THE UNIVERSITY OF CHICAGO

CHEMICALLY-SPECIFIC MOLECULAR SIMULATION OF OLIGONUCLEOSOMAL CHROMATIN STRUCTURE

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

BY
ARIA EESTI CORAOR

CHICAGO, ILLINOIS
DECEMBER 2023

I dedicate this work to my parents, whose unfailing support and acceptance has carried me through the toughest times.

"Science is one land, having the ability to accommodate even more people, as more residents gather in it; it is a treasure that is the greater the more it is shared... Each of us can do his work in his own way, and the common ground does not mean conformity."

—*Wilhelm Ostwald*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My sincerest gratitude to my advisor, Prof. Juan de Pablo, without whose advice and guidance this work would not have been possible. Further, I am most grateful to my committee, including Prof. de Pablo, Prof. Allison Squires, and Prof. Rebecca Willet for their support and direction in these past few years.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE-1746045), and by National Science Foundation Grant No. (CEE-1830969). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We further acknowledge computational resources provided by the Research and Computing Center via the Midway computing cluster at the University of Chicago and the Laboratory Resource Computing Center at Argonne National Laboratories.

ABSTRACT

The manner in which nucleosomes stack to form higher order chromatin is critical to biological regulation of gene expression, but remains poorly understood. Throughout this work, we utilize multiple coarse-grained models to characterize chromatin structure across length-scales, and lend physical interpretation to experimental results. Our simulations reveal that classical, ideal models for chromatin structure are not thermodynamically stable, and that chromatin displays a wide range of heteromorphic structures at equilibrium which span the range between these ideal models. We combine these simulation results with Bayesian maximum a-posteriori estimation to augment the interpretability of related cryo-electron microscopy techniques which have heretofore been used to characterize chromatin fiber structure. We further extend these simulations to fibers of chromatin bearing H4K16 acylations, which are epigenetic marks understood to drive increases in gene expression *in vivo*. Our results demonstrate that, although all H4K16 acylations significantly alter chromatin fiber structure, the resulting structural variation is small compared to the natural heterogeneity present in each fiber at equilibrium. Lastly, we develop and validate a novel method to directly compare and contrast chromatin structural measurements obtained by complementary Hi-C and IF-sSMLM measurements, and demonstrate this method's utility on simulated chromatin structures. In total, our work both adds considerable detail to the field of chromatin folding, and provides clear avenues for future developments of combined computational and experimental techniques.

CHAPTER 1

INTRODUCTION

The study of gene expression, and the factors which influence and regulate it, has remained a central topic in biology for decades.[1,2] It has become increasingly apparent that, beyond the primary sequence of nucleotide bases, the secondary and higher-order structure of chromatin influences gene expression by affecting the availability of DNA to the nucleoplasm and selectively allowing site-specific binding of transcription factors.[3] These higher-order structures are known to be modified by epigenetic changes such as histone modifications, which in turn are known to stably and heritably modify gene expression patterns.[4] However, mechanistic understandings of how these epigenetic changes affect gene expression have frequently eluded researches due to the complexity of the involved biophysical interactions, and the difficulty of probing single-molecule interactions *in vivo* in biological systems.[4] The few mechanistic insights to-date on the influence of molecular interactions on multiscale chromatin structure have enabled the development of new treatments for cancer, as well as better understanding of autoimmunity, neurological and developmental disorders..[5,6]

Classical models for multiscale chromatin organization entail a system of "hierarchical folding", in which the 2-nanometer scale DNA helix wraps around an octomer of histone proteins forming a 10-nanometer scale nucleosome. Nucleosomes stack on each other to form a 30-nanometer scale fiber, which is then folded into successively larger structures.[7–10] Though *in vitro* structure probing techniques have produced evidence for regular organization up to the 30-nanometer scale,[11–13] experiments with greater than four nucleosomes have produced inconsistent results across a range of experimental conditions, and have generally been unable to produce a cohesive picture of chromatin structure at the 30 nanometer scale with molecular detail.[14,15]

Longer samples of DNA and histones, on the scale of several and tens of kilobases, were found to condense *in vitro* in low salt conditions into fibers roughly 30 nanometers wide by

Finch and Klug in 1976.[8] Though these fibers lacked the long-range crystalline order necessary for definitive structure determination through x-ray diffraction techniques, electron-microscope-assisted probing methods showed that the chromatin had arranged into densely packed, left-handed helical fibers, with a helical pitch of 11 nanometers and an approximate width of 30 nanometers. These fibers appeared to have approximately six nucleosomes per turn, and due to the low pitch angle of the helix, the authors argued that this structure represented a solenoidal arrangement of the fiber, where successive nucleosomes contact each other and spiral in an outer shell around the central axis. Several subsequent analyses corroborated Finch and Klug's conclusions. Different groups reported non-variance of the diameter of the fiber with DNA linker length, a unique feature of solenoid-like geometries, and the geometric difficulty of forming fibers at the highest end of the observed diameter range with alternative models.[16–19] The bulk of these studies entailed width and length analyses of fibers reconstituted *in vitro* under low-salt conditions and containing more than 12 nucleosomes.

In recent years, however, competing models of the so-called 30-nm fiber have emerged. These include the two-start zig-zag structure. *In vitro* studies in which the fiber diameter was found to correlate with linker length supported a model for the fiber in which linker DNA crosses through the central fiber axis in every successive nucleosome.[20–22] These results matched those predicted if the 30-nm fiber consisted of stacked tetranucleosomal units, similar to those reported by Schalch *et al.*,[13] which entails straight DNA between nucleosomes and contact between every other nucleosome ($i \pm 2$). More recent high-resolution cryo-electron microscopy studies of six- and twelve-nucleosome arrays have appeared to confirm that this zig-zag persists to larger nucleosome arrays.[23,24]

Discrepancies between these results and those supporting solenoidal structures have been suggested to arise from the choice of linker histone variant (H1 in zig-zags and H5 in solenoids) due to longer nucleosome repeat lengths (NRLs) promoting solenoid structures, and the ad-

dition of divalent cations.[24,25] However, it is not immediately clear which of these structures is most reflective of *in vivo* chromatin, since histone H1, divalent cations, and a broad spectrum of nucleosome repeat lengths are present in human nuclei..[26,27] Furthermore, some evidence suggests that 30 nanometer fibers may exhibit a heteromorphic mix of both structures, with some straight DNA linkers and some bent.[28,29] For these reasons, the structure and relevance of the 30-nanometer fiber has remained a subject of controversy.[23,30,31]

Further clouding inquiry into the structure of chromatin at the 30-nm scale are studies which suggest that, *in vivo*, no "30-nanometer fiber" exists at all. Multiple *in vivo* cryo-electron microscopy studies across several cell lines[18,19,32] revealed irregular fibers 20- to 30-nanometer-wide, in loose arrangements with mostly straight linker DNA and significantly lower linear density than *in vitro* arrays. This view has been further supported by modern cryo-electron microscopy tomography scans over large regions of nuclei, which indicate that the most frequent diameters of chromatin fibers *in vivo* are between 6 and 24 nanometers, and that these numbers are highly irregular.[31] As summarized by Maeshima *et al.*, the presence of 30-nanometer fibers may simply be an artifact of the dilute conditions of *in vitro* studies, which may prevent the interdigitation of adjacent fibers.[33,34] Furthermore, as summarized by Zhou *et al.*, crystallization methods such as vapor diffusion, which have produced tight zig-zag-like structures,[13,23] sample oligonucleosome arrays under artificially tight-packed conditions, thereby producing misleadingly dense arrays which are inconsistent with arrays' ability to cross-link in solution.[30]

The lack of an easily-defined well-ordered oligonucleosome structure constitutes a significant challenge to intuitive interpretation of the effects of biological forces upon local gene structure. If chromatin underwent a binary transition between two pseudo-crystalline helical structures in response to, for example, the addition of a particular epigenetic mark to histones, it would be straightforward to analyze these factors' effects on gene promoter structure, genetic disease structural pathology, etc. The emerging picture of oligonucleosome-scale

chromatin structure as shifting distributions of disordered structures, of varying diameters and helical arrangements, does not lend itself to such straightforward analysis. However, methods to characterize and quantify disorder are well-known within the field of computational materials modeling, offering a possibility to describe the distributions of disordered chromatin states by molecular simulation.

Most simulation work to date on oligonucleosome arrays has focused on fiber-like arrangements, typically using models that rely on approximations of mean electrostatic potentials to account for nucleosome-nucleosome and nucleosome-DNA interactions, and representing linking DNA as a worm-like chain.[35,36] These mesoscale models have been used to extract important trends in fiber organization. Grigoryev *et al.* described an increase in fiber uniformity, increased linear density, and frequency of i $\pm$ 2 contacts upon addition of a linker histone, implying an increase in zig-zag character during condensation. The authors also observed both an increase in i $\pm$ 1 and i $\pm$ 2 contacts upon addition of magnesium, implying both an increase in condensation and increased structural heterogeneity of fibers caused by increased linker DNA flexibility.[28] Perisic *et al.* qualitatively characterized the dependence of fiber structure on NRL and linker histone presence, concluding that three regimes exist of fiber stability: short NRL fibers ($\leq$ 182 bp) fold into tight, non-LH-dependent zig-zags with a narrow width (25nm), medium NRL fibers (191 $\leq$ NRL $\leq$ 209) fail to converge to consistent structures independent of their initialization conditions, but converge to zig-zags upon addition of linker histones, and long NRL fibers (218 $\leq$ NRL) fail to converge to history-independent structures regardless of initialization condition.[37] These results for the conditional instability of the solenoid structure are partially supported by simulations using a non-electrostatic mesoscale model parameterized to match high-resolution nucleosome geometry, which showed similarly convergence of fibers with NRL 169 bp to zig-zag-like arrangements and the difficulty of converging longer NRL arrays into regular fiber arrangements.[38]

Recent work exploring the range of stability of 30-nm fiber structures has also identified a range of conditions in which 30-nm helices are not stable. Simulated oligonucleosome arrays with multiple linker lengths exhibted significant polymorphism and a reduction of fiber regularity,[29] in keeping with theoretical findings which suggest that regular fibers are a unique feature of uniform-NRL arrays.[26,27] Variation in linker length by as few as 5bp has been shown experimentally to induce changes in nucleosomal linker number (effective DNA writhe) by as much as 0.5 turns, inducing significant shifts in fiber structure.[39] These experiments have also shown that changes in linker DNA rigidity resulting from the binding of non-histone proteins also disrupt fiber regularity.[40]

CHAPTER 2

AN INTRINSICALLY DISORDERED MODEL FOR CHROMATIN FIBERS PREDICTED BY MESOSCALE SIMULATION

2.1 Introduction

Closely-spaced arrays of nucleosomes exhibit short-range stacking interactions which induce an effective width along the contour of chromosomal DNA, producing what is generally termed the chromatin fiber. Chromatin fibers of varying dimensions and geometries have been observed forming *in vitro* via 2D and 3D electron microscopy,[8,24,25] X-ray crystallography,[13,23] and *in situ* via cryo-electron microscopy tomography.[31] These fibers are understood to be stabilized by direct internucleosomal contacts such as between H4 tails and nucleosomes' acidic patch, with the most frequent interactions occurring within 5 or fewer internucleosome spacings.[28,41,42] In the light of recent findings that probes of chromatin structure at this length-scale such as ATAC-Seq are strongly correlated with levels of gene expression,[43–45] a fundamental understanding of chromatin fiber structure is essential for elucidating the structure-function relationships inherent in gene expression.

Ideal, zero-entropy models have emerged for the structure of the chromatin fiber, and have enabled researchers to develop intuition regarding the physics underlying the fiber's formation and stability. Though many models have emerged, the two which have most successfully stood the test of time are the 1-start solenoid, with approximately six nucleosomes per helical turn joined by bent linker DNA,[8,16,25] and the 2-start zig-zag, with approximately two nucleosomes per turn joined by straight linker DNA.[21,23,24] Both of these models prescribed that fibers were dominated by specific internucleosome interactions, with the solenoid stabilized by 1-2 interactions and the zig-zag stabilized by 1-3 interactions. However, the relevance of these ideal structures to biological phenomena remains a topic of debate, es-

pecially as the dilute and low-salt conditions optimal for *in vitro* imaging are unlikely to represent the conditions of the nucleoplasm.[33,34]

Despite these models' conceptual simplicity, they have come under significant stress due to a growing body of contrary experimental evidence. 3D electron microscopy measurements have observed that chromatin fibers *in situ* appear to display a wide range of diameters primarily $< 22nm$,[31,46] inconsistent with the $\sim 30nm$ fiber widths observed *in vitro*.[16,25] Furthermore, variants on chromatin conformation capture methods such as Hi-CO and Micro-C have failed to detect the regularity of contact frequencies at the single nucleosome level which is demanded by such ideal fiber models.[47] EM-assisted nucleosome interaction capture (EMANIC) results also demonstrate that nucleosomes interact stochastically with up to at least 6 of their neighbors,[28] indicating that the network of nucleosome interactions is disordered even *in vitro*.

In this work, we use extensive mesoscale molecular dynamics simulations to demonstrate that although the dynamics of local nucleosome rearrangements are slowed by steric collisions, chromatin fibers broadly display smooth, nearly barrier-free free energy landscapes across collective variables which capture folding between the ideal solenoid and ideal zig-zag structures. Our findings quantitatively confirm previous conjectures that the ideal zig-zag structure is dramatically more thermodynamically stable than the ideal solenoid for NRLS 177, 187, 197 and 207, and that the addition of linker histones reverses this relationship. We also show that highly heterogeneous local interactions can lead to relatively consistent fiber structures at the population average level for each given NRL. Lastly, we present a Bayesian framework for the quantitative inference of NRL from *in silico* measurements of fiber parameters based on our previous modeling results.

Previous simulation studies utilizing mesoscale chromatin models have provided great insight into the nature of chromatin fiber structure and its variability with respect to NRL, linker histone model and geometry, nucleosome representation, and other considera-

tions.[36,38,48–51] However, prior attempts at characterizing the relative stability of chromatin fiber structures have generally stopped short of quantifying the free energy differences between the ideal solenoid, ideal zig-zag and disordered intermediate structures.[37,38] Other studies have described an increase in fiber uniformity, linear density, and an increased frequency of $i \pm 2$ contacts upon addition of linker histone.[28] Our work expands upon these findings to more extensively characterize the regime of disordered fiber structures thought to be relevant to *in vivo* chromatin.

2.2 Methods

We represent chromatin using the 1CPN coarse-grained oligonucleosome model,[52] which treats the nucleosome as a single rigid body with a complex internucleosome interaction potential parameterized against the electrostatic and histone tail-mediated inter-nucleosomal potential of the 3SPN.2C-AICG model.[53] DNA is treated as a charged, twistable worm-like chain with a bending persistence length of 50 nm and a twisting persistence length of 75 nm. Electrostatics are treated at the level of Oosawa-Manning theory, combining the Debye-Hückel linearization of the Poisson-Boltzman equation with a fractional reduction in surface charge for DNA bases in order to match the salt-dependence of DNA's persistence length.[52]

Precise quantification of the free energy difference between solenoid and zig-zag structures presents a significant computational challenge, due to the complexity of the highly ordered endpoint structures, and the long timescales ($> 10\,\mu s$, vs. a timestep of $60 fs$) over which the 1CPN model reaches equilibrium. Furthermore, it is not clear *a priori* that either of these crystalline structures are representative of the equilibrium distribution of states of biologically-relevant chromatin. To address both of these concerns, we have chosen to use two methods to analyze this system. First, parallel tempering is used to characterize the equilibrium distribution of structures. Second, an adaptation of the method of Maragliano *et al.* is used to calculate the free energy of oligonucleosome arrays along a transition path

between the solenoid and zig-zag explicitly defined in collective variable space.[54]

The first method accelerates sampling of the equilibrium distribution of states through parallel simulation of the system across a range of temperatures, augmented with temperature-exchange Monte Carlo moves, but is not capable of forcing chromatin into either of the ideal fiber structures alone.[52,55] These simulations are complemented by use of the second method, which, although originally developed for analyzing reaction pathways resulting from convergence of the Finite Temperature String method,[56,57] is able to determine the potential of mean force along any general transition path defined parametrically in multiple collective variables. Since the potential of mean force is constructed through parallel umbrella sampling simulations along the transition path, the method efficiently samples phase space and is capable of strongly biasing the system into even thermodynamically unstable structures through selection an appropriately large spring constant.

2.2.1 Simulation Details

The transition path free energy sampling method used here is implemented using the SSAGES software package for enhanced sampling.[58] For a set of N collective variables, the position of the system in collective variable space is denoted by

$$\boldsymbol{\theta}(\mathbf{r}) = (\theta_1(\mathbf{r}), ..., \theta_N(\mathbf{r})), \tag{2.1}$$

where $\mathbf{r} \in \mathbb{R}^n$ are the Cartesian coordinates of all n particles within the simulation, and $\theta_i(\mathbf{r})$ is the value of the i-th collective variable evaluated for the given configuration $\mathbf{r}$. A transition path is defined as a multivariate parametric function of the path variable $\alpha \in [0, 1]$, which maps this path variable onto corresponding values of collective variables

along the path, and is given by

$$\mathbf{z}(\alpha) = (z_1(\alpha), ..., z_N(\alpha)). \tag{2.2}$$

For this function to describe a reversible thermodynamic path between the solenoid and zig-zag structures, $\mathbf{z}(\alpha)$ must be continuous, differentiable, and chosen so that

$$\mathbf{z}(0) = \boldsymbol{\theta}(\mathbf{r}_{solenoid}) \tag{2.3}$$

$$\mathbf{z}(1) = \boldsymbol{\theta}(\mathbf{r}_{zig-zag}), \tag{2.4}$$

where $\mathbf{r}_{solenoid}$ and $\mathbf{r}_{zig-zag}$ are the Cartesian coordinates of the simulated system when in the ideal solenoid and ideal zig-zag structures, respectively. As in Maragliano *et al.*,[54] a bias is introduced to the potential of the form

$$U_{biased}(\mathbf{r}, \alpha) = U_{unbiased}(\mathbf{r}) + \sum_{j=1}^{N} \frac{k_j}{2}(\theta_j(\mathbf{r}) - z_j(\alpha))^2 \tag{2.5}$$

The potential of mean force along the chosen path is constructed through parallel simulation at multiple values of α, here chosen by Gauss-Legendre quadrature.[59] At all values of α, the mean force in the jth collective variable is calculated as

$$\left|\frac{\partial F(\mathbf{z})}{\partial z_j}\right|_{\mathbf{z}=\mathbf{z}(\alpha)} = k_j \int_0^T (z_j(\alpha) - \theta_j(\mathbf{r}(t)))dt, \tag{2.6}$$

where $F(\mathbf{z})$ represents the Helmholtz free energy of the system at the position z in collective variable-space, and T represents the total simulation time for any given simulation. When this method is applied to a path returned by convergence of the Finite Temperature String method, it can be assumed that this mean force is parallel to the transition path.[54] For any other path $\mathbf{z}(\alpha)$, this assumption is violated, and the reconstructed free energy along the

path must be corrected by integrating only the component of the mean force which is parallel to the path. This can be written as

$$F(\mathbf{z}(\alpha)) - F(\mathbf{z}(\alpha_0)) = \int_{\alpha_0}^{\alpha} \frac{\partial F(\mathbf{z}(\alpha'))}{\partial \alpha'} d\alpha' = \sum_{j=1}^{N} \int_{\alpha_0}^{\alpha} \frac{\partial z_j(\alpha')}{\partial \alpha'} \cdot \left(\frac{\partial F(\mathbf{z})}{\partial z_j}\bigg|_{\mathbf{z}=\mathbf{z}(\alpha')} \right) d\alpha' \tag{2.7}$$

We now define the set of collective variables $\boldsymbol{\theta}(\mathbf{r})$, and the transition path in those variables $\mathbf{z}(\alpha)$. Here we use the set of 22 angles between every ith, i+1st, and i+2nd nucleosome's center of mass (henceforth referred to as 'trinucleosome angles'), as well as the radius of gyration (R_g) of the entire fiber. Variations in these trinucleosome angles control the possibility of forming i $\pm$ 2 contacts, characteristic of the difference between the zig-zag and solenoid structures,[28] while varying R_g induces fibers to condense or decondense, thereby improving sampling.

To construct a transition path $\mathbf{z}(\alpha)$ in collective variable space which avoids steric clashes, we find it convenient to first construct the full structures of fibers intermediate between the ideal solenoid and zig-zag endpoints (i.e., $\mathbf{r}(\alpha) \in \mathbb{R}^n$). Once these intermediate structures are defined, $\mathbf{z}(\alpha)$ is determined straightforwardly. $\boldsymbol{\theta}(\mathbf{r}(\alpha))$ is calculated for each intermediate structure, and $\mathbf{z}(\alpha)$ is calculated by a cubic spline between each value of $\boldsymbol{\theta}(\mathbf{r}(\alpha))$ for all $\alpha \in [0, 1]$.

To actually construct these intermediate fibers, we note that both the ideal solenoid and zig-zag structures are only specific cases of the much larger class of helical fibers which can be geometrically defined by several parameters: helical pitch, helical diameter, azimuthal rotation angle (also referred to as helical angle), nucleosome tilt, and path of the linker DNA[16,24] (Fig. 2.7). Intermediate fibers can be constructed by choosing parameters between those of the ideal solenoid and ideal zig-zag. For simplicity, we choose to linearly

interpolate these helical parameters between the ideal solenoid ($\alpha = 0$) and zig-zag ($\alpha = 1$) (Fig. 2.1). For any choice of helical parameters calculated by this interpolation, we use a custom modification of the 1CPN initialization Python script to calculate the positions and orientations of all nucleosomes in the defined fiber and construct an input LAMMPS data file. LAMMPS simulations are conducted for these input structures for zero timesteps, to verify that no steric clashes occur. The linear interpolation in helical pitch was replaced with a downwards-parabolic interpolation to reduce steric hindrance.

We consider several nucleosome repeat lengths, spanning from 177 to 207 base pairs. To supplement our analysis, unbiased simulations of identical fibers were performed using replica exchange, to sample the equilibrium distribution of states of these fibers. In addition, we tested the effect of adding linker histones to the fiber for each of these repeat lengths. Linker histones are known to promote fiber condensation *in vitro* and alter linker DNA geometry;[25,49,60] they have a qualitative effect on the geometry of entire fibers.[61] For computational expedience in these simulations, the linker histone was represented differently than in prior work using the 1CPN model. The globular head of the linker histone was represented as a single charged bead, fixed to the dyad, interacting with linker DNA via electrostatics and an excluded-volume term. The effect of the C-terminal tail was included by altering the parameters of the anisotropic gaussian interaction of the nucleosome dyad with the linker DNA, which was originally parameterized to account for the H3 tail-DNA interactions. This representation was parameterized to optimally reproduce the force-extension characteristics of the 1CPN dinucleosome bound by the linker histone variant H1.4, using the electrostatic model of Luque *et al.*.[52,60] Though some details regarding linker DNA interactions are lost with this representation, quantities averaged at the level of nucleosome center of mass positions are largely unaffected.

2.3 Results

2.3.1 Chromatin exhibits NRL-dependent, LH-dependent disordered fiber structures

The free energy landscape resulting from our unbiased simulations is shown in Fig. 2.3 for several choices of uniform NRL that are abundant in human cells.[26] These free energy surfaces can be visually classified into one of two categories. The first category, relevant to NRLs 197 and 207 bp, consists of free energy surfaces with only one significant free energy minimum in these collective variables. The second category, relevant to NRLs 167, 177 and 187, exhibits two or more free energy minima. Notably, none of the free energy minima for any of the NRLs sampled without linker histone contained either of the ideal fiber structures. Instead, every free energy minimum lies at a significantly higher radius of gyration and, when visualized, represented a disordered fiber state with transient contacts (Figure 3 insets). For nearly all NRLs sampled here and in both categories, the free energy minima lie in broad basins spanning a large range of trinucleosome angles, typically on the order of 60 degrees with less than 3 kT variation in free energy. This variation is much larger than would be expected if a single, uniform fiber were the most stable configuration, and in fact is more than 50% of the difference between the ideal zig-zag and ideal solenoid themselves (approximately 100 degrees). As such, we conclude that these simulations do not support the existence of *any* uniform fiber structure, and instead support the perspective that these chromatin fibers exist in an ensemble of disordered structures which vary under thermal motion.

The free energy surface of the fiber with NRL 187 and with 100% LH saturation using an electrostatic model for LH 1.4[36] is shown in Figure 2.3f. A free energy minimum appears at a lower radius of gyration than for any of the fibers simulated without linker histone; the corresponding free energy basin clearly includes the collective variable values of the ideal solenoid fiber. However, the free energy basin corresponding to this minimum is extremely

broad in trinucleosome angle, spanning approximately 25-160 degrees with less than $3k_BT$ variation. This behavior is commensurate with experiments which have demonstrated that *in vitro*, linker histone addition induces significant fiber condensation.[16,25,62] However, although these structures are condensed with nucleosomes primarily along the outside of the fibers and DNA located centrally and tightly bound to the LH tails, they are not uniform fibers. Instead, even the LH-bearing fibers display a wide range of trinucleosome angles, indicative of a condensed but heteromorphic fiber. Since these fibers have nearly the same linear density and width as the solenoids, they are consistent with experimental electron microscopy results which to-date have been taken as indicative of a solenoid-like geometry, even though these simulations do not recapitulate an ideal solenoid-like geometry.[25]

While our replica-exchange simulations describe the thermodynamically stable ensemble of chromatin structures, and lend support to the theory of chromatin as intrinsically disordered, they do not answer key questions. First, if chromatin were to be found in a 30-nm-wide fiber, what would the most likely structure be? The free energy differences between the two ideal fiber structures, obtained through the path umbrella sampling method, are shown in Table 1. For fibers with any NRL without linker histones, the ideal zig-zag is more stable than the ideal solenoid, with a free energy difference greater than 75 kT for every NRL. By contrast, the ideal solenoid is found to be more stable than the ideal zig-zag in all fibers with any NRL when linker histones are present, also by free energy differences on the order of 100 kT. Since the linker histones' positive charge is expected to promote more condensed structures where DNA is highly compact, and the ideal solenoid structure is marginally more compact in linear density and radius of gyration than the ideal zig-zag structure, this result is consistent with our physical intuition.

The full free energy curves resulting from umbrella sampling (Fig. 2.2a) with no linker histone exhibit monotonic decreases in free energy from each ideal structure, before leveling out in the broad free energy well identified by the previous unbiased sampling. Note that

Table 2.1: Free energy of the solenoid to zig-zag transition by fiber type (in kT). Negative entries imply the zig-zag is more stable than the solenoid for the corresponding combination of parameters.

NRL (bp)	**177**	**187**	**197**	**207**
-LH	-185.0	-75.7	-113.1	-150.0
+LH	164.4	582.0	715.6	310.5

the path variable, α, depends nonlinearly on the radius of gyration (R_g) and trinucleosome angle; all minima in Figure 2.2a correspond to slightly different CV positions within the same free energy basins described in Figure 3. Though chromatin is not constrained in biological systems to fold along the path used here for umbrella sampling, these results demonstrate that there exists at least one folding path from each ideal fiber structure to a disordered structure along which the free energy is always decreasing. As such, chromatin would not likely encounter any free energy barrier to unfolding from either of these proposed ideal structures. This absence of any energetic barriers to the decondensation of condensed chromatin fiber structures suggests that chromatin is unlikely to remain kinetically trapped in these condensed structures for any appreciable time.

2.3.2 Helical Analysis of Disordered Chromatin

Chromatin which lies in the convex free energy well described in the previous section does not adopt a well-ordered 30-nm fiber conformation. Instead, its structure dynamically changes throughout the simulations, sampling a wide range of configurations which cannot be trivially classified as belonging to either ideal fiber structure. To provide a quantitative analysis of these disordered fibers that is intuitively coherent with historical perspectives of chromatin fiber structure, we calculated the distributions of several structural quantities which are commonly used to define the ideal structures. Since the ideal fiber structures are typically defined in terms of helical parameters, including helical pitch, gyre, and diameter,[16] we

calculated the distributions of these same properties of chromatin fibers from our unbiased simulations for all NRLs. Further explanation of these structural properties is given in Figure 2.7.

Though helical parameters are sufficient to fully define an ideal helical fiber structure, their definition is not clear in any thermal simulation which does not have perfect axis-symmetry. In previous studies, the fiber axis was approximated by parametric polynomial fitting, which allows approximate calculation of fiber length and width.[51] However, it is unclear *a priori* what order of polynomial must be used for each simulated fiber: under-fitting fails to capture the curvature of the fiber, while over-fitting causes the estimated helical axis to reflect the sub-fiber-scale structural variations, capturing zig-zagging of nucleosomes as curvature of the helical axis.

We provide here an alternative method for estimating the fiber helical axis and for calculating the fibers' helical parameters, with minimal assumptions. Working backwards from our goal: a continuous, differentiable interpolating curve can be generated from any set of points that estimate the central helical axis using B-splines. B-splines are piecewise-smooth curves which interpolate all points (knots) in an ordered set by a set of cubic curves, and are both continuous and differentiable at the knots. Higher-order splines could be constructed that are continuous in higher-order derivatives, but cubic splines represent the lowest-order polynomial fit which still returns a curve with the desirable properties of being continuous and differentiable everywhere. As such, if we had a set of points that lie on the helical axis of a simulated chromatin fiber, we could construct a minimal approximation of the axis itself that is continuous and differentiable everywhere by cubic splines.

To create such a set of points along the helical axis, we posit that, for an ideal cylindrical geometry with a constant helical gyre, points along the central axis may be generated by taking the average of some number N of consecutive nucleosomes' centers of mass. As shown in SI1, $N = 6$ is the minimum number of nucleosomes required to accurately estimate the

helical angle of both the ideal solenoid and ideal zig-zag structures, and creates accurate estimates of the helical axis for fibers with any value of helical gyre greater than that of the solenoid (66.4 degrees).

For a random sub-sampling of timesteps taken from the previous unbiased simulations, the helical axis was calculated via this method. The three helical parameters identified previously – pitch, gyre, and diameter – were calculated as follows. The position along the helical axis closest to each nucleosome was tabulated for every nucleosome. Helical pitch between any two consecutive nucleosomes was calculated as the length of the spline between these axis-points. Helical diameter was calculated as twice the length of the normal vector between each nucleosome and its axis-point, plus 7.5 nm to account for the approximate remaining width of the nucleosome beyond its center of mass. Lastly, helical gyre was calculated as the dihedral angle between consecutive normal vectors along the axis. Because sharp turns in the helical axis sometimes occurred, resulting in poor estimation of helical gyre, these normal vectors were corrected by the inverse of the net rotation of the fiber axis tangent between the (i+1)st and ith nucleosome centers (SI1). This procedure returns uniform helical angles of 66 and 193 degrees for the ideal solenoid and ideal zig-zag, respectively, nearly identical to those used to define the ideal structures.

The equilibrium distribution of helical parameters calculated in this manner for simulated chromatin fibers is shown in Figure 2.4. It is apparent from these distributions of both helical parameters that the 1CPN model does not predict any regular helical arrangement. Each distribution of angles spans a range from 50°-200°, including both the 66.4° of the ideal solenoid and the 193° of the ideal zig-zag, whereas regular fiber arrangements would exhibit a narrow distribution peaked around their mean values. Furthermore, the distribution of helical angles is extremely similar across all NRLs, implying very little dependence of solenoid- or zig-zag-character on NRL. Additionally, in the absence of linker histone, the distributions of fiber diameter for all NRLs is nearly gaussian, with standard deviations as

large as 10 nanometers, which is inconsistent with a regular fiber arrangement.

The average diameter of these simulated fibers without linker histones increases linearly with linker length ($R^2 = 0.96$), with a best-fit slope of 0.26 nm per bp of linker DNA (Figure 2.4). It is difficult to directly compare this result to experiments, since it is known that nucleosome arrays reconstituted *in vitro* without linker histone do not form well-defined fibers for all NRLs.[25] However, our data matches quantitatively with the consensus linear relationship between fiber diameter and linker length reported by Athey *et al.* from a review of experimental electron microscopy studies of embedded and sectioned chromatin.[21] The similarity between our results and this body of experimental data lends further support to the general applicability of our analysis, though it is unclear if this linear relationship, originally reported for non-mammalian chromatin, extends to human chromatin *in vivo*.

2.3.3 Comparison to in situ analyses of human chromatin

Unlike the uniform-NRL nucleosome arrays simulated here, biological chromatin exhibits a wide distribution of NRLs.[26] With these helical parameters calculated for all NRL values from 157 to 207, it is possible to make predictions regarding the *in vivo* distribution of helical parameters across a range of NRLs. Although the full human genome exhibits a wide distribution of NRLs,[26] recent nucleosome positioning studies have determined that local regions of the genome have highly uniform NRLs.[63] As such, the ensembles of fiber diameters determined by *in situ* measurement would be expected to be similar to a weighted sum of the individual diameter distributions for uniform arrays of each NRL, weighted by the biological abundance of that NRL. The distributions of fiber parameters calculated previously for each NRL were averaged with weights equal to the estimated *in vivo* abundances of NRLs[26] to construct an ideal fiber diameter distribution in the absence of linker histone. This combined distribution is largely independent of the exact choice of estimated NRL distribution.

The resulting free-energy landscape of helical parameters predicted for biological chro-

matin is shown in Figure 2.5. Significantly, this distribution of predicted *in vivo* fiber diameters (95% CI [11.3 nm, 29.9 nm]) compares favorably with experimental measurements of *in vivo* chromatin fiber diameters at low chromatin volume contents measured by ChromEMT (95% CI [5 nm, 24 nm]).[31] Furthermore, this distribution predicts no energy barriers between locally solenoid-like and zig-zag-like structures, and no local maxima between their corresponding helical angles. The calculated free energy difference between the ideal solenoid and zig-zag structures from this analysis is approximately 2.1 kT per nucleosome. This number compares well to that calculated using umbrella sampling (Table 2.1) with no linker histones, predicting that the zig-zag is 50.4 kT more stable than the solenoid for 24-nucleosome arrays. This per-nucleosome free energy difference is easily overcome thermally, and in the absence of an energy barrier, our results imply that biological chromatin fibers freely and regularly undergo dynamic transitions between locally zig-zag-like and locally solenoid-like structures. This is consistent with prior results, which predict heteromorphic fibers,[28] and supports the claim that fiber-scale chromatin structure is inherently dynamic, and cannot be described by a static set of structure factors.

Since our results predict that chromatin fibers' physical dimensions vary with microscopic structural parameters, they also allow us to reverse this relationship: given a particular experimental measurement of fiber dimensions, e.g. from electron microscopy, we can predict the most likely local NRL of the fiber which would have produced that measurement. We apply the method of maximum a posteriori estimation to predict the most likely NRL of the chromatin fiber This predicted variance of fiber physical dimensions with microscopic structural parameters also allows us to reverse this relationship: given a particular experimental measurement of fiber dimensions, e.g. from electron microscopy, we can predict the most likely local NRL of the fiber which would have produced that measurement. We apply the method of maximum likelihood estimation to predict the NRL most likely to produce fibers of specific values of fiber linear density (nucleosomes per 11nm along the central axis)

and diameter. In this method, the equilibrium probability distribution of fiber structures for every NRL is first weighted by the prior probability of that NRL occurring *in vivo*,[26] as is done in Figure 2.5. Then, the weighted likelihood distribution with highest value is determined for each pair of linear density and fiber diameter. The results of this process are shown in Figure 2.6.

Several trends can be inferred from these results. Fibers with moderate linear density of approximately 2.3 nucleosomes per nm, but with small diameters less than 15nm, primarily exhibit short NRLs less than 170bp. Fibers with both a wide diameter greater than 20nm but with low linear density, 1.75 nucleosomes per nm, predominantly exhibit long NRLs greater than 195bp. Other combinations of fiber geometry exhibit NRLs between 175bp and 190bp. These intermediate NRLs are predicted as most likely for a very wide range of measured structures. This can be understood from the wide distribution of states for these NRLs; our model predicts these fibers to be highly heterogeneous with wide probability distributions, while fibers with most other linker lengths are restricted to narrower ranges of structures which strongly indicate those linker lengths' presence, and are highly unlikely to take on extreme values of linear density or diameter (as can be observed from Figure 2.4). Note, however, that it is unclear if the direct application of this analysis to experimental data is likely to produce meaningful NRL distributions; this analysis is limited by the assumptions of equivalence of local fiber structure between perfectly uniform-NRL and non-uniform-NRL oligonucleosome arrays *in vivo*, and that no LH is bound to the fiber. Improvements could be made upon this analysis by including linker histone effects, which would require accurate experimental measurements of the concentration of bound linker histone in the nucleus, as well as LH binding energies to predict the prior probability of LH being bound to a fiber. Additionally, accurate determination of hyperlocal NRL distributions, such as the dependence of sequential linker lengths on each other, would allow accurate characterization of non-uniform-NRL fiber structures. In order for these results to be scientifically meaningful,

the results described here must be consistent with the established body of experimental data regarding chromatin conformation. To ensure that this work can aid the broader scientific community in interpreting chromatin's structural characteristics, we will directly address key experimental findings which our model may lend valuable insight into. Conveniently, this will highlight common pitfalls which limit the generality of experiments performed under a variety of conditions.

2.4 Discussion

In this study, the free energy landscape of short chromatin fibers with respect to the ideal zig-zag and solenoid conformations were examined. Recent work has challenged both the physiological relevance and thermodynamic stability of these ideal structures.[28,30,31,33] To directly address questions of their stability, we designed a suite of enhanced sampling molecular dynamics simulations to explictly quantify the relative free energy of both ideal structures. Furthermore, we characterized the structural variation within the range of disordered, fiber-like states stable at equilibrium.

