

RESEARCH ARTICLE

An associative memory Hamiltonian model for DNA and nucleosomes

WeiQi Lu¹, José N. Onuchic^{1,2*}, Michele Di Pierro^{3,4*}

1 Center for Theoretical Biological Physics, & Department of Physics and Astronomy, Rice University, Houston, Texas, United States of America, **2** Department of Chemistry, & Department of Biosciences, Rice University, Houston, Texas, United States of America, **3** Department of Physics, Northeastern University, Boston, Massachusetts, United States of America, **4** Center for Theoretical Biological Physics, Northeastern University, Boston, Massachusetts, United States of America

* jonuchic@rice.edu (JNO); dipierro@northeastern.edu (MDP)



Abstract

A model for DNA and nucleosomes is introduced with the goal of studying chromosomes from a single base level all the way to higher-order chromatin structures. This model, dubbed the Widely Editable Chromatin Model (WEChroM), reproduces the complex mechanics of the double helix including its bending persistence length and twisting persistence length, and the temperature dependence of the former. The WEChroM Hamiltonian is composed of chain connectivity, steric interactions, and associative memory terms representing all remaining interactions leading to the structure, dynamics, and mechanical characteristics of the B-DNA. Several applications of this model are discussed to demonstrate its applicability. WEChroM is used to investigate the behavior of circular DNA in the presence of positive and negative supercoiling. We show that it recapitulates the formation of plectonemes and of structural defects that relax mechanical stress. The model spontaneously manifests an asymmetric behavior with respect to positive or negative supercoiling, similar to what was previously observed in experiments. Additionally, we show that the associative memory Hamiltonian is also capable of reproducing the free energy of partial DNA unwrapping from nucleosomes. WEChroM is designed to emulate the continuously variable mechanical properties of the 10nm fiber and, by virtue of its simplicity, is ready to be scaled up to molecular systems large enough to investigate the structural ensembles of genes. WEChroM is implemented in the OpenMM simulation toolkits and is freely available for public use.

OPEN ACCESS

Citation: Lu W, Onuchic JN, Di Pierro M (2023) An associative memory Hamiltonian model for DNA and nucleosomes. PLoS Comput Biol 19(3): e1011013. <https://doi.org/10.1371/journal.pcbi.1011013>

Editor: Tamar Schlick, New York University, UNITED STATES

Received: October 15, 2022

Accepted: March 8, 2023

Published: March 27, 2023

Copyright: © 2023 Lu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files. Source code is available at <https://github.com/DiPierroLab/WeChroM>.

Funding: Work at the Center for Theoretical Biological Physics was supported by the National Science Foundation (NSF) (Grant PHY-2019745). JNO and WL were also supported by the NSF (Grant PHY-2210291) and by the Welch Foundation (Grant C-1792). JNO is a CPRIT Scholar in Cancer Research sponsored by the Cancer Prevention and Research Institute of Texas.

Author summary

The structural ensembles of genes have been so far out of the reach for theoretical and computational investigations because genes are molecular complexes too large to be tackled with even the most efficient computational chemistry approaches and yet too strongly affected by heterogeneous molecular factors to be effectively modeled as a simple polymer. In this work, we develop a computationally efficient, easy-to-use, and widely editable chromatin model designed to address this challenge. Using the framework of Associative

MDP was also supported by the NIGMS of the National Institutes of Health under award number R35GM146852. We thank AMD for providing the computational resources in part used in this research. Additional computing resources were supported in part by the Big-Data Private-Cloud Research Cyberinfrastructure MRI-award funded by NSF under grant CNS-1338099 and by Rice University's Center for Research Computing (CRC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Memory Hamiltonians, this model reproduces the structural and mechanical properties of double-stranded DNA and accounts for the effects of nucleosome-forming histone octamers and other proteins bound to DNA. Our results open the path to studying the structural and mechanical ensembles of genetic systems as large as tens of kilobases of chromatin, i.e., the size of mammalian genes.

Introduction

Deoxyribonucleic acid (DNA) has been studied for decades since the fundamental discovery made by Watson, Crick, Wilkins, and Franklin [1–3]. The repeating units of the DNA molecule, nucleotides, are composed of phosphate, sugar, and a base group; a sequence of covalently bonded nucleotides form single-stranded DNA. The double-helix structure of DNA, formed by two strands of DNA, is maintained by base pairing, planar base stack interactions, and electrostatic interactions [4]. In eukaryotes, DNA is then wrapped around histones to form nucleosomes, which serve as the basic structural unit of DNA packaging. Each nucleosome contains about 147 base pairs of DNA wrapped 1.7 times around a histone octamer. Recent evidence shows that the structure of nucleosomes is very dynamic and irregular [5]. Outer stretches of nucleosomal DNA are observed to spontaneously unwrap and rewrap to the core octamer [6], which can result in DNA strands entering and exiting the nucleosome at flexible angles, and this ultimately has a significant impact on the three-dimensional packing of chromatin. However, we still lack a sufficient understanding of DNA and nucleosome mechanics to ascertain how these features combine with other factors to determine the structural and mechanical properties of higher-order structures.

Numerous theoretical and computational approaches have already been explored to model DNA and chromatin fibers, greatly improving our understanding but still leaving many questions open. This work aims at integrating previous efforts by proposing a novel theoretical approach that focuses on the mechanics of DNA and nucleosomes, yet simple enough to scale up to larger scales.

Detailed, all-atom (AA) simulations have been extensively and successfully used to gain insights into DNA dynamics [7], flexibility [8], and other properties at a fine molecular scale. Fully atomistic models, however, cannot simulate DNA molecules at the level of nucleosome organization or chromatin fiber due to computational limitations, at least in the foreseeable future. Additionally, for very large molecular assemblies, such as even a short segment of chromatinized DNA, the accuracy of AA force fields remains an issue. This is particularly problematic in systems like DNA and histones in which ions play a determinant role. Coarse-grained (CG) models [9–29] for DNA have also proven very successful.

CG approaches reach the time and length scales that are inaccessible to AA simulations. In bottom-up CG models [9–15], effective interactions are determined to eliminate some degrees of freedom from reference atomistic simulations. As a result, CG force fields' ability to reproduce the system's behavior depends on the accuracy of the underlying AA force field, from which it inherits limitations in predictive capabilities, although reducing the computational cost. Top-down CG empirical energy functions [16–29] are instead tuned to match the specific properties of the system as observed through experiments, such as the polymer's flexibility or melting temperature, thus avoiding the pitfalls of atomistic simulations altogether at the expense of universality. CG DNA models have been developed using a variable number of sites to represent each nucleotide: from two [16–19], three [20–24], six [25,26], up to eight [27]. The oxDNA model [16,17] represents each nucleotide as a rigid body with 2 interacting

sites and effectively capture basic structural properties of DNA such as pitch together with mechanical properties as the persistence length. The oxDNA has been successfully used to characterize DNA origami nanostructures [30]. The model by Drukker et al. [18,19] has similar resolution but focuses on the melting behavior and denaturation in double-stranded sequences. The 3-site-per-nucleotide (3SPN.0 [20]/1 [21]/2 [22]/2C [23]) model was developed by de Pablo's group to reproduce the persistence length for both single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA), melting temperatures, and hybridization rate constants. The BioModi Model [24] also chose a 3-site-per-nucleotide representation focusing on modeling the thermodynamics and kinetics of DNA self-assembly. The SIRAH model [25] uses 6 beads per base with explicit solvents and allow modeling interaction of DNA with lipids and proteins. The HiRE-DNA model [27] was extended from its RNA counterpart focusing on DNA self-assembly.

Coarser models are necessary to perform simulations of the chromatin architecture on the size of tens of thousands, even millions of bp [31–50]. A mesoscale model developed by Schlick and coworkers [31,32] employed a rigid nucleosome core particle, flexible histone tails, and 9-bp-per-particle worm-like-chain linker DNA. The model has been used to investigate the relationship between chromatin organization and structural factors, such as nucleosome repeat length and linker histones [33–37]. The 1CPN model proposed by Lequieu et al. [38] uses a rigid nucleosome particle with a twistable wormlike chain representing linker DNA at the resolution of 3 bp per particle. Other coarse-grained models with nucleosome and sub-nucleosome resolution have also been developed in the past few years to connect the molecular and physicochemical data of nucleosomes with the mesoscale characteristics of chromatin, including the “beads-on-a-string” models from Wedemann's group [39–43] and Bajpai et al. [44], the multiscale model from Farr et al. [45], the rigid base-pair model from the van Noort group [46,47], the mesoscopic model from Zhurkin and Norouzi [48,49], and the multi-modal model from Orozco group [50,51]. Most of these mesoscopic models use one interaction site to represent a variable number of nucleotides, relinquishing from explicitly describing of the double helix, which is instead a fundamental determinant of the mechanical properties of DNA, and in particular its supercoiling behavior. As a result, a gap persists in understanding how the mechanical properties of the molecular components of chromatin, DNA, nucleosomes, and DNA-binding proteins affect higher-order chromatin structures.

Traditional CG models involve two types of interactions: bonded potentials among neighbors and non-bonded potentials among any pair of particles. The bonded interactions typically entail the 2-body bonding term, 3-body angle or bending term, 4-body aligning/twisting/dihedral terms, and 4-body to many-body terms to capture base stacking effects. Non-bonded interactions include excluded-volume potentials and electrostatic potentials where the solvent is typically treated implicitly. While following a top-down philosophy, traditional CG models aim to reproduce the physicochemical nature of DNA using interaction potentials motivated by basic chemistry; however, such ambitious effort is often frustrated by a difficult parametrization process and by computationally expensive simulations, due to multi-body interactions and non-bonded electrostatic potentials.

