate atomic-level models of DNA based on a density map using a rigid-body parameter treatment.

Since the cryoEM study came out in 2014, more recent work has produced a large body of information regarding chromatin fiber structure and the binding mode of the linker histone globular domain in nucleosomes. The X-ray crystal structure of a chromatosome (nucleosome core + linker

DNA + linker histone) with 167-bp DNA and the globular domain of chicken histone H5 (gH5) was solved at 3.5 Å in 2015 [46] and that of a chromatosome with 197-bp DNA with full length Xenopus (African clawed frog) linker histone H1.0 was determined at 5.5 Å in 2017 [47]. In both structures, the globular domain of the linker histone binds on the dyad of the nucleosome and interacts with both DNA linkers. On the other hand, an NMR model of a chromatosome containing Drosophila H1 revealed an off-dyad binding mode. These structural differences show that there could be multiple modes of interaction between linker histones and nucleosomes. While these mono-chromatosome structures provide great insight into the interaction of linker histones with nucleosomes, they are not necessarily representative of the environment present in the chromatin in living cells. Without a structural account of the nucleosome arrangements and linker histone interaction in a chromatin fiber, it is difficult to understand the role of chromatin in many biological processes, such as transcription, replication and repair.

2.2 Methods

2.2.1 Fitting nucleosome core particles

The original density map for the 12mer nucleosome array was segmented using the Chimera software [76] into 12 density regions corresponding to each nucleosome. The 1.9 Å resolution NCP crystal structure from Davey et al. (PDB ID: 1KX5) [15] with truncated N-terminal histone tails was then docked into each nucleosome density region (Figure 2.1). We used the `colores` program from the Situs suite to perform rigid-body docking since it utilizes a contour-based fitting protocol shown to improve the docking precision for low-resolution densities [77, 78]. After docking the high-resolution nucleosome structure at every position, we then used the 3DNA software [79, 80] to place reference frames on each base pair of the nucleosomal DNA. We used these reference frames as spatial constraints for the next step, that is, modeling the linker DNA connecting successive nucleosomes.

2.2.2 Modeling flexible DNA linkers

In order to model the flexible DNA linkers we wanted to set up a protocol that fulfills at least three basic criteria. First, the protocol needs to produce DNA with the best possible fit to the density map.

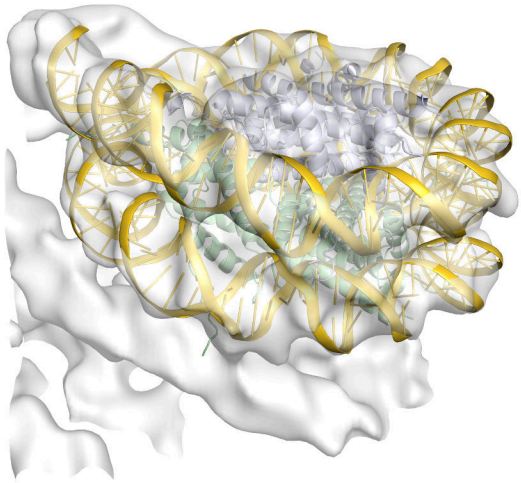


Figure 2.1: Nucleosome core particle docked inside the density map region corresponding to a nucleosome. The docking is performed using the `Situs>colores` program [77] that utilizes a contour-based filter to improve the docking quality.

Second, it should not over-fit DNA to the density map at the expense of producing an energetically implausible model. In other words, the linker DNA needs to be in general agreement with some of the physics-based models previously described in the literature [81, 82]. Lastly, the flexible fitting procedure needs to be very fast. For this reason we set up an optimization algorithm that simultaneously maximizes the density fitting score and minimizes the elastic energy of deformation of a flexible DNA linker.

To compute the elastic energy of deformation we used a coarse-grained elastic polymer model of DNA represented as a collection of base pairs. Each base-pair step is described by a set of six rigid-body parameters, where the first three parameters represent a rotation and the next three represent a translation (Figure 2.2) [83, 84]:

$$\boldsymbol{p} = (p_1, p_2, p_3, p_4, p_5, p_6) \tag{2.1}$$

The elastic energy of deformation for a collection of steps is then calculated as the sum of energies for each step using a simple harmonic potential [81]:

$$\epsilon = \frac{1}{2} \sum_{1=1}^{N-1} \Delta \boldsymbol{p}_{\boldsymbol{i}}^{T} \boldsymbol{F}_{\boldsymbol{i}} \Delta \boldsymbol{p}_{\boldsymbol{i}} \,, \tag{2.2}$$

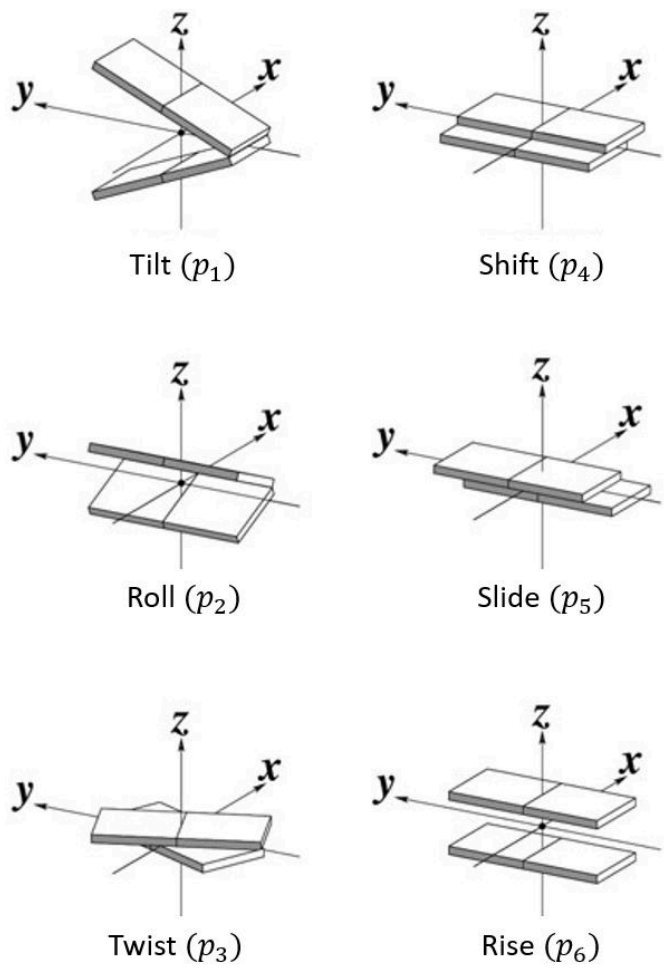


Figure 2.2: Base-pair step parameter representation of local DNA geometry. Figure adapted from Lu and Olson [79].

where N is the total number of base pairs in the collection, $\boldsymbol{F_i}$ is the force constant matrix associated with step i and $\Delta\boldsymbol{p_i} = (\boldsymbol{p_i} - \overline{\boldsymbol{p}})$ is the deviation of the rigid-body parameters at step i from their rest state. In our approach we use an ideal, inextensible, naturally straight DNA model with intrinsic step parameters characteristic of a B-DNA rest state with a helical repeat of 10.5 bp per turn (Eqn. 2.3). Therefore our force constant matrix is the same for every step and does not contain any coupling between modes of deformation (Eqn. 2.4). It is important to note that this was done for simplicity and the model can be very easily implemented to take account of sequence effects.

$$\overline{\boldsymbol{p}} = (0,\ 0,\ 34.2857\,deg,\ 0,\ 0,\ 3.4\,\text{Å})\,, \tag{2.3}$$

$$\boldsymbol{F} = diag[0.0427,\ 0.0427,\ 0.0597\ (k_{\mathrm{B}}T/deg^2),\ 20,\ 20,\ 20\ (k_{\mathrm{B}}T/\text{Å}^2)]\,. \tag{2.4}$$

To compute the density correlation C of the DNA linker with the EM map we use the `collage` program from the Situs suite [78]. The program takes a set of full atoms and an EM density map and computes a correlation score in the range [0,1], with larger values representing better fits. Therefore the total energy term that we aim to minimize in our method is:

$$E = w_1\epsilon - w_2(1 - C)\,, \tag{2.5}$$

where a smaller $(1 - C)$ value now represents a better fit, therefore a "lower energy". The weight terms w_1 and w_2 are used in order to offset the rather large elastic energy values ϵ (in $k_\mathrm{B}T$) of the DNA linkers when they are fitted to the density map, compared to $(1 - C)$ which is unit-less and ranges from [0,1]. This way, the contributions from both terms are comparable and at least of the same order of magnitude. It is worth noting that the general inextensibility of the DNA in our model is achieved by using very large values for the translational force constants, as seen in Eqn. 2.4, instead of disallowing those moves completely. That means in reality, during the minimization steps, the linker DNA can be stretched or horizontally displaced by some small amount, leading to very large values of the elastic energy of deformation when those moves result in a better fit to the EM density.

Powell's method [85], as it is implemented in the Python SciPy library, was used to find a set of base-pair step parameters that minimize the total energy E of the linker. A set of $6(N+1)$ step parameters is required to model a flexible DNA linker of length N bp, in order to account for the base pairs at each end. These pairs represent the terminal base pairs of the linked nucleosomes (i.e. $N+2$ bp linker that includes the constrained terminal base pairs). At each move in the Powell minimization, a full atomic model of the DNA linker is reconstructed from the base-pair step parameters using 3DNA [79] in order to calculate the density correlation score. This procedure is quite costly in terms of computation and is an area that could potentially be improved in the future. The entire minimization algorithm was implemented in Python while the `collage` functionality from Situs [78], which is written in the C programming language, was integrated into the Python code as a set of library functions using the interface generator Swig.

Our optimization method produces DNA linkers with visibly good fits to the EM density maps while maintaining the structural integrity of the DNA. However, due to the general limitation

of minimization algorithms to find only local minima, it is important to have a good starting configuration. In this case, the starting DNA linker must have at least some degree of overlap with the correct linker density; otherwise the optimization can steer the DNA linker towards the wrong density region. A good starting configuration can be achieved in a number of ways. Turning off the density correlation term and generating an elastic energy-optimized DNA linker usually produces a fairly reasonable starting structure (Figure 2.3, pink DNA). This is the approach that we take in most cases in our modeling processes. In some cases however, manual adjustment of this starting structure might be required, especially when linker densities contain highly bent regions. A piecewise construction of the DNA linker was used in this case, where only portions of the linker that overlapped the correct EM density were selected in successive rounds of optimization to obtain an appropriate starting configuration.

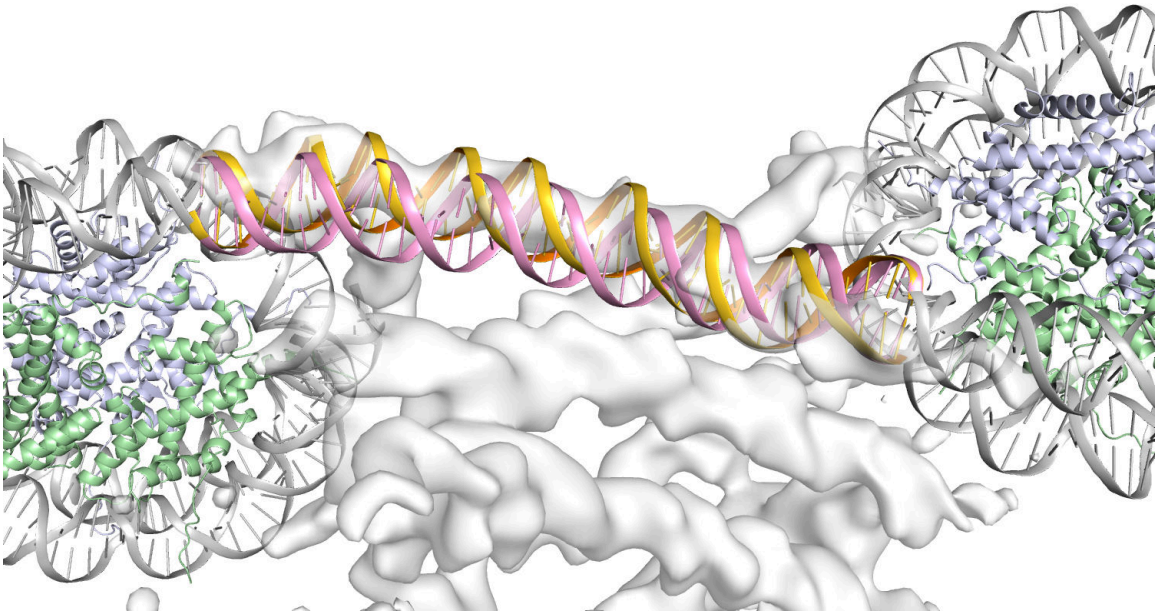


Figure 2.3: An example of a DNA linker, in this case connecting nucleosomes N1 and N2 in the EM density, after elastic minimization (pink) and after the full optimization procedure that minimizes the elastic energy and maximizes the EM density correlation (gold).

We use the local step parameters (tilt, roll, twist, shift, slide, rise) of the optimized DNA linker models to analyze their pathways in the cryoEM density maps. Additionally we define the net local bending (Ω) between successive base pairs given by the square root of the sum of the squares of tilt

(p_1) and roll (p_2),

$$\Omega = \sqrt{p_1^2 + p_2^2}\,. \tag{2.6}$$

2.2.3 Docking the linker histone globular domain

After modeling the nucleosome core particles and the DNA linkers we used the `pdb2vol` program from Situs to create volumetric maps from the docked core histones and modeled DNA. We subsequently used the `voldiff` program to subtract this volume from the original density map. The difference is presumed to correspond to the globular domains of the linker histones. Our approach to model the positioning of the linker histone globular domain in the remaining density regions involves docking high-resolution structures of the linker histone globular domain into each region/position and then analyzing the ensemble of docked models for each position, separately. For this purpose we used all of the high-resolution structures for the globular domain of linker histones H1/H5 (gH1/gH5) that were available in the Protein Data Bank (PDB) [86] as of 2017. The list includes a mix of crystal and solution NMR structures (Table 2.1). Since NMR-determined structures usually contain an ensemble of models, this approach has the benefit of containing a representation of structural heterogeneity in the globular domain and allows for a wider exploration of possible docking models, given the low resolution of the density map. After rigid-body docking every gH1/gH5 model (88 in total) to each position in the density map of the fiber with the `colores` program from Situs, we selected only the best-scoring fit (according to the ranking from Situs) from each case for further analysis.

2.2.4 Contacts between linker histone globular domain and linker DNA

We used the atomic coordinates from the ensemble of docked gH1/gH5 models at each position to analyze the contacts between the linker histone globular domain and nucleosomal DNA, including the DNA linkers entering and exiting the nucleosome. To define a contact, the general practice is to use the van der Waals radius for each atom and assume a contact if the atomic distance is less than the sum of the two atomic radii. In our case, since the resolution of the EM densities is very low and the docking models are very approximate, we use a generous distance of 4.0 Å between any two

Table 2.1: List of high-resolution structures of the globular domains of linker histones H1/H5 downloaded from the Protein Data Bank (PDB). Structures with an asterisk (*) represent nucleosome complexes that contain linker histone globular domains. In the case of 5NL0 the structure was crystallized with full-length linker histone; however, only the globular domain was resolved.

PDB ID	Exp. Method	# Models	Type	Source	Chain Length	Ref.
1GHC	Solution NMR	14	H1	*G. gallus*	75	[87]
1HST	X-ray diffraction	2	H5	*G. gallus*	90	[36]
1UHM	Solution NMR	20	H1	*S. cerevisiae*	78	[88]
1USS	Solution NMR	10	H1	*S. cerevisiae*	88	[89]
1UST	Solution NMR	10	H1	*S. cerevisiae*	93	[89]
1YQA	Solution NMR	10	H1	*S. cerevisiae*	87	[90]
2LSO	Solution NMR	20	H1x	*H. sapiens*	83	NA
4QLC*	X-ray diffraction	1	H5	*G. gallus*	77	[46]
5NL0*	X-ray diffraction	1	H1.0-B	*X. laevis*	196	[47]

atomic centers as the cutoff that defines a contact.

2.3 Results and Discussion

2.3.1 Nucleosome core particle arrangements

After docking the high-resolution nucleosome core structures, we first analyzed the relative arrangements between successive nucleosomes. First of all, due to the near symmetrical nature of the nucleosome core particles (excluding the long tails in the 1KX5 structure), there are two potential orientations of the docked crystal structure for each position. These possibilities are returned as the two best fits by Situs, flipped with respect to one another by a 180° rotation around the nucleosome dyad axis. Since the low resolution of the EM density map does not contain information about the direction of the DNA sequence, we chose the direction that is most represented among the docked nucleosome cores, considering that the high-resolution crystal structure defines the direction of nucleosomal DNA. We then looked at the distances between the terminal base pairs of consecutive nucleosomes (Table 2.2), as a way to validate the density maps and as the first step to modeling the intervening DNA. The periodic nature of the distances in this table show that the fibers are generally arranged in three tetra-nucleosome units, as reported by the authors of the original work [65]. The distances between nucleosomes N4-N5 and between nucleosomes N8-N9, which correspond to the tetra-nucleosome boundaries, are generally longer than the others.

Table 2.2: Distances between terminal base-pairs of consecutive nucleosome core particles docked to the cryoEM density map. The three columns for each construct represent the raw distance d (in Å), the length d_0 of straight B-DNA that is required to span the distance (in base pairs, rounded to integers) and the difference Δd between d_0 and the length of linker DNA in the fiber (considering the NCP contains 147 bp), respectively. It is clear from this table that the distances between nucleosomes N4-N5 and between nucleosomes N8-N9, i.e., between tetra-nucleosome units, in the 12×177 bp density map (colored in red) are significantly larger than what can be accounted for by the described construct.

	12×177 model, 30 bp linker			12×187 model, 40 bp linker		
NCP connection	d (Å)	d_0 (bp)	Δd (bp)	d (Å)	d_0 (bp)	Δd (bp)
1-2	94.2	27	-3	138.2	40	0
2-3	104	30	0	125.3	36	-4
3-4	95.2	27	-3	140.1	40	0
4-5	158.3	46	+16	138.0	40	0
5-6	95.5	27	-3	136.4	39	-1
6-7	103.5	29	-1	124.1	36	-4
7-8	94.8	27	-3	137.6	39	-1
8-9	128.9	37	+7	138.2	40	0
9-10	92.8	26	-4	137.7	39	-1
10-11	108.9	31	+1	124.1	35	-5
11-12	95.5	27	-3	138.9	40	0

It is striking to observe that the distances between tetra-nucleosome units in the case of the 177-bp repeat construct are significantly larger than what a straight B-DNA of the experimentally designed 30-bp linker length could span (Table 2.2, distances shown in red text). This means that given a density map free of potential errors, either the DNA linker is significantly overstretched, or there is some degree of nucleosome repositioning taking place. While those scenarios certainly are possible, the discrepancy may simply reflect an artifact in the EM density from improper modeling and refinement and needs to be investigated further. The 187-bp repeat construct does not appear to have this problem. The distances that are shorter than the construct linker length can be accounted for by DNA deformation, since these DNA linkers do not generally follow a straight path. For this reason, from this point onwards, we deal only with the modeling of the 187-bp repeat fiber.

2.3.2 DNA linker arrangements

Using the method described in Section 2.2.2 we modeled all the DNA linkers in the EM density map of the 187-bp repeat fiber. Our modeling includes the 40-bp linker DNA as well as the first and last 10 bp of DNA on each 147-bp nucleosome or, expressed in terms of each nucleosome dyad (base pair 0) we modeled linkers that connect base pair +63 of one nucleosome with base pair -63 of the following one. This "peeling" allows us to account for any effect that the chromatin fiber arrangement may impose on the wrapping of nucleosomal DNA, instead of placing all of the stress on the DNA linkers. Initially, we build a linker with minimal elastic energy that connects successive nucleosomes. At a second step, we use both the density correlation score and the elastic energy to fit the DNA linkers to the EM density map, as described in methods. Upon visual inspection, the linkers seem to fit the density map very well (Figure 2.4). It can be seen from this figure that there are various regions in the linkers where DNA is significantly deformed from the ideal state to accommodate the EM density and these regions do not appear to be evenly distributed.

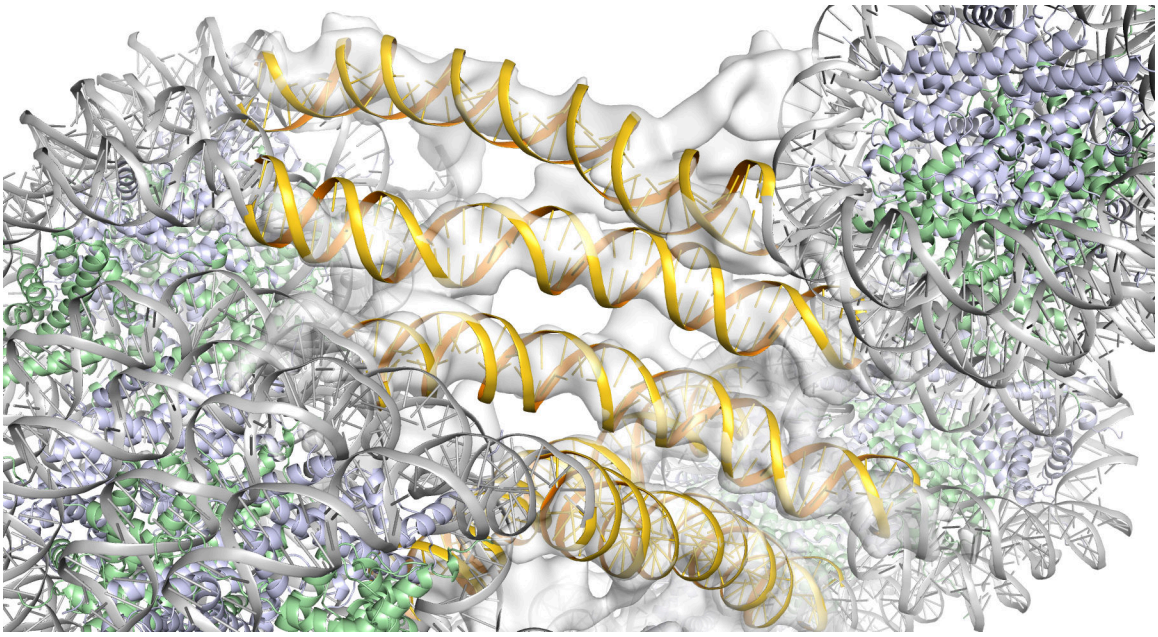


Figure 2.4: Examples of some of the optimized DNA linker models (gold) shown inside their respective EM density map regions (light gray). The corresponding nucleosomes are also shown in gray (DNA), blue and green (histone core). The DNA linkers are modeled to minimize the elastic energy of DNA while maximizing the fit to the EM density. It is clear that the linker DNA is unevenly deformed to fit the density map.

By analyzing the modeled DNA linkers we can learn a lot about the fiber arrangement and tease

out differences between and within each tetra-nucleosome unit. In Figure 2.5 we have plotted the local twist value (p_3) along each DNA linker, after the elastic optimization (top panel) and after full (elastic energy + density correlation) optimization (bottom panel). What is striking about these values is that before the fit to the density map is considered, the modeled DNA linkers have nearly identical local twist patterns (top panel). The values exhibit a somewhat regular fluctuation around the ideal B-DNA intrinsic twist of 34.3°, with the steps tending to overtwist more than undertwist. After the DNA is fitted to the EM density map, however, that uniformity breaks and differences between the DNA linkers emerge (bottom panel). The linker twist profiles are grouped into four categories to reflect the four NCP-NCP steps defined by the tetra-nucleosome repeat nature of the EM density [65]. Each linker group (L-1, L-2, L-3, L-4) is shown in a different color and each linker within a group is plotted with a different line style. Linkers connecting nucleosomes N1-N2, N5-N6 and N9-N10, which are referred as L-1 and colored in red, show the largest deviation from the starting structure, standing out from all others. A possible reason for these differences may be that the constraints imposed by the nucleosome arrangements are not the sole contributors to the linker DNA heterogeneity. Instead, the potential self-repulsion of the DNA, and, more importantly, the presence of the linker histone protein might contribute to this variable accumulation of twist along the DNA. The difference in twist, incurred by EM fitting, compared to elastically optimized linkers is shown in Figure 2.6.

When we plot the net bending of successive base pairs along each DNA linker we can see that, unlike for twist, there are differences among the four groups after elastic optimization, even before considering the EM density fit (Figure 2.7). These differences reflect how the docked nucleosomes are arranged in three tetra-nucleosome units, and impose different spatial constraints on the intervening linkers. These differences in net bending become even greater after fitting the DNA linkers to the EM density. The differences in bending, before and after EM density fitting are shown in Figure 2.8).

From both the twist and the bending profiles, we can see that the modeled DNA linkers are not straight, as they might appear from rough 3D models [65]. The linkers have significant curvature, relatively highly bent regions and a variable twist, compared to relaxed, ideal B-DNA. These changes likely reflect the binding of the linker histone protein, which is known to interact with linker DNA. From this analysis, there is not a clear signal for every linker as to where it might bind the linker

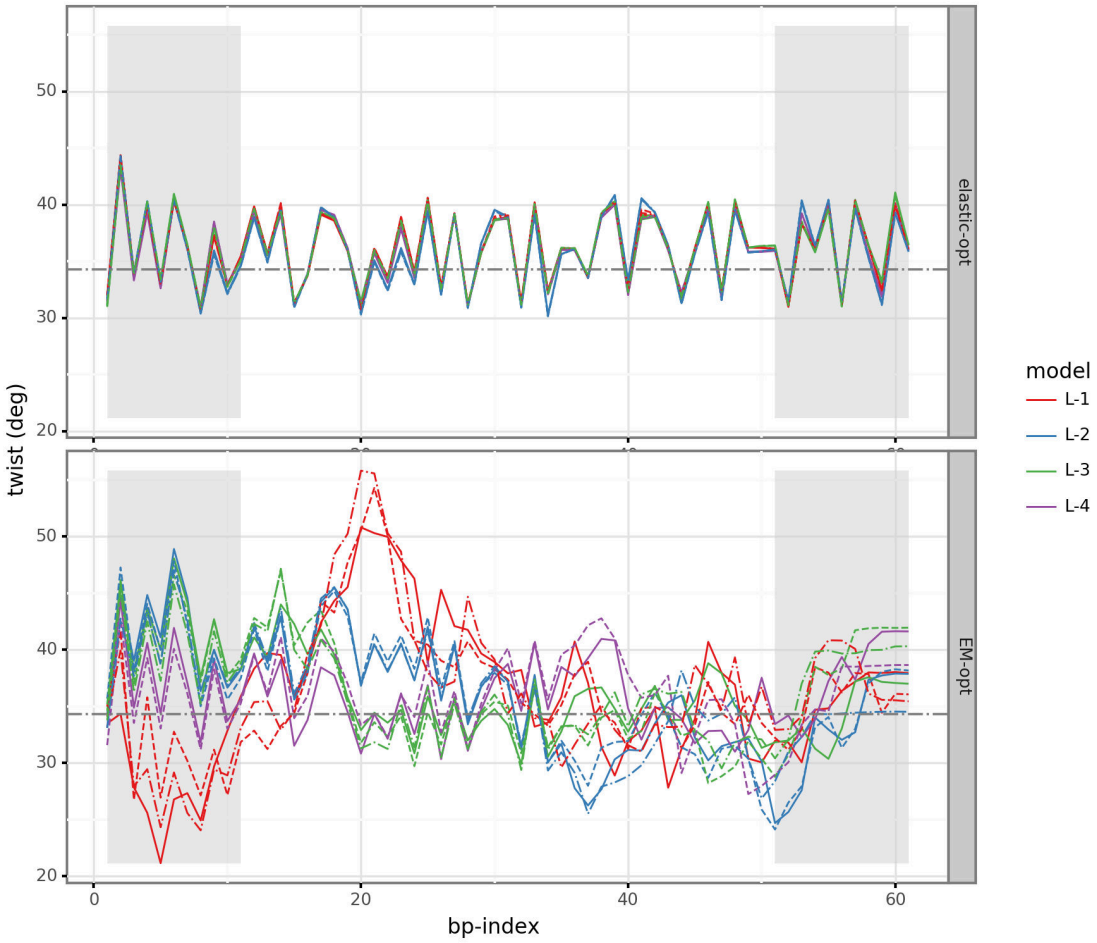


Figure 2.5: Local twist (p_3) along DNA linkers after optimization of the elastic energy (top panel) and after fitting to the EM density map (bottom panel). The graphs are colored to reflect the four linker categories within each of the tetra-nucleosome units, e.g. linkers connecting nucleosomes N1-N2, N5-N6, N9-N10 are shown in red (L-1), linkers N2-N3, N6-N7, N10-N11 are shown in blue (L-2), and so on. Within each group, the values are plotted in different line styles, going from dashed to dash-dot to solid, in order of increasing index. The gray shaded regions highlight the nucleosomal DNA near the exit (left) and entry (right) sites. The elastically optimized DNA linkers have nearly identical twist patterns. After fitting to the EM density, the DNA linkers display regions of strong local twist deviation.

