# 1CPN: A coarse-grained multi-scale model of chromatin

Joshua Lequieu, Andrés Córdoba, and Joshua Moller
*Institute for Molecular Engineering*
*University of Chicago, Chicago, IL 60637 USA*
Juan J. de Pablo
*Institute for Molecular Engineering*
*University of Chicago, Chicago, IL 60637 USA and*
*Materials Science Division*
*Argonne National Laboratory, Argonne, IL 60439 USA*[a)]

(Dated: 15 February 2019)

A central question in epigenetics is how histone modifications influence the 3-d structure of eukaryotic genomes and, ultimately how this 3-d structure is manifested in gene expression. The wide range of length scales that influence 3-d genome structure present important challenges; epigenetic modifications to histones occur on scales of Angstroms, yet the resulting effects of these modifications on genome structure can span microns. There is a scarcity of computational tools capable of providing a mechanistic picture of how molecular information from individual histones is propagated up to large regions of the genome. In this work, a new molecular model of chromatin is presented that provides such a picture. This new model, referred to as 1CPN, is structured around a rigorous multi-scale approach whereby free energies from an established and extensively validated model of the nucleosome is mapped onto a reduced coarse-grained topology. As such, 1CPN incorporates detailed physics from the nucleosome, such as histone modifications and DNA sequence, while maintaining the computational efficiency that is required to permit kilobase-scale simulations of genomic DNA. The 1CPN model reproduces the free energies and dynamics of both single nucleosomes and short chromatin fibers, and is shown to be compatible with recently developed models of the linker histone. It is applied here to examine the effects of linker DNA on the free energies of chromatin assembly and to demonstrate that these free energies are strongly dependent on linker DNA length, pitch and even DNA sequence. The 1CPN model is implemented in LAMMPS and is distributed freely for public use.

## I. INTRODUCTION

The assembly and compaction of eukaryotic DNA into chromatin represents a critical component of cellular function. Modifications to the processes that dictate chromatin folding, such as histone modifications, represent an additional mechanism by which cellular functions such as transcription and replication are regulated. These modifications, known collectively as the epigenome, are central to human health, and a rapidly growing body of literature is finding these processes to be linked with diseases such as cancer[1,2]. Developing a mechanistic understanding of the processes that dictate DNA compaction, and how these processes can be manipulated, is central to molecular biology, biophysics, and, ultimately, human health.

In recent years, genome-wide maps of histone modifications have provided researchers with molecular-level information about chromatin compaction. In particular, it has been found that many histone modifications do not act in isolation, but instead act collectively to dictate cellular function[3]. By analyzing these genome-wide data sets, various groups have identified strong correlations between different patterns in histone modifications and the resulting function of chromatin[4–6]. As foundational as these studies have been, however, they have not yet attempted to translate these correlations into a mechanistic understanding of chromatin that explains how modifications collectively dictate three-dimensional structure and gene expression.

One approach towards such a mechanistic understanding is afforded by so-called "chromatin conformation capture" techniques, which can provide a map of the three-dimensional contacts between different parts of the genome[7]. These measurements have revealed several surprises, such as the fractal structure of chromatin[8,9], and the existence of topologically associated domains[10]. Though initial experiments could only resolve genome contacts with 1000 kb resolution, recent techniques have achieved a resolution of up to one or several nucleosomes[11,12]. Super-resolution microscopy has also been used to directly visualize chromatin compaction, shedding light on multiple, previously unknown subtle features of this process[13–17]. Though both chromatin capture and microscopic techniques have significantly advanced our understanding of large scale structures within the genome, they do not possess sufficient resolution to resolve the interactions between individual nucleosomes. Resolving these interactions is critical, because it is over these length scales that DNA-protein binding occurs, and where histone modifications likely play the largest role[18].

The challenges associated with understanding the scales over which multiple nucleosomes interact together represent a continuation of a venerable body of literature trying to understand the "chromatin fiber"[19,20]. Early work focused on whether this fiber is a "solenoid" one-

[a)]Electronic mail: Corresponding Author: depablo@uchicago.edu

start helix[21–24], where linker DNA between neighboring nucleosomes is bent, or a "zig-zag" two-start helix[25–29], where DNA passes through the center of the fiber. Other work has called the very existence of this fiber *in vivo* into question, suggesting that at physiologically relevant concentrations a well-defined fiber may not even exist[30,31] and, instead, that *in vivo* chromatin is in a liquid-like state[32]. Though progress continues to be made in this field[33,34], understanding chromatin at the nano-scale and how conformations are altered by epigenetic modifications is a significant challenge.

Computational models provide an opportunity to bridge this knowledge gap, by providing a link between intricate, nanometer length-scale processes and those more easily observable in experiments. A particularly successful approach was developed by Arya, Schlick and co-workers, in which a coarse-grained model of the chromatin fiber was used to interpret a wide range of experimental measurements[33,35–37]. In that model, the nucleosome core is represented by a rigid body decorated with optimized charges, and linker DNA is described by a twistable worm-like chain[38]. It contains flexible histone-tails[38], the linker histone[39,40] which mediates inter-nucleosome interactions, and takes into account interactions between nucleosomes and linker DNA. Another successful approach has instead sought to represent the nucleosome with an anisotropic potential, and the conformation of the exiting DNA by a set of angles[41–43]. This model has been used to interpret the experimentally measured force-extension curves of the chromatin fiber[44], and the intrinsic energies of different fiber configurations[45]. In the interest of computational efficiency, both of these models necessarily make assumptions about the nucleosome, namely that with the exception of the histone tails, the nucleosome can be represented as a rigid body. Many of these previous models also rely on Monte-Carlo sampling to provide insights into equilibrium or metastable structures, but the dynamic pathways that connect them have not been investigated.

An alternative strategy is to rely on detailed molecular models of the nucleosome, which have recently emerged as a useful tool to investigate the dynamics and energetics of the histones and their wrapped DNA[46–49]. By employing detailed coarse-grained models of DNA[50] and proteins[51,52], these models make fewer assumptions about the structure of the nucleosome and do not rely on crystal structures to bias sampling. These models have been used to examine a wide range of phenomena associated with the nucleosome, such as unwrapping[47,48], sequence-dependent formation energies[46], and nucleosome repositioning[49]. They are particularly attractive for the study of higher-order structures of chromatin, as they naturally incorporate interactions both within a single nucleosome, and between different nucleosomes. Perhaps most importantly, they offer high resolution and can naturally incorporate histone modifications with amino-acid-level detail[53], thereby providing the means to examine the wide range of histone modifications known to impact chromatin function[3]. In the past, however, computational limitations have restricted the use of these models to small systems consisting of several nucleosomes[53]. A computational approach that can simultaneously capture the physical processes that occur at nanometer scales, while simultaneously permitting access to the longer length and time scales characteristic of 3-d genomic structure could, in principle, unravel the complex relationship between histone modifications and 3-d genomic structure.

In this work, a model of chromatin is presented that relies on a multi-scale approach to systematically incorporate different length scales of chromatin. This model, the "1 **c**ylinder **p**er **n**ucleosome", or 1CPN, represents each nucleosome by a single anisotropic site, and is sufficiently efficient to simulate kilobases of DNA and hundreds of nucleosomes. Notably, the interaction parameters in 1CPN have been rigorously determined from detailed free-energy measurements obtained with the 3SPN-AICG model of the nucleosome[46,47], thereby permitting description of subtle features such as histone modifications and the role of DNA sequence. The 1CPN model is demonstrated to accurately reproduce a wide range of processes within chromatin, including inter-nucleosome free energies of interaction (pair-potentials), the free energies of nucleosome unwrapping, and the sedimentation coefficients of short chromatin fibers. Additionally, the 1CPN model provides a description of the dynamics of chromatin, and can be employed to examine the dynamic folding and rearrangements that are thought to occur within the chromatin fiber.

The remainder of this manuscript is structured as follows. The details of the 1CPN model topology and force field are presented in Section II. Section III discusses the methods used to validate the 1CPN model, including the mapping between 1CPN and the 3SPN-AICG model, the free energy techniques used, and our relative entropy coarse-graining approach. Finally, in Section IV, the results of the 1CPN model are presented along with comparisons to available experimental measurements. Lastly, we note that the 1CPN model is implemented in the LAMMPS simulation package[54] and is freely available at our website.

## II. MODEL

This section provides a detailed description of the 1CPN model topology, the model force field, and the equations of motion governing its dynamics. Before entering into these details, however, a brief outline describes the key features of the 1CPN model and the physics that the 1CPN model seeks to incorporate. Additional details are presented later in the manuscript.

1. **Rigorous validation and mapping between length scales**. Chromatin is a complex hierarchical material, and no single model can be used to

examine all length-scales of chromatin biophysics. For this reason, multi-scale approaches are particularly helpful. The 1CPN model was designed with a topology that can be validated using the detailed 3SPN-AICG model. Further, the 1CPN topology can smoothly map between 3SPN-AICG configurations (and vice versa), thereby allowing for multi-stage simulations where part of the simulation is run at one length scale (and model) of interest, and then continued using another.

2. **Inter-nucleosome Dynamics.** Chromatin is an inherently dynamic structure, where regions of the genome of constantly being expanded or compressed depending on cellular state, or the levels of protein expression. The 1CPN model was parameterized to quantitatively match the the dynamics of chromatin and is well suited to examine the various pathways by which chromatin can rearrange.

3. **Intra-nucleosome Dynamics and nucleosome flexibility.** The dynamics of chromatin go beyond the folding of the chromatin fiber, and extend to the motion of nucleosomes themselves. These dynamics can manifest themselves through fluctuations of nucleosomal DNA[55,56], the incorporation of loops and twist[57,58], repositioning of the nucleosomes themselves[59], or dynamic modification of histone tails[3]. The 1CPN model includes many of these effects, and moves beyond a view where the nucleosome is completely rigid, to one where intra-nucleosome dynamics are considered.

4. **Sequence Dependence.** DNA sequence plays an important role in the position and stability of nucleosomes, and very little work has been performed examining the effect of DNA sequence on the three-dimensional structure of chromatin. The 1CPN model includes a "first-order" approach to this sequence dependence, and can be used to explore the effect of different DNA sequences on the resulting 3-d structure of chromatin.

5. **Computational Efficiency.** In order to produce a model that can access the large length- and time-scales of chromatin assembly, 1CPN was designed with performance in mind. When seeking to incorporate the biophysical phenomena described above, each feature of the model was implemented so as to incur the smallest possible computational expense. The 1CPN model is implemented in LAMMPS, and exhibits good parallelization performance across hundreds of distributed processors.

### A. Model Topology

All sites in 1CPN are anisotropic. Each exhibits 12 degrees of freedom; six corresponding to the site's position and velocity and six corresponding to the site's orientation and angular momentum. Though the use of anisotropic sites incurs increased computational costs, this approach dramatically reduces the total number of force sites needed to represent the chromatin fiber.

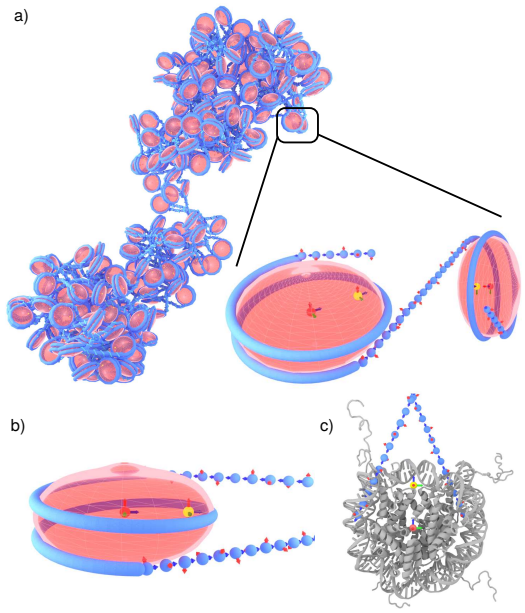


FIG. 1. Topology of the 1CPN model. Nucleosomes are represented by a single anisotropic site (red ellipsoid with blue tube). DNA is represented as a twistable worm like chain (blue spheres), and the dyad is included within the excluded volume of the nucleosome in order to stabilize the entering/exiting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

The orientation of the $i$th site is given uniquely by the quaternion $\boldsymbol{q}_i = (q_{i,0}, q_{i,1}, q_{i,2}, q_{i,3})$, which corresponds to the rotation quaternion that converts from the body-frame to the lab-frame. When discussing the model in the remainder of this section, the orientation of a site is more easily represented by an orthonormal basis $(\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$. This basis is obtained by first defining a reference basis corresponding to the body-frame $(\hat{\boldsymbol{f}}_0, \hat{\boldsymbol{v}}_0, \hat{\boldsymbol{u}}_0)$, and then using the quaternion rotations

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

to obtain the orientation of site $i$ in the lab-frame $(\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{v}}_i, \hat{\boldsymbol{u}}_i)$, where $\boldsymbol{q}^*$ is the conjugate of $\boldsymbol{q}$, and the reference basis can be chosen arbitrarily. We choose $\hat{\boldsymbol{f}}_0 = (1,0,0)$, $\hat{\boldsymbol{v}}_0 = (0,1,0)$, $\hat{\boldsymbol{u}}_0 = (0,0,1)$.