Such detail may not be necessary if we limit our objective to recapitulating just the mechanical properties of DNA. We use an Associative Memory Hamiltonian [52–54] (AMH) to explore the energy landscape of chromatinized DNA, starting from known structures of DNA fragments and protein-DNA complexes. The AMH framework has been successfully employed in protein folding where the energy landscape of a given protein is reconstructed using crystallographic information of short fragments characterized by sequences overlapping that of the protein under investigation [52–54]. We apply this data-driven approach to build a model for DNA and nucleosomes using the 3D structures of B-DNA and nucleosomes. The AMH

implicitly recapitulates all the energy terms already discussed: bending, twisting, base-stacking, and electrostatic energy. Parameters are tuned to reproduce the mechanical properties of DNA, including the twisting and the bending persistence lengths, and their coupling with supercoiling behaviors.

This model, dubbed the Widely Editable Chromatin Model (WEChroM), represents each nucleotide as a single particle and, besides polymer connectivity, uses AMH energy terms for all bonding interactions. By virtue of its simplicity, the WEChroM is computationally efficient and well-suited to simulate large genetic molecular systems up to kilobases of chromatin at base-pair resolution, reaching over to the size of mammalian genes.

We show that the WEChroM accurately reproduces the properties of naked DNA, such as bending and twisting persistence lengths, as well as supercoiling behaviors. Because of the finite energy barrier characteristic of the AMH potentials, the model can also account for the natural emergence of defects in the structure of DNA. Lastly, we introduce an associative memory template for the histone octamer and show that WEChroM recapitulates the free energy profile of nucleosome unwrapping.

Results

An associative memory Hamiltonian for DNA

The AMH was introduced by Friedrichs and Wolynes [56] to study the problem of protein folding. Using crystallographic information, the AMH reconstructs the energy landscape of a protein and encodes the native structure of the target sequence as the global attractor of such a landscape. The energy landscape of the protein is approximated by selecting a subset of the pairs of atoms in the reference protein structure and constraining their distances using a Gaussian well potential. The resulting model of the molecule is roughly harmonic in the folded state but can accommodate configurations with partial or complete unfolding of the protein at a finite energy cost. More than one molecular configuration can be encoded in the Hamiltonian; each one of these configurations is called a “memory”, in analogy with the terminology used in the field of recurrent neural networks [52,56,57].

We adapt the AMH framework to reconstruct DNA interactions and DNA-protein interactions. We start from idealized computational structures for B-DNA and from crystal structures of DNA-protein complexes as memories to construct the energy landscape of DNA and nucleosomes. In the WEChroM, each nucleotide is represented by one particle, as shown in Fig 1. The model is sequence-independent, but this can be easily generalized. The resolution of one nucleotide per particle allows for describing both the bending and twisting properties of DNA without additional rotational coordinates, and it allows the model to closely reproduce the complex mechanical behavior of the double helix, including buckling and the formation of plectonemes.

For naked DNA, the WEChroM potential energy consists of 3 major components: chain connectivity, steric interactions, and associative memory interactions, in this case labeled the double helix term (U_{DH}). The chain connectivity is enforced by harmonic interactions between two neighboring particles on the same strand, mimicking the covalent backbone bonds in a single-stranded DNA. Steric interactions are modeled using a short-range repulsive potential. A complete description of the energy terms modeling chain connectivity and steric interactions is provided in the Appendix B in S1 Document.

Pairwise distances (r_{ij}^m) among nucleotides are captured from computational B-DNA structures and define the associative memory energy term (U_{DH})

$$U_{DH} = - \sum_m \sum_{ij} \lambda \exp - \frac{(r_{ij} - r_{ij}^m)^2}{2\sigma_{ij}^2} \quad (1)$$

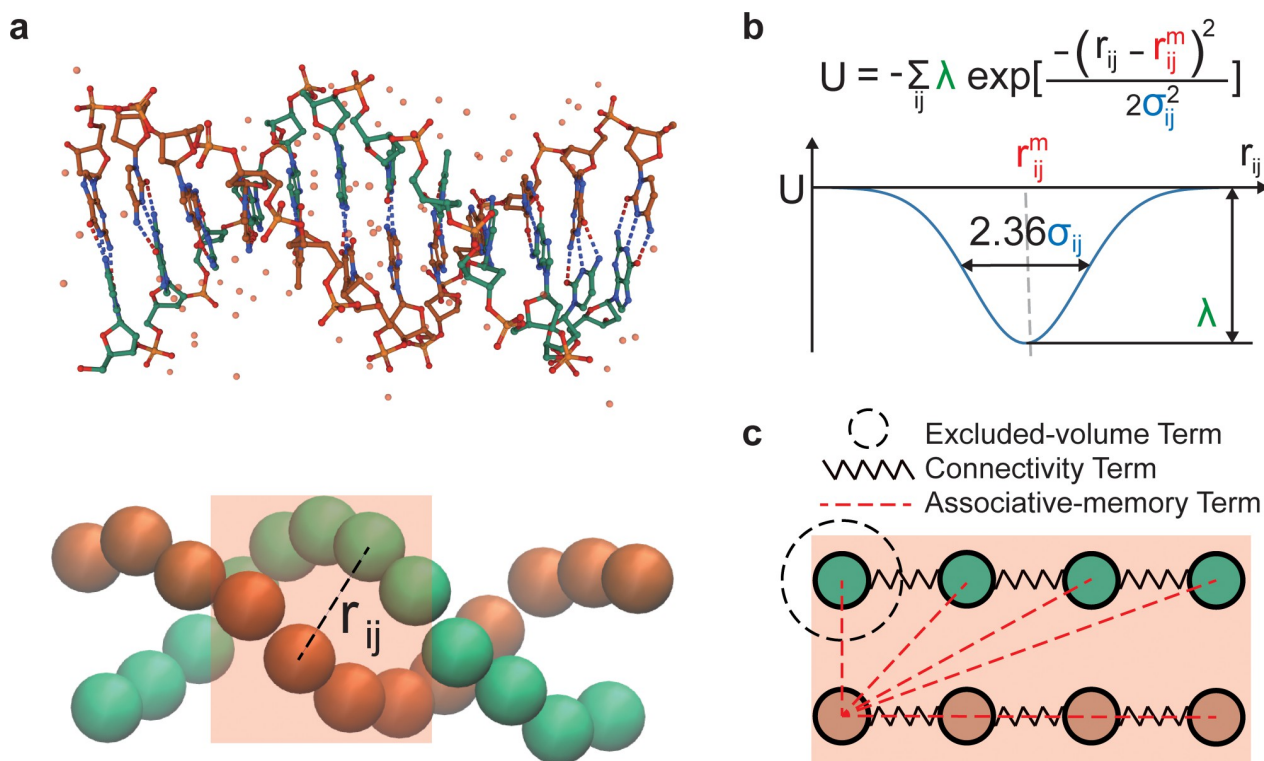


Fig 1. An Associative Memory Hamiltonian for DNA. (a) Comparison of an atomistic representation and a coarse-grained representation of a B-DNA dodecamer (PDB ID: 1BNA [55]). (b) The AMH recapitulates the energy landscape of the DNA molecule using a series of Gaussian wells. The parameters r_{ij}^m are determined using the distances measured in the structural memories, while the energy of interactions, λ and the scale of structural fluctuations, σ , are tuned to reproduce experimental observables. (c) Schematic representation of the energy terms in the model: connectivity (springs), excluded-volume terms (dashed circle), and AMH terms encoding the double helix structure (red dashed lines). An AMH term is applied between any pair of particles inside the box of 4bp. The box is then slid along the DNA, and the procedure is repeated.

<https://doi.org/10.1371/journal.pcbi.1011013.g001>

where r_{ij} is the distance between nucleotides i and j , and where the outer sum runs over memories and the inner sum over all pairs of nucleotides within the memory. The energy λ determines the strength of interactions and it is chosen to be a constant value, regardless of the nucleotide identity. The variance σ_{ij} controls the scale of structural fluctuations. To allow for more flexibility when the interaction distance is larger, σ_{ij} increases proportionally to the square root of the genomic distance between i, j , i.e., $\sigma_{i,i+n} = \sqrt{n}\sigma_0$ (Fig A in S1 Document).

In our model, each memory comprises 4 consecutive bp (Fig 1C). Our analysis indicates that 4bp is the minimal memory size needed for reproducing the mechanics of DNA without creating artifacts. The resulted in interaction patterns among the nucleotides resemble those in Savelyev and Papoian [10,11]. While we were able to reproduce the bending and twisting persistence length of DNA using 3bp memories, this required unreasonably large interaction energies between nucleotide pairs, and we thus rejected the model. We introduce energy terms by sliding the 4-bp memory box along DNA in 1-bp shifts. This results in multiple memory interactions being added between pairs of nucleotides depending on their index difference (Fig A in S1 Document), with pairs of nucleotides closer in genomic distance interacting more strongly.

To create a sequence-independent structural template, we choose 10 random sequences and generate their structures using X3DNA [58]; we then calculate each nucleotide's center of mass and use it to define the distances between nucleotides r_{ij}^m . U_{DH} enshrines the double-helical structure of DNA as the global energy minimum of the WEChroM Hamiltonian while

allowing for structural distortions as well as local defects in the DNA polymer, such as kinks and bubbles.

The current model does not include any energy term enforcing chirality. The correct chirality is set by the initial condition and preserved throughout the simulation. While in principle energy fluctuations could change the global chirality of the system, in practice we did not observe any of these events during the extensive testing of the model. We concluded that a term enforcing chirality was not necessary, at least for the systems under investigation. It is possible, however, that future studies for particular larger molecular systems and different conditions might require such an energy term.