The folding path between the ideal zig-zag and solenoid depicted in Fig. **??** represents a path through structural space which sufficiently constrains the molecular degrees of freedom of the fiber to enforce continuous folding between the endpoints. Previous computational studies have suggested that disordered or heteromorphic fibers are preferred over ideal configurations, with vanishingly few demonstrating a thermodynamically stable solenoid structure.[36,38,48–51] Using a model which was designed to reproduce the free energy characteristics of internucleosome interactions, we have determined that the ideal solenoid is less thermodynamically stable than the ideal zig-zag by $3.2 - 7.7 k_B T$ per nucleosome in the absence of linker histones.

2.4.1 Consistency with Experimental Results

In order for these results to be scientifically meaningful, the results described here must be consistent with the established body of experimental data regarding chromatin conformation. Previous experimental investigations have established that chromatin compaction is dependent upon approximately three critical parameters: linker histone stoichiometry, NRL, and salt. The stoichiometry of the linker histone is the dominant factor, greatly increasing chromatin compaction *in vitro* regardless of other factors.[25,28,38,62] Subsequently, NRL determines the range of feasible fiber structures, with short NRLs enforcing an upper bound on fiber diameters and linear density in the presence of LH *in vitro*.[16,25,28,37] However, *in situ* cryo-electron microscopy tomography studies have not confirmed any binary of chromatin fiber diameters,[31,46] and experiments on *Thyone* and *Necturus* chromatin samples without defined linker histone stoichiometry revealed a systematic, linear relationship between NRL and fiber diameter.[21,22]

Our results are in total consistent with these experimental findings. We do not detect condensed fibers in the absence of linker histones, and, at least for NRL 187, the 1CPN model predicts a highly dense fiber with diameter of approximately 25 nm. However, we do not find evidence supporting that these 30-nm fibers are uniform in helical ordering (i.e. are near-ideal solenoids), finding instead that they exhibit a broad range of trinucleosome angles inconsistent with a uniform solenoid model, as proposed by Routh *et al.*. Furthermore, our findings match surprisingly well fiber diameter measurements of *Thyone* and *Necturus* chromatin, with a linear increase in fiber diameter of $0.23 nm/bp$, compared to previously reported trends of 0.23[22] and 0.31.[21] The average trinucleosome angles reported here are also commensurate with those found for the electrostatic chromatin model by Perisic *et al.*, although we report significantly larger variability of this metric than observed in that study.[37] This may be due to the explicit parametrization of the 1CPN model to quantitatively account for the free energy of nucleosome breathing, absent in the electrostatic model.[52]

Extensive equilibrium simulations of chromatin fibers with LH were not feasible due to the computational expense of this model, although it has already been demonstrated that the 1CPN model recaptures the salt-dependent compaction of chromatin with and without linker histone.[52]

2.4.2 Chromatin is pliable and dynamic, with large conformation changes inducible by small added forces

The free energy surfaces shown in Figure 2.3 imply that chromatin fibers can be deformed significantly without significant expenditure of free energy. For example, only approximately 4kT of free energy would be required to increase the NRL 187 bp, 24 nucleosome fiber's radius of gyration from the 18 nm well to the 32 nm well, an increase of approximately 80%. With the addition of linker histones, this deformability is maintained or even increased: the R_g can be increased from 12 nm to 25 nm by a 4kT deformation, over a 100% increase in polymer size. These small quantities of energy are well within that achievable by the binding of single proteins to chromatin: for example, heterochromatin protein 1's (HP1's) binding affinity to nucleosome surfaces has been reported as approximately -1.52 kT, and two bound HP1 proteins can dimerize with an energy of -4kT.[64,65] Both this high deformability and the intrinsic disorder of the fibers at equilibrium are results of the wide, mostly flat character of the minima of chromatin's free energy surface; the restoring forces driving chromatin to its most stable configuration are weak enough to be overcome both by thermal fluctuations and small applied forces.

This perspective of chromatin as easily deformed is consistent with the insights gained from single-molecule force microscopy experiments.[66–68] Significant experimental efforts have characterized the mechanical response of chromatin fibers to high stress ($>$ 6 pN), under which it has been suggested that individual nucleosomes become unstable and fragment. Multiple authors have also identified a low-force regime (0.5-4 pN) in which the

force-extension curve is linear, which has been associated with the separation of individual inter-nucleosome contacts. Our free energy surfaces, obtained at 0 pN of tensile force, are more relevant to the very-low force regime (< 1 pN, and < 0.5 pN especially) in which experimental data are more scarce, but available. Notably, several studies that reported results in this very-low-force regime found extension of chromatin fibers with 15-25 nucleosomes on the order of tens of nanometers to hundreds of nanometers under applied forces of less than 0.4 pN. There are multiple possible explanations for this phenomenon, and it has been suggested that below 0.5 pN the chromatin fiber may not be fully aligned with the applied force, or that rotational fluctuations and entropic contributions may be present.[68] However, this finding is also at least consistent with the disordered chromatin fiber model described by the free energy surfaces in **Figure 3**, which is expected to exhibit linear extension of up to 100% with a total applied work of the same order of magnitude.

2.5 Conclusions

The results presented here are consistent with recent proposals that the chromatin fiber is dynamic *in vitro*. Our simulations suggest that highly uniform configurations such as the ideal zig-zag or ideal solenoid fibers are not thermodynamically stable, but that the addition of linker histones stabilizes a condensed, non-uniform fiber that would be indistinguishable from these ideal structures under (non-tomographic) electron microscopy. The disordered ensemble of structures described by our simulations also recapitulates the trends of chromatin fiber diameters with NRL determined experimentally for chromatin *in situ* by electron microscopy. These disordered fibers contain mostly straight linker DNA and predominantly $i \pm 2$ and $i \pm 3$ transient contacts, but are highly deformable under thermal motion, unlike previously proposed ideal fiber structures.

We have provided a quantitative thermodynamic description of the relative stability of ideal fiber structures, finding that the zig-zag fiber is more stable than the solenoid

in the absence of linker histone, and that this relationship is reversed in the presence of linker histone. However, both our and others' results suggest that this is not biologically relevant; fiber uniformity is not observed *in vivo*, and is not expected to be stable under these simulations.

We find that uniform-NRL oligomers do exhibit multiple free-energy minima for several NRLs, but that none of these metastable states could serve as a kinetic trap restricting those oligomers to an ideal-fiber-like state. Furthermore, we confirm that for several biologically relevant NRLs, there exists at least one transition path from both the ideal zig-zag and the ideal solenoid structures to the disordered miminum-free-energy structures along which the free energy monotonically decreases, lending further support to the interpretation that chromatin is unlikely to remain kinetically trapped in an ideal-fiber-like structure.

Our simulations also provide a useful platform for understanding the nature of chromatin structure explicitly *in vivo*. Previous mesoscale simulations have largely recapitulated measured *in vitro* "30-nanometer" chromatin fiber diameters, but past models have not been shown to be consistent with ChromEMT experiments, which do not detect such thick fibers. By contrast, our mesoscale model predicts a disordered ensemble of chromatin structures which nearly perfectly matches the NRL-dependent distribution of fiber diameters measured *in situ*, for all NRLs from 167 bp through 207 bp. Though our model is consistent with recent experimental results that show the zig-zag structure to be the most stable of condensed fiber structures - the zig-zag is more stable than the solenoid, our model also indicates that this condensed structure is energetically unfavorable by tens of kT compared to disordered structures.

The disordered chromatin structures described here may be thought of as analogous to a "native state" structure, in that chromatin which is not disrupted by other forces may sample the ensemble of structures described here computationally. When chromatin is influenced by an external event, such as through protein binding or an epigenetic alteration,

chromatin would begin to sample a slightly altered ensemble of structures, with this new ensemble dependent upon the nature of the external event. This manner of conceptualizing chromatin, as sampling ensembles of disordered structures, may provide insights into other complex biological phenomena in the future.

2.6 Figures

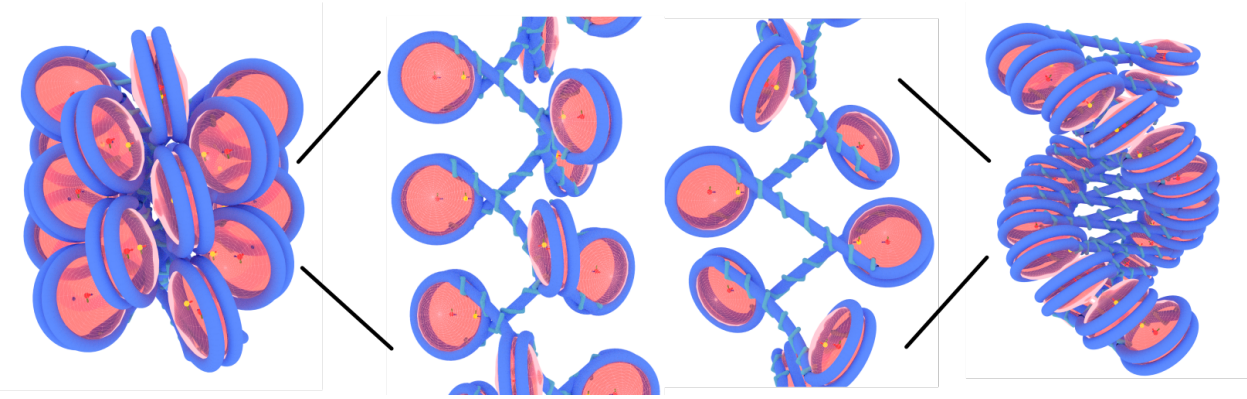

Figure 2.1: The range of reference structures used for umbrella sampling. **Left:** the ideal solenoid. Linker DNAs follow a parabolic path between their ith nucleosome exiting site and their i+1th nucleosome entering site. **Center left:** the intermediate structure at $\alpha = 0.31$. The helical pitch has been increased to reduce steric clashes, and the helical angle has significantly increased. **Center right:** the intermediate structure at $\alpha = 0.62$. The helical pitch is now decreasing to drive fiber condensation. The nucleosomes are now primarily tilted facing in the left-handed-helix direction, as is the case in the zig-zag and not the solenoid. Linker DNAs are nearlly fully extended. **Right:** the ideal zig-zag. Linker DNAs are straight and connect two opposing, tightly-contacting stacks of nucleosomes.

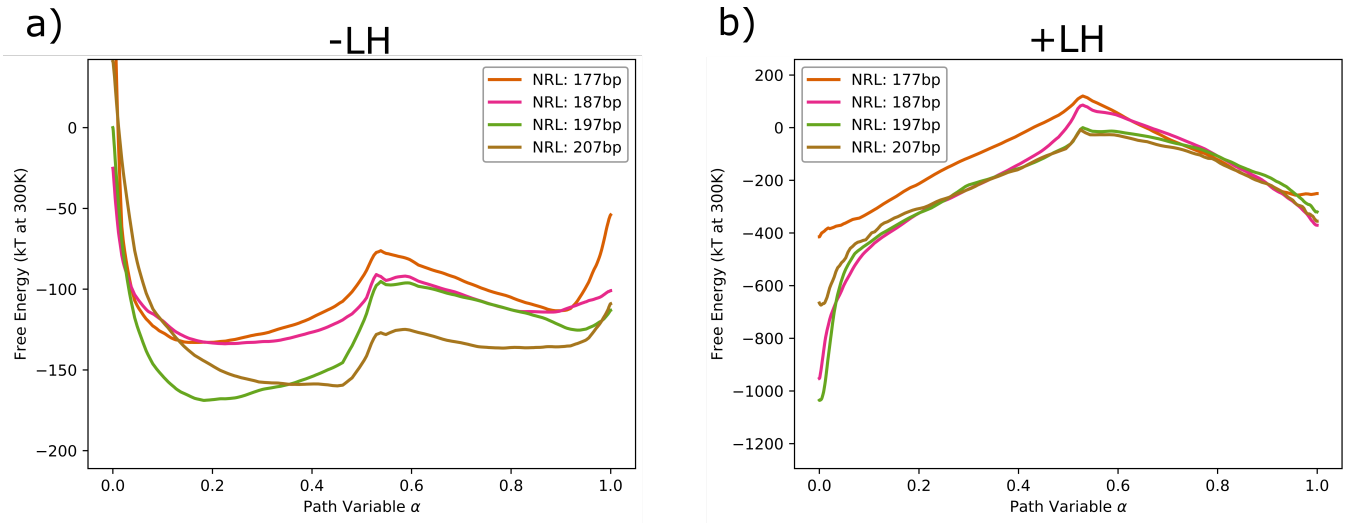

Figure 2.2: **a)** The free energy along the transition path for LH-free arrays. The central peak in each curve is present at the value of the path parameter α at which the R_g hits a maximum. Vertical alignment of each curve is performed for visual effect, and should not be interpreted physically. **b)** The free energy along the reaction coordinate for LH-bound arrays. The central peak is still present in these curves, but occurs at a much higher free energy, hundreds of kT higher than either endpoint. Additionally, the free energy of the fibers decreases monotonically as the transition path approaches either endpoint.

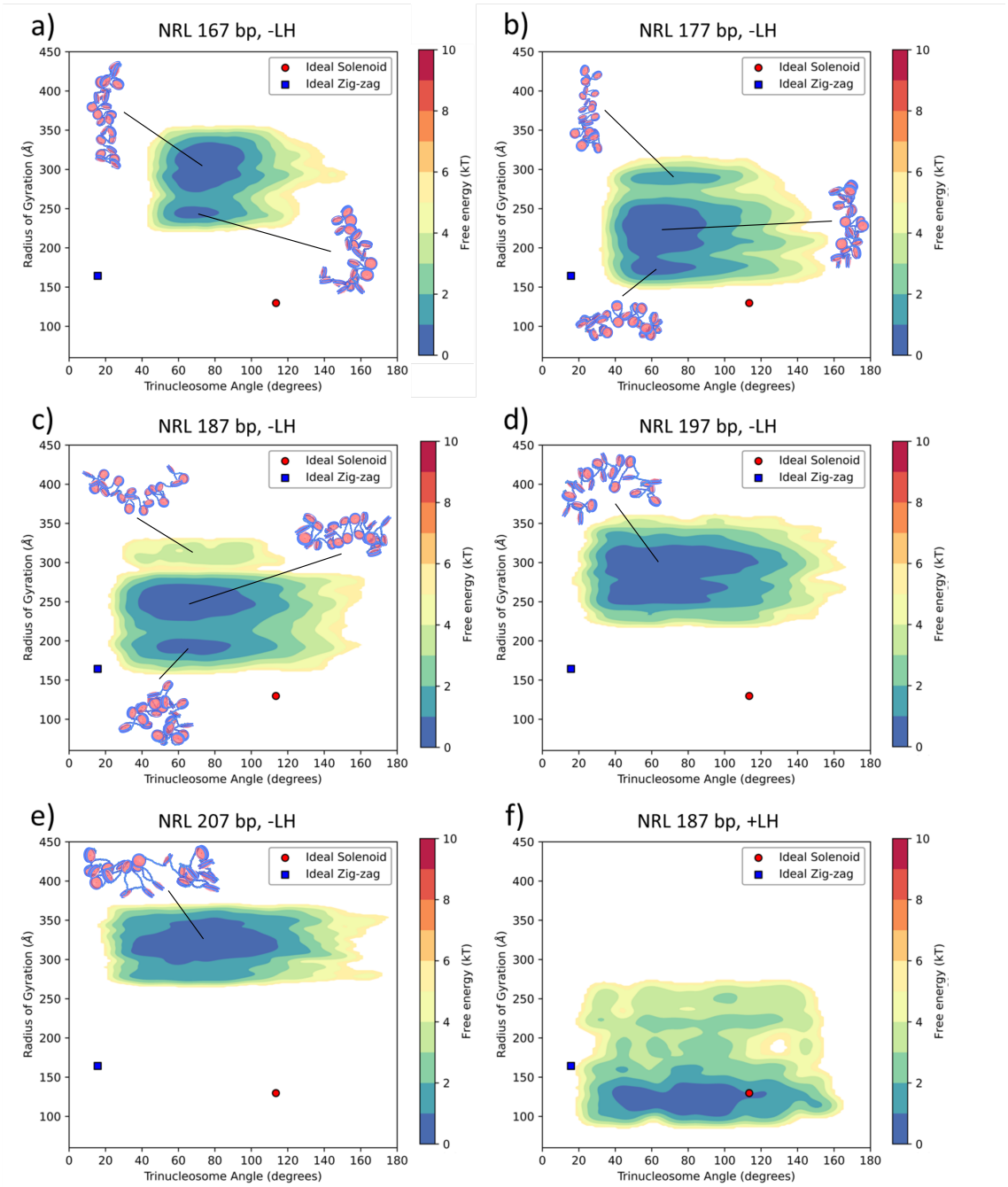

Figure 2.3: The free energy surfaces of 24-nucleosome fibers, obtained by unbiased REMD. **a-e)** The free energy surfaces corresponding fibers with uniform NRL, 167bp-207bp, without linker histone. Insets are visualizations of representative timesteps corresponding to each local minimum of free energy connected via solid black lines. The red circle and blue square represent the collective variable values of the ideal solenoid and ideal zig-zag structures, respectively.[16,24] **f)** The free energy surface correspond to NRL 187bp with linker histone. As discussed in the text, this surface is significantly more rough than those for fibers without LH, and is the only one to include Rg values as low as either of the ideal fiber structures.

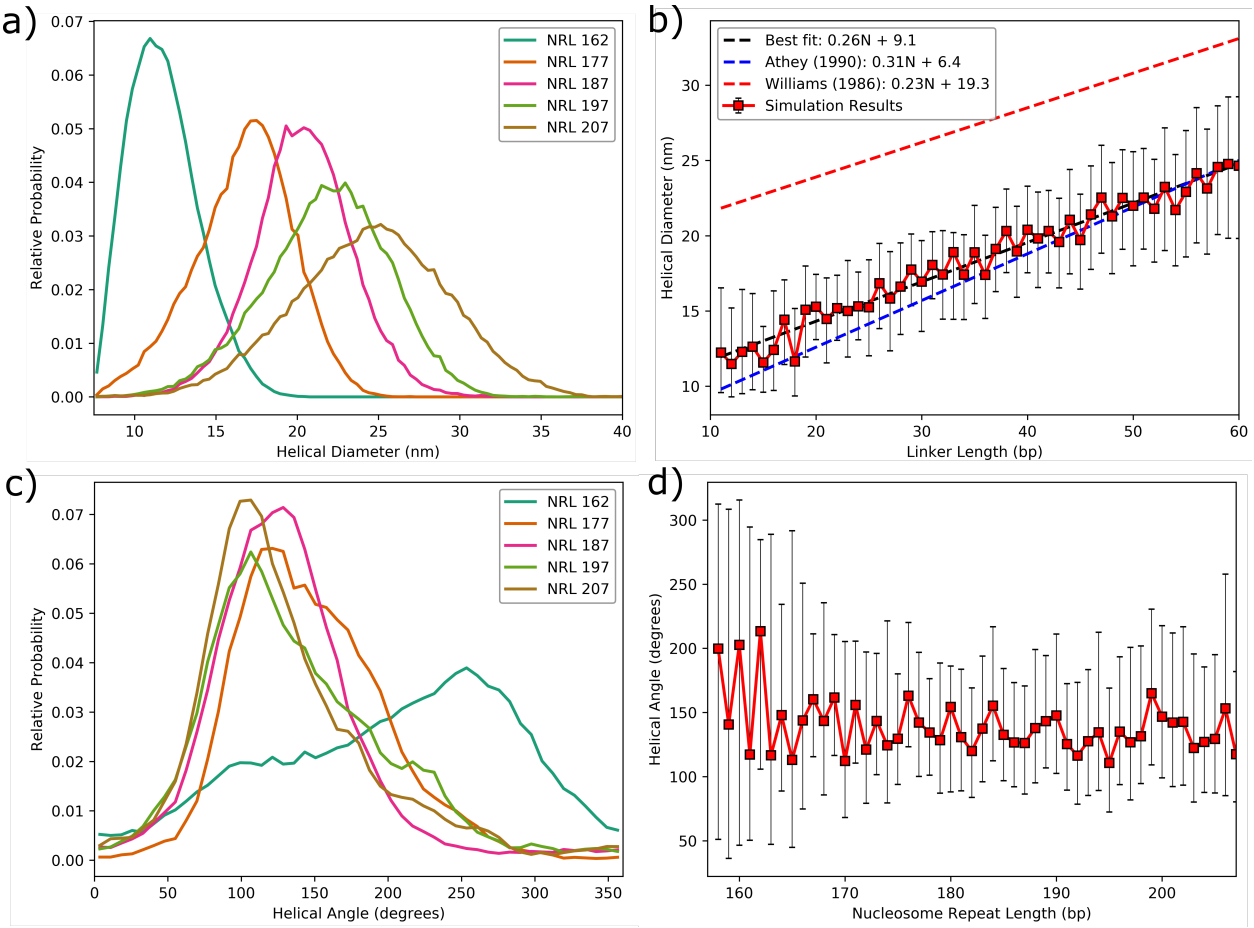

Figure 2.4: **a)** Fiber diameter distributions for several NRLs for LH-free fibers. Fibers of each NRL exhibit highly gaussian distributions of diameters, with longer DNA linkers inducing wider fibers and less uniform distributions (larger variance in each distribution). **b)** Variation of mean fiber diameter with linker length. Taken together, the average diameter for fibers of each NRL varies linearly with linker length ($R^2 = 0.96$), with a best-fit slope of 0.26 nm per bp of linker DNA. Our data matches quantitatively with the linear relationship calculated by Athey *et al.* from a review of EM experimental studies embedded and sectioned chromatin.[21] An alternative relationship determined by Williams *et al.* from experimental x-ray scattering data is also shown.[22] Although the constants of these empirical fits differ significantly, their slopes, which are characteristic of the specific conformation of the chromatin fiber, is well-reproduced by our simulations. Error bars on simulation data represent one standard deviation of the equilibrium distribution. **c,d)** Helical angle distributions for several NRLs for LH-free fibers. Fibers of nearly all NRLs above 163 bp show a tall peak near 120^o, indicating a mode of 3 nucleosomes per turn in a left-handed helix. With NRLs below 163 bp, fibers exhibit a bimodal distribution of helical angles, with a peak near 240^o and a shoulder at 120^o. This can be interpreted as a mixture of left-handed and right-handed helices, both with nearly 3 nucleosomes per turn. Error bars again represent one standard deviation.

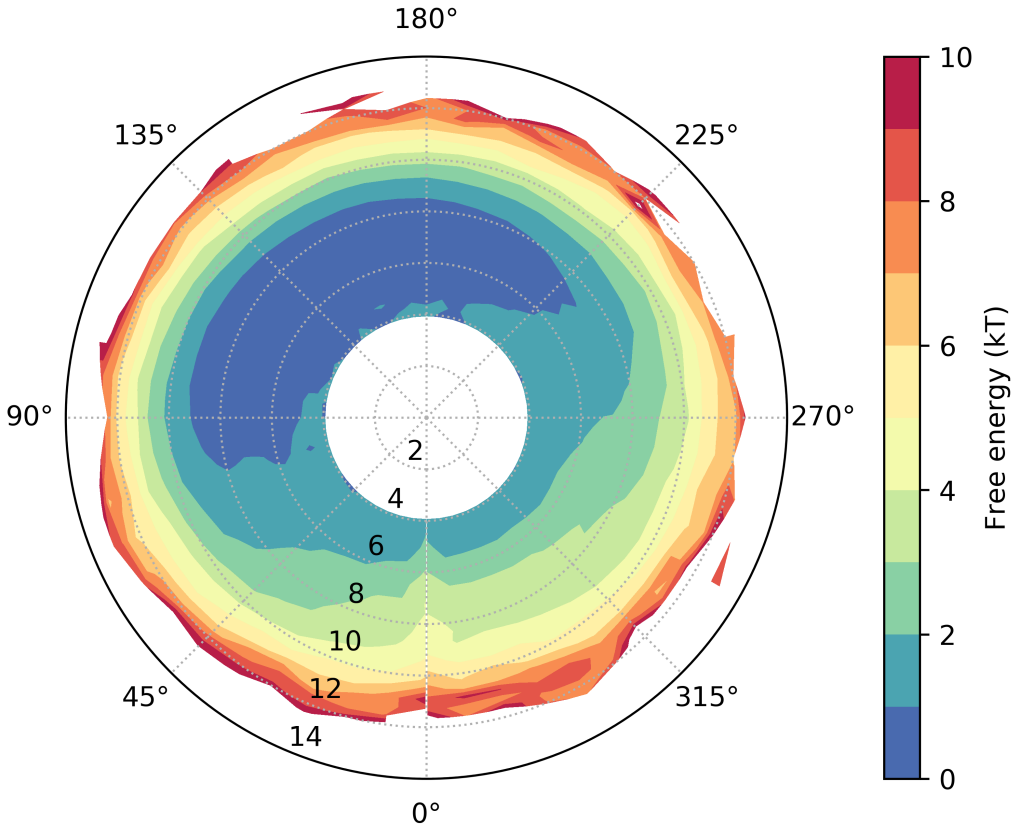

Figure 2.5: The combined free energy surface over helical parameters, weighted by the *in vivo* distribution of NRLs.[26] Angles less than 180^o imply a left-handed helix. Concentric circles around the central axis are labelled with the corresponding helical radius, in nanometers. The free energy surface exhibits a broad, smooth basin centered on a diameter of 14 nm and an angle of 140^o, but which extends from 80^o to 225^o with no energy barriers greater than 1kT across its span. The hole in the distribution at $r < 4nm$ is present due to the correction made for the approximate width of each nucleosome.

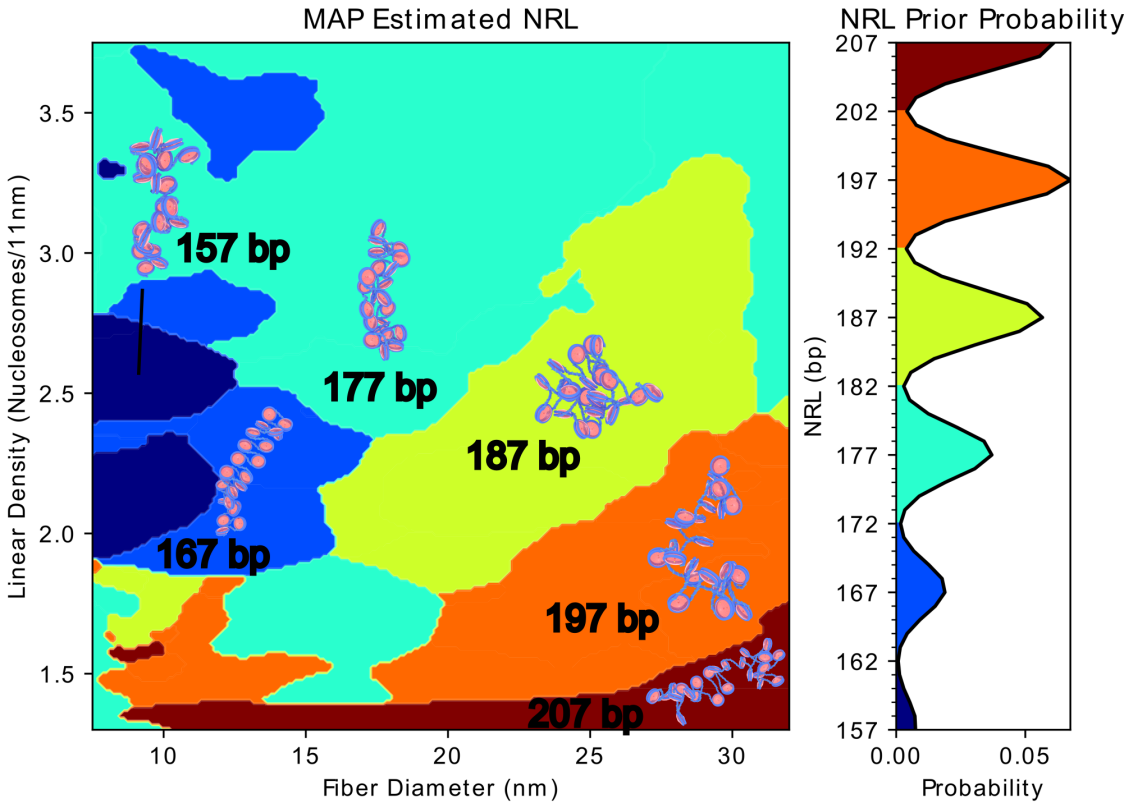

Figure 2.6: Maximum *a posteriori* estimates of the NRL for chromatin fibers of different dimensions. **A:** At any point in this graph, the color corresponds to the NRL with the highest likelihood of having produced that density and diameter, or equivalently, the NRL with the lowest prior-weighted free energy at that density and diameter. Regions are colored according to the value of the NRL, as shown in the color-coded histogram of **B**.

2.7 Supplementary Information

SI1: Calculation of helical parameters Informed by prior approaches where the fiber axis was approximated by polynomial fitting,[51] the helical parameters of equilibrium fibers were calculated as follows. For every successive set of 6 nucleosomes along the fiber, a corresponding average point on the fiber axis was estimated by a weighted average of the set's centers of mass. An unweighted average of N centers of mass closely approximates the location of the fiber axis position for helical angles in the range $\frac{2\pi}{N} \leq \theta_h \leq \frac{2\pi(N-1)}{N}$, but is a poor estimator outside of this range. To achieve higher fidelity in estimation of helical properties from simulation data, a set of symmetric but non-uniform weights were calculated by solving the linear system of equations requiring that the weighted sum of every 6 nucleosomes perfectly estimates the fiber axis for the ideal solenoid, with $\theta_h = 66.4$. The resulting fiber axis estimation scheme accurately estimates the ideal axis location with error less than 20% for all helical angles, and with error less than 5% in the range $50.0° \leq \theta_h \leq 310.0°$.

The resulting points calculated on the fiber axis were joined by cubic splines to produce a continuous representation of the axis. The local helical parameters of every nucleosome in each timestep of the simulations were calculated as follows. All points generated via this method will have a consistent systematic error (i.e., a finite distance from the true helical axis), which decreases for increasing values of N. Any chosen value of N will exhibit periodic points of perfect theoretically accuracy at values of $\theta_{gyre} = \frac{2\pi k}{N} | k \in [1, 2, ..., N-1]$,

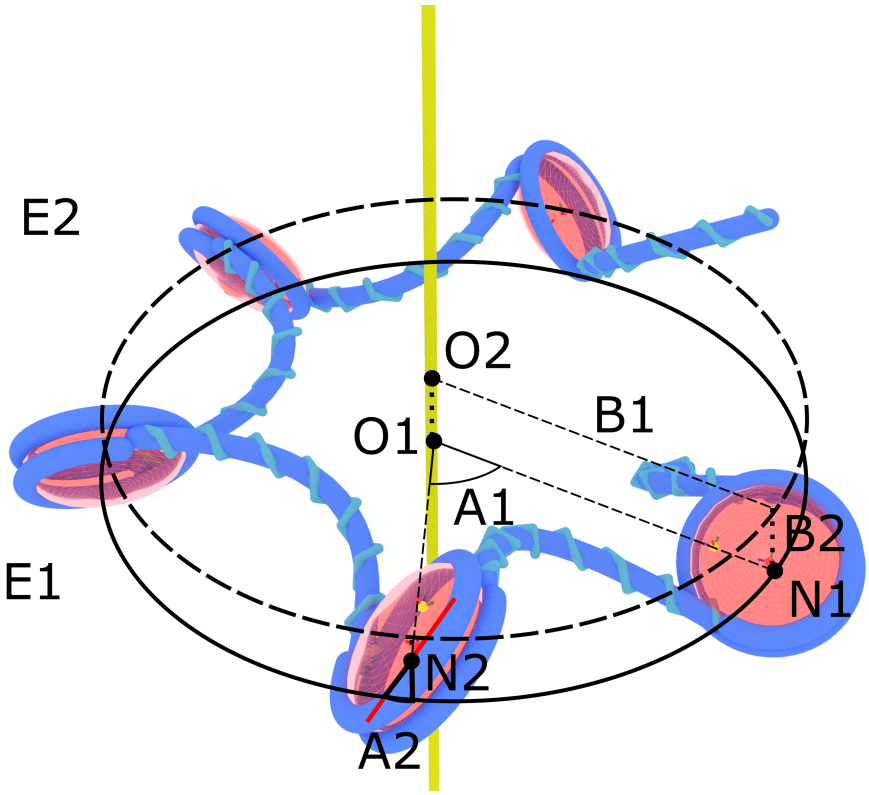

Figure 2.7: A diagram of the helical geometry of chromatin fibers. Nucleosomes spiral in a left-handed helix around a central axis (yellow) in a geometry defined by four parameters: helical pitch, helical gyre, helical radius, and nucleosome tilt. The figure is labelled as follows: ellipsoids E1 and E2 represents the plane perpendicular to the central axis which includes nucleosome center N1 and N2, respectively. The closest points on the central axis to nucleosome N1 and N2 are labelled O1 and O2. The helical radius is defined as the length of segment $\|\overline{O1\,N1}\|$. The helical gyre (or helical angle) is given by the dihedral angle between $\overline{O1\,N1}$ and $\overline{O2\,N2}$, labelled here as A1. The helical pitch is defined as the rise along the central axis per full revolution of 2π radians, and can be calculated by $pitch = \frac{2\pi}{A1}\|\overline{O1\,O2}\|$. The helical tilt is given by the angle between the plane of each nucleosome and the central axis, labelled here as angle A2.

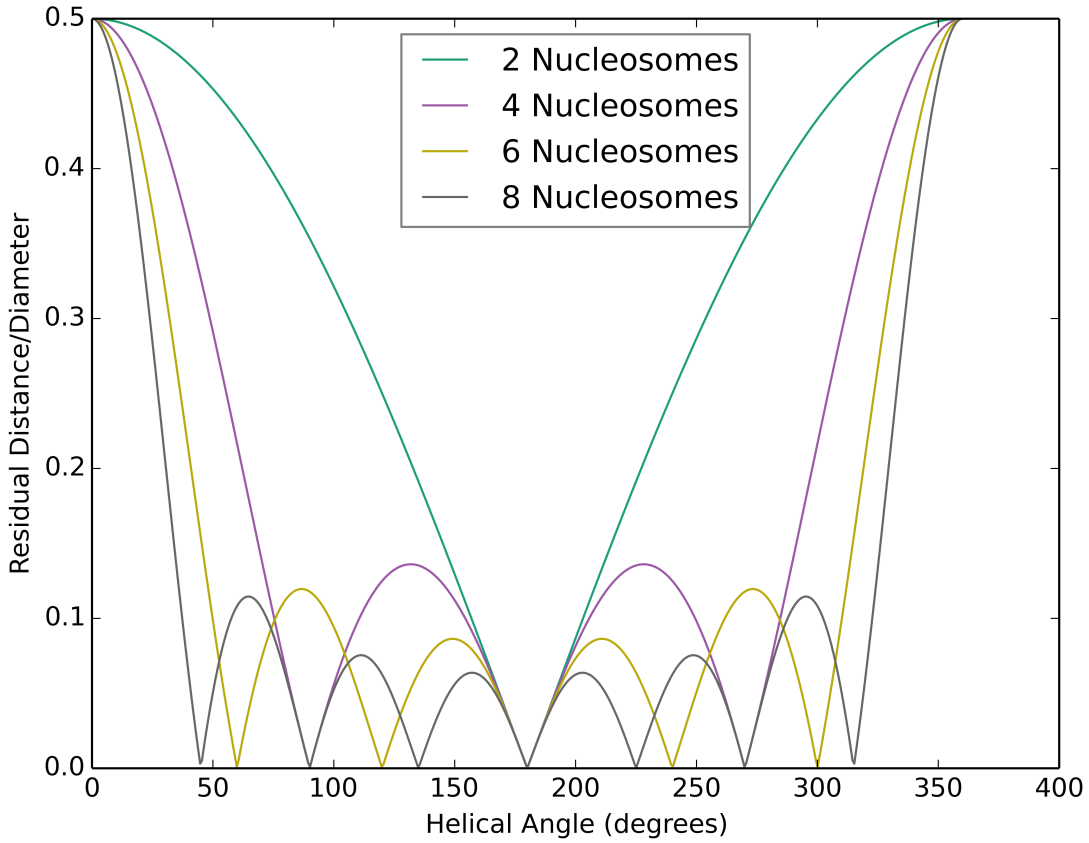

Figure 2.8: Theoretical accuracy of central axis estimation from nucleosome position averaging. These curves are calculated as the distance between the midpoint of N points arrayed uniformly along the unit circle, spaced by the helical angle, and the center of the unit circle. A curve with uniformly zero residual distance is a perfect estimator of the central axis position. Estimates become increasingly accurate as more nucleosomes are used in estimation in the calculation, though this reduces the extent to which the fiber axis tracks the real path of each nucleosome through space.

CHAPTER 3

SIMULATING THE EFFECTS OF ACYLATIONS ON GENE-SCALE CHROMATIN STRUCTURE

3.1 Author Contributions

This chapter includes products of the combined efforts of the authors of [69], which should be considered preliminary until peer review.

Molecular simulations were designed, conducted, and interpreted by A.E.C. and Y.W. under advisement by J.D.P. *In vitro transcription* analysis was performed by S.N. under supervision by R.S. Single molecule force microscopy experiments were conducted by J.F. and N.M. under advisement by H.D. Single molecule FRET probing was performed by H.N. under supervision by B.F. A.E.C. and Y.W. wrote this summary of computational findings to be included in [69], with Y.W. contributing the section on 1CPN re-parametrization.