The molecular representation of the 1CPN model is shown in Figure 1. In the 1CPN model, three different site types are used that correspond to the nucleosomes, the DNA, and the nucleosomal dyad. Nucleosome sites

are represented by a red ellipsoid surrounded by a blue cylinder and are centered on a red sphere. DNA sites are represented by blue spheres, and the dyad sites are represented by yellow spheres within the red ellipsoid. A discretization of one DNA site per three base pairs is chosen and will be discussed in subsequent sections. The orientations of all sites are represented using a local coordinate frame centered at the site's location. The red, green and blue arrows correspond to the $\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{v}}_i, \hat{\boldsymbol{u}}_i$ vectors of the $i$th site, respectively. Note that for the DNA, $\hat{\boldsymbol{v}}$ is omitted for clarity and can be obtained by $\hat{\boldsymbol{v}} = \hat{\boldsymbol{u}} \times \hat{\boldsymbol{f}}$.

A particularly illustrative visualization of the 1CPN model is provided by overlaying it on top of the 1KX5 nucleosome crystal structure[60] (Figure 1c). From this representation, it can be seen that the geometrical position of the DNA sites is well within the nucleosome core particle. This topology was chosen to permit the partial unwrapping of DNA from the histone core, and will be discussed in detail in Section II D. This visualization also makes clear that the dyad site (yellow) is located approximately at the histone dyad. The dyad site is used to incorporate the effect of the H3 tails on the DNA strands entering and exiting the nucleosome, and will be discussed in Section II B 2.

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

### B. Non-bonded Interactions

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

$$U_{\mathrm{nb}} = \sum_{i}^{N} \sum_{j \neq i}^{N} U_{ij}. \tag{1}$$

The interaction potential between two sites, $U_{ij}$, depends on the site types $i$ and $j$. It is given by

$$U_{ij} = \begin{cases} U_{\mathrm{Zewdie}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j\,;\, \sigma_0, \epsilon_0) & \text{if } i \in \text{Nucl.}, j \in \text{Nucl.} \\ U_{\mathrm{Zewdie-LJ}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_i;\, \sigma_0', \epsilon_0') & \text{if } i \in \text{Nucl.}, j \in \text{DNA} \\ U_{\mathrm{Zewdie-LJ}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_j;\, \sigma_0', \epsilon_0') & \text{if } i \in \text{DNA}, j \in \text{Nucl.} \\ U_{\mathrm{elec}}(\boldsymbol{r}_{ij}) & \text{if } i \in \text{DNA}, j \in \text{DNA} \\ U_{\mathrm{aniso,gauss}}(\boldsymbol{r}_{ij}; \hat{\boldsymbol{f}}_i, \hat{\boldsymbol{u}}_i) & \text{if } i \in \text{Dyad}, j \in \text{DNA} \\ U_{\mathrm{aniso,gauss}}(\boldsymbol{r}_{ji}; \hat{\boldsymbol{f}}_j, \hat{\boldsymbol{u}}_j) & \text{if } i \in \text{DNA}, j \in \text{Dyad} \\ 0 & \text{else} \end{cases} \tag{2}$$

where $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$, and $\hat{\boldsymbol{f}}_i$, $\hat{\boldsymbol{u}}_i$, $\hat{\boldsymbol{f}}_j$ and $\hat{\boldsymbol{u}}_j$ are the orientation vectors of sites $i$ and $j$, respectively. The different pair potentials, $U_{\mathrm{Zewdie}}$, $U_{\mathrm{Zewdie-LJ}}$, $U_{\mathrm{elec}}$, and $U_{\mathrm{dyad}}$ are defined in what follows.

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

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

#### 1. Zewdie Potential

Interactions between nucleosome sites are given by the Zewdie potential[61], an extension of the widely used Gay-Berne potential, that has been shown to provide a good representation of the nucleosome[42,43]. The Zewdie potential between sites $i$ and $j$ is given by

$$U_{\mathrm{Zewdie}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j\,;\, \sigma_0, \epsilon_0) = 4\epsilon \left[ \left( \frac{\sigma_0}{r_{ij} - \sigma + \sigma_0} \right)^{12} - \left( \frac{\sigma_0}{r_{ij} - \sigma + \sigma_0} \right)^{6} \right] \tag{3}$$

where $r_{ij} = |\boldsymbol{r}_{ij}|$, $\sigma_0$ and $\epsilon_0$ are parameters, and the functions $\sigma = \sigma(\hat{\boldsymbol{r}}_{\boldsymbol{ij}}, \hat{\boldsymbol{f}}_{\boldsymbol{i}}, \hat{\boldsymbol{f}}_{\boldsymbol{j}})$ and $\epsilon = \epsilon(\hat{\boldsymbol{r}}_{\boldsymbol{ij}}, \hat{\boldsymbol{f}}_{\boldsymbol{i}}, \hat{\boldsymbol{f}}_{\boldsymbol{j}})$, are responsible for the orientation dependence of the potential. Note that $\hat{\boldsymbol{r}}_{ij} = \boldsymbol{r}_{ij}/r_{ij}$. In the Zewdie potential, $\sigma$ and $\epsilon$ are represented by the first six terms of an S-function expansion[62] and are defined as

$$\sigma = \sigma_0[\sigma_{000}S_{000} + \sigma_{cc2}(S_{022} + S_{202}) + \sigma_{220}S_{220} + \sigma_{222}S_{222} + \sigma_{224}S_{224}] \tag{4}$$

$$\epsilon = \epsilon_0[\epsilon_{000}S_{000} + \epsilon_{cc2}(S_{022} + S_{202}) + \epsilon_{220}S_{220} + \epsilon_{222}S_{222} + \epsilon_{224}S_{224}]. \tag{5}$$

In this formulation, the orientation dependence of the potential is embedded in the terms $S_{000}$, $S_{022}$, $S_{220}$, $S_{222}$, and $S_{224}$, which are defined in Table I. The range and strength of the Zewdie parameter is controlled by the parameters $\sigma_0$ and $\epsilon_0$, respectively, while the anisotropic shape of the potential is given by parameters $\sigma_x$ and $\epsilon_x$ for $x = \{000, cc2, 220, 222, 224\}$. A key advantage of this S-function expansion is that these different $\epsilon_x$ and $\sigma_x$ parameters set the energy scales of different directions of approach between particles[61]. For example, the second and third terms $S_{022}$ and $S_{202}$ increase as particle orientations $(\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j)$ align with the inter-particle vector $\boldsymbol{r}_{ij}$. As a result, the value of $\epsilon_{cc2}$ controls the preference of particles to stack column-wise. Similarly, $S_{220}$ is only a function of particle orientation (e.g. $\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j$), and

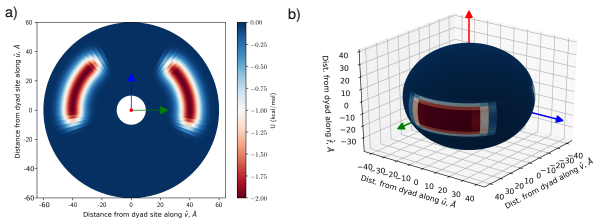


FIG. 2. Pair-wise Dyad Potential. Potential energy of $U_{\text{gauss,aniso}}$ when a) projected onto the $\hat{\boldsymbol{u}}$-$\hat{\boldsymbol{v}}$ plane, b) plotted as a 3-d surface for $r = r_0$. In both panels, the dyad site is located at the origin and its orientation $(\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}})$ given by the red, green and blue arrows, respectively.

so $\epsilon_{220}$ controls the preference of particles to form a nematic phase. Higher order effects, such as quadrupoles, are captured by the terms $S_{222}$ and $S_{224}$.

The flexibility of the expansion makes the Zewdie potential well-suited to represent a complex disk-like biomolecule such as the nucleosome. The 12 parameters in the Zewdie potential are fit to anisotropic pair-potentials obtained using the more detailed 3SPN-AICG model of the nucleosome, and will be discussed in detail in Section III C.

Interactions between DNA and nucleosome sites are also represented using the Zewdie potential. In the 1CPN model, non-bonded interactions between DNA sites are isotropic (i.e. DNA sites are spheres), and a slight modification of the Zewdie potential is required, denoted by $U_{\text{Zewdie-LJ}}$. In $U_{\text{Zewdie-LJ}}$, forces are applied the same as in $U_{\text{Zewdie}}$, but because the DNA sites interact isotropically, the potential is independent of the orientation of the DNA site. In practice, this is achieved by setting $\hat{\boldsymbol{f}}_{DNA} = \hat{\boldsymbol{r}}_{ij}$, $a_0 = a_2$, and then using $U_{\text{Zewdie}}$ as usual. Additionally, due to the different radii of the nucleosome and DNA sites, different values of $\sigma_0$ and $\epsilon_0$ must be chosen. Here we use an arithmetic mixing rule to obtain $\sigma_0' = (\sigma_{DNA} + \sigma_0)/2$. The value of $\epsilon_0$ is chosen to be small, so that nucleosome-DNA interactions are marginally attractive, and instead interact only via an excluded volume. The values of the parameters used in the Zewdie potential are given in Table S3.

Past work using the Zewdie potential[42,43,61] relied on Monte Carlo simulations that, by construction, did not require computation of forces and torques. In order to simulate the Zewdie potential using molecular dynamics (as in the 1CPN model), calculation of the forces and torques is necessary. The derivation of these quantities is provided in the Supporting Information Section S1 A.

### 2. Non-bonded Dyad Interactions

Although the Zewdie potential provides a good representation of the interactions between nucleosomes and DNA in general, it neglects additional, significant interactions that occur within chromatin. Importantly, the role of the histone H3 tail in stabilizing DNA entering and exiting the nucleosome is not taken into account. Previous experimental measurements[63] and simulations[47] indicate that histone tails H3 and H4 stabilize the outer wrap of nucleosomal DNA, and that the histone H3 tail is important for screening repulsive interactions between entering and exiting DNA[64].

In order to incorporate these effects into the 1CPN model, we introduce an additional pair-wise interaction between dyad and DNA sites. Coarse-graining these complex interactions between DNA and histone tails into a single pair potential is challenging, especially because they are highly localized near the entering/exiting DNA. In order to arrive at a suitable functional form, we drew inspiration from the base-pairing interactions in the 3SPN.2 model, where hydrogen bonds between complementary bases were given by an anisotropic Morse potential[65]. Adopting a similar approach for 1CPN, we define the dyad potential as an anisotropic gaussian potential, $U_{\text{gauss,aniso}}$, (Figure 2) that is given by

$$\begin{aligned} U_{\text{gauss,aniso}} &= f(K_\theta, \Delta\theta) f(K_\phi, \Delta\phi)\, U_{\text{gauss}} \\ &= f(K_\theta, \Delta\theta) f(K_\phi, \Delta\phi) \left(-d_0 e^{-(r-r_0)^2/2\sigma^2}\right) \end{aligned} \tag{6}$$

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

$$f(K_\theta, \Delta\theta) = \begin{cases} 1 & -\frac{\pi}{2K_\theta} < \Delta\theta < \frac{\pi}{2K_\theta} \\ 1 - \cos^2(K_\theta \Delta\theta) & \frac{-\pi}{K_\theta} < \Delta\theta < \frac{-\pi}{2K_\theta} \text{ or } \frac{\pi}{2K_\theta} < \Delta\theta < \frac{\pi}{K_\theta} \\ 0 & \Delta\theta < -\frac{\pi}{K_\theta} \text{ or } \Delta\theta > \frac{\pi}{K_\theta} \end{cases} \tag{7}$$

and similarly for $f(K_\phi, \Delta\phi)$, where $\Delta\theta = \theta - \theta_0$, $\Delta\phi = \phi - \phi_0$, and $\theta_0$ and $\phi_0$ are reference angles. Here $\cos\theta = \hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{r}}$ is the polar angle between the normalized bond vector $\hat{\boldsymbol{r}}$ and the dyad orientation vector $\hat{\boldsymbol{u}}$, and $\cos\phi = \hat{\boldsymbol{f}} \cdot \hat{\boldsymbol{r}}$ is the azimuthal angle between $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{f}}$. Although the choice of the other $U_{\text{gauss,aniso}}$ parameters will be discussed in Section IV C, the approximate symmetry of the exiting and entering DNA yields $\phi_0 = 90°$. To aid in interpreting this potential, different projections of $U_{\text{gauss,aniso}}$ are given in Figure 2. It is instructive to compare the locations of the energetic minima in Figure 2a to the overlay of 1CPN and the 1KX5 crystal structure in Figure 1c. Comparison of these two figures gives a qualitative justification for the functional form of $U_{\text{gauss,aniso}}$; this potential accounts for the role of the H3 tail in stabilizing DNA that is entering and exiting the nucleosome. A quantitative comparison will be made by matching free

energies in the 1CPN and 3SPN-AICG models in Section IV C. Before proceeding, we note that the orientation of the DNA site does not enter the potential, and $U_{\text{gauss,aniso}}$ is only a function of the vector connecting the two sites $\boldsymbol{r}$, and the orientation of the dyad site.

### 3. Non-bonded DNA Interactions

Non-bonded interactions between DNA sites are represented by a Debye-Hückel screened electrostatic potential, $U_{\text{elec}}$, given by,

$$U_{\text{elec}} = \frac{q_i q_j e^{-r_{ij}/\lambda_D}}{\epsilon r_{ij}} \tag{8}$$

where $q_i$ and $q_j$ are the charges of sites $i$ and $j$, $r_{ij}$ is the inter-site separation, $\lambda_D$ is the Debye length, and $\epsilon$ is the solution dielectric constant. Each DNA site represents three base-pairs and a charge of $-3$ is given to each DNA bead. The choice of charge per DNA bead was found to give good agreement with experimental measurements of the salt-dependent persistence length of DNA (see Figure 7). The Debye length is defined as

$$\lambda_D = \sqrt{\frac{k_{\text{B}} T \epsilon_0 \epsilon(T,C)}{2 N_A e_c^2 I}} \tag{9}$$

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

$$\epsilon(T,C) = \epsilon(T) a(C), \tag{10}$$

where

$$\epsilon(T) = 249.4 - 0.778 T/K + 7.20 \times 10^{-4} (T/K)^2 \tag{11}$$

and

$$a(C) = 1.000 - 2.551 C/M + 5.151 \times 10^{-2} (C/M)^2 - 6.889 \times 10^{-3} (C/M)^3 \tag{12}$$

as in Ref[65]. The details and justification for this treatment are provided in a previous publication[65].