Similar to modeling double-stranded DNA, we also use the associative memory Hamiltonian framework to describe how proteins influence the structure of chromatin. We model the effect of the histone octamer on DNA by using the crystal structure of a nucleosome core particle (PDB ID: 1kx5 [59]) as the associative memory template. In this case, associative memory terms are introduced among nucleotides and between them and the histone octamer's center of mass. Despite the simplified representation of protein-DNA interactions, our model recapitulates the free energy changes involved in the partial unwrapping of DNA from nucleosomes.

In summary, the AMH is a flexible theoretical framework that allows for introducing structural memories from a vast repertoire of DNA-binding proteins and histone variants and integrating the effect of these molecules at larger scales. This namesake characteristic of the Widely Editable Chromatin Model makes it particularly suitable for studying genetic systems above the nucleosomal scale and below the chromosomal scale, bridging the gap between the atomistic and bottom-up models typically utilized up to the nucleosomal scale, and the polymer physics approaches typical of the chromosomal scale [60]. The WEChroM is thus designed to provide the means for studying the structural ensembles of genes, which are too large for conventional molecular modeling and yet strongly affected by heterogeneous molecular factors to be effectively modeled by simple polymer models.

Bending and twisting persistence lengths

The molecular rigidity of naked DNA is quantified by two persistence lengths: the bending persistence length (L_{bp}) and the twisting persistence length (L_{tp}). Experimentally, the bending persistence length L_{bp} of DNA has been found to be approximately 150 bp (around 50 nm) [63], while the twisting persistence length L_{tp} is believed to be between 75 and 360 bp (or around 25–120 nm) [64–66]. In our model, the bending persistence length L_{bp} is defined by the relaxation of the angle correlation as a function of the genomic distance L :

$$\langle \mathbf{u}(i), \mathbf{u}(i+L) \rangle_i = \exp\left(-\frac{L}{L_{bp}}\right) \quad (2)$$

where $\mathbf{u}(i)$ is tangent to the helical axis at base-pair i . Due to DNA's double helix structure, the angle formed by any two base pairs is oscillating, leading to the following relaxation function for the twisting angle:

$$\langle \mathbf{v}(i), \mathbf{v}(i+L) \rangle_i = \exp\left(-\frac{L}{L_{tp}}\right) \cos(aL + b) \quad (3)$$

where $\mathbf{v}(i)$ is the vector pointing from one strand to the other strand within base pair (i). The twisting persistence length L_{tp} is defined from the exponentially decaying envelope $\exp\left(-\frac{L}{L_{tp}}\right)$; while a, b are the period and phase of the double helix.

By tuning the parameters λ and σ_0 in the double helix term in the AMH, we obtain a range of values for L_{bp} and L_{tp} as shown in Fig 2A and 2B, respectively. The bending persistence length L_{bp} observed in simulations varies between 100 to 500 bp and monotonically increases with increasing λ or decreasing σ_0 . This is expected as larger λ 's lead to stronger associative memory interactions while smaller σ_0 's lead to reduced fluctuations of the distances between nucleotides, both resulting in a stiffer molecule. The twisting persistence length L_{tp} , observed in simulations varies between 89 and 415bp, in good agreement with the experimentally determined range between 75 and 360bp (Fig 2B and Fig 2C in S1 Document). The persistence lengths obtained simulating systems of different sizes (Fig C in S1 Document) are stable, indicating the convergence in the simulations.

Experimental studies have determined that the DNA persistence length depends on temperature, with L_{bp} ranging from 140~165 bp at 277K to approximately 110~120bp at 330K [61,62]. We chose parameter sets characterized by a $L_{bp} \sim 150$ bp at 300K, i.e., consistent with experiments at room temperature. For these parameter sets, we calculate the persistence length L_{bp} as a function of the temperature, ranging from 270K to 340K. Simulations show that the persistence length L_{bp} decreases from 150~180 bp at 275K to 120~140bp at 338K, which is in excellent agreement with the experimentally determined temperature dependence (Fig 3C).

Multiple sets of parameters are compatible with the reported bending and twisting persistence lengths from experiments (highlighted in Fig 2A and 2B and listed in Table B in S1 Document). These different sets of parameters, although having similar persistence lengths, show different features in other aspects. For example, the parameter set with $(\lambda, \sigma) = (0.15, 0.05)$ has a similar L_{bp} and L_{tp} as other data points as shown in Fig 2A, yet the double helix with this parameter set is more prone to forming defects at room temperature. Nonetheless, other parameter sets can produce L_{bp} and L_{tp} that are consistent with the experiments without obvious defects. Such degeneration is caused by the scarcity of experimental constraints and can be removed by introducing additional information on specific genetic systems.

We performed a sensitivity analysis of both bending and twisting persistence length with respect to the energy terms in our model (Fig B in S1 Document). Such an analysis revealed that the stiffness of the double helix is much more sensitive to the strength of the inter-strand interactions than to changes in the intra-strand interactions, with minor increases in the strength of inter-strand bonds leading to a significantly stiffer DNA polymer. This is consistent with the fact that the bending persistence length of a single-stranded DNA (ssDNA) is only approximately 6 bp [67], much smaller than that of double-stranded DNA (dsDNA).

Supercoiling behavior of DNA minicircles

Genomic DNA is often subjected to torsional stress, which can both over- and under-wind the DNA double helix [68], resulting in twisting and coiling of the helix. Supercoiling can be induced by enzymes in order to reduce the volume of chromatin. Therefore, it will also be inevitably introduced when part of the DNA is uncoiled during essential cell processes, like transcription or replication [69,70]. Elucidating the structural properties of DNA molecules in different supercoiling states is therefore essential to improve our understanding of genome functions.

In order to quantitatively characterize supercoiling, we utilize the linking number (Lk), i.e., the total number of times the two single DNA strands coil about one another. The linking number is related to two geometrical properties of the molecule, the twist (Tw , the coiling of the two strands about the helical axis) and the writhe (Wr , the coiling of the helix axis's path around itself). These three properties are related by the simple relation

$$Lk = Wr + Tw \quad (4)$$

More details on how to calculate these quantities are found in Appendix D in S1 Document.