3.2 Introduction

Chromatin structure in the nucleus is strongly influenced by the presence of epigenetic marks, and specifically histone post-translational modifications, on the histones which comprise the nucleosome core particle. These epigenetic marks form the basis for phase separation between heterochromatin and euchromatin at the longest length-scales of chromatin structure, and play an integral role in the epigenetic silencing and activation of gene expression networks.[70–73] These epigenetic marks influence chromatin structure either by directly mediating changes in histone-histone or histone-DNA interactions, or by altering the specifics of nucleosome-ligand binding,[74,75] although it has been recently proposed that this categorical distinction may in part be artificial.[76]

H4K16Ac, or the acetylation of the 16th lysine of histone H4, is one of the most biologi-

cally abundant epigenetic marks in human cells and has been the subject of extensive study to-date.[41,77–81] H4K16 acetylation is enriched at active genes and enhancers in actively-transcribed chromatin compartments, and is understood to play a role in transcription activation.[77,81] This transcriptional activation has been theorized to result from a disruption of contacts between the H4K16 residue and the acidic patch of adjacent nucleosomes, resulting in decompaction of chromatin arrays *in vitro*.[41,62,78,82] This model for the mechanism of H4K16Ac's action is supported by findings that H4K16Ac's presence correlates with increases in chromatin accessibility, and overall decompaction *in vitro*.[62,77,78] This perspective on H4K16Ac's mechanism of biological action has generally persisted despite complicating experimental findings that chromatin *in vivo* is not extensively decompacted by this mark.[76,81,83]

Two closely related epigenetic marks, H4K16 propionylation (H4K16Pr) and H4K16 butyrylation (H4K16Bu), have recently been identified in human cell lines, but their biological roles and mechanisms of action remain unclear.[84–86] H4K16Pr is enriched in cerebrocortical neuroepithelium and hippocampal primordium during critical, H4K16Ac-dependentent stages of embryonic neruodevelopment, likely from KAT8 activity in the presence of excess propionyl-CoA.[87,88] H4K16Bu is deposited on histones in the presence of the p300-CBP complex and excess butyryl-CoA, and has been associated with transcriptionally active genes.[89–91] While it has been suggested that such lysine butyrylations may escape acetylation-dependent histone removal mechanisms, the implications of such effects have not been extensively explored.[89]

In this work, we characterize the range of structural variation induced by H4K16 acylations in kilopasepair-scale chromatin fibers, to shed light on the potential differential mechanisms of their action. We utilize extensive mesoscale molecular dynamics simulations to predict resulting molecular structures, in concert with single molecule structure probing experiments and in vitro transcription assays to both parametrize and validate our simulations.

Ultimately, our work demonstrates that both H4K16 butyrylation and propionylation induce categorically similar structural rearrangements in mesoscale chromatin to that induced by acetylation, with reduced chromatin compaction and nucleosome stacking regularity. Furthermore, we show that the structural differences between chromatin bearing each acylation are relatively small, and as such are unlikely to drive significant differences in biological effects.

Previous simulation and experimental studies have clarified the exact molecular mechanisms by which acetylation disrupts the interaction between the H4K16 residue and adjacent nucleosome acidic patches.[41,92,93] Prior work has also quantified the free energy landscape of nucleosome-nucleosome interactions with and without H4K16 acetylation.**Funke2016a**, [53,94] This work extends such analyses to include the H4K16Pr and H4K16Bu modifications.

3.3 Reparametrization of 1CPN for Acylated Nucleosomes

We performed simulations of isolated chromatin fibers with 12 nucleosomes and a uniform DNA linker length of 30 bp (*i.e.* NRL 177) in the LAMMPS[95] molecular dynamics software using the 1CPN (1-Cylinder-Per-Nucleosome) chromatin polymer model..[52,96] This model allows simulation of kilobasepair-scale segments of chromatin by explicitly treating each nucleosome core as a single rigid body with a complex interaction potential,[97] and treating linker DNA as short segments of a Twistable Stretchable Worm-like Chain (TSWC). 1CPN was explicitly parameterized to reproduce experimental and *in silico* measurements of the internucleosome interaction energy, and the free energy characteristics of linker DNA breathing. Furthermore, this model was designed to be easily integrated with new data on the behavior of nucleosomes bearing specific epigenetic marks,[52,53] and already includes parameters for the explicit treatment of nucleosomes with fully acetylated H4 tails.

We developed a series of additional parameters for the 1CPN model to represent H4K16Ac, H4K16Pr, and H4K16Bu *in silico* to directly reproduce our accompanying experimental force

spectrometry results[69]. These parameters were developed as follows. First, we performed umbrella sampling simulations of two nucleosomes bearing H4K16Ac using the 3SPN[98] (3-sites-per-nucleotide)/AICG[99](atomic-interaction-based coarsed-grained) model, following the procedure established previously for 1CPN parametrization.[52,53] H4K16 acetylation is modeled by neutralizing the positive charge on the corresponding lysine residues. This procedure produces free energy profiles for four independent internucleosome orientations (stacked, rotated stack, side-side, and rotated-side orientations). The resulting stacked free energy profile matches the well depth and minimum poition of the experimental free energy profile within 0.05 kcal/mol and 1 Å(in experiments, the minimum value is 1.292 kcal/mol, and the minimum position is 68.7 Å; in simulations, the minimum value is 1.243 kcal/mol). Zewdie interaction parameters for H4K16Ac nucleosomes were derived from these free energy profiles via least-squares regression[52] and are shown in Figure 3.1.

After obtaining the Zewdie potential parameters for H4K16Ac, we obtain the Zewdie potential parameters for H4K16Pr and H4K16Bu as a combination of both experimental results and simulation results. Upon inspection of the free energy profile of H4K16Ac in comparison to the unmodified profile, it is clear that the only significant differences occur in the stacked orientation, with a corresponding shift in both the well depth and minimum position. All three other orientations (rotated-stacked, side-side, and rotated-side) show no significant difference in well depth or minimum position. This matches our intuition of the mechanism of action of H4K16 acylations, which are thought to interrupt interactions with adjacent nucleosomes' acidic patches in the stacked configuration, and therefore should only cause significant energetic alterations when in the stacked orientation.[41,92,93] In accordance with these findings, we derive the Zewdie parameters for H4K16Pr and H4K16Bu such that the stacked free energy position and minimum are matched with the experimental results, while treating the free energy profiles of the other three orientations as identical to H4K16Ac. Similarly to the unmodified treatment, this also corresponds to a linear transformation of the

original Zewdie potential of H4K16Ac. To make the treatment consistent, we also carry out the transformation for H4K16Ac to fit its experiments, but the effect should be minimal as the difference between the free energy value and position in simulations and in experiments on stacked orientation is very small.

To ensure sampling of both locally- and globally-stable structures, several additional computational techniques were employed. First, all simulations were performed with Replica Exchange Molecular Dynamics (REMD) with 36 parallel replicas, spanning a geometric temperature range from $300K$ to $700K$, enforced via a Langevin thermostat, as described in ref. [52]. Replica exchange moves were proposed every 50,000 timesteps and accepted according to the corresponding Metropolis criterion, as implemented in the LAMMPS REPLICA package. In order to ensure a more complete exploration of possible structures, the Langevin damping coefficient was increased from $2000fs$ to $14144fs$, which was found to hasten simulation convergence while maintaining the correct Maxwell-Boltzmann distribution of instantaneous velocities.

Simulation convergence was identified by convergence of the distribution of sedimentation coefficients within each simulation. For all marks, the average sedimentation coefficient across all replicas converged to a small neighborhood around the long-time average within $\sim$ 5000 timesteps, followed by oscillation around this mean. Analyses were performed on the following 15,000 timesteps, with a resulting production simulation time of $45\mu s$ per replica, or $> 4.5ms$ total simulation time per acylation across all replicas and temperatures. However, it must be noted that this quantification of simulation time may under-represent the extent of possible structural changes observed *in silico*, due to both the use of REMD, and changes to the Langevin damping coefficient, both of which result in faster exploration of structures than would otherwise occur within the nominal simulation time.

Without postprocessing, only 1/36th of the simulated data obeys the canonical distribution of states at $300K$, corresponding to the lowest temperature replicate used in REMD.

To utilize the remaining fraction of the data without biasing our analyses toward unstable and high-temperature structures , all simulated timesteps across all replicas were reweighted to the $300K$ canonical distribution via the Multistate Bennett Acceptance Ratio (MBAR) method, as implemented in PyMBAR.[100] All distributions and averages of simulation properties referenced hereafter are calculated by analysis of these reweighted distributions.

3.4 Results

3.4.1 Linear and Non-linear Structural Variation in Acylated Chromatin

To quantify the structural variation across the acylations observed in simulated chromatin, we calculate a suite of descriptive collective variables (CVs) for all structures from each simulation, each of which provides a unique quantification of chromatin fiber shape and size. The exact collective variables are defined in Table 3.1. These include a suite of 3 CVs which quantify fiber size, 6 variables which quantify fiber surface area accessible to particles of different sizes, and the expected instantaneous FRET efficiency. Specifically, radius of gyration, end-to-end distance, and sedimentation coefficient are all canonical parameters for describing polymer size and chromatin fiber compaction. All surface areas calculated here are Lee-Richards molecular surface areas, and are reported for two categories: Linker DNA surfaces, and nucleosome surfaces. Furthermore, to quantify accessibility across scales of potential ligands, we calculate the accessibility to probe molecules ranging from solvent molecules, to transcription factors, to external nucleosomes. The full distributions of all of these quantities for all acylations are given in [69].

Sedimentation Coefficient, or $S_{20,w}$, is an experimentally-measurable quantification of fiber compaction directly related to the speed of fiber sedimentation during analytical ultracentrifugation experiments. Usefully, $S_{20,w}$ can also be calculated from computational modeling results via the Kirkwood Approximation,[101] and has seen significant prior use in the

literature to compare simulated chromatin fiber structures with experimental equivalents. We show the distributions of $S_{20,w}$ for all acylations in Figure 3.3. Most notably, we find that all acylations exhibit broad, overlapping distributions of $S_{20,w}$, with intra-acylation variation significantly greater than inter-acylation variation. These results concur with prior experiments which demonstrated *in vitro* that 12-nucleosome arrays with NRL 177 sample an extremely wide range of $S_{20,w}$ across at least $50 - 80Sv$, albeit for different salt conditions than studied here.[30] We find that, on average, propionylated fibers are the most compact, followed by unmodified fibers, butyrylated fibers, and acetylated fibers following as the least compact. However, this comparatively small variation alone is unlikely to yield an unambigious structural interpretation of the observed transcriptional variance.

Rather than individually analyze the distributions of other variables introduced in Table 3.1, we may simultaneously identify the key modes of structural variation across chromatin fibers of all acylations via principal component analysis (PCA). PCA is a linear dimensionality reduction technique which identifies the independent modes of correlated motion, called Principal Components, which best describe variation in a set of variables. We perform principal component analysis (PCA) on the sets of CV values for each mark, reweighted to the Canonical Ensemble at $300K$. The loading of each CV corresponding to each principal component of the pooled data is further given in Table 3.2.

Notably, the eigenvalues of the principal components indicate that roughly 70% of all structural variation across all acylations can be explained by linear variation in the first principal component, followed by 15% of all variation explained by the second principal component, with all other principal components contributing less than 15% of the variation of the data set. We find that the first principal component strongly correlates with standard measures of compaction, with nearly equal loadings for R_e, R_g, and $S_{20,w}$. Furthermore, we find that these standard measures of fiber compaction also present with a generalized increase in fiber surface area, as demonstrated by the high loading of $NASA_{Nuc}$.

Unexpectedly, this first principal component has almost no correlation to variation in fret efficiency with a loading of -0.026, implying that this highly dominant form of collective motion is nearly independent of relative donor-acceptor positioning. This counterintuitive result directly implies that nucleosomal rearrangement during overall fiber condensation is at least partially organized, such that the $i \pm 2$ contacts such as between the donor and acceptor nucleosomes rearrange almost completely independently of these global condensation events. Supporting this proposition, we find that simulated FRET efficiency is highly correlated with principal component 2 (loading of 0.604). PC 2 corresponds with a rearrangement of fiber surfaces which preserves overall fiber size and condensation, but significantly increases the accessibility of Linker DNA to the nucleoplasm (loading of 0.703 for $NASA_{DNA}$). Conversely, it can be said that when $i \pm 2$ nucleosomes move closer together within range of FRET, this is highly correlated with the exposure of linker DNA to the surrounding solution.

To better understand how local fiber structures can enable this counter-intuitive independence of donor-acceptor distances from global fiber condensation, we calculate the orientation-dependent potential of mean force between the donor and acceptor nucleosomes for each acylation, as shown in Figure 3.5. To provide a more information-rich description, we calculate distances between the donor and acceptor nucleosomes' centers-of-mass in two orthogonal spatial axes, here termed "X" and "Y", which are defined as follows. The positive Y axis is defined in all simulation timesteps to be distance along the vector from the Donor nucleosome COM through the top face of the nucleosome, perpendicular to the Donor dyad axis. The positive X axis is defined in all simulation timesteps as a projection of the vector pointing directly from the Donor COM to the Acceptor COM onto the plane perpendicular to the Y axis. As such, the position of the Donor COM is always (0,0) in all timesteps, and its top face normal is always identically oriented in the (0,1) direction. By construction, the Acceptor position is defined to always lie at a positive X location, and is generally excluded from the region around (0,0) by strong excluded volume interactions with the donor nucle-

osome. By calculating the free energy profile in these axes (Figure 3.5), we may explicitly identify conserved binding sites, accounting for the aspherical shape of nucleosome particles.

Surprisingly, despite the general lack of long-range order such as that described by the ideal zig-zag or solenoid conformations, we find that the acceptor nucleosome lies in one of at least four well-defined local free energy minima, or binding sites, around the donor nucleosome within a clear first solvation shell. We term these positional sites (A) Stacked, (B) Reverse-stacked, (C) Side-side associated, and (D) Improper-stacked. Though the exact positions of these free energy minima vary by acylation, their schematic position is highly conserved.

We find that the Stacked and Reverse-stacked binding are the most stable configurations for H4K16, H4K16Bu, and H4K16Ac, while H4K16Pr's most stable configuration is side-side associated. Notably, of these four binding sites, both Stacked and Reverse-Stacked are within range of FRET (*i.e.* $R < R_0 \approx 90$Å). The Improper-Stacked orientation, stable in all acylations except H4K16Ac, lies with its free energy minimum at approximately $E_{FRET} = 0.1$, and Side-side associated nucleosomes are significantly beyond the range of FRET ($E_{FRET} \ll 0.1$).

H4K16Ac is the only acylation which exhibits significant disordered structures, with a broad pool of possible states at $\Delta G_0 \approx 6kT$ above the free energy minimum, yet also exhibits the most stable stacked orientation of any acylated fiber. This is unexpected, since H4K16Ac nucleosomes exert the weakest direct attraction of any system studied here, and therefore have the lowest stabilizing forces of site A due to direct donor-acceptor attraction. However, these results indicate that the apparently lower stability of stacked binding in H4K16Pr and H4K16Bu is due to the *increased* stabilization of non-canonical binding structures (C and D), which are stabilized by the many-body interactions with nucleosomes other than the donor and acceptor. Acetylated chromatin fibers maintain sufficient internucleosome attraction to promote donor-acceptor binding, but are *not* sufficiently attractive to evolve higher-order

structures necessary for the stabilization of binding sites B-D.

Furthermore, these results elucidate the origin of our previous finding that expansion of the fiber (PC1) is not correlated with changes in donor-acceptor distances (E_{FRET}) For all acylations, the Acceptor nucleosome generally sits within one of the previously described conserved binding sites, regardless of overall fiber condensation. Furthermore, we find that variation in PC1, *i.e.* overall fiber condensation, does not cause the acceptor nucleosome to shift systematically between these four sites, resulting in the small E_{FRET} loading of PC1.

These findings are further supported by our FRET experiments, which show that H4K16Ac fibers have a larger fraction of acceptor nucleosomes within FRET range of the donor, and closer spatial binding of the acceptor nucleosome than is present in H4K16Pr or H4K16Bu.

3.4.2 Diffusion landscapes

While principal component analysis can identify correlated structural variation within chromatin fibers, it is not capable of identifying whether these principal components dominate the slowest collective motions of the chromatin fibers. To elucidate the fiber kinetics, we show the results of diffusion mapping analysis for each isolation in Figure 3.6. For all acylations, we find that the slowest collective motions, i.e. the first several diffusion eigenvectors, exhibit a wide, stable central pool of structures, surrounded by rare diversions to specific condensed diffusion states. Importantly, we find significant energy barriers, up to $10k_BT$, between the broad diffusive pool and these rare condensed states, implying that these states were only effectively sampled due to the replica exchange molecular dynamics methodology discussed previously, and that transitions between these states may be rarely detected in experiment.

For all acylations, these rare diffusion states are, on average, more compact than the central pool, and also exhibit significant ordering. None of these states, however, can be simply classified as canonical structures such as the zigzag or solenoid proposed in the prior

literature, and instead display a wide array of semi-regularity with mixtures of straight and bent linker DNA. We find 2 such compact states for H4K16Ac fibers, 3 compact states for H4K16Bu, and 4 compact states each for H4K16Pr and H4K16. However, such numbers must not be taken as a quantitative estimate of the total number of condensed states per acylations, since these simulations were not explicitly designed to discover such rare and kinetically-trapped states.

These results have significant similarity with the findings of MFD analysis performed by Kilic *et al.* in ref. [94]. In that study, it was also found that the kinetic landscape of chromatin fibers *in vitro* is dominated by a large, stable pool of disordered states, complemented by rare departures to condensed and ordered states. Furthermore, they identified very long time scales ($100\mu s - 5ms$, with some $> 100ms$) for these order-disorder transitions, which similarly implies significant free energy barriers inherent to these fiber rearrangements. However, we have insufficient data to identify a direct correlation between the diffusive states observed in simulation and the states proposed by Kilic *et al.*, and as such, we cannot claim that the structures we identify *in silico* are identical to those found in experiment.

These diffusion maps also provide some insight as to the effect of acylation on chromatin kinetics. The diffusive landscapes of all acylations have very similar shapes, with all exhibiting the aforementioned rare departures into condensed states. For acetylated fibers only, these rare states directly correspond to significant changes in fret efficiency, which is not exhibited by any other acylation. However, few systematic differences beyond this can be described, and we do not find that departures to these rare, condensed states correlates strongly with other structural parameters beyond the correlations implied by PCA.

3.5 Relation between *in vitro* transcription and structural variation

Our simulations demonstrate straightforwardly the impact of the modified internucleosomal interaction caused by the acylations studied here. Whether quantified by individual compaction parameters such as average sedimentation coefficient or by the first principal component of our data set, we find a direct relationship between interaction well depth and overall compaction, confirming that stronger internucleosomal interactions cause more compact fibers. However, our simulations also demonstrate that the structural variation within each acylation is significantly greater than the variation between different acylations. Furthermore, we find that the DNA origami experiments are in fact consistent with our reported fret measurements, despite their apparent contradiction: we similarly find *in silico* a nonlinear relationship between internucleosomal interaction energy and stability of donor-acceptor pair structures ($i \pm 2$ interactions). Furthermore, the kinetic landscape all of our chromatin simulations qualitatively matches that described in closely related experimental measurements based on Fret efficiency, entailing a predominantly disordered fiber which undergoes rare condensation events in which it becomes temporarily more ordered.

Unlike structural variables which may be calculated at all timesteps within a simulation, and therefore may be compared to experimental measurements, *in vitro* transcription data can only be assessed from the perspective of a single quantity per experiment, and even then is only meaningful as an average over multiple samples. However, we may ask: what combination of structural properties best explains the average variation of transcription rates between each acylation? That is, are there meaningful groups of collective variables which are linearly correlated with transcription rates in our dataset?

Naive inspection of the collective variables introduced earlier does provide some insight into transcription, as all measures rank unmodified, butyrylated, and acetylated fibers in descending order of fiber compaction, which corresponds directly to their transcriptional

ordering. However, none of these variables match the order of magnitude of variation found in transcription rate, nor do any of them correctly place propionylated fibers relative to all other PTMs. To identify if, instead, a combination of these CVs may better explain the variation of transcription rate between acylations, we performed an exploratory analysis to identify potentially meaningful groups of these CVs, combining both a quantitative optimization of variable exponents with a regularization requirement which enforces a qualitative preference for simple groups containing few variables and integer exponents. We report several of the simplest of such collective groups in Table 3.3.

Surprisingly, we find that all seven groups share a nearly identical qualitative interpretation: transcriptional rate is proportional to linker DNA accessibility in competition with the binding of external protein ligands. This model is consistent with the details of our IVT experiments: transcriptional activation by the GAL-VP16 System is initiated by association of the DNA-binding domain of GAL to the chromatin scaffold, which may bear a rate related to the available surface area of linker DNA around the promoter region.[102,103] However, a systematic evaluation of the mechanism underlying these relationships is beyond the scope of this work. This analysis is necessarily qualitative, since this method of collective group discovery is guaranteed to produce results, even if performed on completely randomized data. Instead, we draw attention to the shared qualitative interpretation of these collective groups, which we believe may warrant further study in further research.

3.6 Discussion and conclusions

In our *in vitro* transcription model, we see a nonlinear effect of the addition of methyl groups onto the acyl modification of the H4K16 residue. The addition of a plain acetyl group to this lysine dramatically increases transcription rate, with every additional methyl group further diminishing this transcriptional gain. Our DNA origami experiments demonstrate that these acylations directly reduce the interaction energy between stacked nucleosomes,

with increasing numbers of methyl groups similarly minimizing the magnitude of this effect. Counter-intuitively, however, our FRET efficiency measurements show that this modification of interaction energy does not have a linear effect on the stability of $i \pm 2$ contacts, throwing into question the validity of simple hypotheses correlating transcriptional activity to fiber compaction.

However, our mesoscale simulations further support our claim that this is a true feature of chromatin dynamics, verifying both that direct fiber compaction does not predict the stability of $i \pm 2$ contacts, and suggesting that transcription rate is best explained by a competition between linker DNA accessibility to transcriptional machinery and competitive association by external protein complexes.

Our proposed relationship between transcriptional rate and chromatin molecular surface area is not dissimilar from other relationships proposed in the literature. Almassalha *et al.* found significant association between the lifetime of colorectal cancer patients and a chromatin structural property which they refer to as chromatin fractal dimension, which they interpret as a relationship between transcriptionally available surface area and chromatin density.[104–106] Furthermore, it is in line with findings that genes' expression have a power-law dependence upon the accessibility of their promoters with an exponent of ~ 1.7, and this fit achieves an $R^2 = 0.71$.[107]

3.7 Tables

Collective Variable	Variable Name
$S_{20,w}$	Sedimentation Coefficient (Svedbergs)
R_g	Radius of Gyration (m)
R_e	End-to-End Distance (m)
$NASA_{Nuc}$	Nucleosomal nucleosome-accessible surface area (m^2)
$NASA_{DNA}$	Linker DNA's nucleosome-accessible surface area (m^2)
$SASA_{Nuc}$	Nucleosomal solvent-accessible surface area (m^2)
$SASA_{DNA}$	Linker DNA's solvent-accessible surface area (m^2)
E_{FRET}	*In silico* instantaneous FRET efficiency

Table 3.1: Structural Collective Variables. Each of these structural CVs were used in the principal component analysis. The first three variables each measure overall fiber compaction, while the following four measure the accessibility of different parts of the chromatin fiber to probe molecules of different sizes. The last CV, E_{FRET}, represents the average instantaneous FRET efficiency between probe molecules located on the 4th and 6th nucleosome of the 12 nucleosome fibers, in direct analogy to single molecule fluorescence detection experiments.[69]

CV	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
$S_{20,w}$	-0.441	-0.010	0.023	-0.193	-0.168	-0.015	0.467	0.722
R_g	0.435	0.039	0.021	0.159	-0.058	0.631	-0.330	0.522
R_e	0.368	-0.161	-0.03	-0.085	-0.863	-0.293	0.003	-0.007
$NASA_{Nuc}$	0.433	-0.140	-0.122	0.135	0.117	0.262	0.802	-0.184
$NASA_{DNA}$	0.080	0.703	0.610	0.276	-0.116	-0.091	0.167	-0.024
$SASA_{Nuc}$	0.416	-0.093	-0.047	0.226	0.395	-0.661	-0.033	0.414
$SASA_{DNA}$	0.336	0.291	0.048	-0.876	0.179	-0.019	0.007	0.011
E_{FRET}	-0.026	0.604	-0.778	0.130	-0.092	-0.045	0.013	0.021

Table 3.2: Principal Component loadings. For each principal component, entries in each row represents the component of that PC parallel to each corresponding CV. All 8 PCs represent an orthonormal basis for the chosen collective variable space.

Collective Group	Dimensions	Un	Bu	Pr	Ac
IVT Transcription	-	1	1.05	1.34	1.56
R_g	m	1.00	1.06	1	1.13
R_e	m	1	1.23	1.12	1.49
$1/S_{20,w}$	Sv^{-1}	1.01	1.05	1	1.10
0.5 PC1 - PC2	-	1	1.32	2.98	7.87
$SASA_{DNA}/NASA_{Nuc}$	-	1	1.18	1.20	1.27
$SASA_{DNA} * \left(\frac{R_g R_e}{NASA_{DNA}}\right)^{0.2}$	m^2	1	1.21	1.31	1.49
$\frac{SASA_{DNA}}{NASA_{DNA}}$	-	1	1.27	1.69	1.75

Table 3.3: Transcription-correlated collective groups. Each group represents a function of previously defined CVs which bears quantitative resemblance to measured IVT values. All groups present have been normalized by the smallest value for any acylation for comparison to IVT results. Note that, although theoretical support for any one of these groups has not yet been produced, the qualitative interpretation of all groups are highly similar.

3.8 Figures

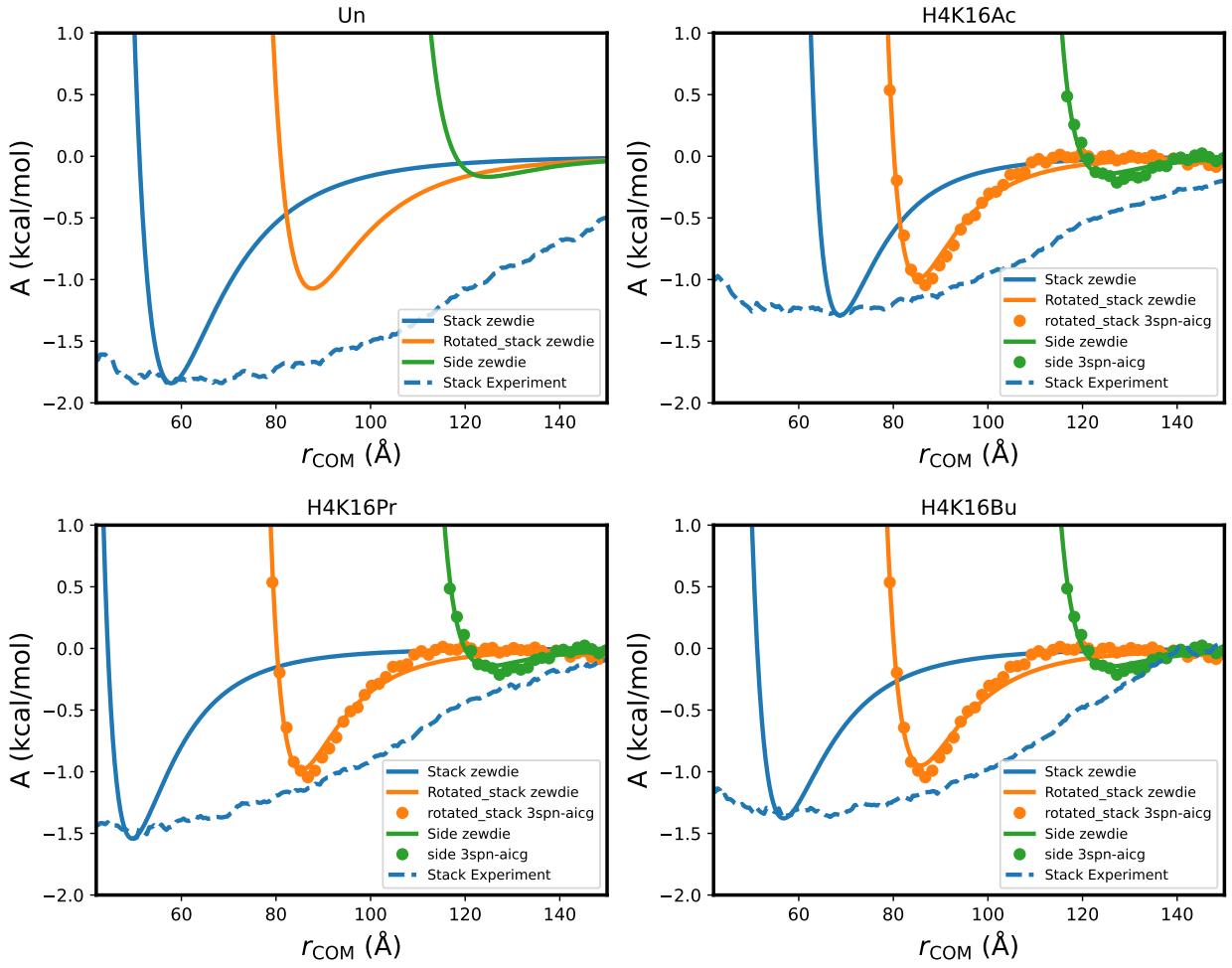

Figure 3.1: Zewdie potential fit to experimental results. For each mark, the free energy profile for three separate orientations is compared to the experimentally-determined free energy profile. As described in text, zewdie parameters were chosen to ensure a match between the modelled and experimental minimum position and energy.

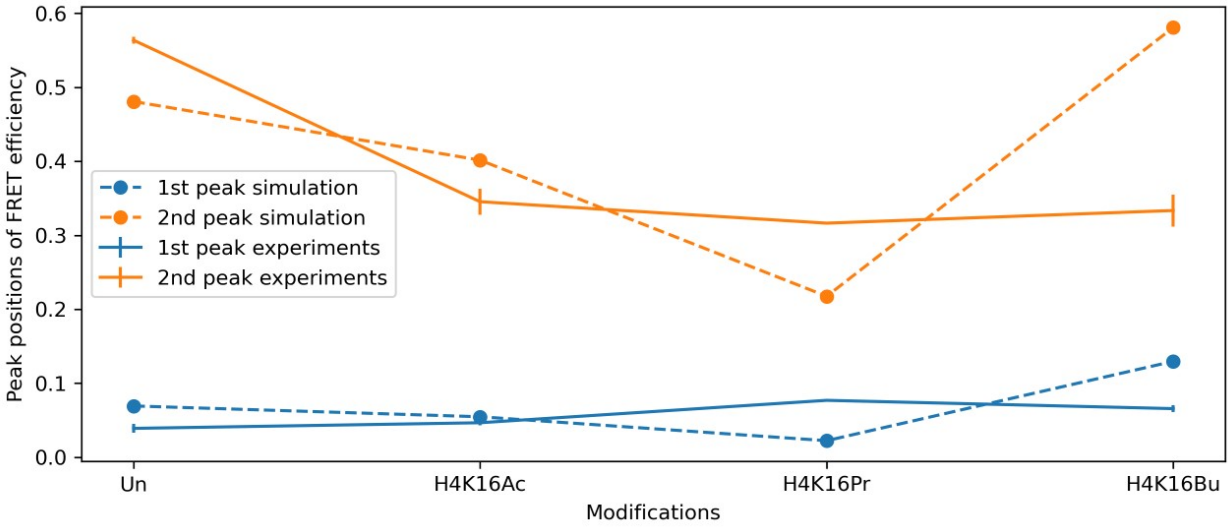

Figure 3.2: Computed E_{FRET} peak positions compared to experiment. Both experimental and computed E_{FRET} distributions were approximated by a 2-fold gaussian mixture model, and the means of both gaussian distributions are reported for each acylation. Statistical uncertainties are reported by vertical lines for experiment and simulation datapoints, though these uncerainties are smaller than the marker for simulated data.

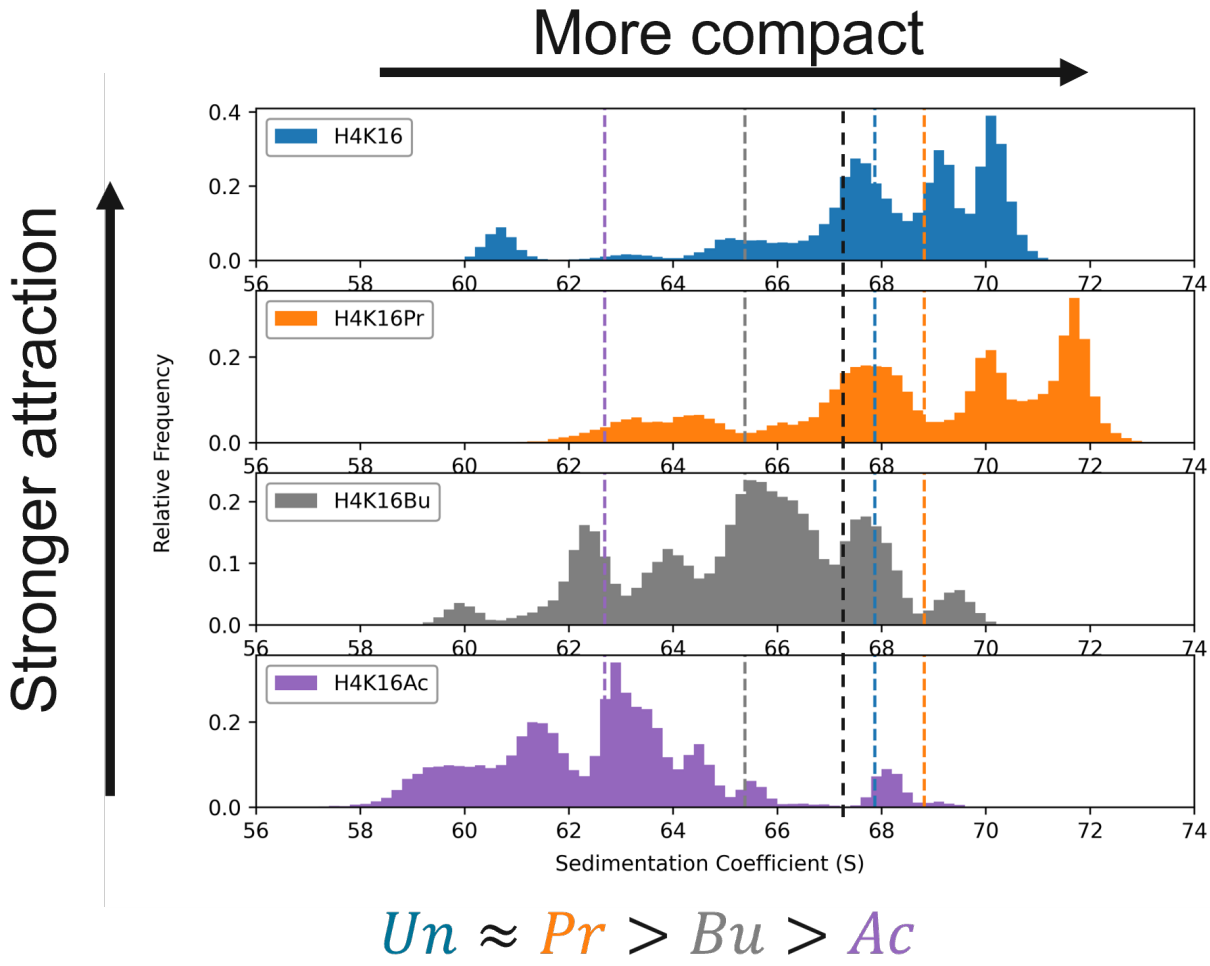

Figure 3.3: Sedimentation coefficient distributions by epigenetic mark. Histograms for each mark are shown, with vertical dashed lines corresponding to the mean value of each distribution. Dashed black line represents the mean $S_{20,w}$ reported in ref. [30] for their central histogram peak, for 12×177 NRL chromatin arrays at $100mM$ $[MgCl_2]$. All distributions represent the combined results of all 36 temperature replicas and 3 random seeds, reweighted to the canonical distribution at 300K by MBAR.