## C. Bonded interactions

The bonded potential, $U_b$, includes interactions between sites, including bonds, angles, and dihedrals, that maintain the topology of the model. Since all sites in 1CPN have both a position and an orientation, the bonded interactions in 1CPN include interactions that constrain both the relative position and relative orientation of bonded sites. In our discussion here, bonded interactions are presented in two parts, first within the DNA molecule, $U_{\text{DNA}}$ and, second, within the nucleosome, $U_{\text{Nucl}}$. The total bonded energy is the sum of these two contributions,

$$U_{\text{b}} = U_{\text{DNA}} + U_{\text{Nucl}}. \tag{13}$$

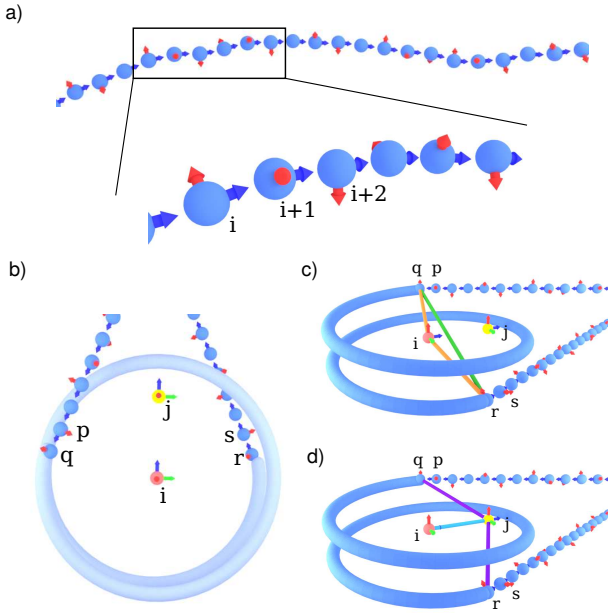


FIG. 3. Bonded Interactions in the 1CPN Model. a) DNA is represented by a twistable worm like chain as defined in Equation 14. The $\hat{\boldsymbol{f}}$ and $\hat{\boldsymbol{u}}$ vectors of each DNA bead are given by the red and blue arrows. b) Bonded interactions within the nucleosome. Indices $i, j$ denote the nucleosome and dyad sites, while $p, q, r, s$ denote the DNA sites involved in the bonded nucleosome interactions. These indices will be used throughout Section II C 2 while discussing the various aspects of the 1CPN force field. c) Bonded interactions within $U_{\text{NDNA}}$ as defined in Equation 20. d) Bonded interactions within $U_{\text{dyad}}$ as defined in Equation 25.

### 1. DNA bonded interactions

Interactions within the DNA molecule are represented by a twistable worm-like chain model[66]. For a DNA chain consisting of $N$ beads (see Figure 3a), $U_{\text{DNA}}$ is given by

$$
\begin{aligned}
U_{\mathrm{DNA}} &= U_{\mathrm{bond}} + U_{\mathrm{angle}} + U_{\mathrm{twist}} + U_{\mathrm{align}} \\
&= \frac{k_{\mathrm{b}}^{\mathrm{D}}}{2\ell_{\mathrm{b},0}} \sum_{i=1}^{N-1} (\ell_i - \ell_{\mathrm{b},0})^2 + \frac{k_\beta}{\ell_{\mathrm{b},0}} \sum_{i=1}^{N-2} \left[1 - \cos(\beta_i)\right] \\
&\quad + \frac{k_w}{\ell_{\mathrm{b},0}} \sum_{i=1}^{N-1} \left[1 - \cos(\omega_i - \omega_0)\right] \\
&\quad + \frac{k_\psi}{\ell_{\mathrm{b},0}} \sum_{i=1}^{N-1} \left[1 - \cos(\psi_i)\right], \qquad (14)
\end{aligned}
$$

where $\ell_i = |\boldsymbol{r}_{i+1} - \boldsymbol{r}_i|$ is the distance between adjacent sites, $k_{\mathrm{b}}^{\mathrm{D}}$ is the force constant, and $\ell_{\mathrm{b},0}$ is the equilibrium bond length; $\cos\beta_i = \hat{\boldsymbol{r}}_{(i)(i+1)} \cdot \hat{\boldsymbol{r}}_{(i+1)(i+2)}$ is the angle between adjacent bonds with force constant $k_\beta$; $\cos w_i = \frac{\hat{\boldsymbol{f}}_{i+1}\cdot\hat{\boldsymbol{f}}_i + \hat{\boldsymbol{v}}_{i+1}\cdot\hat{\boldsymbol{v}}_i}{(1+\hat{\boldsymbol{u}}_i\cdot\hat{\boldsymbol{u}}_{i+1})}$ is the twist around the polymer backbone between adjacent sites, $k_w$ is the corresponding force constant and $w_0$ is the equilibrium twist. The last term, $U_{\mathrm{align}}$ constrains $\hat{\boldsymbol{u}}_i$ of each site to be aligned with the bond vector, $\boldsymbol{r}_{ij}$, where $\cos\psi_i = \hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{(i)(i+1)}$, and $k_\psi$ is the corresponding force constant. In our implementation, we use a modified form of the twistable worm-like chain[67] that accounts for non-zero equilibrium twist (i.e. $\omega_0 \neq 0$). The corresponding forces and torques for this modified potential are given in the Supporting Information Section S1 C.

### 2. Nucleosome-DNA Bonded interactions

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

$$U_{\mathrm{Nucl}} = U_{\mathrm{NDNA}} + U_{\mathrm{dyad}} \qquad (15)$$

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

In the 1CPN model, many of the bonded potentials have the same functional form, and only differ in the value of the underlying parameters. As a consequence, to simplify the notation throughout the following sections, we define the harmonic bonded potentials, $U_{\mathrm{b}}^{\mathrm{h}}$, and the harmonic and cosine angle/orientation potentials, $U_{\mathrm{a}}^{\mathrm{h}}$ and $U_{\mathrm{a}}^{\mathrm{cos}}$, as

$$U_{\mathrm{b}}^{\mathrm{h}}(\ell_{ij}\,;\, k, \ell_0) = \frac{k}{2}(\ell_{ij} - \ell_0)^2 \qquad (16)$$

$$U_{\mathrm{a}}^{\mathrm{h}}(\boldsymbol{x}, \boldsymbol{y}; k, \theta_0) = \frac{k}{2}\left(\theta - \theta_0\right)^2 \qquad (17)$$

$$U_{\mathrm{a}}^{\mathrm{cos}}(\boldsymbol{x}, \boldsymbol{y}; k, \theta_0) = k\left(1 - \cos\left(\theta - \theta_0\right)\right) \qquad (18)$$

$$U_{\mathrm{d}}^{\mathrm{phi}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; k, \phi_0) = k\left(1 - \cos\left(\phi - \phi_0\right)\right) \qquad (19)$$

where $\ell_{ij} = |\boldsymbol{r}_{ij}|$, $\cos\theta = \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}$ with $\hat{\boldsymbol{x}} = \boldsymbol{x}/|\boldsymbol{x}|$ and $k$, $\ell_0$ and $\theta_0$ are parameters corresponding to the force constant, equilibrium length and equilibrium angles, respectively. To further simplify our notation, we label all sites involved in $U_{\mathrm{Nucl}}$ interactions as $i, j, p, q, r, s$ as shown in Figure 3b-d. Throughout the remainder of this section, Equations 16-18, and the indexing convention from Figure 3b-d will be referenced extensively. A detailed diagram describing the various parameters that arise in the model and the relationships between them is provided in the Supporting Information.

The nucleosomal DNA contribution can be subdivided into four terms,

$$U_{\mathrm{NDNA}} = U_{\mathrm{NDNA}}^{\mathrm{geom}} + U_{\mathrm{NDNA}}^{\mathrm{orient}} + U_{\mathrm{NDNA}}^{\mathrm{rotation}} + U_{\mathrm{NDNA}}^{\mathrm{twist}}. \qquad (20)$$

The first term, $U_{\mathrm{NDNA}}^{\mathrm{geom}}$, ensures that the locations of the DNA leaving the nucleosome are consistent with the locations in the 1KX5 crystal structure. This is shown graphically by the orange and green bonds in Figure 3c and is defined as

$$
\begin{aligned}
U_{\mathrm{NDNA}}^{\mathrm{geom}} &= U_{\mathrm{b}}^{\mathrm{h}}(\ell_{iq}; k_{\mathrm{b}}^{\mathrm{N}}, \ell_{0,c}) \\
&\quad + U_{\mathrm{b}}^{\mathrm{h}}(\ell_{ir}; k_{\mathrm{b}}^{\mathrm{N}}, \ell_{0,c}) + U_{\mathrm{b}}^{\mathrm{h}}(\ell_{qr}; k_{\mathrm{b}}^{\mathrm{N}}, \ell_{0,e}) \\
&\quad + U_{\mathrm{a}}^{\mathrm{h}}(\boldsymbol{r}_{qi}, \hat{\boldsymbol{f}}_i; k_\zeta, \zeta) + U_{\mathrm{a}}^{\mathrm{h}}(\boldsymbol{r}_{ir}, \hat{\boldsymbol{f}}_i; k_\zeta, \zeta). \qquad (21)
\end{aligned}
$$

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

$$
\begin{aligned}
U_{\mathrm{NDNA}}^{\mathrm{orient}} &= U_a^h(\boldsymbol{r}_{ij}, \boldsymbol{r}_{jk}; k_\alpha, \alpha) + U_a^h(\boldsymbol{r}_{km}, \boldsymbol{r}_{mn}; k_\alpha, \alpha) \\
&\quad + U_a^h(\hat{\boldsymbol{u}}_j, \hat{\boldsymbol{u}}_k; k_\eta, \eta) + U_a^h(\hat{\boldsymbol{u}}_m, \hat{\boldsymbol{u}}_k; k_\mu, \mu). \qquad (22)
\end{aligned}
$$

The parameters in $U_{\mathrm{NDNA}}^{\mathrm{orient}}$ are determined by comparison to 3SPN-AICG data as described in Section IV C. The third term, $U_{\mathrm{NDNA}}^{\mathrm{rotation}}$ incorporates the ability of nucleosomal DNA to rotate around the histone proteins,

$$U_{\mathrm{NDNA}}^{\mathrm{rotation}} = U_a^{\cos}(\hat{\boldsymbol{f}}_q, \hat{\boldsymbol{f}}_i; k_\theta, \theta) + U_a^{\cos}(\hat{\boldsymbol{f}}_r, \hat{\boldsymbol{f}}_i; k_\theta, \theta), \quad (23)$$

where the parameters $k_\theta$ and $\theta$ are obtained in Section IV D. Lastly, the fourth term $U_{\mathrm{NDNA}}^{\mathrm{twist}}$, corresponds to the ability of DNA twist to be stored *inside* of the nucleosome. The storage of twist within the nucleosome is a common feature of nucleosomal crystal structures[68] and is a hypothesized mechanism by which nucleosomes reposition along a long stretch of DNA. This term is defined as

$$U_{\mathrm{NDNA}}^{\mathrm{twist}} = U_a^h(\hat{\boldsymbol{f}}_q, \hat{\boldsymbol{f}}_r; k_{\mathrm{t}}^{\mathrm{N}}, 0). \qquad (24)$$

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

$$U_{\mathrm{dyad}} = U_{\mathrm{dyad}}^{\mathrm{geom}} + U_{\mathrm{dyad}}^{\mathrm{orient}}. \qquad (25)$$

A graphical representation of the interactions that contribute to $U_{\mathrm{dyad}}$ is shown in Figure 3d. The first term, $U_{\mathrm{dyad}}^{\mathrm{geom}}$, constrains the position of the dyad relative to the

nucleosome and the entering/exiting DNA sites, and is given by

$$U_{\rm dyad}^{\rm geom} = U_{\rm b}^{\rm h}(\ell_{ji};\ell_{0,d}) + U_{\rm b}^{\rm h}(\ell_{jq};\ell_{0,e}) + U_{\rm b}^{\rm h}(\ell_{jr};\ell_{0,e}). \quad (26)$$

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

$$\begin{aligned} U_{\rm dyad}^{\rm orient} = {} & U_a^h(\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j; k_\psi, 90^\circ) + U_a^h(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_j; k_\psi, 90^\circ) \\ & + U_a^h(\boldsymbol{r}_{ij}, \hat{\boldsymbol{f}}_i; k_\psi, 90^\circ) + U_a^h(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j; k_\psi, 0^\circ) \\ & + U_a^h(\boldsymbol{r}_{ij}, \hat{\boldsymbol{u}}_j; k_\psi, 0^\circ) + U_a^h(\boldsymbol{r}_{ij}, \hat{\boldsymbol{u}}_i; k_\psi, 0^\circ), \end{aligned} \quad (27)$$

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

### D. Nucleosomal DNA

An additional parameter in the 1CPN model is the number of DNA base pairs that are coarse-grained into the nucleosome site. In previous models of chromatin, either all 147 base pairs contained within the nucleosome were mapped onto a single nucleosomal site[42,43], or assumed to be a rigid body along with the core histone proteins[38]. The effect of these assumptions is that nucleosomal DNA in these previous models could not unwrap from the histone proteins. However, numerous experimental studies suggest that nucleosomal DNA does indeed unwrap from the histone proteins[55,56], and recent computational work has calculated the free energies of these DNA openings near the ends of the nucleosome to only be several $k_{\rm B}T$. Partial DNA unwrapping significantly changes the angle that DNA leaves the nucleosome and could significantly alter the three dimensional packing of short sections of chromatin.