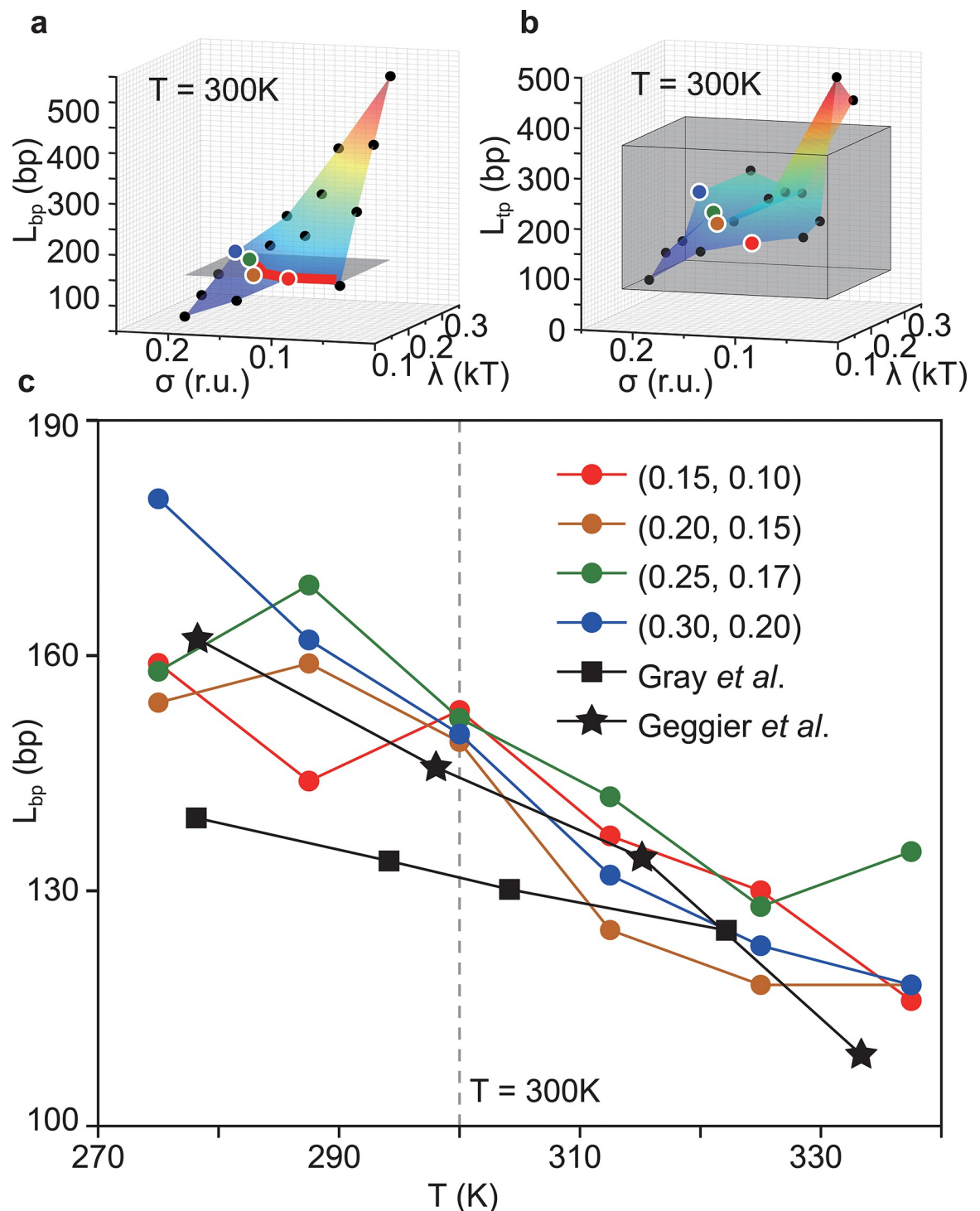


Fig 2. DNA Persistence Length and Parametrization. (a)(b) These figures show how persistence lengths depend on the strength of the energy interactions, λ , and the scale of structural fluctuations, σ . The bending persistence length is shown in panel (a), while the twisting persistence length is shown in panel (b). The σ is in reduced units (r.u.) which is 0.241 nm as explained in the Appendix A in [S1 Document](#). Dots indicate the data points obtained by simulation; colored surfaces indicate their linear interpolation. The horizontal gray plane in (a) shows the experimentally determined bending persistence length of DNA, 150bp. This persistence length is set as the target for the optimization of the parameters. The red line indicates the intersection of the two surfaces, where our model meets the target persistence length value. Similarly, the gray box in (b) is the experimentally determined twisting persistence length, 75–360bp, our target in this case. (c) The figure shows the temperature dependence of bending persistence length. The four colored lines show results obtained using the WEChroM model. Black lines show experimental results published by Gray *et al.* [61] and Geggier *et al.* [62], respectively. As shown, the proposed model agrees well with the experimental observations for several couples of parameters (λ , σ).

<https://doi.org/10.1371/journal.pcbi.1011013.g002>

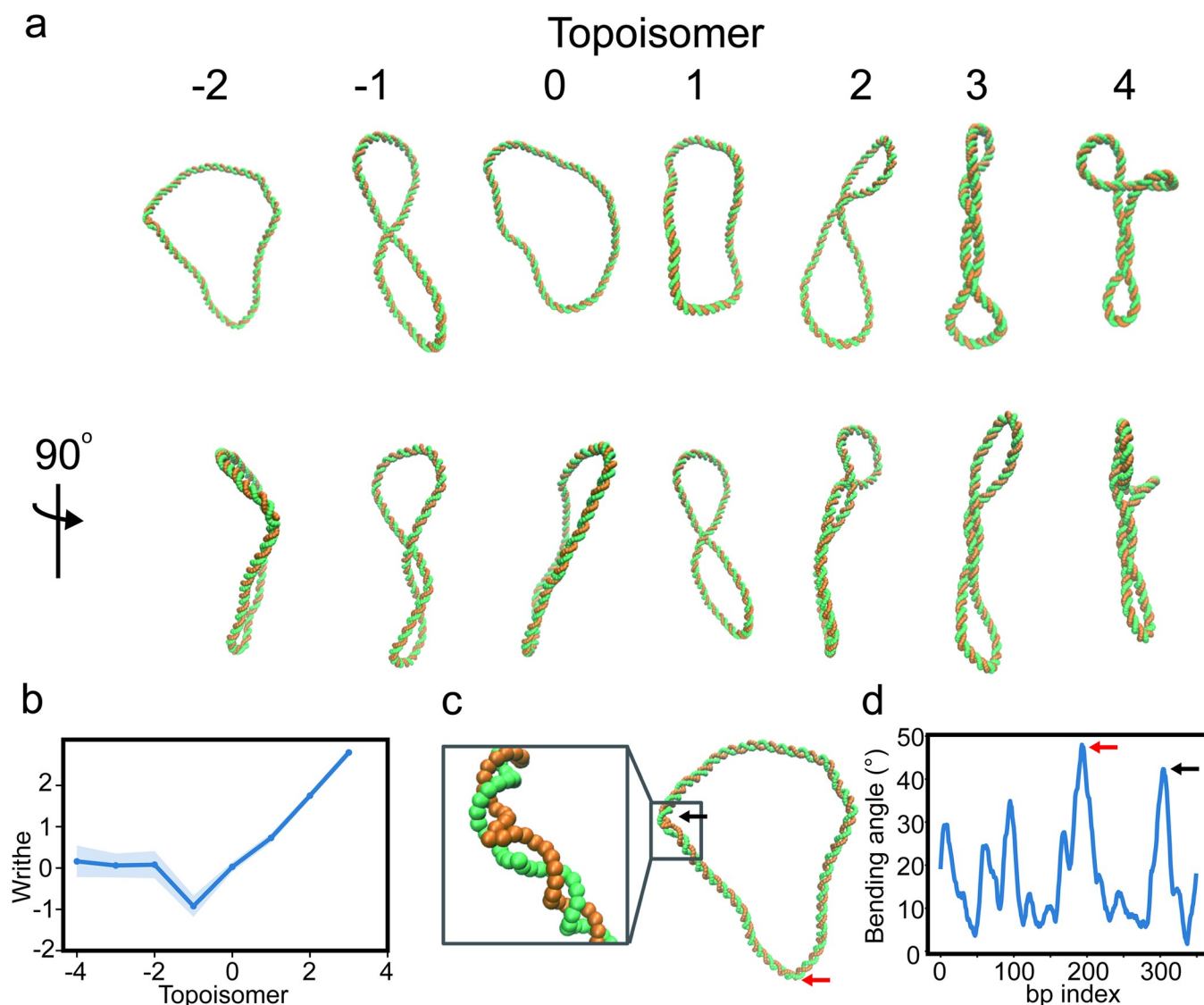


Fig 3. Supercoiling behavior of DNA minicircles. (a) Commonly observed shapes of DNA minicircles (b) Writhe analysis of DNA topoisomers -4 to 4. Positively supercoiled minicircles relax torsional stress forming plectonemes, which manifest in increasing writhe and are visible in the DNA conformations in (a). Moderately negatively supercoiled minicircles also form plectonemes, this time resulting in negative writhe. Strongly negatively supercoiled minicircles do not appear to form plectonemes. (c) Negatively supercoiled DNA relaxes torsional strain through local melting of the DNA double helix. The figure shows an example structure of a topoisomer -2, defects are indicated by the black and red arrows. (d) Bending angle along DNA for the minicircle structure in (c). Local defects are clearly visible at index 304 and 193, once again indicated by the black and red arrows.

<https://doi.org/10.1371/journal.pcbi.1011013.g003>

When the ends of a DNA molecule are covalently ligated to form a circle, the two strands become intertwined and will remain in such a state unless one of the strands is broken, i.e., the linking number is conserved. We performed simulations of supercoiled 350 bp DNA minicircles using the WEChroM model. We prepared 9 topoisomers with both negative and positive supercoiling, ranging from $\Delta Lk = -4$ to $\Delta Lk = 4$; the topoisomer $\Delta Lk = 0$ corresponding to the relaxed DNA minicircle.

Similarly to what is seen in experiments [71], a wide variety of structures is observed in simulations (Fig 3A). Positively or moderately negatively supercoiled minicircles, $\Delta Lk = -1$ to $\Delta Lk = 4$, tend to form plectonemes. These topoisomers transpose twist into writhe, relaxing twisting torsional strain into the bending strain due to writhe (Fig 3B). Surprisingly, strongly

negatively supercoiled DNA minicircles do not show a tendency to form plectonemes. On the contrary, negative torsional stress appears to be more easily relaxed through the formation of defects in the double-helical structure, i.e., by local DNA melting (Fig 3C). Such asymmetric behavior of DNA with respect to supercoiling was previously observed experimentally [72,73], with AFM and Cyro-EM showing that negatively supercoiled minicircles often present melted regions at room temperature while positive ones do not [71,72,74].

As shown, the energy transfer between bending and twisting modes is an essential aspect of DNA mechanics. Such energy transfer is difficult to reproduce when using simpler models that lack details about the three-dimensional nature of DNA. Additionally, the formation of defects in the double-helical structure of DNA also appears to be essential to correctly describe its structural ensembles [75]. Taken together, these two facts highlight the necessity of two crucial design elements of the WEChroM: accounting explicitly for the 3D structure of the DNA double helix and using a non-harmonic potential to recapitulate its empirical energy function.

DNA-protein interactions and nucleosomes

In eukaryotes, DNA is wrapped around histones to form nucleosomes with each nucleosome containing approximately 147 base pairs of DNA wrapped 1.7 times around a histone octamer. Besides compacting chromatin, nucleosomes create a significant barrier to protein binding, in turn contributing to the regulation of essential cellular processes like the gene expression [76]. The outer stretches of nucleosomal DNA are observed to spontaneously unwrap and rewrap to the core octamer [6], significantly impacting the three-dimensional packing of chromatin. A faithful description of nucleosome dynamics is necessary to recapitulate the mechanical behavior of chromatin.

We introduce nucleosomes in the WEChroM by representing the histone octamer with a single particle at its center of mass, then apply AMH interactions between the DNA and histone particles to reproduce the effect of protein-DNA interactions. The DNA-protein interactions $U_{Nucleosome}$ can be formulated as follows:

$$U_{Nucleosome} = U_{center} + U_{neighbor} \quad (5)$$

$$U_{center} = - \sum_{i \in \text{contact DNA}} \lambda_{center} \exp - \frac{(r_i - r_i^m)^2}{2\sigma_{center}^2}, \quad (6)$$

$$U_{neighbor} = - \sum_{ij \in \text{contact DNA}} \lambda_{neighbor} \exp - \frac{(r_{ij} - r_{ij}^m)^2}{2\sigma_{neighbor}^2}, \quad (7)$$

DNA bases in contact with the amino acids of the histone octamer are defined as “contact DNA”. The first part of the nucleosome associative memory U_{center} acts between the center of histone core and contact DNA (Fig 4A). The distances r_i^m are learned from a nucleosome crystal structure. The second potential, $U_{neighbor}$ acts between neighboring contact DNA bases. The energy λ and the variance σ in the two parts were parameterized to match the thermodynamics of the isolated nucleosomes.