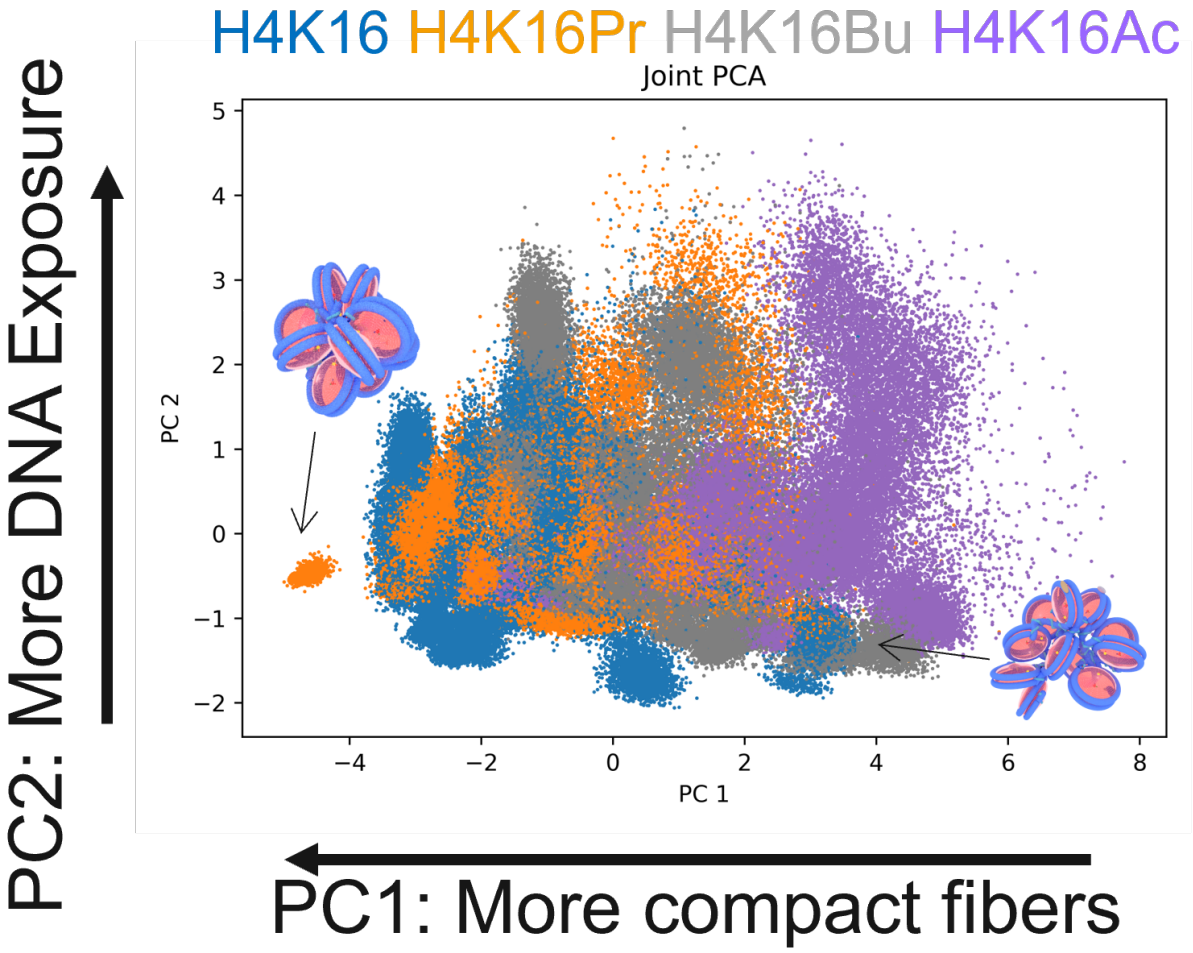

Figure 3.4: Scatterplot of the first two principal components of the acylated chromatin fiber structures. Although all distributions are highly overlapping, fibers with weaker internucleosome interactions (such as H4K16Ac, purple, and H4K16Bu, grey) are located at more positive values of PC1, indicating less compact configurations. There exists a weaker correlation between acylation and PC2, which correlates with regular stacking ordering and accessibility of linker DNA to solution.

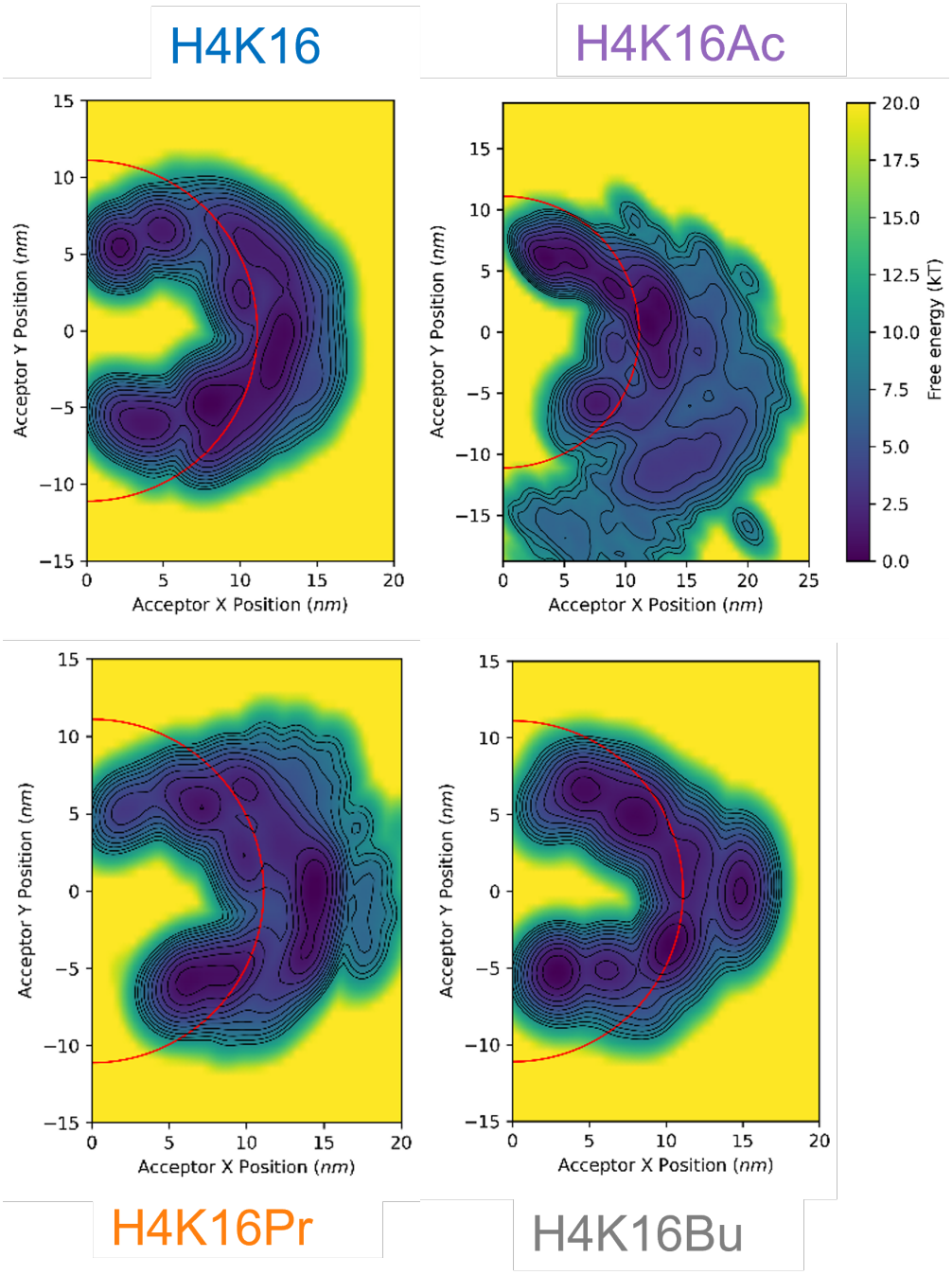

Figure 3.5: Free energy surfaces of donor-acceptor FRET probe-marked nucleosomes. Distance in the Y-axis represents donor-acceptor displacement parallel to the normal vector extending from the top face of the donor nucleosome. Distance in the X-axis represents the component of the donor-acceptor displacement orthogonal to this Y-axis. The red ellipse on each plot represents the level set of $E_{FRET} = 0.1$, which effectively separates "bound" and "unbound" structures as they have been labelled in the corresponding experimental E_{FRET} distributions.

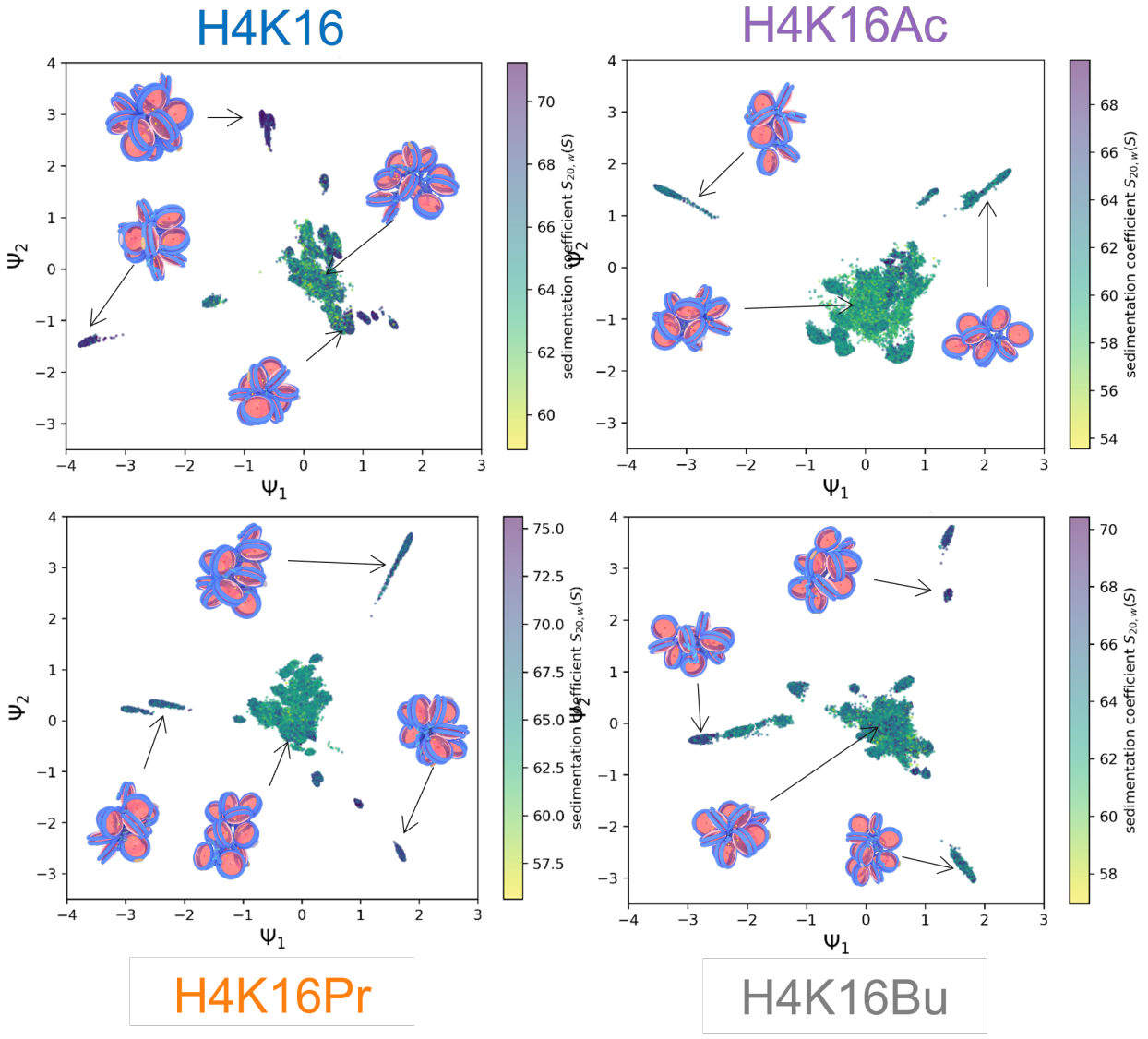

Figure 3.6: Diffusion maps of structural variation for each acylation. Each diffusion map can be described as having a broad, central pool of states which can freely interconvert. Fibers bearing each acylation experiences slow, rare diversions to more condensed states, visible as partially connected trails of points leading away from the central structures. Each scatterplot is colored by sedimentation coefficient to illustrate that these rare structures are more condensed than typical structures. Pictured structures are representative of typical stuctures near their respective local free energy minima.

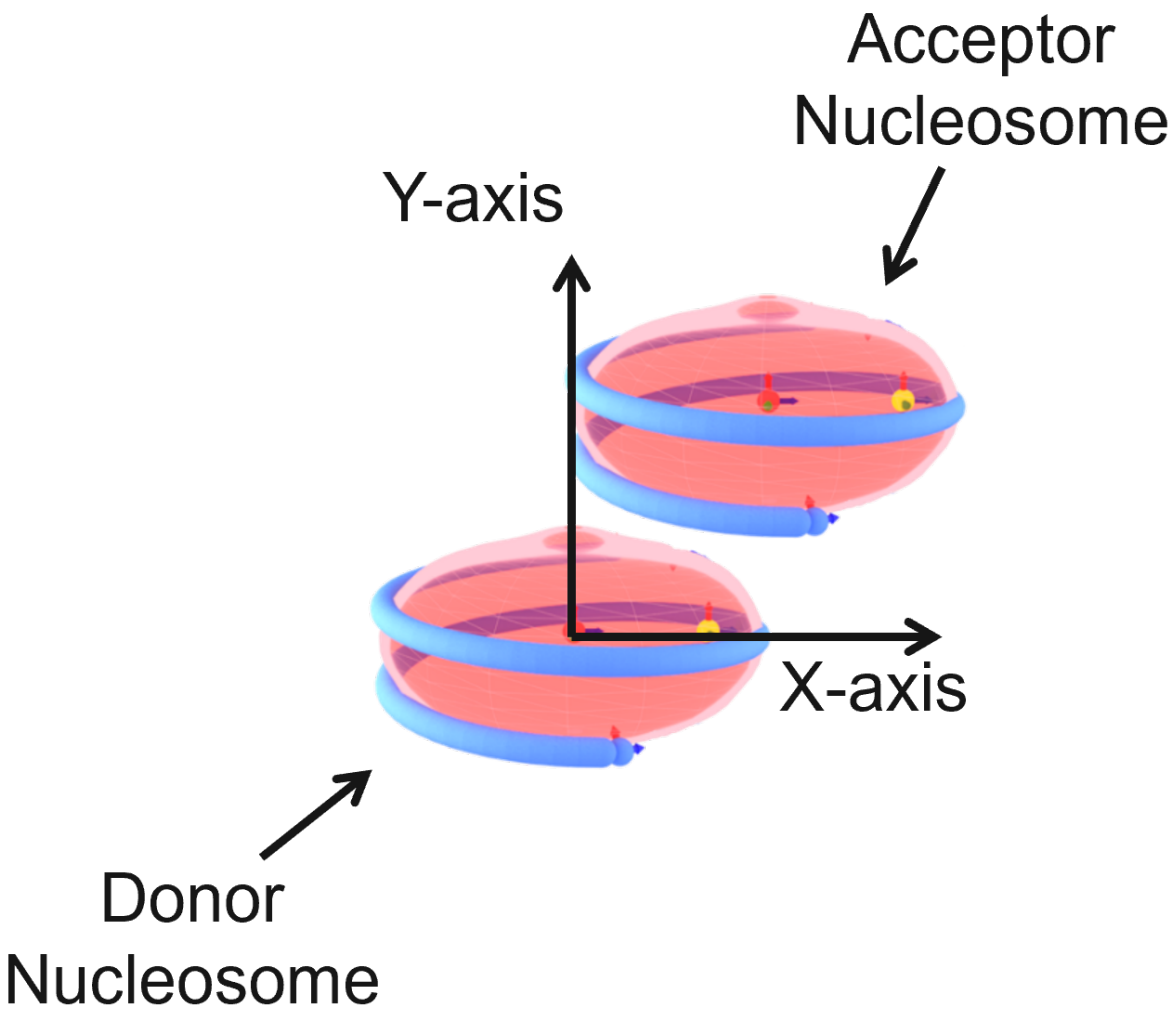

Figure 3.7: Geometry of the 2-body free energy surfaces. The Y axis is defined as the normal of the donor nucleosomal plane, and therefore the orientation of the donor nucleosome is always planar to the X axis. The X axis is defined as the vector in the orthogonal plane to the donor normal which points most directly towards the COM of the acceptor nucleosome. The acceptor nucleosome orientation is not defined by this projection.

CHAPTER 4

IN SILICO METHOD FOR CROSS-COMPARISON OF EPIGENETIC SSMLM IMAGING AND HI-C ANALYSIS

4.1 Introduction

Developments over the past several decades in experimental structure probing methods have led to significant improvements in our understanding of chromatin organization. Specifically, the development and refinement of high-throughput chromatin conformation capture (Hi-C) methods have enabled quantitative insights hierarchical organization of chromatin from the range of inter-chromosomal interactions to sub-kilobasepair interaction patterns.[70,108,109] Hi-C methods have also been instrumental in the development of key biophysical phenomena such as cohesin-dependent loop extrusion, and have played a key role in the establishment of the consensus of chromatin as a fractal polymer.[108,110–112]

However, the immense power of Hi-C techniques have come with drawbacks; increases in Hi-C data resolution require exponential increases in high-throughput sequencing reads in order to resolve the frequency of rare contacts, which in turn requires very large numbers of cells and high sequencing costs.[70,108,109] Furthermore, the method's reliance on population-average structure has precluded the identification of structural features which in single cells deviate strongly from the population-average behavior. Heroic efforts have been made to extend the method to single-cells, with[113–115] However, the overall paucity of read counts have significantly hampered the interpretability of supra-megabasepair structure, which critically depends on the detection of interactions across logarithmic scales of frequency.[116–118] Some groups have made great strides in statistical imputation techniques, by which the sparse dataset of cell-by-cell Hi-C data would be augmented by nonlinear local averaging of similar datapoints via classical statistics and machine learning methods.[119–122] It remains unclear if

the mechanistic insights based on these imputed single-cell datasets are widely applicable or verifiable without dramatic advances in experimental techniques.These inherent difficulties have created significant demand for alternative experimental approaches or data analysis methods which may provide complementary insight into single-cell chromatin structure.

Recent advances in super-resolution optical microscopy techniques[123–125] have provided a naturally complementary approach to understanding chromatin structure in single cells, via either DNA hybridization approaches[126–128] or immunofluorescence/spectroscopic single molecule localization microscopy (IF-sSMLM).[129–133] FISH-based chromatin tracing techniques on the contrary provide spatial information about gene loci across entire chromosomes, but require highly involved sequential fixation, hybridization, and imaging steps in addition to a highly expensive library of oligonucleotide labels, which together appear to have made widespread adoption of these techniques challenging to the scientific community.[134] sSMLM techniques leverage the capabilities of switchable fluorescent antibody tags to localize and discriminate multiple epitopes simultaneously in single cells with typical spatial resolutions below 100 nm, and with the potential to measure time-domain information and with spatial resolution beyond the diffraction limit. Such techniques have already seen applications measuring histone modification dynamics and segregation across heterochromatin and euchromatin domains, among other insights.[129,135–137] However, traditional IF-sSMLM techniques which target epigenetic marks carry the significant drawback that sequence information is generally lost, which creates significant difficulties for both interpreting the effect of observed biophysical phenomena on specific genes, and for comparing any gained single-cell insights to sequencing-based techniques like Hi-C.

In this work, we employ a hybrid top-down, bottom-up approach and develop a technique to extract maximal information from IF-sSMLM data to enable quantitative predictions about limiting Hi-C maps from single-cell sSMLM data. In this approach, we utilize a nucleosome-resolution polymer model for chromatin structure to construct and validate a

direct relationship between the plaid pattern of Hi-C contact maps and fluorescence colocalization observed in IF-sSMLM images. The result of our work is a pipeline which incorporates ChIP-Seq data and IF-sSMLM images to construct a most probable contact map of the underlying chromatin polymer based entirely on definite, analytical formulas and bottom-up polymer modeling, thereby providing preliminary estimates of the results of otherwise expensive and difficult Hi-C experiments for further interpretation. Additionally, we develop a novel machine learning technique, utilizing Convolutional Neural Networks (CNNs), to perform the same projection of IF-sSMLM images to Hi-C characteristics beyond the accuracy limits of the described analytical method. Lastly, we discuss the long-term potential of this class of methods in describing single-cell characteristics and chromatin structure.

4.2 Methods

We derive here our analytical framework for the interpretation of contact-map structure from IF-sSMLM images. The framework we describe here explicitly references a nucleosome-scale polymer model, but is generally applicable to any representation of chromatin structure which includes the positions of individual nucleosomes. We first discuss methods for producing *in silico* contact maps and IF-sSMLM images, and derive formulas for two descriptive statistics which are measurable by both methods: the molar contact ratio, which captures epigenetic mark separation, and the chromatin fractal dimension, which captures the chromatin density heterogeneity. Lastly, we introduce the polymer model for chromatin which we use to produce chromatin structures to which we apply our proposed analysis methods.

4.2.1 Chromatin representation

Suppose we analyze a chromatin fiber consisting in a set of N nucleosomes. Let $\vec{\mathbf{r}}(t) = (\vec{\mathbf{r}}_1(t), \vec{\mathbf{r}}_2(t), ..., \vec{\mathbf{r}}_N(t))$ represent the set of all instantaneous positions of all nucleosomes in the chromatin fiber at instant t, i.e. Monte Carlo step t. For any epigenetic mark α, let the

binary-valued composition vector $\vec{\mathbf{x}}_\alpha$ of size $1 \times N$ represent the index of each nucleosome bearing mark α, with a total of marked nucleosomes termed N_α.

4.2.2 Analysis Methods

4.2.3 Contact Functions and Mapping

Given any such representation of the positions of each nucleosome in a segment of chromatin, we can derive an expression for its corresponding contact map as follows. We define a contact function $h(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t))$ between any two nucleosomes i and j which has the properties:

$$\lim_{\|\vec{\mathbf{r}}_i(t) - \vec{\mathbf{r}}_j(t)\| \to \infty} h(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t)) = 0 \tag{4.1}$$

$$\lim_{\|\vec{\mathbf{r}}_i(t) - \vec{\mathbf{r}}_j(t)\| \to 0} h(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t)) = 1 \tag{4.2}$$

This contact function tracks if any two nucleosomes are in physical contact with each other, and would likely contribute sequencing reads in a Hi-C experiment. After choosing an exact functional form for $h(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t))$, one may calculate an *in silico* contact map by averaging the value of this function for all pairs of nucleosomes over a statistical ensemble of instantaneous structures, or Monte Carlo steps. Since multiple different contact functions have been employed by different authors[71,138,139] without appreciable effect on their results, we employ an efficient grid-based contact function equivalent to that adopted by MacPherson *et al.*:[71]

$$g(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t)) = \begin{cases} 1, \ c(\vec{\mathbf{r}}_i(t)) = c(\vec{\mathbf{r}}_j(t)) \\ \\ 0, \text{ otherwise} \end{cases} \tag{4.3}$$

where $c(\vec{\mathbf{r}}_i(t))$ represents the cell in the 3-dimensional simulation grid which contains the location $\vec{\mathbf{r}}_i(t)$. The cumulative sum of the contact functions between all genomic loci is binned at a pre-defined resolution (here chosen to be 10 kbp), and tracked throughout the

simulation. The final matrix of contact frequencies across the simulated region is plotted on a logarithmic scale to produce a contact map, in a fashion equivalent to the construction of a Hi-C map from Next-Generation Sequencing (NGS) read counts.[108]

4.2.4 Deriving the Molar Contact Ratio from Hi-C Data

In order to determine estimates of the contact maps of chromatin segments based on their IF-sSMLM images, it is necessary to define *which* key characteristics of contact maps are valuable to reproduce. As discussed by prior authors,[70,108,112] one key feature of Hi-C maps is the "plaid pattern" of alternating nuclear subcompartments, which represents the separation of heterochromatin and euchromatin in the nucleus. This plaid pattern is typically quantified by the first principal component of the Hi-C maps for 21 of all 23 chromosomes, and has been observed in single cells,[126] justifying its investigation here. The second primary feature of Hi-C contact maps is the relationship between genomic distance between loci and their contact frequency, typically quantified as a power law relation which has been argued to reflect the fractal globule structure of chromatin.[70,108,112,140,141] Further details of Hi-C contact maps include patterns of chromatin loops and topologically associating domains (TADs). However, the nature of this genomic distance relationship and of TAD structure in single cells remains a point of significant controversy, with experimental evidence indicating that these population-wide characteristics are likely not reflective of single-cell chromatin structures.[110,111,126,127] As such, we have chosen to neglect treatment of TADs in this study, and consider the genomic distance relationship or fractal dimension only as a special case in our analysis. Given this, we seek to define descriptive statistics which can be calculated for both Hi-C contact maps and for IF-sSMLM images and which capture both the plaid pattern and the fractal structure of the underlying chromatin.

We can define a quantification of the Hi-C map plaid pattern which is amenable to estimation from IF-sSMLM as follows. Suppose that the results of a Hi-C experiments are

represented as a matrix of Hi-C contact frequencies $\boldsymbol{A}$ with resolution δ and shape $\frac{N}{\delta} \times \frac{N}{\delta}$. In the ideal case, the experimental data $\boldsymbol{A}$ may be conceptually related to the contact function h defined previously as:

$$\boldsymbol{A}_{i,j} = M_{total} \left\langle h(\vec{\mathbf{r}}_i, \vec{\mathbf{r}}_j) \right\rangle_{exp} \tag{4.4}$$

where M_{total} is the total number of sequencing reads detected, and the average is taken over all digested cells in a Hi-C experiment. This is necessarily a simplification, and does not include sources of variation based on varying read count per cell, sequencing detection efficiency and biases, and others.

Let us define $C(\alpha, \beta)$ as the sum number of contacts detected in this experiment between all genetic loci with marks α and β. $C(\alpha, \beta)$ can be determined from the composition vectors and the experimental data matrix $\boldsymbol{A}$ via the formula

$$C_{exp}(\alpha, \beta) = \vec{\mathbf{x}}_\alpha^\top \boldsymbol{A} \vec{\mathbf{x}}_\beta, \tag{4.5}$$

This same quantity may also be calculated for a simulated chromatin segment. For a given spatial trajectory of T Monte Carlo steps of a segment of chromatin and the contact function g defined in Equation 4.3, we construct a simulated contact matrix $\boldsymbol{A}^{sim}$ as

$$\boldsymbol{A}^{sim}_{i,j} = T \left\langle g(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t)) \right\rangle_{t<T}. \tag{4.6}$$

From this and the composition vectors, we determine

$$C_{sim}(\alpha, \beta) = \vec{\mathbf{x}}_\alpha^\top \boldsymbol{A}^{sim} \vec{\mathbf{x}}_\beta \tag{4.7}$$

$$= \sum_{t=0}^{T} \sum_{i \in \mathbb{N}_\alpha} \sum_{j \in \mathbb{N}_\beta} g(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t)) \tag{4.8}$$

To quantify the Hi-C plaid pattern, we define a quantity which we term the 'molar contact

ratio' between genomic regions marked α and β as:

$$R_\alpha(\beta, \alpha) = \frac{C(\alpha, \beta)}{C(\alpha, \alpha)} \tag{4.9}$$

This quantity, which is defined $R(\beta, \alpha) \in [0, \infty)$, is a measure of the relative frequency of 'self-other' to 'self-self' contacts, and is similar to other dimensionless parameters which quantify the extent of phase separation in chromatin.[142] For a fixed sequence of epigenetic marks, an increase in $R(\beta, \alpha)$ corresponds with a weakening of the plaid pattern, while a decrease in $R(\beta, \alpha)$ causes the plaid pattern to become more prominent.

This ratio can be calculated either from an *in silico* chromatin trajectory using Equation 4.8 or from a Hi-C contact matrix using Equation 4.5, leading to two formulas for its value:

$$R_{sim}(\beta, \alpha) = \frac{\sum_{t=0}^{T} \sum_{i \in \mathbb{N}_\alpha} \sum_{j \in \mathbb{N}_\beta} g(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t))}{\sum_{t=0}^{T} \sum_{i \in \mathbb{N}_\alpha} \sum_{j \in \mathbb{N}_\alpha} g(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t))} \tag{4.10}$$

$$R_{exp}(\beta, \alpha) = \frac{\vec{\mathbf{x}}_\alpha^\top \boldsymbol{A} \vec{\mathbf{x}}_\beta}{\vec{\mathbf{x}_\alpha}^\top \boldsymbol{A} \vec{\mathbf{x}_\alpha}} \tag{4.11}$$

We note that, by construction, $R_{sim}(\beta, \alpha) = R_{exp}(\beta, \alpha)$ exactly if the following assumptions hold:

1. Equation 4.4 holds true; i.e., there exists a contact function $h(\vec{\mathbf{r}}_i, \vec{\mathbf{r}}_j)$ whose average value is linearly related to the number of measured contacts between loci i and j.

2. $g(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t)) = h(\vec{\mathbf{r}}_i, \vec{\mathbf{r}}_j$; *i.e.* the simulated contact function is an exact approximation of the true contact function.

3. The resolution of the experimental contact matrix $\boldsymbol{A}$ and the experimentally-determined sequence of epigenetic marks, $\vec{\mathbf{x}}_\alpha$, are taken at the same resolution as their *in silico* analogs.

4. The *in silico* polymer model exactly reproduces the experimental distribution of pair-

wise contact frequencies.

This molar contact ratio has several useful properties. First, it is internally normalized and scale-independent, so that the absolute number of sequencing reads or resolution of a Hi-C dataset does not systematically affect its characteristic molar contact ratios. For example, both doubling the number of sequencing reads (M_{total}) from a Hi-C experiment and doubling the duration of a simulation (T) would result in doubling both the numerator and denominator of Equation 4.10, yielding the same value for $R_\alpha(\beta, \alpha)$. Secondly, it is defined relative to epigenetic marks whose sequence can be determined by other methods external to Hi-C itself (*e.g.*, ChIP-seq), unlike other compartmentalization measures such as the first principal component of the Hi-C matrix which defines nuclear subcompartments A and B,[108] whose positions along the genome depend upon the values of the Hi-C experiment itself. Lastly, the quantities in its numerator and denominator have direct physical interpretation, namely as "how many times throughout a given experiment do loci i and j collide with each other?", which allows us to calculate the ratio via multiple routes, such as the spatially-based analysis of Equation 4.10 or the sequencing-based analysis of Equation 4.11, and apply physics-based arguments to approximate their value. It is this inherent flexibility and transferability which allows us to estimate $R_\alpha(\beta, \alpha)$ from IF-sSMLM images.

4.2.5 In silico IF-sSMLM Imaging

To derive *in silico* sSMLM images from ensembles of simulated polymer structures, we design a computational procedure which mimics the experimental procedures which are in common across multiple sSMLM studies.[129,143–145] We divide the experimental procedure into three sections which we will replicate: (1) immunofluorescence staining, (2) photophysics approximation, and (3) image collection. Our approach is ultimately similar to that used by other authors[146] for *in silico* SMLM image generation, but enabled us to construct images in an automated, high-throughput manner via a command-line interface (CLI).

During our immunofluorescence staining stage, we select a random subset of marked nucleosomes to become bound with reversibly-switchable fluorescent tags. We treat this by first defining an average fluorophore concentration, ρ_α^f, for use in imaging. For each image of the simulation box with volume V, $\rho_\alpha^f V$ individual epitopes (*i.e.*, nucleosomes with mark α) are randomly selected as "tagged" by fluorophores. Experimentally, the concentration of fluorophores are carefully chosen to produce ρ_α^f values as bounded by practical considerations; low values of ρ_α^f lead to slow image acquisition times, while high values of ρ_α^f may lead to overlapping point-spread functions (PSFs) in individual imaging timesteps, thereby preventing the resolution of individual fluorophores from the diffraction-limited raw data. For each image, we randomly select an average fluorophore concentration, ρ_α^f, from an experimentally-relevant range based on heuristics provided to us by the authors of Brenner *et al.*[143] for each image. This procedure produces a set of fluorophore-tagged nucleosomes, termed here $\mathbb{F}_\alpha$, with cardinality $|\mathbb{F}_\alpha| = \rho_\alpha^f V$ for each epigenetic mark.

Once all fluorescent-tagged nucleosomes are selected, the next step of our imaging procedure is the treatment of the relevant photophysics of sSMLM imaging. In experiments utilizing reversibly-switchable fluorophores, each fluorophore molecule undergoes stochastic transitions between fluorescently active and inactive states before ultimately photobleaching, which irreversibly renders the molecule permanently inactive. While the molecule is in an active state, its fluorescence can be identified by a detector as the appearance of a wide, isolated PSF, while all inactive states are generally indetectable. Each separate PSF identified in a single imaging timestep is termed a single "localization".[146,147] The result of these complex photophysical dynamics is that each fluorophore will produce a random number of detectable fluorescent localizations before photobleaching, which can be empirically described as a Poisson distribution with parameter γ_α dependent on both the intrinsic photophysics of each fluorophore and the experimental details of the measurement.[147]

We assume here that the average number of localizations per fluorophore takes a single,

constant value for each mark and each image, but is constant with respect to all other variables and conditions. For each labelled fluorophore i, we draw the number of observed fluorescent localizations l_α^i from a Gaussian distribution with mean and variance equal to γ_α, and draw an independent value of γ_α for each mark and each image from the reasonable range $\gamma_\alpha \in [10, 20]$. We use the normal approximation to the Poisson distribution to enable testing of the possible influence of non-Poissonian statistics for l_α^i on our analysis pipeline, which can describe the photophysics in some systems.[148,149] Our testing indicated that deviations of the observed photobleaching behavior from Poisson statistics did not yield a measurable reduction in analysis accuracy.

Performing this procedure for all tagged nucleosomes yields the total number of localizations L_α originating from fluorophores bound to mark α as:

$$L_\alpha = \sum_{j \in \mathbb{F}_\alpha} l_\alpha^j = \mathcal{N}\left(\gamma_\alpha \rho_\alpha^f V, \gamma_\alpha \rho_\alpha^f V\right) \tag{4.12}$$

Before image collection can begin, each generated localization j must be assigned both a centroid position $\vec{r'}_j$ and a positional uncertainty σ_j before being recorded on an *in silico* detector. Experimentally, the positional uncertainty is determined by both the width of the PSF and by the number of photons collected from each localization. Since photon-by-photon counting statistics lie outside the scope of this work, we avoid directly modeling the diffraction-limited image and individual photon bursts, and instead sample each uncertainty from a Gaussian distribution, whose parameters are selected from a range chosen to match reasonable values observed in experiments.[143] Once the uncertainty of each localization is determined, the centroid of that localization is sampled from a second Gaussian distribution, centered at the exact corresponding nucleosome position, with width equal to the localization uncertainty. The location and uncertainties of all localizations is then saved explicitly for further analysis.

For the final image collection stage of the *in silico* sSMLM imaging process, we first

prepare an *in silico* "detector" array, 500 pixels by 500 pixels, with two channels, one for each fluorophore type. We imprint this detector array with the uncertainty distribution of each detected localization by adding to it a Gaussian mask, centered at each localization's position and with width given by its positional uncertainty. The final detector array is lastly converted into a single PNG by recasting the continuous-valued RGB channels into 8-bit integers, using OpenCV,[150] to create images analogous to classical experimental sSMLM images for visual inspection and later processing. Exact positions and uncertainties all localizations are also saved.

4.2.6 Estimating the Molar Contact Ratio from sSMLM images

We seek an estimator for the molar contact ratio which can be determined from individual sSMLM images, or their corresponding sets of localizations and uncertainties. However, the locations of individual nucleosomes are not accessible to us, which are required to evaluate prior equations such as 4.8. We notice that equation 4.8 may be rewritten as

$$C_{sim}(\alpha,\beta) = N_\alpha N_\beta \sum_{t=0}^{T} \left\langle g(\vec{\mathbf{r}}_i(t), \vec{\mathbf{r}}_j(t)) \right\rangle_{sim}, \tag{4.13}$$

with the contact function averaged over all simulated pairs of nucleosomes i and j bearing marks α and β, respectively. Inspired by this form, we propose the following estimators, $\hat{C}(\alpha,\beta)$ and $\hat{R}_\alpha(\beta,\alpha)$, based on a transformation of the contact function introduced in Equation 4.3:

$$\hat{C}(\alpha,\beta) = \frac{N_\alpha N_\beta}{L_\alpha L_\beta} \sum_{i\in\mathbb{L}_\alpha} \sum_{j\in\mathbb{L}_\beta} g'(\vec{\mathbf{r}'}_i, \vec{\mathbf{r}'}_j) \tag{4.14}$$

$$\hat{R}_\alpha(\beta,\alpha) = \frac{\hat{C}(\alpha,\beta)}{\hat{C}(\alpha,\alpha)}. \tag{4.15}$$

The contact function $g'(\vec{\mathbf{r}'}_i, \vec{\mathbf{r}'}_j)$ is a grid-base contact function with the same form as

in Equation 4.3, with the prime denoting that the grid used here is a 2-dimensional grid defined at analysis-time over the sSMLM image. This imaging grid need not necessarily use the exact same side length l as the 3D grid which was used in the polymer simulation.

The motivation for the choice of these estimators is straightforward. If in some subregion of an IF-sSMLM image there exist l_α localizations, then

$$\hat{n}_\alpha = \frac{l_\alpha N_\alpha}{L_\alpha} \tag{4.16}$$

is a Maximum Likelihood Estimator for the number of nucleosomes within that same subregion, discounting the contribution of localization uncertainties. Since the grid-based contact function we employ directly relates the subpopulations of colocalized nucleosomes to numbers of internucleosome contacts, this estimator enables a relationship between colocation of fluorescent localizations to be approximately correlated with numbers of internucleosome contacts. The exact contribution of localization uncertainty to contact probability may further be explicitly calculated for any pair of localizations, as their distance is a Rice-distributed random variable.[151] However, the effect of uncertainty is not straightforward when counting contacts among more than two localizations, or positionally correlated localizations such as those studied here, and therefore approximations such as those in equation 4.15 are necessary.

The prefactor before the summation in Equation 4.14 renormalizes the weight of every 'detected' contact to account for the fact that the number of localizations associated with a mark is related to, but not equivalent to, the number of nucleosomes with that mark. As such, so long as the quantity of nucleosomes for any given mark is known (*e.g.*, through previous ChIP-seq and mass spectrometry data), this estimator accounts for variation in the degree of labelling between different fluorophores and temporal variation in localization frequency that may occur due to photobleaching throughout an experiment. More complex methods which link the number of localizations in a sSMLM image with the underlying number of bound target molecules were investigated, but ultimately were not found to significantly improve

on the accuracy of this simple method.

4.2.7 Polymer Model

We treat chromatin as a statistical block copolymer wherein each nucleosome is represented as an explicit bead, and each nucleosome bears exactly one of two possible epigenetic marks, α or β. In this paper, we follow the approach of MacPherson *et al.*[71] and partition the genome into 200-basepair (bp) bins, each representing a separate nucleosome, and represent each of these bins as a single particle. Although more marks or alternative representations could be used, it has been previously shown that this minimal representation is sufficient to induce heterochromatin-euchromatin microphase separation – the primary phenomenon of interest in this study.[71] This representation allows us to represent the sequence of the chromatin polymer as a binary-valued vector.