In order to incorporate the effects of partial DNA unwrapping, the 1CPN model only coarse grains ≈127 base pairs of DNA into the nucleosome, thereby allowing ≈10 base pairs at each end of the nucleosome to unwrap from the histone proteins (see Figure 1c). This choice was found to provide good agreement with free energies of nucleosome unwrapping obtained from the 3SPN-AICG model (see Section IV C). The number of base pairs that can unwrap from the nucleosome in 1CPN is given by parameter, $n_{\rm bp,unwrap}$.

During development of 1CPN, we found it convenient to allow $n_{\rm bp,unwrap}$ to equal 9, 10, or 11 base pairs, depending on the nucleosome repeat length. This choice permits the 1CPN model to examine any nucleosome repeat length, not only values that are even multiples of the DNA discretization of three DNA base pairs per DNA bead. Details about this choice are given in the Supporting Information Section S4.

### E. Brownian Dynamics

Brownian dynamics simulations are used to describe the translational motion of a particle according to

$$m_i d\boldsymbol{v}_i(t) = -\boldsymbol{\zeta}_i \cdot \boldsymbol{v}_i(t)dt + \boldsymbol{B}_i \cdot d\boldsymbol{W}_i(t), \quad (28)$$

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

Similarly, the rotational motion of a particle in solution can be written as[69],

$$[\boldsymbol{R}_i(t)]^{-1}\cdot\boldsymbol{I}_i\cdot d\boldsymbol{\omega}_i(t) = -\boldsymbol{\zeta}_i^{\rm r}\cdot\boldsymbol{\omega}_i(t)dt + \boldsymbol{B}_i^{\rm r}\cdot d\boldsymbol{W}_i(t), \quad (29)$$

$$\frac{d\hat{\boldsymbol{f}}_i(t)}{dt} = \boldsymbol{\omega}_i(t)\times\hat{\boldsymbol{f}}_i(t), \quad (30)$$

$$\frac{d\hat{\boldsymbol{u}}_i(t)}{dt} = \boldsymbol{\omega}_i(t)\times\hat{\boldsymbol{u}}_i(t), \quad (31)$$

$$\frac{d\hat{\boldsymbol{v}}_i(t)}{dt} = \boldsymbol{\omega}_i(t)\times\hat{\boldsymbol{v}}_i(t), \quad (32)$$

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

Here we use freely-draining Brownian dynamics and the friction tensor is assumed to be diagonal, such that $\boldsymbol{\zeta}_i = \frac{m_i}{\lambda_i}\boldsymbol{\delta}$, and $\boldsymbol{\zeta}_i^r = \frac{m_i}{\lambda_i^r}\boldsymbol{\delta}$ where $\lambda_i$ and $\lambda_i^r$ are the so-called scalar translational and rotational damping parameters of the $i$th site.

Since the nucleosome sites in 1CPN have an ellipsoidal shape, the frictional forces that they experience should in principle be anisotropic (i.e. tensorial). In the 1CPN model, we have simplified the dynamics of nucleosome sites by using an effective scalar friction coefficient that was chosen such that the translational diffusion coefficient of the nucleosome particles matches the value obtained with the full tensorial representation (see Section S2 B of the Supporting Information). DNA sites in the 1CPN model are spherical, and their friction coefficient is scalar by construction. Since the mass coarse-grained into the nucleosome sites is approximately 100x larger than that coarse-grained into the DNA sites, the 1CPN model contains different scalar friction coefficients for the different site types. In 1CPN nucleosome sites are assigned a damping time of $\lambda_i = 5.1$ ps, spherical (DNA and dyad) sites are assigned a damping time $\lambda_i = 0.195$ ps, and $\lambda^r = 0.3\lambda_i$.

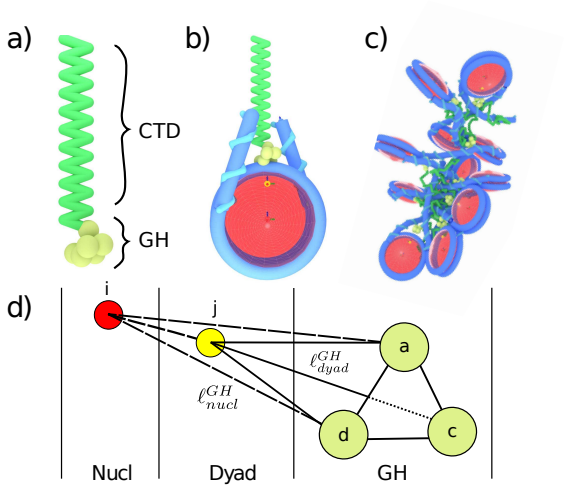


FIG. 4. The Linker Histone in the 1CPN Model. A representation of the a) independent linker histone, b) a representation of the chromatosome particle consisting of the linker histone binding at the dyad axis and a 167 base pair nucleosome core particle, c) a representation of a condensed 12 nucleosome fiber with linker histones included, and d) a schematic representation of the bonds between the nucleosome and dyad sites with the globular head.

### F. Linker Histone

The structure of the chromatin fiber is influenced by DNA-binding proteins, with the H1 or avian variant H5 linker histone being ubiquitous examples[21,70–74]. Sedimentation analyses of small chromatin fibers have demonstrated that chromatin with linker histones produces more condensed fibers than those without[35,72,75–77]. The density of positively-charged residues in the linker histone reduces the electrostatic repulsion and promotes binding of the linker DNA, favoring a more compact structure. Given its direct influence on chromatin structure, a model of the H1/5 linker histone by Luque *et al.*[40] has been adapted to fit the 1CPN model.

The 1CPN linker histone model is at a coarse-grained level of description comparable to that of 1CPN and draws information from atomistic structures and charge distributions. Structurally, it consists of three domains: the central, rigid globular head (GH) domain consisting of $\sim$ 80 amino acid residues, the flexible, highly positively-charged C-terminal domain consisting of $\sim$ 110 residues, and the short N-terminal domain of $\sim$ 25 residues.[78] In its original model definition, the GH region was coarse-grained to be a fixed rigid-body consisting of 6 charged sites. The positions of these sites relative to the nucleosome center of mass are described in Table-S9.

#### 1. Bonded Interactions

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

$$\begin{aligned} U^{\text{geom}}_{\text{GH}} &= U^h_b(\{\boldsymbol{r_{GH}}\}; k_{b,GH}, \{\boldsymbol{\ell_{GH}}\}) \\ &+ U^h_a(\{\boldsymbol{r_{GH}}\}; k_{a,GH}, \{\boldsymbol{\theta_{0,GH}}\}) \\ &+ U^{\text{phi}}_{\text{d}}(\{\boldsymbol{r_{GH}}\}; k_{\phi,GH}, \{\boldsymbol{\phi_{0,GH}}\}) \end{aligned} \tag{33}$$

where $\{\boldsymbol{r_{GH}}\}$ is the set of all coordinates of the rigid GH sites, $\{\boldsymbol{\ell_{GH}}\}$ is the set of lengths between the positions of the rigid sites, and $\{\boldsymbol{\theta_{0,GH}}\}$ and $\{\boldsymbol{\phi_{0,GH}}\}$ are the sets of angles and dihedrals of the rigid structure, respectively. In order to ensure that the GH stays rigid, but not too constrained, a subset of the possible bonds, angles, and dihedral restraints are placed on the system and shown in Table S9.

Although there is discussion regarding the binding location of the linker histone to the nucleosome, we choose to bind the linker histone GH on the dyad axis.[71,79] To accomplish this, we employ two harmonic spring potentials to three sites of the GH. We define the potential, $U^{\text{bind}}_{\text{GH}}$ as:

$$\begin{aligned} U^{\text{bind}}_{\text{GH}} &= \sum_{k=a,c,d} U^h_b(\ell_{jk}; k^{GH}_{dyad}, \ell^{GH}_{dyad}) \\ &+ U^h_b(\ell_{ik}; k^{GH}_{nucl}, \ell^{GH}_{nucl}) \end{aligned} \tag{34}$$

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

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

$$\begin{aligned} U^{\text{geom}}_{\text{CTD}} &= \sum_{i,i+1,i+2}^{N_{\text{CTD}}} U^h_a(\boldsymbol{r}_{i,i+1}, \boldsymbol{r}_{i+1,i+2}; k^{\text{LH}}_{\beta}, \beta^{LH}) \\ &+ \sum_{i,i+1}^{N_{\text{CTD}}} U^h_b(\ell_{i,i+1}; k^{CTD}_b, \ell^{CTD}_b) \end{aligned} \tag{35}$$

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

$$U_{\rm LH}^{\rm bind} = U_b^h(\ell_{LH}; k_b^{CTD}, \ell_b^{CTD}) \tag{36}$$

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

#### 2. Nonbonded Interactions

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

$$U_{\rm LJ}(\boldsymbol{r}_{ij}, \sigma_{ev}, \epsilon_{ev}) = 4\epsilon \left[ \left( \frac{\sigma_0}{r_{ij}} \right)^{12} - \left( \frac{\sigma_0}{r_{ij}} \right)^{6} \right] \tag{37}$$

For all spherical beads, a geometric average of $\sigma_0$ is applied based on the sizes of each species. The energy of interaction ,$\epsilon_0$, is set low to make the interaction purely repulsive. Consistent with the DNA sites, linker histone electrostatics are treated at the level of Debye-Hückel, as described in Section II B 3. The electrostatic interactions are only used between linker histone and DNA. Interactions between the linker histone and the nucleosome are treated similarly as in the DNA with the $U_{\rm Zewdie-LJ}$.

### G. Parameter Values

The parameter values used in the 1CPN model are tabulated in the Supporting Information. These tables also include references to how (or where) each parameter value was obtained. These references consist of both experimental measurements (when available) and free energy comparisons made in this work between the 1CPN and 3SPN-AICG models. Despite the relatively large number of parameters in the 1CPN model, nearly all of them come from experiments or 3SPN-AICG free energy calculations.

### H. Code Availability

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

## III. METHODS

### A. 3SPN-AICG Nucleosome Model

Although this manuscript is primarily concerned with the new 1CPN model, many of the parameters in 1CPN were obtained by mapping 1CPN results to those obtained using the 3SPN-AICG model. For this reason, we provide an overview of the 3SPN-AICG model here, and refer interested readers to the detailed descriptions found in prior publications[46,47,49]. In the 3SPN-AICG model, DNA is represented by the 3SPN.2C model of DNA, the latest version of which[65,80,81] has been extended to incorporates the effects of DNA sequence on both the shape, local flexibility, and global flexibility of the DNA molecule[50]. In the 3SPN.2C model, each base pair is represented by three sites, located at the phosphate, sugar, and the base. This meso-scale description of DNA has proven useful in a wide range of studies, ranging from DNA compaction into a viral capsid[82] to DNA-conjugated nanoparticles[83,84]. The histone proteins are represented by the Atomic Interaction-based Coarse Grained (AICG) model, a Go-based protein model with sequence-specific interactions parameterized from all-atom energies obtained from AMBER[51]. In AICG, each amino acid is represented by a single site, located at the center of mass of the amino acid sidechain. We note that the AICG and 3SPN.2C models were developed independently to be accurate models of proteins and DNA, and consequently, no information about the nucleosome is encoded directly into the model.

Interactions between the 3SPN and AICG models are represented only through electrostatic interactions between the negatively charged phosphate DNA sites and the charged amino acids on the histone protein. As a result, the configurations sampled in the 3SPN-AICG arise naturally as a balance between the elastic energies with the DNA molecule and the electrostatic interactions between the DNA and histones. This has permitted the 3SPN-AICG model to quantitatively match experimental measurements of the tension-dependent unwrapping of the nucleosome without any additional fit parameters[47], and to examine the role sequence-dependent repositioning[49]. In 3SPN-AICG, electrostatic interactions are treated at the level of Debye-Hückel theory, and the solvent implicitly by a Langevin thermostat. All calculations using 3SPN-AICG were performed at 300K at 150mM monovalent salt unless otherwise noted.

### B. Mapping between 3SPN-AICG and 1CPN Models

In order to match the nucleosome-nucleosome interactions energies in the 1CPN and 3SPN-AICG nucleosome models, it was necessary to develop a mapping function between them. In general, such a high reduction in dimensionality is accompanied by a high loss of information. We found, however, that the semi-rigid structure of the histone core enabled a straightforward mapping from the 3SPN-AICG model to the 1CPN model. The position of the 1CPN nucleosome site, $\boldsymbol{r}$, was chosen as the center of mass of the 3SPN-AICG nucleosome. The orientation vectors of the nucleosome site $\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{u}}$ were obtained by defining geometric relationships within the nucleosome, such that $\hat{\boldsymbol{u}}$ points from the center of mass of the histone to the dyad, $\hat{\boldsymbol{f}}$ points along the nucleosomal DNA

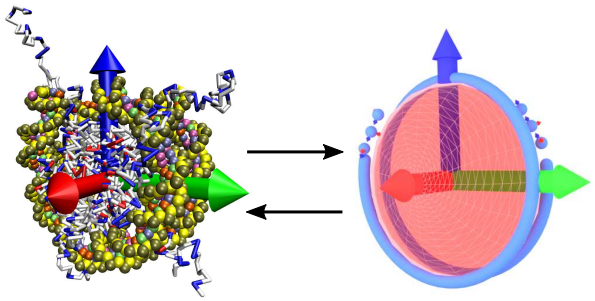

FIG. 5. Mapping between detailed 3SPN-AICG model and coarse-grained 1CPN model.