We investigate the free energy landscape of DNA unwrapping from the nucleosome core, and its dependence on the parameters in our model. The reaction coordinate is chosen as the distance between the two free ends of the DNA, which mimics the nucleosome extension reported in the optical trap experiment [76]. To explore the phase space of unwrapping more rapidly, we used Umbrella Sampling (US) and the Multiple Bennett Acceptance Ratio (MBAR) [77] to calculate the free energy as a function of end-to-end distance.

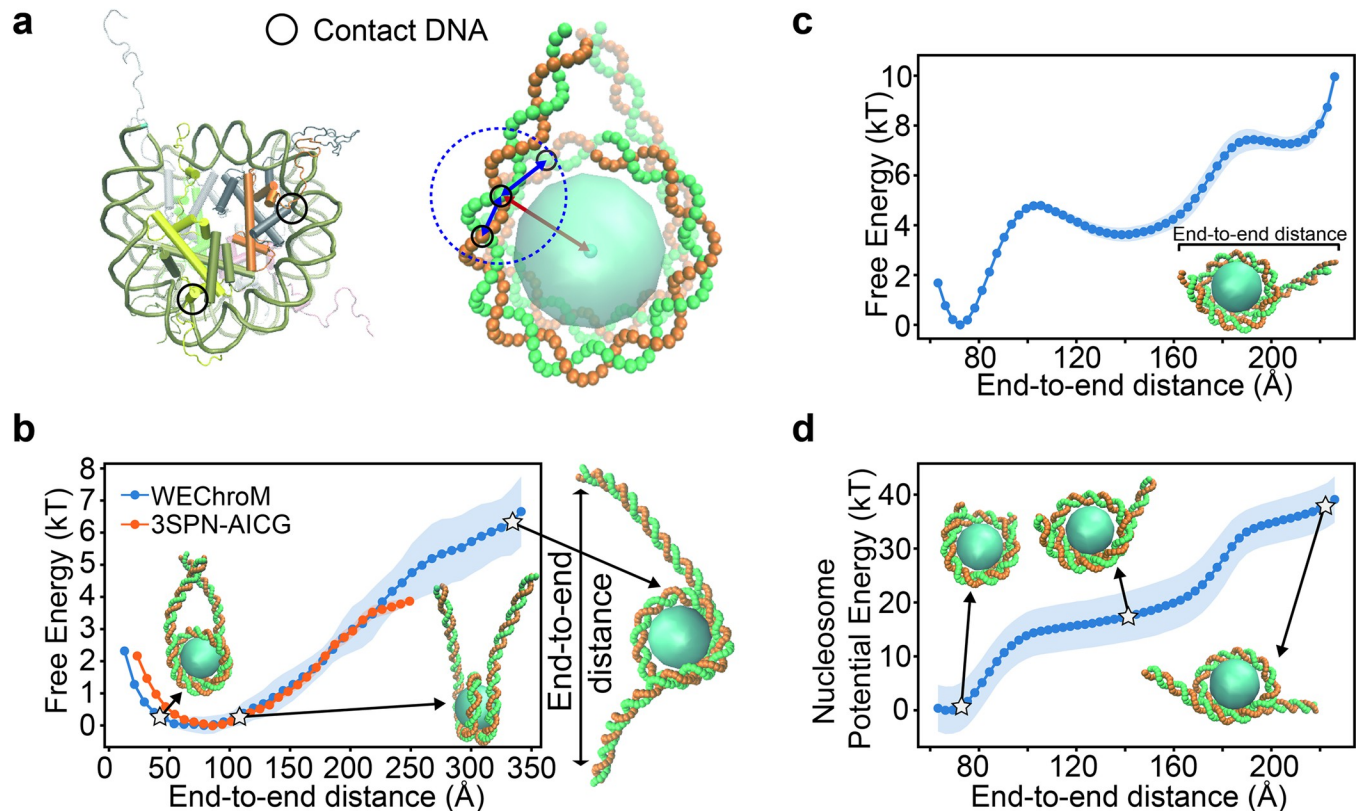


Fig 4. Nucleosome model and unwrapping analysis. (a) A detailed representation of the nucleosome core particle (PDB ID: 1KX5 [59]) (left) and the corresponding coarse-grained WEChroM representation (right). In the left figure, the upper half of the nucleosome particle is shown as solid while the lower half is transparent for clear visualization. The DNA nucleotides in close contact with amino acids are defined as "contact DNA" and subject to the AMH potential modeling histones-to-DNA interactions (black circle). In our model, contact DNA interacts with the histone octamer's center of mass (U_{center} , represented by the red arrow in the figure). Contact DNA also interacts with neighboring contact DNA ($U_{neighbors}$, represented by the black arrows inside the indigo dashed circle). (b) Free energy of partial DNA unwrapping for a system composed of the histone octamer plus 223-bp of DNA; WEChroM model (blue) and 3SPN-AICG model (orange). Simulation data shown for the 3SPN-AICG mode are extracted from ref. [38], figure 8. Examples of nucleosome configurations are shown for end-to-end distances of roughly 45 Å (left), 107 Å (right), 330 Å (outside). (c) Free energy of partial nucleosome unwrapping for a system composed of the histone octamer plus 147-bp of DNA; WEChroM model. (d) Nucleosome AMH energy of the partial nucleosome unwrapping for the 147-bp WEChroM model. Example nucleosome configurations are shown for end-to-end distances of 76 Å (left), 142 Å (middle), 221 Å (right).

<https://doi.org/10.1371/journal.pcbi.1011013.g004>

First we study a 223-bp system, with 147-bp DNA molecules bound to the histone core and 38-bp linker DNA particles on each side. The same system was previously investigated using the 3-Site-per-nucleotide (3SPN) DNA model and the Atomic Interaction-based Coarse Grained (AICG) model of the nucleosome [38,78]. We trained the parameters in $U_{Nucleosome}$ to match the free energy computed with the more detailed 3SPN-AICG model by Lequieu *et al* in ref. [38].

After training, the free energy profile computed with WEChroM is in good agreement with the 3SPN-AICG (Fig 4B right), despite the simplicity of the AMH energy function. Both models place the free energy minimum at around 80 Å and this free energy monotonically increases for the simulated range, 250 Å in the case of 3SPN-AICG and up to 350 Å in the current study. The extension energy in the full range is at the order of magnitude of a few $k_B T$, matching experimental observations that nucleosomes can spontaneously unwrap and rewrap the outer layer of DNA [6].

With this trained set of parameters for WEChroM, we study the unwrapping of DNA from the nucleosome core particle, i.e., the histones octamer wrapped by 147-bp DNA. As before,

the free energy profile is computed using Umbrella Sampling and MBAR. The results are displayed in Fig 4C. Without the flexibility warranted by linker DNA, the free energy displays a two-staged non-monotonical increase as the end-to-end distance stretches to 220 Å, when the outer layer of DNA is completely dissociated from the histone core. For this 147-bp nucleosome, the free energy associated with unwrapping the outer layer is approximately $10 k_B T$, in consistent agreement with the rather scattered values obtained by experiments, which range from about 8 to 20 $k_B T$ [76,79,80].

The AMH potential energy of the two-stage unwrapping process is shown in Fig 4D. Consistently with the free energy, the nucleosome potential energy shows two sharp increases for the end-to-end distance between 80 and 100 Å as well as between 160 and 180 Å. An analysis of the structures associated with these end-to-end distances indicates that for each sharp increase, one end of the DNA is unbound from the histone core; snapshots of these conformations are displayed in Fig 4D.

Altogether, our model suggests that the unwrapping of DNA from the nucleosome core particle is not a gradual—one base-pair at the time—process, instead, the molecular complex accumulates mechanical strain until one arm consisting of tens of base pairs suddenly breaks apart. The presence of molecular cracking in the DNA unwrapping process could be rationalized by considering the high stiffness of double-stranded DNA. On the other hand, because experiments typically require linker DNA to apply the external forces unwrapping the DNA, observing molecular cracking in the wet lab can be challenging, as we have shown that the presence of linker DNA masks this behavior. Overall, our results demonstrate that, beyond double-stranded DNA, the WEChroM model can effectively model the mechanics of single nucleosomes, opening the way to the study of the 10nm fiber and moving to complexes containing full genes.

Discussion

The Widely Editable Chromatin Model (WEChroM) proposes an alternative approach to the modeling of DNA and nucleosomes that can scale up to larger genetic systems such as the 10-nm fiber and genes. The objective of this work is to reproduce the mechanical properties of DNA and nucleosomes with a simple and computationally efficient mathematical model, with the compromise of relinquishing any attempt at reproducing the details of the chemical complexity of the system.

As shown, the model reproduces the bending and twisting persistence lengths, and the temperature dependence of the former. WEChroM is also able to reproduce the effects of positive and negative supercoiling, the formation of plectonemes and of structural defects that relax mechanical stress, and the asymmetric behavior with respect to positive or negative supercoiling, which was previously observed in experiments. Finally, we show that the proposed model is also suitable for reproducing the free energy of partial DNA unwrapping from nucleosomes. Indeed, these results demonstrate that the WEChroM model is a viable strategy to simulate the mechanics of 10-nm fiber, when willing to surrender chemical accuracy.