The potential energy function of our polymer model can be decomposed into bonded and nonbonded potential terms,

$$U(\vec{\mathbf{r}}) = U_b(\vec{\mathbf{r}}) + U_{nb}(\vec{\mathbf{r}}), \tag{4.17}$$

representing respectively the interactions between sequential nucleosomes along the chromosome backbone (*i.e.*, between the jth and the $(j+1)$th nucleosomes), and the net interactions between all other pairs of nucleosomes. We calculate $U_{nb}(\vec{\mathbf{r}})$ using the Theoretically-Informed Coarse Grained (TICG) model for block copolymers,[152] a polymer mean-field technique originally developed to study phase separation in polymer blends and block copolymers. A variation on this model has recently been used by MacPherson *et al.*[71] to model HP1-dependent phase separation behavior in chromatin. Non-bonded interactions in this model are calculated via a field-theoretic Hamiltonian based on the local density of the polymer calculated

over a cubic grid with side length $\sqrt[3]{V_c} = 29.7nm$:

$$U_{nb}(\vec{\mathbf{r}}) = V_c \sum_{c \in \text{grid}} \chi_s(\phi_{\alpha,c}(\vec{\mathbf{r}}) + \phi_{\beta,c}(\vec{\mathbf{r}}))^2 + \chi_{\alpha\beta}\phi_{\alpha,c}(\vec{\mathbf{r}})\phi_{\beta,c}(\vec{\mathbf{r}}) \tag{4.18}$$

where V_c is the volume of cell c, χ_s is the polymer-solvent interaction parameter representing the free energy between nucleosomes and the solvent (nucleoplasm), and $\chi_{\alpha\beta}$ is the Flory-Huggins interaction parameter between monomers of types α and β.[153,154] $\phi_{i,c}$ is the volume fraction of nucleosomes with mark i in grid cell c, which is calculated for each cell as

$$\phi_{\alpha,c}(\vec{\mathbf{r}}) = \sum_{i \leq N} \frac{\nu_{nuc}}{V_c} \mathbf{1}_c(\vec{\mathbf{r}}_i) \tag{4.19}$$

where ν_{nuc} is the volume of an individual nucleosome, and $\mathbf{1}_c(\vec{\mathbf{r}}_i)$ is an indicator function which equals 1 if $\vec{\mathbf{r}}_i$ is within the cell c, and 0 otherwise. Of all terms, only $\chi_{\alpha\beta}$ and χ_s are undetermined constants.

The value of $\chi_{\alpha,\beta}$ determines the propensity of chromatin with different marks to mix, and therefore is of critical importance to the phenomenon of phase separation. If $\chi_{\alpha,\beta}$ is positive, nucleosomes with different epigenetic marks effectively repel each other, leading to a preference for 'self-self' contacts throughout the polymer. If the value of $\chi_{\alpha,\beta}$ is sufficiently large and positive, the energy of this repulsion overcomes the entropic tendency toward mixing of components and, in block co-polymers and in polymer blends, the polymer solution phase-separates into separate regions enriched in different marks (α or β) compared to the bulk composition.[152,155] The representation of the nonbonded interaction of Eqn 4.18 enables the modeling of such nonbonded-interaction-induced phase separations, as well as thermodynamically complete density fluctuations, while remaining fast enough (computationally $\mathcal{O}(N)$) to permit the simulation of chromosome-scale segments of the genome ($\sim$ 100 Mbp

or generally $> 100,000$ nucleosomes).

Following the strategy of MacPherson *et al.*,[71] we model the bonded energy $U_b(\vec{\mathbf{r}})$ as a Discretized Shearable-Stretchable Worm-like Chain (DSSW), which is a variant of the traditional worm-like chain which enables more accurate reproduction of the persistence length characteristics of semi-rigid polymers.[156] The polymer-solvent interaction parameter, χ_s, sets the magnitude of density fluctuations throughout the simulated polymer, but its value is not known exactly for chromatin *in vivo*. Recent ChromEMT results[31] have confirmed that interphase chromatin density varies across a wide range from roughly $\sim 0\%$ to 50% by volume across individual nuclei, indicating that large, positive values of χ_s which would prevent significant density variation would not be physical.[31] We follow the insights of MacPherson *et al.*,[71] who note that a small, positive value of χ_s such as 1 (k_BT) enforces positive compressibility of the polymer, while still allowing the density to still vary significantly between different regions of the polymer. As such, we set χ_s to 1 k_BT for these simulations.

4.2.8 Epigenetic Sequence Generation

Epigenetic marks are not completely randomly arranged along the genome. By a variety of biological mechanisms, nucleosomes with any particular epigenetic mark tend to be found close to other nucleosomes with the same mark – *i.e.*, the probabilities of nearby nucleosomes having any particular epigenetic mark are correlated.[157–159] As the genomic distance between two nucleosomes increases, this correlation decreases in strength.

We reproduce this short-range sequence correlation using the treatment of 'random-block' statistical copolymers developed by Fredrickson *et al.*[160,161] and further developed by Khokhlov *et al.*.[162,163] In this treatment, the copolymer sequence is treated as the product of a Markov-step growth process in which the probability of a specific monomer type (*i.e.*, epigenetic mark) being present at position i is determined by the monomer type present

at position $i-1$, with a Markov transition probability matrix $\mathbf{p}$ given by

$$\mathbf{p} = \begin{bmatrix} p_{\alpha\alpha} & p_{\alpha\beta} \\ p_{\beta\alpha} & p_{\beta\beta} \end{bmatrix}. \tag{4.20}$$

This arises naturally in the case of synthetic copolymers produced by polymerization in a medium containing two or more reactive species, but the statistics which govern these sequences can be used to describe other polymers.[162,163]

Conveniently, for any given an average monomer fraction (*e.g.*, x_α, the mole fraction of monomer type α), this theoretical model has only one degree of freedom.[160] We fix this remaining degree of freedom by specifying λ, the nontrivial eigenvalue of $\mathbf{p}$, which may vary between -1 and 1. This λ, termed the sequence correlation parameter or the sequence blockiness, controls the short-range correlations of the sequence,[160,161] with high values resulting in sequences with successive long repeats of α or β and low or negative values resulting in alternation between the monomer types.[164] Though additional, longer-range sequence correlations can exist and can affect the thermodynamic properties of statistical polymers, it has been found that this model is sufficient to reproduce much of the effects of sequence inhomogeneity on statistical copolymer structure.[162,163]

The exact procedure by which we generate polymer sequences is as follows. First, we determine the elements of the matrix $\mathbf{p}$ which satisfies the given the values of x_α and λ by solving the linear system of constraints (*e.g.*, as given by Rumantysev *et al.*[164]):

$$x_\alpha p_{\alpha\alpha} + (1 - x_\alpha) p_{\alpha\beta} = x_\alpha \tag{4.21}$$

$$p_{\alpha\alpha} + p_{\beta\beta} + 1 = \lambda \tag{4.22}$$

$$p_{\alpha\alpha} + p_{\beta\alpha} = 1 \tag{4.23}$$

$$p_{\alpha\beta} + p_{\beta\beta} = 1 \tag{4.24}$$

The composition of the first monomer is chosen randomly with $p(\alpha) = x_\alpha$. Then, the composition of all remaining monomer units are determined by successive random number generation and comparison to the appropriate entries in the matrix $\mathbf{p}$. In this manner, many different copolymer sequences obeying a specified x_α and λ can be efficiently generated.

4.2.9 Parameter selection

Our chromatin model is, after the above simplifications, fully defined by a set of four undetermined parameters. Three, introduced previously, are $\chi_{\alpha,\beta}$, the Flory-Huggins interaction parameter, x_α, the monomer fraction of α-marked nucleosomes, and λ, the sequence correlation parameter of the chromatin backbone. The fourth, ρ, is the polymer volume fraction which fixes the number of cells in the grid for any chosen number of nucleosomes N, and thereby sets the average volume fraction across all gridcells – i.e.

$$\rho = \left\langle \phi_{\alpha,c}(\vec{\mathbf{r}}) + \phi_{\beta,c}(\vec{\mathbf{r}}) \right\rangle_{c \in grid} \tag{4.25}$$

However, these parameters do not take single, universal values which accurately describe chromatin in all contexts. ρ varies spatially and with the cell cycle,[31] while λ and x_α vary across different epigenetic marks and cell type.[80] $\chi_{\alpha,\beta}$ empirically represents the effects of both direct interactions between nucleosomes marked with α and β, and their indirect interactions mediated by trans-acting ligands such as HP1α,[71,165] whose population cannot be assumed constant across contexts. To capture this diversity in chromatin's characteristics, we generate polymer sequences and perform Monte Carlo simulation for 5,000 discrete combinations of values for each of these parameters chosen from within a range of biologically plausible values.

We determine plausible boundaries for each parameter as follows. x_α is a mole fraction which by definition ranges between 0 and 1; we additionally restrict this range to exclude the

homopolymer end-member systems ($x_\alpha = 0$ and $x_\alpha = 1$), which are not of interest in this study. The range for $\chi_{\alpha,\beta}$ was chosen empirically based on preliminary polymer simulations which identified that the most significant changes in copolymer morphology appeared to occur in the range $\chi_{\alpha,\beta} \in [-5, 10]$; we slightly expanded this range to $\chi_{\alpha,\beta} \in [-15, 15]$ to ensure further structural variation wasn't missed.

We base our bounds on ρ upon the findings of Ou *et al.* in [31], which found that in mitotic U2OS nuclei, the distribution of chromatin volume fraction is approximately gaussian between 12% and 52%, with less than 5% of regions having a density greater than 45%. From these results, we choose the limits of simulated ρ to 10% and 40%, with the slightly more restrictive upper bound due to numerical issues caused by chromatin volume fractions above approximately 40%.

To determine reasonable bounds of λ, we computed Maximum Likelihood Estimates of the value of λ from a set of experimental ChIP-Seq sequences, chosen by convenience as the 12 epigenetic marks listed in the ENCODE database for HCT116 cells. Following the strategy of MacPherson *et al.*,[71] we first pool the Chip-Seq read-counts into 200bp bins, representing approximately individual nucleosomes. We then binarize this Chip-Seq data according to a threshold which we adjust until the fraction of marked nucleosomes equals that determined by mass spectrometry.[80] For each resulting vector of epigenetic mark locations, we determined the Maximum Likelihood Estimate of λ given total fraction of that epigenetic mark. It was found that the values of λ across the 12 chosen epigenetic marks varied between 0.55 and 0.8, a range which we used identically to select candidate parameter values for sequence

generation. The resulting parameter ranges are

$$\chi_{\alpha,\beta} \in [-15, 15] \tag{4.26}$$

$$\rho \in [0.1, 0.45] \tag{4.27}$$

$$x_\alpha \in (0, 1) \tag{4.28}$$

$$\lambda \in [0.55, 0.8] \tag{4.29}$$

For each of the 5,000 simulations, we select individual values for the parameter set $\rho, \chi_{\alpha,\beta}, x_\alpha, \lambda$ by Latin Hypercube Sampling, which randomly samples the parameter space while also ensuring uniform coverage of subspaces and minimizing any spurious random correlation of the parameters.[166]

4.2.10 Monte Carlo Simulation

To ensure structural equilibration of the simulated chromatin polymer, we employ a mixture of four global and local Monte Carlo moves. Following the strategy of MacPherson *et al.*,[71] these moves are 1. Translation, 2. (End) Rotation, 3. Crankshaft, 4. Pivoting. We observe an autocorrelation time of $\tau \approx 800$ cycles for long-range interaction counts (*i.e.*, the product $\phi_i \phi_j$). Simulations are conducted for the lesser of 36,000,000 Monte Carlo cycles or 24 hours wall-time, typically $\sim 360,000$ cycles, or $t_{sim} \approx 450\tau$, to ensure equilibration of slow structural modes.

4.3 Results

4.3.1 Charaterizing the range of simulation results.

To characterize the range of computational results across all 5,000 simulations, we use the spatial contact matrix calculate the observed molar contact ratio for each simulation, and

plot this as a function of each of the four simulation parameters defined previously ($\chi_{\alpha\beta}$, x_α, ρ, and λ). As shown in Figure 4.2A-B, molar contact ratio is most heavily influenced by the heterogeneous interaction strength $\chi_{\alpha\beta}$ and polymer composition x_α, and is much less affected by polymer density and sequence blockiness. Below an interaction strength of approximately $2.5k_BT$ (where k_B is the Boltzmann constant), the molar contact ratio is almost exclusively determined by the composition of the polymer. At and above interaction strengths of $5k_BT$, the molar contact ratio decreases as repulsion increases as expected. However, no discernible trend between molar contact ratio and either polymer density ρ or sequence blockiness λ can be readily determined.

To quantify if a phase separation may be occurring at high interaction strengths, we calculate an order parameter as which we define as:

$$S = R\left(\beta,\alpha\right) R(\alpha,\beta) = \frac{C\left(\alpha,\beta\right)^2}{C(\alpha,\alpha)C(\beta,\beta)} \tag{4.30}$$

For all systems, $S \in [0,1]$. $S = 1$ implies a compositionally homogeneous system, while a $S = 0$ describes a system which is perfectly phase separated, in which there is no spatial overlap between monomers of different types. The values of S for all simulations are shown in Figure 4.2C-D, plotted against the same simulation parameters. As before, the system order is almost wholly determined by the heterogeneous interaction strength $\chi_{\alpha\beta}$ and polymer composition. For simulations where $\chi_{\alpha\beta} < 5$, there exists a wide range of compositions centered on $x_A = 0.5$ for which the system is mostly homogeneous. For $\chi_{\alpha\beta} \geq 5$, the system is heterogeneous and remains separated for all compositions. There is no discernible relationship between this order parameter and either polymer density or sequence blockiness.

4.3.2 Validating Analytical Molar Contact Ratio Prediction

We performed the procedure described previously to estimate the molar contact ratio from sSMLM images derived from each of the 5,000 simulations. For each simulation, 10 evenly-spaced simulation snapshots were selected for imaging, 10 imaging parameters were randomly generated, and an sSMLM image was generated for each pair of frame and imaging parameter, for a total of 500,000 images. To assess the accuracy of this analytical method, we show in Figure 4.3 the value of these predictions plotted against the ground truth contact ratio calculated from the spatial Hi-C Matrix via Equation 4.11. The majority of data points are extremely well-predicted, having very small residuals from the line of zero residuals. However, there also exists a large body of data points for which the analytical prediction vastly overestimates the contact ratio. To explain this, when these data are colored by the order parameter introduced earlier (also show in Figure 4.3), it is revealed that this phase separation order parameter is extremely predictive of the estimation error: the contact ratio is greatly overestimated in strongly phase separated systems, and is accurately predicted in homogeneous systems.

Furthermore, this relationship between prediction error and the order parameter S is highly systematic (*i.e.*, nonrandom), which suggests that knowledge of the order parameter itself may be used to correct the prediction. We show the result of such a correction in in Figure 4.4A, where we scale each predicted contact ratio by S^{γ}, where $\gamma = 0.65$, and is a constant which was tuned manually until the manifold of predictions appeared to collapse onto the line of zero residuals. This demonstrates that *a priori* knowledge of this order parameter can be used to correct the analytical prediction, resulting in extremely high accuracy. We also show the same correction applied using the estimated order parameter $\hat{S}$, defined as

$$\hat{S} = \hat{R}(\beta, \alpha)\hat{R}(\alpha, \beta) \tag{4.31}$$

and derived purely from the IF-sSMLM images. However, as shown in Figure 4.4B, if we have no *a priori* knowledge of the order parameter and must instead estimate it directly from the results of the analytical prediction, the resulting correction performs very poorly.

4.3.3 Machine Learning-based Contact Ratio Prediction

This prediction difficulty in highly phase-separated systems raises a fundamental question: do two-color/ two-mark IF-sSMLM images of these systems simply contain enough information to predict the molar contact ratio beyond the limits achieved analytically in Figure 4.4? That is, can any method of analysis perform significantly better at predicting contact map characteristics than we have achieved here? To begin to answer this question, we design and train a convolutional neural network (CNN) to predict molar contact ratio from sSMLM images, aided by our previous analytical estimation, as well as two other measures: the polymer density ρ and the interaction strength $\chi_{\alpha\beta}$. Though the failure of the CNN would not prove this estimation task impossible, a success of the CNN would prove that sufficient information *is* encoded in the sSMLM images to accurately determine the contact ratio, even if a broadly applicable analytical method still eludes us.

We base the structure of our CNN on that of VGG-16, which was originally developed for general image classification.[167] We replace VGG-16's terminal softmax layer with a fully-connected layer to ensure the CNN performs regression instead of classification. As shown in Figure 4.6A, we also add a secondary input layer after all convolutional layers to provide additional engineered features to the network which may improve prediction. Three alternative versions of this secondary input layer were designed and tested: 'Fully Featured (FF)', 'Minimally Featured (MF)', and 'Non-featured (NF)'. The FF network was provided with a vector of 6 features which were calculated before training and appended to each image in the training set: $[\hat{R}(\beta,\alpha),\hat{R}(\alpha,\beta),\hat{S},x_a,x_b,\lambda]$, under the reasoning that these quantities have the possibility of being estimated when the IF-sSMLM data is combined with ChIP-Seq data.

The MF network included a more restrictive set of only 3 features: $[\hat{R}(\beta,\alpha),\hat{R}(\alpha,\beta),\hat{S}]$, which may all be analytically determined from the IF-sSMLM images via equations 4.14, 4.15 and 4.31 without any additional experiments. Lastly, to verify if the above features improved the network's prediction accuracy, the NF network was designed to have no secondary input layer. All networks used an output layer with four nodes, and the network was trained with a Mean-Squared Error loss function against a set of 4 output values: $R_{sim}(\beta,\alpha)$, $R_{sim}(\alpha,\beta)$, $\chi_{\alpha\beta}$, and ρ, which were derived from the input parameters for each simulation corresponding to each image. All activation functions are rectified linear (ReLU) except for the final linear layer.

Of the 500,000 images which were generated, 400,000 were used for training and 100,000 were used for evaluating test performance. Learning rates were optimized by using a logarithmic sweep from 0.01 to 1e-6, and the learning rate which trained to the smallest minimum testing error was used for further analysis.

After 900 training epochs, the learning-rate-optimized FF, MF and NF networks achieved a whitened RMSE testing error of 0.593, 0.599, and 2.15 respectively, compared to 4.0 expected for a random network. The significant improvement in error of the Minimally Featured network over the Non-Featured network demonstrates that the inclusion of the analytical estimates ($\hat{R}(\beta,\alpha)$,$\hat{R}(\alpha,\beta)$, and $\hat{S}$) as a feature significantly improves the network's ability to learn the correct mapping between image and molar contact ratio. Furthermore, the significantly smaller improvement in RMSE of the Fully-Featured network over the Minimally-Featured network suggests that the inclusion of the additional sequence-based features is not meaningful in improving prediction accuracy of the molar contact ratio. Furthermore, we show in Figure 4.6B that the prediction accuracy of the MF network on the testing set achieves a RMSE on $R_{sim}(\beta,\alpha)$ prediction of less than 0.5 over the majority of data points. The MF network also demonstrates an increase in error when performing predictions on high $R_{sim}(\beta,\alpha)$ images, but this increase is marginal compared to that observed for the analytical

prediction as seen in Figure 4.3.

4.3.4 Predicting Hi-C Matrices from sSMLM Images

Beyond the prediction of the molar contact ratio, these methods may furthermore be used to derive an approximation of the Hi-C matrix itself. This may be valuable in cases such as the analysis of ensembles of single cells for which single-cell Hi-C is not feasible, or to estimate variation of the Hi-C matrix in time when paired with live-cell imaging data.

Inspired by the form of Equation 4.11, we define $\boldsymbol{A}_{pairwise}$ as a sum of contributions from each pair of epigenetic marks:

$$\boldsymbol{A}_{pairwise} = \sum_{\alpha \in \mathbb{M}} \sum_{\beta \geq \alpha} v_{\alpha\beta} \vec{\mathbf{x}}_\alpha \vec{\mathbf{x}}_\beta^\top \tag{4.32}$$

where $v_{\alpha\beta}$ is a scalar constant for each pair of α, β in the ordered set of marks $\mathbb{M}$ such that $\beta \geq \alpha$. We can determine the values of each $v_{\alpha\beta}$ by relating these to a set of measured or predicted molar contact ratios by substituting Eqn 4.32 into Eqn 4.11:

$$R_a(b,a) = \frac{\vec{\mathbf{x}}_a^T \left(\sum_{\alpha \in \mathbb{M}} \sum_{\beta \geq \alpha} v_{\alpha\beta} \vec{\mathbf{x}}_\alpha \vec{\mathbf{x}}_\beta^\top \right) \vec{\mathbf{x}}_b}{\vec{\mathbf{x}}_a^T \left(\sum_{\alpha \in \mathbb{M}} \sum_{\beta \geq \alpha} v_{\alpha\beta} \vec{\mathbf{x}}_\alpha \vec{\mathbf{x}}_\beta^\top \right) \vec{\mathbf{x}}_a} \tag{4.33}$$

$$= \frac{\sum_{\alpha \in \mathbb{M}} \sum_{\beta \geq \alpha} v_{\alpha\beta} \vec{\mathbf{x}}_a^T \vec{\mathbf{x}}_\alpha \vec{\mathbf{x}}_\beta^\top \vec{\mathbf{x}}_b}{\sum_{\alpha \in \mathbb{M}} \sum_{\beta \geq \alpha} v_{\alpha\beta} \vec{\mathbf{x}}_a^T \vec{\mathbf{x}}_\alpha \vec{\mathbf{x}}_\beta^\top \vec{\mathbf{x}}_a} \tag{4.34}$$

For a system with M different marks, this yields a total of $\frac{1}{2}\left(M^2 + M\right)$ simultaneous linear equations and $\frac{1}{2}\left(M^2 + M\right)$ unknown scalars. We solve for all $v_{\alpha\beta}$ values by matrix inversion, and substitute into equation 4.32 to construct each Hi-C Matrix. A typical reconstructed Hi-C matrix is shown in Figure 4.6B. Notably, equation 4.32 will be most accurate in approximating the simulated Hi-C matrix for systems where all genetic loci in the underlying chromatin polymer engage only in pairwise additive interactions. Furthermore, this

equation does not enforce a power law dependence of the contact probability – though it may be trivially modified to enforce one, if there is reason to believe that a particular power law would best describe a given system.

To quantitatively compare the Hi-C matrices resulting from this method with the ground-truth simulated observable, we calculate the Pearson's Correlation Coefficients (PCC) and Stratified Correlation Coefficients (SCC) for all pairs of predicted/observed matrices from the aforementioned polymer simulations which were in the holdout set of the CNN training, *i.e.*, which were not used for training the CNN.[139,168] To establish a meaningful range of SCC and PCC values, we also calculate these correlation coefficients between simulation replicas (*i.e.*, different Monte Carlo initial seeds), and a negative control in which a Hi-C Matrix was constructed with no plaid pattern, but which has the exact genetic distance dependence (*i.e.*, power law) as the ground-truth. To these, we compare both the $\boldsymbol{A}_{pairwise}$ calculated with the analytical contact ratio estimation, and the $\boldsymbol{A}_{pairwise}$ calculated from the CNN contact ratio estimation.

The distributions of these correlation coefficients are shown in Figure 4.6C. Despite the statistical similarity of each simulation performed with a different random seed, these simulations only achieve a median SCC of 0.26, showing a significant influence of finite sampling length on this measure, and stressing the importance of averaging structural properties over multiple random seeds, even after extensive relaxation of polymer simulations. Both the CNN and the analytical prediction achieve median SCCs comparably large (0.14 and 0.16 respectively). Though the Hi-C prediction accuracy of the CNN is unexpectedly lower than that of the analytical prediction, the difference is not statistically significant at this sample size ($p = 0.08$). As has been noted by other authors, the PCC is a poor measurement of Hi-C similarity, and is heavily dominated by the similarity of the genomic distance dependence of the Hi-C matrix.[139,168]

4.4 Discussion

Of primary interest in this system is the propensity of chromatin-like statistical polymers to phase-separate under different biological conditions. Although the average block length of charged historical copolymers has been shown to have significant impact on their phase characteristics, we have found that the limited range of sequence blockiness for biological sequences ($\lambda \in [0.5, 0.9]$) largely prevents this parameter from having any significant effect on phase separation. Instead, we find that phase separation in our system is determined almost exclusively by the heterogeneous interaction parameter $\chi_{\alpha\beta}$ and the monomer fraction x_α, with very little influence from sequence blockiness λ or polymer density ρ. Additionally, The polymer system appears to undergo virtually all structural change which has impact on Hi-C descriptive statistics within the range $-5 < \chi_{\alpha\beta} < 10$, though further investigation would be necessary to determine the exact details such as the critical solution conditions of this polymer system.

Although there exist comparable methods in the literature, we do believe that the analyses performed here provide unique advantages over these alternatives. In the work by Xu *et al.*,[129] the composition of chromatin 'nanoclusters' is analyzed by examination of the self-self Radial Distribution Function (RDF) of each mark. However, this work primarily examines binary 'colocalization' of different epigenetic marks, and does not approach the topic of quantitative comparison to Hi-C experiments.

Bintu *et al.*[127] achieved significantly higher accuracy can be achieved on Hi-C prediction in single cells than is demonstrated here. However, this experimental method requires extensive sequential hybridization and super resolution imaging, which may be to experimentally intensive for many applications. Our analysis methods would be most appropriate when faster, cheaper analysis is needed, and a lower level of accuracy can be tolerated.

We acknowledge that there are several means by which these analysis methods could be improved. Although we have investigated their performance across a wide range of simulation

and imaging conditions, it is not immediately certain how well they generalize to different polymer models. Coarse-grained polymer models for chromatin span a wide range of resolutions and approaches to key structural properties, including the long-range interactions which drive phase separation and the genomic distance dependence of contact frequency (*i.e.*, power law); further work is needed to demonstrate that these methods retain high accuracy in these other systems. Furthermore, our analysis was restricted to chromatin systems in which nucleosomes have exactly one of 2 possible epigenetic marks. Although these methods performed well even on chromatin with compositions far from equimolar, an extension of these methods to systems with multiple marks of interest may be very useful for future scientific inquiry.

This system of analysis reveals a new approach to single cell contact mapping more broadly. Methods which produce Hi-C contact maps at a single cell level have gained significant interest in recent years,[114,122,169] as they provide the hold to reveal cell-to-cell variability in chromatin folding states, a structural basis for transcriptional variation, and more. Classical single-cell Hi-C techniques[113] are limited compared to bulk Hi-C techniques by their inherently low read-counts, which makes it difficult to identify key structural features including TADs and heterochromatin/euchromatin domains. Recently, it has been shown that 3D STORM (Stochastic Optical Reconstruction Microscopy) imaging, combined with high-throughput sequential hybridization with fluorescent probes has Significant promise to reproduce Hi-C contact maps on a single cell basis – bypassing the read-count limit of classical techniques.[127]

This work approaches these techniques via an alternative avenue, quantitatively estimating the strength of heterochromatin/euchromatin phase separation on a single cell basis via sSMLM imaging with no sequential hybridization. When combined with other information available about the genome of interest, this method also yields a single-cell Hi-C contact map, though its reproduction provides useful information about chromatin struc-

ture at very different length scales than other methods – focusing on long-range rather than short-range interactions. We believe that future methods may be able to take advantage of both categories of insights, providing information on long-range chromatin phase separation as demonstrated in this work, in addition to short range contact features analyzable by other methods.

4.5 Conclusions

In this work, we have investigated the variation of Hi-C contact matrices by probing a convenient polymer model, and have characterize the variation of key descriptive statistics regarding these matrices as a function of the undetermined parameters of that model. We have also derived analytical and machine learning based methods for characterizing, and directly reproducing these Hi-C matrices from *in silico* sSMLM images. We see this work as providing a framework in which sSMLM analysis can compliment Hi-C analysis as part of broader scientific inquiry, and we believe that it provides a benchmark for comparison of future analytical methods.

4.6 Figures

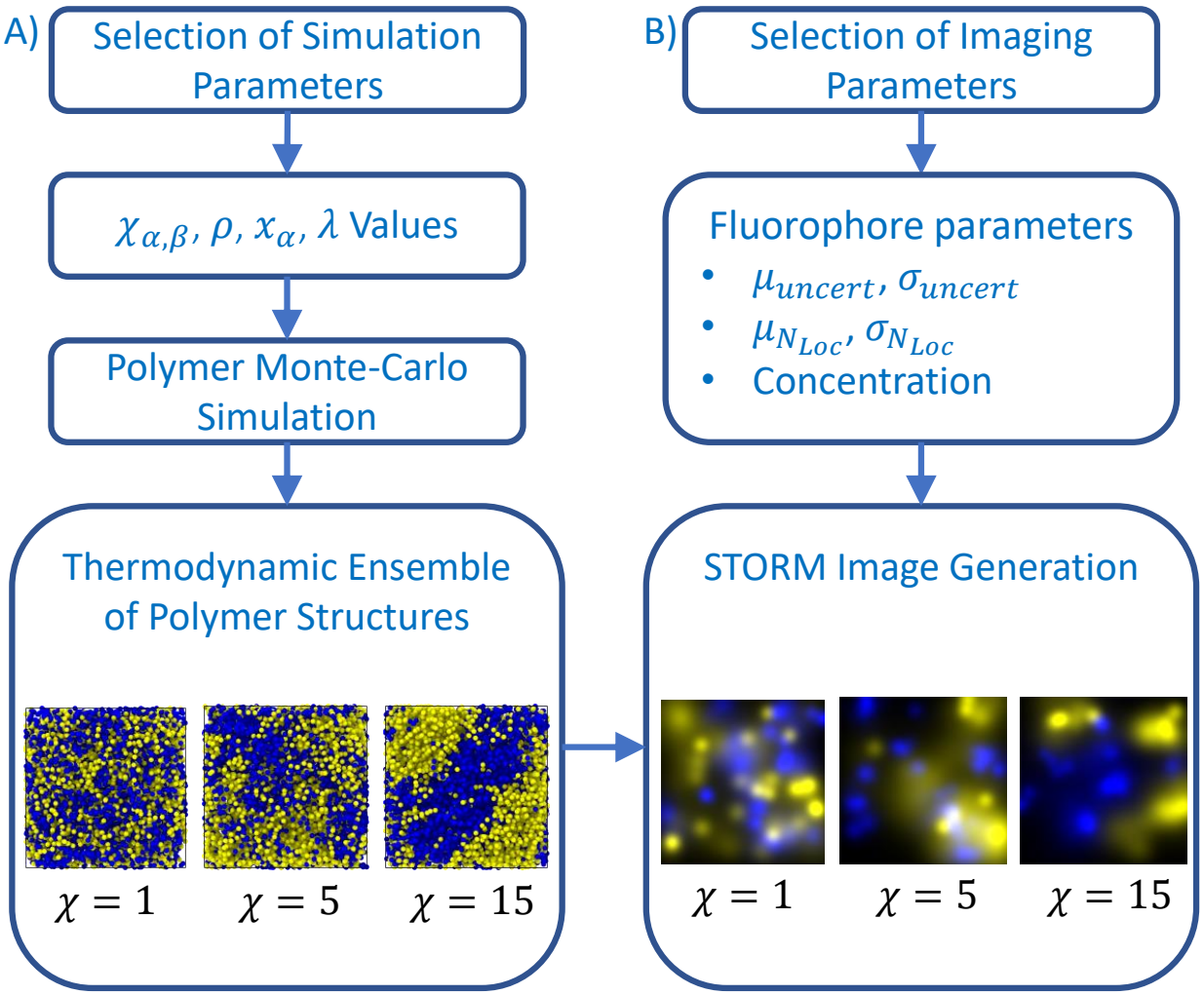

Figure 4.1: The overall simulation and analysis workflow. Separate boxes represent meaningful milestones in the process. **A) Molecular simulation workflow** To characterize the full range of structures feasible in the given polymer model, we first select values for the 4 simulation parameters across a wide biologically-relevant range. Once the values are selected and the polymer sequence is generated, Polymer Monte Carlo simulation is performed for the system until the simulation reaches the specified number of timesteps. Structures from throughout the simulation duration are combined to reflect the full statistical ensemble of chromatin structures. **B) Imaging analysis workflow.** To reflect the range of potential imaging conditions, values for all imaging parameters are selected from a range derived from comparable experiments. Individual simulated chromatin structures are selected for imaging. Lastly, the *in silico* sSMLM imaging procedure is performed, producing super-resolution images for every pair of chromatin structures and imaging parameters.

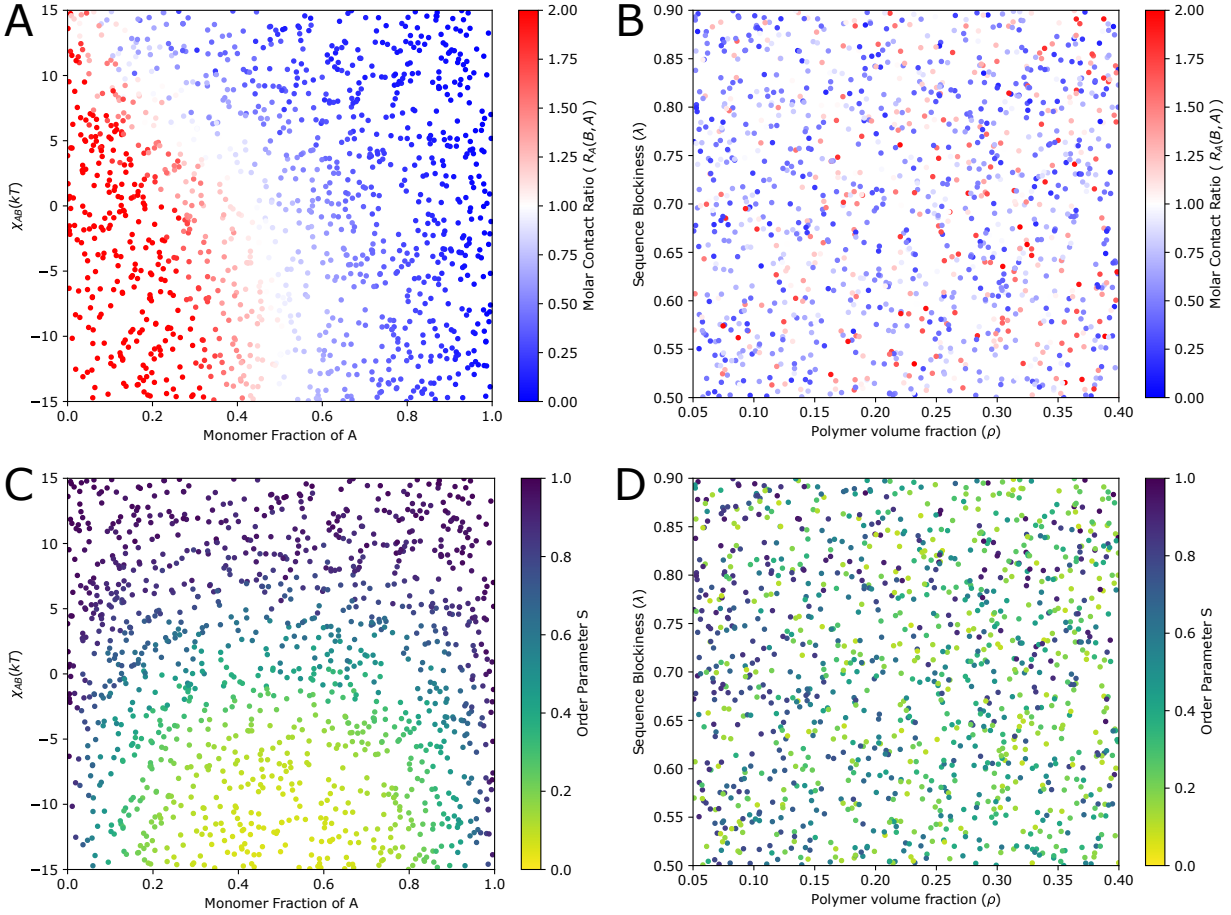

Figure 4.2: Trends of molar contact ratio and order parameter S with simulation parameters. **A)** and **B)** show the trend in the molar contact ratio $R_A(B, A)$ as a function of the four key simulation parameters: χ_{AB}, monomer fraction of a (x_A), Sequence blockiness (λ), and polymer volume fraction (ρ). **A)** clearly shows that contact ratio is strongly controlled by χ_{AB} and x_A. Below approximately $\chi_{AB} = 0$, the contact ratio is below 1 if and only if B is the most abundant monomer species (*i.e.*, $x_A > 0.5$). In the case of repulsive interactions ($\chi_{AB} > 0$), the space of systems with $R_A(B, A)$ expands, as mixing between different monomer types becomes increasingly energetically disfavorable. **B)** shows that there is little, if any, discernable correlation between the molar contact ratio and either ρ or λ. **C)** and **D)** show trends of the order parameter S as functions of the same four simulation parameters. Similarly, χ_{AB} and x_A strongly control this phase-separation order parameter, where the fully homogeneous systems are restricted to systems with attractive A-B interactions$\chi_{AB} < 0$ and near-equimolar quantities of monomers A and B ($x_A \approx 0.5$).