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

TABLE II. Definition of Nucleosome-Nucleosome Orientations for Pair-Potential Calculations

superhelix, and $\hat{\boldsymbol{u}} \times \hat{\boldsymbol{f}} = \hat{\boldsymbol{v}}$.

Once the position and orientation of the nucleosome site are obtained, the dyad site is positioned $\ell_{d,0}$ along $\hat{\boldsymbol{u}}$. DNA sites in 1CPN are positioned along the helical axis of 3SPN DNA corresponding to the central base pair of a three base pair segment. The orientation of these DNA sites are chosen such that $\hat{\boldsymbol{f}}$ points toward the 3SPN minor groove, and $\hat{\boldsymbol{u}}$ points along the helical axis. Additional details of this mapping procedure are described in Section S5 of the Supporting Information.

### C. Nucleosome-Nucleosome Pair Potentials

After mapping the 3SPN-AICG nucleosome model to a position and orientation, the relative orientation of the two nucleosomes $i$ and $j$ is given by six angles that represent all the possible combinations of angles between $\hat{\boldsymbol{f}}_i, \hat{\boldsymbol{f}}_j, \hat{\boldsymbol{r}}_{ij}$ and $\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}$. In the 1CPN model, our choice of the Zewdie potential assumes uniaxial symmetry about $\hat{\boldsymbol{f}}_i$ and $\hat{\boldsymbol{f}}_i$, and therefore the relative orientation of two nucleosomes only depends on the three vectors $\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}$. Accordingly, we avoid configurations where $\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij} \neq 0$ or $\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij} \neq 0$ which are prevented in chromatin fibers by the entering/exiting nucleosomal DNA. The precise definitions of the 3SPN-AICG orientations used in this work are listed in Table II. Note that Orientations C and D gave nearly identical free energies, and therefore Orientation D is omitted from Figure 6 for clarity.

To compute the effective pair-potential, the 3SPN-AICG nucleosomes were constrained to these orientations with strong harmonic angle potentials, and umbrella sampling[85] was performed along the center-of-mass separation, $r$, between the two nucleosomes. The parameters in the 1CPN model's Zewdie potential (see Table S3) were chosen to minimize the total error with the 3SPN-AICG effective pair potentials over all four orientations.

### D. DNA Exit Angle from Nucleosome

As discussed in Section II D, the 1CPN model permits the angle at which DNA exits the nucleosome to change dynamically as DNA unwraps from the histone surface. The energy scales of these interactions in 1CPN were determined by matching the free energies of DNA unwrapping obtained with 1CPN to those obtained using the 3SPN-AICG model (see Section IV C). These free energies were calculated for both 1CPN and 3SPN-AICG by first generating a configuration consisting of a single nucleosome flanked by 38 extra base-pairs on each side (Figure 8a). The DNA end-to-end extension was used as an order parameter to sample the unwrapping of the nucleosome, and the free energy as a function of DNA extension is then determined using Umbrella Sampling and WHAM[85,86]. Note that this procedure is identical to that employed in our previous work with the 3SPN-AICG model in the absence of applied tension to the DNA ends[47].

### E. Nucleosomal DNA Rotation

In order to calculate the free energies of DNA rotation around the nucleosome in the 3SPN-AICG model, we first define an order parameter, $S_{\mathrm{R}}$, that quantifies the rotational position of the nucleosomal DNA. This order parameter has been used in previous work with the nucleosome[46,49] and is described again here. The definition of $S_{\mathrm{R}}$ is given by

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

where $\mathbf{B}$ is a vector from the center of a given base step on the sense strand to its complementary base step on the anti-sense strand, $\mathbf{P}$ is a vector from the center of a base step to the center of the protein, and the angle brackets denote an average over base steps at the -15, -5, +5 and +15 positions relative to the dyad. The positive sign is chosen if $(\boldsymbol{P}\times\boldsymbol{B})\cdot\boldsymbol{D} \leq 0$ (negative if $> 0$), where $\boldsymbol{D}$ is a vector in the 5' to 3' direction along the sense strand. Notably, when $S_{\mathrm{R}} = -\pi/2$, the minor groove is oriented toward the protein core, and when $S_{\mathrm{R}} = \pi/2$, it is oriented away from it. The free energy along $S_{\mathrm{R}}$ was obtained using Metadynamics, an advanced sampling technique where a molecular dynamics trajectory is continuously biased toward unexplored regions of phase-space, thereby ensuring that all of phase space is sampled[87].

DNA rotation is incorporated into the 1CPN model through the choice of parameters in the $U^{\text{rotation}}_{\text{NDNA}}$ potential described in Equation 23. Since the parameter $S_{\text{R}}$ in the 3SPN-AICG model corresponds to $\theta$ in the 1CPN model, the value of $k_\theta$ was obtained by fitting Equation 23 to the 3SPN-AICG rotation data presented in Figure 9. Note that $k_\theta$ listed in Table S6 is half that obtained from the fits in Figure 9, because $U^{\text{rotation}}_{\text{NDNA}}$ is applied to both the entering and exiting DNA sites in 1CPN.

### F. DNA Persistence Length

The DNA molecule is characterized by both a bend persistence length, $\ell_p$, and a twist persistence length, $\ell_t$. Both $\ell_p$ and $\ell_t$ were obtained from simulations of linear chains of 80 DNA beads, corresponding to 240 base pairs. Error bars were obtained from 10 independent simulations. The bend persistence length is extracted from the auto-correlation of $\hat{\boldsymbol{u}}$ vectors along a DNA strand,

$$\langle \hat{\boldsymbol{u}}_0 \cdot \hat{\boldsymbol{u}}_j \rangle = e^{-\ell_{b,0} j/\ell_p}, \tag{38}$$

where $\hat{\boldsymbol{u}}_0$ and $\hat{\boldsymbol{u}}_j$ are the orientations of two DNA sites separated by $j$ beads that point along the polymer backbone, and $\ell_{b,0}$ is the equilibrium bond length given in Eq. 14 (c.f. Figure 3a). The twist persistence length is calculated analogously from the auto-correlation function of the vectors $\hat{\boldsymbol{f}}$ that are perpendicular to the polymer backbone,

$$\langle \hat{\boldsymbol{f}}_0 \cdot \hat{\boldsymbol{f}}_j \rangle = e^{-\ell_{b,0} j/\ell_t}. \tag{39}$$

We note that since our model of DNA contains a nonzero equilibrium twist (i.e. $\omega_0 \neq 0$ in Eq. 14) the $\hat{\boldsymbol{f}}$ vectors rotate around the polymer backbone (c.f. Figure 3a, red arrows), and the auto-correlation of $\hat{\boldsymbol{f}}$ also contains an additional contribution due to the rotating phase of $\hat{\boldsymbol{f}}$. Though this contribution can be resolved explicitly by rotating a bead's reference frame when calculating $\langle \hat{\boldsymbol{f}}_0 \cdot \hat{\boldsymbol{f}}_j \rangle$, we instead opt for the simpler, yet equivalent approach of setting $\omega_0 = 0$ in our simulations used to determine $\ell_t$.

### G. Diffusion Constants

Diffusion constants in the 1CPN model of DNA or chromatin fibers were obtained by tracking the center of mass as a function of time, $r_{\text{cm}}(t)$, and then computing the mean-squared displacement,

$$\langle \Delta R^2(t) \rangle_{\text{eq}} = \langle [\boldsymbol{r}_{\text{cm}}(t) - \boldsymbol{r}_{\text{cm}}(0)]^2 \rangle_{\text{eq}}. \tag{40}$$

The diffusion coefficient is then obtained from the Einstein relation:

$$\mathcal{D} = \lim_{t\to\infty} \frac{\langle \Delta R^2(t) \rangle_{\text{eq}}}{6t}. \tag{41}$$

When calculating the diffusion constant of DNA and chromatin fibers (Figure 10), simulations were performed at 150mM salt, and with a nucleosome repeat length of 207 base pairs. When extracting the diffusion constant from experimental papers[88–91], the diffusivity was not always measured at these salt conditions, and extrapolation was necessary. In general, the diffusivity was approximately constant at salt concentrations $> 50$ mM, and so this extrapolation was assumed to introduce negligible errors.

### H. DNA-mediated nucleosome-nucleosome pair potential

The DNA-mediated nucleosome-nucleosome pair potential (Section IV F) was obtained using umbrella sampling[85] along an order parameter defined by the center-to-center distance, $r$, between the two nucleosome sites. At short separations, inter-nucleosomal DNA can adopt many different configurations, each of which are separated by relatively large energy barriers. As a consequence, diffusion orthogonal to the order parameter was very slow, and obtaining accurate sampling of all possible DNA configurations was slow.

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

### I. Sedimentation Coefficients

Sedimentation coefficients were obtained by first initializing a 12 nucleosome, 207 nucleosome repeat length (NRL) fiber in an extended configuration where the nucleosomes were not in contact. These fibers were then relaxed using Brownian dynamics until they condensed (typically $1 \times 10^9$ time steps, or 20 $\mu s$). This condensed configuration was then used as the initial configuration for a replica exchange simulation at the specified salt concentration with 24 temperatures spaced geometrically between 300K and 700K. Since the energy barriers between different fiber configurations are relatively large, this replica exchange simulation was necessary to accelerate the exploration of many possible fiber conformations, and improve the estimate of the sedimentation coefficient. The replica exchange simulations were typically $\approx 1\times10^8$ steps, or 2 $\mu s$.

The sedimentation coefficient was determined, as described previously[38], according to

$$S_{20,w} = S_1 \left( 1 + \frac{2R}{N'} \sum_i^{N'} \sum_{j>i}^{N'} \frac{1}{R_{ij}} \right) \tag{42}$$

where $N'$ is the number of nucleosomes in the fiber, $R$ is the effective radius of the nucleosomes and is assumed

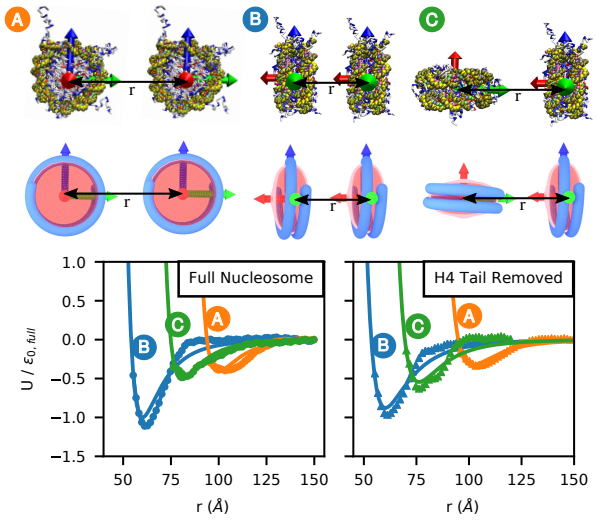


FIG. 6. Pair-Potential between two nucleosomes for several orientations for 3SPN-AICG (points) and 1CPN models (lines).

to be 54.6$\mathring{A}$. $S_1$ is the $S_{20,w}$ of a mono-nucleosome taken as equal to 11.1 Svedberg (S), and $R_{ij}$ is the distance between two nucleosomes. The values of $R$ and $S_1$ were chosen to be consistent with the values used in prior work[38]. The sedimentation coefficient was monitored throughout the replica exchange simulation (at 300K), and was determined as the average once the sedimentation coefficient converged. Error bars represent the standard deviation from four independent simulations.

## IV. RESULTS

### A. Nucleosome-Nucleosome Pair Potential

To parameterize the nucleosome-nucleosome interaction in 1CPN we first compute the anisotropic pair-potential using the 3SPN-AICG nucleosome model (Figure 6). The pair-potential is strongly dependent on the orientation of the nucleosomes, with the energy and length-scales of the interaction ranging from $U/\epsilon_0 \approx 1.0$ to 0.4 and 65$\mathring{A}$ to 100$\mathring{A}$, respectively. The lowest energy configuration corresponds to two nucleosomes stacked face-to-face (Orientation B).

Note that 3SPN-AICG can be used to explore the effects of epigenetic markers, such as the acetylation of the histone tails, on the nucleosome-nucleosome pair potential. As a proof-of-principle, we again calculate the anisotropic pair-potential using 3SPN-AICG, but modify the histone proteins so that the histone H4 tail is removed. The H4 tail is thought to mediate interactions between nucleosomes[19], and modifications to H4 represent an important epigenetic mechanism by which chromatin compaction is controlled. The removal of the H4 tail is observed to decrease the attraction of nucleosomes in Orientation B, and leads to a significant widening of the potential in Orientation B (Figure 6).

Having obtained these free-energy surfaces with the 3SPN-AICG model, we then parameterize the nucleosome-nucleosome interactions energies in 1CPN (see Methods). The resulting pair potentials obtained with 1CPN are shown by the solid lines in Figure 6, and the parameters are reported in Table S4 of the Supplementary Information. The agreement between the pair potentials from 3SPN-AICG (points) and those from 1CPN (lines) is good, with all length and energy scales matched between the two models. Notably, good agreement between the models is obtained for both the "full" nucleosome and for nucleosomes lacking the H4 tail. This result indicates that both the complex orientation-dependent interactions between nucleosomes and the effects of histone modifications on these interactions can be effectively coarse-grained into the 1CPN model's Zewdie potential.

When computing these pair potentials in 3SPN-AICG, the maximum energy of attraction, $\epsilon_{0,full}$, was found to be 7.9 kcal/mol ($\approx 13 k_B T$). This value is in good agreement with the experimental measurements of[92], who measured inter-nucleosome attraction to be $13.2 \pm 0.7 k_B T$, but differs from other experiments that estimate these energies to be $\approx 1 - 3 k_B T$[93–95]. The strength of attraction between nucleosomes continues to be an active area of research, and an extensive comparison between the predictions of the 3SPN-AICG model and experimental measurements will be the topic of a forthcoming manuscript.