The structural ensembles of genes have been so far out of the reach of theoretical and computational investigations because genes are molecular complexes too large to be tackled with even the most efficient CG models and yet too strongly affected by heterogeneous molecular factors to be effectively modeled as simple polymer models. The WEChroM is an attempt to extend the reach of computational modeling with the alternative approach based on AMH. Preliminary benchmarking (see Appendix C in S1 Document) does indicate that the model is efficient enough to allow the study of genetic systems as big as tens of kilobases of chromatin, i.e., the size of mammalian genes.

Because of the extremely large number of combinations of nucleosome types, epigenetic modifications, linker lengths, and bound factors, modeling the chromatin fiber as a polymer composed of nucleosome monomers appears impractical. In fact, we are not likely to see the same exact monomer twice in a system composed of a few kilobases of DNA. WEChroM is also an attempt to solve this conundrum. The namesake characteristic of the Widely Editable Chromatin Model provides the needed flexibility for introducing a vast repertoire of structural memories of DNA binding proteins and histone variants and integrating the effect of these molecules at a larger scale. This characteristic of the AMH approach, which has not been yet exploited, opens the way to the study of structural ensembles of specific segments of chromatin, utilizing the vast amounts of omics data indicating the positioning and the identity of bound factors along DNA are already available for a variety of species, and human in particular.

Materials and methods

Implementation of WEChroM on OpenMM

OpenMM is a toolkit developed for high-performance and extensible molecular dynamics simulations [81]. OpenMM provides a high-level application programming interface (API) for external programs like Python and makes the most efficient use of CPU and GPU hardware capabilities.

WEChroM is implemented in OpenMM with a variety of custom force templates optimized for each term in the Hamiltonian. Compatible hardware platforms include single CPUs, multiple CPUs, and GPUs (with CUDA, OpenCL, and HIP support).

Simulation setup

Atomic-resolution structures for B-DNA (PDB files) were generated using w3DNA 2.0 [58] with its B-DNA fiber model at lengths 52bp, 75bp, 200bp, 350bp, and 1000bp as discussed in the Results section and Fig B in S1 Document. The atomic-resolution structure for the nucleosome (1KX5 [59]) was downloaded from the Protein Data Bank. The atomic-resolution structures were coarse-grained to be used at the WEChroM. Users have the flexibility to provide atomic PDB files other than the above-mentioned ones to coarse grain into a WEChroM molecule. DNA particles and histone protein core particles (in the case of a nucleosome system) in the generated coarse-grained file are automatically recognized by the software. This coarse-grained file are then used as the initial structure of the simulation for straight DNA systems. On the other hand, for minicircle systems, a straight DNA will be twisted to modify the initial twisting number, and then bent into a circle to create the closed open-ring form, which will serve as the initial structure for the simulations.

For naked DNA simulations, WEChroM is ready to run simulations using a potential energy described in the Section “Associative Memory Hamiltonian of DNA” and Appendix B in S1 Document “Connectivity Interactions and Steric Interactions” after the initial preparation. Each term can be turned on and off and vary in strength and setting. A local energy minimization is performed before the formal start of the simulation to prevent unphysical forces produced by the initial positions. The sampling simulation is carried out via Langevin dynamics with a damping coefficient of $\gamma = 1/\tau$, where τ is the time unit. A time step of 0.02τ was used for all the simulations. Only thermodynamical simulations are performed where the dynamics of systems converged in the equilibrium states and the relaxation time was chosen for the best sampling. The temperature is fixed at the unit temperature and assumed to be room temperature. A detailed description of the reduced unit system is provided in the Appendix A in S1 Document. The trajectories are saved in .dcd files. The equilibrium state was

attained after 1×10^6 steps for the 75-bp straight DNA systems and after 1×10^8 steps for 350-bp minicircle systems. The substantial increase in steps required for the minicircles to reach equilibrium is to be expected because our initial structures have a high twisting number, and transitioning to the writhing number necessitates a global reconfiguration. Conversely, the straight DNA system starts near a potential energy minimum configuration and only needs to undergo thermal fluctuations to satisfy persistence length requirements. Each of the trajectories was simulated for 1×10^9 steps, storing a frame every 5×10^3 steps. Analyses are performed after the simulations by reading the trajectory files.

For nucleosome simulations, we employ umbrella sampling to sufficiently explore the configurational space. The DNA end-to-end distance serves as the reaction coordinate. The first or last base pair's center of mass is used to designate the ends of the DNA, and the Euclidean distance between them is used to determine the end-to-end distance. Forty-one umbrella windows are used for the 223 bp system (nucleosome core particle with two arms), and 23 are used for the 147bp system (nucleosome core particle system). In each window, the end-to-end distance is restrained around a center distance using a harmonic biasing potential. We utilized Langevin Dynamics with the same hyperparameters as in the naked DNA system, where the damping coefficient of $1/\tau$ and a time step of 0.02τ , where τ is the unit time in our reduced unit system. Five replicas were used within each umbrella window, and each replica was simulated for 1×10^7 steps storing a frame every 1×10^3 steps. We then utilize the FastMBAR [77] to calculate the free energies and expectations based on the saved trajectories.

Supporting information

S1 Document. WEChroM simulation and software description. Detailed description of the reduced unit system, force field setup, and mechanical property analysis for WEChroM systems.ik (DOCX)

Acknowledgments

We would like to thank Dr. Peter Wolynes, Dr. Wei Lu, Dr. Ailun Wang, Dr. Vinicius Contesoto, and Dr. Antonio Oliveira for their valuable discussions and help during the work. We also thank the Center for Theoretical Biological Physics and AMD for the computational resources.

Author Contributions

Conceptualization: Weiqi Lu, José N. Onuchic, Michele Di Pierro.

Data curation: Weiqi Lu.

Formal analysis: Weiqi Lu.

Funding acquisition: José N. Onuchic, Michele Di Pierro.

Investigation: Weiqi Lu, José N. Onuchic, Michele Di Pierro.

Methodology: Weiqi Lu, José N. Onuchic, Michele Di Pierro.

Project administration: Weiqi Lu, José N. Onuchic, Michele Di Pierro.

Software: Weiqi Lu.

Supervision: José N. Onuchic, Michele Di Pierro.

Validation: Weiqi Lu.

Visualization: Weiqi Lu.

Writing – original draft: Weiqi Lu.

Writing – review & editing: Weiqi Lu, José N. Onuchic, Michele Di Pierro.

References

1. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*. 1953 Apr; 171(4356):737–8. <https://doi.org/10.1038/171737a0> PMID: 13054692
2. Franklin RE, Gosling RG. Molecular Configuration in Sodium Thymonucleate. *Nature*. 1953 Apr; 171(4356):740–1. <https://doi.org/10.1038/171740a0> PMID: 13054694
3. Wilkins MHF, Stokes AR, Wilson HR. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature*. 1953 Apr; 171(4356):738–40.
4. Sun T, Minhas V, Korolev N, Mirzoev A, Lyubartsev AP, Nordenskiöld L. Bottom-Up Coarse-Grained Modeling of DNA. *Front Mol Biosci*. 2021; 8. <https://doi.org/10.3389/fmolb.2021.645527> PMID: 33816559
5. Zhou K, Gaullier G, Luger K. Nucleosome structure and dynamics are coming of age. *Nat Struct Mol Biol*. 2019 Jan; 26(1):3–13. <https://doi.org/10.1038/s41594-018-0166-x> PMID: 30532059
6. Tims HS, Gurunathan K, Levitus M, Widom J. Dynamics of Nucleosome Invasion by DNA Binding Proteins. *J Mol Biol*. 2011 Aug 12; 411(2):430–48. <https://doi.org/10.1016/j.jmb.2011.05.044> PMID: 21669206
7. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T III, et al. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res*. 2010 Jan 1; 38(1):299–313. <https://doi.org/10.1093/nar/gkp834> PMID: 19850719
8. Minhas V, Sun T, Mirzoev A, Korolev N, Lyubartsev AP, Nordenskiöld L. Modeling DNA Flexibility: Comparison of Force Fields from Atomistic to Multiscale Levels. *J Phys Chem B*. 2020 Jan 9; 124(1):38–49. <https://doi.org/10.1021/acs.jpcc.9b09106> PMID: 31805230
9. Sun T, Mirzoev A, Minhas V, Korolev N, Lyubartsev AP, Nordenskiöld L. A multiscale analysis of DNA phase separation: from atomistic to mesoscale level. *Nucleic Acids Res*. 2019 Jun 20; 47(11):5550–62. <https://doi.org/10.1093/nar/gkz377> PMID: 31106383
10. Savelyev A, Papoian GA. Molecular Renormalization Group Coarse-Graining of Polymer Chains: Application to Double-Stranded DNA. *Biophys J*. 2009 May 20; 96(10):4044–52. <https://doi.org/10.1016/j.bpj.2009.02.067> PMID: 19450476
11. Savelyev A, Papoian GA. Chemically accurate coarse graining of double-stranded DNA. *Proc Natl Acad Sci*. 2010 Nov 23; 107(47):20340–5. <https://doi.org/10.1073/pnas.1001163107> PMID: 21059937
12. Kovaleva NA, Koroleva (Kikot) IP, Mazo MA, Zubova EA. The “sugar” coarse-grained DNA model. *J Mol Model*. 2017 Feb 9; 23(2):66. <https://doi.org/10.1007/s00894-017-3209-z> PMID: 28185115
13. He Y, Maciejczyk M, Oldziej S, Scheraga HA, Liwo A. Mean-Field Interactions between Nucleic-Acid-Base Dipoles can Drive the Formation of a Double Helix. *Phys Rev Lett*. 2013 Feb 28; 110(9):098101. <https://doi.org/10.1103/PhysRevLett.110.098101> PMID: 23496746
14. Naômé A, Laaksonen A, Vercauteren DP. A Coarse-Grained Simulation Study of the Structures, Energetics, and Dynamics of Linear and Circular DNA with Its Ions. *J Chem Theory Comput*. 2015 Jun 9; 11(6):2813–26. <https://doi.org/10.1021/acs.jctc.5b00113> PMID: 26575574
15. Maffeo C, Ngo TTM, Ha T, Aksimentiev A. A Coarse-Grained Model of Unstructured Single-Stranded DNA Derived from Atomistic Simulation and Single-Molecule Experiment. *J Chem Theory Comput*. 2014 Aug 12; 10(8):2891–6. <https://doi.org/10.1021/ct500193u> PMID: 25136266
16. Ouldridge TE, Louis AA, Doye JPK. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J Chem Phys*. 2011 Feb 28; 134(8):085101. <https://doi.org/10.1063/1.3552946> PMID: 21361556
17. Snodin BEK, Randisi F, Mosayebi M, Šulc P, Schreck JS, Romano F, et al. Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J Chem Phys*. 2015 Jun 21; 142(23):234901. <https://doi.org/10.1063/1.4921957> PMID: 26093573
18. Drukker K, Schatz GC. A Model for Simulating Dynamics of DNA Denaturation. *J Phys Chem B*. 2000 Jul 1; 104(26):6108–11.
19. Drukker K, Wu G, Schatz GC. Model simulations of DNA denaturation dynamics. *J Chem Phys*. 2001 Jan; 114(1):579–90.