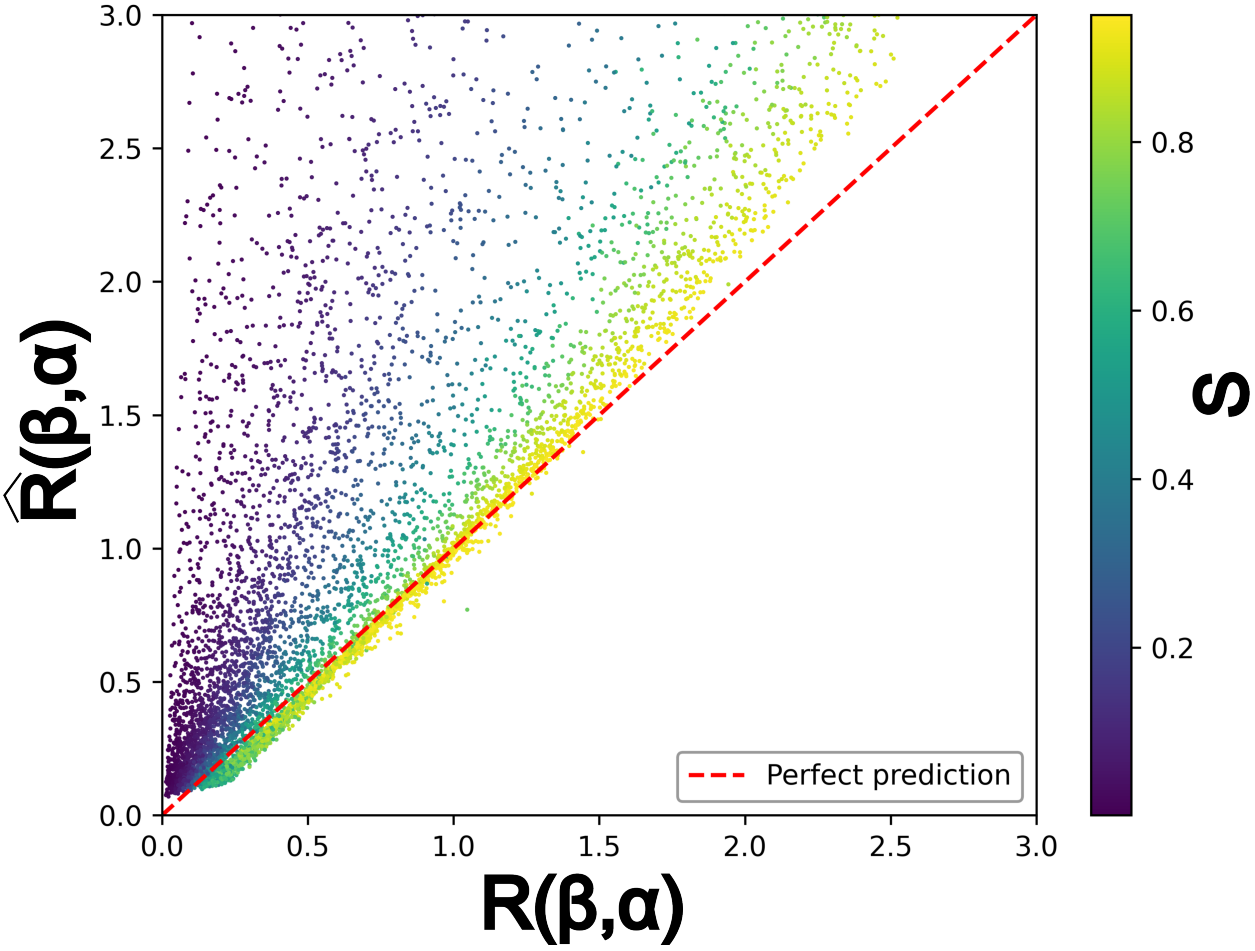

Figure 4.3: Validation scatterplot of the analytical predictions. Each point represents the true contact ratio averaged over each simulation, and the median of all contact ratios analytically estimated from all sSMLM images generated from structures resulting from that simulation. Each point is colored by the value of the order parameter S for the corresponding simulation. As can be clearly seen, the residual between each prediction and the line of perfect prediction (red dashed line) is a nearly monotonic function of this order parameter – high values of S correspond to accurate estimation, and low values of S correspond to overestimation. Systems with compositions $x_\alpha < 0.05$ or $x_\alpha > 0.95$ have been omitted for visual clarity.

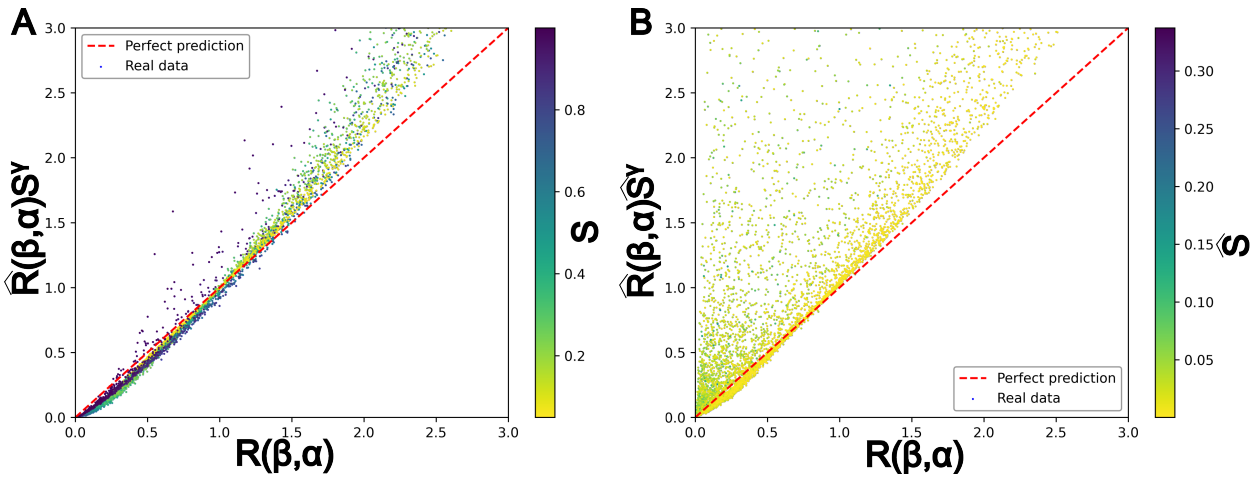

Figure 4.4: Corrected analytical predictions. **A) Correction via ground-truth**. Each analytical prediction, as shown in Figure 4.3, is multiplied by S^{γ}, with S calculated directly from simulated contact frequencies. The resulting correction where $\gamma = 0.65$ is shown. Larger exponents result in underestimation of the contact ratios for highly phase-separated systems, but slightly better estimation for extremely high contact ratio systems (*i.e.*, $R_{\alpha}(\beta, \alpha) \geq 2.0$). **B) Correction without additional characterization**. Each analytical prediction is multiplied by $\hat{S}^{\gamma}$, using the median across all sSMLM images of the analytically-estimated values of $R_{\alpha}(\beta, \alpha)$ and $R_{\beta}(\alpha, \beta)$. Each prediction is colored by this estimated $\hat{S}$. No value of γ resulted in meaningful correction of the predictions, since there is no apparent trend between estimated $\hat{S}$ and prediction accuracy. The correction shown uses $\gamma = 100$.

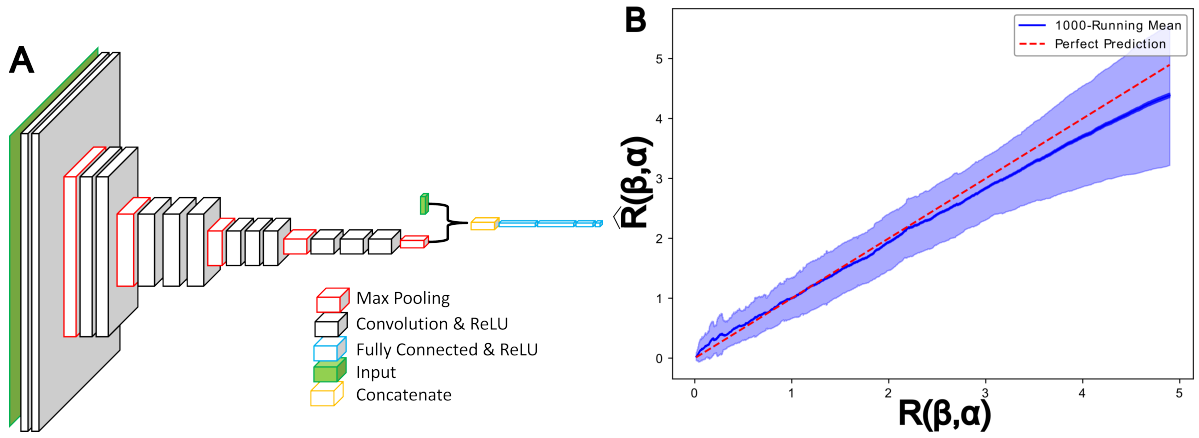

Figure 4.5: Convolutional Neural Network design and validation. **A) Structure of the CNN.** In addition to the convolutional elements which are unchanged from VGG-16,[167] an additional layer of inputs, corresponding to the set $\{x_\alpha, R_\alpha(\beta,\alpha), R_\beta(\alpha,\beta)\}$ for each sSMLM image, is added before the fully-connected layers. Furthermore, the final layer of the CNN is changed from a softmax layer to a fully-connected layer. **B) Validation accuracy.** To demonstrate the reproducibility of the CNN estimation, plot the average of all predictions in a 1000-element sliding window over the validation set, ordered by true contact ratio (solid blue line). An envelope with width $\pm 1\sigma$ within this sliding window is shown in light blue. A 95% confidence interval (t-interval) on the value of the mean of each sliding window is shown in dark blue. For all contact ratios across all simulations, the mean across each sliding window is within $\pm 1\sigma$ of the line of perfect prediction, implying extremely high prediction accuracy across all systems studied here.

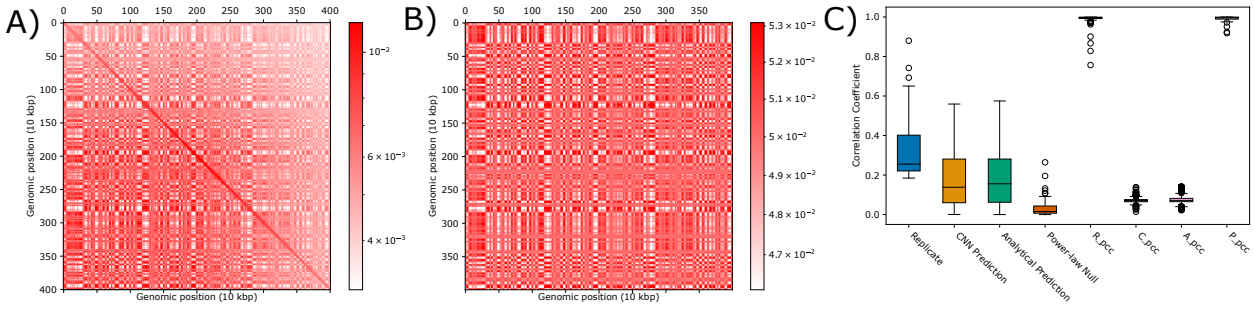

Figure 4.6: Comparison of simulated vs. predicted Hi-C maps. **A) Simulated Hi-C map. B) Typical Hi-C map predicted from STORM analysis.** Since $\boldsymbol{A}_{pairwise}$ does not include a genomic distance dependence, there is no stripe of enriched contact frequency along the main diagonal. Despite this, the overall contact map shows marked similarity with the simulation results shown in A. **C) Correlation coefficients of the Hi-C Matrix prediction.** SCC and PCC values calculated between the consensus simulated Hi-C matrix averaged across three random seeds, and the Hi-C matrices determined by each method discussed in text. Notably, the median SCC between individual simulations is within the interquartile range of both the Hi-C matrices predicted using both the CNN-based and the analytical predictions. All three lie significantly above the 'null' prediction of a Hi-C matrix with no plaid pattern.

CHAPTER 5

CONCLUSIONS

5.1 Summary of contributions

This dissertation investigates the structural insights into chromatin which may be gained through combinations of molecular simulation and experimental data. We have used enhanced sampling techniques to shed light on the thermodynamic stability of mesoscale chromatin fiber structures, and the relationship between stable, disordered structures and the ideal models which they have been theorized to resemble. This work furthermore levered these results to enable the Bayesian inference of nucleosome repeat lengths from *in situ* measurements of chromatin fiber diameters.

Subsequently, we used these insights into the structural disorder of chromatin fibers to characterize the structural effects of histone acylations. Our work shows that H4K16 acylations all have similar impacts on chromatin fiber structures compared to unmodified chromatin, with acetylated, butyrylated, and propionylated chromatin forming increasingly compact fibers in that respective order. Our simulations also show that chromatin fiber dynamics are dominated by slow, rare transitions into more compact fiber arrangements, commensurate with prior experimental reports.

Lastly, we demonstrated a method to directly impute Hi-C contact maps based on IF-sSMLM imaging of chromatin. This technique is highly accurate in reproducing the heterochromatin-euchromatin phase separation, and its accuracy is further enhanced through the use of machine learning techniques. Overall, our work serves as an example of how novel simulation techniques and modes of analysis can significantly enrich scientific insights from experimental data beyond the measurement resolution of those data.

5.2 Future Directions

Significant future directions of research based upon this work include the incorportation of free energetic details of short chromatin fibers into coarse grained polymer models for chromatin. By providing computational mechanisms for the coupling of epigenetic compaction at large scales and short-range structures, researchers may be able to glean insight into the relationship between short-range structural modifications such as the binding of linker histones and sliding of nucleosomes along the DNA backbone with the liquid-liquid phase separations which drive nuclear subcompartmentalization. We believe that the work presented here lays the foundation for such future investigations.

BIBLIOGRAPHY

[1] François Spitz and Eileen E.M. Furlong. "Transcription factors: From enhancer binding to developmental control". In: 13.9 (Sept. 2012), pp. 613–626. ISSN: 14710056. DOI: 10.1038/nrg3207.

[2] Albert J. Keung, J. Keith Joung, Ahmad S. Khalil, and James J. Collins. "Chromatin regulation at the frontier of synthetic biology". In: 16.3 (Mar. 2015), pp. 159–171. ISSN: 14710064. DOI: 10.1038/nrg3900.

[3] Sandy L. Klemm, Zohar Shipony, and William J. Greenleaf. "Chromatin accessibility and the regulatory epigenome". In: *Nature Reviews Genetics* 20.4 (4 Apr. 2019), pp. 207–220. ISSN: 14710064. DOI: 10.1038/s41576-018-0089-8.

[4] C. David Allis and Thomas Jenuwein. "The molecular hallmarks of epigenetic control". In: 17.8 (Aug. 2016), pp. 487–500. ISSN: 14710064. DOI: 10.1038/nrg.2016.59.

[5] Hyun Jung Kim and Suk Chul Bae. "Histone deacetylase inhibitors: Molecular mechanisms of action and clinical trials as anti-cancer drugs". In: 3.2 (2011), pp. 166–179. ISSN: 19438141.

[6] Azam Moosavi and Ali Motevalizadeh Ardekani. "Role of epigenetics in biology and human diseases". In: 20.5 (Nov. 2016), pp. 246–258. ISSN: 1028852X. DOI: 10.22045/ibj.2016.01.

[7] Ada L. Olins and Donald E. Olins. "Spheroid chromatin units (v bodies)". In: *Science* 183.4122 (4122 Jan. 1974), pp. 330–332. ISSN: 00368075. DOI: 10.1126/science.183.4122.330.

[8] J T Finch and A Klug. "Solenoidal model for superstructure in chromatin." In: *Proc. Nat. Acad. Sci. USA*. 73.6 (6 Apr. 1976), pp. 1897–1901. ISSN: 0027-8424. DOI: 10.1073/pnas.73.6.1897.

[9] James D. Watson. *Molecular biology of the gene*. Benjamin/Cummings, 1987, p. 26.

[10] Peter J. Horn and Craig L. Peterson. "Molecular biology: Chromatin higher order folding: Wrapping up transcription". In: 297.5588 (Sept. 2002), pp. 1824–1827. ISSN: 00368075. DOI: 10.1126/science.1074200.

[11] Karolin Luger, Armin W Mader, Robin K Richmond, David F Sargent, and Timothy J Richmond. "Crystal structure of the nucleosome resolution core particle at 2.8 A". en. In: *Nature* 389.6648 (6648 Jan. 1997), pp. 251–260. ISSN: 00280836. DOI: 10.1038/38444.

[12] Timothy J. Richmond and Curt A. Davey. "The structure of DNA in the nucleosome core". In: *Nature* 423.6936 (May 2003), pp. 145–150. ISSN: 00280836. DOI: 10.1038/nature01595.

[13] Thomas Schalch, Sylwia Duda, David F Sargent, and Timothy J Richmond. "X-ray structure of a tetranucleosome and its implications for the chromatin fibre." en. In:

Nature 436.7047 (July 2005), pp. 138–41. ISSN: 1476-4687. DOI: 10.1038/nature03686.

[14] K. Van Holde and J. Zlatanova. "Chromatin higher order structure: Chasing a mirage?" In: 270.15 (Apr. 1995), pp. 8373–8376. ISSN: 00219258. DOI: 10.1074/jbc.270.15.8373.

[15] Donald E. Olins and Ada L. Olins. "Chromatin history: Our view from the bridge". In: 4.10 (Oct. 2003), pp. 809–814. ISSN: 14710072. DOI: 10.1038/nrm1225.

[16] Philip J J Robinson and Daniela Rhodes. "Structure of the '30 nm' chromatin fibre: a key role for the linker histone." In: *Curr. Opin. Struc. Biol.* 16.3 (June 2006), pp. 336–43. ISSN: 0959-440X. DOI: 10.1016/j.sbi.2006.05.007.

[17] J D McGhee, J M Nickol, G Felsenfeld, and D C Rau. "Higher order structure of chromatin: orientation of nucleosomes within the 30 nm chromatin solenoid is independent of species and spacer length". eng. In: *Cell* 33.3 (1983), pp. 831–841. ISSN: 0092-8674. DOI: 10.1016/0092-8674(83)90025-9.

[18] R A Horowitz, D A Agard, J W Sedat, and C L Woodcock. "The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon." en. In: *The Journal of Cell Biology* 125.1 (Jan. 1994), pp. 1–10. ISSN: 0021-9525, 1540-8140. DOI: 10.1083/jcb.125.1.1.

[19] Jan Bednar, Rachel A Horowitz, Sergei A Grigoryev, Lenny M Carruthers, Jeffrey C Hansen, Abraham J Koster, and Christopher L Woodcock. "Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin". en. In: *Proceedings of the National Academy of Sciences* 95.24 (June 1998), pp. 14173–14178. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.95.24.14173.

[20] Michael Smith, Brian Athey, Shawn Williams, and John Langmore. "Radial density distribution of chromatin: evidence that chromatin fibers have solid centers". In: *The Journal of Cell Biology* 110.2 (Jan. 1990), pp. 245–254. ISSN: 0021-9525.

[21] B D Athey, M F Smith, D A Rankert, S P Williams, and J P Langmore. "The diameters of frozen-hydrated chromatin fibers increase with DNA linker length: evidence in support of variable diameter models for chromatin." en. In: *The Journal of Cell Biology* 111.3 (Jan. 1990), pp. 795–806. ISSN: 0021-9525, 1540-8140. DOI: 10.1083/jcb.111.3.795.

[22] S P Williams, B D Athey, L J Muglia, R S Schappe, A H Gough, and J P Langmore. "Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length". eng. In: *Biophysical Journal* 49.1 (1986), pp. 233–248. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(86)83637-2.

[23] Isabel Garcia-Saez, Hervé Menoni, Ramachandran Boopathi, Manu S. Shukla, Lama Soueidan, Marjolaine Noirclerc-Savoye, Aline Le Roy, Dimitrios A. Skoufias, Jan Bed-

nar, Ali Hamiche, Dimitar Angelov, Carlo Petosa, and Stefan Dimitrov. "Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation". In: *Molecular Cell* 72.5 (Dec. 2018), 902–915.e7. ISSN: 10972765. DOI: 10.1016/j.molcel.2018.09.027.

[24] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, and G. Li. "Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units". In: *Science* 344.6182 (6182 Apr. 2014). Tetranucleosome Structure: Zig zag, pp. 376–380. ISSN: 0036-8075. DOI: 10.1126/science.1251413.

[25] Andrew Routh, Sara Sandin, and Daniela Rhodes. "Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure." In: *Proceedings of the National Academy of Sciences of the United States of America* 105.26 (26 July 2008), pp. 8872–7. ISSN: 1091-6490. DOI: 10.1073/pnas.0802336105.

[26] J Widom. "A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells." In: *Proceedings of the National Academy of Sciences of the United States of America* 89.February (1992), pp. 1095–1099. ISSN: 0027-8424. DOI: 10.1073/pnas.89.3.1095.

[27] Bruno Beltran, Deepti Kannan, Quinn Macpherson, and Andrew J. Spakowitz. "Geometrical Heterogeneity Dominates Thermal Fluctuations in Facilitating Chromatin Contacts". In: *Physical Review Letters* 123.20 (Nov. 2019), p. 208103. ISSN: 10797114. DOI: 10.1103/PhysRevLett.123.208103.

[28] Sergei A. Grigoryev, Gaurav Arya, Sarah Correll, Christopher L. Woodcock, and Tamar Schlick. "Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions". In: *Proc. Nat. Acad. Sci. USA.* 106.32 (May 2009), pp. 13317–13322. ISSN: 0027-8424. DOI: 10.1073/pnas.0903280106.

[29] Rosana Collepardo-Guevara and Tamar Schlick. "Chromatin fiber polymorphism triggered by variations of DNA linker lengths". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.22 (June 2014), pp. 8061–8066. ISSN: 10916490. DOI: 10.1073/pnas.1315872111.

[30] Bing Rui Zhou, Jiansheng Jiang, Rodolfo Ghirlando, Davood Norouzi, K. N. Sathish Yadav, Hanqiao Feng, Rui Wang, Ping Zhang, Victor Zhurkin, and Yawen Bai. "Revisit of Reconstituted 30-nm Nucleosome Arrays Reveals an Ensemble of Dynamic Structures". In: *Journal of Molecular Biology* 430.18 (Sept. 2018), pp. 3093–3110. ISSN: 10898638. DOI: 10.1016/j.jmb.2018.06.020.

[31] Horng D Ou, Sébastien Phan, Thomas J Deerinck, Andrea Thor, Mark H Ellisman, and Clodagh C O'Shea. "ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells". en. In: *Science* 357.6349 (Jan. 2017), eaag0025. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aag0025.

[32] A W McDowall, J M Smith, and J Dubochet. "Cryo-electron microscopy of vitrified chromosomes in situ." In: *The EMBO Journal* 5.6 (Jan. 1986), pp. 1395–1402. ISSN: 0261-4189.

[33] Kazuhiro Maeshima, Saera Hihara, and Mikhail Eltsov. "Chromatin structure: does the 30-nm fibre exist in vivo?" In: *Curr. Opin. Cell Biol.* 22.3 (2010), pp. 291–7. ISSN: 1879-0410. DOI: 10.1016/j.ceb.2010.03.001.

[34] Kazuhiro Maeshima, Satoru Ide, Kayo Hibino, and Masaki Sasai. "Liquid-like behavior of chromatin". In: *Current Opinion in Genetics and Development* 37 (2016), pp. 36–45. ISSN: 18790380. DOI: 10.1016/j.gde.2015.11.006.

[35] Qing Zhang, Daniel A Beard, and Tamar Schlick. "Constructing irregular surfaces to enclose macromolecular complexes for mesoscale modeling using the discrete surface charge optimization (DISCO) algorithm". de. In: *Journal of Computational Chemistry* 24.16 (June 2003), pp. 2063–2074. ISSN: 1096-987X. DOI: 10.1002/jcc.10337.

[36] Gaurav Arya, Qing Zhang, and Tamar Schlick. "Flexible Histone Tails in a New Mesoscopic Oligonucleosome Model". In: *Biophys. J.* 91.1 (2006), pp. 133–150. ISSN: 00063495. DOI: 10.1529/biophysj.106.083006.

[37] Ognjen Perišić, Rosana Collepardo-Guevara, and Tamar Schlick. "Modeling studies of chromatin fiber structure as a function of DNA linker length". In: *J. Mol. Biol.* 403.5 (2010), pp. 777–802. ISSN: 00222836. DOI: 10.1016/j.jmb.2010.07.057. arXiv: NIHMS150003.

[38] Nick Kepper, Dietrich Foethke, Rene Stehr, Gero Wedemann, and Karsten Rippe. "Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation." In: *Biophys. J.* 95.8 (Oct. 2008), pp. 3692–705. ISSN: 1542-0086. DOI: 10.1529/biophysj.107.121079.

[39] Tatiana Nikitina, Davood Norouzi, Sergei A. Grigoryev, and Victor B. Zhurkin. "DNA topology in chromatin is defined by nucleosome spacing". In: *Science Advances* 3.10 (Oct. 2017), e1700957. ISSN: 23752548. DOI: 10.1126/sciadv.1700957.

[40] Gaurav Bajpai, Ishutesh Jain, Mandar M. Inamdar, Dibyendu Das, and Ranjith Padinhateeri. "Binding of DNA-bending non-histone proteins destabilizes regular 30-nm chromatin structure". In: *PLoS Computational Biology* 13.1 (Jan. 2017), e1005365. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1005365.

[41] Rosana Collepardo-Guevara, Guillem Portella, Michele Vendruscolo, Daan Frenkel, Tamar Schlick, and Modesto Orozco. "Chromatin Unfolding by Epigenetic Modifications Explained by Dramatic Impairment of Internucleosome Interactions: A Multiscale Computational Study". In: *Journal of the American Chemical Society* 137 (32 Aug. 2015), pp. 10205–10215. ISSN: 0002-7863. DOI: 10.1021/jacs.5b04086.

[42] Viviana I. Risca, Sarah K. Denny, Aaron F. Straight, and William J. Greenleaf. "Variable chromatin structure revealed by in situ spatially correlated DNA cleavage

mapping". In: *Nature* 541.7636 (2017), pp. 237–241. ISSN: 0028-0836. DOI: 10.1038/nature20781.

[43] Longqi Liu, Lizhi Leng, Chuanyu Liu, Changfu Lu, Yue Yuan, Liang Wu, Fei Gong, Shuoping Zhang, Xiaoyu Wei, Mingyue Wang, Lei Zhao, Liang Hu, Jian Wang, Huanming Yang, Shida Zhu, Fang Chen, Guangxiu Lu, Zhouchun Shang, and Ge Lin. "An integrated chromatin accessibility and transcriptome landscape of human preimplantation embryos". In: *Nature Communications* 10 (1 Dec. 2019), pp. 1–11. ISSN: 20411723. DOI: 10.1038/s41467-018-08244-0.

[44] Zhong Xu, Junjing Wu, Jiawei Zhou, Yu Zhang, Mu Qiao, Hua Sun, Zipeng Li, Lianghua Li, Nanqi Chen, Favour Oluwapelumi Oyelami, Xianwen Peng, and Shuqi Mei. "Integration of ATAC-seq and RNA-seq analysis identifies key genes affecting intramuscular fat content in pigs". In: *Frontiers in Nutrition* 9 (Oct. 2022). ISSN: 2296861X. DOI: 10.3389/fnut.2022.1016956.

[45] Karun Kiani, Eric M Sanford, Yogesh Goyal, and Arjun Raj. "Changes in chromatin accessibility are not concordant with transcriptional changes for single-factor perturbations". In: *Molecular Systems Biology* 18 (9 Sept. 2022). ISSN: 1744-4292. DOI: 10.15252/msb.202210979.

[46] Yue Li, Adam Eshein, Ranya K.A. Virk, Aya Eid, Wenli Wu, Jane Frederick, David VanDerway, Scott Gladstein, Kai Huang, Anne R. Shim, Nicholas M. Anthony, Greta M. Bauer, Xiang Zhou, Vasundhara Agrawal, Emily M. Pujadas, Surbhi Jain, George Esteve, John E. Chandler, The-Quyen Nguyen, Reiner Bleher, Juan J. de Pablo, Igal Szleifer, Vinayak P. Dravid, Luay M. Almassalha, and Vadim Backman. "Nanoscale chromatin imaging and analysis platform bridges 4D chromatin organization with molecular function". In: *Science Advances* 7 (1 Jan. 2021). ISSN: 2375-2548. DOI: 10.1126/sciadv.abe4310.

[47] Masae Ohno, Tadashi Ando, David G. Priest, Vipin Kumar, Yamato Yoshida, and Yuichi Taniguchi. "Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs". In: *Cell* 176 (3 Jan. 2019), 520–534.e25. ISSN: 10974172. DOI: 10.1016/j.cell.2018.12.014.

[48] René Stehr, Nick Kepper, Karsten Rippe, and Gero Wedemann. "The effect of internucleosomal interaction on folding of the chromatin fiber." In: *Biophys. J.* 95.8 (Oct. 2008), pp. 3677–91. ISSN: 1542-0086. DOI: 10.1529/biophysj.107.120543.

[49] Rosana Collepardo-Guevara and Tamar Schlick. "The effect of linker histones nucleosome binding affinity on chromatin unfolding mechanisms". In: *Biophys. J.* 101.7 (2011), pp. 1670–1680. ISSN: 00063495. DOI: 10.1016/j.bpj.2011.07.044.

[50] Antoni Luque, Gungor Ozer, and Tamar Schlick. "Correlation among DNA Linker Length, Linker Histone Concentration, and Histone Tails in Chromatin". In: *Biophysical Journal* 110.11 (11 2016), pp. 2309–2319. ISSN: 15420086. DOI: 10.1016/j.bpj.2016.04.024.

[51] Gavin D. Bascom, Taejin Kim, and Tamar Schlick. "Kilobase Pair Chromatin Fiber Contacts Promoted by Living-System-Like DNA Linker Length Distributions and Nucleosome Depletion". In: *Journal of Physical Chemistry B* 121.15 (Apr. 2017), pp. 3882–3894. ISSN: 15205207. DOI: 10.1021/acs.jpcb.7b00998.

[52] Joshua Lequieu, Andrés Córdoba, Joshua Moller, and Juan J. de Pablo. "1CPN: A coarse-grained multi-scale model of chromatin". In: *The Journal of Chemical Physics* 150.21 (June 2019), p. 215102. ISSN: 0021-9606. DOI: 10.1063/1.5092976.

[53] Joshua Moller, Joshua Lequieu, and Juan J. De Pablo. "The Free Energy Landscape of Internucleosome Interactions and Its Relation to Chromatin Fiber Structure". In: *ACS Central Science* 5.2 (Feb. 2019), pp. 341–348. ISSN: 23747951. DOI: 10.1021/acscentsci.8b00836.

[54] Luca Maragliano, Alexander Fischer, Eric Vanden-Eijnden, and Giovanni Ciccotti. "String method in collective variables: Minimum free energy paths and isocommittor surfaces". In: *The Journal of Chemical Physics* 125.2 (July 2006), p. 024106. ISSN: 0021-9606. DOI: 10.1063/1.2212942.

[55] Yuji Sugita and Yuko Okamoto. "Replica-exchange molecular dynamics method for protein folding". In: *Chemical Physics Letters* 314.1-2 (Nov. 1999), pp. 141–151. ISSN: 00092614. DOI: 10.1016/S0009-2614(99)01123-9.

[56] Weinan E And, Weiqing Ren, and Eric Vanden-Eijnden*. "Finite Temperature String Method for the Study of Rare Events†". In: (2005). DOI: 10.1021/JP0455430.

[57] Eric Vanden-Eijnden and Maddalena Venturoli. "Revisiting the finite temperature string method for the calculation of reaction tubes and free energies". In: *The Journal of Chemical Physics* 130.19 (May 2009), p. 194103. ISSN: 0021-9606. DOI: 10.1063/1.3130083.

[58] Hythem Sidky, Yamil J. Colón, Julian Helfferich, Benjamin J. Sikora, Cody Bezik, Weiwei Chu, Federico Giberti, Ashley Z. Guo, Xikai Jiang, Joshua Lequieu, Jiyuan Li, Joshua Moller, Michael J. Quevillon, Mohammad Rahimi, Hadi Ramezani-Dakhel, Vikramjit S. Rathee, Daniel R. Reid, Emre Sevgen, Vikram Thapar, Michael A. Webb, Jonathan K. Whitmer, and Juan J. De Pablo. "SSAGES: Software Suite for Advanced General Ensemble Simulations". In: *Journal of Chemical Physics* 148.4 (Jan. 2018), p. 044104. ISSN: 00219606. DOI: 10.1063/1.5008853.

[59] D. E. Barton, M. Abramovitz, and I. A. Stegun. "Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables." In: *Journal of the Royal Statistical Society. Series A (General)* (1965). ISSN: 00359238. DOI: 10.2307/2343473.

[60] Antoni Luque, Rosana Collepardo-Guevara, Sergei Grigoryev, and Tamar Schlick. "Dynamic condensation of linker histone C-terminal domain regulates chromatin structure". en. In: *Nucleic Acids Res.* 42.12 (June 2014), pp. 7553–7560. ISSN: 13624962. DOI: 10.1093/nar/gku491.

[61] Tamar Schlick and Ognjen Perišić. "Mesoscale simulations of two nucleosome-repeat length oligonucleosomes". In: *Physical Chemistry Chemical Physics* 11.45 (2009), pp. 10729–10737. ISSN: 14639076. DOI: 10.1039/b918629h.

[62] Philip J.J. Robinson, Woojin An, Andrew Routh, Fabrizio Martino, Lynda Chapman, Robert G. Roeder, and Daniela Rhodes. "30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction". In: *Journal of Molecular Biology* 381.4 (4 Sept. 2008), pp. 816–825. ISSN: 00222836. DOI: 10.1016/j.jmb.2008.04.050.

[63] Anton Valouev, Steven M. Johnson, Scott D. Boyd, Cheryl L. Smith, Andrew Z. Fire, and Arend Sidow. "Determinants of nucleosome organization in primary human cells". In: *Nature* 474.7352 (2011), pp. 516–522. ISSN: 00280836. DOI: 10.1038/nature10002.

[64] Peter J. Mulligan, Elena F. Koslover, and Andrew J. Spakowitz. "Thermodynamic model of heterochromatin formation through epigenetic regulation". In: *Journal of Physics Condensed Matter* 27.6 (2015). ISSN: 1361648X. DOI: 10.1088/0953-8984/27/6/064109.

[65] Daniele Canzio, Evelyn Y. Chang, Smita Shankar, Kristopher M. Kuchenbecker, Matthew D. Simon, Hiten D. Madhani, Geeta J. Narlikar, and Bassem Al-Sady. "Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly". In: *Molecular Cell* 41.1 (2011), pp. 67–81. ISSN: 10972765. DOI: 10.1016/j.molcel.2010.12.016.

[66] Y. Cui and C. Bustamante. "Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure". In: *Proceedings of the National Academy of Sciences* 97.1 (2000), pp. 127–132. ISSN: 0027-8424. DOI: 10.1073/pnas.97.1.127. arXiv: NIHMS150003.

[67] Maarten Kruithof, Fan-Tso Chien, Andrew Routh, Colin Logie, Daniela Rhodes, and John van Noort. "Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber." In: *Nature structural & molecular biology* 16.5 (2009), pp. 534–40. ISSN: 1545-9985. DOI: 10.1038/nsmb.1590.

[68] He Meng, Kurt Andresen, and John van Noort. "Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibers". In: *Nucleic Acids Research* 43.7 (Mar. 2015), pp. 3578–3590. ISSN: 0305-1048. DOI: 10.1093/nar/gkv215.

[69] Sandra Nitsch, Aria Coraor, Natalie Möritz, Yiheng Wu, Jonas Funke, Riccardo Alessandri, Harsh Nagpal, Hendrik Dietz, Beat Fierz, Robert Schneider, and Juan de Pablo. "Structural effects of H4K16 Acylations". In: *In preparation* (2024).

[70] Suhas S P Rao, Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, Ido Machol, Arina D. Omer, Eric S. Lander, and Erez Lieberman Aiden. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping". In: *Cell* 159.7 (7 Dec.

2014), pp. 1665–1680. ISSN: 10974172. DOI: 10.1016/j.cell.2014.11.021. arXiv: 1206.5533.

[71] Quinn MacPherson, Bruno Beltran, and Andrew J. Spakowitz. "Bottom–up modeling of chromatin segregation due to epigenetic modifications". en. In: *Proceedings of the National Academy of Sciences of the United States of America* 115.50 (50 Jan. 2018), pp. 12739–12744. ISSN: 10916490. DOI: 10.1073/pnas.1812268115.

[72] Yichao Cai, Ying Zhang, Yan Ping Loh, Jia Qi Tng, Mei Chee Lim, Zhendong Cao, Anandhkumar Raju, Erez Lieberman Aiden, Shang Li, Lakshmanan Manikandan, Vinay Tergaonkar, Greg Tucker-Kellogg, and Melissa Jane Fullwood. "H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions". In: *Nature Communications* 12 (1 Dec. 2021). ISSN: 20411723. DOI: 10.1038/s41467-021-20940-y.

[73] Menno P. Creyghton, Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, Michael A. Lodato, Garrett M. Frampton, Phillip A. Sharp, Laurie A. Boyer, Richard A. Young, and Rudolf Jaenisch. "Histone H3K27ac separates active from poised enhancers and predicts developmental state". In: *Proceedings of the National Academy of Sciences of the United States of America* 107 (50 Dec. 2010), pp. 21931–21936. ISSN: 00278424. DOI: 10.1073/pnas.1016071107.

[74] Brian D. Strahl and C. David Allis. "The language of covalent histone modifications". In: *Nature* 403 (6765 Jan. 2000), pp. 41–45. ISSN: 0028-0836. DOI: 10.1038/47412.

[75] Thomas Jenuwein and C. David Allis. "Translating the Histone Code". In: *Science* 293 (5532 Aug. 2001), pp. 1074–1080. ISSN: 0036-8075. DOI: 10.1126/science.1063127.