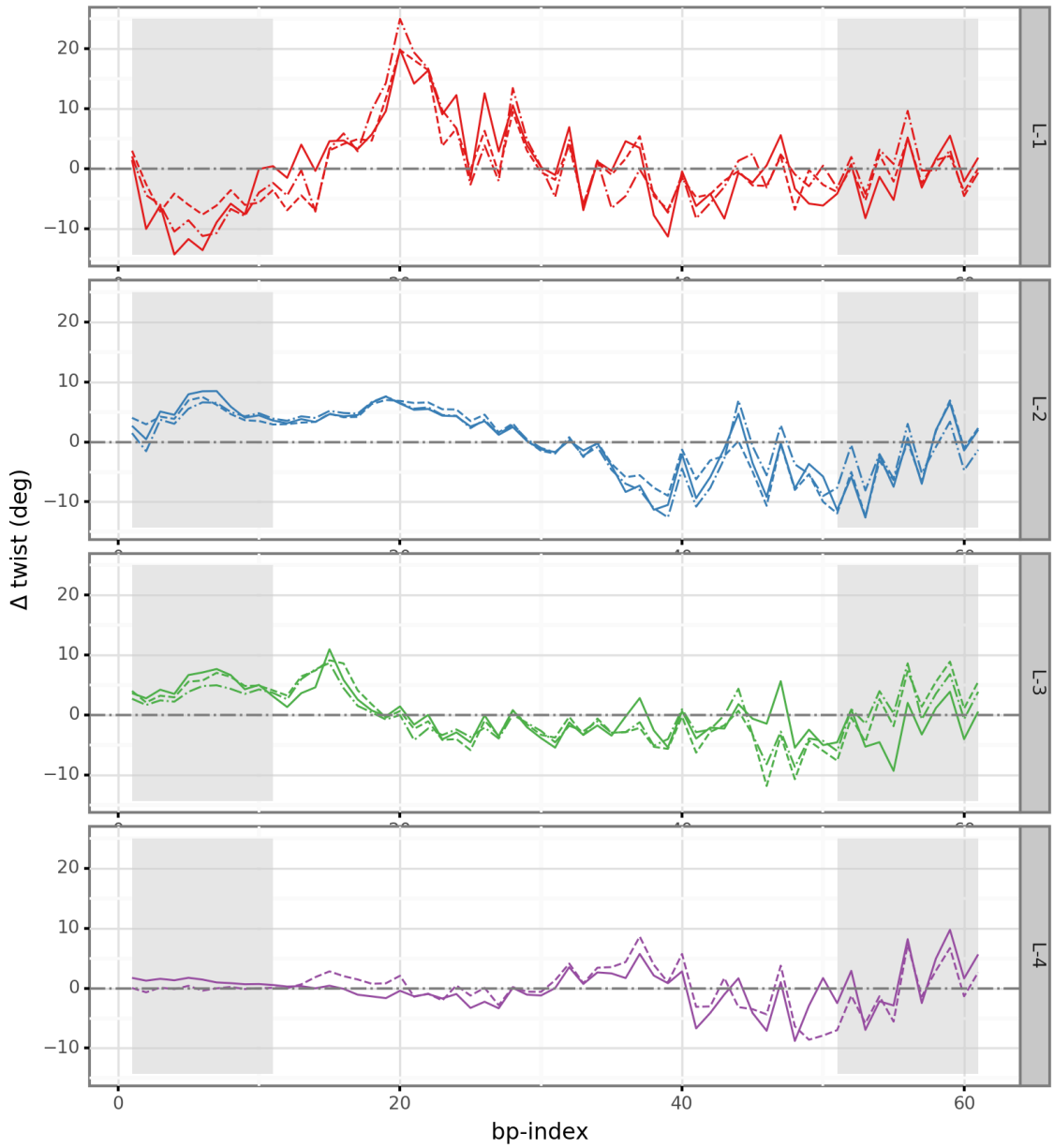


Figure 2.6: Differences in local twist (p_3) along DNA linkers before and after fitting to the EM density map. After fitting to the EM density, the DNA linkers display regions of strong local twist deviation, compared to elastically optimized linkers. Linkers that connect nucleosomes 1 and 2 within each tetramer unit show the largest deformation (L-1, first panel). The colors and line styles are the same as those in Figure 2.5. The gray shaded regions highlight the nucleosomal DNA near the exit (left) and entry (right) sites.

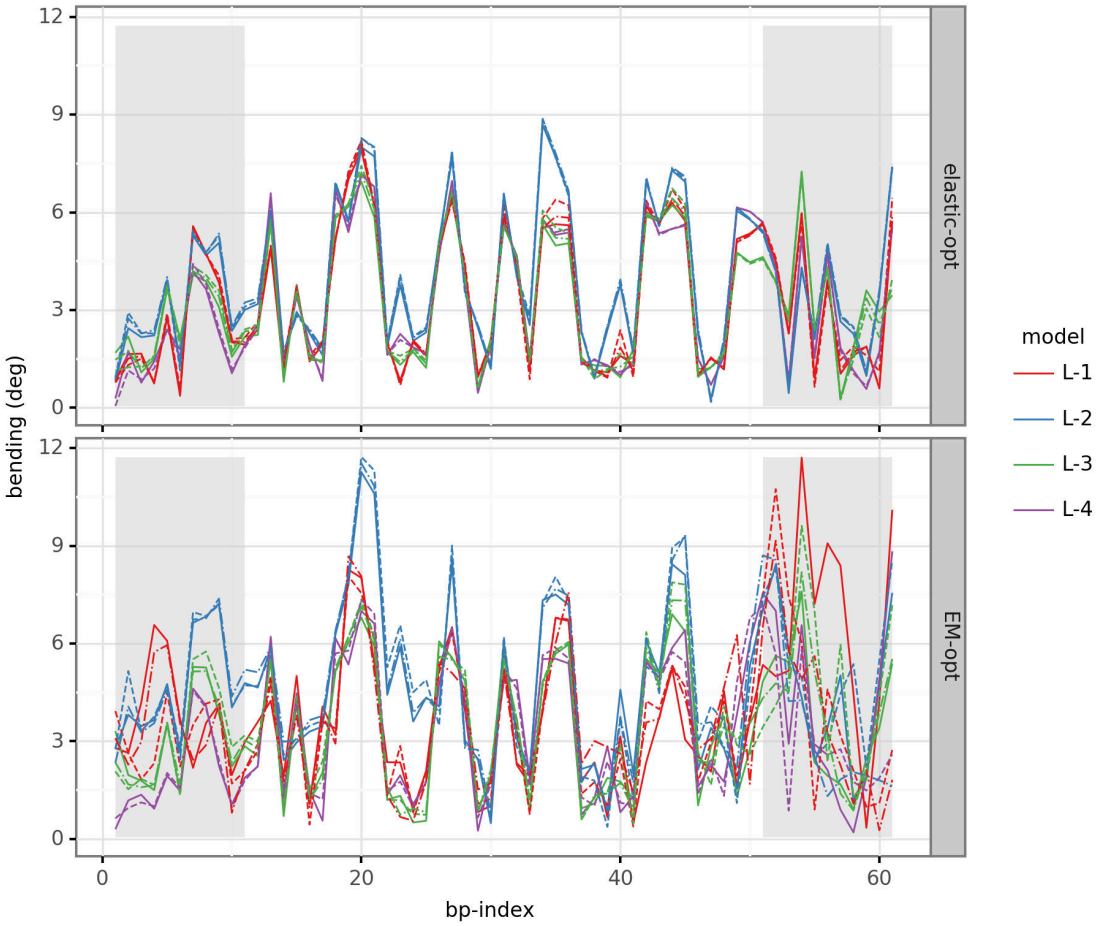


Figure 2.7: Net bending (Ω) of successive base pairs, defined from tilt (p_1) and roll (p_2) values as $\Omega = \sqrt{p_1^2 + p_2^2}$, along DNA linkers after optimization of the elastic energy (top panel) and after fitting to the EM density map (bottom panel). The graphs are colored to reflect the four linker categories within each of the tetra-nucleosome units, e.g. linkers connecting nucleosomes N1-N2, N5-N6, N9-N10 are shown in red (L-1), linkers N2-N3, N6-N7, N10-N11 are shown in blue (L-2), and so on. Within each group, the values are plotted in different line styles, going from dashed to dash-dot to solid, in order of increasing index. The gray shaded regions highlight the nucleosomal DNA near the exit (left) and entry (right) sites. The net bending along the DNA linkers shows that they do not follow straight pathways. The ~5-bp pattern is indicative of curvature. This reflects the constraints induced by the relative orientation of successive nucleosomes.

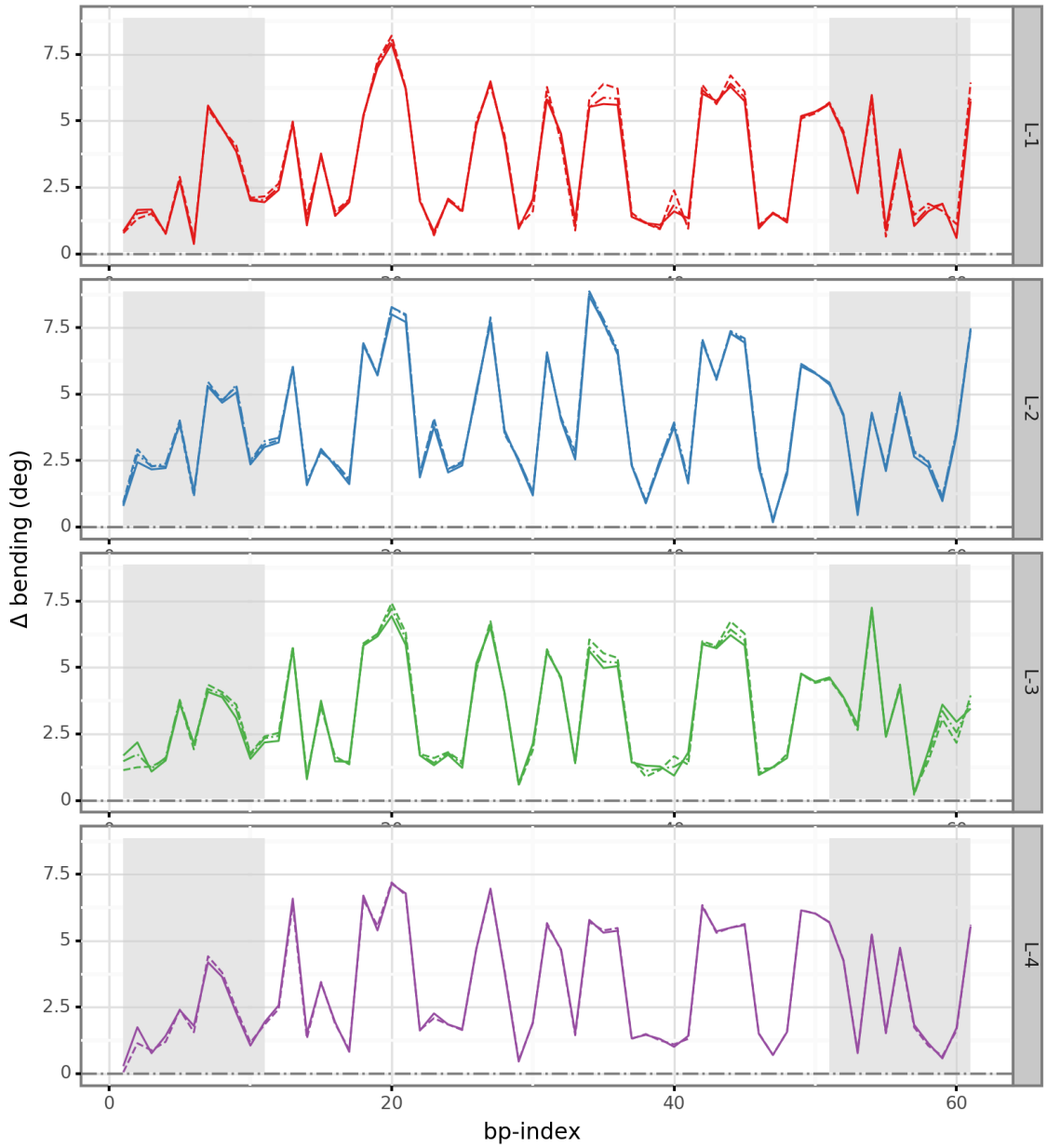


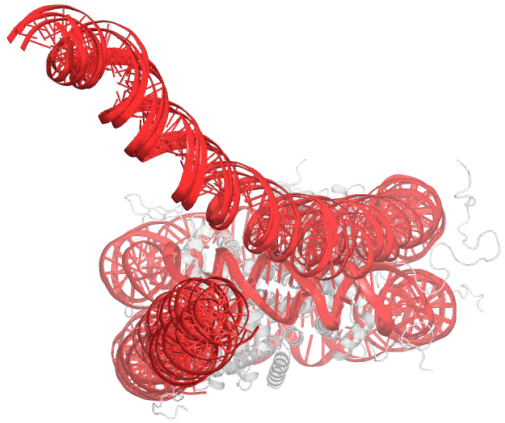
Figure 2.8: Differences in net bending ($\Delta\Omega$) of successive base pairs, along DNA linkers before and after fitting to the EM density map. After fitting to the EM density, the DNA linkers contain regions of higher net bending, compared to elastically optimized linkers. Linkers that connect nucleosomes 2 and 3 within each tetramer unit show the largest bending (L-2, second panel). The colors and line styles are the same as those in Figure 2.7. The gray shaded regions highlight the nucleosomal DNA near the exit (left) and entry (right) sites.

histone, but there are significant differences seen among the linkers to suggest that there may be more than one interaction mode. The linkers connecting nucleosomes N1-N2, N5-N6 and N9-N10, are the ones that show the highest positive twist accumulation at around +10 bp from the exit site (peak at bp-index 20 in Figure 2.5). This location corresponds to one of the presumed interaction sites for the globular domain of the linker histone [65] and the over-twisting here is offset by an under-twisting of the last 10 bp of the DNA in the nucleosome that precedes it (dip at bp index 5 in Figure 2.5). When we compare the DNA pathways of the elastically optimized linkers and the EM-fitted linkers, we can see that some of them incur strong deformations, especially near the entry/exit sites while others do not. For example, L-4 linkers, which connect nucleosomes between tetramer units, show no differences in twist or net bend as they exit the nucleosome (Figures 2.6 and 2.8, bottom panel).

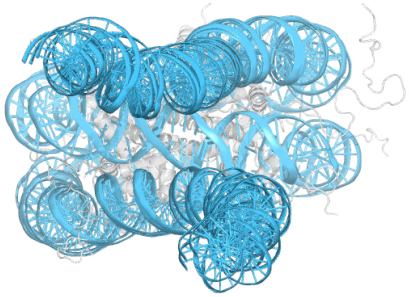
Using molecular visualization of the reconstructed nucleosomes (nucleosome core + linker), it is also apparent that there are four different entry/exit DNA linker arrangements throughout the fiber. Figure 2.9 shows cartoon representations of all the docked nucleosomes and modeled entry/exit DNA linkers in the same view, looking down the dyad axis. The superimposed nucleosomes in any section of the grid have the same position in the tetra-nucleosome unit. The linkers are arranged very differently, based on the position in the tetramer, but follow very similar pathways within each group. The differences in the DNA linker arrangements mean that they present very different environments for linker histone binding. It is therefore possible, according to our cryoEM reconstruction, that there could be multiple binding modes throughout the chromatin fiber.

2.3.3 GH1/H5 docking models

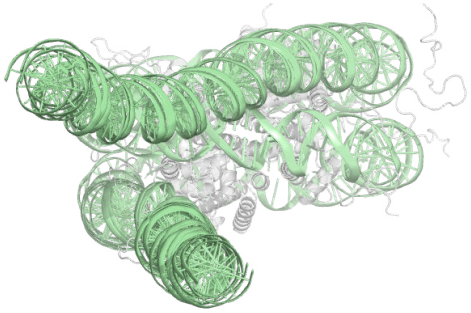
After modeling the nucleosome core particles and DNA linkers to the EM density map, we extracted the remaining density regions, which are assumed to correspond to the linker histone globular domain (Figure 2.10). At this point the remaining regions are very noisy and not well defined, and some of the regions corresponding to adjacent linker histones [65], are difficult to segment from one another. This reflects the low resolution of the density in the first place, the flexibility of the linker histone H5 globular domain (gH5) fold itself, and the potential heterogeneity of its positioning. Our approach to gH5 modeling is complementary to the one described in the original work [65]. We use an ensemble of available high-resolution structures of the globular domains of linker histones H1



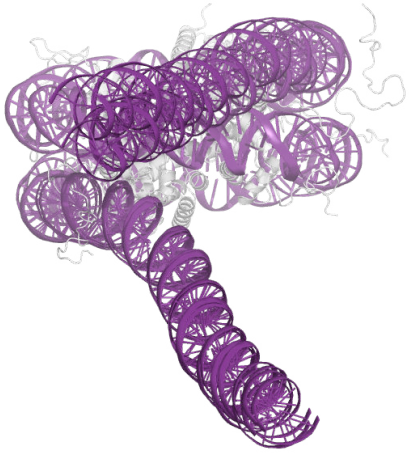
(a) Nucleosomes N1, N5, N9



(b) Nucleosomes N2, N6, N10



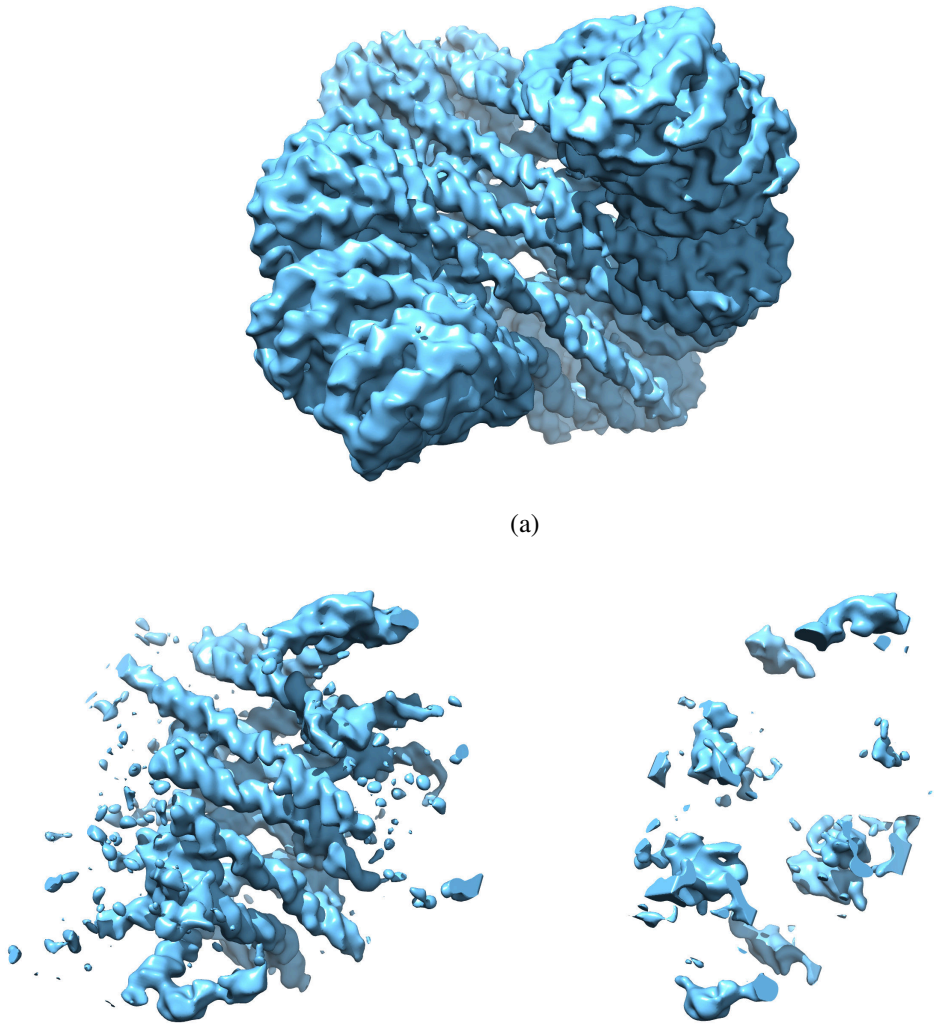
(c) Nucleosomes N3, N7, N11



(d) Nucleosomes N4, N8, N12

Figure 2.9: DNA linkers entering and exiting each nucleosome in the reconstructed chromatin fiber structure superimposed on the nucleosome core particle. The models are shown in the same view, looking down the dyad axis, where DNA enters near the bottom face and exits near the top face of the nucleosome core particle. From this view, we can see that these DNA linker arrangements fall into four different categories, colored accordingly. Nucleosomes N1, N5 and N9 are shown in red (a); N2, N6 and N10 in blue (b); N3, N7 and N11 in green (c); N4, N8 and N12 in purple (d).

or H5 (gH1/gH5, in Table 2.1) to fit into each region individually. It is worth noting that due to the limitations mentioned above, this modeling was done to get a general understanding of the potential binding heterogeneity and interactions of the linker histone globular domain with the nucleosome and not to make any determination about specific docking models.



(a)

(b) (c)

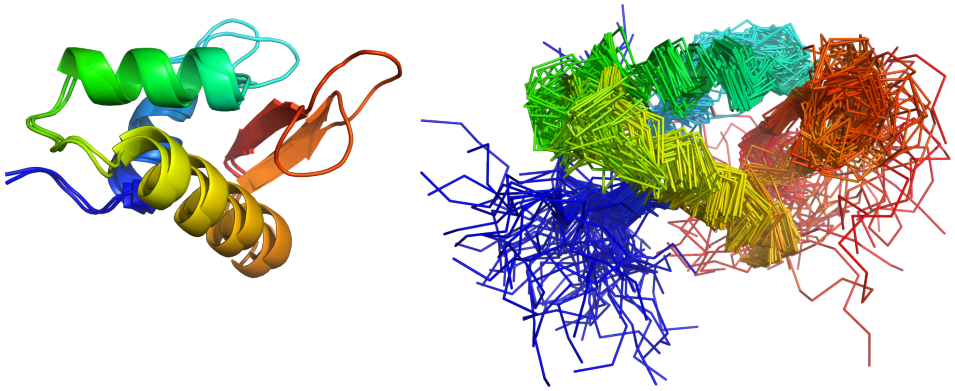
Figure 2.10: EM density map for the 187-bp repeat fiber: (a) as obtained from the EM data bank; (b) after fitting and extracting densities corresponding to nucleosome core particles; (c) after extracting nucleosomes and intervening linker DNA. The density regions shown in (c) are assumed to account for the globular domain of the linker histones. All figures were generated using Chimera [76].

We used a combination of crystal and NMR structures for the globular domain of linker histone H1 and H5, coming from different organisms (Table 2.1). The globular domain is generally conserved

among variants, and even across homologues and species [34, 91], which results in good structural similarity between the models (Figure 2.11). We leverage the structural heterogeneity present across these different variants and within the many NMR models as an inexpensive way to explore the effect of small structural perturbations on the docking model.

Figure 2.11 shows a structural superposition of all the gH1/gH5 structures used here. The left side shows a cartoon representation of the two monomers of the dimeric crystal structure of a chicken erythrocyte linker histone H5 globular domain, the first high-resolution structure of the isolated protein [36]. The protein fold is characterized by a three-helix bundle with a beta hairpin and bears strong similarities to a winged helix motif, which is common among many DNA binding proteins. The biggest difference between these two monomers lies in what is referred to as the "loop 3" region, near the beta sheet that constitutes the "wing" of the motif (orange-reg segment in Figure 2.11). When other linker histone structures are considered, which include an NMR structure of a yeast homologue that displays a more typical "winged helix" DNA binding motif [88], the structural heterogeneity of the ensemble increases within the alpha helical regions as well (Figure 2.11B).

After rigid-body docking every gH1/gH5 model into every site of the remaining EM density regions, we pick only the best-scoring fit for each model (according to the ranking from Situs) for further analysis. We then generate the contacts between the linker histone globular domain and DNA to obtain a general profile for the nature of the interactions across the chromatin fiber. As the data in Figure 2.12 suggest, the globular domain of the linker histone interacts differently with the nucleosomal and linker DNA depending on its location along the fiber. In general, the data suggest that the globular domain interacts in an off-dyad mode, while making contact with both entry and exit DNA linkers, as suggested in the original work [65]. The interaction with nucleosomal DNA (Figure 2.12, central column) around the dyad occurs closer to the exit site (negative bp indexes near dyad are closer to exit linker) in nucleosome classes 1 and 2, and closer to the entry site (positive bp indexes near dyad are closer to entry linker) in classes 3 and 4. These observation are a consequence of the asymmetric interactions of gH1/gH5 with the DNA linkers (Figure 2.12, first and third column). This alternation in the binding side breaks the pseudo-symmetry of the nucleosomes and reflects the tetra-nucleosome repeat of the chromatin fiber, with adjacent linker histones interacting between tetra-nucleosomes but not within [65]. For example, in nucleosomes N1 and N2 the linker histone interacts mainly with the exit DNA linkers, while in N3 and N4 interact mainly with the entry DNA



(a) (b)

Figure 2.11: Superimposed models of the globular domain of linker histones H1/H5 (gH1/gH5) used for docking. Both figures are colored using a rainbow color scheme, transitioning from blue to red along the N-terminus to C-terminus direction. Figures created using PyMOL [48]. (a) Crystal structure of chicken gH5 by Ramakrishnan et al. [36] shown in a cartoon representation. There are some differences between each monomer in the asymmetric unit, especially in the loop connecting the two beta strands (red-orange loop) referred to as the wing region [36] and smaller differences seen in the alpha helical units. (b) All monomeric models used in this study, superimposed on the crystal structure from (a) and shown using a backbone trace representation. It is obvious that the biggest differences among these structures lie in the N- and C-terminal regions, as well as in the wing motif near the C-terminus.

linker. However, the subtle differences between the contacts in nucleosomes N1 and N2 may suggest different binding modes for the entry side model. The same thing can be said for nucleosomes N3 and N4. The categories used to group the chromatosomes (nucleosome core + entry/exit DNA linker + linker histone) are defined the same way as the nucleosomes in the previous section: i.e. chromatosomes N1, N5, N9 constitute group 1, chromatosomes N2, N6, N10 constitute group 2 and so on. The distribution of contacts shown for each chromatosome group is normalized along the entire length of DNA, including nucleosomal DNA around the dyad and the DNA linkers (absolute bar heights over an entire row in Figure 2.12 add up to unity). Since our rigid-body docking of gH1/gH5 structures involves only the EM density and does not concern the nucleosome core particle, many of these resulting models contain clashes, and are therefore purely descriptive. As a result, the very high number of close interactions with the exit DNA linker in the first group of chromatosomes (Figure 2.12, top row) is the main reason behind the apparent low frequency of interactions between the linker histone globular domain and nucleosomal DNA.

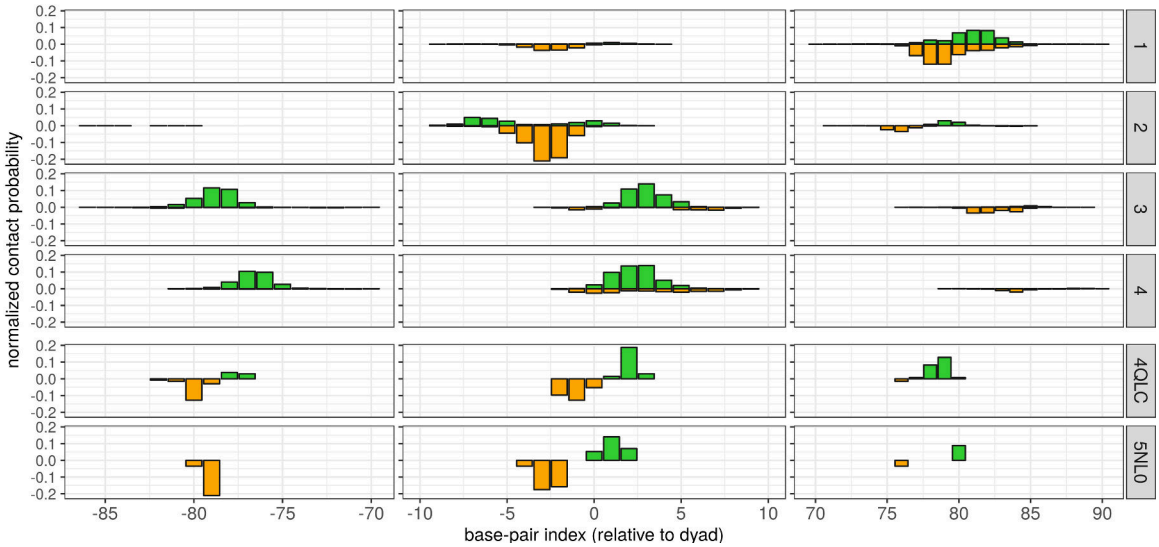


Figure 2.12: Contacts between the gh1/gh5 models and DNA in the chromatosomes of the modeled 187-bp repeat fiber and two high-resolution crystal structures. The contacts are grouped by coding (green, positive bar) and template (orange, negative bar) strands and normalized over the entire length of the DNA, including both linkers and nucleosomal DNA (absolute bar heights over an entire row add up to unity). Numbering is shown relative to the dyad. The top four rows of panels show the contact distributions for each of the four groups of chromatosomes from the modeled EM fiber, whereas the bottom two rows represent the contacts for the two crystal structures with PDB ID: 4QLC [46] and 5NL0 [47], respectively.

For comparison, rows 5 and 6 show the same contact analysis for the two recent chromatosome crystal structures: the first one containing 165 bp of DNA in the presence of the globular domain of chicken H5 (PDB ID: 4QLC [46]) and the second containing 197 bp of DNA in the presence of full length H1.0 from African tree frog (PDB ID: 5NL0 [47]). Both of these structures describe a chromatosome where the globular domain of the linker histone interacts with nucleosomal DNA as well as both DNA linkers in an on-dyad binding mode, very similar to one another. It was suggested through mutation studies and NMR that there are several residues in the globular domain of linker histones that are responsible for the on- or off-dyad character of the interaction [49]. However, it is worth noting that both high-resolution crystal structures published so far represent mono-nucleosomes, which means that the entry/exit linker DNA are not under the same constraints as a nucleosome array.

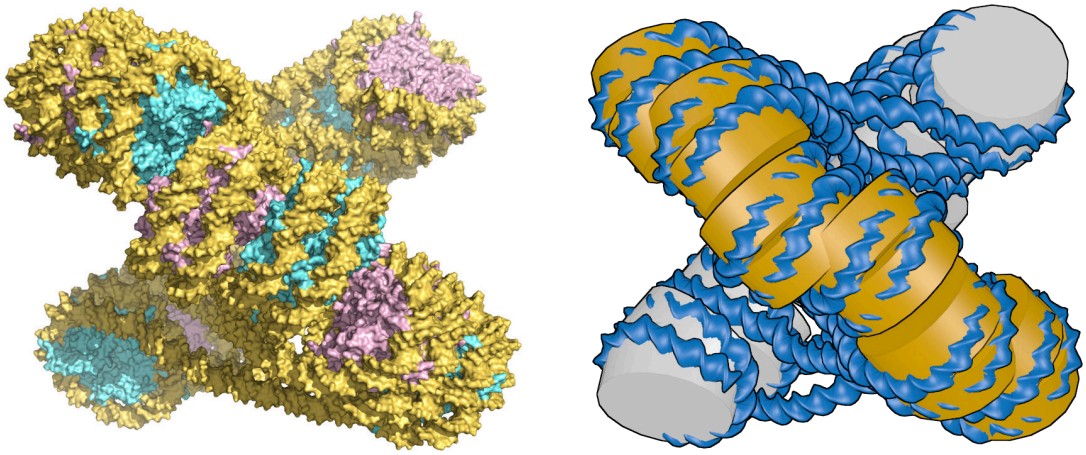
Discussion

Our atomic-level chromatin model from the cryoEM density map of a 12×187 bp fiber suggests that the nucleosomes in the tetramer units may not be homogeneous. While the density map shows a two-start crossed linker structure that is generally considered to have straight DNA linkers (as opposed to the highly bent linkers in a one-start solenoid model), our modeling suggests that the DNA linker pathways are neither straight nor identical throughout the fiber. Figures 2.5–2.8 show that besides the generally small differences among DNA linkers dictated by the different nucleosome orientations in a tetramer unit, the observed deviations may be induced by their interaction with the linker histone. These various arrangements of the DNA linkers may suggest multiple modes of interaction with the linker histone. Besides the interaction with the globular domain, the DNA linkers also interact with the C-terminal domain, which is present in the system but not resolved in the density, due to its unstructured nature.