20. Knotts TA, Rathore N, Schwartz DC, de Pablo JJ. A coarse grain model for DNA. *J Chem Phys*. 2007 Feb 28; 126(8):084901. <https://doi.org/10.1063/1.2431804> PMID: 17343470
21. Sambriski EJ, Schwartz DC, de Pablo JJ. A Mesoscale Model of DNA and Its Renaturation. *Biophys J*. 2009 Mar 4; 96(5):1675–90. <https://doi.org/10.1016/j.bpj.2008.09.061> PMID: 19254530
22. Hinckley DM, Freeman GS, Whitmer JK, de Pablo JJ. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J Chem Phys*. 2013 Oct 14; 139(14):144903. <https://doi.org/10.1063/1.4822042> PMID: 24116642
23. Freeman GS, Hinckley DM, Lequeieu JP, Whitmer JK, de Pablo JJ. Coarse-grained modeling of DNA curvature. *J Chem Phys*. 2014 Oct 28; 141(16):165103. <https://doi.org/10.1063/1.4897649> PMID: 25362344
24. Markegard CB, Fu IW, Reddy KA, Nguyen HD. Coarse-Grained Simulation Study of Sequence Effects on DNA Hybridization in a Concentrated Environment. *J Phys Chem B*. 2015 Feb 5; 119(5):1823–34. <https://doi.org/10.1021/jp509857k> PMID: 25581253
25. Darré L, Machado MR, Brandner AF, González HC, Ferreira S, Pantano S. SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics. *J Chem Theory Comput*. 2015 Feb 10; 11(2):723–39. <https://doi.org/10.1021/ct5007746> PMID: 26575407
26. Dans PD, Zeida A, Machado MR, Pantano S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *J Chem Theory Comput*. 2010 May 11; 6(5):1711–25. <https://doi.org/10.1021/ct900653p> PMID: 26615701
27. Cragolini T, Derreumaux P, Pasquali S. Coarse-Grained Simulations of RNA and DNA Duplexes. *J Phys Chem B*. 2013 Jul 11; 117(27):8047–60. <https://doi.org/10.1021/jp400786b> PMID: 23730911
28. Dans PD, Walther J, Gómez H, Orozco M. Multiscale simulation of DNA. *Curr Opin Struct Biol*. 2016 Apr 1; 37:29–45. <https://doi.org/10.1016/j.sbi.2015.11.011> PMID: 26708341
29. Machado MR, Barrera EE, Klein F, Sónora M, Silva S, Pantano S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J Chem Theory Comput*. 2019 Apr 9; 15(4):2719–33. <https://doi.org/10.1021/acs.jctc.9b00006> PMID: 30810317
30. Snodin BEK, Schreck JS, Romano F, Louis AA, Doye JPK. Coarse-grained modelling of the structural properties of DNA origami. *Nucleic Acids Res*. 2019 Feb 20; 47(3):1585–97. <https://doi.org/10.1093/nar/gky1304> PMID: 30605514
31. Perišić O, Collepardo-Guevara R, Schlick T. Modeling Studies of Chromatin Fiber Structure as a Function of DNA Linker Length. *J Mol Biol*. 2010 Nov 12; 403(5):777–802. <https://doi.org/10.1016/j.jmb.2010.07.057> PMID: 20709077
32. Ozer G, Luque A, Schlick T. The chromatin fiber: multiscale problems and approaches. *Curr Opin Struct Biol*. 2015 Apr 1; 31:124–39. <https://doi.org/10.1016/j.sbi.2015.04.002> PMID: 26057099
33. Arya G, Schlick T. A Tale of Tails: How Histone Tails Mediate Chromatin Compaction in Different Salt and Linker Histone Environments. *J Phys Chem A*. 2009 Apr 23; 113(16):4045–59. <https://doi.org/10.1021/jp810375d> PMID: 19298048
34. Perišić O, Schlick T. Dependence of the Linker Histone and Chromatin Condensation on the Nucleosome Environment. *J Phys Chem B*. 2017 Aug 24; 121(33):7823–32. <https://doi.org/10.1021/acs.jpcc.7b04917> PMID: 28732449
35. Bascom GD, Kim T, Schlick T. Kilobase Pair Chromatin Fiber Contacts Promoted by Living-System-Like DNA Linker Length Distributions and Nucleosome Depletion. *J Phys Chem B*. 2017 Apr 20; 121(15):3882–94. <https://doi.org/10.1021/acs.jpcc.7b00998> PMID: 28299939
36. Bascom GD, Sanbonmatsu KY, Schlick T. Mesoscale Modeling Reveals Hierarchical Looping of Chromatin Fibers Near Gene Regulatory Elements. *J Phys Chem B*. 2016 Aug 25; 120(33):8642–53. <https://doi.org/10.1021/acs.jpcc.6b03197> PMID: 27218881
37. Bascom GD, Myers CG, Schlick T. Mesoscale modeling reveals formation of an epigenetically driven HOXC gene hub. *Proc Natl Acad Sci*. 2019 Mar 12; 116(11):4955–62. <https://doi.org/10.1073/pnas.1816424116> PMID: 30718394
38. Lequeieu J, Córdoba A, Moller J, de Pablo JJ. 1CPN: A coarse-grained multi-scale model of chromatin. *J Chem Phys*. 2019 Jun 7; 150(21):215102. <https://doi.org/10.1063/1.5092976> PMID: 31176328
39. Kepper N, Foethke D, Stehr R, Wedemann G, Rippe K. Nucleosome Geometry and Internucleosomal Interactions Control the Chromatin Fiber Conformation. *Biophys J*. 2008 Oct 15; 95(8):3692–705. <https://doi.org/10.1529/biophysj.107.121079> PMID: 18212006
40. Stehr R, Kepper N, Rippe K, Wedemann G. The Effect of Internucleosomal Interaction on Folding of the Chromatin Fiber. *Biophys J*. 2008 Oct 15; 95(8):3677–91. <https://doi.org/10.1529/biophysj.107.120543> PMID: 18658212