[76] Marc A.J. Morgan and Ali Shilatifard. "Epigenetic moonlighting: Catalytic-independent functions of histone modifiers in regulating transcription". In: *Science Advances* 9 (16 Apr. 2023). ISSN: 2375-2548. DOI: 10.1126/sciadv.adg6593.

[77] Timothy R. Blosser, Janet G. Yang, Michael D. Stone, Geeta J. Narlikar, and Xiaowei Zhuang. "Dynamics of nucleosome remodelling by individual ACF complexes". In: *Nature* 462 (7276 Dec. 2009), pp. 1022–1027. ISSN: 00280836. DOI: 10.1038/nature08627.

[78] Michael Shogren-Knaak, Haruhiko Ishii, Jian-Min Sun, Michael J. Pazin, James R. Davie, and Craig L. Peterson. "Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions". In: *Science* 311 (5762 Feb. 2006), pp. 844–847. ISSN: 0036-8075. DOI: 10.1126/science.1124000.

[79] Anna A. Kalashnikova, Mary E. Porter-Goff, Uma M. Muthurajan, Karolin Luger, and Jeffrey C. Hansen. "The role of the nucleosome acidic patch in modulating higher order chromatin structure". In: *Journal of The Royal Society Interface* 10 (82 May 2013), p. 20121022. ISSN: 1742-5689. DOI: 10.1098/rsif.2012.1022.

[80] Gary LeRoy, Peter A DiMaggio, Eric Y. Chan, Barry M. Zee, M. Andres Blanco, Barbara Bryant, Ian Z. Flaniken, Sherry Liu, Yibin Kang, Patrick Trojer, and Benjamin A. Garcia. "A quantitative atlas of histone modification signatures from human cancer cells". In: *Epigenetics & Chromatin* 6.1 (1 Dec. 2013), p. 20. ISSN: 1756-8935. DOI: 10.1186/1756-8935-6-20.

[81] Gillian C.A. Taylor, Ragnhild Eskeland, Betül Hekimoglu-Balkan, Madapura M. Pradeepa, and Wendy A. Bickmore. "H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction". In: *Genome Research* 23 (12 Dec. 2013), pp. 2053–2065. ISSN: 10889051. DOI: 10.1101/gr.155028.113.

[82] Ying Liu, Chenning Lu, Ye Yang, Yanping Fan, Renliang Yang, Chuan Fa Liu, Nikolay Korolev, and Lars Nordenskiöld. "Influence of histone tails and H4 tail acetylations on nucleosome- nucleosome interactions". In: *Journal of Molecular Biology* 414 (5 Dec. 2011), pp. 749–764. ISSN: 10898638. DOI: 10.1016/j.jmb.2011.10.031.

[83] Aliaksandra Radzisheuskaya, Pavel V. Shliaha, Vasily V. Grinev, Daria Shlyueva, Helene Damhofer, Richard Koche, Vladimir Gorshkov, Sergey Kovalchuk, Yingqian Zhan, Keli L. Rodriguez, Andrea L. Johnstone, Michael C. Keogh, Ronald C. Hendrickson, Ole N. Jensen, and Kristian Helin. "Complex-dependent histone acetyltransferase activity of KAT8 determines its role in transcription and cellular homeostasis". In: *Molecular Cell* 81 (8 Apr. 2021), 1749–1765.e8. ISSN: 10974164. DOI: 10.1016/j.molcel.2021.02.012.

[84] Yue Chen, Robert Sprung, Yi Tang, Haydn Ball, Bhavani Sangras, Sung Chan Kim, John R. Falck, Junmin Peng, Wei Gu, and Yingming Zhao. "Lysine propionylation and butyrylation are novel post-translational modifications in histones". In: *Molecular and Cellular Proteomics* 6 (5 May 2007), pp. 812–819. ISSN: 15359476. DOI: 10.1074/mcp.M700021-MCP200.

[85] Johayra Simithy, Simone Sidoli, Zuo-Fei Yuan, Mariel Coradin, Natarajan V. Bhanu, Dylan M. Marchione, Brianna J. Klein, Gleb A. Bazilevsky, Cheryl E. McCullough, Robert S. Magin, Tatiana G. Kutateladze, Nathaniel W. Snyder, Ronen Marmorstein, and Benjamin A. Garcia. "Characterization of histone acylations links chromatin modifications with metabolism". In: *Nature Communications* 8 (1 Oct. 2017), p. 1141. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01384-9.

[86] Adam F. Kebede, Anna Nieborak, Lara Zorro Shahidian, Stephanie Le Gras, Florian Richter, Diana Aguilar Gómez, Marijke P. Baltissen, Gergo Meszaros, Helena De Fatima Magliarelli, Aaron Taudt, Raphael Margueron, Maria Colomé-Tatché, Romeo Ricci, Sylvain Daujat, Michiel Vermeulen, Gerhard Mittler, and Robert Schneider. "Histone propionylation is a mark of active chromatin". In: *Nature Structural and Molecular Biology* 24 (12 2017), pp. 1048–1056. ISSN: 15459985. DOI: 10.1038/nsmb.3490.

[87] Zhen Han, Hong Wu, Sunjoo Kim, Xiangkun Yang, Qianjin Li, He Huang, Houjian Cai, Michael G. Bartlett, Aiping Dong, Hong Zeng, Peter J. Brown, Xiang Jiao Yang,

Cheryl H. Arrowsmith, Yingming Zhao, and Y. George Zheng. "Revealing the protein propionylation activity of the histone acetyltransferase MOF (males absent on the first)". In: *Journal of Biological Chemistry* 293 (9 Mar. 2018), pp. 3410–3420. ISSN: 1083351X. DOI: 10.1074/jbc.RA117.000529.

[88] Lin Li, Mohammad Ghorbani, Monika Weisz-Hubshman, Justine Rousseau, Isabelle Thiffault, Rhonda E. Schnur, Catherine Breen, Renske Oegema, Marjan M.M. Weiss, Quinten Waisfisz, Sara Welner, Helen Kingston, Jordan A. Hills, Elles M.J. Boon, Lina Basel-Salmon, Osnat Konen, Hadassa Goldberg-Stern, Lily Bazak, Shay Tzur, Jianliang Jin, Xiuli Bi, Michael Bruccoleri, Kirsty McWalter, Megan T. Cho, Maria Scarano, G. Bradley Schaefer, Susan S. Brooks, Susan Starling Hughes, K. L.I. Van Gassen, Johanna M. Van Hagen, Tej K. Pandita, Pankaj B. Agrawal, Philippe M. Campeau, and Xiang Jiao Yang. "Lysine acetyltransferase 8 is involved in cerebral development and syndromic intellectual disability". In: *Journal of Clinical Investigation* 130 (3 Mar. 2020), pp. 1431–1445. ISSN: 15588238. DOI: 10.1172/JCI131145.

[89] Afsaneh Goudarzi, Di Zhang, He Huang, Sophie Barral, Oh Kwang Kwon, Shankang Qi, Zhanyun Tang, Thierry Buchou, Anne Laure Vitte, Tieming He, Zhongyi Cheng, Emilie Montellier, Jonathan Gaucher, Sandrine Curtet, Alexandra Debernardi, Guillaume Charbonnier, Denis Puthier, Carlo Petosa, Daniel Panne, Sophie Rousseaux, Robert G. Roeder, Yingming Zhao, and Saadi Khochbin. "Dynamic Competing Histone H4 K5K8 Acetylation and Butyrylation Are Hallmarks of Highly Active Gene Promoters". In: *Molecular Cell* 62 (2 2016), pp. 169–180. ISSN: 10974164. DOI: 10.1016/j.molcel.2016.03.014.

[90] Shuai Liu, Guanqing Liu, Peifeng Cheng, Chao Xue, Yong Zhou, Xu Chen, Lu Ye, Zhongying Qiao, Tao Zhang, and Zhiyun Gong. "Genome-wide Profiling of Histone Lysine Butyrylation Reveals its Role in the Positive Regulation of Gene Transcription in Rice". In: *Rice* 12 (1 Dec. 2019), p. 86. ISSN: 1939-8425. DOI: 10.1186/s12284-019-0342-6.

[91] Xun Huang, Juan Yan, Min Zhang, Yafang Wang, Yi Chen, Xuhong Fu, Rongrui Wei, Xing ling Zheng, Zhiwei Liu, Xiong Zhang, Hong Yang, Bingbing Hao, Yan yan Shen, Yi Su, Xiaoji Cong, Min Huang, Minjia Tan, Jian Ding, and Meiyu Geng. "Targeting Epigenetic Crosstalk as a Therapeutic Strategy for EZH2-Aberrant Solid Tumors". In: *Cell* 175 (1 Sept. 2018), 186–199.e19. ISSN: 00928674. DOI: 10.1016/j.cell.2018.08.058.

[92] Qinming Chen, Renliang Yang, Nikolay Korolev, Chuan Fa Liu, and Lars Nordenskiöld. "Regulation of Nucleosome Stacking and Chromatin Compaction by the Histone H4 N-Terminal Tail–H2A Acidic Patch Interaction". In: *Journal of Molecular Biology* 429 (13 June 2017), pp. 2075–2092. ISSN: 10898638. DOI: 10.1016/j.jmb.2017.03.016.

[93] Ruihan Zhang, Jochen Erler, and Jörg Langowski. "Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction". In: *Bio-*

physical Journal 112 (3 2017), pp. 450–459. ISSN: 15420086. DOI: 10.1016/j.bpj.2016.11.015.

[94] Sinan Kilic, Suren Felekyan, Olga Doroshenko, Iuliia Boichenko, Mykola Dimura, Hayk Vardanyan, Louise C. Bryan, Gaurav Arya, Claus A.M. Seidel, and Beat Fierz. "Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α". In: *Nature Communications* 9.1 (2018). ISSN: 20411723. DOI: 10.1038/s41467-017-02619-5.

[95] Aidan P. Thompson, H. Metin Aktulga, Richard Berger, Dan S. Bolintineanu, W. Michael Brown, Paul S. Crozier, Pieter J. in 't Veld, Axel Kohlmeyer, Stan G. Moore, Trung Dac Nguyen, Ray Shan, Mark J. Stevens, Julien Tranchida, Christian Trott, and Steven J. Plimpton. "LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales". In: *Computer Physics Communications* 271 (Feb. 2022). ISSN: 00104655. DOI: 10.1016/j.cpc.2021.108171.

[96] Walter Alvarado, Joshua Moller, Andrew L. Ferguson, and Juan J. De Pablo. "Tetranucleosome Interactions Drive Chromatin Folding". In: *ACS Central Science* (2021). ISSN: 23747951. DOI: 10.1021/acscentsci.1c00085.

[97] Habtamu Zewdie. "Computer simulation studies of liquid crystals: A new Corner potential for cylindrically symmetric particles". In: *Journal of Chemical Physics* 108.5 (1998), pp. 2117–2133. ISSN: 00219606. DOI: 10.1063/1.475591.

[98] Daniel M Hinckley, Gordon S Freeman, Jonathan K Whitmer, and Juan J De Pablo. "An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization". In: *The Journal of chemical physics* 139.14 (2013), 10B604_1.

[99] Wenfei Li, Peter G Wolynes, and Shoji Takada. "Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins". In: *Proceedings of the National Academy of Sciences* 108.9 (2011), pp. 3504–3509.

[100] Michael R. Shirts and John D. Chodera. "Statistically optimal analysis of samples from multiple equilibrium states". In: *Journal of Chemical Physics* 129 (12 2008). ISSN: 00219606. DOI: 10.1063/1.2978177.

[101] John G. Kirkwood. "The general theory of irreversible processes in solutions of macromolecules". In: *Journal of Polymer Science Part B: Polymer Physics* 34 (4 Mar. 1996), pp. 597–610. ISSN: 0887-6266. DOI: 10.1002/polb.1996.897.

[102] G. E. Croston, P. J. Laybourn, S. M. Paranjape, and J. T. Kadonaga. "Mechanism of transcriptional antirepression by GAL4-VP16". In: *Genes and Development* 6 (12 A 1992), pp. 2270–2281. ISSN: 08909369. DOI: 10.1101/gad.6.12a.2270.

[103] Angad Garg, Yehuda Goldgur, Beate Schwer, and Stewart Shuman. "Distinctive structural basis for DNA recognition by the fission yeast Zn2Cys6 transcription factor Pho7

and its role in phosphate homeostasis". In: *Nucleic Acids Research* 46 (21 2018), pp. 11262–11273. ISSN: 13624962. DOI: 10.1093/nar/gky827.

[104] L. M. Almassalha, A. Tiwari, P. T. Ruhoff, Y. Stypula-Cyrus, L. Cherkezyan, H. Matsuda, M. A. Dela Cruz, J. E. Chandler, C. White, C. Maneval, H. Subramanian, I. Szleifer, H. K. Roy, and V. Backman. "The global relationship between chromatin physical topology, fractal structure, and gene expression". en. In: *Scientific Reports* 7 (Jan. 2017), p. 41061. ISSN: 20452322. DOI: 10.1038/srep41061.

[105] Luay M. Almassalha, Greta M. Bauer, John E. Chandler, Scott Gladstein, Lusik Cherkezyan, Yolanda Stypula-Cyrus, Samuel Weinberg, Di Zhang, Peder Thusgaard Ruhoff, Hemant K. Roy, Hariharan Subramanian, Navdeep S. Chandel, Igal Szleifer, and Vadim Backman. "Label-free imaging of the native, living cellular nanoarchitecture using partial-wave spectroscopic microscopy". en. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.42 (Jan. 2016), E6372–E6381. ISSN: 10916490. DOI: 10.1073/pnas.1608198113.

[106] Ranya K.A. Virk, Wenli Wu, Luay M. Almassalha, Greta M. Bauer, Yue Li, David VanDerway, Jane Frederick, Di Zhang, Adam Eshein, Hemant K. Roy, Igal Szleifer, and Vadim Backman. "Disordered chromatin packing regulates phenotypic plasticity". In: *Science Advances* 6.2 (2020). ISSN: 23752548. DOI: 10.1126/sciadv.aax6232.

[107] Rebekah R. Starks, Anilisa Biswas, Ashish Jain, and Geetu Tuteja. "Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks". In: *Epigenetics and Chromatin* 12.1 (Feb. 2019), p. 16. ISSN: 17568935. DOI: 10.1186/s13072-019-0260-2.

[108] Erez Lieberman-aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, and Leonid A Mirny. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome". In: *Science* 326.9 October 2009 (5950 2009), pp. 289–293. ISSN: 1095-9203. DOI: 10.1038/nature08398. arXiv: arXiv:1011.1669v3.

[109] Hannah L. Harris, Huiya Gu, Moshe Olshansky, Ailun Wang, Irene Farabella, Yossi Eliaz, Achyuth Kalluchi, Akshay Krishna, Mozes Jacobs, Gesine Cauer, Melanie Pham, Suhas S.P. Rao, Olga Dudchenko, Arina Omer, Kiana Mohajeri, Sungjae Kim, Michael H. Nichols, Eric S. Davis, Dimos Gkountaroulis, Devika Udupa, Aviva Presser Aiden, Victor G. Corces, Douglas H. Phanstiel, William Stafford Noble, Guy Nir, Michele Di Pierro, Jeong Sun Seo, Michael E. Talkowski, Erez Lieberman Aiden, and M. Jordan Rowley. "Chromatin alternates between A and B compartments at kilobase scale for subgenic organization". In: *Nature communications* 14 (1 Dec. 2023), p. 3303. ISSN: 20411723. DOI: 10.1038/s41467-023-38429-1.

[110] Suhas S.P. Rao, Su Chen Huang, Brian Glenn St Hilaire, Jesse M. Engreitz, Elizabeth M. Perez, Kyong Rim Kieffer-Kwon, Adrian L. Sanborn, Sarah E. Johnstone,

Gavin D. Bascom, Ivan D. Bochkov, Xingfan Huang, Muhammad S. Shamim, Jaeweon Shin, Douglass Turner, Ziyi Ye, Arina D. Omer, James T. Robinson, Tamar Schlick, Bradley E. Bernstein, Rafael Casellas, Eric S. Lander, and Erez Lieberman Aiden. "Cohesin Loss Eliminates All Loop Domains". In: *Cell* 171 (2 2017), 305–320.e24. ISSN: 10974172. DOI: 10.1016/j.cell.2017.09.026.

[111] Wibke Schwarzer, Nezar Abdennur, Anton Goloborodko, Aleksandra Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A. Fonseca, Wolfgang Huber, Christian H. Haering, Leonid Mirny, and Francois Spitz. "Two independent modes of chromatin organization revealed by cohesin removal". In: *Nature* 551 (7678 Nov. 2017), pp. 51–56. ISSN: 14764687. DOI: 10.1038/nature24281.

[112] Adrian L. Sanborn, Suhas S.P. Rao, Su Chen Huang, Neva C. Durand, Miriam H. Huntley, Andrew I. Jewett, Ivan D. Bochkov, Dharmaraj Chinnappan, Ashok Cutkosky, Jian Li, Kristopher P. Geeting, Andreas Gnirke, Alexandre Melnikov, Doug McKenna, Elena K. Stamenova, Eric S. Lander, and Erez Lieberman Aiden. "Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes". In: *Proceedings of the National Academy of Sciences of the United States of America* 112 (47 Jan. 2015), E6456–E6465. ISSN: 10916490. DOI: 10.1073/pnas.1518552112.

[113] Takashi Nagano, Yaniv Lubling, Tim J. Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D. Laue, Amos Tanay, and Peter Fraser. "Single-cell Hi-C reveals cell-to-cell variability in chromosome structure". In: *Nature* 502.7469 (7469 2013), pp. 59–64. ISSN: 00280836. DOI: 10.1038/nature12593.

[114] Takashi Nagano, Yaniv Lubling, Eitan Yaffe, Steven W. Wingett, Wendy Dean, Amos Tanay, and Peter Fraser. "Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell". In: *Nature Protocols* 10 (12 Dec. 2015), pp. 1986–2003. ISSN: 17502799. DOI: 10.1038/nprot.2015.127.

[115] Takashi Nagano, Yaniv Lubling, Csilla Várnai, Carmel Dudley, Wing Leung, Yael Baran, Netta Mendelson Cohen, Steven Wingett, Peter Fraser, and Amos Tanay. "Cell-cycle dynamics of chromosomal organization at single-cell resolution". In: *Nature* 547 (7661 July 2017), pp. 61–67. ISSN: 14764687. DOI: 10.1038/nature23001.

[116] Oluwatosin Oluwadare, Max Highsmith, and Jianlin Cheng. *An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data*. Apr. 2019. DOI: 10.1186/s12575-019-0094-0.

[117] Aleksandra A. Galitsyna and Mikhail S. Gelfand. "Single-cell Hi-C data analysis: safety in numbers". In: *Briefings in Bioinformatics* 22 (6 Nov. 2021). ISSN: 14774054. DOI: 10.1093/bib/bbab316.

[118] M. Gridina, A. Taskina, T. Lagunov, A. Nurislamov, T. Kulikova, A. Krasikova, and V. Fishman. "Comparison and critical assessment of single-cell Hi-C protocols". In: *Heliyon* 8 (10 Oct. 2022). ISSN: 24058440. DOI: 10.1016/j.heliyon.2022.e11023.

[119] Jie Liu, Dejun Lin, Galip Gürkan Yardlmcl, and William Stafford Noble. "Unsupervised embedding of single-cell Hi-C data". In: vol. 34. Oxford University Press, July 2018, pp. i96–i104. DOI: 10.1093/bioinformatics/bty285.

[120] Jingtian Zhou, Jianzhu Ma, Yusi Chen, Chuankai Cheng, Bokan Bao, Jian Peng, Terrence J. Sejnowski, Jesse R. Dixon, and Joseph R. Ecker. "Robust single-cell Hi-C clustering by convolution- And random-walk–based imputation". In: *Proceedings of the National Academy of Sciences of the United States of America* 116 (28 2019), pp. 14011–14018. ISSN: 10916490. DOI: 10.1073/pnas.1901423116.

[121] Hyeon Jin Kim, Galip Gürkan Yardımcı, Giancarlo Bonora, Vijay Ramani, Jie Liu, Ruolan Qiu, Choli Lee, Jennifer Hesson, Carol B. Ware, Jay Shendure, Zhijun Duan, and William Stafford Noble. "Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data". In: *PLoS Computational Biology* 16 (9 Sept. 2020). ISSN: 15537358. DOI: 10.1371/journal.pcbi.1008173.

[122] Ruochi Zhang, Tianming Zhou, and Jian Ma. "Multiscale and integrative single-cell Hi-C analysis with Higashi". In: *Nature Biotechnology* 40.2 (2 2022), pp. 254–261. ISSN: 15461696. DOI: 10.1038/s41587-021-01034-y.

[123] Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. "Imaging intracellular fluorescent proteins at nanometer resolution". In: *Science* 313.5793 (2006), pp. 1642–1645. ISSN: 00368075. DOI: 10.1126/science.1127344.

[124] W. E. Moerner and L. Kador. "Optical detection and spectroscopy of single molecules in a solid". In: *Physical Review Letters* 62.21 (1989), pp. 2535–2538. ISSN: 00319007. DOI: 10.1103/PhysRevLett.62.2535.

[125] Stefan W. Hell and Jan Wichmann. "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy". In: *Optics Letters* 19.11 (1994), p. 780. ISSN: 0146-9592. DOI: 10.1364/ol.19.000780.

[126] S. Wang, J.-H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C.-t. Wu, and X. Zhuang. "Spatial organization of chromatin domains and compartments in single chromosomes". In: *Science* 353.6299 (Aug. 2016), pp. 598–602. ISSN: 0036-8075. DOI: 10.1126/science.aaf8084.

[127] Bogdan Bintu, Leslie J. Mateo, Jun Han Su, Nicholas A. Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N. Boettiger, and Xiaowei Zhuang. "Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells". en. In: *Science* 362.6413 (Mar. 2018), eaau1783. ISSN: 10959203. DOI: 10.1126/science.aau1783.

[128] Mengwei Hu and Siyuan Wang. "Chromatin Tracing: Imaging 3D Genome and Nucleome". In: *Trends in Cell Biology* 31 (1 Jan. 2021), pp. 5–8. ISSN: 09628924. DOI: 10.1016/j.tcb.2020.10.006.

[129] Jianquan Xu, Hongqiang Ma, and Yang Liu3. "stochastic optical microscopy (STORM)". In: *Current Protocols in Cytometry* 15.5 (2018), pp. 477–491. ISSN: 1744-8360. DOI: 10.1002/cpcy.23.Stochastic.

[130] Jianquan Xu and Yang Liu. "Imaging Higher-order Chromatin Structures in Single Cells Using Stochastic Optical Reconstruction Microscopy". In: *Bio-protocol* 9 (3 Feb. 2019). ISSN: 23318325. DOI: 10.21769/BioProtoc.3160.

[131] Marcus A. Woodworth, Kenneth K.H. Ng, Aaron R. Halpern, Nicholas A. Pease, Phuc H.B. Nguyen, Hao Yuan Kueh, and Joshua C. Vaughan. "Multiplexed single-cell profiling of chromatin states at genomic loci by expansion microscopy". In: *Nucleic Acids Research* 49 (14 Aug. 2021), E82–E82. ISSN: 13624962. DOI: 10.1093/nar/gkab423.

[132] Jianquan Xu and Yang Liu. "Probing Chromatin Compaction and Its Epigenetic States in situ With Single-Molecule Localization-Based Super-Resolution Microscopy". In: *Frontiers in Cell and Developmental Biology* 9 (June 2021). ISSN: 2296-634X. DOI: 10.3389/fcell.2021.653077.

[133] Liangqi Xie and Zhe Liu. "Single-cell imaging of genome organization and dynamics". In: *Molecular Systems Biology* 17 (7 July 2021). ISSN: 1744-4292. DOI: 10.15252/msb.20209653.

[134] Alexander P. Young, Daniel J. Jackson, and Russell C. Wyeth. "A technical review and guide to RNA fluorescence in situ hybridization". In: *PeerJ* 8 (Mar. 2020), e8806. ISSN: 2167-8359. DOI: 10.7717/peerj.8806.

[135] Yuji Itoh, Esmae J. Woods, Katsuhiko Minami, Kazuhiro Maeshima, and Rosana Collepardo-Guevara. "Liquid-like chromatin in the cell: What can we learn from imaging and computational modeling?" In: *Current Opinion in Structural Biology* 71 (Dec. 2021), pp. 123–135. ISSN: 0959440X. DOI: 10.1016/j.sbi.2021.06.004.

[136] Christine van de Werken, Holger Jahr, Margarida Avo Santos, Cindy Eleveld, Joyce Schuilwerve, Joop S. E. Laven, and Esther B. Baart. "A universal method for sequential immunofluorescent analysis of chromatin and chromatin-associated proteins on chromosome spreads". In: *Chromosome Research* 21 (5 Aug. 2013), pp. 475–489. ISSN: 0967-3849. DOI: 10.1007/s10577-013-9373-9.

[137] Attila Németh. *The Nucleolus*. Ed. by Attila Németh. Vol. 1455. Springer New York, 2016. DOI: 10.1007/978-1-4939-3792-9.

[138] Michele Di Pierro, Bin Zhang, Erez Lieberman Aiden, Peter G. Wolynes, and José N. Onuchic. "Transferable model for chromosome architecture". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.43 (2016), pp. 12168–12173. ISSN: 10916490. DOI: 10.1073/pnas.1613607113.

[139] Yifeng Qi and Bin Zhang. "Predicting three-dimensional genome organization with chromatin states". In: *PLoS Computational Biology* 15.6 (June 2019). Ed. by Jian Ma, e1007024. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1007024.

[140] A Grosberg, Y Rabin, S Havlin, and A Neer. "Crumpled Globule Model of the Three-Dimensional Structure of DNA". In: *EPL (Europhysics Letters)* 23 (5 Jan. 1993), p. 373. ISSN: 0295-5075. DOI: 10.1209/0295-5075/23/5/012.

[141] Leonid A. Mirny. "The fractal globule as a model of chromatin architecture in the cell". In: *Chromosome Research* 19.1 (1 Jan. 2011), pp. 37–51. ISSN: 09673849. DOI: 10.1007/s10577-010-9177-0.

[142] Johannes Nuebler, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A Mirny. "Chromatin organization by an interplay of loop extrusion and compartmental segregation". en. In: *Proceedings of the National Academy of Sciences* 115.29 (29 Jan. 2018), E6697–E6706. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1717730115.

[143] Benjamin Brenner, Ki-Hee Song, Cheng Sun, and Hao F. Zhang. "Improving spatial precision and field-of-view in wavelength-tagged single-particle tracking using spectroscopic single-molecule localization microscopy". In: *Applied Optics* 60.13 (2021), p. 3647. ISSN: 1559-128X. DOI: 10.1364/ao.415275.

[144] Yang Zhang, Yu Zhang, Ki Hee Song, Wei Lin, Cheng Sun, George C. Schatz, and Hao F. Zhang. "Investigating Single-Molecule Fluorescence Spectral Heterogeneity of Rhodamines Using High-Throughput Single-Molecule Spectroscopy". In: *Journal of Physical Chemistry Letters* 12.16 (2021), pp. 3914–3921. ISSN: 19487185. DOI: 10.1021/acs.jpclett.1c00192.

[145] Jianquan Xu, Hongqiang Ma, Jingyi Jin, Shikhar Uttam, Rao Fu, Yi Huang, and Yang Liu. "Super-Resolution Imaging of Higher-Order Chromatin Structures at Different Epigenomic States in Single Mammalian Cells". In: *Cell Reports* 24.4 (4 July 2018), pp. 873–882. ISSN: 22111247. DOI: 10.1016/j.celrep.2018.06.085.

[146] Varun Venkataramani, Frank Herrmannsdörfer, Mike Heilemann, and Thomas Kuner. "SuReSim: Simulating localization microscopy experiments from ground truth models". In: *Nature Methods* 13.4 (2016), pp. 319–321. ISSN: 15487105. DOI: 10.1038/nmeth.3775.

[147] Robert P.J. Nieuwenhuizen, Mark Bates, Anna Szymborska, Keith A. Lidke, Bernd Rieger, and Sjoerd Stallinga. "Quantitative localization microscopy: Effects of photophysics and labeling stoichiometry". In: *PLoS ONE* 10.5 (2015), pp. 1–18. ISSN: 19326203. DOI: 10.1371/journal.pone.0127989.

[148] Yalei Chen, Nathan C. Deffenbaugh, Charles T. Anderson, and William O. Hancock. "Molecular counting by photobleaching in protein complexes with many subunits: Best practices and application to the cellulose synthesis complex". In: *Molecular Biology of*

the Cell 25 (22 Nov. 2014), pp. 3630–3642. ISSN: 19394586. DOI: 10.1091/mbc.E14-06-1146.

[149] David L. Kolin, Santiago Costantino, and Paul W. Wiseman. "Sampling effects, noise, and photobleaching in temporal image correlation spectroscopy". In: *Biophysical Journal* 90 (2 2006), pp. 628–639. ISSN: 00063495. DOI: 10.1529/biophysj.105.072322.

[150] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).

[151] S.O. (Bell Labs) Rice. "Mathematical Analysis of Random Noise Part III: statistical properties of random noise currents". In: 23.1 (1945), pp. 46–156.

[152] Darin Q. Pike, Fraņois A. Detcheverry, Marcus Müller, and Juan J. De Pablo. "Theoretically informed coarse grain simulations of polymeric systems". In: *Journal of Chemical Physics* 131.8 (2009). ISSN: 00219606. DOI: 10.1063/1.3187936.

[153] Paul J. Flory. "Thermodynamics of high polymer solutions". In: *Journal of Chemical Physicsi* 10.51 (1942), pp. 51–61. ISSN: 0973712X. DOI: 10.1063/1.172.

[154] Maurice L. Huggins. "Solutions of long chain compounds". In: 9.5 (1941), p. 440. ISSN: 00219606. DOI: 10.1063/1.1750930.

[155] Hiromi Yamakawa. *Modern Theory of Polymer Solutions*. Harper & Row, Publishers, Inc., 1971.

[156] Elena F Koslover and Andrew J Spakowitz. "Discretizing elastic chains for coarse-grained polymer models Soft Matter". In: *Soft Matter* 9 (2013). DOI: 10.1039/c3sm50311a.

[157] Andrew J. Bannister and Tony Kouzarides. "Regulation of chromatin by histone modifications". In: *Cell Research* 21.3 (2011), pp. 381–395. ISSN: 10010602. DOI: 10.1038/cr.2011.22.

[158] Ksenia Skvortsova, Nicola Iovino, and Ozren Bogdanović. "Functions and mechanisms of epigenetic inheritance in animals". In: *Nature Reviews Molecular Cell Biology* 19 (12 Dec. 2018), pp. 774–790. ISSN: 14710080. DOI: 10.1038/s41580-018-0074-2.

[159] Zibo Zhao and Ali Shilatifard. "Epigenetic modifications of histones in cancer". In: *Genome Biology* 20.1 (2019), pp. 1–16. ISSN: 1474760X. DOI: 10.1186/s13059-019-1870-5.

[160] Glenn H. Fredrickson and Scott T. Milner. "Thermodynamics of random copolymer melts". In: *Physical Review Letters* 67.7 (1991), pp. 835–838. ISSN: 00319007. DOI: 10.1103/PhysRevLett.67.835.

[161] Glenn H. Fredrickson, Scott T. Milner, and Ludwik Leibler. "Multicritical Phenomena and Microphase Ordering in Random Block Copolymers Melts". In: *Macromolecules* 25.23 (23 1992), pp. 6341–6354. ISSN: 15205835. DOI: 10.1021/ma00049a034.

[162] Alexei R. Khokhlov and Pavel G. Khalatur. "Conformation-dependent sequence design (engineering) of AB copolymers". In: *Physical Review Letters* 82.17 (1999), pp. 3456–3459. ISSN: 10797114. DOI: 10.1103/PhysRevLett.82.3456.

[163] A. R. Khokhlov and P. G. Khalatur. "Protein-like copolymers: Computer simulation". In: *Russian Chemical Bulletin* 47.5 (1998), pp. 855–860. ISSN: 10665285. DOI: 10.1007/BF02498152.

[164] Artem M. Rumyantsev, Nicholas E. Jackson, Boyuan Yu, Jeffrey M. Ting, Wei Chen, Matthew V. Tirrell, and Juan J. De Pablo. "Controlling Complex Coacervation via Random Polyelectrolyte Sequences". In: *ACS Macro Letters* 8.10 (Oct. 2019), pp. 1296–1302. ISSN: 21611653. DOI: 10.1021/acsmacrolett.9b00494.

[165] Amy R. Strom, Ronald J. Biggs, Edward J. Banigan, Xiaotao Wang, Katherine Chiu, Cameron Herman, Jimena Collado, Feng Yue, Joan C.Ritland Politz, Leah J. Tait, David Scalzo, Agnes Telling, Mark Groudine, Clifford P. Brangwynne, John F. Marko, and Andrew D. Stephens. "Hp1α is a chromatin crosslinker that controls nuclear and mitotic chromosome mechanics". In: *eLife* 10 (June 2021). ISSN: 2050084X. DOI: 10.7554/eLife.63972.

[166] M. D. McKay, R. J. Beckman, and W. J. Conover. "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code". In: *Technometrics* 21.2 (1979), pp. 239–245. ISSN: 15372723. DOI: 10.1080/00401706.2000.10485979.

[167] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015), pp. 1–14. arXiv: 1409.1556.

[168] Tao Yang, Feipeng Zhang, Galip Gürkan Yardımci, Fan Song, Ross C. Hardison, William Stafford Noble, Feng Yue, and Qunhua Li. "HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient". In: *Genome Research* 27.11 (2017), pp. 1939–1949. ISSN: 15495469. DOI: 10.1101/gr.220640.117.

[169] Jason Ernst and Manolis Kellis. "ChromHMM: Automating chromatin-state discovery and characterization". In: *Nature Methods* 9.3 (2012), pp. 215–216. ISSN: 15487091. DOI: 10.1038/nmeth.1906.

[170] Denes Hnisz, Abraham S Weintraub, Daniel S Day, Anne-laure Valton, Rasmus O Bak, Charles H Li, Johanna Goldmann, Bryan R Lajoie, Zi Peng Fan, Alla A Sigova, Jessica Reddy, Diego Borges-rivera, Tong Ihn Lee, Rudolf Jaenisch, Matthew H Porteus, Job Dekker, and Richard A Young. "Activation of proto-oncogenes by disruption of chromosome neighborhoods". In: *Science* 351.6280 (2016), pp. 1454–1458. ISSN: 1095-9203. DOI: 10.1126/science.aad9024.

[171] P.T Lowary and J Widom. "New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning". In: *Journal of Molecular*

Biology 276.1 (Feb. 1998), pp. 19–42. ISSN: 0022-2836. DOI: 10.1006/JMBI.1997.1494.

[172] Peter V Kharchenko, Artyom A Alekseyenko, Yuri B Schwartz, Aki Minoda, Nicole C Riddle, Jason Ernst, Peter J Sabo, Erica Larschan, Andrey A Gorchakov, Tingting Gu, Daniela Linder-Basso, Annette Plachetka, Gregory Shanower, Michael Y Tolstorukov, Lovelace J Luquette, Ruibin Xi, Youngsook L Jung, Richard W Park, Eric P Bishop, Theresa K Canfield, Richard Sandstrom, Robert E Thurman, David M MacAlpine, John A Stamatoyannopoulos, Manolis Kellis, Sarah C R Elgin, Mitzi I Kuroda, Vincenzo Pirrotta, Gary H Karpen, and Peter J Park. "Comprehensive analysis of the chromatin landscape in Drosophila melanogaster." In: *Nature* 471.7339 (2011), pp. 480–485. ISSN: 0028-0836. DOI: 10.1038/nature09725.

[173] Gabriel E Zentner and Steven Henikoff. "Regulation of nucleosome dynamics by histone modifications." In: *Nat Struct Mol Biol* 20.3 (2013), pp. 259–66. ISSN: 1545-9985. DOI: 10.1038/nsmb.2470.

[174] Joshua P Lequieu, M Hinckley, and Juan J de Pablo. "A molecular view of DNA-conjugated nanoparticle association energies". In: *Soft Matter* 11.10 (2015), pp. 1919–1929. ISSN: 1744-683X. DOI: 10.1039/C4SM02573C.

[175] I. Kulić and H. Schiessel. "Chromatin Dynamics: Nucleosomes go Mobile through Twist Defects". In: *Phys. Rev. Lett.* 91.14 (Oct. 2003), p. 148103. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.91.148103.

[176] G. Arya and T. Schlick. "Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model". In: *Proc. Nat. Acad. Sci. USA.* 103.44 (2006), pp. 16236–16241. ISSN: 0027-8424. DOI: 10.1073/pnas.0604817103.

[177] Rene Stehr, Robert Schopflin, Ramona Ettlg, Nick Kepper, Karsten Rippe, and Gero Wedemann. "Exploring the conformational space of chromatin fibers and their stability by numerical dynamic phase diagrams". In: *Biophys. J.* 98.6 (2010), pp. 1028–1037. ISSN: 00063495. DOI: 10.1016/j.bpj.2009.11.040.

[178] Andrés Córdoba, Daniel M Hinckley, Joshua Lequieu, and Juan J de Pablo. "A Molecular View of the Dynamics of dsDNA Packing Inside Viral Capsids in the Presence of Ions". In: *Biophys. J.* 112.7 (2017), pp. 1302–1315. ISSN: 0006-3495. DOI: https://doi.org/10.1016/j.bpj.2017.02.015.

[179] Scott P. Carmichael and M. Scott Shell. "A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly". In: *J. Phys. Chem. B* 116.29 (2012), pp. 8383–8393. ISSN: 15205207. DOI: 10.1021/jp2114994.

[180] Kevin Struhl and Eran Segal. "Determinants of nucleosome positioning." In: *Nat. Struct. Mol. Biol.* 20.3 (Mar. 2013), pp. 267–73. ISSN: 1545-9985. DOI: 10.1038/nsmb.2506.