In the 1CPN model presented here, we choose $\epsilon_{0,full}$ as an adjustable parameter that sets the energy scale for the nucleosome-nucleosome interactions. This choice allows the 1CPN model to retain the relative energies of the different orientations and histone modifications, yet provides the flexibility to vary the strength of attraction between nucleosomes as additional experimental or simulation become available.

### B. DNA Persistence Length

In addition to the pair-wise interactions between nucleosomes, inter-nucleosome interactions within chromatin are also mediated by the elasticity of the DNA that joins neighboring nucleosomes. To ensure that these elastic contributions are represented faithfully in our model, we calculate the bend and twist persistence lengths of the twistable worm-like chain used in 1CPN (see Methods). Using the DNA parameters from Ref.[41,43] (given in Table S5), our DNA model exhibits a bend persistence lenth, $\ell_p = 50.5 \pm 1.3$ nm and a twist persistence length, $\ell_t = 1072 \pm 164$ nm. Both values are in good agreement with experimental data[97?,98].

Since DNA is a polyanion, the persistence length depends strongly on the concentration of salt in solution.

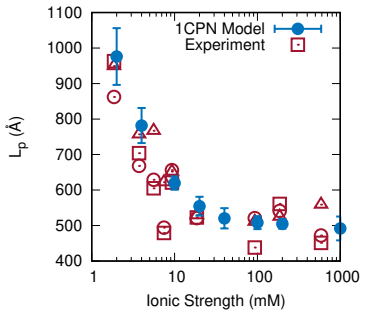


FIG. 7. Persistence Length of 1CPN DNA as a function of ionic strength. Experimental Data is from Ref[96].

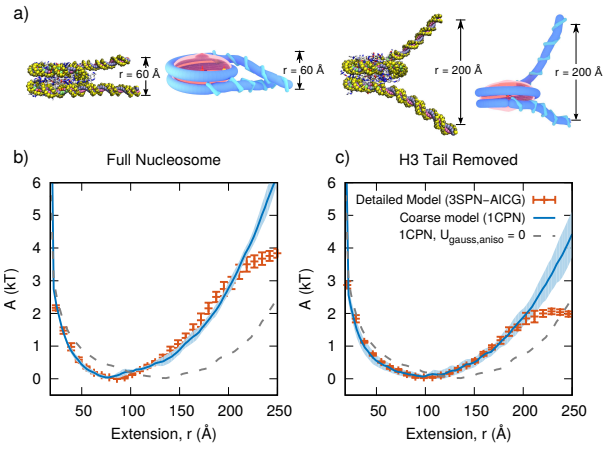


FIG. 8. Free energy of partial nucleosome unwrapping for 3SPN-AICG model (points) and 1CPN model (lines).

Experiments have shown that it can range from ≈ 50 nm at salt concentrations > 100 mM up to ≈ 1000 nm at salt concentrations of 2 mM[96]. In chromatin, this increased rigidity of DNA is thought to cause chromatin fibers to expand at low ionic strength[24,99]. Because of its impact on chromatin structure, we sought to evaluate whether this salt-dependent persistence length of DNA is reproduced in the 1CPN model.

The persistence length as a function of salt was determined using 1CPN for salt concentrations ranging from 2 to 1000 mM, and was compared to experiments by Baumann *et al.*[96] (Figure 7). The agreement between the model predictions and the experimental measurements is reasonable, and indicates that the effective stiffening of DNA with decreased salt is reproduced by 1CPN.

### C. DNA Exit Angle from Nucleosome

The angle at which DNA exits the nucleosome is an important model parameter that dictates how adjacent nucleosomes along a single DNA strand interact. Adjacent nucleosomes are typically separated by 20 to 80 base pairs, a length much shorter than the persistence length of DNA, and therefore bent linker DNA is energetically disfavored[100]. The angle at which DNA exits the nucleosome dictates the boundary conditions of this bent linker DNA, and has a significant effect on the orientations of two associating nucleosomes. Depending on the DNA exit angle, different nucleosome orientations will be either favored or frustrated by the path of the linker DNA.

In previous models of chromatin, this angle was fixed at a constant value, which is either extracted from the nucleosomal crystal structure[38,39] or is chosen as an input parameter in the model[42,43]. In both cases, the resulting configurations of the chromatin fiber are known to be very sensitive to the choice of these angles. These previous approaches also make the assumption that the DNA exit angle is fixed and that partial DNA unwrapping cannot occur.

As mentioned before, in the 1CPN model, we make no such assumptions about the DNA exit angle and instead allow the outermost 10 base pairs of DNA to unwrap from the nucleosome (see Section II D). There is a variable DNA exit angle, and no single value is encoded into the model. In order to parameterize DNA unwrapping in 1CPN, we rely on recent results using the 3SPN-AICG model where the free energies of nucleosome unwrapping were obtained[47]. These free energies, and the forces associated with them were found to be in good agreement with experimental measurements of fully intact nucleosomes[101] and nucleosomes with histone tails removed[63].

The free energies of DNA unwrapping computed using both the 1CPN and the 3SPN-AICG model are consistent with each other over a large range of extensions (Figure 8). Both the DNA-DNA repulsion at low extension ($r < 60$) and the free energetic penalty as the extension is increased to $r = 200\mathring{A}$ are well reproduced by 1CPN. Deviations between the two models only occur at large separations $r > 225$, where more than 10 base pairs of DNA unwrap from the histone surface (see Figure 8a, right). Since the 1CPN model only permits the unwrapping of $\sim 10$ base pairs of DNA, its construction prevents the model from matching the free energies of additional DNA unwrapping. Though 1CPN could be modified to incorporate these larger DNA unwrapping events, our choice of 10 unwrapping base pairs strikes a good balance between matching the free energies of DNA unwrapping and computational efficiency.

As with the nucleosome-nucleosome pair potentials discussed in Section IV A, the 1CPN model can also incorporate the effect of histone modifications on the free energies of DNA unwrapping. The most relevant histone tail to partial DNA unwrapping is the tail of histone H3, which is known to stabilize the outer wrap of nucleosomal DNA[47,63], and to screen repulsive interactions between entering and exiting DNA[64]. We calculate the free energies of DNA unwrapping with the 3SPN-AICG model with the H3 tail removed, and use these free energies to parameterize 1CPN (Figure 8c). As with the free energies

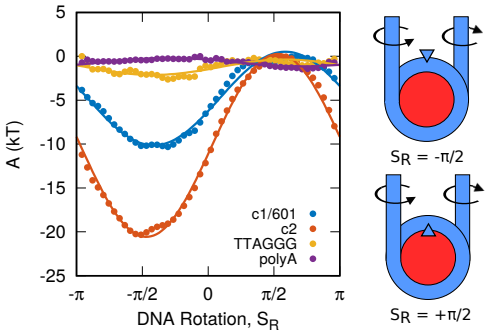


FIG. 9. Free Energy of Nucleosomal DNA rotation for 3SPN-AICG model (points) and 1CPN model (lines).

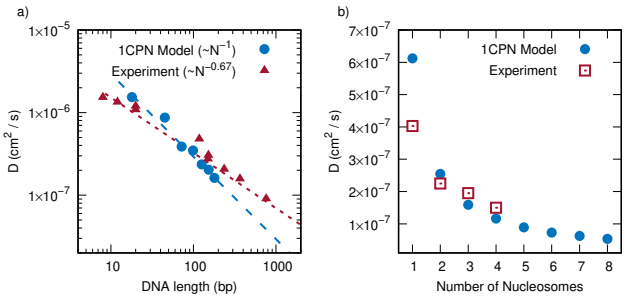


FIG. 10. Comparison 1CPN model dynamics compared to experimental measurements. (a) Diffusion Constant of DNA for different DNA lengths. Experimental data is from Refs[106–110]. (b) Diffusion Constant for chromatin fibers of different lengths. Experimental data is from Refs[88–91].

of full nucleosome, the predictions of the 1CPN model agree with the results obtained using 3SPN-AICG. This finding also underscores the importance of the anisotropic gaussian potential, $U_{\mathrm{aniso,gauss}}$, in governing the exiting configurations of DNA (as discussed in Section II B 2). Whereas inclusion of $U_{\mathrm{aniso-gauss}}$ is sufficient to reproduce the 3SPN-AICG free energies with and without the H3 tail (Figure 8bc, blue line), removal of this potential results in free energies that are too low (Figure 8bc, grey line).

### D. Sequence-Dependent Rotation of Nucleosomal DNA

It has been established that different DNA sequences bind to the histone proteins with different affinity. Different sequences exhibit mechanical properties or equilibrium shapes that make nucleosome formation more or less favorable[46,102]. Certain DNA sequences bind more strongly than others because these sequences contain periodic base-step motifs that energetically prefer to bind to the histone proteins. These periodic base-steps impart a rotational phasing of the DNA molecule, with variations in the rotational position of the histone-bound DNA incuring an energetic penalty.

Using the 3SPN-AICG model, we have quantified the free energies of DNA rotation within the nucleosome for several different DNA sequences (points, Figure 9), as described previously[46]. These sequences were chosen to range from the weakly binding PolyA sequence, to the strongly binding 601 sequence. The orientation of the DNA molecule is given by the order parameter, $S_R$, that describes which side of the DNA molecule is facing the histone core (see Methods). Notably, a value of $S_{\mathrm{R}} = -\pi/2$ corresponds to the minor groove facing the protein, whereas when $S_{\mathrm{R}} = \pi/2$ the minor groove is facing away from it. As expected, these DNA sequences exhibit strong rotational preferences. All three sequences have free energy minima at $\approx -\pi/2$, where the favorable DNA binding motifs are facing the histone protein and maxima at $\approx \pi/2$ these these motifs are facing away from the histone protein. The magnitude of the free energy penalty of this rotation is strongly dependent on sequence, ranging from $\approx 20 k_{\mathrm{B}}T$ for the c2 sequence, to $\approx 1 k_{\mathrm{B}}T$ for the PolyA sequence.

This sequence-dependent rotation is incorporated into the 1CPN model through the term $U^{\mathrm{rotation}}_{\mathrm{NDNA}}$, defined in Equation 23. $U^{\mathrm{rotation}}_{\mathrm{NDNA}}$ is a cosine angle potential (see Equation 18) that contains two parameters $\theta_0$, the equilibrium angle between the nucleosome and the entering/exiting orientation vector $\hat{\boldsymbol{f}}$, and $k_\theta$ the spring constant corresponding to rotation. By choosing different values of $k_\theta$ we can effectively incorporate sequence-dependent rotation into the 1CPN model. The comparison between 1CPN predictions (lines), and the free energies from 3SPN-AICG (points) is satisfactory (Figure 9), and demonstrates that the complex sequence-dependent rotation of nucleosomal DNA is captured in 1CPN.

The rotation of nucleosomal DNA can also manifest itself as DNA twist *stored within* the nucleosome. Many crystal structures of the nucleosome are observed to have over-twisted DNA[68,103,104], and the incorporation of twist into the nucleosome has long been hypothesized to be a mechanism by which the nucleosome can reposition[49,105]. This storage of twist within the 1CPN model is dictated by the value of $k^{\mathrm{N}}_{\mathrm{t}}$, which enters the expression for $U^{\mathrm{twist}}_{\mathrm{NDNA}}$ in Equation 24, and couples the orientation of the DNA entering and exiting the nucleosome. $k^{\mathrm{N}}_{\mathrm{t}}$ dictates the strength of coupling between the entering and exiting DNA and is inversely related to how much twist can be stored within the nucleosome. In our initial analysis with the 1CPN model we have chosen to minimize the possibility of twist storage by choosing $k^{\mathrm{N}}_{\mathrm{t}} = 10$kcal/mol/rad, corresponding to a relatively large energy penalty of $\approx 10 k_B T$ for a single twist defect. However, in cases where DNA twist can be stored in the nucleosome, a value of $k^{\mathrm{N}}_{\mathrm{t}} = 1$kcal/mol/rad is likely to be more appropriate.

### E. Chromatin Dynamics

Chromatin is intrinsically dynamic, and its function depends on the time scales by which the chromatin fiber expands or contracts. Incorporating accurate dynamics into the 1CPN model is therefore an important aspect of the model. We evaluate the accuracy of the 1CPN model by comparing the dynamics of both short DNA strands ($\approx 10 - 100$ bp), and small chromatin fibers ($\approx 1 - 10$ nucleosomes) to available experimental measurements.

The diffusion constant of DNA molecules of different lengths was determined (see Methods) using the 1CPN model (Figure 10a), and found to be in good agreement with experiments[106–110], especially over the range of 20 to 80 base pairs, the typical range of linker DNA separating nucleosomes. We note that because of the "freely-draining" Brownian dynamics employed in the 1CPN model (see Methods), the scaling of the 1CPN model, $N^{-1}$, does not match the scaling measured experimentally, $N^{-0.67}$. The experimental scaling[106,109], $N^{-0.67}$, has been previously justified using the Zimm model of polymer dynamics adjusted to include the stiffness[109], or the helical worm-like chain theory of Yamakawa[111]. Though these better treatments of hydrodynamics could be incorporated into the 1CPN model, the first-order "freely-draining" Brownian dynamics approach provides a reasonable first step to describe dynamics.

The diffusion constant for short chromatin fibers was also measured using the 1CPN model (Figure 10b). Though the diffusivity of a single nucleosome is overestimated in 1CPN, the diffusivity of fibers two to four nucleosomes long is well reproduced in our model. Taken together, these results suggest that the 1CPN model should produce reasonable predictions of dynamic rearrangements within the chromatin fiber.