41. Stehr R, Schöpflin R, Ettig R, Kepper N, Rippe K, Wedemann G. Exploring the Conformational Space of Chromatin Fibers and Their Stability by Numerical Dynamic Phase Diagrams. *Biophys J*. 2010 Mar 17; 98(6):1028–37. <https://doi.org/10.1016/j.bpj.2009.11.040> PMID: 20303860
42. Müller O, Kepper N, Schöpflin R, Ettig R, Rippe K, Wedemann G. Changing Chromatin Fiber Conformation by Nucleosome Repositioning. *Biophys J*. 2014 Nov 4; 107(9):2141–50. <https://doi.org/10.1016/j.bpj.2014.09.026> PMID: 25418099
43. Attou A, Zülske T, Wedemann G. Cohesin and CTCF complexes mediate contacts in chromatin loops depending on nucleosome positions. *Biophys J*. 2022 Dec 20; 121(24):4788–99. <https://doi.org/10.1016/j.bpj.2022.10.044> PMID: 36325618
44. Bajpai G, Jain I, Inamdar MM, Das D, Padinhateeri R. Binding of DNA-bending non-histone proteins destabilizes regular 30-nm chromatin structure. *PLOS Comput Biol*. 2017 Jan 30; 13(1):e1005365. <https://doi.org/10.1371/journal.pcbi.1005365> PMID: 28135276
45. Farr SE, Woods EJ, Joseph JA, Garaizar A, Collepardo-Guevara R. Nucleosome plasticity is a critical element of chromatin liquid–liquid phase separation and multivalent nucleosome interactions. *Nat Commun*. 2021 May 17; 12(1):2883. <https://doi.org/10.1038/s41467-021-23090-3> PMID: 34001913
46. Brouwer T, Pham C, Kaczmarczyk A, de Voogd WJ, Botto M, Vizjak P, et al. A critical role for linker DNA in higher-order folding of chromatin fibers. *Nucleic Acids Res*. 2021 Mar 18; 49(5):2537–51. <https://doi.org/10.1093/nar/gkab058> PMID: 33589918
47. de Jong BE, Brouwer TB, Kaczmarczyk A, Visscher B, van Noort J. Rigid Basepair Monte Carlo Simulations of One-Start and Two-Start Chromatin Fiber Unfolding by Force. *Biophys J*. 2018 Nov 20; 115(10):1848–59. <https://doi.org/10.1016/j.bpj.2018.10.007> PMID: 30366627
48. Norouzi D, Zhurkin VB. Dynamics of Chromatin Fibers: Comparison of Monte Carlo Simulations with Force Spectroscopy. *Biophys J*. 2018 Nov 6; 115(9):1644–55. <https://doi.org/10.1016/j.bpj.2018.06.032> PMID: 30236784
49. Zhurkin VB, Norouzi D. Topological polymorphism of nucleosome fibers and folding of chromatin. *Biophys J*. 2021 Feb 16; 120(4):577–85. <https://doi.org/10.1016/j.bpj.2021.01.008> PMID: 33460599
50. Buitrago D, Labrador M, Arcon JP, Lema R, Flores O, Esteve-Codina A, et al. Impact of DNA methylation on 3D genome structure. *Nat Commun*. 2021 May 28; 12(1):3243. <https://doi.org/10.1038/s41467-021-23142-8> PMID: 34050148
51. Neguembor MV, Arcon JP, Buitrago D, Lema R, Walther J, Garate X, et al. MiOS, an integrated imaging and computational strategy to model gene folding with nucleosome resolution. *Nat Struct Mol Biol*. 2022 Oct; 29(10):1011–23. <https://doi.org/10.1038/s41594-022-00839-y> PMID: 36220894
52. Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J Phys Chem B*. 2012 Jul 26; 116(29):8494–503. <https://doi.org/10.1021/jp212541y> PMID: 22545654
53. Papoian GA, Ulander J, Eastwood MP, Luthey-Schulten Z, Wolynes PG. Water in protein structure prediction. *Proc Natl Acad Sci*. 2004 Mar 9; 101(10):3352–7. <https://doi.org/10.1073/pnas.0307851100> PMID: 14988499
54. Zheng W, Schafer NP, Davtyan A, Papoian GA, Wolynes PG. Predictive energy landscapes for protein–protein association. *Proc Natl Acad Sci*. 2012 Nov 20; 109(47):19244–9. <https://doi.org/10.1073/pnas.1216215109> PMID: 23129648
55. Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, et al. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci*. 1981 Apr; 78(4):2179–83. <https://doi.org/10.1073/pnas.78.4.2179> PMID: 6941276
56. Friedrichs MS, Wolynes PG. Toward Protein Tertiary Structure Recognition by Means of Associative Memory Hamiltonians. *Science*. 1989 Oct 20; 246(4928):371–3. <https://doi.org/10.1126/science.246.4928.371> PMID: 17747919
57. Friedrichs MS, Wolynes PG. Molecular dynamics of associative memory hamiltonians for protein tertiary structure recognition. *Tetrahedron Comput Methodol*. 1990 Jan 1; 3(3):175–90.
58. Li S, Olson WK, Lu XJ. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res*. 2019 Jul 2; 47(W1):W26–34. <https://doi.org/10.1093/nar/gkz394> PMID: 31114927
59. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution††We dedicate this paper to the memory of Max Perutz who was particularly inspirational and supportive to T.J.R. in the early stages of this study. *J Mol Biol*. 2002 Jun 21; 319(5):1097–113.

60. Di Pierro M, Zhang B, Aiden EL, Wolynes PG, Onuchic JN. Transferable model for chromosome architecture. *Proc Natl Acad Sci*. 2016 Oct 25; 113(43):12168–73. <https://doi.org/10.1073/pnas.1613607113> PMID: 27688758
61. Gray HB, Hearst JE. Flexibility of native DNA from the sedimentation behavior as a function of molecular weight and temperature. *J Mol Biol*. 1968 Jul 14; 35(1):111–29. [https://doi.org/10.1016/s0022-2836\(68\)80041-5](https://doi.org/10.1016/s0022-2836(68)80041-5) PMID: 5760559
62. Geggier S, Kotlyar A, Vologodskii A. Temperature dependence of DNA persistence length. *Nucleic Acids Res*. 2011 Mar 1; 39(4):1419–26. <https://doi.org/10.1093/nar/gkq932> PMID: 20952402
63. Baumann CG, Smith SB, Bloomfield VA, Bustamante C. Ionic effects on the elasticity of single DNA molecules. *Proc Natl Acad Sci*. 1997 Jun 10; 94(12):6185–90. <https://doi.org/10.1073/pnas.94.12.6185> PMID: 9177192
64. Moroz JD, Nelson P. Torsional directed walks, entropic elasticity, and DNA twist stiffness. *Proc Natl Acad Sci*. 1997 Dec 23; 94(26):14418–22. <https://doi.org/10.1073/pnas.94.26.14418> PMID: 9405627
65. Noy A, Golestanian R. Length Scale Dependence of DNA Mechanical Properties. *Phys Rev Lett*. 2012 Nov 30; 109(22):228101. <https://doi.org/10.1103/PhysRevLett.109.228101> PMID: 23368161
66. Gao X, Hong Y, Ye F, Inman JT, Wang MD. Torsional Stiffness of Extended and Plectonemic DNA. *Phys Rev Lett*. 2021 Jul 7; 127(2):028101. <https://doi.org/10.1103/PhysRevLett.127.028101> PMID: 34296898
67. Roth E, Glick Azaria A, Girshevitz O, Bitler A, Garini Y. Measuring the Conformation and Persistence Length of Single-Stranded DNA Using a DNA Origami Structure. *Nano Lett*. 2018 Nov 14; 18(11):6703–9. <https://doi.org/10.1021/acs.nanolett.8b02093> PMID: 30352164
68. Fogg JM, Randall GL, Pettitt BM, Sumners DWL, Harris SA, Zechiedrich L. Bullied no more: when and how DNA shoves proteins around. *Q Rev Biophys*. 2012 Aug; 45(3):257–99. <https://doi.org/10.1017/S0033583512000054> PMID: 22850561
69. Seidel R, Dekker C. Single-molecule studies of nucleic acid motors. *Curr Opin Struct Biol*. 2007 Feb 1; 17(1):80–6. <https://doi.org/10.1016/j.sbi.2006.12.003> PMID: 17207989
70. Harada Y, Ohara O, Takatsuki A, Itoh H, Shimamoto N, Kinoshita K. Direct observation of DNA rotation during transcription by Escherichia coli RNA polymerase. *Nature*. 2001 Jan; 409(6816):113–5. <https://doi.org/10.1038/35051126> PMID: 11343125
71. Irobalieva RN, Fogg JM, Catanese DJ, Sutthibutpong T, Chen M, Barker AK, et al. Structural diversity of supercoiled DNA. *Nat Commun*. 2015 Oct 12; 6(1):8440.
72. Bettotti P, Visone V, Lunelli L, Perugini G, Ciaramella M, Valenti A. Structure and Properties of DNA Molecules Over The Full Range of Biologically Relevant Supercoiling States. *Sci Rep*. 2018 Apr 18; 8(1):6163. <https://doi.org/10.1038/s41598-018-24499-5> PMID: 29670174
73. Meng H, Bosman J, van der Heijden T, van Noort J. Coexistence of Twisted, Plectonemic, and Melted DNA in Small Topological Domains. *Biophys J*. 2014 Mar 4; 106(5):1174–81. <https://doi.org/10.1016/j.bpj.2014.01.017> PMID: 24606941
74. Li D, Lv B, wang Q, Liu Y, Zhuge Q. Direct observation of positive supercoils introduced by reverse gyrase through atomic force microscopy. *Bioorg Med Chem Lett*. 2017 Sep 1; 27(17):4086–90. <https://doi.org/10.1016/j.bmcl.2017.07.044> PMID: 28756025
75. Pyne ALB, Noy A, Main KHS, Velasco-Berrelleza V, Piperakis MM, Mitchenall LA, et al. Base-pair resolution analysis of the effect of supercoiling on DNA flexibility and major groove recognition by triplex-forming oligonucleotides. *Nat Commun*. 2021 Feb 16; 12(1):1053. <https://doi.org/10.1038/s41467-021-21243-y> PMID: 33594049
76. Mihardja S, Spakowitz AJ, Zhang Y, Bustamante C. Effect of force on mononucleosomal dynamics. *Proc Natl Acad Sci*. 2006 Oct 24; 103(43):15871–6. <https://doi.org/10.1073/pnas.0607526103> PMID: 17043216
77. Ding X, Vilseck JZ, Brooks CL. Fast Solver for Large Scale Multistate Bennett Acceptance Ratio Equations. *J Chem Theory Comput*. 2019 Feb 12; 15(2):799–802. <https://doi.org/10.1021/acs.jctc.8b01010> PMID: 30689377
78. Lequieu J, Córdoba A, Schwartz DC, de Pablo JJ. Tension-Dependent Free Energies of Nucleosome Unwrapping. *ACS Cent Sci*. 2016 Sep 28; 2(9):660–6. <https://doi.org/10.1021/acscentsci.6b00201> PMID: 27725965
79. Brower-Toland BD, Smith CL, Yeh RC, Lis JT, Peterson CL, Wang MD. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc Natl Acad Sci*. 2002 Feb 19; 99(4):1960–5. <https://doi.org/10.1073/pnas.022638399> PMID: 11854495
80. Meng H, Andresen K, van Noort J. Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibers. *Nucleic Acids Res*. 2015 Apr 20; 43(7):3578–90. <https://doi.org/10.1093/nar/gkv215> PMID: 25779043

81. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLOS Comput Biol. 2017 Jul 26; 13(7):e1005659. <https://doi.org/10.1371/journal.pcbi.1005659> PMID: 28746339