Very recent work has suggested that the unstructured linker histone C-terminal domain (CTD) maintains the disordered state while binding to DNA with very high affinity [92, 93]. The authors propose the possibility that the CTD creates a liquid-like environment and phase-separates with DNA linkers in the interior of the chromatin array to promote fiber compaction. This implies that neither the CTD, nor linker DNA need to adopt particular rigid arrangements in order for this interaction to occur. Another recent work reported the observation of an ultra-high affinity interaction between two disordered proteins while fully maintaining their disordered state, with one of the proteins being the H1-CTD, and suggested this new type of interaction mechanism could be abundant in eukaryotes [94]. Such an interaction could potentially account for the ill-resolved linker histone density in the cryoEM map and the heterogeneity in the DNA linkers we found in the model that we built. Recent tetranucleosome crystal structures also reveal a heterogeneity in nucleosome-nucleosome interactions and higher-order structure arrangements [64].

However, our analysis also reveals a large discrepancy between the expected and observed inter-nucleosome distances in our model for the 12×177-bp density map [65]. Two of the nucleosome arrangements within the tetranucleosome units would require the intervening DNA to be stretched by about 25% or 50%, respectively (Table 2.2). This may reflect an issue in the modeling of the cryoEM density from EM images. Any potential heterogeneity, as we discussed above, could result

in incorrect density modeling if the EM images belonging to different structure are not properly classified. Zhou et al. also argue that the chemical cross-linking between nucleosomes can affect not only their mode of interaction, but also the binding mode of the linker histone [95]. Furthermore, they suggest that cross-linking in general may induce further compaction and adoption of a specific structure of the chromatin fibers that would otherwise be dynamic and heterogeneous. These are critical issues that need to be addressed in future cryoEM work.



(a) (b)

Figure 2.13: Reconstructed chromatin fiber model from cryoEM density map of a 12×187-bp nucleosome array [65]. (left) Surface representation of the approximate atomic-level model of nucleosome core particles and intervening DNA. DNA is shown in yellow, while histone core is shown in blue and pink to distinguish adjacent nucleosomes in each stack. (right) Coarse-grained representation of the same cryoEM fiber model, using a representation described in Chapter 3. The yellow and white colors highlight the two stacks formed by non-consecutive nucleosomes.

The method we presented here can be used to model flexible DNA into any density map at a meso-scale level. With higher resolution cryoEM maps of more chromatin fibers potentially on the way, our approach can be used to model the DNA very efficiently. The base-pair representation of DNA can be converted to full atomic coordinates [79] that can be used for further refinement with other methods, if the density resolution allows it. CryoEM has become a very powerful technique over the past decade and the resolution it can achieve has undergone a revolution. Since chromatin can be notoriously flexible in its natural state and X-ray crystallography is not always possible, cryoEM is poised to be the method of choice for teasing structural details out of long chromatin

fibers. Furthermore, we can use our approximate atomic-level (Figure 2.13) as a basis for coarse-grained simulations of nucleosomal arrays, which are discussed in the next chapters. The different arrangements of the DNA linkers entering and exiting the nucleosomes can be used as models to predict the behavior of chromatin as a result of these different constraints.

As a direct continuation of this work, the ensemble of models for the linker histone globular domain (gH1/gH5) docking on the nucleosome need to be further analyzed. We need to look for the existence of any preferred binding modes and compare them with recent structural data from experiments. With more information coming out very recently on the positioning of the globular domain in the nucleosome [46, 47, 49, 95], it will be interesting to see if any of those models are represented in our docking ensembles. Furthermore, we can use the gH1/gH5 docking models and the DNA linker arrangements from our model as constraints to generate binding ensembles of the linker histone CTD using *de novo* folding methods, such as those built into Rosetta [96]. These efforts are currently underway.

2.4 Conclusions

In this chapter we have built an approximate atomic-level model from the cryoEM density map of a 12×187 bp chromatin fiber. We have presented a method for modeling DNA onto a density map in an efficient and robust manner, potentially as a first step towards full atomic modeling. Our results suggest that the chromatin fiber may contain heterogeneous DNA linker arrangements between nucleosomes and these differences may reflect the interaction between linker DNA and the linker histone. At the same time, our analysis points to potential critical issues in one of the cryoEM density maps. We can use the modeled chromatosomes from our model in coarse-grained simulations of nucleosomal arrays, which are discussed in the following chapters. In future work, we need to further analyze the linker histone globular domain binding models, and use that knowledge to model linker histone C-terminal domain.

Chapter 3

Role of nucleosome folding in global chromatin properties: a meso-scale model of chromatin

3.1 Introduction

Despite decades of research, it still remains unclear how chromatin folds inside a cell nucleus. Structural data from experiments have been sparse, but many computational approaches have been developed to account for these data [2, 4, 61, 97–101]. The proposed structures generally fall into two categories: one-start helices (or solenoidal fibers) with bent linkers, and zig-zag models with straight crossed linkers, usually with 2-start or multi-start helical nucleosome arrangements [102].

Experimental structures of *in vitro* reconstituted nucleosome arrays containing 4–12 nucleosomes and various linker lengths suggest that chromatin adopts a 2-start helix arrangement, with a tetra-nucleosome unit as a repeating motif, forming two stacks of nucleosomes twisted around each other [62, 64, 65, 103]. Measurements from electron microscopy studies and subsequent modeling, however, have suggested that the chromatin fiber diameter is incompatible with a 2-start helix and a solenoid model has been proposed, where interdigitation can allow for face-to-face stacking of non-consecutive nucleosomes [57]. Recent data from a high-resolution genome-wide chromatin conformation map obtained using an ionizing radiation-induced spatially correlated DNA cleavage

mapping technique, suggest that chromatin exists as en ensemble of states, displaying various arrangements of nucleosomes compatible with 2-, 3- and multi-start helical arrangements [104]. An ultracentrifugation study of chromatin fiber compaction found a strong dependence on nucleosome repeat length and solvent conditions [58]. The study suggested that the intrinsic properties of DNA linkers determine the folding of nucleosome arrays and this dependence becomes weaker with increasing repeat length. The multitude of factors that affect chromatin structure has strengthened the view of chromatin as a heterogeneous ensemble of states, instead of a single compact fiber structure [71, 95, 100, 102, 104].

In this study, we investigate the role of nucleosomal DNA folding on chromatin fiber architecture. We have analyzed the nucleosome-bound DNA obtained from numerous available high-resolution structures and investigated how the differential wrapping affects the properties of simulated nucleosome arrays. To the best of our knowledge, this large-scale analysis of nucleosomal DNA from high-resolution structures has not been reported previously. We use a meso-scale treatment of chromatin, previously used in studies of long-range enhancer-promoter communication [2, 3, 5, 6] as well studies of chromatin fiber flexibility and dependence on nucleosome repeat length [4, 105]. We use Markov chain Monte Carlo (MCMC) simulations to generate structural ensembles for various nucleosome arrays. We use several approaches to characterize the fiber geometry by treating the chromatin as a dynamic arrangement of nucleosomes. Fiber compaction results, expressed in terms of estimated sedimentation coefficient, are compared to the experimental data from Correll et al. [58]. We use the treatment from Miyazawa [106, 107] to characterize the fiber geometry in terms of local helical arrangements of nucleosomes.

Unlike previous computational approaches, we use a rigid-body treatment of DNA that allows us to capture details at the level of single base pairs, instead of a bead model that represents several base pairs [98, 108]. We do not impose an explicit interaction energy term for nucleosome stacking [109, 110]; the transient nucleosome contacts occur as a result of electrostatic interactions and constraints imposed by nucleosomal and linker DNA.

Our simulations predict an irregular and dynamic chromatin fiber structure, consisting of a mixture of various local nucleosome arrangements. The overall architecture and flexibility depend on nucleosome repeat length [4, 98, 100]. The range of our predicted fiber geometries is in general agreement with previous computational and experimental work [100, 104]. Our meso-scale model

and simulations predict that the differential folding of DNA around the nucleosome cores, as observed in crystal structures, can lead to different fiber geometries. These fiber geometries are sensitive to partial DNA unwrapping, in a manner that is dependent on the nucleosomal DNA folding, and can be modulated by linker histone binding. Our characterization of fiber geometries in terms of local helical arrangements allows us to describe the fiber as a heterogeneous ensemble of different structures. The predicted fiber geometries, in turn, have implications on larger-scale chromatin properties such as the ability to support long-range communication, which is discussed in Chapter 4.

3.2 Methods

3.2.1 DNA model

In our mesoscale treatment of chromatin, similar to the description in Section 2.2.2, DNA is represented in terms of a set of rigid-body parameters that specify the relative orientation and displacement of successive base pairs [79, 84]. The unbound linker DNA connecting successive nucleosomes is governed by an ideal elastic potential that allows for deformations of double-helical structure consistent with the solution properties of DNA [111]. The equilibrium rest state of the free DNA is that of an ideal, perfectly straight molecule with 10.5 base pairs per turn and 3.6 nm pitch, equivalent to 34.3° helical twist and 0.34 nm vertical displacement at each base-pair step (Eqn. 2.3). The elastic force constants used here are not coupled and their magnitudes are such that a bend of 4.84° or a change in twist of 4.09° raises the energy by 0.5 $k_{\mathrm{B}}T$, the base-pair level equivalents of a bending persistence length of 47.7 nm and a twisting persistence length of 66.5 nm.

The chain is treated as inextensible, with the spacing of successive base pairs held fixed. This treatment of DNA as a collection of base-pairs, each with its own reference frame, is necessary to keep track of the precise position and orientation of all the protein-DNA complexes (e.g. nucleosomes) assembled along the molecule. This "high-resolution" treatment allows us to capture the effects of single base-pair shifts in nucleosome positioning [105], as well as DNA sequence-specific effects, even though the work described here does not consider sequence specificity. The collection of rigid body parameters of the unbound DNA base-pair steps constitutes the primary parameter space in our simulations.

The negative charges along the DNA backbone are coarse-grained by representing the charges

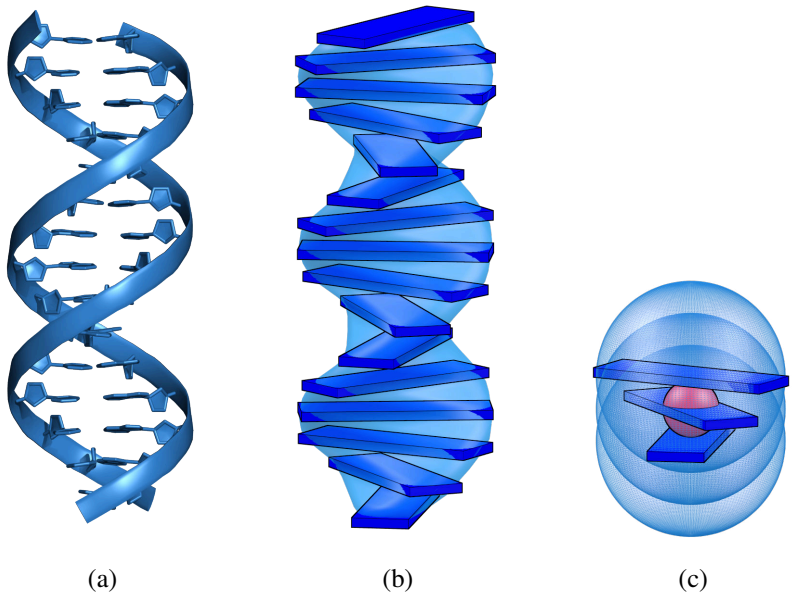


Figure 3.1: Treatment of DNA in the meso-scale model of chromatin: 1.5 turns of DNA with 10 bp/turn shown for illustration. (a) A cartoon representation of a full atomic model, with a ribbon connecting the sugar-phosphate backbone. Image created with PyMOL [48]. (b) DNA is treated as a collection of base pairs, whose reference frames are depicted here by thin blocks. The relative displacement and orientation of every two base pairs is described by a set of six rigid-body parameters. DNA in most figures of nucleosome arrays in this work is illustrated by a smooth twisted ribbon, shown here as a translucent blue volume, which encompasses the base pairs. (c) The combined charge for every three base pairs of DNA is placed at the origin of the central base pair. Every base pair is assigned a spherical excluded volume with diameter equivalent to that of the B-DNA double helix, or 20 Å. Figure adapted from Nizovtseva et al. [6] with permission from Future Medicine.

for every three consecutive base pairs by a single charge placed at the origin of the central base-pair reference frame (Figure 3.1). That means the first charge is placed at the center of base pair 2, the second charge at the center of base pair 5 and so on. When the total number of base pairs is not an integer multiple of 3, the last charge is placed at the center of the last base pair and in that case it only represents the last 1 or 2 base pairs, accordingly. Our coarse-grained system contains charges of –1.5 esu for every three base pairs, or the equivalent of –0.25 esu for every phosphate atom in the backbone. This 75% reduction in phosphate charge reflects the effective charge upon counterion neutralization of B-DNA in a dilute solution, according to Manning's predictions [112].

Since our simulations account for excluded volume effects, we use spheres centered on the origin of each base-pair frame to represent the volume for each base pair. The radius of each sphere is taken as 10 Å, which corresponds to half of the average diameter for an ideal B-DNA molecule.

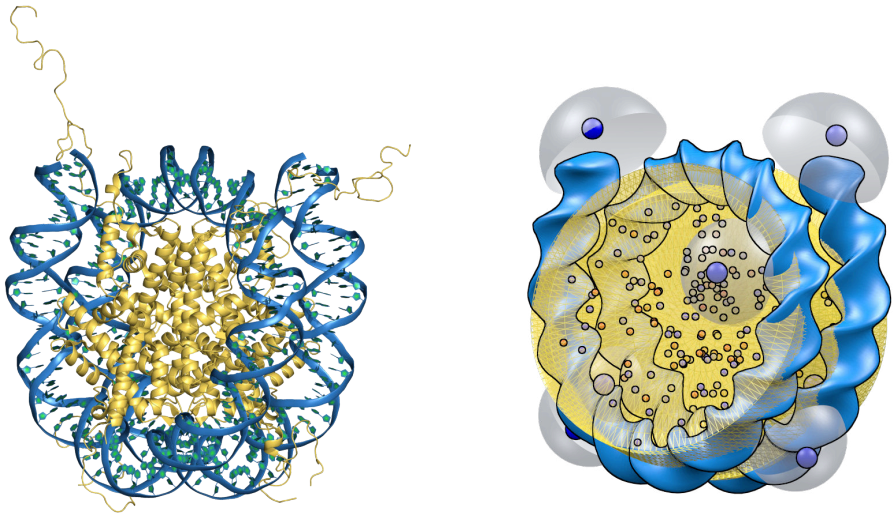
3.2.2 Nucleosome core particle model

The nucleosomes are modeled as rigid bodies associated with a set of frozen DNA base-pair steps and contain a set of charges, based on the atomic details in available high-resolution structures. Initially our model incorporated these details from the currently best-resolved 1.9 Å nucleosome core particle crystal structure (Figure 3.2a) [15]. Other high-resolution structures can be implemented in a similar fashion, which is a topic discussed in subsequent sections.

The shapes and charges of the histone octamer are described in a local nucleosome reference frame and this local frame is connected to the global frame of the DNA molecule via the local coordinates of an anchoring DNA base-pair. A natural choice for this base pair is the central base pair located on the nucleosome dyad of the selected crystal structure with 147 bp of DNA. In the work presented here, we use a nucleosome reference frame with its origin defined by the mean of the DNA base-pair centers. The $\boldsymbol{z}$-axis lies along the superhelical (or cylindrical) axis of the bound DNA, which is defined, in turn, by the line that is equidistant from all base-pair origins. The $\boldsymbol{x}$-axis points towards the dyad and the $\boldsymbol{y}$-axis follows the standard right-handed rule [1, 4]. The choice of this nucleosome frame does not affect the simulations, as long as the dyad base-pair reference frame is defined within it. The excluded volumes of the two tetrameric core histone motifs are approximated by cylinders that generally encompass the locations of the C_α atoms. The center of each cylinder is defined by the average position of the C_α atoms in the tetramer and the dimensions are based on principal component analysis of the atom positions (Figure 3.2b).

The electric charges of the nucleosome histone proteins are split into two groups: those belonging to the trypsin-resistant histone globular domains [25, 113] and those belonging to the histone N-terminal tails, and in the case of H2A histones, also to the C-terminal tails. The charges within the histone core consist of 120 positive charges from the cationic atoms of arginine and lysines (histidines are not treated as protonated at physiological pH) and 66 negative charges from the anionic atoms of aspartic acid and glutamic acid. The full set of 186 formal charges (Figure 3.2b, small spheres inside the cylindrical volumes) is clustered into a smaller subset of coarse-grained charges, for each tetramer individually, on the basis of their inter-atomic distances while preserving the net charge. The number and size of clusters are determined by an optimization procedure that minimizes the difference at representative points between the electrostatic potential (Debye-Hückel)

of the simplified model and the corresponding potential determined for all charged atoms in the octamer core. This is done to reduce the cost of calculating electrostatic interactions, which can be quite expensive. From our investigation this does not seem to affect the outcomes of our simulations, at least in terms of global chromatin fiber properties.



(a) (b)

Figure 3.2: Treatment of nucleosomes in the meso-scale model of chromatin. (a) Cartoon representation of the 1.9 Å crystal structure from Davey et al. (PDB ID: 1KX5 [15]). (b) Coarse-grained representation of the nucleosome by two cylinders, one for each H2A-H2B-H3-H4 histone tetramer in the core. There is a set of fixed point charges inside the cylinder volumes representing the core histone charges and "immobile" portions of trypsin-cleaved histone tails (smaller spheres). A set of charges placed outside the core (larger spheres) approximates the motions of charged residues on "mobile" histone tails. These composite charges are allowed to move inside a half-sphere volume, with respective radii representing the predicted radius of gyration for each tail [114]. The sizes of the spheres shown here are not proportional to the charge magnitudes that they represent. Figure adapted from Nizovtseva et al. [6] with permission from Future Medicine.

The histone tails include 100 positively and 6 negatively charged amino acids. Each histone tail is represented by a set of fixed charges, some of which fall within the coarse-grained volume of the core, and a set of mobile point charges that are allowed to randomly move within a spherical cap anchored at the point where the tail exits the nucleosome core (Figure 3.2b). The dimensions of these spherical caps (9-20 Å) are based on estimates for the radii of gyration of the histone tails obtained from atomic simulations [114]. The number of mobile charges can be coarse-grained at various levels — for the work presented here, the number is taken as half the sum of the charged atoms in the mobile region, and the total charge is maintained by doubling the charge per point. At the

beginning of each simulation, all charges start at the center of each spherical cap. (Figure 3.2b, large blue spheres), and their positions can be sampled or kept fixed in place, based on a user parameter. When a histone tail is not sampled, its total charge is effectively represented by a single point charge of equivalent magnitude. When modeling nucleosomes where the histone N-terminal tails are not resolved in the high-resolution crystal structures, the tails are modeled using the reference structure from Davey et al. [15].

3.2.3 Energy contributions

In our coarse-grained simulations we consider the elastic energy of deformation of the unbound linker DNA steps and the electrostatic interaction between charges. The calculation of the DNA deformation energy is based on the ideal B-DNA model and is described in Chapter 2 (Eqn. 2.2).

Electrostatic interactions are calculated using a Debye-Hückel potential with a Debye length of 6.85 Å, computed on the basis of the ionic strength of the experimental system discussed in Chapter 4, as well as earlier papers from our group [2, 5]. Energy contributions are determined for point charges separated by a distance less than ~80 Å. The interactions are calculated for charges within the DNA chain, between DNA and proteins, between different proteins, but not between charges within a given nucleosome, e.g., not between the tail charges and core charges of a given nucleosome. Electrostatic interactions within DNA are restricted to charges that are more than 10 base pairs away in sequence, or in other words, about a full DNA turn.

During the course of our simulations, sometimes opposite charges manage to find each other at very close distances and any subsequent sampling steps become unfavorable. This is due partially to the way we define our excluded volumes in relation to the position of core charges, but mostly is due to the treatment of histone tails as charges that preferentially lie on the outside of the core volume. On top of this, Monte Carlo methods generally suffer from this issue of "sticky" states with low energy [115]. Since exploring these states is certainly beyond our scope and means, we added a weak 6-12 Lennard Jones potential, to prevent the close approach of charges of opposite sign. We use a very low energy well (0.1 $k_{\mathrm{B}}T$ here) and a distance of 4 Å and only compute this weak interaction for attractive charges within 10 Å of one another.

3.2.4 Simulation protocol

Our simulations use a Markov chain Monte Carlo (MCMC) method based on a Metropolis-Hastings algorithm [116, 117]. In essence, every proposed simulation step adds a perturbation to the last accepted step, a commonly used approach to complicated multi-variate systems like ours. The main parameter space of the system in these simulations is the set of base-pair step parameters of the unbound DNA, that is, all the steps that are not associated with a nucleosome. The first base pair (at the 5′-end of the DNA) defines the global reference frame and the frames of all other elements are calculated from the rigid-body parameters of the DNA steps using standard linear algebra transformations. The positions and orientations of nucleosomes, are calculated from the dyad base-pair frame, which is used as a basis for converting from the local nucleosome frame to the global reference frame. At every simulation step, one DNA linker step is chosen at random and all the angular rigid-body parameters associated with that step are changed, each independently of the other. The proposed increments in the three rotational step parameters ($p_1 - p_3$ in Eqn. 2.1) are given by the product of an assumed maximum amplitude and a random number in the range [–1, +1]. The magnitude of the amplitude is based on the elastic properties of DNA, i.e., the deviation of a step parameter that raises the energy by $k_\mathrm{B}T/2$ (see Section 3.2.1). As mentioned previously, DNA is treated as inextensible by not perturbing any of the translational parameters.

After modifying a DNA step, new reference frames for all the elements downstream of the move are computed, including those of the charges and geometric shapes. After this we check for any collisions between the defined excluded volumes, as well as collisions between any volume with the mobile tail point charges (which fall outside the octamer excluded volumes). We do not check for collisions within a flexible DNA linker (since the likelihood is extremely low) or between the first three base pairs immediately entering and exiting the linker (so as not to constrain the linker movement). At this stage we can also apply any other filters (such as, for example, an end-to-end distance filter if we want to sample looped constructs), so that a proposed configuration which violates those filters can be rejected.

At this point, if the mobile charges of the histone tails are to be sampled during the simulation, a nucleosome is randomly picked with a probability that is specified in the input (p_prot). The value of this parameter is chosen so that configurations are generated with a desired acceptance ratio, here

taken to be ~0.3 following standard guidelines [115, 118]. This acceptance ratio is usually tuned based on a series of 'burn-in' runs of ~500,000-1,000,000 trials, where the acceptance is computed for increasing values of p_{prot}. Otherwise if p_{prot} is set to zero, each histone tail is effectively represented by a single fixed point charge. Finally, moves free of collisions are accepted according to the Metropolis criterion [116], after calculating the total energy of the system. For the results presented in this chapter, we performed about 4–5 million Monte Carlo (MC) trials, after the burn-in period, and collected ensembles of 5,000–8,000 configurations for each construct.

3.2.5 Nucleosomal DNA pathways

Here we have used a simple approach to characterize the mode of DNA wrapping around the histone core in experimental structures of nucleosome core particles (NCP). We have treated the DNA as a collection of base pairs, and obtained the base-pair reference frames as well as the step parameters using the 3DNA software [79, 80]. We make use of the method presented by Clauvelin, Tobias and Olson, to calculate the twist of supercoiling along the DNA pathway [119]. At the same time, we also calculate the angle between the normals of the entry and exit base pairs (Figure 3.3). By flipping the coordinate frame of the base pair near the NCP entry site, we are effectively reporting the angle between the entry and exit linkers assumed to consist of straight ideal B-DNA. The calculated twist of supercoiling, which affects the twist registry between nucleosomes in an array [100], and the angle between the entry/exit base pairs are used to characterize and compare several available high-resolution nucleosome core particles that are used as models in our simulations.

3.2.6 Characterization of chromatin fiber architecture

We report several measurements to characterize the global properties of our simulated chromatin fibers. The sedimentation coefficient, obtained through ultra-centrifugation experiments, is a commonly used measure of the global compaction of a molecule. This value is frequently estimated in computational studies of chromatin [120, 121] using the definition of Hansen et al. [122]. We use the same definition here to calculate the sedimentation coefficient S_N of an array of N nucleosomes:

$$\frac{S_N}{S_1} = 1 + \frac{R}{N} \sum_i \sum_j \frac{1}{R_{ij}} , \tag{3.1}$$

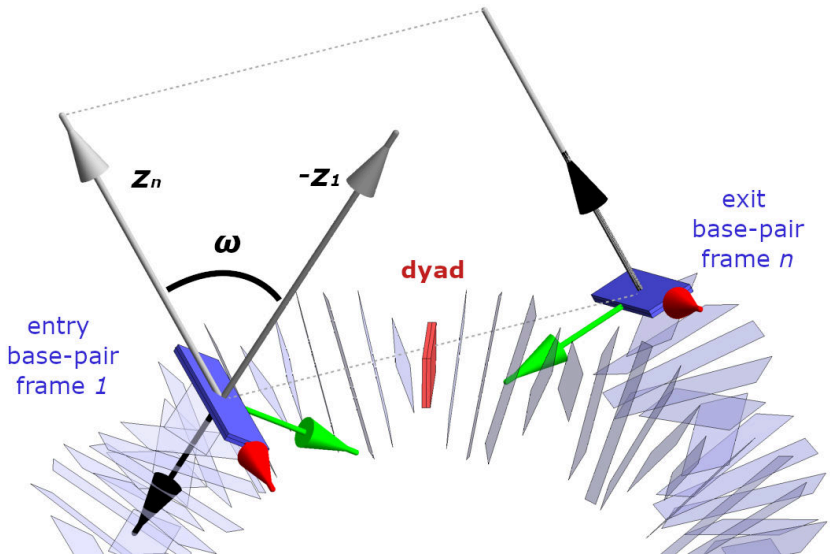


Figure 3.3: Schematic of the computed angle ω between the entry/exit base-pair normals. The entry and exit base pairs are represented by blue rectangles, while the dyad base pair is shown in red. The $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ axes of each frame are represented with green, red and black arrows, respectively. We compute the angle between the negative of the entry frame normal ($-z_1$, darker gray) and the exit frame normal (z_n, lighter gray).

where, $S_1 = 11S$ is the sedimentation coefficient reported for a typical mononucleosome [122], $R = 56$Å is the radius of a single nucleosome, and R_{ij} is the center-to-center distance between nucleosomes i and j in a given configuration.

To characterize the arrangement of nucleosomes in a given chromatin configuration, we determine the internal (virtual) bond lengths, bond angles and torsion angles relating successive nucleosomes. We then use these "backbone" parameters to calculate the local helical parameters for every four consecutive nucleosomes using Miyazawa's equations [106, 107]. We report the helical rotation, helical rise and radius of the repeating monomer unit, which is assumed to be a single nucleosome (Figure 3.4). The helical rotation angle is converted to number of nucleosomes per full 360° turn of the helix for easier interpretation of the nucleosome helical arrangement.

The analysis of compaction and fiber geometry is done for every collected configuration in the simulated ensemble for a given construct and the results are reported as means and standard deviations for each ensemble. Since the calculation of helical parameters for each step requires four monomer units, using the method from Miyazawa, we use a sliding window of four nucleosomes to determine several sets of helical parameters along a given configuration. This means our results on fiber geometry include variations within and between configurations in the ensemble for a particular

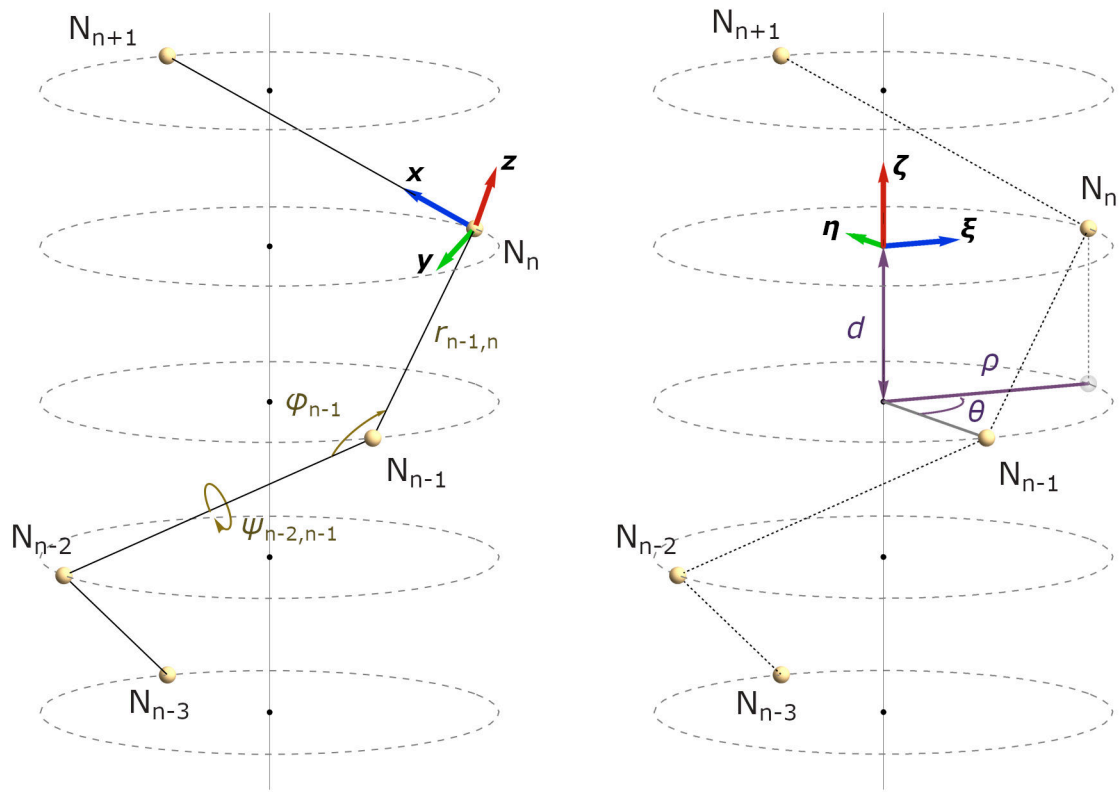


Figure 3.4: Schematic representations of the helical parameters obtained from monomer (nucleosome) positions using the method described by Sugeta and Miyazawa [107]. The vertexes (N_i) represent the reference frame origins of the nucleosomes, which follow a regular helical pathway in this example. The internal coordinates consist of virtual bond lengths ($r_{i,j}$), bond angles ($\phi_{i,j,k}$) and internal rotation (torsion) angles ($\psi_{i,j,k,l}$), obtained from the positions of the nucleosome centers. From these internal coordinates we then calculate the helical parameters (ρ, d, θ): ρ represents the radius, or distance from helical axis $\boldsymbol{\zeta}$; d represents the rise, or translation along helical axis $\boldsymbol{\zeta}$, from one nucleosome to the next; θ represents the angle of rotation around the helical axis $\boldsymbol{\zeta}$ from one nucleosome to the next. In this example, the helical rotation angle θ is 90°, which implies nucleosomes $N_i - N_{i+4}$ are aligned on top of one another along the global helical axis. Since determination of internal rotation angles (ψ) requires four points ($N_{n-3} - N_n$), and the nucleosome arrays in our ensembles do not follow regular arrangements, the local helical parameters in our analysis are calculated for every four consecutive nucleosomes, using a sliding window. Figure adapted from refs. [106, 107].

construct. On the other hand, the average three-dimensional fiber configuration used to visualize a particular construct is built from the mean values of the sampled base-pair step parameters along the DNA sequence in a simulated ensemble.