[181] Brian J. Beliveau, Alistair N. Boettiger, Maier S. Avendaño, Ralf Jungmann, Ruth B. McCole, Eric F. Joyce, Caroline Kim-Kiselak, Frédéric Bantignies, Chamith Y. Fonseka, Jelena Erceg, Mohammed A. Hannan, Hien G. Hoang, David Colognori, Jeannie T. Lee, William M. Shih, Peng Yin, Xiaowei Zhuang, and Chao-ting Wu. "Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes". In: *Nat. Commun.* 6.May (2015), p. 7147. ISSN: 2041-1723. DOI: 10.1038/ncomms8147.

[182] Alistair N Boettiger, Bogdan Bintu, Jeffrey R Moffitt, Siyuan Wang, Brian J Beliveau, Geoffrey Fudenberg, Maxim Imakaev, Leonid A Mirny, Chao-ting Wu, and Xiaowei Zhuang. "Super-resolution imaging reveals distinct chromatin folding for different epigenetic states". In: *Nature* 529.7586 (Jan. 2016), pp. 418–422. ISSN: 1476-4687. DOI: 10.1038/nature16496.

[183] Aram Davtyan, Nicholas P Schafer, Weihua Zheng, Cecilia Clementi, Peter G Wolynes, and Garegin a Papoian. "AWSEM-MD: Protein Structure Prediction Using Coarse-Gained Physical Potentials and Bioinformatically Based Local Structure Biasing." In: *J. Phys. Chem. B* 116 (2012), pp. 8494–8503. ISSN: 15205207. DOI: 10.1021/jp212541y.

[184] Jueng Soo You and Peter A. Jones. "Cancer Genetics and Epigenetics: Two Sides of the Same Coin?" In: *Cancer Cell* 22.1 (2012), pp. 9–20. ISSN: 15356108. DOI: 10.1016/j.ccr.2012.06.008. arXiv: NIHMS150003.

[185] Isabel Jimenez-Useche, Nathan P. Nurse, Yuqing Tian, Bhargav S. Kansara, Daphne Shim, and Chongli Yuan. "DNA methylation effects on tetra-nucleosome compaction and aggregation". In: *Biophysical Journal* 107.7 (2014), pp. 1629–1636. ISSN: 15420086. DOI: 10.1016/j.bpj.2014.05.055.

[186] J Widom and a Klug. "Structure of the 300A chromatin filament: X-ray diffraction from oriented samples." In: *Cell* 43.1 (1985), pp. 207–213. ISSN: 00928674. DOI: 10.1016/0092-8674(85)90025-X.

[187] Van A T Huynh, Philip J J Robinson, and Daniela Rhodes. "A method for the in vitro reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone". In: *Journal of Molecular Biology* 345.5 (2005), pp. 957–968. ISSN: 00222836. DOI: 10.1016/j.jmb.2004.10.075.

[188] Sergei A Grigoryev, Gavin Bascom, Jenna M Buckwalter, Michael B Schubert, Christopher L Woodcock, and Tamar Schlick. "Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes." In: *Proc. Nat. Acad. Sci. USA.* 113.5 (2016), pp. 1238–43. ISSN: 1091-6490. DOI: 10.1073/pnas.1518280113.

[189] Daniel M. Hinckley and Juan J. De Pablo. "Coarse-grained ions for nucleic acid modeling". In: *J. Chem. Theory Comput.* 11.11 (2015), pp. 5436–5446. ISSN: 15499626. DOI: 10.1021/acs.jctc.5b00341.

[190] W Wang, G W Li, C Chen, X S Xie, and X Zhuang. "Chromosome organization by a nucleoid-associated protein in live bacteria". In: *Science* 333.6048 (2011), pp. 1445–1449. ISSN: 0036-8075. DOI: 10.1126/science.1204697.

[191] Jesse R. Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. "Topological domains in mammalian genomes identified by analysis of chromatin interactions". In: *Nature* 485.7398 (2012), pp. 376–380. ISSN: 0028-0836. DOI: 10.1038/nature11082.

[192] Yu P. Kalmykov. "Rotational Brownian motion in an external potential: The Langevin equation approach". In: *J. Mol. Liq.* 69.9 SPEC. ISS. (1996), pp. 117–131. ISSN: 01677322. DOI: 10.1016/S0167-7322(96)90009-4.

[193] Jason Ernst, Pouya Kheradpour, Tarjei S Mikkelsen, Noam Shoresh, Lucas D Ward, Charles B Epstein, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael Coyne, Manching Ku, Timothy Durham, Manolis Kellis, and Bradley E Bernstein. "Mapping and analysis of chromatin state dynamics in nine human cell types." In: *Nature* 473.7345 (2011), pp. 43–9. ISSN: 1476-4687. DOI: 10.1038/nature09906.

[194] Gavin D. Bascom, Karissa Y. Sanbonmatsu, and Tamar Schlick. "Mesoscale Modeling Reveals Hierarchical Looping of Chromatin Fibers Near Gene Regulatory Elements". In: *J. Phys. Chem. B* 120.33 (Dec. 2016), pp. 8642–8653. ISSN: 15205207. DOI: 10.1021/acs.jpcb.6b03197.

[195] Nancy C. Stellwagen, Soffia Magnusdottir, Cecilia Gelfi, and Pier Giorgio Righetti. "Measuring the translational diffusion coefficients of small DNA molecules by capillary electrophoresis". In: *Biopolymers* 58.4 (2001), pp. 390–397. ISSN: 00063525. DOI: 10.1002/1097-0282(20010405)58:4<390::AID-BIP1015>3.0.CO;2-K.

[196] Stephanie Geggier and Alexander Vologodskii. "Sequence dependence of DNA bending rigidity." In: *Proceedings of the National Academy of Sciences of the United States of America* 107.35 (Aug. 2010), pp. 15421–6. ISSN: 1091-6490. DOI: 10.1073/pnas.1004809107.

[197] J. J. Funke, P. Ketterer, C. Lieleg, S. Schunter, P. Korber, and H. Dietz. "Uncovering the forces between nucleosomes using DNA origami". en. In: *Science Advances* 2.11 (Feb. 2016), e1600974–e1600974. ISSN: 2375-2548. DOI: 10.1126/sciadv.1600974.

[198] Giuseppe Chirico and Jörg Langowski. "Kinetics of DNA supercoiling studied by Brownian dynamics simulation". In: *Biopolymers* 34 (1994), pp. 415–433.

[199] W Eimer and R Pecora. "Rotational and Translational Diffusion of Short Rodlike Molecules in Solution - Oligonucleotides". In: *J Chem Phys* 94.3 (1991), pp. 2324–2329. ISSN: 00219606. DOI: 10.1063/1.459904.

[200] Donna L Gosnell and Bruno H. Zimm. "Measurement of diffusion coefficients of DNA in agarose gel". In: *Macromolecules* 26.6 (1993), pp. 1304–1308. ISSN: 0024-9297. DOI: 10.1021/ma00058a018.

[201] Fan Tso Chien and Thijn Van Der Heijden. "Characterization of nucleosome unwrapping within chromatin fibers using magnetic tweezers". In: *Biophysical Journal* 107.2 (2014), pp. 373–383. ISSN: 15420086. DOI: 10.1016/j.bpj.2014.05.036.

[202] Gordon S. Freeman, Daniel M. Hinckley, Joshua P. Lequieu, Jonathan K. Whitmer, and Juan J. de Pablo. "Coarse-grained modeling of DNA curvature". In: *J. Chem. Phys.* 141.16 (Oct. 2014), p. 165103. ISSN: 0021-9606. DOI: 10.1063/1.4897649.

[203] Nick Kepper, Ramona Ettig, Rene Stehr, Sven Marnach, Gero Wedemann, and Karsten Rippe. "Force spectroscopy of chromatin fibers: extracting energetics and structural information from Monte Carlo simulations." In: *Biopolymers* 95.7 (July 2011), pp. 435–47. ISSN: 0006-3525. DOI: 10.1002/bip.21598.

[204] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. "Ionic effects on the elasticity of single DNA molecules". In: *Proc. Nat. Acad. Sci. USA.* 94.12 (1997), pp. 6185–6190. ISSN: 0027-8424. DOI: 10.1073/pnas.94.12.6185.

[205] Song Tan and Curt A. Davey. "Nucleosome structural studies". In: *Curr. Opin. Struc. Biol.* 21.1 (2011), pp. 128–136. ISSN: 0959440X. DOI: 10.1016/j.sbi.2010.11.006.

[206] Chih L. Liu, Tommy Kaplan, Minkyu Kim, Stephen Buratowski, Stuart L. Schreiber, Nir Friedman, and Oliver J. Rando. "Single-nucleosome mapping of histone modifications in S. cerevisiae". In: *PLoS Biology* 3.10 (2005). ISSN: 15449173. DOI: 10.1371/journal.pbio.0030328.

[207] Joshua Lequieu, Andrés Córdoba, Daniel Hinckley, and Juan J. de Pablo. "Mechanical Response of DNA–Nanoparticle Crystals to Controlled Deformation". In: *ACS Cent. Sci.* 2.9 (2016), pp. 614–620. ISSN: 2374-7943. DOI: 10.1021/acscentsci.6b00170.

[208] J C Hansen, J Ausio, V H Stanik, and K E van Holde. "Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1." In: *Biochemistry* 28.23 (1989), pp. 9129–9136. ISSN: 0006-2960. DOI: 10.1021/bi00449a026.

[209] Gero Wedemann and Jörg Langowski. "Computer simulation of the 30-nanometer chromatin fiber." In: *Biophys. J.* 82.6 (June 2002), pp. 2847–59. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(02)75627-0.

[210] C L Woodcock, S a Grigoryev, R a Horowitz, and N Whitaker. "A chromatin folding model that incorporates linker variability generates fibers resembling the native structures." In: *Proc. Nat. Acad. Sci. USA.* 90.19 (1993), pp. 9021–9025. ISSN: 0027-8424. DOI: 10.1073/pnas.90.19.9021.

[211] Hiromi Yamakawa. "Statistical mechanics of helical wormlike chains. I. Differential equations and moments". In: *J. Chem. Phys.* 64.12 (1976), p. 5222. ISSN: 00219606. DOI: 10.1063/1.432197.

[212] Hannah S Tims, Kaushik Gurunathan, Marcia Levitus, and Jonathan Widom. "Dynamics of nucleosome invasion by DNA binding proteins." In: *J. Mol. Biol.* 411.2 (Aug. 2011), pp. 430–48. ISSN: 1089-8638. DOI: 10.1016/j.jmb.2011.05.044.

[213] Philip J J Robinson, Louise Fairall, Van a T Huynh, and Daniela Rhodes. "EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure." In: *Proc. Natl. Acad. Sci. U.S.A.* 103.17 (2006), pp. 6506–6511. ISSN: 0027-8424. DOI: 10.1073/pnas.0601212103.

[214] Steve Plimpton. "Fast Parallel Algorithms for Short-Range Molecular Dynamics". In: *J. Comput. Phys.* 117 (1995), pp. 1–19.

[215] Felix Mueller-Planitz, Henrike Klinker, and Peter B Becker. "Nucleosome sliding mechanisms: new twists in a looped history." In: *Nat. Struct. Mol. Biol.* 20.9 (Sept. 2013), pp. 1026–32. ISSN: 1545-9985. DOI: 10.1038/nsmb.2648.

[216] Daniel M Hinckley, Gordon S Freeman, Jonathan K Whitmer, and Juan J de Pablo. "An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization." In: *J. Chem. Phys.* 139.14 (Oct. 2013), p. 144903. ISSN: 1089-7690. DOI: 10.1063/1.4822042.

[217] Thomas A Knotts, Nitin Rathore, David C Schwartz, and Juan J de Pablo. "A coarse grain model for DNA." In: *J. Chem. Phys.* 126.8 (Mar. 2007), p. 084901. ISSN: 0021-9606. DOI: 10.1063/1.2431804.

[218] Viviana I. Risca and William J. Greenleaf. "Unraveling the 3D genome: Genomics tools for multiscale exploration". In: *Trends in Genetics* 31.7 (2015), pp. 357–372. ISSN: 13624555. DOI: 10.1016/j.tig.2015.03.010.

[219] E J Sambriski, D C Schwartz, and J J de Pablo. "A mesoscale model of DNA and its renaturation." In: *Biophys. J.* 96.5 (Mar. 2009), pp. 1675–1690. ISSN: 1542-0086. DOI: 10.1016/j.bpj.2008.09.061.

[220] Rajeswari S. Edayathumangalam, Philipp Weyermann, Peter B. Dervan, Joel M. Gottesfeld, and Karolin Luger. "Nucleosomes in solution exist as a mixture of twist-defect states". In: *J. Mol. Biol.* 345.1 (2005), pp. 103–114. ISSN: 00222836. DOI: 10.1016/j.jmb.2004.10.012.

[221] Hiromi Yamakawa and Takenao Yoshizaki. "Dynamics of helical wormlike chains. I. Dynamic model and diffusion equation". In: *The Journal of Chemical Physics* 75.2 (1981), pp. 1016–1030. ISSN: 00219606. DOI: 10.1063/1.442065.

[222] Jan Bednar, Rachel A. Horowitz, Jacques Dubochet, and Christopher L. Woodcock. "Chromatin conformation and salt-induced compaction: Three-dimensional structural information from cryoelectron microscopy". In: *Journal of Cell Biology* 131.6 I (1995), pp. 1365–1376. ISSN: 00219525. DOI: 10.1083/jcb.131.6.1365.

[223] Johannes Kästner. "Umbrella sampling". In: *Wiley Interdiscip Rev Comput Mol Sci* 1.6 (Nov. 2011), pp. 932–942. ISSN: 17590876. DOI: 10.1002/wcms.66.

[224] Tsung Han S Hsieh, Assaf Weiner, Bryan Lajoie, Job Dekker, Nir Friedman, and Oliver J. Rando. "Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C". In: *Cell* 162.1 (2015), pp. 108–119. ISSN: 10974172. DOI: 10.1016/j.cell.2015.05.048.

[225] Job Dekker, Marc A Marti-Renom, and Leonid A Mirny. "Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data." In: *Nat. Rev. Genet.* 14.6 (June 2013), pp. 390–403. ISSN: 1471-0064. DOI: 10.1038/nrg3454.

[226] Eden Fussner, Reagan W. Ching, and David P. Bazett-Jones. "Living without 30nm chromatin fibers". In: *Trends Biochem. Sci.* 36.1 (2011), pp. 1–6. ISSN: 09680004. DOI: 10.1016/j.tibs.2010.09.002.

[227] Gu Li, Marcia Levitus, Carlos Bustamante, and Jonathan Widom. "Rapid spontaneous accessibility of nucleosomal DNA." In: *Nat Struct Mol Biol* 12.1 (Jan. 2005), pp. 46–53. ISSN: 1545-9993. DOI: 10.1038/nsmb869.

[228] Alessandro Laio and Francesco L Gervasio. "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science". In: *Rep Prog Phys* 71.12 (Dec. 2008), p. 126601. ISSN: 0034-4885. DOI: 10.1088/0034-4885/71/12/126601.

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

[230] Hidde Tj. Goinga and R Pecora. "Dynamics of Low Molecular Weight DNA Fragments in Dilute and Semidilute Solutions Hidde T j". In: *Macromolecules* 24 (1991), pp. 6128–6138.

[231] J Yao, P T Lowary, and J Widom. "Direct detection of linker DNA bending in defined-length oligomers of chromatin." In: *Proceedings of the National Academy of Sciences of the United States of America* 87.October (1990), pp. 7603–7607. ISSN: 0027-8424. DOI: 10.1073/pnas.87.19.7603.

[232] Le Chang and Shoji Takada. "Histone acetylation dependent energy landscapes in tri-nucleosome revealed by residue-resolved molecular simulations". In: *Scientific Reports* 6.March (2016), p. 34441. ISSN: 2045-2322. DOI: 10.1038/srep34441.

[233] Joshua Lequieu, David C Schwartz, and Juan J. de Pablo. "In silico evidence for sequence-dependent nucleosome sliding". In: *Proc. Nat. Acad. Sci. USA.* 114.44 (2017), E9197–E9205. ISSN: 0027-8424. DOI: 10.1073/pnas.1705685114.

[234] J. Yao, P. T. Lowary, J. Widom, and J. Yao. "Linker DNA Bending Induced by the Core Histones of Chromatin". In: *Biochemistry* 30.34 (1991), pp. 8408–8414. ISSN: 15204995. DOI: 10.1021/bi00098a019.

[235] G Voordouw, Z Kam, N Borochov, and H Eisenberg. "Isolation and Physical Studies of Intact Supercoiled, the open circular and the linear forms of ColE1-Plasmid DNA". In: *Biophysical Chemistry* 8 (1978), pp. 171–189.

[236] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. "Multidimensional free-energy calculations using the weighted histogram analysis method". In: *J. Comput. Chem.* 16.11 (1995), pp. 1339–1350. ISSN: 1096-987X. DOI: 10.1002/jcc.540161104.

[237] Brent Brower-Toland, David A. Wacker, Robert M. Fulbright, John T. Lis, W. Lee Kraus, and Michelle D. Wang. "Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes". In: *J. Mol. Biol.* 346.1 (2005), pp. 135–146. ISSN: 00222836. DOI: 10.1016/j.jmb.2004.11.056.

[238] Shirley Mihardja, Andrew J Spakowitz, Yongli Zhang, and Carlos Bustamante. "Effect of force on mononucleosomal dynamics." In: *Proc. Nat. Acad. Sci. USA*. 103.43 (Oct. 2006), pp. 15871–15876. ISSN: 0027-8424. DOI: 10.1073/pnas.0607526103.

[239] Wenfei Li, Peter G Wolynes, and Shoji Takada. "Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins." In: *Proc. Nat. Acad. Sci. USA*. 108.9 (Mar. 2011), pp. 3504–3509. ISSN: 1091-6490. DOI: 10.1073/pnas.1018983108.

[240] David J Tremethick. "Higher-order structures of chromatin: the elusive 30 nm fiber." In: *Cell* 128.4 (Feb. 2007), pp. 651–4. ISSN: 0092-8674. DOI: 10.1016/j.cell.2007.02.008.

[241] C A Brackley, A N Morozov, and D Marenduzzo. "Models for twistable elastic polymers in Brownian dynamics, and their implementation for LAMMPS." In: *J. Chem. Phys.* 140.13 (Apr. 2014), p. 135103. ISSN: 1089-7690. DOI: 10.1063/1.4870088.

[242] Curt A Davey, David F Sargent, Karolin Luger, Armin W Maeder, and Timothy J Richmond. "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution." In: *J. Mol. Biol.* 319.5 (June 2002), pp. 1097–1113. ISSN: 0022-2836. DOI: 10.1016/S0022-2836(02)00386-8.

[243] Susan S Sorlie and Robert Pecora. "A dynamic light scattering study of four DNA restriction fragments". In: *Macromolecules* 23.2 (1990), pp. 487–497. ISSN: 0024-9297. DOI: 10.1021/ma00204a022.

[244] Bin Zhang, Weihua Zheng, Garegin A. Papoian, and Peter G Wolynes. "Exploring the Free Energy Landscape of Nucleosomes". In: *J. Am. Chem. Soc.* 138.26 (2016), pp. 8126–8133. ISSN: 0002-7863. DOI: 10.1021/jacs.6b02893.

[245] A.J. Stone. "The description of bimolecular potentials, forces and torques: the S and V function expansions". In: *Mol. Phys.* 36.1 (1978), pp. 241–256. ISSN: 0026-8976. DOI: 10.1080/00268977800101541.

[246] Joshua Lequieu, Andrés Córdoba, David C. Schwartz, and Juan J. de Pablo. "Tension-Dependent Free Energies of Nucleosome Unwrapping". In: *ACS Cent. Sci.* 2.9 (2016), pp. 660–666. ISSN: 2374-7943. DOI: 10.1021/acscentsci.6b00201.

[247] Gaurav Arya and Tamar Schlick. "A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments". In: *J. Phys. Chem. A* 113.16 (2009), pp. 4045–4059. ISSN: 10895639. DOI: 10.1021/jp810375d.

[248] Gordon S. Freeman, Joshua P. Lequieu, Daniel M. Hinckley, Jonathan K. Whitmer, and Juan J. de Pablo. "DNA Shape Dominates Sequence Affinity in Nucleosome Formation". In: *Phys. Rev. Lett.* 113.16 (Oct. 2014), p. 168101. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.113.168101.

[249] Biqin Dong, Luay M. Almassalha, Yolanda Stypula-Cyrus, Ben E. Urban, John E. Chandler, The-Quyen Nguyen, Cheng Sun, Hao F. Zhang, and Vadim Backman. "Superresolution intrinsic fluorescence imaging of chromatin utilizing native, unmodified nucleic acids for contrast". In: *Proc. Nat. Acad. Sci. USA.* 113 (2016), p. 201602202. ISSN: 0027-8424. DOI: 10.1073/PNAS.1602202113.

[250] M Scott Shell. "The relative entropy is fundamental to multiscale and inverse thermodynamic problems." In: *J. Chem. Phys.* 129.14 (Oct. 2008), p. 144108. ISSN: 1089-7690. DOI: 10.1063/1.2992060.

[251] Joel M. Gottesfeld, Jason M. Belitsky, Christian Melander, Peter B. Dervan, and Karolin Luger. "Blocking Transcription Through a Nucleosome with Synthetic DNA Ligands". In: *J. Mol. Biol.* 321.2 (Aug. 2002), pp. 249–263. ISSN: 00222836. DOI: 10.1016/S0022-2836(02)00598-3.

[252] Habtamu Zewdie. "Computer-simulation studies of diskotic liquid crystals". In: *Phys. Rev. E* 57.2 (1998), pp. 1793–1805.

[253] Benedetta Dorigo, Thomas Schalch, Alexandra Kulangara, Sylwia Duda, Rasmus R Schroeder, and Timothy J Richmond. "Nucleosome arrays reveal the two-start organization of the chromatin fiber." In: *Science* 306.5701 (2004), pp. 1571–1573. ISSN: 0036-8075. DOI: 10.1126/science.1103124.

[254] Lixiao Wang and Hyuk Yu. "Chain Conformation of Linear Polyelectrolyte in Salt Solutions: Sodium Poly(styrene sulfonate) in Potassium Chloride and Sodium Chloride Solutions". In: *Macrom* 21 (1988), pp. 3498–3501.

[255] Yahli Lorch, Barbara Davis, and Roger D Kornberg. "Chromatin remodeling by DNA bending, not twisting." In: *Proc. Nat. Acad. Sci. USA.* 102.5 (2005), pp. 1329–1332. ISSN: 0027-8424. DOI: 10.1073/pnas.0409413102.

[256] Bryant S Fujimoto and J. Michael Schurr. "Dependence of the torsional rigidity of DNA on base composition". In: *Nature* 344 (1990), pp. 175–178. ISSN: 00219797. DOI: 10.1016/0021-9797(80)90501-9. arXiv: NIHMS150003.

[257] Jan Lipfert, Jacob W J Kerssemakers, Tessa Jager, and Nynke H. Dekker. "Magnetic torque tweezers: Measuring torsional stiffness in DNA and RecA-DNA filaments". In: *Nature Methods* 7.12 (2010), pp. 977–980. ISSN: 15487091. DOI: 10.1038/nmeth.1520.

[258] John K. Eykelenboom, Marek Gierlinski, Zuojun Yue, Nadia Hegarat, Hilary Pollard, Tatsuo Fukagawa, Helfrid Hochegger, and Tomoyuki U. Tanaka. "Live imaging of marked chromosome regions reveals their dynamic resolution and compaction in mitosis". In: *Journal of Cell Biology* 218 (5 May 2019), pp. 1531–1552. ISSN: 15408140. DOI: 10.1083/jcb.201807125.

[259] Tadasu Nozaki, Ryosuke Imai, Mai Tanbo, Ryosuke Nagashima, Sachiko Tamura, Tomomi Tani, Yasumasa Joti, Masaru Tomita, Kayo Hibino, Masato T. Kanemaki, Kerstin S. Wendt, Yasushi Okada, Takeharu Nagai, and Kazuhiro Maeshima. "Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging". In: *Molecular Cell* 67 (2 July 2017), 282–293.e7. ISSN: 10974164. DOI: 10.1016/j.molcel.2017.06.018.

[260] Pablo Aurelio Gómez-García, Stephanie Portillo-Ledesma, Maria Victoria Neguembor, Martina Pesaresi, Walaa Oweis, Talia Rohrlich, Stefan Wieser, Eran Meshorer, Tamar Schlick, Maria Pia Cosma, and Melike Lakadamyali. "Mesoscale Modeling and Single-Nucleosome Tracking Reveal Remodeling of Clutch Folding and Dynamics in Stem Cell Differentiation". In: *Cell Reports* 34.2 (2021). ISSN: 22111247. DOI: 10.1016/j.celrep.2020.108614.

[261] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. "Formation of Chromosomal Domains by Loop Extrusion". In: *Cell Reports* 15 (9 Jan. 2016), pp. 2038–2049. ISSN: 22111247. DOI: 10.1016/j.celrep.2016.04.085.

[262] Orsolya Symmons, Veli Vural Uslu, Taro Tsujimura, Sandra Ruf, Sonya Nassari, Wibke Schwarzer, Laurence Ettwiller, and François Spitz. "Functional and topological characteristics of mammalian regulatory domains". In: *Genome Research* 24 (3 Oct. 2014), pp. 390–400. ISSN: 15495469. DOI: 10.1101/gr.163519.113.

[263] Geoffrey Fudenberg, Nezar Abdennur, Maxim Imakaev, Anton Goloborodko, and Leonid Mirny. "Emerging Evidence of Chromosome Folding by Loop Extrusion". In: *bioRxiv* (Sept. 2018), p. 264648. DOI: 10.1101/264648.

[264] J.D. Watson and F. H. C. Crick. "A Structure for Deoxyribose Nucleic Acid". In: *Nature* 171 (1953), pp. 737–738.

[265] Thomas Cremer and Marion Cremer. "Chromosome territories." In: *Cold Spring Harbor perspectives in biology* 2 (3 2010). ISSN: 19430264. DOI: 10.1101/cshperspect.a003889.

[266] Stephanie Portillo-Ledesma, Lucille H. Tsao, Meghna Wagley, Melike Lakadamyali, Maria Pia Cosma, and Tamar Schlick. "Nucleosome Clutches are Regulated by Chromatin Internal Parameters". In: *Journal of Molecular Biology* 433 (6 2021), p. 166701. ISSN: 10898638. DOI: 10.1016/j.jmb.2020.11.001.

[267] Maria Aurelia Ricci, Carlo Manzo, María Filomena García-Parajo, Melike Lakadamyali, and Maria Pia Cosma. "Chromatin fibers are formed by heterogeneous groups of nu-

cleosomes in vivo". In: *Cell* 160 (6 2015), pp. 1145–1158. ISSN: 10974172. DOI: 10.1016/j.cell.2015.01.054.

[268] J. Besag. "Comments on Ripley's paper". In: *Royal Statistical Society Journal* 39 (1977), pp. 193–195.

[269] B. D. Ripley. "Modelling Spatial Patterns". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.2 (1977), pp. 172–192. DOI: 10.1111/j.2517-6161.1977.tb01615.x.

[270] Norman C. Beaulieu and Kasun T. Hemachandra. "Novel representations for the bivariate Rician distribution". In: *IEEE Transactions on Communications* 59.11 (2011), pp. 2951–2954. ISSN: 00906778. DOI: 10.1109/TCOMM.2011.092011.090171.

[271] Dmitry V Fyodorov, Bing-Rui Zhou, Arthur I Skoultchi, and Yawen Bai. "Emerging roles of linker histones in regulating chromatin structure and function". en. In: *Nature Reviews Molecular Cell Biology* 19.3 (Sept. 2018), pp. 192–206. ISSN: 1471-0080. DOI: 10.1038/nrm.2017.94.

[272] Emily C. Davidson, Adrianne M. Rosales, Anastasia L. Patterson, Boris Russ, Beihang Yu, Ronald N. Zuckermann, and Rachel A. Segalman. "Impact of Helical Chain Shape in Sequence-Defined Polymers on Polypeptoid Block Copolymer Self-Assembly". In: *Macromolecules* 51.5 (Mar. 2018), pp. 2089–2098. ISSN: 15205835. DOI: 10.1021/acs.macromol.8b00055.

[273] Kenneth J. Hanley, Timothy P. Lodge, and Ching I. Huang. "Phase behavior of a block copolymer in solvents of varying selectivity". In: *Macromolecules* 33.16 (Aug. 2000), pp. 5918–5931. ISSN: 00249297. DOI: 10.1021/ma000318b.

[274] Shifan Mao, Quinn MacPherson, and Andrew J. Spakowitz. "Fluctuation Effects in Semiflexible Diblock Copolymers". In: *ACS Macro Letters* 7.1 (Jan. 2018), pp. 59–64. ISSN: 21611653. DOI: 10.1021/acsmacrolett.7b00638.

[275] James McCarty, Kris T. Delaney, Scott P.O. Danielsen, Glenn H. Fredrickson, and Joan Emma Shea. "Complete Phase Diagram for Liquid-Liquid Phase Separation of Intrinsically Disordered Proteins". In: *Journal of Physical Chemistry Letters* 10.8 (Apr. 2019), pp. 1644–1652. ISSN: 19487185. DOI: 10.1021/acs.jpclett.9b00099.

[276] J. T. G. Overbeek and M. J. Voorn. "Phase separation in polyelectrolyte solutions. Theory of complex coacervation". In: *Journal of Cellular and Comparative Physiology* 49.S1 (May 1957), pp. 7–26. ISSN: 0095-9898. DOI: 10.1002/jcp.1030490404.

[277] Artem M. Rumyantsev, Alexey A. Gavrilov, and Elena Yu Kramarenko. "Electrostatically Stabilized Microphase Separation in Blends of Oppositely Charged Polyelectrolytes". In: *Macromolecules* 52.19 (Oct. 2019), pp. 7167–7174. ISSN: 15205835. DOI: 10.1021/acs.macromol.9b00883.

[278] William M. Aumiller, Bradley W. Davis, and Christine D. Keating. *Phase Separation as a Possible Means of Nuclear Compartmentalization*. Jan. 2014. DOI: 10.1016/B978-0-12-800046-5.00005-9.

[279] Ronald Hancock. "A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus". In: *Journal of Structural Biology* 146.3 (June 2004), pp. 281–290. ISSN: 10478477. DOI: 10.1016/j.jsb.2003.12.008.

[280] Alicia N. Schep, Jason D. Buenrostro, Sarah K. Denny, Katja Schwartz, Gavin Sherlock, and William J. Greenleaf. "Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions". In: *Genome Research* 25.11 (Nov. 2015), pp. 1757–1770. ISSN: 15495469. DOI: 10.1101/gr.192294.115.

[281] M. Jordan Rowley, Michael H. Nichols, Xiaowen Lyu, Masami Ando-Kuri, I. Sarahi M. Rivera, Karen Hermetz, Ping Wang, Yijun Ruan, and Victor G. Corces. "Evolutionarily Conserved Principles Predict 3D Chromatin Organization". In: *Molecular Cell* 67.5 (Sept. 2017), 837–852.e7. ISSN: 10974164. DOI: 10.1016/j.molcel.2017.07.022.

[282] Sabine M. Görisch, Malte Wachsmuth, Katalin Fejes Tóth, Peter Lichter, and Karsten Rippe. "Histone acetylation increases chromatin accessibility". In: *Journal of Cell Science* 118.24 (Dec. 2005), pp. 5825–5834. ISSN: 00219533. DOI: 10.1242/jcs.02689.

[283] Uwe Schwartz, Attila Németh, Sarah Diermeier, Josef H Exler, Stefan Hansch, Rodrigo Maldonado, Leonhard Heizinger, Rainer Merkl, and Gernot Längst. "Characterizing the nuclease accessibility of DNA in human cells to map higher order structures of chromatin". In: *Nucleic Acids Research* 47.3 (2019), pp. 1239–1254. ISSN: 13624962. DOI: 10.1093/nar/gky1203.

[284] Jonathan Schmidt, Mário R.G. Marques, Silvana Botti, and Miguel A.L. Marques. "Recent advances and applications of machine learning in solid-state materials science". In: *npj Computational Materials* 5.1 (2019). ISSN: 20573960. DOI: 10.1038/s41524-019-0221-0.

[285] Miguel Reyes, Kianna Billman, Nir Hacohen, and Paul C. Blainey. "Simultaneous Profiling of Gene Expression and Chromatin Accessibility in Single Cells". In: *Advanced Biosystems* 3.11 (Nov. 2019), p. 1900065. ISSN: 2366-7478. DOI: 10.1002/adbi.201900065.

[286] Peter J. Sabo, Michael S. Kuehn, Robert Thurman, Brett E. Johnson, Ericka M. Johnson, Hua Cao, Man Yu, Elizabeth Rosenzweig, Jeff Goldy, Andrew Haydock, Molly Weaver, Anthony Shafer, Kristin Lee, Fidencio Neri, Richard Humbert, Michael A. Singer, Todd A. Richmond, Michael O. Dorschner, Michael McArthur, Michael Hawrylycz, Roland D. Green, Patrick A. Navas, William S. Noble, and John A. Stamatoyannopoulos. "Genome-scale mapping of DNase I sensitivity in vivo using tiling

DNA microarrays". In: *Nature Methods* 3.7 (July 2006), pp. 511–518. ISSN: 15487091. DOI: 10.1038/nmeth890.

[287] Jay R. Hesselberth, Xiaoyu Chen, Zhihong Zhang, Peter J. Sabo, Richard Sandstrom, Alex P. Reynolds, Robert E. Thurman, Shane Neph, Michael S. Kuehn, William S. Noble, Stanley Fields, and John A. Stamatoyannopoulos. "Global mapping of protein-DNA interactions in vivo by digital genomic footprinting". In: *Nature Methods* 6.4 (Mar. 2009), pp. 283–289. ISSN: 15487091. DOI: 10.1038/nmeth.1313.

[288] Alison E. White, Aaron R. Hieb, and Karolin Luger. "A quantitative investigation of linker histone interactions with nucleosomes and chromatin". In: *Scientific Reports* 6.1 (Jan. 2016), pp. 1–14. ISSN: 20452322. DOI: 10.1038/srep19122.

[289] Aaron R. Hieb, Sheena D'Arcy, Michael A. Kramer, Alison E. White, and Karolin Luger. "Fluorescence strategies for high-throughput quantification of protein interactions". In: *Nucleic Acids Research* 40.5 (2012), pp. 1–13. ISSN: 03051048. DOI: 10.1093/nar/gkr1045.

[290] Tamara L Caterino, He Fang, and Jeffrey J Hayes. "Nucleosome linker DNA contacts and induces specific folding of the intrinsically disordered H1 carboxyl-terminal domain." In: *Molecular and cellular biology* 31.11 (June 2011), pp. 2341–8. ISSN: 1098-5549. DOI: 10.1128/MCB.05145-11.

[291] K. Ura, K. Nightingale, and A. P. Wolffe. "Differential association of HMG1 and linker histones B4 and H1 with dinucleosomal DNA: structural transitions and transcriptional repression." In: *The EMBO Journal* 15.18 (Sept. 1996), pp. 4959–4969. ISSN: 02614189. DOI: 10.1002/j.1460-2075.1996.tb00876.x.

[292] Yuhong Fan, Tatiana Nikitina, Jie Zhao, Tomara J. Fleury, Riddhi Bhattacharyya, Eric E. Bouhassira, Arnold Stein, Christopher L. Woodcock, and Arthur I. Skoultchi. "Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation". In: *Cell* 123.7 (Dec. 2005), pp. 1199–1212. ISSN: 00928674. DOI: 10.1016/j.cell.2005.10.028.

[293] Geert Geeven, Yun Zhu, Byung Ju Kim, Boris A. Bartholdy, Seung Min Yang, Todd S. Macfarlan, Wesley D. Gifford, Samuel L. Pfaff, Marjon J.A.M. Verstegen, Hugo Pinto, Marit W. Vermunt, Menno P. Creyghton, Patrick J. Wijchers, John A. Stamatoyannopoulos, Arthur I. Skoultchi, and Wouter de Laat. "Local compartment changes and regulatory landscape alterations in histone H1-depleted cells". In: *Genome Biology* 16.1 (Dec. 2015), p. 289. ISSN: 1474760X. DOI: 10.1186/s13059-015-0857-0.

[294] V. I. Tokar and H. Dreyssé. "Analytical solution of 1D lattice gas model with infinite number of multiatom interactions". In: *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 68.1 (Oct. 2002), p. 5. DOI: 10.1103/PhysRevE.68.011601. arXiv: 0210582 [cond-mat].

[295] Bing Rui Zhou, Jiansheng Jiang, Hanqiao Feng, Rodolfo Ghirlando, T. Sam Xiao, and Yawen Bai. "Structural Mechanisms of Nucleosome Recognition by Linker Histones".

In: *Molecular Cell* 59.4 (Aug. 2015), pp. 628–638. ISSN: 10974164. DOI: 10.1016/j.molcel.2015.06.025.

[296] Timothy J Stasevich, Florian Mueller, David T Brown, and James G McNally. "Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis". eng. In: *The EMBO journal* 29.7 (2010), pp. 1225–1234. ISSN: 1460-2075. DOI: 10.1038/emboj.2010.24.

[297] Karl P. Nightingale, Dmitry Pruss, and Alan P. Wolffe. "A single high affinity binding site for histone H1 in a nucleosome containing the Xenopus borealis 5 S ribosomal RNA gene". In: *Journal of Biological Chemistry* 271.12 (1996), pp. 7090–7094. ISSN: 00219258. DOI: 10.1074/jbc.271.12.7090.