### F. Linker-mediated Nucleosome-Nucleosome Pair Potential

We next proceed to investigate the fundamental building block of the chromatin fiber: two nucleosomes connected by a section of DNA. In Section IV A, we discussed how the the pair-wise interactions between nucleosomes unconnected by DNA depends on their relative orientation and on histone modifications. Here, we examine how these pair potentials are affected when these two nucleosomes are connected by DNA. Nucleosomes are typically separated by 20 to 80 bp of DNA, length scales that are small relative to the bending and twisting persistence length of DNA. Thus, it might be expected that the pair-potential between two nucleosomes connected by DNA would be a strong function of various DNA parameters.

The effect of DNA length is presented in Figure 11b. The DNA length between nucleosomes is varied from 20 to 60 bp (167 to 207 nucleosome repeat length or NRL), and is chosen to be an even multiple of the pitch of DNA (10 bp). From these calculations, three minima in the

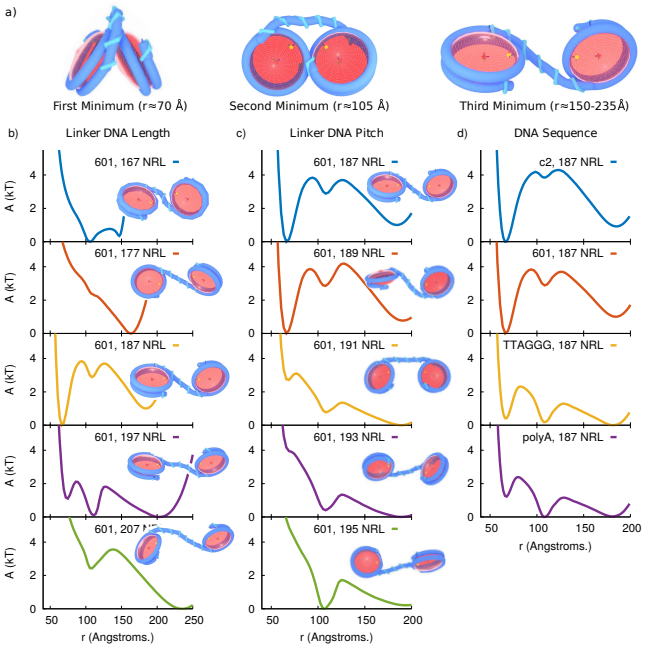


FIG. 11. Effect of linker DNA parameters on two-nucleosome fiber. (a) Molecular configurations for different nucleosome-nucleosome separation distances, $r$. (b) Effect of linker DNA length on two-nucleosome pair-potential for 601 positioning sequence. (c) Effect of linker DNA rotational phase on pair-potential. Nucleosome repeat lengths that differ by 2 bp correspond to equilibrium phase shifts of 72°. (d) Effect of DNA sequence on pair-potential. DNA sequence dependence is incorporated as described in Section IV D.

pair-potentials emerge (Figure 11a). The first minimum corresponds to the two nucleosomes in close face-to-face contact, reminiscent of configuration B in Figure 6, with the DNA strongly bent. The second minimum corresponds to the two nucleosomes aligned side-to-side (configuration A, Figure 6), whereas the third minimum corresponds to the two nucleosomes far apart, with the DNA relatively straight. As a consequence of this relatively straight DNA, the location of the third minimum increases from 150$\mathring{A}$ to 235$\mathring{A}$ as the NRL increases. Although these three minima appear for most NRLs examined, their relative free energies vary significantly as the NRL changes from 167 to 207 bp. For example, the NRL of 177 is observed to stabilize the third minimum, whereas for the 197 NRL the free energies corresponding to each minima are approximately equal.

We also examine the effect of DNA twist on the pair-potentials between two connected nucleosomes (Figure 11c). By varying the NRL by (a single) two base pairs at a time, the angle between the two nucleosomes changes significantly (36°/bp), while the length of the DNA changes only slightly (3.3$\mathring{A}$/bp). The angle between nucleosomes changes how much the DNA must twist to accommodate nucleosome-nucleosome contacts. Slight changes to the nucleosome angle are therefore observed to significantly change the free energies of the first

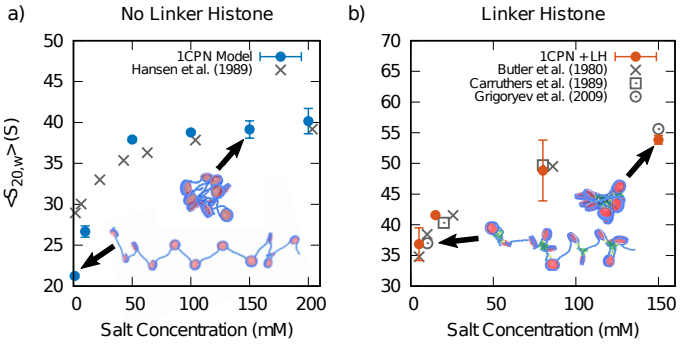


FIG. 12. Sedimentation Coefficient, $\langle S_{20,w} \rangle$, for short chromatin fibers constaining 12 nucleosome, as a function of salt concentration in the (a) absence and (b) presence of the Linker Histone. The nucleosome repeat length of 207 base pairs. the agreement between the 1CPN model (colored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

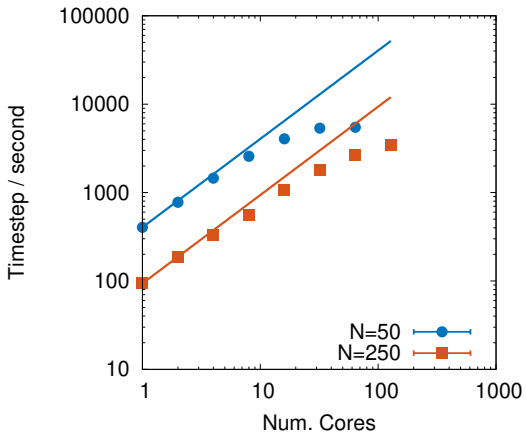


FIG. 13. Performance of 1CPN model for 50 and 250 nucleosome chromatin fibers. Lines indicate ideal scaling. Error bars are smaller than the symbols.

($r = 70\mathring{A}$), and second ($r = 105\mathring{A}$) minima in the pair potential.

At this point, it is important to recall that the amount of twist that can be stored in the linker DNA is balanced by the nucleosome-bound DNA's ability to resist rotation (see Section IV D). For strongly bound DNA sequences (e.g. the 601 sequence), it is expected that twist would be stored in linker DNA. However for weakly bound DNA sequences (e.g. PolyA), this twist might be able to escape the linker DNA by rotating the nucleosome-bound DNA. To examine if this is indeed the case, we can study the effect of DNA sequence on the pair-potential between two nucleosomes (Figure 11c). In agreement with our expectations, we observe that weakly-binding DNA sequences (eg. TTAGGG, PolyA), significanly lower the energy barrier to the second minimum. The relationship between DNA sequence and the 3-d structure of chromatin remains to be explored in detail, and we envision that the 1CPN model will provide a tool for pursuing such future studies.

### G. Chromatin Fiber Sedimentation Coefficients

As mentioned earlier, 1CPN is well suited to examine the structure and dynamics of chromatin fibers. To illustrate this idea, we prepared chromatin fibers consisting of 12 nucleosomes with a nucleosome repeat length of 207 bp and various salt concentrations. Following their relaxation, we computed the sedimentation coefficient (see Methods), and compared our predictions to experimental measurements for an identical system Hansen *et al.*[99] (Figure 12). Sedimentation coefficients have been used extensively to assess the relative compaction of different chromatin fibers, and these calculations represent an important test of 1CPN[35,76,112]. As can be seen in the figure, simulations and experiments agree with each other, particularly for the effects of salt on the compaction and expansion of chromatin.

### H. Performance

We have examined the efficiency of 1CPN as simulations are scaled across multiple processors (Figure 13). For systems of 50 nucleosomes (N=50), the 1CPN model exhibits excellent scaling across 16 processors, whereas in larger systems (N=250), near-ideal scaling is obtained across 64 processors. For a fifty nucleosome fiber, on a single 16 core node, a week of simulation time will produce a $\approx 50\mu s$ trajectory with 1CPN. The efficiency is therefore sufficient for study of a wide range of phenomena associated with chromatin structure and dynamics.

## V. CONCLUSION

A new model of chromatin, 1CPN, has been presented in this work. By relying on a multi-scale approach, 1CPN incorporates physics that occur over nano-meter length scales, such as histone modifications and DNA sequence. The model, however, is computationally efficient and permits simulation of many kilobases of chromatin. The parameterization of 1CPN has relied both on extensive simulations with the detailed 3SPN-AICG model of the nucleosomes, as well as experimental measurements on the structure and dynamics of chromatin. 1CPN has been parameterized to reproduce the many free energies that govern the interactions within chromatin, such as the interactions between nucleosomes, the interaction between DNA and histone tails and how these interactions can be modulated by histone modifications. 1CPN has also been parameterized to reproduce the salt-dependent stiffness of DNA, the effect of DNA sequence on rotation of DNA within the nucleosome, and the dynamics of short chromatin fibers. We incorporate linker histone H1 into 1CPN using the model of Luque et al.[40]. Following this parameterization, 1CPN has been

used to examine the free energies of association between two nucleosomes separated by different lengths of DNA. It is found that the length of this DNA, as well as the relative pitch between two nucleosomes, have a dramatic effect on the interactions between nucleosomes. Finally, it was demonstrated that 1CPN achieves quantitative agreement with experimental measurements of sedimentation coefficients of short chromatin fibers, both in the presence and absence of linker histone H1.

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

## ACKNOWLEDGMENTS

We are grateful for helpful discussions with Marat Andreev, Alec Bowen, and Jonathan Whitmer. The development of multiscale descriptions of DNA is supported by the National Science Foundation. The development of advanced sampling codes required for free energy calculations and the implementation of 1CPN in LAMMPS was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division through the Midwest Integrated Center for Computational Materials (MICCoM). We further acknowledge computational resources provided by the Midway computing cluster at the University of Chicago. A.C. acknowledges financial support from CONICYT under FONDECYT grant number: 11170056.

[1] J. S. You and P. A. Jones, Cancer Cell **22**, 9 (2012), arXiv:NIHMS150003.

[2] D. Hnisz, A. S. Weintraub, D. S. Day, A.-l. Valton, R. O. Bak, C. H. Li, J. Goldmann, B. R. Lajoie, Z. P. Fan, A. A. Sigova, J. Reddy, D. Borges-rivera, T. I. Lee, R. Jaenisch, M. H. Porteus, J. Dekker, and R. A. Young, Science **351**, 1454 (2016).

[3] G. E. Zentner and S. Henikoff, Nat Struct Mol Biol **20**, 259 (2013).

[4] C. L. Liu, T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber, N. Friedman, and O. J. Rando, PLoS Biology **3** (2005), 10.1371/journal.pbio.0030328.

[5] J. Ernst, P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis, and B. E. Bernstein, Nature **473**, 43 (2011).

[6] P. V. Kharchenko, A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle, J. Ernst, P. J. Sabo, E. Larschan, A. A. Gorchakov, T. Gu, D. Linder-Basso, A. Plachetka, G. Shanower, M. Y. Tolstorukov, L. J. Luquette, R. Xi, Y. L. Jung, R. W. Park, E. P. Bishop, T. K. Canfield, R. Sandstrom, R. E. Thurman, D. M. MacAlpine, J. A. Stamatoyannopoulos, M. Kellis, S. C. R. Elgin, M. I. Kuroda, V. Pirrotta, G. H. Karpen, and P. J. Park, Nature **471**, 480 (2011).

[7] J. Dekker, M. A. Marti-Renom, and L. A. Mirny, Nat. Rev. Genet. **14**, 390 (2013).

[8] E. Lieberman-aiden, N. L. V. Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, and L. A. Mirny, Science **326**, 289 (2009), arXiv:arXiv:1011.1669v3.

[9] L. A. Mirny, Chromosome Research **19**, 37 (2011).

[10] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, Nature **485**, 376 (2012).

[11] S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden, Cell **159**, 1665 (2014), arXiv:1206.5533.

[12] T. H. S. Hsieh, A. Weiner, B. Lajoie, J. Dekker, N. Friedman, and O. J. Rando, Cell **162**, 108 (2015).

[13] W. Wang, G. W. Li, C. Chen, X. S. Xie, and X. Zhuang, Science **333**, 1445 (2011).

[14] B. J. Beliveau, A. N. Boettiger, M. S. Avendaño, R. Jungmann, R. B. McCole, E. F. Joyce, C. Kim-Kiselak, F. Bantignies, C. Y. Fonseka, J. Erceg, M. A. Hannan, H. G. Hoang, D. Colognori, J. T. Lee, W. M. Shih, P. Yin, X. Zhuang, and C.-t. Wu, Nat. Commun. **6**, 7147 (2015).

[15] A. N. Boettiger, B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, G. Fudenberg, M. Imakaev, L. A. Mirny, C.-t. Wu, and X. Zhuang, Nature **529**, 418 (2016).

[16] S. Wang, J.-H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C.-t. Wu, and X. Zhuang, Science **353**, 598 (2016).

[17] B. Dong, L. M. Almassalha, Y. Stypula-Cyrus, B. E. Urban, J. E. Chandler, T.-Q. Nguyen, C. Sun, H. F. Zhang, and V. Backman, Proc. Nat. Acad. Sci. USA. **113**, 201602202 (2016).