3.3 Results and discussion

3.3.1 Differential DNA wrapping in high resolution nucleosome structures

As mentioned earlier in methods, most of the work described in previous publications from our lab involves the highest resolution nucleosome structure [15] available in the Protein Data Bank (PDB) [86]. Here we also explore how the differences present in other experimental structures contribute to chromatin fiber folding, when used as NCP models in our simulations. In other words, we want to study the effect that differential DNA wrapping around the nucleosome has on the resulting nucleosomal arrays. For this reason, we have analyzed the nucleosomal DNA pathways for a large number of nucleosome core particles available in the Protein Data Bank [86]. More specifically, we calculated the twist of supercoiling [119] over the 140 base pair steps centered around the dyad of each structure (from –70 to +70), since not all high-resolution structures contain the same length of DNA. We expressed the total twist over the steps in terms of the difference from the expected value for ideal B-DNA of the same length, i.e., 140×360/10.5. We also calculated the angle between the normals of the first and last base pair frames (of the same 141-bp segment), where the normal of the first base pair was flipped. The values of these two quantities for all the analyzed nucleosomes (152 in this analysis) are shown in Figure 3.5.

The different combination of DNA entry/exit angle and relative twist are expected to have an effect on the properties of simulated nucleosome arrays. The relative nucleosome arrangements, their interactions and steric effects in these systems, significantly restrain the pathway of the linker DNA and pose different topological constraints, similar to those in a closed circular DNA [123, 124]. The two main classes of DNA sequences present in the nucleosome structures in the PDB are the human α-satellite sequence and the synthetic Widom 601 sequence [31]. The known examples include a few other derivatives and related sequences, here grouped together in a separate category labeled "other" (Figure 3.5). Our analysis suggests that, as far as the information contained the high-resolution NCP structures in the PDB is concerned, sequence alone is not a predictor for the

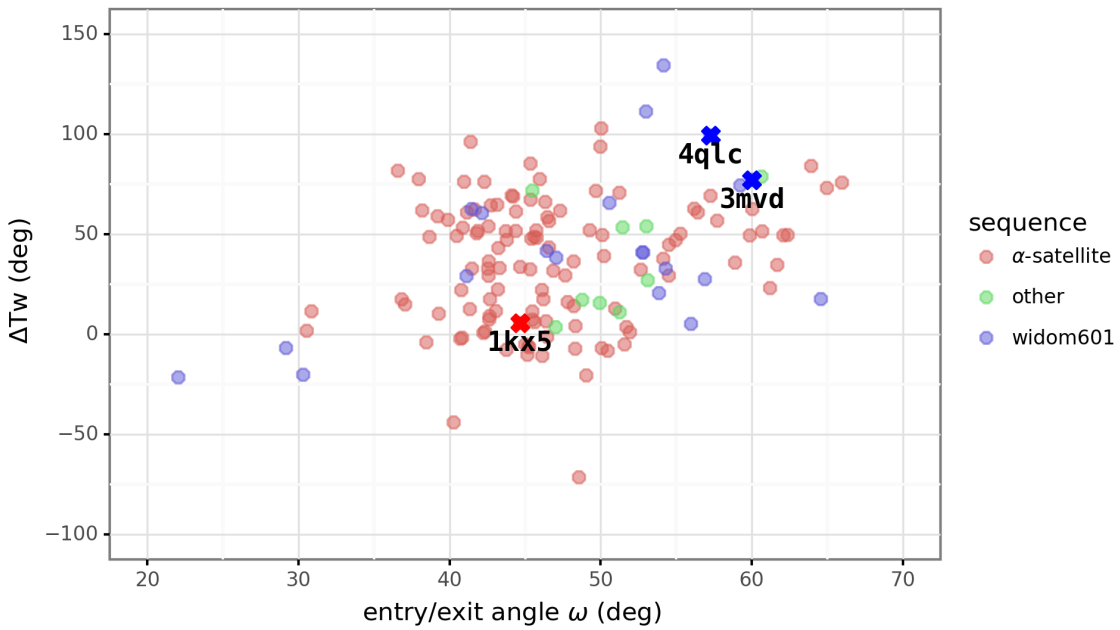


Figure 3.5: The difference in the twist of supercoiling (ΔTw) over the central 140-bp steps compared to ideal B-DNA of the same length is plotted against the DNA entry/exit angle (ω), defined by the normals of base pairs located –70 and +70 bp from the dyad. Our analysis suggests that the different sequences represented in these structures — human α-satellite, Widom's 601 and other sequences, combined in the "others" category — are not predictors for the value pair plotted on this graph. The values for the three nucleosome core particle structures used throughout the rest of this study are denoted by **x** and labeled by their PDB ID.

differential DNA binding to the histone core. This does not necessarily mean that sequence does not affect the binding mode [13, 15, 125], but rather that the differences observed in these experiments could also be attributed to other factors, such as DNA and protein modifications, binding of small molecules or other nucleosome binding proteins, all of which are represented in the PDB, as well as solvent and experimental conditions [126–129]. The combination of these influences can cause DNA to bind differently to the nucleosome and result in very different chromatin fiber structures, as our results here suggest.

In our simulations, we have focused mainly on three nucleosome structures as models for DNA binding to the NCP. Besides the human α-satellite structure already mentioned before (PDB ID: 1KX5 [15]), we have used two other models: the 146-bp Widom 601 NCP structure bound to the RCC1 chromatin factor (PDB ID: 3MVD [125]), a model used in previous computational studies of nucleosome arrays [109, 110, 121]; and the nucleosome assembled with 165 bp containing the Widom 601 positioning DNA sequence and the globular domain of linker histone H5 (PDB ID:

4QLC [46]). In Figure 3.5, these structures are represented by '**x**' markers and labeled accordingly. Our analysis shows that the degree of wrapping of the DNA around the NCP in the 3MVD and 4QLC structures differs significantly from that in the 1KX5 structure in terms of both the twist density around the histone core and the DNA entry/exit angle. The binding of the linker histone in 4QLC, is clearly expected to affect the wrapping of core nucleosomal DNA near the entry/exit sites, but it is hard to separate the effect of linker histone from other factors, such as solvent conditions and crystal packing. With respect to the uptake of supercoiling properties on the central 140 bp of core nucleosomal DNA, however, the 3MVD and 4QLC structures are similar to one another (Figure 3.5).

3.3.2 Compaction profiles of chromatin fibers based on different NCP models

Here we report the global compaction of two different 12-mer nucleosome arrays constructed based on the 1KX5 and 3MVD nucleosome models, for various nucleosome repeat lengths, ranging from 162–212 bp. For each ensemble, we report the average and standard deviation of the estimated sedimentation coefficient values. One thing that is immediately apparent from the spacing-dependent compaction profiles is that the sedimentation coefficients of the two sets of constructs display an oscillatory pattern and appear to be offset from one another by about 2–3 bp (Figure 3.6). Since the 1KX5 and 3MVD structures contain different amounts of DNA bound to the core, we considered only the central 141 bp of DNA in each case in order to make a fair comparison. Therefore the differences here are not a direct result of any extra frozen steps during the simulation, which would result in different DNA linker lengths between the two models for each nucleosome spacing.

Unsurprisingly, the compaction profiles display a periodic character with a repeat of 10–11 bp, which corresponds roughly to the 10.5 bp helical pitch of the ideal DNA model used here. Besides the offset of peaks between the two constructs, the 3MVD-based fibers generally display a slightly lower compaction, except for very short repeat lengths (<168 bp), which implies a somewhat different response to the increasing length of the nucleosome repeat sequence between the two models. A study by Correll et al. [58] used electron microscopy imaging and ultracentrifugation to look at the effect of nucleosome spacing on compaction of 12-mer nucleosome arrays and found that this effect is stronger for short repeat lengths than for longer ones. Their sedimentation coefficient data in 1 mM $MgCl_2$ solution are shown in Figure 3.6 as scatter points. Our results agree with their finding that chromatin fibers with longer repeat lengths are less sensitive to the change in nucleosome spacing,

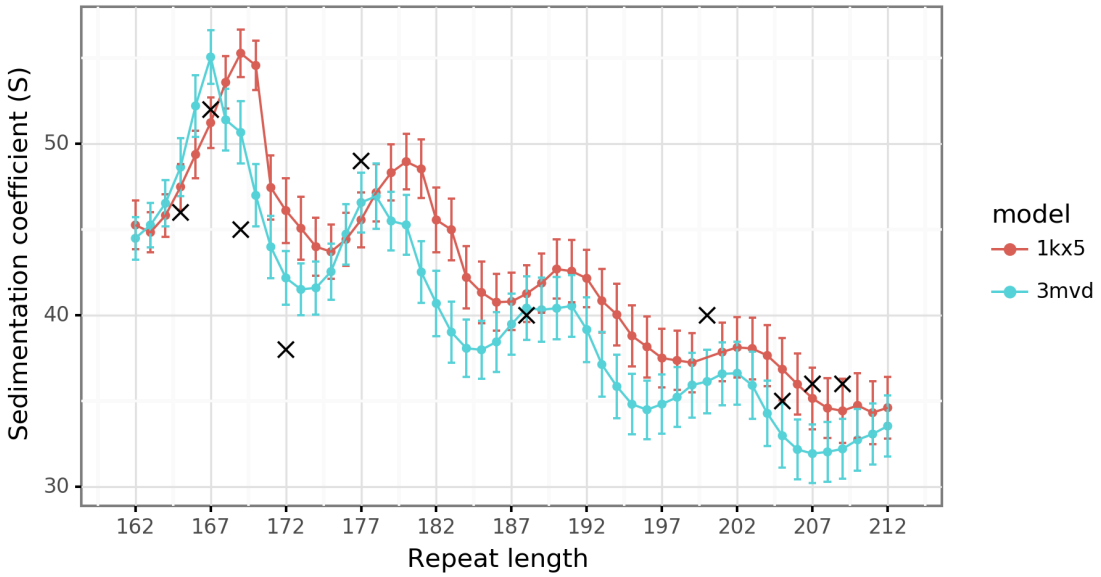


Figure 3.6: Sedimentation coefficients for 12-mer nucleosome arrays built from different NCP models containing the central 141 bp of DNA — 1KX5 (red) and 3MVD (blue). The bars represent the standard deviation of the predicted values for each repeat length. The scatter points with an 'X' marker represent EM measurements of 12-mer arrays performed in 1 mM $MgCl_2$ solution from Correll et al. [58]. Our results show that the predictions from simulated 3MVD-based fibers fit the experimental data better, at least in terms of the pattern of the apparent minima and maxima from the EM study.

compared to chains with shorter repeats. Additionally our estimated sedimentation coefficient values for 3MVD-based fibers are in closer agreement with the data from Correll et al. compared to those from 1KX5-based fibers. These differences in predicted compaction between our fiber models — both the spacing-dependent nature as well as the differences in magnitude — are entirely due to the differences in DNA wrapping in the respective nucleosome core particle structures. The immediate implication of these results is that the different mode of nucleosomal DNA wrapping has an effect on fiber compaction comparable to a change in nucleosome spacing of 2–3 base pairs. Depending on the location along the repeat length parameter space, that 2–3 base-pair shift could result in dramatic changes in nucleosome arrangement (Figure 3.7).

3.3.3 Characterization of chromatin fiber architectures

In order to understand the reason for the differences in global compaction, we analyzed the geometries of the simulated fibers for each NCP model. Our simulations generally predict a zig-zag model where

nucleosomes interact in a non-consecutive manner, i.e., the closest nucleosomes are often second or third (or higher) neighbors in sequence. The ensembles represent a flexible and dynamic fiber adopting a wide range of arrangements that vary with sequence length. In our work, we have chosen to describe the fiber geometry in terms of local helical parameters. The helical rotation angle is converted to number of nucleosomes per turn and the distribution is binned into half integer steps. Figure 3.7 shows the distribution of helical content for both 1KX5-based and 3MVD-based simulated chromatin fibers, while Figure 3.8 shows the helical rise.

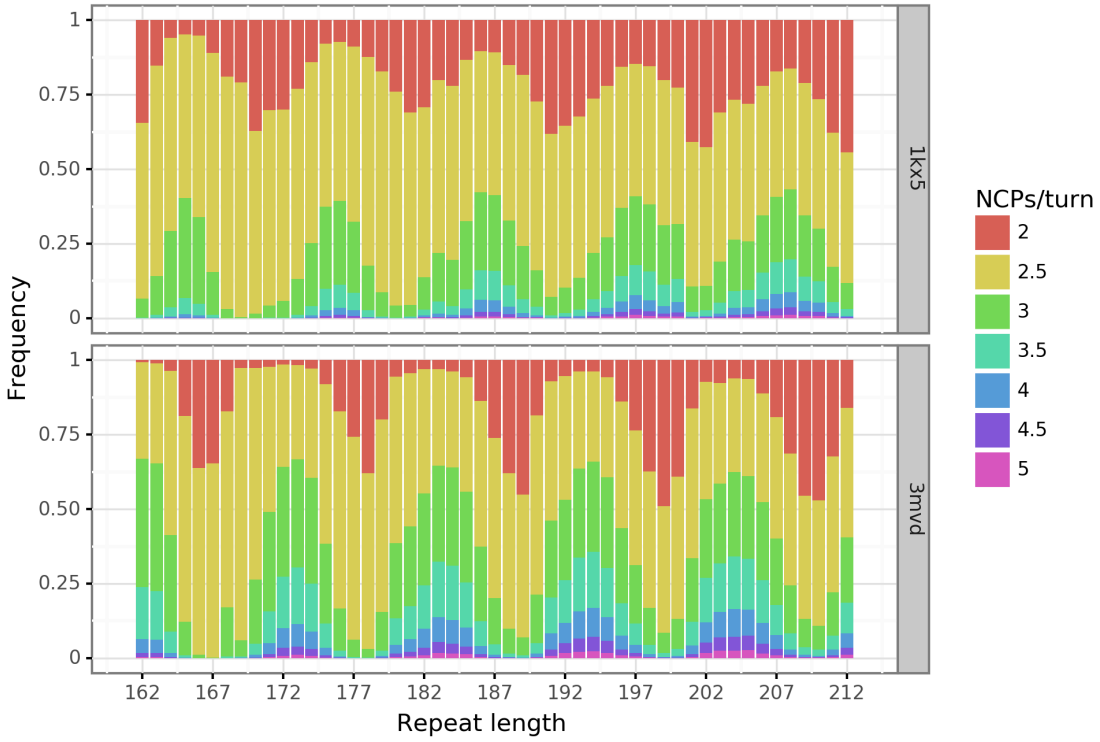


Figure 3.7: Distributions of helical arrangement of the chromatin fibers as a function of nucleosome repeat length, for 12-mer arrays built from two different NCP models with 141 bp of fixed DNA. The helical rotation angle is converted to the number of nucleosomes per turn and the values are binned in half integer steps; for example, 2.5 (yellow) shows frequencies for values in the 2.25–2.75 NCPs/turn range, 3 (green) shows the frequencies for the 2.75–3.25 NCPs/turn range and so on. The first bin represents the 2–2.25 range, since the value cannot be lower than 2 in our helical treatment. For a particular repeat length, the stacked bars show the normalized frequency of each helix type (in terms of nucleosomes per turn) in the ensemble.

The helical rotation angles, as well as the helical rise, display an oscillatory pattern consistent with the repeat of ideal DNA used to model the linkers. With nucleosome repeat length, the fibers adopt different architectures, primarily consisting of a mix of 2-start, 3-start and 5-start (2.5 NCPs/turn

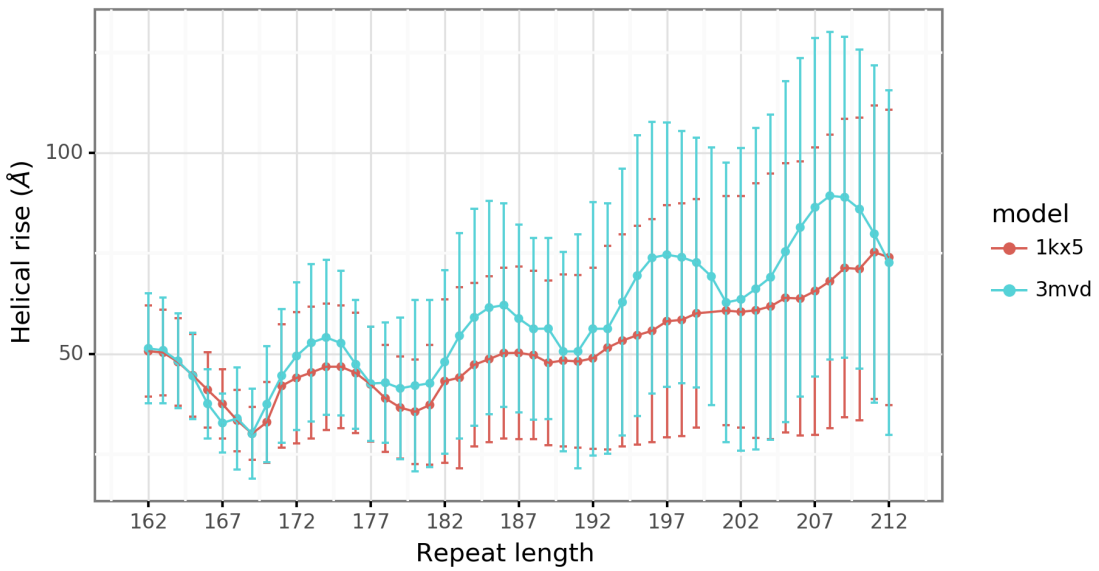


Figure 3.8: Distributions of helical rise for 12-mer nucleosome arrays built from two different NCP models containing 141 bp of DNA. These values represent the vertical displacement along the local central axis from one nucleosome to the next. The bars represent the standard deviation of the predicted values in the ensemble for each repeat length.

or 5 nucleosomes per two turns) models (Figure 3.7). With regard to the notation describing these arrangements, the nomenclature is a little confusing. The helical repeat in a geometric sense, generally defined by how many units are required to complete a full turn of the helix, does not always correspond to the pattern of nucleosome stacking interactions, often used in the literature to describe chromatin architecture. This point is illustrated in Figure 3.9, where the average configuration (based on average step parameters, see Section 3.2.6) for a 167 bp repeat fiber is colored in three different ways to emphasize the arrangement of the second, third and fifth nucleosome. Here we will try to make the distinction whenever it applies, whether we are referring to the helical repeat number or the apparent nucleosome stacking. For example, for a 177 bp repeat length, the fiber based on the 1KX5 NCP model consists of mostly three twisted nucleosome stacks in a loose state, on average, while the 3MVD-based fiber consists primarily of two such stacks in a much tighter arrangement (Figure 3.10), even though they primarily consist of 5-start helices (2.5 NCPs/turn), containing 5 nucleosomes over two turns on average, according to the distributions shown in Figure 3.7.

Figure 3.10 shows the predicted average configurations of simulated fibers with three different nucleosome repeat lengths constructed from the two different NCP models discussed so far. Besides the fact that 3MVD-based fibers look different from 1KX5-based ones with the same nucleosome

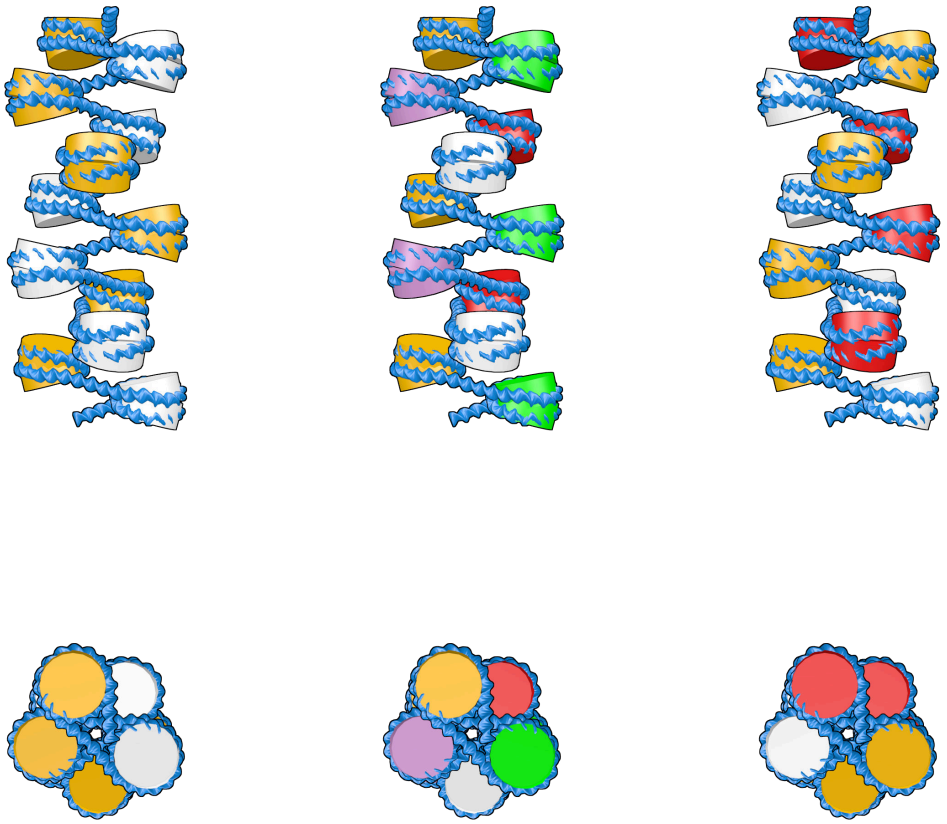
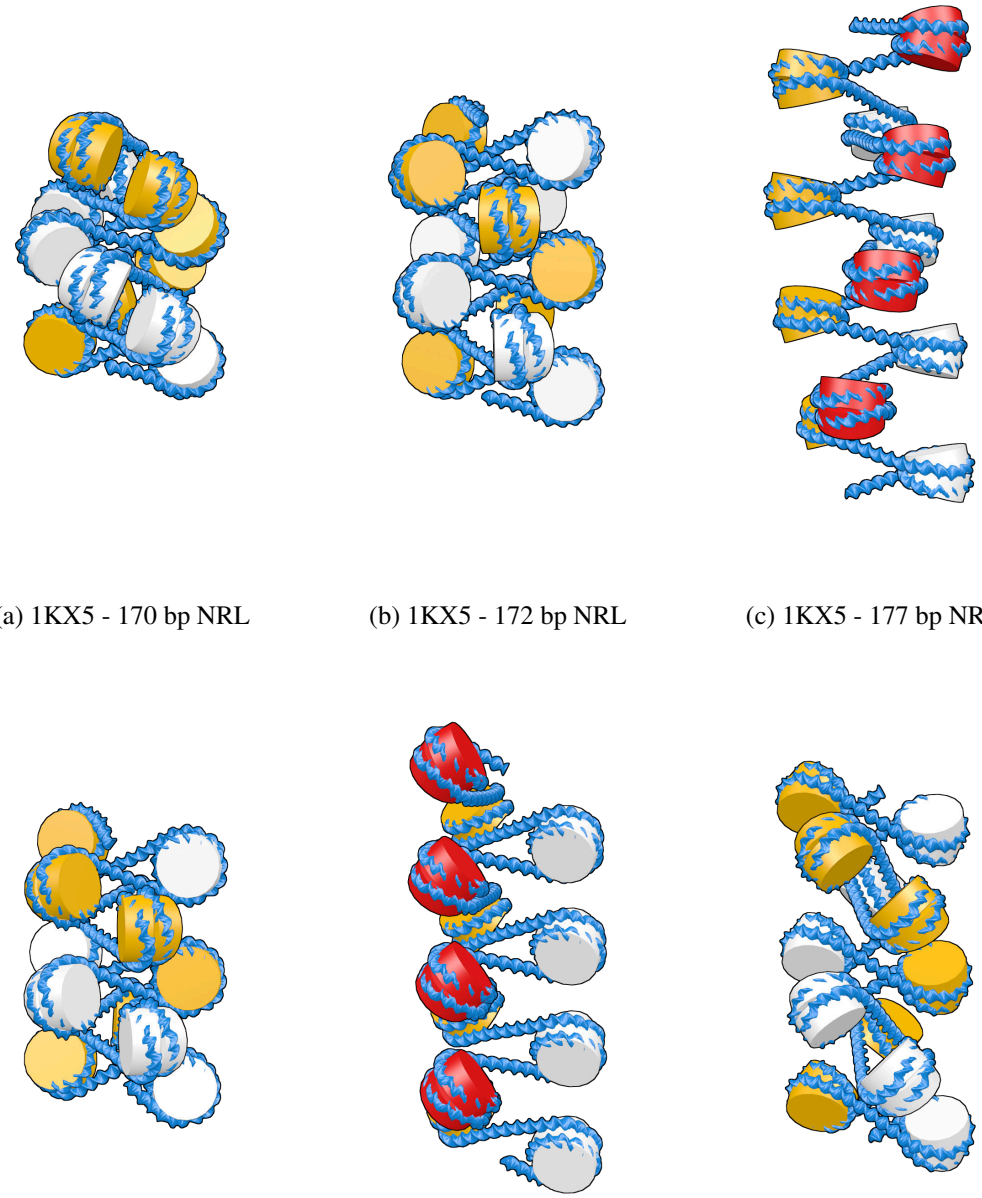


Figure 3.9: Average configuration, based on mean base-pair step parameters over the simulated ensemble, for a 12-mer nucleosome with a 167-bp repeat length built with the 1KX5 nucleosome model. The average structure is aligned along its helical axis and viewed from the front and top. From left to right, we have used the same color for every second, fifth and third nucleosome, respectively to illustrate the different potential notations as a 2-start, 5-start (5 NCPs per 2 turns) and 3-start fiber. In this case, the 5-start coloring in the middle is the most accurate depiction of the helical repeat, in a geometric sense, but does not consider the most likely nucleosome interactions.



(a) 1KX5 - 170 bp NRL (b) 1KX5 - 172 bp NRL (c) 1KX5 - 177 bp NRL

(d) 3MVD - 170 bp NRL (e) 3MVD - 172 bp NRL (f) 3MVD - 177 bp NRL

Figure 3.10: Average configurations, obtained as described in Figure 3.9, for 12-mer nucleosome arrays built from two different NCP models: 1KX5 (a-c) and 3MCD (d-f), and various nucleosome repeat lengths. The 3D arrangements of the nucleosomes are color coded to highlight the different nucleosome "stacking" arrangements as a result of the choice of NCP model and repeat length. For example, for constructs with a 177 bp repeat, the nucleosomes in 1KX5-based fibers are oriented with their cylindrical axis parallel to the fiber axis, while in 3MVD-based fibers they are oriented at an angle. Note the near-identical organization between the 1KX5-172bp and 3MVD-170bp models. In the constructs depicted here, both NCP models contain 145 bp of DNA fixed to the core, so the differences are not due to variations in spacing between the two models.

spacing, it is clear that the 170-bp repeat 3MVD-based construct looks similar to the 1KX5-based model with a 172-bp repeat. This illustrates the larger point observed in Figure 3.7, that the two different construct may adopt generally similar architectures but at different points in the nucleosome repeat length parameter space. These results suggest that some of the discrepancies between several proposed analytical and *in silico* models in the literature can boil down to the choice of the NCP model used. For constructs containing the same repeat length of 177 bp, the nucleosomes are arranged with their cylindrical axis mostly parallel to the fiber axis in the case of 1KX5-based fibers, while the two axes are at an angle in the case of 3MVD-based fibers. This nucleosome inclination angle also displays an oscillatory pattern (not shown here) and together with the helical rotation dictates how nucleosomes are stacked.

Besides the offset in the repeating patterns, the two fibers considered here contain other differences in their response to the change of repeat length. For example, 1KX5-based fibers consist of mostly 2.5 nucleosomes/turn and an even mix of 2 and 3 nucleosomes/turn, while the 3MVD-based fibers appear to be more heterogeneous and more sensitive to the repeat length (Figure 3.7). Looking at the helical rise (Figure 3.8), we can also see that while the values generally trend upwards in both cases, the 3MVD-based fibers display a stronger oscillatory pattern that is present throughout the repeat lengths. These periodic increases in the values of rise are the result of the changes in the angle that DNA linkers adopt with respect to the fiber axis (values not shown here) and can result in a degree of inter-digitation between nucleosomes [130]. The predicted flexibility in helical rise also differs between the two fiber types, while generally getting larger with increasing DNA linker length. These subtle differences reflect the effect that the mode of DNA wrapping in each NCP model presents to the fiber arrangement as an additional constraint. Our results are in general agreement with analytical models proposed by several groups [97, 99, 131], while differing in terms of the phasing of the oscillatory patterns.

Robinson and co-workers used electron microscopy to measure the diameter of different fibers and suggested that the so-called 30-nm fiber can adopt two different structural families characterized by an inter-digitated one-start model [57]. Their diameter measurements are shown in Figure 3.11 as scatter points, along with our predicted diameter values from the simulated fibers. Our results suggest that the diameter of the nucleosome arrays increases with the repeat length, as expected, but in a non-linear fashion [97, 99]. This is due to the combined effect of increased linker length,

changes in fiber helical arrangement and changes in the angle between the DNA linkers and the helical axis (when DNA linkers are perpendicular to the helical axis, the diameter is maximized). While our predictions capture some of the experimental data, they are generally not in agreement and do not support a solenoid fiber.

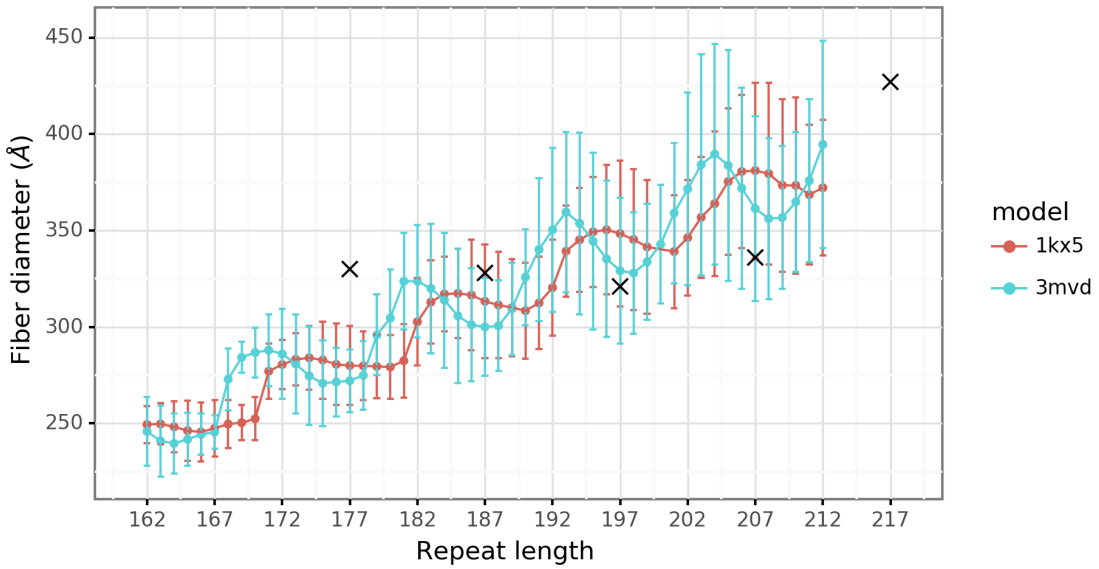


Figure 3.11: Fiber diameters for 12-mer nucleosome arrays built from two different NCP models containing 141 bp of DNA — 1KX5 (red) and 3MVD (blue) — calculated as the sum of twice the helical radius (axis to NCP center) plus the diameter of a typical nucleosome core particle (~110 Å), using the characterization of the chromatin fiber as a helical structure. The bars represent the standard deviation for each ensemble of simulated constructs. The scatter points with an 'X' marker represent EM fiber diameter measurements by Robinson et al. [57].

3.3.4 Role of partial DNA unwrapping in global properties

In the data shown so far, we have treated only the central 141 base pairs (140 bp steps) of the nucleosomal DNA as frozen, while allowing the rest to flex. To study the effect of partial nucleosome unwrapping, we repeated the experiments of the dependence on nucleosome repeat length with 145 fixed base pairs of DNA for each NCP model. Overall the resulting fibers do not undergo dramatic changes, and the pattern of NRL dependence of compaction is maintained for both types of fibers. This again confirms that the differences between the 1KX5- and 3MVD-based chromatin fiber models cannot be attributed simply to the length of flexible DNA, nor the wrapping of the DNA regions near the nucleosome entry and exit sites. By looking at the more subtle differences, we can

see that the effect of unwrapping two base pairs of DNA from either side of the 1KX5 and 3MVD nucleosomes has different effects on the resulting fibers (Figure 3.12). In the case of 1KX5-based fibers it causes a slight shift in the phase of the periodic pattern, where for certain repeat lengths it results in a more compact fiber and in other cases a less compact one. In the case of 3MVD-based fibers, this total of 4-bp unwrapping causes a decrease in compaction across the entire repeat length space. These differences are encoded in the wrapping of the DNA around each nucleosome core.

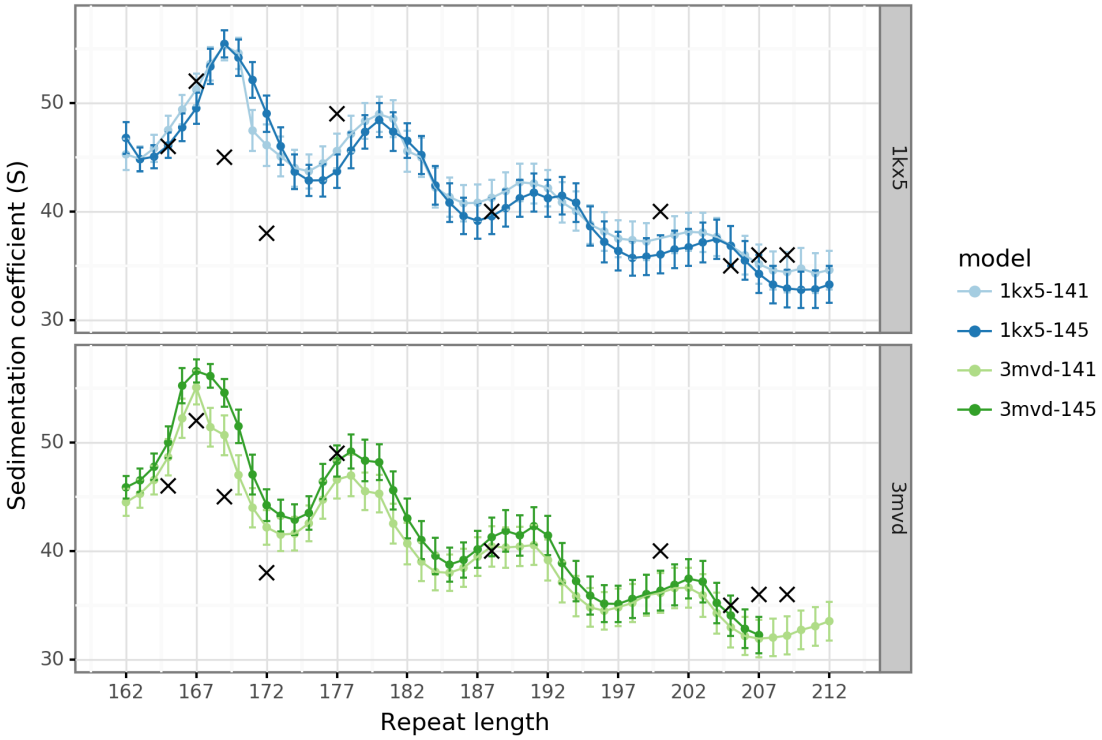


Figure 3.12: Sedimentation coefficients for 12-mer nucleosome arrays built from two different NCP models: 1KX5 (red) and 3MVD (blue), containing 141 and 145 fixed base pairs of DNA for each model. The bars represent the standard deviation of the predicted values for each repeat length. The scatter points with an 'X' marker represent EM measurements of 12-mer arrays performed in 1 mM $MgCl_2$ solution from Correll et al. [58], as in Figure 3.6. This figure shows that allowing perturbation on 2 extra base pairs of DNA from either side of the NCP model during the simulation has a different effect on 3MVD-based constructs compared to 1KX5-based constructs.

To look closely at how this partial unwrapping affects the nucleosome arrangements, we focused on 12-mer arrays of 177-bp nucleosome repeats with different lengths of DNA fixed onto the histone core. As we can see from Figure 3.13, the 1KX5-based fibers display a decrease in overall compaction going from 141 to 143 fixed base pairs. At the same time, with the increase in length of fixed nucleosomal DNA, the 3-start (or higher) helix content increases and reaches a maximum

at 147 bp, the full length of the DNA chain in the crystal structure. The 3MVD-based fibers on the other hand reach minimum compaction at 143 fixed base pairs, which corresponds to maximum 2-start content and minimum 3-start content. These differences reflect the specific binding patterns of DNA to the histone core observed in the crystal structures and illustrate another potential factor that can affect the structural arrangement of the chromatin fibers.

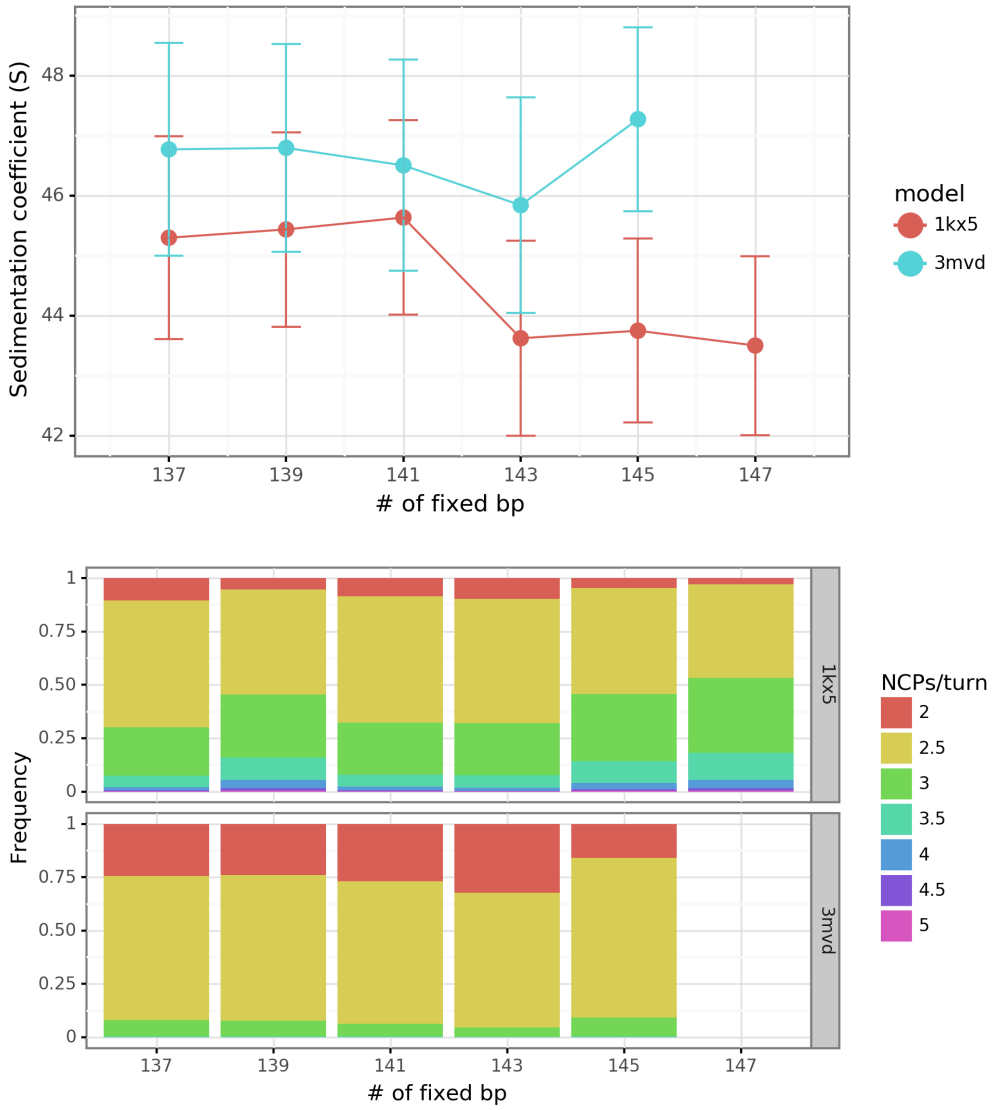


Figure 3.13: (Top) Predicted sedimentation coefficient for 12-mer nucleosome arrays with various levels of partial DNA unwrapping for two different nucleosome models, based on the 1KX5 and 3MVD crystal structures. The bars represent standard deviation of the calculated values. (Bottom) Predicted distribution of helical arrangements for the same constructs.

3.3.5 Insights into role of linker histone protein on chromatin fiber organization

In order to see the effect that linker histone binding has on the arrangement of nucleosomes in an array, we simulated constructs based on the 4QLC nucleosome crystal structure, which contains 165 base pairs of DNA in the presence of the linker histone H5 globular domain (gH5) [46]. The immediate result we see is a dramatic compaction, i.e., increase in sedimentation coefficient, caused by fixing increasing amounts of linker DNA along the pathways described in the crystal structure (Figure 3.14). To demonstrate that the effect is not purely due to the loss of degrees of freedom in the DNA linkers, we set up simulations with the 1KX5 model containing extra fixed B-DNA steps at both entry/exit sites. This results in a loss of flexibility in the simulated fibers, as expected, but the compaction generally remains unchanged. Both sets of simulations were performed on constructs with 177-bp repeats.

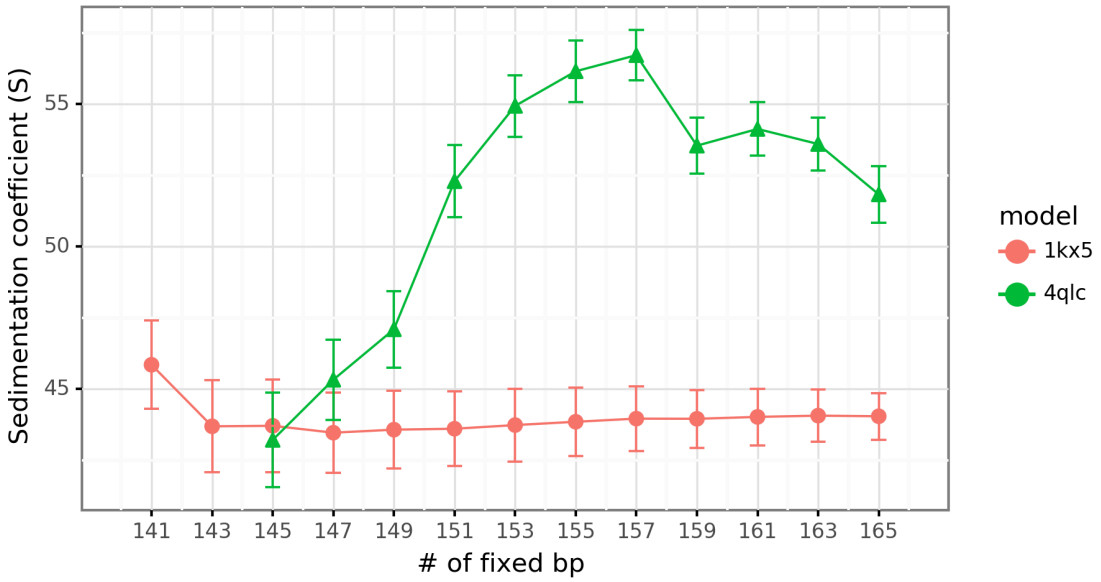


Figure 3.14: Predicted sedimentation coefficient for 12×177-bp nucleosome arrays with various lengths of constrained entry/exit linker DNA. The base pairs of DNA are frozen according to their pathways from the experimental structure in the case of 4QLC-based fibers and according to ideal B-DNA values in the case of 1KX5-based fibers (beyond the 147 bp present in the 1KX5 structure). The graph shows that the extra compaction is not due simply to the reduced number of sampled DNA base pairs, but due to the specific folding of DNA in the case of a nucleosome in the presence of linker histone H1.

The dramatic changes in compaction in the 4QLC-based fibers are a direct consequence of the extreme twist and bend that is accumulated in the DNA linkers due to the binding of the linker

histone, as observed in the crystal structure. By contrast, in the case of 1KX5, where the added DNA is fixed in an ideal relaxed state, the result of wrapping extra base pairs is negligible. While at 145 base pairs the 4QLC-based constructs display a strong 2-start fiber, for increasing amounts of fixed DNA they are mostly arranged in a 5-start (2.5 NCPs/turn) manner (Figure 3.15). The less compact structures appear to be associated with increased heterogeneity and increased 3-start helical content. At the same time, the 1KX5-based fibers remain mostly unchanged with increased amounts of fixed DNA linkers. The increased content of 3-start helical structure appears to be purely the effect of the loss of flexibility.

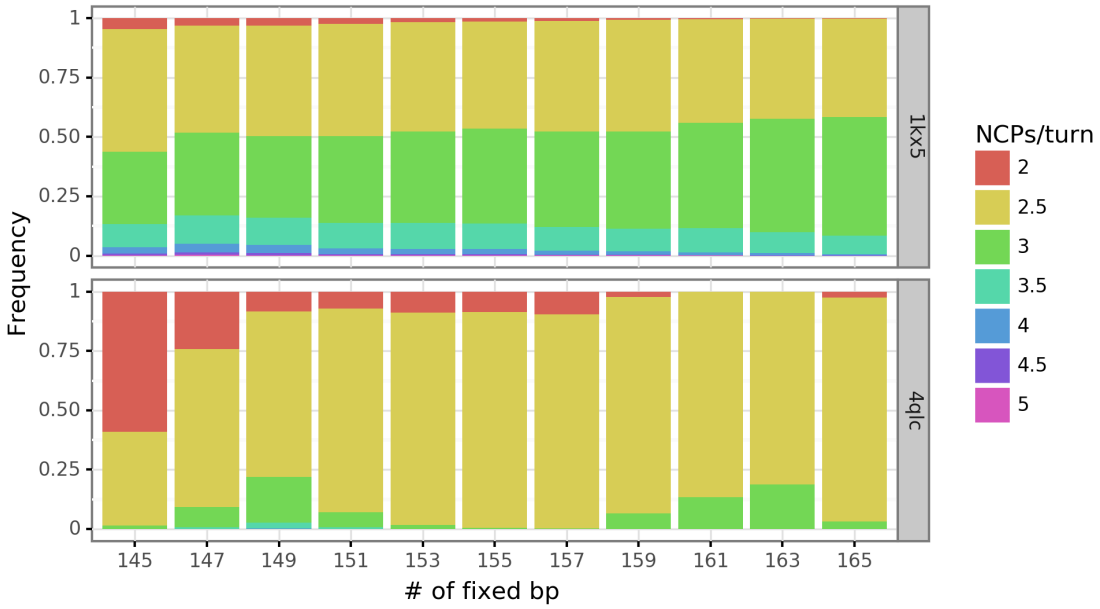
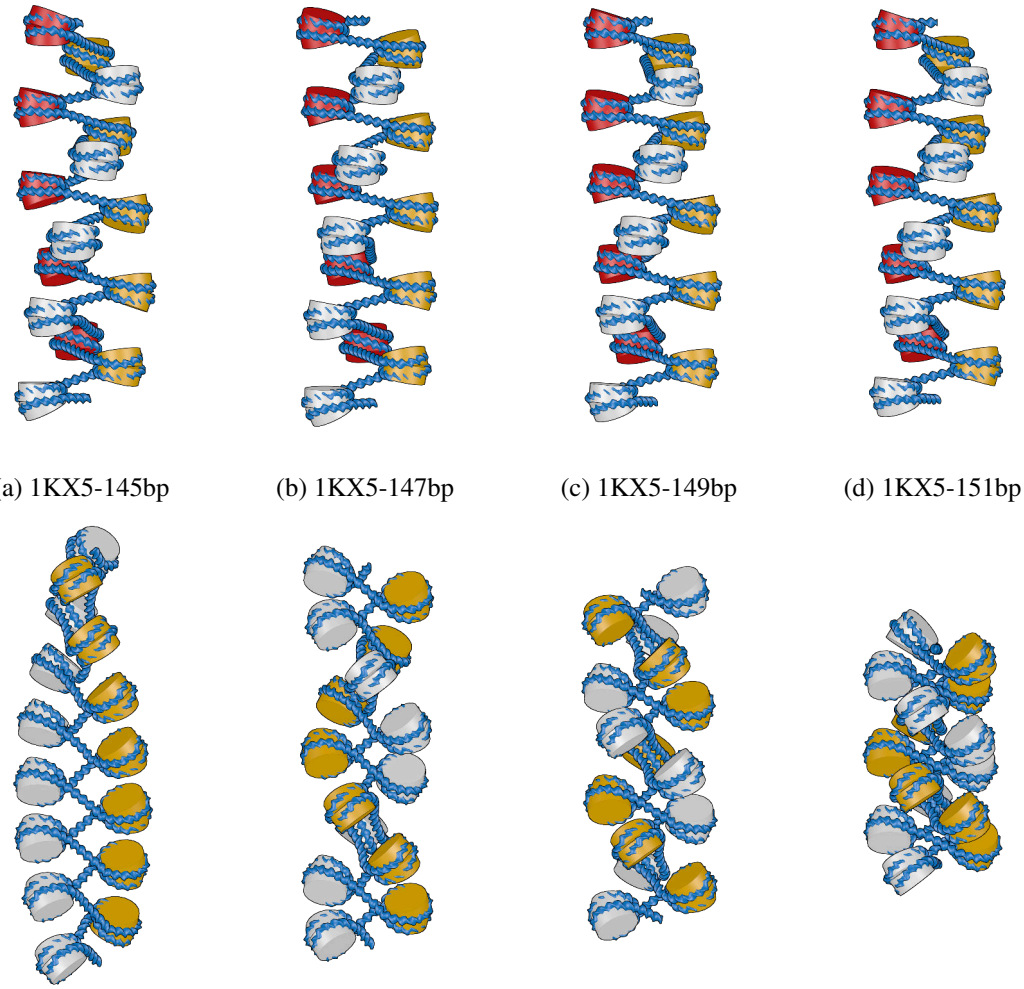


Figure 3.15: Predicted distribution of helical configurations for 12×177-bp nucleosome arrays with various lengths of constrained entry/exit linker DNA. The constructs are the same as those described in Figure 3.14. The figure shows that the arrangement of the DNA entry/exit linkers in the presence of linker histone H1 causes the resulting fiber to adopt a different architecture.

The gH5-containing nucleosome structure used here (4QLC, [46]) was solved in a mono-nucleosome environment and in the absence of the linker histone N- and C-terminal domains, which are not necessarily representative conditions for living cells. However, these simulations illustrate the impact that the linker histone can have on linker DNA and the chromatin structure as a result. Figure 3.16 shows the average configurations of the simulated fibers with 145-151 fixed base pairs of nucleosomal DNA: while the differences in 4QLC-based structures are significant, the 1KX5-based fibers appear unchanged.



(a) 1KX5-145bp (b) 1KX5-147bp (c) 1KX5-149bp (d) 1KX5-151bp

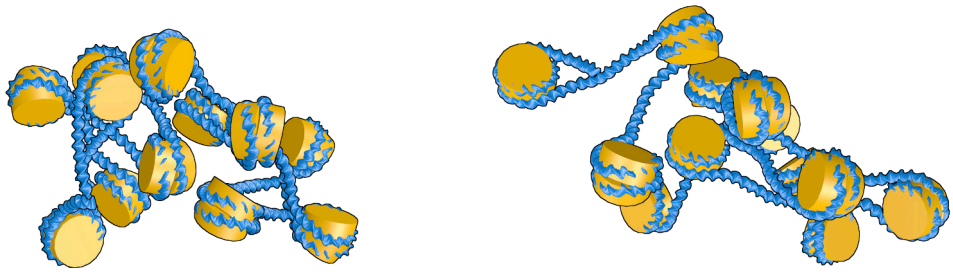
(e) 4QLC-145bp (f) 4QLC-147bp (g) 4QLC-149bp (h) 4QLC-151bp

Figure 3.16: Average configurations for 15×177-bp nucleosome arrays containing increasing amounts of frozen nucleosomal DNA based on the 1KX5 (top row) and 4QLC (bottom row) structures. While the linker histone H5 globular domain-containing 4QLC structure has 165 bp of DNA, 1KX5 contains only 147, therefore the extra frozen base pairs in this case represent rest state ideal B-DNA. These models illustrate that the dramatic effect of the linker histone binding is not due simply to the loss of linker DNA flexibility.

Discussion

Our simulations allows us to describe chromatin as a dynamic ensemble of nucleosome arrangements. For any particular nucleosome repeat length (NRL), the ensemble consists of a heterogeneous mixture of fiber geometries, consistent with previous work on preferred helical arrangements [100]. Unlike studies that focus on the preferences of regular helices, however, our Monte Carlo simulations allow for flexibility within a fiber. In fact, our simulated ensembles consist of irregular fibers that often

fluctuate around an average helical arrangement. In some cases, particularly for fibers with long NRL, the fibers appear completely irregular (Figure 3.17). They contain sharp turns and self-contacts, without any resemblance to the regular average helical structures, which are built based on mean base-pair step parameters.



(a) 12×192-bp array, 1KX5 (b) 12×203-bp array, 3MVD

Figure 3.17: Examples of representative configurations for 12mer nucleosome arrays during the course of the simulation for two different fiber constructs. (a) 12×192-bp array based on 1KX5 structure. (b) 12×203-bp array based on 3MVD structure. The nucleosome models in both cases contain 141 base pairs of DNA fixed to the core. The randomly chosen configurations seem to lack any regular arrangement.

Our simulations predict that fibers with higher content of a 3-start (or higher) helix arrangement tend to be less compact, while those with a higher content of 2-start helix arrangement are more compact. This is in agreement with recent genome-wide observations *in situ* which suggest that fibers with 3-start (or multi-start) structure are associated with more open chromatin, while 2-start structures are more strongly associated with compact chromatin [104]. For constructs with longer repeat lengths, the predicted fibers are very heterogeneous. For shorter repeat lengths, our simulated ensembles consist of approximately 2.5 nucleosomes per turn, on average, with varying amounts of 2-start and 3-start (and multi-start) helices, sensitive to NRL. The range of nucleosome arrangements observed in our simulations is in general agreement with previous predictions from analytical treatments of chromatin fiber geometry [61, 97, 99–101].

Our simulations show that the nature of the sensitivity of fiber geometry to the repeat length, however, is dependent on the choice of nucleosome core particle model. The repeat length for which the 1KX5-based nucleosome arrays achieve minimum compaction and an open 3-start fiber, is offset by about three base pairs, compared to 3MVD-based arrays. For example 1KX5-based arrays with 176-bp NRL and 3MVD-based arrays with 173-bp NRL display similar characteristics. Furthermore,

the results presented in Figure 3.7 indicate that fiber geometry in 3MVD-based constructs with 141 bp fixed nucleosomal DNA may be more sensitive to changes in repeat length, compared to 1KX5-based fibers.

In remains unclear what determines the differential DNA wrapping in the nucleosome crystal structures that leads to the different fiber arrangements in our simulations. It is also unclear whether the differences seen in high-resolution crystal structures are relevant in long chromatin fibers under physiological conditions, where additional constraints are placed on the DNA entering and exiting each nucleosome. In the case of the structures used in our work, the 1KX5 structure contains 147 base pairs of a human α-satellite sequence [15], while the 3MVD structure contains 145 base pairs of the Widom 601 sequence and is bound to a chromatin factor protein [125]. With such disparate systems, it becomes difficult to understand the underlying mechanisms for the observed differences in DNA wrapping. The characters and concentrations of solvent ions also affect DNA structure and binding. Monovalent ions such as K^+ or Na^+ have been shown to have a different effect on DNA, with condensation of sodium ions preferred over potassium ions and the former allowing a closer proximity of two interacting DNA strands [126, 127]. At the same time the presence of divalent ions such as Mg^{2+} and Mn^{2+}, preferentially stabilizes right-handed crossovers of two DNA helices over left-handed crossovers [132]. Explicit modeling of these different ions is not possible in our coarse-grained treatment. However, the implications of all these different factors on chromatin structure remain to be assessed in the future by development of knowledge-based potentials, based on analysis of the local effects from high-resolution structures.

In this work, we have focused on characterization of nucleosome arrays with regular spacing, in order to understand the basic rules that govern chromatin folding. However, a heterogeneous perspective of chromatin as a dynamic ensemble of states is necessary to capture the recent emerging data from *in situ* and *in vivo* experiments. Furthermore nucleosomes *in vivo* are not evenly spaced along the DNA sequence [133, 134], and our computational model allows the study of such systems. Our meso-scale model is quite simple by design, and one of the main benefits of this approach is the ability to run simulations relatively fast and interpret them easily. At the same time it has enough flexibility built-in to capture the large-scale effects of several factors that have been experimentally shown to dictate the organization of chromatin fibers. Unlike other approaches that use Monte Carlo simulations [98, 108], we can investigate changes in nucleosome spacing with single base-

pair precision [105]. We also do not force an interaction through an explicit potential between nucleosomes [109, 110], which favors nucleosome stacking. Our nucleosome interactions arise due to electrostatic interactions and topological constraints imposed by DNA linkers and nucleosomal DNA wrapping.

Our choice for characterization of fiber structure in terms of helical parameters, insensitive to NCP-NCP distance or contacts, allows for an easier interpretation of fiber geometry and its heterogeneity. It also allows us to interpret the implications of chromatin structure on global properties, such as long-range interactions, which are discussed in Chapter 4. However, this treatment rests upon the assumption of a mainly regular helix and one needs to be careful when interpreting the data for highly irregular fibers in terms of ensemble averages, that might not necessarily represent any particular configuration.

3.4 Conclusions

Our simulations suggest that chromatin is an irregular and dynamic structure, consisting of a mixture of local nucleosome arrangements. The overall architecture and flexibility depend on nucleosome repeat length. Our meso-scale model predicts that the differential folding of DNA around the nucleosome cores, as observed in high-resolution crystal structures, can lead to different fiber geometries. These fiber geometries are also sensitive to partial DNA unwrapping, in a manner that is dependent on the nucleosomal DNA folding. Our simulations illustrate the dramatic effect that linker histone binding can assert on chromatin fiber behavior, through the local modulation of DNA linker arrangements entering and exiting the nucleosome.

Chapter 4

Long-range enhancer-promoter communication in chromatin

4.1 Introduction

Complex biological processes often involve multiple regulatory mechanisms and require distant elements on the DNA sequence to come into close proximity. Such is the case with enhancers that regulate transcription in eukaryotes. Enhancers are regulatory DNA sequences that recruit transcription factors and activate transcription at promoter sites, often thousands of bases away in the sequence [6, 69]. This distant activation *in cis*, that is, with both interacting elements on the same chain, requires physical interaction between the regulatory elements at the enhancer and promoter sites and is accompanied by the formation of a large loop. While chromatin has been shown to facilitate enhancer-promoter communication (EPC) at a distance [135], the structural details involving chromatin looping have not been fully elucidated. We want to understand the structural mechanism of this looping and how chromatin fiber architecture affects long-range communication in chromatin.