[18] V. I. Risca and W. J. Greenleaf, Trends in Genetics **31**, 357 (2015).

[19] P. J. J. Robinson and D. Rhodes, Curr. Opin. Struc. Biol. **16**, 336 (2006).

[20] D. J. Tremethick, Cell **128**, 651 (2007).

[21] J. T. Finch and A. Klug, Proc. Nat. Acad. Sci. USA. **73**, 1897 (1976).

[22] J. Widom and a. Klug, Cell **43**, 207 (1985).

[23] J. D. McGhee, J. M. Nickol, G. Felsenfeld, and D. C. Rau, Cell **33**, 831 (1983).

[24] P. J. J. Robinson, L. Fairall, V. a. T. Huynh, and D. Rhodes, Proc. Natl. Acad. Sci. U.S.A. **103**, 6506 (2006).

[25] C. L. Woodcock, S. a. Grigoryev, R. a. Horowitz, and N. Whitaker, Proc. Nat. Acad. Sci. USA. **90**, 9021 (1993).

[26] S. P. Williams, B. D. Athey, L. J. Muglia, R. S. Schappe, a. H. Gough, and J. P. Langmore, Biophys. J. **49**, 233 (1986).

[27] B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder, and T. J. Richmond, Science **306**, 1571 (2004).

[28] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond, Nature **436**, 138 (2005).

[29] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, and G. Li, Science **344**, 376 (2014).

[30] K. Maeshima, S. Hihara, and M. Eltsov, Curr. Opin. Cell Biol. **22**, 291 (2010).

[31] E. Fussner, R. W. Ching, and D. P. Bazett-Jones, Trends Biochem. Sci. **36**, 1 (2011).

[32] K. Maeshima, S. Ide, K. Hibino, and M. Sasai, Current Opinion in Genetics and Development **37**, 36 (2016).

[33] S. A. Grigoryev, G. Bascom, J. M. Buckwalter, M. B. Schubert, C. L. Woodcock, and T. Schlick, Proc. Nat. Acad. Sci. USA. **113**, 1238 (2016).

[34] V. I. Risca, S. K. Denny, A. F. Straight, and W. J. Greenleaf, Nature **541**, 237 (2017).

[35] S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick, Proc. Nat. Acad. Sci. USA. **106**, 13317 (2009).

[36] R. Collepardo-Guevara and T. Schlick, Biophys. J. **101**, 1670 (2011).

[37] G. D. Bascom, K. Y. Sanbonmatsu, and T. Schlick, J. Phys. Chem. B **120**, 8642 (2016).

[38] G. Arya, Q. Zhang, and T. Schlick, Biophysical journal **91**, 133 (2006).

[39] G. Arya and T. Schlick, J. Phys. Chem. A **113**, 4045 (2009).

[40] A. Luque, R. Collepardo-Guevara, S. Grigoryev, and T. Schlick, Nucleic Acids Res. **42**, 7553 (2014).

[41] G. Wedemann and J. Langowski, Biophys. J. **82**, 2847 (2002).

[42] N. Kepper, D. Foethke, R. Stehr, G. Wedemann, and K. Rippe, Biophys. J. **95**, 3692 (2008).

[43] R. Stehr, N. Kepper, K. Rippe, and G. Wedemann, Biophys. J. **95**, 3677 (2008).

[44] N. Kepper, R. Ettig, R. Stehr, S. Marnach, G. Wedemann, and K. Rippe, Biopolymers **95**, 435 (2011).

[45] R. Stehr, R. Schopflin, R. Ettlg, N. Kepper, K. Rippe, and G. Wedemann, Biophys. J. **98**, 1028 (2010).

[46] G. S. Freeman, J. P. Lequieu, D. M. Hinckley, J. K. Whitmer, and J. J. de Pablo, Phys. Rev. Lett. **113**, 168101 (2014).

[47] J. Lequieu, A. Córdoba, D. C. Schwartz, and J. J. de Pablo, ACS Cent. Sci. **2**, 660 (2016).

[48] B. Zhang, W. Zheng, G. A. Papoian, and P. G. Wolynes, J. Am. Chem. Soc. **138**, 8126 (2016).

[49] J. Lequieu, D. C. Schwartz, and J. J. de Pablo, Proc. Nat. Acad. Sci. USA. **114**, E9197 (2017).

[50] G. S. Freeman, D. M. Hinckley, J. P. Lequieu, J. K. Whitmer, and J. J. de Pablo, J. Chem. Phys. **141**, 165103 (2014).

[51] W. Li, P. G. Wolynes, and S. Takada, Proc. Nat. Acad. Sci. USA. **108**, 3504 (2011).

[52] A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, P. G. Wolynes, and G. a. Papoian, J. Phys. Chem. B **116**, 8494 (2012).

[53] L. Chang and S. Takada, Scientific Reports **6**, 34441 (2016).

[54] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).

[55] G. Li, M. Levitus, C. Bustamante, and J. Widom, Nat Struct Mol Biol **12**, 46 (2005).

[56] H. S. Tims, K. Gurunathan, M. Levitus, and J. Widom, J. Mol. Biol. **411**, 430 (2011).

[57] Y. Lorch, B. Davis, and R. D. Kornberg, Proc. Nat. Acad. Sci. USA. **102**, 1329 (2005).

[58] J. M. Gottesfeld, J. M. Belitsky, C. Melander, P. B. Dervan, and K. Luger, J. Mol. Biol. **321**, 249 (2002).

[59] F. Mueller-Planitz, H. Klinker, and P. B. Becker, Nat. Struct. Mol. Biol. **20**, 1026 (2013).

[60] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, J. Mol. Biol. **319**, 1097 (2002).

[61] H. Zewdie, Phys. Rev. E **57**, 1793 (1998).

[62] A. Stone, Mol. Phys. **36**, 241 (1978).

[63] B. Brower-Toland, D. A. Wacker, R. M. Fulbright, J. T. Lis, W. L. Kraus, and M. D. Wang, J. Mol. Biol. **346**, 135 (2005).
[64] G. Arya and T. Schlick, Proc. Nat. Acad. Sci. USA. **103**, 16236 (2006).
[65] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo, J. Chem. Phys. **139**, 144903 (2013).
[66] G. Chirico and J. Langowski, Biopolymers **34**, 415 (1994).
[67] C. A. Brackley, A. N. Morozov, and D. Marenduzzo, J. Chem. Phys. **140**, 135103 (2014).
[68] S. Tan and C. A. Davey, Curr. Opin. Struc. Biol. **21**, 128 (2011).
[69] Y. P. Kalmykov, J. Mol. Liq. **69**, 117 (1996).
[70] C. L. Woodcock, A. I. Skoultchi, and Y. Fan, Chromosome Research **14**, 17 (2006).
[71] J. Bednar, I. Garcia-Saez, R. Boopathi, A. R. Cutter, G. Papai, A. Reymer, S. H. Syed, I. N. Lone, O. Tonchev, C. Crucifix, H. Menoni, C. Papin, D. A. Skoufias, H. Kurumizaka, R. Lavery, A. Hamiche, J. J. Hayes, P. Schultz, D. Angelov, C. Petosa, and S. Dimitrov, Molecular Cell **66**, 384 (2017).
[72] F. Thoma, T. Koller, A. Klug, Koller Th., and A. Klug, Journal of Cell Biology **83**, 403 (1979).
[73] H. Fang, S. Wei, T. H. Lee, and J. J. Hayes, Nucleic Acids Research **44**, 9131 (2016).
[74] S. W. Harshman, N. L. Young, M. R. Parthun, and M. A. Freitas, "H1 histones: Current perspectives and challenges," (2013).
[75] F. Thoma, T. Koller, J. Baldwin, P. Boseley, E. Bradbury, K. Ibel, M. Bellard, P. Oudet, J. Germond, P. Chambon, A. Billett, J. Barry, E. Bolund, C. John, E. Bradbury, S. Danby, H. Rattle, V. Giancotti, S. Bram, S. Bram, H. Ris, K. Brasch, M. Bustin, D. Goldblatt, R. Sperling, J. Compton, R. Hancock, P. Oudet, P. Chambon, E. Dupraw, G. Felsenfeld, J. Finch, A. Klug, J. Finch, M. Noll, R. Kornberg, J. Gall, J. Germond, M. Bellard, P. Oudet, P. Chambon, J. Griffith, K. Hayashi, E. Matsukera, Y. Ohba, D. Hewish, L. Burgoyne, J. Jones, M. Beer, G. Kiryanov, T. Manamshjan, V. Polyakov, D. Fais, J. Chentsov, R. Kornberg, F. Lampert, J. Langmore, J. Wooley, H. Li, J. MacGillivray, A. Cameron, R. Kranze, D. Rickwood, J. Paul, M. Noll, M. Noll, J. Thomas, R. Kornberg, A. Olins, D. Olins, P. Oudet, M. Gross-Bellard, P. Chambon, A. Pooley, J. Pardon, B. Richards, H. Ris, H. Ris, H. Ris, D. Kubai, C. Sahasrabuddhe, K. V. Holde, B. Shaw, T. Herman, R. Kovacic, G. Beaudreau, K. V. Holde, A. Varshavsky, A. Varshavsky, V. Bakayev, G. Georgiev, H. Vollenweider, J. Sogo, T. Koller, H. Vollenweider, T. Koller, J. Sogo, J. Parello, J. Whitlock, R. Simpson, C. Woodcock, C. Woodcock, J. Safer, and J. Stanchfield, Cell **12**, 101 (1977).
[76] A. Routh, S. Sandin, and D. Rhodes, Proceedings of the National Academy of Sciences of the United States of America **105**, 8872 (2008).
[77] P. J. Butler and J. O. Thomas, Journal of Molecular Biology **140**, 505 (1980).
[78] J. Allan, P. G. Hartman, C. Crane-robinson, and F. X. Aviles, Nature **288**, 675 (1980).
[79] V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet, Nature **362**, 219 (1993).
[80] T. A. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo, J. Chem. Phys. **126**, 084901 (2007).
[81] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, Biophys. J. **96**, 1675 (2009).
[82] A. Córdoba, D. M. Hinckley, J. Lequieu, and J. J. de Pablo, Biophys. J. **112**, 1302 (2017).
[83] J. P. Lequieu, M. Hinckley, and J. J. de Pablo, Soft Matter **11**, 1919 (2015).
[84] J. Lequieu, A. Córdoba, D. Hinckley, and J. J. de Pablo, ACS Cent. Sci. **2**, 614 (2016).
[85] J. Kästner, Wiley Interdiscip Rev Comput Mol Sci **1**, 932 (2011).
[86] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **16**, 1339 (1995).
[87] A. Laio and F. L. Gervasio, Rep Prog Phys **71**, 126601 (2008).
[88] J. Yao, P. Lowary, and J. Widom, Proceedings of the National Academy of Sciences **87**, 7603 (1990).
[89] J. Yao, P. Lowary, and J. Widom, Biochemistry **30**, 8408 (1991).
[90] J. Bednar, R. A. Horowitz, J. Dubochet, and C. L. Woodcock, The Journal of cell biology **131**, 1365 (1995).
[91] I. Jimenez-Useche, N. P. Nurse, Y. Tian, B. S. Kansara, D. Shim, and C. Yuan, Biophysical journal **107**, 1629 (2014).
[92] M. Kruithof, F.-T. Chien, A. Routh, C. Logie, D. Rhodes, and J. van Noort, Nature structural & molecular biology **16**, 534 (2009).
[93] Y. Cui and C. Bustamante, Proc. Nat. Acad. Sci. USA. **97**, 127 (2000).
[94] F. T. Chien and T. Van Der Heijden, Biophysical Journal **107**, 373 (2014).
[95] J. J. Funke, P. Ketterer, C. Lieleg, S. Schunter, P. Korber, and H. Dietz, Science Advances **2**, e1600974 (2016).
[96] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proc. Nat. Acad. Sci. USA. **94**, 6185 (1997).
[97] B. S. Fujimoto and J. M. Schurr, Nature **344**, 175 (1990), arXiv:NIHMS150003.
[98] J. Lipfert, J. W. J. Kerssemakers, T. Jager, and N. H. Dekker, Nature Methods **7**, 977 (2010).
[99] J. C. Hansen, J. Ausio, V. H. Stanik, and K. E. van Holde, Biochemistry **28**, 9129 (1989).
[100] J. Widom, Proceedings of the National Academy of Sciences of the United St
[101] S. Mihardja, A. J. Spakowitz, Y. Zhang, and C. Bustamante, Proc. Nat. Acad. Sci. USA. **103**, 15871 (2006).
[102] K. Struhl and E. Segal, Nat. Struct. Mol. Biol. **20**, 267 (2013).
[103] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond, Nature **389**, 251 (1997).
[104] R. S. Edayathumangalam, P. Weyermann, P. B. Dervan, J. M. Gottesfeld, and K. Luger, J. Mol. Biol. **345**, 103 (2005).
[105] I. Kulić and H. Schiessel, Phys. Rev. Lett. **91**, 148103 (2003).
[106] N. C. Stellwagen, S. Magnusdottir, C. Gelfi, and P. G. Righetti, Biopolymers **58**, 390 (2001).
[107] W. Eimer and R. Pecora, J Chem Phys **94**, 2324 (1991).
[108] L. Wang and H. Yu, Macrom , 3498 (1988).
[109] S. S. Sorlie and R. Pecora, Macromolecules **23**, 487 (1990).
[110] H. T. Goinga and R. Pecora, Macromolecules **24**, 6128 (1991).
[111] H. Yamakawa, J. Chem. Phys. **64**, 5222 (1976).
[112] V. A. T. Huynh, P. J. J. Robinson, and D. Rhodes, Journal of Molecular Biology **345**, 957 (2005).