An *in vitro* quantitative assay has been developed by Vasily Studitsky's research group to measure the distant communication rates between an enhancer and its target promoter *in cis* [6, 136, 137]. These experiments are set up in such a manner that communication between the regulatory sites is the rate-limiting step in the transcription reaction and the transcription yield is directly proportional

to the communication rate (Figure 4.1). The experimental results from the *in vitro* studies can be compared to computational predictions to gain insight into the mechanism of enhancer-promoter communication in chromatin.

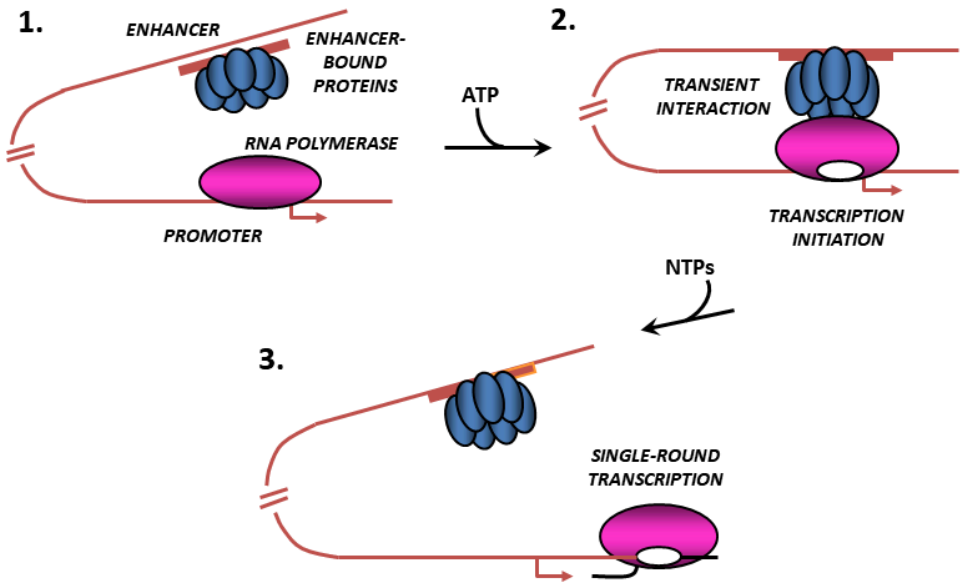


Figure 4.1: Schematic of experimental setup for *in vitro* EPC analysis. 1. RNA polymerase is initially bound at the promoter but cannot initiate transcription, while the NtrC activator protein is bound to the enhancer in an inactive state. 2. The enhancer interacts with the RNA polymerase after fast phosphorylation in the presence of ATP causing looping of the intervening DNA and initiation of transcription. 3. After addition of nucleoside triphosphate (NTP) monomers, the RNA polymerase completes transcription and the yield of RNA transcript is measured. The yield is directly proportional to the enhancer-promoter communication (EPC) rate. Figure adapted from ref. [6] with permission from Future Medicine.

Using this combined approach of *in vitro* studies and simulations of precisely positioned nucleosome arrays, it has been shown that chromatin enhances the efficiency of enhancer-promoter communication compared to naked DNA [135]. While the N-terminal tails increase the enhancement effect, the presence of nucleosome gaps can either enhance or diminish EPC efficiency, depending upon the length of the system [1, 2, 5]. At the same time, it has been shown that distant communication is also affected by nucleosome spacing, with increasingly longer linkers resulting in more efficient EPC for regular arrays with 172, 177 and 207 base-pair repeat lengths [5]. What remains unclear is the precise role of the chromatin fiber architecture in modulating long-range communication, given that the arrangement of nucleosomes is sensitive to the nucleosome repeat length (NRL).

At the same time, while the N-terminal histone tails are essential for efficient for EPC over long distances [2], the role of each tail in this process has not been reported. In this Chapter we aim to address both of these questions from a computational perspective and to compare our predictions against experimental data from our collaborators.

Due to the size of the systems involved, current computational resources preclude atomic simulations of nucleosome arrays. We use the meso-scale model of chromatin, which was discussed in Chapter 3, to simulate regular nucleosome arrays with an activator protein and RNA polymerase holoenzyme bound at enhancer (E) and promoter (P) sites, respectively. We then measure the enhancer-promoter communication in terms of the likelihood of E-P site juxtaposition.

Our computational model suggests that under physiologically-relevant conditions, such as those used in the *in vitro* experiments, chromatin fibers adopt a loose and flexible geometry with roughly three nucleosomes per turn, on average, when the nucleosome core particle (NCP) is modeled from the 1.9 Å crystal structure (PDB ID 1KX5 [15]). Sometimes referred as a three-start helix or a triple-helix, this chromatin fiber arrangement has been proposed before [55, 104], and is distinct from the one-start and two-start models, prevalent in the chromatin literature [102]. Experimental data from *in vitro* assays of EPC display an oscillatory pattern for increasing length of nucleosome array, with a period of three that supports our predicted fiber model. Our results suggest that for E-P constructs with a 13mer nucleosome array, the removal of core histone H3/H4 tails reduces the EPC efficiency more than the removal of H2A/H2B tails, while tailless nucleosomes result in the least efficient EPC.

4.2 Methods

The computational framework for the simulations shown here is the same as that presented in Chapter 3. DNA and the nucleosome core particle models are treated in the same way. Here we present the treatment of the additional components, the promoter-bound RNA polymerase holoenzyme complex and the enhancer-bound activator complex, as well as the treatment of E-P communication.

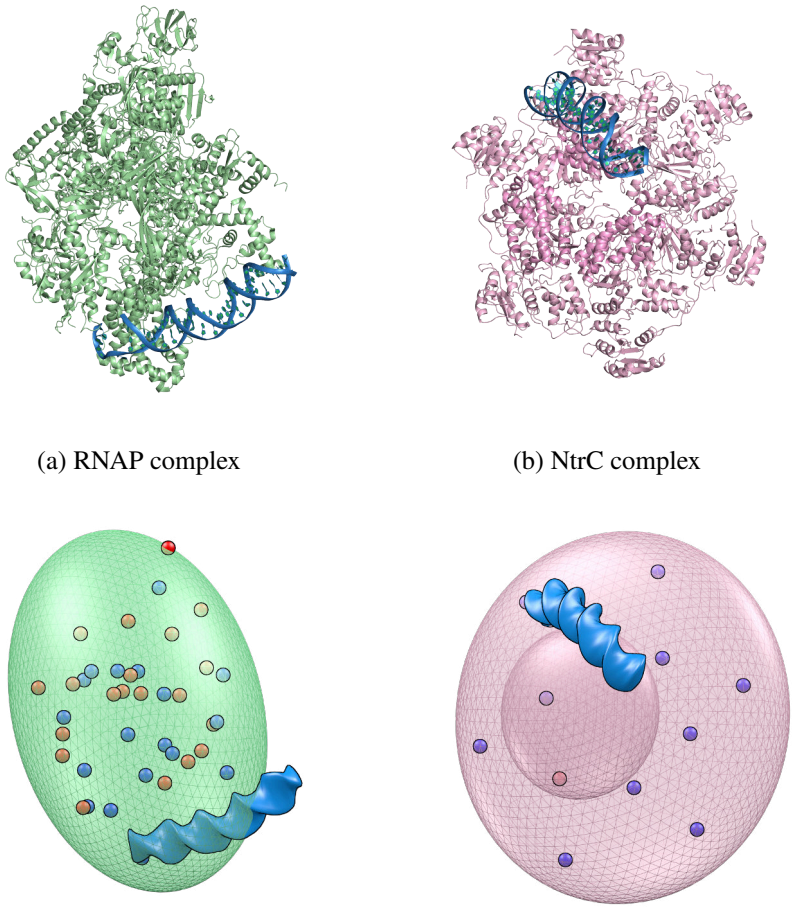
4.2.1 Enhancer and promoter protein complex modeling

The binding of the RNA polymerase (RNAP) holoenzyme complex at the promoter site was modeled based on the high-resolution crystal structures from Murakami et al. [138] and Bae et al. [139] (PDB ID 1L9Z and 4XLP, respectively). The roughly ellipsoid shape of the enzyme (Figure 4.2) is represented by a cylinder with dimensions 13.5×11 nm (height×diameter) in our simulations. This simplified representation is chosen because it is more computationally efficient to compute collisions between cylinders than ellipsoids. The DNA follows the pathway specified in the high-resolution crystal structure and the reference frame of the cylinder is connected to the rest of the fiber through the local coordinates of the first DNA base pair frame, in the same way that DNA is incorporated in the modeling of nucleosomes (see Section 3.2.2). The charged amino acids are clustered and represented by a reduced set of point charges inside the volume in order to reduce computational complexity.

The binding of the bacterial nitrogen regulatory protein C (NtrC) activator at the enhancer DNA site was modeled from two separate components, since no high-resolution crystal structure of the full activated assembly bound to DNA exists to date. The structure for the activated full-length NtrC hexamer was taken from the pseudoatomic model proposed by De Carlo et al. [140]. The binding of the NtrC hexamer to DNA was modeled by docking the crystal structure of the DNA-recognition domain (PDB ID 4FTH) [141] onto the NtrC model. Due to the hexameric nature of the protein complex, there are several docking alternatives for the DNA-binding domain. Since these docking sites are identical from a coarse-grained perspective, the choice does not affect the resulting model. The overall shape of the assembly is approximated by two ellipsoids (Figure 4.2), which are also represented as cylinders with respective dimensions of 3×7 nm and 4.5×11 nm (height×diameter), in our simulations.

It is not possible to position the E-P protein complexes precisely onto the respective promoter and enhancer DNA sequences being modeled in this study, due to the lack of high-resolution structures containing the same DNA sequence. This might affect the results of our predicted EPC rates, since the interaction between the two proteins, which is also not known in structural detail, is sensitive to their rotational phase around the DNA helix [142]. For these reasons, we have also treated the enhancer and promoter bound proteins as neutral spheres positioned at the centers of the respective

regulatory DNA sequences, in order to assess any potential bias introduced by the specific geometric shapes used for excluded volume. This, in effect, averages out any sequence-positioning dependency. Results from both representations of the E-P regulatory proteins will be discussed in the text.



(a) RNAP complex

(b) NtrC complex

(c) RNAP CG model

(d) NtrC CG model

Figure 4.2: Coarse-grained modeling of (a,c) RNA polymerase and (b,d) NtrC activator complexes bound to the *glnAp2* promoter and enhancer sites, respectively, from high-resolution models. The simplified volumes, shown as ellipsoids here, are treated as cylinders of comparable dimensions in the simulations, due to the reduced complexity in detecting collisions. The charges represent a reduced set of those found in the high-resolution structures to lower the computational cost. Figure adapted from ref. [6] with permission from Future Medicine.

4.2.2 Starting configuration set-up

The nucleosome arrays containing the promoter-bound and enhancer-bound proteins were modeled to capture the experimental setup described previously [2, 5]. Specifically, for this study we are

considering arrays containing 4–10, or 13 DNA repeats of 177 base pairs incorporating the Widom 601 [31] nucleosome-positioning sequence. Despite the fact that the *in vitro* studies use the Widom 601 positioning sequence, the unbound DNA (in linker regions) is treated in our simulations as ideal B-DNA. The nucleosomal DNA is modeled from two different nucleosome crystal structures, one with an α-satellite sequence (PDB ID 1KX5 [15]) and the other with a Widom 601 sequence (PDB ID 3MVD [125]). The first nucleosome repeat near the promoter sequence contains a slightly different (truncated) positioning sequence and an extra five base pairs overall (182 bp total) in the biochemical studies [5]. This is reflected in our simulation setup, by a 5-bp longer linker region between nucleosomes 1 and 2 than the linkers between other nucleosomes. The promoter and enhancer regions, before the first nucleosome and after the last have fixed lengths that do not change between different constructs.

Since the value we report is the enhancement of EPC in chromatin compared to naked DNA, mimicking the *in vitro* experiments, we repeat these simulations for nucleosome-free DNA constructs of the same DNA sequence lengths, with the enhancer- and promoter-bound proteins at the same locations. Due to the much larger parameter space of the naked DNA, which necessitates longer simulation times, we often estimate the EPC in naked DNA by only considering the elastic energy of DNA deformation and not the electrostatic or excluded volume terms. These calculations are significantly faster and, especially in the case where the E-P proteins are treated as neutral spherical volumes, the results are a fairly good approximation of the more "explicit" simulations that include excluded volume and electrostatic interactions.

4.2.3 Calculation of enhancer-promoter communication (EPC)

As mentioned above, in the absence of a detailed structural model for the interaction between RNA polymerase and the NtrC activator, we estimate the rate of communication as the looping probability (Figure 4.3), i.e., the frequency of configurations where the E-P center-to-center distance (r) is less than a specified cutoff (r_0), under the assumption that E-P communication is mainly dependent on the distance between the proteins [1, 143]. The enhancement rate (ξ) is then the ratio of the

communication levels in the presence ($P_{\text{chromatin}}$) vs. absence (P_{DNA}) of nucleosomes:

$$EPC \propto P(r < r_0) \tag{4.1}$$

$$\xi = \frac{P_{\text{chromatin}}(r < r_0)}{P_{\text{DNA}}(r < r_0)} \tag{4.2}$$

In view of the dimensions of the E-P protein models used here, we consider r_0 distance cutoffs of 175, 200 and 225 Å as criteria for communication. We report the EPC enhancement and error estimate from the mean and standard deviation, respectively, of the calculated enhancement based on these three cutoffs.

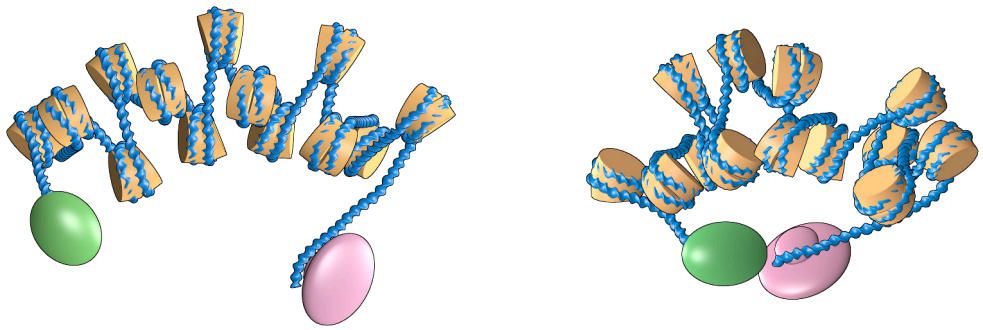


Figure 4.3: Illustration of "open" (left) and "looped" (right) constructs. When the centers of the proteins at the enhancer and promoter sites are closer in distance than the specified cutoff, here taken as 200 Å, the E-P proteins are assumed to be in contact. Figure adapted from ref. [6] with permission from Future Medicine.

Due to the relatively high stiffness of the nucleosome-decorated DNA chains, as well as the freely-jointed chain like behavior of the very long nucleosome depleted DNA, the probabilities of E-P contact are quite low and often necessitate rather long simulations in order to estimate the communication rates efficiently and with reasonably low uncertainty. To address the problem of sampling we employ a series of biased Monte Carlo (MC) simulations where the E-P protein centers are constrained to a small distance r_i and compute the conditional probabilities $P(r < r_i)$, a technique that has been previously used to estimate looping probabilities for short DNA chains [144]. Here we give a brief summary of the method developed by Podtelezhnikov and Vologodskii.

After generating an initial ensemble of configurations, we pick one that has a relatively low E-P distance, usually close to the r_0 distance that defines a contact, as the starting configuration for a series of constrained runs. We then choose a series of distances $r_0 < r_1 < \ldots < r_n$, where r_n is larger

than the maximum E-P separation observed from the initial distribution, as distance cutoffs for a series of constrained simulations. The starting configuration in each set of constrained simulations is usually the last configuration from the previous run. If we define the $P(r_i|r_{i+1})$ as the conditional probability $P(r < r_i|r \leq r_{i+1})$, then we can use the relationship $P(r_i) = P(r_i|r_{i+1})P(r_{i+1})$ and the fact that $P(r_n) = 1$, to compute the probability that the two E-P sites come into close contact $P(r < r_0)$ as:

$$P(r_0) = \sum_{i=0}^{n-1} P(r_i|r_{i+1}) \tag{4.3}$$

When the cutoff intervals are relatively small, the predicted conditional probabilities are high and therefore the error is minimized [144].

In our simulations, we perform 10–20 million Monte Carlo steps, which produce ensembles with approximately 20,000–30,000 accepted configurations. We inspect the progress of the looping probability over the course of the simulation, and depending on how well the ratio converges for $r_1 = 300$Å, we decide whether more sampling is necessary. After an acceptable convergence (through manual inspection), we often rely on one constrained simulation with $r_1 = 300$Å to refine the predicted communication rate by sampling the conditional probability $p(r < 200|r \leq 300)$, as described above.

4.3 Results and discussion

4.3.1 Enhancer-promoter communication profile supports a loose three-start chromatin fiber

From our simulation results, EPC enhancement shows an oscillatory pattern for constructs with increasing number of nucleosome repeats. This oscillatory pattern has a repeat of about three in the case of the 1KX5-based fibers (Figure 4.4, top panel). Qualitatively, the predicted pattern for these fibers is in agreement with the experimental pattern for the increase in transcription activation in chromatin compared to nucleosome-free DNA (Figure 4.4, middle panel). The predicted maximum enhancement values for the 1KX5-based arrays occur in arrays with 6, 9 and 12 nucleosomes between regulatory elements, which agrees with experiment. However, simulations of arrays containing 5, 7 and 10 nucleosomes produce the smallest levels of EPC enhancement, while experimentally, the

lowest levels of transcription activation occur for arrays of 5 and 8 repeats. On the other hand, simulations of arrays with nucleosomes based on the structure from 3MVD produce maximum enhancement values for arrays with 5, 7 and 10 repeats, and minimum EPC enhancement for arrays with 6 and 9 repeats (Figure 4.4, bottom panel). This pattern seems to be out of sync with the experimental results.

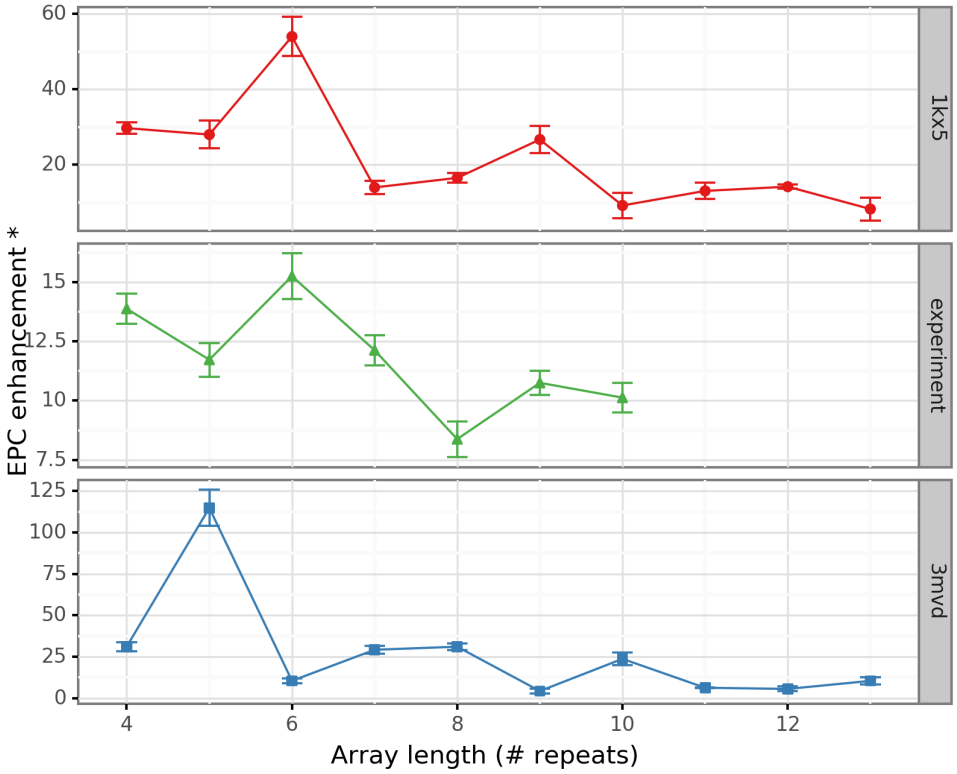


Figure 4.4: EPC enhancement in chromatin constructs with an increasing number of intervening nucleosomes. Predicted enhancement is measured by the relative E-P contact frequency between nucleosome-decorated DNA and naked DNA of the same length. The E-P regulatory proteins in both cases are approximated by cylindrical volumes, positioned at the enhancer and promoter DNA sequences. Chromatin simulations are based on nucleosomal DNA pathways modeled from two different crystal structures (1KX5 and 3MVD), which result in distinct chromatin fiber geometries. *Experimental values represent transcription activation in folds. All constructs contain nucleosome repeats of length 177 bp.

The differences in the predicted EPC patterns between the 1KX5- and 3MVD-based arrays reflect the structure of the chromatin fibers. In order to understand how the fiber geometry leads to the oscillatory pattern of predicted EPC, we have calculated a virtual torsion angle, as a simple measure describing the spatial disposition of the regulatory proteins with respect to the fiber axis. To calculate this angle, we have defined four points located respectively at the center of the protein

attached to the promoter site, the geometric center of the first three nucleosomes, the geometric center of the last three nucleosomes, and the center of the protein attached to the enhancer site (Figure 4.5). We have calculated this angle for each configuration in our accepted ensembles and the summary statistics are shown in Figure 4.6.

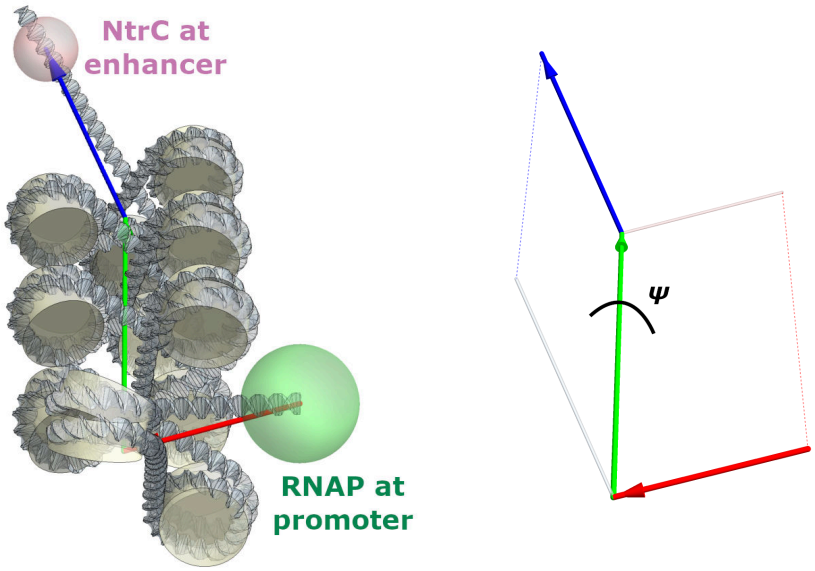


Figure 4.5: Illustration of the four points used to define the virtual torsion angle (ψ) between the two E-P proteins with respect to the fiber axis. The mean coordinates of the first and last three nucleosomes of the array are used respectively as the two central vertices. The first and fourth vertex lie on the centers of the proteins attached at the promoter and enhancer sites, respectively. The angle between the two planes, which are defined by each pair of consecutive vectors, corresponds to the virtual torsion angle.

The graph in the top panel of Figure 4.6 shows that the virtual torsion angle is close to 0°, on average, in simulated arrays with 6, 9 and 12 nucleosomes based on the 1KX5 structure. In other words, the enhancer and promoter regions in these arrays are on the same side of the fiber, on average. The arrangement of the regulatory proteins on the same side of the fiber allows them to get closer in distance and leads to maximum EPC (Figure 4.4). On the other hand, for 3MVD-based fibers, the virtual torsion angle is closest to 0° for arrays with 5 and 10 nucleosomes. The oscillatory pattern with a repeat of roughly three for the 1KX5-based arrays and roughly five for 3VMD-based arrays reflects the average fiber architectures, as discussed in Chapter 3.

While the pattern of predicted EPC values for the 1KX5-based fiber is in general agreement with the experimental values, the magnitudes deviate significantly (Figure 4.4), especially in the

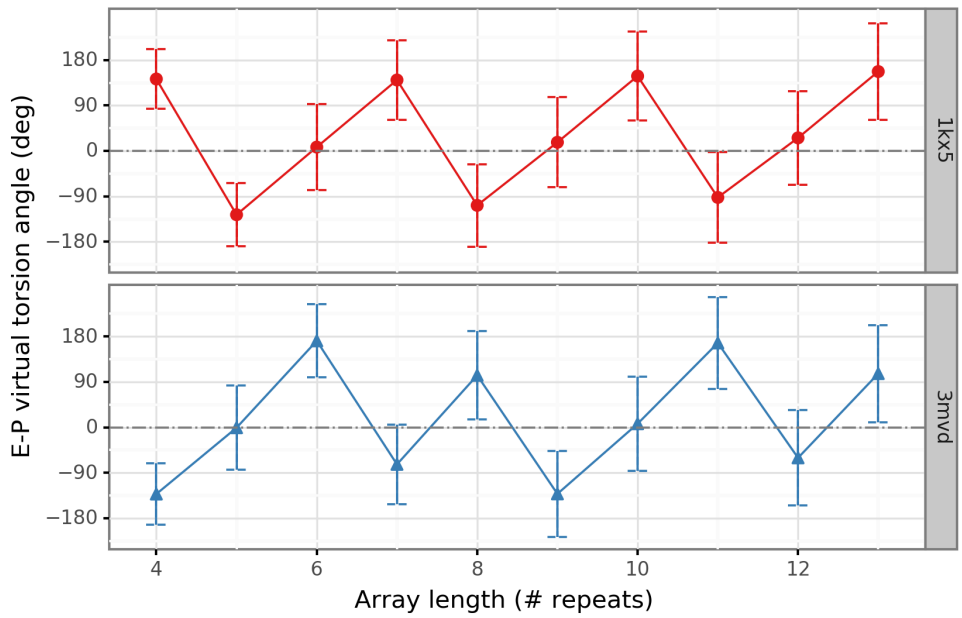


Figure 4.6: Distributions of the virtual torsion angle between the two E-P proteins separated by increasing number of intervening nucleosomes. When the average angle is near 0°, the proteins are roughly on the same side of the fiber, which allows them to get closer. Both systems exhibit oscillatory patterns for the virtual torsion angle, reflecting the internal chromatin fiber geometry. 1KX5-based fibers display a period of three, while 3MVD-based fibers display a period of five.

case of fibers with fewer nucleosomes, where the calculated numerical enhancement is several fold higher. This discrepancy could be due to several factors. One of the likely culprits could be our limited knowledge of the interaction between the regulatory proteins at the enhancer and promoter sites. Our assumption that general proximity leads to interaction does not consider any orientational dependence, which might lower the number of accepted looped configurations [3]. This type of dependence would likely have a bigger impact on shorter arrays, since the E-P proteins in these fibers are generally more rotationally constrained, i.e. their relative orientations are more strongly correlated in shorter constructs than in longer ones. The lack of rotational requirements for this interaction in our model might also introduce other potential problems which are discussed in Section 4.3.3.

4.3.2 Differential role of histone tails in enhancer-promoter communication

We used the 1KX5-based E-P constructs with 13 intervening nucleosomes to study the effect of different histone N-terminal tails on long-range communication. In particular, we investigated the

effects of removing H3/H4 and H2A/H2B tails, separately and in combination. The boundaries for the histone tails were modeled after the trypsin-sensitive regions used in the *in vitro* experimental setup [25, 113]. Our results show that removal of the H2A/H2B tails lowers EPC enhancement slightly but not as much as the removal of the H3/H4 tails (Figure 4.7). Arrays with tailless nucleosomes produced the lowest EPC enhancement. This suggests that H3/H4 tails are the most important for long-range communication. This general trend is supported by experimental values of transcription activation *in vitro*. However, the effect of tails on EPC enhancement in our predictions is not as strong as the effect measured *in vitro*. This magnitude difference between predicted and experimental enhancement could be due to the limitations discussed in the previous section or limitations in histone tail modeling.

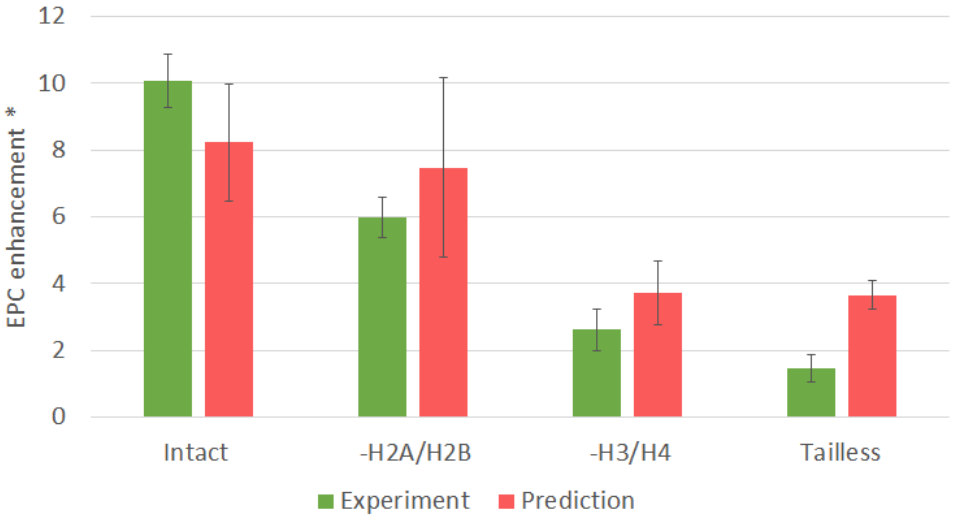


Figure 4.7: EPC enhancement in nucleosome-decorated DNA compared to naked DNA in constructs with 13 nucleosomes between regulatory proteins. Different combinations of histone tails have been truncated from the nucleosome core to measure the effect on EPC. The E-P regulatory proteins in all cases are approximated by cylindrical volumes, positioned at the enhancer and promoter DNA sequences, respectively. *Experimental values represent transcription activation in folds. All constructs contain nucleosome repeats of length 177 bp.

The removal of the tails generally cause an increase in chain extension, as evidenced by a shift in the peaks of the simulated E-P distance distributions towards higher values (Figure 4.8). At the same time, tail removal causes a slight narrowing of this distribution, suggesting a decrease in deformability. In line with the EPC enhancement results, the E-P distance distributions for arrays with tailless nucleosomes and those with truncated H3/H4 tails look nearly identical. The presence

of only the H3/H4 tails, however brings the distribution closer to the one seen for constructs with intact nucleosomes.

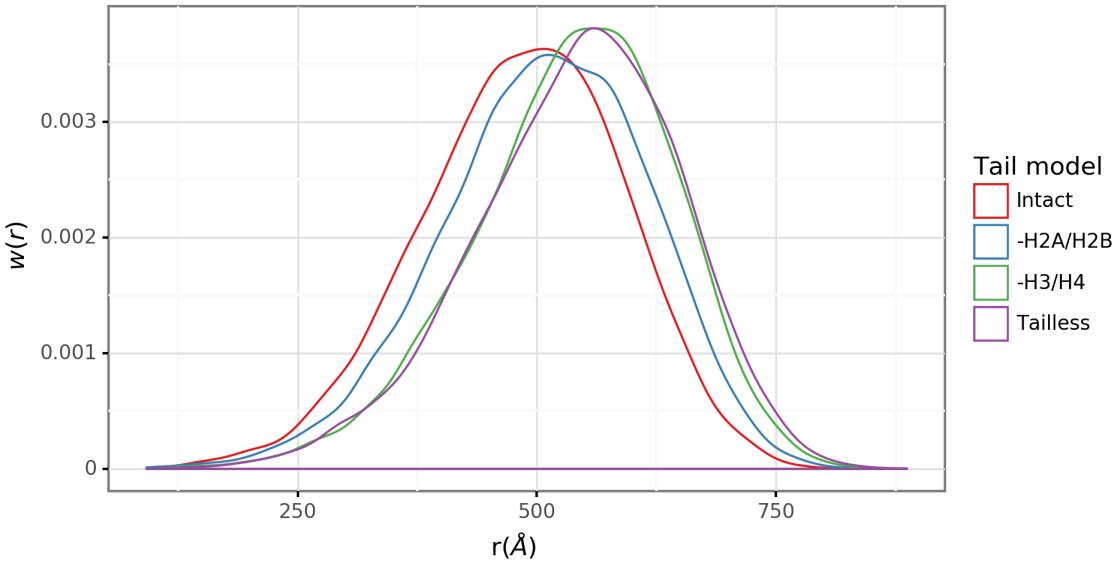


Figure 4.8: Normalized distributions of the E-P distances in constructs with various tail configurations. Removal of the H2A/H2B (blue line) shifts the distribution peak only slightly towards higher values, compared to intact nucleosomes (red line). Removal of the H3/H4 tails (green line) causes a more significant extension and a slight loss of flexibility, indicated by a narrowing of the distribution and has a similar effect to the removal of all core histone tails (purple line).

4.3.3 The effect of E-P coarse-grained representation

As we mentioned earlier, there are discrepancies between the experimental results and predicted EPC enhancement values for constructs with 1KX5-based chromatin, with regard to both the magnitudes and locations of minima. These discrepancies could be due to a number of factors, such as an orientational dependence for the interaction between regulatory proteins at the enhancer and promoter sites, that we do not consider in this study. Furthermore, the center-to-center distance between E-P protein complexes, which is used to define the contacts of regulatory proteins in our analysis, may be affected by the specific shapes of their coarse-grained (CG) representations and positioning on the enhancer/promoter DNA sequences. That is, during the course of the simulations, certain looped configurations may be rejected due to collisions induced by the specific arrangement of the E-P proteins, which depend on their precise location along the DNA. For example, shifting the NtrC along the enhancer sequence by a single base pair would rotate it by about 34° around the DNA helix,

on average. To understand the effects of this potential bias, given that we cannot precisely position the regulatory proteins with base-pair accuracy, we have performed the same set of simulations with a simpler representation of the enhancer-promoter proteins. In these simulations, we represent each protein by a single neutral sphere, positioned at the center of the regulatory sequence. The diameter of the sphere is taken as the lowest dimension of the respective cylindrical representation. It is important to note that, while useful in assessing the effect of E-P coarse-graining and positioning along the DNA, this over-simplified representation may still fail to capture many features of enhancer-promoter interactions. The results from our simulations with "simpler" representation of E-P proteins, which from here on are called spherical, are presented in Figure 4.9.

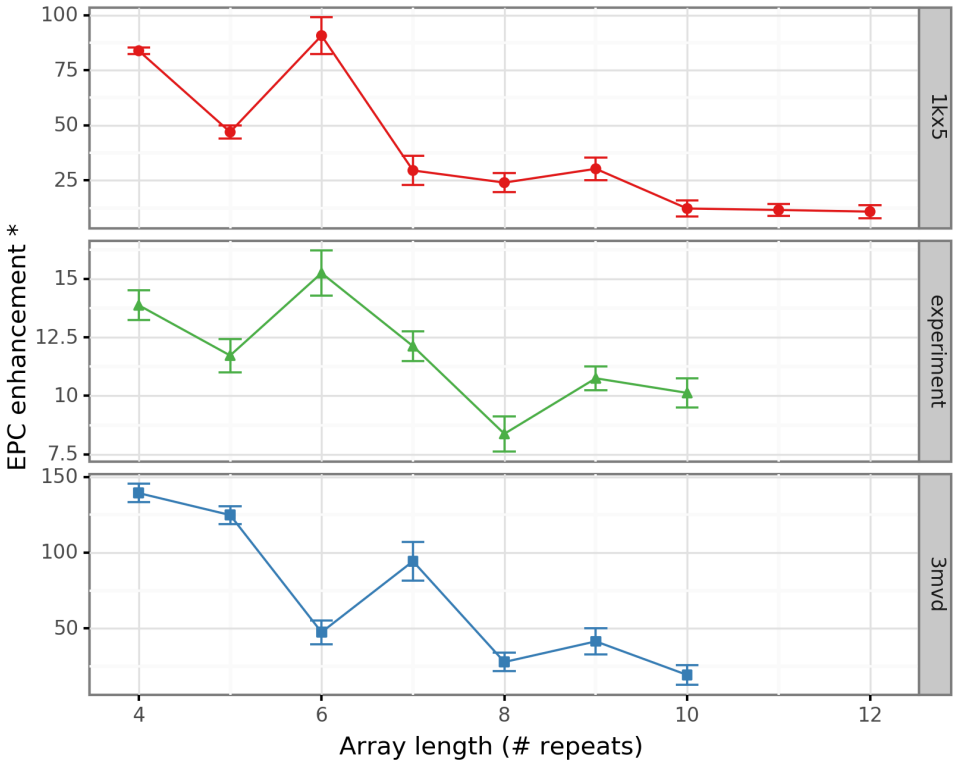


Figure 4.9: EPC enhancement in chromatin constructs with an increasing number of intervening nucleosomes. Predicted enhancement is measured by the relative E-P contact frequency between nucleosome-decorated DNA and naked DNA of the same length. The E-P proteins in this case are represented by neutral spheres located at the centers of the enhancer and promoter DNA sequences. Chromatin simulations are based on nucleosomal DNA pathways modeled from two different crystal structures (1KX5 and 3MVD), which result in distinct chromatin fiber geometries. *Experimental values represent transcription activation in folds. All constructs contain nucleosome repeats of length 177 bp.

The local minimum values of EPC enhancement, predicted to occur for constructs with 5 and

8 intervening 1KX5-based nucleosomes, are in closer agreement with experimental results. The EPC enhancement at 5 is a deeper minimum than in the set of simulations with more "realistic" cylindrical representations of the enhancer- and promoter-bound proteins (Figure 4.4, top panel), while the enhancement pattern for constructs with 10–12 nucleosomes is flattened. The locations of maximum values remain unchanged, at 6 and 9 intervening nucleosomes. Overall, these results bring the general predicted pattern for 1KX5-based E-P constructs in closer agreement with experimental results. The pattern of predicted EPC enhancement for 3MVD-based constructs, on the other hand, changes more dramatically as a result of the spherical E-P representation. Enhancement values for constructs with 8, 9 and 10 3MVD-based nucleosomes (Figure 4.9, bottom panel) are now in better qualitative agreement with experimental results, while the predicted pattern for arrays with 4–7 nucleosomes still appears to be out of sync with experiment.

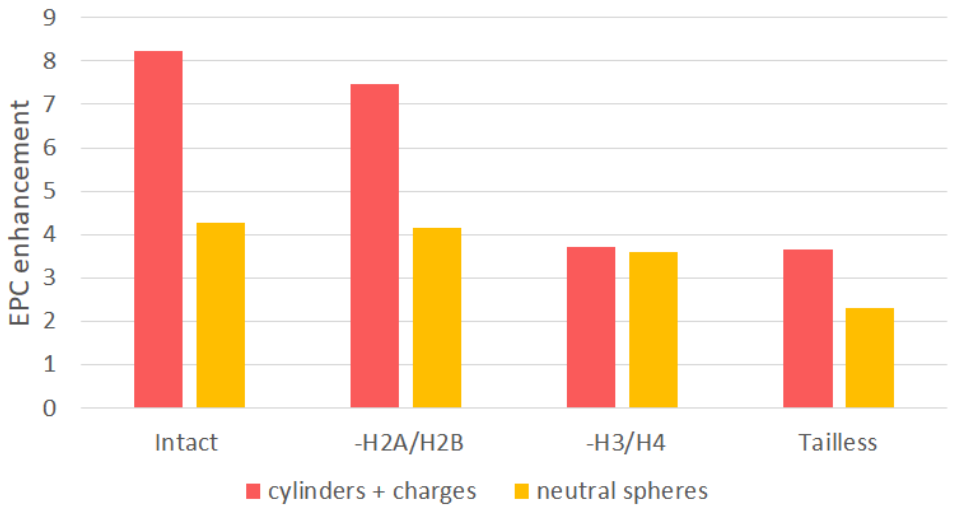


Figure 4.10: EPC enhancement in nucleosome-decorated DNA compared to naked DNA in constructs with 13 nucleosomes between regulatory proteins. Different combinations of histone tails have been truncated from the nucleosome core to measure the effect on EPC. The choice of coarse-grained representation of the regulatory proteins at the enhancer and promoter sites appears to affect the role of tails in EPC enhancement. While constructs with tailless nucleosomes result in lowered EPC compared to those with intact nucleosomes with both E-P models, the predicted enhancement effect is weaker in arrays with spherical representations of the E-P proteins (yellow bars). All constructs contain nucleosome repeats of length 177 bp.

The magnitudes of the predicted communication enhancement, however, still do not agree with experiment. Depending on the number of intervening nucleosomes, the EPC enhancement using the spherical representation of the proteins at the promoter and enhancer sites is either strengthened or

weakened, compared to the values for constructs with a more detailed representation of the regulatory proteins. For constructs with 13 intervening nucleosomes, for example, the predicted enhancement is generally weaker when we do not consider the cylindrical coarse-grained shape of the proteins (Figure 4.10). This implies that the specific mode of interaction between the two regulatory proteins is important for quantitative prediction of the experimentally observed long-range communication. Achieving this would require better modeling of the proteins at the enhancer and promoter sites and their interaction.

4.3.4 Role of nucleosomal DNA wrapping in enhancer-promoter communication

Using the spherical representation of the regulatory proteins, we explored the effect of partial DNA unwrapping from nucleosomes on EPC. In constructs with 1KX5-based nucleosome models, peeling of three base pairs from either side of the nucleosome generally enhances the predicted EPC. However, the level of enhancement depends on the length of the construct (Figure 4.11). The green graph in the lower panel shows that the EPC enhancement for constructs with six intervening 141-bp nucleosomes is nearly unchanged but is now a softer peak, compared to that determined for constructs with 147-bp nucleosomes. The peak for constructs with nine repeats nearly disappears completely. Overall, the predicted EPC enhancement pattern for increasing number of intervening nucleosomes becomes flatter when the DNA is partially unwrapped from the 1KX5 nucleosome model (Figure 4.11). The non-uniform effects of partial nucleosome unwrapping reflect the interplay between increased compaction (see Chapter 3, Section 3.3.4), and changes in the fiber architecture, which lead to different dispositions of the regulatory proteins with respect to the fiber axis.

For E-P constructs with 13 intervening nucleosomes, unwrapping three base pairs from either end of the 1KX5-based nucleosome model strengthens the predicted enhancing role of histone tails on EPC (Figure 4.12). The effect of selective histone tail truncation is considerably stronger in constructs containing 141-bp nucleosomes and longer unbound DNA linkers. Removal of all tails in constructs with 141-bp nucleosomes results in nearly three-fold reduction in EPC. This reduction is in close agreement with the experimental results (Figure 4.7). In contrast, constructs with 147-bp nucleosomes suffer only a two-fold reduction in EPC as a result of tail removal. The predicted reduction in EPC from H2A/H2B and H3/H4 truncation in the case of constructs with 141-bp nucleosomes is also in closer agreement with experiment. These results suggest that partial

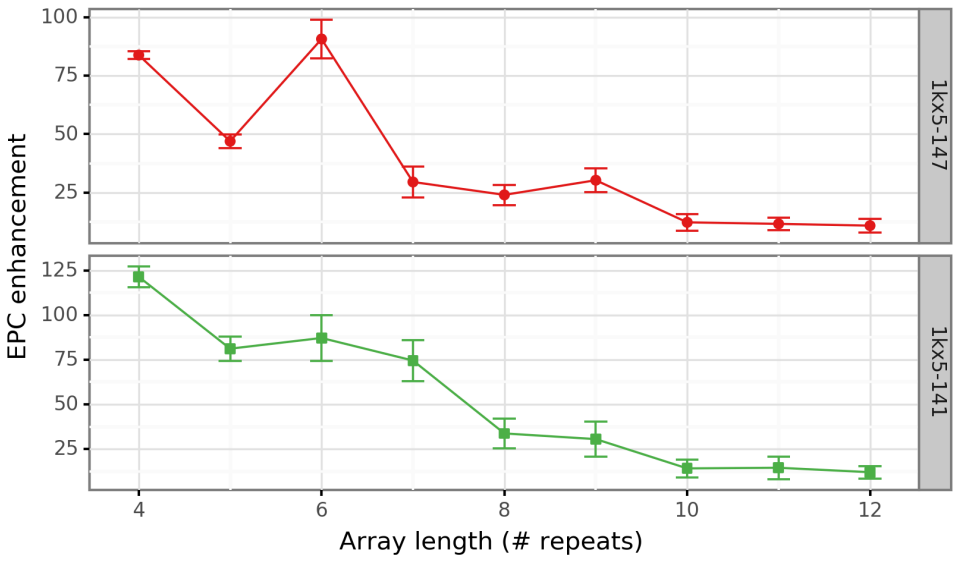


Figure 4.11: EPC enhancement in chromatin constructs with an increasing number of intervening nucleosomes. Predicted enhancement is measured by the relative E-P contact frequency between nucleosome-decorated DNA and naked DNA of the same length. The E-P proteins in this case are represented by neutral spheres located at the centers of the enhancer and promoter DNA sequences. Bound nucleosomal DNA pathways are modeled from the full 147 base pairs (upper graph, red) or the central 141 base pairs (lower graph, green) in the 1KX5 crystal structure [15], while the unbound DNA is subject to sampling. All constructs contain nucleosome repeats of length 177 bp.

nucleosome unwrapping from the nucleosome may have a strong effect on long-range communication that needs to be further investigated.

Discussion

In this work we have used two different nucleosome core particle models that produce distinguishable chromatin fiber geometries (Figure 4.13), according to our meso-scale treatment (see Chapter 3), as a means to investigate the role of fiber geometry on long-range communication. The chromatin fiber structure in E-P constructs with 4–13 intervening nucleosomes dictates whether the regulatory sequences are on the same or opposite sides of the fiber, on average, depending on the number of nucleosome repeats. Experimental data from *in vitro* EPC studies support a three-start helical arrangement of nucleosomes under physiologically-relevant conditions, compatible with our simulated 1KX5-based arrays. Our simulations suggest that the two-start helical arrangement seen in the 3MVD-based arrays is incompatible with experimental EPC data.

A triple-helix structure for chromatin has been proposed as early as 1985 from interpretation of

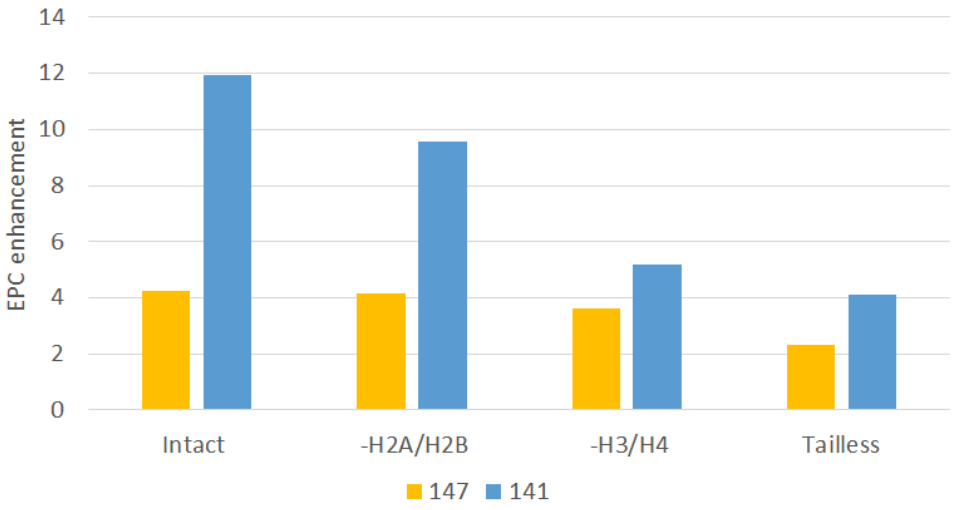
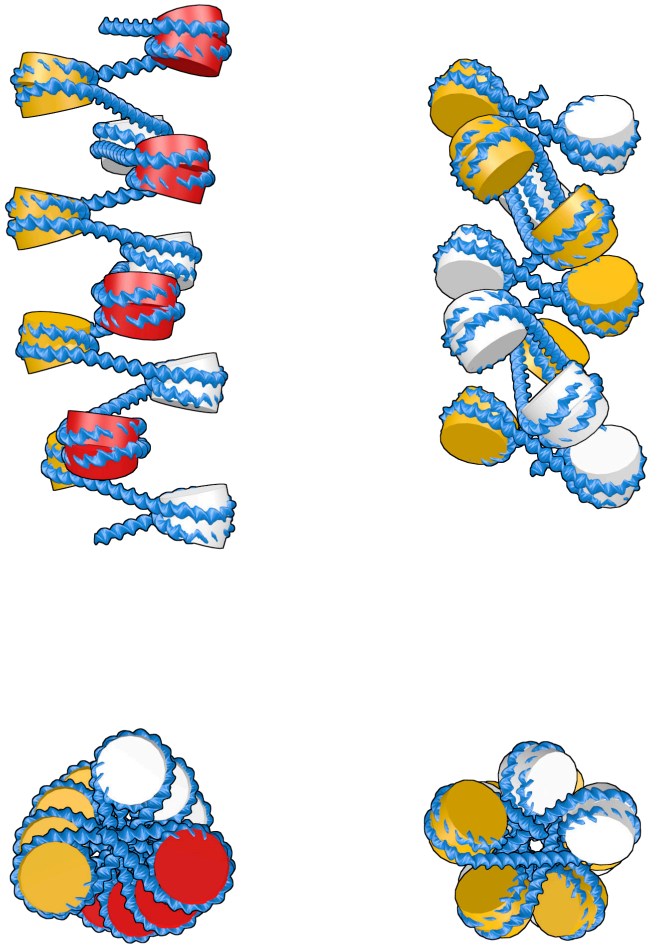


Figure 4.12: Role of partial DNA unwrapping from the nucleosome core on EPC enhancement in constructs with 13 intervening nucleosomes and various tail compositions. EPC is measured as the ratio of the contact frequency between nucleosome-decorated DNA and naked DNA. The E-P proteins in this case are represented by neutral spheres located at the centers of the enhancer and promoter DNA sequences. The predicted loss in EPC, as a result of histone tail removal is stronger in constructs containing 141-bp 1KX5-based nucleosomes than in those containing 147-bp nucleosomes.

flow linear dichroism data [55]. In this model, the faces of the nucleosomes are tilted by about 36° from the fiber axis. This three-start nucleosome arrangement stands in contrast with the two-start zigzag compact chromatin structure observed in X-ray [62, 64, 103], cryoEM [65] and other *in vitro* studies. Greenleaf and collaborators have recently used ionizing radiation-induced spatially correlated cleavage of DNA with sequencing (RICC-seq) to probe chromatin conformation in human cells *in situ* with near single-nucleosome precision [104]. They report that chromatin exists in an ensemble of states, including solenoid, two-start, three-start and multi-start structures. By comparing their results with data predicted from numerous static three-dimensional models [100], they suggest that fibers with three-start (or multi-start) structure are associated with more open chromatin, while two-start structures are more strongly associated with compact chromatin. Our simulated fibers that support *in vitro* EPC experiments are in agreement with an open three-start fiber.

While the role of tails on chromatin structure and compaction has been studied extensively [63, 120, 145–147], there is limited knowledge of their role in long-range communication [2]. Our results on the role of tails suggest that not all tails contribute equally to long-range communication in chromatin, with H3/H4 tails having the biggest effect. In our simulations, this is possibly due to



(a) 1KX5 - 177 bp NRL (b) 3MVD - 177 bp NRL

Figure 4.13: Front and top views of average geometries for nucleosome arrays containing 12×177-bp repeats. The fibers are built with nucleosome models from two different high-resolution structures. (a) The simulated arrays built from 1KX5-based [15] models lead to a more "open" structure with about three nucleosomes per turn, on average. The color coding is chosen to highlight the three-start helical arrangement. (b) 3MVD-based [125] arrays lead to a somewhat more compact structure with about five nucleosomes per turn, on average. In this case, the face-stacking nucleosomes form two columns, highlighted by the white and yellow color coding. An average structure, like the ones shown here, is built from the average base-pair step parameters of the unbound DNA linkers over the simulated ensemble.

the direct interaction of H3 tails with the DNA entering and exiting the nucleosome and the position of H4 tails on the face of the nucleosomes. Both of these factors contribute to small local changes in nucleosome arrangements, that in turn affect the spatial disposition of regulatory elements separated

by nucleosomes. Our results provide hints that the effect of tails may be related to the chromatin fiber arrangement, i.e., for different nucleosomal DNA wrapping (or different nucleosome repeat lengths), the selective removal of histone tails might lead to enhanced EPC. This remains to be studied in future simulations.

Our preliminary results from simulations with partially unwrapped nucleosomal DNA suggest that local dynamics in nucleosomes can play an important role on EPC. Here we only allowed three base pairs from either end of the nucleosome to flex and while the effect on local fiber structure is small, the impact this has on long-range communication is significant. This effect of this partial unwrapping in our simulations is ultimately tied to the chromatin fiber geometry, as evidenced by the differential effect on EPC depending on the number of intervening nucleosomes. Our data suggests an interplay between nucleosome repeat length, partial DNA unwrapping and the presence of N-terminal histone tails that regulates chromatin structure and the efficiency of long-range communication in chromatin. This view is consistent with the role of the linker histone protein, which is known to interact with the DNA entering and exiting the nucleosome and to modulate chromatin structure. It has been shown that nucleosomes undergo spontaneous "breathing" [26, 148, 149] and under external tension can unwrap asymmetrically [150, 151]. The effect of the asymmetric DNA unwrapping of up to 10 bp or more (corresponding to points of interaction between DNA and histone core) from either end of the nucleosome remains to be investigated in future work.

While our results are in qualitative agreement with experiment, the predicted enhancement values for EPC in chromatin compared to naked DNA are generally higher than experimental values by up to 4–5 fold. One of the reasons for the numerical discrepancies could be the orientational dependence of the interaction between the regulatory proteins attached at the enhancer and promoter sites [1, 3]. Our current model does not capture this effect and is limited by the lack of structural data on this interaction. Another potential reason for the quantitative discrepancies could be the consideration of sequence effects, especially in the long DNA segments that contain the regulatory sites. As we have presented in this work, small anisotropic deformations in these regions [105] could have significant effects on long-range communication. We are planning to address both of these concerns in future work.

4.4 Conclusions

Our computational model suggests that under physiologically-relevant conditions, such as those used in the *in vitro* experiments of enhancer-promoter communication, chromatin fibers adopt a loose and flexible geometry with roughly three nucleosomes per turn, on average. Our results from simulations of enhancer-promoter constructs with 4–10 intervening nucleosome repeats, predict an oscillatory pattern for long-range communication enhancement, with increasing number of repeats. This oscillatory pattern has a period of roughly three, and is in agreement with experimental data from *in vitro* assays. Our simulations suggest that in E-P constructs with 13 intervening nucleosomes, the removal of core histone H3/H4 tails reduces the EPC efficiency more than the removal of H2A/H2B tails, while tailless nucleosomes result in the least efficient EPC. The predicted values for EPC, however, are up to five-fold higher than the experimental ones in some cases. This discrepancy may be attributed to deficiencies in the modeling of the regulatory proteins attached to the enhancer and promoter sites, and their interaction, due to the lack of a high-resolution structure of the complex. These issues remain to be addressed in the future, in order to quantitatively predict EPC values observed in experiments.

Chapter 5

Concluding remarks

Dissertation summary

In this dissertation we have presented our investigations of chromatin structure and its implications on biological processes. Our work highlights the challenges in tackling such a complex problem and the necessity for an integrated approach. We have made use of structural modeling, linear programming, Monte-Carlo simulations, data mining, to name just a few methods that are part of our research toolbox. Just as important are the *in vitro* experiments conducted by our collaborators in order to validate our computational findings and help us understand how chromatin structure affects biological processes. Chapter 1 presented a brief history and an overview of the current state of knowledge about chromatin structure and its components.

In Chapter 2 we proposed an efficient method to fit flexible DNA models into density maps of a wide range of resolutions. DNA is modeled as a collection of base pairs that can be easily converted to an approximate atomic-level representation using existing tools [79]. This method can be very useful as rapid advances in cryogenic electron microscopy (cryoEM) have made it possible to obtain larger and more flexible structures with higher resolution than ever before. We have used this approach to generate approximate atomic-level models of recent low-resolution (11 Å) cryoEM density maps of two 12-nucleosome arrays [65]. Our results suggest that the intervening DNA linkers do not follow straight pathways and their arrangements within each tetra-nucleosome unit display structural heterogeneity. These results consequently point to a potential heterogeneity in the positioning of the linker histone globular domain. Our model also reveals potential problems

with one of the cryoEM density maps, where the inter-nucleosome distances seen in the model significantly deviate from those expected from the experimental setup.

In Chapter 3 we demonstrated that differences in the specific mode of DNA wrapping around the nucleosome core, as evidenced by available high-resolution structures in the Protein Data Bank (PDB), has a significant impact on chromatin fiber geometry. We have used a meso-scale representation of DNA to generate structural ensembles of nucleosome arrays to efficiently study the role of factors like nucleosome spacing, nucleosomal DNA folding and partial unwrapping. Our predictions for arrays of nucleosomes modeled from two high-resolution structures (PDB ID: 1KX5 [15], and PDB ID: 3MVD [125]), with different DNA wrapping, show that they adopt different architectures. The rearrangements induced by the nucleosomal DNA folding in each structure are equivalent to those caused by a change in nucleosome spacing by about 2–3 base pairs. For example, a 177-bp repeat nucleosome array modeled from the 1KX5 structure adopts a structure with roughly three nucleosomes per turn (3 NCPs/turn) helical arrangement, on average. An array with the same 177-bp repeat and nucleosomes modeled from the 3MVD structure displays a roughly 2.5 NCPs/turn helical arrangement, on average. These structural differences can have major implications for other processes, such as long-range communication (Chapter 4).

It is important to note: the fact that the two nucleosome structures used in our work contain different sequences (α-satellite vs Widom 601) appears coincidental. From our current analysis, sequence does not dictate the mode of nucleosomal DNA wrapping and the resulting nucleosome array structure. In fact, several other nucleosome structures with an α-satellite sequence also lead to arrays with an architecture similar to that of the Widom 601-containing 3MVD-based fibers. However, it remains unclear what leads to the differential mode of DNA wrapping around the nucleosome core. The α-satellite sequence is vastly more represented in the high-resolution structures in the PDB than the Widom 601 sequence, and this presents a challenge to elucidating the precise role of sequence. As more structures containing different sequences become available, it will be important to analyze their structural differences more carefully. The results described in Chapter 3 represent a first attempt at the large-scale analysis of nucleosome structures.

In Chapter 4 we used our meso-scale model to study the impact of chromatin fiber structure on long-range communication between enhancer and promoter regulatory DNA regions. Our simulations are set up to model the constructs used in the *in vitro* experiments led by our collaborators.

The main takeaway from the work presented here is that our enhancer-promoter communication (EPC) predictions from chromatin fibers that adopt a 3 NCPs/turn helical arrangement, on average, for a 177-bp repeat are in qualitative agreement with experiment. EPC predictions from chromatin fibers with a 177-bp repeat that adopt a 2.5 NCPs/turn helical arrangement, on average, are not in agreement with the *in vitro* EPC assays.

The combined results from Chapters 3 and 4 suggest that nucleosomal DNA folding, and consequently the structures adopted by the resulting nucleosome arrays, may vary among different *in vitro* experiments. These structural variations could arise from the differences in experimental factors, such as ionic content, and have major implications for global chromatin properties. Our work suggests that nucleosome arrays that adopt a 3 NCPs/turn helical structure for a 177-bp repeat (NCP model based on 1KX5) are in better agreement with EPC measurements from RNA transcription assays under physiological conditions (data from our collaborators), while arrays that adopt a 2.5 NCPs/turn helical structure for the same repeat length (NCP model based on 3MVD) are in better agreement with sedimentation velocity measurements [58]. Further research is necessary to investigate the underlying factors that dictate chromatin fiber behavior and the range of structures it may adopt. We hope that our work takes a first step into understanding the discrepancies seen among various chromatin structure studies.

Future considerations

The DNA model used in the work presented here is that of an ideal, naturally straight and isotropically bendable elastic rod. For short DNA segments, such as the linker DNA connecting consecutive nucleosomes, this is a fairly good approximation. However, sequence effects may have an impact on long chains of unbound DNA. For example, consideration of sequence-dependent rest states and deformabilities of individual base-pair steps on a 147-bp Widom 601 nucleosome positioning sequence causes a slight shift in the end-to-end distribution compared to an ideal DNA segment of the same length (Figure 5.1). The sequence-dependent chain shows a slight chain compression and directional bending. While these differences appear small, they may have a significant impact on the enhancer-promoter communication predictions, discussed in Chapter 4, due to their high sensitivity to the end-to-end distance.

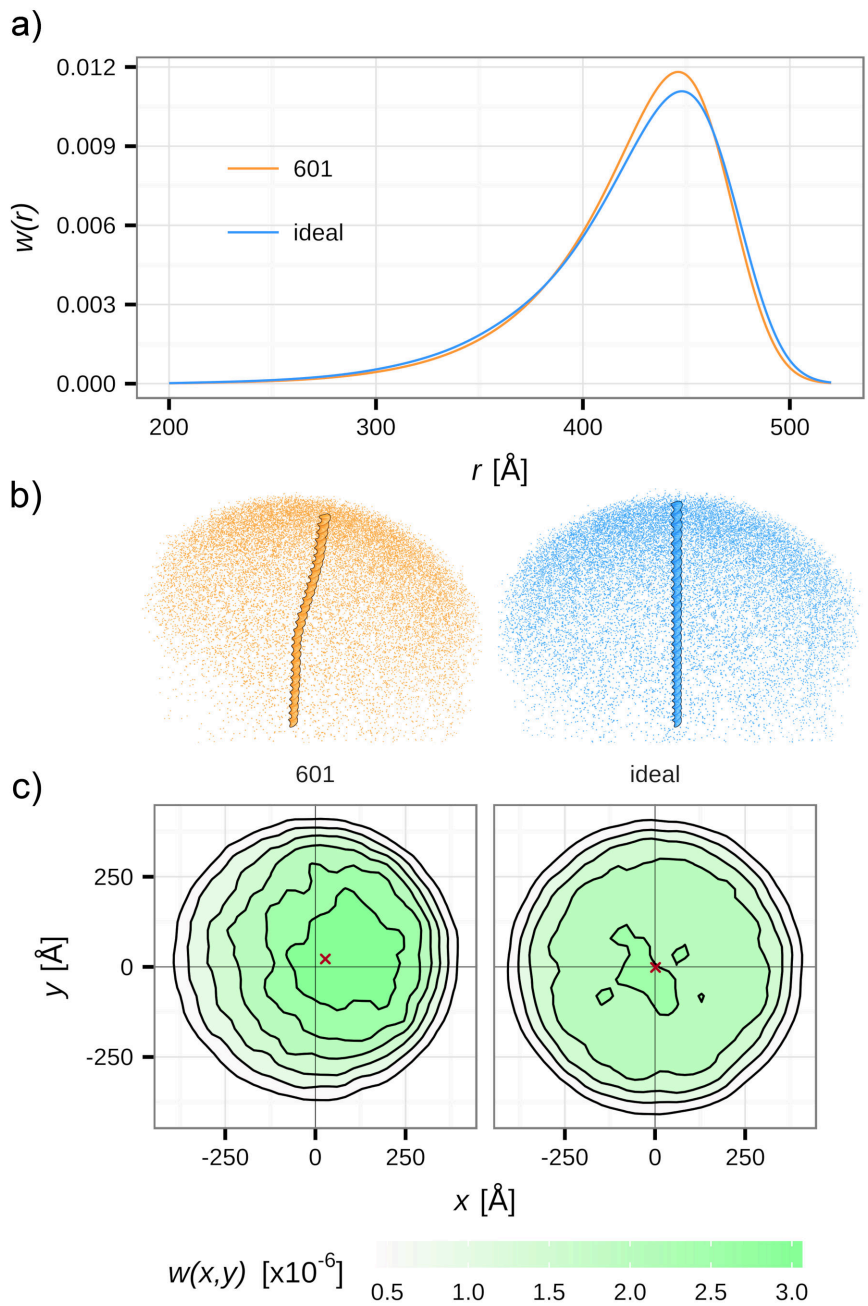


Figure 5.1: Treatment of the Widom 601 nucleosome-positioning sequence (147 bp) in terms of the sequence-dependent deformability and intrinsic structure of individual base-pair steps reveals a global anisotropy that is not present in an ideal DNA model. (a) Distribution $w(r)$ of end-to-end distances r from Monte Carlo simulations suggests that the sequence-dependent chain (gold) is slightly shorter and less flexible relative to its ideal DNA counterpart (blue). (b) These changes reflect subtle differences in the equilibrium rest states (ribbons) and sampled locations (dots) of chain ends. (c) The sequence-dependent deformations of base-pair steps enhance the asymmetry of the end-to-end distribution. This is illustrated by the different offsets of end-to-end densities relative to the average position of the last base-pair (denoted by red $\times$) of the two equilibrium structures in views looking down the normal (z axis) of the first base-pair in each chain. Figure reprinted from ref. [105].

In Chapter 3 we provided some limited insight into how partial DNA unwrapping from the nucleosome core could affect chromatin structure. Natural sequences do not have as strong of an affinity for the nucleosome core as synthetic sequences like the Widom 601 do. The flexibility of the entry/exit DNA linkers makes it necessary to account for nucleosome breathing during the course of the simulation. This is an area which is likely to be crucial for our work moving forward. Given the recent interest and development of experimental approaches using cryoEM to probe partial DNA unwrapping from the nucleosome [30], we are hopeful that future experiments will provide invaluable structural data that we can use to further our studies of chromatin structure.

Another area that remains a challenge for our simulations is the treatment of the linker histone and especially its long unstructured C-terminal domain (CTD). Recent work has suggested a new way of looking at the CTD as adopting a liquid-like behavior and phase-separating with linker DNA in the interior of the chromatin arrays in a dynamic manner [92]. This presents a challenge from a computational perspective, but also an opportunity to not be dependent on an elusive three-dimensional structure of this highly unstructured domain.

Bibliography

1. Zhang, G. *The packaging of DNA in chromatin* PhD thesis (Rutgers University, 2010).
2. Kulaeva, O. I., Zheng, G., Polikanov, Y. S., Colasanti, A. V., Clauvelin, N., Mukhopadhyay, S., Sengupta, A. M., Studitsky, V. M. & Olson, W. K. Internucleosomal interactions mediated by histone tails allow distant communication in chromatin. *J. Biol. Chem.* **287,** 20248–57 (2012).
3. Olson, W. K., Clauvelin, N., Colasanti, A. V., Singh, G. & Zheng, G. Insights into gene expression and packaging from computer simulations. *Biophys. Rev.* **4,** 171–178 (2012).
4. Clauvelin, N., Lo, P., Kulaeva, O. I., Nizovtseva, E. V., Diaz-Montes, J., Zola, J., Parashar, M., Studitsky, V. M. & Olson, W. K. Nucleosome positioning and composition modulate in silico chromatin flexibility. *J. Phys. Condens. Matter* **27,** 064112 (2015).
5. Nizovtseva, E. V., Clauvelin, N., Todolli, S., Polikanov, Y. S., Kulaeva, O. I., Wengrzynek, S., Olson, W. K. & Studitsky, V. M. Nucleosome-free DNA regions differentially affect distant communication in chromatin. *Nucleic Acids Res.* **45,** 3059–3067 (2016).
6. Nizovtseva, E. V., Todolli, S., Olson, W. K. & Studitsky, V. M. Towards quantitative analysis of gene regulation by enhancers. *Epigenomics* **9,** 1219–1231 (2017).
7. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171,** 737–738 (1953).
8. Kornberg, R. Chromatin structure: a repeating unit of histones and DNA. *Science* **184,** 3–6 (1974).
9. Tonna, S., El-Osta, A., Cooper, M. E. & Tikellis, C. Metabolic memory and diabetic nephropathy: potential role for epigenetic mechanisms. *Nat. Rev. Nephrol.* **6,** 332 (2010).
10. Kornberg, R. D. & Thonmas, J. O. Chromatin structure: oligomers of the histones. *Science* **184,** 865–868 (1974).
11. Olins, A. L. & Olins, D. E. Spheroid chromatin units (ν bodies). *Science* **183,** 330–332 (1974).
12. Oudet, P., Gross-Bellard, M. & Chambon, P. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4,** 281–300 (1975).
13. McGinty, R. K. & Tan, S. Nucleosome structure and function. *Chem. Rev.* **115,** 2255–2273 (2015).
14. Luger, K., Mäder, A. & Richmond, R. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **7,** 251–260 (1997).
15. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9Å resolution. *J. Mol. Biol.* **319,** 1097–1113 (2002).

16. Simpson, R. T. Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* **17,** 5524–5531 (1978).
17. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **589,** 2914–2922 (2015).
18. Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E. & Moudrianakis, E. N. The nucleosomal core histone octamer at 3.1 Åresolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci.* **88,** 10148–10152 (1991).
19. Arents, G. & Moudrianakis, E. N. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci.* **92,** 11170–11174 (1995).
20. Wang, X., Moore, S. C., Laszckzak, M. & Ausió, J. Acetylation increases the α-helical content of the histone tails of the nucleosome. *J. Biol. Chem.* **275,** 35013–35020 (2000).
21. Lee, K.-M. & Hayes, J. J. The N-terminal tail of histone H2A binds to two distinct sites within the nucleosome core. *Proc. Natl. Acad. Sci.* **94,** 8959–8964 (1997).
22. Ausio, J., Dong, F. & van Holde, K. Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome. *J. Mol. Biol.* **206,** 451–463 (1989).
23. Böhm, L. & Crane-Robinson, C. Proteases as structural probes for chromatin: The domain structure of histones. *Biosci. Rep.* **4,** 365–386 (1984).
24. Bondarenko, V. A., Steele, L. M., Újvári, A., Gaykalova, D. A., Kulaeva, O. I., Polikanov, Y. S., Luse, D. S. & Studitsky, V. M. Nucleosomes can frm a polar barrier to transcript elongation by RNA polymerase II. *Mol. Cell* **24,** 469–479 (2006).
25. Újvári, A., Hsieh, F. K., Luse, S. W., Studitsky, V. M. & Luse, D. S. Histone N-terminal tails interfere with nucleosome traversal by RNA polymerase II. *J. Biol. Chem.* **283,** 32236–32243 (2008).
26. Li, G., Levitus, M., Bustamante, C. & Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* **12,** 46–53 (2005).
27. Koopmans, W. J., Brehm, A., Logie, C., Schmidt, T. & Van Noort, J. Single-pair FRET microscopy reveals mononucleosome dynamics. *J. Fluoresc.* **17,** 785–795 (2007).
28. Koopmans, W. J., Buning, R., Schmidt, T. & Van Noort, J. spFRET using alternating excitation and FCS reveals progressive DNA unwrapping in nucleosomes. *Biophys. J.* **97,** 195–204 (2009).
29. Fierz, B. Dynamic chromatin regulation from a single molecule perspective. *ACS Chem. Biol.* **11,** 609–620 (2016).
30. Bilokapic, S., Strauss, M. & Halic, M. Histone octamer rearranges to adapt to DNA unwrapping. *Nat. Struct. Mol. Biol.* **25,** 101–108 (2018).
31. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276,** 19–42 (1998).
32. Thoma, F., Koller, T. & Klug, A. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* **83,** 403–427 (1979).
33. Hartman, P. G., Chapman, G. E., Moss, T. & Bradbury, E. M. Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The three structural regions of the histone H1 molecule. *Eur. J. Biochem.* **77,** 45–51 (1977).

34. Happel, N. & Doenecke, D. Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* **431,** 1–12 (2009).

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

36. Ramakrishnan, V., Finch, J. & Graziano, V. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362,** 219–223 (1993).

37. Hussain, S., Goutte-gattat, D., Becker, N., Meyer, S. & Shubhdarshan, M. Single-base resolution mapping of H1–nucleosome interactions and 3D organization of the nucleosome. *Proc. Natl. Acad. Sci.* **107,** 1–6 (2010).

38. Caterino, T. L. & Hayes, J. J. Structure of the H1 C-terminal domain and function in chromatin condensation. *Biochem. Cell Biol.* **89,** 35–44 (2011).

39. Fang, H., Clark, D. J. & Hayes, J. J. DNA and nucleosomes direct distinct folding of a linker histone H1 C-terminal domain. *Nucleic Acids Res.* **40,** 1475–84 (2011).

40. Fang, H., Wei, S., Lee, T.-H. & Hayes, J. J. Chromatin structure-dependent conformations of the H1 CTD. *Nucleic Acids Res.* gkw586 (2016).

41. Zhou, Y. B., Gerchman, S. E., Ramakrishnan, V., Travers, A. & Muyldermans, S. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature* **395,** 402–5 (1998).

42. Fan, L. & Roberts, V. A. Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. *Proc. Natl. Acad. Sci.* **103,** 8384–9 (2006).

43. Brown, D. T., Izard, T. & Misteli, T. Mapping the interaction surface of linker histone H1° with the nucleosome of native chromatin in vivo. *Nat. Struct. Mol. Biol.* **13,** 250–5 (2006).

44. Cui, F. & Zhurkin, V. B. Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to AT-rich and methylated DNA. *Nucleic Acids Res.* **37,** 2818–2829 (2009).

45. Syed, S. H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M. S., Hayes, J. J., Everaers, R., Angelov, D., Bednar, J. & Dimitrov, S. Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. *Proc. Natl. Acad. Sci.* **107,** 9620–5 (2010).

46. Zhou, B.-R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T. S. & Bai, Y. Structural mechanisms of nucleosome recognition by linker histones. *Mol. Cell,* 1–11 (2015).

47. Bednar, J. *et al.* Structure and dynamics of a 197 bp nucleosome in complex with linker histone H1. *Mol. Cell* **66,** 384–397.e8 (2017).

48. Schrodinger, L. *The PyMOL Molecular Graphics System, Version 1.8, Schrodinger, LLC.* 2010.

49. Zhou, B.-R., Feng, H., Ghirlando, R., Li, S., Schwieters, C. D. & Bai, Y. A small number of residues can determine if linker histones are bound on or off dyad in the chromatosome. *J. Mol. Biol.* **428,** 3948–3959 (2016).

50. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci.* **73,** 1897–1901 (1976).

51. Widom, J. Structure of the 300Å chromatin filament: X-ray diffraction from oriented samples. *Cell* **43,** 207–213 (1985).

52. Boulé, J.-B., Mozziconacci, J. & Lavelle, C. The polymorphisms of the chromatin fiber. *J. Phys. Condens. Matter* **27,** 033101 (2015).

53. Worcel, a., Strogatz, S. & Riley, D. Structure of chromatin and the linking number of DNA. *Proc. Natl. Acad. Sci.* **78,** 1461–1465 (1981).

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

55. Makarov, V., Dimitrov, S., Smirnov, V. & Pashev, I. A triple helix model for the structure of chromatin fiber. *FEBS Lett.* **181,** 357–361 (1985).

56. Simpson, R. T., Thoma, F. & Brubaker, J. M. Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: A model system for study of higher order structure. *Cell* **42,** 799–808 (1985).

57. Robinson, P. J. J., Fairall, L., Huynh, V. A. T. & Rhodes, D. EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci.* **103,** 6506–6511 (2006).

58. Correll, S. J., Schubert, M. H. & Grigoryev, S. A. Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J.* **31,** 2416–2426 (2012).

59. Wong, H., Victor, J.-M. & Mozziconacci, J. An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length. *PLoS One* **2,** e877 (2007).

60. Wu, C., Bassett, A. & Travers, A. A variable topology for the 30-nm chromatin fibre. *EMBO Rep.* **8,** 1129–1134 (2007).

61. Depken, M. & Schiessel, H. Nucleosome shape dictates chromatin fiber structure. *Biophys. J.* **96,** 777–784 (2009).

62. Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436,** 138–41 (2005).

63. Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. Chromatin fiber folding: Requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* **327,** 85–96 (2003).

64. Ekundayo, B., Richmond, T. J. & Schalch, T. Capturing Structural Heterogeneity in Chromatin Fibers. *J. Mol. Biol.* **429,** 3031–3042 (2017).

65. Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.-M., Zhu, P. & Li, G. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344,** 376–80 (2014).

66. Sha, K. & Boyer, L. A. *The chromatin signature of pluripotent cells* 1–21 (Harvard Stem Cell Institute, 2008).

67. Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. & Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

68. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).

69. Matharu, N. & Ahituv, N. Minor loops in major folds: enhancer–promoter looping, chromatin restructuring, and their association with transcriptional regulation and disease. *PLOS Genet.* **11,** e1005640 (2015).

70. Eltsov, M., MacLellan, K. M., Maeshima, K., Frangakis, A. S. & Dubochet, J. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc. Natl. Acad. Sci.* **105,** 19732–19737 (2008).

71. Maeshima, K., Rogge, R., Tamura, S., Joti, Y., Hikima, T., Szerlong, H., Krause, C., Herman, J., Seidel, E., DeLuca, J., Ishikawa, T. & Hansen, J. C. Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. *EMBO J.* **35,** e201592660 (2016).

72. Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H. & O'Shea, C. C. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357** (2017).

73. Maeshima, K., Ide, S., Hibino, K. & Sasai, M. Liquid-like behavior of chromatin. *Curr. Opin. Genet. Dev.* **37,** 36–45 (2016).

74. Nozaki, T. *et al.* Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Mol. Cell* **67,** 282–293.e7 (2017).

75. Bintu, B., Mateo, L. J., Su, J.-H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. N. & Zhuang, X. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362,** eaau1783 (2018).

76. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–12 (2004).

77. Chacón, P. & Wriggers, W. Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.* **317,** 375–384 (2002).

78. Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 344–351 (2012).

79. Lu, X.-J. & Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31,** 5108–5121 (2003).

80. Lu, X.-J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **3,** 1213–1227 (2008).

81. Olson, W. K., Gorin, a. a., Lu, X. J., Hock, L. M. & Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci.* **95,** 11163–11168 (1998).

82. Kosikov, K. M., Gorin, A. A., Zhurkin, V. B. & Olson, W. K. DNA stretching and compression: large-scale simulations of double helical structures. *J. Mol. Biol.* **289,** 1301–1326 (1999).

83. Dickerson, R. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* **17,** 1797–1803 (1989).

84. Olson, W. K. *et al.* A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313,** 229–237 (2001).

85. Powell, M. J. D. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* **7,** 155–162 (1964).

86. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–42 (2000).

87. Cerf, C., Lippens, G., Ramakrishnan, V., Muyldermans, S., Segers, A., Wyns, L., Wodak, S. J. & Hallenga, K. Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: Full assignment, tertiary structure, and comparison with the globular domain of histone H5. *Biochemistry* **33,** 11079–11086 (1994).

88. Ono, K., Kusano, O., Shimotakahara, S., Shimizu, M., Yamazaki, T. & Shindo, H. The linker histone homolog Hho1p from *Saccharomyces cerevisiae* represents a winged helix-turn-helix fold as determined by NMR spectroscopy. *Nucleic Acids Res.* **31,** 7199–7207 (2003).

89. Ali, T., Coles, P., Stevens, T. J., Stott, K. & Thomas, J. O. Two homologous domains of similar structure but different stability in the yeast linker histone, Hho1p. *J. Mol. Biol.* **338,** 139–148 (2004).

90. Sanderson, A., Stott, K., Stevens, T. J. & Thomas, J. O. Engineering the structural stability and functional properties of the GI domain into the intrinsically unfolded GII domain of the yeast linker histone Hho1p. *J. Mol. Biol.* **349,** 608–20 (2005).

91. Talbert, P. B. *et al.* A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin* **5,** 7 (2012).

92. Turner, A. L., Watson, M., Wilkins, O. G., Cato, L., Travers, A., Thomas, J. O. & Stott, K. Highly disordered histone H1–DNA model complexes and their condensates. *Proc. Natl. Acad. Sci.* **115,** 11964–11969 (2018).

93. Gibbs, E. B. & Kriwacki, R. W. Linker histones as liquid-like glue for chromatin. *Proc. Natl. Acad. Sci.* **115,** 11868–11870 (2018).

94. Borgia, A. *et al.* Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555,** 61–66 (2018).

95. Zhou, B.-R., Jiang, J., Ghirlando, R., Norouzi, D., Sathish Yadav, K. N., Feng, H., Wang, R., Zhang, P., Zhurkin, V. & Bai, Y. Revisit of reconstituted 30-nm nucleosome arrays reveals an ensemble of dynamic structures. *J. Mol. Biol.* **430,** 3093–3110 (2018).

96. Rohl, C. a., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383,** 66–93 (2004).

97. Ben-Haïm, E., Lesne, A. & Victor, J.-M. Chromatin: A tunable spring at work inside chromosomes. *Phys. Rev. E* **64,** 051921 (2001).

98. Grigoryev, S. A., Arya, G., Correll, S., Woodcock, C. L. & Schlick, T. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc. Natl. Acad. Sci.* **106,** 13317–13322 (2009).

99. Aumann, F., Sühnel, J., Langowski, J. & Diekmann, S. Rigid assembly and Monte Carlo models of stable and unstable chromatin structures: The effect of nucleosomal spacing. *Theor. Chem. Acc.* **125,** 217–231 (2010).

100. Koslover, E. F., Fuller, C. J., Straight, A. F. & Spakowitz, A. J. Local geometry and elasticity in compact chromatin structure. *Biophys. J.* **99,** 3941–3950 (2010).

101. Lanzani, G. & Schiessel, H. Out of register: How DNA determines the chromatin fiber geometry. *EPL* **97,** 38002 (2012).

102. Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **13,** 436–47 (2012).

103. Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R. & Richmond, T. J. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **306,** 1571–1573 (2004).

104. Risca, V. I., Denny, S. K., Straight, A. F. & Greenleaf, W. J. Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. *Nature* **541,** 237–241 (2017).

105. Todolli, S., Perez, P. J., Clauvelin, N. & Olson, W. K. Contributions of sequence to the higher-order structures of DNA. *Biophys. J.* **112,** 416–426 (2017).

106. Miyazawa, T. Molecular vibrations and structure of high polymers. II. Helical parameters of infinite polymer chains as functions of bond lengths, bond angle, and internal rotation angles. *J. Polym. Sci.* **55,** 215–231 (1961).

107. Sugeta, H. & Tatsuo Miyazawa. General method for calculating helical parameters of polymer chains from bond lengths, bond angles, and internal-rotation angles. *Biopolymer* **5,** 673–679 (1967).

108. Perišić, O., Collepardo-Guevara, R. & Schlick, T. Modeling studies of chromatin fiber structure as a function of DNA linker length. *J. Mol. Biol.* **403,** 777–802 (2010).

109. Norouzi, D. & Zhurkin, V. B. Topological polymorphism of the two-start chromatin fiber. *Biophys. J.* **108,** 2591–2600 (2015).

110. Nikitina, T., Norouzi, D., Grigoryev, S. A. & Zhurkin, V. B. DNA topology in chromatin is defined by nucleosome spacing. *Sci. Adv.* **3,** 1–8 (2017).

111. Czapla, L., Swigon, D. & Olson, W. K. Sequence-dependent effects in the cyclization of short DNA. *J. Chem. Theory Comput.* **2,** 685–695 (2006).

112. Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **11,** 179–246 (1978).

113. Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272,** 301–311 (1997).

114. Potoyan, D. A. & Papoian, G. A. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.* **133,** 7405–7415 (2011).

115. Liu, J. S. *Monte Carlo Strategies in Scientific Computing* (Springer New York, New York, NY, 2004).

116. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21,** 1087–1092 (1953).

117. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57,** 97–109 (1970).

118. Roberts, G. O., Gelman, A. & Gilks, W. R. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7,** 110–120 (1997).

119. Clauvelin, N., Olson, W. K. & Tobias, I. Characterization of the geometry and topology of DNA pictured as a discrete collection of atoms. *J. Chem. Theory Comput.* **8,** 1092–1107 (2012).

120. Arya, G. & Schlick, T. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J. Phys. Chem. A* **113,** 4045–59 (2009).

121. Buckwalter, J. M., Norouzi, D., Harutyunyan, A., Zhurkin, V. B. & Grigoryev, S. A. Regulation of chromatin folding by conformational variations of nucleosome linker DNA. *Nucleic Acids Res.* **137,** 10205–10215 (2017).

122. Hansen, J. C., Ausio, J., Stanik, V. H. & van Holde, K. E. Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1. *Biochemistry* **28,** 9129–9136 (1989).

123. Zhang, P., Tobias, I. & Olson, W. K. Computer simulation of protein-induced structural changes in closed circular DNA. *J. Mol. Biol.* **242,** 271–290 (1994).

124. Swigon, D., Coleman, B. D. & Tobias, I. The elastic rod model for DNA and its application to the tertiary structure of DNA minicircles in mononucleosomes. *Biophys. J.* **74,** 2515–2530 (1998).

125. Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467,** 562–566 (2010).

126. Savelyev, A. & Papoian, G. A. Electrostatic, steric, and hydration interactions favor Na^+ condensation around DNA compared with K^+. *J. Am. Chem. Soc.* **128,** 14506–14518 (2006).

127. Savelyev, A. & Papoian, G. A. Inter-DNA electrostatics from explicit solvent molecular dynamics simulations. *J. Am. Chem. Soc.* **129,** 6060–6061 (2007).

128. Savelyev, A. & MacKerell, A. D. Differential impact of the monovalent ions Li^+, Na^+, K^+, and Rb^+ on DNA conformational properties. *J. Phys. Chem. Lett.* **6,** 212–216 (2015).

129. Savelyev, A. & MacKerell, A. D. Differential deformability of the DNA minor groove and altered BI/BII backbone conformational equilibrium by the monovalent ions Li^+, Na^+, K^+, and Rb^+ via water-mediated hydrogen bonding. *J. Chem. Theory Comput.* **11,** 4473–4485 (2015).

130. Wu, C., Mcgeehan, J. E. & Travers, A. A metastable structure for the compact 30-nm chromatin fibre. *FEBS Letters* **590,** 1–8 (2016).

131. Scipioni, A., Turchetti, G., Morosetti, S. & De Santis, P. Geometrical, conformational and topological restraints in regular nucleosome compaction in chromatin. *Biophys. Chem.* **148,** 56–67 (2010).

132. Várnai, P. & Timsit, Y. Differential stability of DNA crossovers in solution mediated by divalent cations. *Nucleic Acids Res.* **38,** 4163–4172 (2010).

133. McGhee, J. D. & Felsenfeld, G. Nucleosome structure. *Annu. Rev. Biochem.* **49,** 1115–1156 (1980).

134. Widom, J. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc. Natl. Acad. Sci.* **89,** 1095–1099 (1992).

135. Rubtsov, M. A., Polikanov, Y. S., Bondarenko, V. A., Wang, Y.-H. & Studitsky, V. M. Chromatin structure can strongly facilitate enhancer action over a distance. *Proc. Natl. Acad. Sci.* **103,** 17690–17695 (2006).

136. Liu, Y., Bondarenko, V. A., Ninfa, A. & Studitsky, V. M. DNA supercoiling allows enhancer action over a large distance. *Proc. Natl. Acad. Sci.* **98,** 14883–8 (2001).

137. Bondarenko, V. A., Liu, Y. V., Jiang, Y. I. & Studitsky, V. M. Communication over a large distance: enhancers and insulators. *Biochem. Cell Biol.* **81,** 241–251 (2003).

138. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science* **296,** 1285–1290 (2002).

139. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. & Darst, S. A. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* **4,** 1–23 (2015).

140. De Carlo, S., Chen, B., Hoover, T. R., Kondrashkina, E., Nogales, E. & Nixon, B. T. The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes Dev.* **20,** 1485–1495 (2006).

141. Vidangos, N. K., Heideker, J., Lyubimov, A., Lamers, M., Huo, Y., Pelton, J. G., Ton, J., Gralla, J., Berger, J. & Wemmer, D. E. DNA recognition by a σ54 transcriptional activator from *Aquifex aeolicus*. *J. Mol. Biol.* **426,** 3553–3568 (2014).

142. Lilja, A. E., Jenssen, J. R. & Kahn, J. D. Geometric and dynamic requirements for DNA looping, wrapping and unwrapping in the activation of *E. coli glnAp2* transcription by NtrC. *J. Mol. Biol.* **342,** 467–478 (2004).

143. Polikanov, Y. S., Bondarenko, V. A., Tchernaenko, V., Jiang, Y. I., Lutter, L. C., Vologodskii, A. & Studitsky, V. M. Probability of the site juxtaposition determines the rate of protein-mediated DNA looping. *Biophys. J.* **93,** 2726–31 (2007).

144. Podtelezhnikov, A. A. & Vologodskii, A. V. Dynamics of small loops in DNA molecules. *Macromolecules* **33,** 2767–2771 (2000).

145. Liu, Y., Lu, C., Yang, Y., Fan, Y., Yang, R., Liu, C. F., Korolev, N. & Nordenskiöld, L. Influence of histone tails and H4 tail acetylations on nucleosome–nucleosome interactions. *J. Mol. Biol.* **414,** 749–764 (2011).

146. Chen, Q., Yang, R., Korolev, N., Liu, C. F. & Nordenskiöld, L. Regulation of nucleosome stacking and chromatin compaction by the histone H4 N-terminal tail–H2A acidic patch interaction. *J. Mol. Biol.* **429,** 2075–2092 (2017).

147. Bascom, G. D. & Schlick, T. Chromatin fiber folding directed by cooperative histone tail acetylation and linker histone binding. *Biophys. J.* **0,** 1–10 (2018).

148. Hayes, J. J. & Hansen, J. C. New insights into unwrapping DNA from the nucleosome from a single-molecule optical tweezers method. *Proc. Natl. Acad. Sci.* **99,** 1752–1754 (2002).

149. Bancaud, A., Conde e Silva, N., Barbi, M., Wagner, G., Allemand, J.-F., Mozziconacci, J., Lavelle, C., Croquette, V., Victor, J.-M., Prunell, A. & Viovy, J.-L. Structural plasticity of single chromatin fibers revealed by torsional manipulation. *Nat. Struct. Mol. Biol.* **13,** 444–50 (2006).

150. Ngo, T. T. M., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A. & Ha, T. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **7,** 10813 (2016).

151. De Bruin, L., Tompitak, M., Eslami-Mossallam, B. & Schiessel, H. Why do nucleosomes unwrap asymmetrically? *J. Phys. Chem. B* **120,** 5855–5863 (2016).